Universidade de Aveiro
Departamento de
Electrónica, Telecomunicações e Informática
2015

Eduardo Daniel
Trujillo Castañeda

Alocação de Recursos para Sistemas Movéis Multi-Utilizador e Multi-Antena

Resource Allocation for Multiple-User Multiple-Antenna Cellular Systems

Universidade de Aveiro
Departamento de
Electrónica, Telecomunicações e Informática
2015

Eduardo Daniel Trujillo Castañeda

Alocação de Recursos para Sistemas Movéis Multi-Utilizador e Multi-Antena

Resource Allocation for Multiple-User Multiple-Antenna Cellular Systems

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requesitos necessários à obtenção do grau de Doutor em Engenharia Electrotécnica, realizada sob a orientação científica do Doutor Atílio Manuel da Silva Gameiro, Professor Associado do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro

Apoio financeiro da FCT e do FSE no âmbito do III Quadro Comunitário de Apoio.
FCT —— SFRH/BD/70130/2010

o júri / the jury

presidente / president

Prof. Doutor Fernando Manuel Bico Marques
Professor Catedrático da Universidade de Aveiro

vogais / examiners committee

Prof. Doutor Rui Miguel Henriques Dias Morgado Dinis
Professor Associado com Agregação, Facultade de Ciências e Technologia, Universidade Nova de Lisboa

Prof. Doutor Manuel Alberto Pereira Ricardo
Professor Associado, Faculdade de Engenharia, Universidade do Porto

Prof. Doutor Luis Filipe Lourenço Bernardo
Professor Auxiliar com Agregação, Facultade de Ciências e Technologia, Universidade Nova de Lisboa

Prof. Doutor Adão Paulo Soares Silva
Professor Auxiliar, Universidade de Aveiro

Prof. Doutor Atílio Manuel da Silva Gameiro
Professor Associado, Universidade de Aveiro (Orientador)

agradecimentos / acknowledgements

Foremost, I would like to thank my supervisor Dr. Atílio Gameiro for giving me the opportunity to be part of his research group and supervising my Ph.D. thesis. He always asked the right questions at the right time and taught me that fundamental knowledge lies in simple facts, *in the details*. I have learned and benefited from his great intuition and deep technical insight. I wish to thank Dr. Ramiro Sámano for guiding my first steps as a Ph.D student. I learned from him many technical and analytical skills required to conduct scientific research. He always clarified my endless doubts and supported me during these exciting and challenging years. My thanks also extend to Dr. Adão Silva for his supervision on the topics related to distributed signal processing and multi-cellular systems. He always supported my work and shared his knowledge which yielded to quite interesting research directions. I want to thank Daniel, Sara, and Carlos whose work quality and passion for research have been essential to motivate and encourage me since I joined the group. I also thank my colleagues at the MOBNET research group and other IT members for the valuable talks and fruitful discussions.
I acknowledge Instituto de Telecomunicações and University of Aveiro for providing excellent working conditions, and Fundação para a Ciência e a Tecnologia (FCT) for its support and funding.

This journey has been a great experience of professional and personal growth. It was enriched by amazing people that, at some point, shared this adventure and gave me countless life lessons. Thanks to Sofía, Diane, Elisabeth, Maryam, Wendy, Frieda, Luisa, Cate, Joana, Jorge, Pedro, Fernando, Danilo, Abbas, Nima, Aderito, Ee, David, Alejandro, Wan, Yang, and José. You guys are the cherry on the cake and make this adventure worth living. Please forgive me if your name should have been listed above and is missing.

Last, but not least, I would like to dedicate this work and to thank my parents, Paula and Rafael, and my sister Berenice for their unconditional love and support.

Eduardo Daniel Castañeda Trujillo
Dicember 2014

Palavras-chave

Sistemas Móveis, Sistemas de Antenas Distribuídos, Sistemas Multi-Utilizador, Pre-codificação linear, selecção de utilizadores, Agrupamento SDMA, Tecnologia de acesso múltiplo, Broadcast de Canais MIMO, Canais de Interferência, MIMO Multi-Utilizador, Pre-codificação Distribuída

Resumo

Esta tese descreve o problema da maximização da taxa de transmissão ou eficiência espectral em sistemas moveis tomando em atenção duas características fundamentais destes, o número de antenas e utilizadores. A fim de resolver este tipo de problema, várias técnicas de alocação de recursos foram estudadas e propostas para diferentes cenários. As antenas nos transmissores estão organizadas em diferentes configurações, podendo ser localizadas ou distribuídas e para estes esquemas, diferentes níveis de cooperação e coordenação entre transmissores foram investigados. Assumindo mais antenas receptoras do que antenas transmissoras, implica que a otimização do sistema seleccione as melhores combinações de transmissor-receptor (problema combinatório), o que pode ser concretizado usando diferentes graus de cooperação entre transmissores. Os modelos de sistemas estudados, podem ser classificados como sistemas limitados por interferência ou sistemas limitados por potência.

Em sistemas limitados por interferência a alocação de recursos é feita independentemente para cada transmissor o que resulta em perda de energia para os receptores não tomados em consideração. Para este tipo de sistemas, é considerado o caso em que a rede de acesso é constituída por antenas distribuídas. Os ganhos obtidos devido ao uso de antenas distribuídas, quer em termos do planeamento de frequências quer da maximização da taxa de transmissão são considerados. Assumindo esquemas multi-utilizador, várias técnicas e algoritmos de transmissão-recepção, alocação de potência e de taxa de transmissão foram desenvolvidos para maximizar a eficiência espectral.

Para sistemas limitados em potência os transmissores alocam os recursos quer de antenas de transmissão quer de recepção conjuntamente. Os transmissores estão equipados com várias antenas e o processamento de sinal é implementado de modo a eliminar a interferência entre utilizadores. Sistemas de célula única e de múltiplas células foram estudados. Para estes foi considerado o problema da maximização de taxa de transmissão o qual foi resolvido heuristicamente, através do desacoplamento do problema em duas partes, uma onde se efectua a seleção de utilizadores e outra onde se considera a alocação de recursos. A seleção de utilizadores é feita em função do tipo de técnicas de pré-codificação implementadas e do nível de informação que o transmissor possui. Os algoritmos de seleção de utilizadores desenvolvidos verificam a compatibilidade espacial entre utilizadores, usando para tal métricas propostas. Cada uma das métricas oferece um trade-off diferente entre a precisão para identificar um utilizador compatível e a complexidade necessária para a implementar. Foram usadas simulações numéricas para avaliar a performance das técnicas de seleção de utilizadores propostas (métricas e algoritmos), performance que foi comparada com as técnicas mais inovadoras.

keywords

Cellular Systems, Distributed Antenna System, Multiuser Systems, Linear Precoding, User Selection, SDMA grouping, Multiple Access Technology, Broadcast MIMO Channels, Interference Channels, Multiuser MIMO, Distributed Precoding

abstract

The thesis addresses the sum rate or spectral efficiency maximization problem in cellular systems with two main components, multiple antennas and multiple users. In order to solve such a problem, several resource allocation techniques are studied and developed for different cellular scenarios. The antennas at the transmitters are arranged in several configurations, i.e., co-located or distributed and for such arrangements different levels of coordination and cooperation between transmitters are investigated. Accounting for more receiver antennas than transmitter antennas implies that system optimization must select the best transmitter-receiver match (combinatorial problem) which can be solved with different degrees of cooperation between transmitters. The system models studied can be classified either as interference limited or as power limited systems.

In *interference limited systems* the resource allocation is carried out independently by each transmitter which yield power leakage to unintended receivers. For this kind of systems, the access network using distributed antenna architectures is examined. The properties of distributed antenna in cellular systems as well as the gains they provide in terms of frequency reuse and throughput are assessed. Accounting for multiple user scenarios, several techniques and algorithms for transmitter-receiver assignment, power allocation, and rate allocation are developed in order to maximize the spectral efficiency.

In *power limited systems* the transmitters jointly allocate resources among transmit and receive antennas. The transmitters are equipped with multiple antennas and signal processing is implemented in order to suppress inter-user interference. Single-cell and multi-cell systems are studied and the problem of sum rate maximization is tackled by decoupling the user selection and the resource allocation (power and precoding) processes. The user selection is a function of the type of precoding technique that is implemented and the level of information that can be processed at the transmitter. The developed user selection algorithms exploit information provided by novel channel metrics which establish the spatial compatibility between users. Each metric provides a different trade-off between the accuracy to identify compatible users, and the complexity required to compute it. Numerical simulations are used to assess the performance of the proposed user selection techniques (metrics and algorithms) whose performance are compared to state-of-the-art techniques.

Contents

List of Figures

List of Tables

List of Acronyms

3GPP	3rd Generation Partnership Project.
AMC	Adaptive Modulation and Coding.
AWGN	Additive White Gaussian Noise.
BC	Broadcast Channel.
BER	Bit Error Rate.
BF	Beamforming.
BLER	Block Error Rate.
BS	Base Station.
CAS	Co-located Antenna System.
CBF	Coordinated Beamforming.
CDF	Cumulative Distribution Function.
CDMA	Code Division Multiple Access.
CoMP	Coordinated Multi-Point.
CSI	Channel State Information.
CU	Central Unit.
DAS	Distributed Antenna System.
DCA	Dynamic Channel Allocation.
DisPC	Distributed Power Control.
DoF	Degrees-of-Freedom.
DPC	Dirty Paper Coding.
DSL	Digital Subscriber Lines.
DVSINR	Distributed Virtual Signal-to-Interference-plus-Noise Ratio.

DZF	Distributed Zero Forcing.
EPA	Equal Power Allocation.
FCA	Fixed Channel Allocation.
FDD	Frequency Division Duplex.
FDMA	Frequency Division Multiple Access.
FFP	Fractional Frequency Planning.
GP	Geometric Programming.
GSO	Gram-Schmidt Orthogonalization.
HetNet	Heterogeneous Network.
IA	Interference Aware.
ICI	Inter-cell Interference.
IFC	Interference Channel.
IPC	Individual Power Constraints.
JT	Joint Transmission or Processing.
LAN	Local Area Network.
LSAP	Linear Sum Assignment Problem.
LTE	Long Term Evolution.
LUT	Look Up Table.
MAC	Multiple Access Channel.
MCS	Modulation and Coding Schemes.
MIMO	Multiple Input Multiple Output.
MISO	Multiple Input Single Output.
MMSE	Minimum Mean Square Error.
MRC	Maximum Ratio Combining.
MU-MIMO	Multi-User Multiple Input Multiple Output.
MU-MISO	Multi-User Multiple Input Single Output.
MUDiv	Multi-User Diversity.
NSP	Null Space Projection.
OFDMA	Orthogonal Frequency Division Multiple Access.

PFS	Proportional Fairness Scheduling.
PRB	Physical Resource Block.
QAM	Quadrature Amplitude Modulation.
QoS	Quality of Service.
QPSK	Quadrature Phase Shift Keying.
RAU	Remote Antenna Units.
RHS	Right Hand Side.
RRM	Radio Resource Management.
RRS	Round Robin Scheduling.
SCA	Successive Convex Approximation.
SDMA	Spatial-Division Multiple Access.
SFR	Soft Frequency Reuse.
SIC	Successive Interference Cancellation.
SINR	Signal to Interference plus Noise Ratio.
SLNR	Signal to Leakage plus Noise Ratio.
SNR	Signal to Noise Ratio.
SU-MIMO	Single-User Multiple Input Multiple Output.
SVD	Singular Value Decomposition.
TDD	Time Division Duplex.
TDMA	Time Division Multiple Access.
THP	Tomlinson-Harashima Precoding.
TPC	Total Power Constraint.
TTI	Transmission Time Interval.
Wi-Fi	Wireless Fidelity.
WiMAX	Worldwide Interoperability for Microwave Access.
ZF	Zero-Forcing.
ZFBF	Zero Forcing Beamforming.
ZFDP	Zero Forcing Dirty Paper.

List of Symbols

$(\cdot)^H$	Hermitian transpose operator
$(\cdot)^T$	Transpose operator
$(x)^+$	$\max\{x, 0\}$
$[\mathbf{A}]_{ij}$	The element a_{ij} of matrix $\mathbf{A}$
α_{pl}	Path-loss exponent
$\setminus$	Set subtraction operator
$\binom{k}{n}$	Number of ways of picking n unordered outcomes from k possibilities
$\lambda_{\max}(\mathbf{A})$	The maximum eigenvalue of matrix $\mathbf{A}$
$\lambda_{\min}(\mathbf{A})$	The minimum eigenvalue of matrix $\mathbf{A}$
$\lambda_i(\mathbf{A})$	The ith eigenvalue of matrix $\mathbf{A}$
$\langle \mathbf{a}, \mathbf{b} \rangle$	Inner product between vectors $\mathbf{a}$ and $\mathbf{b}$
$\mathbb{E}[\cdot]$	Expectation operator
$\mathbb{R}_{++}$	The set of strictly positive real numbers
$\mathbb{R}_{+}$	The set of nonnegative real numbers
$\mathbb{Z}_{+}$	The set of nonnegative integer numbers
$\mathbf{A}$	Upper case bold symbols represent matrices
$\mathbf{A}^{[i]}$	The ith principal submatrix of $\mathbf{A}$ whose ith row and column are removed
$\mathbf{eig}(\mathbf{A})$	The vector that contains all the eigenvalues of the matrix $\mathbf{A}$
$\mathbf{e}_k$	The kth unit coordinate vector, with all elements equal to zero and the kth element equal to one
$\mathbf{I}_n$	Identity Matrix of size n
$\mathbf{x}$	Lower case bold symbols represent vectors
$\mathbf{x} \geq \mathbf{y}$	Componentwise inequality given two vectors $\mathbf{x}$ and $\mathbf{y}$

$\mathcal{CN}(a, b)$	Circular Complex Gaussian random variable with mean a and variance b
$\mid \mathcal{S} \mid$	Cardinality of set $\mathcal{S}$
$\mid \mathbf{A} \mid$	Determinant of the square matrix $\mathbf{A}$
$\parallel \mathbf{x} \parallel_p$	The p norm of vector $\mathbf{x}$
$\rho(\mathbf{A})$	The Perron-Frobenius root (PF-root) or the largest modulus eigenvalue of the matrix $\mathbf{A}$
σ^2	Noise variance
$Conv(\cdot)$	Convex-hull operator
$det(\cdot)$	Determinant operator
$diag(\mathbf{x})$	The diagonal matrix whose main diagonal is $\mathbf{x}$
K	Number of user in the system
K_F	Frequency Reuse Factor
N	Number of distributed antennas or remote antenna units
N_r	Number of antennas at the receiver
N_t	Number of antennas at the transmitter
$null(\mathbf{A})$	The null space of matrix $\mathbf{A}$
P	Available power for transmission
$rank(\cdot)$	Rank operator
$Sp(\mathbf{A})$	Subspace spanned by matrix $\mathbf{A}$
$Sp(\mathbf{A})^{\perp}$	Orthogonal subspace spanned by matrix $\mathbf{A}$
$Tr(\cdot)$	Trace operator
$y_i = [\mathbf{y}]_i$	Entry i of vector $\mathbf{y}$

Chapter 1

Introduction

THIS DISSERTATION is focused on different resource allocation techniques for multiple antenna - multiple user wireless communication systems. The research on this topic is motivated by several open problems in the current literature and standardized technologies. Multiple-antenna systems allow a plethora of mighty signal processing techniques that enhance the overall system performance by efficiently exploiting a multi-dimensional pool of resources. Such a pool can be compound of signal spaces, angles and powers of transmission, time-slots, subcarriers, codes, users, and the like. The optimization over such a large set of resource implies that an ineluctable trade-off between optimality and feasibility must be found by resource allocation policies. On the one hand, optimality can be reached by solving problems mathematically described as the optimization over a set of integer and continuous variables, which may be a thoroughly complex task. Feasibility, on the other hand, implies that suboptimal resource allocation takes place by relaxing and reformulating optimization problems whose solutions can be found by practical reliable algorithms. The research presented in this dissertation is devoted to design close-to-optimal and feasible resource allocation policies for wireless communication systems at the sub-carrier level.

1.1 Motivation

Ubiquitous connectivity of mobile communications is becoming a reality with global technological, social, economical, and environmental impact. Currently, the number of mobile users worldwide reaches almost 7 billion [1] while massive mobile internet access is expected to exceed wired devices access by 2018 [2]. Moreover, the number of networked devices is expected to top 25 billion by 2020 [3], which means that the internet of everything will bring huge challenges in terms of operational capabilities and global standardization. In order to meet the continuous market growth and the emerging data-demanding mobile applications, the deployment of mobile broadband networks (3G and 4G) has been increasing at sustained rates over the last years. Under the umbrella of heterogeneous networks (HetNet) [4], seamlessly connectivity appears to be feasible and current research efforts attempt to operate mobile devices across different wireless technologies, e.g., cellular and Wi-Fi [3]. The fourth generation of mobile communications is already under operation and is achieving maturity from the industry and academic view points. Nevertheless, the future generation of communications will bring service demands that current wireless networks are far from being capable to handle [5]. A group of emerging technologies promise to solve the practical challenges of

current and future wireless communication systems [6, 7]. Among these technologies, signal processing and infra-structure based on multiple-antenna systems, also known as *multiple input - multiple output* (MIMO), will play a crucial role. The deployment of a large number of antennas [8] in cellular architectures will yield extremely densified small serving areas where high data rates can successfully be transmitted [5].

The enormous relevance of multiple-antenna systems lies in the fact that they provide manifold gains, just to name a few: the achievable rates scaled in proportion to the number of deployed antennas, thus the radio spectrum is exploited more efficiently; interference can be mitigated by signal design and power control mechanisms; signal and spatial dimensions can be efficiently used to serve multiple users simultaneously; and MIMO can be used to model and study a large number of scenarios, e.g., cellular systems, distributed antenna systems, or HetNet. The work presented in this thesis is motivated by the mighty characteristics of MIMO, their full integration in current standardized wireless technologies, and the fact that there exists a number of open problems in the field of resources allocation with multiple user - multiple antenna scenarios. The systems studied in the thesis are particularly focused on cellular systems and access network architectures based on distributed antennas.

1.1.1 Resource Allocation

The concept of *resource allocation* in multiple-antenna systems is defined as allocating transmit power among specific users and spatial directions, while meeting a system objective function under a set of power constraints. Finding the optimum resource allocation for a wireless system is a highly complex task, even for a small number of transmitters and receivers. The system optimization must be performed over several dimensions such as antennas, powers, and users, which together form a multi-dimensional pool of resources (a very large solution space). In the literature of wireless communications there exist different approaches to solve the resource allocation problem. One approach consists of modeling the resource allocation as a convex optimization problem so that standard optimization techniques [9, 10] can be used to solve it. By taking this approach one must assume that the optimized variables take continuous values, and that the number of spatial resources at the transmitters and receivers meet certain conditions, i.e., all users can simultaneously use the same medium achieving successful communication. The work presented in this dissertation diverges from that approach due to the following reasons. 1) The sets of antennas and users are finite sets which renders the resource allocation into a combinatorial problem, i.e., the optimal solution is given by exhaustive searching over a search space that grows geometrically. 2) Feasibility of resource allocation is not guaranteed, i.e., it may not exist a solution to the constrained optimization problem. In such a case, verifying the feasibility of the optimization problem is as complex as solving it, therefore, sub-optimal strategies must be adopted.

A sub-discipline of optimization theory known as *heuristic search* provides solutions to problems that either could not be solved any other way or whose solutions take a very long time to be computed. The approach considered in this dissertation is to design, implement, and assess heuristic search algorithms that handle the combinatorial nature and feasibility issues of the resource allocation in multiple-user multiple-antenna systems.

1.1.2 Multiple Antennas

The amount of data rate per unit bandwidth that can be transferred in a given wireless channel is referred in the literature as spectral efficiency, and for a single communication link is fundamentally limited by the available transmit power [11]. Accounting for multiple antennas at the transmitter and the receiver sides provides spatial diversity and naturally yields a MIMO channel. For this kind of channels the spectral efficiency can be enhanced since several independent data streams can be transmitted simultaneously, achieving the so-called *multiplexing gain*. Resource allocation may become a complex task when multiple antennas transceivers are employed because system performance can be improved only if the spatial domain is exploited in a smart way. In multi-user systems, multiplexing gains can be obtained as long as the users are properly separated in space by steering data signals toward specific receivers. In this way, the signal power to intended users is improved whilst interference to other unintended users is partially or completely suppressed. The spatial steering of independent signals consists of manipulating their amplitude and phases (the concept of *beamforming* in classic array signal processing), in order to add them up constructively in desired directions and destructively in undesired directions [12].

Every system requires a particular resource allocation strategy, which can be mathematically described by an objective function either system-oriented (e.g. fairness or sum rate), user-oriented (e.g. user satisfaction), or operative (e.g. power consumption). This means that each system will have a different objective function and set of constraints according to its resource allocation policy. The impact of such a policy is reflected immediately in the way that beamforming must be done, i.e., it does not exist an universal way to achieve optimal beamforming. Since optimality most of the times implies high complexity, sub-optimal beamforming strategies that still achieve multiplexing gains will be studied in this dissertation.

1.1.3 Multiple Users

Provided multiple antenna at the transmitter, a large number of users yields high spectral efficiency and multiplexing gains that scale with the number of antennas at the transmitter. This is because the channels of users located in different positions within a serving area will experience independent fading processes (*user diversity*) [13]. The spatial signature of a user defines its compatibility with other users, i.e., the spatial separability and the effectiveness of the signal steering depend on the characteristics of the multi-user channels. The probability that there exists a subset of spatially compatible users (spatially separable) increases with the total number of competing users, and proper signal steering yields system performance enhancement. Yet, such improvement comes at a price, the so-called *user selection* or scheduling. The resource allocation in multi-user systems requires to select a subset of compatible users in order to achieve several goals: meeting beamforming constraints (based on each particular steering technique), fully exploiting spatial resources, and increasing the power efficiency (inter-user interference suppression). Determining the optimal subset of user is a combinatorial problem that depends on many variables: the number of deployed antennas, the channel quality after beamforming, the power allocation, and the amount of inter-user interference resulting from a given signal steering technique. Moreover, changing the resource allocation (power and signal steering) for a single user will impact the performance of other users, i.e., the optimization problem has either objective or constrained functions that are coupled in their optimized variables. For every beamforming technique (for the sake of

tractability sub-optimal techniques are used in practice) there exists a variety of strategies (optima and sub-optimal) to select the set of compatible users that maximize the system performance. Such strategies depend on the number of deployed antennas, the power regime (noise or interference limited system), and the objective function that must be optimized. The following chapters will present heuristic search algorithms whose objective is to identify a close-to-optimal set of transmitter-receiving pairs or users, that can be simultaneously scheduled taking into account the beamforming structures and power constraints so that the overall system performance is optimized.

1.2 Research Publications

The work presented in this dissertation has produced a number of scientific publications: 1 book chapter, 4 journals, and 5 conference papers listed below.

Ch1 E. Castañeda, A. Silva, and A. Gameiro, User Selection and Precoding Techniques for Rate Maximization in Broadcast MISO Systems, *Contemporary Issues in Wireless Communications*, ed. M. Khatib, InTech, 2014.

J1 E. Castañeda, R. Samano-Robles, and A. Gameiro, Frequency-Reuse Planning of the Down-Link of Distributed Antenna Systems with Maximum Ratio Combining Receivers, *IEEE Latin America Transactions*, vol. 3, pages 1703-1709, April, 2012.

J2 E. Castañeda, R. Samano-Robles, and A. Gameiro, Sum Rate Maximization via Joint Scheduling and Link Adaptation for Interference-Coupled Wireless Systems, *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, p. 268, 2013.

J3 E. Castañeda, A. Silva, R. Samano-Robles, and A. Gameiro, Low-Complexity User Selection for Rate Maximization in MIMO Broadcast Channels with Downlink Beamforming., *The Scientific World Journal.*, vol. 2014, Jan. 2014.

J4 E. Castañeda, A. Silva, R. Samano-Robles, and A. Gameiro, Distributed Linear Precoding and User Selection in Coordinated Multicell Systems, *IEEE Transactions on Vehicular Technology*, 2014 (submitted).

C1 E. Castañeda, Sámano R., Gameiro A., Frequency-Reuse Planning of the Down-Link of Distributed Antenna Systems with Maximum-Ratio-Combining (MRC) Receivers, *IEEE Latin-American Conference on Communications*, Oct., 2011.

C2 E. Castañeda, Sámano R., Gameiro A., Cooperative Scheduling for Distributed Antenna Systems, *IEEE European Signal Processing Conference*, 2012.

C3 E. Castañeda, Sámano R., Gameiro A., Low Complexity Scheduling Algorithm for the Downlink of Distributed Antenna Systems, *IEEE Vehicular Technology Conference*, June, 2013.

C4 E. Castañeda, R. Samano-Robles, and A. Gameiro, Power and Modulation Assignment via Perron-root Optimization for Interference Limited Systems, *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1899-1903, Sept., 2013.

C5 E. Castañeda, A. Silva, R. Samano-Robles, and A. Gameiro, Metrics for Rate Maximization in Multiuser-MISO Systems with Zero-Forcing Beamforming, *IEEE International Conference on Telecommunications*, May, 2014.

1.3 Outline of Dissertation

The dissertation is focused on the design of resource allocation algorithms that jointly perform user selection, beamforming, rate, and power allocation in multiple antenna systems.

The studied scenarios cover distributed antenna systems, single and multiple cell systems for which centralized an semi-centralized algorithms are designed.

Chapter 2 presents a brief overview of the main theoretical concepts that will be studied along the thesis. The general characteristics of multiple antenna - multiple user system are presented, and fundamental concepts such as capacity region and uplink-downlink duality are introduced. The definitions of broadcast channels and interference channels are also provided, which are of central importance to understand the scope of the research and results in this dissertation.

Chapter 3 introduces the concept of distributed antenna system and investigates the advantages of this sort of antenna configurations [J1, C1]. These systems are compared to conventional cellular systems where the transmit antennas are co-located. The objective of a distributed antenna system is to provide diversity and combat interference, and the research objective is to determine the performance gains in terms of frequency reuse planning and spectral efficiency. The main goal of this chapter is to quantify how resilient to interference are systems with distributed antennas, as well as the coverage and throughput enhancement that can be achieved using different transmission schemes.

Chapter 4 introduces several strategies of resource allocation in interference channels. The chapter can be divided in two parts. The first part studies how to associate a given set of transmitters with a larger set of receivers, which is a modeled as an assignment problem [C2, C3]. The second part studies the resource allocation problem once that every transmitter is matched with a unique receiver [J2, C4]. The tackled problem is how to identify which transmitter-receiver pairs (links) can be simultaneously scheduled and which rates must be allocated to them so that the total sum rate is maximized. The problem has a combinatorial nature since the rates that can be allocated are elements of a finite set of valid rates. As a means to solve such a problem, several algorithms are designed exploiting the characteristics of the achievable rate region and estimations of the power consumption.

Chapter 5 is focused on finding a set of close-to-optimal compatible users in order to maximize the achievable sum rates in single cell systems. The chapter can be divided in two parts. The first part is devoted to describe channel metrics that quantitatively measure the spatial compatibility between channels of independent users [Ch1, C5]. In the second part, the combinatorial problem of user selection for the sum rate maximization is addressed [Ch1, J3]. By employing metrics of compatibility and exploiting the structure of linear beamforming techniques, two heuristic search algorithms are designed to find sub-optimal solutions to the user selection problem and an acceptable trade-off between performance and complexity.

Chapter 6 generalizes the user selection problem for coordinated multi-cell systems [J4]. The goal is to maximize the achievable sum rates in scenarios where limited communication (message exchange) between cells is allowed. The chapter extends the channel metrics introduced in Chapter 5 for the multi-cell scenario. Close-to-optimal semi-distributed user selection algorithms are proposed and assessed for systems that operate with linear distributed beamforming. The proposed metrics of compatibility and heuristic search algorithms provide manifold gains: transform the search space of the combinatorial problem reducing its complexity; exploit the properties of each multi-cell beamforming technique in order to improve user selection; exploit metrics of selection that handle different ratio between the number of transmit and receive antennas; and provide upper bounds of the achievable performance in practical multi-cell scenarios.

Chapter 7 concludes the dissertation summarizing the main achieved results and presents future lines of work.

Chapter 2

Overview on Multiple User - Multiple Antenna Systems

This chapter presents a brief overview of the main theoretical concepts that will be studied along the thesis. The notion of multiple antenna - multiple user system is introduced and fundamental concepts such as capacity region and uplink-downlink duality are presented. Accounting for multiple antennas at both the receiver and transmitter sides, increases the degrees of freedom in which resource allocation can be performed. Multiple users provide an extra degree of freedom to the optimization domain and finding optimal operating points for a given global utility function is a highly complex task. The chapter provides a brief description of the kind of optimization problems in systems with a large optimization domain (users, antennas, power allocation, beamforming weights, time-slots, etc.) and describes their relationships with the work presented in the following chapters. Since the work developed in the next chapters is focused on the downlink transmission, the theoretical results covered in this chapter pay particular attention to such regime. It is also established the system model of the downlink transmission in interference channels, i.e., where the individual utility function for a given link (transmitter-receiver pair) is a function of the resource allocation in other concurrent links.

2.1 Introduction

A MULTIPLE-INPUT MULTIPLE OUTPUT (MIMO) system employs multiple antennas at both, transmitter (N_t) and receiver (N_r) sides in order to improve communication performance by means of signal processing techniques. The seminal works of [14, 15] provide a mathematical motivation behind multi-antenna communications, the spectral efficiency [1] measured in bits per second per Hertz (bps/Hz) increases by a factor of $\min(N_t, N_r)$ without increasing the power budget or bandwidth requirements. The signal processing techniques in multi-antenna systems can be classified in two categories, *antenna diversity techniques* and *spatial multiplexing techniques* [17]. On the one hand, antenna diversity provides transmission reliability by transforming a fading wireless channel into a more stable additive white Gaussian noise (AWGN) - like channel without signal degradation due to fading. The probability that multiple statistically independent fading channels experience deep fading is very low, which is exploited by diversity schemes. The antenna diversity techniques can be applied in both

[1] The spectral efficiency can bee seen as the *channel capacity* from the information theoretic point of view. The channel capacity is defined as the maximum rate in which an input message can be reconstructed at the output of a given communication link with a negligible probability of error [16].

the transmission and the reception. Transmission diversity schemes include space diversity, polarization diversity, time diversity, frequency diversity, and angle diversity. Examples of receive diversity schemes are selection combining, maximum ratio combining (MRC), and equal gain combining. On the other hand, spatial multiplexing techniques exploit the degrees-of-freedom (DoF) gains provided by MIMO so that different data streams are simultaneously transmitted, which increases the transmission speed by reusing the spatial dimension provided by multi-antenna transceivers [18].

2.1.1 Multiple User MIMO

In multi-user (MU-MIMO) systems, the available resources (power, bandwidth, antennas, codes, or time-slots) must be assigned simultaneously to K active users. There are two kinds of multiuser channels illustrated in Fig. 2.1: the downlink channel also known as *broadcast channel* (BC) where one transmitter communicates with many receivers; and the uplink channel also called *multiple access channel* (MAC) where many transmitters communicate with a single receiver. There are several differences between BC and MAC. In the former, the transmitted signal is a linear combination of the signals intended for each user affected by the same channel and the transmission is constrained in the total power P. In contrast, in the MAC channel, the signal from different users are affected by different channels and each user k has its individual power constraint P_k [18]. Most communication systems are bi-directional and operate in both BC and MAC channels. In order to avoid interference, each transmission is performed in orthogonal signaling dimensions, which is a signal separation called duplexing [11]. This separation can be done by allocating BC and MAC communications in different time-slots also known as time-division duplexing (TDD), or in separated frequency bands known as frequency-division duplexing (FDD). An advantage of TDD over FDD is that bi-directional channels are typically symmetrical in their channel gains, which allows the estimation of one channel direction if the other is known. In other words, *channel state information* (CSI) can be obtained in TDD, where the reciprocal uplink and downlink channels are time-multiplexed on the same physical wireless channel [19].

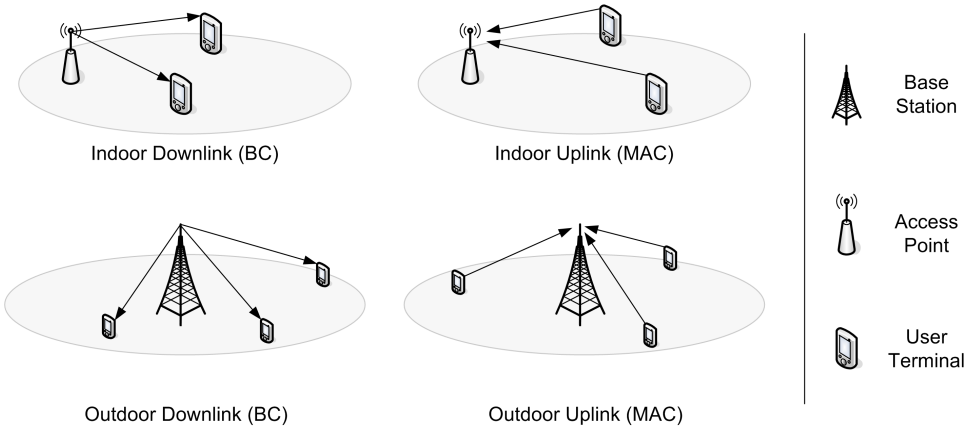

Figure 2.1: BC and MAC for indoor and outdoor scenarios.

In MU-MIMO BC systems, the performance depends on the efficient assignment of the resources at the transmitter or base station (BS). Moreover, if CSI is known at the transmitter the gain is twofold: full spatial degrees of freedom can be attained [18] and the system can

be optimized over a new degree of freedom given by the users. In the literature of MU-MIMO, two dimensions of optimization are studied: *multiplexing diversity* and *multiuser diversity* (MUDiv). The former is a consequence of the independent fading across all MIMO links, which yields a set of parallel spatial channels where different data streams can be transmitted increasing the system capacity [20]. The latter is given when users that are geographically far apart have channels that fade independently at any point in time. For a specific system performance optimization, e.g., sum rate maximization, such an independent fading process is exploited so that the user with specific channel characteristics (e.g., large channel gains and/or spatially uncorrelated channels) will be selected for transmission most likely [21]. Notice that the sum rate is a single number that defines the maximum throughput (amount of error free information successfully transmitted) of the system achieved by the the simultaneously scheduled users, regardless of fairness [11]. MU-MIMO techniques provide manifold gains [22]: multiple antennas attain diversity gain which improves bit error rates (BER); directivity gains are realized by MUDiv since the spatial signatures of the users are uncorrelated, which mitigates inter-user interference; immunity to propagation limitations in single-user MIMO (SU-MIMO) such as rank loss or antenna correlation; and multiplexing gains scale at most N_t (if full CSI is available at the transmitter), which increases the system achievable rates [23].

The achievable rates can be increased when MIMO techniques are combined with signal processing so that independent messages are transmitted to different users by controling power allocation and mitigating/suppressing inter-user interference. Multiple access techniques divide the total signal dimensions into channels which can be assign to different users. The channels are created by a division that can be orthogonal or non-orthogonal along different dimensions. Time-division multiple access (TDMA) and frequency-division multiple access (FDMA) are techniques that create orthogonal channels, whilst code-division multiple access (CDMA) can be orthogonal or non-orthogonal depending on the code design [11]. The spatial dimension provided by the multi-antenna transceivers can be used to create orthogonal or non-orthogonal channelization schemes which is known as spacial-division multiple access (SDMA) [18]. Multiple access techniques are usually combined to exploit different dimensions of optimization. For instance, to achieve orthogonality in time, all the spatial resources at the BS (N_t antennas) are used to communicate with one user (N_r antennas) at a time which is known as SU-MIMO with TDMA. This technique avoids inter-user interference, achieves power gains that scale with N_t, enhances data rates for a single user specially at low signal-to-noise ratio (SNR) regime, and is robust to CSI uncertainty [13, 18, 22]. SDMA exploits CSI at the BS allowing $K > 1$ users to be scheduled at the same time achieving multiplexing gains of at most $\min(N_t, KN_r)$ at the high SNR regime, where the system capacity is limited by the spatial DoF and not by power.

2.2 Capacity of Multi-User MIMO Systems

In multi-user systems, multiple access techniques refer to the assignment of signal dimensions to specific users by exploiting the fact that different users impinge different signatures in time, frequency, code, or space. Since the assignment of the system resources can be given in an infinite number of ways, the capacity of the entire system is given by a *rate region* rather by a single number. The rate region describes all the rates that can be simultaneously supported by the channel with an arbitrary small error probability [11]. The union of the

all the achievable rate vectors under all multi-user transmission strategies is called the *capacity region* of the multi-user system. The fundamental differences between BC and MAC channels, discussed below, imply that their capacity regions are different. However, there is a duality between these channels which allows to find the capacity region of one channel from the capacity region of the other one.

2.2.1 Capacity of Broadcast Channels

The capacity region of the degraded BC is known [16], where the Gaussian channel has a scalar input and a scalar output (single antenna transceivers). In such case, the capacity region is achieved by superposition coding at the transmitter and interference subtraction at the receivers [19]. For the case of non-degraded BC, the capacity region is only known for special cases. Since the BS has knowledge of the data symbols and CSI, the multi-user transmission can be optimized using coding techniques. The achievable *rate region* in BC is based on dirty paper coding (DPC) [24]. The principle behind this optimum coding technique is that the BS knows the interference for a given user and can pre-subtract it before transmission, which yields the capacity of an interference free channel. The seminal work of Caire and Shamai [25] established an achievable rate region for a two-users MIMO scenario and it was proved that a generalization of such a scheme with DPC achieves the entire capacity region of the MIMO BC [26].

Consider a system with a single BC with K users where the $N_r \times N_t$ channel matrix $\mathbf{H}_k$ summarizes the channel gains between the antennas at the BS and each antenna at the kth user. The received signal for the kth user is defined as

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x} + \mathbf{n}_k,$$

where $\mathbf{x}$ is the input to the transmit antennas and $\mathbf{\Sigma}_x = \mathbb{E}[\mathbf{x}\mathbf{x}^H]$ denotes its covariance matrix. It is assumed that the noise vector $\mathbf{n}_k$ is circular symmetric complex Gaussian with covariance $\mathbb{E}[\mathbf{n}\mathbf{n}^H] = \sigma^2\mathbf{I}$. When $N_t > 1$, the BC is called non-degraded, which means that the receivers cannot be ranked by their channel quality since there are multiple gains associated with each transmit-receive antenna pair. Let $\pi(\cdot)$ denote a permutation of the user indices, $\mathbf{\Sigma} = [\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_K]$ denote the set of positive semi-definite covariance matrices meeting a total power constraint, i.e., $Tr(\mathbf{\Sigma}_1 + \ldots + \mathbf{\Sigma}_K) \leq P$. By employing DPC the user $\pi(k)$ is encoded after user $\pi(k-1)$, thus the following rate vector is achievable [11, 26]:

$$\mathbf{R}(\pi, \mathbf{\Sigma}) : R_{\pi(k)} = \log_2 \frac{|\mathbf{I} + \mathbf{H}_{\pi(k)}(\sum_{j \geq k} \mathbf{\Sigma}_{\pi(j)})\mathbf{H}_{\pi(k)}^H|}{|\mathbf{I} + \mathbf{H}_{\pi(k)}(\sum_{j > k} \mathbf{\Sigma}_{\pi(j)})\mathbf{H}_{\pi(k)}^H|}, \quad k \in \{1, \ldots, K\}. \tag{2.1}$$

The BC capacity region is the convex hull of the union of all rate vectors over all permutations and all covariance matrices satisfying the average power constraint:

$$\mathcal{C}_{BC}(P, \mathbf{H}) = Conv\left(\bigcup_{\pi, \mathbf{\Sigma}} \mathbf{R}(\pi, \mathbf{\Sigma})\right). \tag{2.2}$$

The DPC implies that the components of $\mathbf{x} = \mathbf{x}_1, \ldots, \mathbf{x}_K$ are uncorrelated, so that $\mathbf{\Sigma}_1 + \ldots + \mathbf{\Sigma}_K \leq P$. The rate equations defined in (2.1) are neither convex nor concave functions of the covariance matrices. This implies that the computation of the capacity region must be done by exhaustive searching over the entire space of covariance matrices that meet the power

constraint [25]. However, there exists a strong connection between BC and MAC that has been used in the literature to define the capacity region in the downlink direction [11, 18, 26].

2.2.2 Uplink-Downlink Duality

The duality between the DPC for the MIMO BC and the capacity region of the MIMO MAC [18, §10] was established by taking into account the reciprocity of both channels and the duality between scalar Gaussian BC-MAC [27]. Unlike BC, in MAC the rate maximization can be formulated as a concave function of the covariance matrices. In other words, the MAC capacity region can be fully characterized as a convex optimization problem [27]. The connection between both channels allows the DPC region (2.5) to be found using convex optimization techniques. The achievable rate of the kth user computed in (2.1) can be reformulated as a function of $\mathbf{W}_k$ and $\mathbf{G}_k$ the precoder matrix and the receive filter matrix of the kth user $\forall k = 1, \ldots, K$, respectively, with an upper bound established as follows [28]:

$$R_{\pi(k)}(\pi, \mathbf{W}) \leq \log_2 \left| \mathbf{I}_{N_r} + \left(\mathbf{A}_{\pi(k)}\right)^{-1} \mathbf{H}_{\pi(k)} \mathbf{W}_{\pi(k)}^H \mathbf{W}_{\pi(k)} \mathbf{H}_{\pi(k)}^H \right| \tag{2.3}$$

$$= \log_2 \left| \mathbf{I}_{N_t} + \left(\mathbf{B}_{\pi(k)}\right)^{-1} \mathbf{H}_{\pi(k)}^H \mathbf{W}_{\pi(k)} \mathbf{W}_{\pi(k)}^H \mathbf{H}_{\pi(k)} \right| \tag{2.4}$$

where

$$\mathbf{A}_{\pi(k)} = \sigma^2 \mathbf{I}_{N_r} + \sum_{\pi(j) > \pi(k)} \mathbf{H}_{\pi(k)} \mathbf{W}_{\pi(j)}^H \mathbf{W}_{\pi(j)} \mathbf{H}_{\pi(k)}^H$$

$$\mathbf{B}_{\pi(k)} = \sigma^2 \mathbf{I}_{N_t} + \sum_{\pi(j) < \pi(k)} \mathbf{H}_{\pi(j)}^H \mathbf{G}_{\pi(j)} \mathbf{G}_{\pi(j)}^H \mathbf{H}_{\pi(j)}$$

are the interference plus noise components of the downlink and uplink respectively. The expression in (2.4) is equivalent to the uplink rate bound for a MAC, given fixed transmit filters $\mathbf{G}_k$ and a reverse decoding order $\bar{\pi}$. The computation of the optimal precoders that maximize a particular weighted sum rate is highly complex since (2.3) is not convex in $\mathbf{W}$[2]. In fact, such optimal precoders require SINR balancing [31–33] and its evaluation is achieved by computationally demanding algorithms [34, 35]. The equality in (2.4) holds for all users if and only if reciprocal channels are assumed ($\mathbf{H}_{BC} = \mathbf{H}_{MAC}^H$) and the sum power is the same in both cases, i.e., $Tr(\sum_k \mathbf{G}_k \mathbf{G}_k^H) = Tr(\sum_k \mathbf{W}_k \mathbf{W}_k^H)$ which is equivalent to the constraint over the covariance matrices in (2.1). The MIMO MAC is typically subject to a per-user power constraint P_k, and the capacity region of MIMO BC is obtained by taking the convex hull around all possible MAC regions with different per-user powers summing up to the same overall power, i.e., $\sum_k P_k = P$. Mathematically, the BC capacity region can be expressed as the union of the capacity regions for its dual MAC with a pool power constraint as [27]:

$$\mathcal{C}_{BC}(P, \mathbf{H}) = \bigcup_{\{P_1, \ldots, P_K\} : \sum_{k=1}^K P_k = P} \mathcal{C}_{MAC}(P_1, \ldots, P_K; \mathbf{H}), \tag{2.5}$$

[2]The computation of the optimal downlink precoders $\mathbf{W}$ can be reformulated as a number of optimization problems, e.g. semi-definite program, second-order cone program, and Lagrangian dual which transforms the original non-convex problem into a very specific constrained convex optimization problem, cf. [29, 30] and references therein.

where $\mathcal{C}_{BC}$ is the AWGN BC capacity region with total power constraint P and channel matrix $\mathbf{H}$, and $\mathcal{C}_{MAC}$ is the AWGN MAC capacity region with individual power constraints $P_1, \ldots, P_K$ and the same channel matrix. The geometric interpretation of (2.5) is illustrated in Fig. 2.2 for a system with two users where the BC capacity region is formed from the union of the MAC capacity regions with different power allocations between the uplink transmitters that sum up the total power P of the dual BC [11]. The exact shape of the capacity, rate, or another performance region depends on the power constraints, the channel gains, and the correlation between the channel vectors of the scheduled users, which will be elaborated upon in Chapter 4. Understanding the geometry of the BC capacity region is fundamental to achieve operation points where either the weighted sum rate is maximized, power consumption is minimized, time-sharing between user is defined, or any other global system goal is achieved [36][37, §5]. The main results in Chapter 4 are derived from the properties of the BC capacity region under individual SINR constraints, where optimal and sub-optimal power and rate allocation schemes for sum rate maximization are designed. It is worth mentioning that a large amount of theoretical and practical (algorithms) research is built on the basis provided by the uplink-downlink duality and the characteristics of the capacity or achievable SINR region in downlink transmission, cf. [12, 32, 36–38].

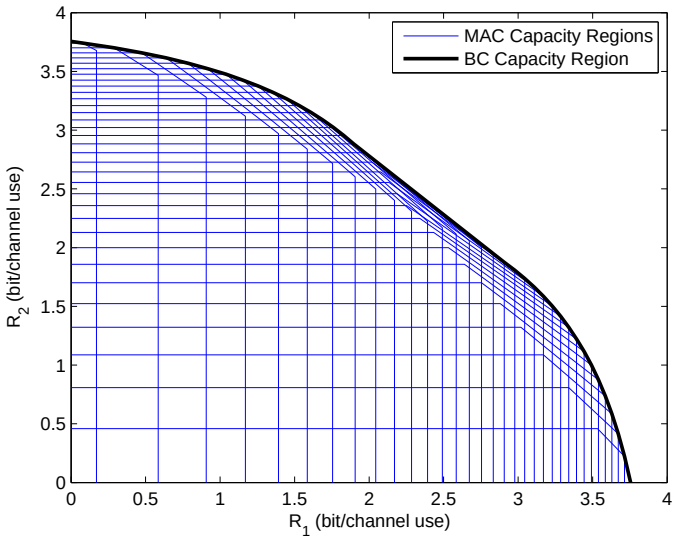

Figure 2.2: Capacity Regions of a two-user MAC and BC for $\mathbf{H}_1 = [1, 0.5]$, $\mathbf{H}_2 = [0.5, 1]$, $P = 10$.

The optimal precoding matrices $\mathbf{W}_k$ in (2.3) can be calculated if the dual transmit filters $\mathbf{G}_k$ are known and viceversa, which is illustrated in Fig. 2.3. This can be achieved computing $\mathbf{B}_k$ $\forall k$ from $\mathbf{G}_k$ and determining $\mathbf{A}_k$ and $\mathbf{W}_k$ iteratively [27, 28]. Uplink-Downlink duality can be also used to calculate capacity regions and precoding matrices $\mathbf{W}_k$ for the MIMO BC transmission with *per-antenna* power constraints [29]. In the uplink the optimal receiver implements successive interference cancellation (SIC) decoding, which achieves the MAC capacity region [18] and can be implemented with reasonable effort [39] by using only receiver-side CSI. The DPC encoding is the dual of SIC decoding but it is mainly a theoretical construct since its implementation requires sophisticated random coding and other signal processing techniques which are highly complex to implement [40]. Nevertheless, the DPC

rate region establishes the fundamental performance limits for MU-MIMO in the downlink transmission. Sub-optimal yet practical non-linear precoding schemes have been proposed in order to reduce complexity of DPC, such as the Tomlinson-Harashima precoding (THP) [18] or vector-perturbation [41] which require highly accurate CSI at the transmitter-side. Linear precoding techniques are also feasible sub-optimal alternatives to DPC and they will be used in Chapter 5 and Chapter 6 to maximize the average sum rate in scenarios where user selection is required, i.e., $N_t < KN_r$.

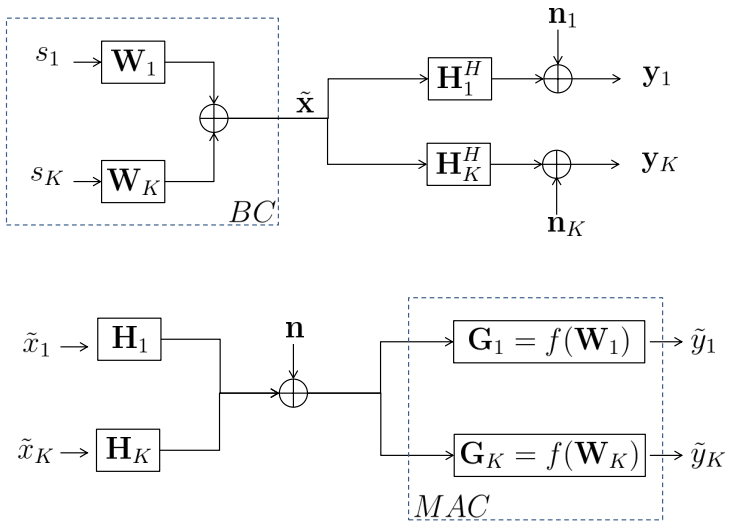

Figure 2.3: Uplink-Downlink Precoding Duality

2.3 Transmission Methods for Broadcast Channels

DPC is a nonlinear process that requires successive encoding and decoding whose complexity is prohibitive in practical systems and alternative SDMA transmission schemes are preferred instead. In the literature, DPC has been interpreted as *beamforming* (BF) [25] which is a transmission scheme where data streams of different users are encoded independently and multiplied by weight vectors $\mathbf{W}$ in order to mitigate mutual interference. BF can be defined as the steering of data signals towards intended/selected users by means of array processing (modifying signal amplitudes and phases) so that the received signal power is increased and the inter-user interference is mitigated [12]. Although practical BF schemes are sub-optimal, several works (e.g., [13, 25]) have shown that they can achieve a large portion of the DPC rate and their performance is close-to-optimal for large N_t and K [40]. The specific selection of the weight vector of a given user may affect the performance of other users, i.e., the achievable SINR of one user is coupled with the other users weight vectors and transmit powers [35]. Since the optimum BF weight vectors must be jointly optimized with power allocation [12], sub-optimal weights based on zero-forcing (ZF) or minimum mean square error (MMSE) can be used [25, 42] in order to reduce complexity.

The joint optimization of the beamforming weights, the power allocation, and the set of links that are scheduled under SDMA transmission is performed for a given global objective function. A common global figure of merit in the downlink transmission is the *total achievable*

sum rate because it quantifies how much total data flow is possible in a BC [43]. The achievable rate defines the number of bits that can be conveyed to a given user with an arbitrary low probability of error [16]. In the literature, the achievable rate is usually given by the Shannon capacity formula, and the performance figures are built on idealized decoding, infinite symbol constellations, and error-control coding over a large number of channel instances, which provides insight w.r.t. the maximum achievable performance and its upper bounds [12]. In the literature of BC systems a large body of research (see Chapter 5 and Chapter 6) is focused on maximizing the sum rate for a specific BF scheme, and the open problems are power allocation and user selection. If the inequality $N_t \geq KN_r$ is fulfilled, power allocation can be computed straightforwardly either by Lagrangian methods (water-filling based on convex optimization [9, 10]) or by equal power allocation (close-to-optimal for the high SNR regime [33, 44–47]). However, when the number of total receive antennas is larger than the number of spatial resources at the BS, i.e., $N_t < KN_r$, user selection is required prior to power allocation if linear precoding is used. For non-linear precoding, studies show that the optimal number of scheduled users with nonzero allocated power at every channel instance is upper bounded by N_t^2 [48]. The selection of the optimum set of users that maximizes the sum rate for a given BF scheme with optimum power allocation is a mixed binary non-convex problem. A mixed binary problem is one where some of the decision variables are constrained to be either 0 or 1. This kind of problems are hard to solve and the computational complexity required to find their solutions will be discussed and illustrated in Chapter 5. In systems where users must be allocated in different radio resources (time-slots or sub-channels) finding the optimum subsets of users under SDMA transmission for each radio resource is a NP-complete problem whose optimum solution can be found via exhaustive search [49]. Recent works reviewed in Chapter 4, Chapter 5, and Chapter 6 have proposed a number of feasible heuristics algorithms that find a sub-optimal yet acceptable solution to the sum rate maximization problem with SDMA communication.

The literature of MU-MIMO has been focused on single-carrier scenarios with ZF-based BF schemes due to their tractability and the fact that some channels characteristics can be used to estimate the reliability of joint transmission for a given set of users. The main objective of joint scheduling and BF is to make better decisions at the medium access control layer by exploiting information from the physical layer (wireless channels knowledge) [50]. Cross-layer scheduling designs usually consider limited information from physical layer and constraints of upper layers related with queuing theory, quality of service (QoS), and delay constraints [51]. The proposed algorithms and literature reviewed in the following chapters addresses the scheduling process in the medium access control layer using exclusively physical layer information, i.e., CSI at the transmitter.

2.4 Cellular MU-MIMO Systems

The fixed deployment of transmitters (base stations or access points) throughout a given area has the objective of providing reliable communication to mobile terminals to a backbone wire network. This kind of infra-structure based networks include cellular systems and wireless local area networks (LAN). Such kind of systems are controlled by a centralized processing unit (CU) in charge of the resource allocation (power, space, time, frequency, and code), provided some known information regarding to the CSI and the interference structure at the transmitter [19]. The premise behind cellular systems is to exploit the power falloff with

distance of signal propagation to reuse the same channel at spatially-separated locations. The cellular TDMA, FDMA, and non-orthogonal CDMA systems can be designed as virtually interference-free by planning the *frequency-reuse distance*. This means that the serving area is divided into non overlapping cells and any cell site within a neighborhood (cluster) cannot use the same frequency channel, which makes the same reused frequency channels sufficiently far apart [52]. Recent work in the area of distributed antenna systems (DAS) [53–56] shows the potential of combining such access network infra-structure with both *frequency reuse* (described in detail in Chapter 3) and *coordinated multi-cell downlink transmission* (briefly described below and in Chapter 6). The integration of these two components is fundamental to enhance performance in terms of throughput and coverage in current cellular systems and in their successors, the HetNet [55]. The characteristics of this kind of frequency separation techniques combined with distributed antenna deployments and different transmission and reception schemes will be further elaborated upon in Chapter 3. The same system model employing distributed antennas at the access network will be used in Chapter 4 and efficient resource allocation algorithms will be tasted for such architectures.

In cellular systems, interference is generated by two sources [57]: other devices in the same cell (intra-cell interference) and co-channel interference from other cells , i.e., inter-cell interference (ICI). In traditional cellular systems, each user belongs to one cell at a time and resource allocation is performed unilaterally by its respective BS. If frequency reuse is employed, the BS can make autonomous resource allocation decisions and be sure that no uncoordinated ICI appears within the cell [12]. In the literature it is common to find single-cell models where the BS is equipped with several antennas and a set of K users compete for radio resources. In such a MU-MIMO scenario the ICI is either negligible or it is assumed to be part of the additive background noise. Therefore, intra-cell interference is the main limiting performance factor, and precoding techniques such the ones describe in Chapter 5 can be used to enhanced the achievable sum rate in BC. In multi-cell systems, the mitigation of ICI is a fundamental problem since the transmit strategy chosen by one BS will affect the reception of the users served by adjacent BSs. Systems with static frequency-reuse (cluster organization) achieve limited spectral efficiency when the user distribution is heterogeneous and suffer from severe ICI [28, 30]. However, dynamic and flexible multi-cell coordination can be performed in a *user-centric* perspective. This implies that the set of base stations that serve a given user is based on the particular needs of this user. Under this approach, each BS will coordinate ICI mitigation to a specific group of users and only send data to it. A BS within a cluster coordinates its resource allocation decisions with the rest of the BSs that affect the same user. In this way, each BS cooperates with its neighbors and dynamically forms a user oriented cluster. Dynamic clustering and user-centric communication are ongoing research topics (e.g., [12, 52] and references therein) which promise to meet the requirements established in the third generation partnership project (3GPP) standards [58].

