UNIVERSIDADE FEDERAL DO CEARÁ

CENTRO DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA

FRANCISCO HUGO COSTA NETO

RADIO RESOURCE MANAGEMENT TECHNIQUES FOR 5G NETWORKS BASED ON MACHINE LEARNING

FORTALEZA

2020

FRANCISCO HUGO COSTA NETO

RADIO RESOURCE MANAGEMENT TECHNIQUES FOR 5G NETWORKS BASED ON MACHINE LEARNING

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Orientador: Prof. Dr. Tarcisio Ferreira Maciel

Coorientador: Prof. Dr. Daniel Costa Araújo

FORTALEZA

2020

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

C872r Costa Neto, Francisco Hugo.
Radio resource management techniques for 5G networks based on machine learning / Francisco Hugo Costa Neto. – 2020.
108 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2020.
Orientação: Prof. Dr. Tarcisio Ferreira Maciel.
Coorientação: Prof. Dr. Daniel Costa Araújo.

1. Gerenciamento de recursos de rádio. 2. Formatação híbrida de feixes. 3. Controle de potência. 4. Aprendizado de máquina. I. Título.

CDD 621.38

FRANCISCO HUGO COSTA NETO

RADIO RESOURCE MANAGEMENT TECHNIQUES FOR 5G NETWORKS BASED ON MACHINE LEARNING

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Aprovada em: 06 de novembro de 2020

BANCA EXAMINADORA

Prof. Dr. Tarcisio Ferreira Maciel (Orientador)
Universidade Federal do Ceará

Prof. Dr. Daniel Costa Araújo (Coorientador)
Universidade de Brasília

Prof. Dr. André Lima Ferrer de Almeida
Universidade Federal do Ceará

Prof. Dr. Cecilio José Lins Pimentel
Universidade Federal de Pernambuco

Prof. Dr. Richard Demo Souza
Universidade Federal de Santa Catarina

À minha querida mãe Ana Cláudia, cujo empenho em me educar sempre veio em primeiro lugar. Aqui estão os resultados dos seus esforços. Com muita gratidão.

AGRADECIMENTOS

I would like to offer my sincere acknowledgment to all the people who over the past few years have contributed to this work. None of this would be possible without the countless exchanges of ideas, suggestions, and conversations that I had with my family, advisors, and friends.

Dear family (Ana, Valmir, and Lui), thank you so much for your unconditional support. Only you understand the challenges that have been overcome so that I could have made it this far. I also thank my beloved wife, Thalita, who has always understood and encouraged me. All your affection, faith, and love gave me the strength to move on. Thank you for being by my side in the most difficult moments - this achievement is also yours.

I am very grateful to my supervisor, Prof. Dr. Tarcisio F. Maciel, for the many opportunities, valuable guidance, and insightful suggestions. Special thanks to my co-supervisor Prof. Dr. Daniel Costa Araujo. Your assistance was extremely relevant to the development of this work. I would also like to thank the many friends who helped me a lot on this journey called doctorate.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. I acknowledge the technical and financial support from CAPES under grant 88882.183549/2018-01 and Ericsson Innovation Center, Brazil, under UFC.40, UFC.44, and UFC.47 Technical Cooperation Contracts Ericsson/UFC.

RESUMO

A quinta geração (5G) de comunicações móveis foi projetada para expandir os recursos das redes sem fio e, consequentemente, fornecer suporte otimizado a vários casos de uso e requisitos de projeto. Em vista disso, conjuntos massivos de antenas de múltiplas-entradas e múltiplas-saídas e a operação na faixa de frequência de ondas milimétricas (mmWave) são importantes soluções técnicas capazes de suportar um aprimoramento expressivo da capacidade de tráfego de dados, uma demanda reconhecidamente relevante de 5G. Nesse contexto, a presente tese investiga técnicas de gerenciamento de recursos de rádio para explorar essas tecnologias e superar seus principais desafios, como condições de propagação hostis, aquisição de informações sobre o estado do canal (CSI) e complexidade de implementação do transceptor. Além disso, as soluções propostas baseiam-se nas principais especificações técnicas do projeto de parceria para a terceira geração (3GPP), com o objetivo de considerar aspectos práticos da implementação. Na primeira parte desta tese, dedicada ao projeto de formatação híbrida de feixes com base no esquema de divisão espacial e multiplexação conjunta, propomos uma estrutura para explorar um feedback limitado da CSI e reduzir a interferência intercelular, considerando diferentes condições de propagação em mmWave. Na segunda parte deste documento, investigamos uma estrutura de controle de potência do enlace de subida compatível com o projeto centrado em feixes da interface aérea da tecnologia de acesso por rádio de 5G. O esquema de sinalização proposto entre estações base permite um controle flexível da potência de transmissão capaz de aumentar a eficiência energética, aprimorando a taxa de dados do sistema e reduzir o consumo de energia, limitando a interferência nas células vizinhas. Esta tese explora diferentes paradigmas de aprendizado de máquina (ML) para otimizar a implantação da rede 5G. Investigamos como o ML pode ajudar na descoberta de propriedades desconhecidas do canal sem fio e no estabelecimento de estratégias bem-sucedidas de RMM a partir do conhecimento determinado pela interação com a rede. Análises numéricas são apresentadas para validar os métodos propostos e demonstrar que, apesar das limitações impostas pelas especificações técnicas do 3GPP, como restrições de hardware e sinalização disponível, as soluções propostas melhoram o desempenho do sistema e atendem a requisitos de engenharia relevantes, como melhoria da taxa de dados e aprimoramento da eficiência energética com reduzidas sobrecarga de sinalização e complexidade computacional.

Palavras-chave: Gerenciamento de recursos de rádio. Formatação híbrida de feixes. Controle de potência. Aprendizado de máquina.

ABSTRACT

The fifth generation (5G) of mobile communications has been envisioned to expand the capabilities of wireless networks and, consequently, to provide optimized support to several use cases and design requirements. In view of this, massive multiple-input multiple-output antenna arrays and the operation at the millimeter wave (mmWave) frequency range are important technical solutions able to support an expressive enhancement of the data traffic capacity, a recognizably relevant demand of 5G. In this context, the present thesis investigates radio resource management (RRM) techniques to explore these technologies and to overcome their main challenges, such as hostile propagation conditions, demanding channel state information (CSI) acquisition, and transceiver implementation complexity. Moreover, the proposed solutions rely on the main technical specifications from the third partnership project (3GPP) aiming to consider practical implementation aspects. In the first part of this thesis, devoted to the hybrid beamforming design based on the joint spatial division and multiplexing scheme, we propose a framework to exploit a limited CSI feedback and to reduce the inter-cell interference considering different mmWave propagation conditions. In the second part of this document, we investigate an uplink power control framework compliant with the beam-centric design of the air interface of 5G radio access technology. The proposed signaling scheme among base stations allows a flexible transmit power control able to increase the energy efficiency by the enhancement of the system data rate and to reduce the power consumption while limiting interference to neighbor cells. This thesis explores different machine learning (ML) paradigms to optimize 5G network deployment. We investigate how ML can help to uncover unknown properties of the wireless channel and establish successful RRM strategies from the knowledge determined by the interaction with the network. Numerical analyses are presented to validate the proposed methods and to demonstrate that, despite the limitations imposed by the 3GPP technical specifications, such as hardware restrictions and available signaling, the proposed solutions improve system performance and achieve relevant engineering requirements, such as data rate improvement and energy efficiency enhancement with reduced signaling overhead and computational complexity.

Keywords: Radio resource management. Hybrid beamforming. Power control. Machine learning.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

UE	User Equipment
UMi	Urban Micro
LOS	Line of Sight
NLOS	Non Line of Sight
RL	Reinforcement Learning
UPC	Uplink Power Control
BS	Base Station
SINR	Signal to Interference-plus-Noise Ratio
NEE	Network Energy Efficiency
5G	Fifth Generation
IMT-2020	International Mobile Communications for 2020 and Beyond
ITU	International Telecommunications Union
eMBB	Enhanced Mobile Broadband
QoE	Quality of Experience
mMTC	Massive Machine Type Communication
MTC	Machine Type Communication
RAN	Radio Access Network
URLLC	Ultra Reliable Low Latency Communication
4G	Fourth Generation
3GPP	Third Generation Partnership Project
NR	New Radio
LTE	Long Term Evolution
IIoT	Industrial Internet of Things
MIMO	Multiple Input Multiple Output
mmWave	Millimeter Wave
RF	Radio Frequency
MU	Multi User
CSI	Channel State Information
TDD	Time Division Duplex
DFT	Discrete Fourier Transform

RRM	Radio Resource Management
ML	Machine Learning
SL	Supervised Learning
UL	Unsupervised Learning
PHY	Physical
MAC	Medium Access Control
PMI	Precoding Matrix Indicator
ADC	Analog to Digital Converter
FDD	Frequency Division Multiplexing
SU	Single User
JSDM	Joint Spatial Division and Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
URA	Uniform Rectangular Array
SDMA	Space Division Multiple Access
RB	Resource Block
TTI	Transmission Time Interval
SNR	Signal to Noise Ratio
OFDM	Orthogonal Frequency Division Multiplexing
AWGN	Additive White Gaussian Noise
CSI-RS	Channel State Information Reference Signal
NMSE	Normalized Mean Square Error
FCM	Fuzzy C-Means
ZF	Zero Forcing
QuaDRiGa	Quasi Deterministic Radio Channel Generator
CDF	Cumulative Distribution Function
SDPC	Soft Dropping Power Control
WRAN	Wireless Regional Area Network
FPC	Fractional Power Control
OLPC	Open Loop Power Control
CLPC	Closed-Loop Power Control
DRL	Deep Reinforcement Learning
DQN	Deep Q-Network

AoD	Angle of Departure
AoA	Angle of Arrival
RS	Reference Signal
SS	Synchronization Signal
PUSCH	Physical Uplink Shared Channel
PUCCH	Physical Uplink Control Channel
PRACH	Physical Random Access Channel
UL-SCH	Uplink Shared Channel
MCS	Modulation and Coding Scheme
TPC	Transmit Power Control
DCI	Downlink Control Information
PBCH	Physical Broadcast Channel
SRI	Sounding Resource Indicator
MDP	Markov Decision Process
SG	Stochastic Game
SRS	Sounding Reference Signal
OSPC	Optimum Solution Power Control
MISO	Multiple Input Single Output
UMa	Urban Macro
EE	Energy Efficiency
SMA	Simple Moving Averaging
RLPC	Reinforcement Learning Power Control
QoS	Quality of Service
6G	Sixth Generation

LISTA DE SÍMBOLOS

B	Number of selected beam pairs
C	Number of cells
D	Number of antenna ports
E	Power behavior
J	Number of user equipment
K	Number of clusters
L	Length of the channel state information reference signal vector
N	Number of PMIs
P	Transmit power
R	Number of RF chains
T_{BS}	Beam sweeping duration
T_D	Transmission duration
T_{SS}	Frame duration
T	Total number of iterations
U	Number of transmit antennas
V	Number of receive antennas
Δ	Required received power according to modulation scheme and channel-coding rate
Γ	Signal to Interference-plus-Noise Ratio
Λ	Antenna element spacing
Ω	Network energy efficiency
Φ	Uplink path loss estimation
Ψ	Total data rate
Υ	Number of scatters
β	Bandwidth
Θ	Coupling matrix
χ	Size of power steps

δ	Transmit power adjustment command
ε	Exploration-exploitation parameter trade off
ν	Learning rate
γ	Signal-to-noise ratio
ι	Number of transmit power control commands
κ	Reference signal index
λ	Signal wavelength
$\mathbf{E}$	Matrix composed of eigenvectors
$\mathbf{H}$	Channel matrix
$\mathbf{R}$	Transmit covariance matrix
$\mathbf{c}$	Centroid of a cluster
$\mathbf{g}$	Codebook column at receiver side
$\mathbf{h}$	Channel vector
$\mathbf{r}$	Received signal in the channel measurement scheme
$\mathbf{s}$	Channel state information reference signal vector
$\mathbf{u}$	Dominant left singular vector
$\mathbf{v}$	Dominant right singular vector
$\mathbf{z}$	Discrete Gaussian noise vector
$\mathscr{A}$	Set of actions
$\mathscr{B}$	Set of suitable beam pairs
$\mathscr{C}$	Cluster
$\mathscr{G}$	Set of user equipment that compose the SDMA group
$\mathscr{S}$	Set of states
$\mathscr{V}$	Complex Gaussian random variable
$\mathscr{X}$	Optimal value of the expected discounted reward
$\mathfrak{R}$	Reward function
μ	Membership grade
ν	Objective function FCM

ω	Energy efficiency
ϕ	Path loss
π	Policy
ψ	Data rate
ρ	Reward function cooperation weight
τ	Transmission type index
θ	Elevation angle
φ	Discount rate
a	Action
b	Beam index
c	Cell index
d	Antenna port index
e	Eigenvalue
f	Exponent of fuzziness
h	Scalar channel
j	User equipment index
k	Scheduled user equipment index
l	Scatter index
o	Random number
q	Frame index
r	Reward
s	State
t	Time step
u	Transmit antenna element index
x	Transmitted signal
y	Received signal
z	Discrete Gaussian noise

SUMÁRIO

1 INTRODUCTION

The Fifth Generation (5G) of mobile communication has been designed to notably expand the capabilities of wireless networks and, consequently, enable a more connected society. The research and development of 5G started almost a decade ago and involves a coordinated global effort from academia and industry to determine technologies able to provide optimized support to a wide range of use cases and requirements [1].

5G networks, differently from its predecessors, are envisioned to support remarkably different types of use cases. According to the International Mobile Communications for 2020 and Beyond (IMT-2020) [2], a set of recommendations from the International Telecommunications Union (ITU) for 5G, the next generation of wireless networks supports three main use cases:

- Enhanced Mobile Broadband (eMBB): it is an extension of mobile broadband under current communication standards; it addresses human-centric connectivity, requiring high data rates and improved Quality of Experience (QoE) [3].
- Massive Machine Type Communication (mMTC): it requires the support to an extensive number of devices; moreover, it also demands the control of a highly dynamic and sporadic Machine Type Communication (MTC) traffic, huge signaling overhead, and Radio Access Network (RAN) congestion [4].
- Ultra Reliable Low Latency Communication (URLLC): it has rigorous requirements of latency, reliability, and availability due to the critical applications involved, such as remote medical surgery, traffic safety control, and industrial manufacturing automation [5].

The major engineering requirements that must be achieved in 5G networks can be briefly summarized as extremely high data rate, high traffic volume, ubiquitous coverage, very low latency, ultra-reliability, high energy efficiency, and a massive number of devices and heterogeneous connections [6, 7, 8]. Therefore, the overall technical goals of 5G networks are set to support these requirements at similar cost and energy consumption levels compared with Fourth Generation (4G) networks.

To address these demands, the Third Generation Partnership Project (3GPP) started years ago the development of technical specifications to allow the employment of more advanced solutions. The first phase of 5G standards constitutes the 3GPP 38 series Release 15 [9]. It comprises the 5G core and a new radio-access technology, which is known as New Radio (NR). The Release 15 is composed of the non-standalone and standalone 5G radio specifications. The first one has the control plane connection to the core network through Long Term Evolution

(LTE), while the second one does not rely on LTE core [10]. Thus, the difference between them primarily affects higher layers and the interface to the core network.

The second phase of 5G is being standardized by 3GPP in Release 16. It brings 3GPP IMT-2020 submission for a full 3GPP 5G system [11]. In addition to enhancing the previous release features, Release 16 focuses on enabling full support for the Industrial Internet of Things (IIoT). Then, the focus of Release 16 is on new features for URLLC, while the main focus of Release 15 was on eMBB [12].

More 5G system enhancements are set to follow in Release 17 in the next years. The scope has recently been approved and scheduled for delivery in 2021 [13]. Release 17 is thought to lead to the introduction of new features for the three main use case families (eMBB, URLLC, mMTC) as well as customizing new use cases emerging in the market [14]. Therefore, the evolution of 5G technical specifications focuses on the enhancement of features introduced in previous releases and on the development of new features to further expand the applicability of 5G networks.

The 5G specifications provide several features to meet the challenging requirements from IMT-2020. Some of the most important technical solutions considered in the deployment of 5G and relevant to the development of this thesis are

- wider transmission bandwidths and, consequently, higher frequency carriers [15];
- massive Multiple Input Multiple Output (MIMO) antenna arrays [16];
- beam-centric design [17].

In the following, we discuss how these technologies motivated and influenced the development of the studies carried out in this thesis.

1.1 Key Technologies

Among the design requirements outlined previously, the need for very high traffic capacity recognizably got relevant research attention in the last years [18, 19, 20]. The development of technologies and architectures to properly serve the continuing demands of customers imposes the joint effort of researchers and regulators to accommodate standards and technologies respecting the spectrum and cost limitations [21, 22, 23].

The exploration of wider bandwidths is an effective and straightforward method to provide the foreseen data rate demands. However, the available bandwidth at lower frequencies is limited, and the spectrum shortage is an unavoidable issue. Consequently, it motivated the use

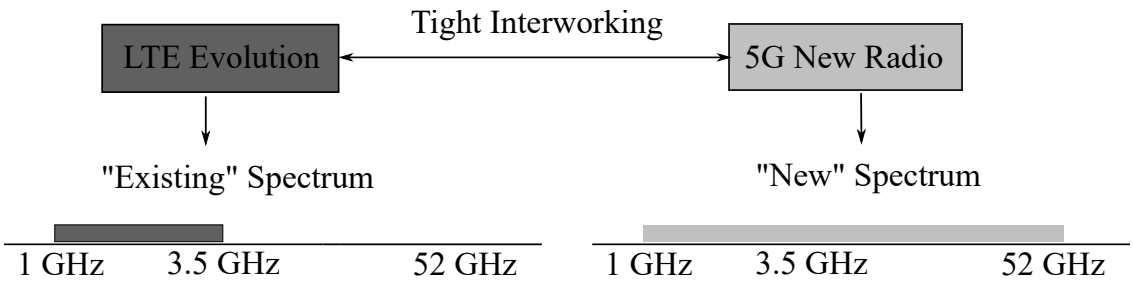
of higher frequencies because of the large amount of free spectrum. [24].

On one hand, LTE supports a maximum bandwidth of 100 MHz, considering carrier aggregation, and is expected to operate below 6 GHz. Specifically, its highest frequencies are 3.5 GHz for licensed spectra and 5 GHz for unlicensed spectra [17]. On the other hand, NR is envisioned to support several channel bandwidths and operate licensed spectra from sub-1 GHz up to 52.6 GHz, as specified in 3GPP Release 15 [25]. These carrier frequencies are divided within two possible ranges. The frequency range 1 (FR1) corresponds to the interval $0.41-7.125$ GHz and the frequency range 2 (FR2) is associated with the carrier frequencies between 24.25 and 52.6 GHz, which belongs to the Millimeter Wave (mmWave) spectrum [26]. The supported channel bandwidths in NR are 100 MHz in FR1 and 400 MHz in FR2 [25]. If even larger bandwidths are to be supported, carrier aggregation can be used [27]. A tight integration of NR and LTE is envisioned in order to efficiently aggregate the traffic between them, as it can be seen in Figure 1.

NR has a scalable waveform numerology to enable diverse services on this wide range of frequencies and bandwidths with different subcarrier spacings. In 3GPP Release 15, five subcarrier spacings are specified: 15 kHz, 30 kHz, 60 kHz, 120, and 240 kHz [25]. The choice of the subcarrier spacing depends on various factors, such as service requirements, hardware impairments, mobility, performance and implementation complexity [26]. A small subcarrier spacing provides a relatively long cyclic prefix in absolute time at a reasonable overhead while higher subcarrier spacing is required to handle the increased phase noise at higher carrier frequencies [27].

Although exploration of the mmWave spectrum at FR2 is very appealing, its implementation involves many challenges. The propagation conditions are hostile because diffraction

Figura 1 – Licensed carrier frequency ranges associated with LTE and NR.



Fonte: Created by the author.

and material penetration incur a greater loss in higher frequencies [28]. Moreover, the mmWave propagation is characterized by fast channel fluctuations. For a given mobile velocity, channel coherence time is linear in the carrier frequency, meaning that it will be very small in the mmWave range. Consequently, from a system level perspective, it implies in a highly intermittent connectivity scenario [29]. Thus, it is crucial to the understanding of mmWave propagation the realization of experimental measurements and the development of empirically-based propagation channel models, as it can be seen in [30, 31, 32] and references therein.

The strong attenuation observed in the mmWave propagation can be compensated by means of advanced multiantenna techniques, which is one of the motivating factors for the beam-centric design in NR. The very small wavelengths of mmWave signals combined with advances in low-power CMOS Radio Frequency (RF) circuits enable large numbers of miniaturized antenna elements to be placed in small dimensions. These multiple antenna systems can be used to form very high gain and electrically steerable arrays to overcome path loss conditions [33]. Therefore, an appropriate beamforming scheme with a great number of antennas allows focusing the signal in a desired direction to overcome the unfavorable channel propagation conditions. Although the spatial resolution increases, it imposes difficulties on initial access, handoff and interference management [34].

The extensive usage of antenna elements, formally called massive MIMO, is considered an essential technology to provide very high data rates, spectral efficiency, enhanced link reliability, and coverage to the next generation of broadband networks [35]. The concept, introduced in [36] and further developed in [37], uses a massive number of antenna elements at the transmitter, which achieves multiplexing and throughput gains of Multi User (MU)-MIMO by serving many UEs within the same time-frequency resource [38, 39, 40, 41].

Massive MIMO systems consider the BS equipped with a number of antennas much larger than the number of active UEs per time-frequency resource. This arrangement smooths out the channel responses due to spatial diversity, and the effects of fast fading are significantly reduced. If the BS has an “infinite” number of antennas the small scale randomness decreases due to law of large numbers [37]. However, the deployment of massive MIMO involves several issues that were not relevant in classical MIMO. The classical solution requires one RF chain per antenna element [42], but it increases implementation complexity, hardware cost, and energy consumption in massive MIMO [43].

Hybrid beamforming has emerged with the promise of having transceivers with a

massive number of antennas connected to very few RF chains. Essentially, hybrid beamforming replaces the fully digital precoding by analog beamforming at RF domain and digital precoding at baseband. The transmitter side uses the baseband digital precoder to process data streams and so produces outputs that are upconverted to the RF segment and mapped via an analog precoder to be transmitted by the BS antenna elements. At the receiver side, an analog combiner merges RF signals from UEs' antennas to create outputs that are downconverted to baseband and further combined, producing a signal for detection/decoding [43].

The air interface of NR is beam-based, which means that the channels and signals can be beamformed for data transmission and control-plane procedures. The large number of antenna elements can be used to extend coverage, enable massive MIMO, and allow interference avoidance by spatial separation [17].

The acquisition of Channel State Information (CSI) is a fundamental feature to support NR beam-centric design. In general, CSI can be acquired by means of codebook-style feedback or through Time Division Duplex (TDD) channel reciprocity [12]. 3GPP Release 15 provides two types of codebook-based CSI feedback. The Type I CSI reporting is similar to the full-dimensional MIMO codebooks in LTE Release 13. It consists in the selection of up to eigth beams from an oversampled grid of Discrete Fourier Transform (DFT) beams. The Type II CSI reporting is similar to the LTE Advanced Release 14 CSI codebook but allowing the selection of more beams and having finer granularity in the amplitude and phase quantization [26]. Compared with Type I, there is also an amplitude scaling factor for each beam which can be wideband or a combination of wideband and frequency selective.

Therefore, the impact of the CSI acquisition on a hybrid beamforming scheme in a mmWave propagation scenario is a relevant research question that we investigate Chapter 2.

Besides the enhancement of the data rate, another relevant 5G design requirement evaluated in this thesis is the improvement of the energy efficiency. The 5G networks must be designed to improve the data rate at similar or lower transmit power consumption [18]. That is, with an increasing number of connected devices, maximizing the data rate by simply increasing the transmission power is not sustainable [44].

Therefore, the UPC constitutes an essential design problem for 5G networks. This important Radio Resource Management (RRM) technique provides mechanisms to increase the energy efficiency by the enhancement of the system data rate and to reduce the power consumption while limiting interference to neighbor cells [45].

4G LTE supports several solutions based on well-founded technical literature to the UPC problem with distinct objectives regarding different deployment scenarios and services [46, 47, 48]. NR UPC follows a structure similar to LTE, i.e., it is based both on signal strength measurements performed by the UE itself (open-loop power control paradigm) as well as by measurements performed bt the BS, which are used to determine transmit power commands that are fed back to the UE as part of the downlink control signaling (closed-loop power control paradigm) [45].

Despite the advances promoted by LTE, the new use cases introduced in 5G, such as eMBB, URLLC, and mMTC, impose challenging requirements compared with the demands of the previous generation [17]. Hence, the 5G radio access technology is expected to expand the capabilities of wireless networks to fulfill these hard demands.

NR has as one innovative aspect the beam-centric design of channels and signals. Therefore, a large number of antenna elements at the transmitters and receivers are used for beamforming to enhance coverage and to improve the interference mitigation [17]. Consequently, the UPC problems in 5G NR can exploit such design flexibility, but will witness more challenging design requirements compared to classical solutions.

Then, the development of a new UPC framework to support 5G NR demands mentioned previously requires more sophisticated tools. In this context, the design of an UPC solution compliant with the beam-centric design paradigm of the NR radio access technology is a relevant research topic and is the focus of Chapter 3.

In the following, it is analyzed how Machine Learning (ML) emerges as a powerful source of mechanisms to optimize 5G network deployment. That is, it is reviewed how ML can help to uncover unknown properties of wireless networks, identify correlations, and suggest novel ways to meet knowledge from the network to meet the ambitious 5G design requirements.

1.2 Machine Learning Applications in 5G

ML is a promising tool to assist network operators in making the 5G vision conceivable as the emergence of new use cases imposes ambitious design requirements and the systems reach notable levels of complexity [49, 50]. It is critical to enhance intelligence in the 5G networks to overcome the increasingly configuration issues and to provide more flexible features such as self-configuration, self-optimization, and self-healing [51, 52, 53]. In addition to the references mentioned previously, the authors of [54, 55, 56, 57] summarize a wide range

of ML-based models along with the corresponding applications in 5G networks.

ML is the set of algorithms which use statistical techniques to allow machines to improve its decision-making capabilities as they acquire more knowledge [50]. That is, it provides techniques able to detect anomalies, forecast future scenarios, get insights on complex systems from the available data, discover hidden patterns, and find potential solutions by interacting with the environment [51, 50]. According to the behavior of the learning agents and the nature of the data used in the learning system, ML algorithms are typically classified into three main categories, namely Supervised Learning (SL), Unsupervised Learning (UL), and RL.

SL algorithms aim to determine a general rule to map inputs into outputs. The main idea is to train a learning model with labeled data indicating the inputs with their corresponding target outputs. In the following, this model is used to determine the most adequate outputs from inputs different of those considered in the training step [58]. Therefore, the objective is to acquire the generalization ability, which refers to the capability to determine appropriate outputs for new inputs [59].

SL algorithms require a large amount of data to train, test, and validate their learning models since the number of samples influences the algorithm robustness [50]. Therefore, from a Physical (PHY) and Medium Access Control (MAC) layer perspective, training a SL model using over-the-air feedback is prohibitively expensive in terms of signaling control overhead [60].

UL algorithms consider the situation where there is not supervisor and the agent learns by itself. That is, the agent collects data through measurements from the system and tries to extract useful knowledge without any external guidance [59]. Therefore, the input information on UL algorithms does not possess labeled data. The UL agent has to find the hidden patterns and relationships in the data based on its own knowledge [50].

UL is also a promising ML paradigm application in 5G networks since its main objective is to make the agent capable of learning without any supervision or human intervention. Compared with SL, UL does not require large training data, which implies a reduced signaling overhead. However, UL learning process is quite complex as compared to SL, which implies a higher computational cost [4].

Among the various tasks which could be performed exploring the UL paradigm, stand out dimensionality reduction and clustering [4]. The manipulation of high-dimensional data is often cumbersome in practical data analysis. Thus, the dimensionality reduction allows reducing dimensionality while preserving intrinsic information contained in the data [59]. Clustering

aims to categorize input samples into cluster without any supervision. Usually, similar samples are supposed to belong to the same cluster, and dissimilar samples are supposed to belong to different clusters [59].

Clustering is frequently used in network design, as it can be seen in [49, 50, 51, 52] and references therein. Therefore, it plays a prominent role throughout this thesis. Chapter 2 investigates different techniques to cluster UEs in a efficient way according to some 5G design criteria, such as data rate enhancement.

The RL paradigm aims to determine the most suitable actions to take in a given system in order to maximize a reward function. The learning agent does not have examples of optimal actions, but it must instead discover them by a trial and error process [58]. At each decision time, the learning agent chooses an action available at a current state, then, the system determines the reward function and the new state. The goal is the determination of the best policy to maximize the reward function by selecting the most proper action in a given state [50].

In the RL paradigm, actions may affect not only the immediate reward but also the next states and the subsequent rewards. Thus, the trial-and-error search and delayed reward are important distinguishing features of RL compared with SL and UL [61]. Another challenge that arises in RL is the trade-off between exploration and exploitation. That is, the learning agent has to exploit what it has already experienced in order to obtain reward, but it also has to explore other actions in order to make better action selections in the following decisions. Neither exploration nor exploitation can be pursued exclusively without failing at the task [61].

RL is a relevant tool to empower networks with autonomous algorithms provided with adaptability and capable of taking advantage of experience when making decisions [52]. Chapter 3 investigates how a learning agent can determine suitable transmit power control commands based on the knowledge acquired with its dynamic interaction with the environment. The proposed UPC strategy yields a proper utilization of limited resources to improve energy efficiency based on the RL paradigm.

1.3 Main Contributions

This thesis investigates RRM techniques in the context of 5G networks and based on ML paradigms. The proposed solutions prioritize the practical implementation aspects, relying on the main technical specifications from 3GPP Release 15. Thus, it is researched how ML techniques can be implemented considering the limitations imposed by these conditions, such as

hardware restrictions and the information available in view of the existing signaling.

Therefore, flexible frameworks are designed to overcome these restrictions and bring relevant 5G engineering requirements, such as as data rate improvement, energy efficiency enhancement, interference mitigation, and reduced signaling overhead.

The first part of this thesis is focused on the development of a hybrid beamforming scheme on a mmWave scenario. The main contributions of this investigation can be summarized as follows:

- design of a hybrid beamforming scheme under limited feedback channels;
- development of a user selection framework based on UL with reduced CSI feedback;
- comprehensive simulation campaign to investigate the performance of the proposed frameworks under different system conditions;
- computational complexity analysis of the proposed user clustering framework.

