RATE SPLITTING FOR INTERFERENCE CHANNELS WITH DEEP REINFORCEMENT LEARNING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

OSMAN NURI IRKIÇATAL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

JANUARY 2024

Approval of the thesis:

RATE SPLITTING FOR INTERFERENCE CHANNELS WITH DEEP REINFORCEMENT LEARNING

submitted by **OSMAN NURI IRKIÇATAL** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** ________________

Prof. Dr. İlkay Ulusoy
Head of Department, **Electrical and Electronics Engineering** ________________

Assoc. Prof. Dr. Ayşe Melda Yüksel Turgut
Supervisor, **Electrical and Electronics Engineering** ________________

Assist. Prof. Dr. Elif Tuğçe Ceran Arslan
Co-supervisor, **Electrical and Electronics Engineering** ________________

Examining Committee Members:

Prof. Dr. Abdullah Aydın Alatan
Electrical and Electronics Engineering, METU ________________

Assoc. Prof. Dr. Ayşe Melda Yüksel Turgut
Electrical and Electronics Engineering, METU ________________

Assist. Prof. Dr. Özlem Tuğfe Demir
Electrical and Electronics Engineering, TOBB ETU ________________

Date: 24.01.2024

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Osman Nuri Irkıçatal

Signature :

ABSTRACT

RATE SPLITTING FOR INTERFERENCE CHANNELS WITH DEEP REINFORCEMENT LEARNING

Irkıçatal, Osman Nuri
M.S., Department of Electrical and Electronics Engineering
Supervisor: Assoc. Prof. Dr. Ayşe Melda Yüksel Turgut
Co-Supervisor: Assist. Prof. Dr. Elif Tuğçe Ceran Arslan

January 2024, 82 pages

In recent advancements within communication systems, the rate-splitting multiple access (RSMA) technique has emerged as a crucial strategy to address interference, a persistent challenge in modern communication systems. This study examines the detailed application of precoding methodologies within RSMA, focusing on the complex environment of multiple-antenna interference channels and leveraging the capabilities of deep reinforcement learning. The primary objective is to optimize precoders and allocate transmit power for both common and private data streams, requiring a nuanced approach involving multiple decision-makers within a continuous action space. To address this challenge, the study proposes the utilization of a multi-agent deep deterministic policy gradient (MADDPG) framework. Within this framework, decentralized agents operate with partial observability but collectively learn from a centralized critic, navigating a multi-dimensional continuous policy space to optimize actions. Simulation outcomes highlight the effectiveness of the proposed rate-splitting method, achieving the information-theoretical upper bound for the sum rate in the single-antenna scenario. Even in multiple-antenna settings, its performance closely approaches this theoretical limit, outperforming benchmarks set by other techniques such as MADDPG without rate-splitting, maximal ratio transmission, zero-forcing, and leakage-based precoding methods. These compelling results emphasize the promising potential of this deep reinforcement learning-driven RSMA approach in communication systems by substantially mitigating interference and optimizing transmission rates and overall system performance.

Keywords: Deep reinforcement learning, Interference channels, Multi agent deep deterministic policy gradient (MADDPG), Rate-splitting multiple access (RSMA).

ÖZ

GİRİŞİM KANALLARI İÇİN DERİN PEKİŞTİRMELİ ÖĞRENME İLE HIZ BÖLÜMÜ

Irkıçatal, Osman Nuri
Yüksek Lisans, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. Ayşe Melda Yüksel Turgut
Ortak Tez Yöneticisi: Dr. Öğr. Üyesi Elif Tuğçe Ceran Arslan

Ocak 2024 , 82 sayfa

İletişim sistemlerindeki son gelişmeler ile, hız bölmeli çoklu erişim tekniği (RSMA) tekniği, çağdaş iletişim sistemlerindeki süregelen bir sorun olan girişimi ele almak için önemli bir strateji olarak ortaya çıkmıştır. Bu çalışma, RSMA içinde ön kodlama yöntemlerinin detaylı bir şekilde uygulanmasını inceleyerek, özellikle çok antenli girişim kanallarının karmaşık alanına odaklanarak, derin pekiştirmeli öğrenmenin yeteneklerinden yararlanmayı amaçlamaktadır. Temel amaç, ortak ve özel olarak adlandırılan her iki tür veri akışı için ön kodlayıcıları ve iletim gücünü optimize etmektir ve bu da sürekli bir eylem alanında çoklu karar vericiyi içeren detaylı bir yaklaşım gerektirir. Bu zorluğu ele almak için çalışma, çoklu ajan derin belirli politika gradyanı (MADDPG) çerçevesinin kullanılmasını önermektedir. Bu çerçeve içinde, dağıtılmış ajanlar kısmi gözlem yeteneğiyle çalışır ancak merkezi bir eleştirmenden birlikte öğrenir, çok boyutlu bir sürekli politika alanında gezinerek eylemleri optimize eder. Simülasyon sonuçları, önerilen hız-bölme yönteminin etkinliğini vurgular ve tek anten senaryosunda toplam hız için bilgi teorik üst sınırına ulaştığını gösterir. Çoklu antenli ortamlarda bile performansı bu teorik üst sınır ile yakındır ve ayrıca hız-bölümü kullanmadan MADDPG, maksimum oranlı iletim, sıfıra zorlama ve sızıntı tabanlı ön kodlama gibi diğer tekniklerin belirlediği referansları aşar. Bu sonuçlar, bu derin öğrenme destekli RSMA yaklaşımının, girişimi önemli ölçüde azaltarak iletişim sistemlerinde iletim hızlarını ve genel sistem performansını iyileştirme potansiyelini vurgular.

Anahtar Kelimeler: Derin pekiştirmeli öğrenme, Çok ajanlı derin belirgin politika gradyanı, Hız bölmeli çoklu erişim (RSMA).

To my family...

ACKNOWLEDGMENTS

I would like to begin by expressing my profound gratitude for Dr. Ayşe Melda Yüksel Turgut, my supervisor, whose unwavering support has been fundamental since the study's inception. Her dedicated time and sincere interest in my work have significantly influenced this study. Without her continual guidance and insightful advice, this research would not have been viable. I feel incredibly fortunate to have been under her mentorship during my graduate studies.

I also want to extend heartfelt appreciation to Dr. Elif Tuğçe Ceran Arslan, my co-advisor, whose exceptional collaboration and immense assistance have been indispensable throughout my research. Her willingness to share extensive knowledge and experiences has significantly advanced my studies. I am deeply grateful for her continual guidance and support in all aspects of my work.

My parents, Hatice Irkıçatal and Cemalattin Irkıçatal, along with my sister, Erva Irkıçatal, and brother, Mehmet Emin Irkıçatal have been the unwavering sources of support on my journey. Their dedication and unwavering belief in my academic pursuits have been the driving force behind my achievements from the very beginning.

I am thankful to all my colleagues whose contributions, too numerous to mention, have been a constant presence throughout this work.

Lastly, I extend my utmost gratitude to my beloved wife, Beyza Nur Irkıçatal. Throughout this journey, your exceptional patience and unwavering support during the most challenging times have been priceless. Your enduring love has been the most precious gift, and words cannot fully convey how deserving you are of the very best in everything.

TABLE OF CONTENTS

CHAPTERS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

BC	Broadcast channel
BS	Base station
CSI	Channel state information
CSIT	Channel state information at transmitter
DDPG	Deep deterministic policy gradient
DRL	Deep reinforcement learning
DQL	Deep Q-learning
DL	Deep learning
DQN	Deep Q-network
D2D	Device-to-device
i.i.d	Independent and identically distributed
IFC	Interference channel
MADDPG	Multi-agent deep deterministic policy gradient
MADQN	Multi-agent deep Q-network
MADDQN	Multi-agent double deep Q-network
MAD3QN	Multi-agent dueling double deep Q-network
MADRL	Multi-agent deep reinforcement learning
MDP	Markov decision process
MIMO	Multiple-input multiple-output
MISO	Multiple-input single output
ML	Machine learning
MRT	Maximal ratio combining
MG	Markov game
OFDM	Orthogonal frequency-division multiplexing

PPO	Proximal policy optimization
RSMA	Rate splitting multiple access
RS	Rate splitting
RIS	Reconfigurable intelligent surface
SISO	Single input single output
SIC	Successive interference cancellation
SLNR	Signal-to-leakage-and-noise ratio
SNR	Signal-to-noise ratio
SINR	Signal-to-interference and noise ratio
TRPO	Trust region policy optimization
TD	Temporal difference
UAV	Unmanned aerial vehicle
UE	User equipment
ZF	Zero forcing
3D	Three dimensional

CHAPTER 1

INTRODUCTION

1.1 Motivation

The imperative for high data rates and reliable connectivity in 6G networks is a direct response to the escalating demands of data-intensive applications and services [7]. Building upon the foundation laid by 5G, the evolution to 6G is fueled by the exigencies of cutting-edge technologies such as augmented and virtual reality, the internet of things (IoT), high-definition video streaming, remote healthcare, and autonomous systems [8]. These applications mandate not only accelerated data transfer but also uninterrupted and reliable connectivity, necessitating advancements beyond the capabilities of existing networks. One of the principal impediments to achieving such connectivity at higher data rates is interference. As 6G networks envisage ultra-densification, characterized by an unprecedented deployment of small cells and connected devices in close proximity, interference becomes a paramount challenge. The physical closeness of transmitters and devices amplifies interference accumulation, compromising the quality of signals and impeding the seamless transmission of data. This ultra-densification, while enhancing network capacity, concurrently intensifies the complexities associated with interference management [9]. To counteract the deleterious effects of interference in this context, state-of-the-art interference management techniques are indispensable. These techniques encompass advanced signal processing methodologies, spectrum-sharing strategies, cognitive radio systems, artificial intelligence-driven algorithms, and innovative modulation schemes [10]. They collectively constitute a robust toolkit for mitigating interference and optimizing spectrum utilization in ultra-dense 6G networks. In essence, the transition from 5G to 6G is propelled by the need to cater to a diverse array of data-intensive applications. The evolution involves not only upgrading existing technologies but also confronting the challenges posed by ultra-densification. This paradigm shift is essential to meet the burgeoning demands of a connected world [11], ensuring that 6G networks deliver the high data rates and reliable connectivity required for the seamless functioning of futuristic applications.