Different forms of ICI control have been proposed in the literature over the last years. An extension of SDMA for multi-cell systems that has received several names, *coordinated multi-point* (CoMP) [28, 58, 59], *network MIMO* [60], or joint signal transmission/processing (JT) [52], attempts to exploit the spatial dimensions serving multiple users (specially cell edge users [12]) while canceling ICI of BSs within a cluster. In this approach, a cluster of BSs is treated as a super cell where mathematical models from the single-cell scenario can be applied straightforwardly [61]. The fact that user data is shared among BSs allows the use of *proactive* interference mitigation within a cluster. A proactive treatment of interference means that coordinated BSs do not tune separately their physical layer/multiple access control layer

parameters but instead coordinate their coding and decoding, exploiting knowledge of global data and CSI [30]. However, to guarantee the high performance for this kind of systems, several conditions must be met [12, 52]: global CSI and data sharing are required, which scales the demands on the channel estimation, feedback links, backhaul networks, and tight BS cooperation[3]; coherent joint transmission and accurate synchronization (increasing the delay spread); and finally, complex centralized resource allocation algorithms which may be infeasible in terms of computation load, delays, and scalability. There is another approach of cooperation defined in the 3GPP standards called *coordinated beamforming* (CBF) which is a form of interference coordination [52, 58]. Interference coordination refers to the partial or total sharing of CSI among BSs to estimate beamforming weights, power allocation, and scheduling without sharing data or signal-level synchronization [30]. CBF implies that each BS has a disjoint set of users to serve with data but selects transmit strategies jointly with all other BSs to reduce ICI. In this approach user data is not exchanged among BSs in a cluster but control information and CSI are exchanged in order to perform simultaneous transmission to an arbitrary number of users [59]. Each user receives data from one BS and its performance is enhanced by jointly designing the beamforming weights which steers the interference toward a specific spatial dimension, e.g., the null space spanned by the unintended user channel. CBF requires less signaling load than CoMP, but in practice it needs accurate CSI (feedback and backhaul requirements) and scheduling coordination. The special case where only one user is served per BS is called the interference channel and it is of particular interest due to the fact that multi-cell coordination can be performed in an user-centric fashion with a low price in terms of message exchanged between BSs.

2.4.1 Interference Channels

The system model of interference channels (IFC) arises in scenarios with at least two transmitters that may belong to independent wireless systems, using concurrently the same spectral band to send information to different users. IFC can be used to characterize cellular, ad hoc, digital subscriber lines (DSL), and cognitive radio systems (illustrated in Fig. 2.4) and it has been an active research field over the last years [63–67]. The capacity region for an arbitrary number of transmitter-receivers pairs equipped with multiple antennas is still an open problem, and it is only well understood for some particular cases. However, valuable progress has been achieved in the field, for instance the works in [65, 66] studied multiple-input single-output (MISO) for a Gaussian IFC with two single antenna users and analytical results showed that: interference may have a structure that can be actually used to improve performance instead of simply adding it to the background noise; and by restricting each receiver to implement single-user detection, transmitter beamforming can be sufficient to achieve all boundary points of the rate region. Single-user detection means that users treat co-channel interference as noise, i.e., they make no attempt to decode and subtract the interference. This is a widely used assumption in the literature of the downlink transmission (this will also be assumed in Chapter 5 and Chapter 6), whilst multi-user detection is well understood and has practical applications in the uplink transmission [57].

The authors in [64] studied MISO IFC and they showed that when multiple antennas are employed at each transmitters, there exists a fundamental need for cooperation in order to

[3]Recent research [62] shows that in the high SNR regime of practical multi-cell systems with clustered cells performing coordinated/cooperative transmission, the achievable system capacity is fundamentally interference limited due to the out of cluster interference.

achieve maximum spectral efficiency when the transmitters belong to different communication systems. Theoretical results show that maximizing the proportional rate allocation, i.e. $\sum_k \log(R_k)\ \forall k$, in scenarios with independent systems (e.g. heterogeneous networks) yields a close-to-optimal joint transmission strategy (beamforming design and power allocation). And when $N_t \geq KN_r$ the maximization of the proportional fairness utility function in [64] is closely connected with a Nash bargaining in the context of game theory and resource allocation theory [68]. However, even for fully centralized systems, authors in [61] showed that the optimal coordinated beamforming is a NP-hard problem for multiple antennas at the transmitter. Moreover, complexity analysis suggest that the computation of global optimal beamforming vectors should be reconsidered (avoided) for general MU-MISO IFC scenario unless the global system goal is described by the max-min problem[4] since such an objective function can be optimized in polynomial time. The authors in [61] analyzed the problem of maximizing the weighted sum rate, proportional rate allocation, harmonic-mean utility, and max-min rate allocation for MISO systems, where only the last one can be solved by standard convex optimization techniques. Nevertheless, for the general MIMO case all four problems are NP-hard [67].

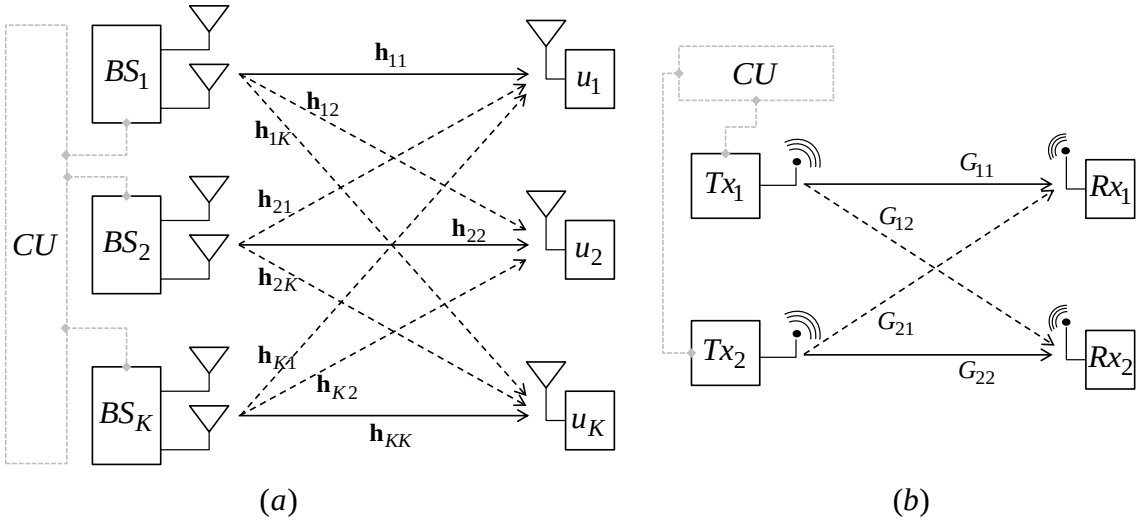

Figure 2.4: Models of Interference Channels. (a) is a K-cell MISO IFC system with signal design capabilities, i.e., beamforming design and power allocation can efficiently suppress interference; (b) is an IFC system with fixed channel gains (e.g., predefined beamforming weights) and interference can be mitigated only by power allocation and scheduling. The solid arrows refer to the useful signal directions, while the dashed arrows refer to interference directions. Coordination through a CU depends on the kind of processing, i.e., centralized or distributed.

In this thesis multi-user IFC scenarios will be studied in two different perspectives for the sum rate maximization problem: 1) one approach that allows rate and power allocation, 2) a second approach that performs signal design, and both approaches implement user selection. Fig. 2.4(b) illustrates a widely studied system model used to solve the QoS (measured by the achievable rate) constrained power minimization [32, 36, 37, 69, 70]. This problem has

[4]This utility function maximizes the fairness over the rate distribution among users since the minimum instantaneous rate is maximized which generally requires to transmit less power to users with good channels.

been studied in several scenarios (MIMO, MISO, etc.) and for both regimes BC and MAC. The algorithms to solve such a problem implement sophisticated optimization techniques to determine the optimal resource allocation, i.e., the optimal set of scheduled links, and their respective beamforming weights and power allocation [67, and references therein]. In Chapter 4 the same problem will be analyzed albeit with a particular set of constraints over the rates that can be assign to each link. Given a set of links in the IFC regime the questions to be answered are, how to decide which links must be active? and which data rates must be allocated to the subset of active links in order to maximize the sum rate?. It will be shown that the problem has a complex combinatorial nature and resource allocation algorithms will be designed so that close-to-optimal solutions for such a complex problem can be found.

Fig. 2.4(a) shows a widely used multi-cell MISO BC/IFC system model [12, 63, 64, 71] where a set of BSs can cooperate/coordinate their transmission in order to find optimal operational points. This implies the joint design of beamforming weights and power allocation to optimize an utility function (maximum sum rate or proportional fairness), which turns the problem NP-hard [61]. The results in [63] show that when the transmitters belong to the same system, e.g., a clustered multi-cell system, the cooperation between them reduces exclusively to control message exchange avoiding user data or CSI sharing. This means that optimal operational points of the system can be achieved by designing beamforming weights based on local CSI. In Chapter 6 close-to-optimal solutions to the sum rate maximization problem are found by using linear beamforming techniques (cf. [63, 72]) in clustered multi-cell MU-MISO systems. The multi-user factor ($N_t < KN_r$) transforms the problem of finding an optimal operational point into a mixed non-convex problem. It will be shown that close-to-optimal operation points can be found by combining sub-optimal yet efficient user selection algorithms and distributed signal design.

Chapter 3

Distributed Antenna Systems and Cellular Architectures

This chapter presents a comparative study of the distributed antenna systems (DAS) and conventional cellular systems with co-located antennas (CAS) and the impact of such architectures in frequency planning. The main characteristics of DAS and a brief discussion about frequency planning are presented. Two performance metrics, the throughput and outage probability are defined for different download transmission schemes. Numerical results show that DAS can increase the spectral efficiency of cellular systems since they can support intensive frequency reuse. This also implies that inter-cell interference can be suppressed in a more efficient way than in conventional CAS systems. The effects of multiple antenna receivers is also studied. Numerical simulations [J1, C1] show that MRC processing is complementary to DAS architecture and it can highly improve performance specially in the low SNR regime.

3.1 Introduction

A cellular distributed antenna system (DAS) consists of several antennas geographically distributed within the cell in order to accomplish several goals: increase coverage area, reduce the access distance to the users, and make a more resilient system to ICI [73]. Each distributed antenna is connected to a central controller via a dedicated link, either by a wireless or an optical connection [74]. This architecture mimics a macroscopic multi-antenna system which can achieve low values of signal correlation [75]. As a consequence, DAS allows the implementation of efficient signal processing schemes which may be inefficient in co-located antenna systems (CAS) due to the severe signal correlation [76]. DAS systems were initially designed to solve indoor coverage problems [77]. Nonetheless, due to the huge potential of such an access architecture, a considerable amount of theoretical and practical research has been done in the last decade in order to find applications in outdoor scenarios (cf., [76, 78] and references therein). As a matter of fact, DAS are tightly integrated into cellular architecture, and current wireless standards already consider access network deployments combined to physical layer technologies like MU-MIMO [55].

DAS systems have been studied in the literature using different performance metrics. In [79, 80] the authors studied the ability of DAS to minimize power consumption considering different antenna configurations. In [75] the authors showed that DAS can highly improve fairness and spectral efficiency. Numerical results showed that since DAS provide macro-

diversity (independent fading channels) the minimum allocated rate per user increases since ICI can be mitigated efficiently. The work in [81] studied adaptive modulation in DAS and showed that significant power savings and capacity gains can be attained w.r.t. conventional systems. Channel capacity and outage probability of DAS with fixed antenna location has been addressed in [78, 82], while the case of random antenna location considering multicell scenarios and universal frequency reuse was studied in [83].

Universal frequency reuse means that the frequency reuse factor (K_F) is equal to 1 which is one of the possible frequency planning configurations in Worldwide Interoperability for Microwave Access (WiMAX) and Long Term Evolution (LTE) systems [84]. K_F specifies the way in which spectral resources (sub-channels) are allocated in a cellular deployment and is a function of the tolerated interference in the system. K_F can be defined as the number of adjacent cells which cannot use the same frequencies for transmission. Its inverse, $1/K_F$, is a factor to indicate how efficiently the bandwidth is used in the cellular system [17]. The spectral efficiency is enhanced when the available bandwidth in the cellular system is reused by neighboring cells, i.e., when the frequency reuse factor is as small as possible [85]. For instance, if K_F is reduced while capacity over a given coverage area remains constant, then less radio resources are required and higher spectral efficiency is achieved.

Despite the recent advances in the field of multiple antenna systems, the influence of frequency reuse planning on the downlink performance of DAS has not been fully investigated in the literature. Moreover, multiuser multi-carrier systems such as Orthogonal Frequency Division Multiple Access (OFDMA) do not have a processing gain as the CDMA systems and a fundamental problem is to achieve both, universal frequency reuse and ICI suppression [84]. In [86] Simonsson explored interference cancellation using different frequency planning techniques and investigated the effects of receivers with multiple antennas to combat ICI for cell-edge users. The results showed that implementing MRC processing at the receivers and frequency planning at cellular level can improve the SINR for users located far from the BS and reduce interference significantly.

This chapter explores the main advantages of DAS over conventional cellular systems in terms of frequency reuse factor. The achieved results shed light on how spatial diversity and macro diversity affect the transmission and reception schemes. A comparative analysis of the different scenarios is provided using outage probability and throughput as the main performance metrics. Numerical results show that DAS overcomes the spectral efficiency of conventional cellular systems and increases throughput figures in deployments where the frequency reuse factor is fixed. Further improvements are achieved when MRC is implemented at the user terminal considering that number of receive antennas is $N_r \geq 2$. An interesting result indicates that in some cases DAS achieves performance levels that conventional systems cannot attain even if its frequency reuse factor is larger that the one used by DAS.

3.2 Frequency Planning

Frequency planning is the distribution of the radio resources over a given service area in cellular systems. This radio resource distribution is essential to achieve high Quality of Service (QoS), efficient power management, ICI suppression, and efficient user scheduling. Its main objective is to maximize the spectral efficiency and its characteristics define the power allocation and scheduling schemes in the access network. The assignment of the radio resources is a key component in the design of cellular systems and can be implemented in two

different approaches [87]: Fixed Channel Allocation (FCA) and Dynamic Channel Allocation (DCA).

3.2.1 Fixed Channel Allocation (FCA)

Under this scheme, a set of radio resources (channels) is permanently allocated in a cellular system. If there is a traffic fluctuation, this scheme may not maintain a minimum required QoS which might yield a capacity degradation. Due to the fact that in cellular systems the traffic can be non-uniform with temporal and spatial fluctuations, a fixed and uniform channel allocation may result in high blocking probability in some cells while others might have a large number of available channels. A frequency planning technique estimates the expected traffic of a set of cells and allocate more radio resources to the cells that will serve more users, i.e., it implements an asymmetric fixed frequency assignment. This approach has been implemented allocating channels in such a way that the average blocking probability in the system is minimized [88].

One approach of the radio resource allocation is the Fractional Frequency Planning (FFP). This scheme is based on the concept of reuse partitioning [89] where the users with the highest signal quality use a lower reuse factor while users with low SINR use a higher reuse factor to mitigate ICI. The most common implementation of FFP in a network is a blend of frequency reuse factors: universal reuse in the cell center ($K_F = 1$) and less intensive frequency reuse at the cell-edge areas ($K_F > 1$). In most of these schemes, high power is allocated to the resources used for cell-edge users in order to guarantee coverage and minimum QoS requirements. Fig. 3.1 shows the radio resources distribution in a deployment with FFP. The radio resources are divided into four frequency segments. The first segment is used to serve users with high SINR at the center area of each cell and the rest of the segments are used considering $K_F = 3$. FFP is a particularly effective strategy to mitigate ICI for the cell-edge users and there exist a large number of variations of it, which employ similar reuse partitioning in sectorized cellular systems [84, §5].

In [90, 91] the authors considered a reuse partitioning scheme as the one shown in Fig. 3.1. The objective of the authors was to classify the Physical Resource Blocks (PRB) with $K_F = 1$ in the center of the cells and $K_F = 3$ at the cell-edge. In this way, more power is allocated to the PRB with $K_F = 3$ increasing the throughput of the users at the cell-edge in OFDMA - LTE systems. A variation of FFP is proposed in [92] where higher power allocation is performed in 2 of the 3 frequency blocks used in the deployment. By assigning two sub-bands to the cell-edge area, the spectral efficiency is improved when two users in the cell-edge of adjacent cells use the same channels but with different data rates.

A similar approach to FFP is the soft frequency reuse (SFR). For a SFR with $K_F = 3$, the total frequency resource is divided into three segments (f_1,f_2,f_3) as shown in Fig. 3.2. In order to reduce ICI generated by the use of intensive frequency reused in the center of the cell, less power is allocated for such radio resources. Each block of radio resources used in the cell-edge areas has higher power levels in order to improve the achievable SINR. Lu et al. [93] implemented a SFR scheme for an OFDMA system that allows the optimization of the radio resource allocation. Different cases of resource management were considered along with SFR: 1) when all the resources of a cell-edge are occupied and there are available resources in the central area; 2) when resources at the central area are not enough to provide a service and there are free resources in the cell-edge area. The proposed solution performs a resource borrowing mechanism that manage radio resources in the BS and is implemented

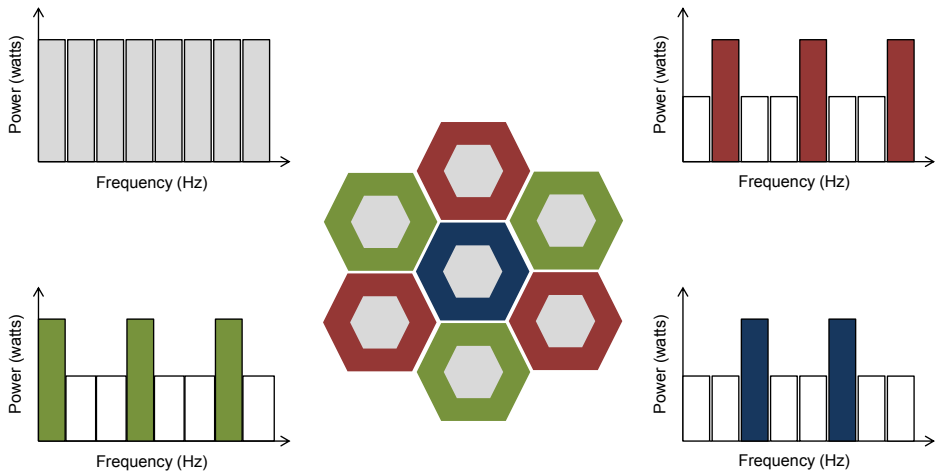

Figure 3.1: Fractional Frequency Planning with conventional cellular deployment

in a sectorized deployment. The work in [83] is a reference for SFR and its integration with the scheduling process. The authors proposed a proportional fairness scheduling (PFS) [94] where the power allocation depends on the frequency reuse. The numerical results show that improving the performance of the cell-edge users may decrease the spectral efficiency of users with better channel conditions. This means that a fair trade-off between center and edge cell achievable data rates must be considered when radio planning designs are implemented in cellular systems.

For networks based on OFDMA such as LTE, the management of radio resources implies also the control of power and data rates. Authors in [95] implemented a different approach for SFR considering a more flexible way to classify the users within the cell. Users are classified according to their required data rates and the number of scheduled users is larger if modulation and coding schemes (MCS) are used, which increases the overall spectral efficiency. In this scenario, the total number of resources are classified in sub-bands that can provide better throughput to the users in both cell zones according to the proposed SFR scheme. Numerical results show that for some scenarios SFR provides a better interference control and a more effective MCS selection compared to universal frequency reuse, which increases the average throughput without degrading the performance at the cell-edge.

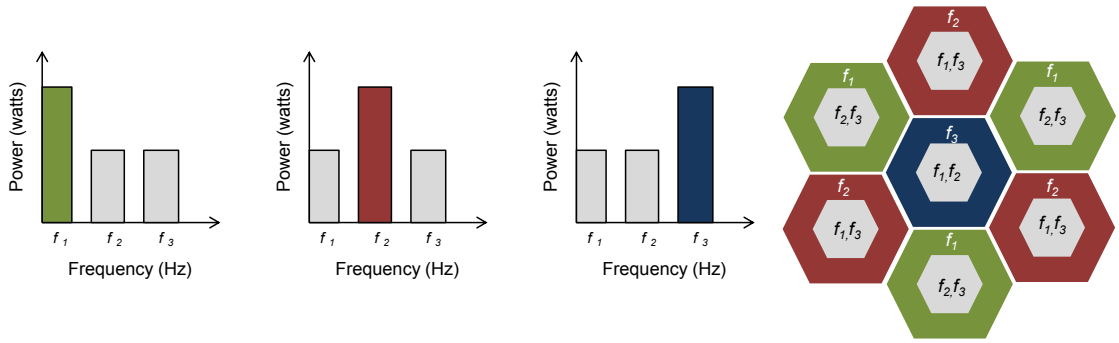

Figure 3.2: Soft Frequency Planning with boundary frequency reuse $K_F = 3$

3.2.2 Dynamic Channel Allocation (DCA)

Under this scheme a set of channels is allocated in a cell temporarily depending on the traffic load or other system metrics. This scheme allows the network to utilize radio resources more efficiently based on certain parameters such as channel or traffic load statistics, improving QoS and optimizing resource allocation. Radio resources are kept in a central pool and are assigned dynamically to cells as new users require service. As soon as the user service is completed, its radio resources return to the central pool. The channels are assigned according to some interference constrains, if more than one channel can be used in a cell, a CU evaluates the cost of using each channel and assigns the one with the minimum cost or higher profit to the system. Each DCA scheme differentiates from each other according to its cost function which may consider the blocking probability, traffic measurements, channel statistics, etc.

In [96] and references therein, the achievable SINR and load information are used to implement DCA scheme which was found to be more efficient in terms of throughput and blocking probability w.r.t. FCA. The DCA scheme accepts a service requirement whenever a channel can be found that satisfies the minimum QoS constraints such as SINR. Numerical results show that DCA outperforms FCA for deployments where there is a non-uniform load among the cells and they suggest that hybrid approaches can be used to reduce the complexity of DCA. If two or more network architectures have to share radio resources, DCA can be also implemented. For instance, in [97] macro-cell and femto-cell OFDMA systems are coordinated and operate in the same bandwidth using adaptive fraction frequency reuse according to a specific access protocol. The DCA schemes can be implemented using distributed or centralized architectures in the access network. In the centralized architecture DCA requires a large amount of signalling [28] and communication between BSs leads to large system latencies [17], while in the distributed approach resource allocation decisions are made by each cell based on the local available information. The distributed approach yields a trade-off between performance and signaling overhead [98], whilst in the centralized approach performs a more accurate signal processing and coordination at a higher complexity price.

3.3 Cellular Architecture and DAS

3.3.1 System Model

Consider the distributed antenna system depicted in Fig. 3.3. Each hexagonal cell has a radius r and consists of a total of $N+1$ radiating nodes (or distributed antennas): one node located at the center of the cell ($n = 0$), and N distributed nodes ($n = 1, \ldots, N$) located at a fixed distance D_r from the center of the cell and spaced at uniform angles given by $\theta_n = \frac{2(n-1)\pi}{N}$. A conventional cellular (CAS) system can be characterized by substituting $N = 0$ in the expressions presented in this section. Two transmission modes will be studied in this chapter. The first one is called blanket transmission with repetition coding, where all the nodes within a cell transmit the same information towards a single user (see Fig. 3.4b). Each node uses a transmit power $\frac{P}{N+1}$, where P is the total transmit power in the cell. In the second transmission mode, called single transmission scheme, only a single antenna is selected for transmission by the criterion of minimizing propagation path-loss (see Fig. 3.4c). The selected node uses full power P while the remaining nodes are deactivated. Each user terminal is equipped with N_r antennas followed by MRC processing. In order to combat ICI, different frequency reuse factors will be evaluated considering that there is only one single user

transmission in any given time/frequency/code dimension. In a hexagonal cellular network, the available frequency reuse factors required to have a symmetric deployment are given by the formula $K_F = (a+b)^2 - ab$, where a and b are two positive integers [85]. For simplicity, only the central cell ($i = 0$) will be used for performance metric calculation and only one tier of 6 outer cells ($i = 1, \ldots, 6$) will be considered as source of interference adopting the system model studied in [78, 99]. All these interfering cells are located at a reuse distance from the central cell given by $D_c = r\sqrt{3K_F}$ [85], and spaced at uniform angles given by $\phi_i = \frac{(i-1)\pi}{3}$. The distance between the nth node in the ith cell of the network and the user of analysis u in the central cell with coordinates (x_u, y_u) is denoted by $d_{i,n}^{(u)}$ and is given by the following expression:

$$d_{i,n}^{(u)} = \sqrt{\Delta_{x_u}(i,n)^2 + \Delta_{y_u}(i,n)^2}, \tag{3.1}$$

where

$$\Delta_{x_u}(i,n) = \delta(i) D_c \cos\phi_i + \delta(n) D_r \cos\theta_n - x_u, \tag{3.2}$$

$$\Delta_{y_u}(i,n) = \delta(i) D_c \sin\phi_i + \delta(n) D_r \sin\theta_n - y_u, \tag{3.3}$$

and $\delta(k)$ is a binary variable which takes a value $\delta(k) = 0$ when $k = 0$, and $\delta(k) = 1$ when $k \neq 0$.

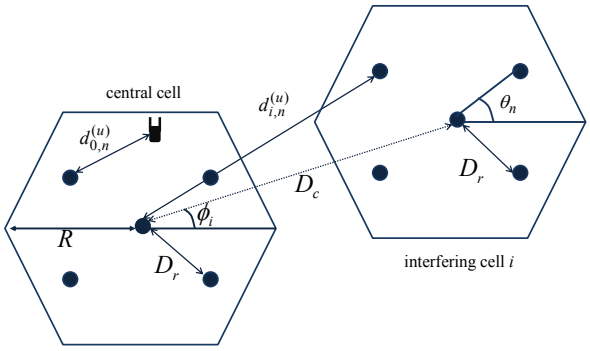

Figure 3.3: Cellular Architecture with DAS and MRC receivers

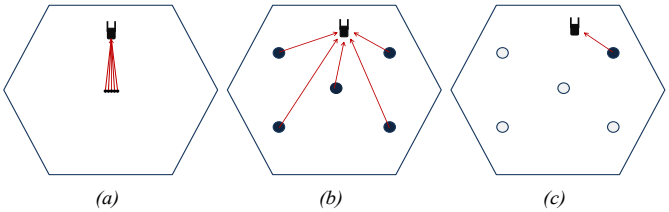

Figure 3.4: Transmission Schemes: a) conventional cellular system, b) blanket transmission where all nodes transmit to one user, c) single transmission where the user is served by the best node.

3.3.2 Propagation and signal models

The channel between the mth antenna of user u in the central cell and the nth node of the ith cell of the network is denoted by $h_{m,i,n}^{(u)}$. Channels of different users and different

distributed antennas are assumed to be statistically independent, Rayleigh distributed, and affected by a propagation path-loss component defined by $L_{i,n}^{(u)} = \left(d_{i,n}^{(u)}\right)^{-\alpha_{pl}}$, where α_{pl} is the path-loss exponent. The channel $h_{m,i,n}^{(u)}$ can be written as $h_{m,i,n}^{(u)} = \sqrt{\frac{P}{N+1} L_{i,n}^{(u)}}\psi_m$, for the blanket transmission scheme, or $h_{m,i,n}^{(u)} = \sqrt{P L_{i,n}^{(u)}}\psi_m$, for the single transmission scheme, where ψ_m is a circularly complex and zero-mean Gaussian random variable with unitary power $\psi_m \sim \mathcal{CN}(0,1)$.

The signal transmitted by the ith cell, denoted by s_i, is assumed to has unitary power $\mathbb{E}[|s_i|^2] = 1$ and the signal received by the mth antenna of user u under blanket transmission is given by:

$$r_{u,m} = \sum_{n=0}^{N} h_{m,0,n}^{(u)} s_0 + \sum_{i=1}^{6}\sum_{n=0}^{N} h_{m,i,n}^{(u)} s_i + \nu_{u,m}, \tag{3.4}$$

while for the single transmission scheme it can be written as

$$r_{u,m} = h_{m,0,z}^{(u)} s_0 + \sum_{i=1}^{6} h_{m,i,q_i}^{(u)} s_i + \nu_{u,m} \tag{3.5}$$

where $z = \arg\max_{n\in\{0,1,\ldots,N\}}\{L_{0,1}^{(u)}, L_{0,2}^{(u)}, \ldots, L_{0,N}^{(u)}\}$ and q_i is an integer randomly selected among $\{0,1,\ldots,N\}$ assuming that the same transmission scheme is used in all the cells. This assumption is used in several works in the literature of distributed antenna systems, e.g., [76, 78, 99]. The term $\nu_{u,m}$ in (3.4) and (3.5) is the noise with normalized power, $\nu_{u,m} \sim \mathcal{CN}(0,1)$. Expressions in (3.4) and (3.5) can be rewritten as

$$r_{u,m} = S_{u,m} + I_{u,m}, \tag{3.6}$$

where $S_{u,m}$ stands for all the contributions of the nodes inside the central cell, and $I_{u,m}$ represents the inter-cell-interference-plus-noise term. Due to the central limit theorem and according to the theoretical work presented in [78, 99], the interference-plus-noise term $I_{u,m}$ can be considered as Gaussian distributed with variance

$$\sigma_{I_{u,m}}^2 = \sum_{i=1}^{6}\sum_{n=0}^{N} \mathbb{E}[|h_{m,i,n}^{(u)}|^2] + 1 \tag{3.7}$$

for the blanket transmission scheme, and

$$\sigma_{I_{u,m}}^2 = \sum_{i=1}^{6} \mathbb{E}[|h_{m,i,q}^{(u)}|^2] + 1 \tag{3.8}$$

for the single transmission scheme, respectively.

Since all channels are Rayleigh-distributed and the interference-plus-noise term is assumed to have Gaussian distribution[1], then the SINR at each antenna element is methematically defined as:

$$\gamma_{u,m} = \frac{|S_{u,m}|^2}{\sigma_{I_{u,m}}^2} \tag{3.9}$$

[1]This means that the term $\sigma_{I_{u,m}}^2$ is seen as a constant for user u and antenna m

and it can be proved that $\gamma_{u,m}$ has exponential distribution [100]:

$$f_{\gamma_{u,m}}(\gamma_{u,m}) = \frac{1}{\overline{\gamma}_{u,m}} e^{-\frac{\gamma_{u,m}}{\overline{\gamma}_{u,m}}} \tag{3.10}$$

where $\overline{\gamma}_{u,m}$ is the average SINR. Since the denominator of the SINR (3.9) is assumed to be a constant, the average SINR can be calculated by applying the expectation operator to the numerator only while using in the denominator the variance of the Gaussian-distributed interference-plus-noise term. This assumption has been used in research works dealing with the analysis of multicell systems, e.g., [100]. The expression of the average SINR for the blanket transmission scheme is given by:

$$\begin{aligned} \overline{\gamma}_{u,m} &= \frac{\mathbb{E}[|S_{u,m}|^2]}{\sigma^2_{I_{u,m}}} \\ &= \frac{\mathbb{E}[|\sum_{n=0}^{N} h^{(u)}_{m,0,n}|^2]}{\sum_{i=1}^{6}\sum_{n=0}^{N}\mathbb{E}[|h^{(u)}_{m,i,n}|]^2 + 1} \\ &= \frac{\sum_{n=0}^{N} \frac{P}{N+1}\left(d^{(u)}_{0,n}\right)^{-\alpha_{pl}}}{\sum_{i=1}^{6}\sum_{n=0}^{N} \frac{P}{N+1}\left(d^{(u)}_{i,n}\right)^{-\alpha_{pl}} + 1} \end{aligned} \tag{3.11}$$

and for the single transmission scheme by:

$$\overline{\gamma}_{u,m} = \frac{\left(d^{(u)}_{0,z}\right)^{-\alpha_{pl}}}{\sum_{i=1}^{6}\left(d^{(u)}_{i,q}\right)^{-\alpha_{pl}} + \frac{1}{P}} \tag{3.12}$$

Since (3.11) and (3.12) are independent of m then the term $\overline{\gamma}_{u,m}$ will be simply denoted by $\overline{\gamma}_u$.

3.4 Performance Metrics

3.4.1 Outage probability

For the uth user in a particular location in the central cell, the outage probability is defined as the probability that the instantaneous SINR, γ_u, is below a required target value γ_T. The outage probability is a metric used to evaluate system coverage performance and can be written in mathematical form as:

$$P^{(u)}_{out} = \Pr\{\gamma_u \leq \gamma_T\}, \tag{3.13}$$

which is equivalent to the cumulative density function (CDF) of the SINR of user u: $F_{\gamma_u}(\gamma_T)$. Using the signal model in the previous section, $P^{(u)}_{out}$ can be calculated in closed-form for a single antenna receiver using the CDF of the exponential distribution [101]:

$$P^{(u)}_{out} = 1 - e^{-\frac{\gamma_T}{\overline{\gamma}_u}}. \tag{3.14}$$

Recall that in this section MRC receivers are considered whose main characteristic is that their output SINR are the sum of the SINRs of the individual antennas, $\gamma_u = \sum_m \gamma_{u,m}$. Assuming no correlation between the receive antennas, N_r exponentially identical and independently distributed random variables are added, which yields a central Chi-square distributed variable with $2N_r$ degrees of freedom whose CDF is given by the following expression [101]:

$$P_{out}^{(u)} = 1 - e^{-\frac{\gamma_T}{\overline{\gamma}_u}} \sum_{k=0}^{N_r-1} \frac{1}{k!} \left(\frac{\gamma_T}{\overline{\gamma}_u} \right)^k \tag{3.15}$$

With the help of (3.15), it is possible to calculate the average outage probability across the cell by integrating the previous expression across the set of all user positions in the cell, denoted by $\mathcal{U}_0$, as follows:

$$P_{out} = \int_{u \in \mathcal{U}_0} P_{out}^{(u)} \Pr\{u\} du, \tag{3.16}$$

where $\Pr\{u\}$ is the probability of occurrence of the user location u. Assuming uniform user distribution across the central cell (all user positions are equally likely) allows to evaluate numerically the above integral as an averaging operation:

$$P_{out} = \frac{1}{N_u} \sum_{u=1}^{N_u} P_{out}^{(u)}, \tag{3.17}$$

where N_u is the number of user positions considered in the numerical evaluation.

3.4.2 Throughput

The throughput of a system can be defined as the long term ratio of the total number of bits correctly received by the users to the total time spent to transmit such information. In systems with fixed transmission intervals, the throughput can be expressed as the ratio of $\mathbb{E}[I_b]$ the average number of bits correctly transmitted per transmission time interval (TTI), divided by the duration of the TTI denoted by L_T:

$$T = \frac{\mathbb{E}[I_b]}{L_T}. \tag{3.18}$$

Since the transmission of information bits in a real network commonly takes place in blocks or packets, the average number of correctly transmitted bits per TTI can be expressed as the product of the correct packet or block reception probability, times the number of bits transported by each packet B_p:

$$\mathbb{E}[I_b] = B_p \Pr\{\text{correct packet reception}\}, \tag{3.19}$$

and the probability of correct packet reception is given by:

$$\Pr\{\text{correct packet reception}\} = \Pr\{t = 1\}, \tag{3.20}$$

where t is a binary random variable with value 0 if the packet is incorrectly received, and $t = 1$ if it is correctly received. The conditional probability of correct packet reception at a particular user location can be approximated by the probability that the SINR of the uth

user achieves a given threshold $\gamma_{T_{mcs}}$. The SINR target $\gamma_{T_{mcs}}$ is associated to a particular modulation and coding scheme being used in the current channel instance, so that:

$$\Pr\{t = 1|u\} = \Pr\{\gamma_u > \gamma_{T_{mcs}}\}. \tag{3.21}$$

This SINR reception model considers that all bits in a packet have been correctly received if the SINR exceeds the threshold $\gamma_{T_{mcs}}$, which is an optimistic assumption. Nevertheless, system level simulations confirm that this SINR packet reception model provides a good approximation to the real system performance using look-up tables (LUTs). It is worth mentioning that this assumption has been extensively used in literature, e.g. [89, 102, 103], and it is used here to simplify the analysis without sacrificing the accuracy of the final results.

The SINR thresholds can be obtained from standard MCS tables of different wireless technologies. In the following, LUTs of WiMAX are used to evaluate the system performance. Table 3.1 shows an example of MCSs used in WiMAX systems [104]. The MCS level 0 corresponds to a SINR lower than the minimum threshold, therefore no service can be provided. Observe that the analysis can be applied to other MCS sets for different wireless technologies.

The global throughput for a single fixed MCS can be expressed as:

$$T = \int_{u \in \mathcal{U}_0} \frac{B_s R_c N_p}{L_T} (1 - F_{\gamma_u}(\gamma_T)) \Pr\{u\} du \tag{3.22}$$

where B_s is the number of bits per symbol, R_c is the coding rate, and N_p is the number of symbols per block. The throughput expression to systems with adaptive MCS is straightforward and is given by:

$$T = \int_{u \in \mathcal{U}_0} \sum_{mcs=0}^{M_{mcs}} \frac{B_s^{(mcs)} R_c^{(mcs)} N_p}{L_T} \times \left(F_{\gamma_u}(\gamma_{T_{mcs+1}}) - F_{\gamma_u}(\gamma_{T_{mcs}}) \right) \Pr\{u\} du, \tag{3.23}$$

where M_{mcs} is the number of available MCS.

Table 3.1: WiMAX modulation and coding schemes

MCS level	BLER	AMC (R_c)	γ_T dB	Bits (B_p)
1	4.10e-3	QPSK - 1/3	-1.14	2
2	4.12e-3	QPSK - 1/2	1.32	2
3	7.15e-3	16QAM - 1/3	6.52	4
4	3.30e-3	16QAM - 4/5	11.67	4

3.5 Numerical Evaluation

This section presents the results of the numerical evaluation of the analytical expressions derived above. The results were obtained by using 10e3 user positions within the cell of analysis. The distributed nodes are located at a distance $D_r = \frac{2}{3}r$ in order to have a uniform distribution across the cell. The number of radiating nodes simulated is $N = 4$ with path-loss exponent $\alpha_{pl} = 4$. The throughput evaluation was performed assuming the following transmission parameters of WiMAX: one single block of $N_p = 7200$ symbols is transmitted to one single user per TTI with a frame duration of $L_T = 5$ (ms).

The total transmit power per cell P used in (3.11) and (3.12), is obtained by defining a reference average SINR value ($SINR_{ref}$) that must be achieved at the edge of the cell of a conventional system without considering interference:

$$\frac{P}{\sigma^2_{I_{u_{ref}}}} = \frac{10^{SINR_{ref}/10}}{r^{\alpha_{pl}}} \tag{3.24}$$

where $\sigma^2_{I_{u_{ref}}} = 1$ is the variance of the interference-plus-noise term of a reference user located at the edge of the cell.

3.5.1 Outage Probability

Fig. 3.5 shows the average outage probability across the cell P_{out} in (3.17) as a function of $SINR_{ref}$ defined in (3.24). One antenna is considered at the receiver ($N_r = 1$) for both conventional and DAS systems. The results are obtained by using the threshold γ_T for 16-QAM modulation, which corresponds to MCS level 4 in Table 3.1. The DAS transmission modes are denoted as DAS_b for blanket transmission, and DAS_s for single transmission. It can be observed that for $SINR_{ref} < 9$(dB), DAS_s with frequency reuse factor $K_F = 1$ attains a smaller outage probability than conventional systems with $K_F = 7$. This means that DAS_s can improve spectrum usage by reducing the required reuse factor to provide a given service. Consider a target value of $P_{out} = 10^{-1}$ in Fig. 3.5. To achieve the target probability, DAS_s with $K_F = 3$ requires $SINR_{ref} = 5$(dB), whilst DAS_b with $K_F = 3$ requires $SINR_{ref} = 20$(dB). The conventional cellular system with frequency reuse factor of $K_F = 7$ achieves the target probability with $SINR_{ref} = 18$(dB) while in all other cases of K_F and $SINR_{ref}$ it fails to reach the target P_{out}. In the particular case of DAS_s with $K_F = 7$, it can be observed that DAS can reach performance levels that conventional architectures cannot achieve, independently of the implemented K_F. This result demonstrates the effectiveness of DAS to improve spectral efficiency of cellular systems with intensive frequency reuse.

Fig. 3.6 shows the average outage probability across the cell P_{out} as a function of $SINR_{ref}$ considering MRC receivers with two antennas ($N_r = 2$). In this case, DAS_s system with $K_F = 1$ achieves a better outage probability than the conventional system for values of $SINR_{ref} < 6$(dB), even if the frequency reuse factor K_F of the conventional system is as large as 21. The improvement in outage probability by implementing MRC at the receiver is more significant for large values of $SINR_{ref}$ w.r.t. results in Fig. 3.5.

3.5.2 Throughput

Fig. 3.7 shows the average throughput T as a function of the reference $SINR_{ref}$ considering different values of the frequency reuse factor K_F and MRC receivers with $N_r = 2$. In order to achieve a target throughput of 4(Mb/s) for $K_F = 3$, DAS_s and DAS_b require a $SINR_{ref}$ of -8(dB) and -2(dB), respectively. In contrast, the conventional system requires $SINR_{ref} > 5$(dB). This implies that DAS can achieve the same throughput figures of the conventional system using lower values of transmit power. For a fixed $SINR_{ref} = -1$(dB) and $K_F = 3$, the average throughput achieved by DAS_s and DAS_b surpasses the average throughput of the conventional system by 150% and 137% respectively. Fig. 3.8 shows the CDF of the user throughput assuming $SINR_{ref} = 0$(dB) and $N_r = 2$ at the receiver. If the minimum required throughput is 2(Mb/s), DAS_s with $K_F = 1$ achieves higher throughput

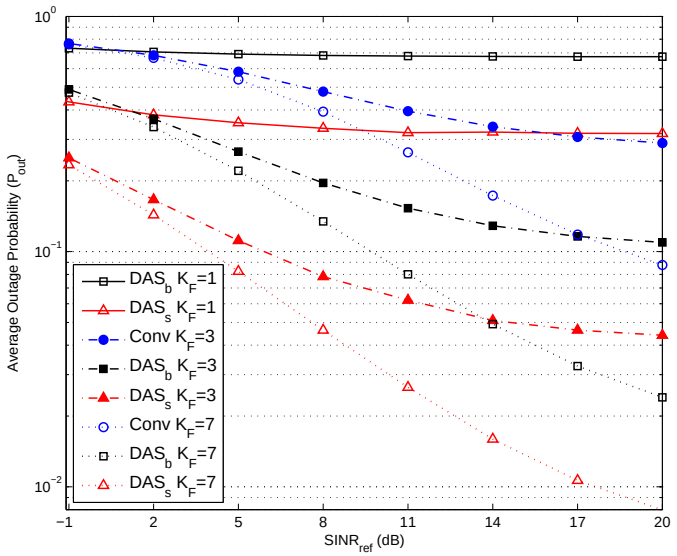

Figure 3.5: Outage Probability P_{out} vs. $SINR_{ref}$ of conventional cellular and distributed systems (blanket DAS_b and single antenna transmission DAS_s) with $N_r = 1$ and different values of K_F.

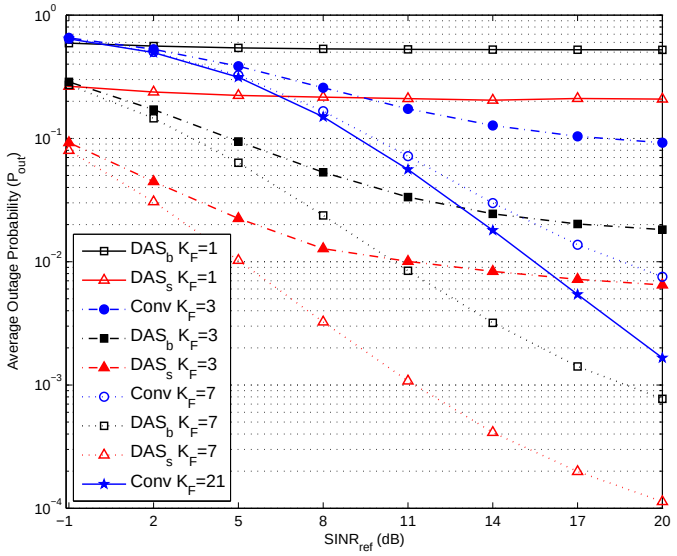

Figure 3.6: Outage Probability P_{out} vs. $SINR_{ref}$ of conventional cellular and DAS (blanket DAS_b and single antenna transmission DAS_s) considering an MRC receiver $N_r = 2$ with uncorrelated channels and different values of K_F.

gains w.r.t. the conventional system with $K_F = 3$ and $K_F = 7$. If the frequency reuse is fixed in both systems for $K_F = 3$, the minimum average throughput attained in DAS_s and DAS_b is 2.9(Mb/s) and 2.4(Mb/s) whilst the conventional system achieves at most 1.3(Mb/s).

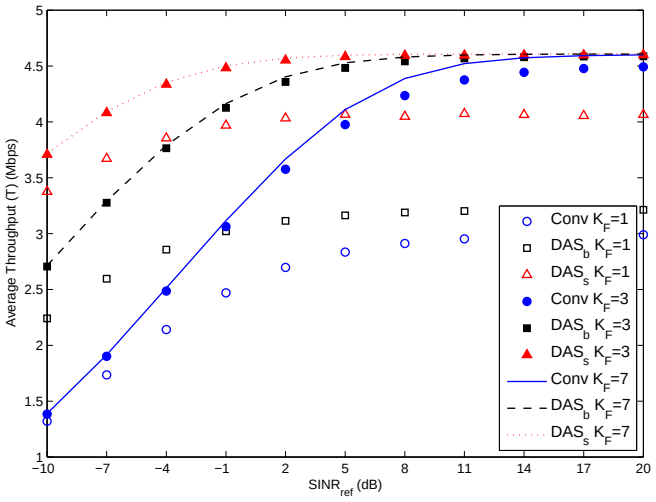

Figure 3.7: Average throughput as a function of $SINR_{ref}$ for conventional and DAS (blanket DAS_b and single DAS_s transmission schemes) using different values of K_F and $N_r = 2$ receiver.

Fig. 3.9 shows the average throughput as a function of the number of antennas N_r at the MRC receiver. The numerical evaluation considered more than 4 antennas at the receiver. Although this configuration is hard to implement in practice due to the characteristics of the user terminals, it illustrates the asymptotical behavior of the achievable throughput as the number of antennas at the receiver increases. The average throughput achieved by DAS_b for $K_F = 1$ with $N_r \in \{1, 2\}$ is similar to the one achieved by conventional system with $K_F = 3$. DAS_s achieves the highest average throughput, specially when the number of antennas at the MRC receiver is $N_r < 4$.

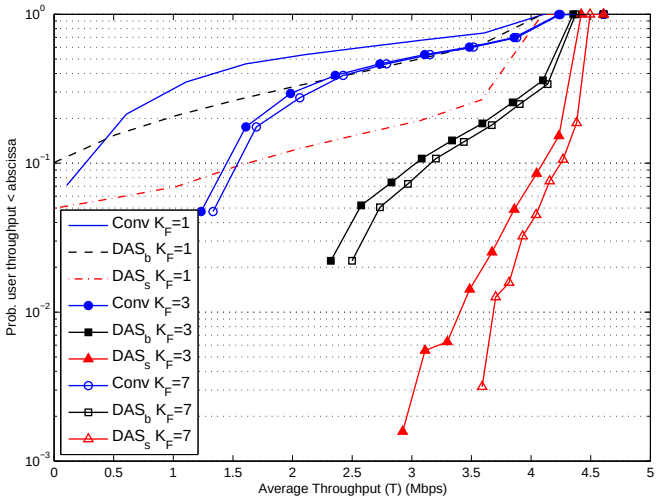

Figure 3.8: CDF of the throughput for conventional and DAS (blanket DAS_b and single DAS_s transmission schemes) for different K_F and $N_r = 2$ receiver.

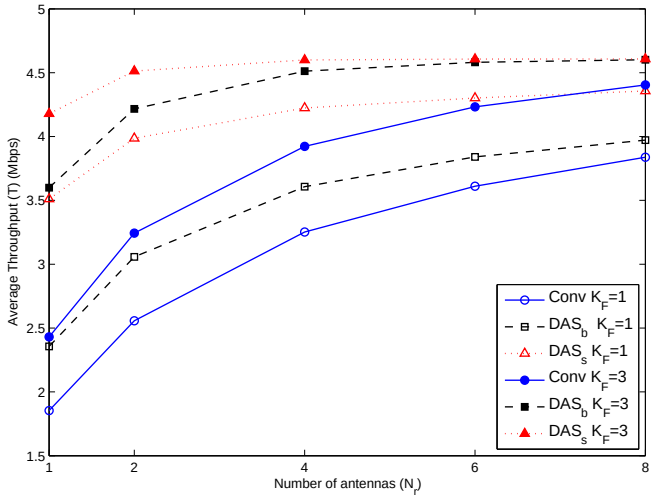

Figure 3.9: Average throughput as a function of N_r for conventional and DAS (blanket DAS_b and single DAS_s transmission schemes) with $SINR_{ref} = 0$(dB) and $K_F \in \{1, 3\}$.

3.6 Conclusions

This chapter presented an analysis of the achievable gains of distributed antenna systems w.r.t. conventional cellular systems, as well as the impact of frequency reuse planning in the achievable outage probability and average throughput. Both cellular systems were enhanced by MRC processing at the mobile terminals considering different values of N_r. Numerical results showed that the distributed antenna architecture can be used to increase the average throughput w.r.t. conventional cellular systems by more than 180% and 220% for blanket (DAS_b) and single transmission scheme (DAS_s), respectively. DAS_s has shown to be the best transmission scheme since it improves P_{out} and T figures with less power consumption and smaller K_F compared to the conventional system. The average throughput can be further improved by MRC processing at the receiver for both access architectures. The distributed antennas architecture can be used to reduce the frequency reuse which enhances the spectrum efficiency and increases the coverage area or cell size with the same power budget. The results show that in some particular cases DASs outperform conventional cellular systems regardless the frequency reuse factor used by the co-located antenna system.

Chapter 4

User Selection and Rate Allocation in Interference Channels

This chapter addresses the sum rate maximization problem for the downlink of a wireless network where multiple transmitter-receiver pairs (links) share the same medium and thus potentially interfere with each other. The rate maximization can be decoupled in two optimization problems: the first one is to find the set of links that can be jointly scheduled [C2, C3], and the other one is to allocate the set of MCSs that maximize the sum rate [J2, C4]. Since the combination of users and available MCSs form a finite set, the rate maximization problem has a combinatorial nature for which sub-optimal yet efficient algorithms are designed so that an acceptable solution to the combinatorial problem can be provided. The user selection problem is reformulated as a matching problem and an efficient algorithm for user-antenna matching is designed. The power and rate allocation problem is tackled by different approaches: casting the allocation problem into an eigenvalue optimization problem, assigning MCSs based on power consumption, and performing a heuristic low-complexity MCS assignment. Numerical results show that the proposed algorithms achieve a good trade-off between sum rate performance and computational complexity.

4.1 Introduction

THE EFFICIENT RESOURCE ALLOCATION in wireless networks is fundamental to fulfill several practical QoS measures such as data rate or outage probability. The number of wireless users and data services has increased dramatically over the last few years and optimization of the resource allocation has become fundamental to guarantee both users and operators satisfaction without increasing system requirements (bandwidth and power budget). Moreover, there are many common scenarios where the set of transmitter-receiver pairs (links) operate simultaneously in a shared medium and interference mitigation techniques must be employed. The QoS is measured in practice by the SINR, and recent works on resource allocation optimization for interference coupled networks [36, 37, 70, 105–109] show the relation between the SINR maximization and the efficient power and rate allocation. Bear in mind that the achievable data rate depends on the SINR which is a global function of all transmit powers in systems where signal processing schemes such as beamforming or signal design are fixed or not implemented. Efficient power control schemes are fundamental to maximize either a global system utility [70, 107–109] or individual rates [32, 36, 105] in networks where interference cannot be eliminated. In multiuser systems with less antennas at the transmitter than the number of competing single antenna users, the system performance

measured by the total sum rate can be optimized by solving two problems sequentially:

1. The selection of a set of users (scheduling) that has the potential to maximize the total sum rate, i.e., finding the optimal transmitter-receiver pairs (links).

2. The efficient rate and power allocation for such a set of users to take into account a given set of system constraints.

In *interference coupled* systems the successful scheduling of a set of interfering links is conditioned to the fulfillment of all individual QoS requirements and power constraints. An efficient scheduling policy must determine the proper set of links for which there exists a power and rate allocation that meets power constraints and SINR requirements. However, finding the set of links that maximizes a given system metric (e.g., total sum rate or fairness) is a combinatorial NP-complete problem [110]. For the particular scenario where the number of users is less than or equal to the number of antennas at the transmitter, a suboptimal solution to the scheduling problem is found in the so called user-removal techniques [38, 110, 111] where the users that violate the power constraints or QoS requirements are found iteratively and temporary dropped.

The works related to *resource allocation* optimization and the ones about *user-removal* have different objectives and both fields have been studied independently. The former assumes that for a given set of links exists an infinite number of solutions to the resource allocation problem and the main objective is to find the allocation that maximizes a specific utility function such as sum rate, power consumption, etc. The latter is concerned about the set of links that can be scheduled whose QoS requirements are fixed and the resource allocation can be achieved by conventional allocation schemes. There is a large number of open issues that must be solved in interference coupled networks and the objective of the work presented in this chapter is two-fold: to provide a solution to the scheduling problem and simultaneously maximize the total sum rate by performing efficient resource allocation. Unlike the available literature, the work presented in this chapter considers that the allowed SINR ranges are constrained to take values from a finite set of thresholds or targets associated with a given set of MCSs. By considering different MCSs the philosophy of user-removal is extended for the case of non-fixed SINR targets and a methodology to find the MCS allocation that maximizes the total sum rate is designed.

4.1.1 Related Works

Over the last 20 years several theoretical [36–38, 69, 112–114] and practical [105, 107, 111, 115] works have been developed to understand and solve the problems of power allocation and utility maximization for cellular, multi-hop, peer-to-peer, satellite, and DSL communications networks. Early works on power control [112, 115–118] designed iterative algorithms under the standard interference function framework [36, 112] in order to guarantee the convergence of the algorithms to a unique and optimal power allocation solution. These works assume that the set of given links is always feasible, i.e., for such set there always exists a solution to the power allocation problem. More recent works [36, 107, 113] studied the relationship between the rate allocation and power control, assuming a high SINR regime. This is a common assumption because for interference coupled systems the mathematical modeling of the resource allocation problem is more tractable [37] and efficient iterative algorithms can be developed. For instance, in [113] the power control problem for rate maximization is

formulated as a convex problem and its solution is found via geometric programming (GP) for wireless networks.

Related works for interference coupled wired (DSL) networks [108, 109, 119] show that suboptimal yet efficient power allocation for rate maximization can be achieved when the non-convex rate maximization problem is approximated by alternative convex objective functions of the transmit powers. In [109] the authors show that relaxed forms of the objective function of the rate maximization problem lead to the convergence of the proposed algorithms and the accuracy of such approximation depends on the convexity properties of the objective functions. In [119] the weighted sum rate maximization problem was extended to the multiuser multi-carrier scenario in interference coupled systems. The idea behind the algorithms presented in [108, 119] is to solve iteratively the original resource allocation problem by optimally solving in each iteration, a relaxed version of the original problem. In each iteration, the local optimal solution bounds the solution of the original problem for a given resource allocation and due to the properties of the relaxed objective function, the convergence to a local optimum is guaranteed. This is known as successive convex approximation (SCA) whose goal is to refine the solution found for the relaxed problem in order to close the gap between the approximated and the optimal resource allocation. A framework to solved general optimization problems under SCA and a comprehensive analysis of state-of-the-art dynamic spectrum management algorithms is presented in [114]. The work in [119] solved the power spectrum management problem implementing a SCA algorithm that exploits the characteristics of the feasible region of resource allocation solutions and in each iteration GP is used to solved the local power assignment problem. The works [37, 70] presented algorithms that solve optimally the joint power and rate allocation problem by exploiting the convex characteristics of the feasible rate region and generalized the resource allocation problem for both high and low SINR regimes.

The mathematical abstraction or representation of an interference coupled network has a strong connection to the theory of irreducible matrices [120, 121] and the Perron-Frobenius theorem [37, 120]. Such mathematical principles are the foundation of several works (e.g., [36–38, 70, 105, 113]) that characterize the feasible rate region and solve optimally the resource allocation problem. The theoretical results derived from the the Perron-Frobenius theorem [37, 38, 70] allow to render the network optimization problem into an eigenvalue problem and to verify the feasibility of both, the set of scheduled links and its corresponding resource allocation. This mathematical tool has been used to solve the rate and power allocation problem for a fixed set of links (e.g., [37, 70, 107]) assuming that the rates and powers can take any real positive value. The main objective of such works is to find the optimum allocation that maximizes a given utility function under a set of power constraints and transmission schemes. The user-removal techniques [38] exploit the Perron-Frobenius theory to identify the infeasibility of a set of links in scenarios where for a given set of power constraints and rate (SINR) requirements, the classical power control algorithms (e.g., [115]) do not converge.

4.1.2 Contributions

The work presented in this chapter is focused on the sum rate maximization problem in multiuser systems considering multiple antennas at the transmitter and that each selected user has individual rate requirements. In Section 4.2 a centralized DAS scenario is considered and an algorithm to define close-to-optimal transmitter-receiver pairs (links) is designed. The problem of assigning one user per distributed antenna is reformulated as a linear sum assignment problem. A low-complexity algorithm is developed in order to find a fast solution

to the matching problem. Once that the set of links has been defined, considering equal power allocation (EPA) and unconstrained individual rates, an algorithm based on the user-removal principle is designed in order to refine the set of scheduled links that improve the system performance. Assuming that the set of close-to-optimal links has been predefined, Section 4.3 focuses on the resource allocation optimization for a general scenario with centralized and coordinated transmitters. The remaining problem is to assign the best rate and power to the scheduled links subject to a set of power constraints and finite number of available rates. Three approaches to solve this problem are presented and numerical results are obtained for a centralized DAS scenario at Section 4.4 which assess the performance of the proposed methodologies.

