THESE

présentée pour obtenir le grade de

Docteur de TELECOM ParisTech

Spécialité: Communication et Electronique

Randa Zakhour

Feedback Limité, Coopération et Coordination dans les Systèmes Cellulaires Multi-antennaires

Soutenue le 22 Avril 2010 devant le jury composé de :

Rapporteurs	Prof. H. Bölcskei, ETH-Zürich
	Prof. P. Loubaton, Université de Marne la Vallée
Examinateurs	Prof. P. Duhamel, Supélec
	Prof. D. Slock, EURECOM
	Prof. W. Utschick, TU München
Directeur de thèse	Prof. D. Gesbert, EURECOM

DISSERTATION

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

Specialization: Communication and Electronics

Randa Zakhour

Aspects of Limited Feedback, Cooperation and Coordination in Multi-antenna Cellular Systems

Defended on the 22nd of April 2010 before the committee composed of:

Reviewers	Prof. H. Bölcskei, ETH-Zürich
	Prof. P. Loubaton, Université de Marne la Vallée
Examiners	Prof. P. Duhamel, Supélec
	Prof. D. Slock, EURECOM
	Prof. W. Utschick, TU München
Thesis supervisor	Prof. D. Gesbert, EURECOM

Abstract

Multi-antenna schemes have been shown to provide remarkable gains in terms of spectral efficiency and extensive research has been dedicated to studying those in detail, discovering new setups in which multi-input multi-output (MIMO) schemes can be used and making them practical. This thesis is concerned with certain aspects of two particular setups in which MIMO techniques may be implemented, which we summarize below.

Limited feedback in the MIMO broadcast channel: We consider the downlink of a single cell where the base station (BS) has multiple antennas. A particular issue when dealing with such a system is that of channel state information at the transmitter (CSIT): this has been shown to play a cardinal role, particularly in the case when users are equipped with single antennas. Given the importance of acquiring that CSIT, and the fact that this acquisition comes at the cost of using resources on the uplink direction and introducing delays, a large body of literature has been concerned with limited feedback schemes for this setup. In this thesis, we contribute two ideas to try to make the most of the available feedback resource.

Coordination and Cooperation in multicell systems: The above model of a single cell corresponds roughly to current cellular designs where frequency planning is used to separate cells that use the same frequency resources so that inter-cell interference is limited. This is however not very efficient and MIMO techniques may be used to allow for increased performance at full reuse. Thus incorporating more antennas at a given BS gives it the possibility to mitigate the interference it causes at users in other cells. This is one of the topics we deal with in this thesis and propose a scheme which, while requiring local CSI only, is shown to perform quite well. MIMO techniques in a multicell environment can moreover extend to multicell processing (MCP), whereby several BSs pool their antennas to essentially act as a giant MIMO transmitter. This, however, has significant costs in terms of backhaul for data and CSI sharing, which threatens the scalability of MCP. We start investigating how to deal with limitations related to both these aspects.

Acknowledgements

My first thanks goes out to my thesis advisor, David Gesbert, who based on a phone interview some four years ago decided that I was PhD material. Without his many insights and guidance, this thesis would not have been possible.

I am grateful to my jury members, for managing to be present whether physically or via video conference, under circumstances one can at best describe as unusual. Not everybody gets to have a volcano erupt and paralyze air travel the week of their defense. I am particularly grateful to Professor Philippe Loubaton for helping improve the quality of some of the proofs, and to Prof. Stephen Hanly with whom I collaborated on the work presented in Chapter 7.

I would also like to thank my friends at EURECOM and in Nice. For their advice, patience, humour, and for teaching me a thing or two about what is really important in life, I am forever grateful. Zuleita, Sara, Daniel, Ikbal, Erhan, Giuliana, Kostas, Antony, Jin Hui, Fadi, Ruben, Samia... Thanks for listening.

I cannot begin to thank my family, for their unconditional love, guidance and support, and for cheering me up whenever I was down.

Contents

List of Figures

Acronyms

Here are the main acronyms used in this document. The meaning of an acronym is usually indicated once, when it first occurs in the text. The English acronyms are also used for the French summary.

ARQ	Automatic Repeat Request
a.s.	almost surely.
AWGN	Additive White Gaussian Noise.
BC	Broadcast Channel.
BS	Base Station.
cdf	cumulative density function.
CBf	Coordinated Beamforming.
CDI	Channel Direction Information.
CQI	Channel Quality Indicator.
CSI	Channel State Information.
CSIR	Channel State Information at Receiver.
CSIT	Channel State Information at Transmitter.
DL	Downlink.
DPC	Dirty Paper Coding.
FDD	Frequency Division Duplex.
IC	Interference Channel.
i.i.d.	independent and identically distributed.
MAC	Multiple Access Channel.
MCP	Multicell Processing.
MMSE	Minimum Mean Square Error.
MRT	Maximum Ratio Transmission.
MSE	Mean Square Error.
MIMO	Multi-Input Multi-Output.
MISO	Multi-Input Single-Output.
MUD	Multiuser Diversity.

NE	Nash Equilibrium.
PA	Power Allocation.
pdf	probability density function.
PSD	Positive Semi-Definite.
RHS	Right Hand Side.
RMT	Random Matrix Theory.
RVQ	Random Vector Quantization.
Rx	Receiver.
SCP	Single Cell Processing.
SDMA	Space Division Multiple Access.
SIC	Successive Interference Cancelation.
SINR	Signal-to-Interference-Noise Ratio.
SISO	Single-Input Single-Output.
s.t.	such that.
TDD	Time Division Duplex.
THP	Tomlinson-Harashima Precoding.
Tx	Transmitter.
wlog	without loss of generality.
wrt	with respect to.
UL	Uplink.
ZF	Zero-Forcing.
ZFBF	Zero-Forcing Beamforming.

Notations

Boldface upper-case letters denote matrices, boldface lower-case letters denote column vectors (unless stated otherwise) and lower-case italics denote scalars. Calligraphic upper-case letters denote sets.

$\mathbb{R}, \mathbb{C}$	The set of all real and complex numbers, respectively.
$\mathbb{R}^+$	The set of positive real numbers.
$\mathbb{C}^{n\times m}$	The set of $n \times m$ matrices with complex-valued entries. If $m = 1$, the index may be dropped.
$X_{i,j}$	The (i, j)th element of the matrix $\mathbf{X}$, if the latter is defined.
x_i	The ith element of vector $\mathbf{x}$, if the latter is defined.
$\mathrm{tr}\mathbf{X}$	trace of the matrix $\mathbf{X}$.
$\lvert\mathbf{X}\rvert$	Determinant of the matrix $\mathbf{X}$.
$\lVert\mathbf{x}\rVert^2$	Squared Euclidean norm of vector $\mathbf{x}$.
$\lvert x\rvert$	Absolute value of x.
$\mathbf{X}^*$	The complex conjugate of matrix $\mathbf{X}$.
$\mathbf{X}^H$	The complex conjugate transpose (Hermitian) of matrix $\mathbf{X}$.
$\mathbf{X}^{\mathrm{T}}$	The transpose of matrix $\mathbf{X}$.
$\mathbf{X}^{-1}$	The inverse of matrix $\mathbf{X}$.
$\mathbf{X}^\dagger$	The pseudo-inverse of matrix $\mathbf{X}$, given by $\mathbf{X}^H \left(\mathbf{X}\mathbf{X}^H\right)^{-1}$.
$\mathrm{diag}\left(\mathbf{x}\right)$	The diagonal matrix with the entries of vector $\mathbf{x}$ along its diagonal.
$\mathbf{X}^{1/2}$	Hermitian square root of the positive semidefinite matrix $\mathbf{X}$.
$\mathbf{I}$ or $\mathbf{I}_n$	Identity matrix and identity matrix of dimension n, respectively.
$\Pr[.]$	Probability.
$\mathbb{E}[.]$	Expectation.
$\sim$	Distributed according to.
$\mathcal{CN}(\mathbf{m}, \mathbf{C})$	Circularly symmetric complex Gaussian random vector of mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$.
$\max, \min$	Maximum and minimum.
max. , min.	Maximize and minimize.

$\lvert\mathcal{A}\rvert$	Cardinality of set $\mathcal{A}$.
$E_1(.)$	The exponential integral, defined as $E_1(x) = \int_x^\infty \frac{e^{-t}}{t} dt$.
$\mathbf{i}$	$\sqrt{-1}$.

Résumé

Introduction

Les systèmes de communication sans fil actuels et ceux en voie de standardisation tentent de satisfaire des demandes de débits de plus en plus élevés, ainsi que de rendre la performance plus équitable à travers la zone de couverture du réseau. Pour atteindre ces objectifs, l'introduction d'antennes multiples en émission et/ou en réception va certainement jouer un rôle important, puisqu'elles permettent d'acrroître la capacité du canal résultant. Dans les systèmes cellulaires, où le spectre radio est limité, des techniques multi-antennaires permettent d'augmenter l'éfficacité spectrale du système et d'améliorer la qualité de service offerte aux utilisateurs. Cependant, que ce soit dans le cas d'une cellule isolée ou dans un système multicellulaire, des aspects réalistes liés à la connaissance du canal à l'émetteur (CSIT) peuvent réduire les gains obtenus par les techniques MIMO. Dans les scénarios à cellules multiples, les concepts de coordination et de coopération entre des liens (cellules) interférant les uns avec les autres ont émergé afin de faire face, voire de bénéficier de l'interférence. Malheureusement, ceux-ci nécéssitent l'échange d'information entre les entités qui coopèrent, ce qui consomme des ressources. Un autre problème est celui de la quantité et du type de données qui doivent être partagées entre les différentes stations de base. Par exemple, si les données de chaque utilisateur ne sont pas partagées, seules des techniques d'évitement d'interférence (interference avoidance) sont possibles, tandis que les données des utilisateurs doivent être partagées complètement pour que leur transmission de façon jointe soit possible. Comme démontré dans cette thèse, des techniques intermédiaires peuvent être considérées.

Tenant compte de ces problèmes, cette thèse traite certains aspects liés à deux scénarios dans lesquels des techniques MIMO peuvent être implémentées, et se centre sur la transmission en voie descendante dans les deux.

Dans un premier temps, nous nous intéresserons au canal broadcast

MIMO à feedback limité : la voie descendante d'une cellule isolée où une station de base dotée de N_t antennes sert $K \geq N_t$ utilisateurs mono-antenne est considérée. Dans un tel système, la connaissance du canal au niveau de l'émetteur (channel state information at the transmitter (CSIT)) joue un rôle particulièrement important. Vue cette importance et le fait que l'acquisition de cette connaissance nécessite l'usage de ressources en voie ascendante, plusieurs travaux de recherche se sont penchés sur le problème de trouver des stratégies de feedback et d'émission sous feedback limité pour ce scénario. Ici, nous contribuons deux idées pour bénéficier au mieux de la ressource dédiée au feedback : la première consiste en une procédure de transmission en deux étapes, la deuxième est une stratégie décentralisée dans laquelle chaque utilisateur varie le nombre de bits utilisés pour quantifier le canal en maintenant un débit moyen.

Dans un deuxième temps, des scénarios multicellulaires sont considérés. Le modèle d'une cellule isolée pourrait correspondre plus ou moins à celui de réseaux cellulaires actuels où la planification des fréquences est telle que les cellules qui utilisent la même bande soient suffisamment distantes pour ne pas trop interférer les unes avec les autres. Ceci n'est cependant pas très efficace et des techniques MIMO peuvent être utilisées pour améliorer la performance en réutilisant la totalité du spectre dans chaque cellule. Ainsi l'incorporation de plusieurs antennes au niveau dune station de base permet d'atténuer l'interférence qu'elle cause au niveau des cellules voisines. Dans cette thèse, une stratégie de précodage a été proposée : celle-ci nécessite une connaissance locale du canal seulement et sa performance dans de nombreux cas est analogue à celle de stratégies centralisées ou itératives. Des techniques MIMO dans un contexte multicellulaire peuvent également être étendues au traitement conjoint des signaux (multicell processing (MCP)) : les différentes stations de base forment ainsi un émetteur multi-antennaire géant. Toutefois, ceci aurait un coût exorbitant en termes de backhaul pour le partage 1) des données à envoyer et 2) de l'information sur le canal. Nous commencerons donc à examiner comment traiter les limitations liées à ces deux aspects.

Systèmes multi-utilisateurs multicellulaires MIMO

On commence par introduire le modèle du canal ainsi que les suppositions/hypothèses adoptées.

- Modèle du canal: Seuls des canaux à bande étroite (à évanouissement uniforme) sont traités. Le canal va en général varier dans le temps.

Cependant, on supposera qu'il reste plus ou moins le même pour une durée suffisamment longue pour que les techniques proposées puissent être appliquées.

- Modèle du récepteur : Tout au long de la thèse nous nous concentrons sur le cas de récepteurs mono-antennaires. On suppose également que chaque récepteur estime son canal parfaitement (CSIR parfait), ce qui peut être accompli de façon efficace grâce aux symboles pilotes émis depuis la station de base. En outre, chaque utilisateur décode son propre signal en traitant les signaux destinés aux autres utilisateurs comme du bruit (single user decoding (SUD)).

- Adaptation de lien idéale : on suppose une adaptation de lien idéale et la formule de capacité de Shannon, $\log_2(1 + SINR)$ bits/sec/Hz, est utilisée comme mesure du débit.

- On suppose qu'il y a toujours des données à transmettre à chaque utilisateur.

Voie de retour limitée sur la voie descendante d'un système MISO multi-utilisateur

Comme signalé dans l'introduction, la présence de CSIT est essentielle pour une bonne performance dans un système multi-utilisateur. Cependant, l'obtention de CSIT nécessite (surtout si les canaux ne sont pas réciproques) l'utilisation d'une voie de retour (feedback), ce qui consomme des resources de la voie ascendante. D'où l'intérêt porté aux systèmes avec feedback limité (voir par exemple, [1] et les références qui y sont contenues, de même que [2]).

Cette partie de la thèse s'intéresse donc à la transmission sur la voie descendante d'un système MISO multi-utilisateur, où une station de base ayant N_t antennes sert K utilisateurs pourvus d'un récepteur mono-antenne, avec $K > N_t$. En particulier, on adopte, comme dans beaucoup de travaux portant sur le feedback limité, une technique de formation de faisceaux (beamforming) linéaire, un précodeur zéro-forcing pour être plus précis. L'objectif de la station de base est de maximiser le débit total des données transmises. Le scénario est présenté dans la Figure 1.

Voie de retour: sélection d'utilisateurs et précodage

Les résultats obtenus se basent sur les observations suivantes:

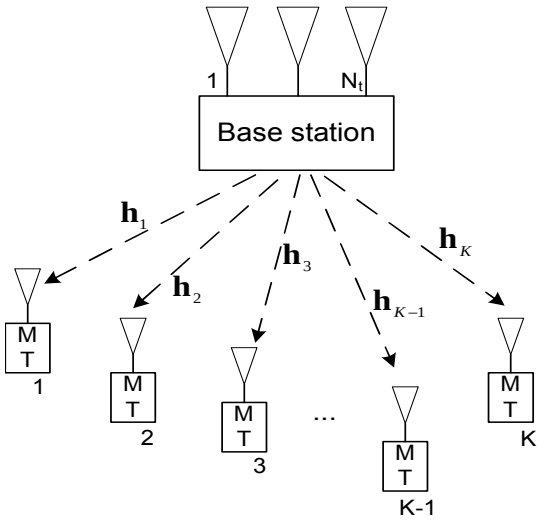


Figure 1 : Canal MIMO Broadcast : une station de base équipée de N_t antennes communique avec K récepteurs à une antenne chacun. $\mathbf{h}_k \in \mathbb{C}^{1\times N_t}$ est le vecteur des coéfficients du canal en voie descendante de l'utilisateur k, $k = 1, \ldots, K$.

- Bien que la sélection optimale d'utilisateurs nécessite la connaissance des canaux des $K >> N_t$ utilisateurs, le précodage final n'a besoin de connaître que les canaux de N_t utilisateurs au plus (ceux qui ont été ordonnancés).

- Bien que la matrice de précodage finale nécessite une connaissance précise des canaux des utilisateurs sélectionnés pour conserver le gain de multiplexage spatial (spatial multiplexing gain), l'algorithme d'ordonnancement peut se contenter d'une représentation moins bonne des canaux des différents utilisateurs.

Alors que dans la plupart des approches antérieures au travail présenté ici sur ce sujet, la ressource dédiée à la voie de retour était utilisée en une seule étape aux deux fins de sélection de l'utilisateur (ou ordonnancement) et de conception de la matrice de précodage, cette thèse a exploré deux pistes qui exploitent les remarques ci-dessus:

- une première approche où le feedback se fait en deux étapes,

- une deuxième approche où chaque utilisateur adapte son feedback à l'état actuel de son canal.

Feedback en deux étapes

Une ressource fixe est allouée à la voie de retour pendant chaque intervalle d'ordonnancement. Quantifiée en bits, cette dernière est de B_{total}. Elle est divisée en deux:

1. pendant la première étape, chaque utilisateur quantifie son canal avec B_1/K, $B_1 = \alpha B_{\text{total}}, \alpha \in [0,\ 1]$, et renvoie cette information à la station de base afin que cette dernière choisisse les utilisateurs à servir. Comme l'objectif est de maximiser le débit total pour une matrice de précodage zero-forcing, et que la station de base utilise la version quantifiée des canaux dont elle dispose comme s'il s'agissait des vrais canaux, l'ensemble $\hat{\mathcal{A}}$ des utilisateurs sélectionnés est obtenu de telle manière que

$$\hat{\mathcal{A}} = \arg \max_{\mathcal{A}\subset\{1,\ldots,K\},|\mathcal{A}|=N_t} \frac{1}{\text{tr}((\hat{\mathbf{H}}_{1,\mathcal{A}}\hat{\mathbf{H}}_{1,\mathcal{A}}^H)^{-1})}, \tag{0.1}$$

   où $\hat{\mathbf{H}}_{1,\mathcal{A}} \triangleq [\hat{\mathbf{h}}_{1,A_1}^{\text{T}} \ldots \hat{\mathbf{h}}_{1,A_{N_t}}^{\text{T}}]^{\text{T}}$, $\hat{\mathbf{h}}_{1,A_k} \in \mathbb{C}^{1\times N_t}$ étant la représentation du canal renvoyée par le $k^{\text{ème}}$ utilisateur de l'ensemble $\mathcal{A} \subset \{1,\ldots,K\}$ dans cette étape.

2. pendant la deuxième étape, les utilisateurs sélectionnés renvoient B_2/N_t, $B_2 = (1-\alpha)B_{\text{total}}$ bits additionnels pour mieux décrire leur canal. La station de base combine l'information obtenue dans les deux étapes en une matrice $\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}$ correspondant aux canaux des utilisateurs sélectionnés qui sera utilisée pour former la matrice de précodage:

$$\hat{\mathbf{W}}_{ZF} = \sqrt{P}\frac{\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}^{\dagger}}{\sqrt{\text{tr}((\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}^H)^{-1})}}. \tag{0.2}$$

Le problème est de déterminer la division optimale, caractérisée par le paramètre α, entre les deux étapes. Pour des utilisateurs dont les canaux sont independants et identiquement distribués et Rayleigh, tels que $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$ pour $k = 1,\ldots,K$, et un modèle de la quantification basé sur la théorie du débit de distorsion [3], une approximation nous mène à choisir α

tel que f_{PL} ci-dessous est maximisé

$$f_{PL} \triangleq \frac{1-\sigma_{e_1}^2}{1+P\sigma_{e_2}^2} + \frac{\sigma_{e_1}^2-\sigma_{e_2}^2}{\log K(1+P\sigma_{e_2}^2)}, \tag{0.3}$$

avec

$$\sigma_{e_1}^2 = 2^{-\alpha B_{total}/(K\times N_t)}, \quad \sigma_{e_2}^2 = 2^{-\frac{B_{total}}{N_t}(\frac{\alpha}{K}+\frac{1-\alpha}{N_t})} \tag{0.4}$$

f_{PL} représente la perte en SNR due à la connaissance imparfaite du canal, $\sigma_{e_1}^2$ est la variance de l'erreur de quantification à la première étape, $\sigma_{e_2}^2$ celle de l'erreur finale.

L'intérêt de cette approche est illustré dans la Figure 2, où l'adaptation du paramètre α permet de bénéficier d'un gain multi-utilisateur quand c'est possible (SNR pas trop élevé) et de se concentrer sur le gain de multiplexage quand ce n'est pas le cas.

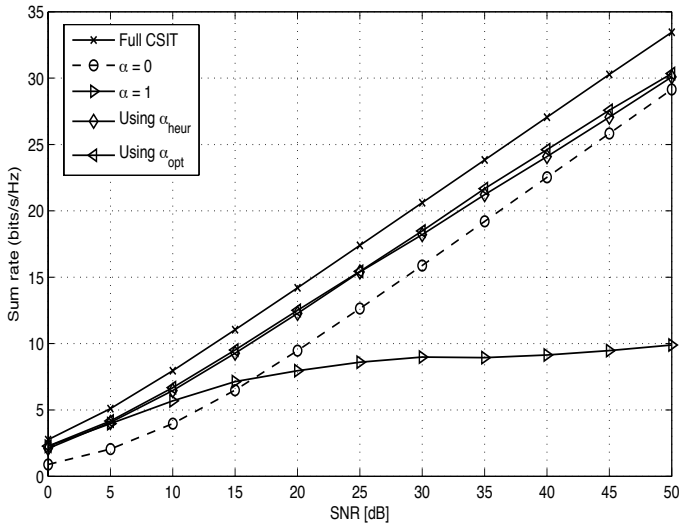


Figure 2: Débits moyens pour $N_t = 2$, $K = 30$, $B_{total} = 120$ bits, et différentes valeurs de α. Les débits obtenus en utilisant α_{heur} qui maximise f_{PL} sont presque aussi élevés que ceux obtenus en utilisant α optimal, α_{opt}.

Feedback Adaptatif Décentralisé

Dans cette deuxième approche, chaque utilisateur adapte son feedback à l'état actuel de son canal. Ainsi, il tient compte de ce dernier pour évaluer

le débit auquel il peut s'attendre et ses chances d'être ordonnancé et décider de manière décentralisée du nombre de bits[1] à utiliser pour renvoyer une version quantifiée de son canal. Ceci est effectué de façon à maximiser le débit moyen de chaque utilisateur (ici aussi, tous les utilisateurs sont supposés avoir une distribution identique du canal) tout en conservant un débit moyen de B_{total}/K bits.

Vue la difficulté d'estimer le débit moyen d'un utilisateur, des approximations sont adoptées, permettant de formuler l'optimisation à faire en fonction d'une série de problèmes d'optimisation fonctionelle convexes. Cette stratégie semi-optimale pour l'approximation faite est comparée à une autre stratégie empirique où chaque utilisateur a à sa disposition un codebook unique pour quantifier son canal, mais qu'il utilise ce dernier de façon selective selon l'état de son canal et l'erreur de quantification (si l'erreur est trop grande, et donc le débit attendu trop faible, l'utilisateur peut choisir de ne pas utiliser la voie de retour).

Dans cette approche, on suppose que les utilisateurs quantifient la direction de leur canal (CDI) [4–6]. Pour un canal $\mathbf{h}$ dont la direction $\tilde{\mathbf{h}} = \mathbf{h}/\|\mathbf{h}\|$ est quantifiée en $\hat{\mathbf{h}}$ tels que $\|\hat{\mathbf{h}}\| = 1$, l'erreur de quantification est donnée par $\sin^2\theta = 1 - |\tilde{\mathbf{h}}\hat{\mathbf{h}}^H|^2$. La fonction de distribution cumulative suivante est utilisée pour approximer celle de l'erreur de quantification :

$$F_{\sin^2\theta}(x) = \begin{cases} \delta^{1-N_t} x^{N_t-1} & 0 \le x \le \delta \\ 1 & x > \delta \end{cases} \tag{0.5}$$

où $\delta \triangleq 2^{-b/(N_t-1)}$, b étant le nombre de bits utilisés pour quantifier $\tilde{\mathbf{h}}$. Cette fonction est une borne supérieure à celle de l'erreur de n'importe quel quantificateur [7] et a été exploitée dans nos dérivations.

Les Figures 3 et 4 permettent de comparer la performance de plusieurs stratégies pour un débit de feedback moyen de 4 bits par utilisateur. Les simulations sont faites pour $N_t = 2$, et pour différentes valeurs du paramètre de semi-orthogonalité ϵ, utilisé dans l'algorithme SUS d'ordonnancement [8] qu'on adopte ici. Les stratégies dans les figures correspondent à:

- une stratégie non-adaptative (marquée 'Fixed Fb' dans les figures), telle que chaque utilisateur renvoie sa direction de canal quantifiée grâce à un codebook unique (de taille égale à 2^4),
- la stratégie adaptative où chaque utilisateur utilise un codebook unique de façon sélective ((marquée 'Single Codebook Adaptive Fb' dans les figures et décrite avec plus de détails dans la Section 3.5.6),

[1] À noter qu'un plus grand nombre de bits correspond à une représentation plus précise et donc moins d'interférence en réception.

- la stratégie adaptative semi-optimale (marquée 'Fully Adaptive Fb' dans les figures) décrite dans la Section 3.5.5, où l'utilisateur choisit entre plusieurs codebooks de tailles différentes, optimisées pour maximiser une approximation du débit moyen,

- la même stratégie que la stratégie adaptative semi-optimale, mais avec les tailles des codebooks tronquées (pour assurer qu'il s'agit de nombres entiers, marquée 'Truncated Fully Adaptive Fb' dans les figures).

Dans la Figure 4, on tente de tenir compte de la signalisation nécessaire pour annoncer à la station de base quel codebook sera utilisé, en montrant une courbe supplémentaire marquée 'fixed Fb (with entropy)', qui correspond à un codebook unique qui est toujours utilisé et dont la taille est 2^{4+e}, où e est l'entropie de la source correspondant la taille des codebooks dans la stratégie adaptative semi-optimale tronquée. Si le nombre d'utilisateurs est trop petit (5 par exemple), une telle stratégie surpasse notre stratégie adaptative, mais pour un plus grand nombre d'utilisateurs (par example 20 comme dans la Figure 4), le gain dû à la stratégie adaptative est significatif.

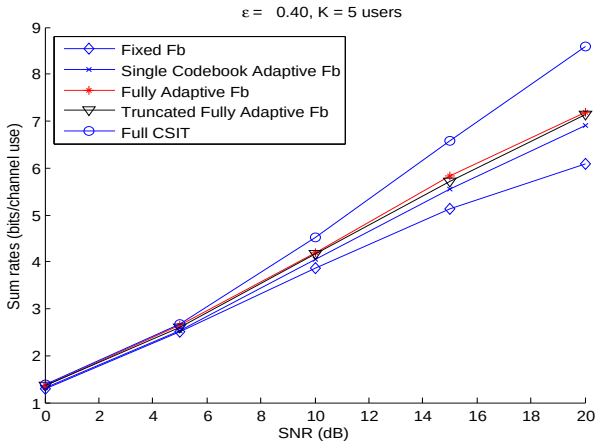


Figure 3: Débits moyens pour $N_t = 2$, $K = 5$, $B_{\text{total}} = 20$ bits, et différentes stratégies adaptatives et non-adaptatives.

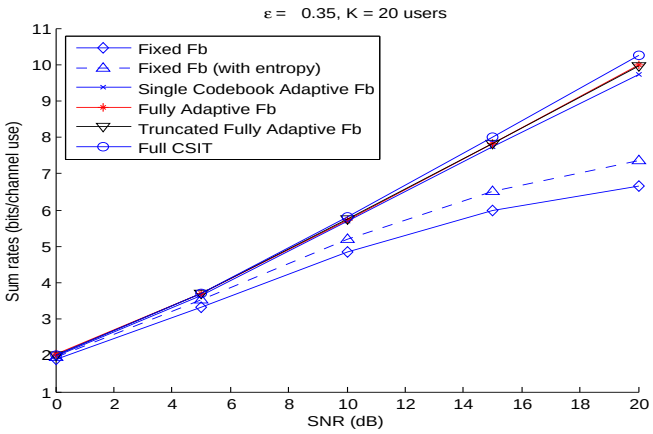


Figure 4: Débits moyens pour $N_t = 2$, $K = 20$, $B_{\text{total}} = 80$ bits, et différentes stratégies adaptatives et non-adaptatives.

Coordination sur le canal à interférence MISO et SINRs virtuels

Dans cette partie, on commence à considérer des configurations multicellulaires. Comme noté dans l'introduction, dans les scénarios impliquant plusieurs émetteurs et récepteurs, la performance obtenue, en termes de débits atteints par exemple, dépend de l'information partagée par les différents nœuds. Ainsi, si les émetteurs ou les récepteurs partagent toutes les données et sont donc capables, respectivement, d'émettre et de décoder de façon conjointe, on obtient un canal de diffusion (Broadcast Channel, BC) et un canal à accès multiples (Multiple Access Channel, MAC). Toutefois, si ce n'est pas le cas, un canal à interférence résulte. Ceci correspond à la situation considérée ici. Plus précisément, on traite la voie descendante dans laquelle les K stations de base sont dotées de $N_t > 1$ antennes chacune, en d'autres termes il s'agit d'un canal à interférence MISO. Nous proposons une stratégie de précodage linéaire, nécéssitant une connaissance locale du canal seulement, basée sur la notion de SINR virtuel: le scénario étudié est illustré par la Figure 5 ci-dessous pour $K = 2$ and $N_t = 3$.

Soit $\mathbf{h}_{j,k}$ le canal de propagation entre l'utilisateur j et la station de base k. On suppose que la station de base k connait, pour $j = 1, \ldots, K$, $\mathbf{h}_{j,k}$, mais pas les $\mathbf{h}_{j,\bar{k}}$ pour $\bar{k} \neq k$, évitant ainsi un échange de toute l'information sur les canaux. Dans la stratégie proposée, on évite également le recours à

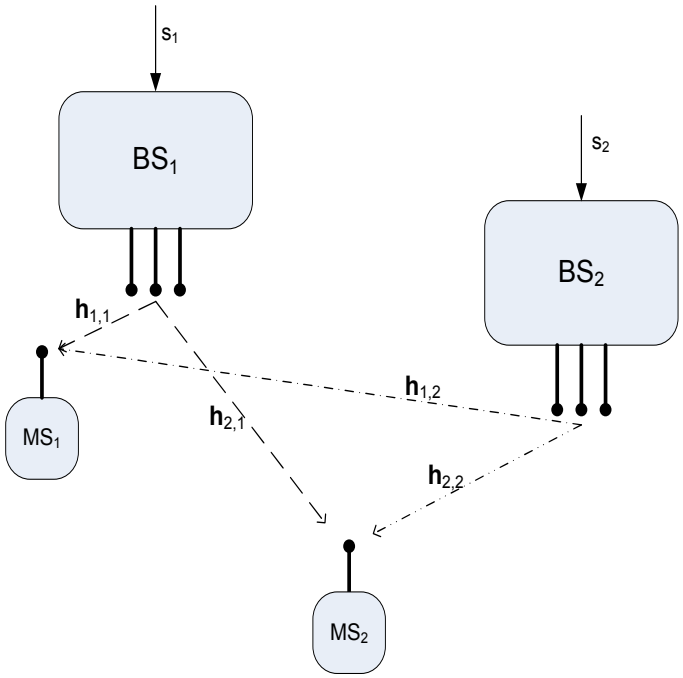


Figure 5: Scénario considéré avec $K = 2$, $N_t = 3$. $\mathbf{h}_{1,1}, \mathbf{h}_{2,1}$ sont connus à la station de base BS_1, $\mathbf{h}_{1,2}, \mathbf{h}_{2,2}$ à la BS_2.

des algorithmes itératifs qu'on retrouve dans la littérature sur le MISO IC.

Soit $\mathbf{w}_k = \sqrt{p_k}\mathbf{u}_k$ le vecteur de précodage destiné à la transmission par la station de base k des symboles (supposés $\mathcal{CN}(0,1)$) de son utilisateur: le vecteur $\mathbf{u}_k$ a une norme égale à 1, p_k est donc la puissance d'émission de cette station de base. On définit un SINR virtuel comme suit:

$$\gamma_k^{virtual} = \frac{p_k|\mathbf{h}_{k,k}\mathbf{u}_k|^2}{\sigma^2 + \sum_{j\neq k}\alpha_{kj}p_k|\mathbf{h}_{j,k}\mathbf{u}_k|^2}.$$

Ce rapport est appelé SINR virtuel car il représenterait le SINR du flux k en voie ascendante virtuelle où les canaux sont les transposées de ceux de la voie descendante réelle, et où i) chaque l'utilisateur j émet avec une puissance $\alpha_{kj}p_k$ ($\alpha_{kk} = 1$, $\alpha_{kj} \geq 0$ sont des paramètres à spécifier pour les autres utilisateurs) et ii) le vecteur $\mathbf{u}_k$ est utilisé pour traiter le signal reçu.

Soit P la puissance maximale de chaque émetteur, on démontre que pour le cas où $K = 2$, $p_k = P$ pour $k = 1, 2$, $\alpha_{12} = \alpha_{21} = 1$, et des vecteurs de précodage qui maximisent les SINR virtuels correspondants résultent en des débits qui sont *optimaux au sens de Pareto*, c'est-à-dire que le débit de l'utilisateur 1 ne peut être augmenté sans diminuer celui de l'utilisateur 2 et vice versa. On prouve également que pour $N_t \geq K > 2$, $p_k = P, k = 1, \ldots, K$, les vecteurs de précodage obtenus maximisant les SINR virtuels pour $\alpha_{jk} = 1, j, k = 1, \ldots, K$ satisfont la paramétrisation nécéssaire à la Pareto optimalité. En outre, nos simulations montrent que sous ces conditions, les débits obtenus sont presque Pareto-optimaux.

L'algorithme proposé impose de toujours émettre à la puissance maximale et de choisir les vecteurs de précodage tel que la métrique suivante, un SINR virtuel, est maximisée:

$$\mathbf{u}_k = \arg\max \frac{|\mathbf{h}_{k,k}\mathbf{u}_k|^2}{\frac{\sigma^2}{P} + \sum_{j \neq k} |\mathbf{h}_{j,k}\mathbf{u}_k|^2}.$$

Sa performance est évaluée en la comparant avec:

- d'autres algorithmes se basant sur la même connaissance locale du canal, à savoir la stratégie égoïste de maximum ratio transmission (MRT) telle que $\mathbf{w}_k^{MRT} = \sqrt{P}\frac{\mathbf{h}_{k,k}^H}{\|\mathbf{h}_{k,k}\|}$ et la stratégie altruiste qui minimise l'interférence causée aux autres utilisateurs (si $K \leq N_t$, il s'agit d'une stratégie zero-forcing). Ceci est illustré dans la Figure 6 pour un système de 7 cellules où la distribution des utilisateurs dans leurs cellules respectives est aléatoire et uniforme et le modèle du canal de propagation comprend des pertes d'espace (path loss) selon le modèle COST-231, shadowing log-normal et affaiblissement rapide (fast fading) de Rayleigh.

- l'algorithme centralisé optimal (dans le sens qu'il consomme le moins de puissance possible) qui donne lieu aux mêmes débits que notre algorithme distribué. Le même modèle du canal que dans la comparaison précédente est utilisé pour obtenir les résultats de la Figure 7. La figure illustre que l'incorporation de plus d'antennes à la station de base augmente l'efficacité de l'algorithme distribué comparé à l'algorithme centralisé, du moins en ce qui concerne la réalisation des débits qui correspondent à la maximisation de SINRs virtuels.

- l'algorithme décentralisé mais itératif proposé dans [9]. Nos résultats montrent que pour le cas où $K \leq N_t$, notre algorithme est au moins

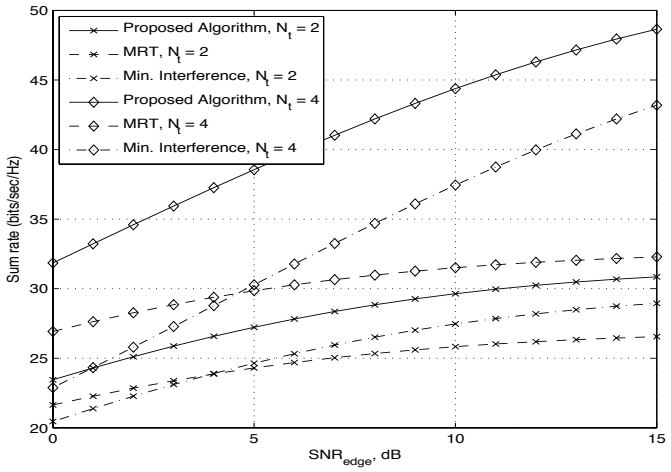


FIgure 6: Somme des débits vs. SNR en bord de cellule (cell-edge SNR) pour $N_t = 2, 4$ et $K = 7$.

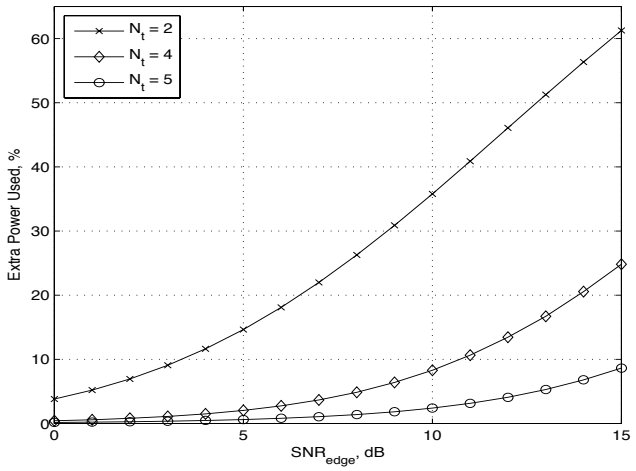
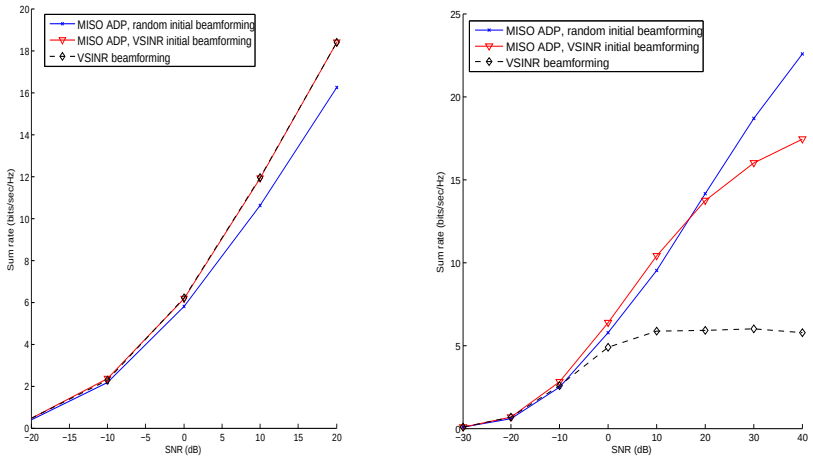


Figure 7: Puissance additionelle résultant de la nature distribuée de notre méthode vs. SNR en bord de cellule pour $N_t = 2, 4, 5$ et $K = 7$.

aussi efficace que l'algorithme itératif si l'objectif est de maximiser la somme des débits. Ceci n'est plus le cas si cette condition n'est pas satisfaite, comme illustré dans la Figure 8.



(a) $K = 2$ (b) $K = 3$

Figure 8: Somme des débits moyens vs. SNR pour des canaux Rayleigh i.i.d. et $N_t = 2$.

Network MIMO Coopératif avec connaissance distribuée des canaux

Dans cette partie et celle qui suit, nous considérons des canaux où les émetteurs peuvent servir un groupe d'utilisateurs mobiles de façon coopérative, en d'autres termes les données des utilisateurs sont partagés par plusieurs émetteurs. Un tel procédé, appelé network MIMO, a été traité dans [10], [11] par exemple. Pour pouvoir implémenter des méthodes de network MIMO en voie descendante, soit les données et l'information sur le canal (CSI) doivent être connues de tous les émetteurs collaborants (pour que ces derniers puissent choisir la matrice de précodage et que le signal d'un utilisateur puisse lui arriver de plusieurs sources), sinon que le CSI soit renvoyé à un processeur central qui conçoit le précodage et informe chaque émetteur de son signal à émettre. Ici, on considère des scénarios où, afin de *limiter ces échanges* et potentiellement réduire les délais qui vont avec, les différents émetteurs ont accès aux données des différents utilisateurs mais pas la to-

talité de l'information sur le canal. Cette situation est illustrée dans la Figure 9 pour N stations de base, équipees de N_t antennes chacune, communiquant avec K stations mobiles.

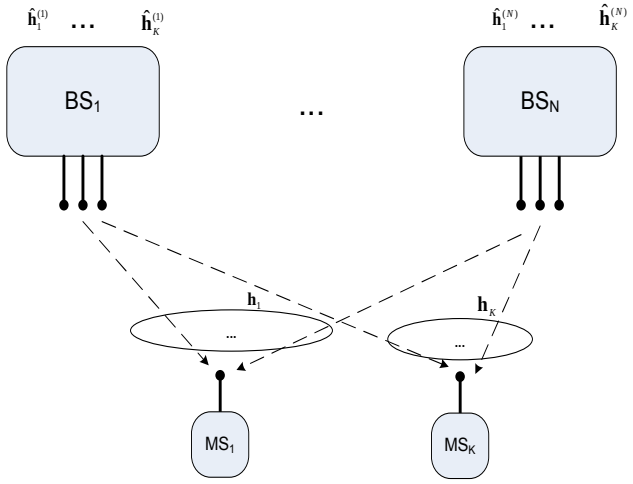


Figure 9: MIMO coopératif avec CSIT imparfaitement partagé.

Layered Virtual SINR

Dans un premier temps, la connaissance du canal est supposée telle que chaque station de base a une connaissance locale du canal, semblable à celle de la partie précédente traitant le canal MISO à interférence. L'approche de maximisation de SINRs virtuels est adaptée à ce nouveau scénario: on dénote cette adaptation, layered virtual SINR (SINR virtuel à plusieurs couches). Soit $\mathbf{w}_k \in \mathbb{C}^{NN_t}$ le vecteur de précodage des symboles de l'utilisateur k. On peut le réécrire en termes des 'sous-vecteurs' correspondant à chaque station de base: ainsi, $\mathbf{w}_k = [\mathbf{w}_{k1}; \ldots; \mathbf{w}_{kN}]$. Chaque $\mathbf{w}_{kj}, k = 1, \ldots, K, j = 1, \ldots, N$ peut être écrit en fonction de la puissance qui lui est allouée p_{kj} et de son vecteur directeur $\mathbf{u}_{kj}$ de norme 1 comme $\mathbf{w}_{jk} = \sqrt{p}_{kj}\mathbf{u}_{kj}$. Le SINR de l'utilisateur k est donc égal à

$$\gamma_k = \frac{|\sum_{j=1}^{N} \mathbf{h}_{k,j}\mathbf{w}_{k,j}|^2}{\sigma^2 + \sum_{\bar{k}=1,\ldots,K,\bar{k}\neq k} |\sum_{j=1}^{N} \mathbf{h}_{k,j}\mathbf{w}_{\bar{k},j}|^2}.$$

Dans le cadre d'une maximisation d'un SINR virtuel, $\mathbf{u}_{jk}$ est tel que:

$$\mathbf{u}_{k,j} = \arg \max_{\|\mathbf{u}\|^2=1} \frac{|\mathbf{h}_{k,j}\mathbf{u}|^2}{\frac{\sigma^2}{p_{kj}} + \sum_{\bar{k}=1,\ldots,K,\bar{k}\neq k} |\mathbf{h}_{\bar{k},j}\mathbf{u}|^2}.$$

Comme la puissance d'une station de base doit être partagée entre les utilisateurs servis, les différents p_{kj} doivent être déterminés tels que, pour $j = 1, \ldots, N$, $\sum_{k=1}^{K} p_{kj} \leq P$. Vue l'intractabilité du problème consistant à choisir les puissances afin de maximiser le débit total par exemple, deux stratégies empiriques sont proposées dont la suivante:

$$p_{kj} = \frac{\|\mathbf{h}_{k,j}\|^2}{\sum_{l=1}^{K} \|\mathbf{h}_{l,j}\|^2} P.$$

Nous évaluons la technique proposée en la comparant au précodage *zero-forcing centralisé optimal* (qui engage à partager le CSI en entier) et le précodage basé sur la maximisation de SINRs virtuels à chaque émetteur de la partie précédente (qui évite le partage du CSI de même que celui des symboles à transmettre, puisque chaque utilisateur est servi par une station de base unique). Ceci est illustré dans la Figure 10, pour différents P et $K = N = 2$; les canaux des 2 utilisateurs sont tels que $\mathbf{h}_{1,1}$ et $\mathbf{h}_{2,2}$ sont $\mathcal{CN}(\mathbf{0}, \mathbf{I})$ alors que $\mathbf{h}_{1,2}$ et $\mathbf{h}_{2,1}$ sont $\mathcal{CN}(\mathbf{0}, \beta\mathbf{I})$, β étant un paramètre de la simulation. Le cas illustré montre que pour un SNR faible, la technique distribuée proposée peut surpasser certaines techniques centralisées. Contrairement au canal à interférence, un β croissant permet d'accroître les débits obtenus au lieu de les diminuer.

Décision d'équipe pour MIMO coopératif avec CSI distribué

Dans un deuxième scénario, on propose un *nouveau* modèle de CSIT distribué qui pourrait correspondre au cas où chaque utilisateur renvoie, en voie de retour, son CSI quantifié à l'aide d'un *codebook hiérarchique*: il en résulte que chaque station de base arrive à décoder la version quantifiée jusqu'à un certain degré de précision seulement. Toutes les stations de base connaissent les statistiques des canaux des différents utilisateurs, de même que les degrés de précision avec lesquelles ces canaux sont connus à chacune d'entre elles.

Le problème de conception du précodage dans ces conditions peut être formulé comme un problème de décision d'équipe: les stations de base sont les membres d'une équipe puisqu'elles ont un objectif commun, maximiser le débit total du réseau par exemple, mais doivent baser leurs décisions

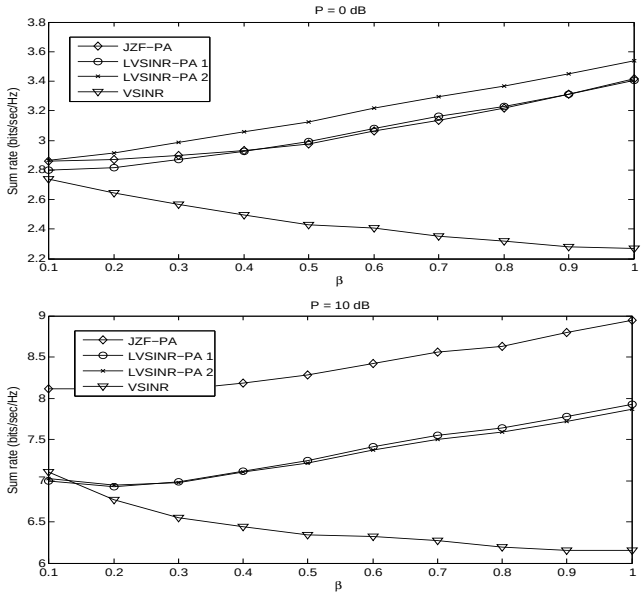


Figure 10: Comparaison de la somme des débits pour $N = K = 2$, $\mathbf{h}_{1,1}$ et $\mathbf{h}_{2,2}$ sont $\mathcal{CN}(\mathbf{0}, \mathbf{I})$; $\mathbf{h}_{1,2}$ et $\mathbf{h}_{2,1}$ sont $\mathcal{CN}(\mathbf{0}, \beta\mathbf{I})$.

individuelles sur des connaissances qui ne sont en général pas communes. On formule le probléme d'optimisation dans un cadre Bayésien et on se propose de trouver les stratégies de beamforming décentralisées de chaque station de base, qui varient uniquement en fonction de leur connaissance locale et permettent de maximiser une utilité moyenne, la moyenne de la somme des débits par exemple.

Pour illustrer cette approche, on considère le cas où 2 utilisateurs sont servis par 2 stations de base. Les canaux sont supposés Rayleigh symétriques: ainsi, $h_{k,k}, k = 1, 2$, sont $\mathcal{CN}(0, 1)$ alors que $h_{k,\bar{k}}, k = 1, 2, \bar{k} = \mod(k, 2) + 1$, sont $\mathcal{CN}(0, \beta)$. Des codebooks hiérarchiques de 6 bits tels que la station de base la plus éloignée connait les deux premiers bits seulement et l'autre les connait tous, sont obtenus par l'algorithme de Lloyd. La performance des stratégies issues de la formulation de décision d'équipe est comparée à différentes bornes supérieures et inférieures:

- borne supérieure: les stations de base partagent leurs connaissances (6 bits par utilisateur) et conçoivent la matrice de précodage conjointement; les débits obtenus sont meilleurs mais le coût en termes de voie

de retour est supérieur.

- borne inférieure: les stations de base ont seulement accès aux canaux quantifiés moins précisément (2 bits par utilisateur) et conçoivent la matrice de précodage conjointement.

- borne inférieure: chaque station de base traite l'information qu'elle a comme si elle était commune aux deux stations de base et conçoit son précodage sous ces conditions; on appelle cette approche 'myope', parce que la station de base fonde ses décisions sur une supposition qui est fausse.

La Figure 11 suggère qu'il vaut mieux baser les décisions sur une connaissance moins bonne mais commune aux différentes entités coopérant que d'avoir chacune d'entre elles décidant de sa stratégie sous la supposition érronée que sa connaissance individuelle est commune à toutes les stations. La formulation Bayésienne surpasse ces deux approches.

Network MIMO sous backhaul limité

Dans cette partie, on suppose que la connaissance du canal est commune et parfaite aux stations de base, et on se concentre sur le problème de backhaul limité dédié au transport des messages des utilisateurs vers celles-ci. Le scénario traité et l'approche de fractionnement des taux proposée sont illustrés dans la Figure 12. En particulier, le taux de l'utilisateur k, dénoté r_k, pour $k = 1, 2$, est divisé en deux parties:

- des messages privés de taux égal à $r_{k,p}$, qui sont acheminés à une seule station de base, l'èmetteur k.

- des messages communs de taux égal à $r_{k,c}$, qui sont acheminés vers plusieurs stations de base qui coopèrent pour les transmettre.

Tenant compte des contraintes sur le backhaul (C_1 et C_2 pour les stations de base 1 et 2, respectivement), $r_1 + r_2 - r_{2,p} \leq C_1$ et $r_1 + r_2 - r_{1,p} \leq C_2$. À noter que si $r_{k,c} = 0$ pour les deux utilisateurs, le système est limité à une transmission coopérative (network MIMO) alors si $r_{k,p} = 0$ pour les deux utilisateurs, on obtient un canal à interference, donc une transmission conventionnelle.

Une région de taux atteignables de même que la conception du précodage pour les atteindre sont dévelopées. On se base notamment sur le fait que le canal entre les deux stations de base et chacun des utilisateurs est analogue

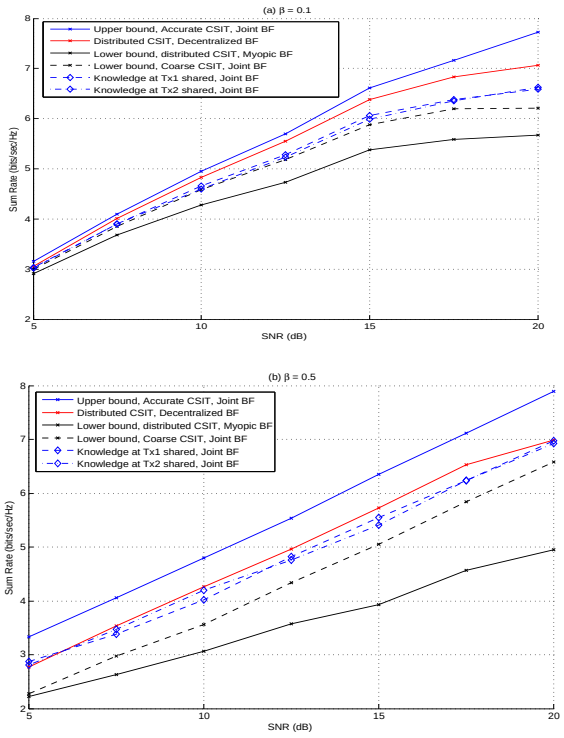


Figure 11: Comparaison de la somme moyenne des débits pour differents β.

à un MAC avec un message commun [70] si la transmission dédiée à l'autre utilisateur est considérée comme du bruit, ce qui est d'ailleurs le cas. La conception du précodage tient compte du fait que l'optimisation correspondante peut être résolue par une succession d'optimisations convexes.

La Figure 13 montre les régions de taux correspondant à l'approche proposée, marquée 'Hybrid IC/Network MIMO', au canal à interférence et au canal network MIMO pour différentes valeurs de la capacité du backhaul, pour $N_t = 2$ et les valeurs des canaux de propagation indiquées dans le titre de cette figure. Comme le montre la figure, si la capacité du backhaul est inférieure à celle du canal à interférence sans limitation du backhaul, les régions de taux de notre approche hybride et du canal à interférence se superposent et toutes deux sont plus grandes que celle de network MIMO. Pour un backhaul croissant, les trois régions deviennent plus grandes (jusqu'à ce que le système ne soit plus limité par le backhaul), la région de network

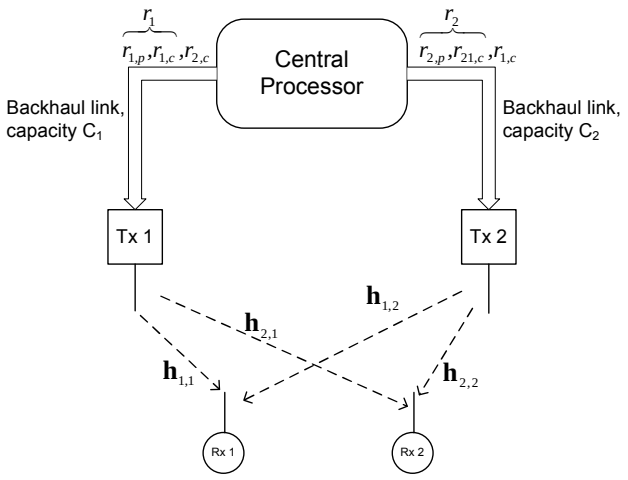


Figure 12: Coopération sous un backhaul de capacité limitée.

MIMO devient plus grande que celle du canal à interférence et s'approche de l'approche hybride jusqu'à ce que ces deux régions deviennent identiques.

Analyse Grands Systèmes et Beamforming

Dans cette dernière partie, on se propose d'appliquer des résultats se basant sur la théorie des grandes matrices aléatoires pour proposer des méthodes simples mais efficaces de précodage et analyser leurs débits sous différents degrés de coopération. Plus précisément, on se limite à un réseau composé de deux cellules, dans lequel chaque station de base est équipée de N_t antennes et chaque cellule contient K utilisateurs; ceci est illustré dans la Figure 14. On étudie le débit maximal qui peut être maintenu vers tous les utilisateurs desservis pour des stratégies de précodage linéaires et les trois cas de figure suivants:

- Single cell Processing (SCP, traitement mono-cellulaire): chacune des stations de base reçoit les données dédiées aux utilisateurs situés dans sa cellule, et connait les canaux entre elle-même et ces derniers. Dans ce cas, le précodage ne peut tenir compte de l'interférence causée au niveau des utilisateurs se trouvant dans l'autre cellule.

- Coordinated Beamforming (CBf, Beamforming Coordonné): ici aussi chacune des stations de base reçoit les données dédiées aux utilisateurs

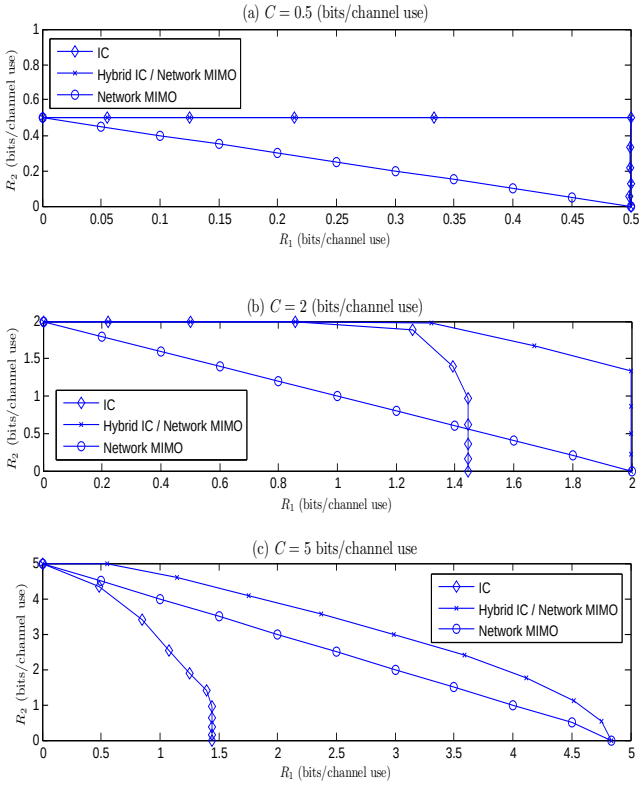


Figure 13: Comparaison de régions de taux pour un SNR de 10dB et des canaux tels que $\mathbf{h}_1 = [-0.3059 + \mathbf{i}0.2314 \quad .0886 - \mathbf{i}0.0132 \quad -.8107 - \mathbf{i}0.416 \quad .8409 - \mathbf{i}0.0964]$, $\mathbf{h}_2 = [-1.1777 + \mathbf{i}0.1235 \quad 0.2034 + \mathbf{i}0.5132 \quad 0.8421 + \mathbf{i}1.5437 \quad -0.0266 + \mathbf{i}0.0806]$.

situés dans sa cellule mais les deux stations de base ont une connaissance commune et parfaite des canaux, ce qui permet un meilleur choix des matrices de précodage qui tient compte de l'interférence dans les deux cellules.