Rate-splitting multiple access (RSMA) presents a flexible and powerful tool in modern communication systems, especially within the realm of interference management. Its usage extends across a spectrum of scenarios where interference poses a significant challenge to reliable and high-rate data transmission. One prominent application lies in addressing the interference challenges inherent in ultra-dense 6G networks [12]. In these densely populated network environments, where transmitters and devices are intricately interwoven in close proximity, interference becomes a critical bottleneck. RSMA's ability to divide transmitted messages into distinct layers—common and private—proves instrumental in navigating these interference networks. By allocating portions of the message to the common layer, which is decoded by all receivers before individual private messages as illustrated in Figure 1.1, successive interference cancellation (SIC) enables a stepwise decoding process. This approach aims to alleviate interference by sequentially eliminating signals from already decoded layers, allowing subsequent receivers to progressively recover their intended information. In multi-user systems, the RSMA network incorporates a decoding order due to SIC. The decoding order plays a crucial role in determining the achievable rate in the network. As users decode and cancel interference successively, the order in which this process occurs affects the overall rate of data transmission in the system. This dynamic feature allows the RSMA network to adapt and optimize its performance based on the specific decoding sequence, ensuring efficient communication in the presence of interference. RSMA ensures a balanced approach to reducing interference without sacrificing the reliability of the intended data streams. This makes RSMA a valuable asset for future communication systems where ultra-densification leads to interference accumulation, enabling these networks to sustain high data rates and reliable connectivity despite challenging interference scenarios. Moreover, RSMA's applicability extends beyond network densification—it finds utility in various contexts, including IoT communications [13], wireless sensor networks, and scenarios demanding secure and efficient data transmission in interference-prone environments. Its flexible approach to handling interference makes RSMA a useful and promising technique for creating smooth, high-rate, and interference-resistant communication systems in various applications and network environments.

However, solving optimization problems for RSMA presents significant challenges due to their inherent non-convex nature. [14] These intricacies become particularly pronounced in scenarios marked by a multitude of users or an expanded array of rate-splitting layers. The crux of the challenge lies in the non-convex nature of these optimization problems. This non-convexity translates into a maze of potential rate-splitting configurations that grows exponentially as the number of users or layers increases. Consequently, the search space for finding the optimal solution expands dramatically, rendering the task of pinpointing the most efficient rate-splitting combination a computationally arduous endeavor. This computational burden escalates significantly, demanding sophisticated algorithms and robust computational resources like MATLAB CVX tools [15] to navigate the exponentially expansive solution space for the optimization problem [16] effectively. Devising strategies to tackle this complexity is paramount, as it directly impacts the practical implementation and efficiency of RSMA in real-world communication systems.

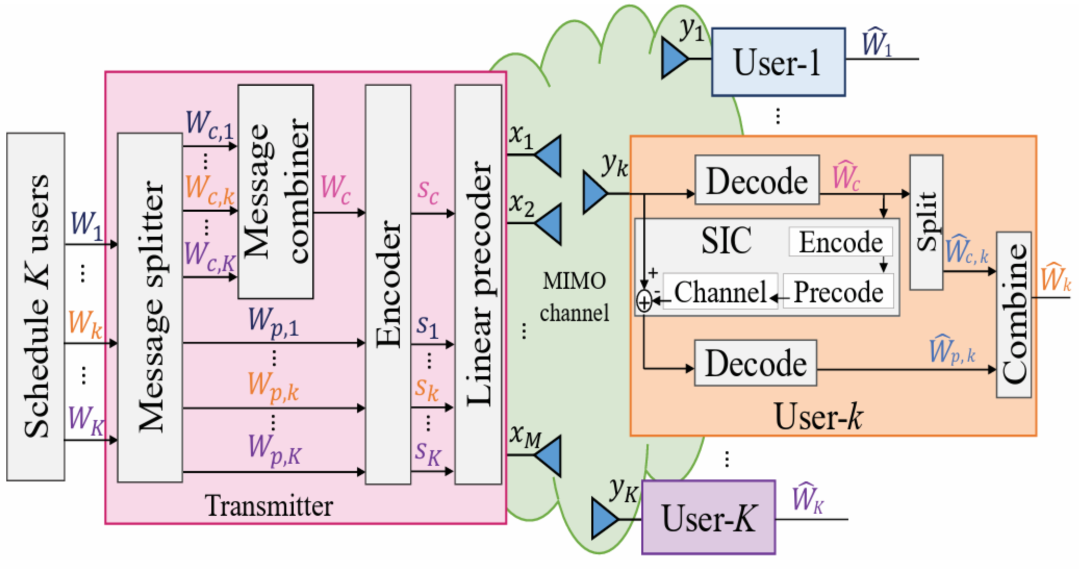

Figure 1.1: Illustration of a typical RSMA scenario [1].

Deep Reinforcement Learning (DRL) stands out as a crucial tool in grappling with the intricate challenges embedded within Rate-Splitting Multiple Access (RSMA), particularly concerning optimization. RSMA's complexity, intertwined with its vast solution space, necessitates sophisticated navigation techniques. DRL, with its capacity to glean optimal strategies through interactions with environments, presents a promising avenue to tackle these intricacies. Within the framework of a Markov decision process (MDP), DRL adapts and learns within RSMA's intricate solution space, decoding complex patterns. This empowers DRL algorithms to gradually optimize rate-splitting strategies in dynamic and convoluted interference scenarios, ultimately augmenting communication systems' efficiency. DRL's utilization not only unravels RSMA's complexities but also propels the evolution of more adaptive and efficient interference management techniques in contemporary communication networks. At its core, DRL offers a robust framework, combinating deep learning and reinforcement learning techniques to learn optimal behaviors by interacting with an environment to achieve specific objectives. With its components—agent, environment, and reward mechanism [17]—DRL systems aim to maximize cumulative rewards over time by learning effective actions in varied environmental states as illustrated in the Figure 1.2. Leveraging deep neural networks, DRL showcases adaptability in handling high-dimensional spaces, making it ideal for scenarios with vast action spaces and dynamic, uncertain environments. This adaptability positions DRL as an invaluable tool in addressing optimization challenges within communication networks, notably in RSMA, where intricate navigation through expansive solution spaces is paramount.

Reinforcement learning, especially DRL, is increasingly employed in solving communication problems in 5G and 6G due to its adaptability to dynamic and complex environments. Unlike traditional machine learning solutions, RL learns from interactions with the environment, making it suitable for scenarios where the system's behavior is not fully known. RL excels in handling uncertainties and incomplete information, crucial features in communication networks. It efficiently optimizes resource allocation, addressing challenges such as power control, precoder optimization and bandwidth allocation [18]. With decentralized decision-making capabilities, RL aligns well with the distributed nature of communication systems. Additionally, RL handles non-stationary and non-linear dynamics, making it effective in capturing the complexities of communication channels. It encourages exploration of novel strate-

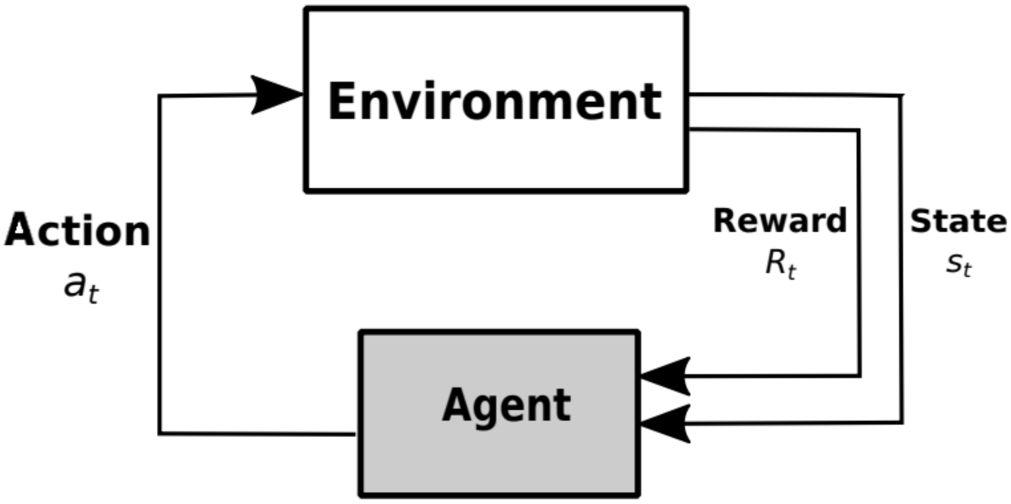

Figure 1.2: The basic reinforcement learning scenario [2].

gies, valuable in the rapidly evolving landscape of 6G. Despite the availability of other machine learning approaches, RL's unique attributes make it a valuable tool for optimizing communication networks for performance, efficiency, and adaptability.

1.2 Related Literature

Interference channels have been a subject of significant research and study within the field of information theory and communications. Over the years, various seminal works and literature have contributed to understanding and addressing interference in communication systems. Shannon's pioneering work in the 1940s and 1950s, particularly in his landmark paper "A Mathematical Theory of Communication," [19] initiated the study of information theory and channel capacity, which forms the theoretical basis for interference channels. Significant contributions in the 1970s and 1980s expanded the understanding of interference channels. While the work [20] by Thomas M. Cover and Abbas El Gamal, delved into the theoretical aspects of multiple access channels and interference, providing fundamental insights into channel capacity and achievable rates in the presence of interference, the work [21] by T. Han and K. Kobayashi, discusses novel strategies or methodologies that enable the transmission of data over interference channels while achieving rates that were previously unattainable using conventional approaches. While the precise capacity region of interference channels remains unknown, seminal works such as [5] and [22] have provided valuable insights. These studies propose upper bounds within approximately one bit of the true capacity region, specifically for single-antenna and multiple-antenna interference channels, respectively. Despite the complexity of determining the exact capacity region, these upper bounds offer crucial benchmarks and indications, shedding light on the potential limits of information transmission in both single and multiple antenna interference scenarios. Such findings contribute significantly to our understanding, guiding further exploration and development of communication strategies in environments with high interference. Another study [4] focuses on Gaussian channels, mathematically modeling noise in communication systems, and after normalization, presents the specific mathematical form of the Gaussian interference channel. The goal is to define theoretical limits on data transmission rates in Gaussian IFC, offering insights into achievable communication rates under strong interference conditions.

Rate-splitting was first proposed by Han and Kobayashi [21] as a novel technique for interference channel. They introduced the concept of rate-splitting as an innovative technique tailored specifically for interference channels. Rate-splitting involves the segmentation of transmitted messages into two distinct layers: a common layer shared by multiple users and a private layer specific to individual receivers. This pioneering idea aimed to address the challenges posed by interference in communication channels by enabling a balance between treating interference as noise and decoding all messages efficiently [12]. By ensuring that all receivers first decode the common messages before extracting private information, rate-splitting offered a novel strategy to optimize communication in the presence of interference. Since its proposal, rate-splitting has remained a cornerstone of study and exploration in information theory, serving as a foundation for numerous advancements in interference management techniques within modern communication networks [23]. Researchers have its applications, complexities, performance bounds, and practical implementations, highlighting its significance [24] in enhancing communication efficiency in environments with high interference.

The application of Deep Reinforcement Learning (DRL) in communication systems, particularly in the context of interference channels and rate-splitting, showcases an evolving landscape of innovative strategies. Leveraging the Markov Decision Process (MDP) framework, researchers have delved into formulating intricate problems like power allocation within communication networks. Introducing a DRL and MADRL scheme rooted in the multi-agent Deep Deterministic Policy Gradient (MADDPG) algorithm which is explained in Appendix A, recent studies, such as [25], have focused on identifying optimal precoders for transmitters. Prior research, exemplified by [3] and [26], has explored precoding challenges in multi-cell multi-user interference channels, though without integrating rate-splitting techniques.