4.2 The User-Antenna Matching Problem

In multiuser MIMO systems where the antennas at the transmitter are co-located (CAS), the most common criterion for selecting users for downlink transmission is based on the spatial correlation between wireless channels (see Chapter 5). In [122] the authors propose a user selection based on graph theory where the criterion to group users served by the same spreading code in a MIMO-CDMA system is defined by their nearly orthogonal spatial signatures. Using graph coloring techniques, users that can create interference between each other are isolated and users with relative low values of interference are served in the same radio resource. A low-complexity algorithm for user selection was proposed in [123] where the set of selected users minimizes the sum of the channel correlation. As the minimum average correlation metric is not enough to guarantee the largest sum rate, an improved version of [123] was proposed in [124] where the scheduling decisions are made evaluating iteratively the correlation and the achievable sum rate of each possible set of users. Optimal and sub-optimal techniques to exploit such channel properties to efficiently optimize the system performance are studied in Chapter 5 and Chapter 6 where signal design is required to fully exploit CSI at the transmitter side.

Although extensive research has been done for user selection in MIMO systems, the selection schemes presented in the literature cannot be applied directly to DAS systems where each distributed antenna serves only one user. In this scenario, the most common scheduling strategy attempts to maximize the system sum rate based on the scheduling of the users with best channel conditions. This is the well known channel-aware scheduling, which exploits instantaneous channel conditions and MUDiv [21]. Authors in [125] implement a channel-aware scheduling by designing a user selection algorithm based on the statistical characteristics of the instantaneous large-scale fading for DAS. Such a greedy algorithm finds the set of dominant users for which the difference between their instantaneous channels lies below a threshold.

In this section, a centralized user selection algorithm is designed to attach each distributed antenna or single-antenna *remote antenna unit* (RAU) to at most one user in order to optimize the system performance. The mathematical formulation of this kind of problem resembles a particular class of matching problem known as linear sum assignment problem (LSAP) [126]. A methodology for user selection is proposed for the downlink transmission in DAS scenarios, which finds a sub-optimal solution of the LSAP by means of a low-complexity matching algorithm.

4.2.1 The Matching Problem Formulation

Consider a single cell wireless communication system with K single-antenna users and one CU equipped with N RAUs as shown in Fig. 4.1(b). The channel between the user k and the RAU n is assumed to be a block fading Rayleigh where the channel gains are constant during each time slot and change independently from slot-to-slot. The channels are affected by a path-loss component, and a shadowing fading component modeled as a log-normal distributed variable with parameter σ_s. The signal received by the user $k \in \{1, \ldots, K\}$ is defined as:

$$y_k = \mathbf{h}_k \mathbf{x} + z_k, \tag{4.1}$$

where $\mathbf{x} = [x_1, \ldots, x_N]^T$ is the transmitted symbol vector by the RAUs, $\mathbf{h}_k \in \mathbb{C}^{1\times N}$ is the channel vector to the kth user, z_k is the additive white Gaussian noise (AWGN) with variance σ_k^2 at the kth user. It is assumed perfect CSI at the CU and $\mathbb{E}[|x_i|^2] = p_i$. By considering equal power allocation (EPA), i.e., that all RAUs have allocated power $p_i = p, \forall i$, the SINR experienced by the user kth in the nth RAU can be expressed mathematically as:

$$SINR_{k,n} = \frac{p|h_{k,n}|^2}{\sum_{j=1, j\neq n}^{N} p|h_{k,j}|^2 + \sigma_k^2}. \tag{4.2}$$

Each RAU feeds back its CSI to the CU where the scheduling processing is performed. The achievable sum rate of the kth user selected to be scheduled in the nth RAU is given by:

$$R_{k,n} = \log(1 + SINR_{k,n}). \tag{4.3}$$

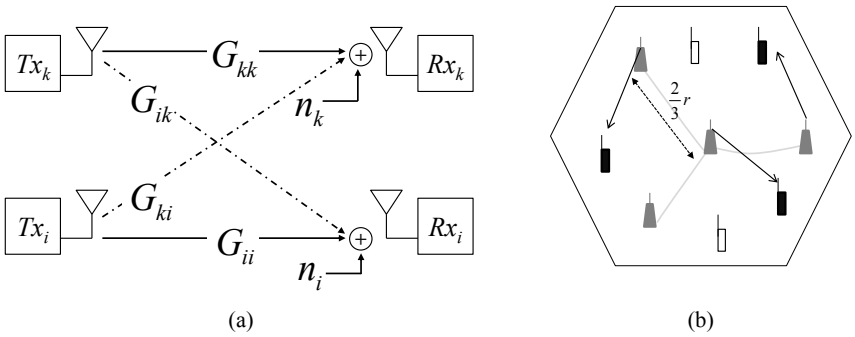

Figure 4.1: (a) System model for two wireless links k and i. (b) Distributed Antenna System Scenario.

The CU must define a set of users that maximizes a specific global function such as the total sum rate. The optimum set of users in this scenario can be found within a search space of size

$$\underbrace{N!\binom{K}{N}}_{\text{For } N \text{ matched antennas}} + \underbrace{N}_{\text{Possible combinations}} \cdot \underbrace{\left(\sum_{n=1}^{N-1} n!\binom{K}{n}\right)}_{\text{Match over a subset of } n \text{ antennas}}, \ \forall K \geq N \tag{4.4}$$

which is prohibitively large even for small values of K. Notice that the search space are all possible combinations that N RAUs can schedule K users having at least one RAU transmitting. In order to reduce the complexity of the user selection, the algorithms proposed in this section analyze the search space of size $N!\binom{K}{N}$ where all RAUs transmit at the same time.

In order to illustrate the users selection in this search space, consider the system in Fig. 4.2 for a single radio resource, consider all K users in the set $V_2 = \{1, 2, 3\}$ to have the same priority, all N RAUs in the set $V_1 = \{a, b, c\}$ to schedule only one user. The links between the users and the RAUs represent the achievable rate given by (4.3) and can be summarized in the matrix $\mathbf{C} \in \mathbb{R}^{K \times N}$. One approach to generate the set of scheduled users is to find iteratively the link with maximum achievable rate, make a match for its corresponding RAU and user, and drop all other links associated with them. This operation can be mathematically represented as a multiplication of the elements in the same row and column of the selected entry by a binary variable defined as:

$$\alpha_{k,n} = \begin{cases} 1 & \text{if row } k \text{ is assigned to column } n \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

Following this procedure the final link match and total sum rate in Fig. 4.2 are $\{ (1, b)\ (2, a)\ (3, c)\ \}$, $\{18\}$, respectively. Nevertheless, among all possible matches, $N! \binom{K}{N} = 6$, the optimum solution is given by $\{(1, b)(2, c)(3, a)\}$ that in this context is Pareto optimal[1] with maximum sum rate $\{21\}$. The user selection in this scenario is a class of weighted matching problem called linear sum assignment (LSAP) that can be defined as follows [126]:

Given a $K \times N$ matrix $\mathbf{C} = (c_{k,n})$, the problem is to match each row to a different column in such a way that the sum of the corresponding entries is maximized. This problem can be mathematically expressed as:

$$\text{maximize} \quad \sum_{k=1}^{K} \sum_{n=1}^{N} \alpha_{k,n} c_{k,n} \tag{4.6}$$

$$\text{subject to} \quad \sum_{n=1}^{N} \alpha_{k,n} = 1 \qquad (k = 1, \ldots, K) \tag{4.7}$$

$$\sum_{k=1}^{K} \alpha_{k,n} = 1 \qquad (n = 1, \ldots, N) \tag{4.8}$$

$$\alpha_{k,n} \in \{0, 1\} \tag{4.9}$$

where $\alpha_{k,n}$ is the weight of the element $c_{k,n}$ defined by (4.5). In literature of combinatorial optimization (e.g., [126, 128]) the optimal solution of (4.6)-(4.9) can be found by several approaches (primal-simplex algorithms, primal-dual algorithms, shortest path algorithms, etc.). In the following section, a low-complexity algorithm that can find a close-to-Pareto solution to the LSAP for the DAS will be designed and assessed.

4.2.2 Greedy and Minimum Rate Loss Matching

The process of scheduling starts when the users report their channel gains to the CU. The achievable rate $R_{k,n}$ of user k is evaluated and it is assumed that RAU n transmits the desired signal and all other RAUs are sources of interference. The terms $R_{k,n}, \forall k \in V_2, \forall n \in V_1$ are collected in the $\mathbf{C} \in \mathbb{R}^{K \times N}$ matrix. The goal is to assign one user to each RAU meeting the constraints (4.7)-(4.9). The entries of the matrix $\mathbf{C}$ are defined by the wireless channels. For

[1] A Pareto optimal allocation is one such that, there does not exist another feasible allocation where at least one user gets a better resource assignment (RAU, carrier, etc), and all others get at least the same resources. This optimal allocation cannot be improved on without hurting at least one user [127].

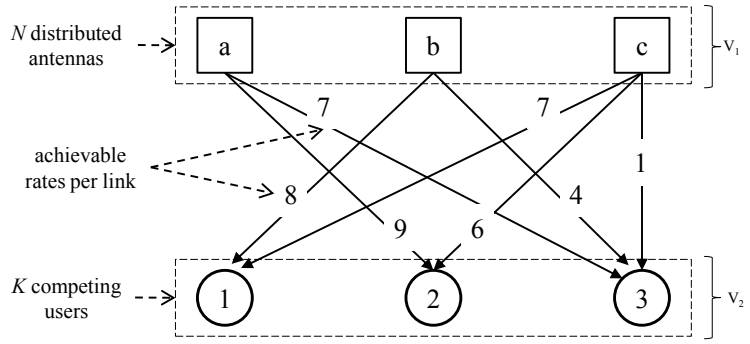

Figure 4.2: A wireless network with $N = 3$ RAUs $\{a, b, c\}$ and $K = 3$ users $\{1, 2, 3\}$. V_1 is the set of all RAUs in the system and V_2 is the set of active users in the system.

---

a given channel realization and user deployment, the information in matrix $\mathbf{C}$ can reflect two different conditions that may take place simultaneously:

1. In the first condition, the matrix $\mathbf{C}$ has dominant elements in V_2 such that every user i has $c_{i,m} \gg c_{i,n}, \forall m, n \in V_1, n \neq m$. In this case the user selection can be performed in a greedy fashion due to the fact that each user has only one link with high achievable rate value.

2. In the second condition, assume that for some elements in V_2 there are several links with large achievable rate such that $c_{i,m} \approx c_{i,n}, \forall m, n \in V_1^{'}, V_1^{'} \subseteq V_1, n \neq m$, i.e., more than one element in V_1 can serve user i.

In the *greedy selection* for each step of the scheduling process, the user with maximum achievable rate k attached to RAU n is selected and all links related to that specific user and RAU are dropped as described in Algorithm 4.1a. This means that when a match between user k and RAU n is found, all remaining matches (with good and poor achievable rate) for k and n are lost in following iterations. Therefore, by being greedy some resources that would be assigned in the future iterations are dismissed. In order to minimize such a loss, a *minimum-rate-loss* algorithm is proposed. In this approach, each time a match is found it minimizes the losses of such selection by finding the match that will provide the maximum amount of available rate for the next iteration.

Define $c^{(T)}$ as the total available rate in the cell computed as:

$$c^{(T)} = \sum_{k} \sum_{n} c_{k,n}, \tag{4.10}$$

where $c_{k,n}$ is the element in row k and column n in $\mathbf{C}$. If a link (i, j) is selected, the amount of total available rate that will be discarded by this selection, is evaluated as:

$$c_{i,j}^{(D)} = \sum_{k} c_{k,j} + \sum_{n \neq j} c_{i,n}. \tag{4.11}$$

The remaining rate in the system once that link (i, j) has been selected, is given by $c_{i,j}^{(A)} = c^{(T)} - c_{i,j}^{(D)}$. For the minimum-rate-loss approach summarized in Algorithm 4.1b, the

Algorithm 4.1a Greedy Matching

1: Find the entry $c^*_{k,n} = \max_{k \in V_2, n \in V_1} c_{k,n}$.
2: Define the match $\mathcal{N} \leftarrow \mathcal{N} + \{(k,n)\}$.
3: Multiply all the entries in row k and column n by their respective weight given by (4.5).
4: Go to Step 1 until all RAUs have been assigned.
5: **return** a match solution $\mathcal{N}$.

Algorithm 4.1b Minimum Rate Loss Matching

1: $\mathbf{C}^{(temp)} \leftarrow \mathbf{C}$
2: Compute $\tilde{\mathbf{C}}$ applying (4.12) to $\mathbf{C}^{(temp)}$.
3: Find the weight $c^*_{k,n} = \max_{k \in V_2, n \in V_1} \tilde{c}_{k,n}$.
4: Define the match $\tilde{\mathcal{N}} \leftarrow \tilde{\mathcal{N}} + \{(k,n)\}$
5: In $\mathbf{C}^{(temp)}$ multiply all the entries in row k and column n by zero.
6: Go to Step 2 until all elements in $\mathbf{C}^{(temp)}$ are zero.
7: **return** a match solution $\tilde{\mathcal{N}}$.

matching is performed over the matrix $\tilde{\mathbf{C}}$ and its elements are evaluated by applying the following transformation:

$$\tilde{c}_{k,n} = c_{k,n} \left(\frac{c^{(A)}_{k,n}}{c^{(D)}_{k,n}} \right), \tag{4.12}$$

where the quotient that multiplies the original term $c_{k,n}$ is the relative gain of the available rate over the rate loss for the (k,n) match. The problem (4.6) is solved by using the outputs of Algorithm 4.1a and Algorithm 4.1b. After applying both algorithms, there are two possible solutions $\mathcal{N}$ and $\tilde{\mathcal{N}}$. The final selected match is the one with maximum sum rate, and the set of potential users that can be scheduled is given by $\mathcal{K} = \{k_1, \ldots, k_{|\mathcal{K}|}\}$ with $|\mathcal{K}| \leq N$.

4.2.3 Rate Maximization with Equal Power Allocation

Solving the user-RAU matching problem is the first step to maximize the sum rate in a multiuser scenario. The second step is to allocate power so that the total system sum rate is maximized. Allocating power to all user in $\mathcal{K}$ may yield a degradation in the system performance due to the fact that one or more links may create more rate degradation to other links than the rate that they actually achieve. In such case, it is necessary to refine the set of selected users and only allocate power to those users that contribute to the sum rate enhancement.

For the set of selected users $\mathcal{K}$ there exist $L = \sum_{n=1}^{N} \binom{|\mathcal{K}|}{n}$ subsets $S^{(l)}_{(\mathcal{K})}, l \in \{1, \ldots, L\}$ that represent all possible ways that elements in $\mathcal{K}$ can be scheduled with the constraint that each user can be served only by its previously assigned RAU. It is possible that the sum rate is maximized by a subset $S^{(l^{opt})}_{(\mathcal{K})}$ such that:

$$(l^{opt}) = \arg \max_{l \in \{1,\ldots,L\}} R(S^{(l)}_{(\mathcal{K})}), \tag{4.13}$$

where the term $R(S^{(l)}_{(\mathcal{K})})$ is the sum rate achieved by subset $S^{(l)}_{(\mathcal{K})}$ and is given by:

$$R(S_{(\mathcal{K})}^{(l)}) = \sum_{i \in S_{(\mathcal{K})}^{(l)}} \log\left(1 + SINR_{k_i,n_{k_i}}\right) \tag{4.14}$$

where k_i is the ith user in subset $S_{(\mathcal{K})}^{(l)}$ and n_{k_i} is its associated RAU.

4.2.3.1 Minimum Rate Contribution Selection

In order to estimate $S_{(\mathcal{K})}^{(l^{opt})}$ without computing all L possible combinations, an iterative process drops users based on the *minimum rate contribution* criterion. This process finds the user for which the downlink transmission generates more interference to the other users compare with his own achievable rate. This is done evaluating the individual rate by (4.3) and calculating the total sum rate by (4.14). Then, it is considered that the user with minimum value for (4.3) is dropped and (4.3) and (4.14) are evaluated for the remaining users. At the end, the user with minimum rate contribution is dropped if the total sum rate is larger when such user is not scheduled. This process is summarized in Algorithm 4.2.

Algorithm 4.2 Minimum Rate Contribution (MinRC)

1: $S_{(\mathcal{K})}^{(e)} \leftarrow \mathcal{K}$
2: $\forall k_i \in S_{(\mathcal{K})}^{(e)}$ eval R_{k_i,n_i} by (4.3).
3: Compute the rate $R(S_{(\mathcal{K})}^{(e)})$ by (4.14).
4: Find the user $k^{(min)} = \min_{k_i \in S(\mathcal{K})} R_{k_i,n_i}$.
5: Define $S_{(\mathcal{K})}^{(tmp)} = S_{(\mathcal{K})}^{(e)} - \{k^{(min)}\}$ and eval $R(S_{(\mathcal{K})}^{(tmp)})$
6: **If** $R(S_{(\mathcal{K})}^{(tmp)}) > R(S_{(\mathcal{K})}^{(e)})$
7: $\quad S_{(\mathcal{K})}^{(e)} \leftarrow S_{(\mathcal{K})}^{(e)} - \{k^{(min)}\}$, and go to Step 2.
8: **else**
9: $\quad$ Scheduled all users in $S_{(\mathcal{K})}^{(e)}$
10: **End**

For the user-RAU matching, the complexity is $O(KN^2)$ considering that each matching implies a process of complexity $O(KN)$ and this process is repeated N times. Nevertheless, for systems where the number of distributed antennas is small and $K \geq N$, the complexity of solving the matching problem by the proposed algorithms is limited. Even if the time complexity of the proposed user assignment algorithm is $O(KN^2)$, its computational complexity (floating point operations) is low compared to the complexity of the algorithms that find the optimal solution for (4.6). The proposed maximization of (4.14) implies a linear search that is given by the cardinality of the set of matched users $O(|\mathcal{K}|)$.

4.2.4 Numerical Results for the Matching Problem

The deployment of the distributed antennas consists of one RAU at the center of the cell and three distributed RAUs uniformly deployed at a distance from the cell center of $\frac{2}{3}$ the cell radius r, illustrated in Fig 4.1(b). The simulation parameters are listed in Table 4.1. In order to quantify the fairness of a given resource allocation policy, several metrics have been proposed in the literature of engineering, economics, and other fields [68, 129]. A fairness

Table 4.1: Simulation Parameters for user selection in the DAS Scenario

Parameters	Values
Cell radius	900 m
Carrier frequency	2.5 GHz
Channel bandwidth	20 MHz
RAU transmit power	43 dBm
Thermal noise power density	-174 dBm
UE noise figure	7 dB
Path loss model	UMi-LoS [130]
Shadow fading standard deviation	$\sigma_s = 3$
Available MCS (M)	8 [50]
Max. num. iterations τ	50
User deployment	uniform

measure is a sequence of mapping $\{f^n : \mathbb{R}_+^n \mapsto \mathbb{R},\ \forall n \in \mathbb{Z}_+\}$ over a given set of n non-negative elements [68]. For a given a vector $\mathbf{x} \in \mathbb{R}_+^n$, let the Gini index $\{F_G(\cdot) \in \mathbb{R}_+ | 0 \leq F_G(\cdot) \leq 1\}$ be the metric to quantify the inequality (unfairness) in the distribution of the rate among selected users, having $F_G(\mathbf{x}) = 0$ as the maximum level of fairness. For K active users in the system, F_G over all channel realizations is given by [129]:

$$F_G = \frac{1}{2K \sum_{k=1}^{K} \overline{R}_k} \left(\sum_{k=1}^{K} \sum_{i=1}^{K} |\overline{R}_k - \overline{R}_i| \right) \tag{4.15}$$

where $\overline{R}_k$ is the average rate achieved by the kth user.

The tested algorithms were analyzed in terms of different performance metrics: the average sum rate, the cardinality of the final set of scheduled users, and the fairness of the average achieved rate. The proposed algorithms to match user-RAU Algorithm 4.1a -Algorithm 4.1b (Alg. 4.1a-b) was compared to both, the generalized Jonker-Volgenant algorithm (JVA) for non-square dense matrices [128] which finds the optimum solution of (4.6)-(4.9), and the dominant user grouping (DUG) algorithm for rate-maximization [125]. The output of the algorithms to make the user-RAU matching was combined in the second phase with three approaches that establish the final subset of scheduled users:

1. $S_{(\mathcal{K})}^{(e)}$ is optimum among all L subsets and maximizes (4.14).

2. $S_{(\mathcal{K})}^{(e)}$ is found by MinRC algorithm.

3. $S_{(\mathcal{K})}^{(e)} = \mathcal{K}$, i.e. all matched users from first phase are scheduled.

Fig. 4.3 shows the average sum rate for different values of K and $N = 4$ achieved by Algorithms 4.1a-b and JVA. The numerical evaluations show that the proposed algorithm Alg. 4.1a-b achieves the same average performance of JVA for the user-RAU matching in the first phase. Fig. 4.3 shows the three approaches to define $S_{(\mathcal{K})}^{(e)}$ and the gap in terms of sum rate when the subset of scheduled users is found by MinRC and when all matched users in the first phase are scheduled. For $K = 10$ the gap in performance between the optimum set compared

to $S_{(\mathcal{K})}^{(e)} = \mathcal{K}$ is up to 15%. Nevertheless, the sum rate loss incurred by not maximizing (4.14) is compensated by scheduling more user which is a gain in terms of fairness.

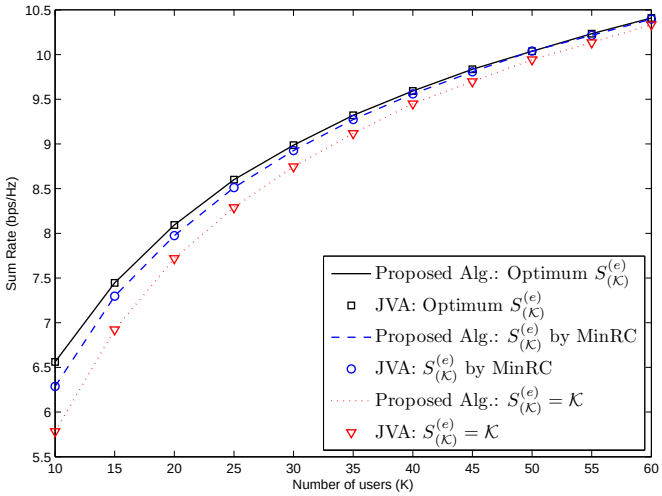

Figure 4.3: Sum rate for $N = 4$, different values of K.

Fig. 4.4 shows the expected value of the cardinality of the subset of scheduled users over N, $\mathbb{E}[|S_{(\mathcal{K})}^{(e)}|/N]$. For lower values of K the percentage of scheduled users in the optimum subset $S_{(\mathcal{K})}^{(e)}$ can be up to 30% below the number of users scheduled when $S_{(\mathcal{K})}^{(e)} = \mathcal{K}$. If the subset $S_{(\mathcal{K})}^{(e)}$ is found by MinRC, it is attained an improvement of 15% in the number of scheduled users compare to the optimum subset and the loss in sum rate is only of 5%. Table 4.2 shows the comparison in terms of sum rate for $S_{(\mathcal{K})}^{(e)} = \mathcal{K}$ between Alg. 4.1a-b and DUG for low values of K. The performance gap between Alg. 4.1a-b and DUG in terms of sum rate is less than 3% but for Alg. 4.1a-b the percentage of scheduled users is 100%. These differences between Alg. 4.1a-b and DUG become negligible for large values of K.

The rate distribution among users measured by (4.15) is shown in Fig. 4.5. The results reflect the contradicting objectives of maximizing (4.14) and scheduling all users found in the first phase. It can be observed that by defining the subset of scheduled users by MinRC, a trade-off between sum rate (a gap of 5% compared with the optimum subset) and fairness is achieved. The CDF of F_G with MinRC is closed to the CDF of the scheme that schedules all user found in the first phase by Alg. 4.1a-b. These results are equal when $\mathcal{K}$ is found by

Table 4.2: Comparison between Alg. 4.1a-b and DUG for $R(S_{(\mathcal{K})}^{(e)})$ and $\mathbb{E}[|S_{(\mathcal{K})}^{(e)}|/N]$ for low values of K and $S_{(\mathcal{K})}^{(e)} = \mathcal{K}$.

	K	10	15	20	25
$R(S_{(\mathcal{K})}^{(e)})$	Alg. 4.1a-b	5.78	6.92	7.71	8.29
	DUG	5.89	6.94	7.72	8.29
$\mathbb{E}[\|S_{(\mathcal{K})}^{(e)}\|/N]$	Alg. 4.1a-b	1	1	1	1
	DUG	0.94	0.98	0.99	0.99

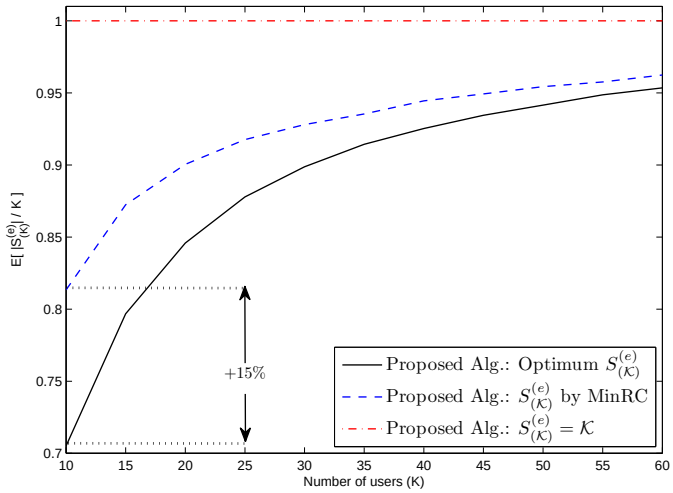

Figure 4.4: Percentage of user assignment $\mathbb{E}[|S^{(e)}_{(\mathcal{K})}|/N]$ for several values of K and different sets $S^{(e)}_{(\mathcal{K})}$.

JVA in the fist phase and the sum rate is maximized by MinRC. For a large number of users ($K \geq 30$) the gap in terms of sum rate between all possible strategies to find $S^{(e)}_{(\mathcal{K})}$ vanishes. Fig. 4.6 shows the CDF of the achieved sum rate for $K = 30$. The gap in performance between the scheduling with MinRC and the optimum scheduling is negligible. Moreover, by scheduling all users found by Alg. 4.1a-b in the first phase, the sum rate gap is only 4% compared to the optimum scheduling with the advantage that 100% of the matched links are active.

4.3 Power Allocation and Link Adaptation

In this section it is assumed that all transmitter-receiver pairs are given, either because $K = N$ or because a previous user selection as the one described in Section 4.2 was previously performed in scenarios where $K > N$. Hereafter, the addressed problem is to maximize the achievable SINR subject to a set of power constraints and considering that the SINR can only take values from a finite set defined by the available MCSs. The optimal system performance can be achieved if the best set of links and their respective SINR are selected, which requires further optimization scheduling of the given links. The joint scheduling and resource allocation that maximizes the sum rate is a complex combinatorial problem that grows exponentially with the number of links and depends strongly on the number of available MCSs. The optimal solution of such a combinatorial problem can be found via exhaustive search by selecting the set of links and MCSs that yield the maximum sum rate. Such procedure has prohibitive complexity and it is required to develop sub-optimal but efficient algorithms that achieve a good trade-off between sum rate performance and complexity. The proposed algorithms merge the objectives of the resource allocation optimization and the user-removal techniques. This integration is achieved by operating over two dimensions of decision, the set of links and the set of available MCSs. Conceptually, the proposed algorithms work in two phases. The first phase establishes whether for a given set of links exists a power and rate allocation that satisfies the SINR requirements imposed by some MCSs under a set of power constraints. In

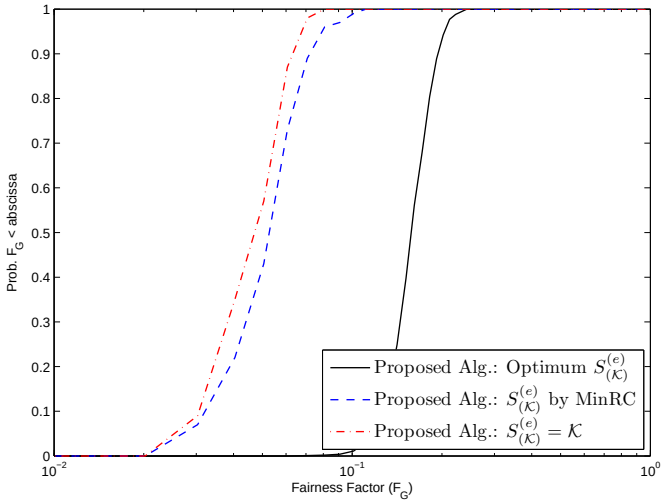

Figure 4.5: CDF of the Fairness Gini Index for $K = 10$ and $N = 4$.

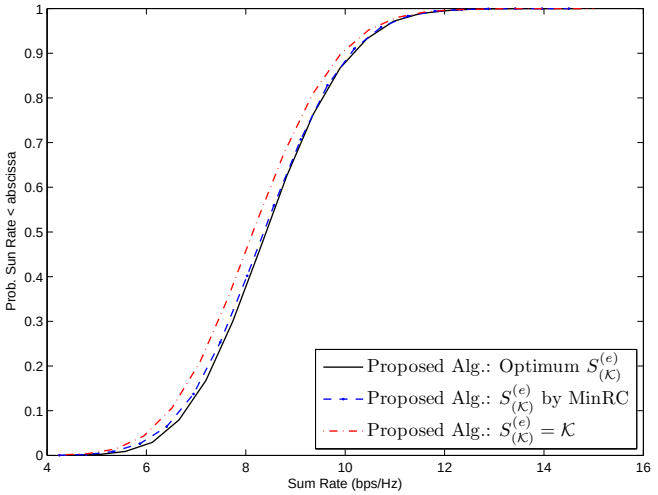

Figure 4.6: CDF of the sum rate for $K = 30$ and $N = 4$. In this case the set $S^{(e)}_{(\mathcal{K})} = \mathcal{K}$ found by Alg. 4.1a-b can be scheduled with only 4% sum rate gap w.r.t. the optimum $S^{(e)}_{(\mathcal{K})}$. As $K \to \infty$ this sum rate gap becomes negligible.

other words, this phase verifies if the set of links and their MCSs are feasible. The second phase modifies either the set of links or the assigned MCSs, based on the feasibility measure provided by the first phase.

In the following section it is shown how this two iterative phases can be designed using the Perron-Frobenius theory by formulating the sum rate maximization problem as an eigenvalue optimization problem. Although, this approach achieves acceptable sum rate and outage figures, it requires the computation of the eigenvalues that characterize the interference coupled network in each iteration. For this reason, alternative solutions that only require either the evaluation of the power consumption or the estimation of the achievable SINR per iteration are designed. Furthermore, it is proposed a low complexity algorithm that performs a fast estimation of the set of links and MCSs that solve the sum rate maximization problem. Numerical results show that despite the fact that the proposed algorithms are suboptimal strategies, they are asymptotically optimal when the number of users in the network grows to infinity. Moreover, it is shown that the proposed algorithms generalize the concept of user-removal for the case of multiple SINR targets, and they are efficient low-complex alternatives to the state-of-the-art user-removal algorithms.

4.3.1 Resource Allocation Problem

Consider the wireless network depicted in Fig. 4.1(a) where the current channel instance is simultaneously being used by K synchronized links. The receivers decode its corresponding data treating interference as white noise, i.e., multiuser detection is not employed. In order to mathematically characterize the interference coupled network, in the following it is adopted the matrix notation and the system model used in recent works [37, 70]. Let p_k be the power used by the kth link and $\mathbf{p}$ the vector that summarized all K powers. The SINR experienced by the kth link given a power allocation $\mathbf{p}$ is [37]:

$$SINR_k(\mathbf{p}) = \frac{p_k G_{kk}}{\sum_{i \neq k}^{K} p_i G_{ki} + \sigma_k^2}, \tag{4.16}$$

where G_{ki} is the power attenuation from the transmitter on link i to the receiver on link k, taking into account propagation loss, fast and slow fading, and G_{kk} is the power loss for the intended transmission at link k. The SINR in (4.16) is a particular form of the SINR since G_{ki} can be given either by a function of the channel magnitude $G_{ki} = f(\mathbf{h}_{ki})$ used in Section 4.2.1 or by the precoded channel $G_{ki} = f(\mathbf{h}_{ki}, \mathbf{w}_{ki})$ as in [131] where $\mathbf{w}_{ki}$ is a precoding weight whose design is studied in Chapter 5 and Chapter 6.

Efficient allocation schemes seek the simultaneous provisioning of individual QoS for multiple wireless links, which implies that each link achieves a SINR that can be maintained above a given threshold or target:

$$SINR_k(\mathbf{p}) \geq \gamma_k. \tag{4.17}$$

The SINR target of the kth link is constrained to take values from a finite set of targets $\gamma_k \in \mathcal{M}_k$ where $\mathcal{M}_k = \{\gamma^{(1)}, \ldots, \gamma^{(M)}\}$, $\gamma^{(m-1)} < \gamma^{(m)}$, and M is the number of available MCSs. A larger value of γ_k implies that link k attempts to maintain a more spectral efficient modulation scheme. Depending on the service, $\mathcal{M}_k$ can be used to assign different priorities among users or to limit their achievable data rates. For instance, the rate demand in link i can be limited while link k can improve its performance having a larger number of available

modulations, $|\mathcal{M}_k| > |\mathcal{M}_i|$ for some $k \neq i$. For sake of simplicity and without loss of generality, it is assumed the same $\mathcal{M}$ for all links, $\mathcal{M}_k = \mathcal{M}, \ \forall k$. The kth link is associated with a modulation index m_k that defines the position of its SINR target in the set $\mathcal{M}$ so that $\gamma_k = \gamma^{(m_k)}$. The discrete set of targets is given by the available set of MCSs supported in the systems, which in practice is defined by the user equipment capabilities and the wireless network technology. The vector of SINR targets is defined as $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)^T$ and all SINR targets will be summarized in a diagonal matrix $\boldsymbol{\Gamma} = diag(\boldsymbol{\gamma})$.

The users requirements in (4.17) can be described in a vector inequality of the form:

$$\mathbf{p} \geq \boldsymbol{\Gamma}\mathbf{V}\mathbf{p} + \boldsymbol{\Gamma}\mathbf{z}, \tag{4.18}$$

where $\mathbf{V}$ is a $K \times K$ nonnegative matrix whose entries are defined as:

$$[\mathbf{V}]_{ki} = \begin{cases} G_{ki}/G_{kk} & \text{if } k \neq i \\ 0 & \text{if } k = i \end{cases} \tag{4.19}$$

and it is assumed that $\mathbf{V}$ is irreducible, which means that each link has at least one interferer[2] [36]. The weighted noise vector $\mathbf{z}$ is defined as:

$$\mathbf{z} = \left(\frac{\sigma_1^2}{G_{11}}, \ldots, \frac{\sigma_K^2}{G_{KK}} \right)^T. \tag{4.20}$$

Consider two sets of power constraints: (a) individual power constraints (IPC), summarized in $\bar{\mathbf{p}}$ so that $\bar{p}_k$ is the maximum available power for the kth link. (b) total power constraints (TPC) so that $\sum_{k=1}^{K} p_k \leq P_t$. For the case of TPC, let $\mathbf{B}$ be a $K \times K$ nonnegative irreducible matrix defined as [37]:

$$\begin{aligned} \mathbf{B} &= \boldsymbol{\Gamma}\mathbf{V} + \frac{1}{P_t}\boldsymbol{\Gamma}\mathbf{z}\mathbf{1}^T \\ &= \boldsymbol{\Gamma}\Big(\mathbf{V} + \frac{1}{P_t}\mathbf{z}\mathbf{1}^T\Big). \end{aligned} \tag{4.21}$$

The matrix $\mathbf{B}$ absorbs the total power constraints as an additional source of interference whose power is inversely proportional to the power P_t. The properties of $\mathbf{B}$ provide fundamental information regarding to the transmission reliability and how interference, powers, and SINR targets are coupled. For the case of IPC, there exists a set of K matrices $\mathbf{B}_k$ and each one absorbs the power constrains of the kth link as an additional source of interference, and they are defined as [37, 38]:

$$\mathbf{B}_k = \boldsymbol{\Gamma}\Big(\mathbf{V} + \frac{1}{\bar{p}_k}\mathbf{z}\mathbf{e}_k^T\Big), \ \forall k \tag{4.22}$$

where $\mathbf{e}_k$ is the kth unit coordinate vector.

In resource allocation theory, one of the fundamental problems is to find the appropriate vector of targets $\boldsymbol{\gamma}$ in order to optimize a given objective function. The set of all attainable target vectors is called feasible SINR target region. The geometry of the feasible SINR target region (Fig. 4.7) is defined by the coupling matrix $\mathbf{V}$, the noise $\mathbf{z}$, the available SINR targets $\boldsymbol{\gamma}$

[2]This implies that all the SINRs are coupled by at least one of the allocated powers.

defined by the MCSs, and the power constraints. Several works that provide a solution to the resource allocation problem exploit the characteristics of the feasible rate region which has a logarithmic relation with the feasible SINR targets. The optimum resource allocation that maximize the system performance lies in the boundary of such feasible region [36, 37, 107]. The information provided by the geometry of the feasible rate region can be exploited to determine not just the optimal resource allocation but also if time-sharing is required. This means that if for the set of links that is attempted to be scheduled does not exist a power allocation that satisfies all links requirements, the set must be split off assigning different time slots to different subset of links [70]. The need of time-sharing depends on the convexity of the rate region, however, such a geometry is not known a priory. This implies that finding the best subsets of users that should use time-sharing is a combinatorial problem that grows exponentially with the number of links[70].

Definition 4.1. *Feasibility conditions*: They are the sets of SINR requirements and power constraints that the resource allocation must fulfill.

Definition 4.2. *Resource feasibility*: A power vector $\mathbf{p}$ is feasible if for such a vector the feasibility conditions are met. A set of targets $\boldsymbol{\gamma}$ or $\boldsymbol{\Gamma}$ is called feasible if all targets in the set can be jointly achieved by power control. A set of links is called feasible if there exists at least one set of targets whose related powers meet the feasibility conditions.

Definition 4.3. *Jain's index of fairness*: For a given a vector $\mathbf{x} \in \mathbb{R}_+^n$, this measure of fairness is defined as follows [132]:

$$J(\mathbf{x}) \triangleq \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n \sum_{i=1}^{n} x_i^2}, \tag{4.23}$$

where $\{J(\cdot) \in \mathbb{R}_+ | \frac{1}{n} \leq J(\cdot) \leq 1\}$ having $J(\mathbf{x}) = 1$ as the maximum level of fairness.

The power allocation and the SINR target (MCS) selection are two sides of the same problem, i.e., the feasibility of $\mathbf{p}$ is given by feasibility of $\boldsymbol{\gamma}$ and vice versa. For each feasible target vector there exists a feasible componentwise power vector (unique up to a scaling factor) that produces such a target vector [36, 37, 70]. From (4.17) and (4.18) it can be observed that the power vector $\mathbf{p}$ and the achievable SINRs depend on the targets $\boldsymbol{\Gamma}$. From power control theory [36, 37, 69] and Perron-Frobenius theory of positive matrices [120, 121], it is known that for feasible targets in (4.17) the power vector can be computed as:

$$\mathbf{p} = [\mathbf{I} - \boldsymbol{\Gamma}\mathbf{V}]^{-1}\boldsymbol{\Gamma}\mathbf{z}. \tag{4.24}$$

Since the power resources are limited in the network any feasible vector $\mathbf{p}$ given by (4.24) must lie within a feasible region limited by a given set of constraints. Define the region of feasible powers considering IPC as:

$$\mathcal{P}^{IC} = \{\mathbf{p} \in \mathbb{R}_{++}^K : \mathbf{p} \ \leq \bar{\mathbf{p}}\}, \tag{4.25}$$

and the region of feasible powers considering TPC as:

$$\mathcal{P}^{TC} = \{\mathbf{p} \in \mathbb{R}_{++}^K : \|\mathbf{p}\|_1 \leq P_t\}, \tag{4.26}$$

where $K = |\mathcal{K}|$ is the cardinality of the subset of links that can be jointly supported $\mathcal{K} \subseteq \bar{\mathcal{K}}$, and $\bar{\mathcal{K}}$ is the set of all available links. The system performance is optimize by solving the

following rate maximization problem over the set of feasible links $\mathcal{K}$ constrained in the joint continuous power and discrete target regions:

$$\begin{aligned} \text{maximize} \quad & \sum_{k\in\mathcal{K},\, \mathcal{K}\subseteq\bar{\mathcal{K}}} w_{\boldsymbol{\gamma}} R(SINR_k(\mathbf{p})) \\ \text{subject to} \quad & \gamma_k \in \mathcal{M}, \qquad \forall k \in \mathcal{K} \\ & \mathbf{p} \in \mathcal{P}, \end{aligned} \tag{4.27}$$

where $R(SINR(\mathbf{p}))$ is the achievable rate associated with a given SINR [50], and $\mathcal{P}$ can be given either by (4.25) or (4.26) depending on the specific network requirements. As the elements of $\boldsymbol{\gamma}$ can only take values from a finite set $\mathcal{M}$, (4.27) is a combinatorial problem whose complexity depends on the size of $\mathcal{M}$ and the number of elements in $\bar{\mathcal{K}}$. The term $w_{\boldsymbol{\gamma}}$ is a priority weight associated with $\boldsymbol{\gamma}$ and the fairness constraints imposed to subset $\mathcal{K}$.

Finding a solution to (4.27) requires the optimization over the set of feasible links that can transmit simultaneously and their respective feasible modulations (and their associated powers). An exhaustive search algorithm attempting to solve (4.27) would require to test all the combinations of links and target vectors. The size of the associated search space $\Omega_{\bar{\mathcal{K}},M}$ is $(M+1)^{|\bar{\mathcal{K}}|}-1$, and several configurations of links sets and target vectors may be infeasible. Looking for the optimal solution in $\Omega_{\bar{\mathcal{K}},M}$ is extremely complex, therefore, iterative algorithms that find a sub-optimal solution to (4.27) are developed. On one hand, the Perron-Frobenius theory is used to reformulate the sum rate maximization problem as an eigenvalue optimization problem. On the other hand, alternative algorithms are designed in order to verify the resource feasibility. Such approaches use the information provided by the power consumption or the achievable SINR, so that the eigenvalue computations are avoided. The fundamental concept behind the proposed algorithms is that each iteration, the resource allocation feasibility is compromised by a link k^*, i.e., by its associated γ_{k^*} and p_{k^*}. The algorithms make decisions in two phases. In the first phase the fulfilment of the feasibility conditions is verified. If the feasibility conditions are violated, the second phase finds k^* and modifies the set of active links $\mathcal{K}$ or the target vector $\boldsymbol{\gamma}$ based on the information that k^* provides.

Hereafter the following notation $\Xi(k,\mathcal{K},\mathbf{V},\mathbf{z},\mathbf{p},\boldsymbol{\gamma})$ is adopted to indicate a drop event of the kth link and consequent actions follow: $\mathcal{K}=\mathcal{K}-\{k\}$, $\mathbf{V}=\mathbf{V}^{[k]}$, $\boldsymbol{\gamma}=\boldsymbol{\gamma}^{[k]}$, $\mathbf{z}=\mathbf{z}^{[k]}$, $\mathbf{p}=\mathbf{p}^{[k]}$, and $\gamma_i=\gamma^{(m_i=M)}\ \forall i\in\mathcal{K}$. The vector $\hat{\mathbf{p}}$ condenses a given set of constraints and is defined as:

$$\hat{\mathbf{p}} = \begin{cases} \bar{\mathbf{p}} & \text{if } \mathcal{P}=\mathcal{P}^{IC}, \\ \underline{\mathbf{p}} & \text{if } \mathcal{P}=\mathcal{P}^{TC}, \end{cases} \tag{4.28}$$

where $\underline{\mathbf{p}}$ is a power vector that distributes equally the total available power among all links, i.e., $\underline{p}_k = P_t/|\mathcal{K}|, \forall k\in\mathcal{K}$.

4.3.2 Optimal Resource Allocation

The optimal solution for problem (4.27) can be given by two different approaches depending on the constraints over the final subset of feasible links. On the one hand, if the size of $\mathcal{K}$ is not constrained, then the optimal solution $\boldsymbol{\gamma}^\star$ is found assuming that the priority weights are simply $w_{\boldsymbol{\gamma}}=1,\forall\boldsymbol{\gamma}$. On the other hand, if the size of $\mathcal{K}$ is constrained to be maximum, then it may exist a set of target vectors with equal sum rates and equal maximum cardinality. In such case, two criteria are considered to reach the optimum allocation: one, by making

$\boldsymbol{\gamma}^\star$ equal to any $\boldsymbol{\gamma}$ with the maximum cardinality and sum rate, which implies $w_{\boldsymbol{\gamma}} = 1, \forall \boldsymbol{\gamma}$. And two, by taking into account the rate distribution among the links in each possible set $\mathcal{K}$, where the optimum target vector $\boldsymbol{\gamma}^\star$ is given by:

$$\boldsymbol{\gamma}^\star = \arg \max_{\{\boldsymbol{\varrho}^{\{\mathcal{K}\}}\} \in \Omega_{\bar{\mathcal{K}},M}} \sum_{k \in \mathcal{K}} w_{\boldsymbol{\varrho}^{\{\mathcal{K}\}}} R([\boldsymbol{\varrho}^{\{\mathcal{K}\}}]_k), \tag{4.29}$$

where $w_{\boldsymbol{\varrho}^{\{\mathcal{K}\}}}$ is a fairness weight associated with the targets $\boldsymbol{\varrho}^{\{\mathcal{K}\}}$ of the subset $\mathcal{K}$ and is given by the Jain's fairness index[3] (4.23) of the data rates attained by $\boldsymbol{\varrho}^{\{\mathcal{K}\}}$. The optimum solution $\boldsymbol{\gamma}^\star$ found by (4.29) maximizes the sum rate times the fairness index, and the final number of active links, i.e., $|\mathcal{K}|$.

4.3.3 Perron-Frobenius root based Optimization

The feasible vector of targets that maximizes the sum rate (4.27) must be in the boundary of the feasible SINR target region. In order to characterize such region a condition for feasibility of the targets is required. From the Perron-Frobenius theory [120, 121], for a positive square matrix $\mathbf{A}$, the PF-root $\rho(\mathbf{A})$ and its associated right eigenvector $\mathbf{x}$ meet: $\mathbf{A}\mathbf{x} \leq \mathbf{x}$, if and only if $\rho(\mathbf{A}) \leq 1$. There exists a direct relation between this property and the mathematical representation of the coupled targets in $\mathbf{B}$. In the context of a interference coupled system described in (4.18), the Perron-Frobenius property means that the SINR targets are jointly achievable if and only if the following necessary and sufficient condition for feasibility is met under total power constraints [37, Theorem 5.68 and Corollary 5.69]:

$$\rho(\mathbf{B}) \leq 1. \tag{4.30}$$

When the system must meet individual power constraints the SINR targets $\boldsymbol{\gamma}$ are jointly achievable if and only if the following necessary and sufficient condition for feasibility is met [37]:

$$\max_{k \in \mathcal{K}} \quad \rho(\mathbf{B}_k) \leq 1. \tag{4.31}$$

Fulfilling (4.30) or (4.31) implies that interference in the system can be mitigated by power control, i.e., the SINR targets are feasible and (4.17) holds with equality. Furthermore, the power allocation vector $\mathbf{p}$ given by (4.24) equals the right eigenvector associated with $\rho(\mathbf{B})$ for TPC or $\rho(\mathbf{B}_k)\, \forall k \in \mathcal{K}$ for IPC [37].

For the case of TPC, the feasible region of targets can be defined as follows:

$$\mathcal{Q}^{TC} = \{\gamma_k = \gamma^{(m_k)} \in \mathcal{M}, \forall k \in \mathcal{K} : \rho(\mathbf{B}) \leq 1\}, \tag{4.32}$$

and similarly for IPC the feasible region is characterized as follows:

$$\mathcal{Q}^{IC} = \{\gamma_k = \gamma^{(m_k)} \in \mathcal{M}, \forall k \in \mathcal{K} : \max_{k \in \mathcal{K}} \rho(\mathbf{B}_k) \leq 1\}. \tag{4.33}$$

Fig. 4.7 illustrates the feasible region $\mathcal{Q}^{TC}$ which is the intersection of sub-levels sets of the spectral radii of all nonnegative irreducible matrices $\mathbf{B}$ [37] taking into account TPC.

[3]Notice that the fairness measured by the weights $w_{\boldsymbol{\varrho}^{\{\mathcal{K}\}}}$ can be also given by the Gini index defined in (4.15). However, the Jain's index is more convenient because the weights in (4.29) increase with fairness and in the worst case they can degrade the achievable rates at most by $\frac{1}{|\mathcal{K}|}$ [68]. In contrast, the Gini index has two drawbacks if it is considered to weight the achievable rates in (4.29). The first one is the domain of the mapping [0,1], and the other one is that $F_G \to 0$ as the fairness of the rate allocation $\boldsymbol{\varrho}^{\{\mathcal{K}\}}$ increases.

The region $\mathcal{Q}^{TC}$ can be completely characterized from its boundary since it is downward comprehensive [37], which means that any $\boldsymbol{\gamma}$ on the boundary of $\mathcal{Q}^{TC}$ defines a set of target vectors that are within the feasible SINR target region. Nonetheless, the point of interest that maximizes the sum rate, is a target vector that solves (4.27) whose components lie in the boundary of $\mathcal{Q}^{TC}$.

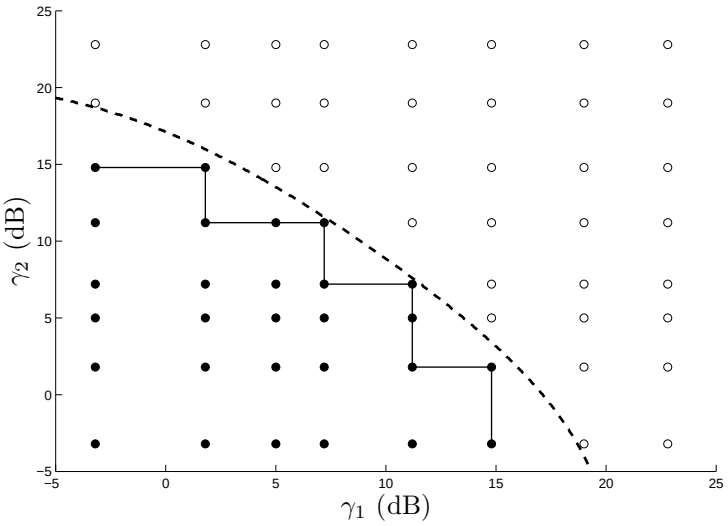

Figure 4.7: For a given $\mathbf{V}$, $\mathbf{z}$, $\mathcal{M}$, and P_t this is an example of the boundaries for the SINR target region for two cases: $\gamma \in \mathbb{R}_{++}$ (dashed line) and $\gamma \in \mathcal{M}$ (solid line). The vectors ($\bullet$) represent the combinations of γ that are jointly achievable for a two-link channel realization. The vectors ($\circ$) cannot be jointly achieved by any feasible power allocation.

The maximization problem can be reformulated as a PF-root optimization over the matrix $\mathbf{B}$ since the feasibility of the targets is given either by (4.30) or by (4.31). The original problem (4.27) can be rewritten as:

$$\begin{aligned} &\text{maximize} && \sum_{k\in\mathcal{K},\ \mathcal{K}\subseteq\bar{\mathcal{K}}} w_{\boldsymbol{\gamma}} R([\boldsymbol{\gamma}]_k) \\ &\text{subject to} && \boldsymbol{\gamma} \in \mathcal{Q}, \end{aligned} \tag{4.34}$$

where the constraint in (4.34) absorbs both constraints in (4.27) and the feasible SINR region $\mathcal{Q}$ is given either by (4.32) or (4.33) depending on the system power constraints.

4.3.3.1 Target Relaxation for a Non-Fixed Set of Links

In order to solve (4.34), it is required to identify which link violates the most the feasible conditions. From the theory of irreducible matrices it is known that if $\mathbf{A}$ is an irreducible square matrix and $\mathbf{A}^{[k]}$ is a proper principal submatrix of $\mathbf{A}$, then $\rho(\mathbf{A}^{[k]}) < \rho(\mathbf{A})$ [120, 121]. In the context of user-removal [38] this property is fundamental to determine the link that must be dropped since it relates the most infeasible link to the minimum PF-root over all principal sub-matrices of $\mathbf{B}$. In our context, the link k^* that compromises the most the resource feasibility is the one whose $\rho(\mathbf{B}^{[k^*]})$ is minimum. The fastest fulfillment of $\boldsymbol{\gamma} \in \mathcal{Q}$ is achieved whether k^* is temporary disconnected or its associated $\boldsymbol{\gamma}_{k^*}$ is relaxed so that the PF-root of the matrix $\mathbf{B}$ is minimized.

Hereafter the fairness weights are set to $w_{\boldsymbol{\gamma}} = 1, \forall \boldsymbol{\gamma}$ since neither the cardinality of the final $\mathcal{K}$ nor the distribution of the achievable rates among the links in $\mathcal{K}$ are considered. The algorithm starts by defining the initial conditions of the targets as $\gamma_k = \gamma^{(m_k=M)}$, $\forall k \in \mathcal{K}$ which may be or not a feasible starting point and is the maximum available target established by a given set $\mathcal{M}$. The feasibility of the target vector is verified either by (4.30) or by (4.31) and if $\boldsymbol{\gamma} \in \mathcal{Q}$ then the algorithm stops and all links in $\mathcal{K}$ transmit simultaneously with the powers defined by (4.24). If $\boldsymbol{\gamma}$ is infeasible a relaxation of the SINR targets is required and in the current iteration the algorithm modifies only the component $[\boldsymbol{\gamma}]_{k^*}$, where the most infeasible link k^* under total power constraints is computed as [38]:

$$k^* = \arg\max_{k \in \mathcal{K}} \left(\frac{1}{\rho(\mathbf{B}^{[k]})} \right), \tag{4.35}$$

and for individual power constraints k^* is found by [38]:

$$k^* = \arg\max_{k \in \mathcal{K}} \left(\min_{i \in \mathcal{K}, i \neq k} \frac{1}{\rho(\mathbf{B}_i^{[k]})} \right). \tag{4.36}$$

For the next iteration the link k^* reduces its target index by one unit $m_{k^*} = m_{k^*} - 1$, and its new SINR target is set to $\gamma_{k^*} = \gamma^{(m_{k^*})}$. In the case that link k^* cannot reduce its minimum target, then it is classified as infeasible or useless. Assigning any positive power to this link will create interference to the other links without achieving its minimum required SINR. At this point, the set of feasible links must be reduced so that $\mathcal{K} = \mathcal{K} - \{k^*\}$, and without the worst link, the algorithm attempts to allocate the maximum target for the remaining users, $m_k = M, \forall k \in \mathcal{K}$. The steps performed to solve (4.34) using the PF-root (PFR) approach are given in Algorithm 4.3.

Algorithm 4.3 PF-root (PFR) based Optimization

1: Set initial values: $\mathcal{K} = \bar{\mathcal{K}}$, $\gamma_k = \gamma^{(m_k=M)}, \forall k \in \mathcal{K}$.
2: **If** $\boldsymbol{\gamma} \in \mathcal{Q}$ by (4.30) or (4.31)
3: Set final vector $\mathbf{p}$ by (4.24), $\boldsymbol{\gamma}$, and Stop.
4: **else**
5: Compute k^* by (4.35) or (4.36)
6: **If** $m_{k^*} > 1$
7: $m_{k^*} = m_{k^*} - 1$, $\gamma_{k^*} = \gamma^{(m_{k^*})}$, Go to Step 2.
8: **else**
9: $\Xi(k^*, \mathcal{K}, \mathbf{V}, \mathbf{z}, \mathbf{p}, \boldsymbol{\gamma})$, Go to Step 2.
10: **End**
11: **End**

Algorithm 4.3 looks for the maximum SINR target vector on the boundary of $\mathcal{Q}$. The target vector $\boldsymbol{\gamma}$ found by this approach is not necessary the same size of the optimal one denoted as $\boldsymbol{\gamma}^\star$, and in such case the feasible regions of both solutions are different. Observe that this algorithm is only suitable for centralized wireless networks as it requires a CU that has a knowledge of global network parameters to compute the PF-root. Under IPC $(|\mathcal{K}|^2 - |\mathcal{K}|)$ PF-root computations are required each iteration and the maximum number of iterations require to find $\mathcal{K}$ is upper bounded by $(\sum_{i=1}^{|\bar{\mathcal{K}}|}((M-1)i+1)) - 1$, which depends on the the number of available MCSs in $\mathcal{M}$ and the size of the initial set $\bar{\mathcal{K}}$.

4.3.3.2 Target Increment for a Fixed Set of Links

In this approach the resource assignment starts with a feasible target vector and the objective is to increase its components as much as possible maintaining feasibility. Since $\mathcal{Q}$ is downward comprehensive [37] the vector $\boldsymbol{\gamma}$ found by Algorithm 4.3 defines a set of vectors which are inside the feasible region $\mathcal{Q}$. However, if the initial conditions of $\boldsymbol{\gamma}$ are set to the minimum attainable target, the unique explicit information about how far these initial conditions are from the boundary of $\mathcal{Q}$ is given by (4.30) or (4.31). Therefore, the problem at hand is the following: how to tighten up the SINR requirements in order to solve (4.34) keeping fixed the initial feasible subset of links $\mathcal{K}$?