The second part of this thesis develops a beam-centric UPC framework aiming at the enhancement of the energy efficiency. It is considered a multi-cell mmWave massive MIMO scenario. The most relevant contributions of this research can be outlined as follows:

- development of a flexible power control strategy that takes into account the coordination among multiple beams and the limitation aspects from 3GPP;
- formulation of UPC framework based on the multi-agent RL paradigm, i.e., design of states, actions, and rewards compliant with the 3GPP specifications;
- development of a signaling scheme that allows cooperation among entities endowed with intelligence in a multi-cell system;
- evaluation of the performance of the proposed UPC framework through numerical simulations.

1.4 Thesis Organization

The content of this thesis is organized in four chapters, including this introductory chapter. The chapters cover different aspects of RRM in mmWave massive MIMO communications systems. Every chapter is meant to be self-contained so that the reader can read them independently without loss of information.

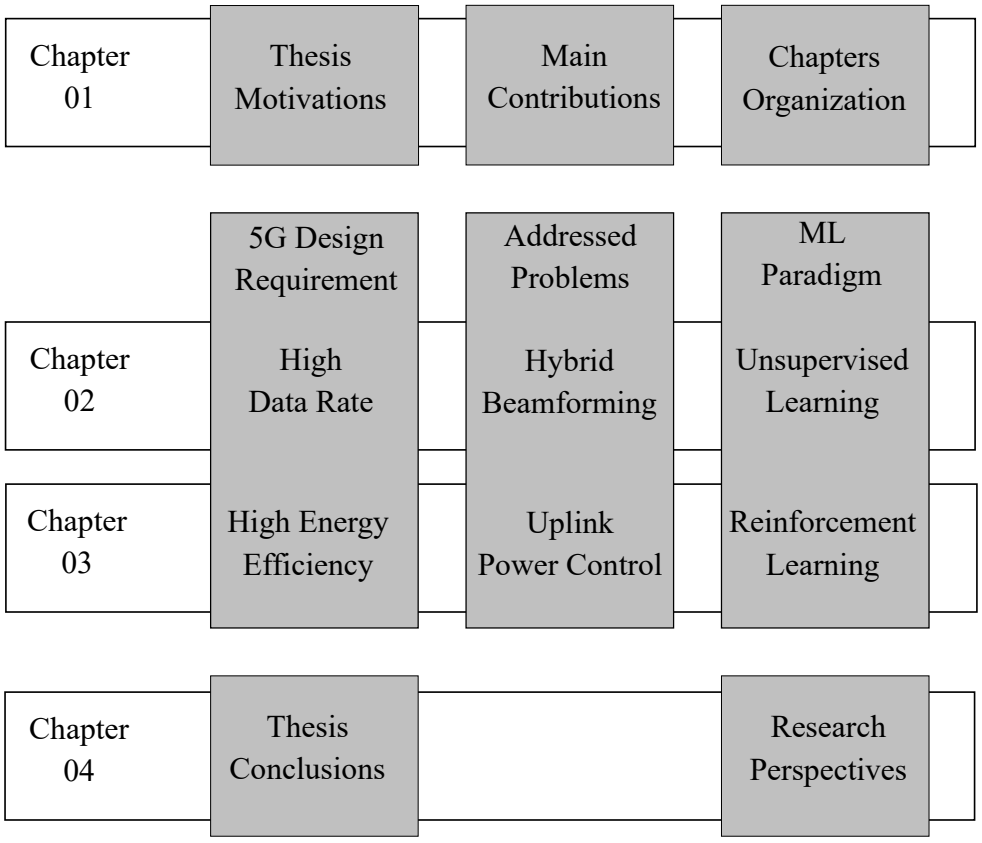
Figure 2 presents the thesis structure, highlighting the goals and the most relevant concepts addressed in each chapter. The main topics covered in Chapters 2 and 3 are described by their main design requirements, the key technologies investigated, and the ML paradigms

employed. In the following, we briefly describe the contents of the remaining chapters.

Chapter 2 proposes a hybrid beamforming design with reduced CSI feedback scheme. We explored a beam sweeping procedure to collect channel measurements and CSI feedback report. Thereby, the BS can perform an adequate estimation of the channel characteristics with reduced signaling overhead. Consequently, it is required short pilot sequences and very few Precoding Matrix Indicators (PMIs) to properly describe channel behavior. Moreover, it is also evaluated different user selection strategies based on an unsupervised learning framework that exploits the CSI provided by the proposed beam sweeping scheme. User selection algorithms based on hard and fuzzy clustering paradigms are compared in terms of computational complexity and ability to explore the reduced CSI to reduce MU interference and improve total data rate.

Chapter 3 proposes an UPC framework compliant with the technical specifications of the 5G wireless networks. The fundamentals of RL are exploited to develop a power control algorithm able to learn a strategy that maximizes the total data rate on the uplink channel and mitigates the neighbor cell interference. The BS uses a set of commands to specify by how much the UE transmit power should change. After implementing such commands, the UE reports a

Figura 2 – Thesis Organization.



Fonte: Created by the author.

set of information to its serving BS, which uses them to predict the next commands to achieve a suitable UE transmit power level. The BS converts the UE reports into rewards according to a predefined cost function, which impacts the long-term behavior of the UE transmit power. Besides the evaluation of the impact of the design parameters on its behavior, the proposed UPC is compared in terms of computational complexity and required signaling with two classical power control solutions.

Chapter 4 summarizes the main conclusions of this thesis and discusses some research perspectives for future works.

1.5 Scientific Production

The content and scientific contributions present in Chapter 2 have been partially published with the following bibliographic information:

(J1) **F. H. Costa Neto**, D. C. Araújo, T. F. Maciel, "Hybrid Beamforming Design based on Unsupervised Machine Learning for Millimeter Wave Systems," *International Journal of Communication Systems*, vol.33, no. 5, Jan. 2020.

(J2) **F. H. Costa Neto**, T. F. Maciel, "SDMA Grouping Based on Unsupervised Learning for Multi-User MIMO Systems," *Journal of Communication and Information Systems*, vol.35, no.1, May 2020. **Invited Paper**

(C1) **F. H. Costa Neto**, D. C. Araújo, T. F. Maciel, "Hybrid Beamforming for Massive MIMO Systems with Limited CSI Feedback," in *Brazilian Symposium on Telecommunications and Signal Processing*, Sep. 2019, pp. 1-5

(C2) **F. H. Costa Neto**, T. F. Maciel, "SDMA Grouping for 5G Systems," in *Brazilian Symposium on Telecommunications and Signal Processing*, Sep. 2019, pp. 1-5

The content and scientific contributions present in Chapter 3 have been partially published with the following bibliographic information:

(J3) **F. H. Costa Neto**, D. C. Araújo, M. P. Mota, T. F. Maciel, A. L. F. Almeida, "Uplink Power Control Framework Based on Reinforcement Learning for 5G Networks," *IEEE Transactions on Vehicular Technology (Under Submission Process)*, 2020.

(P1) **F. H. Costa Neto**, D. C. Araujo, A. L. F. Almeida, R. Timo "Reinforcement Learning Based Uplink Power Control Providing Joint Optimization for Multiple Uplink Power Control Loops in a Distributive Manner", regular patent filed no. PCT / EP2020/062641, May, 2020.

The following publications do not directly contribute to the content of this thesis, but addresses RRM problems in the context of 5G wireless networks:

(C3) M. P. Mota, D. C. Araújo, **F. H. Costa Neto**, F. R. M. Lima, F. R. P. Cavalcanti, "Adaptive Modulation and Coding based on Reinforcement Learning for 5G Networks", in *IEEE GLOBECOM*, Dec. 2019.

(C4) W. V. F. Mauricio, D. C. Araújo, **F. H. Costa Neto**, F. R. M. Lima, , T. F. Maciel, "A Low Complexity Solution for Resource Allocation and SDMA Grouping in Massive MIMO Systems", in *IEEE International Symposium on Wireless Communication Systems*, Aug. 2018.

The following publication addresses RRM problems considering cross-layer optimization, the initial focus of the PhD course:

(J4) **F. H. Costa Neto**, E. B. Rodrigues, D. A. Sousa, T. F. Maciel, F. R. P. Cavalcanti, "QoS-Aware Scheduling Algorithms to Enhance User Satisfaction in OFDMA Systems", *Transactions on Emerging Telecommunications Technologies*, vol. 28, pp. 1-15, Apr. 2017.

This thesis was developed under the context of Ericsson/UFC technical cooperation projects:

- UFC.47: Radio Link Control, Resource and QoS Management for Beam-Based NR. From November/2018 until November/2020
- UFC.44: Hybrid Beamforming and Massive MIMO for 5G Wireless Systems. From October/2016 until October/2018.
- UFC.40: Provision and Quality of Service Control for Wireless Systems of 5th Generation. From September/2014 until September/2016.

Throughout these research projects, the following technical reports have been produced:

(R1) **F. H. Costa Neto**, A. L. F. Almeida, "Beam-Based Uplink Power Control Framework Using Reinforcement Learning", *Third Technical Report (TR03) UFC.47*, Wireless Telecommunications Research Group, Apr. 2020

(R2) **F. H. Costa Neto**, D. C. Araújo, A. L. F. Almeida, "Beam-Based Uplink Power Control Framework Using Reinforcement Learning", *Second Technical Report (TR02) UFC.47*, Wireless Telecommunications Research Group, Nov. 2019.

(R3) **F. H. Costa Neto**, D. C. Araújo, A. L. F. Almeida, "Uplink Power Control Framework Using Reinforcement Learning", *First Technical Report (TR01) UFC.47*, Wireless Tele-

communications Research Group, Apr. 2019.

(R4) **F. H. Costa Neto**, D. C. Araújo, A. L. F. Almeida, T. F. Maciel, "Hybrid Beamforming Design Exploiting Implicit and Explicit CSI in Massive MIMO Systems", *Fourth Technical Report (TR04) UFC.44*, Wireless Telecommunications Research Group, Oct. 2018.

(R5) **F. H. Costa Neto**, D. C. Araújo, A. L. F. Almeida, T. F. Maciel, "Hybrid Beamforming for Massive MIMO Systems with Limited CSI Feedback", *Third Technical Report (TR03) UFC.44*, Wireless Telecommunications Research Group, Mar., 2018.

(R6) **F. H. Costa Neto**, D. C. Araújo, A. L. F. Almeida, T. F. Maciel, "SDMA Grouping for 5G Systems Using Hybrid Beamforming", *Second Technical Report (TR02) UFC.44*, Wireless Telecommunications Research Group, Oct. 2017.

(R7) **F. H. Costa Neto**, D. C. Araújo, A. L. F. Almeida, T. F. Maciel, "SDMA Grouping for 3D MU-MIMO Systems Using Hybrid Beamforming", *Second Technical Report (TR02) UFC.44*, Wireless Telecommunications Research Group, Oct. 2017.

(R8) **F. H. Costa Neto**, D. C. Araújo, A. L. F. Almeida, T. F. Maciel, "Resource Allocation and User Grouping for Massive MIMO", *First Technical Report (TR01) UFC.44*, Wireless Telecommunications Research Group, Mar. 2017.

(R9) **F. H. Costa Neto**, E. B. Rodrigues, T. F. Maciel, "Summary of Radio Resource Management Algorithms", *Fourth Technical Report (TR04) UFC.40*, Wireless Telecommunications Research Group, Oct. 2016.

2 HYBRID BEAMFORMING DESIGN BASED ON UNSUPERVISED LEARNING

In this chapter, it is proposed a hybrid beamforming design with reduced Channel State Information (CSI) feedback. Each Base Station (BS) uses a beam sweeping procedure to provide channel measurements at User Equipment (UEs). Such measurements are fed back to the BS to perform an adequate estimation of the channel characteristics with reduced signaling overhead. Consequently, it is required short pilot sequences and very few Precoding Matrix Indicators (PMIs) to properly describe channel behavior. Furthermore, it is also evaluated different user selection strategies based on an Unsupervised Learning (UL) framework that exploits the channel information provided by the considered beam sweeping scheme. The performance evaluation indicates that the proposed user clustering based on fuzzy c-means can efficiently explore the reduced CSI. The proposed hybrid beamforming scheme successfully reduces the Multi User (MU) interference and achieves significant gains in total data rate as channel conditions and interference environment become more challenging.

The remainder of this chapter is organized as follows. Section 2.1 discusses related works and the main contributions of this chapter. Section 2.2 describes the principal assumptions about the signal model. Section 2.3 presents the proposed channel measurement technique, which is based on beam sweeping. Section 2.4, discusses the proposed hybrid precoding scheme with limited CSI feedback. Finally, performance results are shown in Section 2.5, and the main remarks are drawn in Section 2.6.

2.1 Literature Review and Contributions

Hybrid beamforming is a viable approach for the deployment of massive Multiple Input Multiple Output (MIMO) systems operating at Millimeter Wave (mmWave) carrier frequencies due to the reduction of the number of Analog to Digital Converters (ADCs) and Radio Frequency (RF) chains [62]. Several studies, such as [63] and [64], evaluated how it can provide successful strategies to improve the energy efficiency of Fifth Generation (5G) systems. The authors of [63] investigated optimization solutions based on hybrid beamforming to overcome conventional digital precoding algorithms in terms of energy and cost efficiency. The authors of [64] formulated a joint optimization problem of computation and communication power consumption for MU massive MIMO systems. Unlike the conventional approaches, which only consider the communication power, it is noticed that the energy efficiency of massive MIMO

systems reduces with the increase in the number of antennas and RF chains.

A major challenge of massive MIMO systems is the determination and feedback of CSI among transmit and receive antennas, which requires a considerable amount of spectral resources [65]. Without this information, the system is not capable of delivering very high data rates and consequently, 5G requirements cannot be reached. Several hybrid beamformers have been established under the assumption of perfect CSI at the BS [66, 67, 68, 69]. However, a real system must employ procedures that exploit reference signals and take into account the hybrid structure to estimate the CSI [43]. In Time Division Duplex (TDD) systems, the instantaneous CSI at the transmitter can be achieved using time reciprocity of the uplink and downlink channels. Therefore, a large amount of antenna elements either the UEs or at the BS increases the length of the training sequences, which reduces the spectral efficiency [43]. In Frequency Division Multiplexing (FDD) systems, the overhead is even more intense, since it is required the downlink training and uplink feedback increases with the number of antenna elements [43]. Therefore, it is necessary for the design of hybrid beamforming systems with reduced CSI volumes to relieve the signaling overhead. The authors of [70] first proposed hybrid beamformers using average CSI for a Single User (SU)-MIMO system. For the massive MIMO case, a scheme called Joint Spatial Division and Multiplexing (JSDM) was proposed in [71].

The JSDM scheme provides a two-stage precoder which naturally suits the hybrid beamforming structure. The first stage is an analog beamforming based on the slowly varying second-order channel statistics. The UEs with similar transmit channel covariance eigenspace are grouped together and an analog precoder based on block diagonalization mitigates the inter-group interference [71]. As a consequence, the downlink training can be parallelly conducted in the different groups, which can be thought as virtual sectors. It leads to a reduction of downlink training and uplink feedback overhead proportional to the number of groups since each UE only needs to feedback the intra-group channels. The second stage uses a standard MU-MIMO digital precoder based on the instantaneous knowledge of the resulting reduced dimensional channel matrix to distinguish UEs inside each virtual sector.

The authors of [72] extended the JSDM scheme by the evaluation of different approaches to partition the UE population into groups, namely the k-means algorithm applied to the Grassmanian manifold and the fixed quantization of the Grassmanian manifold based on the minimum chordal distance. In [73], it is evaluated an improved k-means user grouping scheme which instead of chordal distance, considers the weighted likelihood as the metric to

partition UEs. This study is extended in [74] with the evaluation of novel grouping metrics and clustering algorithms, such as k-medoids and hierarchical clustering. The aforementioned works were based on the one cluster scattering model, i.e., the multi-path components arrive at the BS under a very constrained angular range. Therefore, the UEs can be conveniently separated in the first precoding stage since they are associated with multi-paths that are disjoint in the angular domain. However, this channel assumption does not hold in more realistic scenarios.

In our study, we consider a mmWave channel model in an urban environment [75], where we can observe spatially correlated multi-paths. In this context, the transmit channel covariances tend to be partially overlapped among themselves. Consequently, the partitioning of UEs is more difficult. The qualitative principle of JSDM, that establishes the mutual orthogonality between the eigenspaces of the groups served on the same time-frequency resource, reduces the number of feasible partitions, limiting the reduction of signaling and feedback overhead. To overcome this issue, we investigate a more flexible partitioning strategy, called fuzzy clustering, which considers that each UE can have a membership degree in more than one virtual sector. Consequently, in the cases in which it is hard to decide that UEs belong to only one virtual sector, fuzzy clustering can reach a better decision.

Most of the previous works also assumed that the BS has perfect knowledge of the UE channel covariance matrix, which can be accurately learned and tracked since it is slowly-varying over long periods of time. It is considered an explicit CSI feedback, where each UE sends to the BS a representation of a CSI as observed by the UE, such as the transmit covariance matrix. The UE may also perform some additional processing of the observed channel covariance matrix and feedback their dominant eigenvectors.

In our study, we investigate a more realistic scheme of CSI feedback based on the measurement and report framework specified in 5G NR physical layer standard [76]. We consider an implicit feedback scheme, where each UE sends back a set of PMIs to the BS, which point to the index of the codewords selected by the UE in a codebook known at both UE and BS sides, and their respective weights, which are used in the recomposition of estimated channel matrices, respectively.

Motivated by the above discussion, we investigate a low-complexity hybrid beamforming algorithm for downlink massive MIMO mmWave systems and assume the availability of only limited feedback of CSI between UEs and the BS. The main contributions of this chapter can be summarized as follows:

1. evaluation of a channel measurement technique based on beam sweeping;
2. evaluation of the JSDM scheme considering a realistic mmWave channel model and different urban environments;
3. evaluation of UE selection algorithms based on an UL, i.e., different UEs partitioning schemes considering hard and fuzzy clustering.

Notation: bold lowercase and uppercase letters represent column vectors and matrices, respectively. $(\cdot)^T$ and $(\cdot)^H$ stand for transpose and Hermitian of a matrix, respectively. $|\cdot|$, $\|\cdot\|_2$, $\|\cdot\|_F$ denote the absolute value, the Euclidean norm, and the Frobenius norm, respectively. $\mathbf{vec}(\cdot)$ represents the vectorization of a matrix into a column vector. Calligraphic upper-case letters denote sets, and for them $|\cdot|$ denotes set cardinality. $\mathbb{E}\{\cdot\}$ denotes expectation operator.

2.2 Signal Model

We consider a downlink channel of a massive MIMO system employing Orthogonal Frequency Division Multiple Access (OFDMA). The system is composed of one BS that services J UEs. The BS has a Uniform Rectangular Array (URA) composed of U_{v} vertical and U_{h} horizontal antenna elements, where the total number of antennas $U = U_{\mathrm{v}}U_{\mathrm{h}}$, and R RF chains available so that $R \leq U$. Each UE is equipped with a single omnidirectional antenna. In each time slot, K UEs out of the entire set of J UEs are selected to compose a Space Division Multiple Access (SDMA) group $\mathscr{G}$ which receives data sharing a same frequency-time Resource Block (RB) in space. In each resource, the transmitter uses its U antennas to send a data stream to each of the selected UEs. The remaining $(J-K)$ UEs have their data stored in a buffer. Thus, considering an appropriate time interval, all UEs, including the $(J-K)$ UEs not selected in a certain round, will be scheduled later and receive data, i.e., all UEs will empty their buffers and consume all information. The multiple streams are spatially multiplexed using a precoder matrix that mitigates the MU interference. The symbol transmitted to the k-th UE is defined by x_k, where it is assumed that $\mathbb{E}\{x_k x_k^*\} = 1, \forall k \in \{1,2,\cdots,K\}$.

The RB is defined as a set of contiguous and equally spaced subcarriers in the frequency domain and a set of consecutive symbols in the time domain. This group of symbols corresponds to a subframe, that represents the Transmission Time Interval (TTI). We define the subcarrier width and the TTI duration according to the carrier frequency, considering the achievable Signal to Noise Ratio (SNR) subject to phase noise and channel delay spread [77]. More details on this topic are provided in Section 2.5. The downlink channel matrix between

the BS and the k-th UE is denoted by $\mathbf{H}_k \in \mathbb{C}^{U_v \times U_h}$, and its vectorized form is defined as $\mathbf{h}_k = \mathbf{vec}(\mathbf{H}_k) \in \mathbb{C}^{U_v U_h \times 1}$. The channel vector of a given RB is associated with its middle subcarrier and first transmitted Orthogonal Frequency Division Multiplexing (OFDM) symbol in a TTI.

The hybrid beamforming is defined as $\mathbf{f}_k = \mathbf{F}_{\mathrm{RF},k}\mathbf{f}_{\mathrm{BB},k}$, where $\mathbf{F}_{\mathrm{RF},k} \in \mathbb{C}^{U \times R}$ represents the analog precoder matrix while $\mathbf{f}_{\mathrm{BB},k} \in \mathbb{C}^{R \times 1}$ is the digital precoding vector. The analog precoder matrix has elements with equal magnitude and phase defined as constant to a specific range of RBs. The digital precoder is implemented in the baseband domain, so it is RB specific. Therefore, the prior-filtering receive symbol y_k at the k-th selected UE is

$$y_k = \mathbf{h}_k^T \mathbf{f}_k \sqrt{P_k} x_k + \sum_{j \neq k}^{K} \mathbf{h}_k^T \mathbf{f}_j \sqrt{P_j} x_j + z_k, \tag{2.1}$$

where P_k is the transmit power allocated to the stream associated to the k-th UE, the second term on the right-hand side of (2.1) represents the inter-cell interference, and z_k is the Additive White Gaussian Noise (AWGN), defined as $\mathscr{CN}(0, \sigma^2)$, with mean zero and variance σ^2.

2.3 Channel Measurement and CSI Feedback

In order to determine the precoders appropriately, the BS should have knowledge about the CSI. Herein, we exploit the codebook-based CSI feedback Type II framework, specified in 5G NR physical layer [76] to measure and report CSI to the BS. This codebook type has a higher resolution than Type I because it enables the UE to describe the CSI as a linear combination of multiple beams while the type I only considers the strongest beams [78]. This is a very attractive 5G New Radio (NR) feature since we need a very good description of the channel in MU-MIMO scenarios to avoid spectral efficiency losses due to excessive intra-cell interference. Moreover, the codebook structure is a product of two matrices which is similar to the hybrid beamforming. There is a wideband matrix, that could be implemented by the analog beamforming, while the second matrix is designed according to the subband channel, that could be implemented by the digital beamforming.

The CSI acquisition by the BS happens when the UE reports the PMIs and their respective weights. The overall process starts with the BS transmitting Channel State Information Reference Signal (CSI-RS) by means of a beam sweeping procedure. Each UE collects the received signals and estimates the amplitude scaling and the phase rotation of the beams. At each UE, there is a beam grouping selection to report only the best beams and weights to the BS.

We show in the following how to estimate the weights and the theoretical estimation quality.

Let us consider a vector $\mathbf{s} \in \mathbb{C}^{L\times 1}$ that defines a CSI-RS with length L. The vector transmission is repeated over U TTIs, and every time slot is associated to a given beam direction. Since we consider a low mobility scenario, the correlation observed in the channel vectors between the first and U-th TTI is sufficiently high to afford the suitability of this procedure. The received signal $\mathbf{r}_{j,b} \in \mathbb{C}^{L\times 1}$ at the j-th UE corresponds to the b-th beam during the beam sweeping, and its signal model can be written as

$$\mathbf{r}_{j,b} = \sqrt{P_{j,b}}\mathbf{s}\mathbf{h}_j^T \mathbf{w}^*_{b\,\alpha_{j,b}} + \mathbf{z}_j, \tag{2.2}$$

where $P_{j,b}$ is the power associated to the CSI-RS of the j-th UE at the b-th beam, $\mathbf{h}_j \in \mathbb{C}^{U\times 1}$ is the channel vector of the j-th UE, $\mathbf{w}_b \in \mathbb{C}^{U\times 1}$ is the b-th column of the Discrete Fourier Transform (DFT)-based codebook $\mathbf{W} \in \mathbb{C}^{U\times U}$ and $\mathbf{z}_j \in \mathbb{C}^{L\times 1}$ is an AWGN vector distributed as $\mathscr{CN}(\mathbf{0}, \sigma^2\mathbf{1})$. The projection of the channel vector $\mathbf{h}_j$ onto the column $\mathbf{w}_b$ of the codebook $\mathbf{W}$ is represented as the scalar $\alpha_{j,b}$.

The UE wants to estimate $\alpha_{j,b}, \forall b = \{1,2,...,U\}$, to properly decide the group of beams that will represent the CSI. The more beams are chosen, the more precise is the CSI representation, but the overhead increases. Therefore, we need to find a good operational point where the set of reported beams offers enough resolution (and consequently the BS can separate the UEs) yet the overhead has minimal impact on the spectral efficiency.

The estimated channel vector projection $\hat{\alpha}_{j,b}$ of the j-th UE onto the b-th beam $\mathbf{w}_b$ given the codebook and the knowledge of CSI-RS can be written as

$$\hat{\alpha}_{j,b} = \frac{1}{\sqrt{P_{j,b}}\|\mathbf{s}\|^2}\mathbf{s}^H\mathbf{r}_{j,b}. \tag{2.3}$$

The performance of the estimator can be derived as follows

$$\begin{aligned}\mathbb{E}\{|\alpha_{j,b} - \hat{\alpha}_{j,b}|^2\} &= \left(\frac{1}{\sqrt{P_{j,b}}\|\mathbf{s}\|^2}\right)^2 \mathbb{E}\{\|\mathbf{s}^H\mathbf{z}_j\|^2\} = \left(\frac{1}{\sqrt{P_{j,b}}\|\mathbf{s}\|^2}\right)^2 \mathbb{E}\{\mathbf{s}^H\mathbf{z}_j\mathbf{z}_j^H\mathbf{s}\} \\ &= \left(\frac{1}{\sqrt{P_{j,b}}\|\mathbf{s}\|^2}\right)^2 \mathbf{s}^H\mathbb{E}\{\mathbf{z}_j\mathbf{z}_j^H\}\mathbf{s} = \left(\frac{1}{\sqrt{P_{j,b}}\|\mathbf{s}\|^2}\right)^2 \sigma_j^2\mathbf{s}^H\mathbf{I}\mathbf{s} = \frac{\sigma_j^2}{P_{j,b}L}.\end{aligned} \tag{2.4}$$

Eq. (2.4) shows that the performance will increase as the SNR, $P_{j,b}/\sigma_j^2$, of the beam increases. Moreover, the length L of the CSI-RS also impacts on the estimation: the longer it is, the more accurate is the estimation.

The performance of the beam sweeping channel measurement is evaluated in terms of the Normalized Mean Square Error (NMSE) of the j-th UE, which is defined as

$$\text{NMSE}_j = \frac{1}{\mathfrak{C}} \sum_{\mathfrak{c}=1}^{\mathfrak{C}} \frac{\| \alpha - \hat{\alpha}_{j,\mathfrak{c}} \|_2^2}{\| \alpha \|_2^2}, \tag{2.5}$$

where $\mathfrak{C}$ is the number of Monte Carlo runs, $\alpha = [\alpha_{j,1} \cdots \alpha_{j,U}]^T \in \mathbb{C}^{U \times 1}$ is the set of projections of the channel vector $\mathbf{h}_j$ onto the codebook $\mathbf{W}$, and $\hat{\alpha}_{j,\mathfrak{c}} = [\hat{\alpha}_{j,1} \cdots \hat{\alpha}_{j,U}]^T \in \mathbb{C}^{U \times 1}$ is the set of estimated projections of the channel vector onto the codebook $\mathbf{W}$ for a given Monte Carlo run.

The j-th UE measures the projection of its channel $\mathbf{h}_j$ on the beam patterns configured by the DFT-based codebook $\mathbf{W}$ and, among those beam patterns, recommends a suitable set for subsequent data transmission. In the proposed scheme, each UE determines a set of beams $\mathbf{b} = [b_1 \cdots b_N]^T$ with the $N \leq U$ highest weights, since this is the set of beams that provides more information about the estimated instantaneous CSI of the UE. The N selected beams are fed back to the BS as a vector of PMIs together with the vector $\mathbf{a} = [\hat{\alpha}_{j,b_1} \cdots \hat{\alpha}_{j,b_N}]^T$ corresponding to the set of N highest weights associated to the beams' indices.

The BS uses the reported beams and their corresponding weights to estimate the channel by means of a linear combination. Therefore, the estimated channel $\hat{\mathbf{h}}_k \in \mathbb{C}^{U \times 1}$ at the BS is defined as

$$\hat{\mathbf{h}}_j = \sum_{n=1}^{N} \hat{\alpha}_{j,b_n} \mathbf{w}_{\mathbf{b}_n}, \tag{2.6}$$

where $\mathbf{w}_{b_n}$ refers to the b_n-th column of the codebook $\mathbf{W} = \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_U \end{bmatrix}$.

The conventional CSI feedback employed in [71, 79, 72] considers the report of the channel covariance matrix $\mathbf{R}_j$. Given the j-th UE, its channel vector on the t-th TTI is $\mathbf{h}_{t,j} \in \mathbb{C}^{U \times 1}$. Thus, the transmit covariance matrix $\mathbf{R}_j \in \mathbb{C}^{U \times U}$ can be approximated and expressed as

$$\mathbf{R}_j = \frac{1}{\mathfrak{T}} \sum_{t=1}^{\mathfrak{T}} \mathbf{h}_{t,j} \mathbf{h}_{t,j}^H, \tag{2.7}$$

where $\mathfrak{T}$ is the TTI window size and indicates the number of channel vector $\mathbf{h}_{t,j}$ samples considered in the averaging process. Note that the report of the covariance matrix would consume many communication resources. That is, the amount of information associated to the CSI feedback is determined by the number of antenna elements U at the BS and by the number of UEs J in the system. Therefore, the total amount of resources needed for reporting the whole $\mathbf{R}_j$ matrices is proportional to $U^2 J$. If we consider a massive MIMO BS and a high number of UEs, this feedback scheme implies a huge signaling overhead. Consequently, this would potentially decrease system spectral efficiency.

A practical communication system has limited resources and must prioritize carrying data instead of signaling. To provide a CSI feedback with reduced signaling overhead, we exploit the PMI representation of a codebook and their weights instead of explicitly informing the whole covariance matrix. In this case, the signaling amount is determined by the number N of reported PMIs (rather than the number of antennas elements) and by the number of J UEs. Therefore, the proposed CSI feedback scheme provides a signaling load proportional to $2NJ$. Since $N \ll U$, the proposed approach provides a CSI signaling load lower than the conventional second order channel statistic feedback. However, it can be less robust as the device does not estimate the full channel or second order channel statistics. In the following, we evaluate the impact of the proposed CSI report with limited signaling on the hybrid beamforming design.

2.4 Proposed Hybrid Beamforming Design

In the following, we detail the proposed hybrid beamforming design. The JSDM strategy admits the analog precoders dependent only on the second-order channel statistics. In our study, we consider the analog precoder based on the channel measurements as described by Eq. 2.3.