However, addressing the complexities of interference channels necessitates more detailed approaches. Studies, including [27] and [28], have employed DRL methodologies, specifically the proximal policy optimization (PPO) algorithm, to tackle power allocation problems in single-cell communications while incorporating rate-splitting strategies. Other than that, the study [29] investigates how to manage resources and address interference among sensing, energy harvesting, and communication functionalities by using trust region policy optimization (TRPO). Furthermore, recent advancements, such as [30], have utilized Q-learning, namely DQN, techniques to maximize the signal-to-interference-noise ratio (SINR) in multi-access orthogonal frequency division multiplexing (OFDM) networks, illustrating the potential of DRL in enhancing network performance without employing rate-splitting.

Moreover, in the realm of three-dimensional (3D) UAV-based networks, investigations like [31] have highlighted the efficacy of DRL in mitigating interference. Despite these strides, coordinating transmission schemes in interference channels, particularly addressing joint precoding and power allocation problems, remains an intricate challenge. Other research, such as [32] and [33] delve into energy-efficient power and rate allocation. The study [32] centers on managing interference in advanced 6G networks, examining a model within a single-cell single-antenna RSMA network's downlink. In contrast, [33] focuses on wireless device-to-device (D2D) underlaid cellular networks, specifically addressing scenarios where multiple D2D pairs engage in simultaneous wireless information and power transfer.

The studies [34], [35] and [23] explore the integration of RSMA and reconfigurable intelligent surface (RIS) techniques in next-generation networks. They focus on max-

Table 1.1: Some existing works that use a DRL in communication

Papers	DRL Method	Investigated Channel	Optimization Problem	Inclusion of RSMA	Cell/Antenna Configuration
[25]	MADQN, MADDQN, MAD3QN	Broadcast channel	Pilot contamination	✗	Cell-free massive MIMO
[3, 26]	MADDPG/DDPG	Interference channel	Precoder	✗	Multi-user Multi-cell single-cell MISO
[27, 28]	PPO	Broadcast channel	Precoder and power allocation coefficient	✓	Multi-user single-cell SISO
[29]	TRPO	Broadcast channel	Energy harvesting, sensing and communication capabilities	✓	Multi-user single-cell MISO
[30]	DQL	Interference channel	Beamformer and power	✗	Multi-user multi-cell MISO
[31]	DRL	Interference channel	Beamformer and power	✗	Single-user multi-cell MISO
[32]	DQL	Broadcast channel	Energy efficiency	✗	Multi-user single-cell MISO
[33]	MADQL	Broadcast channel	Energy efficiency and power	✗	Multi-user single-cell MISO
[23, 34, 35]	DL/DDPG	Broadcast channel	Precoder and power	✓/✗	Multi-user single-cell MIMO
This work	MADDPG	Interference channel	Precoder and power allocation coefficient	✓	multi-user SISO, multi-user MISO, multi-user MIMO

imizing sum rates in RSMA IoT networks, improving robustness against imperfect channel information in tera-hertz multi-user MIMO systems, and optimizing resource efficiency in cellular networks through joint base stations (BS) and RIS design. Using DRL, hybrid data-model driven schemes, and novel optimization frameworks, these works aim to enhance spectral efficiency while balancing various network metrics. The summary of the existing works in the literature is presented in Table 1.1.

The main idea is to consider the coordination challenges as situations where multiple agents work together within interference channels. These agents must learn and adapt, coordinating their actions with each other to improve communication performance. Remarkably, the proposed approach, a pioneering initiative, introduces a novel application of DRL. It tackles the complex power allocation and precoder design problems within interference channels, specifically integrating rate-splitting techniques. This novel approach stands as a significant advancement, marking the first instance of employing DRL to resolve the power allocation and precoder problem for interference channels with rate-splitting challenges, paving the way for more efficient and adaptive interference management strategies.

1.3 Contributions and Novelties

Existing research on RSMA and the broadcast channel (BC) within wireless communication networks showcases diverse approaches, some integrating learning algorithms while others rely on conventional optimization methods. These studies focus on enhancing efficiency and resource allocation but often overlook the learning aspects associated with decoding orders and channel estimation errors.

Numerous investigations [15], [36] delve into RSMA and BC, addressing resource allocation, power control, and spectral efficiency without employing learning algorithms. Instead, they utilize traditional optimization techniques, heuristic approaches, or game theory principles to optimize system performance.

Conversely, a subset of research [27], [28], [29] within RSMA and BC domains leverages learning algorithms, particularly DRL or other machine learning (ML) methods. These studies explore optimal resource allocation, power distribution, and rate assignment through adaptable AI-based approaches, contributing to improved performance in communication networks.

However, most existing works in RSMA and BC neglect aspects such as learning decoding orders and addressing channel estimation errors. These elements, critical in optimizing network efficiency and reliability, are yet to be thoroughly examined within the context of RSMA and BC paradigms. Therefore, while current research focuses on resource allocation and performance enhancement, there remains untapped potential in exploring learning mechanisms for decoding and mitigating channel estimation errors in inteference channels. The key contributions of this work are summarized as follows:

- The work introduces a novel MADDPG algorithm customized for optimizing precoding and power allocation coefficients in multiple antenna interference channels employing rate-splitting strategies.

- The algorithm's framework allows for centralized learning while enabling decentralized execution, which contributes a decentralized and scalable framework for interference management without the need for constant coordination from a central entity.

- The work compares the performance of the proposed MADDPG algorithm against existing baseline schemes and upper bounds. It showcases the superiority of MADDPG with rate splitting, demonstrating optimal outcomes particularly in scenarios with multiple antennas at base stations and single antenna cases.

- The work investigates the impact of channel estimation errors, and incorporate optimal decoding order selection for common and private messages into the learning algorithm. These steps enhance the algorithm's robustness and broaden its scope of application.

- The study provides a comprehensive analysis of RSMA, outlining its intricacies and challenges, particularly in managing interference and decoding strategies for multiple messages. It introduces and explores the utilization of DRL within RSMA networks, showcasing its potential to optimize resource allocation, address interference, and manage complex decoding strategies efficiently.

1.4 Notation and Outline

In this section, we introduce the notations crucial for mathematical operations used throughout this thesis. To differentiate quantities, lowercase letters (e.g., w) denote scalars, lowercase boldface letters (e.g., $\mathbf{w}$) denote vectors, while uppercase boldface letters (e.g., $\mathbf{W}$) represent matrices. The complex conjugate of scalar w is denoted as w^*, and its magnitude is denoted as $|w|$. Additionally, the determinant, inverse, transpose, and Hermitian of a matrix $\mathbf{W}$ are represented by $|\mathbf{W}|$, $\mathbf{W}^{-1}$, $\mathbf{W}^T$, and $\mathbf{W}^H$, respectively. To specify the entry in the i^{th} row and j^{th} column of a a matrix $\mathbf{W}$, we use the notation $[\mathbf{W}](i,j)$. Moreover, $[\mathbf{W}](i,:)$ and $[\mathbf{W}](:,j)$ denote the i^{th} row and the j^{th} column of $\mathbf{W}$, respectively. Similarly, $[\mathbf{w}](i)$ signifies the i^{th} element of a vector $\mathbf{w}$. For vectors, $||\mathbf{w}||$ represents the Euclidean norm of $\mathbf{w}$. Complex Gaussian random variables with a mean of μ and variance of σ^2 are denoted as $\mathcal{CN}\big(\mu,\sigma^2\big)$, while $\mathcal{CN}\big(\mathbf{w},\mathbf{R}\big)$ symbolizes complex Gaussian random vectors with a mean of $\mathbf{w}$ and a covariance matrix $\mathbf{R}$. Expectation is depicted as $\mathbb{E}(\cdot)$, and the trace operator is denoted by $\mathrm{Tr}\left(\cdot\right)$. For simplicity, network parameters of DRL will be indicated as $Q^{\nu}_{\theta_i}(s, a_1, a_2)$, where θ_i denotes parameterization by another network, specifically the critic network for each agent, and ν represents the policy under consideration.

The remainder of this work follows a structured progression. Chapter 2 delineates the RSMA system model, explicating its key components and architectural underpinnings. In Chapter 3, the focus is directed towards the specialized MADDPG algorithm tailored explicitly for RSMA, detailing its adaptations and intricacies within this context. Chapter 4 introduces benchmark schemes utilized for comparative analysis against the proposed MADDPG algorithm. Moving forward, Chapter 5 synthesizes and presents the findings derived from simulation results, offering insights into the performance evaluations and contrasts between the proposed algorithm and benchmarks. Lastly, Chapter 6 encapsulates the study's conclusions drawn from the preceding sections and outlines potential pathways for future research, providing a comprehensive structure to elucidate the efficacy and implications of the proposed MADDPG approach in RSMA.

CHAPTER 2

SYSTEM MODEL

The general structure of the system model will be given in this chapter. The specific explanation about the configurations, namely SISO, MISO and MIMO, will be presented in the coming sections.

The system model considers an interference channel (IFC) comprising two base stations (BS) each having M_1 and M_2 antennas and two users each having Q_1 and Q_2 antennas respectively paired with each BS_i, $i = 1, 2$. The BS_i sends the message S_i to user equipment, UE_i. The number of messages that can be transmitted in RSMA, U_i^N, restricted by antenna configurations as follows :

$$U_i^N \leq min(M_i, Q_i). \tag{2.1}$$

To enable better interference mitigation, S_i is split into a common and a private part, i.e., S_i^c and S_i^p. The Q_i common and private messages at each BS_i are independently encoded into streams $\boldsymbol{b}_{ic}$ and $\boldsymbol{b}_{ip}$ where $\boldsymbol{b}_{ic}, \boldsymbol{b}_{ip} \in \mathbb{C}^{M_i \times 1}$ and respectively precoded with $\boldsymbol{W}_{ic}$ and $\boldsymbol{W}_{ip}$, where $\boldsymbol{W}_{ic}$ and $\boldsymbol{W}_{ip} \in \mathbb{C}^{M_i \times M_i}$. All messages $\boldsymbol{b}_{1c}$, $\boldsymbol{b}_{1p}$, $\boldsymbol{b}_{2c}$ and $\boldsymbol{b}_{2p}$ are independent from each other, and $\mathbb{E}\{\boldsymbol{b}_{in}\boldsymbol{b}_{in}^*\} = 1$, $i = 1, 2$, $n = c, p$. Then, the transmitted signal of BS_i, $\boldsymbol{x}_i \in \mathbb{C}^{M_i \times 1}$ where $j \neq i$ and $i, j \in \{1, 2\}$, is defined as

$$\boldsymbol{x}_i = \boldsymbol{W}_{ic}\boldsymbol{b}_{ic} + \boldsymbol{W}_{ip}\boldsymbol{b}_{ip}. \tag{2.2}$$