Proposition 4.1. *Consider that $\boldsymbol{\gamma} \in \mathcal{Q}^{IC}$. The element γ_{k^*} that can be increased yielding the minimum increment of the left-hand side of (4.31) is the one whose $\rho(\mathbf{B}_{k^*})$ is minimum.*

Proof. Let i be the element whose $\rho(\mathbf{B}_i)$ is maximum. According to [38, Thm. 9] the maximum achievable $\boldsymbol{\gamma} \in \mathcal{Q}^{IC}$ is defined by $\bar{p}_i$ and the power p_i associated with γ_i is maximum, i.e., $i = \arg\min_k \bar{p}_k - p_k$ [70, Thm. 2]. Let $j = \arg\min_k \rho(\mathbf{B}_k)$, and $\boldsymbol{\gamma}_{(k)} \in \mathcal{Q}^{IC}$ where $\boldsymbol{\gamma}_{(k)} = \boldsymbol{\gamma} + \delta \mathbf{e}_k$ is the target vector whose kth term increases $\delta \in \mathbb{R}_{++}$. Since (4.31) identifies which is the tightest power constraint, i.e., which $\bar{p}_k$ is the hardest to meet, and an increment in $\boldsymbol{\gamma}_{(k)}$ requires an increment in p_k, making $\boldsymbol{\gamma}_{(j)}$ will increase the power p_j associated with the less tight power constraint $\bar{p}_j$. □

In this approach the minimum available target is initially assigned to all links and the priority weights are set to $w_{\boldsymbol{\gamma}} = 1, \forall \boldsymbol{\gamma}$. The links that cannot meet this minimum SINR requirement are found either by (4.35) or by (4.36) and dropped. Once that the initial subset of feasible links $\mathcal{K}$ has been defined, it follows from Proposition 4.1 that the link k^* candidate to increase its SINR requirements is given by:

$$k^* = \arg\min_{k \in \mathcal{K}} \ \rho(\mathbf{B}_k). \tag{4.37}$$

An approximation to (4.37) that avoids the $|\mathcal{K}|$ eigenvalue computations is to define k^* as the link with less power consumption relative to its power constraint:

$$k^* = \arg\max_{k \in \mathcal{K}} \ (\bar{p}_k - p_k)/\bar{p}_k. \tag{4.38}$$

Notice that this simplification applies since $\boldsymbol{\gamma}$ and $\mathbf{p}$ are assumed feasible. In the case of Algorithm 4.3 the vector of powers in the feasible power region $\mathcal{P}$ that satisfies all SINR requirements exists only if (4.30) or (4.31) (according to the system constraints) are met and it cannot be used directly to make previous decision. The tightening of the targets is performed stepwise until the conditions $\boldsymbol{\gamma} \in \mathcal{Q}$ cannot be fulfilled, which implies $|\mathcal{K}|$ PF-root computations each iteration as described in Algorithm 4.4.

4.3.4 Power Consumption based Optimization

In order to solve (4.27) using the tools provided by the Perron Frobenius theory, it is required to optimize the PF-root of the matrix $\mathbf{B}$. In each iteration of Algorithm 4.3, finding the worst link k^* requires $|\mathcal{K}|$ eigenvalue computations. If IPC (4.25) is imposed, finding k^* would require $(|\mathcal{K}|^2 - |\mathcal{K}|)$ PF-root evaluations per iteration. Moreover, the criterion used to evaluate feasibility of the resource allocation (4.30) is also based on the PF-root computation.

Algorithm 4.4 PF-root Incremental Target (PFRIT) based Optimization

1: Set initial values: $\mathcal{K} = \bar{\mathcal{K}}$, $\gamma_k = \gamma^{(m_k=1)}, \forall k \in \mathcal{K}$.
2: **If** $\boldsymbol{\gamma} \in \mathcal{Q}$
3: Define $\mathcal{K}^{tmp} = \mathcal{K}$
4: Compute k^* over $\mathcal{K}^{tmp}$ by (4.37) or (4.38)
5: **If** $m_{k^*} < M$
6: $m_{k^*}^{tmp} = m_{k^*} + 1$, $\boldsymbol{\gamma}^{tmp} = \boldsymbol{\gamma}$, $[\boldsymbol{\gamma}^{tmp}]_{k^*} = \gamma^{(m_{k^*}^{tmp})}$
7: **If** $\boldsymbol{\gamma}^{tmp} \in \mathcal{Q}$
8: Set $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{tmp}$, Go to Step 4.
9: **else**
10: Set $\boldsymbol{\gamma}$, compute $\mathbf{p}$ by (4.24), and Stop. *Convergence*
11: **End**
12: **else**
13: $\mathcal{K}^{tmp} = \mathcal{K}^{tmp} - \{k^*\}$, Go to Step 4.
14: **End**
15: **else**
16: Compute k^* by (4.36)
17: $\Xi(k^*, \mathcal{K}, \mathbf{V}, \mathbf{z}, \mathbf{p}, \boldsymbol{\gamma})$, $\gamma_k = \gamma^{(m_k=1)}\ \forall k \in \mathcal{K}$, Go to Step 2.
18: **End**

An alternative algorithm is developed where the infeasibility of $\boldsymbol{\gamma}$ is determined through the characteristics of $\mathbf{p}$. In [70] it was shown that if the targets are feasible, the link that consumes more power is the one that maximizes the PF-root in (4.30). The link with maximum power requirements compromises the most the resource feasibility. In order to identify the link k^* that has the maximum demand of power, it is used an algorithm that exploits the characteristics of the power vector computed by the Distributed Power Control (DisPC) algorithm [115]. DisPC has been used as part of algorithms that find the optimum power allocation in scenarios where $\boldsymbol{\gamma} \in \mathbb{R}_{++}^{|\mathcal{K}|}$ is a feasible target vector (e.g., [69, 70]). It has been proved that DisPC is a fixed point algorithm [112] that converges to the optimum unique $\mathbf{p}$ always that $\boldsymbol{\gamma}$ is feasible.

If the SINR targets $\boldsymbol{\gamma}$ are infeasible then the PF-root of the matrix $\boldsymbol{\Gamma}\mathbf{V}$ will be greater than 1. Under such conditions for a large number of iterations (t) of the DisPC algorithm, the powers approximate to [118]:

$$\mathbf{p}(t) \approx \rho(\boldsymbol{\Gamma}\mathbf{V})^t C\mathbf{x}, \tag{4.39}$$

where $\mathbf{x}$ is the right eigenvector associated with $\rho(\boldsymbol{\Gamma}\mathbf{V})$ and C is a constant depending on the initial vector $\mathbf{p}(t=1)$ and the coupling matrix $\mathbf{V}$. The power vector goes to infinity as the number of iterations (t) increases due to the fact that $\rho(\boldsymbol{\Gamma}\mathbf{V}) > 1$. As the objective is to allocate the maximum SINR target for all links ($\gamma_k = \gamma^{(M)}, \forall k$), this initial conditions may not be feasible leading to (4.39). This implies that the power vector found by DisPC cannot used directly to select the link that consumes more power. Nevertheless, if the powers are normalized each iteration so that $\|\mathbf{p}(t)\|_1 = 1$, then it can be found within a finite number of iterations τ, a link k^* whose associated power is maximum, i.e., $p_{k^*}(\tau) > p_k(\tau), \forall k \neq k^*$. If feasibility conditions are violated it is required to modify either the set of active links $\mathcal{K}$ or the target γ_{k^*} so that the power consumption of k^* is reduced in the next iteration. The initial conditions $t = 1$ of the vector of normalized powers $\check{\mathbf{p}}$ of the DisPC algorithm are given by the constrained power vector in (4.28), i.e., $\check{\mathbf{p}}(1) = \hat{\mathbf{p}}$. The DisPC algorithm with normalized powers is described in Algorithm 4.5.

Algorithm 4.5 DisPC Algorithm with normalized powers

1: **repeat**
2: $\quad \check{\mathbf{p}}(t+1) = \mathbf{\Gamma V}\check{\mathbf{p}}(t) + \mathbf{z}$
3: $\quad \check{\mathbf{p}}(t+1) = \check{\mathbf{p}}(t+1)/\|\check{\mathbf{p}}(t+1)\|_1$
4: **until** $\|\check{\mathbf{p}}(t+1) - \check{\mathbf{p}}(t)\|_1 \leq \varepsilon$ **or** $t < \tau$

Once that the Algorithm 4.5 has finished, the link that provides more information about resource feasibility is given by the maximum element in the normalized power vector $\check{\mathbf{p}}$:

$$k^* = \arg\max_{k\in\mathcal{K}} \check{p}_k. \tag{4.40}$$

The power vector found by (4.40) does not consider the power constraints (4.25) or (4.26) and they are taken into account once that there exists a power vector defined by (4.24) such that $\mathbf{p} \in \mathbb{R}_{++}^{|\mathcal{K}|}$ which corresponds to a feasible target allocation where no power constraints are imposed. Once that $\mathbf{p} > \mathbf{0}$, the criterion to select the link k^* considering the power constraints is given by:

$$k^* = \begin{cases} \arg\min\limits_{k\in\mathcal{K}} (\bar{p}_k - p_k)/\bar{p}_k & \text{if } \mathcal{P} = \mathcal{P}^{IC}, \\ \arg\max\limits_{k\in\mathcal{K}} p_k & \text{if } \mathcal{P} = \mathcal{P}^{TC}, \end{cases} \tag{4.41}$$

where for IPC, k^* indicates which link consumes more power regarding its individual power constraint. Finding the infeasible link k^* using the DisPC implies a two-time scale algorithm, and two process of optimization over the vector of powers. The first one is finding a $\mathbf{p}$ whose components are in $\mathbb{R}_{++}$ and the second is adjusting $\mathbf{p}$ according the power constraints. The steps to perform the minimum power consumption (MPC) optimization are given in Algorithm 4.6.

4.3.5 Target-to-SINR Ratio based Optimization

An alternative to (4.40)-(4.41) is obtained by using a metric that takes into account the power constraints each iteration. For a given set $\mathcal{K}$ and its respective $\boldsymbol{\gamma}$ and $\mathbf{p}$ it is required to determine how far is the achievable SINR of each link from its associated target. To calculate such a distance, define the function that computes the target-to-SINR ratio ($\gamma_k/SINR_k$) for the kth link as:

$$\psi_k(\boldsymbol{\gamma}, \mathbf{V}, \mathbf{z}, \mathbf{p}) = \frac{[\boldsymbol{\gamma}]_k([\mathbf{Vp}]_k + [\mathbf{z}]_k)}{[\mathbf{p}]_k}. \tag{4.42}$$

In order to maximize the total sum rate, the initial conditions of the SINR targets are set to the maximum modulation available $\gamma_k = \gamma^{(m_k=M)}, \forall k \in \mathcal{K}$. If for such $\boldsymbol{\gamma}$ the power vector $\mathbf{p}$ computed by (4.24) is in the feasible power region $\mathcal{P}$ thus $\boldsymbol{\gamma}$ is feasible. Otherwise, the link k^* with the worst target-to-SINR ratio is given by:

$$k^* = \arg\max_{k\in\mathcal{K}} \psi_k(\boldsymbol{\gamma}, \mathbf{V}, \mathbf{z}, \hat{\mathbf{p}}), \tag{4.43}$$

where $\hat{\mathbf{p}}$ is given by (4.28). The steps performed to solve problem (4.27) using both the MPC approach defined by (4.40)-(4.41) and the target-to-SINR ratio (TSR) approach given by (4.43) are described in Algorithm 4.6.

Algorithm 4.6 MPC/TSR based Optimization

1: Set initial values: $\mathcal{K} = \bar{\mathcal{K}}$, $\gamma_k = \gamma^{(m_k=M)}, \forall k \in \mathcal{K}$.
2: Evaluate $\mathbf{p}$ by (4.24)
3: **If** $\mathbf{p} \in \mathbb{R}_{++}^{|\mathcal{K}|}$
4: **If** $\mathbf{p} \in \mathcal{P}$
5: Set final vectors $\mathbf{p}$, $\boldsymbol{\gamma}$, and Stop.
6: **else**
7: Compute k^* by (4.41) or (4.43), Go to Step 11.
8: **End**
9: **else**
10: Compute k^* by (4.40) or (4.43)
11: **If** $m_{k^*} > 1$
12: $m_{k^*} = m_{k^*} - 1$, $\gamma_{k^*} = \gamma^{(m_{k^*})}$, Go to Step 2.
13: **else**
14: $\Xi(k^*, \mathcal{K}, \mathbf{V}, \mathbf{z}, \mathbf{p}, \boldsymbol{\gamma})$, Go to Step 2.
15: **End**
16: **End**

4.3.6 SINR Target Increment based Optimization

The objective of Algorithm 4.3 and Algorithm 4.6 is to reach a point $\boldsymbol{\gamma}$ in the region of targets $\mathcal{Q}$, ideally on its boundary. Both approaches start with initial conditions that may be out of the feasible region of targets. The number of iterations required for the algorithms to converge depends on the channel realization. When the channel conditions are not favorable many links will achieve low SINR, which implies that the number of iterations required for the algorithms to converge may be large. In order to find a faster way to compute the solution of (4.27) an algorithm that operates in two stages is designed. In the first stage, it finds a target vector $\boldsymbol{\gamma}$ inside the feasible target region. In the second stage, it attempts to reach the closest point to the boundary of the feasible target region from inside out by tightening up the SINR requirements of specific links.

Define $\tilde{\boldsymbol{\Gamma}}$ as the matrix of achievable SINR targets under a given set of power constraints $\hat{\mathbf{p}}$ (4.28) as follows:

$$\tilde{\boldsymbol{\Gamma}} = [diag(\mathbf{V}\hat{\mathbf{p}} + \mathbf{z})]^{(-1)} diag(\hat{\mathbf{p}}). \tag{4.44}$$

Since the elements of the diagonal of $\tilde{\boldsymbol{\Gamma}}$ take values in $\mathbb{R}$, it is necessary to adjust such values to be in the set $\mathcal{M}$ of available SINR targets. The appropriate target value of the kth link lies in one of the SINR target ranges defined by $\mathcal{M} = \{\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(M)}\}$ [50]:

$$\tilde{\gamma}_k \in [\gamma^{(m)}, \gamma^{(m+1)}). \tag{4.45}$$

Therefore, the adjusted target of the kth link over its target interval is defined as:

$$\ddot{\gamma}_k = \min\{\gamma^{(m)}, \gamma^{(m+1)}\}, \tag{4.46}$$

where the target $\ddot{\gamma}_k$ is related to the highest spectral efficiency, i.e., the rate of the selected target (MCS) is the closest but below the achievable capacity of the SINR $\tilde{\gamma}_k$ [50, 106]. The values of the achievable SINR targets from (4.44) are adjusted by (4.46) and the links whose adjusted $\ddot{\gamma}$ are not in the set of available targets $\mathcal{M}$ are dropped. Notice that the computation of $\tilde{\boldsymbol{\Gamma}}$ uses the maximum available power vector $\hat{\mathbf{p}}$ and this lack of initial power control may discard feasible links that could be scheduled.

Once that a target vector $\ddot{\boldsymbol{\gamma}}$ with all its elements in $\mathcal{M}$ and its associated power vector $\mathbf{p} \in \mathcal{P}$ has been defined, the links in $\mathcal{K}$ remain fixed and the first stage is completed. The second stage of the algorithm is focused on tightening up the components of $\ddot{\boldsymbol{\gamma}}$ in order to maximize the final sum rate maintaining resource feasibility. The problem of target increment for a fixed subset of links was addressed in Section 4.3.3.2 using a criterion derived from the Perron Frobenius theory. This criterion is used to define the candidate link k^* that can increase its SINR target and it requires the evaluation of the PF-root of $\mathbf{B}$. In order to avoid the eigenvalue computation such criterion can be simplified by using the minimum power consumption instead. This is intuitive since $\ddot{\boldsymbol{\gamma}}$ and its associated power vector are feasible, and the latter can be used to decide which link can increase its power consumption. The link k^* candidate to tighten its SINR target up is given by:

$$k^* = \begin{cases} \arg\max\limits_{k \in \mathcal{K}} \ (\bar{p}_k - p_k)/\bar{p}_k & \text{if } \mathcal{P} = \mathcal{P}^{IC}, \\ \arg\min\limits_{k \in \mathcal{K}} \ p_k & \text{if } \mathcal{P} = \mathcal{P}^{TC}. \end{cases} \tag{4.47}$$

The steps performed to solve problem (4.27) using the SINR target increment (STI) approach are described in Algorithm 4.7. Notice that more than one link can be discarded in Steps 3 and 4, but $\ddot{\boldsymbol{\Gamma}}$ is not guaranteed to be feasible. In such a case and in order to reduce outage, Steps 21 and 22 are used to drop only one extra link whose associated adjusted target $\ddot{\boldsymbol{\gamma}}$ is minimum.

Algorithm 4.7 SINR Target Increment (STI) based Optimization

1: Set $\mathcal{K} = \bar{\mathcal{K}}$
2: Compute $\tilde{\boldsymbol{\Gamma}}$ by (4.44), and $\ddot{\gamma}_k$ by (4.46)
3: $\mathcal{K}' = \{k : \ddot{\gamma}_k \notin \mathcal{M}\}$
4: $\Xi(k', \mathcal{K}, \mathbf{V}, \mathbf{z}, \hat{\mathbf{p}}, \ddot{\boldsymbol{\gamma}}), \ \forall k' \in \mathcal{K}'$
5: Compute $\mathbf{p}$ by substituting $\ddot{\boldsymbol{\Gamma}} = diag(\ddot{\boldsymbol{\gamma}})$ in (4.24)
6: **If** $\mathbf{p} \in \mathcal{P}$
7: Define $\mathcal{K}^{tmp} = \mathcal{K}$
8: Compute k^* over $\mathcal{K}^{tmp}$ by (4.47)
9: **If** $m_{k^*} < M$
10: $m_{k^*}^{tmp} = m_{k^*} + 1$, $\boldsymbol{\gamma}^{tmp} = \ddot{\boldsymbol{\gamma}}$, $[\boldsymbol{\gamma}^{tmp}]_{k^*} = \gamma^{(m_{k^*}^{tmp})}$
11: Compute $\mathbf{p}^{tmp}$ by using $\boldsymbol{\Gamma}^{tmp} = diag(\boldsymbol{\gamma}^{tmp})$ in (4.24)
12: **If** $\mathbf{p}^{tmp} \in \mathcal{P}$
13: Set $\ddot{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^{tmp}$, Go to Step 8.
14: **else**
15: Set $\boldsymbol{\gamma} = \ddot{\boldsymbol{\gamma}}$, compute $\mathbf{p}$ by (4.24), Stop.
16: **End**
17: **else**
18: $\mathcal{K}^{tmp} = \mathcal{K}^{tmp} - \{k^*\}$, Go to Step 8.
19: **End**
20: **else**
21: $k^* = \min\limits_{k \in \mathcal{K}} [\ddot{\gamma}]_k$
22: $\Xi(k^*, \mathcal{K}, \mathbf{V}, \mathbf{z}, \hat{\mathbf{p}}, \ddot{\boldsymbol{\gamma}})$, Go to Sep 2.
23: **End**

4.4 Performance Evaluation for a DAS Scenario

In order to assess the algorithms a DAS scenario as the one depicted in Fig. 4.1(b) is considered. These type of access interfaces are based on the concept of space diversity and cell splitting in order to improve coverage and spectral efficiency. The deployment of the distributed antennas consists of N RAUs, one at the center of the cell and $N-1$ distributed RAUs uniformly deployed at a distance of $\frac{2}{3}$ the cell radius from the cell center. Consider that the RAUs are coordinated only to control their transmit powers and no signal processing (e.g., beamforming) is used. The channels are modeled as Rayleigh fading and are affected by a path-loss component and a shadowing fading component modeled as a log-normal distributed variable with parameter s_σ.

There are K users uniformly deployed in the cell and $K \geq N$. Therefore, it is necessary to select the initial subset of links $\bar{\mathcal{K}}$ by assigning a different user to each RAU which is accomplished by the RAU-user matching Algorithm 4.1a - 4.1b defined in Section 4.2.2. The algorithm assigns one different user to each RAU deployed within the cell defining in this way the initial set of links $\bar{\mathcal{K}}$. After all RAU-user links have been established, the proposed algorithms find the final subset of feasible links $\mathcal{K}$ and allocate powers and rates. The set of available targets $\mathcal{M}$ is the one presented in Table 4.3. Results are generated by averaging 10e3 channel instances for each value of K and the simulation parameters are listed in Table 4.1.

4.4.1 Examples of the MCS selection

In this subsection two illustrative examples of how the algorithms select the MCSs are presented. For the first example, consider the two-user system depicted in Fig. 4.1(a). The channel gains are given by $G_{11} = 0.8791$, $G_{12} = 0.3999$, $G_{21} = 0.0211$, $G_{22} = 0.8791$, the power constraint is $P_t = 1.4$, the noise power is $\sigma_1^2 = \sigma_2^2 = 10^{-2}$ and the set of available SINR targets ($\mathcal{M}$) is defined in Table 4.3 [50].

Table 4.3: Set of available SINR Targets $\mathcal{M}$ in (dB) and its associated R in (bps/Hz)

Index m	1	2	3	4	5	6	7	8
$\gamma^{(m)}$	-3.2	1.8	5.0	7.2	11.2	14.8	19.0	22.8
$R(\gamma^{(m)})$	0.333	1	1.5	2	3	4	5.14	6.4

Fig. 4.8 shows the evolution of the proposed algorithms over the region of available targets for a scenario with two users considering (4.26). The starting point $\gamma_k = \gamma^{(M)}$ is infeasible for Algorithm 4.3 (PFR) and Algorithm 4.6 (MPC/TSR) and the feasible target vectors are found by different paths. The algorithms stop once that $\mathbf{p} \in \mathcal{P}^{TC}$ or $\boldsymbol{\gamma} \in \mathcal{Q}^{TC}$.

For the second example, consider a scenario with seven active links in $\bar{\mathcal{K}}$ and (4.26) is imposed. Fig. 4.9 shows the evolution of the algorithms for one link k. Since Algorithm 4.3 (PFR) and Algorithm 4.6 (MPC/TSR) start in the maximum available SINR target only target relaxation is performed. However, it is possible that their curves go up reaching the maximum SINR target value. This indicates the dropping of an infeasible link and that the algorithms attempt to allocate the maximum SINR target to the remaining users. In this particular example Algorithm 4.6 either by MPC (4.40)-(4.41) or by TSR (4.43) has one dropping event, whilst Algorithm 4.3 (PFR) finds a feasible $\boldsymbol{\gamma}$ for all links in $\bar{\mathcal{K}}$. This means that the final solution of the algorithms may differ in the number of active links. Algorithm 4.7

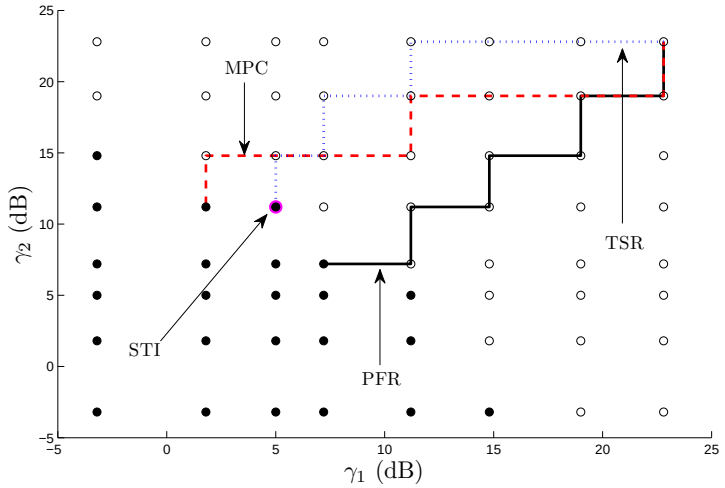

Figure 4.8: Example of the search of the maximum vector of targets $\boldsymbol{\gamma}$.

(STI) starts from a given target value and its objective is to increase such target as much as possible.

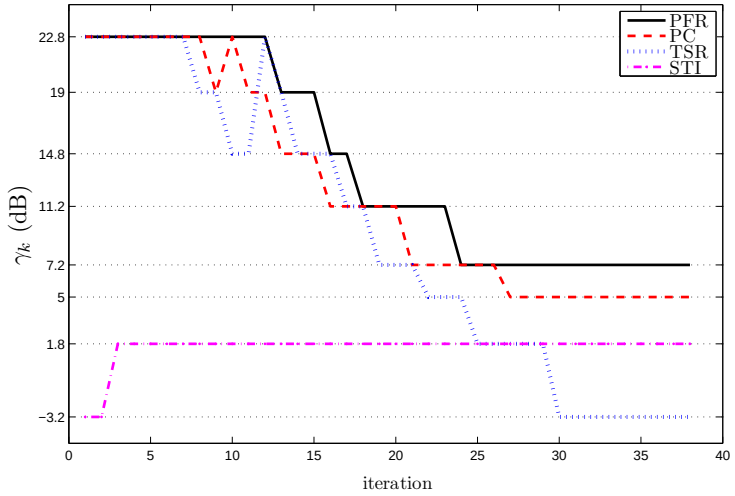

Figure 4.9: MCS selection based on target relaxation and target increment.

4.4.2 PF-root based optimization: Fixed vs Non-Fixed Set

Fig. 4.10 presents the CDF of the sum rate for $K = 8$. Two particular case are shown: (a) for a fixed sum rate target of 15bps, the gap between the proposed algorithms and the optimal allocation is less than 10% which is an acceptable approximation considering all available combinations of links subsets and targets ($|\Omega_{|\bar{\mathcal{K}}|=4,M=8}| = 6560$). (b) for a required probability $Pr = 0.3$, the sum rate gap is less than 10% for both proposed algorithms when

compared to the optimum allocation. The *outage probability* has been used as a metric to assess the performance of different power control algorithms, and it can be defined as the ratio between the number of dropped links to the total number of links [38]. The performance in terms of outage probability is presented in Fig. 4.11. For a fixed outage probability target of 10^{-3} it can be observed that the approach used by Algorithm 4.4 (PFRIT) exploits more efficiently user diversity compared to the approach of Algorithm 4.3 (PFR). Algorithm 4.4 (PFRIT) can be used to minimize outage since it requires $K > 8$ to make $|\mathcal{K}| = N$ and it achieves a performance similar to the optimal allocation for fixed subset $\mathcal{K}$. In contrast, Algorithm 4.3 (PFR) requires $K > 28$ to achieve zero outage. Notice that the optimum allocation for a non-fixed $\mathcal{K}$ requires $K \gg N$ to achieve zero outage. The sum rate deficiency of Algorithm 4.4 (PFRIT) compared to Algorithm 4.3 (PFR) is compensated in terms of fairness since $|\mathcal{K}|$ is also maximized.

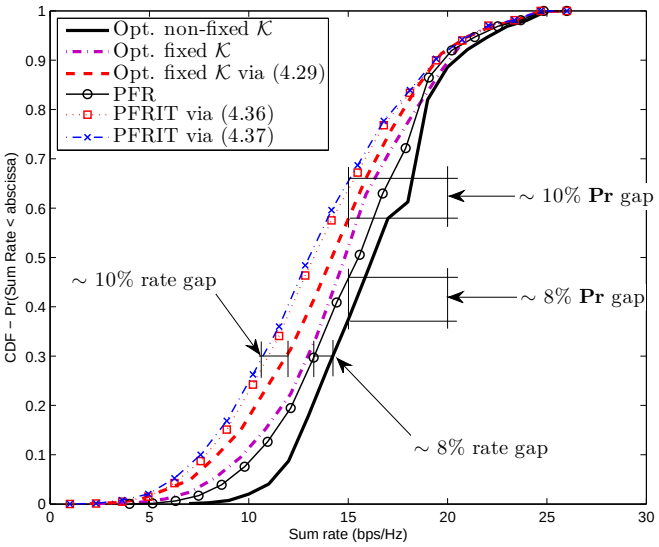

Figure 4.10: Cumulative Distribution Function of the average sum rate for $K = 8$.

4.4.3 Sum Rate and Outage Probability

In order to assess the performance of the algorithms the statistics of the sum rate and outage probability are employed for different DAS scenarios and both set of power constraints (4.25) and (4.26). As K grows, the channel conditions of the users attached to the RAUs are improved and interference can be mitigated more efficiently. Fig. 4.12 and Fig. 4.13 show the total sum rate as a function of the total number of users K considering $N = 7$.

Consider two particular cases of K. For the first case let $K = N = 7$, where in Fig. 4.12 the achieved sum rate of the three algorithms is similar when TPC (4.26) is imposed. The sum rate maximization via PF-root optimization in Algorithm 4.3 (PFR) identifies with more accuracy which link must relax its target which result in an extra gain of 0.73bps/Hz compared to the other approaches. Algorithm 4.7 (STI) achieves the same performance of Algorithm 4.6 (MPC/TSR) since it allocates higher MCSs to its links. However, its performance in terms of outage probability is worst than the other approaches since it may discard feasible links in

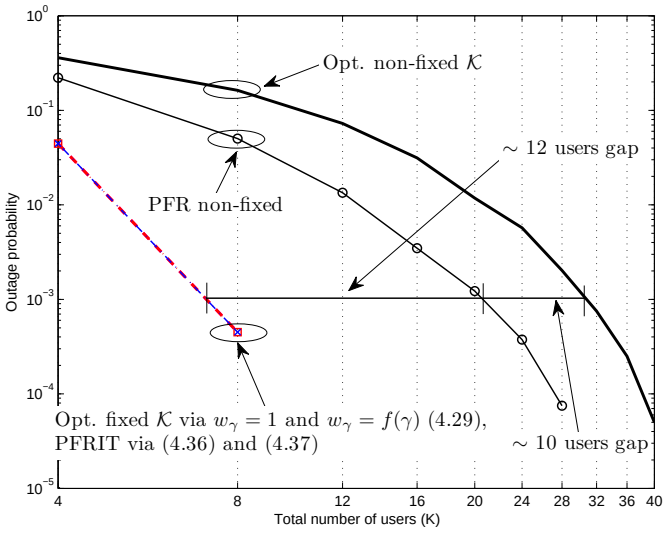

Figure 4.11: Outage probability versus user diversity K.

---

its first stage. In Fig. 4.13, (4.25) is considered and the performance of Algorithm 4.7 (STI) and Algorithm 4.6 via TSR (4.43) are outperformed by Algorithm 4.6 via MPC (4.40)-(4.41), which indicates that when individual power constraints are imposed, target relaxation based on power consumption achieves an extra gain of 0.95bps/Hz compared to the optimization based on the target-to-SINR ratio.

The second case of analysis, $K = 70 \gg N$, provides a rich MUDiv and the performance of all algorithms reach similar values of sum rate for both sets of power constraints. This means that under favorable channel conditions, finding an acceptable solution to (4.27) can be achieved with the low complex Algorithm 4.6 (MPC/TSR) and Algorithm 4.7 (STI) avoiding the PF-root optimization.

The outage probability results are displayed in Fig. 4.14 and Fig. 4.15 for TPC and IPC respectively. If TPC is considered, Algorithm 4.6 (MPC/TSR) has a marginal gap of outage probability compared to Algorithm 4.3 (PFR). In contrast, if IPC is imposed, only Algorithm 4.6 via MPC (4.40)-(4.41) achieves an outage probability closed to the one of Algorithm 4.3 (PFR). In spite of the sum rate achieved by Algorithm 4.7 (STI), it suffers form a larger outage compared to the other schemes. The main advantage of Algorithm 4.7 (STI) is that the number of required iterations for convergence is considerable less compared to the other approaches. Fig. 4.16 shows the average number of iterations required by the algorithms to converge for IPC and similar results are obtained for TPC. If the number of users in the cell is low $K \approx N$, that means that the N RAUs must serve users with worst channel conditions compared to the case where $K \gg N$. In other words, when channel conditions are poor, the feasible target region contains few feasible target vectors making that Algorithm 4.3 and Algorithm 4.6 require more iteration to find a solution to (4.27). This phenomenon is illustrated in Fig. 4.8.

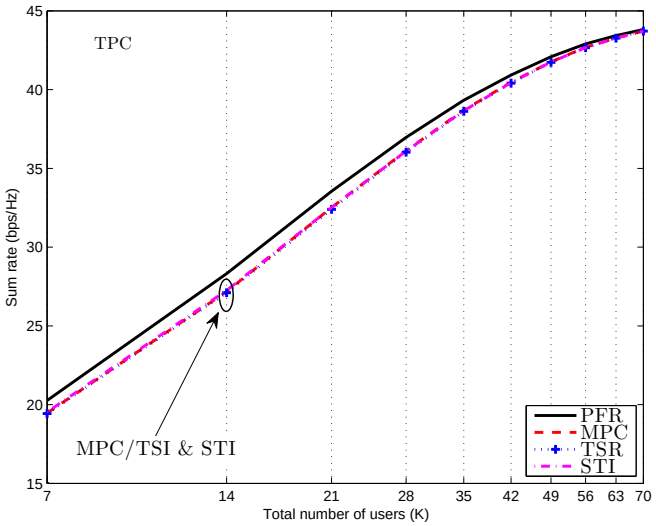

Figure 4.12: Total sum rate as a function of K for $N = 7$, $M = 8$, and $P_t = 43$(dBm).

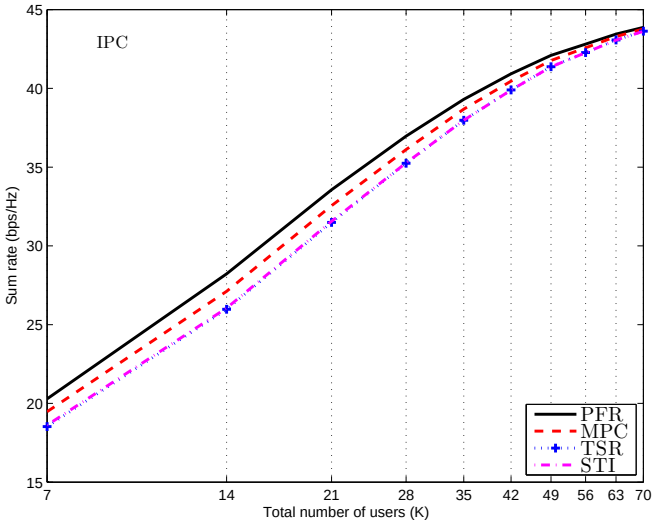

Figure 4.13: Total sum rate as a function of K for $N = 7$, $M = 8$, with $\bar{p}_1 = 39$(dBm) for the central RAU, and $\bar{p}_j = 33$(dBm) for $\forall j \neq 1$, the rest of the RAUs.

4.4.4 Optimal Joint Link Selection and Resource Allocation

In order to quantify the sum rate gap between the proposed algorithms and the optimal solution to (4.27) consider an scenario with $N = 4$, $K = 10$, and the optimal solution is found by exhaustive search. For this particular case the search space contains $|\Omega_{|\bar{\mathcal{K}}|=4,M=8}| = 6560$ combinations. Fig. 4.17 and Fig. 4.18 show the CDF of the sum rate considering TPC and IPC respectively. Algorithm 4.3 and Algorithm 4.6 look for the maximum SINR target vector on the boundary of region of feasible targets $\mathcal{Q}$. The target vector $\boldsymbol{\gamma}$ found by both approaches

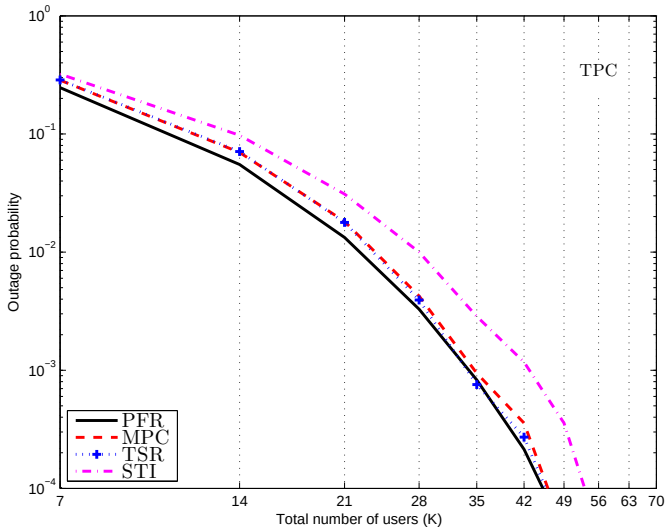

Figure 4.14: Outage probability as a function of K for $N = 7$, $M = 8$, and $P_t = 43$(dBm).

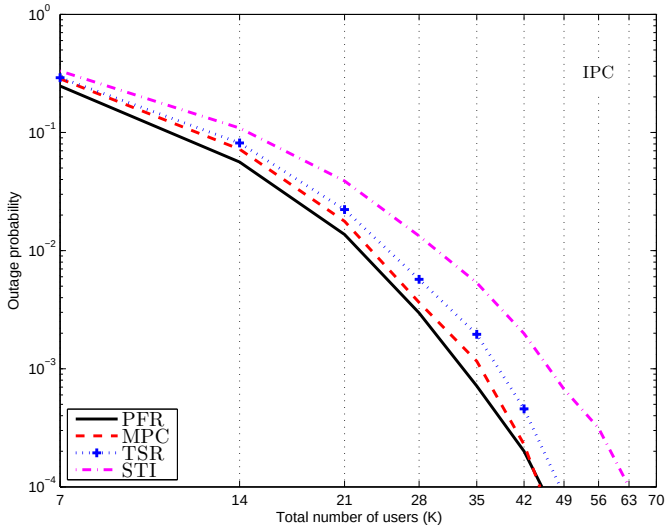

Figure 4.15: Total sum rate as a function of K for $N = 7$, $M = 8$, with $\bar{p}_1 = 39$(dBm) for the central RAU, and $\bar{p}_j = 33$(dBm) for $\forall j \neq 1$, the rest of the RAUs.

---

is not necessary the same size of the optimal one $\boldsymbol{\gamma}^\star$. Since the optimum solution maximizes the sum rate over all combinations in $\Omega_{\bar{\mathcal{K}},M}$ and in Algorithm 4.3 (PFR) and Algorithm 4.6 (MCP/TSR) the stop criteria are triggered once that feasible vectors $\boldsymbol{\gamma}$ or $\mathbf{p}$ have been found. Such criterion to stop the algorithms, maximizes the cardinality of the final set $\mathcal{K}$. The optimum solution of (4.27) will present a larger outage specially when $K \approx N$.

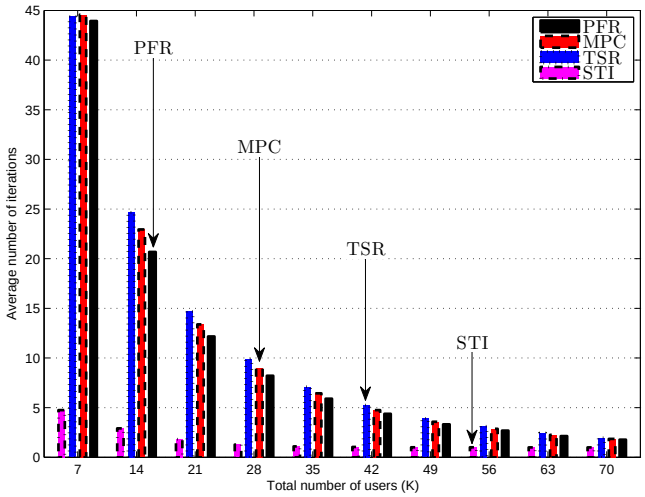

Figure 4.16: Average number of iterations vs K for $N = 7$, $M = 8$, and any definition of $\mathcal{P}$.

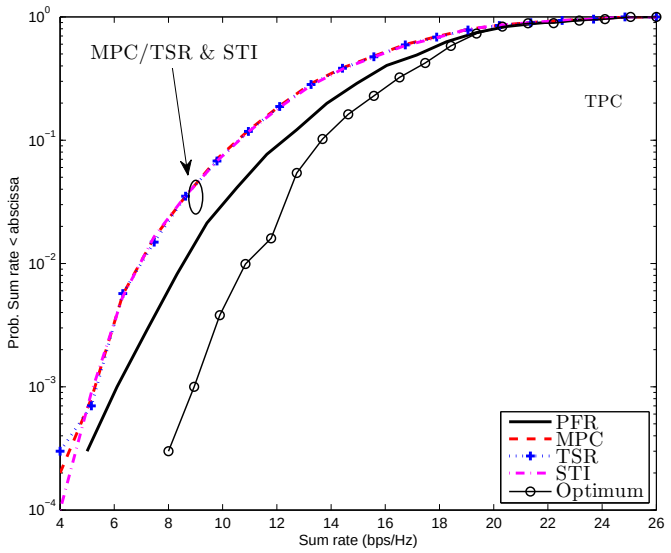

Figure 4.17: CDF of the sum rate for $N = 4$, $M = 8$, $K = 10$, and $P_t = 36.9$(dBm).

4.4.5 Performance Evaluation for the Low-High SINR Regimes

The transition from a noise limited to an interference limited system with $N = 7$, $K = 14$, is displayed in Fig. 4.19 and Fig. 4.20 for TPC and IPC respectively. For IPC, the central RAU has a power constraint $\bar{p}_1$ given by the abscissa and the distributed nodes have $\bar{p}_j = 0.6\bar{p}_1, \forall j \neq 1$ in watts. For both sets of power constraints Algorithm 4.3 (PFR) and Algorithm 4.6 (MCP/TSR) obtain the same outage probability when the system is noise limited. For systems that operate with low transmission powers the resource allocation process can be given by Algorithm 4.6 since its complexity is low compared to Algorithm 4.3 and its

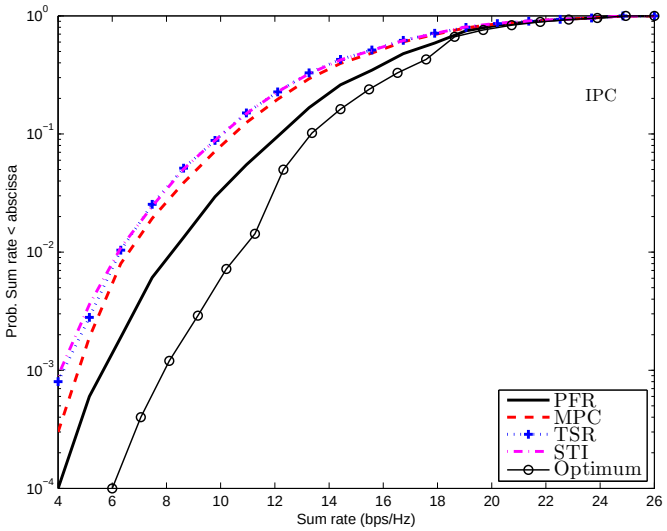

Figure 4.18: CDF of the sum rate for $N = 4$, $M = 8$, $K = 10$, with $\bar{p}_1 = 39$(dBm) for the central RAU, and $\bar{p}_j = 36$(dBm) for $\forall j \neq 1$, the rest of the RAUs.

---

outage performance is similar. When the system is interference limited, the outage probability gap between Algorithm 4.3 (PFR) and Algorithm 4.6 (MCP/TSR) is significant. Nevertheless, Algorithm 4.6 and Algorithm 4.7 (STI) are an alternative to the PF-root optimization approach. Notice that in the interference limited scenario the rate is upper bounded by the rate associated with the maximum target (MCS) in $\mathcal{M}$, specially when there are favorable channel conditions, e.g. $K = 70$ in Fig. 4.12 and Fig. 4.13.

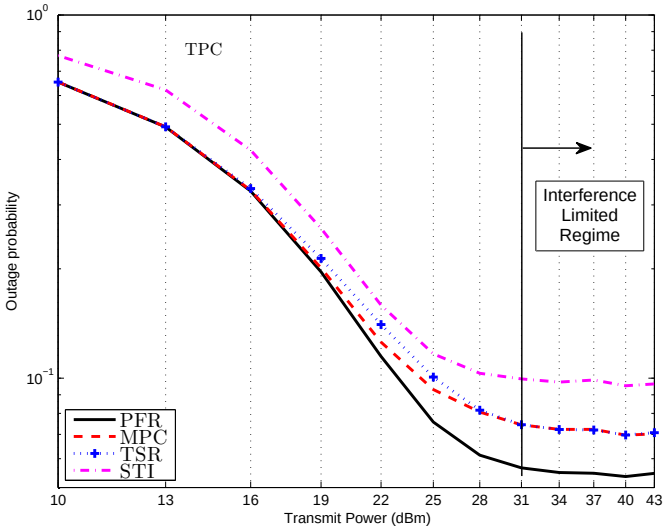

Figure 4.19: Outage probability vs the constrained transmit power P_t for $N = 7$ and $M = 8$.

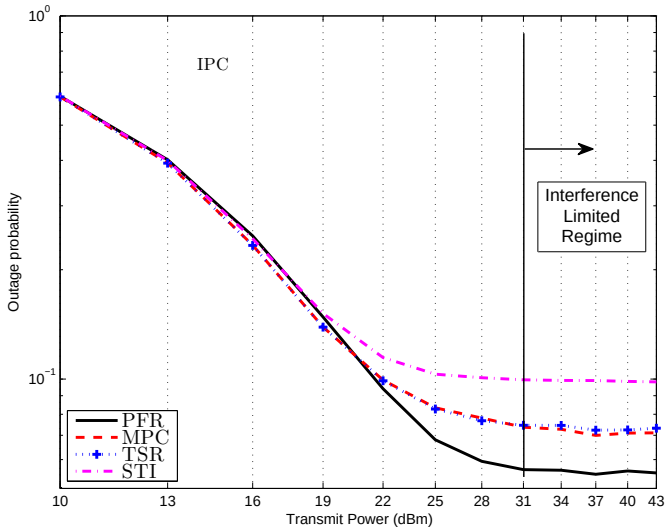

Figure 4.20: Outage probability vs the constrained transmit powers $\bar{\mathbf{p}}$ for $N = 7$ and $M = 8$.

4.4.6 Application for User-Removal

The proposed algorithms have the potential to be used as removal algorithms in the context used in [38], i.e., all links in $\bar{\mathcal{K}}$ have the same unique fixed target and the objective is to find the subset $\mathcal{K}$ with the links that can be simultaneously scheduled. This can be achieved by making $\mathcal{M} = \{\gamma^{(m)}\}$ for any fixed m, which is a particular case of the proposed algorithms. Fig. 4.21 shows the outage probability when a unique fixed target is considered. For this particular case Algorithm 4.3 (PFR) reduces to the Removal Algorithm III-A in [38]. It is worthy to point out that for TPC, Algorithm 4.6 (MCP/TSR) is a low complex alternative to the Removal Algorithm III-A [38] which was claimed to be the unique solver for the case where (4.26) is imposed. The performance of the algorithms is compared to the optimal removal. The outage probability gap between Algorithm 4.6 and the optimal removal is negligible for low values of the fixed target. For large values of the fixed target this gap reflects the inaccuracy of Algorithm 4.6 when selecting the worst link in $\mathcal{K}$ either by the power consumption or by the target-to-SINR ratio criterion. However, this gap is a trade-off given by the low complexity involved in the user-removal process and the accuracy of the selection. The curve from Algorithm 4.7 (STI) is a clear example of the expected outage when more than one link is discarded per iteration. Similar results to Fig. 4.21 are achieved when (4.25) is considered.

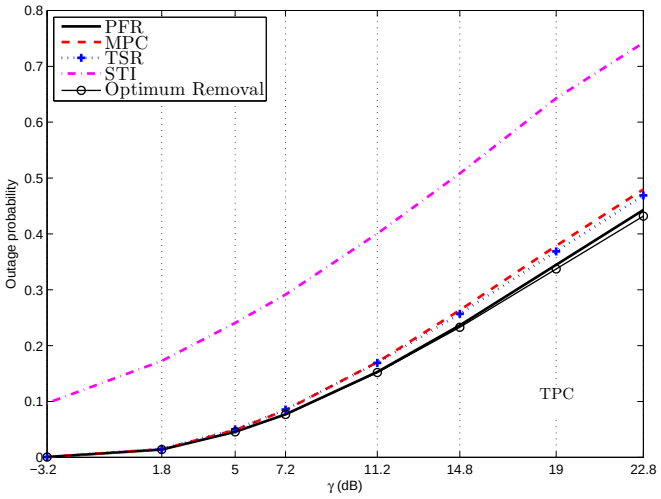

Figure 4.21: Outage probability for a fixed SINR target for all links with $N = 7$ and $P_t = 28.5$(dBm).

4.5 Conclusions

In this chapter the problem of sum rate maximization in interference coupled systems was addressed by solving two subproblems: problem one is the user-antenna matching problem, and problem two is the rate and power allocation for rate maximization. For problem one in Section 4.2, it was designed an algorithm that operates in two phases, one that combines a greedy matching and a minimum-rate-loss matching which provides a first set of potential users. The second phase seeks a refined subset of users in order to maximize the total sum rate. Numerical results show that the proposed matching method can achieve the performance of more complex matching solvers.

For the second problem in Section 4.3, different algorithms were proposed to solve the sum rate maximization problem. This problem has a combinatorial nature since the SINRs are constrained to take values from a finite set. The algorithms find a set of links for which exists a feasible resource allocation that attempts to maximize the sum rate. The proposed algorithms are based on criteria derived from the Perron-Frobenius theory or from the implicit information contained in the power consumption or achievable SINR. Furthermore, a low-complexity fast algorithm was designed for link selection and resource allocation. Such algorithm converges in few iterations and its robustness allows us to achieve acceptable performance under favorable channel conditions.

Numerical results show that the proposed algorithms achieve a good trade-off between complexity and accuracy compared to the optimal solution. The low-complexity algorithms that avoid the PF-root computations are suitable for scenarios with favorable channel conditions (e.g., rich MUDiv). In such scenarios the attainable sum rate is similar to the one achieved by approaches depending on the eigenvalue optimization. It was shown that the proposed algorithms can be used for user-removal under different sets of power constraints whose performance is closed to the one achieved by state-of-the-art algorithms but with significant reduction on computational complexity.

An additional application of the algorithms is to group users for time-sharing scheduling

where transmission is provided only to useful users that can be jointly supported whilst useless users can be served in a later time-slot or handed to another channel or base station.

Chapter 5

User Selection and Signal Design In Single Cell Systems

In this chapter, the sum rate maximization problem in single-cell multiuser MISO broadcast channels is addressed. In the analyzed system, the transmitter can design beamforming weights (Zero-Forcing based) that can remove inter-user interference which scales the system capacity according to the number of transmit antennas. If the number of users is larger than the number of transmit antennas, only as set of users can be scheduled simultaneously and user selection is required. The achievable rates have a bijective relationship with the effective channel gains, i.e., the magnitude of the channels after beamforming [Ch1]. This relation allows to improve the achievable rates when the effective channel gains are optimized. Such an optimization depends on the characteristics of the scheduled users channels, i.e., the magnitude and the spatial compatibility (orthogonality). The sum rate maximization is a mixed optimization problem that operates over two domains: the power allocation and the set of users [J3]. By decoupling those tasks, the user selection can be performed based on channel metrics that indirectly evaluate the effective channel gains. This is relaxation of the original problem where a set of channels optimize a metric-based global objective function. Once that a set of users optimizes a given metric, power allocation is performed by standard convex optimization techniques. This chapter presents a qualitative and quantitative analysis of the metrics that can be used to perform user selection [C5]. Different trade-offs between performance optimization and computational complexity are attained by the proposed optimization algorithms.

5.1 Introduction

The multiuser multiple-input multiple-output (MU-MIMO) systems have a huge potential to attain high data rates in wireless systems [15, 133]. MIMO systems can exploit space-time coding and spatial multiplexing always that CSI is known at the transmitter which increases the overall system capacity. In the wireless scenario of interest (see Fig. 5.1), a transmitter encodes different information for different receivers in a common signal, which is referred in the literature as a Broadcast Channel (see Chapter 2.2). For a classic deployment with one BS equipped with N_t antennas and K single-antenna users, the capacity of a MIMO system increases by a factor of $\min\{N_t, K\}$ times the capacity of a time-division-multiple-access (TDMA) system if the transmitted signals are uncorrelated [15]. The TDMA system cannot exploit the multiple antenna deployment at the BS which leads to a waste of system resources and a degraded performance specially in the high SNR regime. The natural solution

to this problem is to transmit simultaneously to more than one user. A strategy to accomplish this goal is to implement DPC, a nonlinear coding scheme (multiplexing technique) based on coding of known interference [24]. The DPC exploits the full CSI achieving the same capacity of an interference free MIMO BC system [133]. If the number of single antenna users K is larger than N_t, DPC can achieve a linear capacity increase in N_t.

DPC is the optimal capacity maximization scheme in a MIMO BC system. However, it requires a huge computation complexity and feedback information, which rapidly increases with N_t. Two sub-optimal alternatives of DPC were proposed in [25]. The first alternative is the channel inversion or zero-forcing beamforming (ZFBF), which is an orthogonal transmit spatial multiplexing scheme whose main objective is to nullify the mutual interference among users. Despite its simplicity, ZFBF has been shown to achieve the same asymptotic sum capacity of DPC when K is large [40]. The second alternative called zero-forcing dirty paper (ZFDP) is an asymptotically optimal beamforming scheme that combines a QR decomposition of the channel matrix with DPC at the transmitter. In this ranked known interference scheme, the first user is not affected by interference while the second user is only affected by interference coming from the first user and so on.

The capacity maximization using ZFBF (e.g., [25, 40, 134–136]) or ZFDP (e.g., [25, 137, 138]) can be further improved in scenarios where the number of single antenna users is larger than the number of antennas at the BS ($K > N_t$). The users can be seen as an extra dimension of adaptation which is referred in the literature as multiuser diversity (MUDiv). In order to exploit such diversity, it is necessary to select a set of active users whose channel characteristics yield a performance improvement in terms of sum rate when they transmit simultaneously over the same radio resource. The user selection (scheduling) is a medium access control layer process that can use information from the adaptive physical-layer design so that temporal dimension (scheduling) and spatial dimension (multiple antennas) can be fully exploited. The scheduling is a real time process whose computational complexity and implementation affect directly the performance of upper-layers. Finding the set of users that optimizes a given global utility function is a combinatorial problem whose optimal solution is given by exhaustive search and its associated search space grows exponentially with the number of users. The computation of the optimal solution of the scheduling problem is prohibitive for most practical systems (moderate K and N_t). Therefore, it is necessary to find efficient suboptimal scheduling schemes that can provide a good trade-off between performance and complexity.

5.1.1 Related works

A considerable amount of work focused on the asymptotic sum rate of MIMO BC systems with user selection has been done over the last ten years. Several published works (e.g., [40, 134, 136, 137]) presented efficient suboptimal algorithms that attempt to overcome the prohibitively high complexity of exhaustively searching users. Most of the proposed algorithms sub-optimally solve the problem of sum rate maximization by implementing cross-layer designs, where the scheduling decisions are made based on instantaneous CSI or link-level metrics. Since the aforementioned problem can be tackled in different ways, hereafter, it is proposed a classification of the most common structures of the algorithms described in the literature. The classification is based on the methodology followed to solve the mixed convex and combinatorial problem of rate maximization in multiuser MIMO BC systems. This classification is used to make a clear distinction between the metrics used by each class and to

fairly compare the performance achieved by algorithms of different classes.

A class-A algorithm is the one that performs a joint user selection and power allocation optimization. A new user k is added to the set of previously selected users $\mathcal{S}$ only if for a given utility function $U(\cdot)$ the aggregation of k to $\mathcal{S}$ increases the value of the utility function, i.e., $U(\mathcal{S}) < U(\mathcal{S} + \{k\})$. This kind of greedy algorithms [123, 124, 138–141] are highly effective for sum rate maximization. However, they require a high computational processing since the user selection requires the evaluation of the global utility function. Such a evaluation requires both, water-filling power allocation and the Shannon capacity computation for each unselected user in every iteration of the algorithm.

A class-B algorithm operates in two phases. In the first phase a set of users is selected based on specific channel characteristics (metrics) and in the second phase the algorithm evaluates the global utility function for the previously defined set, e.g., [40, 134, 136, 137, 142]. This means that the user selection and the resource allocation (powers and beamforming weights) problems are carried out independently and the sum rate maximization depends upon the channel characteristics of the selected users. Furthermore, the cardinality of the set of selected users is fixed in the first phase and it may be modified during the second phase when the global utility function is evaluated. For instance, if water-filling is performed to evaluate the global utility function, this may result in zero power allocation for some selected users due to: their channel characteristics, the power constraints, and the SNR regime.

In [40] the authors designed a greedy algorithm that performs a semi-orthogonal user selection (SUS) in order to maximize the total sum rate implementing ZFBF. In this class-B algorithm the new selected user maximizes the component of the channel that is orthogonal to the subspace spanned by the channels of the previously selected users. The exact evaluation of that orthogonal component requires the multiplication of the unselected channel vectors by a projector matrix over the subspace defined by channels of the selected users. The authors of [40] showed that the average sum rate of ZFBF combined to their proposed user selection technique achieves asymptotically the average sum rate of DPC when the number of users grows to infinite ($K \to \infty$). Tu and Blum [137] proposed a class-B greedy algorithm for sum rate maximization and ZFDP. The metric for user selection is based on the channel component projected onto the null space spanned by the previously selected user channels. This metric is used to estimate the power degradation that a new user will experienced if it interacts with the orthogonal subspace spanned by the other selected users. A statistical analysis of this methodology was done in [138] where it was shown that the greedy user selection based on null space projections is a suboptimal yet highly efficient way to form groups of quasi-orthogonal users that sub-optimally maximize the sum rate. The main disadvantage of this approach lies in the computation of a projector matrix (null space projector) and its multiplication with the channel vectors of all unselected users each iteration of the algorithm. A similar approach to [137] was presented in [136] for sum rate maximization with ZFBF. The difference between these two approaches lies in the fact that the latter performs singular value decomposition (SVD) in order to compute null space of the selected user channels.