2.4.1 Analog Precoder Design

The pre-beamforming is considered as a generalization of the sectorization, since we partition J UEs into K clusters indicated as $\mathscr{C}_i, \forall i = \{1, \cdots, K\}$ with approximately the same channel characteristics. The central characteristic of the i-th cluster is indicated by its centroid $\mathbf{c}_i$.

The analog precoder has constant amplitude elements, so only phase shifting is performed in the analog domain. Therefore, we only consider the phases of the elements of the centroids. Considering this assumption, the analog precoder $\mathbf{F}_{\text{RF}} \in \mathbb{C}^{U \times K}$ is defined from the centroids of the clusters and can be written as

$$\mathbf{F}_{\text{RF}} = [\tilde{\mathbf{c}}_1 \quad \tilde{\mathbf{c}}_2 \quad \dots \quad \tilde{\mathbf{c}}_K], \tag{2.8}$$

where $\tilde{\mathbf{c}}_i = [\tilde{c}_{i,1} \quad \cdots \quad \tilde{c}_{i,U}]^T \in \mathbb{C}^{U \times 1}$, with each element defined as $\tilde{\mathbf{c}}_{i,u} = \exp(c_{i,u})$

The original approach of JSDM proposed in [72] employs the k-means algorithm to partition the UE population based on channel covariance matrix of all UEs available at the BS, which represents a huge signaling overhead. To overcome this issue, we evaluate a partitioning process with limited CSI feedback. Besides, we also consider a more refined version of the

k-means, called k-means++, and compare it to an algorithm called Fuzzy C-Means (FCM), which considers a different paradigm of clustering based on fuzzy logic. To the best of our knowledge, FCM has not been applied previously with the JSDM framework. The strategy adopted in the partitioning of UEs into clusters determines the characteristics of the centroids, and, consequently, the analog precoder. The behavior of the clustering algorithms and how the centroids are calculated are detailed in the following.

2.4.1.1 Clustering Based on K-means++

The k-means++ method aims to partition UEs into clusters so that each UE belongs to the cluster with the nearest central characteristic. Thereby, UEs with similar transmit channel characteristics are put on the same cluster. The k-means++ replaces the poor initialization step of k-means with a more sophisticated one [80].

Conventionally, the average channel covariance eigenspace is the channel characteristic considered in the UE partitioning strategies. The eigendecomposition of the transmit covariance matrix $\mathbf{R}_j$ can be written as

$$\mathbf{R}_j = \mathbf{E}_j \Delta_j \mathbf{E}_j^{-1}, \tag{2.9}$$

where $\mathbf{E}_j \in \mathbb{C}^{U \times U}$ defines the matrix composed of eigenvectors and $\Delta_j \in \mathbb{C}^{U \times U}$ is the diagonal matrix of eigenvalues. The dominant eigenvector $\tilde{\mathbf{h}}_j \in \mathbb{C}^{U \times 1}$ associated to highest eigenvalue of $\mathbf{R}_j$ is used as a baseline CSI feedback scheme.

Herein, the property that characterizes each UE in the partitioning process performed by the clustering algorithm is the estimated channel $\hat{\mathbf{h}}_j$, defined according to Eq. (2.6). So instead of measuring the covariance matrix, the BS uses the vectors of weights and PMI to represent the UE's CSI. Therefore, we consider that the information used by the BS to partition UEs into clusters is the linear combination of the projection of the instantaneous channel vector on the codebook $\mathbf{W}$.

The vector $\hat{\mathbf{h}}_j$, defined at Eq. (2.6), is associated to each UE and has the same length in comparison to the conventional approach. However, the signaling overhead required to achieve $\hat{\mathbf{h}}_j$ is significantly smaller than reporting the dominant eigenvector.

The first step of the k-means++ algorithm is the initialization of centroids. The conventional approach of the k-means algorithm determines the initial centroids from the characteristic $\hat{\mathbf{h}}_j$ from K UEs randomly sorted out of the entire set of J UEs in the system. This

process is simple, but may cause convergence of the centroids to the local optima. To overcome this issue, the k-means++ considers a more refined initialization of centroids.

In the beginning, i.e., at iteration $t = 0$, an arbitrary first centroid is randomly chosen from the entire set of UEs in the system. Therefore, the first centroid $\mathbf{c}_1^{(t)}$ is defined as the estimated channel $\hat{\mathbf{h}}_j$ (also called characteristic vector) of the sorted UE j. The other $(K-1)$ centroids are determined as the estimated channel of the UEs which have the highest weighted probability function Υ_j defined as

$$\Upsilon_j = \frac{\| \hat{\mathbf{h}}_j - \mathbf{c}_i^{(t)} \|_2^2}{\sum\limits_{j \neq j^\star} \| \hat{\mathbf{h}}_j - \mathbf{c}_i^{(t)} \|_2^2}. \tag{2.10}$$

where i is the index of the closest centroid and $j^\star$ is the first selected UE of the cluster i'.

Then, in the initialization process we select UEs which are furthest away from each other. In the following, the algorithm proceeds as the standard k-means algorithm [80]. Each iteration of the algorithm consists in a clustering assignment followed by a centroid update. In the assignment step, we calculate the distance among the estimated channel of the entire set of UEs and the centroids of clusters $\mathscr{C}_i, \forall i \in \{1, \cdots, K\}$. The UE closest to the centroid of the i-th cluster is determined according to

$$j^\star = \min_j \| \hat{\mathbf{h}}_j - \mathbf{c}_i^{(t)} \|_2^2 . \tag{2.11}$$

The set of components of the i-th cluster is updated, $\mathscr{C}_i^{(t)} = \mathscr{C}_i^{(t-1)} \cup \{j^\star\}$. In the following, the centroid $\mathbf{c}_i$ is updated as the average of the estimated channels of all UEs that belong to the i-th cluster, which can be written as

$$\mathbf{c}_i^{(t)} = \frac{1}{\left|\mathscr{C}_i^{(t)}\right|} \sum_{j \in \mathscr{C}_i^{(t)}} \hat{\mathbf{h}}_j. \tag{2.12}$$

We define a threshold $\varepsilon > 0$ and test at every iteration t if there is no significant change of the centroids in comparison to the previous iteration, i.e.,

$$\sum_{i=1}^{K} \| \mathbf{c}_i^{(t)} - \mathbf{c}_i^{(t-1)} \|_2^2 \leq \varepsilon. \tag{2.13}$$

In the centroid update step, new centroids $\mathbf{c}_i$ are computed for each cluster from the subset of UEs in $\mathscr{C}_i$ using Eq. (2.12). The assignment and centroid update steps are carried out until we reach a convergence. The output of the algorithm is a clustering of the UEs into K disjoint clusters and a set of vectors $\{\mathbf{c}_1, \cdots, \mathbf{c}_K\}$ obtained as the centroids of the clusters.

In the following, we summarize the main procedures of the k-meanss++ clustering in Algorithm 1. Based on that description, the computational complexity of this algorithm is given by $O(KJUT)$. For more details on this analysis, see Appendix A.1.

Algoritmo 1: K-means++ Clustering.

$t \leftarrow 0$;
initialize the first centroid ;
for $i \in \{2,\ldots,K\}$ **do**
    calculate the weighted probability function Υ_j as defined in (2.10) for all UEs in $\{1,\ldots,J\}$ excluding the UEs already chosen;
    associate the centroid $\mathbf{c}_i^{(t)}$ to the estimated channel $\hat{\mathbf{h}}_j$ of the UE j the with highest value of Υ_j;
end
while *condition* (2.13) *holds or the maximum number of iterations is achieved* **do**
    $t \leftarrow t+1$;
    for $j \in \{1,\ldots,J\}$ **do**
        for $i \in \{1,\ldots,K\}$ **do**
            assign j to the closest cluster $\mathscr{C}_i^{(t)}$ according to (2.11) ;
        end
    end
    for $i \in \{1,\ldots,K\}$ **do**
        update centroid $\mathbf{c}_i^{(t)}$ according to (2.12);
    end
end

2.4.1.2 Clustering Based on Fuzzy C-Means

The k-means++ algorithm considers that each UE only belongs to one cluster, i.e., the frontiers among clusters are crisp, clearly defined. However, several factors may complicate the creation of disjoint virtual sectors, such as spatially correlated multi-path and user agglomerations within geographically overlapping areas. To overcome such issues, we investigate a more flexible partitioning strategy, called FCM [81]. This clustering algorithm considers that each UE j in a given iteration t belongs to a cluster $\mathscr{C}_i^{(t)}$ with a given degree $\mu_{ij}^{(t)}$ that characterizes its membership grade, i.e., a given UE can belong to several clusters with a degree of pertinence specified by membership grades between 0 (does not belong to the cluster $\mathscr{C}_i^{(t)}$) and 1 (completely belongs to the cluster $\mathscr{C}_i^{(t)}$). These values are organized in the membership matrix $\mathbf{U}^{(t)} \in \mathbb{R}^{K\times J}$

so that the sum of degrees of pertinence for each UE to all K clusters yields to 1:

$$\sum_{i=1}^{K} \mu_{ij}^{(t)} = 1, \tag{2.14}$$

where $\mu_{ij}^{(t)}$ is the element at the i-th row and j-th column of the matrix $\mathbf{U}^{(t)}$.

The FCM algorithm divides a set of J UEs into K subsets, and determines a cluster center in each group such that a function of dissimilarity measure, i.e., a quadratic error distortion, is minimized [81]. Therefore, it considers the degree of membership as a weighting factor of the Euclidean distance between the UE characteristic vector $\hat{\mathbf{h}}_j$ and the cluster centroid $\mathbf{c}_i^{(t)}$. In a given iteration t, the objective function can be written as

$$v^{(t)} = \left(\mu_{ij}^{(t)}\right)^f \parallel \hat{\mathbf{h}}_j - \mathbf{c}_i^{(t)} \parallel_2^2, \tag{2.15}$$

where $f \geq 1$ is the exponent of fuzziness.

In the first step, the membership matrix $\mathbf{U}^{(t)}$ is initialized with random values between 0 and 1 such that the constraints established in Eq. (2.14) are satisfied. In the following, the algorithm iteratively determines the cluster centroids $\mathbf{c}_i^{(t)}$ and updates the value of the objective function. The centroids are calculated according to

$$\mathbf{c}_i^{(t)} = \frac{\sum_{j=1}^{J} \left(\mu_{ij}^{(t)}\right)^f \hat{\mathbf{h}}_j}{\sum_{j=1}^{J} \left(\mu_{ij}^{(t)}\right)^f}. \tag{2.16}$$

If the objective function is above a tolerance value, $|v^{(t)} - v^{(t-1)}| > \varepsilon'$, the membership matrix $\mathbf{U}^{(t)}$ is updated at each iteration t according to

$$\mu_{ij}^{(t)} = \left[\sum_{k=1}^{K} \left(\frac{\parallel \hat{\mathbf{h}}_j - \mathbf{c}_i^{(t)} \parallel_2}{\parallel \hat{\mathbf{h}}_j - \mathbf{c}_k^{(t)} \parallel_2} \right)^{\frac{2}{f-1}} \right]^{-1}. \tag{2.17}$$

In our proposal, each UE j is assigned to the cluster $\mathscr{C}_i$ with the highest membership degree. Therefore, the number of clusters effectively created in FCM is K', which can be equal to or smaller than $K = R$. In order to simplify our notation, from that point on we will simply denote the number of clusters by K. However, when refering to FCM, one should consider that K is in fact K'.

In the following, we summarize the main procedures of the FCM clustering in Algorithm 2. Based on that description, the computational complexity of this algorithm is given by $O(K^2JUT)$ For more details on this analysis, see Appendix A.2.

Algoritmo 2: FCM Clustering.

$t \leftarrow 0$;
initialize the membership matrix $\mathbf{U}^{(t)}$ with random values in the range $[0,1]$;
while $|\nu^{(t)} - \nu^{(t-1)}| \leq \varepsilon'$ *or the maximum number of iterations is achieved* **do**
    for $i \in \{1,\ldots,K\}$ **do**
        calculate centroid $\mathbf{c}_i^{(t)}$ according to (2.16);
    end
    for $j \in \{1,\ldots,J\}$ **do**
        for $i \in \{1,\ldots,K\}$ **do**
            compute the objective function $\nu^{(t)}$ according to (2.15);
        end
    end
    $t \leftarrow t+1$;
    for $i \in \{1,\ldots,K\}$ **do**
        for $j \in \{1,\ldots,J\}$ **do**
            update the membership matrix $\mathbf{U}^{(t)}$ as defined in (2.17);
        end
    end
    recompute the objective function $\nu^{(t)}$ according to (2.15);
end

2.4.2 Digital Precoder Design

In the following, given the clustering of UEs, we select one UE j from each cluster $\mathscr{C}_i$ that together compose an SDMA group $\mathscr{G}$ containing K UEs. In our study, we consider that the BS randomly selects one UE from each cluster to compose the SDMA group $\mathscr{G}$, ignoring any metric associated to channel characteristics. The random choice is the simplest approach, however disregards any spatial channel compatibility.

We define the group channel matrix $\mathbf{H}_{\mathscr{G}} \in \mathbb{C}^{K\times U}$ based on the channel vectors of the scheduled UEs that compose the SDMA group. It can be written as

$$\mathbf{H}_{\mathscr{G}} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \ldots & \mathbf{h}_K \end{bmatrix}. \tag{2.18}$$

Therefore, given the group channel matrix $\mathbf{H}_{\mathscr{G}}$ and the analog precoder $\mathbf{F}_{\mathrm{RF}}$, the equivalent channel matrix $\mathbf{H}_{\mathrm{eq}} \in \mathbb{C}^{K\times K}$ is given by

$$\mathbf{H}_{\mathrm{eq}} = \mathbf{H}_{\mathscr{G}}\mathbf{F}_{\mathrm{RF}}. \tag{2.19}$$

The digital precoding matrix $\mathbf{F}_{\mathrm{BB}} \in \mathbb{C}^{K\times K}$ is defined as the Zero Forcing (ZF). The ZF precoding is conceived to decorrelate the transmit signals so that the signal at every receiver

output is free of interference. The precoding matrix $\mathbf{F}_{\mathrm{BB}}$ is defined as

$$\mathbf{F}_{\mathrm{BB}} = \frac{\mathbf{H}_{\mathrm{eq}}^{H}(\mathbf{H}_{\mathrm{eq}}\mathbf{H}_{\mathrm{eq}}^{H})^{-1}}{\| \mathbf{H}_{\mathrm{eq}}^{H}(\mathbf{H}_{\mathrm{eq}}\mathbf{H}_{\mathrm{eq}}^{H})^{-1} \|_{F}}. \tag{2.20}$$

The total power constraint is enforced by normalizing the digital and analog filters, such that $\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\sqrt{\mathbf{P}_{\mathscr{G}}}\|_{F}^{2} = P_{\mathrm{RB}}$, where $\mathbf{P}_{\mathscr{G}} \in \mathbb{R}^{K\times K}$ is the diagonal power matrix resulting of the combination of the power matrices of each UE belonging to the SDMA group and P_{RB} is the power available for a given RB. We consider that the number of clusters is equal to the number of RF chains and the total number of streams, i.e., K. Therefore, the dimensions of $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$ are compatible with the dimension of $\mathbf{F}$, so that $\mathbf{F} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}} \in \mathbb{C}^{U\times K}$.

The receive vector of the group $\hat{\mathbf{y}}_{\mathscr{G}} \in \mathbb{C}^{K\times 1}$ is given by

$$\hat{\mathbf{y}}_{\mathscr{G}} = \mathbf{H}_{\mathscr{G}}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\sqrt{\mathbf{P}}_{\mathscr{G}}\,\mathbf{x}_{\mathscr{G}} + \mathbf{z}_{\mathscr{G}}, \tag{2.21}$$

where $\mathbf{x}_{\mathscr{G}} \in \mathbb{C}^{K\times 1}$ is the group symbol vector and the $\mathbf{z}_{\mathscr{G}} \in \mathbb{C}^{K\times 1}$ is the group noise vector.

Defining $\mathbf{Q} = \mathbf{H}_{\mathscr{G}}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\sqrt{\mathbf{P}_{\mathscr{G}}} \in \mathbb{C}^{K\times K}$, the average Signal to Interference-plus-Noise Ratio (SINR) Γ_i perceived by the i-th stream can be calculated as

$$\Gamma_i = \frac{|\mathbf{Q}_{ii}|^{2}}{\sum\limits_{j\neq i}^{K} |\mathbf{Q}_{ij}|^{2} + \sigma_z^{2}}, \tag{2.22}$$

where σ_z^2 is the noise power and $\mathbf{Q}_{ij}$ is the element of $\mathbf{Q}$ at i-th row and j-th column.

The data rate of the i-th stream is calculated according to Shannon's capacity formula [82] and is given by

$$\psi_i = \beta \log_2(1+\Gamma_i), \tag{2.23}$$

where β is the bandwidth of the RB.

2.5 Performance Evaluation

In this section, we evaluate the proposed hybrid beamforming design based on JSDM with limited CSI feedback. For our simulations, we consider a single cell system with a carrier frequency of 28 GHz and a bandwidth of 100 MHz. Based on the 5G numerology proposed in [77], these parameters imply a total set of 125 RBs, each one composed of 12 subcarriers equally spaced of 60 kHz. In our simulations, from the entire set of RBs, we consider a set of 25 active RBs. Furthermore, the number of subframes per frame is 10, each subframe has 14

symbols, and the TTI duration is 0.25 ms. We adopt the 3D Quasi Deterministic Radio Channel Generator (QuaDRiGa) as our channel model [83, 84] (version 2.0.0-664) and references therein. This channel model extends the Wireless World Initiative for New Radio (WINNER) channel model, including new features to make it as realistic as possible. QuaDRiGa supports 3-D propagation, 3-D antenna patterns, time evolving channel traces of arbitrary length, scenario transitions, and variable terminal speeds. General parameters of the simulations are listed in Table 1.

Tabela 1 – General Simulation Parameters

Parameter	**Value**
Min. dist. BS-UE (2D)	20 m
Angle sector	60°
BS height	10 m
UE height	1.5 m
UE track	linear
UE speed	3 km/h
BS antenna model	3GPP-mmWave [75]
BS vertical antennas	8
BS horizontal antennas	8
UE antenna model	omnidirectional
UE antennas	1
Total transmit power	35 dBm
Noise figure	9 dB
Noise spectral density	−174 dBm/Hz
Simulation time	16.25s
Simulation rounds	50

Fonte: Created by the author.

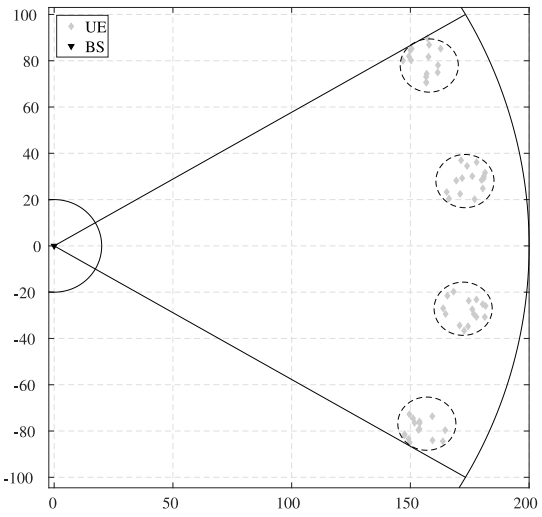
QuaDRiGa is a generic channel model, which uses the same method for generating channel coefficients in different environments. That is, it uses the same method for generating channel coefficients in different environments. Each environment is described by several of individual parameters, such as number of clusters, path loss exponents, shadow fading standard deviation and decorrelation distance, root-mean-square delay spread, angular spread, cross-polarization ratio. Reference [84] provides a detailed description of all these parameters. The main channel model parameters of the Urban Micro (UMi)-Line of Sight (LOS) and UMi-Non Line of Sight (NLOS) environments are provided in [75].

We consider a population of $J = 60$ UEs distributed in a single cell according to

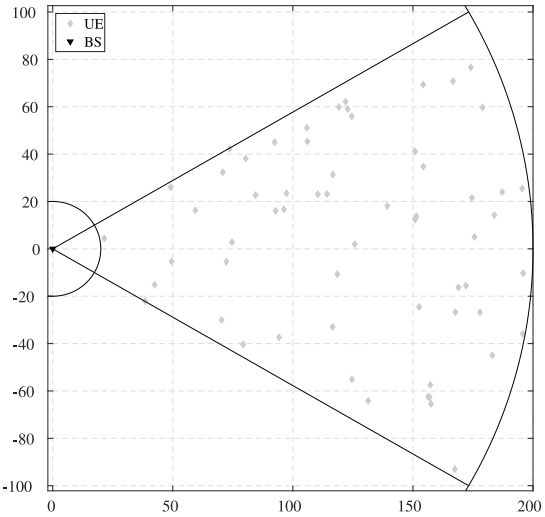
two different possible organizations, as indicated in Figure 3. In the first one, called hotspot disposition, the UEs are distributed into 4 hotspots with the same amount of UEs per hotspot. As it can be seen in Figure 3a, the hotspots are uniformly distributed in a 60° sector and are distant 175 meters from the BS. The UEs are randomly distributed inside each hotspot, which are modeled as circles of radius equal to 25 meters. The second one, called random disposition, considers a random distribution of UEs inside a 60° sector obeying the minimum distance to the BS specified in Table 1. In this disposition, as it can be seen in Figure 3b, the UEs are casually distributed, making partitioning of UEs into clusters much more complex in comparison to the hotspot disposition.

Figura 3 – Schemes of distribution of UEs considered in the evaluation of the clustering algorithms.

(a) Hotspot Disposition.



(b) Random Disposition.

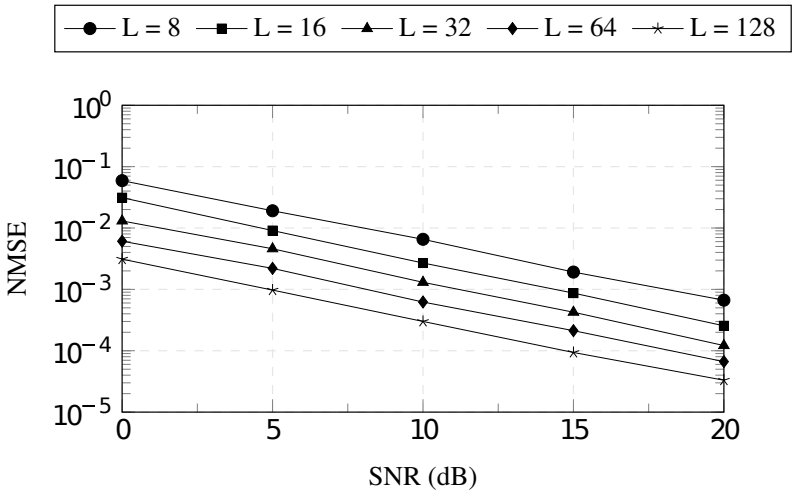


Fonte: Created by the author.

Figure 4 shows the performance of the beam sweeping channel measurement in terms of NMSE for different values of SNR and pilot sequence lengths $L = \{8, 16, 32, 64, 128\}$ in the UMi-LOS scenario. The increase of the pilot sequence length provides the UE more knowledge about the channel characteristics. Consequently, there is a better estimation of channel vector projections onto the codebook entries. We also observe that the performance increases as the SNR of the beam increases. The performance of the estimator is in agreement with the established at Eq. (2.4).

The first scenario of evaluation considers the UMi-LOS channel and the hotspot disposition of UEs indicated by Fig. 3a. The BS uses the k-means++ algorithm to partition

Figura 4 – NMSE of the estimated channel vector for different pilot sequence lengths.

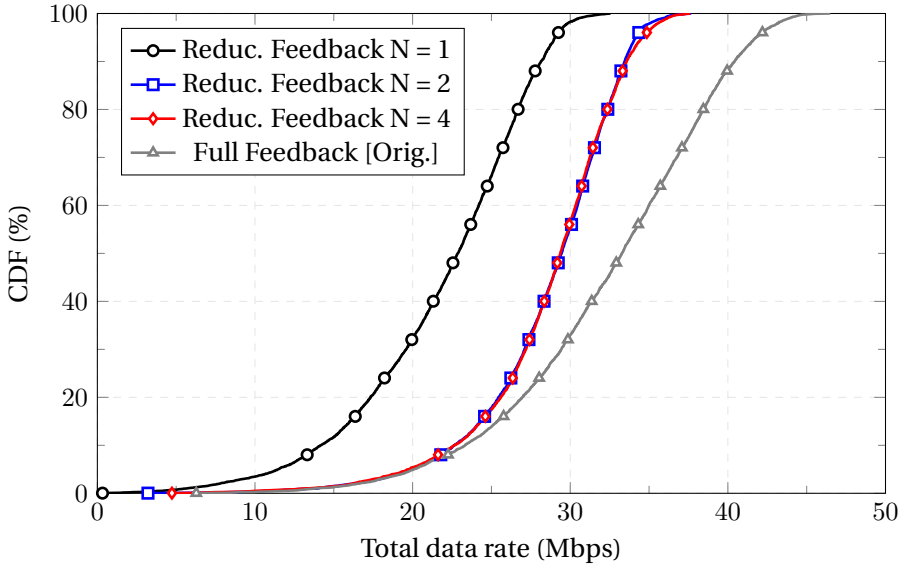


Fonte: Created by the author.

the UEs into a different number of clusters, namely $K = \{2,4,6,8,10\}$. The effect of the limited feedback is evaluated by varying the number of PMIs and weights reported to the BS in $N = \{1,2,4,8\}$. Then, after the determination of the elements of each cluster $\mathscr{C}_i$, the BS randomly schedules one UE from each cluster to compose the SDMA group, i.e., the set of K UEs that will share the same time-frequency resource. In the following, the digital precoder $\mathbf{F}_{\mathrm{BB}}$ is determined according to Eq. (2.20).

Figure 5 shows the Cumulative Distribution Function (CDF) of the total data rate of the proposed solution of reduced feedback (Prop) compared with the conventional signaling

Figura 5 – Evaluation of proposed signaling scheme with different number of reported PMIs in comparison with the original report of CSI based on the covariance matrix.



Fonte: Created by the author.

scheme employed with JSDM based on the covariance matrix [71, 79, 72] . In this initial evaluation, we partition UEs into $K = 4$ clusters using the k-means++ algorithm. We observe an increase in total data rate when the number of PMI changes from $N = 1$ to $N = 2$. Therefore, there is an enhancement from 22.5 Mbps to 29 Mbps at the 50th percentile of the total data rate, which represents a gain of almost 30%. The clustering algorithm divides more effectively UEs according to their channel compatibilities, i.e., the algorithm forms cluster with UEs that have similar channel characteristics. Consequently, the scheduling selects UEs with better estimated channel conditions. However, we observe only a marginal total data rate growth when the number of PMIs is increased from $N = 2$ to $N = 4$ PMIs. Hence, the CSI provided by $N = 4$ PMIs does not provide more meaningful information in comparison with the knowledge supplied by the report of $N = 2$ PMIs in this scenario, since it is considered a UMi-LOS channel, characterized by low angular spread and low rank.

Therefore, the evaluated scheme with limited feedback of CSI reduces significantly the signaling overhead and achieves levels of the total data rate near to the conventional signaling approach. The CSI feedback scheme achieves 90% of total data rate observed in the conventional feedback scheme at the 50th percentile. The conventional JSDM feedback reports the entire channel covariance matrix of each UE $\mathbf{R}_j \in \mathbb{C}^{64\times 64}$, a signaling amount almost a thousand times higher than the reported CSI using the vectors of PMIs and estimated weights ($\mathbf{b}, \mathbf{a} \in \mathbb{C}^{2\times 1}$).

In the following, we expand the evaluation of the behavior of the hybrid beamforming scheme in scenarios with different levels of complexity to partition UEs. That is, we further investigate the impact of different channel conditions and UEs distributions in space.

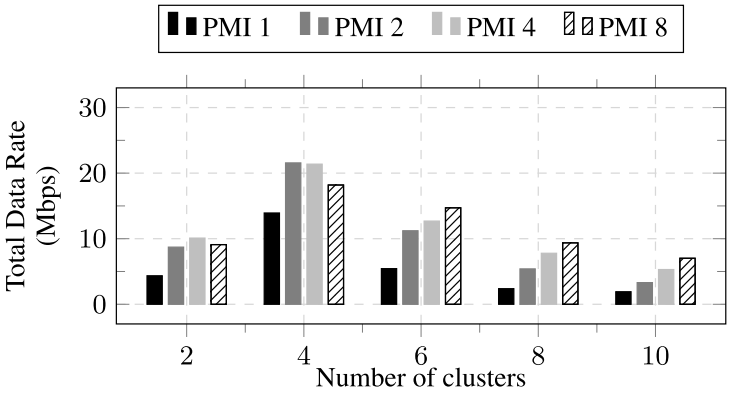
In Figure 6, we show the 10th, 50th and 90th percentiles of the total data rate of the JSDM scheme considering the clustering procedure based on the k-means++ algorithm described in Section 2.4.1.1. The behavior of the total data rate varies according to the number K of clusters considered in the partitioning process and the number N of PMIs. The total data rate increases from $K = 2$ to $K = 4$ when the highest value is achieved. From $K = 6$ there is reduction of the total data rate, i.e., the higher the value of K, the more significant is the reduction. As the number of clusters becomes greater than the number of hotspots artificially established in the disposal of UEs (depicted at Fig. 3a), the process of partitioning becomes more challenging and the interference between the clusters increases. Thus, there is an overlap among the centroids and a poorer conditioning of the equivalent channel $\mathbf{H}_{\text{eq}}$.

Figure 6 also depicts the impact of the CSI feedback on the system performance, i.e.,

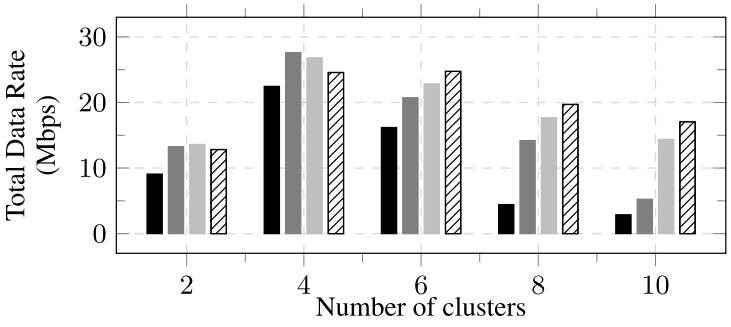
the behavior of the total data rate as a function of the number of PMIs. The CSI feedback with only 1 PMI has the lowest total data rates in all conditions (i.e., for all values of K), since the amount of information provided is insufficient to ensure a proper separation of UEs. Thus, the BS requires more information to perform an efficient clustering. As it can be seen for 10th, 50th and 90th percentiles, there is an improvement of the total data rate when the BS CSI feedback increases from 1 to 2 PMIs. The lowest relative gain between 1 and 2 PMIs occurs when the

Figura 6 – Total data rate in the UMi-LOS channel and hotspot distribution of UEs considering k-means++ as the clustering algorithm.