To satisfy the power constraints at each one of the transmitters, we assume $|\boldsymbol{w}_{ikc}|^2 \leq P_{ik}$ and $|\boldsymbol{w}_{ikp}|^2 \leq P_{ik}$, where $\sum_{k=1}^{M_i} P_{ik}$ is the total power of BS_i, and $\boldsymbol{w}_{ikc}$ and $\boldsymbol{w}_{ikp}$ are representing the k'th column of $\boldsymbol{W}_{ic}$ and $\boldsymbol{W}_{ip}$ respectively. The received signal at user UE_i is then written as

$$\boldsymbol{y_i} = \boldsymbol{H}_i\boldsymbol{x}_i + \boldsymbol{G}_j\boldsymbol{x}_j + \boldsymbol{n_i}. \tag{2.3}$$

Here j indicates the index of the interfering signal, $j = 1, 2$ and $j \neq i$. The channel gain between BS_i and UE_i is indicated as $\boldsymbol{H}_i \in \mathbb{C}^{Q_i \times M_i}$. Similarly, the channel gain between BS_j and UE_i is $\boldsymbol{G}_j \in \mathbb{C}^{Q_i \times M_j}$. The entries in $\boldsymbol{H}_i$ and $\boldsymbol{G}_j$ are independent and identically distributed (i.i.d.) and complex valued random variables. The transmitters BS_i are informed about their outgoing channel gains $\boldsymbol{H}_i$ and $\boldsymbol{G}_i$, while the receivers UE_i know only their incoming channel gains $\boldsymbol{H}_i$ and $\boldsymbol{G}_j$ The noise term at UE_i is denoted with $\boldsymbol{n}_i$. It is circularly symmetric complex Gaussian with mean zero and variance $\sigma_{n,i}^2 \boldsymbol{I}_{Q_i}$, i.e., $\boldsymbol{n}_i \in \mathcal{CN}\big(0, \sigma_{n,i}^2 \boldsymbol{I}_{Q_i}\big)$. Also $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ are independent from each other. For simplicity, we will take $\sigma_{n,i}^2 = N_0$. In the following subsection rate expressions for rate-splitting are explained.

In order to be able to write the achievable rates for RSMA, the decoding order for $\boldsymbol{b}_{ic}$ and $\boldsymbol{b}_{ip}$ have to be determined at both users [1]. Moreover, to attain the best possible achievable rates, one has to consider all possible choices of these decoding orders. For the particular system model we study, we take 2 different decoding orders into consideration for each one of the users. Namely, UE_i either decodes in the order (a) or (b) in 2.4.

$$\begin{aligned} &(a)\boldsymbol{b}_{jc} \rightarrow \boldsymbol{b}_{ic} \rightarrow \boldsymbol{b}_{ip} \\ &(b)\boldsymbol{b}_{ic} \rightarrow \boldsymbol{b}_{jc} \rightarrow \boldsymbol{b}_{ip}. \end{aligned} \tag{2.4}$$

For example, when UE_1 decodes according to the order given in (a) and UE_2 decodes according to (b), the achievable rates at UE_1 are as follows :

$$\begin{aligned} R_{2c}^1 = \log_2 \det\Big(I_{Q_1} + \boldsymbol{G}_2 \boldsymbol{W}_{2c} {\boldsymbol{W}_{2c}}^H {\boldsymbol{G}_2}^H \big(\sum_{n=\{c,p\}} \boldsymbol{H}_1 \boldsymbol{W}_{1n} {\boldsymbol{W}_{1n}}^H {\boldsymbol{H}_1}^H \\ + \boldsymbol{G}_2 \boldsymbol{W}_{2p} {\boldsymbol{W}_{2p}}^H {\boldsymbol{G}_2}^H + N_0 I_{Q_1}\big)^{-1}\Big) \end{aligned} \tag{2.5}$$

$$\begin{aligned} R_{1c}^1 = \log_2 \det\Big(I_{Q_1} + \boldsymbol{H}_1 \boldsymbol{W}_{1c} {\boldsymbol{W}_{1c}}^H {\boldsymbol{H}_1}^H \big(\boldsymbol{H}_1 \boldsymbol{W}_{1p} {\boldsymbol{W}_{1p}}^H {\boldsymbol{H}_1}^H \\ + \boldsymbol{G}_2 \boldsymbol{W}_{2p} {\boldsymbol{W}_{2p}}^H {\boldsymbol{G}_2}^H + N_0 I_{Q_1}\big)^{-1}\Big) \end{aligned} \tag{2.6}$$

$$R_{1p} = \log_2 \det\Big(I_{Q_1} + \boldsymbol{H}_1 \boldsymbol{W}_{1p} {\boldsymbol{W}_{1p}}^H {\boldsymbol{H}_1}^H \big(\boldsymbol{G}_2 \boldsymbol{W}_{2p} {\boldsymbol{W}_{2p}}^H {\boldsymbol{G}_2}^H + N_0 I_{Q_1}\big)^{-1}\Big) \tag{2.7}$$

and the achievable rates at UE_2 are as follows :

$$R_{2c}^2 = \log_2 \det\Big(I_{Q_2} + \boldsymbol{H}_2 \boldsymbol{W}_{2c} {\boldsymbol{W}_{2c}}^H {\boldsymbol{H}_2}^H \big(\sum_{n=\{c,p\}} \boldsymbol{G}_1 \boldsymbol{W}_{1n} {\boldsymbol{W}_{1n}}^H {\boldsymbol{G}_1}^H$$
$$+ \boldsymbol{H}_2 \boldsymbol{W}_{2p} {\boldsymbol{W}_{2p}}^H {\boldsymbol{H}_2}^H + N_0 I_{Q_2} \big)^{-1} \Big) \tag{2.8}$$
$$R_{1c}^2 = \log_2 \det\Big(I_{Q_2} + \boldsymbol{G}_1 \boldsymbol{W}_{1c} {\boldsymbol{W}_{1c}}^H {\boldsymbol{G}_1}^H \big(\boldsymbol{G}_1 \boldsymbol{W}_{1p} {\boldsymbol{W}_{1p}}^H {\boldsymbol{G}_1}^H$$
$$+ \boldsymbol{H}_2 \boldsymbol{W}_{2p} {\boldsymbol{W}_{2p}}^H {\boldsymbol{H}_2}^H + N_0 I_{Q_2} \big)^{-1} \Big) \tag{2.9}$$
$$R_{2p} = \log_2 \det\Big(I_{Q_2} + \boldsymbol{H}_2 \boldsymbol{W}_{2p} {\boldsymbol{W}_{2p}}^H {\boldsymbol{H}_2}^H \big(\boldsymbol{G}_1 \boldsymbol{W}_{1p} {\boldsymbol{W}_{1p}}^H {\boldsymbol{G}_1}^H + N_0 I_{Q_2} \big)^{-1} \Big) \tag{2.10}$$

Since the common messages should be decoded at both receivers, common message rates are actually limited with the minimum of (2.5) and (2.8) and of (2.6) and (2.9). Thus, we define R_{1c} and R_{2c} as

$$R_{1c} = \min(R_{1c}^1, R_{1c}^2) \tag{2.11}$$
$$R_{2c} = \min(R_{2c}^1, R_{2c}^2). \tag{2.12}$$

Then, the rate for UE_i, $i = 1, 2$, can be calculated as

$$R_i = R_{ic} + R_{ip}. \tag{2.13}$$

Note that one can write 4 different sets of achievable rates as in (2.5)-(2.10), considering different combinations of decoding orders listed in (a) and (b) in 2.4. Then, for a given $\beta \in [0, 1]$, that is the given weights of the user rates, and a given decoding order, the objective is to maximize

$$\max_{\boldsymbol{w}_{ikc}, \boldsymbol{w}_{ikp}, \boldsymbol{P}_{ic}, \boldsymbol{P}_{ip}} \quad \beta R_1 + (1-\beta) R_2 \tag{2.14a}$$
$$\text{s.t.} \quad |\boldsymbol{w}_{ikc}|^2 + |\boldsymbol{w}_{ikp}|^2 \leq P_{ik}, \tag{2.14b}$$
$$i = 1, 2$$
$$k = 1, 2, .., M_i$$
$$\sum_{k=1}^{M_i} P_{ik} = P_i \tag{2.14c}$$

where $\boldsymbol{w}_{ikc}$ and $\boldsymbol{w}_{ikp}$ represent the k'th column of $\boldsymbol{W}_{ic}$ and $\boldsymbol{W}_{ip}$ respectively. Also, $\boldsymbol{P}_{ik} = \boldsymbol{P}_{ikc} + \boldsymbol{P}_{ikp}$.

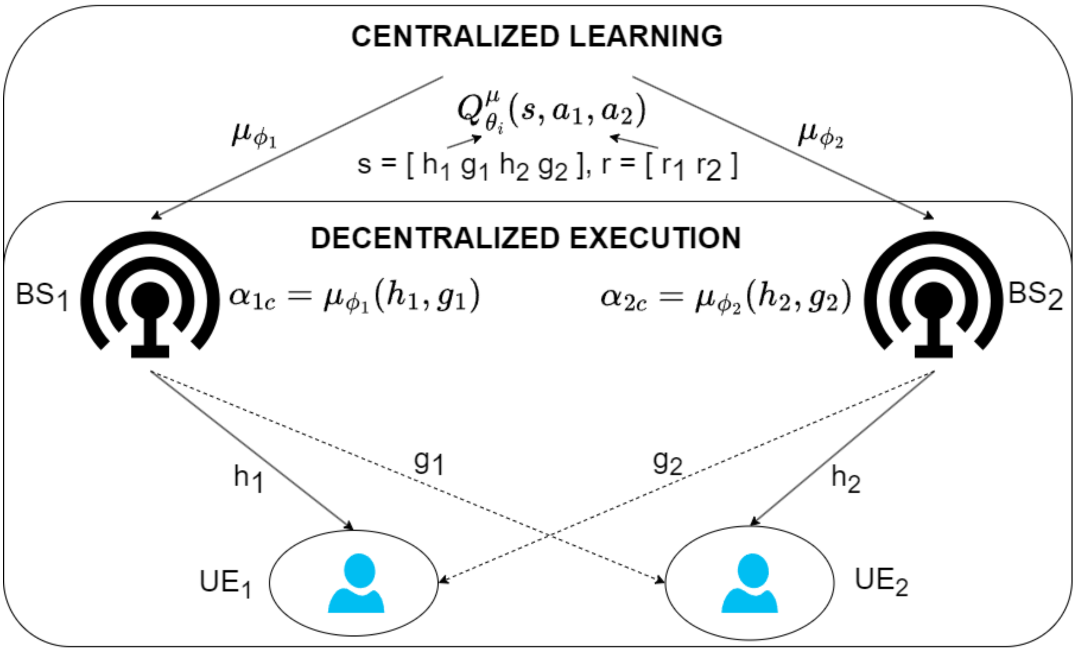
CENTRALIZED LEARNING
$Q^{\mu}_{\theta_i}(s, a_1, a_2)$
μ_{ϕ_1}
μ_{ϕ_2}
s = [h1 g1 h2 g2], r = [r1 r2]
DECENTRALIZED EXECUTION
BS1
$\alpha_{1c} = \mu_{\phi_1}(h_1, g_1)$
$\alpha_{2c} = \mu_{\phi_2}(h_2, g_2)$
BS2
h1
g1
g2
h2
UE1
UE2