Both classes of algorithms require several matrix operations to perform the user selection. Class-A algorithms use matrix inversion in order to perform power allocation per each possible set of selected users. Class-B algorithms require the computation of a projector matrix [143] per iteration and a matrix inversion if the power allocation is based on water-filling. The effectiveness of the beamforming schemes is related to the spatial compatibility between the channels of the selected users. Therefore, class-B algorithms exploit channel metrics that quantitatively measure the spatial compatibility of the selected channels.

5.1.2 Contributions

In this chapter, several metrics that measure the spatial compatibility between users are proposed for Zero-Forcing based beamforming. The relationship between the metrics, their properties, and their applicability for users selection are investigated. A novel metric for spatial compatibility is proposed which is an approximation of the one used in [136, 137]. Such approximation has the advantage that it only requires multiplication of scalars defined by the correlation coefficient between two channels. Based on the studied metrics two user selection strategies are developed. The first strategy is a greedy class-B algorithm that makes scheduling decisions based on the null space projection approximations, i.e., information extracted from the channel norms and the orthogonality between channels. The sum of null space projections can be reformulated as the sum of individual weighted convex functions, and the second strategy for user selection is to model the rate maximization problem as an integer program. In contrast to some related works (e.g., [139]) that only provide a description of the user selection problem as an integer program (due to the high complexity of the problem formulation), a complete mathematical model is derived for the integer constrained program. The solution of the integer program (user set) asymptotically achieves the capacity the optimum user set for moderate values of K. The attained performance of the proposed metrics and algorithms is compared to the performance of state-of-the art algorithms (classes A and B).

5.2 System Model

Consider a single-cell with one BS equipped with N_t co-located antennas and $\mathcal{S}$ is a set of single antenna users ($N_r = 1$) as illustrated in Fig. 5.1. Assuming perfect CSI at the BS and the channel coefficients are modeled as independent random variables, with a zero-mean circularly symmetric complex Gaussian distribution (Rayleigh fading). The signal received by the kth user is given by:

$$\begin{aligned} y_k &= \sqrt{p_k}\mathbf{h}_k\mathbf{w}_k s_k + \sum_{i\neq k}^{K} \sqrt{p_i}\mathbf{h}_k\mathbf{w}_i s_i + n_k \\ &= \mathbf{h}_k\mathbf{x} + n_k, \end{aligned} \tag{5.1}$$

where s_k is the intended symbol for user k, $\mathbf{x} \in \mathbb{C}^{N_t\times 1}$ is the transmitted signal vector from the base station antennas, $\mathbf{h}_k \in \mathbb{C}^{1\times N_t}$ is the channel vector to the user k. Each user ignores the modulation and coding of other users, i.e., it is assumed single-user detection where each user treats the signals intended for other users as interference. $n_k \sim \mathcal{CN}(0, \sigma_n^2)$ is the additive zero mean white Gaussian noise with variance σ_n^2. The entries of the block fading channel $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K]$ and $\mathbf{n} = [n_1, \ldots, n_K]^T$ are normalized so that they have unitary variance, and the transmitter has an average power constraint $\mathbb{E}[\mathbf{x}^H\mathbf{x}] \leq P$. Since the noise has unit variance, P represents the SNR. For linear spatial processing at the transmitter, the BF matrix can be defined as $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_K]$, the symbol vector $\mathbf{s} = [s_1, s_2, \ldots, s_K]^T$ and $\mathbf{P} = diag(p_1, \ldots, p_K)$ is the matrix whose main diagonal contains the powers so that the transmitted signal is given by $\mathbf{x} = \sum_{k=1}^{K} \sqrt{p_k}\mathbf{w}_k s_k$. The SINR of the kth user is

$$SINR_k = \frac{p_k|\mathbf{h}_k\mathbf{w}_k|^2}{\sum_{i\neq k} p_i|\mathbf{h}_k\mathbf{w}_i|^2 + \sigma_n^2}, \tag{5.2}$$

and the maximum instantaneous achievable data rate in bits/Hz of user k for a given block fading channel is given by $r_k = \log_2(1 + SINR_k)$.

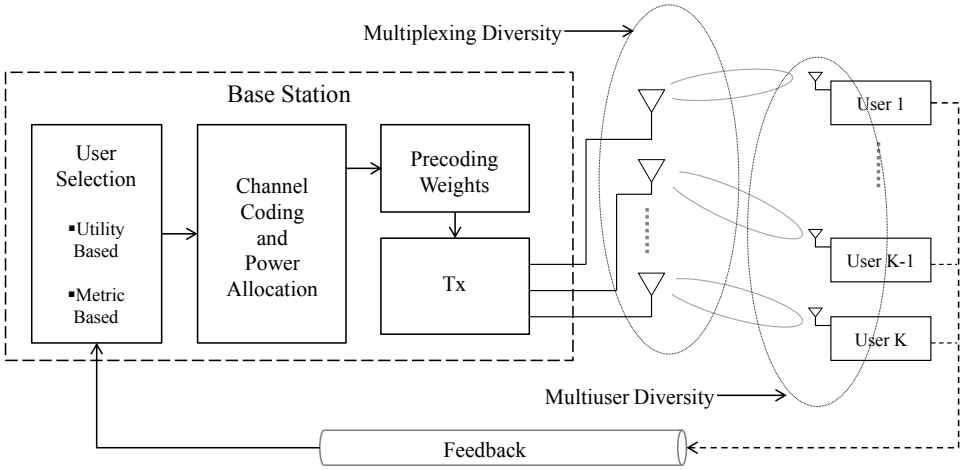

Figure 5.1: MISO Downlink System with N_t antennas at the BS and K single antenna users. The scheduler uses the fedback CSI and for a given set of users the BS performs power allocation and downlink transmission with beamforming.

5.3 The User Selection Problem

The performance of a MIMO system is measured by a global objective function of the individual data rates or SINRs, $U(r_1, \ldots, r_K)$. From the system perspective, it is desirable to optimize $U(\cdot)$ instead of the individual rates r_i $\forall i \in \mathcal{S}$, since the latter are coupled by the transmit powers and the beamforming weights in (5.2). Thus the performance depends on the way that the resources are allocated to each user and how efficiently the inter-user interference is mitigated. Indeed, the enhancement of the system performance can be modeled as a multi-objective optimization problem, where the optimum transmit strategy meets some systems constraints (e.g., power, rates, etc) and maximizes jointly the individual objective targets of the scheduled users [12].

In this chapter the system utility function is modeled as the sum rate using BF under global power constraints. Given $K \leq N_t$ the general optimization problem is given by

$$R^{BF} = \max_{\mathbf{W},\mathbf{P}} \sum_{k=1}^{|\mathcal{S}|} \omega_k r_k \quad \text{subject to } \|\mathbf{W}\mathbf{P}^{\frac{1}{2}}\|_F^2 \leq P \tag{5.3}$$

where ω_k is a priority weight associated to user k defined *a priory* by upper layers of the communications system to take into account QoS, fairness, or another system constraint. If the system prioritizes users, i.e., $\omega_i \neq \omega_j$ $\forall i \neq j$, there exist some sets of feasible $\mathbf{P}$ and $\mathbf{W}$ called Pareto optimal that solves (5.3) whose characteristics depend on the individual weights ω_k. A Pareto optimal allocation is one such that, it does not exist another feasible allocation where at least one user gets a better resource assignment, and all others get at least the same resources. Moreover, an individual rate cannot be improved on without hurting at least one

user. Finding a Pareto solution of (5.3) is a complex problem due to the nature of the optimum $\mathbf{W}$ and $\mathbf{P}$ and each solution depends on the system requirements expressed by the weights ω_k [12]. The computation of optimal beamforming weights $\mathbf{w}_k$ involves SINR balancing [33] and since the weights do not have a closed-form, iterative computational demanding algorithms have been proposed to determine them [34, 35]. As a matter of fact, it is known that problem (5.3) is NP-hard even when all priority weights ω_k are equal [61]. For some specific MISO systems efficient algorithms solve the weighted sum-rate maximization problem in polynomial time, for instance in [131] authors solve the SINR balancing problem through highly efficient algorithms that converge geometrically.

5.3.1 Multiuser scenario

Let $\Omega = \{1, \ldots, K\}$ be the set of all competing users where K is larger than the number of available antennas at the BS, i.e., $|\Omega| = K > N_t$. In order to exploit optimization dimension provided by MUDiv, it is necessary to select a set of active users $\mathcal{S}$ whose channel characteristics maximize the system sum rate when they transmit simultaneously in the same radio resource. Such characteristics are defined by the type of beamforming scheme, the power constraints, the SNR regime, and the deployment characteristics (N_t, K, and N_r). The sum rate maximization (user selection, beamforming, and power allocation) problem can be defined as:

$$\max_{\mathcal{S} \subseteq \Omega : |\mathcal{S}| \leq N_t} \quad R^{BF}(\mathbf{H}(\mathcal{S})), \tag{5.4}$$

where $\mathbf{H}(\mathcal{S})$ is the row-reduced channel matrix containing only the channels of the subset of selected users $\mathcal{S}$, and $R^{BF}(\mathbf{H}(\mathcal{S}))$ is the achievable sum rate for such set of users. Finding the optimum solution of (5.4) requires an exhaustive search over a search space of size $L = \sum_{i=1}^{N_t} \frac{K!}{i!(K-i)!}$, which is the number of all user permutations. Since the computation of the optimal solution of the sum rate maximization problem implies the joint optimization over $\mathbf{W}$, $\mathbf{P}$, and $\mathcal{S}$, the original problem (5.4) can be relaxed by taking one or more of the following actions: 1) by using beamforming weights $\mathbf{w}$ with a defined structure; 2) based on linear beamforming the power allocation can be performed either by the Lagrangian method or by equal power allocation; and 3) based on the structure of the linear beamforming it is possible to design user selection algorithms that exploit information contained in $\mathbf{H}$, achieving a good trade-off between performance and complexity.

In the following section linear beamforming schemes are introduced and it will be assumed that all users are equally prioritized, i.e., $\omega_k = 1 \ \forall k \in \Omega$. Although linear beamforming techniques are sub-optimal schemes compared to DPC, they highly simplify the computation of $\mathbf{w}$ and allow a straightforward allocation of the optimal power (water-filling) [25].

5.4 Linear Precoding Schemes

In this section the structure of the linear beamforming vectors and their main characteristics are described. Since the weight vectors multiply the intended symbols in (5.1) can be seen as a form of precoding, hereafter the terms beamforming and precoding are used interchangeably. The perfect knowledge of the channel vectors of all users at the transmitter is guaranteed through CSI feedback from the mobile stations to the BS. This can be attained through TDD scheme assuming channel reciprocity [12].

5.4.1 Zero Forcing Beamforming (ZFBF)

In Zero Forcing Beamforming (ZFBF) the channel matrix $\mathbf{H}$ at the transmitter is processed so that orthogonal channels between the transmitter and the receiver are created, defining a set of parallel subchannels [25]. The Moore-Penrose pseudoinverse of $\mathbf{H}(\mathcal{S})$ is given by [143]:

$$\bar{\mathbf{W}}(\mathcal{S}) = \mathbf{H}(\mathcal{S})^H(\mathbf{H}(\mathcal{S})\mathbf{H}(\mathcal{S})^H)^{-1}, \tag{5.5}$$

and the ZFBF matrix is given by the normalized column vectors of $\bar{\mathbf{W}}(\mathcal{S})$ as

$$\mathbf{W}(\mathcal{S}) = [\bar{\mathbf{w}}_1/\|\bar{\mathbf{w}}_1\|, \ldots, \bar{\mathbf{w}}_{|\mathcal{S}|}/\|\bar{\mathbf{w}}_{|\mathcal{S}|}\|]. \tag{5.6}$$

For ZFBF scheme the sum rate maximizing power allocation is given by the water-filling algorithm and according to [25] the information rate achieved with optimum $\mathbf{P}$ in (5.3) is given by :

$$R^{ZFBF}(\mathbf{H}(\mathcal{S})) = \sum_{i=1}^{|\mathcal{S}|} (\log(\mu b_i))^+, \tag{5.7}$$

where $b_i = \left\{[(\mathbf{H}(\mathcal{S})\mathbf{H}(\mathcal{S})^H)^{-1}]_{ii}\right\}^{-1}$ is the *effective channel gain*[1] of the ith user and its allocated power is

$$p_i = (\mu b_i - 1)^+, \tag{5.8}$$

and the water level μ is chosen to satisfy

$$\sum_{i\in\mathcal{S}} \left(\mu - \frac{1}{b_i}\right)^+ = P. \tag{5.9}$$

If all users in $\mathcal{S}$ are allocated with nonzero power, the water level has the compact form [46, 138, 144]:

$$\mu = \frac{1}{|\mathcal{S}|}(\|\bar{\mathbf{W}}(\mathcal{S})\|_F^2 + P) = \frac{1}{|\mathcal{S}|}(\sum_{i\in\mathcal{S}} b_i^{-1} + P). \tag{5.10}$$

5.4.2 Zero Forcing Dirty Paper (ZFDP)

Suboptimal throughput maximization in Gaussian BC channels has been proposed in several works [25, 137, 138] based on the QR-type decomposition of the channel matrix $\mathbf{H}(\mathcal{S}) = \mathbf{L}(\mathcal{S})\mathbf{Q}(\mathcal{S})$ obtained by applying Gram-Schmidt orthogonalization (GSO) [121] to the rows of $\mathbf{H}(\mathcal{S})$. $\mathbf{L}(\mathcal{S})$ is a lower triangular matrix and $\mathbf{Q}(\mathcal{S})$ has orthonormal rows. The beamforming matrix given by $\mathbf{W}(\mathcal{S}) = \mathbf{Q}(\mathcal{S})^H$ generates a set of interference channels

$$y_i = l_{ii}\sqrt{p_i}s_i + \sum_{j<i} l_{ij}\sqrt{p_j}s_j + n_i, \quad i = 1, \ldots, k \tag{5.11}$$

while no information is sent to users $k+1, \ldots, K$. In order to eliminate the interference component $I_i = \sum_{j<i} l_{ij}\sqrt{p_j}s_j$ of the ith user, the signals $\sqrt{p_i}s_i$ for $i = 1, \ldots, k$ are obtained by successive dirty-paper encoding, where I_i is noncausally known. This precoding scheme was proposed in [25] and the authors showed that the precoding matrix forces to zero the interference caused by users $j > i$ on each user i. Therefore, this scheme is called zero-forcing

[1] The effective channel gain is the magnitude of the channel after precoding.

dirty-paper (ZFDP) coding. The information rate achieved with optimum $\mathbf{P}$ in (5.3) under the ZFDP scheme is given by [25]:

$$R^{ZFDP}(\mathbf{H}(\mathcal{S})) = \sum_{i=1}^{|\mathcal{S}|} (\log(\mu d_i))^+ , \tag{5.12}$$

where $d_i = |l_{ii}|^2$ and μ is the solution to the water-filling equation

$$\sum_{i\in\mathcal{S}} \left(\mu - \frac{1}{d_i}\right)^+ = P, \tag{5.13}$$

which defines the ith power as $p_i = (\mu d_i - 1)^+$.

5.5 Metrics of Spatial Compatibility

Recall from Section 5.3 that problem (5.4) can be solved by fixing the precoder structure and the power allocation method. Under ZF-based precoding the performance depends on the spatial correlation between the components of $\mathbf{H}(\mathcal{S})$. The more correlated the channels, the higher the power penalty imposed by ZF schemes which yields a degradation of the achievable SINRs and a poor system performance. For this reason, several works in the literature tackled problem (5.4) by optimizing the spatial compatibility between scheduled users. This is accomplished by optimizing a specific metric over the channel matrix $\mathbf{H}(\mathcal{S})$, which can provide information regarding the achievable sum rate. In other words, a metric of spatial compatibility is a function of the CSI at the transmitter $f(\mathbf{H}(\mathcal{S}))$ so that $f : \mathbb{C}^{|\mathcal{S}|\times N_t} \mapsto \mathbb{R}_+$ where the mapping quantifies how profitable is to select $\mathcal{S}$ for transmission.

Different metrics for spatial compatibility have been proposed in the literature and this section presents a unified treatment of the most common metrics used by several state-of-the-art algorithms that solve (5.4). It is worth mentioning that the optimization over a given metric may bring some advantages in terms of computational complexity, for instance, iterative evaluations of $f(\mathbf{H}(\mathcal{S}))$ do not require the computation of the optimum power allocation. Some metrics are given by simple relations between the row vectors in $\mathbf{H}(\mathcal{S})$ which avoids matrix operations. Under certain SNR constraints the user set that solves problem (5.4) achieves maximum multiplexing diversity, i.e., its cardinality is equal to N_t [25]. In such SNR regime, the search space of the solution of problem (5.4) is reduced from L to $L_{N_t} = \binom{K}{N_t}$ and optimization over $f(\mathbf{H}(\mathcal{S}))$ can efficiently find the best set for a given metric.

5.5.1 Null Space Projection

Consider ZFBF and a given set of channels of selected users $\{\mathbf{h}_i\}_{i\in\mathcal{S}}$, the effective channel gain of the ith selected user (see Section 5.4.1) is given by [25]:

$$\begin{aligned} b_i &= \frac{1}{[(\mathbf{H}(\mathcal{S})\mathbf{H}(\mathcal{S})^H)^{-1}]_{i,i}} \\ &= \|\mathbf{h}_i\mathbf{Q}_{\mathcal{V}_i}\|^2 \\ &= \|\mathbf{h}_i\|^2 \sin^2\theta_{\mathcal{V}_i\mathbf{h}_i} \end{aligned} \tag{5.14}$$

where $\tilde{\mathbf{H}}_i = [\mathbf{h}_1, \ldots, \mathbf{h}_{i-1}, \mathbf{h}_{i+1}, \ldots, \mathbf{h}_{|\mathcal{S}|}]$ is the aggregate interference matrix associated with user i, $\mathcal{V}_i = Sp(\tilde{\mathbf{H}}_i)$ is the subspace spanned by the rows of $\tilde{\mathbf{H}}_i$. The matrix $\mathbf{P}_{\mathcal{V}_i} =$

$\tilde{\mathbf{H}}_i^H(\tilde{\mathbf{H}}_i\tilde{\mathbf{H}}_i^H)^{-1}\tilde{\mathbf{H}}_i$ is the orthogonal projector onto $\mathcal{V}_i$ and $\mathbf{Q}_{\mathcal{V}_i} = \mathbf{I} - \mathbf{P}_{\mathcal{V}_i}$ is the projector matrix onto the orthogonal complement of $\mathcal{V}_i$ [143]. The metric in (5.14) is equivalent to the projection of $\mathbf{h}_i$ onto the null space spanned by the rows of $\tilde{\mathbf{H}}_i$ (illustrated in Fig. 5.2), i.e., the null space projection (NSP) operation. It holds that $\|\mathbf{h}_i\mathbf{Q}_{\mathcal{V}_i}\|^2 \leq \|\mathbf{h}_i\|^2$ with equality when $\mathbf{h}_i$ is orthogonal to $Sp(\tilde{\mathbf{H}}_i)$. Notice that $\|\mathbf{h}_i\|^2$ in (5.14) is affected by the weight $\sin^2\theta_{\mathcal{V}_i\mathbf{h}_i}$ which is the squared sine of the angle between the channel vector $\mathbf{h}_i$ and the subspace spanned by the components of $\tilde{\mathbf{H}}_i$. The weight $\sin^2\theta_{\mathcal{V}_i\mathbf{h}_i}$ is referred in the literature as the projection power loss factor [145] since it will affect the effective amount of power that is transmitted over the ith link.

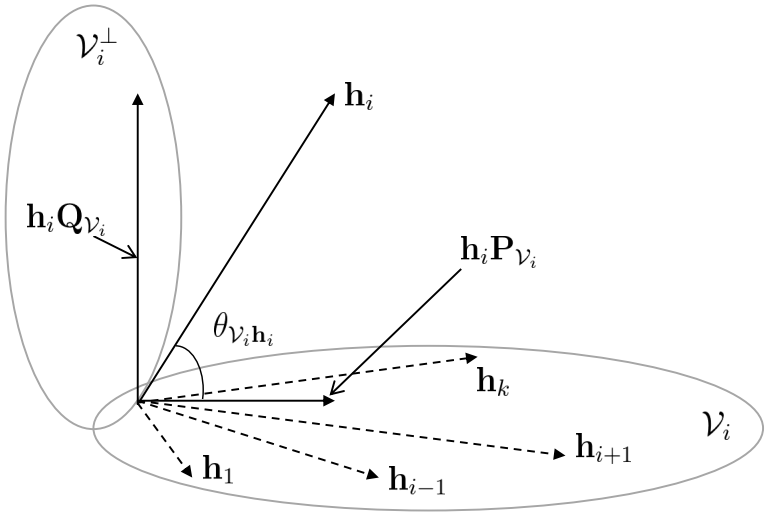

Figure 5.2: The spatial relationship between the components of vector $\mathbf{h}_i$ and $\mathcal{V}_i$.

Using the properties of water-filling and the strong relationship between the achievable rate with the terms b_i, a compact formulation of the maximization of metric (5.14) is elaborated below. Theoretical results in [45, 146] show that for the MISO BC system a meaningful metric to estimate the achievable performance of $\mathcal{S}$ is given by $\prod_{i\in\mathcal{S}} b_i$ under certain constraints over the water level μ for ZFBF and $N_r = 1$. The extension of such results is straightforward for $N_r > 1$ as the product of all effective eigen-direction gains of all users in $\mathcal{S}$ [45]. However, for $N_r = 1$ the performance gap between the sum and the product of the terms b_i is negligible and hereafter it is analyzed the metric $\sum_{i\in\mathcal{S}} b_i$ which is equivalent to a matrix trace operation. Let $\bar{\mathbf{H}}(\mathcal{S}) = \mathbf{H}(\mathcal{S})\mathbf{H}(\mathcal{S})^H$ and $\dot{\mathbf{H}}(\mathcal{S}) = \bar{\mathbf{H}}(\mathcal{S})^{-1}$ be a Wishart matrix [121] and its respective inverse which characterize the interaction of all user channels in $\mathcal{S}$. The set of users that achieves a sub-optimal solution to problem (5.4) by maximizing the sum of the effective channel gains is given by the solution of the following combinatorial problem:

$$\mathcal{S}_\omega = \arg \min_{\mathcal{S}\subset\Omega:|\mathcal{S}|=N_t} Tr(\dot{\mathbf{H}}(\mathcal{S})). \tag{5.15}$$

The optimum set $\mathcal{S}_\omega$ for the NSP metric is unique and it will contain the users that maximize $\sum_i \|\mathbf{h}_i\|^2 \sin^2\theta_{\mathcal{V}_i\mathbf{h}_i}$. The selection of $\mathcal{S}$ based on the NSP in (5.15) yields a close-to-optimal solution to problem (5.4) for a ZF-based precoding under the following conditions: $P > P_0$, $K > N_t$, or large K and N_t. The term P_0 is a critical SNR value that depends on $\mathbf{H}(\mathcal{S})$ for which the cardinality of $\mathcal{S}$ is always maximum, i.e., full spatial multiplexing

is attained [25]. The evaluation of the metric (5.14) is not unique and different algorithms proposed in the literature use different algebraic operations to compute or approximate such a metric. In the following subsection several methods to evaluate the NSP for ZFBF and ZFDP are presented. The application of each one of these methods lies in the complexity required to evaluate them as well as the available information at the BS. Each one of them represents a trade-off between accuracy of the NSP evaluation and the required CSI at the transmitter.

5.5.1.1 Orthogonal Projector for ZFBF

The computation of $\mathbf{Q}_{\mathcal{V}_i}$ is not unique and different forms to evaluate such matrix can be efficiently used in different contexts, i.e., depending on the available CSI at the transmitter, the required computational complexity, and the deployment (K, N_t, N_r). Applying singular value decomposition (SVD) [143] to the matrix $\tilde{\mathbf{H}}_i$ of the ith user yields:

$$\tilde{\mathbf{H}}_i = \mathbf{U}_{\tilde{\mathbf{H}}_i} \mathbf{\Sigma}_{\tilde{\mathbf{H}}_i} [\bar{\mathbf{V}}_{\tilde{\mathbf{H}}_i} \tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_i}]^H, \tag{5.16}$$

where $\tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_i}$ contains the $N_t - r$ basis of the null space of $\tilde{\mathbf{H}}_i$ and $r = rank(\tilde{\mathbf{H}}_i)$. The orthogonal projector matrix is given by $\mathbf{Q}_{\mathcal{V}_i} = \tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_i} \tilde{\mathbf{V}}^H_{\tilde{\mathbf{H}}_i}$ [147] and the set $\mathcal{S}_\omega$ in (5.15) maximizes the objective function $\sum_i \|\mathbf{h}_i \mathbf{Q}_{\mathcal{V}_i}\|^2$.

In some scenarios it is assumed that the BS knows the basis of $Sp(\tilde{\mathbf{H}}_i)$ for any user $i \in \mathcal{S}$. Let $\{\mathbf{v}_j\}_{j=1}^r$ be the column vectors of $\bar{\mathbf{V}}_{\tilde{\mathbf{H}}_i}$ defined in (5.16) and the NSP in (5.14) for the ith user can be computed as follows:

$$\|\mathbf{h}_i \mathbf{Q}_{\mathcal{V}_i}\|^2 = \left\| \mathbf{h}_i \left(\mathbf{I} - \sum_{j=1}^{r} \mathbf{v}_j \mathbf{v}_j^H \right) \right\|^2 . \tag{5.17}$$

The NSP operation in (5.17) can be also computed by applying GSO to $\mathbf{H}(\mathcal{S})$ as in [40] which represents a lower computational complexity than the SVD approach [148]. Using the basis of $Sp(\tilde{\mathbf{H}}_i)$ the magnitude of the NSP operation is also given by $\|\mathbf{h}_i \mathbf{Q}_{\mathcal{V}_i}\|^2 = \|\mathbf{h}_i\|^2 - \|\check{\mathbf{h}}_i\|^2$ where $\check{\mathbf{h}}_i$ is the projection of $\mathbf{h}_i$ onto each one of the orthogonal basis of $\mathcal{V}_i$ as follows [121]:

$$\check{\mathbf{h}}_i = \sum_{j=1}^{r} \frac{\|\mathbf{h}_i\| \cos\theta_{\mathbf{h}_i \mathbf{v}_j}}{\|\mathbf{v}_j\|} \mathbf{v}_j^H, \tag{5.18}$$

where the term $\cos\theta_{\mathbf{h}_i \mathbf{v}_j}$ represents the coefficient of correlation between the vectors $\mathbf{h}_i$ and $\mathbf{v}_j$ and is defined as [143]:

$$\cos\theta_{\mathbf{h}_i \mathbf{v}_j} = \frac{|\mathbf{h}_i \mathbf{v}_j^H|}{\|\mathbf{h}_i\| \|\mathbf{v}_j\|}, \tag{5.19}$$

and the domain of the coefficient is $0 \leq \eta_{\mathbf{h}_i \mathbf{v}_j} = \cos\theta_{\mathbf{h}_i \mathbf{v}_j} \leq 1$ and $\theta_{\mathbf{h}_i \mathbf{v}_j} = \frac{\pi}{2}$ means perfect spatial orthogonality.

The NSP calculation does not necessary require SVD and other matrix operations can be used instead. Using the full channel matrix $\mathbf{H}(\mathcal{S})$ and $\tilde{\mathbf{H}}_i$ for all $i \in \mathcal{S}$ the block matrix determinant formula to compute $\det(\mathbf{H}(\mathcal{S})\mathbf{H}(\mathcal{S})^H)$ reads [149]:

$$\begin{aligned}\det\left(\mathbf{H}(\mathcal{S})\mathbf{H}(\mathcal{S})^H\right) &= \det\left(\begin{bmatrix}\tilde{\mathbf{H}}_i\tilde{\mathbf{H}}_i^H & \tilde{\mathbf{H}}_i\mathbf{h}_i^H \\ \mathbf{h}_i\tilde{\mathbf{H}}_i^H & \mathbf{h}_i\mathbf{h}_i^H\end{bmatrix}\right) \\ &= \det\left(\tilde{\mathbf{H}}_i\tilde{\mathbf{H}}_i^H\right)\|\mathbf{h}_i\mathbf{Q}_{\mathcal{V}_i}\|^2\end{aligned} \tag{5.20}$$

and the NSP is given by $\|\mathbf{h}_i\mathbf{Q}_{\mathcal{V}_i}\|^2 = \det(\mathbf{H}(\mathcal{S})\mathbf{H}(\mathcal{S})^H)/\det(\tilde{\mathbf{H}}_i\tilde{\mathbf{H}}_i^H)$.

The orthogonal projection of $\mathbf{h}_i$ onto $\mathcal{V}_i$, $\forall i \in \mathcal{S}$ defined in (5.14) has a direct relationship with the correlation coefficients defined in (5.19). The normalized power loss experienced by the ith channel when it interacts with the subspace $\mathcal{V}_i$ is called the coefficient of determination given by [143]:

$$\Delta^2_{\mathcal{V}_i\mathbf{h}_i} = \frac{\mathbf{h}_i\mathbf{P}_{\mathcal{V}_i}\mathbf{h}_i^H}{\mathbf{h}_i\mathbf{h}_i^H}, \tag{5.21}$$

where $\Delta^2_{\mathcal{V}_i\mathbf{h}_i}$ measures how much the vector $\mathbf{h}_i$ can be predicted from the channel vectors of $\tilde{\mathbf{H}}_i$. Notice that from (5.14) and (5.21) the projection power loss factor of $\mathbf{h}_i$ due to the projection onto the null space of $\mathcal{V}_i$ is equivalent to $1 - \Delta^2_{\mathcal{V}_i\mathbf{h}_i}$ which can be evaluated as follows [143]:

$$1 - \Delta^2_{\mathcal{V}_i\mathbf{h}_i} = (1 - \eta^2_{\mathbf{h}_i\pi(1)})(1 - \eta^2_{\mathbf{h}_i\pi(2)|\pi(1)}) \ldots (1 - \eta^2_{\mathbf{h}_i\pi(k)|\pi(1)\ldots\pi(k-1)}), \tag{5.22}$$

where $\pi(k)$ is the kth ordered element of $\tilde{\mathbf{H}}_i$ and $\eta_{\mathbf{h}_i\pi(k)|\pi(1)\ldots\pi(k-1)}$ is the partial correlation between the channel vector $\mathbf{h}_i$ and the selected vector associated with $\pi(k)$ eliminating the effects due to $\pi(1), \pi(2), \ldots, \pi(k-1)$. Using multiple regression analysis it is possible to evaluate $\|\mathbf{h}_i\mathbf{Q}_{\mathcal{V}_i}\|^2 = \|\mathbf{h}_i\|^2(1 - \Delta^2_{\mathcal{V}_i\mathbf{h}_i})$ by extracting the partial correlation coefficients from the correlation coefficients choosing one user order π out of $(|\mathcal{S}| - 1)!$ permutations of the users in $\mathcal{S}$ [143].

A different approach can be applied if for a given set of channels $\{\mathbf{h}_j\}_{j\in\mathcal{S}}$ the orthogonal projector matrix of each channel is known so that for $N_r = 1$, and $\mathbf{Q}_j = \mathbf{I} - \mathbf{h}_j^H(\mathbf{h}_j\mathbf{h}_j^H)^{-1}\mathbf{h}_j$ and for $N_r > 1$ SVD can be used to calculate $\mathbf{Q}_j$. From [150] the following result holds:

$$\mathbf{Q}_{\mathcal{V}_i} = \left(\prod_{j\neq i, j\in\mathcal{S}} \mathbf{Q}_j\right)^n, \; n \to \infty \tag{5.23}$$

which establishes that the orthogonal projector matrix onto $\mathcal{V}_i$ can be approximated by recurrently projecting onto independent orthogonal subspaces such that their intersection strongly converges to $\mathbf{Q}_{\mathcal{V}_i}$ as n grows.

The NSP measured by $Tr(\dot{\mathbf{H}}(\mathcal{S}))$ has been used by several user selection algorithms either by avoiding the exhaustive search required to solve (5.15) or by using relaxed forms to compute (5.14). Table 5.1 summarizes different forms of computing the NSP applied in the literature of users selection.

5.5.1.2 Orthogonal Projector for ZFDP

In Section 5.4.2 the ZFDP beamforming was described and it was mentioned that the received signal of user i contains an interference component from all $j < i$ users, i.e., the previously encoded users. Given an specific encoding order $\pi(i)$, $i \in \{1, \ldots, |\mathcal{S}|\}$ (a permutation

Table 5.1: NSP for the ith user of $\mathcal{S}$ in Algorithms for User Selection

$\|\mathbf{h}_i\mathbf{Q}_{\mathcal{V}_i}\|^2$	Evaluation	System	NSP Approach
$\|\mathbf{h}_i\tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_i}\tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_i}^H\|^2$	Exact (5.14)	MISO	Iterative selection of $\mathcal{S}$ based on SVD [136, 137]
$\|\mathbf{h}_i\mathbf{Q}_{\mathcal{V}_i}\|^2$	Exact (5.17)	MISO	Selection based on GSO [40, 151]
$\|\mathbf{h}_i\|^2(1-\Delta^2_{\mathcal{V}_i\mathbf{h}_i})$	Approx. (5.19)	MISO	Selection based on correlation coefficients [152]
$\|\mathbf{h}_i\left(\prod_{j\neq i}\mathbf{Q}_j\right)^n\|^2$	Approx. (5.23)	MIMO	Tree search using SVD and $n\in\{1,2,3\}$ [12, 153]

of the users in $\mathcal{S}$), for this scheme the beamforming vectors $\mathbf{w}_{\pi(i)}$ are computed either by a QR-type decomposition or by GSO as follows [154]:

$$\mathbf{w}_{\pi(i)} = \frac{\mathbf{T}_i\mathbf{h}_{\pi(i)}^H}{\|\mathbf{T}_i\mathbf{h}_{\pi(i)}^H\|}, \tag{5.24}$$

where

$$\mathbf{T}_i = \begin{cases} \mathbf{I} & \text{for } i=1, \\ \mathbf{T}_{i-1} - \mathbf{w}_{\pi(i-1)}\mathbf{w}_{\pi(i-1)}^H & \text{for } i=2,\ldots,q, \end{cases} \tag{5.25}$$

and $i=1,\ldots,q$ with $q=rank(\mathbf{H}(\mathcal{S}))$. The matrix $\mathbf{T}_i$ is the orthogonal projector matrix onto $Sp(\mathbf{h}_{\pi(j<i)})$ the subspace spanned by all previously encoded users for which $\mathbf{h}_{\pi(j<i)}\mathbf{w}_{\pi(i)}=0$. Some authors (e.g., [154, 155]) use the following expression as an objective function over the channel matrix $\mathbf{H}(\mathcal{S})$ for user selection and ZFDP:

$$f(\mathbf{H}(\mathcal{S})) = \sum_{i=1}^{|\mathcal{S}|}\|\mathbf{h}_i\mathbf{T}_i\|^2. \tag{5.26}$$

Observe that user selection and sum rate maximization based on (5.26) implicitly depend on one particular selected encoding order π out of $|\mathcal{S}|!$ different valid orders. Since different orders yield different values of (5.26), Tejera et al. [154] proposed a method to perform the successive encoding optimizing of the order π. Such an optimum order is attained by an algorithm that evaluates (5.14) each iteration for every successive encoded user. An alternative suboptimal approach can be employed as in [137] where π is defined by the descending order of the effective channel gains of the users in $\mathcal{S}$.

5.5.2 Approximation of the NSP for ZFBF

The objective function of problem (5.15) can be further relaxed by using a lower bound of $Tr(\dot{\mathbf{H}}(\mathcal{S}))$ in order to avoid the computation of the inverse matrix $\dot{\mathbf{H}}(\mathcal{S})$. Considering the definition of trace

$$Tr(\dot{\mathbf{H}}(\mathcal{S})) = \sum_i \lambda_i^{-1}(\bar{\mathbf{H}}(\mathcal{S})), \tag{5.27}$$

and using the arithmetic-geometric mean inequality over the eigenvalues of $\dot{\mathbf{H}}(\mathcal{S})$ it holds [16]:

$$|\mathcal{S}|\left(\prod_i \lambda_i(\dot{\mathbf{H}}(\mathcal{S}))\right)^{1/|\mathcal{S}|} \leq Tr(\dot{\mathbf{H}}(\mathcal{S})). \tag{5.28}$$

Since $|\mathcal{S}|$ is constant and independent of the selected channels for all L_{N_t} user permutations, the lower bound on the objective function of (5.15) can be simplified as:

$$\prod_i \lambda_i(\dot{\mathbf{H}}(\mathcal{S})) = \det(\bar{\mathbf{H}}(\mathcal{S}))^{-1}. \tag{5.29}$$

A suboptimal but less computational demanding way to find a set of users that solves problem (5.4) is by defining the set that solves the following combinatorial problem:

$$\mathcal{S}_\xi = \arg \max_{\mathcal{S}\subset\Omega:|\mathcal{S}|=N_t} \det(\bar{\mathbf{H}}(\mathcal{S})), \tag{5.30}$$

where the optimized objective function only requires to compute the determinant of a matrix product. Observe that the lower bound of metric (5.29) is highly related with metric (5.20) and the performance degradation of the former arises because the terms $\det(\tilde{\mathbf{H}}_i\tilde{\mathbf{H}}_i^H)$ $\forall i \in \mathcal{S}$ are neglected.

In [152] it was presented a greedy algorithm where the metric for user selection is based on an approximation of (5.22) and the correlation coefficients are used instead of the partial correlation coefficients. Such relaxation neglects the channel gain degradation due to the terms $\pi(1), \pi(2), \ldots, \pi(k-1)$. Given channel matrix $\mathbf{H}(\mathcal{S})$, the metric that approximates $\sum_{i\in\mathcal{S}} b_i$ is defined as follows [152]:

$$\Lambda(\mathbf{H}(\mathcal{S})) = \sum_i \|\mathbf{h}_i\|^2 \prod_{j\neq i} \sin^2\theta_{\mathbf{h}_i\mathbf{h}_j} \quad \forall i,j \in \mathcal{S} \tag{5.31}$$

Using this metric a suboptimal solution to problem (5.4) is given by the set of users that solve the following combinatorial problem:

$$\mathcal{S}_\Lambda = \arg \max_{\mathcal{S}\subset\Omega:|\mathcal{S}|=N_t} \Lambda(\mathbf{H}(\mathcal{S})). \tag{5.32}$$

5.5.3 ϵ-orthogonality

Several user selection algorithms (e.g., [134, 152, 155, 156]) attempt to create groups of quasi-orthogonal users based on the information provided by the coefficient of correlation (5.19). A set of channels $\{\mathbf{h}_i\}_{i\in\mathcal{S}}$ is called ϵ-orthogonal if $\cos\theta_{\mathbf{h}_i\mathbf{h}_j} < \epsilon$ for every $i,j \in \mathcal{S}$ [134]. Some works addressed problem (5.4) by scheduling the set of user with minimum ϵ-orthogonality. The metric is measured over the normal channels [156] or over the eigenvectors computed by SVD [134]. If the orthogonality among channels in $\mathbf{H}(\mathcal{S})$ is the only metric used to define $\mathcal{S}$ (regardless of the channel gains), the sum rate maximization problem (5.4) can be sub-optimally solved by the set of users that minimizes the ϵ-orthogonality among all L_{N_t} possible sets. Such a set can be formally described as:

$$\mathcal{S}_\epsilon = \arg \min_{\mathcal{S}\subset\Omega:|\mathcal{S}|=N_t} \left(\max_{i,j\in\mathcal{S}} \cos\theta_{\mathbf{h}_i\mathbf{h}_j} \right). \tag{5.33}$$

Some works in the literature define $\mathcal{S}_\epsilon$ as the set with minimum sum of correlation coefficients $\sum_i \sum_{j\neq i} \cos\theta_{\mathbf{h}_i\mathbf{h}_j}$ $\forall i,j \in \mathcal{S}$ [134, 155] or as the set with minimum average correlation coefficient [157]. Observe that these objective functions are based on pairwise metrics and they can be negatively biased by few terms η with large values neglecting the remaining coefficients with relatively small values. An alternative utility function that can identify such

large terms is the geometric mean over all correlation coefficients in $\mathcal{S}$ since it would assign the same priority to each term. In multi-carrier systems with beamforming transmission [156, 157], problem (5.33) is solved for each carrier which represents an user grouping process based on (5.19). In MU-MIMO systems the user grouping problem based on (5.33) either for scheduling time slots, sub-carriers, or both, can be modeled as a graph coloring problem [49, 158] or as a graph clique problem [142] and complexity reduction is the main objective of the proposed grouping algorithms.

5.5.4 Orthogonality Defect

The orthogonality defect derived from Hadamard's inequality [16] measures how close a basis is to orthogonal. Given the matrix $\mathbf{H}(\mathcal{S})$ the orthogonality defect is defined as:

$$\delta(\mathbf{H}(\mathcal{S})) = \frac{\prod_{i=1} \|\mathbf{h}_i\|^2}{\prod_{i=1} \lambda_i(\bar{\mathbf{H}}(\mathcal{S}))}, \tag{5.34}$$

and $\delta(\mathbf{H}(\mathcal{S})) = 1$ if and only if the elements of $\mathbf{H}(\mathcal{S})$ are pairwise orthogonal. The metric defined in (5.34) reflects the orthogonality of the set $\{\mathbf{h}_i\}_{i\in\mathcal{S}}$ and it has been proposed to evaluate the compatibility between wireless channels so that spatial multiplexing gain is maximized [134, 159]. The original problem (5.4) can be suboptimally solved by the set of users that minimizes the orthogonality defect which is formally described as:

$$\mathcal{S}_\delta = \arg \min_{\mathcal{S}\subset\Omega:|\mathcal{S}|=N_t} \delta(\mathbf{H}(\mathcal{S})). \tag{5.35}$$

Observe that problem (5.35) uses a weighted version of the utility function of problem (5.30) where the weight is defined by the inverse of $\prod_{i=1} \|\mathbf{h}_i\|^2$.

5.5.5 Condition number

The ZF-based beamforming methods are in general power inefficient since the spatial direction of $\mathbf{w}_i$ is not matched to $\mathbf{h}_i$ $\forall i \in \mathcal{S}$. Inverting an ill-conditioned channel matrix $\mathbf{H}(\mathcal{S})$ yields a significant power penalty and a strong SINR degradation at the receivers. In numerical analysis, a metric to measure the invertibility of a matrix is given by the condition number. In MIMO system this metric is used to measure how the eigenvalues of a channel matrix spread out and to indicate multipath richness for a given channel realization. The less spread of the eigenvalues, the larger the achievable capacity in the high SNR regime. For the matrix $\mathbf{H}(\mathcal{S})$ the condition number is defined as [160]:

$$\kappa(\mathbf{H}(\mathcal{S})) = \frac{|\lambda_{\max}(\bar{\mathbf{H}}(\mathcal{S}))|}{|\lambda_{\min}(\bar{\mathbf{H}}(\mathcal{S}))|}, \tag{5.36}$$

and $\kappa(\mathbf{H}(\mathcal{S}))$ assess the multiplexing gain of a MIMO system [160]. Another definition of the condition number is given by the product $\|\mathbf{A}\|\|\mathbf{A}^{-1}\|$ for a given nonsingular square matrix $\mathbf{A}$ [121] and (5.36) generalizes that metric for any matrix $\mathbf{H}(\mathcal{S}) \in \mathbb{C}^{|\mathcal{S}|\times N_t}$ where $|\mathcal{S}| \leq N_t$. If the condition number is small, the matrix $\mathbf{H}(\mathcal{S})$ is said to be well-conditioned which implies that as $\kappa(\mathbf{H}(\mathcal{S})) \to 1$ the total achievable sum rate in the MISO systems under ZFBF can achieve a large portion of the sum rate of the inter-user interference free scenario. Problem (5.4) can be sub-optimally solved by a set of users with the minimum condition number and such set is formally described as:

$$\mathcal{S}_{\kappa} = \arg \min_{\mathcal{S} \subset \Omega : |\mathcal{S}| = N_t} \kappa(\mathbf{H}(\mathcal{S})). \tag{5.37}$$

Another metric to estimate matrix condition is given by the Demmel condition number. Several applications in MIMO systems have been proposed in recent years for this metric, e.g., link adaptation, coding, and beamforming [161]. The Demmel condition number can be seen as the ratio between the total energy of the channels of $\mathbf{H}(\mathcal{S})$ over the magnitude of the smallest eigenvalue of $\bar{\mathbf{H}}(\mathcal{S})$ and is given by the following expression[161]:

$$\kappa_D(\mathbf{H}(\mathcal{S})) = \frac{Tr(\bar{\mathbf{H}}(\mathcal{S}))}{\lambda_{\min}(\bar{\mathbf{H}}(\mathcal{S}))}, \tag{5.38}$$

where $Tr(\bar{\mathbf{H}}(\mathcal{S})) = \|\mathbf{H}(\mathcal{S})\|_F^2$, i.e., the Frobenius norm is related with the overall energy of the channel. Using (5.38) the set of users that sub-optimally solves (5.4) is given by the solution of the following combinatorial problem:

$$\mathcal{S}_{\kappa_D} = \arg \min_{\mathcal{S} \subset \Omega : |\mathcal{S}| = N_t} \kappa_D(\mathbf{H}(\mathcal{S})). \tag{5.39}$$

5.6 Power Projection Based User Selection

In this section it is designed a user selection algorithm that sub-optimally solves the sum rate maximization problem. This design only considers the physical layer information, whereas the application level delay effects are not considered. It is assumed that all users have infinite information to transmit when they are scheduled. The generalization of the user selection problem is modeled as an integer program and the sub-optimality of the selection metrics is discussed.

5.6.1 Iterative Power Projection (IPP)

Based on the fact that (5.14) has a fundamental connection with the coefficients of correlation as described in Section 5.5.1, this subsection presents an algorithm that attempts to find a quasi-orthogonal set of users $\mathcal{S}$ using exclusively the information provided by the channel norms and a measure of the orthogonality between two user channels given by (5.19). Fig. 5.3 exemplifies the required information used to find the set $\mathcal{S}$, and for two selected users i and j the figure shows the physical components that affect the interaction with a third unselected user k.

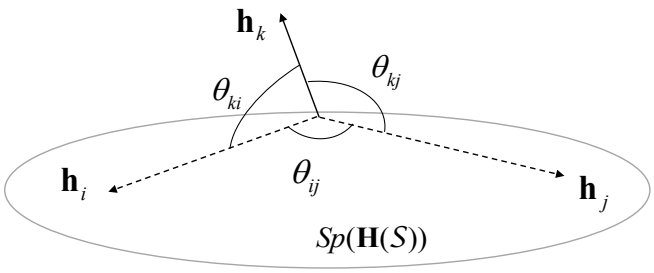

Figure 5.3: Interaction of two selected users i and j with third unselected user k

In order to start the users selection process, it is assumed that the base station knows the coefficients of correlation for all users in $\Omega = \{1, \ldots, K\}$, which requires $(K^2 - K)/2$ computations of (5.19) since $\eta_{\mathbf{h}_i\mathbf{h}_j} = \eta_{\mathbf{h}_j\mathbf{h}_i}$ and the computation of the coefficients (inner product and vector norm operations) can be done within time $O(K)$. For sake of notation let $\varrho_{ij} = 1 - \eta^2_{\mathbf{h}_i\mathbf{h}_j}$, $\hat{\varrho}_{ij} = 1 - |\eta_{\mathbf{h}_i\mathbf{h}_j}|$ and define the following geometric and arithmetic means for the elements ϱ associated to user $i \in \Omega$ as:

$$M_{g(i)} = \left(\prod_{j \neq i, j \in \Omega} \varrho_{ij} \right)^{\frac{1}{|\Omega|-1}} \leq \left(\frac{1}{|\Omega| - 1} \right) \sum_{j \neq i, j \in \Omega} \varrho_{ij}, \tag{5.40}$$

where $M_{g(i)}$ is a lower bound of the arithmetic mean of the projection power loss factors of user i. The first selected user is the one that maximizes the product of its channel magnitude by the average projection onto all other users:

$$i^* = \arg\max_{i \in \Omega} \|\mathbf{h}_i\|^2 M_{g(i)}, \tag{5.41}$$

and the sets of selected and unselected users are updated, $\mathcal{S} = \{i^*\}$ and $\Omega = \Omega - \{i^*\}$ respectively. Selecting the first user by (5.41) has the goal of assigning priority weights to the channel norms, i.e., users with large channel norms are penalized if their associated correlation coefficients have a large variance. Furthermore, the geometric mean $M_{g(i)}$ minimizes the bias created by the terms ϱ with very large or small values, which would be neglected if the arithmetic mean of the projection power loss factors were considered in (5.41).

The following user to be selected must maximize two criteria at the same time. On the one hand, it must maximize its own projected power which is affected by the coefficients ϱ of the already selected users in $\mathcal{S}$. The effective projected power of the user $i \in \Omega$ is given by:

$$\psi_i = \|\mathbf{h}_i\|^2 \prod_{j \in \mathcal{S}} \varrho_{ij}. \tag{5.42}$$

On the other hand, the users in $\mathcal{S}$ have already achieved the effective channel gains:

$$\phi_j = \|\mathbf{h}_j\|^2 \prod_{k \neq j, k \in \mathcal{S}} \varrho_{jk}, j \in \mathcal{S}. \tag{5.43}$$

For a new user candidate $i \in \Omega$, its aggregation to the set $\mathcal{S}$ implies a reduction of the total sum of effective channel gains of the selected users ($\sum_{j \in \mathcal{S}} \phi_j$) by the factors ϱ associated with the new selected user. Using the arithmetic and geometric means, lower bounds of the average effective channel gains of the selected users in (5.43) can be defined for the ith unselected user as follows:

$$\prod_{j \in \mathcal{S}} \phi_j \varrho_{ij} \leq \left(\frac{1}{|\mathcal{S}|} \sum_{j \in \mathcal{S}} \phi_j \varrho_{ij} \right)^{|\mathcal{S}|} \leq \left(\frac{1}{|\mathcal{S}|} \sum_{j \in \mathcal{S}} \phi_j \right)^{|\mathcal{S}|}. \tag{5.44}$$

The total effective projection power (an approximation to the effective channel gain) $\dot{\varphi}_i$ of the unselected user i takes into account the two required criteria. 1) the average projection power over the elements in $\mathcal{S}$ computed for the lower bound in (5.44), and 2) the projection power of user $i \in \Omega$ (5.42):

$$\dot{\varphi}_i = \underbrace{\left(\prod_{j\in\mathcal{S}} \phi_j \varrho_{ij}\right)}_{\text{gain } \forall j \in \mathcal{S}} \underbrace{\left(\|\mathbf{h}_i\|^2 \prod_{j\in\mathcal{S}} \varrho_{ij}\right)}_{\text{gain } \forall i \in \Omega}$$
$$= \underbrace{\left(\prod_{j\in\mathcal{S}} \phi_j\right)}_{\text{constant } \forall i \in \Omega} \left(\prod_{j\in\mathcal{S}} \varrho_{ij}\right) \left(\|\mathbf{h}_i\|^2 \prod_{j\in\mathcal{S}} \varrho_{ij}\right). \tag{5.45}$$

By taking the square of the product of the terms ϱ_{ij} both effects are considered, the impact of the selected users over user i and the power degradation that the users in $\mathcal{S}$ will have if user i is selected. Since the effective projected power of the selected users remains constant for all users in Ω, the metric in (5.45) can be normalized as follows:

$$\varphi_i = \|\mathbf{h}_i\|^2 \prod_{j\in\mathcal{S}} \varrho_{ij}^2. \tag{5.46}$$

Given $\mathcal{S}$, the next selected user is found using the metric defined in (5.46) as:

$$i^* = \arg\max_{i\in\Omega} \varphi_i, \tag{5.47}$$

where the selection of the locally optimum $\varphi(n)$ in a given iteration n is conditioned on the choice of $\varphi(1), \ldots, \varphi(n-1)$.

As $N_t, K \to \infty$ the number of total operations to solve problem (5.4) becomes computationally demanding and an optimization of the set Ω can be performed. By using (5.47) recursively, each iteration requires the comparison of $|\Omega|$ elements in order to select the user whose projection power is maximum. Considering that the cardinality of the final set must be N_t, without dynamically modifying Ω each iteration, the selection process would require a total of $L_c = N_t(K - (N_t - 1)/2)$ comparison operations. It is worth mentioning that the algorithms proposed in [134, 136, 137] also require L_c comparison operations times the elements of $\mathcal{S}$. However, the computational complexity is quite different since each comparison implemented in those works requires a matrix multiplication whilst the metric used in (5.46) is a multiplication of real positive numbers.

In [40, 162, 163] after a new user i is added to $\mathcal{S}$, the set of unselected user Ω is reduced by keeping the users whose correlation factors are above a threshold α_{th}, i.e., $\Omega(n) = \{j \in \Omega(n-1) : \eta_{\mathbf{h}_i\mathbf{h}_j} < \alpha_{th}\}$, where n stands for the iteration number, and i is the selected user of iteration $n-1$. This sub-selection within the algorithm has the drawback that the value of the parameter α_{th} is fixed which might result in a drastic reduction of the size of Ω and the degradation of the multiuser diversity. According to [40] there exists an optimum value of the threshold α_{th} for each value of K and N_t but the mathematical relationship between these terms is not given in a closed form. The statistical dependence of the average throughput regarding to α_{th} has been established only for the case where the cardinality of the set of selected users is constrained to be 2, i.e., $|\mathcal{S}| = 2$ in [145].

A dynamic optimization of the set Ω can be performed by considering two criteria to discard users at each iteration. The first criterion is related to the statistics of the projection powers regarding the users that have been selected. The second criterion weights the first one

based on the number of active users and the number of antennas N_t. Define the arithmetic mean of the projected powers given the new selected user i^* as:

$$M_{a(i^*)} = \frac{1}{|\Omega|} \sum_{j \in \Omega} \|\mathbf{h}_j\|^2 \varrho_{i^* j}. \tag{5.48}$$

Notice that the power projection computation is performed considering only the power projection loss factors associated with i^* and each term of the sum in (5.48) is the multiplication of two real numbers. The metric defined in (5.48) is used to discard users whose projection powers are below the arithmetic mean which results in a reduction of the number of comparisons for the next iteration. Nevertheless, when the number of total users is low ($K \approx N_t$) the number of users in Ω should not be reduced drastically in order to preserve enough multiuser diversity and to achieve full spatial multiplexing. Define a weight factor based on the number of antennas N_t and the size of the sets $\mathcal{S}$ and Ω as follows:

$$w_{(N_t,\mathcal{S},\Omega)} = 1 - \left(\frac{N_t - |\mathcal{S}|}{|\Omega|} \right)^{\frac{1}{N_t - |\mathcal{S}|}}. \tag{5.49}$$

The objective of $w_{(N_t,\mathcal{S},\Omega)}$ is to scale $M_{a(i^*)}$ in iteration n taking into account the spatial resources available at the BS and the current size of Ω. Given the new selected user i^* and weighting (5.48) times (5.49), the modified set of users that will compete to be scheduled in the next iteration $n+1$ is defined as:

$$\Omega(n+1) = \left\{ j \in \Omega(n) : \|\mathbf{h}_j\|^2 \varrho_{i^* j} \geq w_{(N_t,\mathcal{S},\Omega)} M_{a(i^*)} \right\}. \tag{5.50}$$

The procedure to generate the quasi-orthogonal set of user that solves problem (5.4) is described in Algorithm 5.1.

Algorithm 5.1 Iterative Power Projection (IPP)

1: $\Omega = \{1, \ldots, K\}$, $\mathcal{S} = \emptyset$, $n = 0$
2: **while** $|\mathcal{S}| < N_t$ **do**
3: **if** $n = 0$ **then**
4: Compute i^* by (5.41)
5: **else**
6: Compute i^* by (5.47)
7: **end if**
8: $n = n + 1$, $\mathcal{S}(n) = \mathcal{S}(n-1) \cup \{i^*\}$, $\Omega(n) = \Omega(n-1) - \{i^*\}$
9: Update $\Omega(n)$ by (5.50)
10: **end while**
11: Power Loading Principle: water-filling

5.6.2 An Integer Linear Program (ILP) Approach

The optimization performed in Algorithm 5.1 can be described as a greedy search over a tree structure [164] where the root is given by the element of Ω that preservers a higher average projected power (5.41). Similar approaches are implemented in [40, 134, 136, 137] considering the user with the maximum channel norm as the root of the search tree. The greedy Algorithm 5.1 makes a sequence of decisions in order to optimize the metric in (5.47). However, this local optimization might not lead to a global optimal solution. Moreover,

since the first user is found by (5.41), the correlation of such a user with the future selected users is neglected when $\mathcal{S}$ is initialized. A general mathematical model of the interaction of all elements in $\mathcal{S}$ that exploits the metrics used in (5.41) and (5.47) can be designed. The structure of (5.41) and (5.47) maximize the squared channel norm weighted by the product of the correlation coefficients. Such an operation can cast a relaxed version of the user selection problem (5.4) into an integer programming problem.