- Multicell Processing (MCP, traitement multicellulaire, network MIMO): les stations de base ont toutes deux accès aux données et à l'information des canaux de tous les utilisateurs. Elles peuvent donc mettre leurs antennes en commun pour servir les utilisateurs.

Notre modèle des canaux est tel que les canaux entre chaque station de base et chaque utilisateur sont indépendants les uns des autres. En

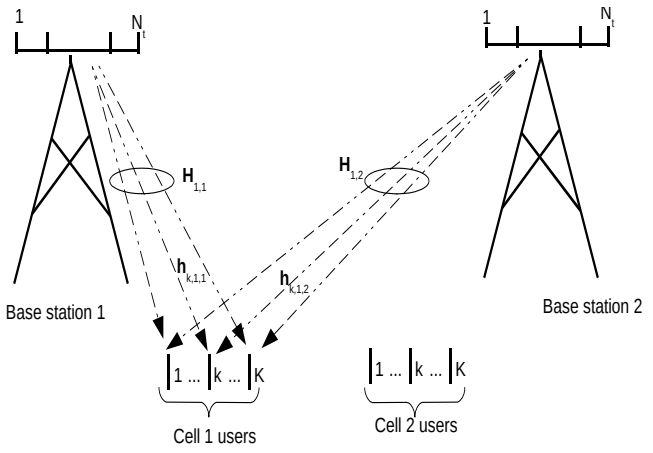


Figure 14: Modèle du système analysé

plus, le canal entre un utilisateur k, $k = 1, \ldots, K$, dans la cellule j, $j = 1, 2$ et la station de base j $\mathbf{h}_{k,j,j}$ est $\mathcal{CN}(\mathbf{0}, \mathbf{I})$, alors que le canal entre cet utilisateur et la station de base $\bar{j}, \bar{j} =$ mod $(j, 2) + 1$ est $\mathcal{CN}(\mathbf{0}, \epsilon\mathbf{I})$. Ce modèle est semblable au modèle de Wyner en voie descendante pour deux cellules [12, 13].

Dans chacun des cas de figure considérés, on formule le problème d'optimisation dont l'objectif est de déterminer le débit maximal qui peut être maintenu vers tous les utilisateurs du système. Dans chaque cas, ce problème peut être résolu par une méthode de dichotomie. Pour une valeur donnée du débit et par conséquent du SINR à maintenir, on peut formuler un problème d'optimisation *convexe*, c'est-à-dire, en supposant que le problème est faisable, dont le problème dual a la même valeur optimale. En particulier, la résolution des problèmes en voie descendante peut être liée à la résolution de problèmes duals correspondant à des voies ascendantes virtuelles. On se base essentiellement sur les travaux de [14, 15] pour formuler les problèmes duals ainsi que leurs solutions et celles des problèmes primaux en fonction des coéfficients de Lagrange correspondant aux contraintes de SINR et de puissance maximale par station de base: les premiers peuvent être interprétés comme les puissances de transmission, les seconds comme les puissances du bruit à chaque station de base en voie ascendante. Considérant la limite dans laquelle N_t et K tendent vers l'infini alors que le rapport K/N_t tend vers une constante qu'on dénote la charge de la cellule (cell loading) β simplifie la résolution des problèmes duals et permet un characterisation des valeurs optimales et des vecteurs de précodage asymptotiquement opti-

maux. Ainsi, pour une puissance d'émission maximale égale à P, le SINR maximal qui peut être réalisé à chaque utilisateur pour les différents cas de figures, de même que les vecteurs de précodage asymptotiquement optimaux qui permettent de l'atteindre sont résumés dans le théorème suivant.

Theorem 1. *Sous réserve d'une contrainte de puissance par station de base égale à P, les SINR maximaux qui peuvent être asymptotiquement (quand $N_t, K \to \infty, K/N_t \to \beta$) maintenus à chaque utilisateur pour β fixe pour chacun des cas de figure considérés sont définis par les équations de point fixe suivantes:*

$$\gamma^*_{SCP} = \frac{1}{\beta} \frac{1}{\frac{\sigma^2}{P} + \epsilon + \frac{1}{1+\gamma^*_{SCP}}}.$$
$$\gamma^*_{Coord} = \frac{1}{\beta} \frac{1}{\frac{\sigma^2}{P} + \frac{1}{1+\gamma^*_{Coord}} + \frac{\epsilon}{1+\epsilon\gamma^*_{Coord}}}.$$
$$\gamma^*_{MCP} = \frac{1}{\beta} \frac{1}{\frac{\sigma^2}{(1+\epsilon)P} + \frac{1}{1+\gamma^*_{MCP}}}.$$

Ces SINR peuvent être respectivement réalisés par les vecteurs de précodage asymptotiquement optimaux suivants:

$$\mathbf{w}^{SCP}_{kj} = \sqrt{\frac{P}{\beta N_t}} \frac{\hat{\mathbf{w}}^{SCP}_{kj}}{\|\hat{\mathbf{w}}^{SCP}_{kj}\|}, \quad \hat{\mathbf{w}}^{SCP}_{kj} = \left(\mathbf{I}_{N_t} + \frac{1}{\beta \left(\frac{\sigma^2}{P} + \epsilon \right) N_t} \sum_{\bar{k} \neq k} \mathbf{h}^H_{\bar{k},j,j} \mathbf{h}_{\bar{k},j,j} \right)^{-1} \mathbf{h}^H_{k,j,j},$$

$$\mathbf{w}^{Coord}_{kj} = \sqrt{\frac{P}{\beta N_t}} \frac{\hat{\mathbf{w}}^{Coord}_{kj}}{\|\hat{\mathbf{w}}^{Coord}_{kj}\|}, \quad \hat{\mathbf{w}}^{Coord}_{kj} = \left(\mathbf{I}_{N_t} + \frac{P}{\beta\sigma^2 N_t} \sum_{\bar{j},\bar{k},(\bar{j},\bar{k}) \neq (j,k)} \mathbf{h}^H_{\bar{k},j,j} \mathbf{h}_{\bar{k},j,j} \right)^{-1} \mathbf{h}^H_{k,j,j},$$

$$\mathbf{w}^{MCP}_{kj} = \sqrt{\frac{P}{\beta N_t}} \frac{\hat{\mathbf{w}}^{MCP}_{kj}}{\|\hat{\mathbf{w}}^{MCP}_{kj}\|}, \quad \hat{\mathbf{w}}^{MCP}_{kj} = \left(\mathbf{I}_{2N_t} + \frac{P}{\beta\sigma^2 N_t} \sum_{\bar{j},\bar{k},(\bar{j},\bar{k}) \neq (j,k)} \tilde{\mathbf{h}}^H_{\bar{k},j} \tilde{\mathbf{h}}_{\bar{k},j} \right)^{-1} \tilde{\mathbf{h}}^H_{k,j}$$

La simplicité des expressions obtenues dans le théorème précédent permet de comparer aisément la performance liée aux trois degrés de coopération étudiés, et donc d'évaluer si le coût additionnel associé à plus de coopération en vaut la peine, en évitant de recourir à des simulations de Monte Carlo. La validité des résultats asymptotiques pour un système de dimensions finies est démontrée dans la Figure 15: pour $K = 3, N_t = 4$, $\frac{P}{\sigma^2} = 10$ et ϵ dans $[.01; .1; .5; .8; 1]$, on résout les problèmes d'optimisation pour

différents échantillons indépendants des canaux (générés selon le modèle adopté) et on obtient les moyennes des débits réalisés qu'on compare aux débits qui résultent de l'analyse asymptotique. Même pour un aussi faible nombre d'antennes, l'analyse grands systèmes produit une bonne approximation.

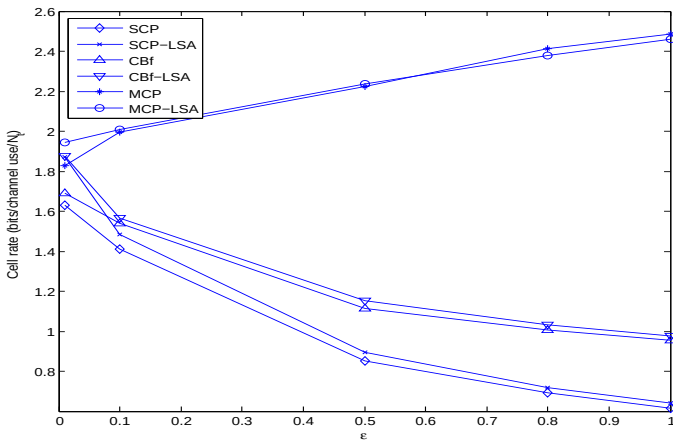


Figure 15: Comparaison pour SNR =10 dB, entre les débits moyens normalisés par cellule résultant des optimisations pour un système de dimensions finies ($K = 3, N_t = 4$) et ceux résultant de l'analyse grands systèmes avec $\beta = 0.75$.

L'analyse grands systèmes permet également d'obtenir des stratégies de précodage simples mais efficaces, comme le montre le théorème ci-dessus. L'applicabilité de ces stratégies pour un système de dimensions finies permettrait d'éviter les algorithmes complexes qui doivent être exécutés chaque fois que le canal change. Ceci est vérifié dans la Figure 16.

Conclusions et Extensions possibles

Différents aspects liés à une voie de retour limitée, à la coopération et la coordination dans un réseau cellulaire ont été étudiés dans cette thèse.

Conclusions

Dans le contexte de la voie descendante d'une cellule isolée dans laquelle une station de base équipée d'antennes multiples sert plusieurs utilisateurs, deux approches différentes ont été proposées pour bénéficier d'une

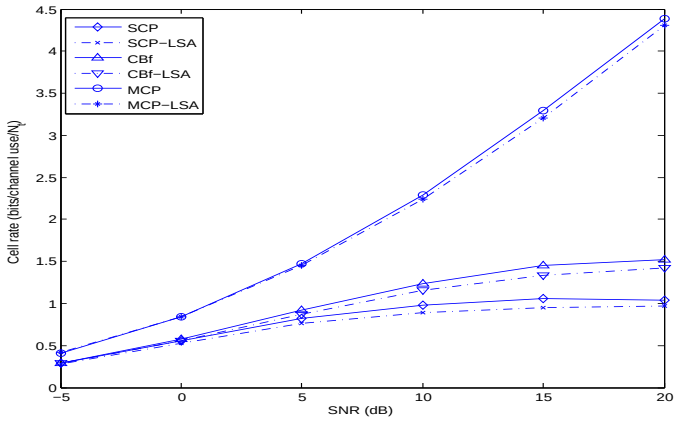


Figure 16: Comparaison pour $\epsilon = 0.5$, entre les débits moyens normalisés par cellule réalisés en appliquant les beamformers asymptotiquement optimaux pour un système de dimensions finies ($K = 3, N_t = 4$) et ceux résultant de l'analyse grands systèmes avec $\beta = 0.75$.

ressource limitée dédiée au feedback, à savoir une approche de feedback en deux étapes et une approche de feedback adaptatif. On a montré que ces dernières améliorent considérablement la performance comparée à celle d'une approche où tous les utilisateurs renvoient leur CSI avec la même précision à chaque intervalle d'ordonnancement. Pour la technique adaptative, il semble qu'une technique d'adaptation qui utilise un codebook unique mais où les utilisateurs renvoient leur CSI de manière sélective offre un bon compromis en complexité et en performance, par rapport à des techniques adaptatives presque optimales mais beaucoup plus complexes.

Dans un contexte multicellulaire, la coordination [2] pour une configuration où chaque station de base est munie d'antennes multiples et sert un seul utilisateur a été abordée. Un algorithme simple non-itératif qui ne requiert qu'une information locale des canaux a été étudiée et les conditions dans lesquelles cet algorithme a une performance proche de celle des algorithmes centralisés ou itératifs en termes de somme des débits ont été caractérisées.

En outre, la coopération [3] a été étudiée sous deux contraintes différentes.

[2]Signifiant que les émetteurs ont chacun ses propres utilisateurs, mais conçoivent leur précodage de manière à tenir compte de l'interférence qu'ils causent chez les utilisateurs dans les autres cellules.

[3]Dans le sens que les émetteurs mettent en commun leurs ressources pour servir leurs utilisateurs.

La première contrainte consiste en une connaissance distribuée des canaux: les émetteurs coopérant ont accès aux données de l'utilisateur, mais fondent leurs décisions de précodage sur des informations différentes à propos des canaux; dans ce contexte, la notion de codebooks hiérarchiques de quantification de CSI a été proposée comme moyen d'ordonner les connaissances des différents émetteurs du canal d'un utilisateur donné.

La deuxième contrainte consiste en un backhaul de capacité fixe, ce qui limite la quantité de données dédiées à chaque utilisateur qui peut être partagée par les différents émetteurs: ici, un fractionnement des taux (rate splitting) a été proposé. Cela permet une meilleure utilisation des ressources du backhaul que dans un système où les utilisateurs reçoivent leurs messages à partir d'une station de base unique (canal à interférence) ou d'une configuration network MIMO dans laquelle les messages doivent être envoyés à tous les émetteurs. Une alternative fondée sur une conception centralisée du signal à transmettre suivies de l'utilisation du backhaul pour transmettre à chaque station de base une version quantifiée du signal qu'elle doit transmettre a également été étudiée. Selon l'état des canaux, cette alternative peut résulter en une meilleure performance. Son applicabilité dépendra de la structure du réseau.

Enfin, trois différents niveaux de coopération ont été étudiés pour une configuration de deux cellules où chaque station de base a plusieurs antennes et sert simultanément plusieurs utilisateurs dans sa 'zone de couverture'. Dans une première configuration non-coopérative, chaque station de base sert égoïstement son propre ensemble d'utilisateurs, en ignorant les interférences qu'elle génère à l'extérieur de sa cellule. La deuxième configuration est une configuration coordonnée, où chaque station de base ne sert toujours que son propre ensemble d'utilisateurs, sauf que le précodage est conçu conjointement afin que les interférences inter-cellulaires soient prises en considération. Enfin, une configuration totalement coopérative est considérée où les données des utilisateurs sont acheminées vers les deux stations de base qui mettent en commun leurs antennes pour servir les utilisateurs ensemble. Pour un modèle de canal simple, les résultats d'une analyse grands systèmes permettent de caractériser la performance des différents systèmes, ce qui simplifie leur comparaison. Une telle approche permet également de déterminer la charge optimale des cellules, à savoir le nombre d'utilisateurs à servir.

Extensions Possibles

Le travail de recherche effectué au cours de cette thèse peut être étendu de plusieurs manières:

- L'étude relative au feedback limité peut être dévelopée en tenant en compte des modèles plus réalistes des canaux.
- Le feedback limité dans le cas multicellulaire mérite plus d'étude. La majorité du travail de recherche dans cette thèse dans le cadre d'un réseau multicellulaire présupposait un CSIT soit complet, soit local mais parfait.
- Pour la configuration de CSIT distribué, il faudrait quantifier si l'épargne en termes de ressources dédiées à la signalisation est telle que la perte en performance en vaut la peine.
- Les résultats des Chapitres 6 et 7 pourrait être étendus au-delà de deux cellules. La question relative à l'acheminement des données des utilisateurs dans un grand réseau où plusieurs stations de base coopèrent pour servir les utilisateurs mais où le backhaul est limité nécéssite un cadre de traitement plus général que celui présenté ici.
- Les résultats du Chapitre 7 peuvent également être étendus à des modèles de canal plus généraux. En outre, notre approche suppose des canaux identiquement distribués et ne tient pas compte de l'ordonnancement au niveau de chaque station de base. Traiter ceci serait aussi d'un grand intérêt.

Chapter 1

Introduction

1.1 Background and Motivation

Future and currently standardized wireless communications systems aim to meet the ever increasing demand for high data rate communication services, as well as to make performance fairer across the coverage area in the sense that users at the cell edge are not too disadvantaged relative to users closer to a base station. To achieve these targets, multiple antennas at the transmit and/or receive side are certain to play a significant role, thanks to the remarkable capacity enhancements with respect to the single-antenna case that these enable. In cellular systems, where the radio spectrum is scarce, multi-antenna approaches promise to increase spectral efficiency and improve the quality of service (QoS) for users. However, whether in the case of a single isolated cell or in that of a multicell system, realistic issues related to how much channel state information (CSI) is available at the transmitters may limit the gains due to multiple input multiple output (MIMO) techniques. For multicell setups, the concepts of coordination and cooperation among interfering links (cells) have emerged in an effort to cope with, or even exploit, interference. Unfortunately both these typically require information exchange between the cooperating entities, which consumes resources. An additional issue is how much and what type of data should be shared between the different base stations (BSs). For instance, if no user data is shared, schemes relying on interference avoidance may be possible,

whereas if user data is fully shared joint transmission becomes possible. As shown in the thesis, intermediate schemes may be considered depending on the system constraints.

With these issues in mind, in this thesis we study two different multi-user wireless systems, focusing on transmission in the downlink (DL) direction in both of them.

The first part of the thesis considers a single cell scenario, equivalent to a MIMO broadcast channel (BC), where the transmitter is equipped with multiple antennas whereas receivers only have a single antenna each. The potential gains of MIMO transmission techniques over conventional single-antenna transmission strategies were highlighted by the pioneering works by Foschini and Gans [16] and Telatar [17]. While most of the initial research activities in the field focused on single-user (SU) MIMO where multiple spatial channels are allocated to a single user, recently there has been a shift to multi-user configurations [1]. In addition to the *spatial multiplexing* gain enabled in single links by incorporating more antennas at both ends of the transmission, such setups introduce an additional gain termed *multi-user diversity* (MUD) gain, made possible by appropriate choice of which users are scheduled, provided the scheduling entity has enough CSI [18]. In fact, and unfortunately, unlike the SU MIMO where the benefit of CSI at the transmitter (CSIT) is marginal in general, in a MIMO BC, it is essential for achieving these gains. In a time division duplexing (TDD) system channel knowledge at the transmitter could be obtained through channel estimation in the uplink (UL), exploiting channel reciprocity. In a frequency division duplexing (FDD) system on the other hand, reciprocity no longer holds and the transmitter must rely on UL feedback (FB) from the users to obtain CSIT. The trade-off between gains and cost associated with CSI feedback guided the research carried out in the first part of this thesis in Chapter 3.

The second part of the thesis considers multicell configurations. Two different transmission approaches to dealing with interference in such a setting are considered:

- The first equates the system to a multiple input single output (MISO) interference channel (IC) and relies on BSs coordinating their transmissions so as to avoid generating too much interference at each other's users. Optimal coordination seems to require either centralized design or equivalently full CSI sharing which allows the different BSs to reach consistent transmission decisions, or alternatively, an iterative scheme, whereby each transmitter updates its design based on local information until convergence. Trying to avoid these two options without too

much performance deterioration has lead to the results in Chapter 4.

- The second, termed network MIMO, consists of several base stations jointly serving a group of users, thus resulting in a BC (albeit with different power constraints than the traditional BC); however, in practice, such a setup is significantly different than a single cell setup in that the data to all the users would need to be duplicated and routed to all cooperating base stations, which may be prohibitive. CSI may also need to be shared. The costs associated with this motivate the design of schemes which can achieve some of the network MIMO gains without the full exchange of information, including the work presented in Chapters 5 and 6.

1.2 Outline of the dissertation

In general terms, the common thread to the thesis is the issue of how best to deal with a limited resource in a multi-user multiple antenna cellular system, whether it be the resource dedicated to CSI feedback in the single cell, or the backhaul links for CSI or data sharing in the multicell case. The outline of the thesis is as follows.

Chapter 1 presents the motivation of this dissertation, giving its outline and listing its contributions. Chapter 2 gives an overview of the different setups considered in this thesis, the main working assumptions and how they relate to some of the existing literature. Chapter 3 deals with limited CSI Feedback in MU-MISO Broadcast channels: a single cell setup is considered in which a base station with N_t antennas serves a group of K users with the objective of sum rate maximization, under limited CSI feedback in the form of total per scheduling period or average FB rate. Chapter 4 deals with coordinated beamforming on the MISO IC using what we refer to as the VSINR Framework, VSINR denoting virtual signal to interference plus noise ratio (SINR). This corresponds to a multicell setup where a given user's data is transmitted from a single base station. We consider the case where each transmitter has limited CSI. Chapter 5 deals with cooperative MISO with distributed CSIT. This also corresponds to a multicell setup. However now data may reach a given user from different transmitters who jointly beamform to serve all users in the system. Here, limited CSI at each transmitter and full data sharing is assumed. The novelty of the adopted approach lies in the notion of distributed CSI, so that the knowledge at a given transmitter may not be the same as that at others. Chapter 6 treats multicell MISO cooperation when a noiseless, limited-capacity backhaul dedicated to

routing the user data to the transmitters is taken into consideration. Here, full CSIT is assumed to be shared by all transmitters, but the backhaul is constrained in such a way that it may not be optimal for the user messages to be fully shared. Chapter 7 compares different levels of data and CSI sharing in a multicell environment using random matrix theory (RMT) results for large systems. Finally, Chapter 8 concludes and suggests further related research directions.

1.3 Research contributions

1.3.1 Chapter 3

The main contributions of this chapter are two different approaches to facing a limited feedback resource constraint in a single cell multi-user MIMO (MU-MIMO) system: a two-stage FB approach and an adaptive FB idea. Results are derived for both under the assumption of Rayleigh independent identically distributed (i.i.d.) fading channels. These were presented in the following publications:

- R. Zakhour and D. Gesbert, "A two-stage approach to feedback design in MU-MIMO channels with limited channel state information", *Proceedings of IEEE 18th IEEE Annual International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC 2007)*, Athens, September 3-7, 2007.

- R. Zakhour and D. Gesbert, "Adaptive feedback rate control in MIMO broadcast systems with user scheduling", *Proceedings of IEEE Information Theory and Applications Workshop (ITA 2008)*, San Diego, January 28-30, 2008.

- R. Zakhour and D. Gesbert,"Adaptive feedback rate control in MIMO broadcast systems", *Proceedings of IEEE Information Theory Workshop (ITW 2008)*, Porto, May 5-9, 2008.

1.3.2 Chapter 4

In the context of the MISO IC, we propose a simple one-shot algorithm for coordinated beamforming based on local CSI at each transmitter based on what we term VSINR maximization. One of the contributions of this chapter consists in showing how the result of such an algorithm relates to the Pareto boundary of the MISO IC. In the case of two links, it is shown that the

resulting scheme yields a rate pair which belongs to the Pareto boundary. The sum-rate performance of such a scheme is shown through numerical simulations. Numerical simulations also illustrate the power efficiency of such an approach for more than two cells.

The following are the related publications:

- R. Zakhour and D. Gesbert, "Coordination on the MISO interference channel using the virtual SINR framework", *Proceedings of the International ITG Workshop on Smart Antennas (WSA'09)*, Berlin, February 16-18, 2009.
- R. Zakhour, K. M. Ho and D. Gesbert, "Distributed beamforming coordination in multicell MIMO channels", *Proceedings of IEEE 69th Vehicular Technology Conference (VTC 2009-Spring)*, Barcelona, April 26-29, 2009.

1.3.3 Chapter 5

In this chapter, we consider the network MIMO setup and introduce the concept of distributed CSIT for precoding over this network model. We relate the problem of distributed beamforming design for joint data transmission in multicell scenarios to team-decision problems in which team members wish to maximize a common utility but do not share the same information upon which to base their decisions [19]. The contributions of this chapter are

- A heuristic approach for precoding when each base station is equipped with multiple antennas, based on an extension of the work done in Chapter 4.
- A more systematic and general approach to precoding with distributed CSIT based on team decision problems.
- A feedback design strategy, which establishes a hierarchy of the CSI at the different cooperating base stations and simplifies distributed precoding design.

Part of the work in this chapter appears in:

- R. Zakhour and D. Gesbert, "Team decision for the cooperative MIMO channel with imperfect CSIT sharing", *Proceedings of IEEE Information Theory and Applications Workshop (ITA 2010)*, San Diego, January 31-February 5, 2010.

- R. Zakhour and D. Gesbert, "Distributed multicell-MIMO precoding using the layered virtual SINR framework", accepted for publication in *Transactions on Wireless Communications*.

1.3.4 Chapter 6

This chapter deals with network MIMO with a capacity-limited backhaul. Specifically, we propose to split the data rate intended for each user into a set of subrates, each known at a subset of the transmitters, resulting in a setup which is a hybrid between an interference channel and a network MIMO channel where all transmitters share all of the user messages. An achievable rate region is characterized for the two transmitter two receiver case. Moreover, we adapt a 'quantize-and-forward' type of scheme, initially proposed in [20] for dirty-paper coding (DPC) based cooperation in a Wyner model network, to our setup and compare it to our proposal.

Part of this work was published in

- R. Zakhour and D. Gesbert, "On the Value of Data Sharing in Constrained-Backhaul Network MIMO", *Proceedings of 2010 International Zurich Seminar on Communications (IZS 2010)*, Zurich, March 3-5, 2010.

1.3.5 Chapter 7

The work presented in this Chapter was initiated while the author was visiting Prof. Hanly's group at the University of Melbourne. Here, different levels of cooperation are compared in a two-cell setup. Specifically,

- in a first non-cooperative scheme, each base station independently maximizes the minimum rate of its worst user,
- in a second scheme, the base stations jointly coordinate their beamforming, also with the objective of maximizing the minimum rate of the worst case user,
- in a final fully cooperative scheme, the base stations pool their antenna together to jointly serve their users, with the same objective as above.

For a symmetric channel model, we find closed form solutions for the optimal beamforming strategies to use, and the optimal achievable rates in the above problems, for the large system case where we let the number of antennas at each transmitter, N_t, and the number of users in each cell, K, grow

unbounded such that their ratio $\frac{K}{N_t}$ tends to a constant β. Interestingly, the optimal beamforming strategies may be related to the regularized zero-forcing strategy. The closed form expressions allow for the cell load (i.e. how many users to serve jointly) to be optimized, for the different cooperative strategies used.

The work carried out in this Chapter was compiled into:

- R. Zakhour and S. V. Hanly, "Base station cooperation on the downlink: Large system analysis", submitted to *Transactions on Information Theory*, June 2010, arXiv:1006.3360.

Chapter 2

Multi-user Multicell MIMO Systems

This chapter presents the system assumptions adopted throughout this thesis, along with a brief discussion with some of these. The different Multi-user (MU) MIMO cellular setups considered are described, along with the related issues studied.

2.1 System Assumptions

2.1.1 Channel Model and Assumptions

Only narrowband (flat-fading) channels will be dealt with in this thesis. The proposed methods could be applied on a per subcarrier basis in OFDM systems, assuming separate power constraints on each.

The wireless channel is in general time-varying. We assume that it will however remain more or less the same for a long enough duration (i.e. the coherence time is long enough relative to the symbol duration) for the schemes we consider to be feasible. A common model is the block fading channel model, which remains constant for a duration of T symbols, then changes independently of its current value. This is the model adopted in this thesis.

The variation in the channel is due to several factors that occur on different time scales, as detailed below.

Single Input Single Output (SISO) channel model

Let h denote a channel coefficient between a given transmit-receive antenna pair. It can be modeled as

$$h = \sqrt{PL}\sqrt{\varsigma}\bar{h}, \tag{2.1}$$

where

- $PL \in \mathbb{R}_+$ stands for the path loss, the attenuation of the signal caused by distance between transmitter and receiver. Pathloss is in general modeled as

  $$PL = \beta d^{-\mu}, \tag{2.2}$$

  where μ denotes the path-loss exponent, β the path-loss constant and d the distance from transmitter to receiver, expressed in suitable units.

- $\varsigma \in \mathbb{R}_+$ is the large-scale fading (shadowing) component due to the impediment caused by buildings and large obstacles. It is normally modeled as log-normally distributed random variable.

- $\bar{h} \in \mathbb{C}$ is the fast-fading (multipath, small-scale) component. This is due to the presence of scatterers around the transmitters and receivers. A common model when there is no line of sight component to the transmission is the Rayleigh fading model, whereby

  $$\bar{h} \sim \mathcal{CN}(0, 1). \tag{2.3}$$

  This is the small-scale fading model we adopt throughout this thesis. Moreover, in our analysis in the single-cell case, channels are assumed to be i.i.d. Rayleigh fading, and the effects of path-loss and large-scaling are ignored. This is not the case in the multicell setup.

Multiple Input Single Output (MISO) channel model

When multiple antennas are considered at the transmitter and/or receiver, the channel coefficients may be spatially correlated, depending on the antenna configuration. Correlation will be ignored throughout this thesis, though taking this into consideration will have an effect on both performance (capacity, bit error rates, etc.) and feedback requirements. Thus letting $\mathbf{h}_{k,j} \in \mathbb{C}^{1\times N_t}$ denote the channel coefficients vector from transmitter j to user k, it will be modeled as

$$\mathbf{h}_{k,j} = \sqrt{PL_{k,j}}\sqrt{\varsigma_{k,j}}\bar{\mathbf{h}}_{k,j}, \tag{2.4}$$

where $PL_{k,j}$ and $\varsigma_{k,j}$ are that user's path-loss and large-scale fading components, and $\bar{\mathbf{h}}_{k,j} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_t})$. When no path-loss or shadowing is considered, as in Chapter 3, $\mathbf{h}_{k,j} = \bar{\mathbf{h}}_{k,j}$.

2.1.2 Receiver Model

Throughout the thesis we focus on the single-antenna receiver case. Moreover, we assume perfect channel state information at the receiver (CSIR), i.e. users are assumed to be able to estimate their channels perfectly. CSIR can be obtained relatively efficiently from pilot symbols in DL channels, as the terminals share a common pilot channel. For the multicell case where data is received from a single base station, users need not estimate the channel to the non-serving base stations but should be able to measure the interference power generated. When data is jointly transmitted from multiple base stations, users need to estimate the channel to all the base stations, so some coordination of the training needs to be done.

Moreover, we assume single user decoding (SUD), in the sense that each user decodes its signal while treating other users' transmissions as noise. The rates achieved will depend on the transmission scheme used, which will be specified later.

2.1.3 Ideal link adaptation

Ideal link adaptation is assumed and the continuous-rate, continuous-power Shannon capacity formula is calculated as user throughput measure. As the focus is not on the receiver side processing, we consider that interference due to messages intended for other users is treated as noise. Since full CSIT is never assumed, these rates cannot be known at the transmitter(s) and some feedback from the receivers (possibly via Automatic Repeat Request(ARQ)) is needed to adapt the rate accordingly. We will not take this feedback into consideration, and therefore the throughput results we show are effectively upper bounds on the actual rates.

2.1.4 Infinite backlogged users

The queues for each user's data are assumed to never be empty. Thus, if a user is scheduled, the base station will have data to transmit to it. As the focus has essentially been on throughput maximization or on maintaining fairness by maximizing the minimum achievable rate in the network, queue state information and traffic arrival processes are neglected.

2.2 MU-MIMO Channels

Consider the discrete-time complex baseband MIMO channel with N transmitters, each equipped with N_t antennas, and K single-antenna receivers[1]. This is defined by

$$\mathbf{y}(i) = \mathbf{H}\mathbf{x}(i) + \mathbf{n}(i), \quad i = 1, \ldots, n \tag{2.5}$$

where $\mathbf{x}(i) \in \mathbb{C}^{NN_t}$ is the transmitted vector at time i, $\mathbf{y}(i), \mathbf{z}(i) \in \mathbb{C}^K$ are the corresponding received and noise vectors, and $\mathbf{H} \in \mathbf{C}^{K \times NN_t}$ is the full channel matrix

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_{1,1} & \ldots & \mathbf{h}_{1,N} \\ & \ddots & \\ \mathbf{h}_{K,1} & \vdots & \mathbf{h}_{K,N} \end{bmatrix}, \tag{2.6}$$

$\mathbf{h}_{k,j} \in \mathbb{C}^{1 \times N_t}$ being the channel coefficient vector from transmitter j to user k, as introduced in Section 2.1.1. In the latter, $\mathbf{h}_k$ will be used to denote a given user's full (from all base stations) channel. Moreover, when we want to emphasize that a user is associated with a given cell, users will be instead indexed by two indices (user and cell) instead of just a single index (user only). The noise vector sequence $\{\mathbf{n}(i)\}$ is i.i.d. with components $n_k(i) \sim \mathcal{CN}(0, \sigma^2)$ for $k = 1, \ldots, K$ and $i = 1, \ldots, n$.

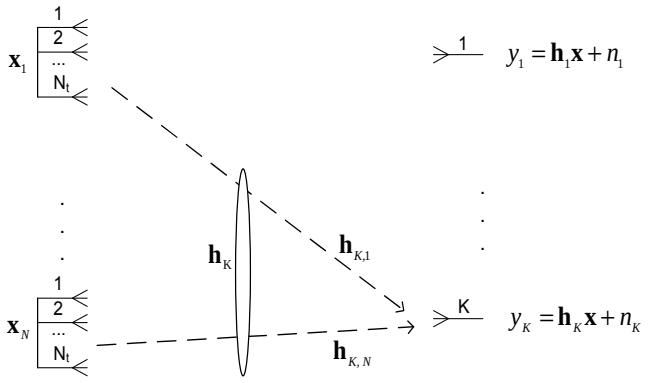


Figure 2.1: General MU-MIMO channel.

Figure 2.1 shows model (2.5), dropping the time index dependence. Such a model describes several situations of interest in data communications in

[1]We are adopting an exposition along the lines of the introduction in [21].

general, depending on the constraints put on the transmitters and the receivers and on the assumptions about the channel matrix. We describe some of these and see how each appears in the context of this dissertation.

2.2.1 MIMO Gaussian BC: Single cell or Multicell with full network MIMO

If all the transmitters in the above model (2.5) are allowed to cooperate and the receivers are constrained to decode their signals independently, a vector Gaussian BC is obtained. Such a setup arises in the downlink of a wireless system where the base station is equipped with an antenna array. In the multicell cooperation context, it represents the extreme case where all base stations pool their antennas together to serve their users jointly, the latter each decode their intended received signals independently. These two cases are represented in Figures 2.2 and 2.3, respectively.

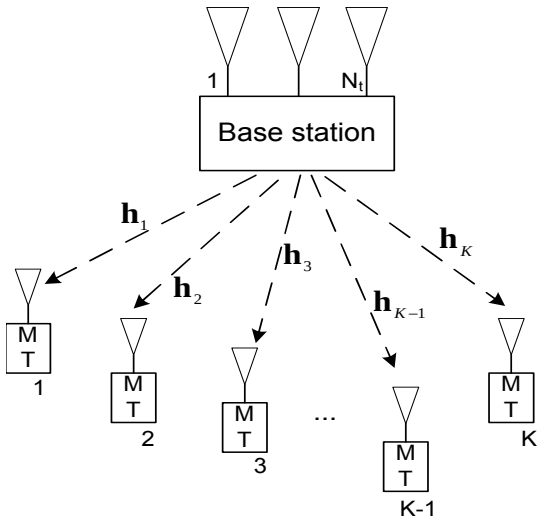


Figure 2.2: MIMO Broadcast Channel, Single Cell Setup with N_t antennas at the transmitter and K single antenna receivers. $\mathbf{h}_k \in \mathbb{C}^{1\times N_t}$ denotes the channel coefficients vector of the DL channel from the base station to user k, $k = 1, \ldots, K$.

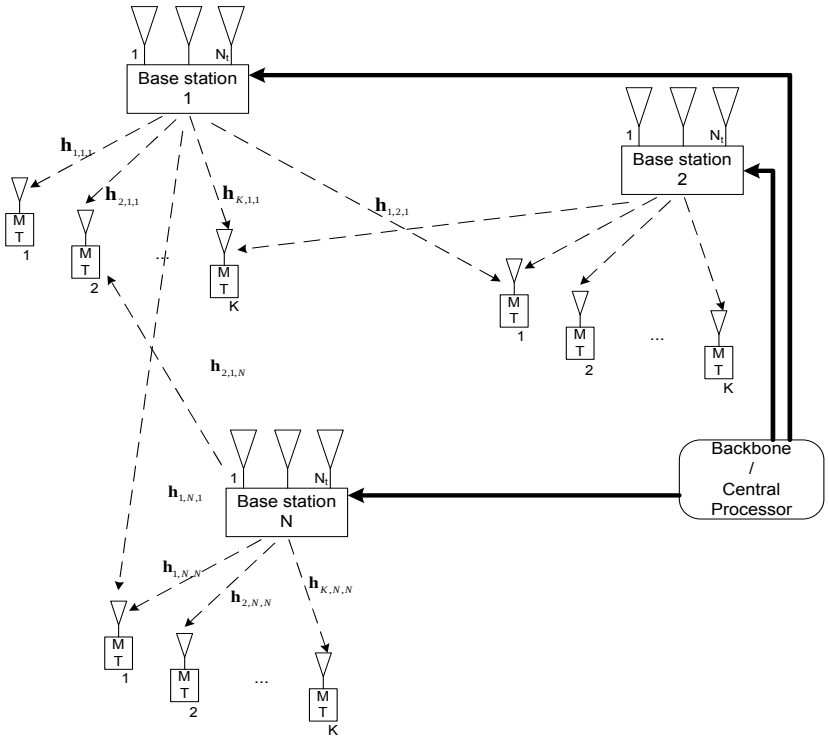


Figure 2.3: MIMO Broadcast Channel, Multicell Setup with N_t antennas at each base station and K single antenna receivers in each cell. $\mathbf{h}_{k,j,\bar{j}} \in \mathbb{C}^{1\times N_t}$ denotes the channel coefficients vector of the DL channel from base station $\bar{j}$ to user k in cell j, $k = 1,\ldots,K$, $j,\bar{j} = 1,\ldots,N$.

Importance of Channel State Information

In general, integration of multiple antennas at the transmitters and receivers in a communications system leads to enhanced capacity through spatial diversity and multiplexing gains. In multi-user configurations, a further capacity increase due to MUD is attainable through judicious scheduling [18]. This is however only possible if enough is known at the transmitter about the receivers. In fact, in the single-user case[2], CSIT contributes little to achieving the multiplexing gain of N_t:at high SNR, equal power allocation over the transmit antennas tends to be optimal and therefore CSIT is no

[2]Assuming the receiver has at least as many antennas as the transmitter.

longer necessary [3]. For MU-MIMO, on the other hand, CSIT is crucial to guarantee the full multiplexing gain, especially in the case of single-antenna users considered here. This has lead to a multitude of works dealing with limited CSI feedback, some of which we review in Chapter 3, along with this thesis' contributions to dealing with constrained feedback resources.

For the case of multicell cooperation, more questions related to CSIT come up. How is it obtained? How is it shared? How is the related transmission design done? These are some of the questions that motivate the work in Chapter 5.

Linear precoding schemes

Dirty paper coding (DPC) has been shown to be the capacity achieving transmission scheme [22] for the Gaussian MIMO BC. However, due to the complexity of such a scheme, we choose, in this thesis, to focus on linear precoding based schemes. Thus if transmitter j is serving a group $\mathcal{S}$ of users, its transmit signal may be written as

$$\mathbf{x}_j = \mathbf{W}_{j,\mathcal{S}}\mathbf{s}_{\mathcal{S}}, \tag{2.7}$$

where $\mathbf{W}_{j,\mathcal{S}} \in \mathbb{C}^{N_t \times |\mathcal{S}|}$ is the precoding matrix applied to the user symbols vector, $\mathbf{s}_{\mathcal{S}} \in \mathbf{C}^{|\mathcal{S}|}$, which consists of i.i.d. $\mathcal{CN}(0,1)$: thus, we assume Gaussian signaling throughout this thesis. In general, transmitter j's signal is assumed to be subject to an average power constraint P_j such that

$$\begin{aligned}\mathbb{E}\left[\mathbf{x}_j^H\mathbf{x}_j\right] &= \mathbb{E}\left[\mathbf{s}_{\mathcal{S}}^H\mathbf{W}_{j,\mathcal{S}}^H\mathbf{W}_{j,\mathcal{S}}\mathbf{s}_{\mathcal{S}}\right] \\ &= \mathbb{E}\mathrm{Tr}\left[\mathbf{s}_{\mathcal{S}}\mathbf{s}_{\mathcal{S}}^H\mathbf{W}_{j,\mathcal{S}}^H\mathbf{W}_{j,\mathcal{S}}\right] = \mathrm{Tr}\left[\mathbf{W}_{j,\mathcal{S}}^H\mathbf{W}_{j,\mathcal{S}}\right] \leq P_j. \end{aligned} \tag{2.8}$$

The design of the precoding matrix, and therefore the achievable rates at the different users, will depend on the CSIT. It also depends on the system objective. For example, in a single cell setup, given full CSIT, one could, in increasing order of complexity, perform

- simple zero-forcing (ZF) with equal power allocation, whereby up to N_t users are served on interference-free links obtained by orthogonalizing their transmissions and allocating each user equal power. Such a scheme was shown to achieve optimal performance, in terms of sum rate scaling, for large K in a single cell setup with users having i.i.d. Rayleigh fading channels [8].

[3] For i.i.d. circularly symmetric complex Gaussian channel coefficients

- a weighted sum rate maximizing zero-forcing whereby the power allocation is optimized to maximize a weighted sum rate, given a zero-forcing precoding setup.
- a weighted sum rate maximization, whereby not only the power allocation but the whole precoding matrix is optimized.

A multitude of works has dealt with optimizing various objective functions, too many to survey in the context of this thesis. Some will be mentioned when appropriate.

2.2.2 MISO IC: Multicell Setup

If both the transmitters and the receivers are not allowed to transmit cooperatively and transmitters and receivers form disjoint pairs ($N = K$), (2.5) represents a Gaussian interference channel. In a multicell setup, this corresponds to the case where each base station serves a single user. This is represented in Figure 2.4.

When $N_t > 1$, beamforming may be applied at each transmitter to its user's signal: in [23], for example, a beamforming scheme to ensure a set of Signal to Interference plus Noise Ratios (SINRs) at the different users is described. In general, this relies on the fundamental concept of uplink-downlink duality, and requires a series of iterations.

The different base stations make beamforming decisions that affect not only their own users, but also those in other cells, generating inteference. What sort of strategies can they use? If we limit the exchange of CSIT knowledge at each transmitter, how is performance affected? These are some of the questions we tackle in Chapter 4.

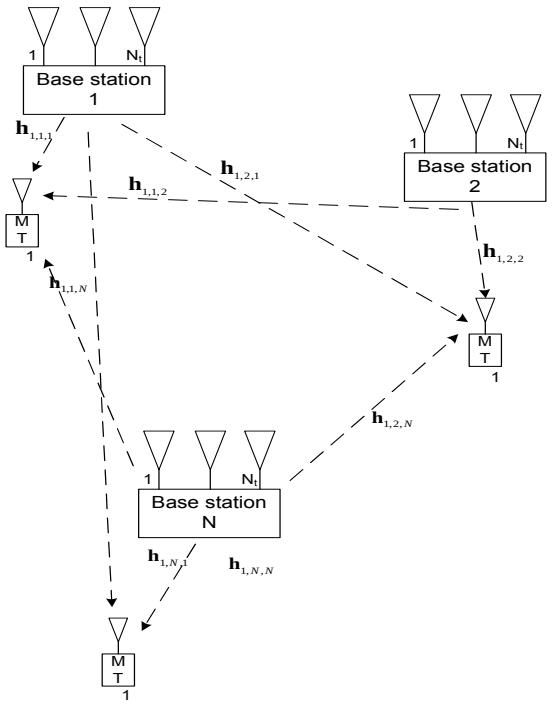


Figure 2.4: MISO IC, Multicell Setup with N_t antennas at each BS and K single antenna receivers in each cell. $\mathbf{h}_{k,j,\bar{j}} \in \mathbb{C}^{1\times N_t}$ denotes the channel coefficients vector of the DL channel from BS $\bar{j}$ to user k in cell j, $k = 1,\ldots,K$, $j,\bar{j} = 1,\ldots,N$.

Chapter 3

Limited CSI Feedback in Single Cell MISO Broadcast channels

3.1 Introduction

In this chapter, the problem of limited feedback in a downlink MU-MIMO system is treated and two different ways to use an available fixed average feedback rate are proposed.

The importance of CSIT in a MIMO BC was highlighted in Chapter 2. However, acquiring CSIT consumes system resources (especially in an FDD system, where reciprocity cannot be exploited, and particularly so in wideband OFDMA systems where the amount of feedback scales with the total bandwidth). A lot of recent research has focused on systems with partial CSIT (limited feedback) so as to circumvent this problem (see [1] and references therein, and the special issue [2], in particular [24]).

Most approaches to MU-MIMO transmission under partial CSI have centered on linear precoding, which is much simpler than nonlinear processing (required for implementing the optimal DPC scheme, for example), this simplicity coming at the cost of tolerable performance loss, especially in the large number of users case [8]. Limited feedback approaches may be categorized according to what type of linear precoding is done: typically ei-

ther orthogonal random beamforming (ORBF) or zero-forcing beamforming (ZFBF) are assumed.

ORBF, based on opportunistic beamforming ideas of [25] was introduced in [26]: an orthogonal set of beams is randomly generated and the users with the best SINR on each of the beams are selected for transmission. The feedback in this case thus consists, for a given user, of the index of its best beam and the possibly quantized SINR on that beam, instead of its entire channel information coefficients vector. This scheme achieves the same rate scaling with the number of users as the DPC approach but performs badly for not so large number of users in the system, and several publications have been devoted to improving RBF for cases where the number of users is finite ([6, 27] for example), as well as to analyzing the effect of assigning a finite number of bits for quantizing the SINR information ([28] for example).

On the other hand, when ZFBF is used, the fed-back CSI is used to design a ZF precoding channel matrix: when perfect CSI is available, such a precoding matrix eliminates inter-user interference. For such a scheme, the same capacity scaling as the optimal one, in terms of multiplexing and MUD gains, is possible *provided the feedback rate is linearly scaled with SNR in dB* [4, 5, 29]. Here the CSI normally consists of a quantized version of the channel direction, along with a scalar channel quality indicator (CQI), often an estimate of the SINR.

One can also categorize limited feedback approaches according to whether *all* [4–6] or *only an appropriate subset* [30–32] of the users feed back their CSI. In the latter case, for a given maximum feedback rate, criteria are established so that the number of users that feed back their quantized channel is kept limited. In both cases, when feedback is done, a fixed-size codebook is resorted to for the CSI quantization.

3.1.1 Feedback for Scheduling vs. Beamforming

The work carried out as part of this thesis on the topic of limited feedback in the MU-MIMO downlink is based on the following observations:

- While optimal scheduling may require information about all $K >> N_t$ users in the system, the final precoding requires CSIT of at most N_t of those users, since only a subset of the users ends up being scheduled at any given scheduling period.

- While the final precoder design relies on accurate channel information to allow for a fine spatial separation of the selected users (so as to

maintain spatial multiplexing gain), the scheduling algorithm might get away with a poorer representation of the channel.

While in most of the approaches considered prior to the work presented here on this topic, the feedback resource providing CSIT was used in a single shot for both purposes of user selection (or scheduling) *and* precoding matrix design to serve the selected users, this thesis has considered exploiting the above remarks in two ways:

1. A first approach proposes that the allotted feedback bits be split among two tasks (scheduling followed by precoder design) so as to maximize the attainable sum rate under a fixed total feedback constraint. An analysis is presented that leads to an algorithm for choosing the feedback split factor. The value of the proposed ideas is further confirmed by simulations. Note that an approach based on two stages was also presented in [27], where in the first stage users feed back their CSI assuming an ORBF scheme, followed by another stage for the scheduled users. Even though the approach does reduce the feedback, most of the fed back values are unquantized real variables.

2. A second approach considers a single *adaptive* stage of feedback, whereby users feed back the CSI with different degrees of accuracy depending on their current channel state. They do this so as to meet an average feedback rate constraint. Thus, whenever a user consider it is less likely to be scheduled or that when scheduled it expects its rate not to be too high, it will feed back less accurately then when this is not the case: the feedback design should take into account the specificity of the scheduler and not only that of the beamforming algorithm. This adaptation process follows the basic intuition that a user ought to spend more feedback resources when the expected return in terms of downlink rate is larger, akin to e.g. power control in fading channels where more power is invested when the channel is better. This approach is shown to include as a special case schemes where only an appropriate set of users feeds back their CSI. In fact, our simulations show that this scheme, for the simulated setups, results in little loss compared to the more flexible general scheme proposed.

The remainder of this chapter is organized as follows. A brief summary of the system model is given in Section 3.2; recall that assumptions which are common to the whole thesis were detailed in the previous chapter. Section 3.3 is dedicated to detailing our first two-stage approach; section 3.5

deals with our second adaptive single stage approach. Numerical results are shown in Section 3.6. Finally, Section 3.7 concludes the contributions of this chapter and gives directions for future related research.

3.2 System Model

We consider a MU-MIMO channel, where a transmitter equipped with N_t antennas communicates with $K \geq N_t$ single-antenna receivers: this is shown in Figure 2.2. The received signal at user k, denoted $y_k \in \mathbb{C}$ can be written as:

$$y_k = \mathbf{h}_k \mathbf{x} + n_k \tag{3.1}$$

where $\mathbf{x} \in \mathbb{C}^{N_t \times 1}$ is the transmitted signal vector, $\mathbf{h}_k \in \mathbb{C}^{1 \times N_t}$ and $n_k \in \mathbb{C}$ represent the channel vector and the noise at the kth user, respectively. Both the channel vector entries and the noise are assumed to be i.i.d. $\mathcal{CN}(0,1)$. $\mathbf{x}$ is subjected to an average power constraint P such that $\mathbb{E}[\mathbf{x}^{\mathrm{H}}\mathbf{x}] = P$.

3.2.1 CSI Assumptions and Acquisition

When feeding back their CSI, we ignore the cost of learning the UL channel. This could be assumed to be obtained almost for free, if each user is anyhow sending information on the UL during each frame and the same resource (channel) is used for both the UL traffic and the feedback information. Of course the more of the resource is assigned to feedback, the less UL traffic will be possible.

The exact way in which feedback is done (CSI quantization) depends on the exact scheme considered, and we postpone the related discussion to the corresponding sections below.

3.2.2 Scheduling and Beamforming

As the scheduling scheme and exact precoding depend again on the way the feedback is carried out, these are defined more precisely in the related sections below. In both approaches, a ZF precoding matrix will be adopted but the corresponding power allocation differs between the two.

Note that whether for scheduling or precoding, the quantized channels will be treated as if they were accurate, i.e. no attempt is made to take more robust decisions.

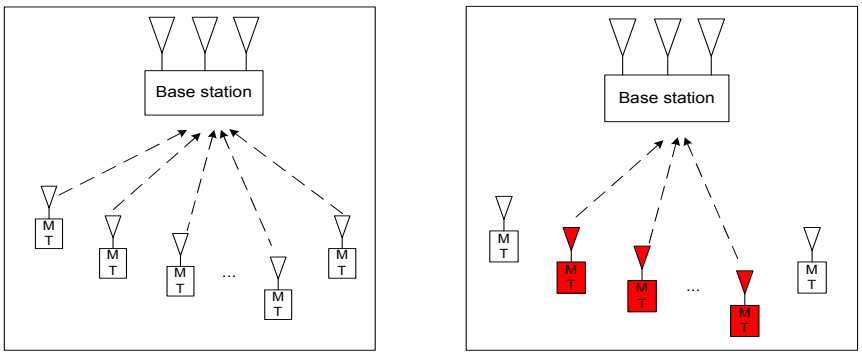
Thus, the transmit signal can be represented as:

$$\mathbf{x} = \mathbf{W}_{ZF}\mathbf{s}, \tag{3.2}$$

where $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ is the vector of independent complex Gaussian transmit symbols dedicated to the $N_s \leq N_t \leq K$ scheduled users, such that $\mathbb{E}[\mathbf{s}^{\mathrm{H}}\mathbf{s}] = \mathrm{I}_{N_s}$. In general, we will assume that $N_s = N_t$. $\mathbf{W}_{ZF}$ is such that, under full CSIT, interference among the scheduled users is eliminated. The average power constraint at the base station implies that:

$$\mathrm{Tr}\left[\mathbf{W}_{ZF}\mathbf{W}_{ZF}^{H}\right] \leq P. \tag{3.3}$$

3.3 MU-MIMO with 2-Stage Feedback



(a) All K users feed back some CSI

(b) The scheduled N_t users feed back further CSI.

Figure 3.1: Two stage scheme overview.

In this approach, we assume a fixed resource allocated for feedback at each scheduling interval. This feedback resource is split between two stages, as shown in Figure 3.1: a first stage, as in figure 3.1(a), in which all users feed back some CSI and a result of which a group of users will be scheduled, and a second stage, as in figure 3.1(b), in which those scheduled users will feed back a refinement on their CSI and a result of which is the design of the precoding matrix to be used for transmission to those users.

We represent the fixed resource by an average feedback rate in bits, which we denote B_{total} . Letting $\alpha \in [0, 1]$ denote the splitting factor between those two stages, $B_1 = \alpha B_{\text{total}}$ and $B_2 = (1 - \alpha)B_{\text{total}}$ bits will be dedicated to the scheduling and precoding matrix design stages, respectively. In each stage, all users are allocated the same amount of feedback, thus each user is allocated B_1/K bits to quantize his channel in the first stage, and each

selected user gets B_2/N_t bits to quantize the refinement in the second stage. This is summarized in Figure 3.2.

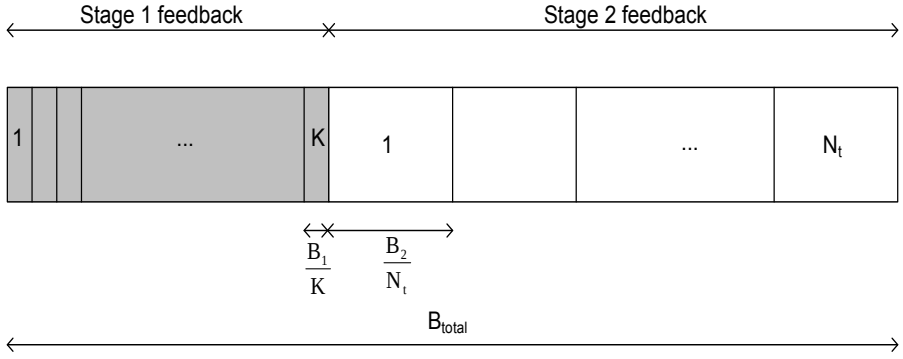


Figure 3.2: Feedback rate split for the proposed two-stage scheme.

3.3.1 Scheduling and Precoding Matrix

Full CSIT case

Before describing the scheduling and the beamforming design in the proposed two stage limited feedback setup, we summarize these for the full CSIT case. For simplicity, we assume that N_t users are scheduled. Denote by $\mathcal{A} \subset \{1, \ldots, K\}$ that set ($|\mathcal{A}| = N_t$), and let $\mathbf{H}_{\mathcal{A}} \in \mathbb{C}^{N_t \times N_t} \triangleq [\mathbf{h}_{A_1}^{\mathrm{T}} \ldots \mathbf{h}_{A_{N_t}}^{\mathrm{T}}]^{\mathrm{T}}$ be the aggregate channel of the users in that set, , A_i being the ith entry in the set $\mathcal{A}$. The corresponding zero-forcing beamforming matrix is given by

$$\mathbf{W}_{ZF} \triangleq \sqrt{P}\mathbf{H}_{\mathcal{A}}^{\dagger} / \sqrt{\mathrm{tr}((\mathbf{H}_{\mathcal{A}}\mathbf{H}_{\mathcal{A}}^{H})^{-1})}. \tag{3.4}$$

The ZF process effectively transforms the MIMO channel into N_t parallel subchannels with equal power gain. Consequently, for each $k \in \mathcal{A}$, the received signal (3.1) becomes:

$$y_k = \sqrt{\frac{P}{\mathrm{tr}((\mathbf{H}_{\mathcal{A}}\mathbf{H}_{\mathcal{A}}^{H})^{-1})}} s_k + n_k. \tag{3.5}$$

The achievable sum rate for a given $\mathcal{A}$ is thus equal to:

$$SR_{ZF\text{-}CSIT} = N_t \log_2 \left(1 + \frac{P}{\mathrm{tr}((\mathbf{H}_{\mathcal{A}}\mathbf{H}_{\mathcal{A}}^{H})^{-1})}\right), \tag{3.6}$$

and the scheduling rule that maximizes (3.6) is:

$$\mathcal{A}_{opt} = \arg \max_{\mathcal{A}\subset\{1,\ldots,K\},|\mathcal{A}|=N_t} \frac{1}{\mathrm{tr}((\mathbf{H}_{\mathcal{A}}\mathbf{H}_{\mathcal{A}}^H)^{-1})}. \tag{3.7}$$

Two-Stage Limited Feedback

The two steps of scheduling (finding the optimal set $\mathcal{A}$) and precoding matrix design are mapped into two feedback stages. In the first stage, each of the K receivers feeds back a "coarse" quantized version of its channel vector, $\hat{\mathbf{h}}_{1,k}$, $k = 1,\ldots,K$; a group of users, denoted by $\hat{\mathcal{A}}$, is selected as in (3.7), but with the real channels replaced by their quantized versions. Thus,

$$\hat{\mathcal{A}} = \arg \max_{\mathcal{A}\subset\{1,\ldots,K\},|\mathcal{A}|=N_t} \frac{1}{\mathrm{tr}((\hat{\mathbf{H}}_{1,\mathcal{A}}\hat{\mathbf{H}}_{1,\mathcal{A}}^H)^{-1})}, \tag{3.8}$$

where similar to the definition of $\mathbf{H}_{\mathcal{A}}$, $\hat{\mathbf{H}}_{1,\mathcal{A}} \triangleq [\hat{\mathbf{h}}_{1,A_1}^{\mathrm{T}} \ldots \hat{\mathbf{h}}_{1,A_{N_t}}^{\mathrm{T}}]^{\mathrm{T}}$.