Figure 2.1: System architecture for MADDPG with rate-splitting for SISO case

2.1 RSMA System Formulation for the SISO Case

In the SISO scenario, we expect that the number of messages transmitted in the RSMA structure is limited to one. This limitation is due to the conditions specified in Equation (2.1), where M_1, M_2, Q_1, and Q_2 are all set to 1. Then, the transmitted signal of BS_i, $x_i \in \mathbb{C}^{1\times 1}$ where $j \neq i$ and $i, j \in \{1, 2\}$, is defined as

$$x_i = \sqrt{\alpha_{ic}} b_{ic} + \sqrt{\alpha_{ip}} b_{ip}. \tag{2.15}$$

Since there is no need to determine any precoder vector, $\boldsymbol{w}_i$, in Equation 2.2, α_{ic} and α_{ip} become P_{ic} and P_{ip} where $\alpha_{in} \in [0, 1]$, $n = c, p$, indicates the power ratio of the encoded data b_{in}. Note that $\alpha_{ic} + \alpha_{ip} = P_i$. The received signal at user UE_i is then written as

$$y_i = h_i x_i + g_j x_j + n_i. \tag{2.16}$$

Here j indicates the index of the interfering signal, $j = 1, 2$ and $j \neq i$. Then for the decoding order of $\boldsymbol{b}_{2c} \rightarrow \boldsymbol{b}_{1c} \rightarrow \boldsymbol{b}_{1p}$, the achievable rates at UE_1 can be written as follows :

$$R_{2c}^{1} = \log\left(1 + \frac{|\boldsymbol{g}_2|^2\alpha_{2c}}{\sum_{n=\{c,p\}}|\boldsymbol{h}_1|^2\alpha_{1n} + |\boldsymbol{g}_2|^2\alpha_{2p} + N_0}\right) \tag{2.17}$$

$$R_{1c}^{1} = \log\left(1 + \frac{|\boldsymbol{h}_1|^2\alpha_{1c}}{|\boldsymbol{h}_1|^2\alpha_{1p} + |\boldsymbol{g}_2|^2\alpha_{2p} + N_0}\right) \tag{2.18}$$

$$R_{1p} = \log\left(1 + \frac{|\boldsymbol{h}_1|^2\alpha_{1p}}{|\boldsymbol{g}_2|^2\alpha_{2p} + N_0}\right) \tag{2.19}$$

and for the decoding order $\boldsymbol{b}_{2c} \to \boldsymbol{b}_{2c} \to \boldsymbol{b}_{1p}$, achievable rates at UE_2 are

$$R_{2c}^{2} = \log\left(1 + \frac{|\boldsymbol{h}_2|^2\alpha_{2c}}{\sum_{n=\{c,p\}}|\boldsymbol{g}_1|^2\alpha_{2n} + |\boldsymbol{h}_2|^2\alpha_{2p} + N_0}\right) \tag{2.20}$$

$$R_{1c}^{2} = \log\left(1 + \frac{|\boldsymbol{g}_1|^2\alpha_{1c}}{|\boldsymbol{g}_1|^2\alpha_{1p} + |\boldsymbol{h}_2|^2\alpha_{2p} + N_0}\right) \tag{2.21}$$

$$R_{2p} = \log\left(1 + \frac{|\boldsymbol{h}_2|^2\alpha_{2p}}{|\boldsymbol{g}_1|^2\alpha_{1p} + N_0}\right). \tag{2.22}$$

Here $|\boldsymbol{x}|^2$ means $\boldsymbol{x}\boldsymbol{x}^H$. The general system architecture that summarizes the whole network structure is given in Figure 2.1. Then, the rates are calculated exactly the same as (2.11)-(2.13).

Note that one can write 4 different sets of achievable rates as in (2.17)-(2.22), considering different combinations of decoding orders listed in (a) and (b) in 2.4. Then, for a given $\beta \in [0, 1]$ and a given decoding order, the objective is to maximize

$$\max_{\alpha_{ic}, i=1,2} \quad \beta R_1 + (1-\beta)R_2 \tag{2.23a}$$

$$\text{s.t.} \quad \alpha_{ic} + \alpha_{ip} \leq P_i, \qquad i = 1, 2. \tag{2.23b}$$

2.2 RSMA System Formulation for the MISO Case

Considering the MISO case, it can be anticipated that messages that will be transmitted through the communication channel in RSMA structure are restricted by 1 according to (2.1). Even if we take M_1 and M_2 as 3, (2.1) is restricted by Q_1 and Q_2 for each users. For simplicity, these antenna numbers are taken as 1 for this case.

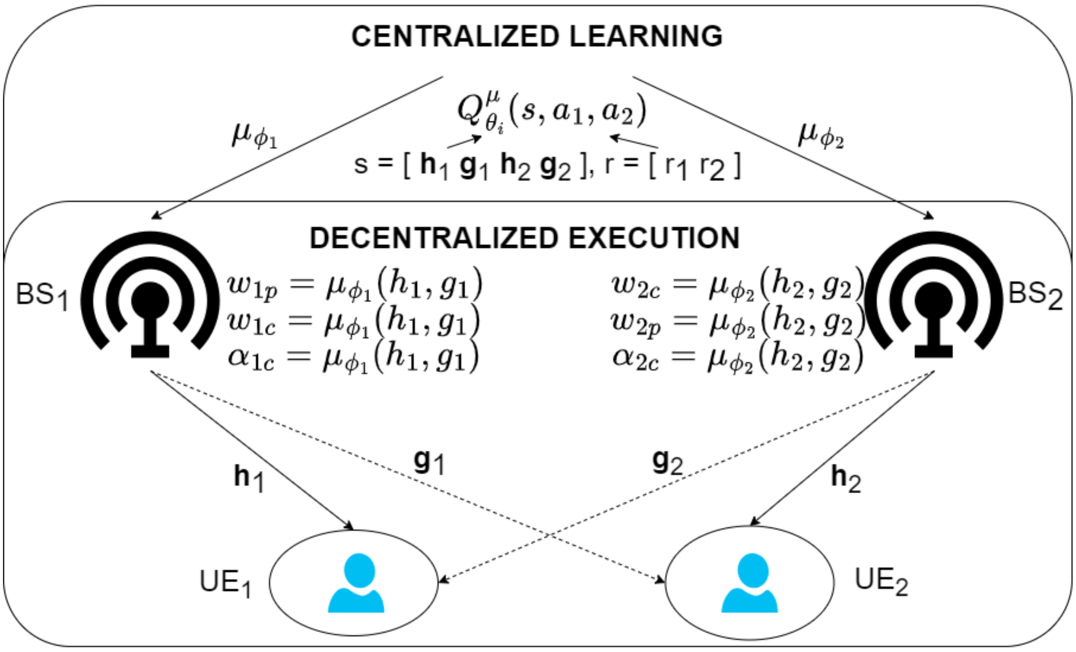
CENTRALIZED LEARNING
$Q^{\mu}_{\theta_i}(s, a_1, a_2)$
s = [$\mathbf{h}_1$ $\mathbf{g}_1$ $\mathbf{h}_2$ $\mathbf{g}_2$], r = [r_1 r_2]
μ_{ϕ_1}
μ_{ϕ_2}
DECENTRALIZED EXECUTION
BS1
$w_{1p} = \mu_{\phi_1}(h_1, g_1)$
$w_{1c} = \mu_{\phi_1}(h_1, g_1)$
$\alpha_{1c} = \mu_{\phi_1}(h_1, g_1)$
$w_{2c} = \mu_{\phi_2}(h_2, g_2)$
$w_{2p} = \mu_{\phi_2}(h_2, g_2)$
$\alpha_{2c} = \mu_{\phi_2}(h_2, g_2)$
BS2
$\mathbf{h}_1$
$\mathbf{g}_1$
$\mathbf{g}_2$
$\mathbf{h}_2$
UE1
UE2

Figure 2.2: System architecture for MADDPG with rate-splitting for MISO case

Then, the transmitted signal of BS_i, $\boldsymbol{x}_i \in \mathbb{C}^{3\times 1}$ where $j \neq i$ and $i, j \in \{1, 2\}$, is defined as

$$\boldsymbol{x}_i = \sqrt{\alpha_{ic}}\boldsymbol{w}_{ic}b_{ic} + \sqrt{\alpha_{ip}}\boldsymbol{w}_{ip}b_{ip}. \tag{2.24}$$

To satisfy the power constraints at each one of the transmitters, we assume $|\boldsymbol{w}_{ic}|^2 \leq P_i$ and $|\boldsymbol{w}_{ip}|^2 \leq P_i$, where P_i is the total power of BS_i. Also $\alpha_{in} \in [0, 1]$, $n = c, p$, indicates the power ratio of the encoded data b_{in}. Note that $\alpha_{ic} + \alpha_{ip} = 1$. The received signal at user UE_i is then written as

$$y_i = \boldsymbol{h}_i\boldsymbol{x}_i + \boldsymbol{g}_j\boldsymbol{x}_j + n_i. \tag{2.25}$$