Define the interaction of the user $i \in \Omega$ with the rest of the users as a function f_i considering the structure of (5.46) as:

$$f_i = \|\mathbf{h}_i\|^2 \prod_{j \neq i} \varrho_{ij}^2, \ \forall i, j \in \Omega, \tag{5.51}$$

and by applying a change of variables, define the function $\tilde{f}_i$ as

$$\tilde{f}_i = a_i + \sum_{j \neq i} b_{ij}, \tag{5.52}$$

where

$$a_i = 2\log(\|\mathbf{h}_i\|), \tag{5.53}$$

and

$$b_{ij} = 2\log(\varrho_{ij}). \tag{5.54}$$

The next goal is to maximize the total sum of the projected powers which is a function of two factors, the orthogonality between the selected channels and the amount of remaining power after a projection. Therefore, (5.4) can be reformulated as the maximization of $\sum_i \tilde{f}_i$ subject to $|\mathcal{S}| = N_t$. In order to introduce such constraint, define the following binary variable y_i as:

$$y_i = \begin{cases} 1 & \text{if user } i \text{ is selected} \\ 0 & \text{otherwise} \end{cases} \tag{5.55}$$

In the same way it can be defined a set of binary variables x_{ij} that relate the common coefficient ϱ_{ij} of two users as:

$$x_{ij} = \begin{cases} 1 & \text{if both users } i \text{ and } j \text{ are selected} \\ 0 & \text{otherwise} \end{cases} \tag{5.56}$$

The mathematical model for the user selection problem based on the channel norms and correlation coefficients is given by:

$$\begin{aligned} \text{maximize} \quad & \sum_i a_i y_i + 2 \sum_i \sum_{j=i+1} b_{ij} x_{ij} \\ \text{subject to} \quad & \sum_i y_i = N_t \\ & y_i + y_j \leq 1 + x_{ij}, \forall i, j \\ & x_{ij} \leq y_i, \quad \forall i, j \\ & x_{ij} \leq y_j, \quad \forall i, j \\ & y_i \in \{0, 1\}, \quad \forall i \\ & x_{ij} \in \{0, 1\}, \quad \forall i, j \end{aligned} \tag{5.57}$$

variables y_i, x_{ij}

where (5.57) is a binary programming problem that generalizes the objective function optimized by Algorithm 5.1. The advantage of this formulation is that the order in which the users are selected has no impact in the orthogonality of the elements of $\mathbf{H}(\mathcal{S})$, i.e., the negative effects of selecting local optimum users in each iteration is canceled. The solution to the user selection problem is given by the binary variables y_i and power allocation based on water-filling is performed as described in Section 5.4. Observe that a conversion from $\tilde{f}_i$ to f_i is not required because the set of scheduled users is defined as $\mathcal{S} = \{i \in \Omega : y_i = 1\}$. Since the objective function is convex and the constraints are given by affine functions, this problem can be solved by the pseudo-dual simplex method [10] for integer programs or using standard optimization packages [165, 166]. Moreover, problem (5.57) always has a feasible solution because the only constraint that might lead to infeasibility is the equality constraint that is always met due to the fact that $K \geq N_t$. Problem (5.57) is a relaxed version of (5.4) and finds a suboptimal solution to the user selection problem owing to the nature of the coefficients b_{ij} which is discussed below.

5.6.3 Sub-optimality of IPP and ILP

The NSP found by (5.14) has a direct relationship with the correlation coefficients η of the users in $\mathcal{S}$ and the channel vector $\mathbf{h}$ of the candidate user in Ω. It can be observed that the product that scales the squared channel norm of user i in (5.46) contains all the information of the correlation coefficients of elements of $\mathcal{S}$ which resembles the product (5.22). However, (5.46) considers redundant information about how all elements in $\mathbf{H}(\mathcal{S})$ interact with $\mathbf{h}$, which results in a suboptimal evaluation of (5.22). As a matter of fact, the NSP approximation in (5.46) is an upper bound of the NSP which is elaborated upon in Section 6.4.4. Notice that as K grows, the probability that the basis of $Sp(\mathbf{H}(\mathcal{S}))$ can describe a new candidate user's channel $\mathbf{h}$ decreases. Therefore, the gap between the product of the correlation coefficients and the product of the partial correlation coefficients reduces as well. This characteristic is used in [134] to prove that when $K \to \infty$ the performance of a SVD-based scheduling algorithm that generates a quasi-orthogonal set of user by approximating (5.21) achieves asymptotical optimal user selection performance.

The optimum metric for user selection varies according to the precoding scheme that is implemented. For the case of ZFDP, since (5.46) considers redundant information when all terms ϱ are multiplied, lack of accuracy can be compensated by the elimination of the non-causally known interference. In the case of ZFBF the orthogonality among selected channels plays a more important role in terms of sum rate maximization. In order to compensate the lack of knowledge of the partial correlation coefficients in (5.22), larger values of the power loss factors are considered. This means that user selection described in Algorithm 5.1 for ZFBF is performed using $\hat{\varrho}_{ij}$ instead of ϱ_{ij}. Due to the fact that $\hat{\varrho}_{ij} \leq \varrho_{ij}$ (with equality when the channels are uncorrelated) the projection power loss factor increases its value, and in this way the poor orthogonality between channels has a higher impact when the squared channel norms are scaled in (5.46).

5.7 Performance Evaluation

In this section two types of numerical evaluations are presented. The first type of evaluation is a comparison between the metrics defined in Section 5.5 and its capability to identify the set of users that maximizes the sum rate. In the second type of evaluation, the proposed user selection algorithms of Section 5.6 are compared to several state-of-the-art algorithms, namely the semi-orthogonal user selection (SUS) [40] with threshold parameter α_{th}, and the NSP based approach [136, 137]. The upper bound of the sum rate is given by the expected value of the solution of (5.4) found by exhaustive search. In order to highlight the contribution of multiuser diversity, the performance is compared to two simplistic user selection approaches, one based on the maximum channel gain (MCG) criterion (selecting the N_t users with largest channel norms), and a second approach performing round robin scheduling (RRS).

The performance of Algorithm 5.1 (IPP) is also compared to two greedy class-A algorithms, one proposed by Dimic and Sidiropoulos (D-S) [138], and the other proposed by Karachontzitis and Toumpakaris (K-T) [124]. The solution of the integer linear program (ILP) optimization in (5.57) is presented and used as an upper bound of the performance of Algorithm 5.1 (IPP) and compared to the optimum solution of (5.4). The simulations consider complete CSI, fading channels are generated following a complex Gaussian distribution with unit variance, and the average sum rate is given in bps/Hz. Since the system performance is evaluated via Shannon capacity by means of (5.7) and (5.12), the results are independent of the specific implementation on the coding and modulation schemes, which provides a general design insight.

5.7.1 Optimality of the Channel Metrics

The results of this section compare the average sum rate achieved when optimal user selection is performed over different metrics of $\mathbf{H}(\mathcal{S})$. The curves displayed in Fig. 5.4 are obtained by optimally solving the combinatorial problems previously introduced in Section 5.5, whose solutions are employed to evaluate (5.4). Notice that these results are upper bounds of the average sum rate for each metric, which implies that any class-B user selection algorithm can achieve at most the same performance for its optimized metric. The sum rate achieved by $\mathcal{S}_\omega$ defined by (5.15) approaches the optimal sum rate achieved by $\mathcal{S}^\star$ paying a computational cost of matrix product and inversion operations for each possible set $\mathcal{S}$. Due to the properties of water-filling the sum rate is maximized when the terms b_i have larger and uniform values. For low values of K there exists a performance gap between $\mathcal{S}^\star$ and $\mathcal{S}_\omega$ because the probability that the terms b_i have large uniform values is small. As K grows this probability increases and the performance gap vanishes. The average sum rate achieved by $\mathcal{S}_\xi$ defined by (5.30) is a lower bound of the NSP metric (5.14) for all K with a performance gap not larger than 3% w.r.t. $\mathcal{S}^\star$. This result suggest that (5.29) is a good candidate metric for user selection since it achieves a good trade-off between complexity and performance.

The orthogonality defect $\delta(\mathbf{H}(\mathcal{S}))$ cannot identify with accuracy the set that maximizes the sum rate because this metric only reflects the degradation of the product of the eigenvalues of $\bar{\mathbf{H}}(\mathcal{S})$ with respect to the channel gains. Consider that the optimum solution of (5.35) yields $\delta(\mathbf{H}(\mathcal{S}_\delta)) = 1$. This result may indicate that $\mathbf{H}(\mathcal{S}_\delta)$ forms an orthogonal basis yet minimizing the orthogonal defect ignoring the maximum achievable denominator in (5.34) leads to a poor user selection in terms of sum rate maximization. Numerical results shows that this effect worsens as K grows. The optimum channel matrix $H(\mathcal{S}^\star)$ does not achieve perfect

ϵ-orthogonality or pairwise orthogonality among its components. The sum rate attained with the solution of problems (5.35), (5.37), and (5.39) degrades as K grows. This is due to the fact that for high multiuser diversity the probability of finding a set of users that minimizes the dispersion of the eigenvalues of $\bar{\mathbf{H}}(\mathcal{S})$ increases but this does not imply that the maximum sum of effective channel gains or transmitted power is achieved. Metric (5.31) improves its performance as K grows since the probability that $H(\mathcal{S}_\Lambda)$ achieves both large $\sum_i \|\mathbf{h}_i\|^2$ and pairwise uncorrelated channels increases with K. The set $\mathcal{S}_\epsilon$ found by metric (5.19) may be not unique since two different sets may containing the same maximum η but their achievable sum rates may be quite different.

Fig. 5.5 illustrates the average sum rate achieved by each metric as a function of the SNR. Observe that the performance gap between the different metrics and the optimum solution increases up to a critical SNR and after that point the performance gap between curves remains constant. This indicates that the metrics performance depends on the SNR regime. The average sum rate is similar for other ZF-based beamforming techniques such as successive-ZF or ZF dirty-paper [25].

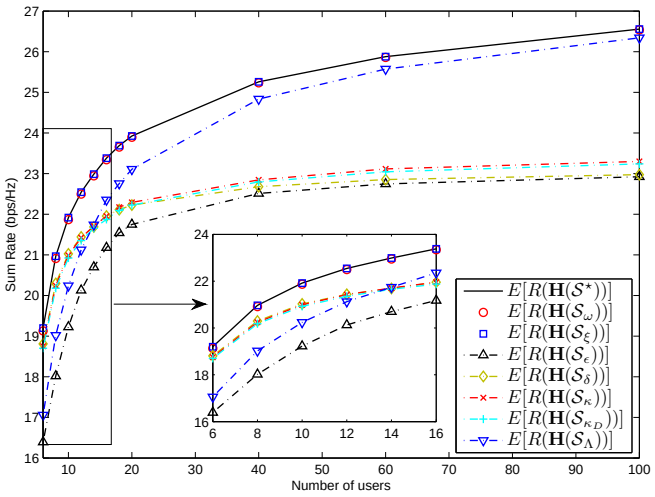

Figure 5.4: Average sum rate as a function of the number of users (K) with SNR=18(dB) and $N_t = 4$.

Table 5.2 summarizes the operations performed over the eigenvalues of $\bar{\mathbf{H}}(\mathcal{S})$ and the matrix operations required to compute each metric. Keep in mind that to solve any of the combinatorial problems in Section 5.5 the time complexity of the sort operation $O(L_{N_t} \log(L_{N_t}))$ is scaled by L_{N_t} times the complexity required to compute each metric. For metrics based on the coefficient of correlation (5.19) and (5.31) the associated time complexity is $O(N_t)$ since only inner products and vector norms are required.

5.7.2 Throughput (R) vs number of active users (K)

In Fig. 5.6 and Fig. 5.7, it is compared the throughput performance of different user selection strategies and Algorithm 5.1 w.r.t. the number of competing users K. The performance of ZFBF is highly susceptible to the characteristics of the set of selected users $\mathcal{S}$. IPP algorithm performs the user selection exploiting the information of the terms $\hat{\varrho}$. Since $\hat{\varrho}_{ij} \leq \varrho_{ij}$, the consequence is a more drastic reduction in the power projection in (5.46) due to the value

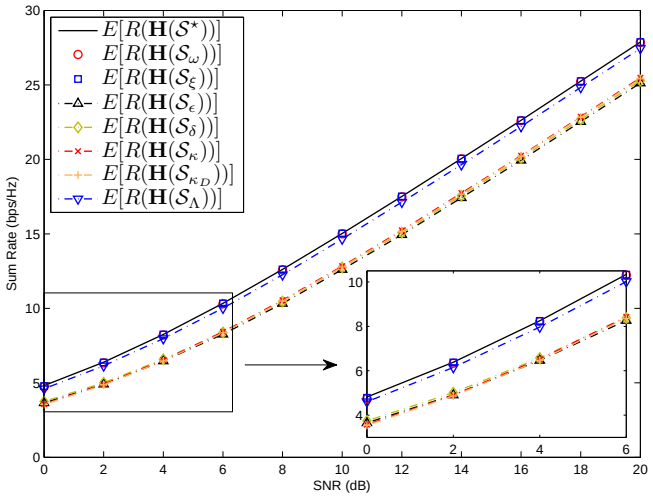

Figure 5.5: Average sum rate as a function of the SNR with $K = 20$ and $N_t = 4$.

Table 5.2: Metric Properties. λ_i is the ith eigenvalue of $\bar{\mathbf{H}}$, $f(\mathbf{H})$ is a function of λ. $\mathbf{HH}^H$, $\mathbf{H}^{-1}$, $\mathbf{eig}(\mathbf{H})$ are matrix product, inverse, and eigenvalue decomposition operations.

$f(\mathbf{H})$	$Tr(\dot{\mathbf{H}})$	$\det(\bar{\mathbf{H}})$	$\delta(\mathbf{H})$	$\kappa(\mathbf{H})$	$\kappa_D(\mathbf{H})$
$f(\lambda)$	$\sum_i \lambda_i^{-1}$	$\prod_i \lambda_i$	$\frac{\prod_i \|\mathbf{h}_i\|^2}{\prod_i \lambda_i}$	$\frac{\lambda_{\max}}{\lambda_{\min}}$	$\frac{\sum_i \lambda_i}{\lambda_{\min}}$
$\mathbf{HH}^H$	✓	✓	✓	-	-
$\mathbf{H}^{-1}$	✓	-	-	✓	-
$\mathbf{eig}(\mathbf{H})$	-	-	-	-	✓

of the correlation coefficient $\eta_{\mathbf{h}_i\mathbf{h}_j}$. Fig. 5.6 shows that IPP achieves a considerable portion of the average sum rate of the optimum selection, for instance when $K = 5$ the performance gap regarding the optimum user selection is about 11%. For $K = 10$, IPP achieves 90% of the optimum users selection performance. The parameter α_{th} is a threshold whose function is to drop users with correlation factor below to it as described in Subsection 5.6.1. In this particular example $\alpha_{th} = 1$ in order to guarantee that the set constraint in (5.4) is not violated. The objective of IPP algorithm is to achieve the performance of the greedy user selection based on the NSP. The performance of the IPP algorithm has an asymptotic behavior regarding the NSP approach as K grows. For $K = 20$, IPP achieves roughly 97% of the sum rate of the NSP based algorithms [136, 137].

A comparison of the IPP algorithm to the ILP optimization shows that the latter exploits more efficiently the user diversity as K grows. It is interesting that for $K \geq 20$ the ILP optimization achieves a better performance than the NSP approach in Fig. 5.6. This result suggest that there exists a critical value of K for which the user selection of the ILP optimization overcomes the selection performed using the metric defined in (5.14). For $K = 20$, the performance gap between the optimum user selection and the ILP optimization is less than 5%. This means that for given deployment N_t, there exists a finite value K_0 for which $\forall K > K_0$ the sum rate gap between the exhaustive search and the model (5.57) is negligible. However, the complexity of computing the solution of (5.57) grows exponentially with K

which is impractical for online implementations, but it is still an acceptable approximation to (5.4) compared to the likely large search space size of the optimum solution for moderate values of K.

The performance of the IPP is determined by the precoding scheme that is used. For ZFDP in Fig. 5.7, it can be observed that IPP performs as well as SUS but there is still a performance gap compared to the NSP approach. For $K = 20$, IPP achieves the same performance of the greedy selection of [124] and 98% and 99% of the sum rate of the optimum selection and the NSP approach respectively. For ZFDP and $K \geq 8$, the ILP optimization achieves better performance than IPP but is not effective enough to reach the performance of the NSP approach for low values of K. Nevertheless, for $K = 20$, the ILP optimization achieves 98% of the sum rate of the optimum selection. IPP shows an asymptotic performance as $K \to \infty$ with respect to the NSP approach and the optimum selection for both precoding schemes.

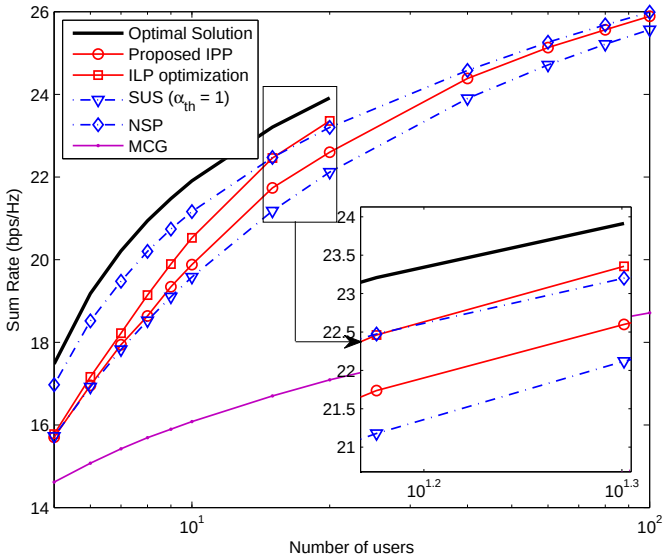

Figure 5.6: Average Sum Rate as a function of the number of users K for the ZFBF scheme with $SNR = 18$(dB) and $N_t = 4$.

5.7.3 Sum rate (R) vs SNR (P)

For Zero-Forcing-based beamforming and a given SNR (P) the maximum sum rate under the constraint $|\mathcal{S}| \leq N_t$ in (5.4), might be achieved by a set of selected users of cardinality strictly less than $rank(\mathbf{H}(\mathcal{S}))$ [25]. Nevertheless, from the properties of water-filling power allocation in (5.7), there exists a finite value P_0 (which depends on $\mathbf{H}(\mathcal{S})$) for which $\forall P \geq P_0$, (5.4) is solved by a subset of cardinality N_t. Notice that since the greedy class-A algorithms in [138] and [124] obey the constraint $|\mathcal{S}| \leq N_t$, the sum rate that they achieve for $P < P_0$ is higher than the capacity of the optimal solution in (5.4) but the number of scheduled users may be less than in a class-B algorithm. This phenomenon can be observed in Fig. 5.8 where for a given number of user $K = 10$, the value of $P_0 \approx 10$(dB) and the optimum solution of (5.4) is always better than the solution of the algorithms in D-S [138] and K-T [124]. It is

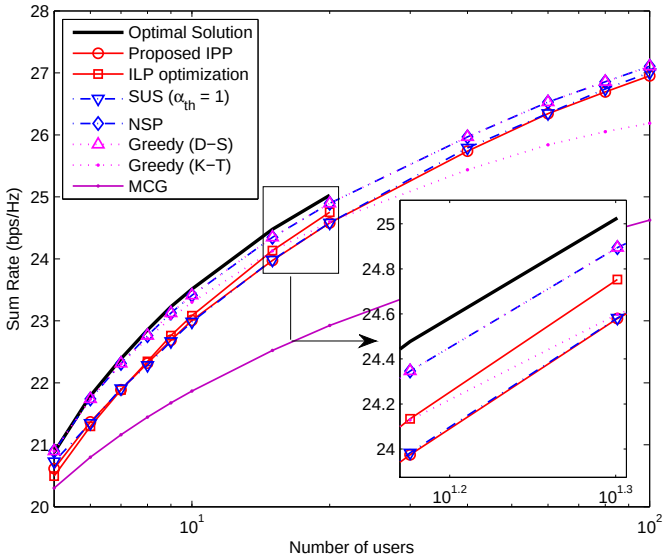

Figure 5.7: Average Sum Rate as a function of the number of users K for the ZFDP scheme with $SNR = 18$(dB) and $N_t = 4$.

worthy to point out that the optimum user selection here presented is found in a search space of size $\binom{K}{N_t}$ for a class-B algorithm, whilst the search space of a class-A algorithm [124, 138] has size of $\sum_{n=1}^{N_t}\binom{K}{n}$, which has no constraints on the minimum number of selected users. Therefore, the optimum solution shown in the results is valid only for class-B algorithms and presenting results for class-A algorithms attempts to highlight the difference between both classes. Consider the high SNR regime ($10 \leq P \leq 20$) in Fig. 5.8, the performance gap between IPP and the optimum solution ranges from 14% to 9% and for the NSP approach the performance gap goes form 9% to 4% in the same SNR range. For the case of ZFBF, the ILP optimization achieves a better approximation to NSP than the IPP approach. However, in the case of ZFDP in Fig. 5.9, the performance gap between IPP and the ILP optimization is about 1%, and both approaches achieve roughly 98% of the optimum selection capacity for SNR of 20(dB). An interesting fact is that the MCG selection achieves 93% the optimum selection capacity for $K = 10$ and $P = 20$(dB) under ZFDP. This indicates that for the high SNR regime, channel gains play a more important role for the user selection process in scenarios where nonlinear precoding is implemented. The performance of a class-B algorithm depends on the multiuser diversity and the SNR regime.

5.7.4 Cardinality of $\mathcal{S}$ and Ω

The cardinality of the set $\mathcal{S}$ is conditioned by the class of the algorithm that is implemented, its parameters, and the type of precoding that is used. In Fig. 5.10 it is analyzed the average value of the ratio $|\mathcal{S}|/N_t$ for (a) ZFBF and (b) ZFDP. Such ratio indicates if full spatial multiplexing is achieved. In the case of ZFBF, notice that both class-A algorithms S-D and K-T [124, 138] require $K \geq 20$ in order to achieve the maximum cardinality of $\mathcal{S}$. To exemplify the inconvenience of designing an algorithm dependent of non-dynamic parameters, notice that setting a wrong value to the parameter α_{th} of the SUS algorithm might lead to

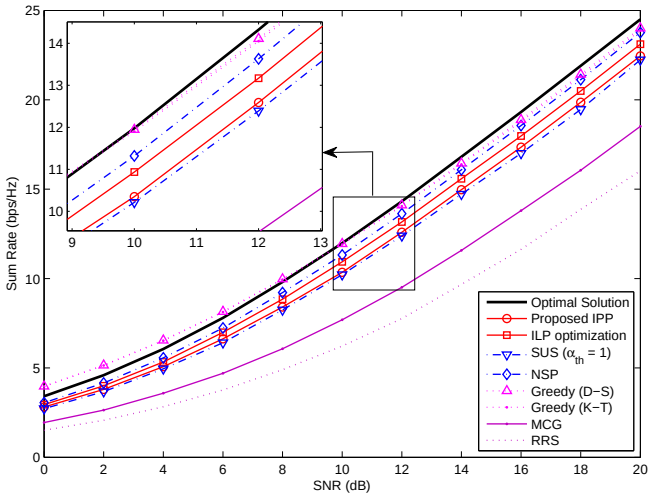

Figure 5.8: Average Sum Rate as a function of the SNR for ZFBF scheme with $K = 10$ and $N_t = 4$.

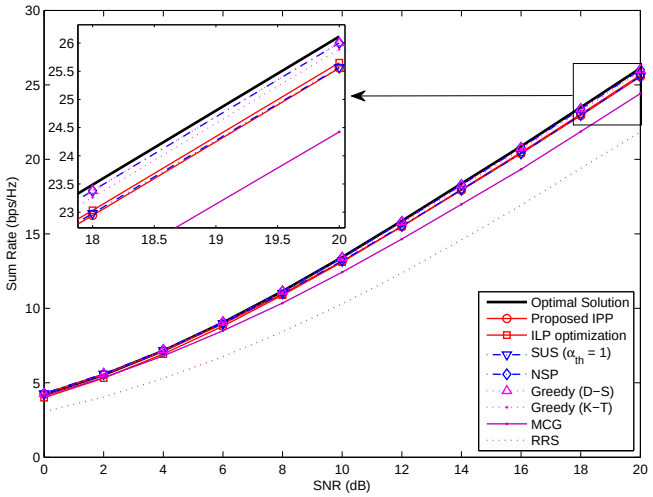

Figure 5.9: Average Sum Rate as a function of the SNR for ZFDP scheme with $K = 10$ and $N_t = 4$.

---

a degradation of both the cardinality of the set of selected users and the sum rate. For the case of ZFDP observe that the robustness of the precoder allows us to schedule N_t users in both classes of algorithms. This has a direct impact in the achieved fairness owing to the cardinality of $|\mathcal{S}| = N_t$.

The reduction of the set Ω per iteration becomes relevant for high values of K and N_t. The effects of (5.50) on the cardinality of the set of unselected users Ω are presented in Fig. 5.11 for (a) $N_t = 3$ and (b) $N_t = 4$. The figures show the average number of users kept in the set Ω each iteration of Algorithm 5.1 for different number of users. The first iteration always considers all K users to find the initial selected user. As the size of $\mathcal{S}$ increases the number of required users to achieve $|\mathcal{S}| = N_t$ reduces and (5.49) takes into account such decrement to assign the proper priority to $M_{a(i^*)}$.

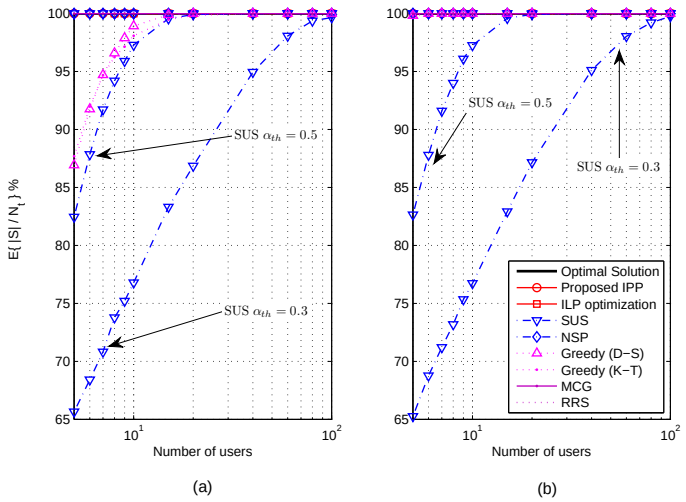

Figure 5.10: The metric $\mathbb{E}\{|\mathcal{S}|/N_t\}$ measures the degree of spatial multiplexing that is exploited for each scheduling algorithm considering $SNR = 18$(dB) (a) ZFBF and (b) ZFDP.

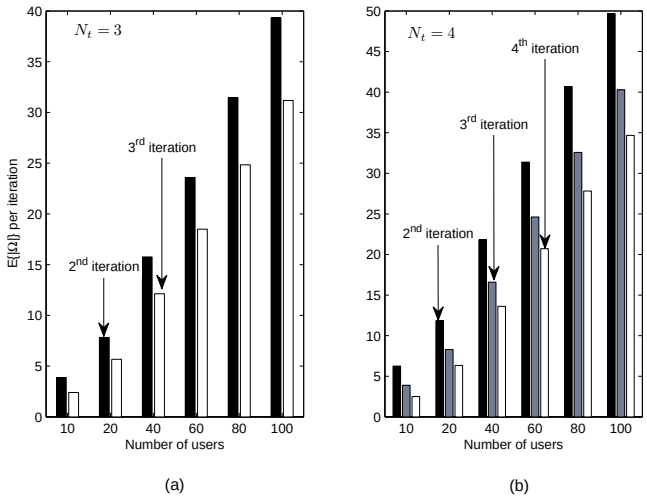

Figure 5.11: Average cardinality of the set of unselected users ($\mathbb{E}\{|\Omega|\}$) each iteration of the IPP algorithm with $SNR = 18$(dB) (a) $N_t = 3$ and (b) $N_t = 4$.

5.7.5 Complexity Analysis

The complexity required to solve (5.4) can be analyzed in two parts. The first one is the complexity required to implement each one of the precoders and the second one is the complexity of IPP. For the case of ZFBF, the precoding requires a $N_t \times N_t$ matrix inversion $\mathbf{W}$ and for ZFDP the evaluation of the beamforming weights requires a QR-type decomposition. For both coding schemes, this process is carried out after IPP finished the user selection process. The most costly operation in IPP is the evaluation of $(K^2 - K)/2$ inner products to define the correlation coefficients that can be done in time $O(K)$. Since this values does

not change along the selection process, they must be computed once and can be stored in memory. Notice that the evaluation of (5.41) requires a time $O(K)$ since only multiplications of real positive numbers are required and an sort operation performed in time $O(K\log_2(K))$. For the case where the set Ω reduces in one element per iteration and a total of N_t iterations are required the total complexity is $O(KN_t + N_tK\log_2(K))$. However, for the following iterations the time complexity of (5.47) is a function of the set of unselected user modified according the statistics of the projection power given by $M_{a(i^*)}$ and $w_{(N_t,\mathcal{S},\Omega)}$. This implies that each iteration will required a time $O(|\Omega|(1+\log_2(|\Omega|))) \approx O(|\Omega|)$ and Ω changes each iteration according to (5.50). The solution of (5.57) requires the optimization over $L_{ILP} = \frac{1}{2}K(K+3)$ binary variables in the objective function. This means that a total of $2^{L_{ILP}}$ configurations of those variables are available and the number of valid configurations depends on the constraints imposed over the binary variables. Regardless the existence of pseudo-polynomial algorithms that solve integer programs avoiding the evaluation of all configurations [10], real time computation of the solution of (5.57) is prohibitive for large values of K. Table 5.3 summarizes the time complexity of different user selection algorithms.

The proposed algorithms assume complete CSI. However, in practical systems it is challenging to guarantee this condition. Even if channel estimation is very accurate, it may exist an error in the CSI at the transmitter due to mobility and feedback delays. Several works (e.g., [40, 136, 155]) showed that outdated CSI may alter the orthogonality of the selected channels which degrades the performance of Zero-Forcing based transmission schemes. The authors in [155] showed that a significant fraction of the sum rate with complete CSI can be achieved if the ratio between the outdated channel at the transmitter and the estimation error is kept above a threshold. Therefore, as the frame lengths are designed so that the magnitude of the real channels and the errors due to outdated estimates maintain a given average ratio, the proposed user selection techniques are effective.

Table 5.3: Complexity comparison of user selection algorithms

Class A		Class B		
[138]	[124]	SUS [40]	NSP [137]	IPP
$O(KN_t^3)$	(KN_t^2)	$O(KN_t^3)$	$O(KN_t^3)$	$O(KN_t)$

5.8 Conclusions

In this chapter, the problem of sum rate maximization was addressed for multi-user single-cell systems. This fundamental problem in wireless communications is studied by decoupling user selection and beamforming design. Although they are independent tasks, the user selection must consider the characteristics of the implemented beamforming scheme which depends on the properties of the channel matrix of a selected set of users. The aim of the chapter was twofold: to study metrics of the channels that are closely related to the achievable sum rate, and to design algorithms that select users based on such metrics. It was developed a low complexity algorithm that finds a quasi-orthogonal set of users that maximizes the total sum rate for BC channels using ZFBF and ZFDP schemes. A fundamental relation between the correlation coefficients and the null space projection was used to perform user selection. The proposed IPP algorithm approximates the projected power using products of the correlation

coefficients, a metric that is based exclusively on the physical characteristics of the multi-user channels. The IPP algorithms was compared to different state-of-the-art algorithms and numerical results show small performance gaps between the optimum user selection and the proposed one.

By approximating the NSP as a product of correlation coefficients and transforming individual metrics per each competing user an integer programming model was derived, whose performance approximates the performance of the exhaustive search. The results obtained by numerical simulation indicate that the proposed scheduling designs can profit from fundamental information that characterizes the relation between wireless channels without implementing computational demanding operations.

Chapter 6

User Selection and Signal Design In Multi-Cell Systems

In this chapter the problem of sum rate maximization in multi-user multi-cell systems is tackled. A set of adjacent BS form a cluster in order to perform coordinated transmission to cell-edge users, and coordination is carried out through a CU. However, the message exchange between BSs and the CU is limited to scheduling control signaling and no user data or CSI exchange is allowed. In the considered multi-cell coordinated approach, each BS has its own set of cell-edge users and transmits only to one intended user while interference to non-intended users at other BSs is suppressed by signal steering (precoding). Two distributed linear precoding schemes are studied, Distributed Zero Forcing (DZF) and Distributed Virtual Signal-to-Interference-plus-Noise Ratio (DVSINR). Considering multiple users per cell and the backhaul limitations, the BSs rely on local CSI to solve the user selection problem [J4] . First, it is investigated how the SNR regime and the number of antennas at the BSs affect the effective channel gain (the magnitude of the channels after precoding) and its relationship with MUDiv. Considering that user selection must be based on the type of implemented precoding, metrics of compatibility (estimations of the effective channel gains) are designed which can be computed from local CSI at each BS and reported to the CU for scheduling decisions. Based on such metrics, user selection algorithms are designed so that a set of users that potentially maximizes the sum rate can be found. Numerical results show the effectiveness of the proposed metrics and algorithms for different configurations of users and antennas at the base stations.

6.1 Introduction

THE PERFORMANCE OF COORDINATED DOWNLINK TRANSMISSION with linear precoding in multi-antenna multi-cell systems has been an active area of research over the last years. Recent works (e.g. [28, 52, 84] and references therein) have shown that cooperation and coordination between clustered BSs improve rates, coverage, and efficiently suppress ICI which specially benefits cell-edge users [60]. BS coordination involves message exchange between neighboring cells and according to the level of coordination, multi-cell systems have been classified in three groups [12, 52, 167, 168]: interference aware (IA), joint processing/transmission (JT), and coordinated beamforming (CBF). In IA there is no information exchange among BSs, each transmitter serves its own set of users, and transmission parameters are adjusted in a selfish fashion by measuring ICI [52]. In contrast, in JT systems it is assumed that CSI and user data are globally available, full coordination is attainable though a CU, and each user

receives data from a group of coordinated BSs (cluster). The JT system can be interpreted as a broadcast channel [12] with distributed antennas and several radio resource management (RRM) tasks (e.g., scheduling, power control, precoding design, data queue control, etc.) extended from the single-cell systems can be applied (e.g., [28, 52, 60, 67, 72, 84, 169, 170]). However, such extensions must take into account backhaul rate limitations, CSI acquisition, joint transmission, and other system constraints [168]. In CBF the BSs need only data of the users in their own cells and they do not require to know the precoders and traffic of other BSs. The shared information is related to scheduling control signaling and partial or full CSI in order to mitigate spatial ICI. The BSs design precoding vectors towards the scheduled users so that the gain is two-fold: increasing the signal strength at the receivers and suppressing interference in the adjacent cells [168]. Efficient RRM schemes can be implemented under CBF using local CSI [171, 172] which relaxes the wideband backhaul and synchronization requirements [84].

Regardless the type of coordination between neighboring BSs, the inter-cluster interference problem arises if multiple clusters are taken into account, which can be dealt in two ways. The most straightforward way is to apply the principle of cellular planning (see Chapter 3.2) with frequency reuse [84, §5]. Using different radio resources in adjacent clusters (it can be dynamically allocated) mitigates or eliminates the inter-cluster interference. A second approach to reuse radio resource among different clusters is by means of inter-cluster coordination, where adjacent clusters implement interference mitigation techniques for the users at the edges of the clusters (e.g. [60]). For sake of simplicity and modeling tractability this chapter will consider a single cluster network with B BSs for the single carrier case.

6.1.1 Related Works

Depending on the system utility function that is optimized, there exist different strategies to achieve optimal power allocation and precoding design assuming that global CSI is known, and that the number of antennas at the transmitters can server all competing users (cf. [52, 67] for an in-depth survey). In the scenario where each BS serves only one user and CSI is not exchanged among BSs, the system model can be referred to as *interference channel* [12]. Recent works characterize its achievable rate region and jointly perform power allocation and precoding design (e.g., [61, 64, 67, 167]) under the assumption that the intended user of each BS has been previously selected by some procedure. However, for multi-user multi-cell scenarios each BS must select one user from its own pool of users before proceeding with precoding calculations. In this scenario the sum rate maximization is a complex combinatorial problem because the number of users is larger than the number of available spatial resources (antennas) and global CSI may not be available. The global performance is highly sensitive to the set of scheduled users, since selecting user k at BS b relies on local CSI and the associated channels of k at other BSs modify the precoders characteristics. Additionally, the multiplexing gain and ICI suppression depend upon the number of antennas at the BSs [168].

In the literature of single-cell MU-MISO systems with precoding based on Zero Forcing (ZF), the sum rates maximization problem is commonly tackled by decoupling the user selection from the power allocation and precoding design. As discussed in Chapter 5.5, the user selection can be performed based on the NSP (e.g., [40, 45, 137]) or an approximation of it (e.g., [152, 153]). The NSP provides an accurate measure of the *effective channel gain* (the channel magnitude after precoding), so that the user channels selected based on such a metric are spatially compatible or quasi-orthogonal. For ZF precoding, this means that the users

selected using the NSP can provide a close-to-optimal solution to the sum rate maximization problem in multi-user scenarios. Recent works on multi-cell systems have proposed extensions from single-cell user selection algorithms assuming that partial or global CSI is available at the CU (e.g., [12, 28, 60, 169, 170]). These extensions [28, 60] are centralized algorithms that exploit the concept of NSP to improve sum rates relying on global CSI at the scheduler. If global CSI is not available, distributed precoding and scheduling can still be implemented. For instance, LTE-Advanced standard [59, 84] considers distributed linear precoding such as signal-to-leakage-plus-noise ratio (SLNR) [12, 173] and ZF whose computation requires to know local CSI and the set of intended users. One strategy for joint distributed precoding and scheduling is to limit the exchange of CSI such that the clustered BSs jointly select users in a sequential fashion, i.e., the first BS selects its users and broadcast its decision, then the second BS selects its user based on the decision made by the first one and so on [84]. Another approach has been introduced in [172] where users selection, precoding design, and power allocation are treated as decoupled problems but their parameters are jointly updated at the CU. Results show that distributed RRM schemes with limited message exchange between BSs can improve system performance.

6.1.2 Contributions

In the considered system model, a set of adjacent BSs form a cluster and they coordinate their transmission strategies through a CU in order to serve a set of cell-edge users and mitigate ICI. The clustered BSs adopt the CBF transmission scheme where the data for an intended user is transmitted from one BS, whereas the impairments from the ICI are mitigated by coordinated precoding. Two distributed linear precoding schemes will be used: Distributed Zero Forcing (DZF) and Distributed Virtual Signal-to-Interference-plus-Noise Ratio (DVSINR derived from SLNR). It is assumed that each BS has its own set of intended users, no user data or CSI is exchanged between BSs, and the shared information between BSs and the CU is for scheduling control. In each scheduling instance the clustered BSs attempt to maximize the sum rate by selecting a set of users with particular characteristics. Optimizing the performance in the described scenario is a challenging task since global CSI is not available and the backhaul connection with the CU only supports scheduling control information. Moreover, selecting the best set of users whose channel characteristics maximize the sum rate is a combinatorial problem whose complexity grows exponentially with the number of BSs and users per cell [172].

To solve the user selection problem, and taking into account that the BSs implement either DZF or DVSINR, the key contributions of this chapter are summarized as follows.

- Initially, it is discussed how the instantaneous and average effective channel gains of DZF and DVSINR depend on the signal-to-noise-ratio (SNR) regime, the number of antennas at the BSs, and MUDiv. This insight of the precoder schemes is used to establish in which way local CSI must be process at each BS. Then precoder-based metrics of user compatibility are designed, i.e., depending on the type of precoding, a mapping from the CSI of each user to a real number is developed. The proposed metrics are estimations of the achievable effective channel gains and provide different levels of accuracy and complexity. The metrics are designed to operate in different system configurations based on the number of transmit antennas and BSs, i.e., an interference limited or a power limited system.

- The scheduling process must be perform at the CU by using the metrics reported by the BSs. This goal is accomplished by developing algorithms for user selection with different levels of complexity that properly combine the reported metrics. Once that a set of users has been selected, the decision is informed to the BSs and they compute either DZF or DVSINR based on the local CSI of the selected users.

- A pre-selection methodology is designed in order to reduce the number of competing users per BS. The method is a ranking-based per-antenna selection that preserves MU-Div in CBF systems, achieving considerable gains in terms of complexity reduction with marginal performance loss. Numerical results show that the proposed metrics and algorithms for user selection can achieve a large portion of the optimal sum rate (the benchmark is a fully centralized system) by exploiting local CSI with limited message exchange between BSs and the CU.

6.2 System Setup and Problem Formulation

Consider a multi-user multi-cell clustered network where a group of B adjacent BSs form a cluster. Each BS has N_t antennas, all users in the network are equipped with a single antenna, and define $\epsilon_D \triangleq \max\{N_t-(B-1),0\}$. The BSs only exchange messages of scheduling control through a CU and precoding design is performed at each BS using local CSI. The joint user selection and precoding design are performed for cell-edge users located in the cell-edge area defined by B BSs. The users are deployed within a circular area that spans a radius r_{coop} (a fraction of the cell radius r) illustrated in Fig. 6.1. The bth BS has one index set of edge users $\mathcal{S}_b$ and it only transmits data to one user in this set. Consider that $\mathcal{S}_b \cap \mathcal{S}_j = \emptyset$, $\forall j \neq b$ and the transmitted signal from BS b to user $k \in \mathcal{S}_b$ is: $\mathbf{x}_b = \sqrt{P_b}\mathbf{w}_b s_b$. P_b is the transmitted power, $\mathbf{w}_b \in \mathbb{C}^{N_t \times 1}$ is the unit norm precoder and s_b is the transmitted data symbol with $\mathbb{E}[|s_b|^2] = 1$, $\mathbb{E}[\|\mathbf{x}_b\|^2] = P_b$, and $P_b \leq P$ where P is the maximum available power. The received signal of user k is given by:

$$y_{bk} = \sqrt{P_b}\mathbf{h}_{bk}^H \mathbf{w}_b s_b + \sum_{j=1, j\neq b}^{B} \sqrt{P_j}\mathbf{h}_{jk}^H \mathbf{w}_j s_j + n_k, \tag{6.1}$$

where $\mathbf{h}_{bk} \sim \mathcal{CN}(0, \varrho_{bk}^2 \mathbf{I})$ of size $N_t \times 1$ is a flat Rayleigh fading propagation channel between user k and BS b and ϱ_{bk}^2 is the long-term channel power gain. The term $n_k \sim \mathcal{CN}(0, \sigma_n^2)$ is the noise. The receivers treat co-terminal interference as noise and the instantaneous signal-to-interference-plus-noise ratio (SINR) of user $k \in \mathcal{S}_b$ is defined as:

$$SINR_{bk} = \frac{P_b |\mathbf{h}_{bk}^H \mathbf{w}_b|^2}{\sum_{j=1, j\neq b}^{B} P_j |\mathbf{h}_{jk}^H \mathbf{w}_j|^2 + \sigma_n^2}. \tag{6.2}$$

6.2.1 Problem Formulation

In a cluster with B BSs, there exist $L = \prod_{b=1}^{B} |\mathcal{S}_b|$ user permutations with B users that can be chosen for simultaneous transmission. Each user in $\mathcal{S} = \bigcup_{b=1}^{B} \mathcal{S}_b$ has a unique index and all BSs know which indices belong to each BS. Let $\mathcal{G}_l$ $\forall l \in \{1, \ldots, L\}$ be a set of B users where each user is served by one BS and the users indices in the set l are the same

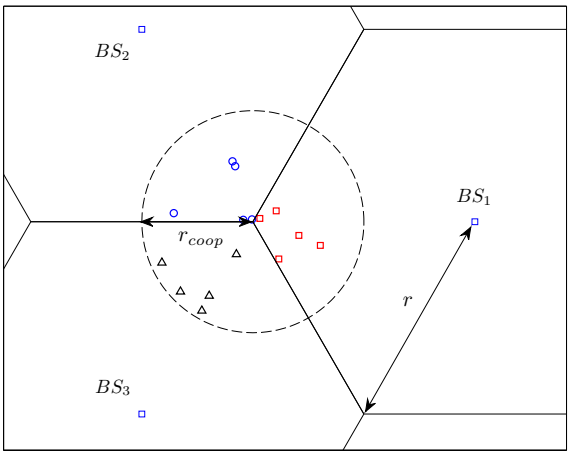

Figure 6.1: Deployment with $B = 3$, cell radius $r = 1$(km), and cell-edge cooperation area of radius $r_{coop} = 300$(m).

---

for all BSs. The set $\mathcal{G}_l$ has an associated channel matrix at the bth BS which is given by $\mathbf{H}_b^{(l)} \triangleq \{\mathbf{h}_{bi} : i \in \mathcal{G}_l\}$, i.e., all the local channels of the users grouped in $\mathcal{G}_l$. The sum rate maximization problem in the multi-user multi-cell scenario is defined as:

$$\begin{aligned} \underset{l\in\{1,\ldots,L\}}{\text{maximize}} \quad & \sum_{b=1}^{B} \log_2\left(1 + SINR_{bk}\right) \\ \text{subject to} \quad & \|\mathbf{w}_b^{(type)}(\mathbf{H}_b^{(l)})\|^2 = 1, \quad \forall b \in \{1,\ldots,B\} \end{aligned} \tag{6.3}$$

where $SINR_{bk}$ is defined in (6.2) for the user $k \in \{\mathcal{S}_b \cap \mathcal{G}_l\}$.

The precoding vectors $\mathbf{w}_b^{(type)}(\mathbf{H}_b^{(l)})$ $\forall b$ are functions of $\mathbf{H}_b^{(l)}$ at each BS for the given set l, and $type \in \{DZF, DVSINR\}$ is the implemented precoding technique which will be defined in the next section. The main objective is to find the set l that solves problem (6.3) which can be attained by taking advantage of the properties of $\mathbf{w}_b^{(type)}$. Such properties are used to exploit the local CSI and the SNR regime in order to evaluate the effective channel gains, i.e., $|\mathbf{h}_{bk}^H \mathbf{w}_b|^2$ $\forall k, b$, which are tightly related with the achievable rate.

6.3 Distributed Linear Precoding

In this section the properties and structure of two precoding techniques DZF and DVSINR are investigated. The main goal is to define underlying characteristics of the precoders which depend on the SNR regime and N_t, and quantify how those characteristics affect the instantaneous and average effective channel gains.

6.3.1 Distributed Zero Forcing (DZF)

Zero-forcing is a classic precoding strategy which removes the inter-user interference constrained by $B \leq N_t$. The conditions to achieve near Pareto-optimal rates with distributed ZF for the two-BS scenarios were presented in [64] and for B BSs generalized expressions to

compute $\mathbf{w}_b$ are provided in [12, 171]. Let $\tilde{\mathbf{H}}_{bk}$ be the aggregate interference matrix of user k given by:

$$\tilde{\mathbf{H}}_{bk} = [\mathbf{h}_{b1}, \ldots, \mathbf{h}_{b(k-1)}, \mathbf{h}_{b(k+1)}, \ldots, \mathbf{h}_{bB}], \tag{6.4}$$

and each term $\mathbf{h}_{bi}$ $\forall i \neq k$ corresponds to the channel between BS b and the user $i \in \mathcal{S}_i$. The matrix $\tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_{bk}} = null(\tilde{\mathbf{H}}_{bk})$ contains ϵ_D column vectors that are candidates to form $\mathbf{w}_b$ since they will produce zero interference to the other users in $\tilde{\mathbf{H}}_{bk}$. If $\epsilon_D > 1$ the elements of $\tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_{bk}}$ can be linearly combined to form the precoding vector as follows [171]:

$$\mathbf{w}_b^{(DZF)} = \tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_{bk}} \frac{(\mathbf{h}_{bk}^H \tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_{bk}})^H}{\|\mathbf{h}_{bk}^H \tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_{bk}}\|}, \tag{6.5}$$

and the received signal at user k has its phase aligned so that $\mathbf{h}_{bk}^H \mathbf{w}_b \in \mathbb{R}_+$.

Proposition 6.1. *The expected value of the effective channel gain of user $k \in \mathcal{S}_b$ served by BS b using DZF precoding with $\mathbf{w}_b$ given by (6.5) under constraint $N_t \geq B$ is defined as follows:*

$$\mathbb{E}\left[|\mathbf{h}_{bk}^H \mathbf{w}_b|^2\right] = \frac{\epsilon_D}{N_t} \mathbb{E}\left[\|\mathbf{h}_{bk}\|^2\right] \tag{6.6}$$

Proof. See Appendix A.1 □

6.3.2 Distributed Virtual SINR (DVSINR)

The ideal precoder technique would be able to balance between signal power maximization and interference power minimization and a heuristic way to find such balance is reached by maximizing the SLNR [12]. In [72] the authors show that it is possible to achieve Pareto-optimal rates in multi-cell transmission when the precoding vectors are given by:

$$\mathbf{w}_b^\star = \arg \max_{\|\mathbf{w}\|^2=1} \frac{\upsilon_{bk} |\mathbf{h}_{bk}^H \mathbf{w}|^2}{\sum_{j \neq k} \upsilon_{bj} |\mathbf{h}_{bj}^H \mathbf{w}|^2 + \frac{\sigma_n^2}{P_b}}, \tag{6.7}$$

where $\upsilon_{bk} \in (0, 1)$ which is a heuristic extension of the SLNR precoding [12, 173]. Then $\mathbf{w}_b^\star$ $\forall b$ are linear combinations of the maximal ratio transmission and ZF precoders and the coefficients υ_{bk} that optimally maximize the sum rate can be only computed with global CSI. If maximum ICI is accounted[1] ($\upsilon_{bk} = 1$ $\forall b, k$) the precoders that solve the virtual SINR maximization problem (6.7) are given by [72]:

$$\mathbf{w}_b^{(DVSINR)} = \frac{\mathbf{D}_{bk} \mathbf{h}_{bk}}{\|\mathbf{D}_{bk} \mathbf{h}_{bk}\|}, \tag{6.8}$$

where $\mathbf{D}_{bk} = \mathbf{C}_{bk}^{-1}$, $\mathbf{C}_{bk} \triangleq \rho_b^{-1} \mathbf{I}_{N_t} + \tilde{\mathbf{H}}_{bk} \tilde{\mathbf{H}}_{bk}^H$ is a $N_t \times N_t$ positive-definite Hermitian matrix and $\rho_b = \frac{P_b}{\sigma_n^2}$. The following result describes the relation between the eigenvalues of $\mathbf{D}_{bk}$ and the expected value of the effective channel gain.

[1] The authors in [167] showed that the coefficients υ_{bk} can define user weights that may represent, for instance, user priority.

Proposition 6.2. *The effective channel gain of the user $k \in \mathcal{S}_b$ served by the bth BS under DVSINR precoding constrained by $N_t \geq B$ can be approximated as follows:*

$$\mathbb{E}\left[|\mathbf{h}_{bk}^H \mathbf{w}_b|^2\right] \approx \mathbb{E}\left[\|\mathbf{h}_{bk}\|^2 J(\mathbf{eig}(\mathbf{D}_{bk}))\right] \tag{6.9}$$

where $\mathbf{w}_b$ is defined in (6.8), $J(\cdot)$ is the Jain's fairness index defined in (4.23), and $\mathbf{eig}(\mathbf{D}_{bk})$ is the vector that contains all the eigenvalues of matrix $\mathbf{D}_{bk}$.

Proof. See Appendix A.2 □

Proposition 6.3. *For DVSINR, given the matrix $\tilde{\mathbf{H}}_{bk} \in \mathbb{C}^{N_t \times (B-1)}$ and its corresponding $\mathbf{D}_{bk} \in \mathbb{C}^{N_t \times N_t}$ under constraint $N_t \geq B$ it holds that*

$$\lim_{\rho_b \to \infty} \left(J(\mathbf{eig}(\mathbf{D}_{bk})) - \frac{\epsilon_D}{N_t} \right) = 0, \tag{6.10}$$

which implies that $\exists \rho_0$ and $\forall \rho_b \geq \rho_0$ the expected value of the effective channel gain is upper bounded as follows:

$$\mathbb{E}\left[|\mathbf{h}_{bk}^H \mathbf{w}_b|^2\right] \leq \frac{\epsilon_D}{N_t} \mathbb{E}\left[\|\mathbf{h}_{bk}\|^2\right] \tag{6.11}$$

Proof. See Appendix A.3 □

The ICI for DVSINR is nonzero and for the high SNR regime the interference components in the denominator of (6.2) are usually neglected [72, 167]. The following result provides an approximation of the power that is leaked from clustered BSs using DVSINR precoding.

Proposition 6.4. *For DVSINR and $N_t \geq B$, the magnitude of the interference or leakage from the jth BS over the channel $\mathbf{h}_{jk} \in \tilde{\mathbf{H}}_{ji}$ $\forall j \neq b$ in the denominator of (6.2) for the user $k \in \mathcal{S}_b$ served by the bth BS can be approximated as follows:*

$$\mathbb{E}\left[|\mathbf{h}_{jk}^H \mathbf{w}_j|^2\right] \approx \mathbb{E}\left[\frac{\|\mathbf{h}_{jk}\|^2}{\epsilon_D N_t (\rho_j \lambda_{\min}(\tilde{\mathbf{H}}_{ji}^H \tilde{\mathbf{H}}_{ji}) + 1)^2} \right], \tag{6.12}$$

where $\mathbf{w}_j$ is a function of the matrix $\tilde{\mathbf{H}}_{ji}$ associated to the user i served by the jth BS.

Proof. See Appendix A.4 □

6.4 The Multicell User Selection

This section discusses the multi-user channel characteristics that affect the performance of DZF and DVSINR as well as their relation with MUDiv. Then problem (6.3) is addressed by decoupling[2] the user selection task from the precoding design. Based on the implemented precoding technique novel metrics to determine spatial compatibility (functions of the local CSI) are designed which will be used to develop user selection algorithms at the CU. The objective is to find a set of receivers whose associated channels can provide a close-to-optimal solution to the sum rate maximization problem.

[2]Decoupling both problems is a technique used in single-cell systems under precoding schemes with defined structure such as Zero Forcing (e.g., [40, 45, 137, 152, 153]) which can achieve close-to-optimal performance.

6.4.1 Linear Precoding and User Selection

Consider that $k \in \{\mathcal{S}_b \cap \mathcal{G}_l\}$ and let $\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l) \in \mathbb{C}^{N_t \times (B-1)}$ be the aggregate interference matrix of k which contains all channels of $\mathbf{H}_b^{(l)}$ except $\mathbf{h}_{bk}$.

6.4.1.1 DZF

This scheme is defined if $N_t \geq B$ and achieves $|\mathbf{h}_{bi}^H \mathbf{w}_b^{(DZF)}|^2 = 0$, i.e., zero inter-user interference $\forall i \in \mathcal{G}_l \setminus \{k\}$. From Proposition 6.1 observe that the average value of $|\mathbf{h}_{bk}^H \mathbf{w}_b^{(DZF)}|^2$ depends on N_t and ϵ_D. As the latter grows the effective channel gain is enhanced and consequently the SINR. However, results in Appendix A.1 show that the instantaneous effective channel gain is a function of the angle between $\mathbf{h}_{bk}$ and the basis of $\mathcal{V}_{bk} = Sp(\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l))$, the subspace spanned by the i.i.d. columns of $\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)$. The selected user at BS b needs to meet two conditions in order to improve system performance regardless the SNR regime: 1) maximize the local $|\mathbf{h}_{bk}^H \mathbf{w}_b^{(DZF)}|^2$ which means that channel magnitude and spatial compatibility (quasi-orthogonality) must be optimized jointly; 2) the channels $\{\mathbf{h}_{jk}\}_{j=1,j\neq b}^{B}$ of user k at other BSs should be spatially compatible to the other selected channels so that $|\mathbf{h}_{ji}^H \mathbf{w}_j^{(DZF)}|^2$, $\forall j \neq b$, $\forall i \neq k$ is also maximized. Moreover, $\forall k \in \mathcal{S}_b$ there exist $L_{bk} = \prod_{j=1,j\neq b}^{B} |\mathcal{S}_j|$ different precoders and a set $\mathcal{G}_l$ maximizing $|\mathbf{h}_{bk}^H \mathbf{w}_b^{(DZF)}|^2$ at BS b is, in general, not the best set at other BSs.