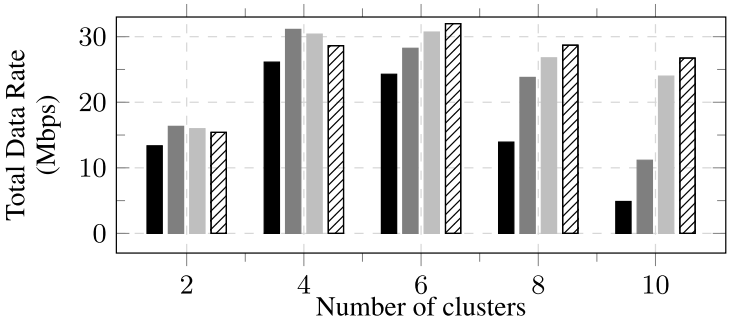
(a) 10th percentile.



(b) 50th percentile.



(c) 90th percentile.



Fonte: Created by the author.

partitioning of UEs is done into 4 clusters, despite the highest absolute value. This happens because the scenario is biased towards partitioning into 4 hotspots. As partitioning becomes more challenging, the information added with that second PMI makes partitioning more efficient. Hence the higher relative rate gains are observed for the highest values of K.

However, we cannot assume that the increase in the number of PMIs will always improve the total data rate or that it will happen in the same way to all number of clusters. As it can be seen, the increase of signaling to 4 or 8 PMIs may result in different variations of the average total data rate, which depends on the number of clusters and the number of original hotspots. There is a decrease of the total data rate with a higher signaling when the number of clusters is smaller than or equal to the number of hotspost ($K \leq 4$) because the additional information provided with 4 or 8 PMIs does not improve the clustering process. The increase in the amount of PMIs implies a greater amount of beams being combined. It is observed an increment of the frequency of selection of certain beams, which reduces the differences between the centroids of the clusters. Consequently, it increases the interference among clusters, which is not compensated by the additional information provided. On the other hand, we observe that the increase in the number of PMIs to 4 or 8 improves the total data rate when the number of clusters is higher than the number of hotspots ($K > 4$). In this case, a greater number of PMIs is required in order to make a more adequate partitioning of users since the number of clusters does not correspond to the most suitable number of subsets of UEs, i.e., the number of hotspots. Therefore, it is a more challenging task and requires more information.

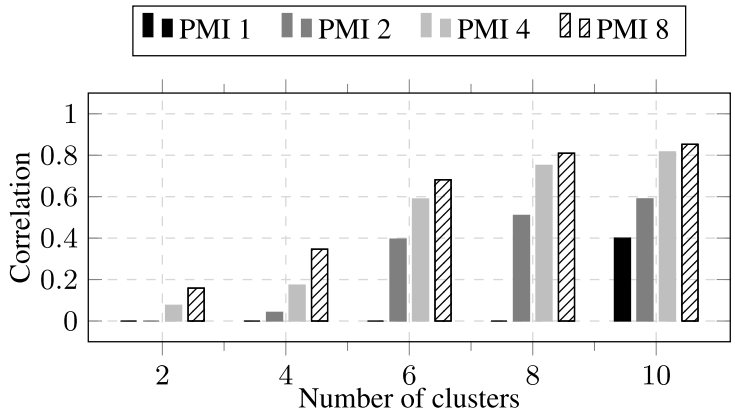
Figure 7 indicates the average of the maximum correlation among the centroids, calculated using the expression $|\mathbf{c}_i^T \mathbf{c}_{i'}|/(||\mathbf{c}_i||\,||\mathbf{c}_{i'}||), \forall i, i' = \{1, \cdots, K\}$, considering different number of clusters $K = \{2,4,6,8,10\}$ and different number of PMIs $N = \{1,2,4,8\}$. The increase in the number of PMIs implies that a larger number of linearly combined columns of the codebook is used to describe the instantaneous channel. Consequently, the beams become wider, increasing the superposition among the centroids and consequently the interference among the UEs that compose the SDMA group. Therefore, the increment of the interference reduces the SINR and, consequently, the total data rate. We observe that the highest total data rate is achieved when the number of clusters is equal to the number of hotspots, i.e., $K = 4$. As we move away from this value and consider $K > 4$, there is an increase in the correlation among the centroids, and, consequently, interference arises among UEs sharing time-frequency resources. Figure 7 indicates an increase in the correlation among the centroids when we try to partition

the population into more clusters, since the possibility of intersecting beams, even narrow ones, increases with the number of clusters. Therefore, the combination of this interference and the insufficient amount of information compromises the clustering at the BS and results in extremely reduced total data rates compared to other CSI report configurations.

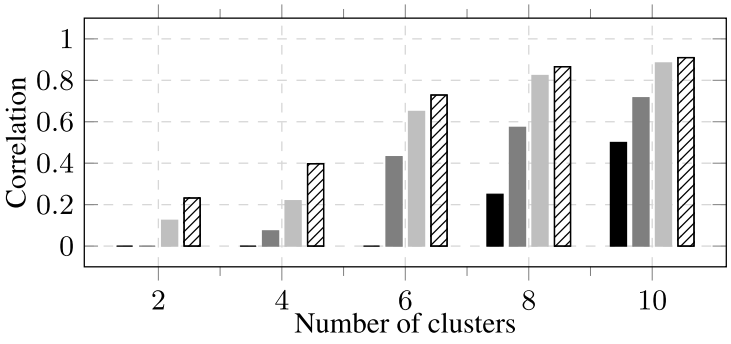
The second scenario of evaluation still considers the hotspot disposition of UEs. However, we now analyze the k-means++ clustering algorithm with limited CSI considering the

Figura 7 – Correlation in the UMi-LOS channel and hotspot distribution of UEs considering k-means++ as the clustering algorithm.

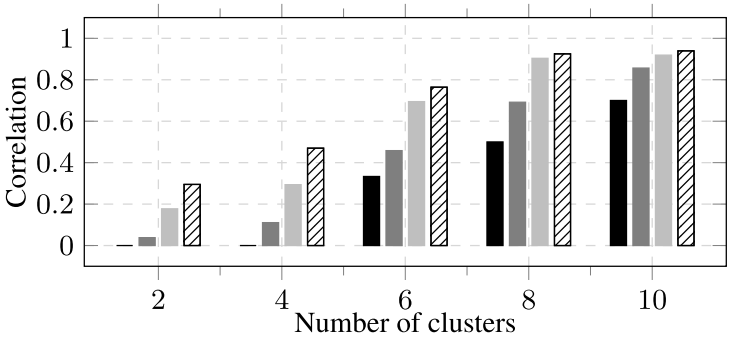
(a) 10th percentile.



(b) 50th percentile.

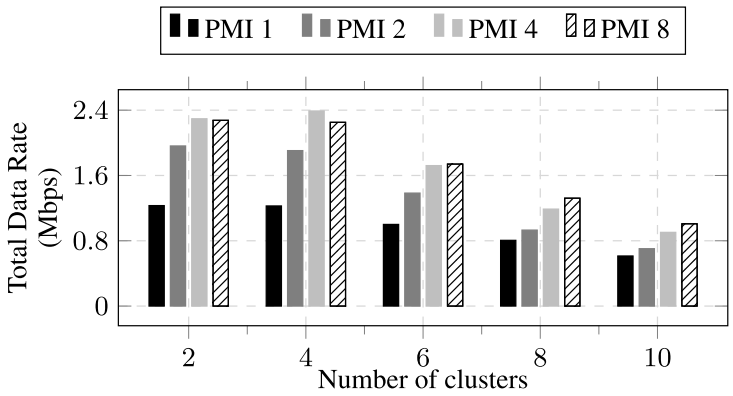


(c) 90th percentile.



Fonte: Created by the author.

Figura 8 – 50th percentile of the total data rate in UMi-NLOS channel and hotspot distribution of UEs considering k-means++ as the clustering algorithm.



Fonte: Created by the author.

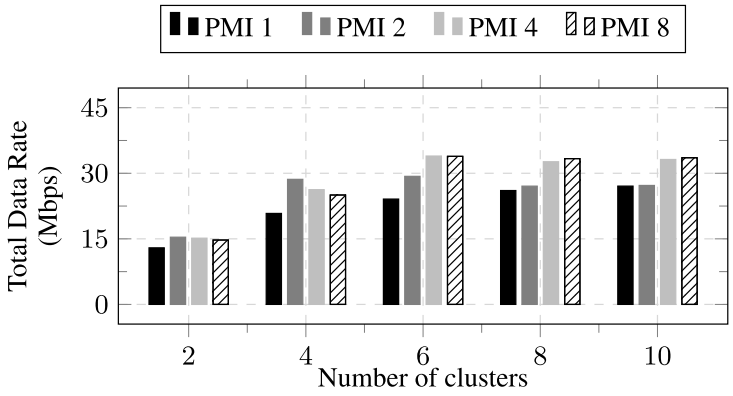
UMi-NLOS channel. This condition turns the partioning of UEs more difficult, since we can observe spatially correlated multi-paths. For simplicity sake, from now on we only analyze the 50th percentile because the trends observed in the other percentiles are maintained. Figure 8 shows the 50th percentile of the total data rate considering different number of clusters, namely $K = \{2,4,6,8,10\}$, and different amounts of PMIs and weights reported to the BS, namely $N = \{1,2,4,8\}$. The NLOS environment considers the effect of more clusters of scatterers and higher angular spread as indicated in [75]. The average total data rate is smaller compared to the LOS scenario and more PMIs are required to achieve suitable total data rates. We had already mentioned the inability to obtain adequate information with only 1 PMI. That is, the lack of information and less channel directivity burdens the clustering algorithm, consequently the total data rate has lower levels. If we consider the feedback of 2 PMIs, the average total data rate increases, as observed previously in the UMi-LOS scenario. In the UMi-LOS the enhancement is of the order of 90%, while in the UMi-NLOS it is observed an average improvement of 35%. This reduction in the 50th percentile of the total data rate is a consequence of the more challenging channel conditions. In the UMi-LOS scenario, we observe a limitation on the improvement of the total data rate with the increase of the amount of PMIs, i.e., considering more information feedback does not turns the partitioning process more efficient. In UMi-NLOS scenario, the increase in the signaling to 4 PMIs represents an improvement in the 50th percentile of the total data rate. It is also important to note that the gain limitation with increased signaling for 8 PMIs is still observed. Therefore, the value of PMIs to provide useful information to the BS perform clustering is directly influenced by channel conditions, i.e., harsh scattering environments need more information.

Previously, we considered a set of 60 UEs organized into 4 hotspots. We evaluate the k-means++ clustering algorithm with a variable number of clusters, namely $K = \{2,4,6,8,10\}$, and the maximum value of total data rate is achieved when $K = 4$. Therefore, the information provided allows the creation of clusters when the UE population is organized into spatially uniform circles. To further evaluate the information supplied by the proposed channel measurement based on beam sweep to the BS, the third scenario considers the UE randomly distributed inside an angular sector of 60°. Thus, the number of clusters is not known in advance.

Figure 9 depicts the 50th percentile of the total data rate considering different number of clusters, namely $K = \{2,4,6,8,10\}$, and different amounts of PMIs and weights reported to the BS, namely $N = \{1,2,4,8\}$, in a UMi-LOS scattering environment. The total data increases as the number of clusters rises until $K = 6$. Above this value, the total data rate stabilizes. Therefore, the BS receives enough information to partition the UEs properly with a limited interference until this clustering requirement. The increase in the number of clusters in which we divide the UEs makes the process more difficult so that the total data rate holds, but the rate per UE in the SDMA is reduced. Hence, as the value of K increases more information is required to partition UEs properly, which explains the highest total data rates achieved with higher amounts of PMIs. For $K = 2$, the 50th percentile of the total rate for the different PMI values are practically the same. Therefore, the increase in the signaling does not improve the system performance. Partitioning of UEs in this scenario can be done satisfactorily with only 1 PMI. At the other extreme, when we try to partition the UEs into $K = 10$ clusters, there is a 20% gain in the total rate as a result of a better knowledge of the UEs' channel conditions.

In the previous scenarios, we included different conditions to make the partitioning

Figura 9 – 50th percentile of the total data rate in the UMi-LOS channel and random distribution of UEs considering k-means++ as the clustering algorithm.



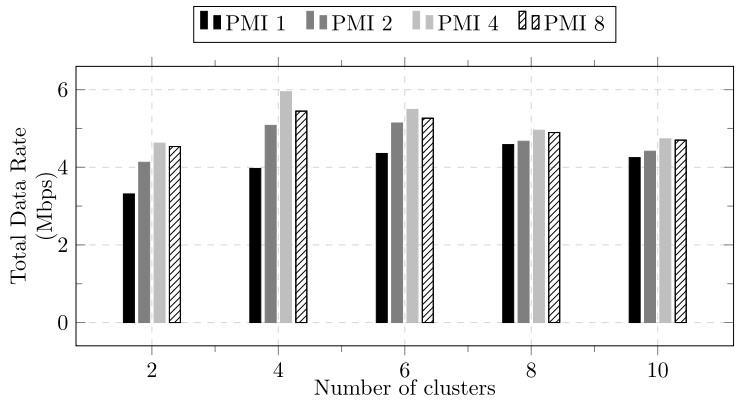
Fonte: Created by the author.

of UEs into clusters harder. We added each of these adversities and evaluate their impact on the k-means++ clustering algorithm. In the following, we consider these conditions simultaneously, i.e., UMi-NLOS channel and a random distribution of UEs in the sector of the cell and consider a more robust clustering algorithm. In addition to the k-means++, we evaluate the FCM with an exponent of fuzziness $f = 2$, that considers a paradigm of clustering based on the fuzzy logic.

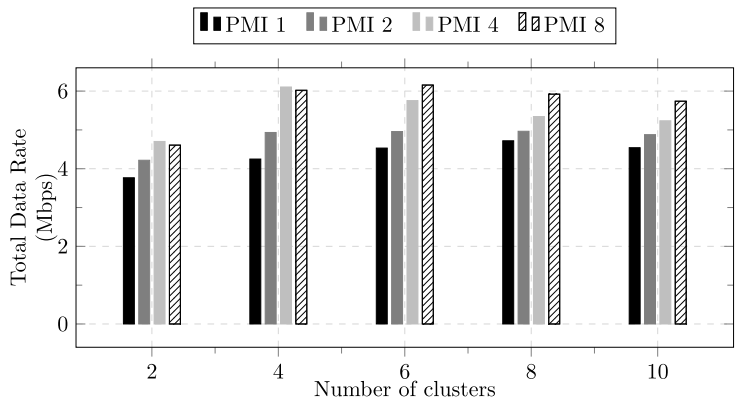
Figure 10 compares the k-means++ and FCM clustering algorithms in terms of the 50th percentile of the total data rate considering different number of clusters, namely $K = \{2,4,6,8,10\}$, and different amounts of PMIs and weights reported to the BS, namely $N = \{1,2,4,8\}$. As it can be seen in Figure 10a, the k-means++ algorithm shows a behavior very similar to the described previously, i.e., the need of information increases as the difficulty to partition UEs increases. However, the total data rate reduces as the number of clusters increases, even if more PMIs are considered. That results indicates a limitation of the k-means++ algorithm to create mutually orthogonal clusters in a channel characterized by spatially correlated

Figura 10 – Comparison for the 50th percentile of the total data rate in the UMi-NLOS scenario for different clustering algorithms considering the UEs randomly distributed.

(a) k-means++ clustering algorithm.



(b) FCM clustering algorithm.



Fonte: Created by the author.

multi-paths.

The FCM algorithm employs a more flexible partitioning strategy so that in the most challenging condition, when $K = 10$, there is an increment of 30% of the total data rate in comparison to k-means++, as it can be seen in Figure 10b. The FCM algorithm keeps the total rate at stable levels, while k-means++ presents a performance degradation with the increase in the number of clusters. Moreover, it exploits the information feedback more efficiently – the best rate-signaling trade-offs are achieved with 4 PMIs. Therefore, the FCM algorithm allows achieving the highest performance in terms of clustering results, especially when the clusters are not well separated due to the random distribution of UE and they are overlapped as a consequence of the scattering environment.

2.6 Chapter Remarks

In this study, we investigated a limited CSI feedback scheme to allow the implementation of a hybrid beamforming scheme based on an UL framework. We use a beam sweeping channel measurement collect the best PMIs and report them to BS. The hybrid beamforming design based on JSDM considered for comparison requires the report of the covariance matrix of each UE to the BS, so that the total amount of resources needed for reporting the whole channel covariance matrix is proportional to U^2J, where U is the number of antennas and J is the number of UEs in the system. The proposed CSI feedback scheme, on its turn, requires a signaling load proportional only to $2NJ$, where N is the number of reported PMIs. Since $N \ll U$, it is observed a meaningful signaling reduction, as well as an adequate representation of the channel statistics for the hybrid beamforming design in massive MIMO systems when compared to the conventional report of channel covariance matrix.

Moreover, we also evaluate different user selection strategies based on an UL framework that exploits the CSI provided by the evaluated beam sweeping scheme. Our results indicate that the user selection based on fuzzy c-means is able to efficiently utilize the reduced CSI. The proposed hybrid beamforming scheme reduces the multi-user interference and achieves significant gains in total data rate as the channel conditions and interference environment become more challenging. The performance evaluation indicates that the proposed user clustering based on fuzzy c-means can efficiently explore the reduced CSI. It achieves an enhancement of 25% in the total data rate compared with k-means++ on the most challenging conditions.

3 UPLINK POWER CONTROL DESIGN BASED ON REINFORCEMENT LEARNING

In this chapter, it is proposed an Uplink Power Control (UPC) framework compliant with the technical specifications of the Fifth Generation (5G) wireless networks. We apply the fundamentals of Reinforcement Learning (RL) to develop a power control algorithm able to learn a strategy that mitigates the inter-cell interference and, consequently, enhances the total data rate on the uplink channel. The Base Station (BS) uses a set of commands to specify by how much the User Equipment (UE) transmit power should change. After implementing such commands, the UE reports a set of information to its serving BS, and this, in turn, predicts the next commands to achieve a suitable UE transmit power level. The BS converts the UE reports into rewards according to a predefined reward function, which impacts the long-term behavior of the UE transmit power. The simulation results indicate a near-optimum performance of the proposed framework in terms of total transmit power, total data rate, and network energy efficiency. It provides a self-exploratory power control strategy that overcomes Soft Dropping Power Control (SDPC) with similar signaling levels.

The remaining of this chapter is organized as follows. Section 3.1 discusses related works and the main contributions of this chapter. Section 3.2 describes the main system model assumptions and Section 3.3 describes the 5G New Radio (NR) specifications to deploy UPC. Section 3.4 reviews the fundamental concepts of RL. Then, Section 3.5 presents the proposed framework for UPC. Section 3.6 describes the comparison algorithms. Simulation results are discussed in Section 3.7. Finally, the main remarks of this chapter are drawn in Section 3.8.

3.1 Literature Review and Contributions

The application of the RL into UPC problems constitute a self-organized solution capable of finding autonomously suitable transmit power levels. Consequently, it can reduce inter-cell interference and increase system total data rate properly. It is a suitable technique and has been relevant in numerous studies in the field. In [85], the authors proposed a power control framework to manage the interference in a cognitive radio network. They modeled the wireless network as a multi-agent system, where the agents interact directly with the environment and learn a strategy, also called policy, to manage the power levels. In their model, the BSs represent the decision-maker entities which manage the radio resources allocated to their associated UEs. However, they focused on the downlink operation and presented only preliminary results without

considerations about Third Generation Partnership Project (3GPP) specifications. We highligth that the RL-based UPC solution proposed here is compatible with 5G NR specifications, such as required signaling, power control commands, and hardware constraints.

In [86], the authors considered a more realistic cognitive radio network, modeled as a Wireless Regional Area Network (WRAN) compliant with the 802.22 [87] standards. Therein, they considered the downlink and uplink operation and situations of complete and incomplete information about the environment. The results indicated that a multi-agent RL system could automatically learn a policy to successfully manage the interference, without introducing signaling overhead into the system.

In [88], a decentralized UPC based on the multi-agent RL combined with a Fractional Power Control (FPC) mechanism was proposed for Long Term Evolution (LTE)-based multi-tier networks. In their study, each UE decides the transmit power based on the channel conditions, namely, the uplink path loss. In this framework, each UE learns independently the transmit power without the need to wait for control signaling from an associated BS. It is shown that the solution reduces the signaling in uplink transmission. However, due to the limited computational resources of the UEs, it also reduces the processing capability of the decision-maker entity. In our approach, the entity endowed with intelligence is the BS, which has enough computational capacity to control several power control mechanisms working in parallel and in real-time.

The authors of [89] also investigated a learning-based power control based on an FPC mechanism in LTE systems. They presented a data-driven framework to model the interference patterns in Orthogonal Frequency Division Multiple Access (OFDMA)-based networks. Based on the measurement of these interference patterns, the proposed learning algorithms defined an optimal setting of the cell-specific power control parameters. Therein, the authors assumed that all path loss variables must be interpreted in a time-scale sense so that it averages the effect of fast-fading. In other words, the UPC mechanism proposed in [89] can compensate for path loss and large-scale variations such as shadowing, but does not adequately handles fast-fading. In scenarios where these effects are prominent, this simplification may render an inappropriate representation of the channel conditions, restricting the success of that UPC solution.

The works mentioned previously ([88, 89]) are based on the FPC mechanism. They employed the Open Loop Power Control (OLPC) paradigm, i.e., they defined the transmit power according to large-scale channel conditions, namely, path loss measurements. The conventional FPC solution identifies UEs based only on the path loss, which is not entirely proper in dense

deployment scenarios. Such approaches do not consider interference conditions while allocating power to the UEs, usually resulting in high interference situations [90].

To overcome these issues, in our proposal we perform transmit power adjustment according to the Closed-Loop Power Control (CLPC) paradigm. That is, we consider the impact in the system of the power commands taken previously. Moreover, we establish a cooperation among BSs to improve the mitigation of the inter-cell interference. This is achieved by means of a multi-agent RL approach. In [91], the authors proposed a UPC framework in which the UEs intelligently learn the best action, i.e., the selection of the transmit power from a pre-specified power set. In their model, the decision-maker performs a CLPC mechanism that perceives as a useful policy the actions that improve the Signal to Interference-plus-Noise Ratio (SINR) levels above a given threshold. The authors used a Deep Reinforcement Learning (DRL) method, called Deep Q-Network (DQN) [92]. Differently from conventional Q-learning, which uses a lookup table to store knowledge, the DQN employs a deep neural network to represent this information. The lookup table approach is shown to be more computationally efficient than the neural network. On the other hand, the neural network has reduced memory requirements compared to the lookup table approach [86]. In [93], the authors also designed a distributed UPC framework based on a DQN. The proposed strategy learns a policy that guides transmitters to adjust their power levels according to the CLPC paradigm under practical constraints, such as delayed information exchange and incomplete Channel State Information (CSI).

The DRL-based techniques have achieved remarkable attention in the last years, as it can be seen in [94, 95, 96]. They can provide suitable learning strategies in complex and broad-scale networks, where RL may not be able to discover an optimal strategy in a reasonable time. In this model, the neural network is trained frequently based on distinct experiences obtained during the interactions with the environment, which incurs in high computational complexity.

However, the system model assumptions of our study result in a UPC problem where the advantages of DRL do not become good enough to outweigh the critical disadvantage of its use. Hence, our RL-based solution presents lower signaling and lower computational complexity than DRL-based techniques. Moreover, RL-based techniques do not require training, being able to provide real-time learning. Therefore, we turn our attention to a multi-agent RL-based solution combined with the CLPC paradigm that is compliant with the 5G NR technical specifications [97, 98, 99, 100, 101, 102, 103].

The main difference of the NR UPC framework compared to its predecessor LTE is

on the use of multiple control loops associated with beams. Specifically, each control loop may be associated with a specific pair of transmit and receive beams. For instance, one electronic device may have a beam associated with two Multiple Input Multiple Output (MIMO) layers (or even more) so that the device can manage multiple control loops at the same time. Without proper coordination among the loops, the multiple processes produce sub-optimal power solutions [17].

To the best of our knowledge, despite the relevance of the beam-centric power control to the management of interference in NR networks, we have not found in the literature studies investigating this problem considering practical implementation aspects. Therefore, it is a relevant research topic the development of flexible power control strategies to take into account the coordination among multiple beams and the technical specifications from 3GPP. We summarize the main contributions of this chapter as follows:

1. development of a UPC framework compliant with the technical specifications from 3GPP release 15;
2. formulation of a beam-based transmit power control based on the principles of Machine Learning (ML), specifically on RL paradigm;
3. development of a signaling scheme to allow cooperation among the entities endowed with intelligence in a NR multi-cell system;
4. comparison of the proposed UPC frameworks with two classical algorithms, namely the optimum solution power control (OSPC) and the soft dropping power control (SDPC), in terms of total transmit power, total data rate, and network energy efficiency.

3.2 System Model

3.2.1 Overall Scenario and Channel Model

We consider a multi-cell wireless network consisting of C cells. Each cell has one BS equipped with V antennas and serves J UEs equipped with U antennas each. We assume uplink transmission, and all cells share the same frequency band. The UEs inside a cell are synchronized with their corresponding BS and periodically measure their associated beam pairs. The scenario under consideration assumes that each UE is served by only one BS. Figure 11 represents the considered scenario.

The discrete received signal model at the BS of the c-th cell due to the j-th UE is

represented as

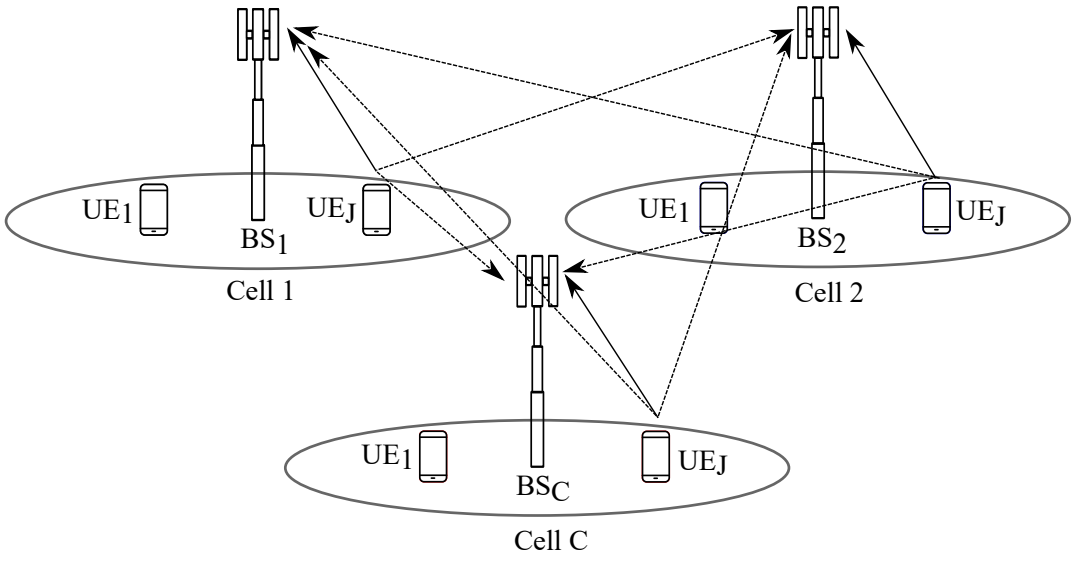
$$y_{j,c} = \underbrace{\bar{\mathbf{g}}_{j,c}^{H}\mathbf{H}_{j,c}\bar{\mathbf{f}}_{j,c}P_{j,c}x_{j,c}}_{\text{useful signal}} + \underbrace{\sum_{\substack{j'\neq j \\ j'=1}}^{J}\bar{\mathbf{g}}_{j,c}^{H}\mathbf{H}_{j,c}\bar{\mathbf{f}}_{j',c}P_{j',c}x_{j',c}}_{\text{intra-cell interference}}$$
$$+ \underbrace{\sum_{j'=1}^{J}\sum_{\substack{c'\neq c \\ c'=1}}^{C}\bar{\mathbf{g}}_{j,c}^{H}\mathbf{H}_{j,c}\bar{\mathbf{f}}_{j',c'}P_{j',c'}x_{j',c'}}_{\text{inter-cell interference}} + \underbrace{\bar{\mathbf{g}}_{j,c}^{H}\mathbf{z}}_{\text{filtered noise}}, \tag{3.1}$$

where $\bar{\mathbf{g}}_{j,c} \in \mathbb{C}^{V\times 1}$ is the receive beamforming vector, $\mathbf{H}_{j,c} \in \mathbb{C}^{V\times U}$ is the Millimeter Wave (mmWave) channel matrix, $\bar{\mathbf{f}}_{j,c} \in \mathbb{C}^{U\times 1}$ is the transmit beamforming vector, $P_{j,c}$ is the transmit power, $x_{j,c}$ is the transmitted symbol, and $\mathbf{z} \in \mathbb{C}^{V\times 1}$ is the Additive White Gaussian Noise (AWGN) vector with zero mean and variance σ^2.

We assume a narrow band block-fading MIMO channel, which is constant within a time-frequency resource block. The channel follows a geometric model with limited number Υ of scatterers [104]. Each scatter contributes with a single path between UE and BS. Therefore, the channel matrix $\mathbf{H}_{j,c} \in \mathbb{C}^{V\times U}$ between the j-th UE and the BS of c-th cell can be written as

$$\mathbf{H}_{j,c} = \sqrt{\phi_{j,c}}\sum_{l=1}^{\Upsilon}\mathscr{V}_l\mathbf{w}_{\text{BS}}(\vartheta_{l,j,c}^{UE},\theta_{l,j,c}^{UE})\mathbf{w}_{\text{UE}}{}^{H}(\vartheta_{l,j,c}^{BS},\theta_{l,j,c}^{BS}), \tag{3.2}$$

Figura 11 – Multi-cell wireless network. Solid lines indicate the useful signals and dashed lines indicate the interference signals.



Fonte: Created by the author.

where $\phi_{j,c}$ denotes the path loss between the j-th UE and the BS of the c-th cell and $\mathscr{V}_l$ is the complex gain of the l-th scatter. The azimuth $\vartheta^{UE}_{l,j,c}, \vartheta^{BS}_{l,j,c} \in [0, 2\pi]$ and the elevation $\theta^{UE}_{l,j,c}, \theta^{BS}_{l,j,c} \in [0, \pi]$ are the Angle of Departure (AoD) and Angle of Arrival (AoA) at the BS and UE, respectively.