In the second stage, users in $\hat{\mathcal{A}}$ send back refinements of their channels (e.g. quantized versions of $\mathbf{h}_k - \hat{\mathbf{h}}_{1,k}$). The new channel estimates $\hat{\mathbf{h}}_{2,k}$ are used to design the ZF precoding matrix $\hat{\mathbf{W}}_{ZF}$:

$$\hat{\mathbf{W}}_{ZF} = \sqrt{P}\frac{\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}^{\dagger}}{\sqrt{\mathrm{tr}((\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}^H)^{-1})}}, \tag{3.9}$$

where $\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}} \triangleq [\hat{\mathbf{h}}_{2,\hat{A}_1}^{\mathrm{T}} \ldots \hat{\mathbf{h}}_{2,\hat{A}_{N_t}}^{\mathrm{T}}]^{\mathrm{T}}$.

Due to quantization error, interference will not be entirely eliminated by the ZF-precoder. Thus the received signal vector will be given by:

$$\mathbf{y} = \sqrt{P}\mathbf{H}_{\hat{\mathcal{A}}}\frac{\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}^{\dagger}}{\sqrt{\mathrm{tr}((\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}^H)^{-1})}}\mathbf{s} + \mathbf{n}. \tag{3.10}$$

Each user channel $\mathbf{h}_{\hat{A}_i}$ may be written as the sum of its final quantized version and the corresponding error term,

$$\mathbf{h}_{\hat{A}_i} = \hat{\mathbf{h}}_{2,\hat{A}_i} + \mathbf{e}_{2,\hat{A}_i}. \tag{3.11}$$

Letting $\mathbf{E}_{2,\hat{\mathcal{A}}} \triangleq [\mathbf{e}_{2,\hat{A}_1}^{\mathrm{T}} \ldots \mathbf{e}_{2,\hat{A}_{N_t}}^{\mathrm{T}}]^{\mathrm{T}}$, $\mathbf{H}_{\hat{\mathcal{A}}} = \hat{\mathbf{H}}_{2,\hat{\mathcal{A}}} + \mathbf{E}_{2,\hat{\mathcal{A}}}$ and (3.10) becomes:

$$\mathbf{y} = \frac{\sqrt{P}}{\sqrt{\mathrm{tr}((\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}^H)^{-1})}}\mathbf{s} + \frac{\sqrt{P}\mathbf{E}_{2,\hat{\mathcal{A}}}\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}^{\dagger}}{\sqrt{\mathrm{tr}((\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}^H)^{-1})}}\mathbf{s} + \mathbf{n}. \tag{3.12}$$

Since the second term in (3.12) corresponds to the deviation from the scaled identity matrix obtained when ZF is perfect (cf. (3.5)), we use the following performance metric to approximate the achieved sum-rate:

$$SR_{ZF\text{-}Q2} = \sum_{i=1}^{N_t} \log_2(1 + SINR_{\hat{A}_i}) \tag{3.13}$$

where

$$SINR_{\hat{A}_i} = \frac{P}{\mathrm{tr}((\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}^H)^{-1}) + P\|(\mathbf{E}_{2,\hat{\mathcal{A}}}\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}^\dagger)_i\|^2}, \tag{3.14}$$

$(.)_i$ denoting the i^{th} row of a given matrix.

3.4 Splitting factor Optimization

The performance of the above scheme will depend on how the split between the two stages is carried out. In this section, for the assumed i.i.d. Rayleigh fading model, and adopting a quantization model from rate distortion theory, we find an approximate way to solve this problem.

3.4.1 Quantization Model

For user k, entries in the channel vector $\mathbf{h}_k$ are i.i.d. $\mathcal{CN}(0,1)$ r.v.'s. To model their quantization we adopt an ideal model from rate-distortion theory [3], corresponding to the successive refinement of a Gaussian with mean squared-error as distortion measure. The resulting achievable distortions per vector entry at each stage in terms of α and B_{total}, are:

$$\sigma_{e_1}^2 = 2^{-\alpha B_{\text{total}}/(K\times N_t)} \tag{3.15}$$

$$\sigma_{e_2}^2 = 2^{-\frac{B_{\text{total}}}{N_t}(\frac{\alpha}{K}+\frac{1-\alpha}{N_t})}, \tag{3.16}$$

Furthermore, entries in the first and second stage quantized channel vectors, $\hat{\mathbf{h}}_{1,k}$ and $\hat{\mathbf{h}}_{2,k}$ respectively, are i.i.d. and related to each other and to the true CSIT by the following distributions (where $j = 1, \ldots, N_t$):

$$\hat{h}_{1,k,j} \sim \mathcal{CN}(0, 1-\sigma_{e_1}^2) \tag{3.17}$$

$$\hat{h}_{2,k,j}|\hat{h}_{1,k,j} \sim \mathcal{CN}(\hat{h}_{1,k,j}, \sigma_{e_1}^2 - \sigma_{e_2}^2) \tag{3.18}$$

$$h_{k,j}|\hat{h}_{2,k,j} \sim \mathcal{CN}(\hat{h}_{2,k,j}, \sigma_{e_2}^2). \tag{3.19}$$

Short discussion of the model

The above model is based on

1. an MMSE distortion measure for the quantization, not necessarily the optimal distortion measure [33] for MU-MIMO system, and

2. the corresponding rate distortion model, which would only hold if the number of quantized elements becomes quite large. If the number of antennas is large enough, this may be a good enough approximation. Otherwise, one could introduce a penalty factor (similar to the SNR gap introduced in the $\log(1 + SNR)$ expression for capacity when using specific modulation and coding schemes [34].) to account for the difference in the real quantization.

In Appendix 3.A, we show how a two-stage channel estimation in a TDD (and therefore reciprocal) system can be shown to follow a similar successive refinement model. The expressions for $\sigma^2_{e,1}$ and $\sigma^2_{e,2}$ may be formulated in terms of the total number of time slots allocated to feedback. Essentially the whole derivation of this chapter would hold, up to Equation (3.31), where the optimization would be done using the modified expressions from the Appendix.

Even for an FDD system, one could consider a so called analog feedback [35, 36] and a similar model may be shown to apply.

3.4.2 Extreme cases

Before tackling the optimum splitting factor and the resulting performance estimation, we analyze the extreme cases of α being either 0 or 1. Their comparison serves as justification for the adopted rate splitting approach. Both cases correspond to having a single quantization stage, but differ in the following manner:

- $\alpha = 0$: N_t randomly selected users feed back their channels to enable the design of the precoding matrix.

- $\alpha = 1$: all users feed back their channels; scheduling and precoding matrix design are done based on the quantized channel. Quantization model aside, this is equivalent to the strategy in [4] and most existing work.

To compare performance as a function of feedback rate in both schemes, we estimate the ergodic sum rate by averaging (3.13) over the quantized

channel of the selected users and the corresponding quantization error statistics (we drop the 2 from $SR_{ZF\text{–}Q2}$ since we only have a single stage).

Taking the expectation over the quantization error statistics first, we are able to bound $\overline{SR}_{ZF\text{–}Q} \triangleq \mathbb{E}_{\hat{\mathbf{H}}_{\hat{\mathcal{A}}},\mathbf{E}_{\hat{\mathcal{A}}}} SR_{ZF\text{–}Q}$ as:

$$\begin{aligned}&\mathbb{E}_{\lambda_{\min}} \log\left(1+\frac{\lambda_{\min}}{N_t(1/P+\sigma_e^2)}\right)\\ &< \frac{\overline{SR}_{ZF\text{–}Q}}{N_t} < \mathbb{E}_{\lambda_{\min}}\left[\log\left(1+P[\lambda_{\min}+\sigma_e^2]\right)\right] - \mathrm{eE}_1\left(\frac{1}{P\sigma_e^2}\right),\end{aligned} \tag{3.20}$$

where $\lambda_{\min}$ corresponds to the smallest eigenvalue of $\hat{\mathbf{H}}_{\hat{\mathcal{A}}}\hat{\mathbf{H}}_{\hat{\mathcal{A}}}^H$ and $\mathrm{eE}_1(x) \triangleq e^x \mathrm{E}_1(x)$. A sketch of the derivation of these bounds is provided in Appendix 3.B.

But $\overline{SR}_{ZF\text{–}Q}$ and its bounds will differ for the two α's since:

- $\sigma_{e,\alpha=0}^2 \leq \sigma_{e,\alpha=1}^2$ (since $K \geq N_t$)
- The statistics of the quantized channels of users in $\hat{\mathcal{A}}$ are different, because of differences in both quantization error and scheduling schemes.

First extreme: $\alpha = 0$

Since users are selected randomly, the entries of $\hat{\mathbf{H}}_{\hat{\mathcal{A}}}$ are all i.i.d. $\mathcal{CN}(0, 1-\sigma_e^2)$. Consequently (using results in [37]), bounds on $\overline{SR}_{ZF\text{–}Q}$ (cf. (3.20)) are given by

$$\begin{aligned}\mathrm{eE}_1\left(\frac{N_t^2}{Pc_0}\right) <& \frac{\overline{SR}_{ZF\text{–}Q}}{N_t}\\ &< \log(1+P\sigma_e^2) + \mathrm{eE}_1\left(\frac{N_t}{Pc_0}\right) - \mathrm{eE}_1\left(\frac{1}{P\sigma_e^2}\right)\\ &< \mathrm{eE}_1\left(\frac{N_t}{Pc_0}\right) + \gamma_{EM},\end{aligned} \tag{3.21}$$

where

$$c_0 = \frac{1-\sigma_e^2}{1+P\sigma_e^2} \tag{3.22}$$

and γ_{EM} is the Euler-Mascheroni constant.

Lemma 1. *For fixed σ_e^2, the given scheme has a multiplexing gain of 0.*

Proof. For fixed σ_e^2, $Pc_0 \to \frac{1-\sigma_e^2}{\sigma_e^2}$ as $P \to \infty$: both upper and lower bound will tend to constants confirming the lemma. □

Theorem 2. *A necessary and sufficient condition for the given scheme to maintain the multiplexing gain of N_t under quantization error is to have $\sigma_e^2 = O(1/P)$.*

Proof. This can be shown by calculating the limit defining the multiplexing gain. Note that a similar condition is obtained in [29]. □

Remark 1. *Comparing upper and lower bounds in* (3.21)*, c_0 can be identified as a scaling factor which quantifies power loss with respect to perfect channel knowledge.*

Second extreme: $\alpha = 1$

We lower bound $\lambda_{\min}$'s cumulative distribution function (cdf) by that of the maximum of $N_1 \triangleq \binom{K}{N_t}$ i.i.d. exponentially distributed r.v.s of mean $(1-\sigma_e^2)/N_t$ (effectively ignoring the dependencies between the N_1 possible sets of scheduled users). Similarly, we upper bound it by that of choosing the maximum out of $N_2 \triangleq \left\lfloor \frac{K}{N_t} \right\rfloor$ with the same distribution (only considering disjoint groups). Thus

$$\left(1 - e^{-\frac{N_t x}{1-\sigma_e^2}}\right)^{N_1} \leq F_{\lambda_{\min}}(x) \leq \left(1 - e^{-\frac{N_t x}{1-\sigma_e^2}}\right)^{N_2}. \tag{3.23}$$

Using these cdfs and applying Jensen's inequality to the upper bound (concavity of the log) (cf. (3.20)), $\overline{SR}_{ZF\text{–}Q}$ is bounded as

$$\sum_{k=1}^{N_2} (-1)^{k+1} \binom{N_2}{k} \mathrm{eE}_1\left(\frac{kN_t^2}{Pc_0}\right) < \frac{\overline{SR}_{ZF\text{–}Q}}{N_t} < \gamma_{EM} + \log\left(1 + \frac{Pc_0}{N_t} H_{N_1}\right), \tag{3.24}$$

where H_{N_1} denotes the N_1-th harmonic number. These bounds can be used to reach conclusions about the scaling and power loss with respect to the perfect CSIT case that are similar to those made when $\alpha = 0$.

3.4.3 Optimal α

Comparing both cases, in order to maintain the same approximate power loss with respect to perfect CSIT for a fixed P, $\alpha = 1$ would necessitate K/N_t times higher feedback rate than $\alpha = 0$, the rate for the former being however greater due to MUD. Fig. 3.3 shows for the same B_{total}, the achievable sum rates in both cases. At low SNR, the accuracy of the quantization for $\alpha = 1$ is sufficient to achieve some of the MUD gains. This no longer holds at

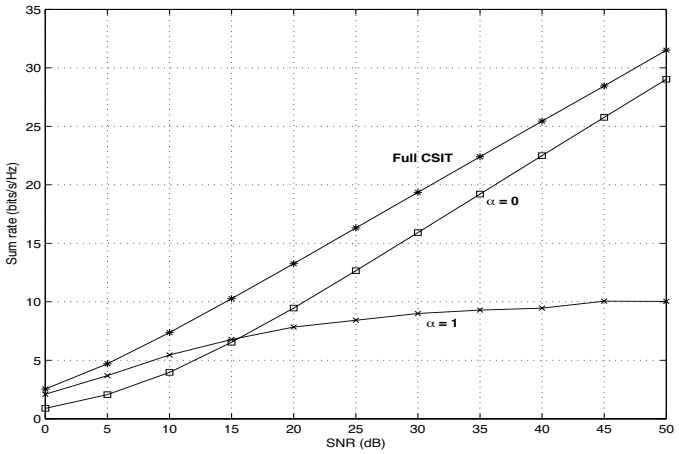


Figure 3.3: Average sum rates for $N_t = 2, K = 20, B_{\text{total}} = 80$ bits.

higher SNR, as the system becomes interference-limited. Thus the simple binary scheme of switching between $\alpha = 0$ and $\alpha = 1$ depending on N_t and K would lead to better performance than either separately. Further improvement would be expected from letting α vary within $[0, 1]$.

The optimal α, α_{opt} is defined as:

$$\alpha_{opt} = \arg\max_{\alpha} \overline{SR}_{ZF\text{–}Q2}, \tag{3.25}$$

where $\overline{SR}_{ZF\text{–}Q2} \triangleq \mathbb{E}_{\hat{\mathbf{H}}_1, \hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}, \mathbf{E}_{2,\hat{\mathcal{A}}}} [SR_{ZF\text{–}Q2}]$ (cf. (3.13)). Similarly to (3.59), this expectation can be rewritten as:

$$N_t \mathbb{E}_{\hat{\mathbf{\Lambda}}_{2,\hat{\mathcal{A}}}, \mathbf{E}_{2,\hat{\mathcal{A}}}} \log \left(1 + \frac{1}{\sum \lambda_{2,\hat{\mathcal{A}},j}^{-1} (\frac{1}{P} + |\mathbf{E}_{2,\hat{\mathcal{A}},i,j}|^2)} \right), \tag{3.26}$$

where $\hat{\mathbf{\Lambda}}_{2,\hat{\mathcal{A}}}$ is the diagonal matrix of eigenvalues of $\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}} \hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}^H$, denoted $\lambda_{2,\hat{\mathcal{A}},j}, j = 1, \ldots, N_t$, the joint distribution of which depends on that of $\hat{\mathbf{H}}_1$ and $\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}$, alternatively on $\hat{\mathbf{H}}_1$ and $\mathbf{E}_{12,\hat{\mathcal{A}}} \triangleq \hat{\mathbf{H}}_{2,\hat{\mathcal{A}}} - \hat{\mathbf{H}}_{1,\hat{\mathcal{A}}}$. Letting $\lambda_{2,\min}$ be the minimum eigenvalue of $\hat{\mathbf{H}}_{2,\hat{\mathcal{A}}} \hat{\mathbf{H}}_{2,\hat{\mathcal{A}}}^H$, we can bound $\overline{SR}_{ZF\text{–}Q2}$ (cf. (3.20))

as follows:

$$N_t \mathbb{E}_{\lambda_{2,\min}} \log\left(1 + \frac{P\lambda_{2,\min}}{N_t(1+P\sigma_{e_2}^2)}\right) < \overline{SR}_{ZF\text{-}Q2} < N_t\left[\gamma_{EM} + \mathbb{E}_{\lambda_{2,\min}} \log\left(1 + \frac{P\lambda_{2,\min}}{1+P\sigma_{e_2}^2}\right)\right]. \tag{3.27}$$

Unfortunately we are unable to bound the distribution of $\lambda_{2,\min}$. Instead we approximate the achievable sum rate by its upper bound obtained by applying Jensen's inequality, i.e. by bringing the expectation in (3.27) inside the logarithm.

Further noting that

$$\mathbb{E}_{\hat{\mathbf{H}}_{1,\hat{\mathcal{A}}},\mathbf{E}_{12,\hat{\mathcal{A}}}} \mathrm{tr}\left((\hat{\mathbf{H}}_1 + \mathbf{E}_{12})(\hat{\mathbf{H}}_1 + \mathbf{E}_{12})^H\right) = \mathbb{E}_{\hat{\mathbf{H}}_{1,\hat{\mathcal{A}}}} \mathrm{tr}\left(\hat{\mathbf{H}}_1\hat{\mathbf{H}}_1^H\right) + N_t^2(\sigma_{e_1}^2 - \sigma_{e_2}^2) \tag{3.28}$$

and recalling that the trace is the sum of the eigenvalues, we are intuitively lead to approximate the desired expectation by:

$$\mathbb{E}_{\hat{\mathbf{H}}_{1,\hat{\mathcal{A}}},\mathbf{E}_{12,\hat{\mathcal{A}}}} \lambda_{2,\min} \approx \mathbb{E}_{\hat{\mathbf{H}}_{1,\hat{\mathcal{A}}}} \lambda_{1,\min} + c(\sigma_{e_1}^2 - \sigma_{e_2}^2) \tag{3.29}$$

for some finite c.

Guided by the results of the discussion of the two extreme cases and our knowledge of the perfect CSIT case, approximating the expectation of (3.29) leads us to define the following power loss factor with respect to the ideal case:

$$f_{PL} \triangleq \frac{1-\sigma_{e_1}^2}{1+P\sigma_{e_2}^2} + \frac{\sigma_{e_1}^2 - \sigma_{e_2}^2}{\log K(1+P\sigma_{e_2}^2)}, \tag{3.30}$$

where $\log K$ is the expected MUD gain. The first term in the summation is the loss due to the first stage quantization while the second term is that caused by the second stage.

α is thus selected based on the following heuristic:

$$\alpha_{heur} = \arg\max_{\alpha} f_{PL} \tag{3.31}$$

Plugging in the values from (3.15) and (3.16), and introducing $\beta =$

$2^{-\alpha B_{\text{total}}/N_t}$ and letting $c_{PL} = 2^{-\frac{B_{\text{total}}}{N_t^2}}$:

$$\begin{aligned} f_{PL} &= \frac{1-\beta^{\frac{1}{K}}}{1+Pc_{PL}\beta^{\frac{1}{K}-\frac{1}{N_t}}} + \frac{\beta^{\frac{1}{K}} - c_{PL}\beta^{\frac{1}{K}-\frac{1}{N_t}}}{\log K\left(1+Pc_{PL}\beta^{\frac{1}{K}-\frac{1}{N_t}}\right)} \\ &= \frac{1-\left(1-\frac{1}{\log K}\right)\beta^{\frac{1}{K}}}{1+Pc_{PL}\beta^{\frac{1}{K}-\frac{1}{N_t}}} - \frac{c_{PL}\beta^{\frac{1}{K}-\frac{1}{N_t}}}{\log K\left(1+Pc_{PL}\beta^{\frac{1}{K}-\frac{1}{N_t}}\right)}. \end{aligned} \tag{3.32}$$

Taking the derivative with respect to β and setting it to 0 yields:

$$\underbrace{\frac{\left(1-\frac{1}{\log K}\right)}{K}}_{c_0>0,\text{ if } K\geq 3} + \underbrace{\frac{\left(1-\frac{1}{\log K}\right)Pc_{PL}}{N_t}}_{c_1>0,\text{ if } K\geq 3}\beta^{\frac{N_t-K}{KN_t}} + \underbrace{\left(P+\frac{1}{\log K}\right)c_{PL}\frac{N_t-K}{N_tK}}_{c_2\leq 0}\beta^{-\frac{1}{N_t}} = 0 \tag{3.33}$$

The left-hand side of the above equation can be rewritten as a polynomial, which by Descartes' rule, has a single positive root. Thus finding the optimal split in terms of our heuristic can be reduced to establishing whether that positive root corresponds to an α between 0 and 1. If it does not, the optimal α is 1.

3.5 Adaptive Feedback Rate Allocation

An alternative to the above approach is to move the feedback decision away from the BS, which in the above two-stage approach decides which of the users feeds back more information, to the users, who based on their current channel state decide whether to feed back and how accurately.

In [31], *multiuser aware* limited feedback is proposed so that the optimal sum rate scaling is obtained in the large number of users limit. The authors design codebooks for channel norm and channel direction quantization based on orthonormal vector sets for the latter. A similar approach is proposed in [32] with zero-forcing instead of orthogonal beamforming. As is quite common in limited feedback MIMO setups, the channel direction information (CDI), which for user k's channel is given by $\frac{\mathbf{h}_k}{\|\mathbf{h}_k\|}$, is quantized. The authors propose a feedback scheme whereby a user feeds back its quantized CDI if the quantization error is below a certain threshold, and the channel norm lies in a given range. In the large number of users regime, they give

bounds for the threshold values such that the number of users feeding back their CSI is on average

$$\mathbb{E}\left[N_{\rho,\sigma}\right] = 2^{r+1} N_t \log K(1 + o(1)) + O(1/K), \tag{3.34}$$

where 2^r is the size of the codebook used for CDI quantization, r can be as low as $\log_2 N_t$ for a codebook consisting of a single orthonormal basis.

Both these schemes however do not account for the signaling required for notifying the base station of feedback. Another approach based on channel-aware contention is proposed in [38]. The amount of feedback is shown to be a constant and to achieve the asymptotic sum-capacity. The feedback period of the proposed protocol consists of a series of N_t sub-periods, where in each slot users whose channel satisfies a certain slot-dependent condition compete to access the feedback channel. In the large number of user case, ignoring the downlink part of the signaling (which, at a minimum, includes broadcasting the quantized CDI of the user selected in the preceding slot), a feedback rate of about $N_t(\log_2 K + b)$ bits, where b is the size of the codebook used for feedback.

In this part of the thesis, we have investigated an alternative approach to these for the not too large number of users, whereby users *adapt* their feedback to the current channel state. As stated in the introduction, users have more incentive to send a more accurate version of their channel back to the base station when they are more likely to be scheduled and to get a good rate. Thus, we try to design a feedback scheme which adapts the feedback rate, subject to an average feedback rate, so as to maximize the expected sum rate, taking into consideration the scheduling scheme considered, the other user statistics and the current instantaneous CSI at the user in question.

3.5.1 User Selection and Precoding Scheme

The key concepts developed in this section, i.e. the benefits of a channel quality dependent feedback rate, could be developed for a different precoding scenario however, to simplify derivations, we adopt ZFBF with uniform power allocation, used in [8, 29, 39, 40] for example. Thus, for N_t users scheduled, the transmit signal $\mathbf{x}$ in (3.2) is given by:

$$\mathbf{x} = \sum_{k=1}^{N_t} \mathbf{w}_k s_k, \tag{3.35}$$

where $\|\mathbf{w}_k\|^2 = \frac{P}{N_t}$ and satisfy the zero-forcing condition, i.e. $\hat{\mathbf{H}}_{-k}\mathbf{w}_k = 0, \forall k = 1, \ldots, N_t$, $\hat{\mathbf{H}}_{-i} \in \mathbb{C}^{(K-1)\times N_t}$ is the matrix formed by concatenating the channels of all the scheduled users, except the k-th one.

The scheduling scheme tries to maximize the estimated sum rate achieved by ZFBF to a subset of users, based on the fed back CSI: in the optimal case, this would be done through exhaustive search over all groups of up to N_t users; a suboptimal scheme would use a greedy algorithm such as those of [8, 40]. The details of the scheduling algorithm depend on the CSI feedback. For our scheduling probability derivations, discussed in Appendix 3.C, we assume the semiorthogonal user selection (SUS) algorithm in [8] is used, summarized as follows: at stage i, $i \leq N_t$, users are semiorthogonal to the already selected group are eligible for selection; among those, the one with the highest norm, after projection onto the null space of already selected users' channels, is chosen. Semiorthogonality is defined as follows: two unit-norm column vectors $\mathbf{u}$ and $\mathbf{v}$ are semiorthogonal if $\mathbf{u}^H\mathbf{v} \leq \epsilon$, where $\epsilon < 1$ needs to be specified to the algorithm.

3.5.2 CSI Quantization

We assume each user feeds back (i) a quantized CDI, and (ii) a CQI [1], which allows the base station to estimate the quality of the user's channel as well as that of the quantized CDI [4, 5, 39]. As our scheme takes into consideration the quality of the CDI in the feedback decision, we will restrict the CQI to being the channel norm squared.

In what follows, $\tilde{\mathbf{h}}$ will be used to denote the true channel direction, i.e. $\tilde{\mathbf{h}} \triangleq \frac{\mathbf{h}}{\|\mathbf{h}\|}$, while $\hat{\mathbf{h}}$ will represent the quantized version of the CDI $\tilde{\mathbf{h}}$.

Codebooks for the CDI consist of a set of unit-norm N_t-dimensional vectors. Given $\hat{\mathbf{h}}$, the corresponding quantization error is defined as $\sin^2\theta$, where $\theta \triangleq \angle(\tilde{\mathbf{h}}, \hat{\mathbf{h}})$, the angle between the true and quantized channel directions. Thus $\sin^2\theta = 1 - |\tilde{\mathbf{h}}\hat{\mathbf{h}}^H|^2$. Its cumulative distribution function (cdf), for any quantizer, may be upper-bounded by [7]:

$$F_{\sin^2\theta}(x) = \begin{cases} \delta^{1-N_t} x^{N_t-1} & 0 \leq x \leq \delta \\ 1 & x > \delta \end{cases} \tag{3.36}$$

where $\delta \triangleq 2^{-b/(N_t-1)}$, b being the number of bits used for quantization. This distribution will be exploited in some of our derivations. Though it is an upper bound only, one can verify through simulations that for codebooks generated using Lloyd's algorithm [41] it does provide quite a good approximation of the true distribution.

[1] The CQI is a scalar. It is assumed unquantized in our derivations, though it will also be quantized in practice. Studies show that performance changes very little when as few as 2 bits are allocated to it.

A common CQI measure is the following SINR estimate given by [4, 5, 39]:

$$\hat{\gamma} \triangleq \frac{\frac{P}{N_t}\|\mathbf{h}\|^2 \cos^2\theta}{1 + \frac{P}{N_t}\|\mathbf{h}\|^2 \sin^2\theta} \tag{3.37}$$

This choice is simple and its accuracy in reflecting the achievable SINR under the given transmission scheme increases with the number of users in the system, as the probability of actually finding users whose quantized channels are almost orthogonal to each other increases. If such a group of users is actually found, the zero-forcing vector corresponding to a user will be matched to its quantized channel, and the interference seen would be solely due to the quantization error, as would the reduction in desired signal power, which is exactly what is expressed by the SINR model (3.37).

3.5.3 Ergodic Rates

For the system described in the previous section, the system performance measure is the expected sum rate, which may be expressed as:

$$\mathbb{E}SR \triangleq \sum_{k=1}^{K} \mathbb{E}_{\mathbf{h}_i, i\in\{1,\ldots,K\}} \log_2(1+\gamma_k) = \sum_{k=1}^{K} \mathbb{E}R_k \tag{3.38}$$

where γ_k is the actual SINR after precoding and scheduling at user k [2]. The achieved SINR will depend, among other factors, on the fedback CSI. We introduce the notion of *adaptive feedback rate* with the aim of enhancing the rates attained: users feed back their CSI with different accuracies depending on their current channel state, while maintaining an average feedback rate constraint, i.e. consuming a fixed amount of feedback resources on average. As argued in the introduction, under such a constraint, since not all users will end up being scheduled, it makes sense for those who expect to be to give more information, especially since this means that, should they effectively be scheduled, the transmitter will be able to reduce the interference they receive from other users' signals better, thereby achieving higher SINR and consequently higher $\mathbb{E}SR$. The interesting questions are (i) how should this adaptation be done so as to maximize $\mathbb{E}SR$, (ii) is there a channel quality level below which it is not worth quantizing and feeding back the channel at all?

As the feedback decisions are to be made at each user independently, i.e. (3.38) is to be maximized in a distributed fashion, and users' channels (both

[2] Note that by definition of SINR, $\gamma_k = 0$ if a user is not scheduled.

actual and quantized) are independent of each other, an obvious simplification is to have each user performing the adaptation so as to maximize its own expected rate, under an individual (the same for all) average feedback rate constraint $\bar{b}$. Letting B_{total} denote the total average feedback rate as in the previous section. $\bar{b} = \frac{B_{\text{total}}}{K}$.

Moreover, we remove the dependence on the other users' quantized channels (which depends on the final result of the bit rate allocation) by approximating the SINR as in (3.37). Though this SINR provides an upper bound on the achieved rate when N_t users are scheduled, under equal power allocation, as the related literature shows, it gives a reasonable indicator of the rates achieved.

As different users' channels are i.i.d., the approximate expected rate at any user will be the same, assuming similar feedback strategies, i.e. $\mathbb{E}R_i = \mathbb{E}R, \forall i = 1, \ldots, K$ (cf. (3.38)). The feedback rate is adaptive, so $\mathbb{E}R$ will depend on whether the user decides to feed back his CSI, how accurately he does so, and on how the scheduling probability and resulting rate are affected by this. Thus user k follows this set of steps:

1. Select the codebook to use for quantization, as a function of the CQI, $\|\mathbf{h}_k\|^2$;
2. Quantize the channel given the selected codebook;
3. Decide, given the expected rate to be achieved by feeding back, whether it is worth it.

Selecting the codebook is a function of the CQI alone, since we consider Rayleigh fading channels where channel directions are isotropically distributed on the complex unit sphere; moreover, user channels are independent and we assume that individual codebooks are designed so that on average the quantization error does not depend on the direction; this holds for example for random vector quantization (RVQ).

We define $b(.)$ and $y_{\max}(.)$ as the functions of the channel norm squared which specify the size of the codebook to use, and the maximum quantization error for feedback, respectively.

Lemma 2. *The expected rate at any given user may be approximated in terms of the joint statistics of the channel norm and quantization error by:*

$$\mathbb{E}R = \int_0^\infty P_{S|\|\mathbf{h}\|^2}(x) \int_0^{y_{\max}(x)} \log_2\left(1 + \hat{\gamma}(x,y)\right) f_{\sin^2\theta|\|\mathbf{h}\|^2}(y|x) dy f_{\|\mathbf{h}\|^2}(x) dx, \tag{3.39}$$

where

- $\sin^2\theta$ *is the random variable representing the quantization error (see section 3.5.2).*
- x *and* y *denote the variables in the integration with respect to* $\|\mathbf{h}\|^2$ *and* $\sin^2\theta$*, respectively.*
- $\hat{\gamma}(x,y)$ *is the value of the SINR estimate given in* (3.37) *, when* $\|\mathbf{h}\|^2 = a$ *and* $\sin^2\theta = y$*.*
- $P_{S|\|\mathbf{h}\|^2}(.)$ *is the probability of being scheduled, conditioned on* $\|\mathbf{h}\|^2$*.*

The probability of being scheduled and the pdf of the quantization error have not yet been specified. The latter will in reality depend on the codebook being used: in what follows, it is reduced to a dependence on the codebook's size alone by exploiting the cdf in (3.36). The scheduling probability on the other hand, depends on the scheduling scheme used, and on the overall feedback decision strategy. Intuitively, it is a non-decreasing function of the CQI, since a higher CQI would mean a higher rate. This is discussed in Appendix 3.C.

3.5.4 Feedback Optimization

A valid adaptive feedback strategy specifies the functionals $b(.)$ and $y(.)$ as defined in the previous section, while meeting the average feedback bit rate constraint

$$\int_0^\infty \int_0^{y_{\max}(x)} b(x) f_{\sin^2\theta|\|\mathbf{h}\|^2}(y|x)dy f_{\|\mathbf{h}\|^2}(x)dx \leq \bar{b} = \frac{B_{\text{total}}}{K}. \tag{3.40}$$

3.5.5 Simplified Water-filling Solution

The optimal solution of the problem would require a dynamic programming formulation of the problem, due to its nonconvexity. This is done in [42] for example in the context of adaptive transmission with finite code rates: this could be related to our setup in the case where the sizes of the different codebooks available are pre-specified.

Instead of pursuing such an approach, we simplify the problem statement by letting $y_{\max}(x) = \bar{y}_{\max}, \forall x \in [0,\infty)$, $\bar{y}_{\max}$ being a value to be determined. We refer to the solution obtained as a 'water-filling' kind of solution in the sense that strictly positive codebook sizes will be allocated only if the channel quality is above a certain level.

One can show that, for the optimal $b(.)$, the above definition of $y_{\max}(.)$ will divides the range of channel norms squared into two parts:

- $[0, x_{\max}]$ for which $b(x)$ is such that $2^{\frac{b(x)}{N_t-1}} \geq y_{\max}$, and
- $[x_{\max}, \infty)$ for which $b(x)$ is such that $2^{\frac{b(x)}{N_t-1}} < \bar{y}_{\max}$.

The average feedback bit rate will be split among them, let $\bar{b}_1$ be the average feedback bit rate allocated to $[0, x_{\max}]$ and $\bar{b}_2$ the average feedback bit rate allocated to $[x_{\max}, \infty)$. For the optimal allocation, $\bar{b}_1 + \bar{b}_2 = \bar{b}$.

For fixed $x_{\max}$, $\bar{y}_{\max}$ and $\bar{b}_1$, relaxing the integer constraint on $b(.)$, the optimal $b(.)$ is determined by solving two subproblems.

$$\begin{aligned}
\max. \quad & \int_0^{x_{\max}} P_{S|\|\mathbf{h}\|^2}(x) \int_0^{\bar{y}_{\max}} \log_2\left(1 + \frac{\frac{P}{N_t}x(1-y)}{1+\frac{P}{N_t}xy}\right)(N_t-1)2^{b(x)}y^{N_t-2}dy f_{\|\mathbf{h}\|^2}(x)dx \\
\text{s.t.} \quad & \int_0^{x_{\max}} \int_0^{\bar{y}_{\max}} b(x) f_{\sin^2\theta|\|\mathbf{h}\|^2}(y|x)dy f_{\|\mathbf{h}\|^2}(x)dx \leq \bar{b}_1 \\
& 2^{-\frac{b(x)}{N_t-1}} \geq \bar{y}_{\max}, \forall x \in [0, x_{\max}] \\
& b(x) \geq 0, \forall x \in [0, x_{\max}].
\end{aligned} \tag{3.41}$$

and

$$\begin{aligned}
\max. \quad & \int_{x_{\max}}^{\infty} P_{S|\|\mathbf{h}\|^2}(x) \int_0^{2^{-\frac{b(x)}{N_t-1}}} \log_2\left(1 + \frac{\frac{P}{N_t}x(1-y)}{1+\frac{P}{N_t}xy}\right)(N_t-1)2^{b(x)}y^{N_t-2}dy f_{\|\mathbf{h}\|^2}(x)dx \\
\text{s.t.} \quad & \int_{x_{\max}}^{\infty} b(x) f_{\|\mathbf{h}\|^2}(x)dx \leq \bar{b}_2, \\
& b(x) \geq -(N_t-1)\log_2 \bar{y}_{\max}, \forall x \in [x_{\max}, \infty).
\end{aligned} \tag{3.42}$$

These are solved in Appendices 3.D and 3.E respectively. The optimal bit rate allocation, for given $\bar{y}_{\max}$ satisfies

$$b(x) = \begin{cases} 0, & P_{S|\|\mathbf{h}\|^2}(x) \leq \lambda_1 \\ \frac{1}{\log 2}\left[\frac{P_{S|\|\mathbf{h}\|^2}(x)}{\lambda_1} - 1\right], & \lambda_1 < P_{S|\|\mathbf{h}\|^2}(x) < x_t \\ (1-N_t)\log_2 \bar{y}_{\max}, & x_t \leq P_{S|\|\mathbf{h}\|^2}(x) \leq x_{\max} \\ \arg_b P_{S|\|\mathbf{h}\|^2}(x)2^b \int_0^{2^{-\frac{b}{N_t-1}}} \frac{\frac{P}{N_t}x}{1+\frac{P}{N_t}xy}y^{N_t-1}dy - \lambda_2 = 0, & x \geq x_{\max} \end{cases}, \tag{3.43}$$

where $x_t = \min\left(x_{\max}, \lambda_1(1 + (1 - N_t)\log \bar{y}_{\max})\right)$, is the first channel norm for which feedback occurs, and λ_1, λ_2 are such that

$$\int_0^{x_{\max}} b(x) 2^{b(x)} \bar{y}_{\max}^{N_t - 1} f_{\|\mathbf{h}\|^2}(x) dx = \bar{b}_1 \tag{3.44}$$

$$\int_{x_{\max}}^{\infty} b(x) f_{\|\mathbf{h}\|^2}(x) dx = \bar{b}_2 \tag{3.45}$$

This may be solved numerically by two line searches over $\bar{y}_{\max}$ and $x_{\max}$.

3.5.6 Step-Functions' Solution

A simplification of the above approach is to restrict the set of codebooks to just a single one. Let b_{fixed} denote the size of said codebook. Then if $\|\mathbf{h}\|^2 \geq x_t$ and $\sin^2\theta < \bar{y}_{\max}$, where x_t and $\bar{y}_{\max}$ are to be optimized over, the user will feedback his CSI.

The objective function in (3.39) can be rewritten as:

$$\int_{x_t}^{\infty} P_{S|\|\mathbf{h}\|^2}(x) f_{\|\mathbf{h}\|^2}(x) \int_0^{\bar{y}_{max}} \log_2\left(1 + \frac{\frac{Px}{N_t}(1-y)}{1 + \frac{Px}{N_t} y}\right) f_{\sin^2\theta}(y) dy dx, \tag{3.46}$$

whereas the feedback rate constraint is given by:

$$\begin{aligned} & b_{fixed} F_{\sin^2\theta}(\bar{y}_{max}) \left[1 - F_{\|\mathbf{h}\|^2}(x_t)\right] \\ & = b_{fixed} 2^{b_{fixed}} (\bar{y}_{max})^{N_t - 1} \left[1 - F_{\|\mathbf{h}\|^2}(x_t)\right] \leq \bar{b}. \end{aligned} \tag{3.47}$$

The optimization problem is not convex in general. However, due its low dimensionality, an exhaustive search approach can be used to solve it. For fixed x_t, one could try different values of b_{fixed} (up to some maximum), use the feedback rate constraint to find the corresponding $\bar{y}_{max}$, then compute the resulting expected rate.

3.6 Simulation Results

Simulations were run to investigate the performance of the heuristic feedback rate splitting for the two-stage feedback approach. [3]

Fig. 3.4 compares the performance for a 30-user system for different values of α. As can be seen, our scheme effectively provides a smooth transition between the two extreme cases, and comparing to the α_{opt} case (found

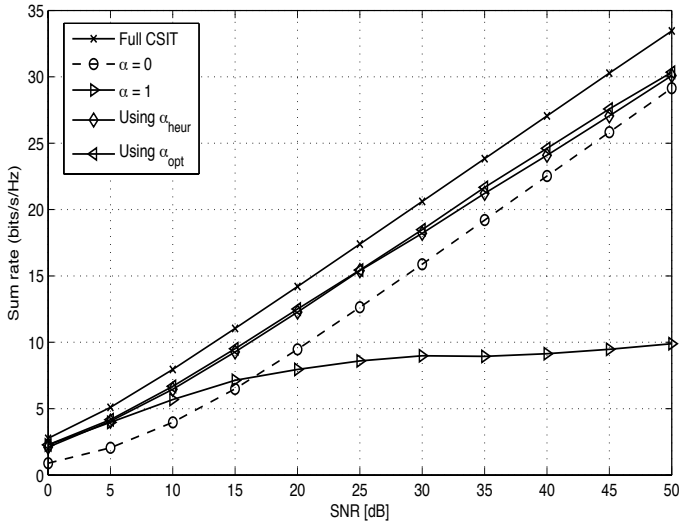


Figure 3.4: Average sum rates for $N_t = 2$, $K = 30$, $B_{\text{total}} = 120$ bits, and different schemes. α_{heur} provides a smooth transition between extreme α's, with tolerable loss with respect to α_{opt}.

by exhaustive search) the loss due to the non-optimality of the heuristic is reasonable.

Figures 3.5 and 3.6 compare, for an average feedback rate per user of 4 bits, the performance of several schemes. To generate the quantization errors, the model given by (3.36) is used, and non-integer codebook sizes are allowed. The simulations are run $N_t = 2$, and for different values of the semi-orthogonality parameter ϵ. The schemes in the figures are:

- A non-adaptive scheme, where each user always feeds back his CDI using a fixed codebook,
- The adaptive scheme described in Section 3.5.6, where each user uses a single codebook to quantize his CDI, but only does so if the quantization and channel quality are good enough,
- The adaptive scheme described in Section 3.5.5, where the user switches between several codebooks.

[3] As scheduling according to (3.8) is too computationally intensive, we resort to the SUS algorithm of [8] instead.

- The above scheme where the sizes of the codebooks are truncated.

Most of our comparisons do not account for the signaling required to announce which codebook is being used. This will actually depend on how the feedback is designed. In an attempt to take that into account, we show an additional curve marked 'Fixed Fb (with entropy)' in Figure 3.6: this is obtained by computing the entropy of the source corresponding to the sizes of the different codebooks used when the adaptive scheme's codebook sizes are truncated to the nearest integer, and increasing the size of the fixed size feedback codebook by that amount. For few users (5 for example, not shown here), the gain from this is quite significant and the resulting codebook performs better than an adaptive scheme. For a larger number of users, this enhances performance but the achievable sum rate still saturates much earlier than with the adaptive scheme.

Depending on how one signals which codebook is currently used, the simpler single codebook adaptive scheme may be preferable. Note, however, that the fully adaptive scheme provides more information as to the channel norm, so using it may lessen the feedback required for CQI feedback, which was assumed to be known perfectly in this study.

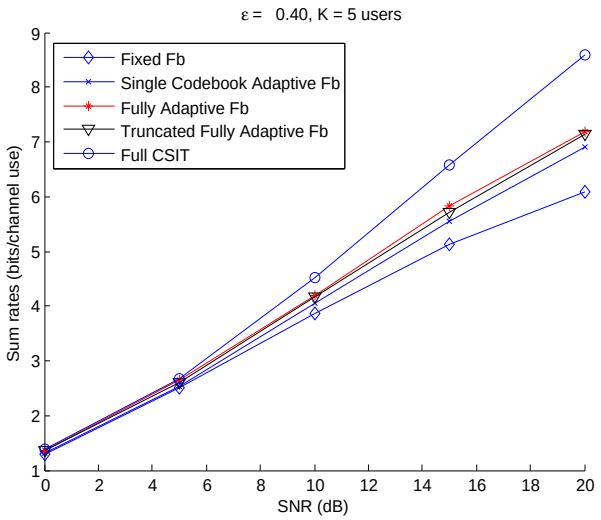


Figure 3.5: Average sum rates for $N_t = 2$, $K = 5$, $B_{\text{total}} = 20$ bits, and different adaptive and non-adaptive schemes.

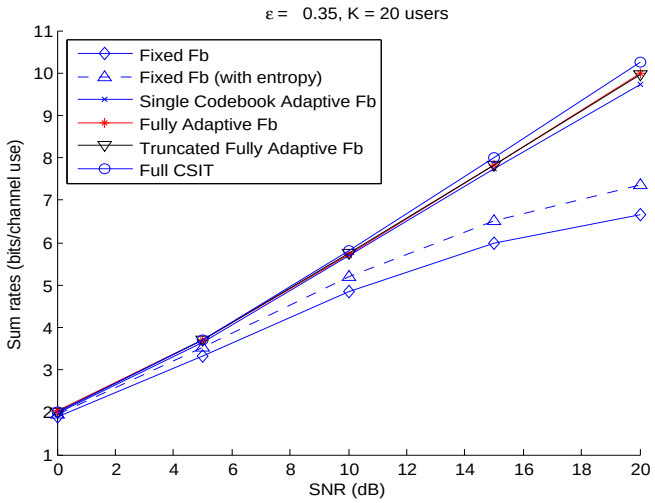


Figure 3.6: Average sum rates for $N_t = 2$, $K = 20$, $B_{\text{total}} = 80$ bits, and different adaptive and non-adaptive schemes.

3.7 Conclusion and future works

In this chapter, a two-stage resource allocation scheme was proposed for the multiuser MIMO broadcast channel under feedback rate constraint: the available feedback rate is split between the scheduling phase where all receivers feed back a coarse quantization of their channel state information (CSIT), and the precoding phase where the selected receivers feed back refined versions of their CSI for precoding matrix design. The optimum splitting of the available feedback rate between the two stages was investigated in a Rayleigh fading channel, and a heuristic algorithm was derived and tested.

Another feedback scheme, where each user adapts its feedback to its current channel state, is also proposed: thus, depending on the current channel norm, a different size codebook is used for feedback; the better the channel, the larger the codebook. A 'water-filling' kind of solution to the codebook size optimization problem is obtained.

Our focus has not been on the design of the quantization schemes but rather on illustrating how modifying the feedback strategies themselves can improve performance, although more judicious design would improve this further.

We have assumed that all users have the same statistics and equal weights as far as maximizing the sum rate; the average feedback rate is thus the same for all; if this does not hold, different users could be allocated different portions of the average feedback rate. How this could be done warrants further research.

3.A Two stage channel estimation

In a reciprocal system, a training based CSI acquisition scheme may be considered. We modify the scheme proposed in [35] for our two stage approach. Thus in a first stage, users all transmit simultaneously during some training interval $\tau_1 \geq K$. Let $\mathbf{S}_1$ denote the $K \times \tau_1$ matrix of training signals, then $\mathbf{S}_1$ may be written as

$$\mathbf{S} = \sqrt{\tau_1 P_{ul}}\Phi_1, \tag{3.48}$$

where P_{ul} is the transmit power on the UL, and Φ_1 is a $K \times \tau_1$ unitary matrix so that

$$\Phi_1 \Phi_1^H = \mathbf{I}_K. \tag{3.49}$$

The $N_t \times \tau_1$ signal matrix received at the BS is then expressed as

$$\mathbf{X}_1 = \sqrt{\tau_1 P_{ul}}\mathbf{H}^T \Phi_1 + \mathbf{Z}_1, \tag{3.50}$$

where $\mathbf{H}$ is the $K \times N_t$ concatenated channel of all the users in the system and $\mathbf{Z}_1$ is the $N_t \times \tau_1$ matrix of noise samples at the UL receiver, whose elements are i.i.d. $\mathcal{CN}(0, \sigma_{BS}^2)$. Post-multiplying this by Φ_1^H, we get

$$\tilde{\mathbf{X}}_1 = \mathbf{X}_1 \Phi_1^H = \sqrt{\tau_1 P_{ul}}\mathbf{H}^T + \tilde{\mathbf{Z}}_1, \tag{3.51}$$

where the elements of the new noise matrix $\tilde{\mathbf{Z}}_1$ are also i.i.d. $\mathcal{CN}(0, \sigma_{BS}^2)$.

Let $\mathcal{A}$ of cardinality N_t denote the set of scheduled users. Extracting only the columns which correspond to those users from $\tilde{\mathbf{X}}_1$

$$\tilde{\mathbf{X}}_{1,\mathcal{A}} = \sqrt{\tau_1 P_{ul}}\mathbf{H}_{\mathcal{A}}^T + \tilde{\mathbf{Z}}_{1,\mathcal{A}}. \tag{3.52}$$

In the second stage, of duration $\tau_2 \geq N_t$, the above scheme is repeated with only the users in $\mathcal{A}$ transmitting. Denoting by $\mathbf{X}_2$ the receive signal in the second stage, post-multiplying it by the corresponding unitary matrix Φ_2 yields

$$\tilde{\mathbf{X}}_{2,\mathcal{A}} = \sqrt{\tau_2 P_{ul}}\mathbf{H}_{\mathcal{A}}^T + \tilde{\mathbf{Z}}_{2,\mathcal{A}}, \tag{3.53}$$

where the elements of noise matrix $\tilde{\mathbf{Z}}_2$ are also i.i.d. $\mathcal{CN}(0, \sigma_{BS}^2)$ and independent of those in $\tilde{\mathbf{Z}}_1$.

Let $\hat{\mathbf{H}}_{1,\mathcal{A}}$ denote the estimate obtained from $\tilde{\mathbf{X}}_{1,\mathcal{A}}$ alone, and $\hat{\mathbf{H}}_{2,\mathcal{A}}$ denote that obtained from both $\tilde{\mathbf{X}}_{1,\mathcal{A}}$ and $\tilde{\mathbf{X}}_{2,\mathcal{A}}$, the following holds for the (n,k)th element of the matrix:

$$[\mathbf{H}_{\mathcal{A}}]_{nk}|[\hat{\mathbf{H}}_{2,\mathcal{A}}]_{nk} \sim \mathcal{CN}\left([\hat{\mathbf{H}}_{2,\mathcal{A}}]_{nk}, \sigma_{e,2}^2\right), \tag{3.54}$$

$$[\hat{\mathbf{H}}_{2,\mathcal{A}}]_{nk}|[\hat{\mathbf{H}}_{1,\mathcal{A}}]_{nk} \sim \mathcal{CN}\left([\hat{\mathbf{H}}_{1,\mathcal{A}}]_{nk}, \sigma_{e,1}^2 - \sigma_{e,2}^2\right), \tag{3.55}$$

$$[\hat{\mathbf{H}}_{1,\mathcal{A}}]_{nk} \sim \mathcal{CN}\left(0, 1 - \sigma_{e,1}^2\right) \tag{3.56}$$

where

$$\sigma_{e,1}^2 = \frac{1}{1 + \frac{P_{ul}}{\sigma_{BS}^2}\tau_1} \tag{3.57}$$

$$\sigma_{e,2}^2 = \frac{1}{1 + \frac{P_{ul}}{\sigma_{BS}^2}(\tau_1 + \tau_2)} \tag{3.58}$$

Let τ denote the number of slots dedicated to total feedback, $\tau = \tau_1 + \tau_2$, the split is to be optimized while keeping the constraint that $\tau_1 \geq K$, $\tau_2 \geq N_t$.

3.B Bounds on Two-stage Limited Feedback

The expectation $\overline{SR}_{ZF\text{–}Q}$ can be reformulated as shown in equation (3.59) below,

$$\begin{aligned}
\overline{SR}_{ZF\text{–}Q} &= \mathbb{E}_{\hat{\mathbf{H}}_{\hat{\mathcal{A}}}, \mathbf{E}_{\hat{\mathcal{A}}}} \sum_{i=1}^{N_t} \log_2(1 + SINR_{\hat{A}_i}) \\
&= N_t \mathbb{E}_{\hat{\mathbf{H}}_{\hat{\mathcal{A}}}, \mathbf{E}_{\hat{\mathcal{A}}}} \log\left(1 + \frac{1}{\frac{\mathrm{tr}((\hat{\mathbf{H}}_{\hat{\mathcal{A}}}\hat{\mathbf{H}}_{\hat{\mathcal{A}}}^H)^{-1})}{P} + \|(\mathbf{E}_{\hat{\mathcal{A}}}\hat{\mathbf{H}}_{\hat{\mathcal{A}}}^{\dagger})_i\|^2}\right) \\
&\overset{(a)}{=} N_t \mathbb{E}_{\hat{\mathbf{\Lambda}}_{\hat{\mathcal{A}}}, \hat{\mathbf{U}}, \mathbf{E}_{\hat{\mathcal{A}}}} \log\left(1 + \frac{1}{\frac{\mathrm{tr}(\hat{\Lambda}_{\hat{\mathcal{A}}}^{-1})}{P} + \mathbf{E}_{\hat{\mathcal{A}},i}\hat{\mathbf{U}}\hat{\Lambda}_{\hat{\mathcal{A}}}^{-1}\hat{\mathbf{U}}^H\mathbf{E}_{\hat{\mathcal{A}},i}^H}\right) \\
&\overset{(b)}{=} N_t \mathbb{E}_{\hat{\mathbf{\Lambda}}_{\hat{\mathcal{A}}}, \mathbf{E}_{\hat{\mathcal{A}}}} \log\left(1 + \frac{1}{\sum_{j=1}^{N_t} \lambda_{\hat{\mathcal{A}},j}^{-1}(\frac{1}{P} + |\mathbf{E}_{\hat{\mathcal{A}},i,j}|^2)}\right),
\end{aligned} \tag{3.59}$$

where in (a) $\hat{\mathbf{H}}_{\hat{\mathcal{A}}}$ is replaced by its eigenvalue decomposition ($\hat{\mathbf{\Lambda}}_{\hat{\mathcal{A}}}$ is the diagonal matrix containing the eigenvalues $\{\lambda_{\hat{\mathcal{A}},j}\}$ of $\hat{\mathbf{H}}_{\hat{\mathcal{A}}}\hat{\mathbf{H}}_{\hat{\mathcal{A}}}^H$; $\hat{\mathbf{U}}$ is unitary), and (b) is obtained by noting that $\mathbf{E}_{\hat{\mathcal{A}},i}\hat{\mathbf{U}}$ has the same statistics as $\mathbf{E}_{\hat{\mathcal{A}},i}$[4].

[4] $\mathbf{E}_{\hat{\mathcal{A}},i}$ corresponds to the error vector associated with the quantization of user i's channel, user i being an arbitrary user in $\hat{\mathcal{A}}$.

$D \triangleq \sum_{j=1}^{N_t} \lambda_{\hat{\mathcal{A}},j}^{-1}(1/P + |\mathbf{E}_{\hat{\mathcal{A}},i,j}|^2)$ in (3.59) may be bounded as [43]:

$$\frac{1/P + |\mathbf{E}_{\hat{\mathcal{A}},i,j_{\min}}|^2}{\lambda_{\min}} < D < \frac{\sum_{j=1}^{N_t}(1/P + |\mathbf{E}_{\hat{\mathcal{A}},i,j}|^2)}{\lambda_{\min}}, \tag{3.60}$$

where $\lambda_{\min}$ is the smallest eigenvalue in the summation and $j_{\min}$ the index of the corresponding entry in the $\mathbf{E}_{\hat{\mathcal{A}},i}$ vector. $\overline{SR}_{ZF\text{–}Q}$ is thus bounded as in equation (3.61). The sum of N_t squared norms of i.i.d. $\mathcal{CN}(0,\sigma_e^2)$ r.v.'s in the lower bound's denominator, and $|\mathbf{E}_{\hat{\mathcal{A}},i,j}|^2$ in the upper bound are replaced by a Gamma(N_t,σ_e^2) distributed r.v., γ_{N_t} and a Gamma$(1,\sigma_e^2)$ distributed r.v., γ_1, respectively. Once these changes of variable made, applying Jensen's inequality to the lower bound, and averaging over γ_1 in the upper bound, then upper bounding the result, yields (3.20).

$$\begin{aligned}\mathbb{E}_{\lambda_{\min},\mathbf{E}_{\hat{\mathcal{A}}}} \log\left(1 + \frac{\lambda_{\min}}{\sum_{j=1}^{N_t}(\frac{1}{P} + |\mathbf{E}_{\hat{\mathcal{A}},i,j}|^2)}\right) &< \frac{\overline{SR}_{ZF\text{–}Q}}{N_t} \\ < \mathbb{E}_{\lambda_{\min},\mathbf{E}_{\hat{\mathcal{A}}}} \log\left(1 + \frac{\lambda_{\min}}{\frac{1}{P} + |\mathbf{E}_{\hat{\mathcal{A}},i,j_{\min}}|^2}\right).\end{aligned} \tag{3.61}$$

3.C Probability of being scheduled

For any greedy scheduling algorithm, the probability of a given user being scheduled given its CQI, can be written as:

$$P_{S|\|\mathbf{h}\|^2}(x) = \sum_{i=1}^{N_t} \Pr\left[S_i \middle| \|\mathbf{h}\|^2 = X\right], \tag{3.62}$$

where S_i denotes the event of being the ith scheduled user.