Here j indicates the index of the interfering signal, $j = 1, 2$ and $j \neq i$. The channel gain between BS_i and UE_i is indicated as $\boldsymbol{h}_i \in \mathbb{C}^{1\times M}$. Similarly, the channel gain between BS_j and UE_i is $\boldsymbol{g}_j \in \mathbb{C}^{1\times M}$. The entries in $\boldsymbol{h}_i$ and $\boldsymbol{g}_j$ are independent and identically distributed (i.i.d.) and complex valued random variables. Then for the same configurations, the achievable rates at UE_1 can be written as follows :

$$R_{2c}^1 = \log\left(1 + \frac{\left|\boldsymbol{g}_2\boldsymbol{w}_{2c}\right|^2\alpha_{2c}}{\sum_{n=\{c,p\}}\left|\boldsymbol{h}_1\boldsymbol{w}_{1n}\right|^2\alpha_{1n} + \left|\boldsymbol{g}_2\boldsymbol{w}_{2p}\right|^2\alpha_{2p} + N_0}\right) \tag{2.26}$$

$$R_{1c}^1 = \log\left(1 + \frac{\left|\boldsymbol{h}_1\boldsymbol{w}_{1c}\right|^2\alpha_{1c}}{\left|\boldsymbol{h}_1\boldsymbol{w}_{1p}\right|^2\alpha_{1p} + \left|\boldsymbol{g}_2\boldsymbol{w}_{2p}\right|^2\alpha_{2p} + N_0}\right) \tag{2.27}$$

$$R_{1p} = \log\left(1 + \frac{\left|\boldsymbol{h}_1\boldsymbol{w}_{1p}\right|^2\alpha_{1p}}{\left|\boldsymbol{g}_2\boldsymbol{w}_{2p}\right|^2\alpha_{2p} + N_0}\right) \tag{2.28}$$

and the achievable rates at UE_2 are

$$R_{2c}^2 = \log\left(1 + \frac{\left|\boldsymbol{h}_2\boldsymbol{w}_{2c}\right|^2\alpha_{2c}}{\sum_{n=\{c,p\}}\left|\boldsymbol{g}_1\boldsymbol{w}_{1n}\right|^2\alpha_{2n} + \left|\boldsymbol{h}_2\boldsymbol{w}_{2p}\right|^2\alpha_{2p} + N_0}\right) \tag{2.29}$$

$$R_{1c}^2 = \log\left(1 + \frac{\left|\boldsymbol{g}_1\boldsymbol{w}_{1c}\right|^2\alpha_{1c}}{\left|\boldsymbol{g}_1\boldsymbol{w}_{1p}\right|^2\alpha_{1p} + \left|\boldsymbol{h}_2\boldsymbol{w}_{2p}\right|^2\alpha_{2p} + N_0}\right) \tag{2.30}$$

$$R_{2p} = \log\left(1 + \frac{\left|\boldsymbol{h}_2\boldsymbol{w}_{2p}\right|^2\alpha_{2p}}{\left|\boldsymbol{g}_1\boldsymbol{w}_{1p}\right|^2\alpha_{1p} + N_0}\right). \tag{2.31}$$

The general system architecture that summarizes the whole network structure is given in Figure 2.2. Then, the rates are calculated exactly the same as (2.11)-(2.13).

Note that one can write 4 different sets of achievable rates as in (2.26)-(2.22), considering different combinations of decoding orders listed in (a) and (b) in very first part of System Model on page 15. Then, for a given $\beta \in [0, 1]$ and a given decoding order, the objective is to maximize

$$\max_{\boldsymbol{w}_{ic},\boldsymbol{w}_{ip},\alpha_{ic},i=1,2} \quad \beta R_1 + (1-\beta)R_2 \tag{2.32a}$$

$$\text{s.t.} \quad \alpha_{ic}|\boldsymbol{w}_{ic}|^2 + (1-\alpha_{ic})|\boldsymbol{w}_{ip}|^2 \leq P_i, \quad i = 1, 2. \tag{2.32b}$$

2.3 RSMA System Formulation for the MIMO Case

The general derivations of the expressions are given through page 2.2-2.14c for MIMO case. One can see the general system architecture in Figure 2.3.

MIMO RSMA outperforms SISO and MISO RSMA configurations in several ways. With multiple antennas at both ends, MIMO RSMA enables simultaneous transmission of multiple data streams, which can be seen in equation (2.2), enhancing spectral

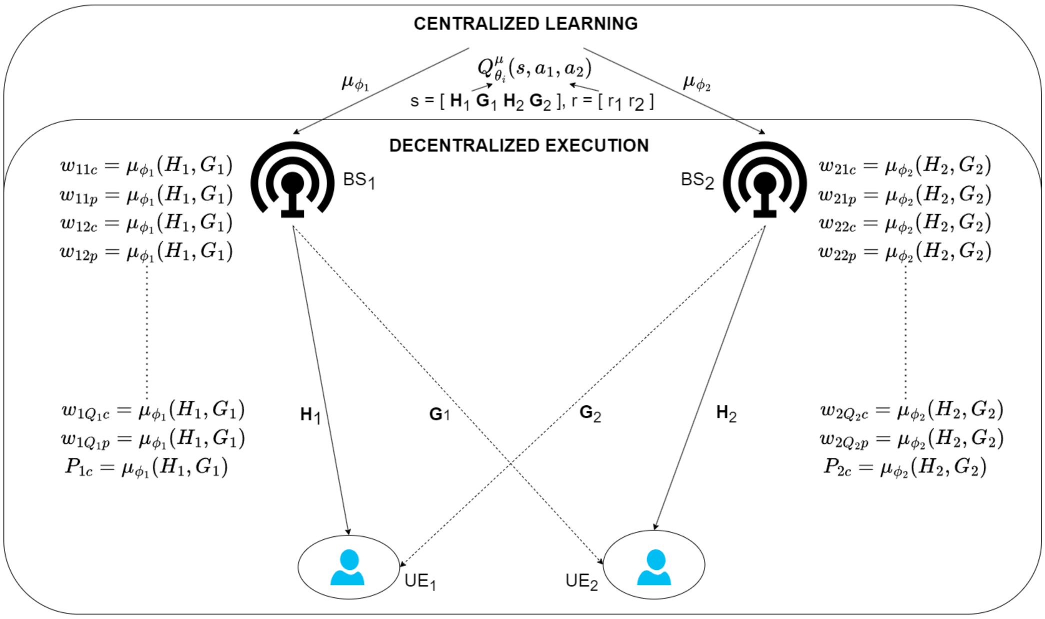
CENTRALIZED LEARNING
μ_{ϕ_1}
$Q^{\mu}_{\theta_i}(s, a_1, a_2)$
s = [H1 G1 H2 G2], r = [r1 r2]
μ_{ϕ_2}
DECENTRALIZED EXECUTION
$w_{11c} = \mu_{\phi_1}(H_1, G_1)$
$w_{11p} = \mu_{\phi_1}(H_1, G_1)$
$w_{12c} = \mu_{\phi_1}(H_1, G_1)$
$w_{12p} = \mu_{\phi_1}(H_1, G_1)$
BS1
BS2
$w_{21c} = \mu_{\phi_2}(H_2, G_2)$
$w_{21p} = \mu_{\phi_2}(H_2, G_2)$
$w_{22c} = \mu_{\phi_2}(H_2, G_2)$
$w_{22p} = \mu_{\phi_2}(H_2, G_2)$
$w_{1Q_1c} = \mu_{\phi_1}(H_1, G_1)$
$w_{1Q_1p} = \mu_{\phi_1}(H_1, G_1)$
$P_{1c} = \mu_{\phi_1}(H_1, G_1)$
H1
G1
G2
H2
$w_{2Q_2c} = \mu_{\phi_2}(H_2, G_2)$
$w_{2Q_2p} = \mu_{\phi_2}(H_2, G_2)$
$P_{2c} = \mu_{\phi_2}(H_2, G_2)$
UE1
UE2

Figure 2.3: System architecture for MADDPG with rate-splitting for MIMO case

efficiency and offering higher data rates. It harnesses spatial diversity and multiplexing gains, effectively managing interference and improving reliability against channel fading. The system's adaptability and robustness to varying channel conditions further highlight its superiority, allowing for adaptive strategies and increased overall flexibility in wireless communication setups.

However, its complexity escalates notably compared to SISO and MISO RSMA systems. The increased complexity primarily stems from the need to handle multiple antennas at both the transmitter and receiver ends, resulting in heightened signal processing demands and computational requirements. MIMO RSMA involves intricate spatial processing, necessitating sophisticated algorithms for beamforming, precoding, and decoding across multiple antenna elements. Moreover, managing interference and decoding orders becomes more challenging with the increased dimensionality and complexity introduced by multiple antennas. Therefore, learning based solution will be presented in the next chapter.

CHAPTER 3

MADDPG FOR PRECODING AND POWER ALLOCATION COEFFICIENTS OPTIMIZATION

In this section, we propose to use multi-agent deep reinforcement learning algorithm with decentralized policies and joint action optimization in order to solve the average sum-rate maximization problem defined in Chapter 2. Specifically, we adopt the MADDPG algorithm [6], which is an extension of the well-known deep deterministic policy gradient (DDPG) algorithm [37] tailored specifically for multi-agent systems. It is a powerful algorithm that has been successfully applied to challenging tasks in signal processing and communication areas [38], [25].

MADDPG employs centralized training and decentralized execution, where all agents share a common critic network to facilitate joint action optimization, and decentralized execution, where each agent independently executes its learned policy based on local observations. Specifically, the critic network in MADDPG takes as input not only the local observations and actions of an individual agent, but also the observations and actions of all other agents in the system. By doing so, the critic can learn a centralized value function that takes into account the joint actions of all agents. This centralized value function can then be used to train each agent's policy network.

On the other hand, during execution or deployment, each agent only has access to its own local observations and actions, namely, each agent i at the BS_i chooses precoding vectors and power allocation coefficient based on local information characterized by outgoing channels only. This is known as decentralized execution, as each agent acts independently based on its own observations and policies without relying on information from other agents. By decoupling execution from learning, MADDPG is able to handle complex multi-agent systems, where agents have limited or incomplete

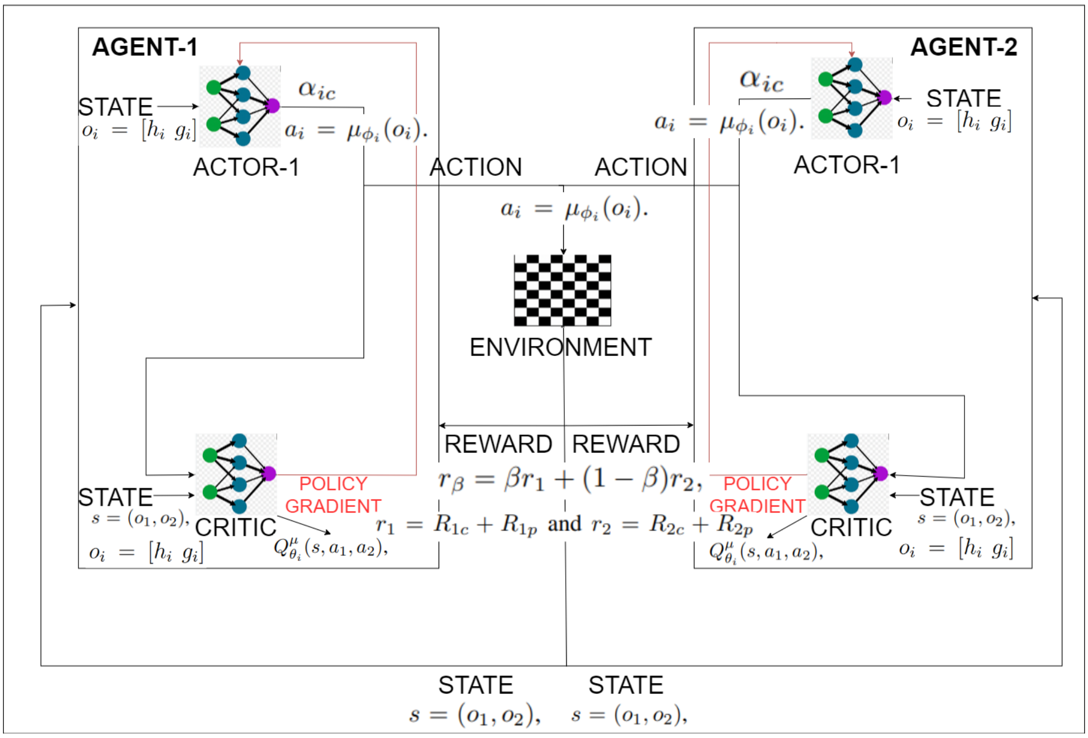

Figure 3.1: MADDPG algorithm structure for SISO case

information about the system as a whole.