We assume Uniform Rectangular Arrays (URAs) at BSs and UEs. There are V_v vertical antenna elements and V_h horizontal antennas elements at each BS, such that $V = V_v V_h$. Similarly, each UE has U_v vertical antenna elements and U_h horizontal antennas elements, such that $U = U_v U_h$ The array response at a BS and at a UE are expressed, respectively, as

$$\mathbf{w}_{\text{BS}}(\vartheta^{BS}_{l,j,c}, \theta^{BS}_{l,j,c}) = \frac{1}{\sqrt{V}}[1, \ldots, e^{\left((V_v-1)\frac{2\pi\Lambda}{\lambda}\cos\theta^{BS}_{l,j,c} + (V_h-1)\frac{2\pi\Lambda}{\lambda}\sin\vartheta^{BS}_{l,j,c}\sin\theta^{BS}_{l,j,c}\right)}] \tag{3.3}$$

$$\mathbf{w}_{\text{UE}}(\vartheta^{UE}_{l,j,c}, \theta^{UE}_{l,j,c}) = \frac{1}{\sqrt{U}}[1, \ldots, e^{\left((U_v-1)\frac{2\pi\Lambda}{\lambda}\cos\theta^{UE}_{l,j,c} + (U_h-1)\frac{2\pi\Lambda}{\lambda}\sin\vartheta^{UE}_{l,j,c}\sin\theta^{UE}_{l,j,c}\right)}] \tag{3.4}$$

where Λ is the antenna element spacing and λ is the signal wavelength.

The receive and transmit beamforming follow the so-called hybrid structure, which are defined as $\bar{\mathbf{g}}_{j,c} = \tilde{\mathbf{G}}\mathbf{u}$ and $\bar{\mathbf{f}}_{j,c} = \tilde{\mathbf{F}}\mathbf{v}$, respectively. We define $\tilde{\mathbf{G}} \in \mathbb{C}^{V\times B}$ and $\tilde{\mathbf{F}} \in \mathbb{C}^{U\times B}$ as the truncated receive and transmit beam codebooks representing a set of B selected beam pairs. These beams are chosen from the set of suitable beam pairs $\mathscr{B}$.

The receive and transmit codebooks are modeled according to the Discrete Fourier Transform (DFT) matrices $\mathbf{G} \in \mathbb{C}^{V\times V}$ and $\mathbf{F} \in \mathbb{C}^{U\times U}$, respectively. The beams are be associated with the column vectors of the receive and transmit codebooks, $\mathbf{g}_{b_{Rx}}$ and $\mathbf{f}_{b_{Tx}}$, respectively . They can be written as

$$\mathbf{g}_{b_{Rx}} = \frac{1}{\sqrt{V}}\left[1, \quad e^{-2\pi\frac{1}{V}}, \quad \cdots, \quad e^{-2\pi\frac{(b_{Rx}-1)}{V}}\right]^T, \tag{3.5}$$

$$\mathbf{f}_{b_{Tx}} = \frac{1}{\sqrt{U}}\left[1, \quad e^{-2\pi\frac{1}{U}}, \quad \cdots, \quad e^{-2\pi\frac{(b_{Tx}-1)}{U}}\right]^T, \tag{3.6}$$

where $\mathbf{g}_{b_{Rx}} \in \mathbb{C}^{V\times 1}$ is the b_{Rx}-th column of the receive beam codebook $\mathbf{G}$ and $\mathbf{f}_{b_{Tx}} \in \mathbb{C}^{U\times 1}$ is the b_{Tx}-th column of the transmit beam codebook $\mathbf{G}$.

The vectors $\mathbf{u} \in \mathbb{C}^{B\times 1}$ and $\mathbf{v} \in \mathbb{C}^{B\times 1}$ correspond to the dominant left and right singular vectors of the equivalent channel $\hat{\mathbf{H}}_{j,c} \in \mathbb{C}^{B\times B}$, defined as

$$\hat{\mathbf{H}}_{j,c} = \tilde{\mathbf{G}}^H \mathbf{H}_{j,c} \tilde{\mathbf{F}}. \tag{3.7}$$

Therefore, the singular value vectors $\mathbf{u}$ and $\mathbf{v}$ define the weights that combine the transmit and receive beams, respectively.

Each individual uplink transmission is carried out from a specific antenna port, the identity of which is known by the system [17]. Each antenna port has its own specific Reference Signal (RS), which is used by the device to estimate the CSI. Therefore, the d-th antenna port associated with the selected beam pairs B of the link between the j-th UE and the BS of the c-th cell can be written as

$$h_{d,j,c} = \mathbf{u}^H \hat{\mathbf{H}}_{j,c} \mathbf{v}. \tag{3.8}$$

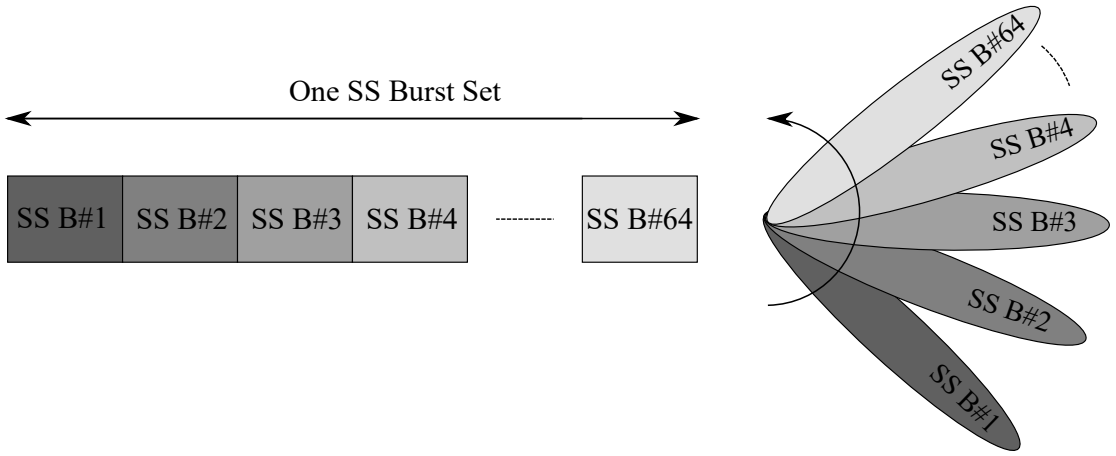
3.2.2 Beam Sweeping Scheme

We employ the beam sweeping scheme proposed in [105] to determine the set $\mathscr{B}$ of the most suitable transmit-receive beam pairs. The suitability of a beam pair is determined according to the connectivity provided by the transmitter and receiver beam directions and it does not necessarily correspond to transmitter and receiver beams that are physically pointing directly to each other [17].

The beam sweeping operation covers a spatial area with a set of beams according to pre-specified intervals and directions. It is carried out an exhaustive search in a set of directions (each one identified by a beamforming vector) that covers the whole angular space. The BS sequentially transmits Synchronization Signal (SS) blocks, that compose a SS burst set, and each SS block can be mapped to a certain angular direction, as it can be seen in Fig. 12.

The evaluation of the quality of the received signal at the UE is based on the Signal to Noise Ratio (SNR), i.e., the average of the received power on SS divided by the noise power.

Figura 12 – Model of multiple time-multiplexed SS blocks within an SS burst set period for a higher frequency band.



Fonte: Created by the author.

The selected beam pairs for t-th signaling period between the BS and the j-th UE of the c-th cell are defined according to the B highest values of SNR, which can be written as

$$\gamma_{t,b_{Rx},b_{Tx}} = \frac{\|\mathbf{g}_{b_{Rx}}^H \mathbf{H}_{j,c,q} \mathbf{f}_{b_{Tx}}\|}{\sigma^2}. \tag{3.9}$$

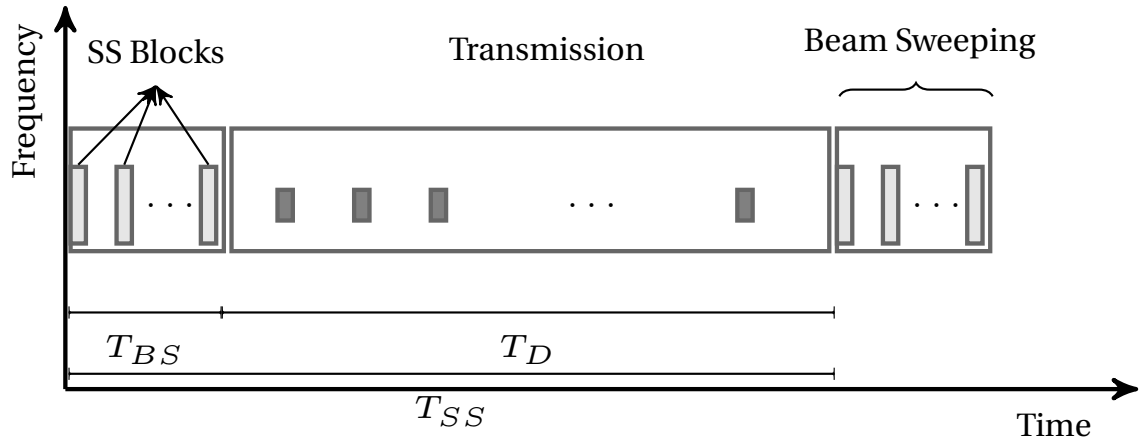
We consider a signaling period, of duration T_{SS}, divided into two time windows, as shown in Fig. 13. The first one contains a set of SS blocks with a duration T_{BS}. It performs a beam sweeping procedure, i.e, the search of the best set of transmit-receive beam pairs. The second window is dedicated to data transmission using the selected beam pairs. During this period, of duration T_D, each UE reports periodically the measured SNR to the BS. The BS measures all possible combinations of transmit and receive beams from the codebooks $\mathbf{F}$ and $\mathbf{G}$, respectively, during the transmission of the SS blocks. We assume that the AoDs and AoAs are constant over the beam sweeping period T_{BS}, which varies from 5 to 160 ms [17].

We define the SINR of the d-th antenna port described by Eq. (3.8) between the j-th UE and the BS of the c-th cell as

$$\Gamma_{d,j,c} = \frac{P_{d,j,c} |\bar{\mathbf{g}}_{d,j,c}^H \mathbf{H}_{j,c} \bar{\mathbf{f}}_{d,j,c}|^2}{\sum\limits_{j'=1}^{J} \sum\limits_{\substack{c'=1 \\ c' \neq c}}^{C} \sum\limits_{\substack{d'=1 \\ d' \neq d}}^{D} P_{d',j',c'} |\bar{\mathbf{g}}_{d,j,c}^H \mathbf{H}_{j,c} \bar{\mathbf{f}}_{d',j',c'}|^2 + \sigma^2}, \tag{3.10}$$

where the intra-cell interference mentioned in Eq. (3.1) is not considered since we assume a single active UE per time-frequency resource.

Figura 13 – Model of signaling period.



Fonte: Created by the author.

3.3 NR Specifications to Uplink Power Control

NR UPC is the set of procedures that manage the transmit power of uplink physical channels, namely Physical Uplink Shared Channel (PUSCH), Physical Uplink Control Channel (PUCCH), and Physical Random Access Channel (PRACH), to guarantee suitable communication [101]. We seek to determine the minimum transmit signal power necessary for appropriate decoding of the information conveyed through the physical channel [17]. Furthermore, the UPC procedures must also limit the interference to the other uplink transmissions.

Several factors, such as the channel attenuation, noise, and interference at the receiver side, influence the transmit power mechanisms. Moreover, modulation and coding schemes also affect the transmit power since the number of bits per sub carrier changes due to the link adaptation [17]. All these factors have distinct impacts on the power control behavior, that is subject to the characteristics of each uplink physical channel.

The power control behavior is designed to ensure that the transmit power does not exceed the power limit denoted as P_{CMAX} [97]. Then, the transmit power control of an uplink physical channel for a given carrier has a generic structure which can be written as

$$P = \min\{P_{\mathrm{CMAX}}, E(\cdot)\}, \tag{3.11}$$

$E(\cdot)$ is an expression that characterizes the behavior of the transmit power control. The parameters of $E(\cdot)$ are determined according to the characteristics of the uplink physical channel.

In addition to the maximum per-carrier transmit power, there is a limit on the maximum UE output power denoted here as P_{TMAX}. For a UE configured for NR transmission on multiple carriers, the P_{CMAX} of each carrier needs to be scaled down to ensure that the eventual transmit power of the UE over all configured uplink carriers does not exceed the maximum allowed value.

The transmit power control expressions of the uplink physical channels are thoroughly detailed in [101]. The PUSCH is used for the transmission of Uplink Shared Channel (UL-SCH) data and control information. Thus, compared to PRACH and PUCCH, it presents a relation between power control and link adaptation that allows more flexibility. Consequently, the power control of the PUSCH has a greater scope of mechanisms and encompasses what can be done in PUCCH and PRACH. Therefore, the UPC framework developed in this chapter considers the PUSCH expression (3.12) as the baseline power control. It can be concisely written,

in dBm scale, as

$$P = \min\{P_{\text{CMAX}}, P_0(\tau) + \xi(\tau)\Phi(\kappa) + 10\log_{10}(2^{\mu}\beta) + \Delta + \delta(\upsilon)\}, \tag{3.12}$$

where

- $P_0(\cdot)$ is the target received power;
- τ indicates the transmission type, a network configurable parameter which specifies the factors that compose the target received power;
- $0 < \xi(\cdot) \leq 1$ is the path loss compensation factor;
- $\Phi(\cdot)$ is the estimation of the uplink path loss;
- κ is a RS index, e.g., CSI-RS or SS block;
- μ is the subcarrier spacing parameter;
- β is the bandwidth of the resource assignment;
- Δ models the required received power according to number of resource bits per resource element according to the modulation scheme and channel coding rates;
- $\delta(\cdot)$ is the power adjustment due to the closed loop power control;
- υ determines the closed loop process index.

Eq. (3.12) indicates that the PUSCH power control is composed by two mechanisms, namely OLPC, which defines the support for fractional path loss compensation, where the device estimates the uplink path loss, and CLPC, which determines explicit power control commands provided by the network.

In the OLPC, the UE transmit power is adjusted according to estimates of the uplink path loss based on downlink measurements, as it can be seen by expression $P_0(\tau) + \xi(\tau)\Phi(\kappa)$. The transmit power is adjusted to achieve the target $P_0(\tau)$, which is a network configurable parameter regulated to provide a target data rate, given the noise and interference levels at the receiver. The UE estimates the downlink path loss $\Phi(\tau)$ using the RS index τ for the active downlink. A UE does not simultaneously maintain more than four path loss estimates per transmission [101].

The parameter $\xi(\tau)$ determines the compensation level of the path loss $\Phi(\kappa)$. The full path loss compensation, which is determined when we assume $\xi(\tau) = 1$, ensures that the received SINR matches the requirement for the Modulation and Coding Scheme (MCS) selected by the network, assuming that the UE transmit power does not reach its maximum value. The fractional path loss compensation, controlled by the parameter $0 < \xi(\tau) < 1$, requires a relatively lower transmit power, implying less interference to the other cells. However, the received power,

and consequently, the SINR, decreases as the path loss increases. The data rate is also reduced to compensate this effect by switching to a lower rate MCS. The benefit of fractional path loss compensation is reduced interference to neighbor cells. This comes at the price of more considerable variations in the service quality, with reduced data rate availability for UEs closer to the cell border. Consequently, in our study, we assume full path loss compensation to ensure service quality stability, requiring a CLPC mechanism to manage the uplink interference.

The transmit power should be proportional to the bandwidth assigned for the transmission, as indicated by the term $10\log_{10}(2^{\mu}\beta)$ in Eq. (3.12). The transmit power must be proportional to the size of the resource block with a 15 kHz numerology, where the subcarrier width is defined as $\Delta f = 2^{\mu} \cdot 15$ kHz. In our study, we assume a fixed bandwidth for the PUSCH transmission. Therefore, this term can be omitted from the power control expression.

The term Δ models the impact of the number of bits per resource element and channel-coding rates on the transmit power. This model is expressed as

$$\Delta = 10\log(2^{1.25\zeta} - 1) + 10\log(\rho), \tag{3.13}$$

where ζ is the number of information bits per resource element, ρ models the impact of data transmission on PUSCH, and $\rho = 1$ when the PUSCH includes UL-SCH data. In our study, we assume that the PUSCH received power is matched to a certain MCS given by the selected value of $P_0(\tau)$. In this case, according to [106], we must turn off the Δ function by setting its value to zero.

In the CLPC, the UE transmit power is adjusted according to power control commands provided by the BS. This regulation is determined based on prior network measurements of the received uplink power [17]. The term $\delta(\upsilon)$ is the power control command and represents the closed loop solution. Such commands are carried out in the Transmit Power Control (TPC) field within uplink scheduling grants (Downlink Control Information (DCI) formats $0-0$, $0-1$, and $2-2$ [101]). Each power control commands consists of 2 bits corresponding to four different steps: -1 dB, 0 dB, $+1$ dB, and $+3$ dB. These steps are associated with TPC command field values 0, 1, 2 and 3, respectively. Each command specifies the value in dB that a UE should add to its current transmit power.

The parameters κ and υ in Eq. (3.12) are directly associated with the selected beam pairs. The uplink path loss estimate $\Phi(\kappa)$ should reflect the path loss of the antenna port associated with a set κ of B uplink beam pairs to be used for the PUSCH transmission. If the uplink path loss estimation is not provided by higher layer parameters, the UE calculates

$\Phi(\kappa)$ using a RS index from the correspondent SS/Physical Broadcast Channel (PBCH) block. Therefore, the beam sweeping procedure also provides the necessary information to estimate uplink path loss. Each UE is limited to monitor up to $\kappa_{\max} = 4$ parallel path loss estimation processes [101].

The network also provides a mapping from the possible Sounding Resource Indicator (SRI) values provided in the scheduling grant to the different values of κ. After a beam management process to determine the best set of beam pairs, and consequently, the correspondent antenna ports, the path loss estimate is then used in the power control expression.

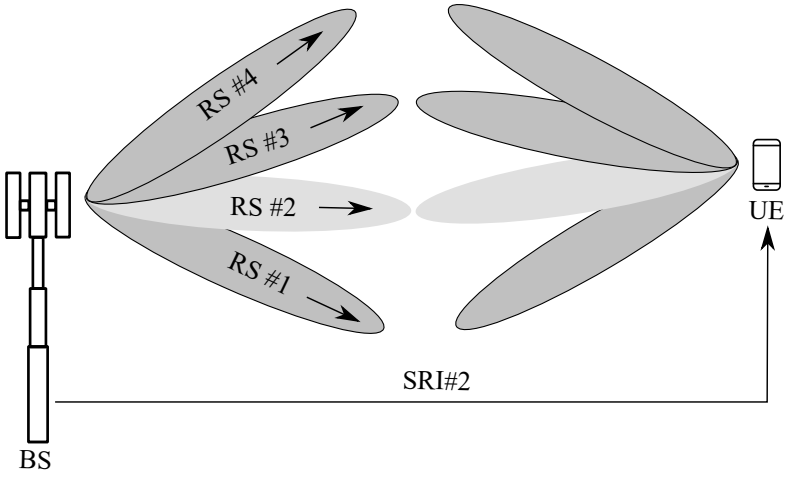
Figure 14 represents this procedure for a case with 4 parallel path loss estimation processes. In this example, the UE is scheduled for PUSCH transmission with the SRI of the scheduling grant set to the antenna port number 2. Hence, the transmit power of the scheduled PUSCH transmission is determined based on the path loss estimate based on the measurements on *RS*#2. Consequently, the parameters κ and υ are directly associated to the antenna port index d.

Therefore, based on the assumptions previously described, the transmit power of the d-th antenna port of the j-th UE in the c-th cell is represented as a simplification of the Eq. (3.12) and can be written as

$$P^{\text{PUSCH}}_{d,j,c} = \min\{P_{\text{CMAX}}, P_0(\tau) + \xi_{j,c}(\tau)\Phi_{j,c}(d) + \delta_{j,c}(d)\}. \tag{3.14}$$

In the following, we detail the main concepts of RL used in the development of the proposed UPC framework.

Figura 14 – Signaling scheme of the multiple power estimation process.



Fonte: Created by the author.

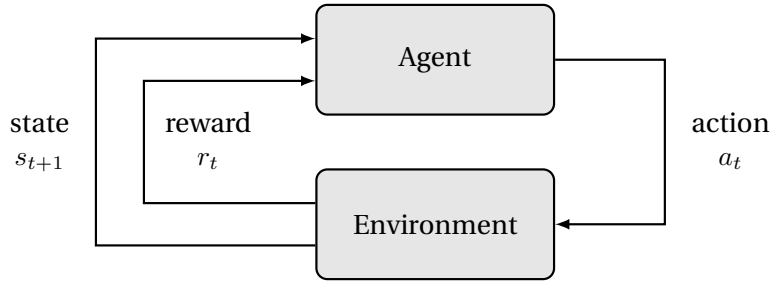
3.4 Fundamentals of Reinforcement Learning

The Markov Decision Process (MDP) model represents the theoretical basis of RL [61]. It provides a mathematical framework to describe scenarios where the decision-maker entity, hereafter referred to as an agent, does not have precise control of all outcomes, i.e., the environment behavior is partially deterministic and partially random [49]. An MDP model is a discrete-time stochastic optimal control problem described in terms of the tuple $\{\mathscr{S}, \mathscr{A}, \Pi, \mathfrak{R}\}$, where $\mathscr{S}$ is the finite set of environment states, $\mathscr{A}$ is the finite set of actions available to the agents, $\Pi : \mathscr{S} \times \mathscr{A} \times \mathscr{S} \rightarrow [0,1]$ is a state transition probability function, and $\mathfrak{R} : \mathscr{S} \times \mathscr{A} \times \mathscr{S} \rightarrow \mathbb{R}$ is a reward function.

Therefore, an MDP model describes the relationship between the agent and the environment using the concepts of state, action, transition function, and reward in a sequence of time steps [107]. The state $s_t \in \mathscr{S}$ in a discrete time step t is a value (or a set of values) that models the information that the agent has about the environment. The action $a_t \in \mathscr{A}$ is an adjustment parameter used by the agent to interact with the environment. The state transition function $\Pi(s_t, a_t, s_{t+1})$ defines the probability of transition from a state s_t to state s_{t+1} according to the action a_t taken by the agent. The reward r_t is a scalar function which indicates the immediate payoff from taking an action a_t in a state s_t [61].

The interaction between the agent and the environment can be modeled as the transition from state s_t to s_{t+1} restricted to the set of all possible states $\mathscr{S}$. The transition is a consequence of an action a_t chosen in a set of available actions $\mathscr{A}$ and associated with a reward r_{t+1}. The behavior of the agent is described by its policy, i.e., how the agent chooses its actions according to the state of the environment. Figure 15 represents the interaction among these elements.

Figura 15 – Interaction between RL elements.



Fonte: Created by the author.

The goal of MDP is the determination of the best policy, i.e., the most appropriate selection of actions according to the state s of the environment that maximizes its expected discounted reward. The optimal value associated with the discounted rewards achieved from an initial state s_0 can be written as

$$\mathscr{X}(s) = \max_{\pi} \left[\mathbb{E}\left(\sum_{t=1}^{\infty} \varphi^t r_t \right) \right], \tag{3.15}$$

where $\pi : \mathscr{S} \rightarrow \mathscr{A}$ represents the mapping between actions and states used to determine r_t; $\varphi \in [0,1)$ is the discount factor, a parameter that encodes an increasing uncertainty about the rewards. It bounds the sum which otherwise might grow indefinitely.

The Bellman's principle of optimality establishes that an optimal policy has the property that whatever the initial state s_0 and initial action a_0 are, the remaining actions must constitute an optimal policy with regard to the state resulting from the first action a_0 [108]. Therefore, the optimal value function is unique and can be described as

$$\mathscr{X}(s) = \max_{a} \left\{ \mathbb{E}[r(s,a,s')] + \varphi \sum_{s' \in \mathscr{S}} \Pi(s,a,s') \mathscr{X}(s') \right\}. \tag{3.16}$$

Hence, the optimal policy can be specified as the determination of the action a that maximizes the expected discounted sum from a given state s, which can be defined as

$$\pi^{\star}(s) = \arg\max_{a} \left\{ \mathbb{E}[r(s,a,s')] + \varphi \sum_{s' \in \mathscr{S}} \Pi(s,a,s') \mathscr{X}(s') \right\}. \tag{3.17}$$

In this work, we consider a model-free approach based on RL, since we cannot deduce the state transition probability function due to the dynamics of the wireless environment. The agent must find the desirable policy by taking into consideration the value of a state-action value function $Q : \mathscr{S} \times \mathscr{A} \rightarrow \mathbb{R}$. The function $Q(s_t, a_t)$ determines the overall expected discounted reward when starting in a state s_t and selecting an action a_t.

Among the several RL algorithms, we highlight the Q-learning algorithm, which is an off-policy temporal difference algorithm initially proposed in [109]. This algorithm works by updating an estimate of the state-action value function based on the iterations of the agent with the environment. The state-action values are updated according to

$$Q(s_{t+1}, a_{t+1}) = (1-\eta) Q(s_t, a_t) \; + \eta \left[r_{t+1} + \varphi \max_{a_{t+1} \in \mathscr{A}} Q(s_{t+1}, a_{t+1}) \right], \tag{3.18}$$

where $\eta \in [0,1]$ is the learning rate and $\varphi \in [0,1]$ is the discount factor, which trades off the instantaneous and future rewards.

The learning process occurs through the balance between exploration, i.e., the sample of unseen parts of the state-action space, and exploitation of the accumulated knowledge [61]. We consider an adaptive ε-greedy algorithm strategy. Every time an agent takes an action a_t, it has a probability ε_t to be random (exploration) and a probability $(1-\varepsilon_t)$ to select an action a_t based on previous experience (exploitation). The value of ε_t is gradually reduced over time from an initial value $\varepsilon_{\max}$ until it reaches a minimum value $\varepsilon_{\min}$. The agent has to store the state-action values to be able to learn from the interactions with the environment. In our study, we build a matrix $\mathbf{Q} \in \mathbb{R}^{|\mathscr{S}|\times|\mathscr{A}|}$ to store the $Q(s_t,a_t)$ values. This mechanism has low requirements of computational memory in discrete state and action sets, and do not require any complex computational operation of training to store the acquired knowledge.

The generalization of the MDP to the multi-agent case is called Stochastic Game (SG) [110]. In a system composed of I agents, the SG is described by the tuple $\{\mathscr{S}', \mathscr{A}', \Pi', \mathfrak{R}'\}$, where $\mathscr{S}' = \mathscr{S}_1 \times \cdots \times \mathscr{S}_I$ is the joint finite set of environment states; $\mathscr{A}' = \mathscr{A}_1 \times \cdots \times \mathscr{A}_I$ is the joint finite set of actions available to the agents; $\Pi' : \mathscr{S}' \times \mathscr{A}' \times \mathscr{S}' \to [0,1]$ is the joint state transition probability function; $\mathfrak{R}' : \mathscr{S}' \times \mathscr{A}' \times \mathscr{S}' \to \mathbb{R}$ is a reward function. In our study, we assume a model-free distributed Q-learning algorithm for multi-agent RL. We adopt the decentralized learning paradigm, i.e., each agent acts independently, without a centralized coordination among them. Each agent i maintains a local policy π_i and a local state-action value function $Q_i(s_t^i, a_t^i)$ based on the iteration of the agent depending only on its own action. We summarize the interactions between the i-th agent and the environment in Algorithm 3.

Algoritmo 3: Pseudo code of the multi-agent Q-learning algorithm.

```
let t = 0;
sort the initial state;
while stop condition not reached do
    sort an uniform random number o ∈ [0,1];
    determine ε_t;
    if (o < ε_t) then
        select randomly an action a_t^i ∈ A_i;
    else
        select an action a_t^i ∈ A_i which has the maximum Q_i(s_t^i, a_t^i);
    end
    execute the action a_t^i observe the new state s_{t+1}^i;
    calculate the associated reward r_t^i(s_t^i, a_t^i, s_{t+1}^i);
    update the matrix Q_i with the Q_i(s_t^i, a_t^i) value calculated according to Eq. (3.18);
    t = t + 1;
end
```

3.5 Proposed Uplink Power Control Framework

Our main contribution in comparison to the previous solutions mentioned in Section 3.1 is the joint power optimization of multiple antenna ports per UE using an RL-based technique since we consider multiple-antenna transmission/reception.

In our framework, each cell regards its BS as an agent, and the remaining of the system (other BSs, UEs, and UL-SCH) represents the environment. The behavior of each agent depends not only on the associated BS's actions, but also on the actions taken by neighboring BSs, since all cells use the same frequency and suffer from inter-cell interference. In other words, the states, actions, and rewards from different cells that are coupled, and influence each other.

The state of the BS of the c-th cell at the t-th iteration is a tuple of D powers associated with each antenna port $s_{t,c} = \{P_{1,c}, \cdots, P_{D,c}\}$. The index j is omitted for simplification purposes, since we consider only one UE per resource. These powers are modeled as discrete values in dBm and are limited to minimum and maximum values, i.e., $s_{t,c} \in S_c = \{\{P_{1,c}^{\min}, \cdots, P_{D,c}^{\min}\}, \cdots, \{P_{1,c}^{\max}, \cdots, P_{D,c}^{\max}\}\}$. The power limits for the d-th antenna port are defined according to

$$P_{d,c}^{\min} = P_{d,c}^{\text{SINR}} - \xi_c(\tau)\Phi_c(d), \tag{3.19}$$

$$P_{d,c}^{\max} = P_{\text{CMAX}} - \xi_c(\tau)\Phi_c(d), \tag{3.20}$$

where $P_{d,c}^{\text{SINR}}$ is the power required to ensure minimum SINR at the d-th antenna port of the c-th cell, $\Gamma_{d,c}^{\min}$.

The action of the c-th agent at the t-th iteration is a set of TPC commands, defined as $a_{t,c} = \{\delta_1, \cdots, \delta_D\}$. The TPC command δ_d is associated with the d-th antenna port and limited to the options $\{-1, 0, +1, +3\}$ defined in [101, Table 7.1.1-1].

We consider a reward function based on a performance indicator of the network to quantify the effects of power variation of the antenna ports. The proposed reward function is a convex sum of the cell's data rates, and can be written as

$$r_{t,c} = \rho_c \sum_{j=1}^{J}\sum_{d=1}^{D} \beta \log_2(1+\Gamma_{d,j,c}) + \sum_{\substack{c'=1 \\ c'\neq c}}^{C} \rho_{c'} \left(\sum_{j=1}^{J}\sum_{d=1}^{D} \beta \log_2(1+\Gamma_{d,j,c'}) \right), \tag{3.21}$$

where β is the resource block's bandwidth and $\rho_c \in [0,1]$ is a weight factor that determines how much the data rate of the c-th cell impacts the reward. These values are also shared among cells

and their sum equals to one, i.e., $\sum_{c=0}^{C} \rho_c = 1$. Hence, the reward function depends on the SINR level $\Gamma_{d,j,c}$ and on the level of cooperation among the neighbor cells.