Assuming no feedback is done for the given channel norm, the probability of being scheduled will be zero. Otherwise, finding closed form solutions for $\Pr\left[S_i \middle| \|\mathbf{h}\|^2 = X\right]$ for any number of antennas appears to be quite tedious. For $N_t = 2$, the first user selected will be the one with the highest fed back CQI; whereas the second user selected (if any) will be, out of the ϵ-orthogonal users, the one who reports back the highest projected channel norm. To obtain tractable expressions for the scheduling probability, we approximate the probability of being scheduled second by that of having the highest channel norm among the ϵ orthogonal users, instead of having the highest projected channel norm.

Now given our feedback strategy, feedback occurs for channel norm squared X if:

$$\sin^2\theta \leq y_{\max}(X) \tag{3.63}$$

Consequently, the probability of feedback given that the channel norm is less than X:

$$P_{fb}(X) = \int_0^X f_{\|\mathbf{h}\|^2}(x) F_{\sin^2\theta}\left(y_{\max}(x)\right) dx, \tag{3.64}$$

and the overall probability of feedback:

$$P_{fb} = P_{fb}(\infty), \tag{3.65}$$

where:

$$F_{\sin^2\theta}\left(y(x)\right) = \begin{cases} 2^{b(x)}y & 0 \leq y \leq 2^{-b(x)} \\ 1 & \text{o.w.} \end{cases} \tag{3.66}$$

Thus,

$$\Pr\left[S_1 \middle| \|\mathbf{h}\|^2 = X\right] = \Pr\left[\text{fb with the highest channel norm} \middle| \|\mathbf{h}\|^2 = X\right] \tag{3.67}$$

and

$$\Pr\left[S_2 \middle| \|\mathbf{h}\|^2 = X\right] = \Pr\left[\text{fb with the highest norm in } \mathcal{S}_{\epsilon,1} \middle| \|\mathbf{h}\|^2 = X\right], \tag{3.68}$$

where $\mathcal{S}_{\epsilon,1}$ is the set of users ϵ-orthogonal to the first user selected.

We denote the first probability by P_1, the second being P_2.

$$P_1 = F_{\sin^2\theta}\left(y_{\max}(x)\right) \sum_{i=0}^{K-1} \binom{K-1}{i} (1 - P_{fb})^{K-1-i} P_{fb}(X)^i. \tag{3.69}$$

To compute P_2, we require the probability of two users' fed back directions being ϵ-orthogonal. Assuming independent codebooks and isotropically distributed channel directions and codewords, then for any two users k and l, $k \neq l$, their quantized CDI, $\hat{\mathbf{h}}_k$ and $\hat{\mathbf{h}}_l$ will be such that $|\hat{\mathbf{h}}_k \hat{\mathbf{h}}_l|^2$ follows a $\beta(1,1)$ distribution. Consequently, the probability that they are ϵ-orthogonal is given by

$$p_\epsilon = \int_0^\epsilon f_{\cos^2\theta}(v) dv = \epsilon. \tag{3.70}$$

$$
\begin{aligned}
P_2 &= F_{\sin^2\theta}\left(y_{\max}(x)\right) p_\epsilon \binom{K-1}{1} \int_X^\infty f_{\|\mathbf{h}\|^2}(t) F_{\sin^2\theta}\left(y(t)\right) \{ \\
&\sum_{i=1}^{K-2} \binom{K-2}{i} (1-P_{fb})^{K-2-i} \sum_{j=0}^{i} \binom{i}{j} p_\epsilon^j P_{fb}(t)^{i-j} (1-p_\epsilon)^{i-j} P_{fb}(X)^j dt \; \} \\
&= F_{\sin^2\theta}\left(y_{\max}(x)\right) \sum_{i=1}^{K-2} \frac{(K-1)!\,(1-P_{fb})^{K-2-i}}{(K-2-i)!} \{ \\
&\sum_{j=0}^{i} \frac{p_\epsilon^{j+1}(1-p_\epsilon)^{i-j}}{j!(i-j)!} P_{fb}(X)^j \left[\int_X^\infty f_{\|\mathbf{h}\|^2}(t) F_{\sin^2\theta}\left(y(t)\right) P_{fb}(t)^{i-j} \right] dt \; \}
\end{aligned}
\tag{3.71}
$$

3.D Sub-problem (3.41) Solution

Sub-problem (3.41) is equivalent to the following (functional) convex optimization problem over $u(x) = 2^{b(x)}$

$$
\begin{aligned}
\max. \quad & \int_0^{x_{\max}} P_{S|\|\mathbf{h}\|^2}(x) u(x) f_{\|\mathbf{h}\|^2}(x) dx \\
\text{s.t.} \quad & \int_0^{x_{\max}} u(x) \log u(x) f_{\|\mathbf{h}\|^2}(x) dx \leq \frac{\bar{b}_1 \log(2)}{\bar{y}_{\max}^{N_t-1}} \\
& u(x) \leq \bar{y}_{\max}^{1-N_t}, \forall x \in [0, x_{\max}] \\
& u(x) \geq 1, \forall x \in [0, x_{\max}].
\end{aligned}
\tag{3.72}
$$

Let λ_1 be the Lagrangian coefficient associated with the modified average bit rate constraint in (3.72), $\mu_1(x)$ and $\eta_1(x)$ the Lagrangian multipliers associated with the remaining two constraints. The corresponding Euler-Lagrange equation is given by:

$$
P_{S|\|\mathbf{h}\|^2}(x) f_{\|\mathbf{h}\|^2}(x) - \lambda_1 f_{\|\mathbf{h}\|^2}(x) \left(\log u(x) + 1\right) - \mu_1(x) + \eta_1(x) = 0. \tag{3.73}
$$

It is enough to focus on x's with strictly positive $f_{\|\mathbf{h}\|^2}(x)$'s. Normalizing the above equation by $f_{\|\mathbf{h}\|^2}(x)$ and letting $\mu_1'(x) = \frac{\mu_1(x)}{f_{\|\mathbf{h}\|^2}(x)}$ and $\eta_1'(x) = \frac{\eta_1(x)}{f_{\|\mathbf{h}\|^2}(x)}$, we get:

$$
P_{S|\|\mathbf{h}\|^2}(x) - \lambda_1 \left(\log u(x) + 1\right) - \mu_1'(x) + \eta_1'(x) = 0. \tag{3.74}
$$

The optimal $u(x)$ has the following form:

$$u(x) = \begin{cases} 1 & P_{S|\|\mathbf{h}\|^2}(x) \leq \lambda_1 \\ \bar{y}_{\max}^{1-N_t} & P_{S|\|\mathbf{h}\|^2}(x) > \lambda_1 \left(\log \bar{y}_{\max}^{1-N_t} + 1\right) \\ \exp\left(\frac{P_{S|\|\mathbf{h}\|^2}(x)}{\lambda_1} - 1\right) & \text{otherwise.} \end{cases} \tag{3.75}$$

The optimal λ_1 is equal to 0, if $F\left(x_{\max}\right) \log\left(\bar{y}_{\max}^{1-N_t}\right) < b_1 \log(2)$, in which case the constraint is not met with equality, so that the extra bit rate should be shifted to the other range. If the constraint is met with equality, define $x_{\lambda_1,\min}$ such that $P_{S|\|\mathbf{h}\|^2}(x_{\lambda_1,\min}) = \lambda_1$, $x_{\lambda_1,\max}$ such that $P_{S|\|\mathbf{h}\|^2}(x_{\lambda_1,\max}) = \left(\log \bar{y}_{\max}^{1-N_t} + 1\right)\lambda_1$,

$$\begin{aligned} &\int_{x_{\lambda_1,\min}}^{x_{\lambda_1,\max}} \left(\frac{P_{S|\|\mathbf{h}\|^2}(x)}{\lambda_1} - 1\right) \exp\left(\frac{P_{S|\|\mathbf{h}\|^2}(x)}{\lambda_1} - 1\right) f_{\|\mathbf{h}\|^2}(x)dx \\ &+ \int_{x_{\lambda_1,\max}}^{x_{\max}} \bar{y}_{\max}^{1-N_t} \log \bar{y}_{\max}^{1-N_t} f_{\|\mathbf{h}\|^2}(x)dx = \frac{\bar{b}_1 \log(2)}{\bar{y}_{\max}^{N_t-1}} \end{aligned} \tag{3.76}$$

3.E Sub-problem (3.42) solution

Sub-problem (3.42) can also be shown to be convex.

Proof. Let I_x denote the inner integral in the objective function:

$$I_x = \int_0^{2^{-\frac{b(x)}{N_t-1}}} \log_2\left(1 + \frac{\frac{P}{N_t}x(1-y)}{1+\frac{P}{N_t}xy}\right)(N_t-1)2^{b(x)}y^{N_t-2}dy. \tag{3.77}$$

One can show that

$$\begin{aligned} \frac{dI_x}{db(x)} &= 2^{b(x)} \int_0^{2^{-\frac{b(x)}{N_t-1}}} \frac{\frac{P}{N_t}x}{1+\frac{P}{N_t}xy} y^{N_t-1}dy \\ &= -u_x^{N_t-1}\left[(-1)^{N_t}\log\left(1+\frac{1}{u_x}\right) + \sum_{i=1}^{N_t-1}(-1)^i u_x^{i-N_t}\frac{1}{N_t-i}\right], \end{aligned} \tag{3.78}$$

where $u_x = \frac{1}{\frac{P}{N_t}x2^{-\frac{b(x)}{N_t-1}}}$.

$$\frac{d^2 I_x}{db(x)du_x}$$
$$= -(N_t-1)u_x^{N_t-2}\left[(-1)^{N_t}\log\left(1+\frac{1}{u_x}\right)+\sum_{i=1}^{N_t-1}(-1)^i u_x^{i-N_t}\frac{1}{N_t-i}\right]$$
$$-u_x^{N_t-1}\left[(-1)^{N_t}\left[\frac{1}{1+u_x}-\frac{1}{u_x}\right]-\sum_{i=1}^{N_t-1}(-1)^i u_x^{i-N_t-1}\right] \quad (3.79)$$

Let $v_x=\frac{1}{u_x}$ and consider the function

$$f(v_x)=(N_t-1)\left[(-1)^{N_t}\log(1+v_x)+\sum_{i=1}^{N_t-1}(-1)^i v_x^{N_t-i}\frac{1}{N_t-i}\right]$$
$$+\left[(-1)^{N_t}\left[\frac{1}{v_x+1}-1\right]-\sum_{i=1}^{N_t-1}(-1)^i v_x^{N_t-i}\right] \quad (3.80)$$

$$f'(v_x)=(N_t-1)\left[(-1)^{N_t}\frac{1}{1+v_x}+\sum_{i=1}^{N_t-1}(-1)^i v_x^{N_t-i-1}\right]$$
$$+\left[-(-1)^{N_t}\frac{1}{(v_x+1)^2}-\sum_{i=1}^{N_t-1}(-1)^i(N_t-i)v_x^{N_t-i-1}\right]$$
$$=(-1)^{N_t}\frac{1}{1+v_x}\left[(N_t-1)-\frac{1}{v_x+1}\right]+\sum_{i=1}^{N_t-1}(-1)^i(i-1)v_x^{N_t-i-1}$$
$$=\frac{v_x^{N_t-1}}{(1+v_x)^2}\geq 0 \quad (3.81)$$

Thus f is increasing in v_x, for $v_x\geq 0$. Its minimum value will be at 0 and is equal to 0. As a result,

$$\frac{d^2 I_x}{db(x)du_x}\leq 0. \quad (3.82)$$

Finally, since $\frac{d^2 I_x}{db(x)^2}=\frac{d^2 I_x}{db(x)du_x}\frac{du_x}{db(x)}$ and $\frac{du_x}{db(x)}=\frac{1}{\frac{P}{N_t}x2^{-\frac{b(x)}{N_t-1}}}\frac{\log 2}{N_t-1}\geq 0$, then

$$\frac{d^2 I_x}{db(x)^2}\leq 0, \quad (3.83)$$

which completes the proof of the concavity of the objective function.

□

The corresponding Euler-Lagrange equation is equal to

$$P_{S|\|\mathbf{h}\|^2}(x)\frac{dI_x}{db(x)} - \lambda_2 + \eta_2'(x) = 0, \tag{3.84}$$

where λ_2 is the Lagrange coefficient associated with the feedback rate constraint, $\eta_2(x)$ is that associated with the lower bound on $b(x)$; $\eta_2'(x) = \frac{\eta_2(x)}{f_{\|\mathbf{h}\|^2}(x)}$.

As the first term in the above equation is increasing in x for fixed $b(x)$ and decreasing in $b(x)$ for fixed x, if

$$P_{S|\|\mathbf{h}\|^2}(x_{\max})\bar{y}_{\max}^{1-N_t}\int_0^{\bar{y}_{\max}} \frac{\frac{P}{N_t}x_{\max}}{1+\frac{P}{N_t}x_{\max}y}y^{N_t-1}dy - \lambda_2 \leq 0, \tag{3.85}$$

then there must exist an x_{thres} such that for $x \in [x_{\max}, x_{thres}]$, $b(x) = (1-N_t)\log_2 \bar{y}_{\max}$, and for $x \geq x_{thres}$,

$$P_{S|\|\mathbf{h}\|^2}(x)2^{b(x)}\int_0^{2^{-\frac{b(x)}{N_t-1}}} \frac{\frac{P}{N_t}x}{1+\frac{P}{N_t}xy}y^{N_t-1}dy - \lambda_2 = 0. \tag{3.86}$$

x_{thres} is determined so that the bit rate constraint is met with equality. For it to be finite, $(1-N_t)\log_2 \bar{y}_{\max}\left[1-F\|\mathbf{h}\|^2(x)(x_{\max})\right] < \bar{b}_2$.

To ensure x_{thres} coincides with $x_{\max}$, the following must hold

$$P_{S|\|\mathbf{h}\|^2}(x_{\max})\bar{y}_{\max}^{1-N_t}\int_0^{\bar{y}_{\max}} \frac{\frac{P}{N_t}x_{\max}}{1+\frac{P}{N_t}x_{\max}y}y^{N_t-1}dy - \lambda_2 = 0. \tag{3.87}$$

Moreover, the bit rate constraint should be met with equality.

$$\bar{b}_2 = \int_{x_{\max}}^{\infty} b(x)f_{\|\mathbf{h}\|^2}(x)dx. \tag{3.88}$$

Chapter 4

Coordination on the MISO Interference Channel using the Virtual SINR Framework

4.1 Introduction

In this chapter, we begin considering multicell setups. As noted in the introductory chapter, in scenarios involving multiple transmitters and receivers, such as a multicell scenario, the performance attained, in terms of rates achieved for example, will depend on how much information may be shared at the nodes involved. Thus, if either all transmitters or all receivers share their entire data and as a result perform joint transmission or joint decoding respectively, the situation will be equivalent to a BC and a MAC, respectively. However, if this is not the case, then an interference channel (IC) is obtained. This is the situation considered in the present chapter, as sharing data may put too much strain on the backhaul of the system. More precisely we deal with the downlink direction and propose a transmission strategy based on the so-called "virtual SINR (VSINR) framework", explained below. In Chapters 5 and 6, we deal with the case where data sharing is allowed but taking into consideration the necessary CSIT shar-

ing load and the data backhaul constraint, respectively. This includes an extension of the VSINR framework to the case where data sharing is allowed.

Assuming each transmitter has multiple antennas and each receiver a single antenna, the setting is the MISO interference channel, considered for example in [23, 44] (the more general MIMO IC, which corresponds to receivers also having multiple antennas, is considered in [45, 46], among others). In particular, [44] and subsequent publications [47, 48] of the same authors have focused on the case of two transmitters and full CSIT. Considering the scenario from the viewpoint of game theory, with transmitters as players, a parametrization of the Pareto boundary of the rate region was found, and different algorithms suggested for finding different points on the boundary. [49] provides a parametrization of the Pareto boundary in a more general case.

Here we argue that it may not always be reasonable to assume that all the CSI is shared by all transmitters, and consider the case where each transmitter has local channel knowledge: it only knows the channel between itself and all receivers that are within its range. In a TDD system, this information may be gained from those users' transmission in the uplink. If reciprocity may not be assumed, one could consider that each receiver feeds back his full CSI to his serving BS which is partially shared with other base stations, thereby saving on signaling. This scenario has been tackled in [50] where an iterative method is proposed to achieve rates at all receivers involved that are higher than those achieved without cooperation. In contrast, what we develop in the present work is a one-shot algorithm. Given the local information at each transmitter, we propose a simple transmission scheme based on having each transmitter maximize what we refer to as a VSINR. For certain choices of parameters, the latter can be seen as the SINR achieved in the uplink if the same filters were used, in the TDD case, or in the virtual uplink (see [23]) in case there is no actual reciprocity.

The rest of the chapter is organized as follows. Section 4.2 defines the system model and performance measures considered. Section 4.3 introduces the VSINR framework. The approach of maximizing a virtual SINR at each transmitter is justified by relating it to uplink-downlink duality and to the full-CSIT results, in Section 4.4 below. Based on the given analysis, Section 4.5 states the proposed algorithm. Simulations in Section 4.6 show the value of the proposed algorithm in realistic scenarios.

4.2 System Model

We consider the MISO interference channel where K transmitters (e.g. base stations in a cellular system) with $N_t \geq 2$ antennas each, each communicate with a single receiver (mobile terminal) having a single antenna. This is illustrated in Figure 4.1 for $K = 2$, $N_t = 3$.

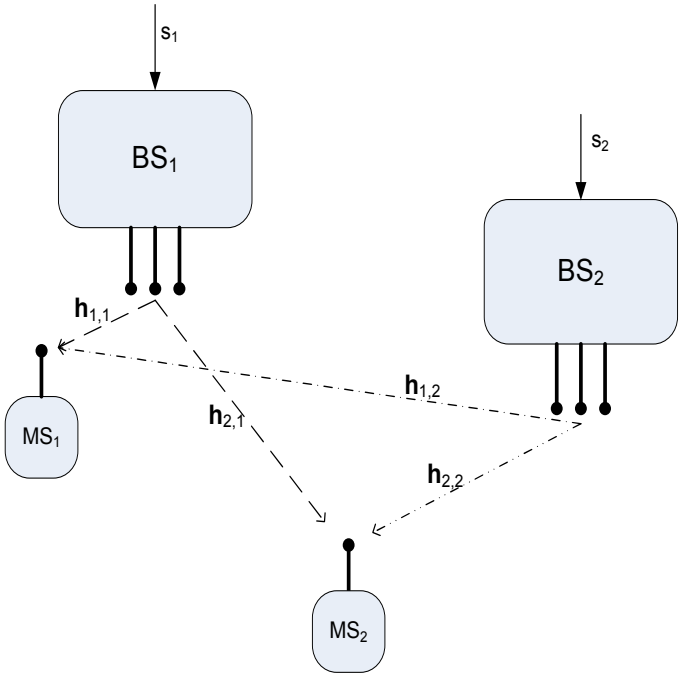


Figure 4.1: Scenario considered for $K = 2$, $N_t = 3$. $\mathbf{h}_{1,1}, \mathbf{h}_{2,1}$ are known at BS_1, $\mathbf{h}_{1,2}, \mathbf{h}_{2,2}$ at BS_2.

The signal transmitted by base station k, $\mathbf{x}_k$ is given by:

$$\mathbf{x}_k = \mathbf{w}_k s_k = \sqrt{p_k}\mathbf{u}_k s_k, \tag{4.1}$$

where $s_k \sim \mathcal{CN}(0, 1)$ is the transmit symbol intended for user k, $\mathbf{u}_k$ is the unit-norm beamforming vector used to carry this symbol and p_k is the transmit power used. A power constraint holds at each transmitter whereby $p_k \leq P$, P being the peak transmit power at each of the base stations.

The signal received at user k will be given by:

$$y_k = \sum_{j=1}^{K} \sqrt{p_j} \mathbf{h}_{k,j} \mathbf{u}_j s_j + n_k \tag{4.2}$$

where $\mathbf{h}_{k,j} \in \mathbb{C}^{1 \times N_t}$ is the channel between that user and base station j, $n_k \sim \mathcal{CN}(0, \sigma^2)$ is the noise at the considered receiver. The rate achieved at user k is given by:

$$R_k = \log_2(1 + \gamma_k) \tag{4.3}$$

where the SINR γ_k is equal to:

$$\gamma_k = \frac{p_k |\mathbf{h}_{k,k} \mathbf{u}_k|^2}{\sigma^2 + \sum_{j \neq k} p_j |\mathbf{h}_{k,j} \mathbf{u}_j|^2} \tag{4.4}$$

The rate region $\mathcal{R}$ is defined as the set of rates that may be achieved simultaneously at the different base stations, given the power constraints at each base station. I.e.,

$$\begin{aligned} \mathcal{R} = \{(R_1, \ldots, R_K) \in \mathbb{R}_+^K \\ | R_k \text{ as in } (4.3), p_k \leq P, \ \forall k \in \{1, \ldots, K\}\} \end{aligned} \tag{4.5}$$

Beamforming under distributed CSIT

We assume limited CSIT at each transmitter in that each transmitter only knows the channel between itself and all users [1], but not the channel between the other transmitters and the users. We would like to achieve a set of rates which is as close as possible to the boundary of $\mathcal{R}$, while yielding the best sum rate possible. Moreover, we would like to do so in a distributed fashion, relying only on *locally available* CSI as just defined, which leads us to an optimization problem solved at each base station within the VSINR framework, detailed in the next section.

4.3 Virtual SINR

In its most general form, a VSINR at base station k is defined as the ratio between the useful signal power received at its served user and the sum

[1] Strictly speaking, each transmitter only needs to know the channels between itself and users that are close enough to suffer from interference.

of noise plus a weighted sum of the interference powers it causes at the remaining users. From an intuitive point of view, it makes sense to want to leverage between the benefit at a given cell's own user and the harm caused at other cell's users if the cells wish to cooperate. Thus:

$$\gamma_k^{virtual} = \frac{p_k|\mathbf{h}_{k,k}\mathbf{u}_k|^2}{\sigma^2 + \sum_{j\neq k}\alpha_{kj}p_k|\mathbf{h}_{j,k}\mathbf{u}_k|^2}, \tag{4.6}$$

where $\alpha_{kj} \in \mathbb{R}_+, j,k = 1,\ldots,K$ are a given set of weights. Recall that $\|\mathbf{u}_k\| = 1$. This can be seen as the SINR achieved on the uplink of a system where at the kth base station, receive vector $\mathbf{w}_k$ is used to process the received signal, mobile station k transmits its signal with power p_k, and mobile $j, \forall j \neq k$ transmit with power $\alpha_{kj}p_k$: the 'virtual uplink' was first introduced in [23] in the context of downlink power control and beamforming in a multicell environment. Uplink-downlink duality is revisited in Chapter 7.
When transmitting at full power, Equation (4.6) becomes:

$$\gamma_k^{virtual} = \frac{|\mathbf{h}_{k,k}\mathbf{u}_k|^2}{\frac{1}{\rho} + \sum_{j\neq k}\alpha_{kj}|\mathbf{h}_{j,k}\mathbf{u}_k|^2}, \tag{4.7}$$

where $\rho = \frac{P}{\sigma^2}$.

As the objective is to have a distributed algorithm which relies only on information local to each base station, we propose that each transmitter solve a VSINR maximization problem, which can be stated as follows:

$$\mathbf{w}_k = \arg\max_{\|\mathbf{u}\|^2=1} \frac{|\mathbf{h}_{k,k}\mathbf{w}_k|^2}{\frac{1}{\rho} + \sum_{j\neq k}\alpha_{kj}|\mathbf{h}_{j,k}\mathbf{u}_k|^2}. \tag{4.8}$$

This is justified in the following section.

4.4 Analysis

As first noted in [23], the same rate region may be achieved in the UL (for a reciprocal channel, in the virtual UL otherwise) and DL directions using the same set of vectors for receive and transmit beamforming respectively, but with different power levels in both directions that satisfy the same total power constraint. This is one form of what is referred to as uplink-downlink duality. In our later analysis, we will show that due to the power constraints at each transmitter a modified version of uplink-downlink duality needs to be considered, which can be related to Lagrangian duality (see [14, 51] for example).

4.4.1 Virtual SINR Maximization as Pareto Boundary Achieving Strategy

Transmission in the MISO IC may be viewed as a game, where each of the transmitters is a player trying to achieve a certain goal, wich may be a selfish one such as trying to maximize his own rate or an altruistic one such as trying to maximize his own rate without causing any interference at any other user. One can then define the Pareto boundary of the channel, as the set of Pareto-optimal rate-tuples: thus, a given tuple belongs to the Pareto boundary if it is not possible to increase any rate within that tuple without decreasing at least one of the others. As shown in [44, 49], rates on the Pareto boundary of the MISO interference channel are achieved by transmitting at full power (at least for the case where $N_t \geq K$).

We reproduce Proposition 1 and Corrolary 2 from [49][2], which characterize the beamforming vectors along the Pareto boundary for the general case when $N_t \geq K \geq 2$ and for the special case of $K = 2, N_t \geq 2$, and relate these characterizations to the solution of a virtual SINR maximization problem as follows.

Proposition 1 ([49]). *Let k be given and fixed. Suppose that $\mathbf{h}_{j,k}$ are linearly independent for $j = 1, \ldots, K$ and that $\mathbf{h}_{j,k}^H \mathbf{h}_{j',k} \neq 0$ for all $j, j', j \neq j'$. Then if $\mathbf{u}_k$ is a beamforming vector on the Pareto boundary, there exist complex numbers $\{\xi_{jk}\}$ such that*

$$\mathbf{u}_k = \sum_{j=1}^{K} \xi_{jk} \mathbf{h}_{j,k}^H \tag{4.9}$$

and

$$\|\mathbf{u}_k\|^2 = 1, \quad p_k = P. \tag{4.10}$$

Proposition 2. *Maximizing* (4.7) *over $\mathbf{u}_k$ such that $\|\mathbf{u}_k\| = 1$ for any selection of $\{\alpha_{kj}\}$ yields a beamforming vector choice that satisfies proposition 1.*

Proof. Maximizing (4.7) can be reformulated as solving a generalized eigenvalue problem. The unique solution such that the norm constraint is satisfied (up to a scalar rotation) is given by

$$\mathbf{u}_k = \frac{\left(\frac{1}{\rho}\mathbf{I} + \sum_{j \neq k} \alpha_{kj} \mathbf{h}_{j,k}^H \mathbf{h}_{j,k}\right)^{-1} \mathbf{h}_{k,k}^H}{\| \left(\frac{1}{\rho}\mathbf{I} + \sum_{j \neq k} \alpha_{kj} \mathbf{h}_{j,k}^H \mathbf{h}_{j,k}\right)^{-1} \mathbf{h}_{k,k}^H \|} \tag{4.11}$$

[2] with appropriate changes to match our notation

Such a $\mathbf{u}_k$ belongs to the set specified in Proposition 1. □

Note 1. *The converse however is not true. The authors in [49] note that wlog one of the ξ_{jk} may be constrained to be real-valued. A beamformer given by* (4.11) *imposes the following constraint on the ξ_{jk}*

$$\xi_{kk}\|\mathbf{h}_{k,k}\|^2 + \sum_{j\neq k} \xi_{jk}\mathbf{h}_{k,k}\mathbf{h}_{j,k}^H \in \mathbb{R}^+. \tag{4.12}$$

For $K > 2$, one can easily find a set of ξ_{jk} which do not satisfy this constraint, even if the constraint that one of them be real-valued is imposed. As shown below, this is not true when $K = 2$. However, as Theorem 3 states, any point on the Pareto boundary, as long as $K \leq N_t$, can indeed be achieved by maximizing a VSINR for appropriate choices of the α_{jk}'s.

Corollary 1 ([49])**.** *Any point on the Pareto boundary for $K = 2$ is achievable with the beamforming strategy:*

$$\mathbf{u}_k(\lambda_k) = \frac{\lambda_k \mathbf{u}_k^{NE} + (1-\lambda_k)\mathbf{u}_k^{ZF}}{\|\lambda_k \mathbf{u}_k^{NE} + (1-\lambda_k)\mathbf{u}_k^{ZF}\|}, k = 1, 2 \tag{4.13}$$

for some set of real-valued *parameters λ_k, $0 \leq \lambda_k \leq 1$, $k = 1, 2$, where*

$$\mathbf{u}_k^{NE} = \frac{\mathbf{h}_{k,k}^H}{\|\mathbf{h}_{k,k}\|} \quad \textit{and} \quad \mathbf{u}_k^{ZF} = \frac{\Pi_{\bar{k}k}^{\perp}\mathbf{h}_{k,k}^H}{\|\Pi_{\bar{k}k}^{\perp}\mathbf{h}_{k,k}^H\|} \tag{4.14}$$

are the Nash Equilibrium (NE) or Maximum Ratio Transmission (MRT) and ZF solutions, respectively. $\Pi_{\bar{k}k}^{\perp}$ is the projection matrix onto the null space of $\mathbf{h}_{\bar{k},k}$, $\Pi_{\bar{k}k}^{\perp} = \mathbf{I}_{N_t} - \frac{\mathbf{h}_{\bar{k},k}\mathbf{h}_{\bar{k},k}^H}{\|\mathbf{h}_{\bar{k},k}\|^2}$.

Theorem 3. *Any point on the Pareto boundary may be attained by solving the VSINR optimization problem, as given in* (4.8)*, for an appropriate choice of $\alpha_{kj} \in \mathbb{R}_+, j, k = 1, \ldots, K$, provided $K \leq N_t$.*

Proof. Details are given in Appendix 4.A for the two-link case, $N_t \geq 2$, by relating the solution to Corollary 1. Appendix 4.B generalizes this result for $N_t \geq K \geq 2$. □

4.4.2 Achieving a particular point on the Pareto Boundary for the Two Link Case

In general, a point on the Pareto boundary is obtained by maximizing a weighted sum rate for some set of weighting coefficients. Solving this in general requires either sharing the full CSIT or devising some iterative algorithm

whereby the transmitters exchange values of the Lagrange coefficients of the dual problem. For a specific choice of α_{kj}, in the two link case, attaining a point on the Pareto boundary is guaranteed in one shot.

Theorem 4. *The rate pair obtained by beamforming using the solutions to problem* (4.8) *with* $\alpha_{12} = \alpha_{21} = 1$ *lies on the Pareto boundary of the rate region.*

Proof. Appendix 4.C proves this. □

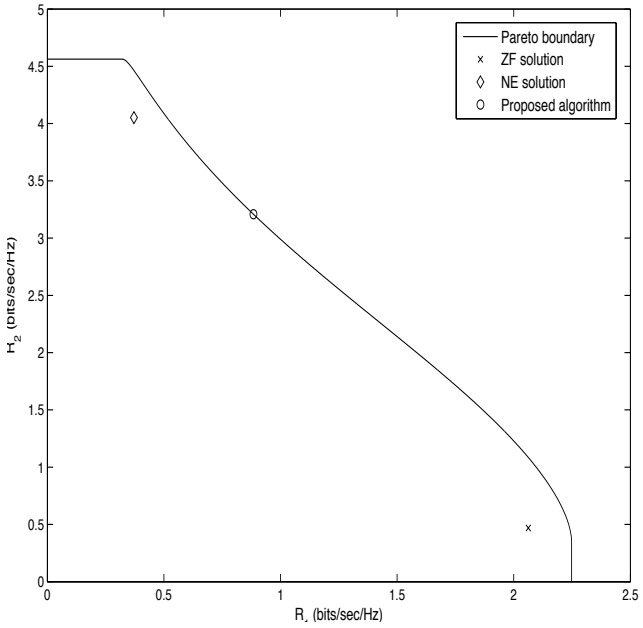


Figure 4.2: Pareto rate boundary, MRT, ZF and $\alpha_{12} = \alpha_{21} = 1$ points for a channel instance sampled from a channel with i.i.d. $\mathcal{CN}(0, 1)$ coefficients, $N_t = 3$, $K = 2$.

This is illustrated in Figure 4.2, which also shows the rate pairs corresponding to the NE (or MRT) and ZF solutions, which correspond to the most selfish and the most altruistic strategies, respectively, and whose beamforming vectors are given by (4.14) above.

4.5 Proposed Algorithm

The performance of the set of precoding vectors obtained in a distributed way by maximizing a virtual SINR at each of the transmitters will depend on the α_{ij}'s selected. Motivated by Theorem 4 above, we propose to set all of these to 1. Thus at base station k:

$$\mathbf{u}_k = \arg \max_{\|\mathbf{u}\|^2=1} \frac{|\mathbf{h}_{k,k}\mathbf{u}|^2}{\frac{1}{\rho} + \sum_{j \neq k} |\mathbf{h}_{j,k}\mathbf{u}|^2}. \tag{4.15}$$

4.5.1 Two-link case: Further Analysis

Theorem 4 states that the rate pair achieved by setting $\alpha_{12} = \alpha_{21} = 1$ lies on the Pareto boundary of the rate region, but says nothing about the achieved sum rate. An idea of the performance is gained by analyzing the SINR's at low and high SNR, and comparing them with the optimal strategies at those extreme regimes, which are known to be the NE solution and the ZF solution, respectively.
The resulting SINR's from applying our algorithm in the two-link case are given by, where the parameters involved are defined in Appendix 4.A:

$$\gamma_k = \rho \frac{(a_k + b_k(1+c_k))^2}{a_k + b_k(1+c_k)^2} \frac{a_{\bar{k}} + b_{\bar{k}}(1+c_{\bar{k}})^2}{(1+c_{\bar{k}})(a_{\bar{k}} + b_{\bar{k}}(1+c_{\bar{k}}))} \tag{4.16}$$

At the ZF solution:

$$\gamma_k^{ZF} = \rho b_k \tag{4.17}$$

At the NE solution:

$$\gamma_k^{NE} = \rho \frac{a_k + b_k}{1 + \frac{a_{\bar{k}}}{a_{\bar{k}} + b_{\bar{k}}} c_{\bar{i}}} \tag{4.18}$$

At low SNR, $1 + c_k \approx 1$, and both γ_k and γ_k^{NE} may be approximated by $\rho(a_k + b_k)$. On the other hand, at high SNR, $\gamma_k \approx \rho b_k = \gamma_k^{ZF}$. Thus, at both extremes, this approach performs as good as the best out of these two schemes.

4.6 Numerical Results

Figure 4.3 illustrates the performance of our approach in a more realistic scenario, the parameters of which are specified in Table 4.1. User locations in a cell follow a uniform distribution.

Parameter	Value
Path loss model	Cost-231, small/medium city
Number of cells	3, 7
Cell radius	1000 m
Transmit antenna gain, G_{tx}	16 dB
Shadowing mean	0 dB
Shadowing variance	10 dB
Receive antenna gain, G_{rx}	6 dB
Edge SNR	0-15 dB

Table 4.1: Simulation setup parameters

We show the average sum rates achieved and compare them with the selfish scheme corresponding to MRT, and the altruistic scheme corresponding to minimizing the total interference caused to other users. Our scheme clearly surpasses both. As the figure shows the results for the 7-cell case, and $N_t < 7$ for all simulations, interference caused can never be eliminated completely and eventually the rates would saturate. However, for the cell edge SNR range considered the performance gains are still quite significant. Note that, when $N_t \geq K$, interference can be eliminated completely and at high SNR the rates achieved with our scheme and the interference minimizing solution would differ by at most a constant (in favor of our scheme).

4.6.1 Comparison with Full CSIT Case

We would like to have an idea of the loss due to the distributed nature of our algorithm. The simplest way to define loss is in terms of total power consumption: in our algorithm, all transmitters always use full power. Alternatively, if full CSIT was available at all transmitters or these were allowed to share channel information, then it may be possible to achieve the same rates at all users with a lower total transmit power.

Figure 4.4 illustrates the power loss due to the distributed nature of our scheme, again for the 7-cell case. More precisely, for different number of antennas, the minimum power needed, under the individual power constraints at each base station, to achieve the same rates as those achieved by our distributed algorithm is computed and the figure illustrates the difference between this power and the power consumed by our scheme as a percentage of the total power available across the system. As the number of antennas at each transmitter increases, the difference decreases: this is because with more degrees of freedom afforded by the higher number of antennas, even

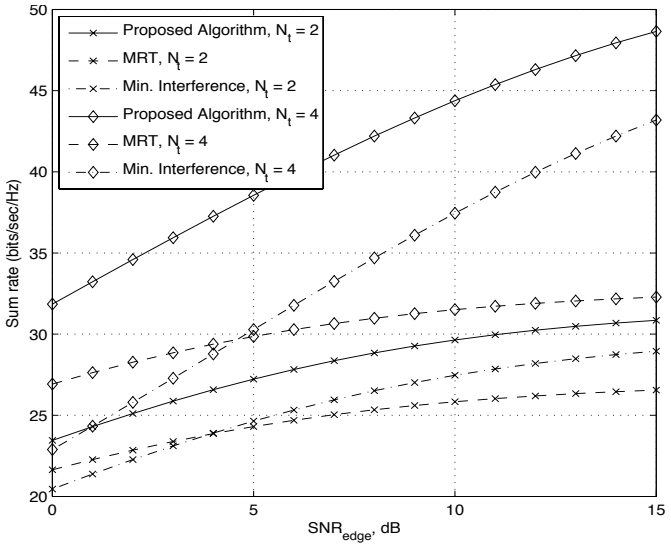


Figure 4.3: Sum rates vs. cell-edge SNR for $N_t = 2, 4$ for the 7-cell case.

in our distributed algorithm, more efficient power use is done automatically thereby reducing the benefit of centralized knowledge, at least for achieving the same rates as our algorithm. From our other simulations, not shown here, for the 3-cell case, for $N_t \geq 3$, this difference is almost negligible, which leads us to conjecture that we are quite close, if not on, the Pareto boundary in this case, for most channel instances.

4.6.2 Comparison to Iterative Schemes

A final evaluation of the simple scheme proposed is to compare it with one of the iterative schemes in the literature aimed at maximizing some objective function. We thus compare it to the asynchronous distributed pricing (ADP) scheme [9]: put briefly, this is an iterative scheme which allows the distributed optimization of a sum utility; it relies on each receiver announces an 'interference price', based upon which the different transmitters update their beamforming strategies so as to maximize a local approximation of the sum utility.

Figure 4.5 compares the average sum rates obtained by the VSINR optimization to those obtained by applying the ADP algorithm (with utility function the sum rate) from [9] with i) random initial beamforming and ii) initial beamforming based on maximizing VSINRs at each transmitter, for

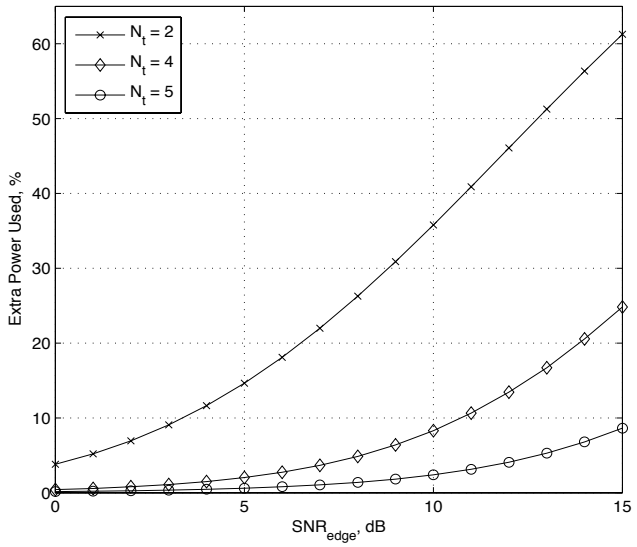
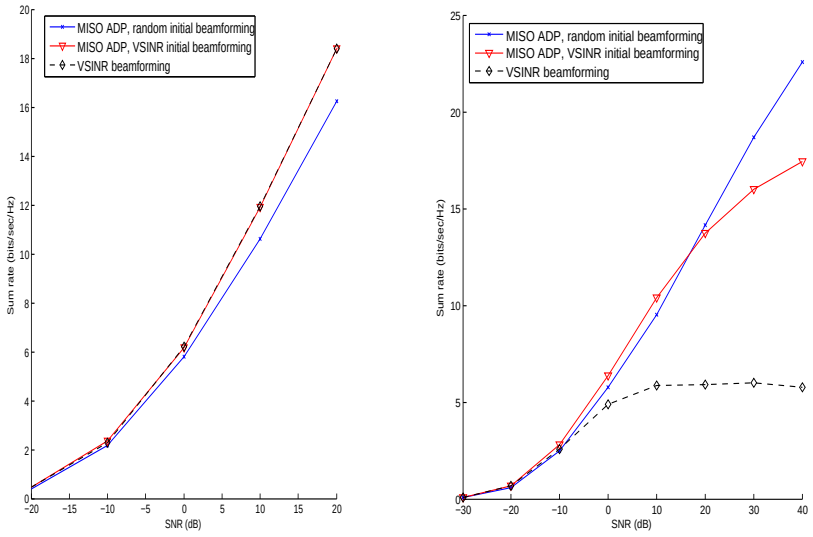


Figure 4.4: Extra power consumed due to the distributed nature of our scheme vs. cell-edge SNR for $N_t = 2, 4, 5$ for the 7-cell case.

i.i.d. $\mathcal{CN}(0,1)$ channel coefficients, for $N_t = 2$ and $K = 2$ and $K = 3$. For $K = 2$, there is quite little to gain (on average at least) from applying ADP, and starting at the VSINR beamforming outperforms randomly generating the initial beamforming. For $K = 3$, there are too few degrees of freedom at each transmitter so that the VSINR beamforming performs much worse than the iterative algorithm, and is not always a good choice for initial beamforming. In fact, the iterative algorithm tends to turn off one of the transmitters at higher SNR.

4.7 Conclusion

In this Chapter, the MISO interference channel with local CSIT at each transmitter was considered. Maximizing a virtual SINR was proposed as a beamforming strategy which uses only the local information and avoids iterative algorithms which require information exchange between different transmitters. The validity of such a scheme was discussed by relating it to existing literature on the subject and its Pareto optimality for a two link case was shown. In general, if the objective is sum rate maximization, this scheme performs quite well if there are at least as many antennas at each



(a) $K = 2$ (b) $K = 3$

Figure 4.5: Average sum rate vs. SNR for i.i.d. Rayleigh channels and $N_t = 2$.

transmitter as there are active links. When this is no longer the case, such a scheme, while still performing better than MRT or interference minimizing transmit strategies, performs much worse than the optimal scheme. In such a case the sum rate is maximized by having a fraction of the users transmit at less than full power. In fact at high SNR, some links would simply turn off. The degrees of freedom afforded by a scheme that relies on beamforming alone is in fact $\min(N_t, K)$ so that if $N_t < K$, different interference avoidance schemes may be needed to be able to serve more concurrent users: interference alignment in the time or frequency domain (for a frequency selective channel) [52], or by restricting signals to be real instead of complex-valued [53] for example would be needed to achieve higher degrees of freedom [46]. Attaining these degrees of freedom does not seem to be possible with local CSI alone [52], not without resorting to iterative schemes or perhaps by devising some distributed power control scheme.

4.A Proof of Theorem 3

To simplify expressions, in what follows $\bar{k}$ is used to denote the 'other' user/base station index (i.e., $\bar{k} = \mod(k,2)+1$, for $k \in \{1,2\}$).
We show that the rate region achieved by the parametrization given in Corrolary 1 of the two-link Pareto boundary can also be achieved by varying the α's in their feasible region ($\mathbb{R}_+$) and maximizing the corresponding virtual SINRs.
Maximizing the virtual SINR of (4.7):

$$\mathbf{u}_k = \arg \max_{\|\mathbf{u}\|^2=1} \frac{|\mathbf{h}_{k,k}\mathbf{u}|^2}{1/\rho + \alpha_{k\bar{k}}|\mathbf{h}_{\bar{k},k}\mathbf{u}|^2}. \tag{4.19}$$

Proposition 3. *The solution of problem* (4.19) *can be written as:*

$$\mathbf{u}_k = \sqrt{\zeta_k}\frac{\Pi_{\bar{k}k}\mathbf{h}_{k,k}^H}{\|\Pi_{\bar{k}k}\mathbf{h}_{k,k}^H\|} + \sqrt{1-\zeta_k}\frac{\Pi_{\bar{k}k}^{\perp}\mathbf{h}_{k,k}^H}{\|\Pi_{\bar{k}k}^{\perp}\mathbf{h}_{k,k}^H\|} \tag{4.20}$$

where $0 \le \zeta_k \le 1, k = 1,2$ *and* $\Pi_{\bar{k}k}$ *and* $\Pi_{\bar{k}k}^{\perp}$ *are the projection matrices onto* $\mathbf{h}_{\bar{k},k}$ *and its the null space, respectively:*

$$\Pi_{\bar{k}k} = \frac{\mathbf{h}_{\bar{k},k}^H\mathbf{h}_{\bar{k},k}}{\|\mathbf{h}_{\bar{k},k}\|^2}, \quad \Pi_{\bar{k}k}^{\perp} = \mathbf{I}_{N_t} - \Pi_{\bar{k}k}. \tag{4.21}$$

Proof. Similar to that of Proposition 1 in [48]. □

Define:

$$\begin{aligned} a_k &= \|\Pi_{\bar{k}k}\mathbf{h}_{k,k}^H\|^2 \\ b_k &= \|\Pi_{\bar{k}k}^{\perp}\mathbf{h}_{k,k}^H\|^2 \\ c_k &= \rho\|\mathbf{h}_{\bar{k},k}\|^2 \end{aligned} \tag{4.22}$$

Proposition 4. ζ_k *that solves* (4.19) *is given by:*

$$\zeta_k = \frac{a_k}{a_k + b_k(1+\alpha_{k\bar{k}}c_k)^2} \tag{4.23}$$

Proof. With $\mathbf{u}_k$ as in (4.20),

$$\begin{aligned} |\mathbf{h}_{k,k}\mathbf{u}_k|^2 &= \left|\sqrt{\zeta_k}\frac{\mathbf{h}_{k,k}\Pi_{\bar{k}k}\mathbf{h}_{k,k}^H}{\|\Pi_{\bar{k}k}\mathbf{h}_{k,k}^H\|} + \sqrt{1-\zeta_k}\frac{\mathbf{h}_{k,k}\Pi_{\bar{k}k}^{\perp}\mathbf{h}_{k,k}^H}{\|\Pi_{\bar{k}k}^{\perp}\mathbf{h}_{k,k}^H\|}\right|^2 \\ &= \left(\sqrt{a_k\zeta_k} + \sqrt{b_k(1-\zeta_k)}\right)^2 \end{aligned} \tag{4.24}$$

Similarly,

$$|\mathbf{h}_{\bar{k},k}\mathbf{u}_k|^2 = \zeta_k \|\mathbf{h}_{\bar{k},k}\|^2 \tag{4.25}$$

Thus the virtual SINR is equal to:

$$\gamma_k^{virtual} = \frac{\rho\left(\sqrt{a_k\zeta_k} + \sqrt{b_k(1-\zeta_k)}\right)^2}{1 + \alpha_{k\bar{k}}\zeta_k c_k} \tag{4.26}$$

One can easily verify that this ratio is maximized for the value specified in (4.23). □

Proposition 5. *In terms of the NE and ZF beamforming vectors,* (4.20) *can be rewritten as:*

$$\mathbf{u}_k = \frac{\lambda_k \mathbf{u}_k^{NE} + (1-\lambda_k)\mathbf{u}_k^{ZF}}{\|\lambda_k \mathbf{u}_k^{NE} + (1-\lambda_k)\mathbf{u}_k^{ZF}\|} \tag{4.27}$$

where

$$\lambda_k = \frac{1}{\sqrt{\frac{a_k}{a_k+b_k}\left(\frac{1}{\zeta_k}-1\right)} + \left(1 - \sqrt{\frac{b_k}{a_k+b_k}}\right)} \tag{4.28}$$

Proof. $\mathbf{w}_k$, as expressed by (4.20), can be rewritten in terms of $\mathbf{w}_k^{NE}$ and $\mathbf{w}_k^{ZF}$ as:

$$\mathbf{u}_k = \sqrt{\zeta_k}\sqrt{\frac{a_k+b_k}{a_k}}\mathbf{u}_k^{NE} + \left[\sqrt{1-\zeta_k} - \sqrt{\zeta_k}\sqrt{\frac{b_k}{a_k}}\right]\mathbf{u}_k^{ZF} \tag{4.29}$$

We need to show that this is in fact of the form given in (4.27), i.e. that the following equalities hold for some λ_k:

$$\frac{\lambda_k^2}{\lambda_k^2\frac{a_k}{a_k+b_k} + \left(1-\lambda_k+\lambda_k\sqrt{\frac{b_k}{a_k+b_k}}\right)^2} = \zeta_k\frac{a_k+b_k}{a_k}, \text{and}$$

$$\frac{1-\lambda_k}{\sqrt{\lambda_k^2\frac{a_k}{a_k+b_k} + \left(1-\lambda_k+\lambda_k\sqrt{\frac{b_k}{a_k+b_k}}\right)^2}} = \sqrt{1-\zeta_k} - \sqrt{\zeta_k}\sqrt{\frac{b_k}{a_k}}$$

where we replaced the denominator of (4.27) by its value in terms of the parameters defined in (4.22).

One can verify that λ_k as given by (4.28) above satisfies both these equations. □

Combining propositions 4 and 5, we complete the proof. Plugging (4.23) into (4.28), we get:

$$\lambda_k = \frac{1}{\alpha_{k\bar{k}} c_k \sqrt{\frac{b_k}{a_k + b_k}} + 1} \tag{4.30}$$

Clearly this is a decreasing function of $\alpha_{k\bar{k}}$. It is easy to check that for $\alpha_{k\bar{k}} = 0$, $\lambda_k = 1$ and that as $\alpha_{k\bar{k}} \to \infty$, $\lambda_k \to 0$.

4.B Minimizing total transmit power subject to SINR and per transmitter power constraints

Consider the problem of minimizing total transmit power such that γ_k^t, $k = 1, \ldots, K$ are achieved at each of the users, while meeting the individual power constraints at each transmitter.

$$\begin{aligned} &\text{minimize } \alpha \sum_{k=1}^{K} P \\ &\text{subject to } \mathbf{w}_k^H \mathbf{w}_k \leq \alpha P, \quad k = 1, \ldots, K \\ &\qquad \frac{|\mathbf{h}_{k,k} \mathbf{w}_k|^2}{\sigma^2 + \sum_{j \neq k} |\mathbf{h}_{k,j} \mathbf{w}_j|^2} \geq \gamma_k^t, \quad k = 1, \ldots, K \end{aligned} \tag{4.31}$$

This can easily be shown to be equivalent to a convex optimization problem (see [51, 54] for example).

The corresponding Lagrangian is equal to

$$\alpha \sum_{k=1}^{K} (1 - \lambda_k) P + \sum_{k=1}^{K} \mathbf{w}_k^H \left[\lambda_k \mathbf{I} + \sum_{j \neq k} \mu_j \mathbf{h}_{j,k}^H \mathbf{h}_{j,k} - \mu_k \frac{1}{\gamma_k^t} \mathbf{h}_{k,k}^H \mathbf{h}_{k,k} \right] \mathbf{w}_k + \sum_{k=1}^{K} \mu_k \sigma^2, \tag{4.32}$$

where $\lambda_k \geq 0$ denotes the Lagrange coefficient associated with power constraint k, and $\mu_k \geq 0$ denotes the Lagrange coefficient associated with SINR constraint k.

The dual problem is that of maximizing

$$\sigma^2 \sum_{k=1}^{K} \mu_k \tag{4.33}$$

subject to

$$\left[\lambda_k \mathbf{I} + \sum_{j \neq k} \mu_j \mathbf{h}_{j,k}^H \mathbf{h}_{j,k} - \frac{\mu_k}{\gamma_k^t} \mathbf{h}_{k,k}^H \mathbf{h}_{k,k}\right] \tag{4.34}$$

being positive semi-definite (PSD), and $\sum_{k=1}^{K}(1-\lambda_k)P = 0$. In other words, the dual problem is equivalent to maximizing over $\hat{\mathbf{w}}_k$, $\mu_k \geq 0$ and $\lambda_k \geq 0$

$$\sigma^2 \sum_{k=1}^{K} \mu_k \tag{4.35}$$

such that

$$\frac{\mu_k |\mathbf{h}_{k,k} \hat{\mathbf{w}}_k|^2}{\lambda_k \|\hat{\mathbf{w}}_k\|^2 + \sum_{j \neq k} \mu_j |\mathbf{h}_{j,k} \hat{\mathbf{w}}_k|^2} \leq \gamma_k^t \tag{4.36}$$

and $\sum_{k=1}^{K}(1 - \lambda_k)P = 0$.

For the above constraint to hold for any $\hat{\mathbf{w}}_k$, it must hold for the one that maximizes its right hand side. Moreover, one could show that at the optimum of the dual, the constraints (4.36) must be met with equality. Thus optimal $\hat{\mathbf{u}}_k$ could be the MMSE filters given by

$$\left((1 + \lambda_k)\, \mathbf{I} + \sum_{j \neq k} \mu_j \mathbf{h}_{j,k}^H \mathbf{h}_{j,k}\right)^{-1} \mathbf{h}_{k,k}^H. \tag{4.37}$$

Going back to the primal problem, from the KKT conditions we have that

$$\left[\lambda_k \mathbf{I} + \sum_{j \neq k} \mu_j \mathbf{h}_{j,k}^H \mathbf{h}_{j,k} - \mu_k \frac{1}{\gamma_k^t} \mathbf{h}_{k,k}^H \mathbf{h}_{k,k}\right] \mathbf{w}_k = \mathbf{0}. \tag{4.38}$$

In other words,

$$\mathbf{w}_k = \frac{\mu_k \mathbf{h}_{k,k} \mathbf{w}_k}{\gamma_k^t} \left[\lambda_k \mathbf{I} + \sum_{j \neq k} \mu_j \mathbf{h}_{j,k}^H \mathbf{h}_{j,k}\right]^{-1} \mathbf{h}_{k,k}^H, \tag{4.39}$$

and is therefore, up to a scalar, equal to $\hat{\mathbf{w}}_k$.

Thus, any solution for the above power minimizaton problem subject to SINR constraints is up to a scalar the solution of an equivalent uplink SINR maximization problem as expressed by the left-hand side of (4.36). This

includes the SINR values that lie on the Pareto boundary. Since we know that in this case the optimal $\mathbf{u}_k$ have norm 1 and therefore the optimal $\mathbf{w}_k$ will have norm P, these will indeed be the solutions of maximizing a virtual SINR as given by (4.7) such that

$$\frac{\lambda_k}{c_k} = \frac{1}{\rho} \text{ and } \frac{\mu_j}{c_k} = \alpha_{kj}, \quad j \neq k \tag{4.40}$$

for some $c_k > 0$. Moreover, since the duality gap is zero and the optimal $\alpha = 1$,

$$K\rho = \sum_{k=1}^{K} \mu_k. \tag{4.41}$$

4.C Proof of Theorem 4 for two links

For any $\mathbf{u}_1, \mathbf{u}_2$ of the form (4.20), one can show that SINRs are given by:

$$\gamma_k = \rho \frac{(\sqrt{a_k \zeta_k} + \sqrt{b_k(1-\zeta_k)})^2}{1 + \zeta_{\bar{i}} c_{\bar{i}}} \tag{4.42}$$

A rate pair is Pareto optimal if one cannot increase one of the rates without necessarily decreasing the other. Note that any point on Pareto boundary has to have the corresponding (ζ_1, ζ_2) pair in the region defined by $\zeta_k \in \left[0, \frac{a_k}{a_k + b_k}\right], k = 1, 2$: this is so since for higher ζ_k it is always possible to achieve higher useful signal at user k while causing less interference at user $\bar{i}$ (cf. (4.42)).
Denote by $\gamma_k^{1,1}$ the SINR values achieved by setting $\alpha_{12} = \alpha_{21} = 1$. To show that the corresponding rates belong to the Pareto boundary, we solve the following optimization problem:

$$\begin{aligned} &\text{maximize } \gamma_1 \\ &\text{such that } 0 \leq \zeta_k \leq \frac{a_k}{a_k + b_k}, \quad k = 1, 2 \\ &\qquad \gamma_2 \geq \gamma_2^{1,1} \end{aligned} \tag{4.43}$$

This can be formalized as the following convex optimization problem:

$$\begin{aligned} &\text{minimize } -t \\ &\text{such that } 0 \le \zeta_k \le \frac{a_k}{a_k+b_k}, \quad i=1,2 \\ &\qquad t \ge 0 \\ &\qquad \gamma_2^{1,1}\left(1+\zeta_1 c_1\right) - \rho\left(\sqrt{a_2\zeta_2}+\sqrt{b_2(1-\zeta_2)}\right)^2 \le 0 \\ &\qquad t\left(1+\zeta_2 c_2\right) - \rho\left(\sqrt{a_1\zeta_1}+\sqrt{b_1(1-\zeta_1)}\right)^2 \le 0 \end{aligned} \tag{4.44}$$

This problem is strictly feasible and consequently Slater's condition for strong duality holds [55].