The forthcoming sections present specific algorithmic details. While we address decoding order estimation in this algorithm, we avoid excessive repetition of similar elements in the algorithm specifics. Nonetheless, we are exploring two scenarios: one utilizing decoding order estimation, while the other employs an exhaustive search. The sole divergence between them lies in the count of actor networks integrated into the system. To incorporate decoding order estimation, an additional actor network is necessary, maintaining the same inputs as the others but generating a discrete output that specifies the decoding sequence.

3.1 MADDPG Algorithm Construction in the SISO case

In the realm of the multi-agent actor-critic based reinforcement learning, we need to define environment, actions, states and rewards. For each decoding order, we consider an environment with two agents. The overall system for the SISO case is summarized in Fig. 3.1.

In this system, each agent i at BS_i chooses a precoding vector a power allocation coefficient α_{ic} based on the local observation $o_i = [\boldsymbol{h}_i \ \boldsymbol{g}_i]$, $i \in \{1, 2\}$. We construct an actor network for the the power allocation coefficient α_{ic}, and a critic network to evaluate the performance of policies for each agent i. Let $\mu = \{\mu_{\phi_1}(o_1), \mu_{\phi_2}(o_2)\}$ denote the set of policies parameterized by $\phi = \{\phi_1, \phi_2\}$. The agents will choose their actions $a_i = [\alpha_{ic}]$ according to the partial state o_i by following a deterministic policy $a_i = \mu_{\phi_i}(o_i)$. To ensure sufficient exploration, we also add a noise vector, whose entries are i.i.d. according to $\mathcal{N}(0, \sigma_N^2)$ to the deterministic action $a_i = \mu_{\phi_i}(o_i)$. Then, the gradient of the expected reward $J(\phi_i)$ for each agent i can be computed as

$$\nabla_{\phi_i} J(\phi_i) = \mathbb{E}\left[\nabla_{a_i} Q^{\mu}_{\theta_i}(s, a_1, a_2)|_{a_i=\mu_{\phi_i}(o_i)} \nabla_{\phi_i}\mu_{\phi_i}(o_i)\right], \tag{3.1}$$

where $Q^{\mu}_{\theta_i}(s, a_1, a_2)$ represents the state action value function parameterized by the critic network with θ_i for each agent. It takes as input the state information $s = (o_1, o_2)$, i.e., channel gains for all users, and the actions $a = (a_1, a_2)$ of all agents, and outputs the Q-value for agent i.

Each agent receives a collaborative reward, denoted by r_β, which is a function of the environmental state and actions taken according to state observation.

$$r_\beta = \beta r_1 + (1 - \beta) r_2, \tag{3.2}$$

where $r_1 = R_{1c} + R_{1p}$ and $r_2 = R_{2c} + R_{2p}$ for the case of RSMA, and β and $1 - \beta$ denote the given weights of the user rates, defined in (2.23a). MADDPG uses this rate expression to maximize the total discounted return, which is given by $\mathrm{R} = \sum_{t=0}^{\infty} \gamma^t r^{\mu}_{\beta,t}$ where $r^{\mu}_{\beta,t}$ is the average sum-rate reward obtained under policy μ at time t, and $\gamma \in [0, 1]$ denotes the discount factor.

The critic network estimates the Q-value function $Q^{\mu}_{\theta_i}(s, a_1, a_2)$, which is the expected cumulative reward starting from state $s = (h_1, g_1, h_2, g_2)$ and taking a joint action a under policies $\mu = \{\mu_{\phi_1}, \mu_{\phi_2}\}$. MADDPG algorithm employs an experience replay buffer which records the experiences of all agents and stores tuples $< s, a_1, a_2, r_\beta, s' >$. A mini-batch of B experience tuples $< s_j, a_{1,j}, a_{2,j}, r_{\beta,j}, s'_j >_{j=1}^{B}$ are randomly sampled from the replay buffer $\mathcal{D}$, where $s'_j = (o'_{1,j}, o'_{2,j})$ and $r_{\beta,j}$

denote the next state and reward observed after actions $a_{1,j}$, $a_{2,j}$ are taken at state $s_j = (o_{1,j}, o_{2,j})$, respectively.

In addition, target networks from the DQN algorithm [39] are adopted to provide stability between actor and critic updates. Target actor networks and critic networks are denoted by μ^- and Q^- and parameterized by ϕ^- and θ^-, respectively.

We update the critic network by minimizing the mean-squared temporal difference (TD) error $\mathcal{L}(\theta_i)$ in sampled mini-batch.

$$\mathcal{L}(\theta_i) = \frac{1}{B}\sum_{j=1}^{B}\left(y_j - Q^{\mu}_{\theta_i}(s_j, a_{1,j}a_{2,j})\right)^2, \tag{3.3}$$

where the TD target y_j is computed as

$$y_j = r_{\beta,j} + \gamma Q^{\mu^-}_{\theta_i}(s'_j, \mu_{\phi_1^-}(o'_{1,j}), \mu_{\phi_2^-}(o'_{2,j})). \tag{3.4}$$

Then, we update the actor network for each agent i by using the deterministic policy gradient as

$$\nabla_{\phi_i} J(\phi_i) \approx \frac{1}{B}\sum_{j=1}^{B}\nabla_{a_i} Q^{\mu}_{\theta_i}(s_j, a_{1,j}, a_{2,j})\nabla_{\phi_i}\mu_{\phi_i}(o_i). \tag{3.5}$$

The target networks are then updated softly to match actor and critic parameters

$$\phi_i^- \leftarrow \tau\phi_i + (1-\tau)\phi_i^- \qquad\qquad \theta_i^- \leftarrow \tau\theta_i + (1-\tau)\theta_i^-, \tag{3.6}$$

where $0 < \tau < 1$ is a hyper-parameter controlling the update rate. Finally, we select the best sum-rate over all decoding orders. This is known as exhaustive search. However, as stated earlier, we also integrated the decoding order estimation to this algorithm for all antenna configurations.

3.2 MADDPG Algorithm Construction in the MISO case

Since the number of BS is chosen as 2, we consider an environment with two agents. The overall system is summarized in Fig. 3.2.

In this system, each agent i at BS_i chooses a precoding vector $\boldsymbol{w}_i = [\boldsymbol{w}_{ic}\ \boldsymbol{w}_{ip}]$, different from SISO case, and a power allocation coefficient α_{ic} based on the local

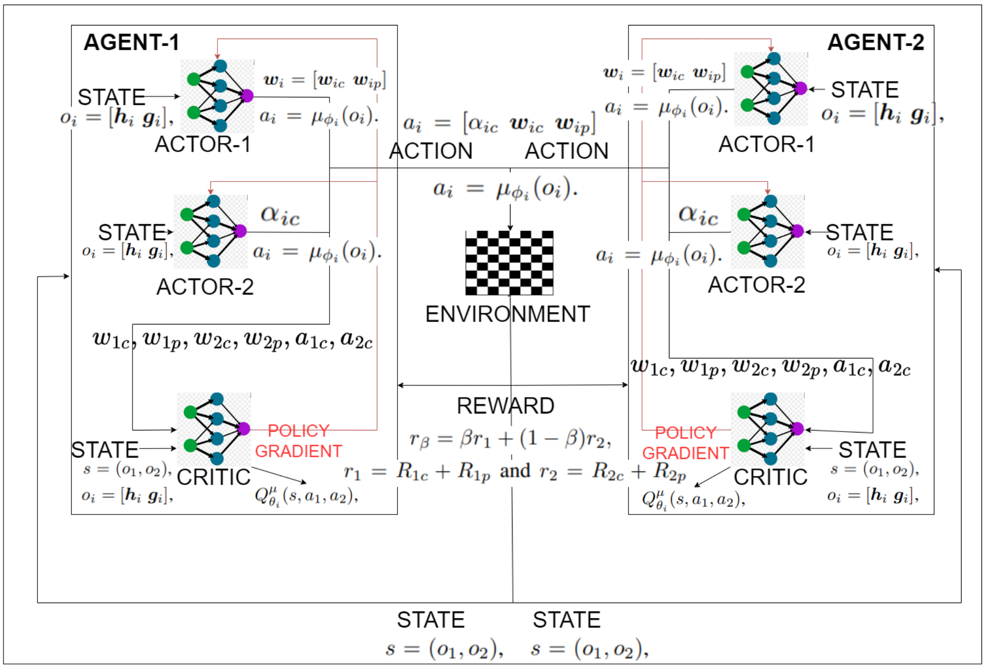
AGENT-1
STATE
$o_i = [\boldsymbol{h}_i\ \boldsymbol{g}_i]$,
ACTOR-1
$\boldsymbol{w}_i = [\boldsymbol{w}_{ic}\ \boldsymbol{w}_{ip}]$
$a_i = \mu_{\phi_i}(o_i)$.
STATE
$o_i = [\boldsymbol{h}_i\ \boldsymbol{g}_i]$,
ACTOR-2
α_{ic}
$a_i = \mu_{\phi_i}(o_i)$.
$w_{1c}, w_{1p}, w_{2c}, w_{2p}, a_{1c}, a_{2c}$
STATE
$s = (o_1, o_2)$,
$o_i = [\boldsymbol{h}_i\ \boldsymbol{g}_i]$,
CRITIC
POLICY GRADIENT
$Q^{\mu}_{\theta_i}(s, a_1, a_2)$,
$a_i = [\alpha_{ic}\ \boldsymbol{w}_{ic}\ \boldsymbol{w}_{ip}]$
ACTION
ACTION
$a_i = \mu_{\phi_i}(o_i)$.
ENVIRONMENT
REWARD
$r_\beta = \beta r_1 + (1-\beta) r_2$,
$r_1 = R_{1c} + R_{1p}$ and $r_2 = R_{2c} + R_{2p}$
AGENT-2
$\boldsymbol{w}_i = [\boldsymbol{w}_{ic}\ \boldsymbol{w}_{ip}]$
$a_i = \mu_{\phi_i}(o_i)$.
STATE
$o_i = [\boldsymbol{h}_i\ \boldsymbol{g}_i]$,
ACTOR-1
α_{ic}
$a_i = \mu_{\phi_i}(o_i)$.
STATE
$o_i = [\boldsymbol{h}_i\ \boldsymbol{g}_i]$,
ACTOR-2
$w_{1c}, w_{1p}, w_{2c}, w_{2p}, a_{1c}, a_{2c}$
POLICY GRADIENT
CRITIC
STATE
$s = (o_1, o_2)$,
$o_i = [\boldsymbol{h}_i\ \boldsymbol{g}_i]$,
$Q^{\mu}_{\theta_i}(s, a_1, a_2)$,
STATE
$s = (o_1, o_2)$,
STATE
$s = (o_1, o_2)$,