6.4.1.2 DVSINR

This scheme does not impose a constraint on N_t but its capability to combat inter-user interference depends on it. For a given user set $\mathcal{G}_l$, Proposition 6.2 shows that in the low and medium SNR regimes the expected value of $|\mathbf{h}_{bk}^H \mathbf{w}_b^{(DVSINR)}|^2$ depends on the magnitude of $\mathbf{h}_{bk}$ and the characteristics of $\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)$. In particular, the magnitude of each i.i.d. vector in $\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)$ and its singular values which directly modifies $J(\mathbf{eig}(\mathbf{D}_{bk}))$, cf. Appendix A.2. In the high SNR, Proposition 6.3 indicates that the expected value of $|\mathbf{h}_{bk}^H \mathbf{w}_b^{(DVSINR)}|^2$ is limited by ϵ_D and N_t similar to DZF. Since the impact of $\mathbf{D}_{bk}$ in the effective channel gain is dominated by ϵ_D eigenvalues associated to the null space of $\mathcal{V}_{bk}$, the selected user at each BS should meet the same conditions previously described for DZF. From results in Appendices A.2 and A.3 it can seen that at the low SNR the eigenvalues of $\mathbf{D}_{bk}$ have similar magnitudes and the BS can select its user selfishly based on the channel magnitudes regardless the characteristics of $\mathcal{V}_{bk}$. At medium SNR the user selection is more complicated since the instantaneous effective channel gain is modified by the weighted basis of $\mathcal{V}_{bk}$ where the weights are functions of ρ_b and ϵ_D, cf. Appendix A.2. Notice that because $|\mathbf{h}_{bi}^H \mathbf{w}_b^{(DVSINR)}|^2 \neq 0$, $\forall i \in \mathcal{G}_l \setminus \{k\}$ the achievable SINR (6.2) strongly depends on N_t and B. If $\epsilon_D > 0$ (power limited scenario) the amount of leaked power from BS j to the user k served by BS b is inversely proportional to $\epsilon_D N_t$. When $\rho_j \to \infty$ the leakage is also scaled by ρ_j^{-2} according to Proposition 6.4 and inter-user interference vanishes. The expression (6.12) reveals that for a fixed ρ_j the leakage is minimized if $\lambda_{\min}(\tilde{\mathbf{H}}_{ji}^H \tilde{\mathbf{H}}_{ji})$ is maximized, which occurs if the i.i.d. vectors in $\tilde{\mathbf{H}}_{ji}$ are quasi-orthogonal. For user selection purposes, at BS b the best set $\mathcal{G}_l$ should meet two conditions: 1) $\mathbf{h}_{bk}$ is quasi-orthogonal to $\mathcal{V}_{bk}$ (similar to DZF), and 2) the elements in $\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)$ are quasi-orthogonal. If $\epsilon_D = 0$ (interference limited scenario) a strategy for user selection based only on local CSI is hard to define because the channels of all user in $\mathcal{G}_l$ are coupled in the SINR

expression. In other words, accurate user selection in such scenario requires CSI exchange between BSs.

6.4.2 Metric for user selection: $N_t \geq B$

Due to the fact that global CSI is not available at the CU, centralized user selection (e.g., [28, 170, 174]) cannot be performed. In order to design semi-distributed user selection it is required to define the type of *scheduling control information* exchanged between the BSs and the CU. The metric g_{bl} is a function of the local CSI $\mathbf{H}_b^{(l)}$ so that $g_{bl} : \mathbb{C}^{N_t \times B} \mapsto \mathbb{R}_+$. Such mapping computes an approximation of $|\mathbf{h}_{bk}^H \mathbf{w}_b|^2$, i.e., it quantifies how profitable is to select the set $\mathcal{G}_l$ for transmission at the bth BS. Let $\mathbf{P}_{\mathbf{h}_{bk}} = \tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)(\tilde{\mathbf{H}}_{bk}^H(\mathcal{G}_l)\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l))^{-1}\tilde{\mathbf{H}}_{bk}^H(\mathcal{G}_l)$ be the projector matrix onto $\mathcal{V}_{bk}$, and $\mathbf{Q}_{\mathbf{h}_{bk}} = \mathbf{I}_{N_t} - \mathbf{P}_{\mathbf{h}_{bk}}$ the projector matrix onto the orthogonal complement of $\mathcal{V}_{bk}$ [143]. The proposed metric to estimate $|\mathbf{h}_{bk}^H \mathbf{w}_b|^2$ is given by:

$$g_{bl} = \|\mathbf{Q}_{\mathbf{h}_{bk}} \mathbf{h}_{bk}\|^2 + \alpha_{bk} \|\mathbf{P}_{\mathbf{h}_{bk}} \mathbf{h}_{bk}\|^2, \tag{6.13}$$

where α_{bk} is a function of the type of precoding scheme.

For DZF $\alpha_{bk} = 0$ for all ρ_b since the precoder takes the form $\mathbf{w}_b^{(DZF)} = \mathbf{Q}_{\mathbf{h}_{bk}} \mathbf{h}_{bk} / \|\mathbf{Q}_{\mathbf{h}_{bk}} \mathbf{h}_{bk}\|$ which is the direction of the projection of $\mathbf{h}_{bk}$ onto $Sp(\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l))^{\perp}$ (cf. Appendix A.1). For the case of DVSINR, $\mathbf{w}_b^{(DVSINR)} = \mathbf{h}_{bk} / \|\mathbf{h}_{bk}\|$ as $\rho_b \to 0$, i.e., the precoder is given by the matched filter and one must have $\alpha_{bk} = 1$ in order to meet $g_{bl} = |\mathbf{h}_{bk}^H \mathbf{w}_b^{(DVSINR)}|^2 = \|\mathbf{h}_{bk}\|^2$. When $\rho_b \to \infty$ the precoder is given by $\mathbf{w}_b^{(DVSINR)} = \mathbf{Q}_{\mathbf{h}_{bk}} \mathbf{h}_{bk} / \|\mathbf{Q}_{\mathbf{h}_{bk}} \mathbf{h}_{bk}\|$ and one must have $\alpha_{bk} = 0$ so that $g_{bl} = |\mathbf{h}_{bk}^H \mathbf{w}_b^{(DVSINR)}|^2 = \|\mathbf{Q}_{\mathbf{h}_{bk}} \mathbf{h}_{bk}\|^2$. Therefore, α_{bk} must change depending on the SNR regime and the characteristics of the i.i.d. vectors in $\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)$.

Proposition 6.5. *The effective channel gain $|\mathbf{h}_{bk}^H \mathbf{w}_b^{(DVSINR)}|^2$ is given by a nonlinear combination of the orthonormal basis of both $Sp(\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l))$ and $Sp(\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l))^{\perp}$. The metric (6.13) is an approximation of $|\mathbf{h}_{bk}^H \mathbf{w}_b^{(DVSINR)}|^2$ and a heuristic definition of the weight α_{bk} is given by:*

$$\alpha_{bk} = \frac{1}{(\rho_b \lambda_{\max}(\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)^H \tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)) + 1)^2}. \tag{6.14}$$

Proof. See Appendix A.5 □

6.4.3 Metric for user selection: $N_t < B$

In this scenario $\epsilon_D = 0$, DZF is not defined [72], and DVSINR precoding can be implemented but inter-user interference is unavoidable. Moreover, metric (6.13) does not provide information for user selection or cannot be computed. If $B - N_t = 1$ then $\mathbf{P}_{\mathbf{h}_{bk}} = \mathbf{I}_{N_t}$ and no useful information is extracted from (6.13). If $B - N_t > 1$ the matrix $\tilde{\mathbf{H}}_{bk}^H(\mathcal{G}_l)\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)$ is ill-conditioned[3] and $\mathbf{P}_{\mathbf{h}_{bk}}$ is no longer a projector matrix. Therefore, it is necessary to define a metric of the form $g_{bl} = \|\mathbf{h}_{bk}\|^2 f(\mathbf{H}_b^{(l)}, \rho_b)$. The function $f(\mathbf{H}_b^{(l)}, \rho_b)$ must become 1 as $\rho_b \to 0$ whilst its value should change according to the strength of $\mathbf{h}_{bk}$ and its spatial relation with $\mathcal{V}_{bk}$ as $\rho_b \to \infty$.

[3] Observe that $\tilde{\mathbf{H}}_{bk}^H(\mathcal{G}_l)\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)$ is a matrix of size $B-1 \times B-1$ which has N_t non-zero eigenvalues. When $B - N_t > 1$ the ratio $\lambda_{\max}(\tilde{\mathbf{H}}_{bk}^H(\mathcal{G}_l)\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l))/\lambda_{\min}(\tilde{\mathbf{H}}_{bk}^H(\mathcal{G}_l)\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)) \to \infty$ and the matrix is close to singular [121].

Consider two grouped users $k \in \{\mathcal{S}_b \cap \mathcal{G}_l\}$, $i \in \{\mathcal{S}_j \cap \mathcal{G}_l\}$, and define $m_b^{(ki)} = \|\mathbf{h}_{bk}\|^2/\|\mathbf{h}_{bi}\|^2$. $m_b^{(ki)}$ can provide a coarse estimation of the channel strength of the users regarding the bth BS. If $m_b^{(ki)} \approx 1$ this may suggest either that k and i are close to each other at the cell-edge, or that k is far from BS b and transmission over channel $\mathbf{h}_{bk}$ could be affected by strong interference. For $m_b^{(ki)} \gg 1$, k may be close to BS b or i is either far from BS b or experiencing deep fading. If $m_b^{(ki)} < 1$ fading is strong in $\mathbf{h}_{bk}$ and transmission may not be feasible. In order to quantify how strong and reliable for transmission is $\mathbf{h}_{bk}$ using local CSI, define the mapping $M_{bk}^{(l)}$ as:

$$M_{bk}^{(l)} = \frac{\|\mathbf{h}_{bk}\|^2}{\prod_{j \in \mathcal{G}_l \setminus \{k\}} \|\mathbf{h}_{bj}\|^{2/(B-1)}}, \tag{6.15}$$

where the denominator is the geometric mean of the squared norms of the column vectors of $\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)$. Using the geometric mean has two objectives: collecting in a single quantity the strength of the channels $\{\mathbf{h}_{bj}\}_{j \in \mathcal{G}_l \setminus \{k\}}$ and considering the effects of each magnitude equally[4] in the averaging operation.

It is required to estimate the spatial compatibility between all the elements of $\mathbf{H}_b^{(l)}$, the degradation due to correlation in $\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)$ and the effects of ρ_b. Define the metric for spatial compatibility as

$$\zeta_{bk}^{(l)} = \frac{\left|\det\left((\mathbf{H}_b^{(l)})^H \mathbf{H}_b^{(l)}\right)\right|}{\left|\det\left(\rho_b^{-1}\mathbf{I}_{N_t} + \tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)(\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l))^H\right)\right|}, \tag{6.16}$$

which is the ratio between the volume of a $B \times B$ matrix over the the volume of a $N_t \times N_t$ matrix. Recall that the determinant measures the volume spanned by the columns of a matrix. The more orthogonal the column vectors of a matrix, the larger the value of its determinant [121].

The heuristic metric for user selection is defined as

$$g_{bl} = \|\mathbf{h}_{bk}\|^2 \left(\alpha_{bk} + (1 - \alpha_{bk}) M_{bk}^{(l)} \zeta_{bk}^{(l)}\right), \tag{6.17}$$

where α_{bk} is given by (6.14). Observe that in the low SNR $\alpha_{bk} \to 1$ which yields $g_{bl} \approx \|\mathbf{h}_{bk}\|^2$. In the high SNR $\alpha_{bk} \to 0$ and the selection metric is $g_{bl} \approx \|\mathbf{h}_{bk}\|^2 M_{bk}^{(l)} \zeta_{bk}^{(l)}$.

6.4.4 NSP Approximation

The NSP operation $\|\mathbf{Q}_{\mathbf{h}_{bk}} \mathbf{h}_{bk}\|^2 = \|\mathbf{h}_{bk}\|^2 \sin^2 \theta_{\mathcal{V}_{bk}\mathbf{h}_{bk}}$ can be approximated by using the inner products of the elements of $\mathbf{H}_b^{(l)}$ which reduces the processing effort required to compute metrics (6.13) or (6.17). The term $\theta_{\mathcal{V}_{bk}\mathbf{h}_{bk}}$ is the angle between $\mathbf{h}_{bk}$ and the subspace $\mathcal{V}_{bk}$. For two independent channels $\mathbf{h}_{bk}$ and $\mathbf{h}_{bi}$ at the same BS, the spatial compatibility between them can be measured by the coefficient of correlation defined as [143]:

$$\eta_{\mathbf{h}_{bk}\mathbf{h}_{bi}} = \frac{|\langle \mathbf{h}_{bk}, \mathbf{h}_{bi} \rangle|}{\|\mathbf{h}_{bk}\| \|\mathbf{h}_{bi}\|}, \tag{6.18}$$

[4]This is not the case for the arithmetic mean since the magnitudes $\{\|\mathbf{h}_{bj}\|^2\}_{j \in \mathcal{G}_l \setminus \{k\}}$ may have a large variance in which case the smallest magnitudes would be neglected.

which was already introduced in Section 5.5 equation (5.19) for the single-cell scenario and is presented here for convenience. The coefficient $\sin^2\theta_{\mathcal{V}_{bk}\mathbf{h}_{bk}}$ that scales $\|\mathbf{h}_{bk}\|^2$ in a NSP operation can be computed as [143]:

$$\sin^2\theta_{\mathcal{V}_{bk}\mathbf{h}_{bk}} = (1-\eta^2_{\mathbf{h}_{bk}\pi(1)})\dots(1-\eta^2_{\mathbf{h}_{bk}\pi(i)|\pi(1)\dots\pi(i-1)}), \tag{6.19}$$

which is the same metric defined in (5.22) extended to the multi-cell scenario. If the correlation coefficients (6.18) are used instead of the partial correlation coefficients in (6.19) a suboptimal evaluation of $\sin^2\theta_{\mathcal{V}_{bk}\mathbf{h}_{bk}}$ can be computed. Using this approximation of the NSP, the reported metric to the CU by the bth BS for the user $k \in \{\mathcal{S}_b \cap \mathcal{G}_l\}$ is given by:

$$g_{bl} = \|\mathbf{h}_{bk}\|^2 \prod_{i\neq k, i\in\mathcal{G}_l} \sin^2\theta_{\mathbf{h}_{bk}\mathbf{h}_{bi}}. \tag{6.20}$$

Observe that metric (6.20) can be computed even if $N_t < B$ since (6.18) is independent of B and exists for all $N_t \geq 2$. If $N_t \geq B$ metric (6.20) is an upper bound of the NSP. This means that $\|\mathbf{h}_{bk}\|^2$ is scaled by a coefficient larger than $\sin^2\theta_{\mathcal{V}_{bk}\mathbf{h}_{bk}}$ which prioritizes the channel magnitude over the spatial compatibility when the user selection is performed. The relationship between the real and the approximated expected values of the NSP is presented in the following proposition.

Proposition 6.6. *For $N_t \geq B$ it holds that the average value of metric (6.20) is an upper bound of the average metric (6.13) with $\alpha_{bk} = 0$, i.e., the NSP, so that:*

$$\mathbb{E}\left[\|\mathbf{h}_{bk}\mathbf{Q}_{\mathbf{h}_{bk}}\|^2\right] \leq \mathbb{E}\left[\|\mathbf{h}_{bk}\|^2 \prod_{i\neq k, i\in\mathcal{G}_l} \sin^2\theta_{\mathbf{h}_{bk}\mathbf{h}_{bi}}\right] \tag{6.21}$$

Proof. See Appendix A.6 □

6.4.5 Exhaustive Search Selection over the Metrics

The optimal solution of (6.3) can be only found by exhaustive searching over the achievable rates of the sets $\mathcal{G}_l$ $\forall l \in \{1,\dots,L\}$. This task that requires global CSI at the CU and the computation of BL precoders in order to accurately evaluate the L possible achievable sum rates. A sub-optimal solution to (6.3) can be found by not exchanging full CSI with the CU but instead by reporting the metrics computed by (6.13), (6.17), or (6.20). Assuming that all BSs know the L ordered sets, the bth BS computes the metrics g_{bl} $\forall l$ and report them to the CU where the set that is chosen to perform coordinated transmission solves the following problem:

$$l^\star = \arg\max_{l\in\{1,\dots,L\}} \prod_{b=1}^{B} g_{bl}. \tag{6.22}$$

Bearing in mind that g_{bl} attempts to estimate the effective channel gains, the rationale behind the product in (6.22) is that for MISO transmission a set of users maximizing the product of their effective channel gains also achieves maximum sum rate [45]. In the studied scenario, taking the product of the metrics assigns the same priority to each independent metric g_{bl} $\forall b$. This means that the computation of $l^\star$ is not biased by a dominant metric $g_{bl} \gg g_{jl}$ $\forall j \neq b$ for a given set l, which would be only beneficial to BS b. Once that $l^\star$ has been found, the BSs use the matrices $\mathbf{H}_b^{(l^\star)}$ $\forall b$ to locally compute the precoders which are used to sub-optimally solve (6.3).

6.4.6 Search Space Pruning

Previously, it was discussed that solving (6.22) does not require global CSI but reporting L metrics from each BS to the CU. If the number of cell-edge users is large ($|\mathcal{S}| \gg BN_t$) computing the metrics for all user permutations L may become prohibitive. In single-cell systems authors in [13] showed that for fixed N_t and single-antenna users, the system capacity under spatial division multiple access scales by $N_t \log(\log(|\mathcal{S}_b|))$ at the bth BS. This result means that MUDiv provides a marginal contribution to the capacity enhancement unless $|\mathcal{S}_b| \to \infty$. Similar conclusions extend to multi-cell systems operating in JT mode (e.g., [169, 174]). Numerical results in [169] show that MUDiv is beneficial for BS cooperation when only a fraction of the total number of users is considered to participate in the selection process. For a multi-cell JT system employing ZF precoding [174], BN_t transmit antennas can serve at most the same number of single-antenna users, and low-complexity user selection algorithms can be extended from single-cell systems [40, 45, 137, 152, 153].

In the considered CBF scenario the objectives are to achieve multiplexing gain, decrease the solution space size of problem (6.22) by selecting a small fraction of competing users from $\mathcal{S}$, and preserve MUDiv when selecting the competing users. In order to find a subset $\hat{\mathcal{S}}_b \subseteq \mathcal{S}_b$ at BS b, let $\mathbf{h}_{bk} = [h_{bk1}, \ldots, h_{bkN_t}]^T$ be the channel of the user $k \in \mathcal{S}_b$ where h_{bkn} is the channel component of the nth transmit antenna, and consider the following. 1) For DZF efficient user selection must be focused on finding quasi-orthogonal users regarding the SNR regime (see Section 6.4.1). 2) For DVSINR efficient user selection in the low SNR is determined by the channel magnitude (see Section 6.4.2). 3) In the high SNR the effective channel gains of DZF and DVSINR are similar (cf. Proposition 6.1 and Proposition 6.3) and efficient user selection must find spatially orthogonal users. A fast way to find a set of quasi-orthogonal users in JT systems is by applying a ranking-based per-antenna selection as in [174]. The idea behind such selection is that for two user, k and i having $|h_{bkn}| > |h_{bkn'}|$ $\forall n' \neq n$, $|h_{bim}| > |h_{bim'}|$ $\forall m' \neq m$, and $\forall n \neq m$, their inner product decreases as the magnitude of each dominant antenna n and m increases, i.e., they become quasi-orthogonal.

It is required that the channel of the selected user $k \in \{\mathcal{S}_b \cap \mathcal{G}_l\}$ of BS b to be as orthogonal as possible w.r.t. the channels in $\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)$. Therefore, the per-antenna ranking can be used for pre-selecting the users with maximum per-antenna channel magnitude. In this way a user $k \in \hat{\mathcal{S}}_b$ will have a dominant antenna (spatial direction) n and it is likely that channels in $\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)$ do not have per-antenna channel magnitudes similar or closed to $|h_{bkn}|$ at the same antenna n due to path-loss effects, which guarantees certain degree of orthogonality.

Define the dominant user for the antenna n at BS b as

$$k_{bn} = \arg\max_{i \in \mathcal{S}_b} \; |h_{bin}|, \tag{6.23}$$

and let the user with the best channel magnitude be

$$k_{b(\max)} = \arg\max_{i \in \mathcal{S}_b} \; \|\mathbf{h}_{bi}\|, \tag{6.24}$$

where the subset of users that will participate in the selection process at the bth BS is defined as

$$\hat{\mathcal{S}}_b = \{k_{bn}\}_{n=1}^{N_t} \cup \{k_{b(\max)}\}. \tag{6.25}$$

This user pre-selection reduces the search space size since only considers the strongest users per spatial direction per BS. Including $k_{b(\max)}$ in the set $\hat{\mathcal{S}}_b$ guarantees that for DVSINR the

strongest user will be considered for selection. Observe that the index $k_{b(\max)}$ can be one or more of the indices k_{bn} $\forall n$ which may be repeated as well, and $|\hat{\mathcal{S}}_b|$ can be at most $N_t + 1$. Assuming that $|\mathcal{S}_b| \geq N_t + 1$, $\forall b$ the number of reported metrics per BS L_r that will be used to solve (6.22) is bounded as follows:

$$L_r = \prod_{b=1}^{B} |\hat{\mathcal{S}}_b| \leq (N_t + 1)^B \leq L = \prod_{b=1}^{B} |\mathcal{S}_b| \tag{6.26}$$

and notice that L_r is independent of $|\mathcal{S}_b|$ $\forall b$.

6.5 Numerical Results

In this section the performance of the joint distributed linear precoding and user selection is illustrated numerically. The results are obtained using the deployment described in Section 6.2 with $B = 3$, cell radius $r = 1$km, and cell-edge cooperation area of radius $r_{coop} = 300$m. For simplicity all BSs have the same number of users K. The long-term channel power gain is proportional to $1/d_{bk}^4$ where d_{bk} is the distance between user k and BS b. It is assumed perfect CSI at each BS, the average sum rate is given in bps/Hz, and the results are averaged over 10,000 channel realizations. The results are computed by assigning $P_b = P$ for all $b \in \{1, \ldots, B\}$ and the same SNR regime at the cell border to all BSs, i.e., $\rho = P/\sigma_n^2$.

The system performance benchmark is given by the optimal solution of problem (6.3) which is achieved by global CSI at the CU and is referred to as O-GCSI. In order to solve problem (6.22) two strategies are implemented: 1) considering all L user permutations, and 2) applying the search space pruning with L_r user permutations. For scenarios where $N_t \geq B$, the results obtained for (6.13) are referred to as O-MUS (metric of user selection) when L is consider, or R-MUS if L_r is used. Similarly, metric (6.20) is referred to as O-NSPA (NSP approximation) for L and R-NSPA for L_r. If $N_t < B$ the results for (6.17) are referred to as O-MUS2 and R-MUS2 for L and L_r respectively. In order to highlight how the proposed metrics exploit MUDiv their performance is compared to a selfish user selection where each BS transmits to its strongest user (maximum channel norm) referred to as Max-SNR.

6.5.1 Sum rate vs SNR

The average sum rate as function of ρ (dB) for DZF and DVSINR is displayed in Fig. 6.2 and Fig. 6.3 respectively. In Fig. 6.2 for the case $N_t = 3$, $B = 3$, and a target rate of 13bps/Hz the O-NSPA requires about 1dB extra to achieve the target compared to O-GCSI. For a target ρ of 10dB the O-NSPA has a gap about 1bps/Hz compared to O-MUS. The simulated scenarios considered $K = 10$ users per BS, O-GCSI, O-MUS, O-NSPA require to evaluate $L = 10^3$ metrics per BS while R-MUS, R-NSPA require $L_r \leq 4^3$ for $N_t = 3$. For $\rho = 10$dB R-MUS and R-NSPA achieve 96% and 91% of the optimal performance O-GCSI, which shows the effectiveness of the search space pruning for CBF systems under DZF precoding. The performance gap between R-MUS and R-NSPA w.r.t. O-GCSI is about 1% and 2.5% for $N_t = 4$ and $\rho = 10$dB. Notice that as N_t grows metrics (6.13) and (6.20) converge to the same value which is mainly determined by the squared channel norm, cf. Proposition 6.6. This effect can be seen for $N_t = 4$ and $\rho = 10$dB where the performance gap between O-GCSI and Max-SNR is about 8%.

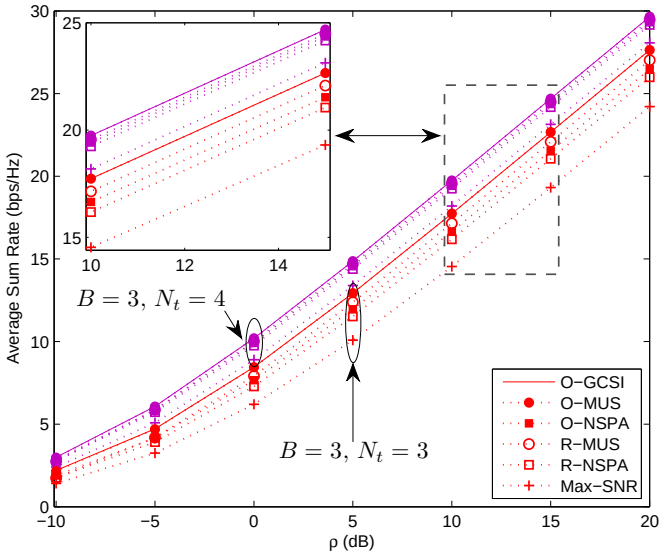

Figure 6.2: Average Sum Rate as a function of ρ(dB) for DZF precoding with $K = 10$, $B = 3$ and $N_t \in \{3, 4\}$.

Fig. 6.3 show results for DVSINR with $B = 3$ and $N_t \in \{3, 2\}$. In the case $N_t \geq B$ the performance of O-MUS and O-GCSI are quite closed and due to the heuristic nature of metric (6.13) and its parameter (6.14), the results for O-MUS are sub-optimal in the middle SNR range. The performance gap of O-MUS is less than 3% in the SNR range $\rho \in [-10, 10]$ and such a gap vanishes for other values of ρ. In the case of O-NSPA for $N_t \geq B$, it achieves up to 96% of the rate of O-GCSI in the whole SNR range. For $N_t = 3$ and $\rho = 10$dB, the performance gap between R-MUS and R-NSPA w.r.t. O-GCSI is about 3% and 5%.

In Section 6.4 was discussed that for the interference limited scenario, $N_t < B$, the SINR (6.2) of user $k \in \mathcal{S}_b$ depends on all its channels $\{\mathbf{h}_{jk}\}_{j=1}^{B}$. However, BS b only knows $\mathbf{h}_{bk}$ and an accurate user selection must take into account both the effective channel gain over the intended and unintended (leakage) channels, unlikely the case $N_t \geq B$. The figure shows that considering all user permutations L for metric (6.17), O-MUS2, is highly efficient in the low SNR regime and it achieves up to 91% of the sum rate of O-GCSI when $\rho = 20$dB. In contrast, O-NSPA cannot exploit MUDiv efficiently and only achieves 78% of the O-GCSI performance at the same SNR. Accounting for the search space pruning, R-NSPA and R-MUS2 attain 79% and 73% of the O-GCSI performance, respectively. These results show the effectiveness of the propose metric (6.17) and highlight the fact that one relies on L metrics per BS in order to achieve acceptable performance and compensate the lack of CSI knowledge of other BSs.

6.5.2 Sum rate vs K

Fig. 6.4 shows the average sum rate as a function of the number of users (multiuser diversity) K for DZF and $\rho = 10$dB. The figure illustrates the average sum rate of two scenarios where $N_t = B$ and $N_t > B$. In the CBF scenario, numerical results show that the set of users maximizing the product of their effective channel gains (computed with local CSI)

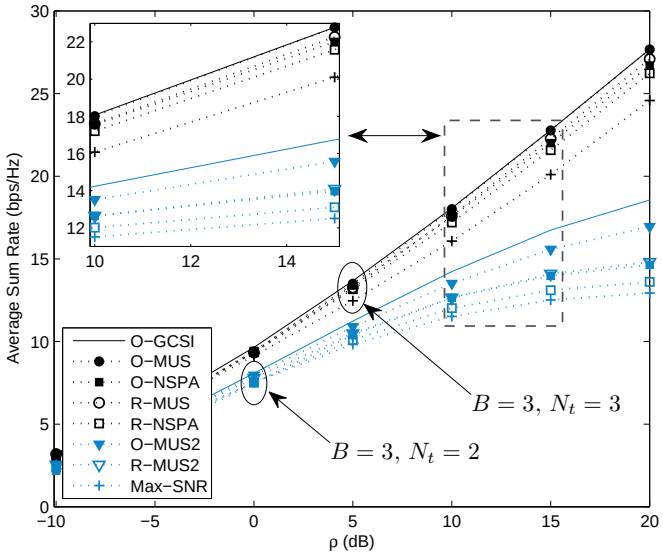

Figure 6.3: Average Sum Rate as a function of ρ(dB) for DVSINR precoding with $K = 10$, $B = 3$ and $N_t \in \{2, 3\}$.

---

achieve maximum sum rate for DZF[5]. For $K = 12$ O-MUS and O-GCSI overlap, O-NSPA attains up to 94% of the sum rate of O-GCSI for $N_t = 3$ and 98% for $N_t = 4$. The performance gap between O-MUS and O-NSPA reduces considerably by only adding one extra antenna per BS. This implies a large gain in terms of computational complexity when metric (6.20) is used instead of (6.13) for system configurations where $N_t > B$. The performance of R-MUS and R-NSPA illustrates the benefits of the proposed search space pruning. For $K = 12$, $N_t = 3$, and $B = 3$ the sum rate gap between R-MUS and O-MUS is less than 4% while the gap between R-NSPA and O-NSPA is less than 3% but the gain in terms of computational load is remarkable since $L_r \leq 64 < L = 1728$.

Fig. 6.5 shows the sum rate as a function of K for DVSINR and $\rho = 10$dB. The figure shows the performance for three scenarios with fixed $B = 3$ and $N_t \in \{4, 3, 2\}$. For $N_t = 4$ and $K = 12$, O-MUS and O-NSPA achieve 99% and 98% respectively of the benchmark sum rate. For $N_t = 3$ and $K = 12$, O-MUS and O-NSPA achieve 99% and 97% of the O-GCSI performance respectively. When $N_t = 2$, the O-MUS2 and O-NSPA achieve 94% and 89% of the optimal sum rate, respectively. The computational gains due to search space pruning are larger when $N_t \geq B$ and slightly reduce when $N_t < B$. For the latter scenario the performance gap between O-MUS2 and R-MUS2 is about 7% while the gap between O-NSPA and R-NSPA is about 6% for $K = 12$.

The results in all scenarios illustrate the effectiveness of the proposed metrics, two of them capturing more accurately spatial compatibility of the multi-user channels given by (6.13) and (6.17), and a third one given by (6.20) less computationally demanding and independent of the relation between B and N_t.

[5]Results in [45] showed that in single cell scenarios under Zero Forcing based precoding, the users that maximize the product of the effective channel gains achieve also maximum sum rate.

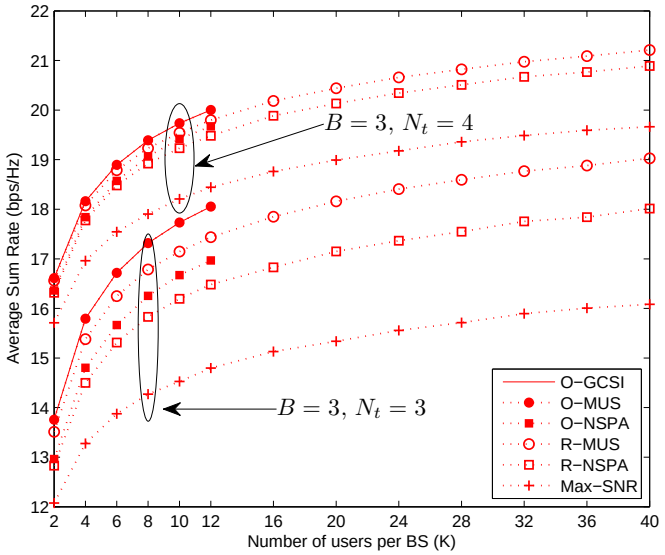

Figure 6.4: Average Sum Rate as a function of the number of users per BS (K) for DZF with $\rho = 10(\text{dB})$, $B = 3$, and $N_t \in \{3, 4\}$.

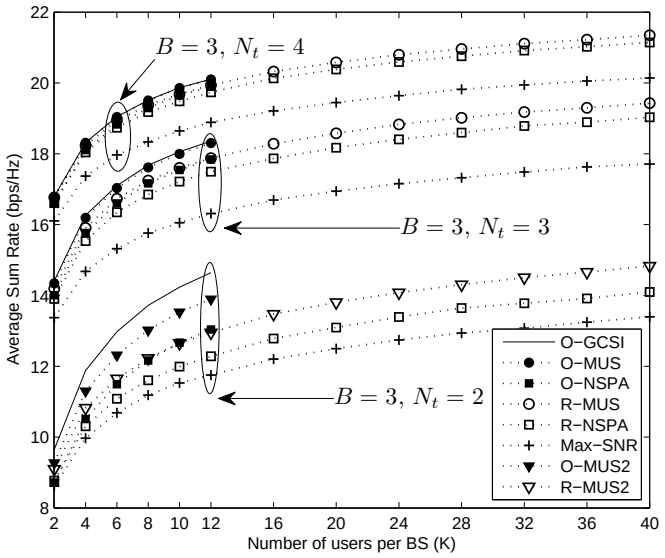

Figure 6.5: Average Sum Rate as a function of the number of users per BS (K) for DVSINR with $\rho = 10(\text{dB})$, $B = 3$, and $N_t \in \{2, 3, 4\}$.

6.6 Conclusions

In this chapter the sum rate maximization problem was addressed in multi-cell systems for the CBF mode with limited message exchange between BSs. Considering that CSI is not globally available two distributed linear precoding schemes with defined structures were adopted, DZF and DVSINR. Their characteristics and associated effective channel gains were studied.

It was shown that user selection must be perform based on the precoding technique that is implemented and channel metrics were designed in order to exploit local CSI for scheduling purposes. The objective of the metrics was to provide an estimation of the achievable effective channel gains for each precoder technique. For interference limited scenarios a metric for DVSINR was designed which allows to efficiently identify spatially compatible users. A third metric was proposed to avoid dependance on the number of antennas at the BSs and to reduce computational processing since only requires the computation of inner products. Finally, a method for search space pruning was developed which dramatically reduces the number of metrics reported from the BSs to the CU and preserves MUDiv. The proposed algorithm and metrics for user selection were assessed by simulations and numerical results show their potential to improve performance in coordinated multi-user multi-cell systems with limited message exchange between BSs.

Chapter 7

Conclusions

The thesis in a nutshell: The research presented in the dissertation is a sequence of steps whose final goal is to design user selection, signal design, and resource allocation techniques over distributed antenna systems. The initial step was to quantify the potential gains of DAS in terms of frequency reuse and study its resilience against interference. After establishing the DAS scenarios the multi-user component was introduced. In such scenario two problems were addressed, selecting one single-antenna user per distributed single-antenna and allocating resources to the links in such a way that sum rates were maximized. In order to exploit the advantages of MIMO, accounting for distributed multiple-antenna transmitters is mandatory. However, the user selection and resource allocation previously developed cannot be directly applied to the desired scenario due to the characteristics of the MU-MIMO channels. Therefore, a transition step was required whose objective was to design efficient user selection and resource allocation techniques for a single-transmitter with multiple antennas (modeled as a single-cell system). Finally, once that efficient user selection and resource allocation algorithms have been designed for a single-transmitter with $N_t > 1$, they were extended in order to operate over distributed multiple-antenna transmitter systems (modeled as a multi-cell system).

THIS CHAPTER summarizes the main results in this thesis and presents open problems and future lines of work. The following discussions endeavor to answer the question: *how does the research presented in this dissertation fit in the context of current cellular technologies and scientific research?*. By putting the proposed work and achieved results into perspective and taking into account recent references in the literature, the intention of the author is to provide a global view and the scope of this dissertation. The main discussed subject is resource allocation in wireless communication systems. The field of resource allocation is tightly linked to the following concepts: frequency planning, transmitter-receiver association, rate allocation, power control, user selection/scheduling, and cross-layer design. The studied system models consist of multiple antenna transmitters (co-located or distributed) and receivers, deployed in cellular scenarios. The main contributions of this dissertation are in the fields of user selection and link adaptation. In terms of user selection, efficient heuristic algorithms were designed for single-cell or multi-cell scenarios with the objective to identify the set of users that maximizes the achievable sum rate. Channel-aware user selection in a mobile communication system faces the question: how to share the radio resources available in the system between different users in order to achieve efficient resource utilization?; from the system perspective (operator), it is desirable to serve as many users as possible, while still satisfying the set of

QoS requirements that may exist (user perspective). Link adaptation refers to the dynamic allocation of rate, modulation, and power as a function of the wireless channels variations (radio-link quality). This concept is closely related to scheduling since a given set of users with good channels (spatially compatible and with large magnitudes) can utilize radio resources more efficiently. Centralized and semi-distributed algorithms were designed in order to provide efficient rate and power allocation in distributed antenna systems.

Chapter 3 Distributed Antenna Systems and Cellular Architectures [J1, C1]

■ Many techniques for interference mitigation in wireless systems are based on the concept of orthogonality. By transmitting over orthogonal (non-overlapping) dimensions either in space, time, frequency, or code, it is possible to avoid interference. One of the techniques used in cellular systems to combat ICI is to allocate frequency bands in non-overlapping coverage areas so that, close-to-orthogonal frequency allocation is achieved (frequency planning). Although universal frequency reuse is implemented by practical cellular systems (e.g. LTE), frequency planning will be necessary in future wireless systems since it is essential to mitigate ICI, specially in high frequency bands (3.5GHz and above) [5, 58].

■ In cellular systems where users are scheduled in orthogonal dimensions, the ICI is the main performance limiting factor. By deploying distributed antennas and employing different transmission schemes it is possible to optimize resources allocation and improve throughput figures. DAS can increase coverage, reduce power consumption, and provide resilience against ICI. Two transmission strategies were studied: blanket transmission which employs all distributed antennas to serve a user, and single-antenna transmission where the user is connected to the closest distributed antenna. It is worth noting that both schemes have been recently extended for MU-MIMO DAS scenarios, either integrating DAS with the CoMP transmission approach [53, 56] (see Chapter 2.4) or with frequency planning [54]. This also highlights the recent adoption of DAS within state-of-the-art transmission schemes and frequency planning strategies [55, 56].

■ The work presented in the chapter is focused on answering the question: which are the gains in terms of frequency reuse by deploying DAS w.r.t. CAS?. To answer that question two performance metrics are used, outage probability and throughput, which are functions of the achievable SINR. The analysis also considers that there is finite set of allowed SINR levels defined by a given set of MCSs. The results can be interpreted in two ways. On the one hand, DAS can considerable enhance performance for both transmission strategies compared to CAS, if frequency reuse is fixed. On the other hand, DAS allows a more aggressive frequency reuse since similar performance to CAS can be attained with a smaller frequency reuse factor.

■ *Lessons learned. i*) Integrating DAS as part of the access network infra-structure provides coverage and capacity gains in cellular systems. The key aspect to achieve such gains is the diversity that distributed antennas provide to the cellular architecture. *ii*) Different transmission schemes require different resource allocation policies, i.e., each scheme requires a particular assignment of resources among the active distributed antennas. *iii*) DAS is an upcoming technology that can increase the spectral efficiency in current cellular systems and its deployment and operation in future communications systems is tangible.

■ *Research Extensions.* *i*) HetNet [4] may include DAS either as part of the macro-cells, the micro-cells, or even both. Investigating more transmission schemes in such scenarios is necessary to understand the potential of DAS as part of HetNet, and to design efficient resource allocation policies. *ii*) Considering MU-MIMO scenarios implies, in general, heterogeneous (umbalanced) data load distribution within the coverage areas. Optimization algorithms should be designed in order to improve the area spectral efficiency measured in bps/Hz/m^2, i.e., optimizing HetNet demands to maximize the average data rate per unit bandwidth per unit area supported by a network [5].

Chapter 4 User Selection and Rate Allocation in Interference Channels [J2, C2, C3, C4]

■ Resource allocation is a fundamental component of communication networks which manages and optimizes the utilization of limited resources such as power and bandwidth. One of the key objectives of resource allocation is to provide reliable connection and maximize a global system objective (e.g. fairness or sum rates). A necessary operation to maximize performance is to mitigate interference (inter-user, inter-cell, and intra-cell) which can be done by controlling power allocation and the set of links that use the same wireless medium concurrently. By considering a distributed single-antennas system (DAS deployment) and several single-antenna users, a number of question questions related to resource allocation arise: how to perform simultaneous transmissions?, what transmission scheme should be used?, how to assign a user to each distributed antenna?, given a set of rate and power constraints, how to perform resource allocation in order to maximize the sum rate?, how to maximize the number of scheduled users?. The objective of this chapter was to address such questions.

■ One of the characteristics of DAS is that the antennas are geographically far apart and spread throughout the coverage area. Because of this reason, one can consider that each antenna can serve as an independent transmitter that coordinates its resource allocation though a central processing unit. In this way, a coordinated single-antenna transmission scheme can be employed and each distributed antenna must serve users that are close to it. When multiple users are uniformly deployed, it is needed to match each antenna with the user that will increase the total sum rate and minimize the interference for other matched users. Such a matching operation can be modeled as an assignment problem, whose optimal solution can be found by exhaustive searching over a finite set of feasible solutions. To address this problem, heuristic search algorithms were designed and assessed with positive results in terms of optimality and complexity.

■ Once that each distributed antenna has been matched with a user, it is necessary to determine the maximum rate supported in each link meeting a set of power constraints. Since the allowed rates are defined by a given set of MCSs the problem at hand becomes combinatorial. In the resource allocation problem feasibility issues emerge because it may not be possible to allow all the distributed antennas to transmit. The system model and the power constraints can be expressed in a matrix form and feasibility can be found by intrinsic properties of such a matrix. Based on the Perron-Frobenius theory of non-negative matrices or heuristic measures of power consumption, heuristic search algorithms were designed in order to solve the rate allocation problem and to handle feasibility issues.

■ *Lessons learned.* *i*) Understanding the structure of the solution space (SINR target region) is fundamental to avoid infeasible resource allocation and to solve such problem efficiently. *ii*) Resource allocation algorithms cannot fully optimize fairness and sum rates at the same time, which are fundamentally opposite goals. In this context, fairness implies to maximize the number of active links while maximizing the sum rate may require to keep several transmitters disconnected. *iii*) User removal techniques are required to guarantee feasibility and resource allocation policies should employ them according to the global system objective.
■ *Research Extensions.* *i*) The power allocation framework introduced in this chapter is general and flexible enough to be used along with beamforming design and user selection for distributed multiple-antenna transmitter systems. Although some works have been produced in this direction [67], considering the case where the allowed rates (defined by some MCS) and/or powers are given in finite sets still requires further research. *ii*) Modeling the user selection problem in distributed multiple-antenna transmitter systems as an assignment problem is also an interesting topic for future work. Due to the fact that the assignment problem operates over a matrix that describes the dynamics of the system, channel metrics can be used to capture such dynamics which can be used for efficient user selection.

Chapter 5 User Selection and Signal Design In Single Cell Systems [Ch1, J3, C5]

■ The topic of user selection has drawn the attention of academia over the last ten years and several practical (greedy and iterative) algorithms have been published. Even more important is the fact that current standards in the industry, namely LTE Releases 11 [175] and Wi-Fi/IEEE 802.11ac, already allow simultaneous transmission of different user streams in the same channel at the same time. This means that beamforming techniques will massively be under operation in the years to come. It also implies that there is huge room for the design and implementation of practical and efficient user selection algorithms, specially necessary in crowded MU-MIMO scenarios.
■ Most of the literature of user selection measures directly or indirectly the compatibility (spatial separability) between multi-user channels in order to group and schedule subset of quasi-orthogonal users. There are several metrics (functions of the multi-user channel) in the literature that have been used to accomplish that objective. One of the achieved goals in this thesis is the unified treatment of the most common metrics used to measure channel compatibility, i.e., it was shown which are the relationships that held between the metrics and why they should or should not be used for user selection.
■ The combinatorial nature of the user selection problem jeopardizes optimum spectral and power efficiency as well as multiplexing gains. However, it is possible to handle such a problem by iteratively choosing users whose channels meet certain criterion. The user selection can be performed jointly with power selection or by decoupling both processes. If the latter approach is followed, then channel metrics play a fundamental role in the way that the set of selected users is found.
■ *Lessons learned.* *i*) If the beamforming technique is based on ZF, the optimum metric of selection is given by the NSP since it immediately provides information about the effective channel gains (at a price of full CSI at the transmitter). Other studied metrics have acceptable performance only for low number of users, i.e.,

they cannot fully exploit MUDiv. *ii*) Metrics based on the NSP were derived so that the user selection is performed with less computational complexity at a price of optimality (unavoidable trade-off). *iii*) It is possible to cast the user selection problem as a binary programming problem with unique solution. However, the solution of such problem can be only used as a benchmark since solving it becomes computationally costly as the number of users increases. *iv*) User selection can be performed based on the coefficients of correlation between multi-user channels. Although a sub-optimal user selection is attained, it requires few operations to compute the coefficients and the processing would be the same if limited CSI is considered, i.e., the transmitter knows quantized CSI.

■ *Research Extensions. i*) The efficiency of the proposed metrics and user selection algorithms can be tested in scenarios with quantized CSI and for practical beamforming schemes. *ii*) The proposed user selection algorithms can be extended and tested over multiple carrier scenarios. *iii*) Considering MU-MIMO scenarios in standardized wireless technologies, implies that the selection of spatially compatible user needs to be integrated with physical resource block (minimum resource unit in time and frequency) allocation. This means that current opportunistic scheduling techniques [51] must be extended to provide support to multiple users sharing the same physical resource blocks. *iv*) Heterogeneous systems is an important direction to extend the proposed algorithms. For BC systems where the users have more than one antenna [33, 140, 153, 176–181], the generalization of ZFBF is given by the Block Diagonalization (BD) scheme [42]. Since the performance of BD depends on the spatial compatibility of the MU-MIMO channels, fast and efficient user selection is necessary to maximize sum rate figures. *v*) In MU-MIMO scenarios when the ith user is equipped with N_{r_i} antennas, its signal space spans several dimension and mutual inter-user interference can be measured either in the angular or the subspace domain [176, 179]. Metrics for user selection in this kind of systems generalize either a function of $\theta_{\mathcal{V}_i \mathbf{h}_i}$ (illustrated in Fig. 5.2) or a measure of spatial compatibility for multidimensional subspaces. An interesting direction for future research is to identify in which scenarios multidimensional metrics (e.g., the *principal angles between subspaces*, the *subspace collinearity*, the *chordal distance*, and the *geometrical angle between subspaces* [179, 181]) can be used to design reliable user selection algorithms.

Chapter 6 User Selection and Signal Design In Multi-Cell Systems [J4]

■ The cooperative transmission in multi-cell system fundamentally seeks to mitigate interference for users located within the boundaries of neighbor cells. The use of distributed linear precoding schemes allows to mitigate the inter-user interference which directly enhances achievable rates and can potentially improve fairness if local user selection (per-base-station) is considered. There are different approaches for cooperation in multi-cell systems and each one of them achieves different gains at different cost in terms of shared information, signal processing, and other practical considerations (see Chapter 2.4).

■ One of the current challenges, in both academia and industry, is to design cooperative transmission schemes (for cell-edge users) with resilience to multi-cell synchronization issues, and limited backhaul rates. The proposed user selection algorithms exploit distributed linear precoding, which lie in the category of inter-

ference coordination techniques since the level of shared information between BSs is very limited. By exploiting local CSI it is possible to perform precoding-oriented user selection, which means that the proposed heuristic metrics and search algorithms fully exploit the characteristics of the implemented precoding schemes.

■ *Lessons learned. i*) It is possible to design channel metrics as a function of the SNR regime, the available number of antennas per BS, and the local CSI, so that accurate measures of compatibility between cell-edge users can be computed. The derived metrics exploit the fundamental structure of each precoding technique, i.e., they are based on the effective channel gains. In the case of DVSINR precoding, even for the $N_t < B$ scenario, the user selection based on the proposed metric achieves more than 90% of the optimum selection performance. *ii*) Power allocation strategies can be simplified since the precoders are unique for each set of selected users, which relaxes processing tasks at the BSs. *iii*) Multi-cell coordination allows a variety of efficient user selection algorithms even if the shared information between BSs is limited. Heuristic search algorithms can be designed so that user selection can be performed in a semi-distributed fashion. *iv*) Optimal and close-to-optimal performance can be achieved by properly processing the channel metrics at the CU. The proposed processing method is highly effective for DZF in any SNR regime, and only sub-optimal for DVSINR in middle rages of SNR. *v*) Low-complex (sub-optimal) algorithms can be further designed in order to compute a limited number of metrics per BS for user selection purposes. Such algorithms can overcome the performance of simplistic but practical scheduling approaches with a low price in terms of complexity and message exchange.

■ *Research Extensions. i*) Although the metrics and heuristic algorithms achieve acceptable performance, they have been designed considering full CSI at each BS. The work can be extended taking into account quantized or statistical CSI and specific constraints over the backhaul rates. *ii*) The proposed algorithms shed light on how to perform efficient user selection assuming that all the competing users have always data to be transmitted. The analysis can be further extended by considering that the users have different data loads. Under this scenario meaningful and more realistic optimization can be performed, and it would be possible to design algorithms that optimize the performance perceived by both, the system operator and the users. *iii*) The studied coordinated transmission only considered a single cluster with B BSs. Another direction in which this work can be extended is by assessing and adapting the proposed algorithms to multi-cluster systems where ICI is not negligible and scheduling schemes can potentially mitigate interference [62]. *iv*) Fairness issues can be taken into account by implementing different local selection strategies (see Chapter 6.4.6) and performing user selection over deployed central and edge users. *v*) According to some works (e.g., [138, 153]) for single-cell scenarios is not efficient to use all N_t antennas to perform SDMA. Investigating whether such limitations exist in CBF under optimal or equal power allocation is also an interesting topic of future research.

Appendices

Appendix A

Interference Channels in Multi-Cell Systems

A.1 Proof of Proposition 6.1

In the following, the notation is slightly modified and the user and BS subindices are omitted whenever the context avoids uncertainty. Consider the channel of the served user $\mathbf{h}$, its precoding vector $\mathbf{w}$ defined in (6.5), and its aggregate interference matrix $\tilde{\mathbf{H}}$, define $\tilde{\mathbf{V}} = null(\tilde{\mathbf{H}})$ the matrix that contains the orthonormal vectors $\{\tilde{\mathbf{v}}_i\}_{i=1}^{\epsilon_D}$ and $\rho = P/\sigma_n^2$. The effective channel gain is given by:

$$
\begin{aligned}
|\mathbf{h}^H\mathbf{w}|^2 &= \|\mathbf{h}\|^2 \cos^2\theta_{\mathbf{hw}} \\
&= |<\mathbf{h},\mathbf{w}>|^2 \\
&\overset{(a)}{=} |Tr(\mathbf{h}\mathbf{w}^H)|^2 \\
&= \frac{|Tr(\mathbf{h}\mathbf{h}^H\tilde{\mathbf{V}}\tilde{\mathbf{V}}^H)|^2}{\|\mathbf{h}^H\tilde{\mathbf{V}}\|^2} \\
&\overset{(b)}{=} \frac{|Tr(\tilde{\mathbf{V}}\tilde{\mathbf{V}}^H\mathbf{h}\mathbf{h}^H)|^2}{|\mathbf{h}^H\tilde{\mathbf{V}}\tilde{\mathbf{V}}^H\mathbf{h}|} \\
&= \frac{|Tr(\tilde{\mathbf{V}}\tilde{\mathbf{V}}^H\mathbf{h}\mathbf{h}^H)|^2}{|Tr(\tilde{\mathbf{V}}\tilde{\mathbf{V}}^H\mathbf{h}\mathbf{h}^H)|} \\
&= |Tr(\tilde{\mathbf{V}}\tilde{\mathbf{V}}^H\mathbf{h}\mathbf{h}^H)| \\
&\overset{(c)}{=} |\mathbf{h}^H\tilde{\mathbf{V}}\tilde{\mathbf{V}}^H\mathbf{h}| \\
&= |\mathbf{h}^H(\tilde{\mathbf{V}}\tilde{\mathbf{V}}^H)^H(\tilde{\mathbf{V}}\tilde{\mathbf{V}}^H)\mathbf{h}| \\
&\overset{(d)}{=} \|\tilde{\mathbf{V}}\tilde{\mathbf{V}}^H\mathbf{h}\|^2 \\
&\overset{(e)}{=} \| \left(\sum_{i=1}^{\epsilon_D} \tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i^H \right) \mathbf{h}\|^2 \\
&= \sum_{i=1}^{\epsilon_D} \|\tilde{\mathbf{v}}_i^H\mathbf{h}\tilde{\mathbf{v}}_i\|^2
\end{aligned}
$$

$$= \sum_{i=1}^{\epsilon_D} < \tilde{\mathbf{v}}_i^H \mathbf{h} \tilde{\mathbf{v}}_i, \tilde{\mathbf{v}}_i^H \mathbf{h} \tilde{\mathbf{v}}_i >$$
$$\overset{(f)}{=} \|\mathbf{h}\|^2 \sum_{i=1}^{\epsilon_D} \cos^2 \theta_{\mathbf{h}\tilde{\mathbf{v}}_i}$$

where (a) is due to the fact that $\|\mathbf{w}\| = 1$ and $< \mathbf{h}, \mathbf{w} >= Tr(\mathbf{h}\mathbf{w}^H)$. (b) is given by substituting $\mathbf{w}^H$ into (a) properties of the trace and outer product [121]. (c) is done by expressing the denominator in (b) in the form of the numerator. (d) The basis of the null space projection $\tilde{\mathbf{V}}$ can used to compute the projection matrix $\mathbf{Q}_\mathbf{h} = \tilde{\mathbf{V}}\tilde{\mathbf{V}}^H$ [147, §2.6]. The projector matrix is an idempotent matrix, i.e., $\mathbf{Q}_\mathbf{h} = \mathbf{Q}_\mathbf{h}\mathbf{Q}_\mathbf{h}^H = \mathbf{Q}_\mathbf{h}^H\mathbf{Q}_\mathbf{h}$ [143]. (e) is a decomposition of the form $\tilde{\mathbf{V}}\tilde{\mathbf{V}}^H = \sum_{i=1}^{\epsilon_D} \tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i^H$. In (f), given the orthonormal vectors $\{\tilde{\mathbf{v}}_i\}_{i=1}^{\epsilon_D}$, the projection of $\mathbf{h}$ onto $Sp(\tilde{\mathbf{H}})^\perp$ can be computed by the sum of the individual projections onto each one of the orthonormal basis [143]. As $\|\mathbf{h}\|^2$ and $\cos^2\theta_{\mathbf{h}\tilde{\mathbf{v}}_i}$ are independent variables [182], the expected value of the effective channel gain is

$$\mathbb{E}\left[|\mathbf{h}^H\mathbf{w}|^2\right] = \mathbb{E}\left[\|\mathbf{h}\|^2\right]\mathbb{E}\left[\sum_{i=1}^{\epsilon_D}\cos^2\theta_{\mathbf{h}\tilde{\mathbf{v}}_i}\right].$$

Given $\mathbf{h}, \tilde{\mathbf{v}}_i \in \mathbb{C}^{N_t \times 1}$ define the random variable υ_i as

$$\upsilon_i = \cos^2\theta_{\mathbf{h}\tilde{\mathbf{v}}_i}. \tag{A.1}$$

According to [182] the cumulative probability function of υ_i is given by $F_{\upsilon_i}(\upsilon_i) = 1-(1-\upsilon_i)^{N_t-1}$ and the expected value of the random variable is $\mathbb{E}[\upsilon_i] = \int_0^1 \upsilon_i f_{\upsilon_i}(\upsilon_i)d\upsilon_i = \frac{1}{N_t}$.