Let $\mu_i, i=1,\ldots,3$ be the Langrange multipliers associated with the positivity constraints, $\xi_k, k=1,2$ the Lagrange multipliers associated with the upper bounds on the ζ_k, and $\lambda_k, k=1,2$ the Lagrange multipliers associated with the SINR constraints, the corresponding Karush-Kuhn-Tucker (KKT) conditions [55] are given by:

$$\begin{aligned} &-1-\mu_3+\lambda_2(1+\zeta_2 c_2)=0 \\ &-\mu_1+\xi_1+\lambda_1\gamma_2^{1,1}c_1 = \lambda_2\rho\left(a_1-b_1+\sqrt{a_1b_1}\left(\sqrt{\frac{1-\zeta_1}{\zeta_1}}-\sqrt{\frac{\zeta_1}{1-\zeta_1}}\right)\right) \\ &-\mu_2+\xi_2+\lambda_2 t c_2 = \lambda_1\rho\left(a_2-b_2+\sqrt{a_2b_2}\left(\sqrt{\frac{1-\zeta_2}{\zeta_2}}-\sqrt{\frac{\zeta_2}{1-\zeta_2}}\right)\right) \\ &\quad \mu_1,\mu_2,\mu_3,\lambda_1,\lambda_2,\xi_1,\xi_2 \ge 0 \\ &\quad \mu_k\zeta_k=0, \quad \xi_k\left[\zeta_k-\frac{a_k}{a_k+b_k}\right]=0, k=1,2, \mu_3 t=0 \\ &\quad \lambda_1\left[\gamma_2^{1,1}\left(1+\zeta_1 c_1\right)-\rho\left(\sqrt{a_2\zeta_2}+\sqrt{b_2(1-\zeta_2)}\right)^2\right]=0 \\ &\quad \lambda_2\left[t\left(1+\zeta_2 c_2\right)-\rho\left(\sqrt{a_1\zeta_1}+\sqrt{b_1(1-\zeta_1)}\right)^2\right]=0 \end{aligned} \tag{4.45}$$

For $\zeta_k, k=1,2$ given by (4.23), with $\alpha_{k\bar{k}}=1$, one can verify that these values, together with the values of t and the Lagrange multipliers given in equation (4.46) below provide a consistent solution of the KKT conditions. This guarantees optimality. Noting that the optimal value of problem (4.43)

is indeed that achieved by our algorithm completes the proof.

$$\begin{aligned}
&\mu_1 = \mu_2 = \mu_3 = 0, \xi_1 = \xi_2 = 0, \\
&\lambda_2 = \frac{1}{1+\zeta_2 c_2}, \\
&\lambda_1 = \frac{1}{1+\zeta_2 c_2} \frac{(a_1 + b_1(1+c_1))^2 (a_2 + b_2(1+c_2)^2)}{(a_2 + b_2(1+c_2))^2 (a_1 + b_1(1+c_1)^2)}, \\
&t = \gamma_1^{1,1}.
\end{aligned} \tag{4.46}$$

Chapter 5

Cooperative Network MIMO with imperfect CSIT sharing

5.1 Introduction

In this and the following chapter, we move away from the interference channel to a channel where joint MIMO precoding across distant transmitters to cooperatively serve the set of mobile users is possible, i.e. user data is shared by several transmitters. Such a scheme, also referred to as so-called network MIMO, was dealt with in [10], [11] for example. This situation is illustrated in Fig. 5.1.

In the downlink scenario, implementation of network MIMO requires both data and CSI to be shared by the transmitters, or to be fed back to some central processor which designs the transmission and informs the base stations of which precoding solutions shall be used. Data and CSI sharing comes, however, at a cost of delay, feedback and backhaul resources. One way to reduce the delays or have a more efficient use of the backhaul is to reduce how much CSI needs to be shared. Thus, assuming that the user data is conveniently routed to all concerned transmitters, but assuming the transmitters obtain local CSIT only, we obtain a MIMO channel with a novel CSIT model where the different transmitters do not have the same vision of the downlink channel. To the best of our knowledge this problem has not yet been investigated, despite its strong relevance in practical situations:

in fact, previous work on multi-transmitter MIMO precoding assumes that either perfect [10], [11] or limited CSIT [56] is available and shared among all transmitters. Instances of such a new distributed CSIT situation are described below.

A first case occurs in the context of a reciprocal system, in which CSIT is acquired from uplink transmissions. Transmitters thus only learn the channels between themselves and the users, but not those between the other transmitters and the users. As mentioned before, full CSI sharing may be too costly, but perhaps some statistical information about the unknown links may be affordable. This is essentially the same level of CSI assumed in Chapter 4 in the context of the MISO interference channel. In fact our initial investigation in the context of distributed CSIT is for this setup and extends the VSINR idea to the joint transmission case. This is presented in Section 5.3 below.

In a second instance of a feedback model, we consider that the different receivers broadcast the CSI estimates (which they have obtained over the downlink) over the air, and each transmitter attempts to decode the said information independently (see the framework proposed in [57]). Thus, depending on the distance between a receiver and each base station, a given base station may decode successfully the totality or part of the CSI fed back. In another instance of partially shared CSIT, a station may decode completely the CSI feedback of a subset of users, and forward subquantized versions of it to neighboring bases. Note that both approaches lead to i) a reduction in CSI exchange, and ii) different representations of the same channel at the different transmitters.

Ideally, the different transmitters would have to conciliate their views in order to design a consistent set of precoding vectors that will maximize a performance metric at the user side, despite possible differences in their estimated CSIT. This problem can be categorized as a so-called"team-decision problem" or a decentralized statistical decision making problem [19, 58]. More generally, in such problems, i) each decision maker (here, transmitter) has different but correlated information about the underlying uncertainty in the channel state, and ii) the transmitters need to act in a coordinated manner in order to realize the common payoff (which could be for example, the average sum rate). Such a scheme offers the possibility for reduction in communication requirements, at the expense of performance reduction [59], yet is expected to perform better than a framework where the decision makers simply ignore the differences in their view of the channel state.

The contributions in this chapter are as follows:

- For the TDD CSI acquisition scheme, we propose a suboptimal solution for the MISO case based on the virtual SINR scheme introduced in the previous chapter. We refer to this approach as a *layered* virtual SINR scheme, as it relies on viewing the broadcast channel with distributed CSIT as a superposition of interference channels.
- The more general team decision problem corresponding to cooperative network MIMO with imperfect CSIT sharing is presented.
- We propose a new distributed CSIT framework for quantized feedback, which allows a simpler formulation of the above team decision problem.
- We investigate the best transmit strategy (here beamforming) to be adopted by the transmitters in this framework, detailing it for the 2×2 case.
- We study via Monte Carlo simulations the performance gap to a scenario with centralized, yet still inaccurate, CSIT.

5.2 System Model

Consider a set of N base stations communicating with K mobile stations. We assume the transmitters have $N_t \geq 1$ antennas each, whereas receivers have a single antenna each.

We use the channel notation introduced in Chapter 2, so that $\mathbf{h}_{k,j} \in \mathbb{C}^{1\times N_t}$ corresponds to the channel between transmitter j and receiver k and $\mathbf{h}_k \stackrel{\triangle}{=} [\mathbf{h}_{k,1} \ \ldots \ \mathbf{h}_{k,N}] \in \mathbb{C}^{1\times NN_t}$, receiver k's whole channel. $\mathbf{h}_{k,j} \in \mathcal{CN}\left(\mathbf{0}, \sigma_{kj}^2 \mathbf{I}_{N_t}\right)$.

The signal received at mobile station k is given by:

$$y_k = \mathbf{h}_k \mathbf{x} + n_k, \tag{5.1}$$

where $\mathbf{x} \in \mathbb{C}^{NN_t\times 1}$ is the concatenated transmit signal sent by all transmitters and $n_k \sim \mathcal{CN}(0, \sigma^2)$ is the noise at that receiver.

Multi-transmitter cooperative processing in the form of joint linear precoding with per-transmitter power constraints is adopted. Thus, $\mathbf{x}$ can be expressed as:

$$\mathbf{x} = \mathbf{W}\mathbf{s} = \sum_{k=1}^{K} \mathbf{w}_k s_k, \tag{5.2}$$

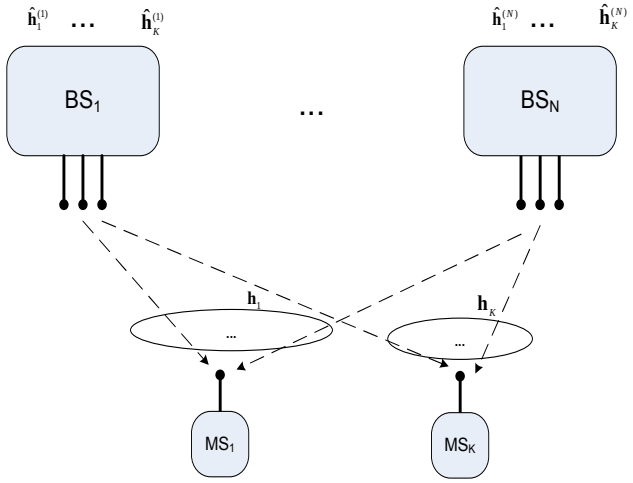


Figure 5.1: Cooperative MIMO channel with imperfect CSIT sharing setup, with N base stations, K mobile stations.

where $\mathbf{s} \in \mathbb{C}^{K\times 1}$ is the vector of transmit symbols, its entries are assumed to be independent and $\mathcal{CN}(0,1)$. The precoding matrix is $\mathbf{W} = [\mathbf{w}_1 \ \dots \mathbf{w}_K] \in \mathbf{C}^{NN_t\times K}$, where $\mathbf{w}_k = [\mathbf{w}_{k,1}; \dots; \mathbf{w}_{k,N}]$ is the beamforming vector corresponding to user k's symbol, $\mathbf{w}_{k,j} \in \mathbb{C}^{N_t\times 1}$ corresponding to Transmitter j's precoding. Defining $\mathbf{V}_j$ as $[\mathbf{w}_{1,j} \ \dots \ \mathbf{w}_{K,j}]$, i.e. as the precoding matrix used at transmitter j, $\mathbf{W}$ may be alternatively written as

$$\begin{bmatrix} \mathbf{V}_1 \\ \dots \\ \mathbf{V}_N \end{bmatrix}. \tag{5.3}$$

The signal transmitted by BS_j, $\mathbf{x}_j$, is thus

$$\mathbf{x}_j = \sum_{k=1}^{K} \mathbf{w}_{k,j} s_k = \sum_{k=1}^{K} \sqrt{p_{kj}}\mathbf{u}_{k,j} s_k, \tag{5.4}$$

where $\mathbf{w}_{k,j} = \sqrt{p_{kj}}\mathbf{u}_{k,j}$, $\|\mathbf{u}_{k,j}\| = 1$, and p_{kj} is the power allocated by transmitter j to serving user k.

Its power constraint is given by:

$$\|\mathbf{V}_j\|_{\mathrm{F}}^2 = \sum_{k=1}^{K} \|\mathbf{w}_{k,j}\|^2 = \sum_{k=1}^{K} p_{kj} \leq P, \ \forall j = 1, \ldots, N. \tag{5.5}$$

Finally, the rate achieved at user k is equal to

$$r_k = \log_2(1 + \gamma_k), \tag{5.6}$$

where the SINR γ_k is equal to

$$\gamma_k = \frac{|\sum_{j=1}^{N} \mathbf{h}_{k,j}\mathbf{w}_{k,j}|^2}{\sigma^2 + \sum_{\bar{k}=1,\ldots,K,\bar{k}\neq k} |\sum_{j=1}^{N} \mathbf{h}_{k,j}\mathbf{w}_{\bar{k},j}|^2}. \tag{5.7}$$

As stated in the introduction, the user information symbols $\mathbf{s}$ are routed to the multiple cooperating base stations. However, the CSI is not fully shared, and the design of the precoding will need to take this into consideration. Details of the distributed CSI knowledge follow.

5.2.1 Distributed CSIT

Previous work on multi-transmitter MIMO precoding assumes that either i) perfect CSIT is shared and available at all transmitters [10, 60], [11] or ii) limited CSIT is available, yet common to all transmitters e.g. [56]. Here we argue that a more general and realistic setup is one where the CSI feedback is designed in such a way that different transmitters end up with different representations of the channel: for instance it is likely that users which are relatively closer to some transmitters will be able to convey more precise information about their channel state to these in an FDD system, whereas in a TDD system in which CSI is acquired from estimating the channel on the uplink, the different transmitters will naturally only be able to learn the channels between themselves and the users but not those from the other transmitters to the users.

The benefit of such distributed CSIT schemes is a reduction in signaling with respect to the scheme where all transmitters must achieve the same state of CSI knowledge, hence potentially a greater scalability of multi-transmitter MIMO cooperation.

The distributed CSI model for quantized channel feedback is shown in Figure 5.2, where transmitter j's knowledge of $\mathbf{h}_k$ is represented by its quantized version $\hat{\mathbf{h}}_k^{(j)}$. For the TDD channel estimation case, one can think of $\hat{\mathbf{h}}_k^{(j)}$ as being equal to $[\mathbf{0}, \ldots, \mathbf{h}_{k,j} \ldots, \mathbf{0}]$: $\mathbf{h}_{k,j}$ is known perfectly at transmitter j but none of the other channel coefficients in $\mathbf{h}_k$, for all k (sufficiently close to that transmitter).

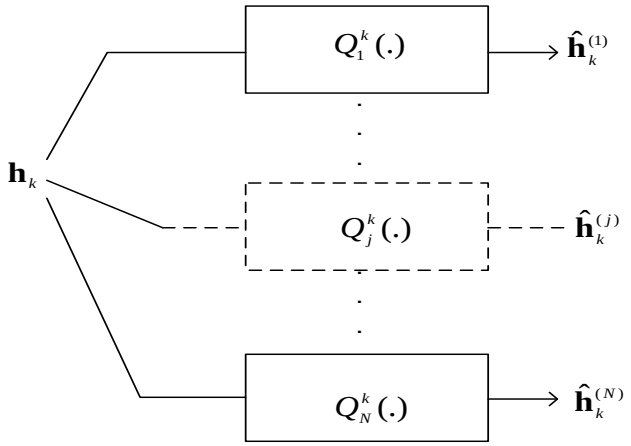


Figure 5.2: Distributed CSI model: each CSI vector is seen through a different quantization filter at each base station.

5.3 Joint precoding with local CSIT: Virtual SINR approach

In this section, we focus on a similar local CSIT setup as in the previous chapter. Also similarly to that chapter, we assume $N_t \geq 2$. The difference with that chapter is the fact that user data is known at all transmitters, thereby enabling joint transmission.

5.3.1 Multicell MIMO: Layered Virtual SINR

Given the local channel knowledge assumption, it is not possible for base stations to jointly design the whole beamforming matrix to carry symbols s_k, $k = 1, \ldots, K$ (cf. (5.4)). They can however still cooperate to transmit to the users in the system, as they have access to their data. If $N = K$, one way to think of the thus defined channel is as a superposition of K K-link interference channels, where in the lth interference channel, BS_j serves $MS_{\ \mathrm{mod}\,(j+l-2,K)+1}$; the difference from a regular superposition of ICs is that the data being transmitted to each user is effectively the same.

Guided by this view of the channel, we propose to use the concept of

virtual SINR maximization introduced in the previous chapter to design the beamforming vectors at each BS. What this means is that BS j will design K virtual SINR beamformers, one per user, i.e. it applies algorithm (4.15) from Chapter 4, K times, thereby obtaining $\mathbf{u}_{jk}$, $j \in \{1,\ldots,N\}, k \in \{1,\ldots,K\}$ defined in (5.4) as:

$$\mathbf{u}_{k,j} = \arg \max_{\|\mathbf{u}\|^2=1} \frac{|\mathbf{h}_{k,j}\mathbf{u}|^2}{\frac{\sigma^2}{p_{kj}} + \sum_{\bar{k}=1,\ldots,K,\bar{k}\neq k} |\mathbf{h}_{\bar{k},j}\mathbf{u}|^2}. \tag{5.8}$$

This is what we refer to as the *layered virtual SINR* (LVSINR) maximizing beamforming solution. Note that we replace P in (4.15) by p_{kj} in the above equation, the power allocated by transmitter j to user k. A heuristic method for how to determine the p_{kj} follows in the next subsection. We first give some intuitions to justify our approach.

Maximizing a virtual SINR effectively balances between the useful signal at the target user and the interference power generated at others. Moreover, one can do so while ensuring that, assuming synchronization in the system as we do here, the useful signal arriving at a given user from the different base stations does so constructively. In addition, such beamforming vectors satisfy the characterization given in [61] for Pareto optimality.

5.3.2 Power Allocation (PA)

Here we address the question of how to allocate the power between the different data streams at each base station. To find an answer, we first note the following:

Proposition 6. *Given the layered virtual SINR scheme, if $N_t \geq K$, full power at each transmitter should always be used, i.e.:*

$$\sum_{k=1}^{K} p_{kj} = P, \quad \textit{for } j = 1,\ldots,N. \tag{5.9}$$

Proof. The proof is quite similar to that in the MISO IC case (see the proof of Proposition 1 in [49]) and relies on the fact that given that $N_t \geq 2$, it is always possible to increase the rate of a given user by focusing more power orthogonally to the other users' channels. □

One would theoretically like to determine the PA so as to maximize some expected performance metric, the sum rate for example, given the channel knowledge at each transmitter and the adopted LVSINR transmission

scheme. This proves to be quite unwieldy however, and we instead resort to heuristics which, though suboptimal, have the added benefit of not requiring the statistical knowledge of the links between the other transmitters and the users.

1. *Statistical power splitting:*
   An intuitive approach is to split the power according to the following rule:
   $$p_{kj} = P\frac{\sigma_{kj}^2}{\sum_{\bar{k}=1}^{K}\sigma_{\bar{k}j}^2} \tag{5.10}$$
   The intuition behind this is that a transmitter should allocate more power to the users that will benefit more from it, i.e. the ones it has stronger links to. Thus if the strength of the link to a user is quite low, there will be little use of allocating it any power.

2. *Channel aware power splitting:*
   Pushing further the idea proposed in (5.10), another approach would be to split the power according to the instantaneous channel strength, as these are available locally, i.e.:
   $$p_{kj} = P\frac{\|\mathbf{h}_{k,j}\|^2}{\sum_{\bar{k}=1}^{K}\|\mathbf{h}_{\bar{k},j}\|^2} \tag{5.11}$$

Complexity-wise, using (5.10) implies power splitting to be recomputed when the channel statistics change which would normally occur at much slower rate than the instanteneous change in the channel, which (5.11) follows. Moreover, although the power splitting (5.11) is not designed to maximize any fairness criterion, it turns out that it coincidentally provides an interference fair solution for $N = K = 2$, where interference fairness is defined as a measure of the difference between the interference powers incurred at each user: in fact, the interference power at both users is the same, which leads us to the following lemma.

Lemma 3 (Interference fairness in the two link case)**.** *The power splitting* (5.11) *is strictly interference fair for* $N = K = 2$*.*

Proof. The interference power at user k is equal to (cf. the SINR expression in (5.7)):
$$I_k = \left|\sum_{j=1}^{2}\sqrt{p_{\bar{k}j}}\mathbf{h}_{k,j}\mathbf{w}_{\bar{k}j}\right|^2. \tag{5.12}$$

Using the PA in (5.11) and the corresponding layered virtual SINR beamforming vectors, the terms in the above summation may be written as:

$$\sqrt{p_{\bar{k}j}}\mathbf{h}_{kj}\mathbf{w}_{\bar{k}j} = \frac{\sqrt{P\beta_j}\cos\theta_j e^{\sqrt{-1}\angle \mathbf{h}_{k,j}\mathbf{h}_{\bar{k},j}^H}}{\sqrt{1+\frac{P\beta_j}{\sigma^2}\sin^2\theta_j\left(2+\frac{P\beta_j}{\sigma^2}\right)}}, \quad k=1,2, \tag{5.13}$$

where $\cos^2\theta_j = \frac{|\mathbf{h}_{k,j}\mathbf{h}_{\bar{k},j}^H|^2}{\|\mathbf{h}_{k,j}^H\|^2\|\mathbf{h}_{\bar{k},j}^H\|^2}$ and $\beta_j = \frac{\|\mathbf{h}_{k,j}^H\|^2\|\mathbf{h}_{\bar{k},j}^H\|^2}{\sum_{l=1}^2\|\mathbf{h}_{l,j}^H\|^2}$. Thus

$$\sqrt{p_{\bar{k}j}}\mathbf{h}_{k,j}\mathbf{w}_{\bar{k}j} = \sqrt{p_{kj}}\mathbf{h}_{\bar{k},j}\mathbf{w}_{kj}e^{\sqrt{-1}\pi} \tag{5.14}$$

and, as a result, we can verify that $I_1 = I_2$, thereby completing the proof. □

5.3.3 Numerical Results

We compare the performance of our layered virtual SINR approach (LVSINR in the figures) to a fully centralized scheme, namely joint zero-forcing (i.e. both base stations pool their antennas together and do downlink zero-forcing) with optimal PA (JZF-PA in the figures): note that this is discussed under per-antenna power constraints in [62,63], among others; one can show that in the case considered here as well, if the channels' pseudo-inverse is used for zero-forcing (see discussion in [63]), this is a convex optimization problem which is relatively easy to solve. We further compare our results to a fully distributed case, where no data is shared among the transmitters, which uses the VSINR-based algorithm from Chapter 4.

Figure 5.3 illustrates performance in terms of sum rate for a symmetric channel: we define a symmetric channel as one where the variances of the channel coefficients of the links between MS_k and BS_k, $k=1,2$ (direct links) are equal and fixed at 1, whereas the variances of the channel coefficients of the links between MS_k and $BS_{\bar{k}}$, for $k=1,2$ (cross-links) are also equal and their value β is varied. The results are shown for $N_t = 2$ and for two power values: 0 and 10 dB. We note that at lower SNR, our distributed scheme performs as well as or even better, depending on the PA scheme used than centralized zero-forcing even with PA (for the number of antennas considered); this is no longer the case at higher SNR. Moreover, at low SNR, the instantaneous heuristic (PA 2 in the figures) performs significantly better than the statistics-based heuristic (PA 1 in the figures) whereas the performances of both are comparable at higher SNR. More importantly, for any SNR regime, the higher the β, the stronger the cross-link and the more beneficial the cooperation. In fact, using the VSINR algorithm leads to lower

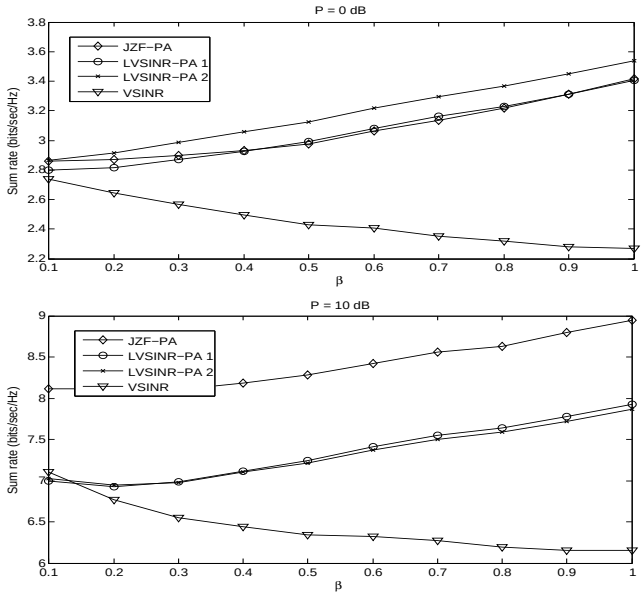


Figure 5.3: Performance comparison for symmetric case for different SNR values: $\sigma_{11}^2 = \sigma_{22}^2 = 1$, $\sigma_{12}^2 = \sigma_{21}^2 = \beta$.

sum rate for higher β, as in the absence of cooperation more interference is generated, whereas the opposite occurs for LVSINR and JZF-PA.

Figure 5.4 shows the performance comparison for a given arbitrary set of variances of the different links. Here too, one can see that as long as the "cross-links" are relatively strong there are rate gains to be obtained by sharing the data. Moreover, as the number of antennas increases the gap between our scheme and the JZF-PA scheme decreases. Only the statistics-based heuristic PA is shown in this figure.

Finally, Figure 5.5 illustrates LVSINR for $K = 4$ users served by $N = 3$ base stations with $N_t = 4$, such that σ_{jk}^2, $j = 1, \ldots, 3, k = 1, \ldots, 4$, is given by the jth entry in the kth column of the matrix

$$\begin{bmatrix} 1 & .5 & .05 \\ .2 & 1 & .2 \\ .05 & .5 & 1 \\ .3 & .3 & .08 \end{bmatrix}. \tag{5.15}$$

As there are fewer users to serve than there are base stations, comparing to a MISO IC (and applying the VSINR scheme) is no longer possible.

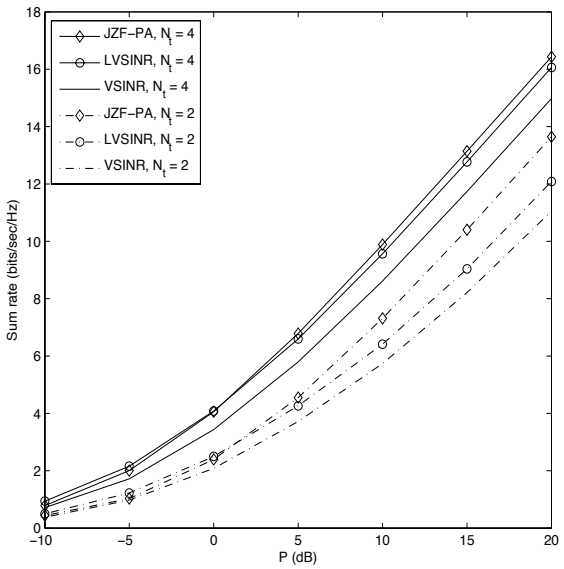


Figure 5.4: Performance comparison for asymmetric case: $\sigma_{11}^2 = .5, \sigma_{12}^2 = .3, \sigma_{21}^2 = .3, \sigma_{22}^2 = 1$, for different number of antennas.

5.4 Decentralized Beamforming

Though in the previous section, we assumed that statistical knowledge of the unknown links may be available at each transmitter, we did not actually make use of it. We now turn to a different distributed CSIT model more suitable for the case where users feed back quantized versions of their channels. Assuming each transmitter has, in addition to its own local information, statistical knowledge about the CSI at other transmitters, allows for a Bayesian formulation of the problem, which may be related to team decision theory [19, 58].

The N transmitters may be viewed as members of a team which need to take decisions in order to attain a common payoff, but who do not have access to the same information. Transmitter j chooses $\mathbf{V}_j$ based on its local CSI, $\hat{\mathbf{H}}^{(j)} \triangleq \left[\hat{\mathbf{h}}_1^{(j)}; \ldots; \hat{\mathbf{h}}_K^{(j)}\right]$ and the extra statistical information it has. Thus decisions at different transmitters are based on possibly different information. However, the performance (SINR, rate, BER for example) depends on all of these decisions. This is taken into account by a Bayesian formulation. As described below, we assume the quantization codebooks to

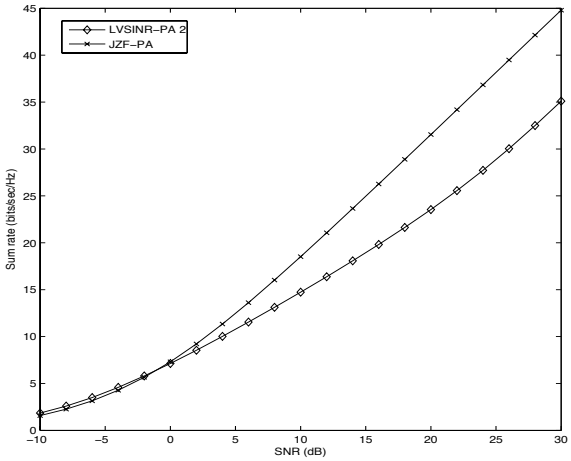


Figure 5.5: Sample performance comparison for 3 cells, $N_t = 4$ and $K = 4$.

be hierarchical. This offers additional structure to the problem and allows for some simplification in the problem statement.

Hierarchical CSI structure

To provide help in solving the problem, we propose that the true channel corresponding to user k, $\mathbf{h}_k$, be quantized using a hierarchical codebook, such that different transmitters know the channel up to different levels of said codebook. Each base station knows, in addition:

- Each user's channel statistics,
- The hierarchical codebook for each user,
- the hierarchy in their knowledge; this is detailed below.

For user k, $k = 1, \ldots, K$, we define a degrees of accuracy mapping

$$L_k : \{1, \ldots, N\} \mapsto \{0, l_{k,max}\}, \tag{5.16}$$

which maps each transmitter to the number of bits it can decode from user k's feedback information, in other words to its level of knowledge in user k's hierarchical codebook; $l_{k,max}$ corresponds to the most accurate level (the hierarchical codebook has $2^{l_{k,max}}$ codewords), whereas 0 bits means only

statistical knowledge of that user's channel. Thus transmitter j can decode quantized $\mathbf{h}_k$ up to accuracy level $L_k(j)$ yielding estimate $\hat{\mathbf{h}}_k^{(j)}$. One interesting advantage of this hierarchical information structure is that,if $L_k(j_1) > L_k(j_2)$, then transmitter j_1 knows exactly what is known by transmitter j_2, i.e. $\hat{\mathbf{h}}_k^{(j_2)}$, in addition to its own estimate $\hat{\mathbf{h}}_k^{(j_1)}$. On the other hand, transmitter j_2 does not know precisely what is decoded by transmitter j_1 however it does know that $\hat{\mathbf{h}}_k^{(j_1)}$ must belong to the subset of codewords located in the Voronoi region centered at $\hat{\mathbf{h}}_k^{(j_2)}$. This is illustrated in Figure 5.6, where $L_k^{-1}(l)$ denotes the inverse of the accuracy mapping and gives the set of users (if any) which decode the feedback information of user k up to accuracy level l.

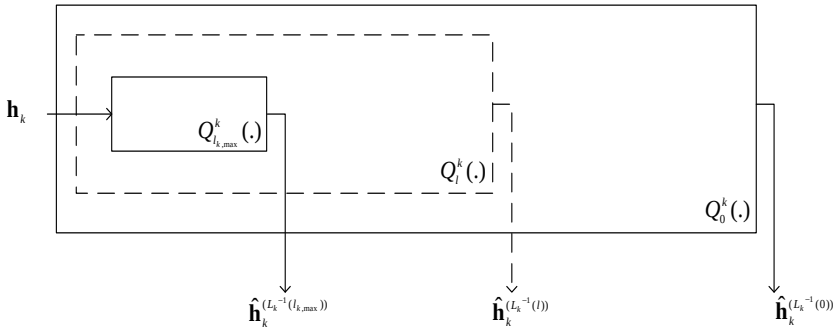


Figure 5.6: Distributed hierarchical CSI model: the quantization codebooks are designed to be hierarchical to offer additional structure.

5.4.1 Bayesian Formulation

We define the common goal of the considered team of N transmitters as the maximization of the expected value of a sum utility function, the sum rate for example. This utility function U is a function of the true channel states $\mathbf{h}_1, \ldots, \mathbf{h}_K$, as well as the decisions made at each transmitter, the beamforming vectors $\mathbf{w}_j$, since linear precoding is considered. We can write the objective function, to be maximized, as:

$$\mathcal{U} = \mathbb{E}\left[U\left(\mathbf{H}, \mathbf{V}_1\left(\hat{\mathbf{H}}^{(1)}\right), \ldots, \mathbf{V}_N\left(\hat{\mathbf{H}}^{(N)}\right)\right)\right], \tag{5.17}$$

where $\mathbf{H} \triangleq [\mathbf{h}_1; \ldots; \mathbf{h}_K]$, and the dependence of the decisions at each agent (transmitter) as a function of his knowledge is made explicit.

The sum utility may be written as

$$U\left(\mathbf{H}, \mathbf{V}_1\left(\hat{\mathbf{H}}^{(1)}\right), \ldots, \mathbf{V}_N\left(\hat{\mathbf{H}}^{(N)}\right)\right)$$
$$= \sum_{k=1}^{K} U_k\left(\mathbf{h}_k, \mathbf{V}_1\left(\hat{\mathbf{H}}^{(1)}\right), \ldots, \mathbf{V}_N\left(\hat{\mathbf{H}}^{(N)}\right)\right) \tag{5.18}$$

Restricting ourselves to deterministic decisions, in the sense that there will be a single $\mathbf{V}_i$ corresponding to each state of channel knowledge at transmitter j, $\hat{\mathbf{H}}^{(j)}$, $\mathcal{U}$ can be expanded into:

$$\mathcal{U} = \int \ldots \int d\mathbf{H} f_{\mathbf{H}}(\mathbf{H})\, U\left(\mathbf{H}, \tilde{\mathbf{V}}_1(\mathbf{H}), \ldots, \tilde{\mathbf{V}}_N(\mathbf{H})\right), \tag{5.19}$$

where

$$\tilde{\mathbf{V}}_j(\mathbf{H}) \triangleq \mathbf{V}_j\left(\hat{\mathbf{H}}^{(j)} \middle| \mathbf{H}\right)$$

is the beamforming strategy at transmitter j given the local knowledge at that transmitter corresponding to a true channel $\mathbf{H}$.

5.4.2 Global Optimization

A globally optimal set of beamforming decisions consists of sets of beamforming matrices $\{\mathbf{V}_j\}$, $j = 1, \ldots, N$, (one set per user, consisting of as many matrices as there are possible states of knowledge at that user), which jointly maximize $\mathcal{U}$. As stated in [58], [64] for example, it is often intractable to find the globally optimal strategies at the different team members. In such cases, a suboptimal solution may be obtained by finding strategies that are *person-by-person optimal*, as specified next.

5.4.3 Person-by-person Optimization

Person-by-person optimal strategies are such that for each team member, his strategy is optimal given the other team members' strategies. Clearly, the globally optimal strategies are person-by-person optimal, but the converse is in general not true. In our particular setup of distributed CSIT, an optimal strategy for transmitter j, given that the other transmitters' strategies are fixed, may be characterized, for a local channel knowledge equal to $\hat{\mathbf{H}}^{(j)}$, as follows:

$$\mathbf{V}_j^*\left(\hat{\mathbf{H}}^{(j)}\right)$$
$$= \arg \max_{\|\mathbf{V}_j\|_{\mathrm{F}}^2 \le P} \int \ldots \int d\mathbf{H} f_{\mathbf{H}|\hat{\mathbf{H}}^{(j)}}\left(\mathbf{H}|\hat{\mathbf{H}}^{(j)}\right) \tilde{U}(\mathbf{H}, \mathbf{V}_j) \tag{5.20}$$

where

$$\tilde{U}\left(\mathbf{H},\mathbf{V}_j\right) \triangleq U\left(\mathbf{H},\tilde{\mathbf{V}}_1\left(\mathbf{H}\right),\ldots,\mathbf{V}_j,\ldots,\tilde{\mathbf{V}}_N\left(\mathbf{H}\right)\right). \tag{5.21}$$

Since $\hat{\mathbf{H}}^{(j)}$ corresponds to a quantized version of the channel, we define $\mathbf{R}\left(\hat{\mathbf{H}}^{(j)}\right)$ the Voronoi region corresponding to this state of knowledge at transmitter j:

$$f_{\mathbf{H}|\hat{\mathbf{H}}^{(j)}}\left(\mathbf{H}|\hat{\mathbf{H}}^{(j)}\right) = \begin{cases} \frac{1}{\Pr\left[\mathbf{R}\left(\hat{\mathbf{H}}^{(j)}\right)\right]} f_{\mathbf{H}}\left(\mathbf{H}\right) & \mathbf{H} \in \mathbf{R}\left(\hat{\mathbf{H}}^{(j)}\right) \\ 0 & \mathbf{H} \notin \mathbf{R}\left(\hat{\mathbf{H}}^{(j)}\right) \end{cases}, \tag{5.22}$$

where

$$\Pr\left[\mathbf{R}\left(\hat{\mathbf{H}}^{(j)}\right)\right] = \int \ldots \int_{\mathbf{R}\left(\hat{\mathbf{H}}^{(j)}\right)} d\mathbf{H} f_{\mathbf{H}}. \tag{5.23}$$

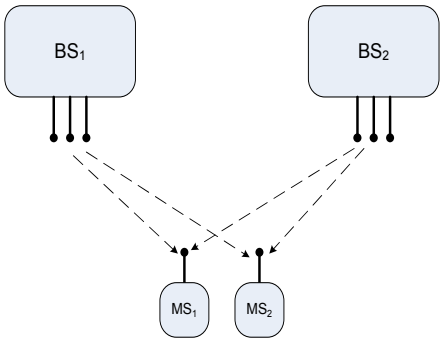
Thus, (5.20) is equivalent to:

$$\mathbf{V}_j^*\left(\hat{\mathbf{H}}^{(j)}\right) = \arg \max_{\|\mathbf{V}_j\|_{\mathrm{F}}^2 \leq P} \int \ldots \int_{\mathbf{R}\left(\hat{\mathbf{H}}^{(j)}\right)} d\mathbf{H} f_{\mathbf{H}}\left(\mathbf{H}\right) \tilde{U}\left(\mathbf{H},\mathbf{V}_j\right). \tag{5.24}$$

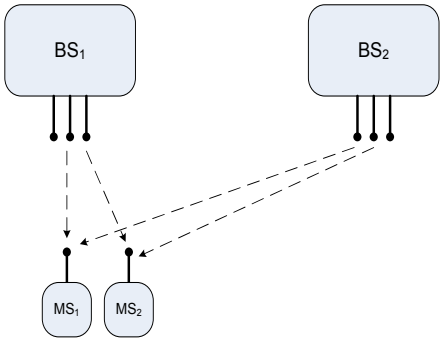
Such a person-by-person optimization approach may be useful if the number of decisions to be determined is too large, or if the knowledge at the different transmitters does not satisfy our hierarchical assumption. For the case when $K = N = 2$, which we consider next, we formulate the problem in a way so as to try to find the globally optimal transmitter strategies.

5.4.4 Decentralized Beamforming, for $K = N = 2$

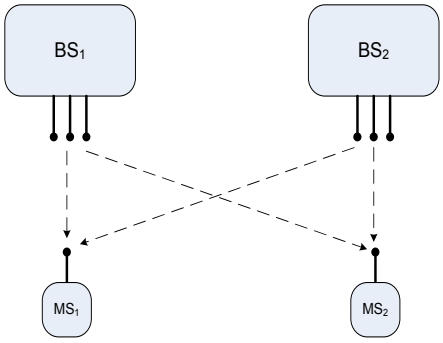
To simplify exposition of the solution to the problem, we focus on the $K = N = 2$ case. The hierarchy in the knowledge at the two transmitters, and as a result the corresponding beamforming strategies, fall into one of three cases, which may be characterized as follows:
Common knowledge: In this case, $L_1(1) = L_1(2)$ and $L_2(1) = L_2(2)$. It corresponds to the *traditional* assumption under limited CSIT, where both transmitters have the same knowledge. This arises, for instance, when users are at the cell edge, as represented in Figure 5.7(a). This is equivalent to having centralized beamforming decisions being made.
Degraded knowledge: In this case, $L_1(1) \geq L_1(2)$ and $L_2(1) \geq L_2(2)$, or $L_1(1) \leq L_1(2)$ and $L_2(1) \leq L_2(2)$. In other words, one of the transmitters has a better representation of both channels, and will adapt its beamforming



(a) Two users at cell edge



(b) Two users in the same cell



(c) Two users inside respective cells

Figure 5.7: Different cell setups corresponding to different CSI hierarchies.

on a finer scale than the other transmitter. Such a situation would arise, for example, if the two users being served lie in the same 'cell', as in Figure 5.7(b).

Symmetric knowledge: Here, $L_1(1) > L_1(2)$ and $L_2(1) < L_2(2)$, or $L_1(1) < L_1(2)$ and $L_2(1) > L_2(2)$. So one of the transmitters has a better representation of the channel of a given user and a worse one for the other user, with the reverse occuring at the other transmitter. This corresponds, for instance, to the base stations serving users each situated within their own 'cell', as in Figure 5.7(c). As will be detailed below one needs to jointly optimize sets of beamforming decisions at the two transmitters corresponding to a given common coarse state of channel knowledge.

We now focus on the symmetric case where $L_1(1) > L_1(2)$ and $L_2(1) < L_2(2)$: this represents the more common setup among the ones described and shown in Figure 5.7 above and is also the more challenging to formulate; the remaining cases can be dealt with in a similar manner. We characterize each user's quantized CSI by a pair $\mathbf{i}_1 = (i_{1,2}, i_{1,1})$ for user 1, and another $\mathbf{i}_2 = (i_{2,1}, i_{2,2})$ for user 2. The first index in each pair corresponds to the coarse knowledge (hence is shared by both users), i.e. the index of the codeword in the coarsest codebook, to which the channel is quantized, $Q_{L_k^{-1}(\min_j L_k(j))}(\mathbf{h}_k)$ (see Figure 5.6), and the second index provides the missing bits to locate the finer codeword around the coarsest one, $Q_{L_k^{-1}(\max_j L_k(j))}(\mathbf{h}_k)$. Given the structure of the distributed CSI, the beamforming matrix decisions may be parametrized in terms of these indices, so that $\mathbf{V}_1$ varies with $(\mathbf{i}_1, i_{2,1})$, whereas $\mathbf{V}_2$ is a function of $(i_{1,2}, \mathbf{i}_2)$.

Taking this into consideration, we expand (5.19) into

$$\sum_{i_{1,2}=1}^{2^{L_1(2)}} \sum_{i_{2,1}=1}^{2^{L_2(1)}} S\left(i_{1,2}, i_{2,1}\right). \tag{5.25}$$

In (5.25), $S\left(i_{1,2}, i_{2,1}\right)$ is given by

$$\sum_{i_{1,1}=1}^{I_1} \sum_{i_{2,2}=1}^{I_2} \int_{\mathbf{R}_1(\mathbf{i}_1)} \int_{\mathbf{R}_2(\mathbf{i}_2)} d\mathbf{h}_1 d\mathbf{h}_2 f_{\mathbf{H}}\left(\mathbf{H}\right) U\left(\mathbf{H}, \mathbf{V}_1\left(\mathbf{i}_1, i_{2,1}\right), \mathbf{V}_2\left(i_{1,2}, \mathbf{i}_2\right)\right), \tag{5.26}$$

where $I_1 = 2^{L_1(1)-L_1(2)}$, $I_2 = 2^{L_2(2)-L_2(1)}$, $\mathbf{R}_1(\mathbf{i}_1)$ and $\mathbf{R}_2(\mathbf{i}_2)$ correspond to the Voronoi regions associated with the indexed codewords.

It is easy to verify that the beamforming decisions for each $S\left(i_{1,2}, i_{2,1}\right)$ term may be optimized separately. For given $i_{1,2}$ and $i_{2,1}$, we optimize the

corresponding $S(i_{1,2}, i_{2,1})$. To simplify notation we remove the dependence on $i_{1,2}$ and $i_{2,1}$ from the expressions. The problem is thus:

$$\begin{aligned}
\max. \sum_{i_{1,1}=1}^{I_1} \sum_{i_{2,2}=1}^{I_2} \int_{\mathbf{R}_1(i_{1,1})} \int_{\mathbf{R}_2(i_{2,2})} d\mathbf{h}_1 d\mathbf{h}_2 & \\
[f_{\mathbf{H}}(\mathbf{H}) U(\mathbf{H}, \mathbf{V}_1(i_{1,1}), \mathbf{V}_2(i_{2,2}))] & \quad (5.27)\\
\text{s.t. } \|\mathbf{V}_1(i_{1,1})\|_F^2 \leq P, i_{1,1} = 1, \ldots, I_1 & \quad (5.28)\\
\|\mathbf{V}_2(i_{2,2})\|_F^2 \leq P, i_{2,2} = 1, \ldots, I_2. & \quad (5.29)
\end{aligned}$$

Recalling the separable nature of our utility function (refer to equation (5.18)), this can be reformulated as:

$$\begin{aligned}
\max. \sum_{i_{1,1}=1}^{I_1} \sum_{i_{2,2}=1}^{I_2} \sum_{k=1}^{2} \Pr\left[\mathbf{R}_{\bar{k}}(i_{\bar{k},\bar{k}})\right] \int_{\mathbf{R}_k(i_{k,k})} d\mathbf{h}_k & \\
[f_{\mathbf{h_k}}(\mathbf{h}_k) U_k(\mathbf{h}_k, \mathbf{V}_1(i_{1,1}), \mathbf{V}_2(i_{2,2}))] & \\
\text{s.t. } \|\mathbf{V}_1(i_{1,1})\|_F^2 \leq P, i_{1,1} = 1, \ldots, I_1 & \\
\|\mathbf{V}_2(i_{2,2})\|_F^2 \leq P, i_{2,2} = 1, \ldots, I_2, & \quad (5.30)
\end{aligned}$$

where $\bar{k} = \mod(k, 2) + 1$ and

$$\Pr\left[\mathbf{R}_{\bar{k}}(i_{\bar{k},\bar{k}})\right] = \int_{\mathbf{R}_{\bar{k}}(i_{\bar{k},\bar{k}})} d\mathbf{h}_{\bar{k}} f_{\mathbf{h}_{\bar{\mathbf{k}}}}(\mathbf{h}_{\bar{k}}), \quad (5.31)$$

is the probability of user $\bar{k}$'s channel being quantized to the codeword indexed by the pair $(i_{\bar{k},k}, i_{\bar{k},\bar{k}})$.

Application to sum rate maximization

The above problem may be approximately solved via a projected gradient ascent method. Moreover, to avoid integration, we resort to approximations. As we deal with sum rate maximization in our illustrative examples, the

following approximation is plugged into problem formulation (5.30) above:

$$\begin{aligned}
&\int_{\mathbf{R}_k(i_{k,k})} d\mathbf{h}_k U_k\left(\mathbf{h}_k, \mathbf{V}_1\left(i_{1,1}\right), \mathbf{V}_2\left(i_{2,2}\right)\right) \\
&= \int_{\mathbf{R}_k(i_{k,k})} d\mathbf{h}_k \log_2\left(1+\frac{\left|\mathbf{h}_k\mathbf{w}_k\left(i_{1,1}, i_{2,2}\right)\right|^2}{\sigma^2+\left|\mathbf{h}_k\mathbf{w}_{\bar{k}}\left(i_{1,1}, i_{2,2}\right)\right|^2}\right) \\
&\approx \Pr\left[\mathbf{R}_k(i_{k,k})\right] \log_2\left(1+\frac{\mathbf{w}_k\left(i_{1,1}, i_{2,2}\right)^H \mathbf{C}_k^{(i_{k,k})} \mathbf{w}_k\left(i_{1,1}, i_{2,2}\right)}{\sigma^2+\mathbf{w}_{\bar{k}}\left(i_{1,1}, i_{2,2}\right)^H \mathbf{C}_k^{(i_{k,k})} \mathbf{w}_{\bar{k}}\left(i_{1,1}, i_{2,2}\right)}\right),
\end{aligned} \tag{5.32}$$

where $\mathbf{C}_k^{(i_{k,k})} = \mathbb{E}\left[\mathbf{h}_k\mathbf{h}_k^H \middle| \mathbf{h}_k \in \mathbf{R}_k(i_{k,k})\right]$, and $\mathbf{w}_k\left(i_{1,1}, i_{2,2}\right)$, $k = 1, 2$ is obtained from $\mathbf{V}_1\left(i_{1,1}\right)$ and $\mathbf{V}_2\left(i_{2,2}\right)$ by extracting the appropriate entries as defined in our system model. A similar approximation was used in [65,66] for example. The quality of this approximation increases and becomes asymptotically optimal with the size of the codebook.

5.4.5 Reference Schemes

Simple upper and lower bounds to the proposed schemes correspond to joint beamforming based on the more accurate (unachievable in a distributed CSIT system) and the least accurate (achievable) CSIT, respectively. Another decentralized scheme which attempts to use the local channel knowledge would be for each base station to design its transmission assuming all the other base stations share the same knowledge as itself. This is much simpler than the proposed decentralized scheme, and has similar complexity to joint beamforming design based on the coarse CSIT.

5.4.6 Numerical Results

To illustrate the gains from such decentralized scheme, we show the average sum rates achieved for a symmetric $K = N = 2$, $N_t = 1$ channel, where Rayleigh fading is assumed and the covariance matrix of user 1's channel is given by $[1\ 0; 0\ \beta]$, that of user 2 by $[\beta\ 0; 0\ 1]$, β being a simulation parameter modeling the strength of the 'cross links'. We also vary the number of bits used for the different quantization levels.

The hierarchical codebooks are designed using Lloyd's algorithm: first the coarse codebook, then for each codeword in it, the corresponding finer codebook.

Figure 5.8 compares the proposed decentralized scheme to the upper and lower bounds given in 5.4.5 for $L_1(2) = L_2(1) = 2$ and $L_1(1) = L_2(2) = 6$. We label the scheme which attempts to use local channel knowledge as if it were shared 'myopic beamforming (BF)'. Thus, the upper bound scheme would require $2(L_1(1) + L_2(2)) = 24$ bits of CSIT being shared, whereas the schemes based on distributed CSIT would require $L_1(1) + L_2(2) + L_1(2) + L_2(1) = 16$ bits. The benefit of the second layer of CSI over the more coarse shared representation of the channel depends on the SNR and on the value of β. At low SNR and for β low, there is little use for the extra information. The performance of myopic BF, even though it relies on more information than the joint beamforming relying on coarse CSI, is significantly worse, highlighting the importance of coordinated action. For reference, we also plot the performance that would be obtained if the knowledge at transmitter i, $i = 1, 2$ were indeed common to both transmitters and joint beamforming would result; clearly this yields more gain than joint beamforming based on coarse CSI.

5.5 Conclusions

In this chapter, two different distributed CSIT setups were tackled to investigate the performance that may be achieved if one foregoes sharing the channel state information fully among cooperating transmitters. The first setup arises in the context of TDD systems where CSI is gained from uplink transmission and therefore the transmitters each end up with mutually exclusive pieces of the CSI puzzle. For transmitters with $N_t \geq 2$ we investigate extending the VSINR framework applied in the previous chapter to the MISO IC to the case of joint transmission. The second setup corresponds to transmitters being able to decode the fed back quantized CSI up to different levels of accuracy. A hierarchical quantization model for this case was proposed and a Bayesian formulation for the decentralized beamforming design was provided, based on team decision theory.

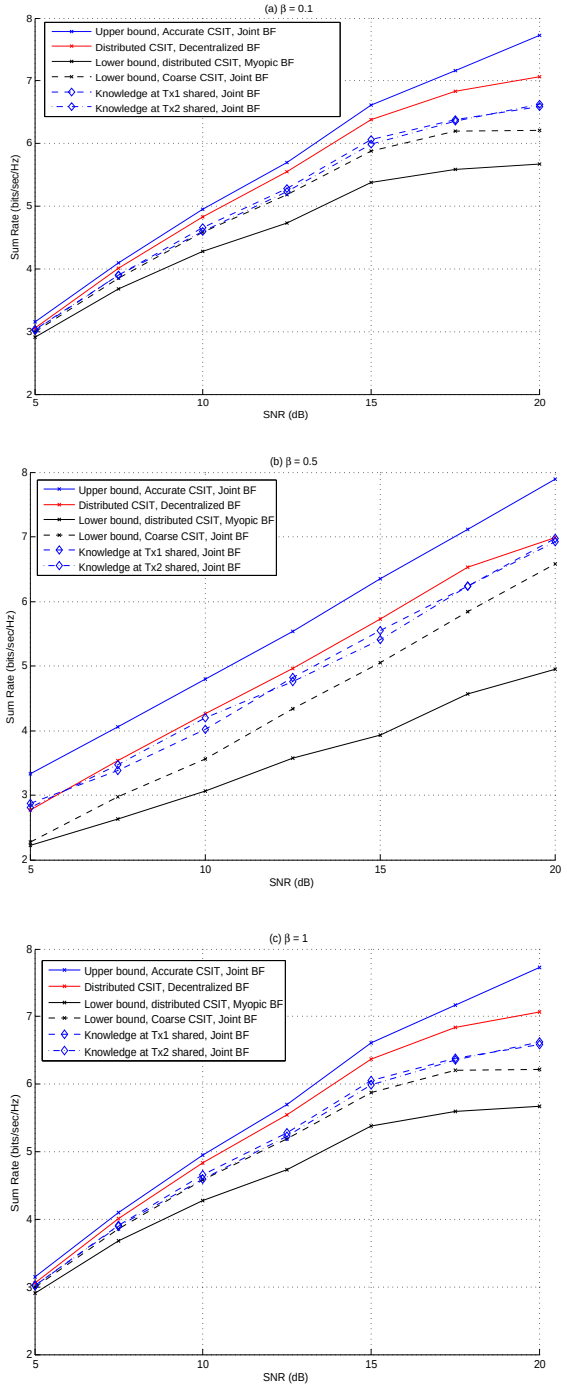


Figure 5.8: Sum Rate Comparison for $L_1(2) = L_2(1) = 2, L_1(1) = L_2(2) = 6$ bits and different β.

Chapter 6

Optimized data sharing in Network MIMO with finite backhaul capacity

6.1 Introduction

In this chapter, we tackle the issue of limited backhaul capacity in a cooperative multicell setup. As noted in the introductory chapters, multicell processing (multicell MIMO, network MIMO) as proposed in [10], [11] for example requires full data sharing. This subsumes high capacity backhaul links, which may not always be feasible, or even simply desirable, in certain applications. In fact, under limited backhaul rate constraints, data sharing consumes a precious fraction of the backhaul capacity which must be compensated by the capacity gain induced by the network MIMO channel over the classical interference channel. Imposing finite capacity constraints on the backhaul links brings with it a set of interesting research questions, in particular:

- Given the backhaul constraints, assuming that *not all* traffic is shared across transmitters, i.e assuming a certain part remains private to each transmitter, what kind of rates can we expect to achieve? What is the capacity region of the resulting multicell channel? In fact, the

multicell MIMO channel under finite backhaul no longer corresponds to a MIMO broadcast channel, nor does it correspond to the so-called interference channel.

- How useful is data sharing when backhaul constraints are present? In other words, how do the rates achieved with a data sharing-, and therefore joint transmission enabling-scheme compare to those achieved without data sharing, when the backhaul is limited?

These questions have lead to a number of recent interesting research efforts. To cite a few, in [67] and [20], joint encoding for the downlink of a cellular system is studied under the assumption that the base stations are connected to a central unit via finite-capacity links. The authors investigate different transmission schemes and ways of using the backhaul capacity in the context of a modified version of Wyner's channel model. One of their main conclusions is that "central encoding with oblivious cells", whereby quantized versions of the signals to be transmitted from each base station, computed at the central unit, are sent over the backhaul links, is shown to be a very attractive option for both ease of implementation and performance, unless high data rates are required. If this is the case, the base stations need to be involved in the encoding, i.e. at least part of the backhaul link should be used for sending the messages themselves not the corresponding codewords.

In [68], an optimization framework, for an adopted backhaul usage scheme, is proposed for the downlink of a large cellular system. A so-called joint transmission configuration matrix is defined: this specifies which antennas in the system serve which group of users. The backhaul to each base station is used to either carry quantized versions of the transmit signals computed centrally similarly to the central encoding with oblivious cells scheme in [20], except that a more realistic system model is assumed, alternatively, the backhaul is used to carry uncoded binary user data. The numbers of bits per user and per antenna are optimized, such that users served by the same set of antennas are allocated the same number of bits.

In [69], a more information-theoretic approach is taken and a two-cell setup is considered in which, in addition to links between the network and each base station, a finite-capacity link connects the two multi-antenna base stations: the authors view the thus formed channel as a superposition of an interference channel and a broadcast channel. The backhaul is used to share the data to be jointly transmitted: this could be in the form of the full messages, or of quantized versions of the signals to be transmitted, depending on whether the data is coming from the network directly or shared over

the link between the base stations. The schemes proposed lead to nonconvex problems which make it difficult to find the rate and power parameters that come into play, as well as to characterize the optimum beamforming vectors to use, and the suboptimal scheme of maximum ratio transmission is resorted to.

In this chapter, our contributions are as follows:

- We also consider a setup in which the backhaul is between the network and each of the base stations, and focus on how to use this given backhaul to serve the users in the system. We focus on the two-cell problem.
- We specify a transmission scheme whereby superposition coding is used to transmit signals to each user: this allows us to formulate a continuum between full message sharing between base stations (network MIMO) and the conventional network with single serving base stations (IC); the data rate is in fact split between two distinct forms of data to be received by the users, a private form to be sent by the 'serving' base alone and a common form to be transmitted via multiple bases.
- We express the corresponding rate region in terms of the backhaul constraints and the beamforming vectors used to carry the different signals.
- We reduce finding the boundary of the aforementioned region to solving a set of convex optimization problems. We compare the rates achieved in such a hybrid scheme to those obtained for network MIMO and the IC and illustrate the gains related to moderate sharing levels in certain realistic situations.
- We also adapt the "central encoding with oblivious cells" in [20] to the channel model and linear precoding transmission scheme we use to enable comparison with the proposed rate splitting approach.

6.2 System Model

The system considered is shown in Figure 6.1. In this preliminary study, we focus on a two transmitter two receiver setup. We assume a noiseless backhaul link of capacity C_j between the central processor (CP) or the backbone network, and transmitter j, for $j = 1, 2$: it will be used to transmit the messages for each user. We distinguish between different types of messages:

- private messages will be sent from the CP to only one of the transmitters, and

- shared or common messages, which are sent from the CP to both transmitters, and are consequently jointly transmitted. Note that this notion of a common message is different from that commonly used in the context of interference channels for example, as they do not correspond to messages to be decoded by both receivers, but rather to messages to be sent by both transmitters.