This expression represents a weighted average of the system's capacity taking into account all cells. In turn, the weights of this average represent the degree of importance that the agent considers for a given cell. The first term of Eq.(3.21) represents the performance indicator parameter related to the relationship between UEs and BS of a given cell. In other words, it describes the direct impact of the entity endowed with intelligence (the BS is the decision maker) on the associated UEs in the cell where it is the main entity. The second term of Eq.(3.21) measures the impact of the agent decision on the data rate of the other cells of the network. Therefore, a given action could be measured as beneficial if the data rate of the other cell increases. Thus, strategies that result in a reduction of the multi-user interference become interesting to the agents, even if the data rate of its cell remains the same or presents a reduction.

After an initial parameter configuration to define the power limits described by Eqs. (3.19) and (3.20), the process of taking actions and calculating rewards is carried out. The mapping between the states, actions and rewards is initially performed in an exploratory fashion. This means the BSs and UEs exchange commands and rewards, in order to learn the relationship between the set of actions $\mathscr{A}$ and states $\mathscr{S}$ to the observed rewards described by Eq. (3.18). Once this mapping is completed, the BS exploits it to choose the appropriate TPC command and send it to each UE. If any change in the scenario occurs, the mapping is updated, and another solution is provided. Such an update requires minimal signaling.

We assume an adaptive ε-greedy method [109] to update ε values. That is, the value of ε is gradually reduced over discrete time steps t according to

$$\varepsilon_t = \frac{\varepsilon_{\max}}{1 + \mathfrak{z}t}, \tag{3.22}$$

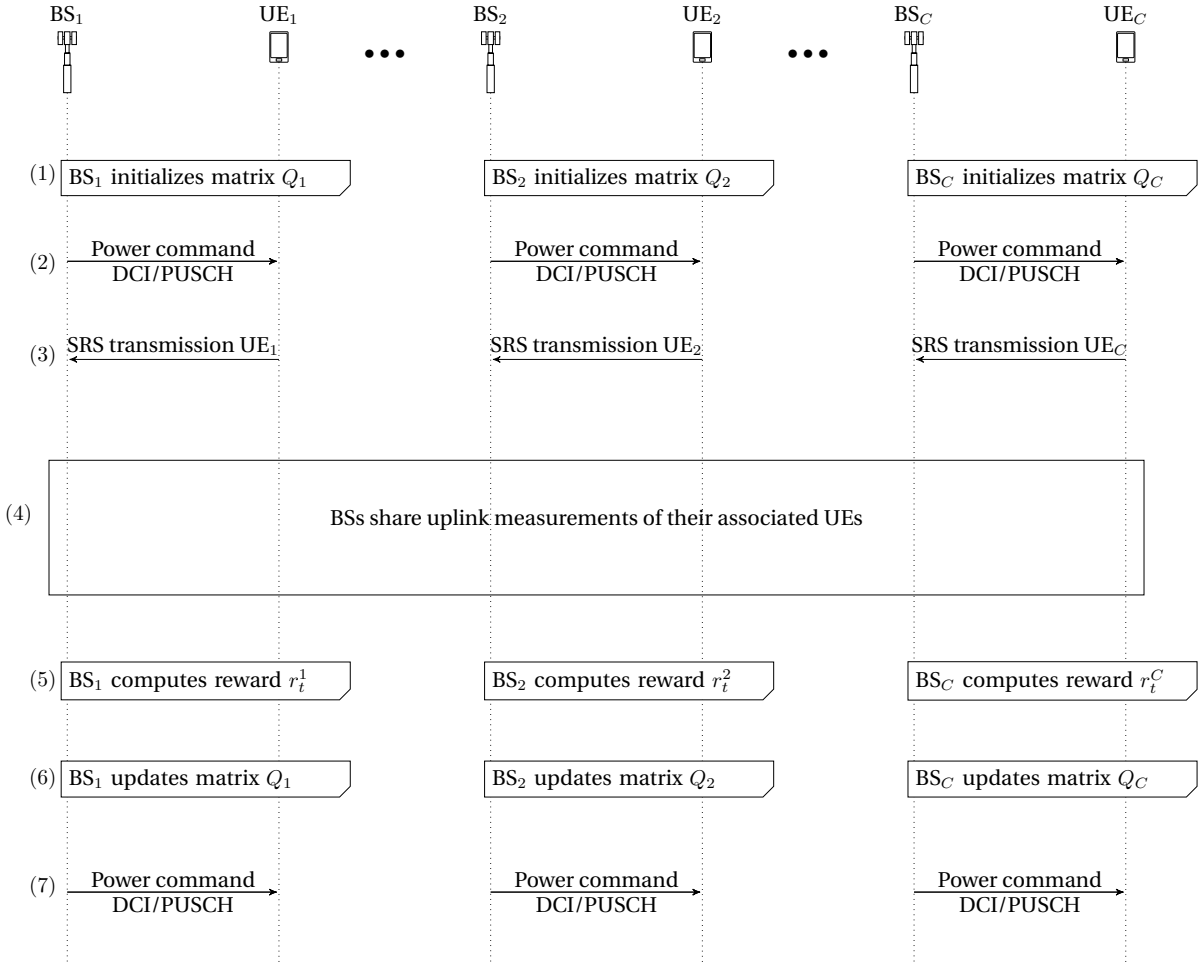
where $\varepsilon_{\max}$ is the initial exploration rate, and $\mathfrak{z}$ is a fixed parameter that guarantees a given value to ε at the t-th iteration. In the beginning of the each experiment, each agent explores intensely its state-action space and updates its matrix $\mathbf{Q}_c$, i.e., $[\mathbf{Q}_c]_{s_t,a_t} = Q(s_{t,c}, a_{t,c})$. Thus, the algorithm starts with a relative high exploration rate, which is reduced at each time step. In the end, the probability of exploitation is higher than a minimum predetermined threshold, e.g., 90%.

3.5.1 Signaling Scheme

Figure 16 represents the signaling scheme of the proposed UPC framework. Each cell has one BS that serves one UE per resource. At the step (1) , each cell c defines the dimensions and the initial assessments of the matrix $Q_c \in \mathbb{R}^{|\mathscr{S}_c| \times |\mathscr{A}_c|}$, which stores the state-action values $Q(s_{t,c}, a_{t,c})$ resultant of the interactions of the c-th BS with the environment. Therefore, it is defined the spaces of states and actions, i.e., the sets $\mathscr{S}_c$ and $\mathscr{A}_c$, respectively.

The cardinality of the set of states $\mathscr{S}_c$ is defined according to the number of power intervals and the number of antenna ports. The number of power intervals is a function of the

Figura 16 – Representation of the proposed UPC framework considering reception/transmission of measurement/information from/to another BS.



Fonte: Created by the author.

step size among transmit power levels and the power limits $P_{d,c}^{\min}$ and $P_{d,c}^{\max}$, and can be written as

$$|\mathscr{S}_c| = \left(\frac{P_{d,c}^{\max} - P_{d,c}^{\min}}{\chi} \right)^D \tag{3.23}$$

where χ is the size of the power step.

The cardinality of the set of actions $\mathscr{A}_c$ is a function of the number of TPC commands and antenna ports, since each action a_t^c is defined as a set of TPC commands sent to the antenna ports. Thus, it can be written as

$$|\mathscr{A}_c| = \iota^D \tag{3.24}$$

where ι is the number of TPC commands.

At the step (2), each BS sends a set of TPC commands to the associated UE. We assume uplink scheduling grants according to the DCI format $0-1$, where there are 2 bits reserved to adjust PUSCH transmission power [17]. We consider this format since it supports multi-antenna fields, like number of antenna ports, SRI, and Sounding Reference Signal (SRS) request. These values are defined according to the ε-greedy algorithm. Hence, each agent determines an action modeled by the tuple $a_c = \{\delta_1, \cdots, \delta_D\}$ which defines the update of the UEs' transmit power levels.

Then, at the step (3), each BS observes the new power of its associated UE according to the SRS transmission. Then, at step (4), the BSs share their uplink measurements based on the SRS transmissions of the associated UE. Based on this measurements, at step (5), each BS calculates the reward associated with the action taken at time step (2). Then, each BS updates the mapping between the spaces of actions and rewards at time step (6). Finally, at step (7), based on the updated mapping, each BS determines the next TPC commands, i.e., the next update of the power of the associated UE.

3.5.2 Pseudo Code of the Proposed UPC Framework

In the following, Algorithm 4 summarizes the main steps of the proposed RL-based framework. These instructions are carried out independently by each agent. The main loop (instructions between lines 3 and 9) determines the sequential selection of actions according to the epsilon-greedy policy. The instruction with the greatest computational complexity is defined at line 9, as it requires the search for the maximum value in the matrix $\mathbf{Q}_c$. We define the stop condition when t is equal to the total number of iterations T . This approach is also considered

in the comparison algorithms. Thus, the computational complexity of the proposed framework based on the pseudo-code Algorithm 4 is $O(TC|\mathscr{S}_c||\mathscr{A}_c|)$. For more details see Appendix B.

3.6 Comparison Algorithms

We compare the performance of the proposed UPC framework with two classical solutions found in the literature.

3.6.1 Optimal Solution Power Control

The Optimum Solution Power Control (OSPC) algorithm aims at minimizing the transmit power with SINR constraints and considering fixed transmit and receive beam pairs. This strategy has been initially studied in [111] and extended from single data stream Multiple Input Single Output (MISO) systems to multiple data streams MIMO systems in [112]. These studies investigated a joint optimization of precoder and transmit power. In our study, we consider the transmit and receive beams determined by the beam sweeping scheme described in Section 3.2.2. Thus, the OSPC algorithm is focused on the determination of the optimal transmit

Algoritmo 4: Proposed RL-based UPC framework.

let $t = 0$;
sort the initial power levels of each cell $s_{t,c} = \{P_{1,c}, \cdots, P_{D,c}\} \in \mathscr{S}_c$;
while *stop condition not reached* **do**
    sort an uniform random number $o \in [0,1]$;
    determine ε_t according to Eq. (3.22);
    if $(o < \varepsilon_t)$ *or* $(t = 0)$ **then**
        select randomly an action $a_{t,c} = \{\delta_1, \cdots, \delta_D\} \in \mathscr{A}_c$;
    else
        select the action $a_{t,c} = \{\delta_1, \cdots, \delta_D\} \in \mathscr{A}_c$ with the maximum $Q(s_{t,c}, a_{t,c})$;
    end
    compute the transmit power update defined by the action $a_{t,c}$;
    verify if the antenna port power limits $P_{d,c}^{\min}$ and $P_{d,c}^{\max}$ are respected;
    execute the transmit power updates;
    calculate the associated reward $r_{t,c}(s_{t,c}, a_{t,c}, s_{t+1,c})$ according to Eq.(3.21);
    update the matrix $\mathbf{Q}_c$ with the $Q(s_{t,c}, a_{t,c})$ value according to Eq. (3.18);
    $t = t + 1$;
end

power. This optimization problem can be formally written as

$$\begin{aligned} &\max \min_{d,c} \frac{\Gamma_{d,c}}{\Gamma_{d,c}^{\text{Target}}} \\ &\text{s.t} \quad \sum_{d=1}^{D} P_{d,c} \leq P_{\text{CMAX}} \end{aligned} \tag{3.25}$$

where $\Gamma_{d,c}^{\text{Target}}$ is the SINR target at the d-th antenna port of the UE at the c-th cell. If the ratio $\frac{\Gamma_{d,c}}{\Gamma_{d,c}^{\text{Target}}} \geq 1$, then the SINR targets are feasible. Otherwise, we have infeasible targets and is necessary lowering the SINR target values [112]. If the SINR target is not adjusted, we observe an increase of the transmit power to achieve the target since the ratio is monotonically increasing with $P_{d,c}$. Based on our simulations, the highest feasible SINR target at the considered scenario is defined as $\Gamma_{d,\,c}^{\text{Target}} = \Gamma_{d,\,c}^{\text{max}} = 6$ dB.

The authors of [111, 112] demonstrated that, given the optimal transmit power vector at the c-th cell, defined as $\mathbf{p}_c = [P_{1,c}, \cdots, P_{D,c}]^T \in \mathbb{R}^{D\times 1}$, Eq. (3.25) can be rewritten as

$$\mathbf{p}_c \frac{\Gamma_{d,c}^{\text{Target}}}{\Gamma_{d,c}} = \mathfrak{D}_c \Theta_c^T \mathbf{p}_c + \mathfrak{D}_c \boldsymbol{\sigma} \tag{3.26}$$

where $\mathfrak{D}_c = \text{diag}\left\{ \frac{\Gamma_{1,c}^{\text{Target}}}{\|\bar{\mathbf{g}}_c^H \mathbf{H}_c \bar{\mathbf{f}}_c\|^2}, \cdots, \frac{\Gamma_{D,c}^{\text{Target}}}{\|\bar{\mathbf{g}}_c^H \mathbf{H}_c \bar{\mathbf{f}}_c\|^2} \right\}$, $\boldsymbol{\sigma}$ is a vector of noise powers on the antenna ports, and Θ_c, called coupling matrix, is defined as

$$[\Theta_c]_{b_{Rx},b_{Tx}} = \begin{cases} \|\mathbf{g}_{b_{Rx}}^H \mathbf{H}_{j,c,q} \mathbf{f}_{b_{Tx}}\|^2, & \text{if} \quad b_{Tx} \neq b_{Rx}. \\ 0, & \text{otherwise.} \end{cases} \tag{3.27}$$

Multiplying both sides of Eq. (3.26) by the vector $\mathbf{1}$ and considering the constraint of Eq. (3.25), simplified to $||\mathbf{p}_c||_1 = P_{\text{CMAX}}$, we achieve

$$\frac{\Gamma_{d,c}^{\text{Target}}}{\Gamma_{d,c}} = \frac{1}{P_{\text{CMAX}}} \mathbf{1}^T \mathfrak{D}_c \Theta_c^T \mathbf{p}_c + \frac{1}{P_{\text{CMAX}}} \mathbf{1}^T \mathfrak{D}_c \boldsymbol{\sigma} \tag{3.28}$$

We can combine Eqs (3.26) and (3.28) to compose the following eigensystem

$$\Xi_c \mathbf{p}_c^{\text{Ext}} = \frac{\Gamma_{d,c}^{\text{Target}}}{\Gamma_{d,c}} \mathbf{p}_c^{\text{Ext}}, \tag{3.29}$$

where $\mathbf{p}_c^{\text{Ext}}$ is the extended transmit power vector, defined as $\mathbf{p}_c^{\text{Ext}} = [P_{1,c}, \cdots, P_{D,c}, 1]^T \in \mathbb{R}^{(D+1)\times 1}$, and Ξ_c is the extended coupling matrix, defined as

$$\Xi_c = \begin{bmatrix} \mathfrak{D}_c \Theta_c^T & \mathfrak{D}_c \boldsymbol{\sigma} \\ \dfrac{1}{P_{\text{CMAX}}} \mathbf{1}^T \mathfrak{D}_c \Theta_c^T & \dfrac{1}{P_{\text{CMAX}}} \mathbf{1}^T \mathfrak{D}_c \boldsymbol{\sigma} \end{bmatrix}. \tag{3.30}$$

The ratio $\dfrac{\Gamma_{d,c}^{\text{Target}}}{\Gamma_{d,c}}$ is the reciprocal eigenvalue of the nonnegative extended coupling matrix Ξ_c. It represents meaningful values if the conditions $\mathbf{p}_c^{\text{Ext}} > 0$ and $\dfrac{\Gamma_{d,c}^{\text{Target}}}{\Gamma_{d,c}} > 0$ are fulfilled. According to the Perron-Frobenius theorem [113], for any nonnegative real matrix there exists a nonnegative vector $\mathbf{p}_c^{\text{Ext}} \geq 0$ and an eigenvalue $e_{\max}$ such that

$$\Xi_c \mathbf{p}_c^{\text{Ext}} = e_{\max} \mathbf{p}_c^{\text{Ext}}. \tag{3.31}$$

Thus, the maximal eigenvalue $e_{\max} = \dfrac{\Gamma_{d,c}^{\text{Target}}}{\Gamma_{d,c}}$ and the associated eigenvector are always nonnegative. No other eigenvalue fulfills the positivity requirement since the relationship among transmit power and the ratio $\dfrac{\Gamma_{d,c}^{\text{Target}}}{\Gamma_{d,c}}$ is monotonic [114]. Based on that, they provide the optimal solution to the problem described by Eq. (3.25).

Therefore, the optimal transmit powers are the first D components of the dominant eigenvector of Ξ_c, scaled such that its last component equals one [111, 112]. In our study, to ensure a fair comparison between the algorithms, we limited the transmit power values to the same intervals defined in the proposed framework. That is, $P_{d,c} \in [P_{d,c}^{\min}, P_{d,c}^{\max}]$ and $\sum_{d=1}^{D} P_{d,c} \leq P_{\text{CMAX}}$. Algorithm 5 summarizes the main commands of the OSPC scheme. These instructions are carried out independently by each agent. The main loop (instructions between lines 3 and 9) determines the sequential selection of actions according to the eigen system discussed previously. The computational complexity of this strategy based on the pseudo-code Algorithm 5 is $O(TCD^2VU) + O(TCD^3)$. For more details see Appendix B.

3.6.2 Soft Dropping Power Control

Initially proposed in [115], the so called SDPC is an iterative power allocation algorithm which promotes a self-regulation of the target SINR. Figure 17 describes the behavior of the target SINR $\Gamma_{d,c}^{\text{Target}}$ according to the transmit power $P_{d,c}$. The target SINR varies from a

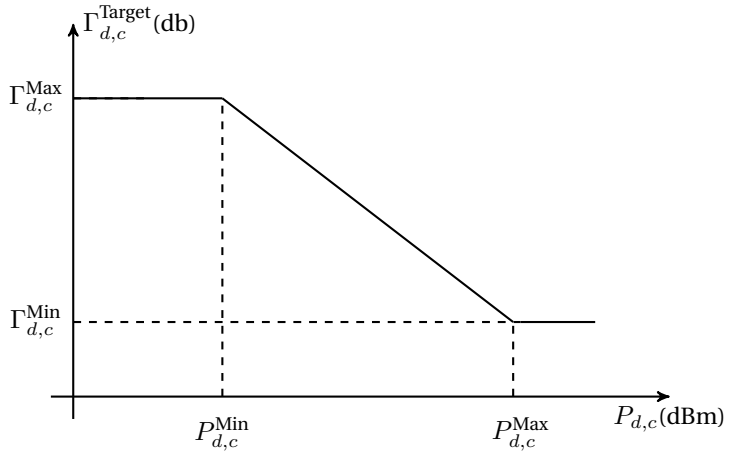
Algoritmo 5: Optimal Solution Power Control.

```
let t = 0;
define SINR targets Γ^Target_{d,c};
while stop condition not reached do
    define matrix Ξ_c;
    solve eigensystem described by Eq. (3.29);
    verify if the antenna port power limits P^min_{d,c} and P^max_{d,c} are respected;
    t ← t + 1;
end
```

maximum value $\Gamma_{d,c}^{\text{Max}}$ to a minimum value $\Gamma_{d,c}^{\text{Min}}$ as the transmit power goes from a minimum $P_{d,c}^{\text{Min}}$ to a maximum value $P_{d,c}^{\text{Max}}$. The target SINR gradually decreases as the required transmit power rises. This behavior increases the probability of determining a feasible power allocation such that the target SINR of all antenna ports can be reached. The antenna ports with worse quality, which demand higher power, aim at lower SINR values while antenna ports with better channel quality, which demand lower power, aim at higher SINR values.

The target SINR in linear scale $\hat{\Gamma}_{d,c}^{\text{Target}}$ of the d-th antenna port at the c-th cell is

Figura 17 – Behavior of the target SINR as a function of a variable transmit power.



Fonte: Created by the author.

defined according to

$$\hat{\Gamma}_{d,c}^{\text{Target}} = \begin{cases} \hat{\Gamma}_{d,c}^{\text{Max}}, \text{ if } \hat{P}_{d,c} \leq \hat{P}_{d,c}^{\text{Min}} \\ \hat{\Gamma}_{d,c}^{\text{Max}} \left(\dfrac{\hat{P}_{d,c}}{\hat{P}_{d,c}^{\text{min}}} \right)^{\varepsilon}, \text{ if } \hat{P}_{d,c}^{\text{Min}} < \hat{P}_{d,c} < \hat{P}_{d,c}^{\text{Max}} \\ \hat{\Gamma}_{d,c}^{\text{Min}}, \text{ if } \hat{P}_{d,c} \geq \hat{P}_{d,c}^{\text{Max}} \end{cases} \tag{3.32}$$

where $\hat{\Gamma}_{d,c}^{\text{Min}}$ and $\hat{\Gamma}_{d,c}^{\text{Max}}$ are the minimum and the maximum SINR in linear scale, respectively; $\hat{P}_{d,c}^{\text{Min}}$ and $\hat{P}_{d,c}^{\text{Max}}$ denote the minimum and the maximum transmit power in linear scale, respectively; $\hat{P}_{d,c}(t+1)$ is the transmit power in linear scale; and $\varepsilon = \dfrac{\log_{10}(\Gamma_{d,c}^{\text{Min}}/\Gamma_{d,c}^{\text{Max}})}{\log_{10}(P_{d,c}^{\text{Max}}/P_{d,c}^{\text{Min}})}$.

Therefore, the transmit power $\hat{P}_{d,c}(t+1)$ associated with the UE from the c-th cell and the d-th antenna port at the $(t+1)$-th iteration is updated as follows

$$\hat{P}_{d,c}(t+1) = \hat{P}_{d,c}(t) \left[\frac{\Gamma_{d,c}^{\text{Target}}(t)}{\Gamma_{d,c}(t)} \right]^{\varpi}, \tag{3.33}$$

where ϖ is a feedback parameter that controls the fraction of the difference between the target and the current SINRs that should be compensated at each iteration. It is defined as $\varpi = (1-\varepsilon)^{-1}$ to ensure convergence [116, 117].

Algorithm 6 summarizes the main commands of the SDPC scheme. At each Monte Carlo simulation, we randomly sort the initial power, i.e., $P_{d,c}^{\text{PUSCH}}(0) \in [P_{d,c}^{\text{min}}, P_{d,c}^{\text{max}}]$. According to our simulation results, the feasible SINR limits are $\Gamma_{d,c}^{\text{min}} = 0$ dB and $\Gamma_{d,c}^{\text{max}} = 6$ dB.

Then, each antenna port has its transmit power iteratively updated according to Eq. (3.33) until the stop condition is reached. In our simulations, it is defined as the total number of iterations. If the achieved power $P_{d,c}$ is over $P_{d,c}^{\text{Max}}$ or under $P_{d,c}^{\text{Min}}$, it is constrained as follows

$$P_{d,c}(t+1) = \min\{P_{d,c}^{\text{Max}}, \max\{P_{d,c}, P_{d,c}\}\}. \tag{3.34}$$

Algorithm 6 summarizes the main steps of the SDPC algorithm. These instructions are carried out independently by each agent. The computational complexity of this strategy based on the pseudo-code Algorithm 6 is $O(TCD)$. For more details see Appendix B.

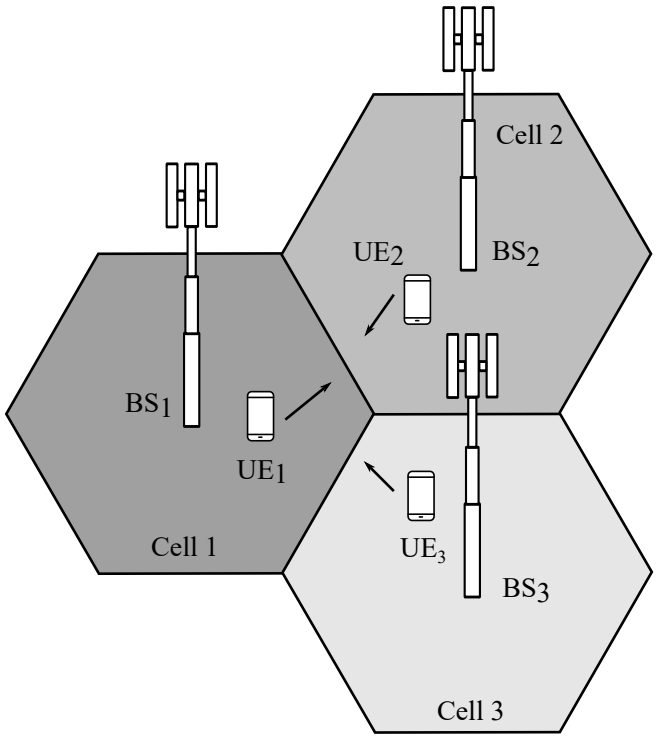
Algoritmo 6: Soft Dropping Power Control.

let $t = 0$;
determine the initial power levels of each antenna port $\hat{P}_{d,c} \in [\hat{P}_{d,c}^{\min}, \hat{P}_{d,c}^{\max}]$;
define SINR limits $\hat{\Gamma}_{d,c}^{\min}$ and $\hat{\Gamma}_{d,c}^{\max}$;
while *stop condition not reached* **do**
    calculate target SINR according to Eq. (3.32);
    update transmit power according to Eq. (3.33);
    verify if transmit power limits are respected;
    t ← t + 1;
end

3.7 Performance Evaluation

In this section, we evaluate the performance of the proposed UPC framework. We consider a simulation scenario with three cells, where each cell has one BS, one associated UE, and two antenna ports. The UEs are randomly positioned in the central region of their respective cells. Each UE moves along a linear track towards the cell border at a constant speed and a random direction, as it can be seen in Fig. 18.

Figura 18 – Simulation scenario.



Fonte: Created by the author.

The cells share the same frequency band composed of 12 subcarriers, where orthogonal uplink transmission is assumed. The subcarrier spacing is 120 kHz, since we assumed a numerology based on $\mu = 3$, and the frequency carrier is 28 GHz. In this numerology, one time slot has a duration of 0.125ms. In this work, a time slot is referred to as an iteration.

The path loss follows the Urban Macro (UMa)-Non Line of Sight (NLOS) model [75, Table 7.4.1-1]. The shadowing is modeled as log-normal distribution with a standard deviation of 4 dB. The noise power is modeled as $10\log_{10}(290 \cdot 10^{-23} \cdot \beta)$ dBm, where β is the bandwidth [118]. The main simulation parameters are listed in Table 2.

We evaluate the proposed RL-based UPC framework in terms of (i) uplink transmit power, (ii) data rate, and (ii) energy efficiency.

The data rate of the c-th cell, based on the Shannon's capacity formula, can be

Tabela 2 – General Simulation Parameters

Parameter	**Value**
Inter site distance	200 m
Minimum distance BS-UE (2D)	25 m
Angle sector	60°
BS height	15 m
UE height	1.5 m
UE track	linear
UE speed	5 km/h
BS antenna model	omnidirectional
BS antennas	8×8
UE antenna model	omnidirectional
UE antennas	2×2
Max. transmit power per carrier	24 dBm
Carrier frequency	28 GHz
Bandwidth	1.44 MHz
Number of subcarriers	12
Subcarrier spacing	120 kHz
Number of subframes	10
Number of symbols	14
Azimuth angle range	$[-60^\circ, 60^\circ]$
Elevation angle range	$[60^\circ, 120^\circ]$
Number of paths	10
Simulation rounds	100

Fonte: Created by the author.

expressed as

$$\psi_c = \sum_{d=1}^{D} \beta \log_2(1+\Gamma_{d,c}), \tag{3.35}$$

the index j is omitted for simplification purposes, since we consider only one UE per resource. Thus, the SINR defined by Eq. (3.10) as $\Gamma_{d,j,c}$ is simplified to $\Gamma_{d,c}$. Moreover, the total data rate is defined as

$$\Psi = \sum_{c=1}^{C} \sum_{d=1}^{D} \beta \log_2(1+\Gamma_{d,c}). \tag{3.36}$$

The Energy Efficiency (EE) of the c-th cell is defined as the ratio of its achievable data rate over the uplink transmit power. It can be written as

$$\omega_c = \frac{\sum_{d=1}^{D} \beta \log_2(1+\Gamma_{d,c})}{\sum_{d=1}^{D} P_{d,c}}. \tag{3.37}$$

The Network Energy Efficiency (NEE) is defined as ratio between the system achievable data rate over the total transmitted power, it is expressed as

$$\Omega = \frac{\sum_{c=1}^{C} \sum_{d=1}^{D} \beta \log_2(1+\Gamma_{d,c})}{\sum_{c=1}^{C} \sum_{d=1}^{D} P_{d,c}}. \tag{3.38}$$

The performance evaluation of the proposed RL-based UPC framework is organized into three parts. In the first part, we analyze the impact of the main RL design parameters. Then, in the second part, we evaluate the impact of the reward function on the main system performance metrics. Finally, in the third part, we compare the proposed framework with its most appropriate design parameter setting with classical power control algorithms.

3.7.1 RL Design Parameters

Without loss of generality, we assume a UE power class 3, i.e., the maximum uplink transmit power per carrier is defined as $P_{\text{CMAX}} = 24\,\text{dBm}$ [119]. Therefore, each antenna port can assume one value from a total of 25 discrete power levels, separated in steps of 1 dBm.

The state of the agent of the BS at the c-th cell can be written as $s_{t,c} \in S_c = \{\{P_{1,c}^{\min}, P_{2,c}^{\min}\}, \cdots, \{P_{1,c}^{\max}, P_{2,c}^{\max}\}\}$. Consequently, each agent has a total of $|\mathscr{S}_c| = 25^2 = 625$ states. The maximum and minimum power limits are described by Eqs. (3.19) and (3.20). Notice

that the sum of the powers of the antenna ports cannot exceed the maximum power P_{CMAX} for each UE.

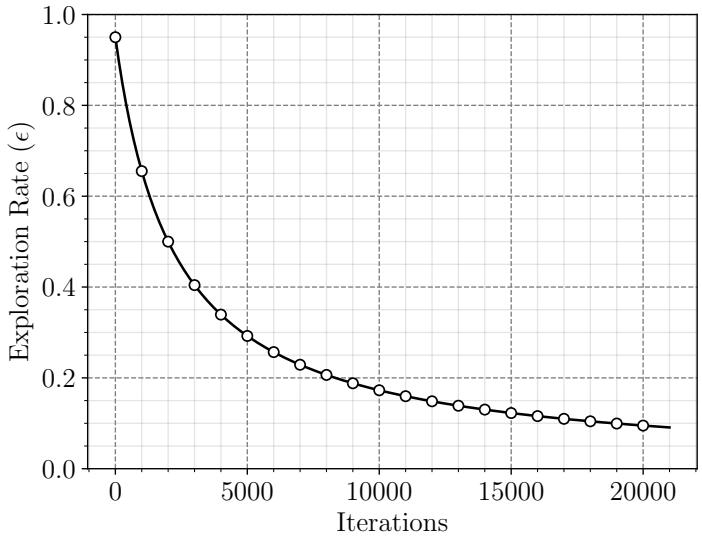
The Q-learning algorithm determines the best policy for adjusting the transmit power of each antenna port, defined by $\{\delta_1, \delta_2\}$. Recall that we consider TPC in the set $\{-1, 0, +1, +3\}$ [101, Table 7.1.1-1]. Therefore, each agent has an action space with a total of $|\mathscr{A}_c| = 4^2 = 16$ actions.