A.2 Proof of Proposition 6.2

For the sake of notation, consider the channel of the intended user as $\mathbf{h}$ its aggregate interference matrix $\tilde{\mathbf{H}}$, $\mathbf{w}$ defined in (6.8), its associated matrix $\mathbf{D}$, and $\rho = P/\sigma_n^2$, so that:

$$\begin{aligned}\mathbb{E}\left[|\mathbf{h}^H\mathbf{w}|^2\right] &\overset{(a)}{=} \mathbb{E}\left[\frac{|\mathbf{h}^H\mathbf{D}\mathbf{h}|^2}{\|\mathbf{D}\mathbf{h}\|^2}\right] \\ &\overset{(b)}{=} \mathbb{E}\left[\frac{|Tr(\mathbf{D}\mathbf{h}\mathbf{h}^H)|^2}{|Tr(\mathbf{D}\mathbf{D}^H\mathbf{h}\mathbf{h}^H)|}\right]\end{aligned} \tag{A.2}$$

where (a) obeys the definition of (6.8) and (b) is given by properties of the trace and outer product [121]. Let $\tilde{\mathbf{H}} = \mathbf{U}_{\tilde{\mathbf{H}}}\mathbf{\Sigma}_{\tilde{\mathbf{H}}}\mathbf{V}_{\tilde{\mathbf{H}}}^H$ be the SVD of the aggregate interference matrix and the unitary matrix $\mathbf{U}_{\tilde{\mathbf{H}}}$ is formed by the vectors $\{\mathbf{u}_i\}_{i=1}^{N_t}$. The matrix $\mathbf{D}$ can be decomposed as:

$$\mathbf{D} = \sum_{i=1}^{N_t-\epsilon_D} \lambda_i(\mathbf{D})\mathbf{u}_i\mathbf{u}_i^H + \sum_{j=N_t-\epsilon_D+1}^{N_t} \rho\mathbf{u}_j\mathbf{u}_j^H = \mathbf{D}_P + \mathbf{D}_Q \tag{A.3}$$

The effective channel gain is

$$
\begin{aligned}
|\mathbf{h}^H\mathbf{w}|^2 &= \frac{|\mathbf{h}^H\mathbf{D}\mathbf{h}|^2}{\|\mathbf{D}\mathbf{h}\|^2} \\
&= \frac{|\mathbf{h}^H\mathbf{D}_P\mathbf{h} + \mathbf{h}^H\mathbf{D}_Q\mathbf{h}|^2}{\|\mathbf{D}_P\mathbf{h}\|^2 + \|\mathbf{D}_Q\mathbf{h}\|^2} \\
&= \|\mathbf{h}\|^2\beta_{\mathbf{h}}
\end{aligned} \tag{A.4}
$$

where

$$
\beta_{\mathbf{h}} = \frac{\left(\sum_{i=1}^{N_t-\epsilon_D} \lambda_i(\mathbf{D})\cos^2\theta_{\mathbf{hu}_i} + \rho\sum_{j=N_t-\epsilon_D+1}^{N_t}\cos^2\theta_{\mathbf{hu}_j}\right)^2}{\sum_{i=1}^{N_t-\epsilon_D} \lambda_i^2(\mathbf{D})\cos^2\theta_{\mathbf{hu}_i} + \rho^2\sum_{j=N_t-\epsilon_D+1}^{N_t}\cos^2\theta_{\mathbf{hu}_j}}. \tag{A.5}
$$

Consider the decomposition of the effective channel gain (A.4) and $\beta_{\mathbf{h}}$ in (A.5), thus $|Tr(\mathbf{D}\mathbf{h}\mathbf{h}^H)| = \|\mathbf{h}\|^2\sum_{i=1}^{N_t}\lambda_i(\mathbf{D})\cos^2\theta_{\mathbf{hu}_i}$. Notice that each eigenvalue in (A.5) is affected by a variable of the form (A.1) and recall that $\mathbb{E}\left[\cos^2\theta_{\mathbf{hu}_i}\right] = \frac{1}{N_t}\ \forall i$. The following relation holds: $\mathbb{E}\left[|Tr(\mathbf{D}\mathbf{h}\mathbf{h}^H)|\right] = \mathbb{E}\left[\|\mathbf{h}\|^2\frac{Tr(\mathbf{D})}{N_t}\right]$. For the denominator in the RHS of (A.2) a similar results is obtained for its average value: $\mathbb{E}\left[|Tr(\mathbf{D}\mathbf{D}^H\mathbf{h}\mathbf{h}^H)|\right] = \mathbb{E}\left[\|\mathbf{h}\|^2\frac{Tr(\mathbf{D}\mathbf{D}^H)}{N_t}\right]$. The expected value[1] in the RHS of (A.2) can be approximated by considering that the expected value of the squared cosines affects all eigenvalues $\lambda_i(\mathbf{D})\ \forall i$ and the eigenvalues are independent of the correlation coefficients and $\|\mathbf{h}\|^2$ so that:

$$
\begin{aligned}
\mathbb{E}\left[\frac{|Tr(\mathbf{D}\mathbf{h}\mathbf{h}^H)|^2}{|Tr(\mathbf{D}\mathbf{D}^H\mathbf{h}\mathbf{h}^H)|}\right] &\approx \mathbb{E}\left[\frac{\left(\frac{1}{N_t}Tr(\mathbf{D})\|\mathbf{h}\|^2\right)^2}{\frac{1}{N_t}Tr(\mathbf{D}\mathbf{D}^H)\|\mathbf{h}\|^2}\right] \\
&= \mathbb{E}\left[\|\mathbf{h}\|^2\frac{\left(\sum_{i=1}^{N_t}\lambda_i(\mathbf{D})\right)^2}{N_t\sum_{i=1}^{N_t}\lambda_i^2(\mathbf{D})}\right] \\
&= \mathbb{E}\left[\|\mathbf{h}\|^2 J(\mathbf{eig}(\mathbf{D}))\right]
\end{aligned}
$$

A numerical example of $\mathbb{E}\left[|\mathbf{h}^H\mathbf{w}|^2/\|\mathbf{h}\|^2\right]$ and its approximation $\mathbb{E}\left[J(\mathbf{eig}(\mathbf{D}))\right]$ are presented in Fig. A.1 for $B = 3$ and $N_t \in \{3, 4, 6\}$. Notice that the curves are normalized regarding to $\|\mathbf{h}\|^2$ in order to exclusively illustrate the relations between the eigenvalues of $\mathbf{D}$.

A.3 Proof of Proposition 6.3

Let $\tilde{\mathbf{H}} = \mathbf{U}_{\tilde{\mathbf{H}}}\mathbf{\Sigma}_{\tilde{\mathbf{H}}}\mathbf{V}_{\tilde{\mathbf{H}}}^H$ be the SVD of the aggregate interference matrix of the served user. And let $\hat{\mathbf{H}} = \tilde{\mathbf{H}}\tilde{\mathbf{H}}^H = \mathbf{U}_{\hat{\mathbf{H}}}\mathbf{\Sigma}_{\hat{\mathbf{H}}}\mathbf{V}_{\hat{\mathbf{H}}}^H$ be the SVD of the Hermitian matrix $\hat{\mathbf{H}}$. The diagonal matrix that contains the eigenvalues of $\hat{\mathbf{H}}$ can be defined from the eigenvalues of the aggregate interference matrix as

$$
\mathbf{\Sigma}_{\hat{\mathbf{H}}} = \mathbf{\Sigma}_{\tilde{\mathbf{H}}}\mathbf{\Sigma}_{\tilde{\mathbf{H}}}^H,
$$

[1]For a fixed B and $N_t \geq B$, if N_t is very large, the central limit theorem can be invoked and say that the values of the numerators and denominators will be very close to the average. For low values of N_t we cannot invoke this but simulations have shown that it still verifies.

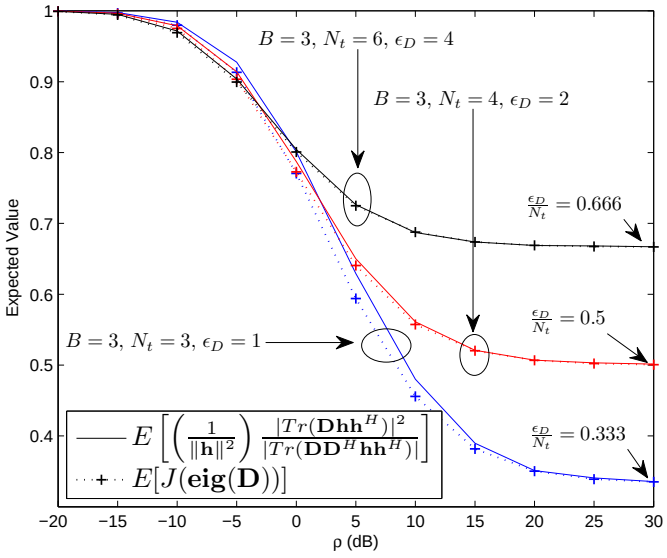

Figure A.1: Normalized values of the effective channel gain of DVSINR precoder and its approximation for $B = 3$ and $N_t \in \{3, 4, 6\}$.

---

and the eigenvalues of the matrix $\mathbf{D}$ are given by:

$$\lambda_i(\mathbf{D}) = \frac{1}{\rho^{-1} + [\mathbf{\Sigma}_{\hat{\mathbf{H}}}]_{ii}}. \tag{A.6}$$

Due to the fact that $\lambda_{\min}(\hat{\mathbf{H}})$ is equal to zero with multiplicity ϵ_D, $\lambda_{\max}(\mathbf{D}) = \rho$ with multiplicity ϵ_D. This means that $N_t - \epsilon_D$ eigenvalues of $\mathbf{D}$ are bounded as $\rho \to \infty$ and ϵ_D are not. The Jain's index of $\mathbf{eig}(\mathbf{D})$ is such that:

$$\lim_{\rho\to\infty} \frac{\left(\sum_{i=1}^{N_t} \lambda_i(\mathbf{D})\right)^2}{N_t \sum_{i=1}^{N_t} \lambda_i^2(\mathbf{D})} = \frac{\epsilon_D}{N_t},$$

which is illustrated in Fig. A.1.

A.4 Proof of Proposition 6.4

In order to simplify the notation let $\mathbf{h}_1$ be the channel of the user served in the local BS, with its associated matrices $\tilde{\mathbf{H}}_1$ and $\mathbf{D}_1 = \mathbf{D}_{1P} + \mathbf{D}_{1Q}$ as in (A.3). And let $\mathbf{h}_2 \in \tilde{\mathbf{H}}_1$ be a channel vector used to create the precoding vector $\mathbf{w}_1$. The interference term $|\mathbf{h}_2^H \mathbf{w}_1|^2$ in (6.2) for DVSINR can be unfolded as follows:

$$
\begin{aligned}
|\mathbf{h}_2^H\mathbf{w}_1|^2 &= \frac{|\mathbf{h}_2^H\mathbf{D}_1\mathbf{h}_1|^2}{|Tr(\mathbf{D}_1\mathbf{D}_1^H\mathbf{h}_1\mathbf{h}_1^H)|} \\
&\overset{(a)}{=} \frac{\left|\sum_{i=1}^{N_t-\epsilon_D}\lambda_i(\mathbf{D}_1)\mathbf{h}_2^H\mathbf{u}_i\mathbf{u}_i^H\mathbf{h}_1\right|^2}{\|\mathbf{h}_1\|^2\sum_{i=1}^{N_t}\lambda_i^2(\mathbf{D}_1)\cos^2\theta_{\mathbf{h}_1\mathbf{u}_i}} \\
&\overset{(b)}{=} \frac{\left|\sum_{i=1}^{N_t-\epsilon_D}\lambda_i(\mathbf{D}_1)\langle\mathbf{h}_2,\mathbf{u}_i\rangle\langle\mathbf{u}_i,\mathbf{h}_1\rangle\right|^2}{\|\mathbf{h}_1\|^2\sum_{i=1}^{N_t}\lambda_i^2(\mathbf{D}_1)\cos^2\theta_{\mathbf{h}_1\mathbf{u}_i}} \\
&\overset{(c)}{\leq} \|\mathbf{h}_2\|^2\frac{\left(\sum_{i=1}^{N_t-\epsilon_D}\lambda_i(\mathbf{D}_1)\cos\theta_{\mathbf{h}_2\mathbf{u}_i}\cos\theta_{\mathbf{h}_1\mathbf{u}_i}\right)^2}{\sum_{i=1}^{N_t}\lambda_i^2(\mathbf{D}_1)\cos^2\theta_{\mathbf{h}_1\mathbf{u}_i}}
\end{aligned}
$$

where the numerator in (a) only takes into account the basis and eigenvalues of $\mathbf{D}_{1P}$ in (A.3) since the basis of $\mathbf{D}_{1Q}$ contain the null space of $\mathbf{h}_2$. The result in (c) obeys the triangle inequality [121] since the terms $\langle\mathbf{h}_2,\mathbf{u}_i\rangle\langle\mathbf{u}_i,\mathbf{h}_1\rangle$ in (b) are complex numbers, and by taking their associated norms and coefficients of correlation their absolute values are already computed, cf. (6.18). In order to define an upper bound of $\mathbb{E}\left[|\mathbf{h}_2^H\mathbf{w}_1|^2\right]$ notice that the eigenvalues of $\mathbf{D}_1$ in the denominator of (c) are affected by an independent random variables of the form (A.1) with expected value $\frac{1}{N_t}$. And the average sum of the eigenvalues of $\mathbf{D}_{1P}$ in the numerator is affected by $\mathbb{E}\left[(\cos\theta_{\mathbf{h}_1\mathbf{u}_i}\cos\theta_{\mathbf{h}_2\mathbf{u}_i})^2\right] = \frac{1}{N_t^2}$. Observe that the random variables defined by the cosine functions are independent of each other and independent of $\|\mathbf{h}_2\|^2$ and $\mathbf{eig}(\mathbf{D}_1)$. Therefore, the expected value of the leakage is upper bounded as follows:

$$
\mathbb{E}\left[|\mathbf{h}_2^H\mathbf{w}_1|^2\right] \leq \mathbb{E}\left[\frac{\|\mathbf{h}_2\|^2}{N_t}\frac{|Tr(\mathbf{D}_{1P})|^2}{Tr(\mathbf{D}_1\mathbf{D}_1^H)}\right]. \tag{A.7}
$$

By inspecting (A.7) it can be said that in the particular case where $N_t < B$ the component $\mathbf{D}_{1Q}$ is not present in $\mathbf{D}_1$ according to (A.3), and the upper bound is given by:

$$
\mathbb{E}\left[|\mathbf{h}_2^H\mathbf{w}_1|^2\right] \leq \mathbb{E}\left[\|\mathbf{h}_2\|^2 J(\mathbf{eig}(\mathbf{D}_{1P}))\right].
$$

For the scenarios where $N_t \geq B$ the trace ratio in (A.7) can be approximated by dividing the largest squared eigenvalue of the numerator given by $\lambda_{\max}^2(\mathbf{D}_{1P}) = (\rho^{-1}+\lambda_{\min}(\tilde{\mathbf{H}}_1^H\tilde{\mathbf{H}}_1))^{-2}$ over the largest squared eigenvalue in the denominator $\lambda_{\max}^2(\mathbf{D}_1) = \rho^2$ which has multiplicity ϵ_D. By considering only these eigenvalues in the ratio, the contribution of the other eigenvalues is ignored and the approximated expected value of the leakage is given by:

$$
\begin{aligned}
\mathbb{E}\left[|\mathbf{h}_2^H\mathbf{w}_1|^2\right] &\approx \mathbb{E}\left[\frac{\|\mathbf{h}_2\|^2\lambda_{\max}^2(\mathbf{D}_{1P})}{N_t\epsilon_D\lambda_{\max}^2(\mathbf{D}_1)}\right] \\
&= \mathbb{E}\left[\frac{\|\mathbf{h}_2\|^2}{N_t\epsilon_D(\rho\lambda_{\min}(\tilde{\mathbf{H}}_1^H\tilde{\mathbf{H}}_1)+1)^2}\right].
\end{aligned} \tag{A.8}
$$

A numerical example of the upper bound[2] and the approximation of $\mathbb{E}\left[|\mathbf{h}_2^H\mathbf{w}_1|^2\right]$ is presented in Fig. A.2 for $B = 3$ and $N_t \in \{3,4\}$. Notice that in the high SNR the

[2]Extensive simulations were computed to verify the results for several values of B and N_t. Numerical results show that $\forall\rho \geq -20(\text{dB})$ as $\frac{\epsilon_D}{N_t} \to 1$, specifically when $N_t > B(B-2)+2$, inequality (A.7) does not hold and (A.8) under estimate the average leakage.

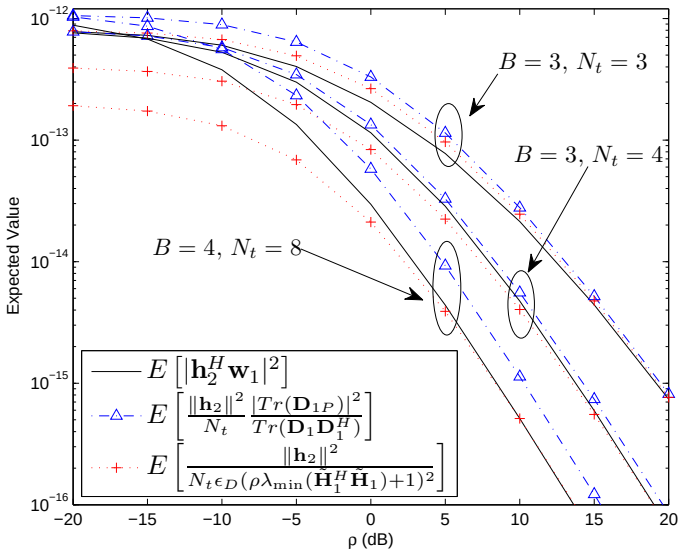

Figure A.2: Approximation, and exact value of the average leakage $\mathbb{E}\left[\left|\mathbf{h}_2^H \mathbf{w}_1\right|^2\right]$ for $B \in \{3, 4\}$ and $N_t \in \{3, 4, 8\}$.

gap between the leakage and its approximation (A.8) decreases. For instance, in the case $N_t = B = 3$, $\epsilon_D = 1$ and there is a dominant eigenvalue in each matrix $\mathbf{D}_{1P}$ and $\mathbf{D}_1$, i.e., $\epsilon_D \rho^2 > \lambda^2_{\max}(\mathbf{D}_{1P}) > \lambda^2_{\min}(\mathbf{D}_{1P})$. In scenarios where $N_t > B$ the approximation (A.8) under estimates the leakage in the low SNR. The reason is that in this regime all the eigenvalues of $\mathbf{D}_1$ have similar magnitudes (cf. Fig. A.1) but the denominator is already multiplied by ϵ_D, the number of dominant eigenvalues of $\mathbf{D}_1$ in the high SNR.

A.5 Proof of Proposition 6.5

From (A.3) it can be observed that the basis of both $Sp(\tilde{\mathbf{H}})$ and $Sp(\tilde{\mathbf{H}})^{\perp}$ are combined when forming the precoder. The value of the effective channel gain is a function of ρ and for the low SNR regime $\mathbf{D}_P$ is dominant while in the high SNR regime $\mathbf{D}_Q$ is the dominant term of $\mathbf{D}$. The exact interaction between of the vectors $\{\mathbf{u}_i\}_{i=1}^{N_t}$ and $\mathbf{h}$ is given by $\beta_{\mathbf{h}}$. Observe that the components of $\mathbf{D}_Q$ in (A.5) compute the magnitude of the projection of $\mathbf{h}$ onto each basis of $Sp(\tilde{\mathbf{H}})^{\perp}$, the exact NSP component scaled by ρ. The components of $\mathbf{D}_P$ do not represent the exact projection of $\mathbf{h}$ onto $Sp(\tilde{\mathbf{H}})$ because each one of the basis is affected by a different eigenvalue $\lambda_i(\mathbf{D})$. The term $|\mathbf{h}^H\mathbf{w}|^2$ combines a component of $\mathbf{h}$ onto $Sp(\tilde{\mathbf{H}})^{\perp}$ and weighted components of $\mathbf{h}$ onto the basis of $Sp(\tilde{\mathbf{H}})$.

The intuition behind the heuristic metric (6.13) is that the effective channel gain is bounded as follows:

$$\|\mathbf{Q}_{\mathbf{h}}\mathbf{h}\|^2 \leq |\mathbf{h}^H\mathbf{w}|^2 \leq \|\mathbf{h}\|^2.$$

This means that one can always take into account the magnitude of $\|\mathbf{Q}_{\mathbf{h}}\mathbf{h}\|^2$ and the component $\|\mathbf{P}_{\mathbf{h}}\mathbf{h}\|^2$ should be modified by a monotonic decreasing function of ρ with values in the range $[0, 1]$. By observing that $\beta_{\mathbf{h}}$ is the ratio of the squared combination of the eigenvalues of $\mathbf{D}$ over the combination of its squared eigenvalues, the function (6.14) is defined by the

quotient of $\lambda^2_{\min}(\mathbf{D}) = (\rho^{-1} + \lambda_{\max}(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}))^{-2}$ over $\lambda^2_{\max}(\mathbf{D}) = \rho^2$. The objective of such ratio is to measure how much the maximum and minimum eigenvalues of $\mathbf{D}$ spread out as a function of ρ. Observe that as $\rho \to 0$ the value of (6.14) goes to 1 and when $\rho \to \infty$ the function goes to zero.

A.6 Proof of Proposition 6.6

Let $\tilde{\mathbf{V}}_{bk}(\mathcal{G}_l) = null(\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l))$ be the matrix that contains the orthonormal basis of the null space of $\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)$ and $\tilde{\mathbf{v}}_i$ is its ith column vector with $i \in \{1, \ldots, \epsilon_D\}$. The NSP can be computed as $\|\mathbf{h}_{bk}\mathbf{Q}_{\mathbf{h}_{bk}}\|^2 = \|\mathbf{h}_{bk}\|^2 \sum_{i=1}^{\epsilon_D} \cos^2\theta_{\mathbf{h}_{bk}\tilde{\mathbf{v}}_i}$ (see Appendix A.1). Recall that $\|\mathbf{h}_{bk}\|^2$ and $\cos^2\theta_{\mathbf{h}_{bk}\tilde{\mathbf{v}}_i}$ are independent variables [182] and the factors of the product in the RHS of (6.21) are independent. Assuming that the components of $\tilde{\mathbf{H}}_{bk}(\mathcal{G}_l)$ are independent, it holds that:

$$\mathbb{E}\left[\prod_{i\neq k, i\in\mathcal{G}_l} \sin^2\theta_{\mathbf{h}_{bk}\mathbf{h}_{bi}}\right] = \left(1 - \frac{1}{N_t}\right)^{(B-1)}$$

and $\frac{\epsilon_D}{N_t} \leq \left(1 - \frac{1}{N_t}\right)^{(B-1)}$ with equality when $B = 2$.

References

[1] International Telecommunication Union (ITU). The World in 2014: ICT Facts and Figures, 2014. Press Release.

[2] Cisco. Visual Networking Index, 2014. white paper at Cisco.com.

[3] International Telecommunication Union (ITU). The State of Broadband 2014: Broadband for all, 2014. white paper at broadbandcommission.org.

[4] A Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T.A Thomas, J.G. Andrews, P. Xia, H.S. Jo, H.S. Dhillon, and T.D. Novlan. Heterogeneous Cellular Networks: From Theory to Practice. *IEEE Communications Magazine*, 50(6):54–64, June 2012.

[5] J.G. Andrews, S. Buzzi, Wan Choi, S.V. Hanly, A Lozano, AC.K. Soong, and J.C. Zhang. What Will 5G Be? *IEEE Journal on Selected Areas in Communications*, 32(6):1065–1082, June 2014.

[6] FP7 European Project 317669 METIS (Mobile and wireless communications Enablers for the Twenty-twenty Information Society), 2012. https://www.metis2020.com/.

[7] F Boccardi, R.W. Heath, A Lozano, T.L. Marzetta, and P. Popovski. Five Disruptive Technology Directions for 5G. *IEEE Communications Magazine*, 52(2):74–80, February 2014.

[8] F. Rusek, D. Persson, Buon Kiong Lau, E.G. Larsson, T.L. Marzetta, O. Edfors, and F. Tufvesson. Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays. *IEEE Signal Processing Magazine*, 30(1):40–60, Jan 2013.

[9] S.P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[10] Singiresu Rao. *Engineering Optimization: Theory and Practice*. John Wiley & Sons, Inc., 2009.

[11] Andrea Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.

[12] E. Bjornson and E. Jorswieck. *Optimal Resource Allocation in Coordinated Multi-Cell Systems*. Foundations and Trends(r) in Communications and Information. Now Publishers Incorporated, 2013.

[13] M. Sharif and B. Hassibi. A Comparison of Time-Sharing, DPC, and Beamforming for MIMO Broadcast Channels With Many Users. *IEEE Trans. on Commun.*, 55(1):11–15, Jan 2007.

[14] G.J. Foschini and M.J. Gans. On Limits of Wireless Communications in a Fading Environment when using Multiple Antennas. *Wireless Personal Communications*, 6(3):311–335, 1998.

[15] Emre Telatar. Capacity of Multi-Antenna Gaussian Channels. *European Transactions on Telecommunications*, 10(6):585–595, 1999.

[16] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley, 2006.

[17] W. Yang Y. Cho, J. Kim and C. G. Kang. *MIMO-OFDM wireless communications with MATLAB*. John Wiley and Sons, 2010.

[18] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[19] A. Osseiran, J.F. Monserrat, and W. Mohr. *Mobile and Wireless Communications for IMT-Advanced and Beyond*. Wiley, 2011.

[20] Lizhong Zheng and D.N.C. Tse. Diversity and Multiplexing: a fundamental tradeoff in Multiple-Antenna Channels. *IEEE Transactions on Information Theory*, 49(5):1073–1096, May 2003.

[21] P. Viswanath, D.N.C. Tse, and R. Laroia. Opportunistic beamforming using dumb antennas. *IEEE Transactions on Information Theory*, 48(6):1277–1294, Jun 2002.

[22] D. Gesbert, M. Kountouris, R.W. Heath, Chan-Byoung Chae, and T. Salzer. Shifting the MIMO Paradigm. *IEEE Signal Processing Magazine*, 24(5):36–46, Sept 2007.

[23] D.N.C. Tse, P. Viswanath, and Lizhong Zheng. Diversity-Multiplexing Tradeoff in Multiple-Access Channels. *IEEE Transactions on Information Theory*, 50(9):1859–1874, Sept 2004.

[24] M. H M Costa. Writing on dirty paper (corresp.). *IEEE Transactions on Information Theory*, 29(3):439–441, 1983.

[25] G. Caire and S. Shamai. On the achievable throughput of a multiantenna Gaussian broadcast channel. *IEEE Transactions on Information Theory*, 49(7):1691–1706, 2003.

[26] H. Weingarten, Y. Steinberg, and S. Shamai. The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel. *IEEE Transactions on Information Theory*, 52(9):3936–3964, Sept 2006.

[27] N. Jindal, S. Vishwanath, and A Goldsmith. On the Duality of Gaussian Multiple-Access and Broadcast Channels. *IEEE Transactions on Information Theory*, 50(5):768–783, May 2004.

[28] P. Marsch and G.P. Fettweis. *Coordinated Multi-Point in Mobile Communications: From Theory to Practice*. Cambridge University Press, 2011.

[29] Wei Yu and Tian Lan. Transmitter Optimization for the Multi-Antenna Downlink With Per-Antenna Power Constraints. *IEEE Trans. on Signal Processing*, 55(6):2646–2660, June 2007.

[30] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and Wei Yu. Multi-Cell MIMO Cooperative Networks: A New Look at Interference. *IEEE Journal on Selected Areas in Communications*, 28(9):1380–1408, December 2010.

[31] H. Boche and M. Schubert. A General Duality Theory for Uplink and Downlink Beamforming. In *IEEE Vehicular Technology Conference*, volume 1, pages 87–91 vol.1, 2002.

[32] M. Schubert and H. Boche. Solution of the Multiuser Downlink Beamforming Problem with Individual SINR Constraints. *IEEE Trans. on Vehicular Technology*, 53(1):18 – 28, jan 2004.

[33] B.C. Lim, W.A. Krzymien, and C. Schlegel. Efficient Sum Rate Maximization and Resource Allocation in Block-Diagonalized Space-Division Multiplexing. *IEEE Transactions on Vehicular Technology*, 58(1):478–484, Jan 2009.

[34] S.S. Christensen, R. Agarwal, E. Carvalho, and J.M. Cioffi. Weighted Sum-Rate Maximization using Weighted MMSE for MIMO-BC Beamforming Design. *IEEE Transactions on Wireless Communications*, 7(12):4792–4799, December 2008.

[35] Mats Bengtsson and Bjorn Ottersten. Handbook of Antennas in Wireless Communications. chapter Optimum and Suboptimum Transmit Beamforming. CRC Press, 2002.

[36] M. Schubert and H. Boche. *QoS-based Resource Allocation and Transceiver Optimization*. Foundation and Trends(r) in Communications and Information Theory, Hanover, USA, 2006.

[37] S. Stanczak, M. Wiczanowski, and H. Boche. *Fundamentals of Resource Allocation in Wireless Networks: Theory and Algorithms*. Springer, Berlin, Germany, 2009.

[38] H. Mahdavi-Doost, M. Ebrahimi, and A.K. Khandani. Characterization of SINR Region for Interfering Links With Constrained Power. *IEEE Transactions on Information Theory*, 56(6):2816 –2828, june 2010.

[39] N.I Miridakis and D.D. Vergados. A Survey on the Successive Interference Cancellation Performance for Single-Antenna and Multiple-Antenna OFDM Systems. *IEEE Communications Surveys & Tutorials*, 15(1):312–335, First 2013.

[40] Taesang Yoo and A. Goldsmith. On the Optimality of Multiantenna Broadcast Scheduling Using Zero-Forcing Beamforming. *IEEE Journal on Selected Areas in Communications*, 24(3):528 – 541, march 2006.

[41] C.B. Peel, B.M. Hochwald, and a.L. Swindlehurst. A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser Communication - Part I: Channel Inversion and Regularization. *IEEE Transactions on Communications*, 53(1):195–202, January 2005.

[42] Q.H. Spencer, A.L. Swindlehurst, and M. Haardt. Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels. *IEEE Transactions on Signal Processing*, 52(2):461–471, Feb 2004.

[43] Nihar Jindal and Andrea Goldsmith. Dirty-paper coding versus TDMA for MIMO broadcast channels. *IEEE Transactions on Information Theory*, 51(5):1783–1794, 2005.

[44] Jing Jiang, R Michael Buehrer, and W.H. Tranter. Greedy Scheduling Performance for a Zero-Forcing Dirty-Paper Coded System. *IEEE Transactions on Communications*, 54(5):789–793, May 2006.

[45] P.W.C. Chan and R.S. Cheng. Capacity Maximization for Zero-Forcing MIMO-OFDMA Downlink Systems with Multiuser Diversity. *IEEE Transactions on Wireless Communications*, 6(5):1880–1889, May 2007.

[46] Jianqi Wang, D.J. Love, and M.D. Zoltowski. User Selection With Zero-Forcing Beamforming Achieves the Asymptotically Optimal Sum Rate. *IEEE Transactions on Signal Processing*, 56(8):3713–3726, Aug 2008.

[47] Le-Nam Tran and Een-Kee Hong. Multiuser Diversity for Successive Zero-Forcing Dirty Paper Coding: Greedy Scheduling Algorithms and Asymptotic Performance Analysis. *IEEE Transactions on Signal Processing*, 58(6):3411–3416, June 2010.

[48] Wei Yu and Wonjong Rhee. Degrees of Freedom in Wireless Multiuser Spatial Multiplex Systems with Multiple Antennas. *IEEE Transactions on Communications*, 54(10):1747–1753, Oct 2006.

[49] F. Shad, T.D. Todd, Vytas Kezys, and J. Litva. Dynamic slot allocation (DSA) in indoor SDMA/TDMA using a smart antenna basestation. *IEEE/ACM Transactions on Networking*, 9(1):69–81, Feb 2001.

[50] A. Perez-Niera and M. Campalans. *Cross-Layer Resource Allocation in Wireless Communications: Techniques and Models from PHY and MAC Layer Interaction.* Academic Press, Oxford, UK, 2010.

[51] Arash Asadi and Vincenzo Mancuso. A Survey on Opportunistic Scheduling in Wireless Communications. *IEEE Communications Surveys & Tutorials*, 15(4):1671–1688, 2013.

[52] D.H.N. Nguyen and T Le-Ngoc. *Wireless Coordinated Multicell Systems: Architectures and Precoding Designs.* SpringerBriefs in computer science. Springer, 2014.

[53] Xiujun Zhang, Yin Sun, Xiang Chen, Shidong Zhou, Jing Wang, and Ness B. Shroff. Distributed Power Allocation for Coordinated Multipoint Transmissions in Distributed Antenna Systems. *IEEE Transactions on Wireless Communications*, 12(5):2281–2291, May 2013.

[54] Jie Zhang, Rong Zhang, Guangjun Li, and L. Hanzo. Distributed Antenna Systems in Fractional-Frequency-Reuse-Aided Cellular Networks. *IEEE Transactions on Vehicular Technology*, 62(3):1340–1349, March 2013.

[55] Jr. R. W. Heath, Steven Peters, Yi Wang, and Jiayin Zhang. A Current Perspective on Distributed Antenna Systems for the Downlink of Cellular Systems. *IEEE Communications Magazine*, (April):161–167, 2013.

[56] Jingon Joung, Yeow Khiang Chia, and Sumei Sun. Energy-Efficient, Large-Scale Distributed-Antenna System (L-DAS) for Multiple Users. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):954–965, Oct 2014.

[57] J.G. Andrews. Interference Cancellation for Cellular Systems: A Contemporary Overview. *IEEE Wireless Communications*, 12(2):19–29, April 2005.

[58] E. Dahlman, S. Parkvall, and J. Skold. *4G: LTE/LTE-Advanced for Mobile Broadband*. Elsevier Science, 2nd edition, 2014.

[59] Daewon Lee, Hanbyul Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana. Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges. *IEEE Communications Magazine*, 50(2):148–155, February 2012.

[60] Jun Zhang, Runhua Chen, J.G. Andrews, A. Ghosh, and R.W. Heath. Networked MIMO with clustered linear precoding. *IEEE Transactions on Wireless Communications*, 8(4):1910–1921, April 2009.

[61] Ya-Feng Liu, Yu-Hong Dai, and Zhi-Quan Luo. Coordinated Beamforming for MISO Interference Channel: Complexity Analysis and Efficient Algorithms. *IEEE Transactions on Signal Processing*, 59(3):1142–1157, March 2011.

[62] A Lozano, R.W. Heath, and J.G. Andrews. Fundamental Limits of Cooperation. *IEEE Transactions on Information Theory*, 59(9):5213–5226, Sept 2013.

[63] E.A. Jorswieck, E.G. Larsson, and D. Danev. Complete Characterization of the Pareto Boundary for the MISO Interference Channel. *IEEE Trans. on Signal Processing*, 56(10):5292–5296, Oct 2008.

[64] E.G. Larsson and E.A. Jorswieck. Competition Versus Cooperation on the MISO Interference Channel. *IEEE Journal on Selected Areas in Commun.*, 26(7):1059–1069, September 2008.

[65] R.H. Etkin, D.N.C. Tse, and Hua Wang. Gaussian Interference Channel Capacity to Within One Bit. *IEEE Transactions on Information Theory*, 54(12):5534–5562, Dec 2008.

[66] Xiaohu Shang, Biao Chen, and H.V. Poor. Multiuser MISO Interference Channels With Single-User Detection: Optimality of Beamforming and the Achievable Rate Region. *IEEE Transactions on Information Theory*, 57(7):4255–4273, July 2011.

[67] Mingyi Hong and Zhi-Quan Luo. Academic Press Library in Signal Processing: Communications and Radar Signal Processing. volume 2, chapter 8. Signal Processing and Optimal Resource Allocation for the Interference Channel. Elsevier Science, 2013.

[68] T. Lan and M. Chiang. An Axiomatic Theory of Fairness in Resource Allocation. Technical report, George Washington University, 2011. http://www.princeton.edu/~chiangm/fairness.pdf [Online; accessed September 30, 2014].

[69] Mung Chiang, Prashanth Hande, Tian Lan, and Chee Wei Tan. *Power Control in Wireless Cellular Networks*. Foundations and Trends(r) in Networking, 2008.

[70] C. W. Tan, M. Chiang, and R. Srikant. Fast Algorithms and Performance Bounds for Sum Rate Maximization in Wireless Networks. *IEEE/ACM Trans. on Networking*, 21(3):706–719, 2012.

[71] R. Mochaourab and E. Jorswieck. Optimal Beamforming in Interference Networks with Perfect Local Channel Information. *IEEE Trans. on Signal Processing*, 59(3):1128–1141, March 2011.

[72] E. Bjornson, R. Zakhour, D. Gesbert, and B. Ottersten. Cooperative Multicell Precoding: Rate Region Characterization and Distributed Strategies With Instantaneous and Statistical CSI. *IEEE Transactions on Signal Processing*, 58(8):4298–4310, 2010.

[73] Shidong Zhou, Ming Zhao, Xibin Xu, Jing Wang, and Yan Yao. Distributed Wireless Communication System: A New Architecture for Future Public Wireless Access. *IEEE Communications Magazine*, 41(3):108–113, Mar 2003.

[74] D. Wake, M. Webster, G. Wimpenny, K. Beacham, and L. Crawford. Radio Over Fiber for Mobile Communications. In *IEEE International Topical Meeting on Microwave Photonics*, pages 157–160, Oct 2004.

[75] I. Toufik and R. Knopp. Wideband Channel Allocation in Distributed Antenna Systems. In *IEEE 64th Vehicular Technology Conference*, pages 1–5, Sept 2006.

[76] Honglin Hu, Yan Zhang, and Jijun Luo. *Distributed Antenna Systems: Open Architecture for Future Wireless Communications*. Auerbach Publications, 2007.

[77] A. A M Saleh, A.J. Rustako, and R. Roman. Distributed Antennas for Indoor Radio Communications. *IEEE Transactions on Communications*, 35(12):1245–1251, December 1987.

[78] Wan Choi, J.G. Andrews, and Chaehag Yi. Capacity of Multicellular Distributed Antenna Networks. In *International Conference on Wireless Networks, Communications and Mobile Computing*, volume 2, pages 1337–1342 vol.2, June 2005.

[79] A. Obaid and H. Yanikomeroglu. Reverse-link power control in CDMA Distributed Antenna Systems. In *IEEE Wireless Communications and Networking Confernce*, volume 2, pages 608–612 vol.2, 2000.

[80] E. Kudoh and F. Adachi. Power and Frequency Efficient Virtual Cellular Network. In *IEEE Vehicular Technology Conference*, volume 4, pages 2485–2489 vol.4, April 2003.

[81] M.V. Clark, III Willis, T., L.J. Greenstein, Jr. Rustako, A., V. Erceg, and R. Roman. Distributed versus Centralized Antenna Arrays in Broadband Wireless Networks. In *IEEE Vehicular Technology Conference*, volume 1, pages 33–37 vol.1, 2001.

[82] Lin Dai, Shi dong Zhou, and Yan Yao. Capacity with MRC-based Macrodiversity in CDMA Distributed Antenna Systems. In *IEEE Global Telecommunications Conference*, volume 1, pages 987–991 vol.1, Nov 2002.

[83] Xunyong Zhang, Chen He, Lingge Jiang, and Jing Xu. Inter-cell interference coordination based on softer frequency reuse in OFDMA cellular systems. In *International Conference on Neural Networks and Signal Processing*, pages 270 –275, june 2008.

[84] E Hossain, DI Kim, and Vijay K. Bhargava. *Cooperative Cellular Wireless Networks*. Cambridge University Press, 2011.

[85] T. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall, 1996.

[86] A. Simonsson. Frequency Reuse and Intercell Interference Co-Ordination In E-UTRA. In *IEEE Vehicular Technology Conference*, pages 3091–3095, April 2007.

[87] I Katzela and M. Naghshineh. Channel Assignment Schemes for Cellular Mobile Telecommunication Systems: A Comprehensive Survey. *IEEE Commun. Surveys & Tutorials*, 3(2):10–31, 2000.

[88] Winner i, d4.3. identification, definition and assessment of cooperation schemes between rans, June 2005.

[89] F Khan. *LTE for 4G Mobile Broadband: Air Interface Technologies and Performance*. Cambridge University Press, 2009.

[90] A. Lodhi, A. Awad, T. Jeffries, and P. Nahi. On re-use partitioning in LTE-FDD systems. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 910 –913, sept. 2009.

[91] M. Porjazoski and B. Popovski. Analysis of Intercell interference coordination by Fractional frequency reuse in LTE. In *International Conference on Software, Telecommunications and Computer Networks*, pages 160 –164, sept. 2010.

[92] Rizwan Ghaffar and Raymond Knopp. Fractional frequency reuse and interference suppression for OFDMA networks. In *International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, pages 273 –277, 31 2010-june 4 2010.

[93] Zhaoxin Lu, Hui Tian, Qiaoyun Sun, Bo Huang, and Shuqin Zheng. An Admission Control Strategy for Soft Frequency Reuse Deployment of LTE Systems. In *IEEE Consumer Communications and Networking Conference*, pages 1 –5, jan. 2010.

[94] V.K.N. Lau. Proportional Fair Space - Time Scheduling for Wireless Communications. *IEEE Transactions on Communications*, 53(8):1353–1360, Aug 2005.

[95] M. Al-Shalash, F. Khafizov, and Zhijun Chao. Interference constrained soft frequency reuse for uplink ICIC in LTE networks. In *IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, pages 1882 –1887, sept. 2010.

[96] Y. Argyropoulos, S. Jordan, and S.P.R. Kumar. Dynamic Channel Allocation in Interference-Limited Cellular Systems with Uneven Traffic Distribution. *IEEE Transactions on Vehicular Technology*, 48(1):224 –232, jan 1999.

[97] Ruihong An, Xin Zhang, Gen Cao, Ruiming Zheng, and Lin Sang. Interference Avoidance and Adaptive Fraction Frequency Reuse in a Hierarchical Cell Structure. In *IEEE Wireless Communications and Networking Conference*, pages 1 –5, april 2010.

[98] Martin Döttling, Werner Mohr, and Afif Osseiran. *Radio Technologies and Concepts for IMT-Advanced*. John Wiley and Sons, 2009.

[99] Wan Choi and J.G. Andrews. Downlink performance and capacity of distributed antenna systems in a multicell environment. *IEEE Transactions on Wireless Communications*, 6(1):69–73, Jan 2007.

[100] A. Czylwik and T. Matsumoto. Downlink Beamforming for Frequency-Duplex Systems in Frequency-Selective Fading. In *IEEE Vehicular Technology Conference Proceedings*, volume 2, pages 695–699 vol.2, 2000.

[101] J.G. Proakis. *Digital Communications*. McGraw-Hill Higher Education, 1995.

[102] V. Naware, G. Mergen, and Lang Tong. Stability and delay of finite-user slotted ALOHA with multipacket reception. *IEEE Transactions on Information Theory*, 51(7):2636–2656, July 2005.

[103] F. Baccelli, B. Blaszczyszyn, and P. Muhlethaler. An ALOHA Protocol for Multihop Mobile Wireless Networks. *IEEE Transactions on Information Theory*, 52(2):421–436, Feb 2006.

[104] WiMAX Forum. Wimax system level evaluation methodology, v.0.0.1, 2006.

[105] H. Boche and M. Schubert. Perron-root Minimization for Interference-Coupled Systems with Adaptive Receive Strategies. *IEEE Trans. on Communications*, 57(10):3173 –3164, october 2009.

[106] Anders Gjendemsjo, GeirE Oien, Henrik Holm, Mohamed-Slim Alouini, David Gesbert, Kjell Hole, and Pal Orten. Rate and Power Allocation for Discrete-Rate Link Adaptation. *EURASIP Journal on Wireless Communications and Networking*, 2008(1), 2008.

[107] Mung Chiang and J. Bell. Balancing Supply and Demand of Bandwidth in Wireless Cellular Networks: Utility Maximization over Powers and Rates. In *IEEE INFOCOM*, volume 4, pages 2800 – 2811, march 2004.

[108] P. Tsiaflakis, M. Diehl, and M. Moonen. Distributed Spectrum Management Algorithms for Multiuser DSL Networks. *IEEE Transactions on Signal Processing*, 56(10):4825–4843, 2008.

[109] P. Tsiaflakis and F. Glineur. A novel class of iterative approximation methods for DSL spectrum optimization. In *IEEE International Conference on Communications*, pages 3154–3159, 2012.

[110] M. Andersin, Z. Rosberg, and J. Zander. Gradual Removals in Cellular PCS with Constrained Power Control and Noise. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, volume 1, pages 56 –60, sep 1995.

[111] J. Zander. Performance of Optimum Transmitter Power Control in Cellular Radio Systems. *IEEE Transactions on Vehicular Technology*, 41(1):57 –62, feb 1992.

[112] R.D. Yates. A Framework for Uplink Power Control in Cellular Radio Systems. *IEEE Journal on Selected Areas in Communications*, 13(7):1341 –1347, sep 1995.

[113] Mung Chiang. *Geometric Programming for Communication Systems*. Foundation and Trends(r) in Communications and Information Theory, Hanover, USA, 2005.

[114] Tao Wang and L. Vandendorpe. Successive Convex Approximation based Methods for Dynamic Spectrum Management. In *IEEE Int. Conf. on Commun.*, pages 4061–4065, 2012.

[115] G.J. Foschini and Z. Miljanic. A Simple Distributed Autonomous Power Control Algorithm and its Convergence. *IEEE Transactions on Vehicular Technology*, 42(4):641 –646, nov 1993.

[116] R.D. Yates, S. Gupta, C. Rose, and S. Sohn. Soft dropping power control. In *IEEE Vehicular Technology Conference*, volume 3, pages 1694 –1698 vol.3, may 1997.

[117] Xiaoxin Qiu and K. Chawla. On the Performance of Adaptive Modulation in Cellular Systems. *IEEE Transactions on Communications*, 47(6):884 –895, jun 1999.

[118] C. W. Sung and K.-K. Leung. A Generalized Framework for Distributed Power Control in Wireless Networks. *IEEE Transactions on Information Theory*, 51(7):2625 –2635, july 2005.

[119] Tao Wang and L. Vandendorpe. On the SCALE Algorithm for Multiuser Multicarrier Power Spectrum Management. *IEEE Transactions on Signal Processing*, 60(9):4992–4998, 2012.

[120] R. B. Bapat and T. E. S. Raghavan. *Nonnegative Matrices and Applications*. Cambridge University Press, Cambridge, UK, 1997.

[121] J. Gentle. *Matrix algebra: Theory, Computations, and Applications in Statistics*. Springer, 2007.

[122] E. Driouch and W. Ajib. A Graph Theory Based Scheduling Algorithm for MIMO-CDMA Systems using Zero Forcing Beamforming. In *IEEE Symposium on Computers and Communications*, pages 674–679, July 2008.

[123] Tianxiong Ji, Chunhui Zhou, Shidong Zhou, and Yan Yao. Low Complex User Selection Strategies for Multi-User MIMO Downlink Scenario. In *IEEE Wireless Communications and Networking Conference*, pages 1532 –1537, march 2007.

[124] S. Karachontzitis and D. Toumpakaris. Efficient and Low-Complexity User Selection for the Multiuser MISO Downlink. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 3094 –3098, sept. 2009.

[125] Xinying Gao, Anxin Li, Yuan Yan, and H. Kayama. Dominant Users Grouping Algorithm for Multiple RAUs-UEs Coordination in DAS System. In *IEEE Vehicular Technology Conference*, pages 1 –5, sept. 2010.

[126] R. Burkard, M. Dell'Amico, and S. Martello. *Assignment Problems.* Siam, 2009.

[127] Z. Han and K. J. R. Liu. *Resource Allocation for Wireless Networks: Basics, Techniques, and Applications.* Cambridge University Press, 2008.

[128] R. Jonker and A. Volgenant. A Shortest Augmenting Path Algorithm for Dense and Sparse Linear Assignment Problems. *Computing*, 38:325–340, 1987.

[129] A. W. Marshall and I. Olkin. *Inequalites: Theory of Majorization and Its Applications.* Academic Press., 1979.

[130] ITU-R. *M.2135-1. Guidelines for evaluation of radio interface technologies for IMT-Advanced*, 2009.

[131] D.W.H. Cai, T.Q.S. Quek, and Chee-Wei Tan. A Unified Analysis of Max-Min Weighted SINR for MIMO Downlink System. *IEEE Trans.s on Signal Processing*, 59(8):3850–3862, Aug 2011.

[132] R. Jain, W. Hawe, and D. Chiu. A Quantitative measure of fairness and discrimination for resource allocation in Shared Computer Systems. Technical report, DEC-TR-301, 1984.

[133] S. Vishwanath, N. Jindal, and A. Goldsmith. Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels. *IEEE Transactions on Information Theory*, 49(10):2658–2668, 2003.

[134] A. Bayesteh and A.K. Khandani. On the User Selection for MIMO Broadcast Channels. *IEEE Transactions on Information Theory*, 54(3):1086–1107, 2008.

[135] A. Wiesel, Y.C. Eldar, and S. Shamai. Zero-Forcing Precoding and Generalized Inverses. *IEEE Transactions on Signal Processing*, 56(9):4409–4418, 2008.

[136] Li-Chun Wang and Chu-Jung Yeh. Scheduling for Multiuser MIMO Broadcast Systems: Transmit or Receive Beamforming? *IEEE Trans. on Wireless Communications*, 9(9):2779–2791, 2010.

[137] Zhenyu Tu and R.S. Blum. Multiuser diversity for a dirty paper approach. *IEEE Communications Letters*, 7(8):370–372, 2003.

[138] G. Dimic and N.D. Sidiropoulos. On Downlink Beamforming with Greedy User Selection: Performance Analysis and a Simple New Algorithm. *IEEE Transactions Signal Processing*, 53(10):3857 – 3868, oct. 2005.

[139] V.K.N. Lau. Optimal Downlink Space-Time Scheduling Design With Convex Utility Functions - Multiple-Antenna Systems With Orthogonal Spatial Multiplexing. *IEEE Transactions on Vehicular Technology*, 54(4):1322–1333, July 2005.

[140] Zukang Shen, Runhua Chen, J.G. Andrews, R.W. Heath, and B.L. Evans. Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization. *IEEE Transactions on Signal Processing*, 54(9):3658 –3663, sept. 2006.

[141] R.C. Elliott and W.A. Krzymien. Downlink Scheduling via Genetic Algorithms for Multiuser Single-Carrier and Multicarrier MIMO Systems With Dirty Paper Coding. *IEEE Transactions Vehicular Technology*, 58(7):3247 –3262, sept. 2009.

[142] Taesang Yoo and A. Goldsmith. Sum-rate Optimal Multi-Antenna Downlink Beamforming Strategy based on Clique Search. In *IEEE Global Telecommun. Conf.*, volume 3, 2005.

[143] H. Yanai, K. Takeuchi, and Y. Takane. *Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition*. Springer, 2011.

[144] Shengchun Huang, Hao Yin, Haoming Li, and V.C.M. Leung. Decremental User Selection for Large-Scale Multi-User MIMO Downlink with Zero-Forcing Beamforming. *IEEE Wireless Communications Letters*, 1(5):480–483, October 2012.

[145] H.-C. Yang and M.-S. Alouini. *Order Statistics in Wireless Communications: Diversity, Adaptation, and Scheduling in MIMO and OFDM Systems*. Cambridge University Press, 2011.

[146] Yuwei Shi, Qiyue Yu, Weixiao Meng, and Zhongzhao Zhang. Maximum Product of Effective Channel Gains: An Innovative User Selection Algorithm for Downlink Multi-User Multiple Input and Multiple Output. *Wireless Communications and Mobile Computing*, 2012.

[147] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Matrix Computations. Johns Hopkins University Press, 1996.

[148] Runhua Chen, R.W. Heath, and J.G. Andrews. Transmit Selection Diversity for Unitary Precoded Multiuser Spatial Multiplexing Systems With Linear Receivers. *IEEE Transactions on Signal Processing*, 55(3):1159–1171, March 2007.

[149] A. Razi, D.J. Ryan, I.B. Collings, and Jinhong Yuan. Sum rates, rate allocation, and user scheduling for multi-user MIMO vector perturbation precoding. *IEEE Transactions on Wireless Communications*, 9(1):356–365, January 2010.

[150] Israel Halperin. The product of projection operators. *Acta Sci. Math.(Szeged)*, 23:96–99, 1962.

[151] Junling Mao, Jinchun Gao, Yuanan Liu, and Gang Xie. Simplified Semi-Orthogonal User Selection for MU-MIMO Systems with ZFBF. *IEEE Wireless Communications Letters*, 1(1):42–45, February 2012.

[152] Eduardo Castañeda, Adao Silva, Ramiro Samano-Robles, and Atilio Gameiro. Low-complexity User Selection for Rate Maximization in MIMO Broadcast Channels with Downlink Beamforming. *The Scientific World Journal*, 2014:1–13, 2013.

[153] M. Fuchs, G. Del Galdo, and M. Haardt. Low-Complexity Space - Time - Frequency Scheduling for MIMO Systems With SDMA. *IEEE Transactions on Vehicular Technology*, 56(5):2775–2784, Sept 2007.

[154] P. Tejera, W. Utschick, G. Bauch, and J.A. Nossek. Subchannel Allocation in Multiuser Multiple-Input Multiple-Output Systems. *IEEE Trans. on Information Theory*, 52(10):4721–4733, Oct 2006.

[155] T.F. Maciel and A. Klein. On the Performance, Complexity, and Fairness of Suboptimal Resource Allocation for Multiuser MIMO-OFDMA Systems. *IEEE Transactions Vehicular Technology*, 59(1):406–419, Jan 2010.

[156] Y.J.A. Zhang and K.B. Letaief. An efficient resource-allocation scheme for spatial multiuser access in MIMO/OFDM systems. *IEEE Trans. on Communications*, 53(1):107–116, Jan 2005.

[157] Wen-Ching Chung, Li-Chun Wang, and Chung-Ju Chang. A low-complexity beamforming-based scheduling to downlink OFDMA/SDMA systems with multimedia traffic. *Wireless Networks*, 17(3):611–620, 2011.

[158] E. Driouch and W. Ajib. Efficient Scheduling Algorithms for Multiantenna CDMA Systems. *IEEE Transactions on Vehicular Technology*, 61(2):521–532, Feb 2012.

[159] P. Marsch. *Coordinated Multi-Point under a Constrained Backhaul and Imperfect Channel Knowledge*. Beiträge aus der Informationstechnik. Jörg Vogt Verlag, 2010.

[160] C. Oestges and B. Clerckx. *MIMO Wireless Communications: From Real-World Propagation to Space-Time Code Design*. Elsevier Science, 2007.

[161] Caijun Zhong, M.R. McKay, T. Ratnarajah, and Kai-Kit Wong. Distribution of the demmel condition number of wishart matrices. *IEEE Transactions on Communications*, 59(5):1309–1320, May 2011.

[162] Peng Lu and Hong-Chuan Yang. Sum-rate Analysis of Multiuser MIMO System with Zero-Forcing Transmit Beamforming. *IEEE Transactions on Communications*, 57(9):2585–2589, 2009.

[163] S. Sigdel and W.A. Krzymien. Efficient User Selection and Ordering Algorithms for Successive Zero-Forcing Precoding for Multiuser MIMO Downlink. In *IEEE Vehicular Technology Conference*, pages 1–6, 2009.

[164] TH Cormen, CE Leiserson, RL Rivest, and C Stein. *Introduction to Algorithms*. MIT Press, July 2001.

[165] Inc. CVX Research. CVX: Matlab Software for Disciplined Convex Programming, version 2.0. http://cvxr.com/cvx, August 2012.

[166] J. Löfberg. YALMIP : a toolbox for modeling and optimization in MATLAB. In *Computer Aided Control Systems Design, 2004 IEEE International Symposium on*, pages 284–289, 2004.

[167] Seok-Hwan Park, Haewook Park, Hanbae Kong, and Inkyu Lee. New Beamforming Techniques Based on Virtual SINR Maximization for Coordinated Multi-Cell Transmission. *IEEE Transactions on Wireless Communications*, 11(3):1034–1044, March 2012.

[168] Chenyang Yang, Shengqian Han, Xueying Hou, and AF. Molisch. How do we design CoMP to achieve its promised potential? *IEEE Wireless Communications*, 20(1):67–74, February 2013.

[169] Hoon Huh, A.M. Tulino, and G. Caire. Network MIMO With Linear Zero-Forcing Beamforming: Large System Analysis, Impact of Channel Estimation, and Reduced-Complexity Scheduling. *IEEE Transactions on Information Theory*, 58(5):2911–2934, 2012.

[170] Binglai Niu, V.W.S. Wong, and R. Schober. Downlink Scheduling with Transmission Strategy Selection for Multi-Cell MIMO Systems. *IEEE Transactions on Wireless Communications*, 12(2):736–747, February 2013.

[171] Adao Silva, Reza Holakouei, and Atilio Gameiro. Power allocation strategies for distributed precoded multicell based systems. *EURASIP J. Wireless Commun. and Networking*, 2011(1):1, 2011.

[172] Wei Yu, Taesoo Kwon, and Changyong Shin. Multicell Coordination via Joint Scheduling, Beamforming, and Power Spectrum Adaptation. *IEEE Transactions on Wireless Communications*, 12(7):1–14, July 2013.

[173] M. Sadek, A. Tarighat, and A.H. Sayed. A Leakage-Based Precoding Scheme for Downlink Multi-User MIMO Channels. *IEEE Trans. on Wireless Communications*, 6(5):1711–1721, 2007.

[174] B. Khoshnevis, Wei Yu, and Y. Lostanlen. Two-Stage Channel Quantization for Scheduling and Beamforming in Network MIMO Systems: Feedback Design and Scaling Laws. *IEEE Journal on Selected Areas in Communications*, 31(10):2028–2042, October 2013.

[175] Chaiman Lim, Taesang Yoo, Bruno Clerckx, Byong Lee, and Byonghyo Shim. Recent Trend of Multiuser MIMO in LTE-advanced. *IEEE Communications Magazine*, (March):127–135, 2013.

[176] Cheng Wang and R.D. Murch. Adaptive downlink multi-user MIMO wireless systems for correlated channels with imperfect CSI. *IEEE Transactions on Wireless Communications*, 5(9):2435–2446, September 2006.

[177] Runhua Chen, Zukang Shen, J.G. Andrews, and R.W. Heath. Multimode Transmission for Multiuser MIMO Systems With Block Diagonalization. *IEEE Transactions on Signal Processing*, 56(7):3294–3302, July 2008.

[178] S. Sigdel and W.A. Krzymien. Simplified Fair Scheduling and Antenna Selection Algorithms for Multiuser MIMO Orthogonal Space-Division Multiplexing Downlink. *IEEE Transactions on Vehicular Technology*, 58(3):1329–1344, March 2009.

[179] Xinping Yi and E.K.S. Au. User Scheduling for Heterogeneous Multiuser MIMO Systems: A Subspace Viewpoint. *IEEE Transactions on Vehicular Technology*, 60(8):4004–4013, Oct 2011.

[180] Kyeongjun Ko and Jungwoo Lee. Multiuser MIMO User Selection Based on Chordal Distance. *IEEE Transactions on Communications*, 60(3):649–654, March 2012.

[181] A. Adhikary and G. Caire. Joint Spatial Division and Multiplexing: Opportunistic Beamforming and User Grouping. *IEEE Transactions on Information Theory*, May 2013. Submitted for publication.

[182] Chun Kin Au-Yeung and D.J. Love. On the performance of random vector quantization limited feedback beamforming in a MISO system. *IEEE Transactions on Wireless Communications*, 6(2):458–462, 2007.