Thus for user k, the message rate r_k will be split across $r_{k,p}$ and $r_{k,c}$, where $r_{k,c}$ and $r_{k,p}$ refer to the common and the private rates for that user, respectively:

$$r_k = r_{k,p} + r_{k,c}. \tag{6.1}$$

Assumptions Full CSIT is assumed to be available at both transmitters, since we want to focus on the cost of sharing data. Chapter 4 focuses on the problem of CSIT sharing.

Notation In what follows, $\bar{k} = \mod(k,2)+1, k = 1,2$ and is used to denote the *other* transmitter/receiver depending on the context.

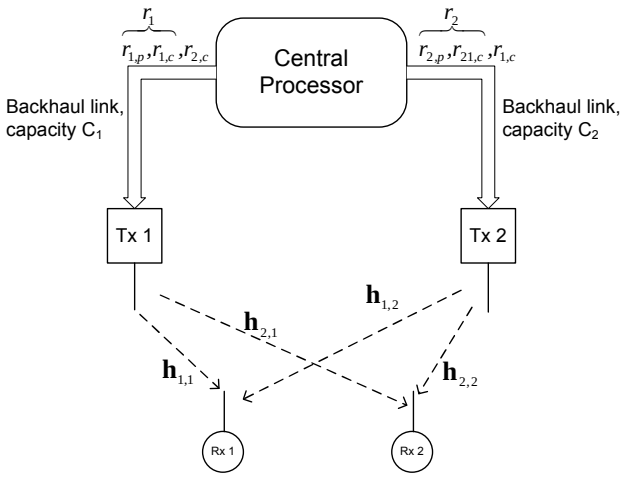


Figure 6.1: Constrained backhaul setup.

6.3 Proposed Backhaul Usage

Backhaul link k with finite capacity C_k serves to carry both private and common messages for user k as well as the common messages for user $\bar{k}$, so that the following constraint applies:

$$C_k \geq r_{k,p} + r_{k,c} + r_{\bar{k},c}, \quad k = 1, 2. \tag{6.2}$$

Using (6.1), this constraint can be rewritten as:

$$C_k \geq r_k + r_{\bar{k}} - r_{\bar{k},p}, \quad k = 1, 2. \tag{6.3}$$

Since $r_{k,p} \geq r_k$, we have that

$$r_1 + r_2 \leq C_1 + C_2. \tag{6.4}$$

6.3.1 Over the air transmission

Given that for each user, part of his messages will be known at both cooperating nodes, whereas another part will only be known at one of them, the transmission will consist of the superposition of the precoded symbols corresponding to each of the two categories of messages: 'private symbols' will arrive from the transmitter than knows them only, whereas 'common symbols' will be jointly precoded by both transmitters.

The transmitted signal may be modeled as follows:

$$\begin{aligned}\mathbf{x} = &\begin{bmatrix} \mathbf{w}_{1,c} & \mathbf{w}_{2,c} \end{bmatrix} \begin{bmatrix} s_{1,c} \\ s_{2,c} \end{bmatrix} + \begin{bmatrix} \mathbf{w}_{1,p} \\ \mathbf{0} \end{bmatrix} s_{1,p} \\ &+ \begin{bmatrix} \mathbf{0} \\ \mathbf{w}_{2,p} \end{bmatrix} s_{2,p},\end{aligned} \tag{6.5}$$

where $\mathbf{x} \in \mathbb{C}^{2N_t}$ is the transmitted signal, such that the first N_t elements are transmitter 1's transmit signal, the remaining N_t elements are transmitter 2's signal. $\mathbf{w}_{k,c} \in \mathbb{C}^{2N_t}, k = 1, 2$ are the beamforming vectors carrying user k's common message symbols $s_{k,c}$, and $\mathbf{w}_{k,p} \in \mathbb{C}^{N_t}, k = 1, 2$ are the beamforming vectors carrying user k's private message symbols $s_{k,p}$. Gaussian signaling is assumed, so that $s_{1,p}, s_{1,c}, s_{2,p}, s_{2,c}$ are all $\mathcal{CN}(0, 1)$.

Per base station power constraints $P_j, j = 1, 2$ imply that:

$$\|\mathbf{D}_j \mathbf{w}_{1,c}\|^2 + \|\mathbf{D}_j \mathbf{w}_{2,c}\|^2 + \|\mathbf{w}_{j,p}\|^2 \leq P_j, \tag{6.6}$$

where $\mathbf{D}_j$ is a matrix whose only non-zero elements are elements $(N_t - 1)j + 1 : jN_t$ of the diagonal and equal 1.

The signal received at receiver k will be given by (see (6.5)):

$$\begin{aligned} y_k &= \mathbf{h}_k \mathbf{x} + n_k = \left[\begin{array}{cc} \mathbf{h}_{k,1} & \mathbf{h}_{k,2} \end{array} \right] \mathbf{x} + n_k \\ &= \mathbf{h}_k \mathbf{w}_{1,c} s_{1,c} + \mathbf{h}_k \mathbf{w}_{2,c} s_{2,c} + \mathbf{h}_{k,1} \mathbf{w}_{1,p} s_{1,p} \\ &+ \mathbf{h}_{k,2} \mathbf{w}_{2,p} s_{2,p} + n_k \end{aligned} \tag{6.7}$$

6.3.2 Background: MAC with Common Message

Given our assumptions,the channel between the two transmitters and user k can be viewed as a MAC with a common message [70], where the receiver noise is replaced by receiver noise plus interference due to transmission to user $\bar{k}$. Denoting by σ_k^2 this power, i.e.

$$\sigma_k^2 = \sigma^2 + \left| \mathbf{h}_{k,\bar{k}} \mathbf{w}_{\bar{k},p} \right|^2 + \left| \mathbf{h}_k \mathbf{w}_{\bar{k},c} \right|^2, \tag{6.8}$$

the following rate region is achievable:

$$\begin{aligned} r_{k,p} &\leq \log_2 \left(1 + \frac{\left| \mathbf{h}_{k,k} \mathbf{w}_{k,p} \right|^2}{\sigma_k^2} \right), \\ r_k = r_{k,p} + r_{k,c} &\leq \log_2 \left(1 + \frac{\left| \mathbf{h}_{k,k} \mathbf{w}_{k,p} \right|^2 + \left| \mathbf{h}_k \mathbf{w}_{k,c} \right|^2}{\sigma_k^2} \right). \end{aligned} \tag{6.9}$$

Proof. This follows from results obtained for the two-transmitter MAC with a common message [70]. More details are given in Appendix 6.A. □

6.3.3 Particular Cases

The transmission scheme introduced here covers the two particular cases of:

- an interference channel, which is obtained by forcing $r_{k,p} \equiv r_k, k = 1, 2$, and
- a multicell MIMO channel, obtained by forcing $r_{k,p} \equiv 0, k = 1, 2$.

6.4 Achievable Rate Region

An achievable rate region $\mathcal{R}$ is the set of $(r_1, r_{1,p}, r_2, r_{2,p})$, as specified above, that satisfy the specified backhaul and power constraints.

One way to obtain the rate region boundary is to use the rate profile notion from [71]: a rate profile specifies how the total rate is split between

the users. Points on the boundary of the rate region are thus obtained by solving the following problem for α discretized over $[0, 1]$:

$$
\begin{aligned}
\text{max. } & r \\
\text{s.t. } & r_1 \geq \alpha r \\
& r_2 \geq (1-\alpha) r \\
& r_1 + r_2 - r_{2,p} \leq C_1, \quad r_1 + r_2 - r_{1,p} \leq C_2 \\
& r_k \leq \log_2 \left(1 + \frac{|\mathbf{h}_{k,k}\mathbf{w}_{k,p}|^2 + |\mathbf{h}_k \mathbf{w}_{k,c}|^2}{\sigma^2 + \left|\mathbf{h}_{k,\bar{k}}\mathbf{w}_{\bar{k},p}\right|^2 + \left|\mathbf{h}_k \mathbf{w}_{\bar{k},c}\right|^2} \right), k = 1, 2, \qquad (6.10) \\
& r_{k,p} \leq \log_2 \left(1 + \frac{|\mathbf{h}_{k,k}\mathbf{w}_{k,p}|^2}{\sigma^2 + \left|\mathbf{h}_{k,\bar{k}}\mathbf{w}_{\bar{k},p}\right|^2 + \left|\mathbf{h}_k \mathbf{w}_{\bar{k},c}\right|^2} \right), k = 1, 2, \qquad (6.11) \\
& \|\mathbf{w}_{k,p}\|^2 + \|\mathbf{D}_k \mathbf{w}_{k,c}\|^2 + \|\mathbf{D}_k \mathbf{w}_{\bar{k},c}\|^2 \leq P_k, k = 1, 2 \qquad (6.12)
\end{aligned}
$$

We solve this problem using a bisection method over r, determining the feasibility is detailed in subsection 6.4.1 .

6.4.1 Establishing feasibility of a given rate

Assume sum rate r and α to be fixed. Thus, $r_1 = \alpha r$, $r_2 = (1-\alpha)r$. Establishing feasibility of a given rate pair hinges on the following two remarks:

- For r_k to be supported, it cannot possibly exceed C_k, and
- Sharing information whenever possible outperforms not doing so. Thus a rate pair is not achievable unless it is so for the minimum possible private message rates $r_{k,p}$, $k = 1, 2$. Given the backhaul constraints, these are:

$$(r_{k,p})_{min} = \max(0, r_1 + r_2 - C_{\bar{k}}), k = 1, 2. \qquad (6.13)$$

Thus if $r_1 > C_1$ or $r_2 > C_2$, the rate pair is necessarily infeasible. Otherwise, we check if rate tuple $\left(r_1, (r_{1,p})_{min}, r_2, (r_{2,p})_{min}\right) \in \mathbb{R}^4$ is achievable [1]. Section 6.4.2 below shows one way to establish feasility of a given tuple $(r_1, r_{1,p}, r_2, r_{2,p})$ and the corresponding beamforming.

[1] Note that in our simulations, since not sharing messages yields a simpler beamforming scheme, we first check for the feasibility of rate tuple (r_1, r_1, r_2, r_2).

6.4.2 Feasibility of $(r_1, r_{1,p}, r_2, r_{2,p})$

Assume r_1, r_2, $r_{1,p}$ and $r_{2,p}$ are fixed. As the objective is to determine the feasibility of this set of rates, this can be done by solving any optimization problem subject to constraints (6.10), (6.11) and (6.12), in particular the total power minimization problem, which can be formulated as:

$$\begin{aligned}
\text{min.} \quad & \sum_{k=1}^{2} \left[\|\mathbf{w}_{k,c}\|^2 + \|\mathbf{w}_{k,p}\|^2 \right] \\
\text{s.t.} \quad & 2^{r_k} - 1 \leq \frac{|\mathbf{h}_{k,k}\mathbf{w}_{k,p}|^2 + |\mathbf{h}_k\mathbf{w}_{k,c}|^2}{\sigma^2 + \left|\mathbf{h}_{k,\bar{k}}\mathbf{w}_{\bar{k},p}\right|^2 + \left|\mathbf{h}_k\mathbf{w}_{\bar{k},c}\right|^2}, k = 1, 2, \\
& 2^{r_{k,p}} - 1 \leq \frac{|\mathbf{h}_{k,k}\mathbf{w}_{k,p}|^2}{\sigma^2 + \left|\mathbf{h}_{k,\bar{k}}\mathbf{w}_{\bar{k},p}\right|^2 + \left|\mathbf{h}_k\mathbf{w}_{\bar{k},c}\right|^2}, k = 1, 2, \\
& \|\mathbf{D}_k\mathbf{w}_{k,c}\|^2 + \|\mathbf{D}_k\mathbf{w}_{\bar{k},c}\|^2 + \|\mathbf{w}_{k,p}\|^2 \leq P_k, k = 1, 2.
\end{aligned}$$

We can transform the above problem into an equivalent convex optimization problem.

- If $r_{k,p} \equiv 0$ or $r_k \equiv r_{k,p}$, we can reduce the problem as follows:
  - If $r_{k,p} \equiv 0$, the corresponding constraint becomes redundant, and $\mathbf{w}_{k,p} = \mathbf{0}$.
  - If $r_k \equiv r_{k,p}$, then $\mathbf{w}_{k,c} = \mathbf{0}$ at the optimum and we can remove the constraint corresponding to r_k.

  In both cases, the remaining constraint can be transformed into a second-order cone constraint as in [14, 54, 72].

- Otherwise, further manipulation is needed to obtain the equivalent convex problem. Consider the inequalities related to user k's rates. Imposing the decoding order to be common message, then private message (alternatively, given the MAC with common message structure that we have, dirty paper coding could be used to encode the private message to ensure it is received without interference), both inequalities must be met with equality at the optimum. Combining these two equations, we get:

$$\frac{2^{r_{k,p}} - 1}{2^{r_k} - 2^{r_{k,p}}} |\mathbf{h}_k\mathbf{w}_{k,c}|^2 = |\mathbf{h}_{k,k}\mathbf{w}_{k,p}|^2 . \tag{6.14}$$

Further noting that $\mathbf{h}_k\mathbf{w}_{k,c}$ and $\mathbf{h}_{k,k}\mathbf{w}_{k,p}$ being real does not restrict the solution, we obtain the following equivalent convex problem:

$$\begin{aligned}
\text{min.} \quad & \sum_{k=1}^{2} \left[\|\mathbf{w}_{k,c}\|^2 + \|\mathbf{w}_{k,p}\|^2\right] \\
\text{s.t.} \quad & \sqrt{\frac{2^{r_k} - 2^{r_{k,p}}}{2^{r_{k,p}} - 1}}\mathbf{h}_{k,k}\mathbf{w}_{k,p} = \mathbf{h}_k\mathbf{w}_{k,c}, k = 1, 2 \\
& \sqrt{2^{r_{k,p}} - 1}\left\|\left[\ \sigma \quad \mathbf{h}_{k,\bar{k}}\mathbf{w}_{\bar{k},p} \quad \mathbf{h}_k\mathbf{w}_{\bar{k},c}\ \right]\right\| \leq \mathbf{h}_{k,k}\mathbf{w}_{k,p}, k = 1, 2 \\
& \|\mathbf{D}_k\mathbf{w}_{k,c}\|^2 + \|\mathbf{D}_k\mathbf{w}_{\bar{k},c}\|^2 + \|\mathbf{w}_{k,p}\|^2 \leq P_k, k = 1, 2
\end{aligned}$$

6.4.3 Numerical Results

Figure 6.2 shows the rate regions corresponding to the proposed scheme, labelled Hybrid IC/Network MIMO, to the interference channel ($r_{k,c} = 0, k = 1, 2$) and to the network MIMO scheme ($r_{k,p} = 0, k = 1, 2$) for different values of the backhaul capacity, for a particular channel instance with $N_t = 2$. As seen in this figure, for low backhaul capacity, the rate regions corresponding to the hybrid scheme and the IC overlap and are both significantly larger than the network MIMO region. As the backhaul capacity increases, all 3 regions become larger (up to the point where the system is no longer constrained by the backhaul), the network MIMO region becomes larger than the IC region and closer to the Hybrid scheme's achievable region, until eventually these two regions overlap (this is not shown in the figure but occurs for $C \approx 10$ bits/channel use for the sample channel).

Finally, Figure 6.3 illustrates the average common to total rate ratio, when the objective is to maximize the minimum rate (i.e. find the point on the rate region boundary whose rate profile is characterized by $\alpha = .5$), for a symmetric Rayleigh fading channel such that $\mathbf{h}_{k,k} \sim \mathcal{CN}(0, \mathbf{I}_{N_t})$ and $\mathbf{h}_{k,\bar{k}} \sim \mathcal{CN}(0, \beta\mathbf{I}_{N_t})$, when $N_t = 1$. The ratios are plotted as a function of C for two different values of β. Unless the power of the cross links is zero, the rate region of unconstrained network MIMO setup will be strictly larger than that of the unconstrained IC. Thus if the backhaul is large enough, data sharing will allow for higher rates to be achieved (up to the limit of the unconstrained network MIMO channel).

6.5 Comparison with a Quantized backhaul

Depending on the network setup, it may also be possible to move the processing away from the base stations and assume these to be oblivious of the

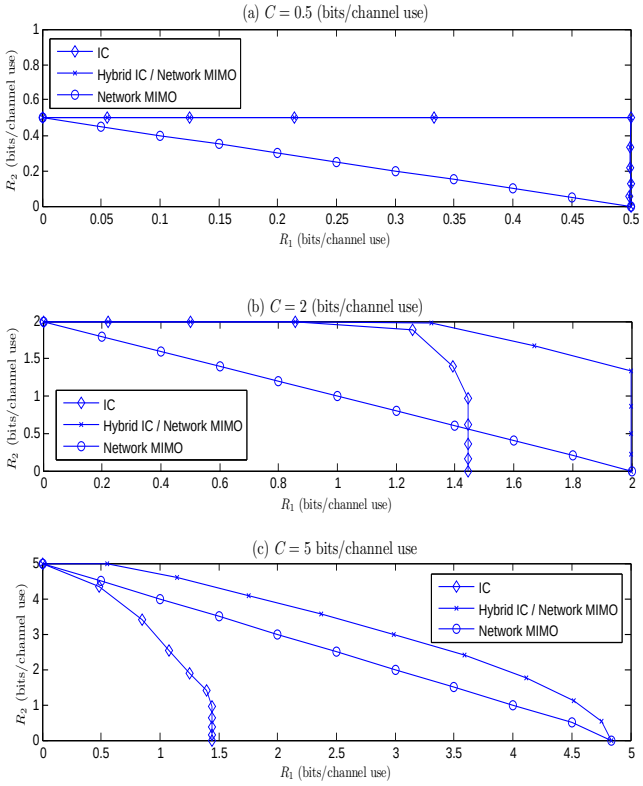


Figure 6.2: Sample Rate Regions Comparison for 10dB SNR for $\mathbf{h}_1 = [-0.3059 + \mathbf{i}0.2314 \quad .0886 - \mathbf{i}0.0132 \quad -.8107 - \mathbf{i}0.416 \quad .8409 - \mathbf{i}0.0964]$, $\mathbf{h}_2 = [-1.1777 + \mathbf{i}0.1235 \quad 0.2034 + \mathbf{i}0.5132 \quad 0.8421 + \mathbf{i}1.5437 \quad -0.0266 + \mathbf{i}0.0806]$.

encoding scheme used [20]: we modify the scheme proposed in [20], for a DPC scheme applied to a Wyner-type model, to our setup. Thus, in this case, the backhaul is used to carry quantized versions of the signals to be transmitted by each transmitter.

We focus our derivations on the case where $N_t = 1$. Without quantization and for a linear precoding scheme, the signal to be transmitted from both base stations can be written as

$$\mathbf{x} = \mathbf{w}_1 s_1 + \mathbf{w}_2 s_2 \in \mathbb{C}^{2N_t} = \mathbb{C}^2 \tag{6.15}$$

Let x_j be the signal to be transmitted from base station j. Given the precoding used,

$$x_j \sim \mathcal{CN}\left(0, |w_{1j}|^2 + |w_{2j}|^2\right), \tag{6.16}$$

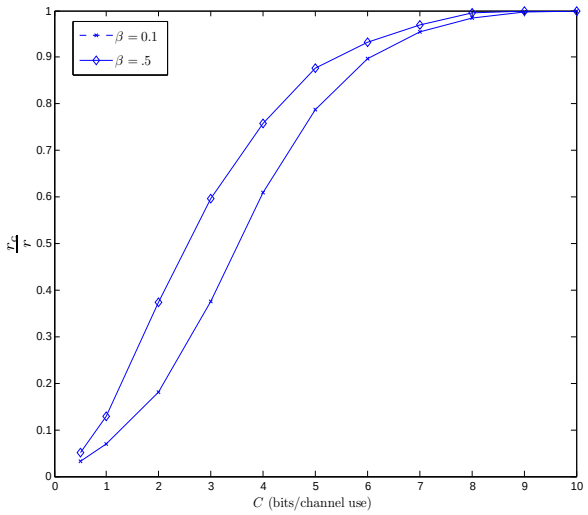


Figure 6.3: Average common to total rate Ratios vs. Backhaul capacity for Max Min Rates, $N_t = 1$, SNR = 10dB, $\mathbf{h}_{k,k} \sim \mathcal{CN}(0, \mathbf{I}_{N_t})$ and $\mathbf{h}_{k,\bar{k}} \sim \mathcal{CN}(0, \beta \mathbf{I}_{N_t})$.

where $\mathbf{w}_k = [w_{k1}; w_{k2}]$ is the precoding vector carrying user k's symbols, as per our previous notation.

Similar to [20], we adopt, for each x_j, a forward test channel of the form

$$\hat{x}_j = x_j + q_j, \tag{6.17}$$

where x_j and q_j are independent complex Gaussian random variables with zero means and variances $|w_{1j}|^2 + |w_{2j}|^2$ and $\sigma_{q_j}^2$, respectively. q_j models the quantization noise.

Given the backhaul constraints,

$$I(\hat{x}_j, x_j) = \log_2 \left(1 + \frac{\sigma_{x_j}^2}{\sigma_{q_j}^2}\right) = \log_2 \left(1 + \frac{|w_{1j}|^2 + |w_{2j}|^2}{\sigma_{q_j}^2}\right) \leq C_j \tag{6.18}$$

Thus

$$\sigma_{q_j}^2 \geq \frac{|w_{1j}|^2 + |w_{2j}|^2}{2^{C_j} - 1} \tag{6.19}$$

As we would like to minimize the quantization noise, this should be met with equality. Also, given the power constraints, we have that

$$\sigma_{\hat{x}_j}^2 = |w_{1j}|^2 + |w_{2j}|^2 + \sigma_{q_j}^2 \leq P_j. \tag{6.20}$$

Combining (6.19) and (6.20), we get

$$\left(|w_{1j}|^2 + |w_{2j}|^2\right)\left(1 + \frac{1}{2^{C_j}-1}\right) = \left(|w_{1j}|^2 + |w_{2j}|^2\right)\frac{2^{C_j}}{2^{C_j}-1} \leq P_j. \quad (6.21)$$

The signal received at user k is thus[2]

$$y_k = \sum_{j=1}^{2} h_{k,j}\left[x_j + q_j\right] + n_k = \sum_{j=1}^{2} h_{k,j} w_{kj} s_k + z_k \quad (6.22)$$

where

$$z_k = n_k + \sum_{j=1}^{2} h_{k,j}\left[w_{\bar{k}j} s_{\bar{k}} + q_j\right]. \quad (6.23)$$

Thus

$$r_k \leq \log_2\left(1 + \frac{\left|\sum_{j=1}^{2} h_{k,j} w_{kj}\right|^2}{\sigma^2 + \left|\sum_{j=1}^{2} h_{k,j} w_{\bar{k}j}\right|^2 + \sum_{j=1}^{2} |h_{k,j}|^2 \sigma_{q_j}^2}\right) \quad (6.24)$$

$$= \log_2\left(1 + \frac{\left|\sum_{j=1}^{2} h_{k,j} w_{kj}\right|^2}{\sigma^2 + \left|\sum_{j=1}^{2} h_{k,j} w_{\bar{k}j}\right|^2 + \sum_{j=1}^{2} |h_{k,j}|^2 \frac{|w_{1j}|^2 + |w_{2j}|^2}{2^{C_j}-1}}\right) \quad (6.25)$$

6.5.1 Rate region boundary

Similar to Section 6.4, the achievable rate region boundary corresponding to the above scheme may be obtained by varying a parameter α in $[0, 1]$ corresponding to the rate split between user 1 and 2, and solving a sum rate optimization problem as follows

$$\begin{aligned} &\text{max. } r \\ &\text{s.t. } r_1 \geq \alpha r, \quad r_2 \geq (1-\alpha) r \\ &\qquad (r_1, r_2) \in \mathcal{R}_{Quant}, \end{aligned} \quad (6.26)$$

where $\mathcal{R}_{Quant}$ is defined as the set of (r_1, r_2) pairs satisfying (6.25) and the power constraints (6.21).

The above problem (6.26) can also be solved with a line search over r. Specifying r determines both r_1 and r_2. The constraint (6.25) can then be transformed into a convex constraint on $\mathbf{w}_1$ and $\mathbf{w}_2$. The feasibility of a given (r_1, r_2) pair can thus once again be established by solving a convex optimization problem.

[2]Recall $\bar{k}$ is used to denote the 'other' user/base station.

6.5.2 Numerical Results

Figures 6.4 and 6.5 compare the two different ways of using the backhaul for two different channel instances: depending on the strength of the interfering links and on the backhaul constraints, one or the other scheme will be better. In Figure 6.4, the interfering links are quite strong (for user 2, the link from transmitter 1 is stronger than its 'direct link' from transmitter 2), hence the benefit of joint transmission, even if accompanied by quantization noise.

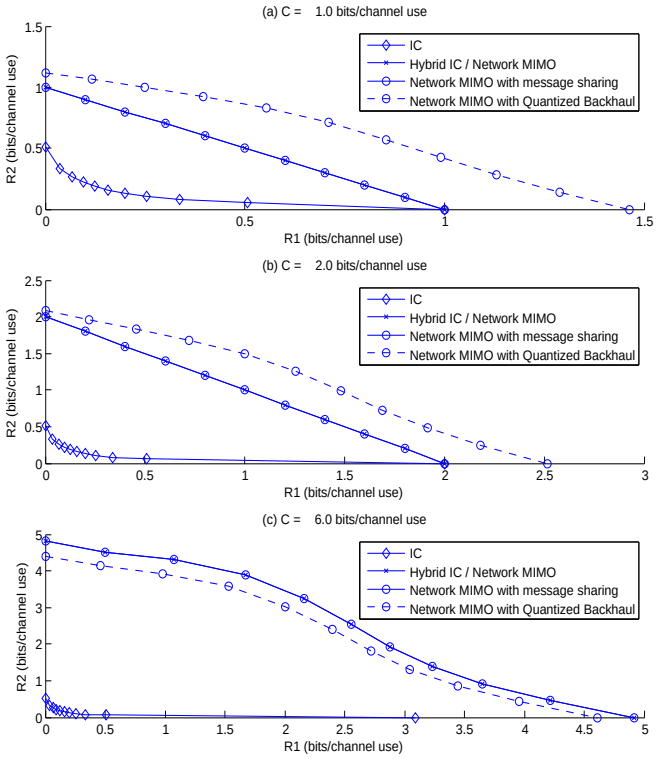


Figure 6.4: Sample Rate Regions Comparison for 10dB SNR, $\mathbf{h}_1 = [-0.3059 - \mathbf{i}0.8107 \quad 0.0886 + \mathbf{i}0.8409]$ and $\mathbf{h}_2 = [-1.1777 + \mathbf{i}0.8421 \quad 0.2034 - \mathbf{i}0.0266]$.

6.6 Conclusion

In this chapter, we proposed to use the backhaul capacity to convey different types of messages: private messages transmitted from the serving

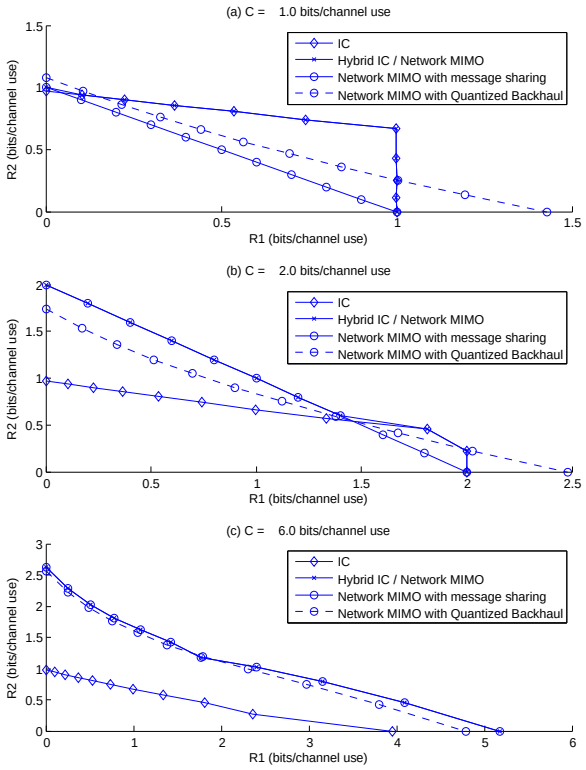


Figure 6.5: Sample Rate Regions Comparison for 10dB SNR, $\mathbf{h}_1 = [-0.8152 + \mathbf{i}0.8872 \quad -0.2150 - \mathbf{i}0.6359]$ and $\mathbf{h}_2 = [0.3489 - \mathbf{i}0.2163 \quad -0.2714 + \mathbf{i}0.1499]$.

base station only, and common messages jointly transmitted from several base stations. A corresponding achievable rate region for the two-cell setup was characterized and simulations have illustrated the benefit of the hybrid scheme considered. This approach to cooperation was also compared to a scheme in which a central processor designs the network MIMO transmission and then forwards to each BS a quantized version of the designed signal, which the latter then transmits.

6.A MAC with common message

For convenience, we reproduce the following result from [73], initially obtained by Slepian and Wolf [70], where $I(.;.)$ denotes mutual information and $I(.;.|.)$ denotes conditional mutual information:

Theorem 5 (MAC with a common message)**.** *The sources put out statistically independent messages W_0, W_1, W_2 with nR_0, nR_1, nR_2 bits, respectively. The message W_0 is seen by both encoders and is called the* common *message, whereas W_1 and W_2 appear only at the respective encoders 1 and 2, i.e. are* private *to those encoders. Encoder 1 maps (w_0, w_1) to a sequence $x_1^n \in \mathcal{X}_1^n$, encoder 2 maps (w_0, w_2) to a sequence $x_2^n \in \mathcal{X}_2^n$, and the channel $P_{Y|X_1,X_2}(.)$ puts out a sequence $y^n \in \mathcal{Y}^n$. Consider a distribution $P_{UX_1X_2Y}$ that factors as $P_U P_{X_1|U} P_{X_2|U} P_{Y|X_1X_2}$. The following rate region, denoted $\mathcal{R}(P_U, P_{X_1|U}, P_{X_2|U})$, is achievable:*

$$\begin{aligned} R_1 &< I(X_1; Y|X_2, U), \\ R_2 &< I(X_2; Y|X_1, U), \\ R_1 + R_2 &< I(X_1, X_2; Y|U), \\ R_0 + R_1 + R_2 &< I(X_1, X_2; Y). \end{aligned} \tag{6.27}$$

One can further restrict attention to $|\mathcal{U}| \leq \min(|\mathcal{Y}| + 3, |\mathcal{X}_1||\mathcal{X}_2| + 2)$. The capacity of the thus defined MAC is the union of such regions,

$$\mathcal{C}_{\text{MAC}} = \bigcup_{P_U, P_{X_1|U}, P_{X_2|U}} \mathcal{R}(P_U, P_{X_1|U}, P_{X_2|U}). \tag{6.28}$$

With $R_2 = 0$, i.e. (6.27) simplifies to:

$$\begin{aligned} R_1 &< I(X_1; Y|X_2, U), \\ R_0 + R_1 &< I(X_1, X_2; Y). \end{aligned} \tag{6.29}$$

Recall that in the setup considered,

$$y_k = \mathbf{h}_{k,1}\mathbf{x}_{k,1} + \mathbf{h}_{k,2}\mathbf{x}_{k2} + z_k, \tag{6.30}$$

where the receiver noise $z_k \sim \mathcal{CN}(0, \sigma_k^2)$.

Thus, using (6.29), the following rates will be achievable (implicit is the conditioning on the CSIR, which is assumed perfect):

$$r_{k,p} < I(\mathbf{x}_{kk}; y_k|\mathbf{x}_{k\bar{k}}, s_{k,c}) \tag{6.31}$$

$$r_{k,p} + r_{k,c} < I(\mathbf{x}_{kk}, \mathbf{x}_{k\bar{k}}; y_k), \tag{6.32}$$

One can easily verify that the mutual information expressions are the ones in (6.9).

Chapter 7

Large System Analysis for Beamforming Design

7.1 Introduction

Throughout the last few chapters, consideration of interference has lead us to examine the system level architectural issue of cooperation between base stations in the network. We have considered the MISO interference channel and looked at a couple of cooperative schemes that take into consideration different CSIT at the transmitters, as well as the backhaul constraints. In this chapter, we consider obtaining intuitions about the system level performance of different cooperative strategies based on large system analysis.

When multiple antennas are incorporated at each BS, we can tradeoff the maximization of the rates of the in-cell users (ignoring interference), with the minimization of the interference spilled over into the other cells. If enough cooperation is enabled, these two objectives can go hand in hand.

A large body of research has recently dealt with cooperation and coordination in multicell systems. Many papers are concerned with developing new algorithms, to meet various proposed performance objectives (e.g. transmit power minimization under given SINR constraints for coordinated beamforming [14, 15], or minimum SINR maximization for network MIMO [10]). Others analyze a given scheme under a specific channel model.

One can distinguish between setups where base stations each serve a

different group of users, and cases where all the transmitters jointly transmit to all users in the system, the so-called multicell processing (MCP), macrodiversity or network MIMO setups. For MCP, a classical model for performance analysis is Wyner's model [12, 13]. For example, [13] characterize the sum rate for single-antenna base stations pooling their antennas to perform ZF to the users in the system: they consider a circular variant of the infinite linear Wyner model for both non-fading and fading setups, with scheduling based on local channel statistics. Results are derived for the number of cooperating BSs tending to infinity, and scaling results are further obtained by letting the number of users per cell do the same.

In the present chapter, we focus on optimizing linear precoding under different states of CSI and data sharing. Thus, we consider that each base station may serve more than one user (even in the case where there is no cooperation at all, unlike in the MISO interference channel case) and formalize the problem of maximizing the minimum network-wide achievable rate, i.e. rate balancing, under the following three different setups:

- Single Cell Processing (SCP) in which each BS has perfect CSI about mobiles in its cell, but no knowledge about the channels to the mobiles in other cells. The BS can control the interference between the mobiles in its own cell, but not the interference spill-over to other cells.

- Coordinated beamforming (CBf), in which the beamforming decisions at each BS take into account the impact of interference on the other cells. In this approach, each BS has CSI about the channels to its own users, but it also knows the channels to the mobiles in the other cells (similar to the CSI assumption in Chapter 4).

- MCP, in which the BSs cooperate to jointly precode signals to all the mobiles in all cells. Through cooperation, the BSs know the channels from all BSs to all the mobiles in the whole network.

In the first two setups, user data is routed to a single BS only, whereas in the MCP case, it is routed to all cooperating BSs.

We specialize our derivations to considering only two cells. Nevertheless, such an approach could be extended to larger networks, if one wishes to perform system-wide optimization of a large cellular network.

7.1.1 Asymptotic approach

Random matrix theory has received a lot of attention in the communications literature recently [74], particularly large system results where the dimen-

sions of some of the system parameters, such as number of users, the length of the chip sequence in CDMA systems, the number of transmit and receive antennas in MIMO links, are allowed to grow towards infinity at similar rates: such an approach often allows for a compact characterization of performance, while enables asymptotic system optimization. The attractive feature of such these results is that the asymptotic expressions turn out to be good estimates for even relatively low values of the parameters [75–79].

We are thus lead to apply random matrix theory results to our two cell model and take a large system approach in which the number of antennas at the base station N_t and that of mobiles in each cell K both grow large together, while the ratio $\frac{K}{N_t}$, which we refer to as cell loading, goes to a finite constant, denoted β. We start by formulating the solution to the rate balancing optimization problems in the finite system size. A key to obtaining the optimum and corresponding optimal beamforming strategies relies on expressing the problem in terms of its *dual*.

7.1.2 Duality

Many duality results have been noted in the context of MIMO communications. In the context of the capacity region of the Gaussian BC, duality results are given in [80] and [81] for example. When linear beamforming is considered, as in this chapter, a series of duality results were also obtained, and used to come up with iterative algorithms for beamforming design. As stated in Chapter 4, one of the first references to UL-DL duality is [23], in the context of joint power control and beamforming schemes in the downlink of a multiple antenna system employing simple linear transmission strategies followed by single user receivers such that the SINR at each mobile is above a target value; the proposed algorithm achieves a feasible solution for the DL if there is one and minimizes the total transmitted power in the network.

In [82], the authors show that not only is a feasible solution obtainable from duality results, but also the optimal solution for the DL total power minimization problem under given target SINR constraints. They derive their results from Lagrangian duality, whereby the transmit powers of the dual UL are found to correspond to the Lagrange coefficients associated with SINR constraint.

[83] also showed that the achievable rate regions are the same in the UL and the DL subject to a sum power constraint, and that effective bandwidth results derived for the UL can thus also be applied to the DL. A number of iterative algorithms was also derived based on duality. In [84], the authors consider among other things, the problem of maximizing the sum of effective

bandwidths in the UL and by duality, the DL. For the two-user case, a closed form solution may be obtained, whereas the more general problem can be formulated as a convex optimization problem and solved via an interior point method for example. Related problems such as SINR balancing are also considered. As noted, the mean effective bandwidth provides an indication of how good the transmission is in a mean sense.

In [51], Lagrangian duality is also used to show the UL-DL duality in a single-cell DL and to derive alternative algorithms for solving the power minimization problem. In [14, 15], a similar approach generalizes UL-DL duality to the multicell setting. We will use a similar derivation below.

7.1.3 Contributions

For the adopted channel model, we solve the dual and primal optimization problems in the asymptotic regime. This allows us to characterize asymptotic performance as a function of cell loading β, and the strength of the 'interfering' links ϵ. This provides us with a simple mathematical model that enables the optimization of system parameters, as well as the comparison of the different schemes: specifically, we show how the effect of interference at a user is captured by an 'effective interference' term, as in [75]. We also find that the large system beamforming matrices may be related to transmit beamforming based on regularized ZF (RZF), as described in [85]. The large system analysis of regularized beamforming for a single isolated cell setup was performed in [86], where the limiting SINR is calculated as the number of antennas and the number of users grow large with their ratio going to a finite constant. Finally, Monte Carlo simulations show how closely the obtained expressions approximate the finite-size system performance.

7.2 System Model

Our model, as illustrated in Figure 7.1, has two cells, each with a base station (BS) equipped with an array of N_t antennas. and K single-antenna mobiles. We use a flat fading channel model, and adopt the following notation regarding channel coefficient vectors:

- $\mathbf{h}_{k,\bar{j},j} \in \mathbb{C}^{1\times N_t}$ denotes the channel vector from BS j to cell $\bar{j}$'s user k;
- $\tilde{\mathbf{h}}_{k,\bar{j}} = [\mathbf{h}_{k,\bar{j},1} \quad \mathbf{h}_{k,\bar{j},2}] \in \mathbb{C}^{1\times 2N_t}$.

The data symbols to be transmitted to each user are assumed zero-mean unit variance complex circularly symmetric Gaussian, $\mathcal{CN}(0,1)$, and user k

in cell j's symbol is denoted $s_{k,j}$; $\mathbf{s}_j = [s_{1,j} \ \dots \ s_{K,j}]^T$ and $\mathbf{s} = [\mathbf{s}_1^T \ \mathbf{s}_2^T]^T$. The received signal at user k in cell j is given by

$$y_{k,j} = \sum_{\bar{j}=1}^{2} \mathbf{h}_{k,j,\bar{j}} \mathbf{x}_{\bar{j}} + n_{k,j}, \tag{7.1}$$

where $\mathbf{x}_{\bar{j}} \in \mathbb{C}^{N_t}$ denotes BS $\bar{j}$'s signal, which consists of the linearly precoded symbols of the users it is serving, and which is subject to the average power constraint $\mathbb{E}\left[\mathbf{x}_{\bar{j}}^H \mathbf{x}_{\bar{j}}\right] \leq P$ and $n_{k,j} \sim \mathcal{CN}(0, \sigma^2)$ is the noise at the receiver.

Channels between a user and his serving BS are i.i.d. for all users, and $\mathcal{CN}(0,1)$, whereas channels between users and the interfering BS are i.i.d. and $\mathcal{CN}(0,\epsilon)$. Thus ϵ controls the level of interference between neighboring cells.

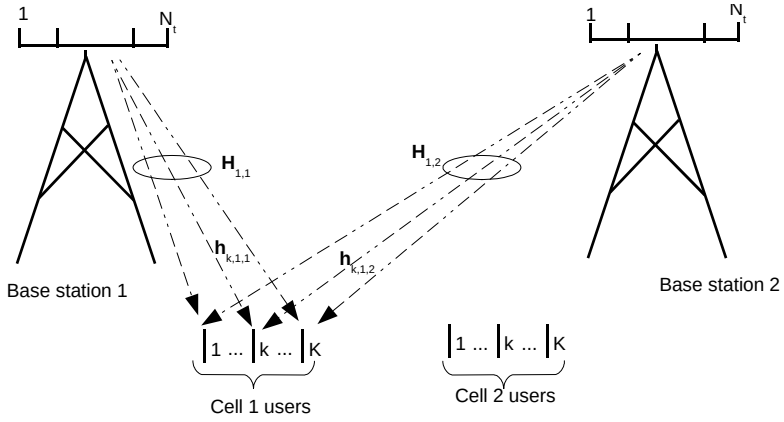


Figure 7.1: System model

7.3 Cooperation and Linear Beamforming Schemes

We consider the problem of maximizing the network-wide minimum achievable rate for three different degrees of cooperation and coordination between the cells. The optimization problem to be solved in each case is presented, along with a sketch of the solution. Similar problems have been treated in the literature [14, 54] and can be reduced to solving a series of convex optimization problems. In the following, when BS j is being discussed, unless otherwise specified, $\bar{j} = \mod(j, 2) + 1$, the index of the other BS.

7.3.1 SCP

This is the conventional case where each BS serves its own users without bothering about what is happening in the other cell. We assume full re-use of time and spectrum across cells. In cell j the problem to be solved is

$$\begin{aligned} \max._{\gamma,\mathbf{w}_{kj},k=1,\ldots,K} \quad & \gamma \\ \text{s.t.} \quad & \text{SINR}_{k,j} \geq \gamma, \quad k = 1,\ldots,K \\ & \sum_{k=1}^{K} \|\mathbf{w}_{kj}\|^2 \leq P. \end{aligned} \tag{7.2}$$

The SINR at user k in cell j is given by

$$\text{SINR}_{k,j} = \frac{|\mathbf{h}_{k,j,j}\mathbf{w}_{kj}|^2}{\sigma_{k,j}^2 + \sum_{\bar{k}=1}^{K} |\mathbf{h}_{k,j,j}\mathbf{w}_{\bar{k}j}|^2}, \tag{7.3}$$

where $\sigma_{k,j}^2$ is noise plus other cell interference term given by:

$$\sigma_{k,j}^2 = \sigma^2 + \sum_{\bar{k}=1}^{K} |\mathbf{h}_{k,j,\bar{j}}\mathbf{w}_{\bar{k}\bar{j}}|^2. \tag{7.4}$$

Such a scheme may require an iterative procedure, since the beamforming at each base station will influence the interference at the other, and therefore its beamforming as well. We focus on finding the maximum SINR that can be met simultaneously in both cells. We obtain this using a bisection method. For fixed γ we obtain the beamforming vectors by minimizing total transmit power subject to SINR constraints on all users in the cell. If the optimum value is less than P for both base stations, γ is attainable.

$$\begin{aligned} \min._{\mathbf{w}_{jk},j=1,\ldots,K} \quad & \sum_{k=1}^{K} \|\mathbf{w}_{kj}\|^2 \\ \text{s.t.} \quad & \text{SINR}_{k,j} \geq \gamma, \quad k = 1,\ldots,K. \end{aligned} \tag{7.5}$$

7.3.2 Coordinated Beamforming

In this approach, each BS sends data to its own users only, as in SCP, but here the CSI is shared between the BSs and the interference generated to

users in other cells can be taking into consideration. (7.2) becomes:

$$\begin{aligned} \max._{\gamma,\mathbf{w}_{kj},k=1,\ldots,K,j=1,2} \quad & \gamma \\ \text{s.t.} \quad & \mathrm{SINR}_{k,j} \geq \gamma, \quad k=1,\ldots,K, j=1,2 \\ & \sum_{k=1}^{K} \|\mathbf{w}_{kj}\|^2 \leq P, \quad j=1,2. \end{aligned} \tag{7.6}$$

The SINR at user k in cell j is given by:

$$\mathrm{SINR}_{k,j} = \frac{|\mathbf{h}_{k,j,j}\mathbf{w}_{kj}|^2}{\sigma^2 + \sum_{\bar{j}=1}^{2}\sum_{\bar{k}=1,(\bar{k},\bar{j})\neq(k,j)}^{K} |\mathbf{h}_{k,j,\bar{j}}\mathbf{w}_{\bar{k}\bar{j}}|^2}. \tag{7.7}$$

Here too, the above problem may be solved by the bisection method. To determine feasibility of a given γ, we solve (as in [14]) the following problem:

$$\begin{aligned} \min._{\phi>0,\mathbf{w}_{kj},k=1,\ldots,K,j=1,2} \quad & \phi \sum_{j=1}^{2} P \\ \text{s.t.} \quad & \mathrm{SINR}_{k,j} \geq \gamma, \quad k=1,\ldots,K, j=1,2 \qquad (7.8) \\ & \sum_{k=1}^{K} \mathbf{w}_{kj}^H \mathbf{w}_{kj} \leq \phi P, \quad j=1,2. \qquad (7.9) \end{aligned}$$

7.3.3 MCP

This is the case where the base stations cooperate fully: both CSI and data is available at all transmitters, who pool their antennas together to serve the users jointly.

$$\begin{aligned} \max._{\gamma,\mathbf{w}_{kj},j=1,\ldots,K,k=1,2} \quad & \gamma \\ \text{s.t.} \quad & \mathrm{SINR}_{k,j} \geq \gamma, \quad k=1,\ldots,K, j=1,2 \qquad (7.10) \\ & \sum_{j=1}^{2}\sum_{k=1}^{K} \|\mathbf{E}_{\bar{j}}\mathbf{w}_{kj}\|^2 \leq P, \bar{j}=1,2 \qquad (7.11) \end{aligned}$$

$\mathbf{E}_j, j=1,2$ is a diagonal matrix which selects the elements of each beamforming vector corresponding to base station j (i.e. its non-zero diagonal elements occupy locations $(j-1)N_t+1$ to jN_t).

The MCP SINR at user k in cell j is given by:

$$\mathrm{SINR}_{k,j} = \frac{|\tilde{\mathbf{h}}_{k,j}\mathbf{w}_{kj}|^2}{\sigma^2 + \sum_{\bar{j}=1}^{2}\sum_{\bar{k}=1,(\bar{k},\bar{j})\neq(k,j)}^{K} |\tilde{\mathbf{h}}_{k,j}\mathbf{w}_{\bar{k}\bar{j}}|^2}. \tag{7.12}$$

Once again, the above problem may be solved by the bisection method. To determine feasibility of a given γ, we solve (similar to [51]):

$$\begin{aligned}
\text{min.} \quad & \phi \sum_{j=1}^{2} P \\
\text{s.t.} \quad & \text{SINR}_{k,j} = \frac{|\tilde{\mathbf{h}}_{k,j}\mathbf{w}_{kj}|^2}{\sigma^2 + \sum_{\bar{j},\bar{k},(\bar{j},\bar{k})\neq(j,k)} |\tilde{\mathbf{h}}_{k,j}\mathbf{w}_{\bar{k}\bar{j}}|^2} \geq \gamma, \quad k = 1,\ldots,K, j = 1,2 \\
& \sum_{\bar{j}=1}^{2}\sum_{\bar{k}=1}^{K} \|\mathbf{E}_j \mathbf{w}_{\bar{k}\bar{j}}\|^2 \leq \phi P. \qquad (7.13)
\end{aligned}$$

7.4 Large system analysis

As noted in the introduction, we resort to large system analysis in which the number of antennas at each base station and the number of mobiles in each cell grow to infinity, their ratio $\frac{K}{N_t}$, which we also refer to as the cell loading, converging to a constant β. For all cases, we consider the dual of the original (primal) problem [55]: the optimal beamforming vectors in each case can be shown (details are given in Appendix 7.A, 7.B and 7.C for single cell processing, coordinated beamforming and joint transmission) to be of the form:

$$\mathbf{w}_{kj}^{SCP} = \sqrt{\delta_{kj}^{SCP}} \left(\mathbf{I}_{N_t} + \sum_{\bar{k}\neq k} \lambda_{\bar{k}j}^{SCP} \mathbf{h}_{\bar{k},j,j}^H \mathbf{h}_{\bar{k},j,j} \right)^{-1} \mathbf{h}_{k,j,j}^H$$

$$\mathbf{w}_{kj}^{Coord} = \sqrt{\delta_{kj}^{Coord}} \left[\mu_j^{Coord} \mathbf{I}_{N_t} + \sum_{(\bar{j},\bar{k})\neq(j,k)} \lambda_{\bar{k}\bar{j}}^{Coord} \mathbf{h}_{\bar{k},\bar{j},j}^H \mathbf{h}_{\bar{k},\bar{j},j} \right]^{-1} \mathbf{h}_{k,j,j}^H$$

$$\mathbf{w}_{kj}^{MCP} = \sqrt{\delta_{kj}^{MCP}} \left[\sum_{\bar{j},\bar{k},(\bar{j},\bar{k})\neq(j,k)} \lambda_{\bar{k}\bar{j}}^{MCP} \tilde{\mathbf{h}}_{\bar{k},\bar{j}}^H \tilde{\mathbf{h}}_{\bar{k},\bar{j}} + \sum_{\bar{j}=1}^{2} \mu_{\bar{j}}^{MCP} \mathbf{E}_{\bar{j}} \right]^{-1} \tilde{\mathbf{h}}_{k,j}^H$$

where the μ_j's and $\lambda_{k,j}$'s correspond to the Lagrange coefficients in the dual problems and may be interpreted as noise powers and UL transmit powers on dual ULs, respectively. Large system analysis is undertaken to obtain a simplification, while still reaching meaningful conclusions for the finite-size system. We thus solve the asymptotic dual problems, obtaining asymptotically optimal noise and UL power levels. This is done in Appendix

7.D, 7.E and 7.F. For feasible γ, we find the optimal dual parameters in terms of the system parameters, namely ϵ and the cell loading parameter β. We then obtain the corresponding maximum feasible γ for each scheme. Finally, we characterize the optimum β in terms of total network throughput maximization while maintaining a given SINR across all users. This is done in the following series of theorems.

7.4.1 Asymptotically Optimal Beamformers

Theorem 6 (Asymptotically optimal beamforming for SCP)**.** *Assume* $\beta\left(\frac{\gamma}{1+\gamma}+\epsilon\gamma\right)<1$. *To asymptotically achieve SINR* γ *at each mobile terminal in the two-cell network as* $N_t\to\infty$ *with* $\frac{K}{N_t}\to\beta>0$, *beamforming vectors of the following form are asymptotically optimal:*

$$\mathbf{w}_{kj}^{SCP}=\sqrt{\frac{\bar{p}}{N_t}}\frac{\hat{\mathbf{w}}_{kj}^{SCP}}{\|\hat{\mathbf{w}}_{kj}^{SCP}\|}, \tag{7.14}$$

where
$$\hat{\mathbf{w}}_{kj}^{SCP}=\left(\mathbf{I}_{N_t}+\frac{\bar{\lambda}}{N_t}\sum_{\bar{k}\neq k}\mathbf{h}_{\bar{k},j,j}^{H}\mathbf{h}_{\bar{k},j,j}\right)^{-1}\mathbf{h}_{k,j,j}^{H} \tag{7.15}$$

and $\bar{p}$, $\bar{\lambda}$ *and* γ *are linked to each other via*

$$\bar{p}=\frac{\sigma^2\gamma}{1-\beta\left(\frac{\gamma}{1+\gamma}+\gamma\epsilon\right)}, \tag{7.16}$$

and
$$\gamma=\frac{1}{\frac{1}{\bar{\lambda}}+\frac{\beta}{1+\gamma}}=\frac{1}{\frac{\sigma^2}{\bar{p}}+\epsilon\beta+\frac{\beta}{1+\gamma}}. \tag{7.17}$$

Proof. See Appendix 7.D. □

Corollary 2. *Subject to per base station power constraint* P, *the maximum asymptotic network-wide achievable SINR for a given cell loading factor* β *is the unique solution to the following fixed point equation:*

$$\gamma_{SCP}^{*}=\frac{1}{\beta}\frac{1}{\frac{\sigma^2}{P}+\epsilon+\frac{1}{1+\gamma_{SCP}^{*}}}. \tag{7.18}$$

Note 2 (Relation to regularized zero-forcing)**.** *Note that the above optimal beamformers asymptotically correspond to a precoding matrix at each of the two transmitters given by*

$$\mathbf{W}_j^{SCP} = c_j \left[\mathbf{I}_{N_t} + \frac{\bar{\lambda}}{N_t} \mathbf{H}_{j,j}^H \mathbf{H}_{j,j} \right]^{-1} \mathbf{H}_{j,j}^H, \tag{7.19}$$

where c_j is such that the power constraint is met with equality at BS_j, and $\mathbf{H}_{j,j}$ is the concatenation of the channels between the users in cell j and their serving base station. This is the regularized zero-forcing scheme proposed in [85] and studied in the asymptotic regime in [86].

Theorem 7 (Asymptotically optimal beamforming for CBf)**.** *Assume $\beta \left(\frac{\gamma}{1+\gamma} + \frac{\epsilon\gamma}{1+\epsilon\gamma} \right) < 1$. To asymptotically achieve SINR γ at each mobile terminal in the two-cell network as $N_t \to \infty$ with $\frac{K}{N_t} \to \beta > 0$, beamforming vectors of the following form are asymptotically optimal:*

$$\mathbf{w}_{kj}^{Coord} = \sqrt{\frac{\bar{p}}{N_t}} \frac{\hat{\mathbf{w}}_{kj}^{Coord}}{\|\hat{\mathbf{w}}_{kj}^{Coord}\|}, \tag{7.20}$$

$$\textit{where} \qquad \hat{\mathbf{w}}_{kj}^{Coord} = \left(\mathbf{I}_{N_t} + \frac{\bar{\lambda}}{N_t} \sum_{\bar{j},\bar{k},(\bar{j},\bar{k})\neq(j,k)} \mathbf{h}_{\bar{k},j,j}^H \mathbf{h}_{\bar{k},j,j} \right)^{-1} \mathbf{h}_{k,j,j}^H, \tag{7.21}$$

and $\bar{p}$, $\bar{\lambda}$ and γ are related to each other by

$$\frac{\bar{p}}{\bar{\lambda}} = \sigma^2, \tag{7.22}$$

$$\textit{and} \qquad \gamma = \frac{1}{\frac{1}{\bar{\lambda}} + \frac{\beta}{1+\gamma} + \frac{\beta\epsilon}{1+\epsilon\gamma}} = \frac{1}{\frac{\sigma^2}{\bar{p}} + \frac{\beta}{1+\gamma} + \frac{\beta\epsilon}{1+\epsilon\gamma}}. \tag{7.23}$$

Proof. See Appendix 7.E. □

Corollary 3. *Subject to per base station power constraint P, the maximum asymptotic network-wide achievable SINR for a given cell loading factor β is the unique solution to the following fixed point equation:*

$$\gamma_{Coord}^* = \frac{1}{\beta} \frac{1}{\frac{\sigma^2}{P} + \frac{1}{1+\gamma_{Coord}^*} + \frac{\epsilon}{1+\epsilon\gamma_{Coord}^*}}. \tag{7.24}$$

Theorem 8 (Asymptotically optimal beamforming for MCP)**.** *Suppose* $\frac{\beta\gamma}{1+\gamma} < 1$. *To asymptotically achieve SINR* γ *at each mobile terminal in the two-cell network as* $N_t \to \infty$ *with* $\frac{K}{N_t} \to \beta > 0$, *beamforming vectors of the following form are asymptotically optimal:*

$$\mathbf{w}_{kj}^{MCP} = \sqrt{\frac{\bar{p}}{N_t}} \frac{\hat{\mathbf{w}}_{kj}^{MCP}}{\|\hat{\mathbf{w}}_{kj}^{MCP}\|}, \tag{7.25}$$

where

$$\hat{\mathbf{w}}_{kj}^{MCP} = \left(\mathbf{I}_{2N_t} + \frac{\bar{\lambda}}{N_t} \sum_{\bar{j},\bar{k},(\bar{j},\bar{k})\neq(j,k)} \tilde{\mathbf{h}}_{\bar{k},j}^H \tilde{\mathbf{h}}_{\bar{k},j} \right)^{-1} \tilde{\mathbf{h}}_{k,j}^H \tag{7.26}$$

and $\bar{p}$, $\bar{\lambda}$ *and* γ *are linked to each other via*

$$\frac{\bar{p}}{\bar{\lambda}} = \sigma^2, \tag{7.27}$$

$$\textit{and} \qquad \gamma = \frac{1}{\frac{1}{(1+\epsilon)\bar{\lambda}} + \frac{\beta}{1+\gamma}} = \frac{1}{\frac{\sigma^2}{\bar{p}(1+\epsilon)} + \frac{\beta}{1+\gamma}}. \tag{7.28}$$

Proof. See Appendix 7.F. □

Corollary 4. *Subject to per base station power constraint* P, *the maximum asymptotic network-wide achievable SINR for a given cell loading factor* β *is the unique solution to the following fixed point equation:*

$$\gamma_{MCP}^* = \frac{1}{\beta} \frac{1}{\frac{\sigma^2}{(1+\epsilon)P} + \frac{1}{1+\gamma_{MCP}^*}}. \tag{7.29}$$

7.4.2 Effective interference

The optimal SINR expressions above are striking in how they capture the effect of interference in the three different beamformers. Indeed, they supply a simple "effective interference" characterization, which can be used to directly check if a particular target SINR can be achieved.