Our initial purpose is to evaluate the impact of the main Q-learning design parameters, namely, exploration rate, learning rate, and discount factor in the network performance in terms of NEE. In the practice, this tuning process will be required at the beginning of the operation of the UPC to determine the most appropriate set of design parameters in terms of NEE.

We assume an adaptive ε-greedy algorithm where the value of the exploration rate (ε) is gradually reduced over discrete time steps t according to Eq. (3.22). Therefore, in the proposed simulation setting, ξ is a fixed parameter that guarantees $\varepsilon_t = 0.50$ when $t = 2{,}000$. Figure 19 describes the behavior of the proposed adaptive ε-greedy algorithm. In the beginning of each experiment, the agent explores intensely the state-action space and updates its Q-table $\mathbf{Q}_c$. In the end, the probability of exploitation is higher than 90%.

Equation (3.18) indicates the role of the learning rate (η) and the discount factor (φ) on the determination of the $Q(s_{t,c}, a_{t,c})$ values. We rewrite it to emphasize the impact of these

Figura 19 – Proposed behavior of the exploration rate according to the adaptive ε-greedy algorithm described by Eq. (3.22).



Fonte: Created by the author.

parameters:

$$\underbrace{Q(s_{t+1},a_{t+1})}_{\text{new value}} = \underbrace{Q(s_t,a_t)}_{\text{old value}} + \eta \left[r_{t+1} + \varphi \underbrace{\max_{a_{t+1}\in\mathscr{A}} Q(s_{t+1},a_{t+1})}_{\text{estimate of the optimal future value}} - Q(s_t,a_t) \right], \quad (3.39)$$

where the learning rate determines how much the most recent information replaces older information and the discount factor determines the impact of estimate of optimal future values on the new Q value.

Figure 20 shows the average NEE achieved with the proposed RL-based UPC framework with different design parameter settings. We consider the reward function with $\rho_c = \rho_{c'} = 0.5$. Figure 20b indicates the behavior of the proposed adaptive ε-greedy algorithm described by Eq. (3.22). In comparison, we evaluate a fixed ε-greedy algorithm, whose results are depicted at Fig. 20a. In this case, the exploration rate remains constant at $\varepsilon = 0.50$. We consider distinct values of learning rate $\eta \in [0,1]$ and discount factor $\varphi \in [0,1)$.

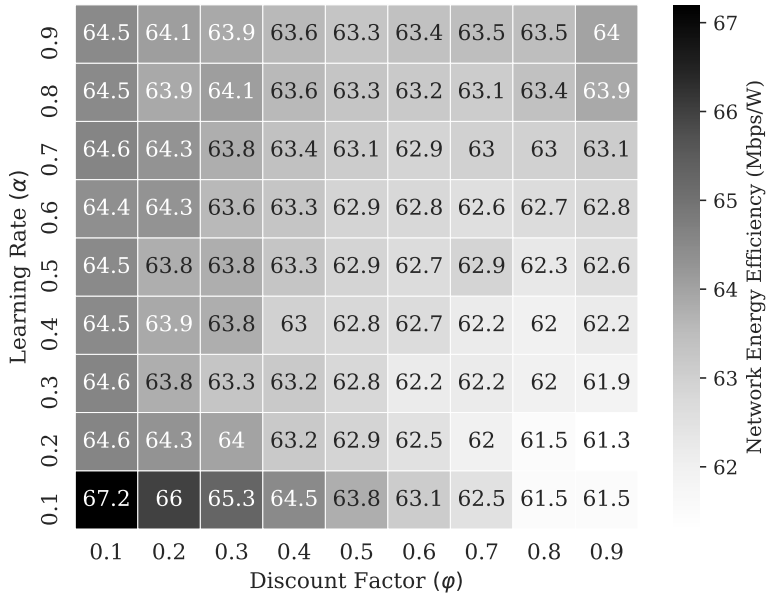
The adaptive ε-greedy algorithm achieves NEE levels higher than the fixed ε-greedy algorithm for most of the evaluated configurations since it enhances the quality of the learned policies due to the balancing between exploration and exploitation. In the beginning, the exploration rate has higher values, which allows the agent to explore the spaces of actions and states. Consequently, the agent is able to discover the impact of different decisions. The level of interaction with the environment reduces as the agent analyzes the impact of its decisions. At the end of the simulation, with lower exploration rate values, the agent takes the most suitable actions. Therefore, when evaluating the settings with the highest levels of energy efficiency, there is a gain of 75% for the adaptive ε-greedy algorithm.

Both ε-greedy algorithms achieve the highest values of NEE with similar ranges of learning rate and discount factor values, namely $\eta \to 0$ and $\varphi \to 0$. Small values of learning rate means agents prioritizing the exploitation of prior knowledge while small values of discount factor mean an agent tends to consider the current reward value. Thus, this setting of design parameters results in a learning system is strongly influenced by the acquired knowledge and the reward function.

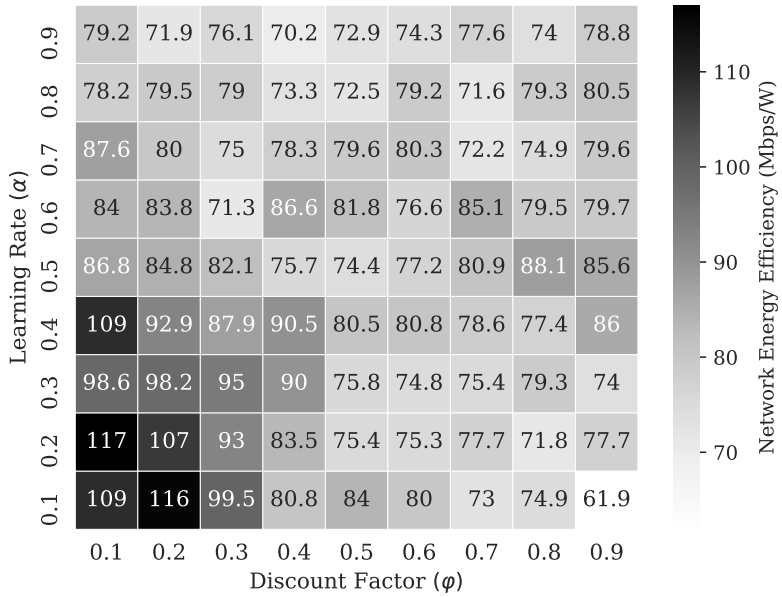
Therefore, according to the presented results the most appropriate design parameter setting to increase the NEE corresponds to set $\eta = 0.20$ and $\varphi = 0.10$. Table 3 summarizes the main Q-learning parameters considered in the performance evaluation provided in the following.

Figura 20 – Average NEE achieved with the proposed RL-based UPC framework with different design parameter settings.

(a) Fixed ε-greedy



(b) Adaptive ε-greedy



Fonte: Created by the author.

Tabela 3 – Main RL Design Parameters

Parameter	**Value**
Number of iterations	21,000
Discount factor (φ)	0.10
Learning rate (η)	0.20
Initial exploration rate ($\varepsilon_{\max}$)	0.95
Number of states	625
Number of actions	16

Fonte: Created by the author.

3.7.2 Evaluation of the Proposed Reward Function

In the following, using the parameter setting defined in Table 3, we assess the proposed power control solution in terms of the total transmitted power and the achieved data rate. Our analysis is focused on the impact of the reward function in the proposed power management policy.

In order to reduce the degrees of freedom in our problem, we rewrite the reward function as a convex sum of data rates, where the first term contains the data rate of the cell whose agent directly manages the transmit power and the second term accounts the data rate of the other cells in the system. This adaptation aims to simplify the determination of the best set of design parameters and to reduce the sensitivity of the learning system. Thus, the reward functions of the agents in our simulation model can be written as

$$r_{t,c} = \rho \sum_{d=1}^{D} \beta \log_2(1+\Gamma_{d,c}) + (1-\rho) \sum_{\substack{c'=1 \\ c' \neq c}}^{C} \sum_{d=1}^{D} \beta \log_2(1+\Gamma_{d,c'}), \tag{3.40}$$

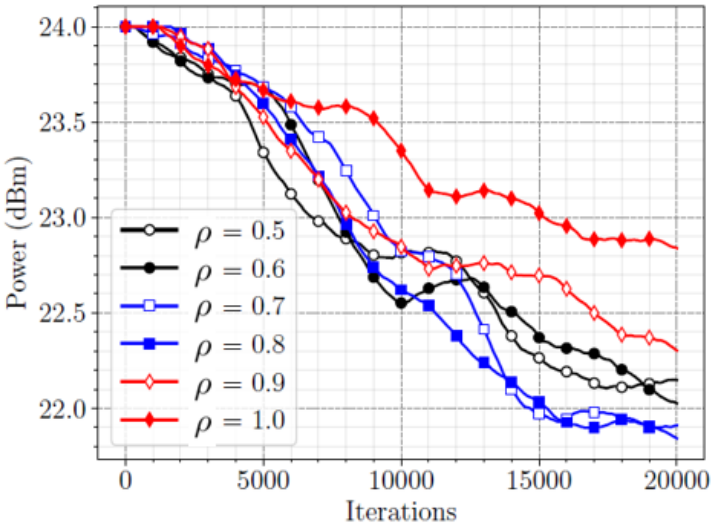
where ρ is the design parameter that regulates how the agent behavior is impacted by the gains or losses of the data rates of the other cells. Moreover, it also determines the level of cooperation among the agents in the learning process.

Figure 21a examines the total transmit power (in dBm) as a function of the number of iterations. We apply a Simple Moving Averaging (SMA) with a window of 500 iterations to smooth the curves. All BSs operate their agents simultaneously according to the reward functions described by Eq. (3.40). The proposed UPC framework decreases the total transmission power over the iterations in comparison with the initial level from 1 dBm to 2 dBm, varying according to the value of ρ. Therefore, the learning process resulting from the interaction with the system promotes an energy efficient resource management policy.

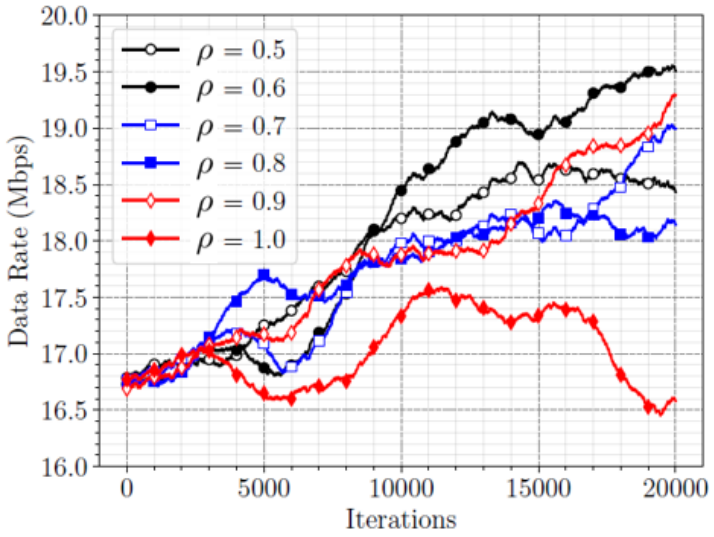
The behavior of the total transmit power varies according to the value of the design parameter ρ. On the one hand, when $\rho = 1.0$, the algorithm attempts to maximize the data rate of each cell and this parameter is influenced solely by its transmission power. Therefore, the algorithm increases the powers of all UEs indistinctly. Each UE updates its powers autonomously, without any explicit observation of how its behavior affects the remaining of the system. Consequently, this parameter configuration reaches the highest power levels, namely 24 dBm. On the other hand, when $\rho < 1.0$, each agent maximizes its own expression of the

Figura 21 – Evaluation of the proposed RL-based UPC design considering different values of the design parameter ρ.

(a) Total transmit power



(b) Total data rate



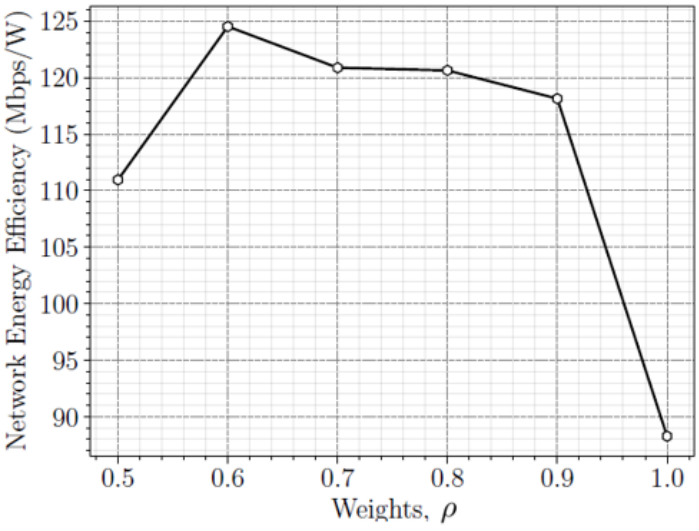
Fonte: Created by the author.

weighted sum of the data rates, thus allowing to measure the impact of its decisions in the other cells. Figure 21a reveals power management policies with data rates of other cells allow UE transmissions at a lower power level than the power level obtained by policies with $\rho = 1.0$. In comparison with $\rho = 1.0$, there is a reduction of 1 dBm in the total transmit power when $\rho = 0.8$.

Figure 21b shows the total data rate (in Mbps) as a function of the number of iterations. The overall system data rate decreases when $\rho = 1.0$ because UEs tend to transmit t maximum power, thus reducing the SINR. The cooperation among agents is capable of significantly improving the system conditions. The curves show that $\rho = 0.9$ promotes an enhancement of 15% of the total data rate in comparison with $\rho = 1$. The best performance of the system in terms of total rate is seen when $\rho = 0.6$, which increases in 20% the total data rate in comparison with the worst case.

Figure 22 indicates the average NEE in the last $2,000$ iterations achieved by the proposed RL-based UPC considering different values of ρ. We observe higher levels of the NEE when $\rho < 1$. Considering this parameter setting, the proposed UPC framework reduces the transmit power and increases the total data rate. However, it requires an exchange of information between cells. That is, the enhancement of the network efficiency in at least 20% comes at the cost of a signaling exchange among BSs. The NEE with $\rho = 0.6$ is 40% higher than when $\rho = 1$. Therefore, the design parameter $\rho = 0.60$ achieves the highest NEE in the considered scenario.

Figura 22 – Network energy efficiency achieved by the proposed RL-based UPC considering different values of the design parameter of cooperation among agents ρ.



Fonte: Created by the author.

Note that the determination of the most suitable design parameter value ρ to obtain the highest levels of NEE depends on the simulation scenario, being predominant the UE's location within the cell. On one hand, UEs located at the cell edge experience high inter-cellular interference. In this case, lower values of ρ represent the best choice, as this configuration seeks to balance reward's objectives more cooperatively, leading to lower levels of interference. On the other hand, UEs close to the BS are less affected. In this case, higher ρ values can be chosen without the negative effects described above, since the increase in power does not significantly increase the levels of inter-cellular interference.

We have performed an exhaustive search for the parameter ρ to determine its behavior in a given scenario and to validate the proposed technique. However, from a practical point of view, the development of an adaptive strategy for this parameter would be worth to investigate. This is one of our prospects for future works.

3.7.3 Comparison with Classical Algorithms

In the following, we compare the performance of the proposed UPC framework, hereafter referred as Reinforcement Learning Power Control (RLPC), with two classical solutions found in the literature and described in Section 3.6.

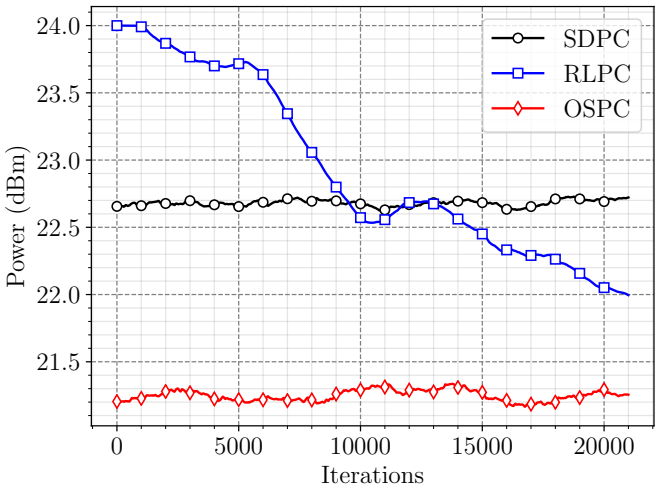
Figure 23a shows the behavior of the total transmit power (in dBm) as a function of the number of iterations. In the initial iterations, the RLPC presents the highest total transmit power. It exceeds the levels obtained by SDPC by 0.5 dBm. However, this disadvantage is reversed as the agents interact with the environment and the knowledge acquired is used in decision making. At the end of the simulation, the RLPC outperforms the SDPC, with a significant reduction of the total transmit power (1.5 dBm), and its transmit power levels approach that observed with OSPC.

The proposed RLPC has a slower convergence, but finds a power solution able to reduce the interference and increase the SINR. Consequently, we observe a continuous increment of the total data rate, that enhances 20% compared with SDPC and approaches the optimal solution, as it can be seen in Figure 23b.

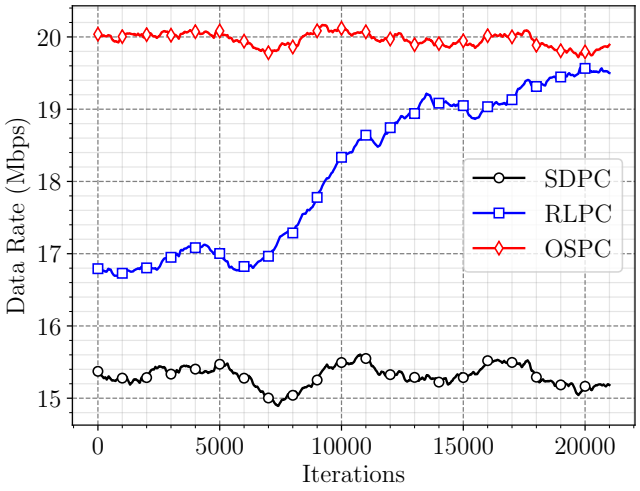
Figure 24 depicts the behavior of the NEE as a function of the number of iterations. The SDPC is not able to learn new strategies from the interaction with the environment. Thus, it cannot approach to the optimal solution. The RLPC provides a self-exploratory energy-efficient solution which enhances its network energy efficient approximately 95%, achieving 75% of the

Figura 23 – Comparison of the proposed RL-based UPC design with classical algorithms.

(a) Total transmit power



(b) Total data rate

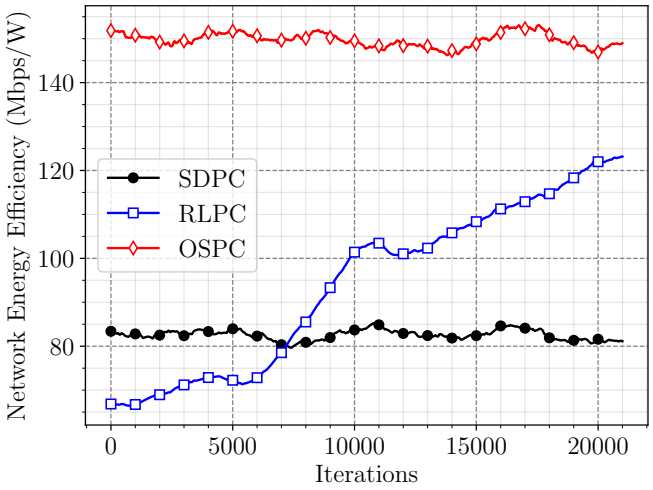


Fonte: Created by the author.

performance of the optimal solution.

The RLPC has computational complexity higher than that of SDPC, since it requires the determination of the largest value of the Q table, while SDPC only requires a comparison of two scalar values. The computational disadvantage of RLPC is compensated by the significant reduction in transmit power and the substantial increase in the data rate in comparison with SDPC. OLPC obtains the best results in all analyzed parameter settings. This occurs at a high computational cost resulting from the eigendecomposition operation involving a high-dimensional matrix. In addition, OLPC requires an intense signaling, since the entire channel

Figura 24 – Comparison of the proposed RL-based UPC design with classical algorithms in terms of NEE.



Fonte: Created by the author.

matrix, precoders, decoders and SINR targets must be informed at each iteration. The RLPC algorithm involves a much lower signaling level, requiring only the parameter ρ (on the beginning of the process) and the data rate. Table 4 summarizes the main comparison aspects of the algorithms under analysis.

Tabela 4 – Comparison Among Uplink Power Control Algorithms.

Algorithm	**Signaling**	**Complexity**
SDPC	Low	Low
RLPC	Low	High
OSPC	High	High

Fonte: Created by the author.

3.8 Chapter Remarks

The proposed UPC framework based on multi-agent RL for a 5G NR network provides a self-exploratory energy-efficient solution. It enabled the system to learn a suitable set of transmit power commands to enhance the total data rate on the uplink channel under neighbor cell interference mitigation.

Based on the principles of the Q-learning algorithm, it is defined that each BS is an entity endowed with intelligence, that dynamically interacts with the environment determi-

ning transmit power commands executed by their associated UEs. After implementing such commands, each UE reports a Quality of Service (QoS) indicator to its corresponding BS. In addition, the BSs share among themselves their respective QoS indicators. This set of information characterizes the knowledge acquired by each learning agent. This knowledge is mapped into a reward, according to a predefined cost function, which impacts the long-term behavior of the transmit power.

Simulation results show that the proposed UPC framework provides a QoS similar to the conventional solutions while significantly reducing the UE's power consumption. Moreover, higher sum rates and increased NEE are attained when the framework considers the cooperation among BSs.

4 CONCLUSIONS AND PERSPECTIVES

This thesis investigated Radio Resource Management (RRM) techniques in the context of Fifth Generation (5G) networks and based on Machine Learning (ML) paradigms. It considered practical implementation aspects, relying on the main technical specifications from the Third Generation Partnership Project (3GPP) Release 15. Despite the limitations imposed by these conditions, such as hardware restrictions and available signaling, the proposed solutions met relevant engineering design requirements with reduced signaling overhead and computational complexity. The main contributions and research perspectives for each chapter are summarized in the following.

Chapter 2 described a hybrid beamforming design under limited feedback requirements. The Base Station (BS) performs a beam sweeping procedure to collect Channel State Information (CSI) as a function of Precoding Matrix Indicators (PMIs). The hybrid beamforming design based on Joint Spatial Division and Multiplexing (JSDM), considered for comparison, requires the report of the covariance matrix of each User Equipment (UE) to the BS so that the total amount of resources needed for reporting the whole channel covariance matrix is proportional to U^2J, where U is the number of transmit antennas and J is the number of UEs in the system. The proposed solution, on its turn, requires a signaling load proportional only to $2NJ$, where N is the number of reported PMIs. Since $N \ll U$, it is observed a meaningful signaling reduction, as well as an adequate representation of the channel statistics for the hybrid beamforming design in a Millimeter Wave (mmWave) massive Multiple Input Multiple Output (MIMO) system when compared to the conventional report of the channel covariance matrix.

Chapter 2 also investigated how the Unsupervised Learning (UL) paradigm provides clustering algorithms able to exploit the channel information provided by the codebook-based Type II CSI feedback. The simulations results indicated that the partitioning UEs based on Fuzzy C-Means (FCM) was able to explore more efficiently the reduced CSI feedback in the most challenging channel conditions. However, this algorithm has higher computational complexity than the conventional solution based on k-means++. The proposed hybrid beamforming scheme reduced the Multi User (MU) interference and achieved significant gains in total data rate as the channel conditions and interference environment become more challenging.

Some of the research perspectives for the studies performed at Chapter 2 include the investigation of other UL paradigms other than centroid-based clustering, such as connectivity-based clustering, density-based clustering, and grid-based clustering. Moreover, the application

of dynamic clustering algorithms to update the number of clusters according to channel conditions or system requirements could bring flexibility to the network operation. We also consider as a relevant perspective the investigation of more advanced scheduling algorithms, considering the optimization of some design parameters, such as data rate and power.

Chapter 3 proposed an Uplink Power Control (UPC) framework compliant with the technical specifications from 3GPP Release 15. Based on the fundamentals of multi-agent Reinforcement Learning (RL), it was developed a beam-based power control algorithm able to learn a strategy that mitigates the inter-cell interference and, consequently, enhances the total data rate on the uplink channel.

Based on the Q-learning algorithm, it was defined that each BS is an entity endowed with intelligence, that dynamically interacts with the environment determining the transmit power commands executed by their associated UEs. After implementing such commands, the proposed signaling scheme defines that each UE reports a Quality of Service (QoS) indicator to its corresponding BS. In addition, the BSs share among themselves their respective QoS indicators. This set of information characterizes the knowledge acquired by each learning agent. This knowledge is mapped into a reward, according to a predefined cost function, which impacts the long-term behavior of the transmit power. The simulation results indicate a near-optimum performance of the proposed framework in terms of total transmit power, total data rate, and network energy efficiency. The proposed signaling scheme provided a self-exploratory transmit power control strategy that overcomes the classical Soft Dropping Power Control (SDPC) algorithm with similar signaling levels.

Some of the research perspectives for the studies performed in Chapter 3 include the development of an adaptive update of the RL parameters learning rate and discount rate and the investigation of an offline step to initialize Q-table and speed up the convergence of the proposed UPC algorithm. Another relevant research aspect is the application of Deep Reinforcement Learning (DRL) algorithms into the proposed UPC framework, considering the practical limitations imposed by the standard.

As the 5G standard is being completed, academia and industry began to consider the development of the next generation of wireless networks. The Sixth Generation (6G) is expected to achieve even higher data rates and require broader frequency ranges. ML will occupy an even more prominent position in the development of 6G networks as the scenarios will become extremely complex and the need for computational efficiency is expected to increase

significantly. Therefore, the content of this thesis contribute to the standardization of ML embedded communications to advance the ML techniques into 6G.

REFERÊNCIAS

1 Wang, C. et al. Cellular Architecture and Key Technologies for 5G Wireless Communication Networks. *IEEE Communications Magazine*, v. 52, n. 2, p. 122–130, fev. 2014.

2 ITU-RM2083. *Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond*. [S.l.], 2015.

3 Tullberg, H. et al. The METIS 5G System Concept: Meeting the 5G Requirements. *IEEE Communications Magazine*, v. 54, n. 12, p. 132–139, dez. 2016.

4 Sharma, S. K.; Wang, X. Toward Massive Machine Type Communications in Ultra-Dense Cellular IoT Networks: Current Issues and Machine Learning-Assisted Solutions. *IEEE Communications Surveys Tutorials*, v. 22, n. 1, p. 426–471, maio 2020.

5 Chen, H. et al. Ultra-Reliable Low Latency Cellular Networks: Use Cases, Challenges and Approaches. *IEEE Communications Magazine*, v. 56, n. 12, p. 119–125, set. 2018.

6 ICT-317669-METIS. *Scenarios, Requirements, and KPIs for 5G Mobile and Wireless System*. [S.l.], 2013.

7 ICT-317669-METIS. *Updated Scenarios, Requirements, and KPIs for 5G Mobile and Wireless System with Recommendations for Future Investigations*. [S.l.], 2015.

8 ITU-RM2410. *Minimum Requirements Related to Technical Performance for IMT-2020 Radio Interface(s)*. [S.l.], 2017.

9 3GPP. *3GPP Specification Serie 38*. 2016. Disponível em: <http://www.3gpp.org/DynaReport/38-series.htm>.

10 3GPP. *3GPP Release 15*. Disponível em: <http://www.3gpp.org/release-15>.

11 3GPP. *3GPP Release 16*. Disponível em: <http://www.3gpp.org/release-16>.

12 Ghosh, A. et al. 5G Evolution: A View on 5G Cellular Technology Beyond 3GPP Release 15. *IEEE Access*, v. 7, p. 127639–127651, set. 2019.

13 3GPP. *3GPP Release 17*. Disponível em: <http://www.3gpp.org/release-17>.

14 Peisa, J. et al. 5G Evolution: 3GPP Releases 16 & 17 Overview. *Ericsson Technology Review*, p. 1–14, fev. 2020. Disponível em: <https://www.ericsson.com/493cdb/assets/local/reports-papers/ericsson-technology-review/docs/2020/5g-nr-evolution.pdf>.

15 Rappaport, T. S. et al. Millimeter Wave Mobile Communications for 5G Cellular: It Will Work! *IEEE Access*, v. 1, p. 335–349, maio 2013.

16 Lu, L. et al. An Overview of Massive MIMO: Benefits and Challenges. *IEEE Journal of Selected Topics in Signal Processing*, v. 8, n. 5, p. 742–758, abr. 2014.

17 Dahlman, E.; Parkvall, S.; Skold, J. *5G NR: The Next Generation Wireless Access Technology*. [S.l.]: Academic Press, 2018. v. 1. ISBN 978-01-2814-323-0.

18 Andrews, J. G. et al. What Will 5G Be? *IEEE Journal on Selected Areas in Communications*, v. 32, n. 6, p. 1065–1082, jun. 2014. ISSN 0733-8716.

19 Agiwal, M.; Roy, A.; Saxena, N. Next Generation 5G Wireless Networks: A Comprehensive Survey. *IEEE Communications Surveys Tutorials*, v. 18, n. 3, p. 1617–1655, 2016.

20 Olwal, T. O.; Djouani, K.; Kurien, A. M. A Survey of Resource Management Toward 5G Radio Access Networks. *IEEE Communications Surveys Tutorials*, v. 18, n. 3, p. 1656–1686, 2016.

21 Osseiran, A. et al. Scenarios for 5G Mobile and Wireless Communications: The Vision of the METIS Project. *IEEE Communications Magazine*, v. 52, n. 5, p. 26–35, maio 2014. ISSN 0163-6804.

22 ERICSSON. *5G Radio Access*. [S.l.], 2016. Disponível em: <https://www.ericsson.com/en/white-papers/5g-radio-access--capabilities-and-technologies>.

23 ERICSSON. *5G Systems*. [S.l.], 2017. Disponível em: <https://www.ericsson.com/en/white-papers/5g-systems--enabling-the-transformation-of-industry-and-society>.

24 Gupta, A.; K. Jha, R. A Survey of 5G Network: Architecture and Emerging Technologies. *IEEE Access*, v. 3, n. 1, p. 1206–1232, jul. 2015. ISSN 2169-3536.

25 3GPP. *Release Description*. [S.l.], 2018.

26 Zaidi, A. et al. *5G Physical Layer: Principles, Models, and Technology Components*. [S.l.]: Academic Press, 2018.

27 Parkvall, S. et al. NR: The New 5G Radio Access Technology. *IEEE Communications Standards Magazine*, v. 1, n. 4, p. 24–30, dez. 2017.