It is natural to compare the schemes directly using the limiting SINR expressions. This is accomplished in the following theorem, where $SNR \triangleq \frac{P}{\sigma^2}$.

Theorem 9. *Let $\gamma^*_{SCP}, \gamma^*_{Coord}, \gamma^*_{MCP}$ denote the SINR's under SCP, coordinated beamforming, and MCP, respectively. Then*

$$\gamma^*_{SCP} < \gamma^*_{Coord} < \gamma^*_{MCP}. \tag{7.30}$$

At signal to noise ratio, SNR, and interference level, ϵ, denote the effective interference at target SINR, γ, by:

$$I_{\mathit{eff}}\left(SNR, \epsilon, \gamma\right) = \begin{cases} \beta\left(1 + \frac{SNR}{1+\gamma} + \epsilon SNR\right) & \mathit{SCP} \\ \beta\left(1 + \frac{SNR}{1+\gamma} + \frac{\epsilon SNR}{1+\epsilon\gamma}\right) & \mathit{CBf} \\ \beta\left(1 + \frac{SNR}{1+\gamma} + \frac{\epsilon SNR}{1+\gamma}\right) & \mathit{MCP} \end{cases} \tag{7.31}$$

Then the feasibility of γ in the case of SCP, or CBf, is equivalent to satisfaction of the inequality $\dfrac{SNR}{I_{\mathit{eff}}\left(SNR, \epsilon, \gamma\right)} > \gamma$, and in the MCP case, it is equivalent to $\dfrac{(1+\epsilon)SNR}{I_{\mathit{eff}}\left(SNR, \epsilon, \gamma\right)} > \gamma$.

Proof. Follows closely the proof of Proposition 3.2 in [75]. □

We note the close parallel with the notion of effective interference that arises in the large system analysis of linear UL multiuser receivers [75].

7.4.3 Asymptotically optimal cell loading

The above theorems characterized the optimal SINR under the various schemes considered for fixed cell loading β. Figure 7.2 shows, for $\epsilon = .5$, the achievable rates per cell per antenna for different values of the SNR for the coordinated beamforming scheme. Clearly there are system-level gains to be attained by selecting the optimal β. This conclusion holds for the other schemes as well. Our next step is to characterize this optimum loading. This corresponds to finding the β that maximizes the normalized (by the number of antennas) rate per cell r, i.e. solving the following problem:

$$\text{maximize}_\beta \; r = \beta \log(1 + \gamma^*) \tag{7.32}$$

with γ^* characterized by the appropriate fixed-point equation (cf. Eqs (7.18), (7.24) and (7.29)).

Theorem 10 (Characterization of the optimum β for SCP)**.** *If*

$$\epsilon + \frac{\sigma^2}{P} \geq 1, \tag{7.33}$$

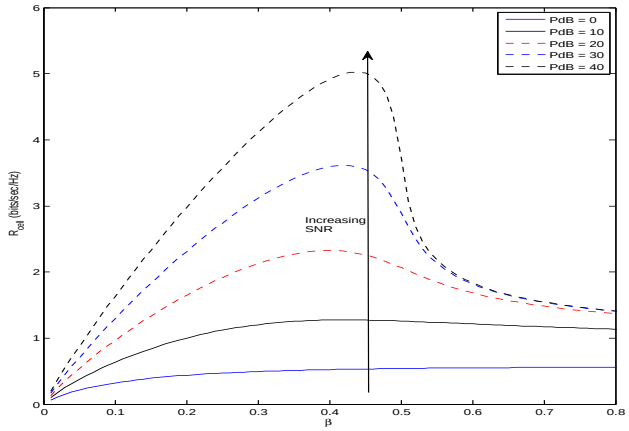


Figure 7.2: Achievable rates by asymptotically optimal coordinated beamforming for $\epsilon = .5$ and different SNR values.

$\beta^* \to \infty$. *Otherwise, the optimum occurs at a finite* β^* *which may be found by a line search.*

Proof. Refer to Appendix 7.G. □

Theorem 11 (Optimal cell loading for CBf)**.** *If* $\frac{\sigma^2}{P} + \epsilon - 2\epsilon^2 - 1 \geq 0$*, one can increase* β *indefinitely. Otherwise, there is a finite value at which* r *is maximized.*

Proof. The proof follows along similar lines to that of Theorem 10, although the algebra is more tedious. Refer to Appendix 7.H. □

Theorem 12 (Characterization of the optimum β for MCP)**.** *If*

$$\frac{\sigma^2}{P} \geq (1+\epsilon), \tag{7.34}$$

$\beta^* \to \infty$. *Otherwise, the optimum occurs at a finite* β^* *that may be found by a line search.*

Proof. (7.29) is the same as (7.18) with $\epsilon + \frac{\sigma^2}{P} \leftarrow \frac{\sigma^2}{P(1+\epsilon)}$. Performing this substitution in (7.33) yields the result. □

The above results define for each scheme a noise-limited region, where cell loading can be increased indefinitely; however, this leads to ever decreasing rates per user, not to mention that more user channels need to be learned.

7.5 Performance Results

How do these schemes compare with each other and with other approaches from the literature? Clearly, coordinated beamforming requires more CSI than SCP, and MCP involves BS cooperation, so it is not surprising that the SINR's are ordered as in (7.30); in this section, we obtain numerical results to provide a quantitative comparison between these schemes in different scenarios. Intercell interference can be avoided altogether by applying the classic principle of re-use partitioning. In this section, we consider the time division (TD) scheme in which each base station is given a separate time-slot, which we shall also call "1/2-reuse".

7.5.1 When is 1/2-reuse SCP better than CBf?

Let β_{TD} be the cell loading in the 1/2-reuse scheme in which each BS transmits half the time. To compare with CBf (a full re-use scheme), let $\beta := \beta_{Coord} := \beta_{TD}/2$. Then the rate for the TD scheme is

$$r_{TD}(\beta) = \beta \log\left(1 + \gamma^*_{TD}(\beta)\right), \tag{7.35}$$

$$\text{with} \quad \gamma^*_{TD}(\beta) = \frac{1}{2\beta} \frac{1}{\frac{\sigma^2}{2P} + \frac{1}{1+\gamma^*_{TD}(\beta)}} = \frac{1}{\beta} \frac{1}{\frac{\sigma^2}{P} + \frac{2}{1+\gamma^*_{TD}(\beta)}}. \tag{7.36}$$

The rate using coordinated beamforming is given by:

$$r_{Coord}(\beta) = \beta \log\left(1 + \gamma^*_{Coord}(\beta)\right) \tag{7.37}$$

$$\text{with} \quad \gamma^*_{Coord} = \frac{1}{\beta} \frac{1}{\frac{\sigma^2}{P} + \frac{1}{1+\gamma^*_{Coord}(\beta)} + \frac{\epsilon}{1+\epsilon\gamma^*_{Coord}(\beta)}} \tag{7.38}$$

Thus, $r_{Coord}(\beta) > r_{TD}(\beta)$ if $\epsilon < 1$, and $r_{Coord}(\beta) < r_{TD}(\beta)$ if $\epsilon > 1$. It follows that coordinated beamforming is only useful when $\epsilon < 1$; otherwise, it is better to partition the cells with a reuse factor of 1/2. Of course, if base station association is performed properly, ϵ should be less than 1.

7.5.2 Nulling interference using pure ZF

Finally, we compare the schemes proposed to non-optimized ZF beamforming in the different setups. Thus we can consider two forms of pure ZF in the context of single cell processing. SCP-ZF stands for ZF the same-cell

interference, with the BS oblivious to other cell interference. Generalized ZF (GZF) is when the BSs zero force the interference in the two-cell system. It is clear that SCP-ZF will do better than GZF when the level of intercell interference is low, but the situation will reverse when the intercell interference is high, and the numerical results in the subsection below confirm this. In the MCP setting, the two BSs can jointly zero force all the interference in the two-cell system, and we denote this case by "macro-ZF".

7.5.3 Numerical results

Figures 7.3-7.5 compare the different schemes by varying the cell loading β. We notice that when ϵ is small, CBf gains little over SCP, but offers significant gains compared to pure ZF or 1/2-reuse. When ϵ is small, SCP-ZF is superior to GZF, as expected. When ϵ is large, but < 1 (*e.g.* $\epsilon = 0.8$), coordinated beamforming gains significantly over SCP, but does not gain much compared to 1/2-reuse, or to GZF (if the loading can be perfectly optimized). Note that in this case, the relevant comparison is with GZF.

When $\epsilon < 1$, CBf is always better, for appropriately selected β, than SCP, SCP-ZF, and GZF. Compared to the ZF schemes, it performs better across a wider range of cell loadings. In large networks, it avoids the intractable frequency planning problem associated with fractional re-use schemes.

In the two-cell model, MCP offers the most gain when ϵ is large. Even when ϵ is small, as in Figure 7.3, the gains over CBf, 1/2-reuse, and single cell ZF schemes, respectively, are significant, and in Figure 7.5 they are higher still, because ϵ is larger in that case. Unlike the other schemes, MCP improves with increasing ϵ, but it requires coordination between the base stations and the cost in terms of backhaul capacity not accounted for in this chapter.

Finally, we investigate the applicability of the asymptotic results to a finite system. In a first step, for $K = 3, N_t = 4$, $\frac{P}{\sigma^2} = 10$ and ϵ taking values in $[.01; .1; .5; .8; 1]$, we solve the optimization problems described in Section 7.3 for different independent samples of the channel and obtain the corresponding average rates. Even for such a low number of antennas, the large system analysis (LSA) results seem to provide quite a good approximation. The results are shown in Figure 7.6 below.

The optimization problems in 7.3 are time-consuming, particularly for the SCP case, which requires iterations between the optimization at the two transmitters until convergence. Thus, we consider applying the asymptotically optimal beamforming vectors from Section 7.4 (slightly modified so as not to break the per transmitter power constraint for the MCP scheme) in

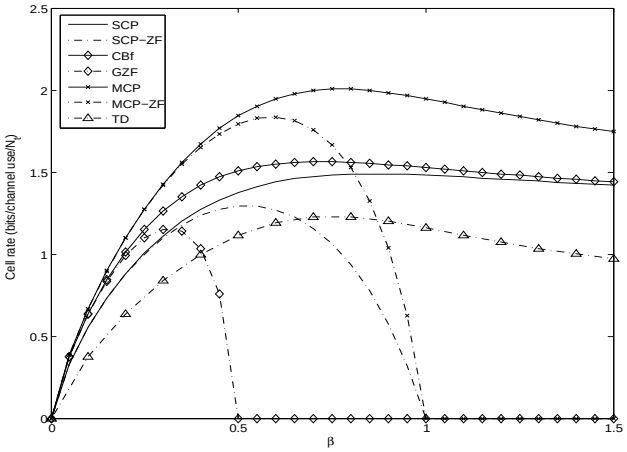


Figure 7.3: Effect of cell loading on rate achieved for SNR = 10dB, $\epsilon = .1$

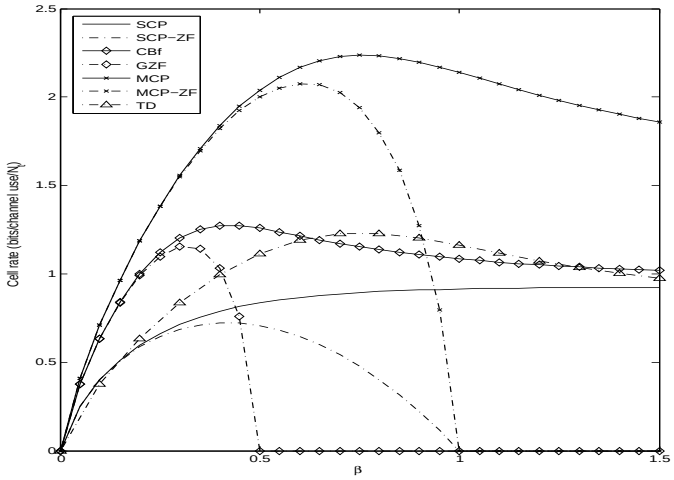


Figure 7.4: Effect of cell loading on rate achieved for SNR = 10dB, $\epsilon = .5$

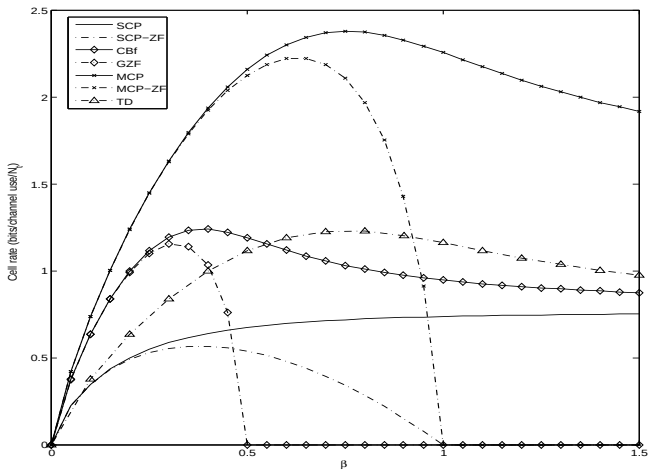


Figure 7.5: Effect of cell loading on rate achieved for SNR = 10dB, $\epsilon = .8$

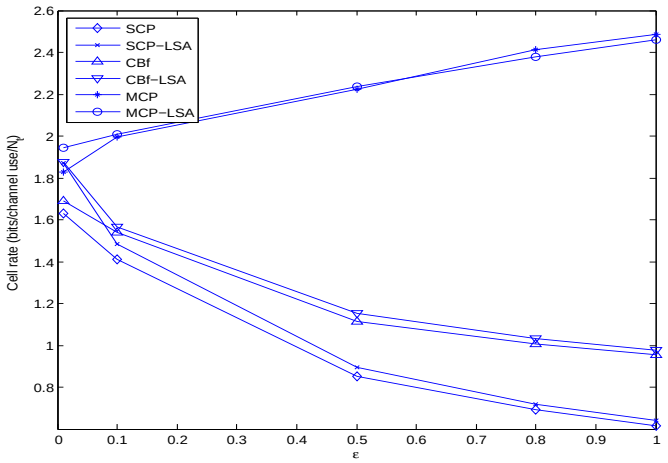


Figure 7.6: Comparison between average normalized cell rates achieved by the optimization problems in Section 7.3 and the large system analysis results for $K = 3, N_t = 4$ and SNR = 10 dB.

the finite system case. The results are shown in Figure 7.7, again for $N_t = 4$ and $K = 3$.

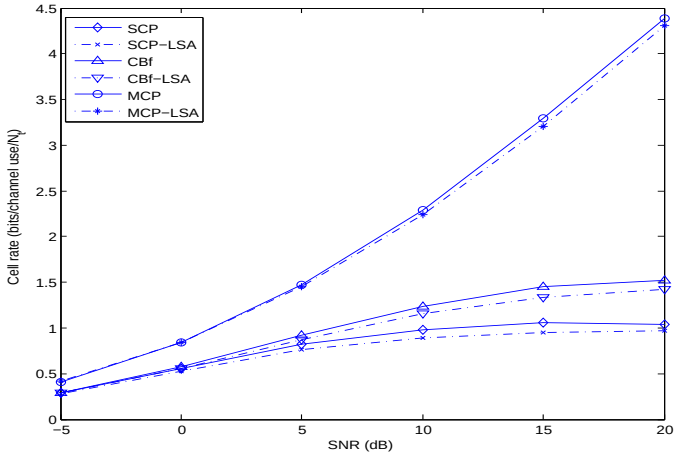


Figure 7.7: Comparison between average normalized cell rates achieved by applying the asymptotically optimal beamformers in Section 7.4 to the finite system case and the large system analysis results for $K = 3, N_t = 4$ and $\epsilon = 0.5$.

7.6 Conclusions

This chapter provided an asymptotic analysis of a two cell interfering network in which the number of antennas at the base stations, and the number of users in the cells, both grow large together. Schemes that balance rates across users in the system for three levels of cooperation, namely SCP, CBf and MCP, were compared. MCP is shown to offer significant rate gains, if we are willing to accommodate the increased coordination and communication between base stations required by this scheme, which is not accounted for here. We characterize and compare the limiting SINR's and bitrates of the three schemes. In particular, we formulate a notion of *effective interference* and show how it can be used to decide if a given SINR target is feasible. We also optimize the cell loading. In most scenarios, there is an optimal cell loading, which can be computed easily from the asymptotic formulas. However, if the noise level is sufficiently high, and the interference level not too great, the rate can be an increasing function of the system loading. We characterize exactly when the system has a finite optimal loading and when

it does not. In Section 7.5, we characterize exactly when the coordinated beamforming scheme with full re-use beats half reuse, and how it compares with unoptimized zero forcing schemes.

The validity of the obtained results for finite size systems was verified via Monte Carlo simulations.

7.A SCP Dual Problem

Rewriting the SINR constraint corresponding to user k in cell j as

$$\frac{1}{\gamma}\left|\mathbf{h}_{k,j,j}\mathbf{w}_{kj}\right|^2 \geq \sum_{\bar{k}\neq k}\left|\mathbf{h}_{k,j,j}\mathbf{w}_{\bar{k}j}\right|^2 + \sigma_{k,j}^2, \tag{7.39}$$

the Lagrangian of the modified primal (7.5) is given by

$$\sum_k \lambda_{kj}\sigma_{k,j}^2 + \sum_k \mathbf{w}_{kj}^H \left[\mathbf{I}_{N_t} - \frac{\lambda_{kj}}{\gamma}\mathbf{h}_{k,j,j}^H\mathbf{h}_{k,j,j} + \sum_{\bar{k}\neq k}\lambda_{\bar{k}j}\mathbf{h}_{\bar{k},j,j}^H\mathbf{h}_{\bar{k},j,j}\right]\mathbf{w}_{kj},$$

where λ_{kj}, $k = 1,\ldots,K$ denote the Lagrange multipliers associated with the cell's users' SINR constraints.
The corresponding dual problem will be given by

$$\begin{aligned}\max.{}_{\lambda_{kj},k=1,\ldots,K} \quad & \sum_{k=1}^{K}\lambda_{kj}\sigma_{k,j}^2 \\ \text{s.t.} \quad & \mathbf{I}_{N_t} - \frac{\lambda_{kj}}{\gamma}\mathbf{h}_{k,j,j}^H\mathbf{h}_{k,j,j} + \sum_{\bar{k}\neq k}\lambda_{\bar{k}j}\mathbf{h}_{\bar{k},j,j}^H\mathbf{h}_{\bar{k},j,j} \succeq 0, \quad k = 1,\ldots,K.\end{aligned}$$

If feasible, this will have the same optimal values as the dual uplink problem:

$$\begin{aligned}\min.{}_{\hat{\mathbf{w}}_{kj},\lambda_{kj},k=1,\ldots,K} \quad & \sum_k \lambda_{kj}\sigma_{k,j}^2 \\ \text{s.t.} \quad & \max_{\hat{\mathbf{w}}_{kj}} \frac{\lambda_{kj}\left|\mathbf{h}_{k,j,j}\hat{\mathbf{w}}_{kj}\right|^2}{\hat{\mathbf{w}}_{kj}^H\hat{\mathbf{w}}_{kj} + \sum_{\bar{k}\neq k}\lambda_{\bar{k}j}\left|\mathbf{h}_{\bar{k},j,j}\hat{\mathbf{w}}_{kj}\right|^2} \geq \gamma, \quad k = 1,\ldots,K.\end{aligned}$$

The optimal $\hat{\mathbf{w}}_{kj}$, up to a scalar multiplication are given by

$$\hat{\mathbf{w}}_{kj} = \left(\mathbf{I}_{N_t} + \sum_{\bar{k}\neq k}\lambda_{\bar{k}j}\mathbf{h}_{\bar{k},j,j}^H\mathbf{h}_{\bar{k},j,j}\right)^{-1}\mathbf{h}_{k,j,j}^H, \tag{7.40}$$

whereas the optimal $\mathbf{w}_{kj}$ for the primal problem are

$$\mathbf{w}_{kj} = \sqrt{\delta_{kj}}\hat{\mathbf{w}}_{kj} = \sqrt{\frac{p_{kj}}{N_t}}\frac{\hat{\mathbf{w}}_{kj}}{\|\hat{\mathbf{w}}_{kj}\|}, \tag{7.41}$$

where $\delta_{kj} \in \mathbb{R}_+$ is a scaling factor, and $\frac{p_{kj}}{N_t}$ is the transmit power allocated to $\mathbf{w}_{kj}$. From the SINR constraints:

$$\frac{p_{kj}}{N_t\gamma}\frac{|\mathbf{h}_{k,j,j}\hat{\mathbf{w}}_{kj}|^2}{\|\hat{\mathbf{w}}_{kj}\|^2} - \sum_{\bar{k}\neq j}\frac{p_{\bar{k}j}}{N_t}\frac{\left|\mathbf{h}_{k,j,j}\hat{\mathbf{w}}_{\bar{k}j}\right|^2}{\|\hat{\mathbf{w}}_{\bar{k}j}\|^2} = \sigma_{k,j}^2, \quad k = 1,\ldots,K. \tag{7.42}$$

Assuming the optimal λ_{kj}'s determined, the $\frac{p_{kj}}{N_t}$ can be determined from this set of equations.

7.B CBf Dual Problem

The SINR constraint corresponding to user k in cell j can be rewritten as:

$$\frac{1}{\gamma}|\mathbf{h}_{k,j,j}\mathbf{w}_{kj}|^2 \geq \sigma^2 + \sum_{\bar{j},\bar{k},(\bar{k},\bar{j})\neq(k,j)}\left|\mathbf{h}_{k,j,\bar{j}}\mathbf{w}_{\bar{k}\bar{j}}\right|^2. \tag{7.43}$$

The Lagrangian of the modified primal (7.9) is given by:

$$\begin{aligned}&\sum_{j=1}^{2}\sum_{k=1}^{K}\mathbf{w}_{kj}^H\left[\mu_j\mathbf{I}_{N_t} - \frac{\lambda_{kj}}{\gamma}\mathbf{h}_{k,j,j}^H\mathbf{h}_{k,j,j} + \sum_{(\bar{j},\bar{k})\neq(j,k)}\lambda_{\bar{k}\bar{j}}\mathbf{h}_{\bar{k},\bar{j},j}^H\mathbf{h}_{\bar{k},\bar{j},j}\right]\mathbf{w}_{kj}\\ &+\phi\sum_{j=1}^{2}(1-\mu_j)P + \sum_{j,k}\lambda_{kj}\sigma^2,\end{aligned} \tag{7.44}$$

where λ_{kj}, is the Lagrange multiplier associated with cell j's user k's SINR constraint, and μ_j, $j = 1, 2$ those associated with the BSs' power constraint.

The Lagrangian dual is given by:

$$\begin{aligned}\max._{\lambda_{kj}\geq 0,\mu_j\geq 0} \quad & \sum_{j,k}\lambda_{kj}\sigma^2\\ \text{s.t.} \quad & \mu_j\mathbf{I}_{N_t} - \frac{\lambda_{kj}}{\gamma}\mathbf{h}_{k,j,j}^H\mathbf{h}_{k,j,j} + \sum_{(\bar{j},\bar{k})\neq(j,k)}\lambda_{\bar{k}\bar{j}}\mathbf{h}_{\bar{k},\bar{j},j}^H\mathbf{h}_{\bar{k},\bar{j},j} \succeq 0,\\ & \qquad k = 1,\ldots,K, j = 1,2\\ & \sum_{j=1}^{2}(1-\mu_j)P = 0.\end{aligned}$$

This can be shown to be equivalent to a dual uplink problem with uncertain noise (cf. [51]):

$$\begin{aligned}\min._{\lambda_{kj}\geq 0,\mu_j\geq 0} \quad & \sum_{k,j}\lambda_{kj}\sigma^2 \\ \text{s.t.} \quad & \mu_j\mathbf{I}_{N_t}-\frac{\lambda_{kj}}{\gamma}\mathbf{h}_{k,j,j}^H\mathbf{h}_{k,j,j}+\sum_{(\bar{j},\bar{k})\neq(j,k)}\lambda_{\bar{k}\bar{j}}\mathbf{h}_{\bar{k},\bar{j},j}^H\mathbf{h}_{\bar{k},\bar{j},j}\preceq 0, \\ & k=1,\ldots,K, j=1,2 \\ & \sum_{j=1}^{2}(1-\mu_j)P=0.\end{aligned}$$

Or equivalently

$$\begin{aligned}\min._{\lambda_{kj}\geq 0,\mu_j\geq 0,\hat{\mathbf{w}}_{kj}} \quad & \sum_{k,j}\lambda_{kj}\sigma^2 \\ \text{s.t.} \quad & \gamma\leq\frac{\lambda_{kj}|\hat{\mathbf{w}}_{kj}\mathbf{h}_{k,j,j}|^2}{\hat{\mathbf{w}}_{kj}^H\left[\mu_j\mathbf{I}_{N_t}+\sum_{(\bar{j},\bar{k})\neq(j,k)}\lambda_{\bar{k}\bar{j}}\mathbf{h}_{\bar{k},\bar{j},j}^H\mathbf{h}_{\bar{k},\bar{j},j}\right]\hat{\mathbf{w}}_{kj}}, \\ & k=1,\ldots,K, j=1,2 \\ & \sum_{j=1}^{2}(1-\mu_j)P=0.\end{aligned}$$

The optimal beamforming vectors on the dual uplink, assuming feasibility, are (up to a scaling factor):

$$\hat{\mathbf{w}}_{kj}=\left[\mu_j\mathbf{I}_{N_t}+\sum_{(\bar{j},\bar{k})\neq(j,k)}\lambda_{\bar{k}\bar{j}}\mathbf{h}_{\bar{k},\bar{j},j}^H\mathbf{h}_{\bar{k},\bar{j},j}\right]^{-1}\mathbf{h}_{k,j,j}^H \tag{7.45}$$

whereas the optimal beamforming vectors on the downlink are of the form

$$\mathbf{w}_{kj}=\sqrt{\delta_{kj}}\hat{\mathbf{w}}_{kj}=\sqrt{\frac{p_{kj}}{N_t}}\frac{\hat{\mathbf{w}}_{kj}}{\|\hat{\mathbf{w}}_{kj}\|}, \tag{7.46}$$

the δ_{kj} and $\frac{p_{kj}}{N_t}$ being defined as in the SCP case. From DL SINR constraints, for $j=1,2, k=1,\ldots,K$,

$$\frac{p_{kj}}{N_t\gamma}\frac{|\mathbf{h}_{k,j,j}\hat{\mathbf{w}}_{kj}|^2}{\|\hat{\mathbf{w}}_{kj}\|^2}-\sum_{\bar{j},\bar{k},(\bar{k},\bar{j})\neq(k,j)}\frac{p_{\bar{k}\bar{j}}}{N_t}\frac{\left|\mathbf{h}_{k,j,\bar{j}}\hat{\mathbf{w}}_{\bar{k}\bar{j}}\right|^2}{\|\hat{\mathbf{w}}_{\bar{k}\bar{j}}\|^2}=\sigma^2. \tag{7.47}$$

7.C MCP Dual Problem

The SINR expression may be rewritten similarly to (7.43) and the Lagrangian of the modified primal (7.13) is equal to

$$\phi P \sum_{j=1}^{2}(1-\mu_j) + \sum_{j=1}^{2}\sum_{k=1}^{K}\lambda_{kj}\sigma^2 + \sum_{j=1}^{2}\sum_{k=1}^{K}\mathbf{w}_{kj}^H\left[\sum_{\bar{j},\bar{k},(\bar{j},\bar{k})\neq(j,k)}\lambda_{\bar{k}\bar{j}}\tilde{\mathbf{h}}_{\bar{k},\bar{j}}^H\tilde{\mathbf{h}}_{\bar{k},\bar{j}} - \frac{\lambda_{kj}}{\gamma}\tilde{\mathbf{h}}_{k,j}^H\tilde{\mathbf{h}}_{k,j} + \sum_{\bar{j}=1}^{2}\mu_{\bar{j}}\mathbf{E}_{\bar{j}}\right]\mathbf{w}_{kj}.$$

The KKT conditions are given by

$$\sum_{j=1}^{2}(1-\mu_j) = 0,$$

$$\left[\sum_{\bar{j},\bar{k},(\bar{j},\bar{k})\neq(j,k)}\lambda_{\bar{k}\bar{j}}\tilde{\mathbf{h}}_{\bar{k},\bar{j}}^H\tilde{\mathbf{h}}_{\bar{k},\bar{j}} - \frac{\lambda_{kj}}{\gamma}\tilde{\mathbf{h}}_{k,j}^H\tilde{\mathbf{h}}_{k,j} + \sum_{\bar{j}=1}^{2}\mu_{\bar{j}}\mathbf{E}_{\bar{j}}\right]\mathbf{w}_{kj} = \mathbf{0},$$

$$\lambda_{kj}\left[\sigma^2 + \sum_{\bar{j},\bar{k},(\bar{j},\bar{k})\neq(j,k)}|\tilde{\mathbf{h}}_{k,j}\mathbf{w}_{\bar{k}\bar{j}}|^2 - \frac{1}{\gamma}|\tilde{\mathbf{h}}_{k,j}\mathbf{w}_{kj}|^2\right] = 0,$$

$$\mu_j\left[\sum_{\bar{j}=1}^{2}\sum_{\bar{k}=1}^{K}\|\mathbf{E}_j\mathbf{w}_{\bar{k}\bar{j}}\|^2 - \phi P\right] = 0.$$

The dual problem is equivalent to

$$\begin{aligned}\max._{\mu_j\geq 0,\lambda_{kj}\geq 0,\hat{\mathbf{w}}_{kj}} \quad & \sum_{j=1}^{2}\sum_{k=1}^{K}\lambda_{kj}\sigma^2 \\ \text{s.t.} \quad & \gamma \geq \max_{\hat{\mathbf{w}}_{kj}} \frac{\lambda_{kj}|\tilde{\mathbf{h}}_{k,j}\hat{\mathbf{w}}_{kj}|^2}{\hat{\mathbf{w}}_{kj}^H\left[\sum_{\bar{j},\bar{k},(\bar{j},\bar{k})\neq(j,k)}\lambda_{\bar{k}\bar{j}}\tilde{\mathbf{h}}_{\bar{k},\bar{j}}^H\tilde{\mathbf{h}}_{\bar{k},\bar{j}} + \sum_{\bar{j}=1}^{2}\mu_{\bar{j}}\mathbf{E}_{\bar{j}}\right]\hat{\mathbf{w}}_{kj}}, \\ & \qquad k=1,\ldots,K, j=1,2 \\ & \sum_{j=1}^{2}(1-\mu_j) = 0. \end{aligned} \tag{7.48}$$

Plugging in the optimal $\hat{\mathbf{w}}_{jk}$

$$\hat{\mathbf{w}}_{kj} = \left[\sum_{\bar{j},\bar{k},(\bar{j},\bar{k})\neq(j,k)} \lambda_{\bar{k}\bar{j}} \tilde{\mathbf{h}}_{\bar{k},\bar{j}}^H \tilde{\mathbf{h}}_{\bar{k},\bar{j}} + \sum_{\bar{j}=1}^{2} \mu_{\bar{j}} \mathbf{E}_{\bar{j}} \right]^{-1} \tilde{\mathbf{h}}_{k,j}^H, \tag{7.49}$$

the target SINR constraint becomes:

$$\gamma \geq \lambda_{kj} \tilde{\mathbf{h}}_{k,j} \left[\sum_{\bar{j},\bar{k},(\bar{j},\bar{k})\neq(j,k)} \lambda_{\bar{k}\bar{j}} \tilde{\mathbf{h}}_{\bar{k},\bar{j}}^H \tilde{\mathbf{h}}_{\bar{k},\bar{j}} + \sum_{\bar{j}=1}^{2} \mu_{\bar{j}} \mathbf{E}_{\bar{j}} \right]^{-1} \tilde{\mathbf{h}}_{k,j}^H. \tag{7.50}$$

Also, as the duality gap is zero if the primal is feasible, at the optimum

$$\phi \sum_{j=1}^{2} P = \sum_{j=1}^{2} \sum_{k=1}^{K} \lambda_{kj} \sigma^2. \tag{7.51}$$

Moreover, as in the previous two cases, the beamforming vectors in the original problem and in the dual problem may be related as follows:

$$\mathbf{w}_{kj} = \sqrt{\delta_{kj}} \hat{\mathbf{w}}_{kj} = \sqrt{\frac{p_{kj}}{N_t}} \frac{\hat{\mathbf{w}}_{kj}}{\|\hat{\mathbf{w}}_{kj}\|}. \tag{7.52}$$

Plugging these into the DL SINR constraints provides the solution for the $\frac{p_{kj}}{N_t}$:

$$\frac{p_{kj}}{N_t \gamma} \frac{|\tilde{\mathbf{h}}_{k,j} \hat{\mathbf{w}}_{kj}|^2}{\|\hat{\mathbf{w}}_{kj}\|^2} - \sum_{(\bar{k},\bar{j})\neq(k,j)} \frac{p_{\bar{k},\bar{j}}}{N_t} \frac{|\tilde{\mathbf{h}}_{k,j} \hat{\mathbf{w}}_{\bar{k}\bar{j}}|^2}{\|\hat{\mathbf{w}}_{\bar{k}\bar{j}}\|^2} = \sigma^2. \tag{7.53}$$

7.D Large System Analysis for SCP

Before proceeding, we note that in [87], alternative proofs of the convergence of the optimal dual uplink powers in the asymptotic regime to constants for the SCP, CBf and MCP cases are provided.

Using the optimal uplink beamformers yields SINR values given by:

$$\mathrm{SINR}_{k,j} = \lambda_{kj} \mathbf{h}_{k,j,j} \left(\mathbf{I} + \sum_{\bar{k}\neq k} \lambda_{\bar{k}j} \mathbf{h}_{\bar{k},j,j}^H \mathbf{h}_{\bar{k},j,j} \right)^{-1} \mathbf{h}_{k,j,j}^H. \tag{7.54}$$

Thus the uplink power levels satisfy the following fixed point equation

$$\lambda_{kj} = \frac{\mathrm{SINR}_{k,j}}{\mathbf{h}_{k,j,j}\left(\mathbf{I} + \sum_{\bar{k}\neq k} \lambda_{\bar{k}j}\mathbf{h}_{\bar{k},j,j}^H \mathbf{h}_{\bar{k},j,j}\right)^{-1}\mathbf{h}_{k,j,j}^H}. \tag{7.55}$$

Introducing $\bar{\lambda}_{kj} = N_t\lambda_{kj}$, this is equivalent to

$$\bar{\lambda}_{kj} = \frac{\mathrm{SINR}_{k,j}}{\frac{1}{N_t}\mathbf{h}_{k,j,j}\left(\mathbf{I} + \frac{1}{N_t}\sum_{\bar{k}\neq k} \bar{\lambda}_{\bar{k}j}\mathbf{h}_{\bar{k},j,j}^H \mathbf{h}_{\bar{k},j,j}\right)^{-1}\mathbf{h}_{k,j,j}^H} \triangleq I_{kj}^{SCP}\left(\bar{\boldsymbol{\lambda}}\right), \tag{7.56}$$

where $\bar{\boldsymbol{\lambda}} = \left\{\bar{\lambda}_{kj}\right\} \in \mathbb{R}_+^K$.

We can show that for any k, $I_{kj}^{SCP}\left(\bar{\boldsymbol{\lambda}}\right)$ is an interference function [1], as:

1. Clearly, $I_{kj}^{SCP}(\bar{\boldsymbol{\lambda}}) > 0$ as long as $SINR_{kj} > 0$.
2. If $\bar{\boldsymbol{\lambda}}^{(1)} \leq \bar{\boldsymbol{\lambda}}^{(2)}$, the eigenvalues of $\left(\mathbf{I} + \frac{1}{N_t}\sum_{\bar{k}\neq k} \bar{\lambda}_{\bar{k}j}\mathbf{h}_{\bar{k},j,j}^H \mathbf{h}_{\bar{k},j,j}\right)^{-1}$ will be larger than or equal to those of $\left(\mathbf{I} + \frac{1}{N_t}\sum_{\bar{k}\neq k} \bar{\lambda}'_{\bar{k}j}\mathbf{h}_{\bar{k},j,j}^H \mathbf{h}_{\bar{k},j,j}\right)^{-1}$ and consequently $I_{kj}^{SCP}(\bar{\boldsymbol{\lambda}}^{(1)}) \leq I_{kj}^{SCP}(\bar{\boldsymbol{\lambda}}^{(2)})$.
3. For $\alpha > 1$,

$$\begin{aligned} I_{kj}^{SCP}(\alpha\bar{\boldsymbol{\lambda}}) &= \frac{\alpha\mathrm{SINR}_{k,j}}{\frac{1}{N_t}\mathbf{h}_{k,j,j}\left(\frac{1}{\alpha}\mathbf{I} + \frac{1}{N_t}\sum_{\bar{k}\neq k} \bar{\lambda}_{\bar{k}j}\mathbf{h}_{\bar{k},j,j}^H \mathbf{h}_{\bar{k},j,j}\right)^{-1}\mathbf{h}_{k,j,j}^H} \\ &< \frac{\alpha\mathrm{SINR}_{k,j}}{\frac{1}{N_t}\mathbf{h}_{k,j,j}\left(\mathbf{I} + \frac{1}{N_t}\sum_{\bar{k}\neq k} \bar{\lambda}_{\bar{k}j}\mathbf{h}_{\bar{k},j,j}^H \mathbf{h}_{\bar{k},j,j}\right)^{-1}\mathbf{h}_{k,j,j}^H} = \alpha I_{kj}^{SCP}(\bar{\boldsymbol{\lambda}}). \end{aligned} \tag{7.57}$$

Thus, by analogy with the standard power control algorithm in [88], if the set of SINR targets is feasible, then for any initial $\bar{\boldsymbol{\lambda}}$ vector, an iterative algorithm that at the nth stage, updates $\bar{\lambda}_{jk}$ as

$$\bar{\lambda}_{jk}^{(n)} = I_{kj}^{SCP}(\bar{\boldsymbol{\lambda}}^{(n-1)}),$$

will converge to the unique fixed point of the K set of equations. In particular, let $\bar{\boldsymbol{\lambda}}^{(0)}$ be $\mathbf{0}$. Let γ denote the common SINR target (i.e. $\mathrm{SINR}_{k,j} = \gamma$ for all k) and focus on the asymptotic regime when $N_t, K \to \infty$ and $\frac{K}{N_t} \to \beta < \infty$:

[1] As defined in [88], an interference function $I(\bar{\boldsymbol{\lambda}})$ satisfies the following properties: i)[positivity] $I(\bar{\boldsymbol{\lambda}}) > 0$ for all $\bar{\boldsymbol{\lambda}} \geq 0$ ii) [monotonicity] if $\bar{\boldsymbol{\lambda}} \leq \bar{\boldsymbol{\lambda}}'$ then $I(\bar{\boldsymbol{\lambda}}) \leq I(\bar{\boldsymbol{\lambda}}')$, and iii)[scalability] $\alpha I(\bar{\boldsymbol{\lambda}}) > I(\alpha\bar{\boldsymbol{\lambda}})$ for $\alpha > 1$.

- At the first iteration,

$$\bar{\lambda}_{jk}^{(1)} = I_{kj}^{SCP}(\bar{\boldsymbol{\lambda}}^{(0)}) = I_{kj}^{SCP}(\mathbf{0}) = \frac{\gamma}{\frac{1}{N_t}\|\mathbf{h}_{k,j,j}\|^2}.$$

  In the asymptotic regime, this will tend to $\frac{\gamma}{\beta}$. I.e. at the end of the first iteration, as the system grows large, all scaled dual power levels will be the same. Let $\bar{\lambda}_j^{(1)} = \frac{\gamma}{\beta}$.

- At the second iteration,

$$\begin{aligned}\bar{\lambda}_{jk}^{(2)} &= I_{kj}^{SCP}(\bar{\boldsymbol{\lambda}}^{(1)}) = I_{kj}^{SCP}(\bar{\lambda}^{(1)}\mathbf{I}_K) \\ &= \frac{\gamma}{\frac{1}{N_t}\mathbf{h}_{k,j,j}\left(\mathbf{I} + \bar{\lambda}^{(1)}\frac{1}{N_t}\sum_{\bar{k}\neq k}\mathbf{h}_{\bar{k},j,j}^H\mathbf{h}_{\bar{k},j,j}\right)^{-1}\mathbf{h}_{k,j,j}^H}. \end{aligned} \tag{7.58}$$

  Making use of Lemma 1 in [89], and of the fact that the entries of $\mathbf{h}_{k,j,j}$ are $\mathcal{CN}(0,1)$, one can show that as $N_t, K \to \infty$, such that $\frac{K}{N_t} \to \beta$, the denominator in (7.58) tends to

$$\frac{1}{\bar{\lambda}^{(1)}}\frac{1}{N_t}\mathrm{Tr}\left(\frac{1}{\bar{\lambda}_j^{(1)}}\mathbf{I} + \frac{1}{N_t}\sum_{\bar{k}\neq k}\mathbf{h}_{\bar{k},j,j}^H\mathbf{h}_{\bar{k},j,j}\right)^{-1}. \tag{7.59}$$

  This in turn tends to $\frac{1}{\bar{\lambda}_j^{(1)}}m\left(-\frac{1}{\bar{\lambda}_j^{(1)}}\right)$, where $m(z)$ is the Stieltjes transform of the asymptotic eigenvalue distribution function of $\frac{1}{N_t}\sum_{k=1}^K\mathbf{h}_{k,j,j}^H\mathbf{h}_{k,j,j}$,

$$m(z) = \frac{1}{-z + \frac{\beta}{1+m(z)}}. \tag{7.60}$$

$$\text{Thus,} \qquad \bar{\lambda}_{jk}^{(2)} \to \frac{\gamma}{\frac{1}{\bar{\lambda}_j^{(1)}}m\left(-\frac{1}{\bar{\lambda}_j^{(1)}}\right)} \tag{7.61}$$

  We denote the left-hand side of (7.61) by $\bar{\lambda}_j^{(2)}$.

- Similar to the above derivation, we can show that at the i-th iteration,

$$\bar{\lambda}_j^{(i)} = \frac{\gamma\bar{\lambda}_j^{(i-1)}}{m\left(-\frac{1}{\bar{\lambda}_j^{(i-1)}}\right)}. \tag{7.62}$$

As the above scheme is known to converge, in the analyzed large system setting, $\bar{\lambda}_{jk}$ tends to the same value for each k, which we denote $\bar{\lambda}_j$, so that

$$\gamma = m\left(-\frac{1}{\bar{\lambda}_j}\right). \tag{7.63}$$

Thus to achieve target SINR γ for all users in both cells in the considered asymptotic regime, the average virtual uplink power will be the same in both cells $\bar{\lambda}_1 = \bar{\lambda}_2 = \bar{\lambda}$ and must satisfy the equation

$$\frac{1}{\gamma} m\left(-\frac{1}{\bar{\lambda}}\right) = 1. \tag{7.64}$$

Using (7.60), this allows us to write $\bar{\lambda}$ as

$$\bar{\lambda} = \frac{1}{\frac{1}{\gamma} - \frac{\beta}{1+\gamma}}. \tag{7.65}$$

We now turn to the DL primal problems. These are not independent of each other, we now turn to the interference plus noise power at user k in cell j, $\sigma^2_{k,j}$ as specified by (7.4). $\sigma^2_{k,j}$ can be written as

$$\sigma^2_{k,j} = \sigma^2 + \sum_{\bar{k}=1}^{K} \frac{p_{kj}}{N} |\mathbf{h}_{k,j,\bar{j}} \hat{\mathbf{w}}_{\bar{k}\bar{j}}|^2 / \|\hat{\mathbf{w}}_{\bar{k}\bar{j}}\|^2. \tag{7.66}$$

But we can show that

$$\begin{aligned} \frac{1}{N}|\mathbf{h}_{k,j,\bar{j}} \hat{\mathbf{w}}_{\bar{k}\bar{j}}|^2 &\sim \frac{\epsilon}{N} \mathbf{h}_{\bar{k},\bar{j},\bar{j}} \left(\mathbf{I} + \frac{\bar{\lambda}}{N} \sum_{l \neq \bar{k}} \mathbf{h}^H_{l,\bar{j},\bar{j}} \mathbf{h}_{l,\bar{j},\bar{j}}\right)^{-2} \mathbf{h}^H_{\bar{k},\bar{j},\bar{j}} \\ &\sim \frac{\epsilon}{N} \mathrm{Tr}\left(\mathbf{I} + \frac{\bar{\lambda}}{N} \sum_{l \neq \bar{k}} \mathbf{h}^H_{l,\bar{j},\bar{j}} \mathbf{h}_{l,\bar{j},\bar{j}}\right)^{-2} \to \epsilon \, \frac{d}{d\bar{\lambda}} m\left(-\frac{1}{\bar{\lambda}}\right). \end{aligned} \tag{7.67}$$

Similarly, it can be shown that $\frac{1}{N}\|\hat{\mathbf{w}}_{kj}\|^2$ (cf. Eq. (7.40)) will converge as follows:

$$\frac{1}{N}\|\hat{\mathbf{w}}_{kj}\|^2 \to \frac{d}{d\bar{\lambda}} m\left(-\frac{1}{\bar{\lambda}}\right). \tag{7.68}$$

Thus, asymptotically, $\sigma^2_{k,j}$ is independent of the user index k, and we will therefore write it as σ^2_j.

Combining the asymptotics given by (7.67) and (7.68) with

$$\begin{aligned}
&\frac{1}{N^2}\left|\mathbf{h}_{k,j,j}\hat{\mathbf{w}}_{kj}\right|^2 \\
&\sim \left|\frac{1}{N}\mathbf{h}_{k,j,j}\left[\mathbf{I}+\frac{1}{N}\bar{\lambda}\sum_{\bar{k}\neq k}\mathbf{h}_{\bar{k},j,j}^H\mathbf{h}_{\bar{k},j,j}\right]^{-1}\mathbf{h}_{k,j,j}^H\right|^2 \\
&\to \left(\frac{1}{\bar{\lambda}}m\left(-\frac{1}{\bar{\lambda}}\right)\right)^2,
\end{aligned} \tag{7.69}$$

and

$$\begin{aligned}
&\frac{1}{N}\left|\mathbf{h}_{k,j,j}\hat{\mathbf{w}}_{\bar{k}j}\right|^2 \\
&\sim \frac{1}{N}\frac{\mathbf{h}_{k,j,j}\left[\mathbf{I}+\frac{1}{N}\bar{\lambda}\sum_{l\neq\bar{k},k}\mathbf{h}_{l,j,j}^H\mathbf{h}_{l,j,j}\right]^{-2}\mathbf{h}_{k,j,j}^H}{\left(1+\frac{\bar{\lambda}}{N}\mathbf{h}_{k,j,j}\left[\mathbf{I}+\frac{1}{N}\bar{\lambda}\sum_{l\neq\bar{k},k}\mathbf{h}_{l,j,j}^H\mathbf{h}_{l,j,j}\right]^{-1}\mathbf{h}_{k,j,j}^H\right)^2} \\
&\to \frac{\frac{d}{d\bar{\lambda}}m\left(-\frac{1}{\bar{\lambda}}\right)}{\left(1+m\left(-\frac{1}{\bar{\lambda}}\right)\right)^2},
\end{aligned} \tag{7.70}$$

the DL SINR constraint equation (7.42) for $k=1,\ldots,K$, becomes

$$\frac{p_{kj}}{\gamma}\frac{\left(\frac{1}{\bar{\lambda}}m\left(-\frac{1}{\bar{\lambda}}\right)\right)^2}{\frac{d}{d\bar{\lambda}}m\left(-\frac{1}{\bar{\lambda}}\right)}-\frac{1}{\left(1+m\left(-\frac{1}{\bar{\lambda}}\right)\right)^2}\sum_{\bar{k}\neq k}\frac{p_{\bar{k}j}}{N}=\sigma_j^2. \tag{7.71}$$

Let P_j denote the total transmit power on cell j's DL. Thus

$$P_j=\sum_{k=1}^{K}\frac{p_{k,j}}{N}. \tag{7.72}$$

Since each user's power will be asymptotically insignificant, (7.71) becomes

$$\frac{p_{kj}}{\gamma}\frac{\left(\frac{1}{\bar{\lambda}}m\left(-\frac{1}{\bar{\lambda}}\right)\right)^2}{\frac{d}{d\bar{\lambda}}m\left(-\frac{1}{\bar{\lambda}}\right)}-\frac{1}{\left(1+m\left(-\frac{1}{\bar{\lambda}}\right)\right)^2}P_j=\sigma_j^2, \tag{7.73}$$

so that, as $N \uparrow \infty$, $p_{k,j}$ is also asymptotically independent of k. Since

$$\frac{d}{d\bar{\lambda}} m\left(-\frac{1}{\bar{\lambda}}\right) = \frac{m\left(-\frac{1}{\bar{\lambda}}\right)}{\frac{1}{\bar{\lambda}} + \frac{\beta}{\left(1+m\left(-\frac{1}{\bar{\lambda}}\right)\right)^2}} \frac{1}{\bar{\lambda}^2}, \tag{7.74}$$

(7.73) is equivalent to

$$\frac{p_{kj}}{\gamma} m\left(-\frac{1}{\bar{\lambda}}\right) \left[\frac{1}{\bar{\lambda}} + \frac{\beta}{\left(1+m\left(-\frac{1}{\bar{\lambda}}\right)\right)^2}\right] - \frac{P_j}{\left(1+m\left(-\frac{1}{\bar{\lambda}}\right)\right)^2} = \sigma_j^2.$$

Summing over the k and normalizing by N, and further noting that $\gamma = m\left(-\frac{1}{\bar{\lambda}}\right)$, this reduces to

$$\frac{P_j}{\bar{\lambda}} = \beta\sigma_j^2. \tag{7.75}$$

On the other hand, from (7.66), (7.67), and (7.68) we have that

$$\frac{\sigma_j^2}{\sigma^2 + \epsilon P_{\bar{j}}} \to 1, \quad \text{as } N \uparrow \infty, \quad \text{for } j = 1, 2.$$

Therefore, P_1 and P_2 converge to the same value and $P_j \to \frac{\beta\sigma^2\bar{\lambda}}{1-\epsilon\beta\bar{\lambda}}$. Substituting $\bar{\lambda}$ by its value in (7.65),

$$P_j \to \frac{\sigma^2\gamma\beta}{1-\beta\left(\frac{\gamma}{1+\gamma}+\gamma\epsilon\right)} \qquad j = 1, 2. \tag{7.76}$$

Finally, since $p_{k,j}$ is asymptotically independent of k,

$$p_{k,j} \to \frac{\sigma^2\gamma}{1-\beta\left(\frac{\gamma}{1+\gamma}+\gamma\epsilon\right)} \qquad j = 1, 2, \tag{7.77}$$

Denoting by $\bar{p}$ this common value, we get (7.16).

7.E Large System Analysis for Coordinated Beamforming

The dual uplink SINR is equal to

$$\text{SINR}_{k,j} = \lambda_{kj}\mathbf{h}_{k,j,j}\left[\mu_j\mathbf{I}_{N_t} + \sum_{(\bar{j},\bar{k})\neq(j,k)} \lambda_{\bar{k}\bar{j}}\mathbf{h}_{\bar{k},\bar{j},j}^H\mathbf{h}_{\bar{k},\bar{j},j}\right]^{-1} \mathbf{h}_{k,j,j}^H. \tag{7.78}$$

As in the SCP case, defining $\bar{\lambda}_{kj} = N_t \lambda_{kj}$, $k = 1, \ldots, K$, $j = 1, 2$, these satisfy the following fixed point equation

$$\bar{\lambda}_{kj} = \frac{\mathrm{SINR}_{k,j}}{\frac{1}{N_t}\mathbf{h}_{k,j,j}\left[\mu_j \mathbf{I}_{N_t} + \frac{1}{N_t}\sum_{(\bar{j},\bar{k})\neq(j,k)} \bar{\lambda}_{\bar{k}\bar{j}} \mathbf{h}_{\bar{k},\bar{j},j}^H \mathbf{h}_{\bar{k},\bar{j},j}\right]^{-1} \mathbf{h}_{k,j,j}^H} \triangleq I_{kj}^{CBf}\left(\bar{\boldsymbol{\lambda}}, \mu_j\right), \tag{7.79}$$

which, for given μ_j, is an interference function. Following a similar approach to that in Appendix 7.D, letting γ be the common SINR target for all k, j, in the considered asymptotic regime, and starting with $\bar{\boldsymbol{\lambda}}^{(0)} = \mathbf{0}$:

- At the first iteration,

$$\bar{\lambda}_{kj}^{(1)} = I_{kj}^{CBf}\left(\mathbf{0}, \mu_j\right) = \mu_j \frac{\gamma}{\frac{1}{N_t}\|\mathbf{h}_{k,j,j}\|^2}. \tag{7.80}$$

  In the asymptotic regime, this will tend to $\mu_j \frac{\gamma}{\beta}$. Let $\bar{\lambda}_j^{(1)} = \mu_j \frac{\gamma}{\beta}$.

- At the second iteration,

$$\begin{aligned}\bar{\lambda}_{kj}^{(2)} &= I_{kj}^{CBf}\left(\bar{\boldsymbol{\lambda}}^{(1)}, \mu_j\right) \\ &= \frac{\gamma}{\frac{1}{N_t}\mathbf{h}_{k,j,j}\left[\mu_j \mathbf{I}_{N_t} + \frac{1}{N_t}\sum_{(\bar{j},\bar{k})\neq(j,k)} \bar{\lambda}_{\bar{j}}^{(1)} \mathbf{h}_{\bar{k},\bar{j},j}^H \mathbf{h}_{\bar{k},\bar{j},j}\right]^{-1} \mathbf{h}_{k,j,j}^H}\end{aligned}$$

  In the large system regime, we can apply Lemma 1 in [89] to show that the denominator of the right-hand side of the above equation tends to $\frac{1}{N_t}\mathrm{Tr}\left[\mu_j \mathbf{I}_{N_t} + \frac{1}{N_t}\sum_{(\bar{j},\bar{k})\neq(j,k)} \bar{\lambda}_{\bar{j}}^{(1)} \mathbf{h}_{\bar{k},\bar{j},j}^H \mathbf{h}_{\bar{k},\bar{j},j}\right]^{-1}$.

  Asymptotically $\frac{1}{N_t}\sum_{(\bar{j},\bar{k})\neq(j,k)} \bar{\lambda}_{\bar{j}}^{(1)} \mathbf{h}_{\bar{k},\bar{j},j}^H \mathbf{h}_{\bar{k},\bar{j},j}$ is equivalent to a matrix $\frac{1}{N_t}\mathbf{X}\mathbf{T}\mathbf{X}^H$ where $\mathbf{X}$ a $N_t \times 2K$ matrix of i.i.d. complex entries with unit variance, and $\mathbf{T}$ be a $2K \times 2K$ diagonal matrix such that its first K diagonal elements are $\bar{\lambda}_j^{(1)}$ and its last K diagonal elements are equal to $\epsilon\bar{\lambda}_{\bar{j}}^{(1)}$. Then applying Theorem II.1 in [76] for example, (also theorem 4.1 in [75], both taken from [90] originally), its empirical distribution function will almost surely converge to a nonrandom distribution function G, whose Stieltjes transform m_G is the unique (pointwise) solution to

$$m_j^{(1)}(z) = \frac{1}{-z + \frac{\beta\bar{\lambda}_j^{(1)}}{1+\bar{\lambda}_j^{(1)} m_j^{(1)}(z)} + \frac{\beta\epsilon\bar{\lambda}_{\bar{j}}^{(1)}}{1+\epsilon\bar{\lambda}_{\bar{j}}^{(1)} m_j^{(1)}(z)}}. \tag{7.81}$$

Now

$$\lim_{N_t\to\infty}\frac{1}{N_t}\mathrm{Tr}\left[\frac{1}{N_t}\mathbf{XTX}^H+\mu_j\mathbf{I}_{N_t}\right]^{-1}$$
$$=\lim_{N_t\to\infty}\frac{1}{N_t}\sum_{i=1}^{N_t}\frac{1}{l_i+\mu_j}=\mathbb{E}_l\left[\frac{1}{l+\mu_j}\right]=m_j^{(1)}(-\mu_j) \tag{7.82}$$

where $\{l_i\}$ are the eigenvalues of $\frac{1}{N_t}\mathbf{XTX}^H$ and the expectation is taken with respect to the limiting probability density function G. The last equality holds from the definition of the Stieltjes transform. Thus $\bar{\lambda}_{kj}^{(2)}$ will tend to $\frac{\gamma}{m_j^{(1)}(-\mu_j)}$; let $\bar{\lambda}_j^{(2)}$ denote this quantity.

- Similar to the above derivation, we can show that at the i-th iteration, $\bar{\lambda}_{kj}^{(i)}$ will tend to

$$\bar{\lambda}_j^{(i)}=\frac{\gamma}{m_j^{(i-1)}(-\mu_j)}, \tag{7.83}$$

where

$$m_j^{(i-1)}(z)=\frac{1}{-z+\frac{\beta\bar{\lambda}_j^{(i-1)}}{1+\bar{\lambda}_j^{(i-1)}m_j^{(i-1)}(z)}+\frac{\beta\epsilon\bar{\lambda}_{\bar{j}}^{(i-1)}}{1+\epsilon\bar{\lambda}_{\bar{j}}^{(i-1)}m_j^{(i-1)}(z)}}. \tag{7.84}$$

Again, as the above scheme converges, $\bar{\lambda}_{kj}$ will tend to the same value for each k, which we denote $\bar{\lambda}_j$, such that

$$\bar{\lambda}_j=\frac{\gamma}{m_j\left(-\mu_j\right)}, \tag{7.85}$$

$$m_j(z)=\frac{1}{-z+\frac{\beta\bar{\lambda}_j}{1+\bar{\lambda}_j m_j(z)}+\frac{\beta\epsilon\bar{\lambda}_{\bar{j}}}{1+\epsilon\bar{\lambda}_{\bar{j}}m_j(z)}}. \tag{7.86}$$

The dual problem thus simplifies to:

$$\begin{aligned}\max.\quad & \beta\sigma^2\left(\bar{\lambda}_1+\bar{\lambda}_2\right)\\ \text{s.t.}\quad & \gamma\le\bar{\lambda}_j m_j(-\mu_j)\\ & \sum_j(1-\mu_j)=0.\end{aligned} \tag{7.87}$$

The SINR constraints will be met with equality at the optimum. Thus

$$\gamma = \bar{\lambda}_1 U_1 = \bar{\lambda}_2 U_2, \tag{7.88}$$

where $U_j = m_j\left(-\mu_j\right)$ such that

$$U_j = \frac{1}{\mu_j + \frac{\beta\bar{\lambda}_j}{1+\bar{\lambda}_j U_j} + \frac{\beta\epsilon\bar{\lambda}_{\bar{j}}}{1+\epsilon\bar{\lambda}_{\bar{j}} U_j}}. \tag{7.89}$$

Thus

$$\frac{\gamma}{\bar{\lambda}_j} = \frac{1}{\mu_j + \frac{\beta\bar{\lambda}_j}{1+\gamma} + \frac{\beta\epsilon\bar{\lambda}_{\bar{j}}}{1+\epsilon\bar{\lambda}_{\bar{j}}\frac{\gamma}{\bar{\lambda}_j}}}. \tag{7.90}$$

This can be rewritten as

$$\mu_j = \frac{\bar{\lambda}_j}{\gamma} - \frac{\beta\bar{\lambda}_j}{1+\gamma} - \frac{\beta\epsilon\bar{\lambda}_{\bar{j}}}{1+\epsilon\bar{\lambda}_{\bar{j}}\frac{\gamma}{\bar{\lambda}_j}} \geq 0. \tag{7.91}$$

Adding up (7.91) for $j = 1$ and $j = 2$, we get

$$\mu_1 + \mu_2 = -\beta\frac{\bar{\lambda}_1 + \bar{\lambda}_2}{1+\gamma} - \frac{\beta\epsilon\bar{\lambda}_2}{1+\epsilon\bar{\lambda}_2\frac{\gamma}{\bar{\lambda}_1}} - \frac{\beta\epsilon\bar{\lambda}_1}{1+\epsilon\bar{\lambda}_1\frac{\gamma}{\bar{\lambda}_2}} + \frac{\bar{\lambda}_1 + \bar{\lambda}_2}{\gamma} \leq 2. \tag{7.92}$$

Holding $t = \bar{\lambda}_1 + \bar{\lambda}_2$ fixed, the above constraint is made less likely to be broken by maximizing $\frac{\bar{\lambda}_2}{1+\epsilon\bar{\lambda}_2\frac{\gamma}{\bar{\lambda}_1}} + \frac{\bar{\lambda}_1}{1+\epsilon\bar{\lambda}_1\frac{\gamma}{\bar{\lambda}_2}}$. This can be rewritten as

$$\frac{(t-\bar{\lambda}_1)\bar{\lambda}_1}{\bar{\lambda}_1 + \epsilon(t-\bar{\lambda}_1)\gamma} + \frac{\bar{\lambda}_1(t-\bar{\lambda}_1)}{(t-\bar{\lambda}_1) + \epsilon\bar{\lambda}_1\gamma}.$$

Taking the first order derivative with respect to $\bar{\lambda}_1$:

$$\begin{aligned}&\frac{\epsilon(t-\bar{\lambda}_1)^2\gamma - \bar{\lambda}_1^2}{\left(\bar{\lambda}_1 + \epsilon(t-\bar{\lambda}_1)\gamma\right)^2} + \frac{(t-\bar{\lambda}_1)^2 - \bar{\lambda}_1^2\epsilon\gamma}{\left((t-\bar{\lambda}_1) + \epsilon\bar{\lambda}_1\gamma\right)^2} \\ &= t^3\epsilon\gamma(1+\epsilon\gamma)\frac{t-2\bar{\lambda}_1}{\left((t-\bar{\lambda}_1) + \epsilon\bar{\lambda}_1\gamma\right)^2\left(\bar{\lambda}_1 + \epsilon(t-\bar{\lambda}_1)\gamma\right)^2}.\end{aligned}$$

The optimum $\bar{\lambda}_1$ is thus $t/2$ and consequently $\bar{\lambda}_1 = \bar{\lambda}_2 = \bar{\lambda}$. For the dual problem to be bounded and the primal problem feasible, μ_j should be positive, consequently, from (7.91),

$$\frac{1}{\gamma} - \frac{\beta}{1+\gamma} - \frac{\beta\epsilon}{1+\epsilon\gamma} > 0. \tag{7.93}$$

In this case, $\bar{\lambda}$ must satisfy

$$\bar{\lambda} = \frac{1}{\frac{1}{\gamma} - \frac{\beta}{1+\gamma} - \frac{\beta\epsilon}{1+\epsilon\gamma}}. \tag{7.94}$$

The optimum μ_k will also be such that $\mu_1 = \mu_2 = \mu$ such that $\mu = 1$. Finally $U_1 = U_2 = U = \frac{\gamma}{\bar{\lambda}}$, so that $\gamma = \frac{1}{\frac{1}{\bar{\lambda}} + \frac{\beta}{1+\gamma} + \frac{\beta\epsilon}{1+\epsilon\gamma}}$.

Now consider, for $j = 1, 2$, $k = 1, \ldots, K$, equation (7.47), and let $m(z)$ denote (7.86) specialized to $\bar{\lambda}_1 = \bar{\lambda}_2 = \bar{\lambda}$, so that

$$m(z) = \frac{1}{-z + \frac{\beta\bar{\lambda}}{1+\bar{\lambda}m(z)} + \frac{\beta\epsilon\bar{\lambda}}{1+\epsilon\bar{\lambda}m(z)}}. \tag{7.95}$$

Since $\mu_j \to \mu = 1$ $j = 1, 2$, and $\bar{\lambda}_j \to \bar{\lambda}$ $j = 1, 2$, and with $m(z)$ defined in (7.95), we have:

$$\frac{1}{N^2} \left|\mathbf{h}_{k,j,j}\hat{\mathbf{w}}_{kj}\right|^2 \to \left[m\left(-\mu\right)\right]^2, \tag{7.96}$$

$$\frac{1}{N}\|\hat{\mathbf{w}}_{kj}\|^2 \to \frac{d}{d\mu} m\left(-\mu\right), \tag{7.97}$$

$$\frac{1}{N}\left|\mathbf{h}_{k,j,j}\hat{\mathbf{w}}_{\bar{k}j}\right|^2 \to \frac{\frac{d}{d\mu} m\left(-\mu\right)}{\left(1 + \bar{\lambda}m(-\mu)\right)^2}, \tag{7.98}$$

$$\frac{1}{N}\left|\mathbf{h}_{k,j,\bar{j}}\hat{\mathbf{w}}_{\bar{k}\bar{j}}\right|^2 \to \frac{\epsilon\frac{d}{d\mu} m\left(-\mu\right)}{\left(1 + \epsilon\bar{\lambda}m(-\mu)\right)^2}. \tag{7.99}$$

In the large N limit, (7.47) thus becomes

$$\frac{p_{kj}}{\gamma}\frac{\left(m\left(-\mu\right)\right)^2}{\frac{d}{d\mu}m\left(-\mu\right)} - \frac{\sum_{\bar{k}\neq k}\frac{p_{\bar{k}j}}{N}}{\left(1 + \bar{\lambda}m(-\mu)\right)^2} - \frac{\epsilon\sum_{\bar{k}}\frac{p_{\bar{k}\bar{j}}}{N}}{\left(1 + \epsilon\bar{\lambda}m(-\mu)\right)^2} = \sigma^2.$$

Since a given user's power is asymptotically negligible and defining P_j as in (7.72), this equation becomes

$$\frac{p_{kj}}{\gamma}\frac{\left(m\left(-\mu\right)\right)^2}{\frac{d}{d\mu}m\left(-\mu\right)} - \frac{P_j}{\left(1 + \bar{\lambda}m(-\mu)\right)^2} - \frac{\epsilon P_{\bar{j}}}{\left(1 + \epsilon\bar{\lambda}m(-\mu)\right)^2} = \sigma^2,$$

so that, as in the SCP case, $p_{k,j}$ will also be asymptotically independent of k. Summing over k and normalizing by N, we obtain the limiting equation for P_j:

$$\frac{P_j}{\gamma}\frac{\left(m\left(-\mu\right)\right)^2}{\frac{d}{d\mu}m\left(-\mu\right)} - \frac{\beta P_j}{\left(1 + \bar{\lambda}m(-\mu)\right)^2} - \frac{\epsilon\beta P_{\bar{j}}}{\left(1 + \epsilon\bar{\lambda}m(-\mu)\right)^2} = \beta\sigma^2.$$

Hence

$$\begin{aligned} P_j \rightarrow & \frac{\beta\sigma^2}{\frac{1}{\gamma}\frac{(m(-\mu))^2}{\frac{d}{d\mu}m(-\mu)} - \frac{\beta}{\left(1+\bar{\lambda}m(-\mu)\right)^2} - \frac{\epsilon\beta}{\left(1+\epsilon\bar{\lambda}m(-\mu)\right)^2}} \\ & \overset{(a)}{=} \frac{\sigma^2\beta}{\mu\frac{1}{\gamma}m\left(-\mu\right) + \beta\frac{\bar{\lambda}\frac{1}{\gamma}m(-\mu)-1}{(1+\bar{\lambda}m(-\mu))^2} + \epsilon\beta\frac{\bar{\lambda}\frac{1}{\gamma}m(-\mu)-1}{(1+\epsilon\bar{\lambda}m(-\mu))^2}} \\ & \overset{(b)}{=} \beta\sigma^2\bar{\lambda} \overset{(c)}{=} \frac{\beta\sigma^2\gamma}{1-\beta\left(\frac{\gamma}{1+\gamma}+\frac{\epsilon\gamma}{1+\epsilon\gamma}\right)}, \end{aligned} \tag{7.100}$$

where (a) follows by plugging in the value of the derivative $\frac{d}{d\mu}m\left(-\mu\right)$, (b) from using the identity $\bar{\lambda}m(-\mu)=\gamma$, $\mu=1$ and (c) by additionally using (7.95). Similarly,

$$p_{kj} \rightarrow \bar{p} = \sigma^2\bar{\lambda} = \frac{\sigma^2\gamma}{1-\beta\left(\frac{\gamma}{1+\gamma}+\frac{\epsilon\gamma}{1+\epsilon\gamma}\right)}. \tag{7.101}$$

Equation (7.100) shows that asymptotically, there is a solution to the primal problem if and only if $\beta\left(\frac{\gamma}{1+\gamma}+\frac{\epsilon\gamma}{1+\epsilon\gamma}\right)<1$.

7.F Large System Analysis for MCP

At the optimum, the inequality in (7.50) will be met with equality for all users. Thus

$$\begin{aligned} \bar{\lambda}_{kj} &= \frac{\gamma}{\frac{1}{N_t}\tilde{\mathbf{h}}_{k,j}\left[\sum_{\bar{j},\bar{k},(\bar{j},\bar{k})\neq(j,k)}\frac{1}{N_t}\bar{\lambda}_{\bar{k}\bar{j}}\tilde{\mathbf{h}}_{\bar{k},\bar{j}}^H\tilde{\mathbf{h}}_{\bar{k},\bar{j}} + \sum_{\bar{j}=1}^{2}\mu_{\bar{j}}\mathbf{E}_{\bar{j}}\right]^{-1}\tilde{\mathbf{h}}_{k,j}^H} \\ &\triangleq I_{kj}^{MCP}(\boldsymbol{\lambda}), \end{aligned} \tag{7.102}$$

which can be shown to be an interference function for given μ_1,μ_2. The following analysis holds for any μ_1,μ_2, including the optimal ones.
At iteration i for the standard power control algorithm starting at the all zero vector, $\bar{\lambda}_{kj}^{(i)}$ is equal to

$$\frac{\gamma}{\frac{1}{N_t}\tilde{\mathbf{h}}_{k,j}\left[\frac{1}{N_t}\sum_{\bar{j},\bar{k},(\bar{j},\bar{k})\neq(j,k)}\bar{\lambda}_{\bar{j}}^{(i-1)}\tilde{\mathbf{h}}_{\bar{k},\bar{j}}^H\tilde{\mathbf{h}}_{\bar{k},\bar{j}} + \sum_{\bar{j}=1}^{2}\mu_{\bar{j}}\mathbf{E}_{\bar{j}}\mathbf{E}_{\bar{j}}\right]^{-1}\tilde{\mathbf{h}}_{k,j}^H}. \tag{7.103}$$

The denominator above can be rewritten as

$$\frac{1}{N_t}\tilde{\mathbf{h}}_{k,j}\left[\frac{1}{N_t}\tilde{\mathbf{H}}_{k,j}^H\bar{\mathbf{L}}_{k,j}^{(i-1)}\tilde{\mathbf{H}}_{k,j}+\mathbf{M}\right]^{-1}\tilde{\mathbf{h}}_{k,j}^H$$
$$=\frac{1}{N_t}\tilde{\mathbf{h}}_{k,j}\mathbf{M}^{-1/2}\left[\frac{1}{N_t}\mathbf{M}^{-1/2}\tilde{\mathbf{H}}_{k,j}^H\bar{\mathbf{L}}_{k,j}^{(i-1)}\tilde{\mathbf{H}}_{k,j}\mathbf{M}^{-1/2}+\mathbf{I}_{2N_t}\right]^{-1}\mathbf{M}^{-1/2}\tilde{\mathbf{h}}_{k,j}^H$$

where $\mathbf{M}=\sum_{\bar{j}=1}^{\bar{2}}\mu_{\bar{j}}\mathbf{E}_{\bar{j}}$ is a diagonal matrix, and $\tilde{\mathbf{H}}_{k,j}$ consists of the concatenation of all user channels except user k in cell j, whereas $\bar{\mathbf{L}}_{k,j}^{(i-1)}$ is the diagonal matrix with the first K elements along the diagonal equal to $\bar{\lambda}_{\bar{j}}^{(i-1)}$ and the second K elements equal to $\bar{\lambda}_{\bar{j}}^{(i-1)}$.
Since the entries in $\tilde{\mathbf{h}}_{k,j}$ have different variances, we resort to Lemma 14 in [91] to show that the expectation over $\tilde{\mathbf{h}}_{k,j}$ of the above expression will tend to

$$\frac{1}{N_t}\sum_{l=1}^{2}|d_{k,l}|^2\sum_{i=(l-1)N_t+1}^{lN_t}\left(\left[\frac{1}{N_t}\mathbf{M}^{-1/2}\tilde{\mathbf{H}}_{k,j}^H\bar{\mathbf{L}}_{k,j}^{(i-1)}\tilde{\mathbf{H}}_{k,j}\mathbf{M}^{-1/2}+\mathbf{I}_{2N_t}\right]^{-1}\right)_{ii}$$

and its variance is upper bounded by

$$\frac{C_1}{N_t}\lambda_{\max}^2\left(\left[\frac{1}{N_t}\mathbf{M}^{-1/2}\tilde{\mathbf{H}}_{k,j}^H\bar{\mathbf{L}}_{k,j}^{(i-1)}\tilde{\mathbf{H}}_{k,j}\mathbf{M}^{-1/2}+\mathbf{I}_{2N_t}\right]^{-1}\right).$$

where $|d_{k,1}|^2=\frac{1}{\mu_1}$, resp. $\frac{\epsilon}{\mu_1}$ and $|d_{k,2}|^2=\frac{\epsilon}{\mu_2}$, resp. $\frac{1}{\mu_2}$ if the user is in cell 1, resp. cell 2, and where C_1 is a constant that depends only on the fourth moment of the entries in our vector and $\lambda_{\max}(\mathcal{B})$ is $\mathcal{B}$'s largest eigenvalue.

We now apply Theorem 13 in [91]. $\frac{1}{N_t}\mathbf{M}^{-1/2}\tilde{\mathbf{H}}_{k,j}^H\bar{\mathbf{L}}_{k,j}^{(i-1)}\tilde{\mathbf{H}}_{k,j}\mathbf{M}^{-1/2}$ has the limiting bounded function $v^{(i-1)}$ given by:

$$v^{(i-1)}(x,y)=\begin{cases}\frac{\bar{\lambda}_1^{(i-1)}}{\mu_1} & 0\le x<1, 0\le y<\beta\\ \frac{\bar{\lambda}_1^{(i-1)}\epsilon}{\mu_2} & 1\le x<2, 0\le y<\beta\\ \frac{\bar{\lambda}_2^{(i-1)}\epsilon}{\mu_1} & 0\le x<1, \beta\le y<2\beta\\ \frac{\bar{\lambda}_2^{(i-1)}}{\mu_2} & 1\le x<2, \beta\le y<2\beta\end{cases}\tag{7.104}$$

Then for each $a,b\in[0,2]$, $a<b$ and $z\in\mathbb{C}^+$

$$\frac{1}{N_t}\sum_{i=\lfloor aN_t\rfloor}^{i=\lfloor bN_t\rfloor}\left[\frac{1}{N_t}\mathbf{M}^{-1/2}\tilde{\mathbf{H}}_{k,j}^H\bar{\mathbf{L}}_{k,j}^{(i-1)}\tilde{\mathbf{H}}_{k,j}\mathbf{M}^{-1/2}-z\mathbf{I}_{2N_t}\right]_{ii}^{-1}\xrightarrow{\mathcal{P}}\int_a^b u^{(i-1)}(x,z)dx$$

where $u^{(i-1)}(x,z)$ satisfies the equation

$$u^{(i-1)}(x,z) = \frac{1}{-z + \int_0^{2\beta} \frac{v^{(i-1)}(x,y)dy}{1+\int_0^2 u^{(i-1)}(w,z)v^{(i-1)}(w,y)dw}}$$

Moreover, a.s., the empirical eigenvalue distribution of $\frac{1}{N_t}\mathbf{M}^{-1/2}\tilde{\mathbf{H}}_{k,j}^H\bar{\mathbf{L}}_{k,j}^{(i-1)}\tilde{\mathbf{H}}_{k,j}\mathbf{M}^{-1/2}$ converges weakly to a limiting distribution G^* whose Stieltjes transform

$$m^{(i-1)}(z) = \int_0^\infty \frac{1}{\lambda - z} dG^*(\lambda) \tag{7.105}$$

is given by $\int_0^1 u^{(i-1)}(x,z)dx$.

Finally, [2]

$$\frac{1}{N_t}\sum_{i=\lfloor aN_t\rfloor}^{i=\lfloor bN_t\rfloor}\left[\frac{1}{N_t}\mathbf{M}^{-1/2}\tilde{\mathbf{H}}_{k,j}^H\bar{\mathbf{L}}_{k,j}^{(i-1)}\tilde{\mathbf{H}}_{k,j}\mathbf{M}^{-1/2} - z\mathbf{I}_{2N_t}\right]_{ii}^{-2} \xrightarrow{\mathcal{P}} \frac{d}{dz}\int_a^b u^{(i-1)}(x,z)dx.$$

$$\begin{aligned}
&u^{(i-1)}(x,z)\\
&= \frac{1}{-z + \int_0^{2\beta} \frac{v^{(i-1)}(x,y)dy}{1+\int_0^2 u^{(i-1)}(w,z)v^{(i-1)}(w,y)dw}}\\
&= \begin{cases}
\dfrac{1}{-z+\frac{\frac{\bar\lambda_1^{(i-1)}{}^{\beta}}{\mu_1}}{1+\frac{\bar\lambda_1^{(i-1)}}{\mu_1}U_1^{(i-1)}(z)+\frac{\bar\lambda_1^{(i-1)}{}^{\epsilon}}{\mu_2}U_2^{(i-1)}(z)}+\frac{\beta\frac{\bar\lambda_2^{(i-1)}{}^{\epsilon}}{\mu_1}}{1+\frac{\bar\lambda_2^{(i-1)}{}^{\epsilon}}{\mu_1}U_1^{(i-1)}(z)+\frac{\bar\lambda_2^{(i-1)}}{\mu_2}U_2^{(i-1)}(z)}} & 0 \le x < 1\\
\dfrac{1}{-z+\frac{\beta\frac{\bar\lambda_1^{(i-1)}{}^{\epsilon}}{\mu_2}}{1+\frac{\bar\lambda_1^{(i-1)}}{\mu_1}U_1^{(i-1)}(z)+\frac{\bar\lambda_1^{(i-1)}}{\mu_2}\epsilon U_2^{(i-1)}(z)}+\frac{\beta\frac{\bar\lambda_2^{(i-1)}}{\mu_2}}{1+\frac{\bar\lambda_2^{(i-1)}{}^{\epsilon}}{\mu_1}U_1^{(i-1)}(z)+\frac{\bar\lambda_2^{(i-1)}}{\mu_2}U_2^{(i-1)}(z)}} & 1 \le x < 2
\end{cases}
\end{aligned}$$

where

$$U_1^{(i-1)}(z) = \frac{1}{-z + \frac{\frac{\bar\lambda_1^{(i-1)}{}^{\beta}}{\mu_1}}{1+\frac{\bar\lambda_1^{(i-1)}}{\mu_1}U_1^{(i-1)}(z)+\frac{\bar\lambda_1^{(i-1)}{}^{\epsilon}}{\mu_2}U_2^{(i-1)}(z)} + \frac{\beta\frac{\bar\lambda_2^{(i-1)}{}^{\epsilon}}{\mu_1}}{1+\frac{\bar\lambda_2^{(i-1)}{}^{\epsilon}}{\mu_1}U_1^{(i-1)}(z)+\frac{\bar\lambda_2^{(i-1)}}{\mu_2}U_2^{(i-1)}(z)}}.$$

$$U_2^{(i-1)}(z) = \frac{1}{-z + \frac{\beta\frac{\bar\lambda_1^{(i-1)}{}^{\epsilon}}{\mu_2}}{1+\frac{\bar\lambda_1^{(i-1)}}{\mu_1}U_1^{(i-1)}(z)+\frac{\bar\lambda_1^{(i-1)}{}^{\epsilon}}{\mu_2}U_2^{(i-1)}(z)} + \frac{\beta\frac{\bar\lambda_2^{(i-1)}}{\mu_2}}{1+\frac{\bar\lambda_2^{(i-1)}{}^{\epsilon}}{\mu_1}U_1^{(i-1)}(z)+\frac{\bar\lambda_2^{(i-1)}}{\mu_2}U_2^{(i-1)}(z)}}.$$

[2] This holds since $\left[\frac{1}{N_t}\mathbf{H}^H\mathbf{H} - z\mathbf{I}\right]_{ii}^{-2} = \frac{d}{dz}\left[\frac{1}{N_t}\mathbf{H}^H\mathbf{H} - z\mathbf{I}\right]_{ii}^{-1}$

As a result, $\bar{\lambda}_{k1}^{(i)}, k = 1, \ldots, K$ becomes $\frac{\gamma}{\frac{1}{\mu_1}U_1^{(i-1)}(-1)+\frac{\epsilon}{\mu_2}U_2^{(i-1)}(-1)}$, whereas $\bar{\lambda}_{k2}^{(i)}, k = 1, \ldots, K$ becomes $\frac{\gamma}{\frac{\epsilon}{\mu_1}U_1^{(i-1)}(-1)+\frac{1}{\mu_2}U_2^{(i-1)}(-1)}$.

It follows that, asymptotically, users in cell $j, j = 1, 2$ will all tend to have the same scaled dual uplink transmit power, denote this $\bar{\lambda}_j$. From the above analysis and Equation (7.102), $\bar{\lambda}_1$ and $\bar{\lambda}_2$ are the unique solutions to the fixed point equation:

$$\begin{aligned} \bar{\lambda}_1 &= \frac{\gamma}{\frac{1}{\mu_1}U_1 + \frac{\epsilon}{\mu_2}U_2}, \\ \bar{\lambda}_2 &= \frac{\gamma}{\frac{\epsilon}{\mu_1}U_1 + \frac{1}{\mu_2}U_2}, \end{aligned} \tag{7.106}$$

with $U_1 = U_1(-1)$ and $U_2 = U_2(-1)$ such that

$$\begin{aligned} U_1(z) &= \frac{1}{-z + \frac{\frac{\bar{\lambda}_1\beta}{\mu_1}}{1+\frac{\bar{\lambda}_1}{\mu_1}U_1(z)+\frac{\bar{\lambda}_1\epsilon}{\mu_2}U_2(z)} + \frac{\beta\frac{\bar{\lambda}_2\epsilon}{\mu_1}}{1+\frac{\bar{\lambda}_2^{(i-1)}\epsilon}{\mu_1}U_1(z)+\frac{\bar{\lambda}_2}{\mu_2}U_2(z)}}. \\ U_2(z) &= \frac{1}{-z + \frac{\beta\frac{\bar{\lambda}_1\epsilon}{\mu_2}}{1+\frac{\bar{\lambda}_1}{\mu_1}U_1^{(i-1)}(z)+\frac{\bar{\lambda}_1\epsilon}{\mu_2}U_2(z)} + \frac{\beta\frac{\bar{\lambda}_2}{\mu_2}}{1+\frac{\bar{\lambda}_2\epsilon}{\mu_1}U_1(z)+\frac{\bar{\lambda}_2}{\mu_2}U_2(z)}}. \end{aligned}$$

The dual problem now becomes

$$\begin{aligned} \text{maximize } & \sum_{k=1}^{2}\sigma^2\bar{\lambda}_k \\ \text{s.t. } & \sum_{k=1}^{2}(1-\mu_k) = 0 \\ & \gamma = \bar{\lambda}_1\left[\frac{1}{\mu_1}U_1 + \epsilon\frac{1}{\mu_2}U_2\right], \qquad \gamma = \bar{\lambda}_2\left[\epsilon\frac{1}{\mu_1}U_1 + \frac{1}{\mu_2}U_2\right]. \end{aligned}$$

The last two equalities can be rewritten as

$$U_1 = \gamma\frac{\mu_1}{1-\epsilon^2}\left(\frac{1}{\bar{\lambda}_1} - \frac{\epsilon}{\bar{\lambda}_2}\right) \geq 0, \quad U_2 = \gamma\frac{\mu_2}{1-\epsilon^2}\left(\frac{1}{\bar{\lambda}_2} - \frac{\epsilon}{\bar{\lambda}_1}\right) \geq 0.$$

Plugging these into the Stieltjes transform fixed-point equations yields

$$\frac{\gamma}{1-\epsilon^2}\left(\frac{1}{\bar{\lambda}_1} - \frac{\epsilon}{\bar{\lambda}_2}\right)\left[\mu_1 + \frac{\beta}{1+\gamma}\left[\bar{\lambda}_1 + \bar{\lambda}_2\epsilon\right]\right] = 1 \tag{7.107}$$

$$\frac{\gamma}{1-\epsilon^2}\left(\frac{1}{\bar{\lambda}_2} - \frac{\epsilon}{\bar{\lambda}_1}\right)\left[\mu_2 + \frac{\beta}{1+\gamma}\left[\bar{\lambda}_1\epsilon + \bar{\lambda}_2\right]\right] = 1. \tag{7.108}$$

This can be rewritten as

$$\mu_1 = \frac{1}{\gamma}\bar{\lambda}_1\bar{\lambda}_2\frac{1-\epsilon^2}{\bar{\lambda}_2-\epsilon\bar{\lambda}_1} - \frac{\beta}{1+\gamma}\left[\bar{\lambda}_1+\bar{\lambda}_2\epsilon\right] \geq 0$$
$$\mu_2 = \frac{1}{\gamma}\bar{\lambda}_1\bar{\lambda}_2\frac{1-\epsilon^2}{\bar{\lambda}_1-\epsilon\bar{\lambda}_2} - \frac{\beta}{1+\gamma}\left[\bar{\lambda}_1\epsilon+\bar{\lambda}_2\right] \geq 0.$$

Summing the above two equations, we get

$$\mu_1+\mu_2 = \frac{1}{\gamma}(1-\epsilon^2)\bar{\lambda}_1\bar{\lambda}_2\left[\frac{1}{\left(\bar{\lambda}_2-\epsilon\bar{\lambda}_1\right)}+\frac{1}{\left(\bar{\lambda}_1-\epsilon\bar{\lambda}_2\right)}\right] - \frac{\beta(1+\epsilon)}{1+\gamma}\left[\bar{\lambda}_1+\bar{\lambda}_2\right]$$
$$= (1+\epsilon)\left(\bar{\lambda}_2+\bar{\lambda}_1\right)\left[\frac{1}{\gamma}\frac{\bar{\lambda}_1\bar{\lambda}_2\left(1-\epsilon\right)^2}{\left(\bar{\lambda}_2-\epsilon\bar{\lambda}_1\right)\left(\bar{\lambda}_1-\epsilon\bar{\lambda}_2\right)} - \frac{\beta}{1+\gamma}\right] \leq 2$$

Fixing $t = \bar{\lambda}_1+\bar{\lambda}_2$, the constraint is more likely to be satisfied by minimizing $\left[\frac{1}{\gamma}\frac{\bar{\lambda}_1\bar{\lambda}_2(1-\epsilon)^2}{\left(\bar{\lambda}_2-\epsilon\bar{\lambda}_1\right)\left(\bar{\lambda}_1-\epsilon\bar{\lambda}_2\right)} - \frac{\beta}{1+\gamma}\right]$, in other words by minimizing

$$\frac{\bar{\lambda}_1(t-\bar{\lambda}_1)}{\left(t-(1+\epsilon)\bar{\lambda}_1\right)\left((1+\epsilon)\bar{\lambda}_1-\epsilon t\right)}. \tag{7.109}$$

Taking the derivative with respect to $\bar{\lambda}_1$, we get $\frac{-\epsilon t^2(t-2\bar{\lambda}_1)}{\left(t-(1+\epsilon)\bar{\lambda}_1\right)^2\left((1+\epsilon)\bar{\lambda}_1-\epsilon t\right)^2}$. Thus $\bar{\lambda}_1 = \bar{\lambda}_2 = \bar{\lambda}$ and the optimum will occur for $\bar{\lambda}$ satisfying

$$(1+\epsilon)\bar{\lambda}\left[\frac{1}{\gamma}-\frac{\beta}{1+\gamma}\right] = 1. \tag{7.110}$$

Note that if $\frac{1}{\gamma}-\frac{\beta}{1+\gamma}$ is negative, the dual problem will be unbounded and the primal thus infeasible. Whereas for a feasible dual, the optimal $\mu_1 = \mu_2 = \mu = 1$.

Now turning to the DL power levels, recall the DL SINR constraint (7.53). Similar arguments to those used for the SCP and CBf cases show that the $p_{k,j}$s in each cell are asymptotically independent of k. Moreover, since the optimal μs are the same, all the $p_{k,j}$s converge to the same value. Defining $\bar{p}$ as this common asymptotic power level, and dropping indices wherever variables are equal in both cells, (7.53) becomes

$$\frac{\bar{p}}{\gamma}\frac{\frac{(1+\epsilon)}{\mu}U^2}{\frac{dU(z)}{dz}|_{z=-1}} - \frac{\bar{p}\beta\frac{(1+\epsilon)}{\mu}}{\left(1+\frac{\bar{\lambda}}{\mu}(1+\epsilon)U\right)^2} = \sigma^2, \tag{7.111}$$

which is equivalent to

$$\frac{\bar{p}}{\gamma}\frac{(1+\epsilon)U}{\mu}\left[1+\frac{\frac{\bar{\lambda}}{\mu}(1+\epsilon)\beta}{\left(1+\frac{\bar{\lambda}}{\mu}(1+\epsilon)U\right)^2}\right]-\frac{\bar{p}\beta\frac{(1+\epsilon)}{\mu}}{\left(1+\frac{\bar{\lambda}}{\mu}(1+\epsilon)U\right)^2}$$
$$=\sigma^2,$$

where we also used the short-hand U for $U(-1)$. Using the fact that $U(-1)=\frac{\mu\gamma}{\lambda(1+\epsilon)}$ and $\mu=1$, this reduces to $\frac{\bar{p}}{\lambda}=\sigma^2$.

7.G Proof of Theorem 10

Using (7.18), r simplifies to:

$$\frac{1+\gamma^*}{\gamma^*\left(\frac{\sigma^2}{P}+\epsilon\right)(1+\gamma^*)+\gamma^*}\log\left(1+\gamma^*\right) \tag{7.112}$$

Defining $\eta=\frac{\sigma^2}{P}+\epsilon$ and taking the derivative with respect to γ^*:

$$-\frac{1+\eta\left(1+\gamma^*\right)^2}{\gamma^*\eta\left(1+\gamma^*\right)+\gamma^*)^2}\log\left(1+\gamma^*\right)+\frac{1}{\gamma^*\eta\left(1+\gamma^*\right)+\gamma^*} \tag{7.113}$$

It has the same sign as:

$$-\log\left(1+\gamma^*\right)+\frac{\gamma^*\eta\left(1+\gamma^*\right)+\gamma^*}{1+\eta\left(1+\gamma^*\right)^2} \tag{7.114}$$

It is 0 at 0, and $-\infty$ at $+\infty$.
Differentiating the latter:

$$-\frac{1}{1+\gamma^*}+\frac{\eta+1+\eta\left(\eta+1\right)\left(1+\gamma^*\right)^2-2\eta\gamma^{*2}}{\left(1+\eta\left(1+\gamma^*\right)^2\right)^2} \tag{7.115}$$

This will have the same sign as:

$$-\left(1+\eta\left(1+\gamma^*\right)^2\right)^2+\left(1+\gamma^*\right)\left(\eta+1+\left(\eta+1\right)\eta\left(1+\gamma^*\right)^2-2\eta\gamma^{*2}\right)$$
$$=-\gamma^{*4}\eta^2-\gamma^{*3}\eta\left[1+3\eta\right]-\gamma^{*2}\eta\left[1+3\eta\right]-\gamma^*\left[\eta^2-1\right] \tag{7.116}$$

The constant term above is non-negative if $\eta-1\geq 0$. If this is the case, then the derivative will be non-positive. And the function in (7.114) is decreasing. Otherwise, the function in (7.114) is increasing then decreasing. This knowledge combined with that of its values at 0 and ∞ implies the rate will be increasing then decreasing, thus proving the theorem.

7.H Proof of Theorem 11

The normalized achievable rate per cell is equal to

$$\begin{aligned} r = \beta \log\left(1+\gamma^*\right) &= \frac{1}{\gamma^*} \frac{1}{\left[\frac{\sigma^2}{P} + \frac{1}{1+\gamma^*} + \frac{\epsilon}{1+\epsilon\gamma^*}\right]} \log\left(1+\gamma^*\right) \\ &= \frac{\left(1+\gamma^*\right)\left(1+\epsilon\gamma^*\right)}{\left[\frac{\sigma^2}{P}\gamma^*\left(1+\gamma^*\right)\left(1+\epsilon\gamma^*\right) + \gamma^*\left(1+\epsilon\gamma^*\right) + \epsilon\gamma^*\left(1+\gamma^*\right)\right]} \log\left(1+\gamma^*\right) \end{aligned} \tag{7.117}$$

Differentiating with respect to γ^*, the sign of the derivative is that of:

$$\begin{aligned} &\left(1+\epsilon\gamma^*\right)\left[\frac{\sigma^2}{P}\gamma^*\left(1+\gamma^*\right)\left(1+\epsilon\gamma^*\right) + \gamma^*\left(1+\epsilon\gamma^*\right) + \epsilon\gamma^*\left(1+\gamma^*\right)\right] \\ &- \left[\left(\epsilon\gamma^{*2}+1\right)\left[1+\epsilon\right] + \frac{\sigma^2}{P}\left(1+\gamma^*\right)^2\left(1+\epsilon\gamma^*\right)^2 + 4\epsilon\gamma^*\right]\log\left(1+\gamma^*\right) \end{aligned}$$

At zero, it is zero. At ∞, it is negative.

Equivalently, we can study:

$$\frac{\left(1+\epsilon\gamma^*\right)\left[\frac{\sigma^2}{P}\gamma^*\left(1+\gamma^*\right)\left(1+\epsilon\gamma^*\right) + \gamma^*\left(1+\epsilon\gamma^*\right) + \epsilon\gamma^*\left(1+\gamma^*\right)\right]}{\left[\left(\epsilon\gamma^{*2}+1\right)\left[1+\epsilon\right] + \frac{\sigma^2}{P}\left(1+\gamma^*\right)^2\left(1+\epsilon\gamma^*\right)^2 + 4\epsilon\gamma^*\right]} - \log\left(1+\gamma^*\right) \tag{7.118}$$

The sign of the derivative of (7.118) can be shown to be the opposite of that of:

$$\begin{aligned} &\gamma^{*7}a^2\epsilon^4 + \gamma^{*6}\epsilon^3\left(3a^2\epsilon + 4a^2 + 2a\epsilon\right) + \gamma^{*5}\epsilon^2\left(3a^2\epsilon^2 + 12a^2\epsilon + 6a^2 + 4a\epsilon^2 + 6a\epsilon\right) \\ &+ \gamma^{*4}\epsilon\left(a^2\epsilon^3 + 12a^2\epsilon^2 + 18a^2\epsilon + 4a^2 + 3a\epsilon^3 + 8a\epsilon^2 + 9a\epsilon - 2\epsilon\left(\epsilon+1\right)\right) \\ &+ \gamma^{*3}\left(4a^2\epsilon^3 + 18a^2\epsilon^2 + 12a^2\epsilon + a^2 + a\epsilon^4 + 13a\epsilon^2 + 6a\epsilon + \epsilon^2 - \epsilon^4 - 16\epsilon^3\right) \\ &+ \gamma^{*2}\left(6a^2\epsilon^2 + 12a^2\epsilon + 3a^2 + 3a\epsilon^2 + 10a\epsilon + a + \epsilon - \left(17\epsilon^3 + 8\epsilon^2 + 4a\epsilon^3\right)\right) \\ &+ \gamma^*\left(4a^2\epsilon + 3a^2 + 6a\epsilon + a - \left(2a\epsilon^3 + 3a\epsilon^2 + 9\epsilon^3 + 4\epsilon^2 + 3\epsilon\right)\right) \\ &+ a^2 + 2a\epsilon - 2a\epsilon^2 - 2\epsilon^3 - \epsilon^2 - 1, \end{aligned} \tag{7.119}$$

where $a = \frac{\sigma^2}{P}$.

One can factor the constant term into:

$$(a+\epsilon+1)(a+\epsilon-2\epsilon^2-1) \tag{7.120}$$

Its sign is thus determined by the second factor.

7.H.1 $a + \epsilon - 2\epsilon^2 - 1 \geq 0$

Defining $\epsilon_{0,1} = \frac{1-\sqrt{8a-7}}{4}, \epsilon_{0,2} = \frac{1+\sqrt{8a-7}}{4}$, one can show that this term will be positive if $a \geq 1$, $0 \leq \epsilon \leq \epsilon_{0,2}$, or $\frac{7}{8} \leq a \leq 1$, $\epsilon_{0,1} \leq \epsilon \leq \epsilon_{0,2}$. One can further establish that if this is the case, all of the other coefficients of the polynomial (7.119) will be positive [3], and consequently (7.118) will be decreasing. As it is 0 at the origin, it will always be negative, and consequently in this case, the optimum will lie at $\gamma^* \to 0$: this corresponds to $\beta^* \to \infty$.

7.H.2 $a + \epsilon - 2\epsilon^2 - 1 < 0$

In this case, a strictly positive zero of (7.119) is guaranteed to exist. Assuming it is unique (which we think it always will be), (7.118) will be increasing then decreasing. We know that at 0 it is zero, and at ∞, it is $-\infty$, Thus, (7.117) will have a unique maximum for finite $\gamma^* > 0$.

[3] This is done, for example, by using the fact that for $a \geq \frac{7}{8}$, $0 \leq \epsilon \leq \epsilon_{0,2}$, $\epsilon \leq \sqrt{a}$ and $\epsilon \leq a$.

Chapter 8

Conclusions and Future Work

In this PhD dissertation different aspects of limited feedback, cooperation and coordination were investigated.

8.1 Conclusions

In the context of a single isolated cell in which a BS with multiple antennas serves several users on the downlink, we proposed two different approaches to dealing with a limited feedback resource, namely a two-stage feedback approach and an adaptive feedback approach. These were shown to significantly enhance performance relative to a scheme in which all the users always feed back their CSI with the same accuracy at each scheduling interval. It appears that an adaptive scheme that relies on a single codebook but feeds back from it in a selective manner offers a good tradeoff in complexity and performance, compared to a more complex adaptive scheme.

In the multicell context, coordination, in the sense that each transmitter serves its own set of users but does not ignore how his transmission affects users in other cells, for a setup where each BS has multiple antennas and serves a single user, was tackled. A simple one-shot algorithm requiring only local CSIT was investigated and conditions under which it is expected to yield good performance in terms of system sum rate were characterized.

With enough antennas at the transmitter, resorting to an iterative algorithm to maximize the system sum rate is found to be of limited value.

Moreover, cooperation, in the sense that transmitters may jointly serve users, was looked into under two different constraints. The first is distributed CSI, i.e. the cooperating transmitters have access to the user data but base their precoding decisions on different CSI knowledge; in this context, the notion of hierarchical CSI quantization codebooks was proposed as a means of establishing some order in the way the different transmitters perceive a given user's channel. The second is a finite-capacity backhaul, which limits how much of the user data may be shared between the different transmitters: here, rate splitting was proposed so that the messages intended for a given user fall under two categories: private messages which are routed to a single BS to transmit, and common messages which are routed to several cooperating BSs. This allows for better use of the backhaul relative to a scheme where users receive their messages from a single BS (IC) or a network MIMO setup in which messages are replicated and routed to the transmitters. An alternative based on performing central processing on the core network side and routing quantized versions of the transmit signals to the different BSs is also investigated. Depending on the channel states, it could lead to better rate performance. Its feasibility in practice would however depend on the actual network structure.

Finally, three different levels of cooperation were investigated for a two-cell setup where each BS has several antennas and simultaneously serves multiple users in its 'coverage area'. In a first non-cooperative setup, each BS selfishly serves its own set of users, ignoring the interference it generates outside of its cell. In a second coordinated setup, each BS still only serves its own set of users, except that the precoding is designed jointly so that inter-cell interference is accounted for. Finally, in a fully cooperative setup, user data is routed to both BSs which pool their antennas together to serve the users. For a simple channel model, large system results are used to characterize the performance of the different schemes, thereby simplifying their comparison. Such an approach also makes it possible to determine the optimal cell loading, i.e. the number of users to serve at a given time.

8.2 Future Work

The work presented in this PhD dissertation can be extended in many ways:

- The limited feedback work carried out can be expanded to take into account more realistic channels.

- Limited feedback in the multicell case requires further investigation. Most of our multicell investigation has assumed either full CSIT, or local but perfect CSIT.

- In the distributed CSIT setup, quantifying the savings in terms of signaling relative to the loss in terms of performance warrants further investigation.

- The results in chapters 6 and 7 could be extended beyond two cells. The question of how to route traffic in a large network where BSs cooperate but where the backhaul links are finite-capacity seems to require a more general framework than the one investigated here.

- The results in Chapter 7 could also be extended to more general channel models. Using results based on large system analysis to simplify the beamforming design, particularly in a multicell setup would limit the need to exchange instantaneous CSI, only statistical CSI exchange would be required between the cells. Our approach also does not take into consideration scheduling at each BS, it assumes a set of i.i.d. users, which are all meant to be served simultaneously, are present in the cell. Treating this is another research direction of interest.

Bibliography

[1] D. Gesbert, M. Kountouris, R. Heath, C.-B. Chae, and T. Sälzer, "From single user to multiuser communications: Shifting the MIMO paradigm," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 36–46, Sept. 2007.

[2] "Special issue on exploiting limited feedback in tomorrow's wireless communication networks," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 1341–1365, 2008.

[3] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, pp. 269–275, Mar. 1991.

[4] T. Yoo, N. Jindal, and A. J. Goldsmith, "Finite-rate feedback MIMO broadcast channels with a large number of users," in *Proc. IEEE Int. Symp. on Information Theory*, July 2007.

[5] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna broadcast channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1478–1491, Sept. 2007.

[6] K. Huang, J. Andrews, and R. H. Jr., "Orthogonal beamforming for SDMA downlink with limited feedback," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Apr. 2007.

[7] S. Zhou, Z. Wang, and G. Giannakis, "Quantifying the power loss when transmit beamforming relies on finite-rate feedback," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1948 – 1957, July 2005.

[8] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[9] D. Schmidt, C. Shi, R. Berry, M. Honig, and W. Utschick, "Distributed resource allocation schemes: Pricing algorithms for power control and beamformer design in interference networks," *IEEE Signal Processing Magazine, special issue on "Game Theory in Signal Processing and Communications*, vol. 26, no. 5, pp. 53–63, Sept. 2009.

[10] K. Karakayali, G. Foschini, R. Valenzuela, and R. Yates, "On the maximum common rate achievable in a coordinated network," in *Proc. IEEE Int. Conf. on Communication*, 2006.

[11] O. Somekh, O. Simeone, O. Bar-Ness, and A. Haimovich, "Distributed multi-cell zero-forcing beamforming in cellular downlink channels," in *Proc. IEEE Global Telecomm. Conf.*, 2006.

[12] S. Jing, D. N. C. Tse, J. B. Soriaga, J. Hou, J. E. Smee, and R. Padovani, "Multicell downlink capacity with coordinated processing," *EURASIP Journal on Wireless Communications and Networking*, 2008.

[13] O. Somekh, O. Simeone, Y. Bar-Ness, A. M. Haimovich, and S. Shamai, "Cooperative multicell zero-forcing beamforming in cellular downlink channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3206 –3219, July 2009.

[14] H. Dahrouj and W. Yu, "Coordinated beamforming for the multi-cell multi-antenna wireless system," in *Proc. Conference on Information Sciences and Systems 2008, CISS 2008*, 2008, pp. 429–434.

[15] ——, "Coordinated beamforming for the multicell, multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, May 2010.

[16] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.

[17] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecomm.*, vol. 10, pp. 585–596, Nov. 1999.

[18] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE Int. Conf. on Communication*, June 1995.

[19] R. Radner, "Team decision problems," *Ann. Math. Statist.*, vol. 33, no. 3, 1962.

[20] O. Simeone, O. Somekh, H. Poor, and S. Shamai (Shitz), "Downlink multicell processing with limited-backhaul capacity," *EURASIP Journal on Advances in Signal Processing*, May 2009.

[21] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1691–1706, 2003.

[22] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.

[23] F. Rashid-Farrokhi, K. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1437–1450, Oct. 1998.

[24] D. J. Love, R. W. Heath Jr, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. A. G. Editors), "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 1341–1365, 2008.

[25] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, June 2002.

[26] 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, Feb. 2005.

[27] M. Kountouris, D. Gesbert, and T. Sälzer, "Enhanced multiuser random beamforming: Dealing with the not so large number of users case," *IEEE J. Sel. Areas Commun.*, October 2008.

[28] J. Diaz, O. Simeone, and Y. Bar-Ness, "Asymptotic analysis of reduced-feedback strategies for MIMO Gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1308–1316, Mar. 2008.

[29] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, Nov. 2006.

[30] D. Gesbert and M. S. Alouini, "How much feedback is multi-user diversity really worth?" in *Proc. IEEE Int. Conf. on Communication*, 2004.

[31] K. Huang, R. Heath, and J. Andrews, "Multi-user aware limited feedback for MIMO systems," *IEEE Trans. Signal Processing*, Jan. 2007.

[32] C. Swannack, G. W. Wornell, and E. Uysal-Biyikoglu, "MIMO broadcast scheduling with quantized channel state information," in *Proc. IEEE Int. Symp. on Information Theory*, July 2006.

[33] D. Love, R. H. Jr., W. Santipach, and M. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Communications Magazine*, vol. 42, no. 10, Oct. 2004.

[34] J. M. Cioffi, "EE 379A - digital communication: Signal processing: Lecture notes," 2007-2008.

[35] T. L. Marzetta and B. M. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. Signal Processing*, vol. 54, pp. 1268–1278, Apr. 2006.

[36] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO downlink made practical: achievable rates with simple channel state estimation and feedback schemes," 2007, submitted to IEEE Trans. Information Theory, Nov. 2007, arXiv:0711.2642.

[37] A. Edelman, "Eigenvalues and condition numbers of random matrices," Ph.D. dissertation, MIT, Mathematics Department, 1989.

[38] R. Agarwal, C.-S. Hwang, and J. Cioffi, "Opportunistic feedback protocol for achieving sum-capacity of the MIMO broadcast channel," in *Proc. IEEE Vehicular Technology Conf.*, Sept.-Oct. 2007.

[39] M. Kountouris, R. de Francisco, D. Gesbert, D. Slock, and T. Sälzer, "Efficient metrics for scheduling in MIMO broadcast channels with limited feedback," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, April 2007.

[40] G. Dimic and N. Sidiropoulos, "On downlink beamforming with greedy user selection: Performance analysis and simple new algorithm," *IEEE Trans. Signal Processing*, Oct. 2005.

[41] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer Academics, 1992.

[42] L. Lin, R. Yates, and P. Spasojevic, "Adaptive transmission with finite code rates," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1847 – 1860, May 2006.

[43] C. Swannack, E. Uysal-Biyikoglu, and G. W. Wornell, "MIMO broadcast scheduling with limited channel state information," in *Proc. 43rd Annu. Allerton Conf. Communications, Control and Computing*, IL, United States, Sept. 2005.

[44] E. Larsson and E. Jorswieck, "Competition and collaboration on the miso interference channel," in *Proc. 45rd Annu. Allerton Conf. Communications, Control and Computing*, IL, United States, 2007.

[45] G. Scutari, D. P. Palomar, and S. Barbarossa, "Asynchronous iterative water-filling for Gaussian frequency selective interference channels," *IEEE Trans. Inf. Theory*, vol. 54, pp. 2868–2878, July 2008.

[46] V. R. Cadambe and S. A. Jafar, "Interference alignment and the degrees of freedom for the K user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, p. 34253441, Aug. 2008.

[47] E. Larsson and E. Jorswieck, "Competition versus collaboration on the MISO interference channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1059–1069, Sept. 2008.

[48] E. Jorswieck and E. Larsson, "The MISO interference channel from a game-theoretic perspective: A combination of selfishness and altruism achieves Pareto optimality," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Mar./Apr. 2008.

[49] ——, "Complete characterization of the Pareto boundary for the MISO interference channel," *IEEE Trans. Signal Processing*, vol. 56, no. 10, pp. 5292–5296, Oct. 2008.

[50] K. M. Ho and D. Gesbert, "Spectrum sharing in multiple antenna channels: A distributed cooperative game theoretic approach," in *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Communications*, Sept. 2008.

[51] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Processing*, vol. 55, pp. 2646–2660, June 2007.

[52] J. Thukral and H. Bölcskei, "Interference alignment with limited feedback," in *Proc. IEEE Int. Symp. on Information Theory*, June 2009.

[53] F. Negro, S. P. Shenoy, D. T. M. Slock, and I. Ghauri, "Interference alignment limits for K-user frequency-flat MIMO," in *Eusipco 2009, 17th European Signal Processing Conference*, Aug. 2009.

[54] A. Wiesel, Y. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Processing*, vol. 54, no. 1, Jan. 2006.

[55] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[56] M. Kobayashi, M. Debbah, and J. Belfiore, "Outage efficient strategies in network MIMO with partial CSIT," in *Proc. IEEE Int. Symp. on Information Theory*, 2009.

[57] A. Papadogiannis, E. Hardouin, and D. Gesbert, "Decentralising multicell cooperative processing on the downlink : a novel robust framework," *EURASIP Journal on Wireless Communications and Networking, Special Issue on Broadband Wireless Access*, Aug. 2009.

[58] Y.-C. Ho, "Team decision theory and information structures," *Proceedings of the IEEE*, vol. 68, no. 6, June 1980.

[59] J. Tsitsiklis, "Decentralized detection," in *Advances in Signal Processing*, H. V. Poor and J. B. Thomas, Eds. JAI Press, 1993, vol. 2, pp. 297–344.

[60] K. Karakayali, R. Yates, G. Foschini, and R. Valenzuela, "Optimum zero-forcing beamforming with per-antenna power constraints," in *Proc. IEEE Int. Symp. on Information Theory*, June 2007.

[61] E. Björnson, R. Zakhour, D. Gesbert, and B. Ottersten, "Cooperative multicell precoding: Rate region characterization and distributed strategies with instantaneous and statistical CSI," *IEEE Trans. Signal Processing*, 2010, accepted for publication.

[62] F. Boccardi and H. Huang, "Optimum power allocation for the MIMO-BC zero-forcing precoder with per-antenna power constraints," in *Proc. 40th Annual Conference on Information Sciences and Systems 2006*, Mar. 2006.

[63] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses." *IEEE Trans. Signal Processing*, vol. 56, no. 9, pp. 4409–4418, 2008.

[64] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," in *The 23rd IEEE Conference on Decision and Control*, Dec. 1984.

[65] M. Kobayashi and G. Caire, "Joint beamforming and scheduling for a multi-antenna downlink with imperfect transmitter channel knowledge," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, Sept. 2007.

[66] D. Hammarwall, M. Bengsston, and B. Ottersten, "Acquiring partial csi for spatially selective transmission by instantaneous channel norm feedback," *IEEE Trans. Signal Processing*, no. 3, Mar. 2008.

[67] S. Shamai (Shitz), O. Simeone, O. Somekh, and H. Poor, "Joint multi-cell processing for downlink channels with limited-capacity backhaul," in *Proc. Information Theory and Applications Workshop*, 2008.

[68] P. Marsch and G. Fettweis, "A framework for optimizing the downlink performance of distributed antenna systems under a constrained backhaul," in *Proc. 13th European Wireless Conference (EW'07)*, Paris, France, April 2007.

[69] ——, "On base station cooperation schemes for downlink network MIMO under a constrained backhaul," in *Proc. IEEE Global Telecomm. Conf.*, Nov.-Dec. 2008.

[70] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, pp. 1037–1076, Sept. 1973.

[71] M. Mohseni, R. Zhang, and J. Cioffi, "Optimized transmission for fading multiple-access and broadcast channels with multiple antennas," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, Aug. 2006.

[72] S. Vorobyov, A. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Processing*, vol. 51, no. 2, Feb. 2003.

[73] G. Kramer, "Topics in multi-user information theory," *Foundations and Trends in Communications and Information Theory*, 2007.

[74] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, 2004.

[75] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 641–657, Mar. 1999.

[76] H. Dai and H. Poor, "Asymptotic spectral efficiency of multicell MIMO systems with frequency-flat fading," *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 2976–2988, Nov. 2003.

[77] A. Tulino, A. Lozano, and S. Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491 – 2509, july 2005.

[78] D. Aktas, M. Bacha, J. Evans, and S. Hanly, "Scaling results on the sum capacity of cellular networks with MIMO links," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3264 – 3274, july 2006.

[79] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim, "On the capacity achieving covariance matrix for Rician MIMO channels: An asymptotic approach," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1048 –1069, Mar. 2010.

[80] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658 – 2668, Oct. 2003.

[81] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 361 –374, Feb. 2006.

[82] E. Visotsky and U. Madhow, "Optimum beamforming using transmit antenna arrays," in *Proc. IEEE Vehicular Technology Conf.*, vol. 1, July 1999, pp. 851–856 vol.1.

[83] D. Tse and P. Viswanath, "Downlink-uplink duality and effective bandwidths," in *Proc. IEEE Int. Symp. on Information Theory*, 2002.

[84] H. Boche and M. Schubert, "Effective bandwidth maximization for uplink/downlink multi-antenna systems," in *Proc. IEEE Int. Conf. on Communication*, vol. 5, May 2003.

[85] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[86] V. Nguyen and J. Evans, "Multiuser transmit beamforming via regularized channel inversion: A large system analysis," in *Proc. IEEE Global Telecomm. Conf.*, Nov.-Dec. 2008.

[87] R. Zakhour and S. Hanly, "Base station cooperation on the downlink: Large system analysis," submitted to IEEE Trans. Information Theory, June 2010, arXiv:1006.3360.

[88] R. D. Yates, "A framework for uplink power control in cellular radio systems," vol. 13, pp. 1341–1347, 1996.

[89] J. Evans and D. N. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2059–2078, Sept. 2000.

[90] J. Silverstein and Z. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *J. Multivariate Anal.*, vol. 54, no. 2, pp. 175–192, 1995.

[91] S. V. Hanly and D. N. C. Tse, "Resource pooling and effective bandwidths in CDMA networks with multiuser receivers and spatial diversity," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1328–1351, May 2001.