28 Li, Q. C. et al. 5G Network Capacity: Key Elements and Technologies. *IEEE Vehicular Technology Magazine*, v. 9, n. 1, p. 71–78, mar. 2014. ISSN 1556-6072.

29 Rangan, S.; Rappaport, T. S.; Erkip, E. Millimeter-Wave Cellular Wireless Networks: Potentials and Challenges. *Proceedings of the IEEE*, v. 102, n. 3, p. 366–385, fev. 2014.

30 Rappaport, T. S. et al. Wideband Millimeter-Wave Propagation Measurements and Channel Models for Future Wireless Communication System Design. *IEEE Transactions on Communications*, v. 63, n. 9, p. 3029–3056, maio 2015.

31 Rappaport, T. S. et al. Overview of Millimeter Wave Communications for Fifth-Generation (5G) Wireless Networks-With a Focus on Propagation Models. *IEEE Transactions on Antennas and Propagation*, v. 65, n. 12, p. 6213–6230, dez. 2017. ISSN 0018-926X.

32 Chen, J.; Ge, X.; Ni, Q. Coverage and Handoff Analysis of 5G Fractal Small Cell Networks. *IEEE Transactions on Wireless Communications*, v. 18, n. 2, p. 1263–1276, fev. 2019. ISSN 1536-1276.

33 Akdeniz, M. R. et al. Millimeter Wave Channel Modeling and Cellular Capacity Evaluation. *IEEE Journal on Selected Areas in Communications*, v. 32, n. 6, p. 1164–1179, jun. 2014.

34 Chin, W. H.; Fan, Z.; Haines, R. Emerging Technologies and Research Challenges for 5G Wireless Networks. *IEEE Wireless Communications*, v. 21, n. 2, p. 106–112, abr. 2014. ISSN 1536-1284.

35 Larsson, E. G. et al. Massive MIMO for Next Generation Wireless Systems. *IEEE Communications Magazine*, v. 52, n. 2, p. 186–195, fev. 2014.

36 T. L. Marzetta. How Much Training is Required for Multiuser Mimo? In: *Asilomar Conference on Signals, Systems and Computers*. [S.l.: s.n.], 2006. p. 359–363. ISSN 1058-6393.

37 T. L. Marzetta. Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. *IEEE Transactions on Wireless Communications*, v. 9, n. 11, p. 3590–3600, nov. 2010. ISSN 1536-1276.

38 Caire, G.; Shamai, S. On the Achievable Throughput of a Multiantenna Gaussian Broadcast Channel. *IEEE Transactions on Information Theory*, v. 49, n. 7, p. 1691–1706, jul. 2003. ISSN 0018-9448.

39 Vishwanath, P.; Tse, D. N. C. Sum Capacity of the Vector Gaussian Broadcast Channel and Uplink-Downlink Duality. *IEEE Transactions on Information Theory*, v. 49, n. 8, p. 1912–1921, ago. 2003. ISSN 0018-9448.

40 Vishwanath, S.; Jindal, N.; Goldsmith, A. Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels. *IEEE Transactions on Information Theory*, v. 49, n. 10, p. 2658–2668, out. 2003. ISSN 0018-9448.

41 Gesbert, D. et al. Shifting the MIMO Paradigm. *IEEE Signal Processing Paradigm*, v. 24, n. 5, p. 36–46, set. 2007. ISSN 1053-5888.

42 Rusek, F. et al. Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays. *IEEE Signal Processing Magazine*, v. 30, n. 1, p. 40–60, jan. 2013. ISSN 1053-5888.

43 Molisch, A. F. et al. Hybrid Beamforming for Massive MIMO: A Survey. *IEEE Communications Magazine*, v. 55, n. 9, p. 134–141, set. 2017.

44 Buzzi, S. et al. A Survey of Energy-Efficient Techniques for 5G Networks and Challenges Ahead. *IEEE Journal on Selected Areas in Communications*, v. 34, n. 4, p. 697–709, abr. 2016.

45 Simonsson, A.; Furuskar, A. Uplink Power Control in LTE - Overview and Performance, Subtitle: Principles and Benefits of Utilizing rather than Compensating for SINR Variations. In: *IEEE Vehicular Technology Conference*. [S.l.: s.n.], 2008. p. 1–5.

46 H. Zhang and N. Prasad and S. Rangarajan and S. Mekhail and S. Said and R. Arnott. Standards-Compliant LTE and LTE-A Uplink Power Control. In: *IEEE International Conference on Communications*. [S.l.: s.n.], 2012. p. 5275–5279.

47 Lee, Y. L. et al. Recent Advances in Radio Resource Management for Heterogeneous LTE/LTE-A Networks. *IEEE Communications Surveys Tutorials*, v. 16, n. 4, p. 2142–2180, 2014. ISSN 2373-745X.

48 Ku, G.; Walsh, J. M. Resource Allocation and Link Adaptation in LTE and LTE Advanced: A Tutorial. *IEEE Communications Surveys Tutorials*, v. 17, n. 3, p. 1605–1633, 2015. ISSN 2373-745X.

49 Jiang, C. et al. Machine Learning Paradigms for Next-Generation Wireless Networks. *IEEE Wireless Communications*, v. 24, n. 2, p. 98–105, jul. 2017. ISSN 1536-1284.

50 Morocho-Cayamcela, M. E.; Lee, H.; Lim, W. Machine Learning for 5G/B5G Mobile and Wireless Communications: Potential, Limitations, and Future Directions. *IEEE Access*, v. 7, p. 137184–137206, 2019. ISSN 2169-3536.

51 Li, R. et al. Intelligent 5G: When Cellular Networks Meet Artificial Intelligence. *IEEE Wireless Communications*, v. 24, n. 5, p. 175–183, out. 2017. ISSN 1536-1284.

52 Klaine, P. V. et al. A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks. *IEEE Communications Surveys Tutorials*, v. 19, n. 4, p. 2392–2431, jul. 2017. ISSN 1553-877X.

53 Calabrese, F. D. et al. Learning Radio Resource Management in RANs: Framework, Opportunities, and Challenges. *IEEE Communications Magazine*, v. 56, n. 9, p. 138–145, set. 2018.

54 Sezer, O. B.; Dogdu, E.; Ozbayoglu, A. M. Context-Aware Computing, Learning, and Big Data in Internet of Things: A Survey. *IEEE Internet of Things Journal*, v. 5, n. 1, p. 1–27, nov. 2018.

55 Kibria, M. G. et al. Big Data Analytics, Machine Learning, and Artificial Intelligence in Next-Generation Wireless Networks. *IEEE Access*, v. 6, p. 32328–32338, maio 2018. ISSN 2169-3536.

56 Sun, Y. et al. Application of Machine Learning in Wireless Networks: Key Techniques and Open Issues. *IEEE Communications Surveys Tutorials*, v. 21, n. 4, p. 3072–3108, 2019. ISSN 2373-745X.

57 Hussain, F. et al. Machine Learning for Resource Management in Cellular and IoT Networks: Potentials, Current Solutions, and Open Challenges. *IEEE Communications Surveys Tutorials*, v. 22, n. 2, p. 1251–1275, jan. 2020.

58 C. M. Bishop. *Pattern Recognition and Machine Learning*. [S.l.]: Springer, 2006. ISBN 9780387310732,0387310738.

59 M. Sugiyama. *Introduction to Statistical Machine Learning*. [S.l.]: Morgan Kaufmann, 2016. ISBN 978-0-12-802121-7.

60 Shafin, R. et al. Artificial Intelligence-Enabled Cellular Networks: A Critical Path to Beyond-5G and 6G. *IEEE Wireless Communications*, v. 27, n. 2, p. 212–217, mar. 2020.

61 Sutton, R.; Barto, A. G. *Reinforcement Learning: An Introduction*. MIT Press, 2018. (Adaptive Computation and Machine Learning series). ISBN 9780262039246. Disponível em: <https://books.google.com.br/books?id=6DKPtQEACAAJ>.

62 Ahmed, I. et al. A Survey on Hybrid Beamforming Techniques in 5G: Architecture and System Model Perspectives. *IEEE Communications Surveys Tutorials*, v. 50, n. 4, p. 3060–3097, 2018. ISSN 1553-877X.

63 Zi, R. et al. Energy Efficiency Optimization of 5G Radio Frequency Chain Systems. *IEEE Journal on Selected Areas in Communications*, v. 34, n. 4, p. 758–771, abr. 2016. ISSN 0733-8716.

64 Ge, X. et al. Joint Optimization of Computation and Communication Power in Multi-User Massive MIMO Systems. *IEEE Transactions on Wireless Communications*, v. 17, n. 6, p. 4051–4063, jun. 2018. ISSN 1536-1276.

65 Björnson, E.; Larsson, E. G.; Marzetta, T. L. Massive MIMO: Ten Myths and One Critical Question. *IEEE Communications Magazine*, v. 54, n. 2, p. 114–123, fev. 2016. ISSN 0163-6804.

66 Zhang, X.; Molisch, A. F.; Kung, S. Variable-Phase-Shift-Based RF-Baseband Codesign for MIMO Antenna Selection. *IEEE Transactions on Signal Processing*, v. 53, n. 11, p. 4091–4103, nov. 2005. ISSN 1053-587X.

67 Ayach, O. E. et al. Spatially Sparse Precoding in Millimeter Wave MIMO Systems. *IEEE Transactions on Wireless Communications*, v. 13, n. 3, p. 1499–1513, mar. 2014. ISSN 1536-1276.

68 Alkhateeb, A. et al. Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems. *IEEE Journal of Selected Topics in Signal Processing*, v. 8, n. 5, p. 831–846, out. 2014. ISSN 1932-4553.

69 Alkhateeb, A.; Leus, G.; Heath, R. W. Limited Feedback Hybrid Precoding for Multi-User Millimeter Wave Systems. *IEEE Transactions on Wireless Communications*, v. 14, n. 11, p. 6481–6494, nov. 2015.

70 Sudarshan, P. et al. Channel Statistics-Based RF Pre-Processing with Antenna Selection. *IEEE Transactions on Wireless Communications*, v. 5, n. 12, p. 3501–3511, dez. 2006.

71 Adhikary, A. et al. Joint Spatial Division and Multiplexing - The Large-Scale Array Regime. *IEEE Transactions on Information Theory*, v. 59, n. 10, p. 6441–6463, out. 2013.

72 Nam, J. et al. Joint Spatial Division and Multiplexing: Opportunistic Beamforming, User Grouping and Simplified Downlink Scheduling. *IEEE Journal of Selected Topics in Signal Processing*, v. 8, n. 5, p. 876–890, out. 2014.

73 Xu, Y. et al. User Grouping and Scheduling for Large Scale MIMO Systems with Two-Stage Precoding. In: *International Conference on Communications*. [S.l.: s.n.], 2014. p. 5197–5202. ISSN 1938-1883.

74 Xu, Y.; Yue, G.; Mao, S. User Grouping for Massive MIMO in FDD Systems: New Design Methods and Analysis. *IEEE Access*, v. 2, p. 947–959, ago. 2014. ISSN 2169–3536.

75 3GPP. *Study on Channel Model for Frequencies from 0.5 to 100 GHz* . [S.l.], 2017. V.15.0.0.

76 3GPP. *Study on New Radio Access Technology - Physical Layer Aspects (Release 14)*. [S.l.], 2017.

77 Zaidi, A. A. et al. Waveform and Numerology to Support 5G Services and Requirements. *IEEE Communications Magazine*, v. 54, n. 11, p. 90–98, nov. 2016.

78 Onggosanusi, E. et al. Modular and High-Resolution Channel State Information and Beam Management for 5G New Radio. *IEEE Communications Magazine*, v. 56, n. 3, p. 48–55, mar. 2018. ISSN 0163-6804.

79 Adhikary, A. et al. Joint Spatial Division and Multiplexing for mm-Wave Channels. *IEEE Journal on Selected Areas in Communications*, v. 32, n. 6, p. 1239–1255, jun. 2014. ISSN 0733-8716.

80 Arthur, D.; Vassilvitskii, S. k-means++: The Advantages of Careful Seeding. In: *ACM-SIAM Symposium on Discrete Algorithms*. [S.l.: s.n.], 2007. p. 1027–1035. ISBN 978-0-898716-24-5.

81 Dunn, J. C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, v. 3, p. 32–57, nov. 1973.

82 Cover, T. M.; Thomas, J. A. *Elements of Information Theory*. 2nd. [S.l.]: Wiley-Interscience, 2006. ISBN 0-471-06259-6.

83 Jaeckel, S.; Raschkowski, L.; Thiele, L. QuaDRiGa: A 3-D Multi-Cell Channel Model With Time Evolution for Enabling Virtual Field Trials. *IEEE Transactions on Antennas and Propagation*, v. 62, n. 6, p. 3242–3256, jun. 2014.

84 Jaeckel, S. et al. *Quasi Deterministic Radio Channel Generator - User Manual and Documentation*. Fraunhofer Heinrich Hertz Institute Wireless Communications and Networks Einsteinufer 37, 10587 Berlin, Germany, 2017. Disponível em: <http://www.quadriga-channel-model.de>.

85 Galindo-Serrano, A.; Giupponi, L. Distributed Q-Learning for Interference Control in OFDMA-Based Femtocell Networks. In: *IEEE Vehicular Technology Conference*. [S.l.: s.n.], 2010. p. 1–5. ISSN 1550-2252.

86 Galindo-Serrano, A.; Giupponi, L. Distributed Q-Learning for Aggregated Interference Control in Cognitive Radio Networks. *IEEE Transactions on Vehicular Technology*, v. 59, n. 4, p. 1823–1834, maio 2010. ISSN 1939-9359.

87 Cordeiro, C.; Challapali, K.; Birru, D. IEEE 802.22: An Introduction to the First Wireless Standard based on Cognitive Radios. *Journal of Communications*, v. 01, n. 01, abr. 2006.

88 Dzulkifly, S. et al. Decentralized Q-Learning for Uplink Power Control. In: *IEEE International Workshop on Computer Aided Modelling and Design of Communication Links and Networks*. [S.l.: s.n.], 2015. p. 54–58.

89 Deb, S.; Monogioudis, P. Learning-Based Uplink Interference Management in 4G LTE Cellular Systems. *IEEE/ACM Transactions on Networking*, v. 23, n. 2, p. 398–411, abr. 2015. ISSN 1558-2566.

90 Kaleem, Z.; Ahmad, A.; Rehmani, M. H. Neighbors' Interference Situation-Aware Power Control Scheme for Dense 5G Mobile Communication System. *Telecommunication Systems*, v. 67, n. 3, p. 443–450, mar. 2018. ISSN 1572-9451.

91 Li, X. et al. Intelligent Power Control for Spectrum Sharing in Cognitive Radios: A Deep Reinforcement Learning Approach. *IEEE Access*, v. 6, p. 25463–25473, out. 2018. ISSN 2169-3536.

92 Mnih, V. et al. Human-Level Control Through Deep Reinforcement Learning. *Nature*, v. 518, p. 529–533, fev. 2015.

93 Nasir, Y. S.; Guo, D. Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks. *IEEE Journal on Selected Areas in Communications*, v. 37, n. 10, p. 2239–2250, out. 2019. ISSN 1558-0008.

94 C. Zhang and P. Patras and H. Haddadi. Deep Learning in Mobile and Wireless Networking: A Survey. *IEEE Communications Surveys Tutorials*, v. 21, n. 3, p. 2224–2287, 2019. ISSN 2373-745X.

95 Luong, N. C. et al. Applications of Deep Reinforcement Learning in Communications and Networking: A Survey. *IEEE Communications Surveys Tutorials*, v. 21, n. 4, p. 3133–3174, 2019. ISSN 2373-745X.

96 Ahmed, K. I.; Tabassum, H.; Hossain, E. Deep Learning for Radio Resource Allocation in Multi-Cell Networks. *IEEE Network*, v. 33, n. 6, p. 188–195, nov. 2019. ISSN 1558-156X.

97 3GPP. *General Description*. [S.l.], 2018. V15.6.0.

98 3GPP. *Services Provided by the Physical Layer*. [S.l.], 2020.

99 3GPP. *Physical Channels and Modulation*. [S.l.], 2020. V15.8.0.

100 3GPP. *Multiplexing and Channel Coding*. [S.l.], 2020. V15.8.0.

101 3GPP. *Physical Layer Procedures for Control*. [S.l.], 2020. V15.8.0.

102 3GPP. *Physical Layer Procedures for Data*. [S.l.], 2020. V15.8.0.

103 3GPP. *Physical Layer Measurements*. [S.l.], 2020. V15.6.0.

104 Alkhateeb, A. et al. Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems. *IEEE Journal of Selected Topics in Signal Processing*, v. 8, n. 5, p. 831–846, 2014. ISSN 1932-4553, 1941-0484.

105 Giordani, M. et al. A Tutorial on Beam Management for 3GPP NR at mmWave Frequencies. *IEEE Communications Surveys Tutorials*, v. 21, n. 1, p. 173–196, 2019. ISSN 2373-745X.

106 Dahlman, E.; Parkvall, S.; Skold, J. *4G, LTE-Advanced Pro and The Road to 5G*. 3rd. ed. [S.l.]: Academic Press, 2016. v. 1. ISBN 978-0-12-804575-6.

107 C. J. C. H. Watkins, and P. Dayan. Technical Note: Q-Learning. *Machine Learning*, v. 8, n. 3, p. 279–292, maio 1992. ISSN 1573-0565.

108 Bellman, R. *Dynamic Programming*. Princeton, NJ: Princeton University Press, 1957.

109 WATKINS, C. J. C. H. *Learning from Delayed Rewards*. Tese (phdthesis) — Kings's College, maio 1989.

110 Busoniu, L.; Babuska, R.; De Schutter, B. Multi-agent Reinforcement Learning: An Overview. In: ____. *Innovations in Multi-Agent Systems and Applications - 1*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 183–221. ISBN 978-3-642-14435-6.

111 Schubert, M.; Boche, H. Solution of the Multiuser Downlink Beamforming Problem with Individual SINR Constraints. *IEEE Transactions on Vehicular Technology*, v. 53, n. 1, p. 18–28, jan. 2004.

112 Khachan, A. M.; Tenenbaum, A. J.; Adve, R. S. Linear Processing for the Downlink in Multiuser MIMO Systems with Multiple Data Streams. In: *IEEE International Conference on Communications*. [S.l.: s.n.], 2006. p. 1–6.

113 C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. [S.l.]: Society for Industrial and Applied Mathematics, 2001.

114 W. Yang and G. Xu. Optimal Downlink Power Assignment for Smart Antenna Systems. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*. [S.l.: s.n.], 1998. v. 6, p. 3337–3340.

115 R.D. Yates and S. Gupta and C. Rose and S. Sohn. Soft Dropping Power Control. In: *IEEE Vehicular Technology Conference*. [S.l.: s.n.], 1997. p. 1–5.

116 Tarcisio Ferreira Maciel. *Suboptimal Resource Allocation for Multi-User MIMO-OFDMA Systems*. Tese (phdthesis) — Technische Universitat Darmstadt, set. 2008.

117 Yuri Victor Lima de Melo. *Power Control and Energy Efficiency Strategies for D2D Communications Underlying Cellular Networks*. Dissertação (mathesis) — Federal University of Ceara, jul. 2015.

118 T. S. Rappaport. *Wireless Communications: Principles and Practice*. [S.l.]: Prentice Hall, 1996. v. 1. ISBN 0-13-375536-3.

119 3GPP. *User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone*. [S.l.], 2018.

120 Cormen, T. et al. *Introduction to Algorithms*. 3rd. ed. [S.l.]: MIT Press, 2009. ISBN 978-0-262-03384-8.

APÊNDICE A – COMPUTATIONAL COMPLEXITY OF CLUSTERING ALGORITHMS

We evaluate the computational complexity of the applied algorithms based on an asymptotic analysis using the Big-O notation [120]. The time complexity evaluation is performed according to the pseudo-codes of the considered algorithms described in Algorithms 1 and 2.

A.1 K-means++ Clustering

According to pseudo-code presented in Algorithm 1, in line 2 it is defined the initialization of the first centroid $\mathbf{c}_1^{(t)}$. The computational complexity of this operation is $O(U)$, since the elements of estimated channel $\hat{\mathbf{h}}_j \in \mathbb{C}^{U\times 1}$ are copied to the centroid vector $\mathbf{c}_1^{(t)} \in \mathbb{C}^{U\times 1}$. The initialization of the other centroids $\mathbf{c}_i^{(t)} \in \{2,\ldots,K\}$ is performed from line 3 to line 6. A loop of $(K-1)$ iterations is defined in line 3. At each iteration, in line 4 is calculated the weighted probability function Υ_j which has complexity $O(JU)$. In the following, line 5 determines the association of the estimated channel $\hat{\mathbf{h}}_j$ of UE j with highest Υ_j to the centroid $\mathbf{c}_i^{(t)}$ which has complexity $O(U)+O(J)$. Therefore, the initialization of centroids has a computational complexity defined as $(K-1)(O(JU)+O(U)+O(J)) = O(KJU)$.

After the initialization of centroids, the algorithm performs iterations composed of a clustering assignment followed by a centroid update until the termination condition is achieved. The group assignment step associates each UE j to cluster $\mathscr{C}_i$ according to (11) and has complexity $O(KJU)$. The centroid update is defined according to (12) as the average of the estimated channels of all UEs that belongs to each cluster and has complexity $O(KJU)$. Assuming a limit of T iterations, the computational complexity of the commands between lines 7 and 17 is defined as $T(O(KJU)+O(KJU)) = O(KJUT)$.

Therefore, the computational complexity of the k-means++ clustering algorithm based on the pseudo-code Algorithm 1 is $O(KJU)+O(KJUT) = O(KJUT)$.

A.2 FCM Clustering

According to pseudo-code presented in Algorithm 2, in line 2 it is defined the initialization of the membership matrix $\mathbf{U}^{(t)}$ with random values in the range $[0,1]$. The computational complexity of this command is $O(KJ)$ since $\mathbf{U}^{(t)} \in \mathbb{R}^{K\times J}$. In the following, the algorithm iteratively determines the cluster centroids and the value of the objective function. The definition of

centroids $\mathbf{c}_i \in \mathbb{C}^{U\times 1}$ is performed according (16) in a loop of K iterations and has a computational complexity of $K(O(JU)+O(J)+O(1)) = O(KJU)+O(KJ) = O(KJU)$.

After the definition of centroids, in line 9 is calculated the objective function according to (15). This operation involves the squared norm which has a complexity $O(U)$ and must be repeated in two loops of J and K iterations, respectively. Therefore, the computational complexity of the operation between lines 7 and 11 is $KJ(O(U)) = O(KJU)$.

In the following, the membership matrix is updated according to (17). The operation defined in line (11) has computational complexity $K(O(U)+O(U)) = O(KU)$ and is repeated KJ due to the loops defined in line 13 and 14. Therefore, the computational complexity associated with this update operation is given by $KJ(O(KU)) = O(K^2JU)$.

Assuming a limit of T iterations, the computational complexity of the commands between lines 3 and 19 is defined as $T(O(K^2JU)+O(KJU)) = O(K^2JUT)$. Therefore, the computational complexity of the FCM clustering algorithm based on the pseudo-code Algorithm 2 is $O(KJU)+O(K^2JUT) = O(K^2JUT)$.

APÊNDICE B – COMPUTATIONAL COMPLEXITY OF POWER CONTROL ALGORITHMS

B.1 Computational Complexity of the RLPC ALgorithm

Line 2 in Algorithm 1 defines the initialization of power levels of all D antenna ports in each cell. These powers are modeled as discrete values and are limited to minimum and maximum values ($P_{d,c}^{\min}$ and $P_{d,c}^{\max}$) determined by Eqs. (3.19) and (3.20), respectively. The number of possible power values is a function of the step size χ among transmit power levels and the power limits. Thus, initialization of power levels requires the sort of D values in a space of possible states (powers) with cardinality given by Eq. (3.23). A loop of C iterations is required to initialize the power levels on all cells. Therefore, the computational complexity related to this line is $O\left(\sum_{c=1}^{C}|\mathscr{S}_c|\right) = O\left(\sum_{c=1}^{C}\left(\frac{P_{d,c}^{\max}-P_{d,c}^{\min}}{\chi}\right)^{D}\right)$.

For simplification purposes, we consider the space of states in all cells with the same size $|\mathscr{S}_c| = \left(\frac{P_{d,c}^{\max}-P_{d,c}^{\min}}{\chi}\right)^{D}$. Then, the computational complexity becomes $O(C|\mathscr{S}_c|)$.

The main loop involves the commands between lines 4 and 16. This loop is executed T times until the stop condition is determined. Line 4 determines the sort of a random number. This operation has complexity $CO(1) = O(C)$, since it is repeated by all agents. In the following, line 5 requires the calculation of the learning rate ε_t according to Eq. (3.22). This operation involves the calculation of scalar values, which has complexity $CO(1) = O(C)$. In one hand, if the condition $(e < \varepsilon_t)$ or $(t = 0)$ is true, it is sorted a random set of actions from the space of action, as defined in line 7, which has computational complexity $O\left(\sum_{c=1}^{C}|\mathscr{A}_c|\right) = O(C\iota^D)$ (ι is the number of TPC commands). It is assumed that all agents have the same space of actions. On the other hand, if the condition $(e < \varepsilon_t)$ or $(t = 0)$ is false, we select an action based on the evaluation of the maximum element of row $Q(s_{t,c},:)$. Therefore, the operation defined in line 9 has computational complexity $CO(|\mathscr{A}_c| = O(C|\mathscr{A}_c)) = O(C\iota^D)$.

Line 11 defines the computation of transmit power update of the antenna ports in all cells. Thus, the computational complexity of this operation is $CDO(1) = O(CD)$. The verification of the antenna port power limits in line 12 has computational complexity $CO(D) = O(CD)$ and the execution of transmit power updates in line 13 has complexity $CO(D) = O(CD)$. The calculation of the reward associated to the performed actions has computational complexity $CO(D) = O(CD)$. Finally, the update of the matrix Q defined in line 15 has computational

complexity $CO(|\mathscr{S}_c||\mathscr{A}_c|) = O(C|\mathscr{S}_c||\mathscr{A}_c|)$.

The computational complexity of the proposed RL-based uplink PC framework based on the analyzed pseudo-code is $O(C|\mathscr{S}_c|) + T(O(C) + O(|\mathscr{A}_c|) + O(CD) + O(|\mathscr{S}_c||\mathscr{A}_c|)) \rightarrow O(TC|\mathscr{S}_c||\mathscr{A}_c|)$.

B.2 Computational Complexity of the OSPC ALgorithm

Line 2 in Algorithm 2 defines the SINR targets of all D antenna ports in all C cells. This command has computational complexity $CD \cdot O(1) = O(CD)$. The next command at line 4 specifies the extended coupling matrix of each cell, which requires the operations defined at Eq. (3.27) and Eq.(3.30). The determination of the coupling matrix Ψ_c at Eq.(3.30) requires the manipulation of beamforming vectors and channel matrices, namely $\overline{\mathbf{w}}_{b_r} \in \mathbb{C}^{U\times 1}$, $\bar{\mathbf{f}}_{b_t} \in \mathbb{C}^{V\times 1}$, and $\mathbf{H} \in \mathbb{C}^{U\times V}$, respectively. The operation is repeated D^2 times since it is created a squared matrix with dimension $D \times D$. Thus, the computational complexity is $D^2 \cdot (O(U) + O(UV) + O(V)) \rightarrow O(D^2U) + O(D^2VU) + O(D^2V) \rightarrow O(D^2UV)$, where V and U are the number of receive and transmit antennas, respectively.

Besides the coupling matrix Ψ_c, the definition of the extended coupling matrix Λ_c also requires the specification of the auxiliary matrix $\mathfrak{D}_c$. This matrix requires the computational of the norm of the vector $\overline{\mathbf{w}}_{b_r}^H \mathbf{H} \bar{\mathbf{f}}_{b_t}$ to determine the D elements of its main diagonal. Therefore, it has computational complexity $D \cdot (O(U) + O(VU) + O(V)) \rightarrow O(DU) + O(DVU) + O(DV) \rightarrow O(DVU)$.

The extended coupling matrix Λ_c is composed by different blocks of matrices. The determination of $\mathfrak{D}_c \Psi_c^T$ has computational complexity $O(D^2)$, since it requires the multiplication of matrices of dimension $D \times D$. The determination of $\frac{1}{P_{\text{PCMAX}}} \mathbf{1}^T \mathfrak{D}_c \Psi_c^T$ has computational complexity $O(D^2)$ since it involves the multiplication of arrays with dimensions $D \times 1$ and $D \times D$. The determination of $\mathfrak{D}_c \sigma$ also has computational complexity $O(D^2)$ since it involves again the multiplication of arrays with dimensions $D \times 1$ and $D \times D$. The determination of $\frac{1}{P_{\text{PCMAX}}} \mathbf{1}^T \mathfrak{D}_c \sigma$ has computational complexity $O(D^2)$ since it involves the multiplication of arrays with dimensions $D \times 1$ and $D \times D$. The composition of the extended coupling matrix has computational complexity $O(D^2) + O(D^2) + O(D^2) + O(D^2) \rightarrow O(D^2)$.

Thus, the definition of the extended coupling matrix has computational complexity $T \cdot C \cdot O(D^2VU) + O(DVU) + O(D^2) \rightarrow O(TCD^2VU) + O(TCDVU) + O(CTD^2) \rightarrow O(TCD^2VU)$, where T is the total number of iterations.

In addition, line 5 requires the eigendecomposition of the extended coupling matrix Λ_c, so the computational complexity is $T \cdot C \cdot O((D+1)^3) = O(TCD^3)$. The verification of the power limits at line 6 has computational complexity $T \cdot C \cdot D \cdot O(1) = O(TCD)$. The implementation of the power commands has computational complexity $T \cdot C \cdot D \cdot O(1) = O(TCD)$.

Thus, the total number of operations in big O notation is $O(CD) + O(TCD^2VU) + O(TCD^3) + O(CTD) \rightarrow O(TCD^2VU) + O(TCD^3)$.

B.3 Computational Complexity of the SDPC ALgorithm

Line 2 in Algorithm 3 defines the initial transmit power of all D antenna ports in all C cells. As observed previously, this command has computational complexity $CD \cdot O(1) = O(CD)$. The main loop involves the commands between lines 4 and 8. This loop is executed T times until the stop condition is determined. Line 4 calculates the target SINR $\Gamma_{c,d}^{\text{target}}$ according to Eq. (3.32). This operation has computational complexity $C \cdot D \cdot O(1) = O(CD)$.

Line 5 determines the update of the transmit power in all antenna ports of each cell according to Eq. (3.33). This command has computational complexity $C \cdot D \cdot O(1) = O(CD)$. Line 7 defines the execution of the transmit power updates, which has computational complexity $C \cdot D \cdot O(1) = O(CD)$.

Therefore, the computational complexity of the soft dropping power control based on the previous pseudo-code is $O(CD) + T\left(O(CD) + O(CD) + O(CD)\right) \rightarrow O(TCD)$.