Figure 3.2: MADDPG algorithm structure for MISO case

observation $o_i = [\boldsymbol{h}_i\ \boldsymbol{g}_i]$, $i \in \{1, 2\}$. We construct an actor network for the precoding vector $\boldsymbol{w}_i$, the power allocation coefficient α_{ic}, and a critic network to evaluate the performance of policies for each agent i. Let $\mu = \{\mu_{\phi_1}(o_1), \mu_{\phi_2}(o_2)\}$ denote the set of policies parameterized by $\phi = \{\phi_1, \phi_2\}$. The agents will choose their actions $a_i = [\alpha_{ic}\ \boldsymbol{w}_{ic}\ \boldsymbol{w}_{ip}]$ according to the partial state o_i by following a deterministic policy $a_i = \mu_{\phi_i}(o_i)$. To ensure sufficient exploration, we also add a noise vector, whose entries are i.i.d. according to $\mathcal{N}(0,\ \sigma_N^2)$ to the deterministic action $a_i = \mu_{\phi_i}(o_i)$. Then, the gradient of the expected reward $J(\phi_i)$ for each agent i can be computed same as 3.1. The remaining derivations and formulas, spanning from 3.2 to 3.6, are consistent across all sections except for the algorithm hyperparameters and the state representation $s = (\boldsymbol{h}_1, \boldsymbol{g}_1, \boldsymbol{h}_2, \boldsymbol{g}_2)$.

3.3 MADDPG Algorithm Construction in MIMO case

We examine an environment containing two agents for every decoding order. The entire system is outlined in Figure 3.3. In this setup, each agent i at BS_i chooses a precoding vector $\boldsymbol{W}_i = [\boldsymbol{W}_{ic}\ \boldsymbol{W}_{ip}]$ and a power allocation coefficient $\boldsymbol{P}_{ic}$ based on the

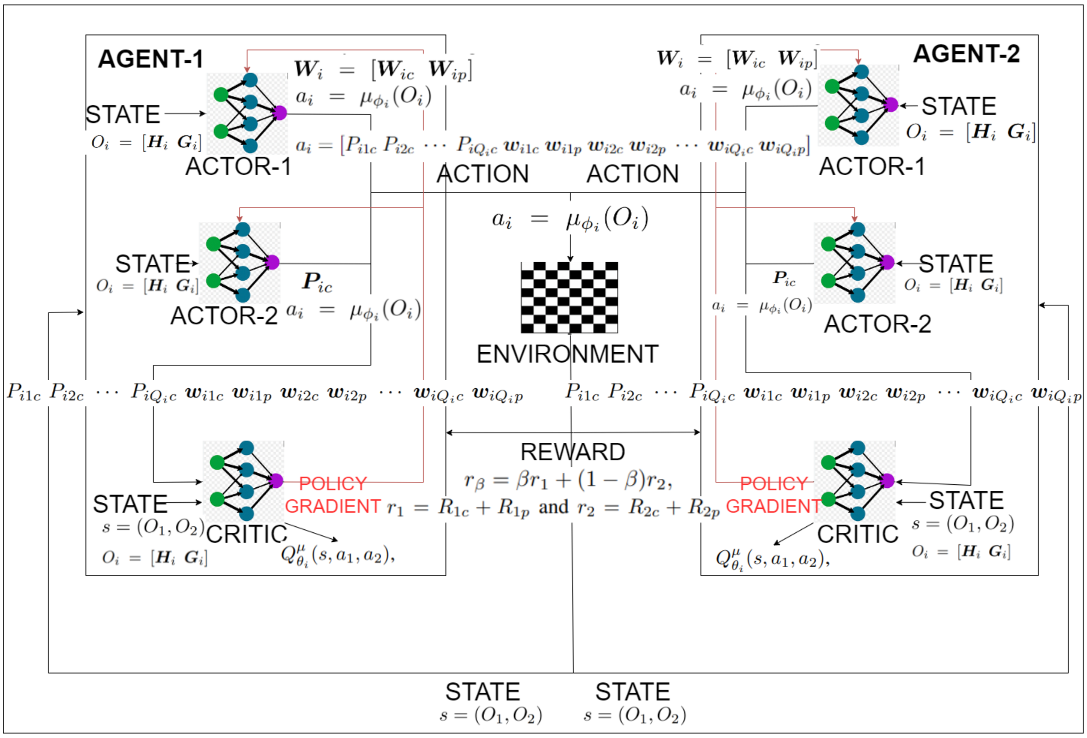
AGENT-1
$W_i = [W_{ic}\ W_{ip}]$
$a_i = \mu_{\phi_i}(O_i)$
STATE
$O_i = [H_i\ G_i]$
ACTOR-1
$a_i = [P_{i1c}\ P_{i2c} \cdots P_{iQ_ic}\ w_{i1c}\ w_{i1p}\ w_{i2c}\ w_{i2p} \cdots w_{iQ_ic}\ w_{iQ_ip}]$
ACTION
ACTION
AGENT-2
$W_i = [W_{ic}\ W_{ip}]$
$a_i = \mu_{\phi_i}(O_i)$
STATE
$O_i = [H_i\ G_i]$
ACTOR-1
$a_i = \mu_{\phi_i}(O_i)$
STATE
$O_i = [H_i\ G_i]$
P_{ic}
ACTOR-2
$a_i = \mu_{\phi_i}(O_i)$
ENVIRONMENT
P_{ic}
$a_i = \mu_{\phi_i}(O_i)$
STATE
$O_i = [H_i\ G_i]$
ACTOR-2
$P_{i1c}\ P_{i2c} \cdots P_{iQ_ic}\ w_{i1c}\ w_{i1p}\ w_{i2c}\ w_{i2p} \cdots w_{iQ_ic}\ w_{iQ_ip}$
$P_{i1c}\ P_{i2c} \cdots P_{iQ_ic}\ w_{i1c}\ w_{i1p}\ w_{i2c}\ w_{i2p} \cdots w_{iQ_ic}\ w_{iQ_ip}$
REWARD
$r_\beta = \beta r_1 + (1-\beta)r_2,$
POLICY
GRADIENT
$r_1 = R_{1c} + R_{1p}$ and $r_2 = R_{2c} + R_{2p}$
POLICY
GRADIENT
STATE
$s = (O_1, O_2)$
CRITIC
$O_i = [H_i\ G_i]$
$Q^{\mu}_{\theta_i}(s, a_1, a_2),$
CRITIC
STATE
$s = (O_1, O_2)$
$O_i = [H_i\ G_i]$
$Q^{\mu}_{\theta_i}(s, a_1, a_2),$
STATE
$s = (O_1, O_2)$
STATE
$s = (O_1, O_2)$

Figure 3.3: MADDPG algorithm structure for MIMO case

local observation $O_i = [\boldsymbol{H}_i\ \boldsymbol{G}_i]$, $i \in \{1, 2\}$. We construct Q_i actor networks for the precoding vectors $\boldsymbol{w}_{ik}$ and an actor network for the the power allocation coefficients P_{ikc}, and a critic network to evaluate the performance of policies for each agent i. Let $\mu = \{\mu_{\phi_1}(O_1), \mu_{\phi_2}(O_2)\}$ denote the set of policies parameterized by $\phi = \{\phi_1, \phi_2\}$. The agents will choose their actions $a_i = [P_{i1c}\ P_{i2c} \cdots P_{iQ_ic}\ \boldsymbol{w}_{i1c}\ \boldsymbol{w}_{i1p}\ \boldsymbol{w}_{i2c}\ \boldsymbol{w}_{i2p} \cdots \boldsymbol{w}_{iQ_ic}\ \boldsymbol{w}_{iQ_ip}]$ according to the partial state O_i by following a deterministic policy $a_i = \mu_{\phi_i}(O_i)$. To ensure sufficient exploration, we also add a noise vector, whose entries are i.i.d. according to $\mathcal{N}(0, \sigma_N^2)$ to the deterministic action $a_i = \mu_{\phi_i}(O_i)$. Then, the gradient of the expected reward $J(\phi_i)$ for each agent i can be computed same as 3.1. The remaining derivations and formulas, spanning from 3.2 to 3.6, are consistent across all sections except for the algorithm hyperparameters, the state representation $s = (\boldsymbol{H}_1, \boldsymbol{G}_1, \boldsymbol{H}_2, \boldsymbol{G}_2)$ and observations, O_i.

3.4 MADDPG Algorithm Summary

MADDPG is an extension of the DDPG algorithm designed for cooperative or competitive scenarios involving multiple interacting agents. It combines the actor-critic framework of DDPG with multiple actor and critic networks, each agent having its own policy to select actions based on local observations. This approach allows agents to learn and interact in environments where their actions affect not only their local rewards but also the global system performance. Agents iteratively update their policies using experiences sampled from their own interactions, enabling decentralized execution based on centralized learning. This algorithm facilitates coordination and learning in scenarios with multiple decision-making entities while leveraging experience sharing to improve overall performance in complex environments. The specific MADDPG algorithm for MIMO Sum-Rate Maximization is given in Algorithm 1. To obtain other antenna configurations, one needs to replace the particular parts with equations derived in Section 3.1 and Section 3.2.

Algorithm 1 MADDPG for Sum-Rate Maximization

```
Initialize actor networks $\mu_{\phi_i}(O_i)$ and critic networks $Q_{\theta_i}(s_i, a_{i,1}, a_{i,2})$ with weights
$\theta_i$ and $\phi_i$
Initialize target networks $\mu^-$ and $Q^-$ with weights $\theta_i^- \leftarrow \theta_i$ and $\phi_i^- \leftarrow \phi_i$
Initialize replay buffer $\mathcal{D}$
for $episode = 1, \ldots, E$ do
    for $t = 1, \ldots, T$ do
        for each agent $i$ do
            Observe partial state $O_i = (\boldsymbol{H}_i, \boldsymbol{G}_i)$
            Select action $a_i = \mu_{\phi_i}(O_i)$,
            Execute action with exploration noise:
              $a_i = \mu_{\phi_i}(O_i) + \mathcal{N}(0,\, \sigma_N^2)$.
            Observe reward $r_i$.
        end for
        Observe the sum-rate reward $r_\beta = \beta r_1 + (1-\beta) r_2$ and state
          $s = (O_1\ O_2) = (\boldsymbol{H}_1, \boldsymbol{G}_1, \boldsymbol{H}_2, \boldsymbol{G}_2)$ and next state $s'$
        Add transition $(s, a_1, a_2, r_\beta, s')$ to $\mathcal{D}$
        Sample a minibatch of $B$ transitions:
          $(s_j, a_{1,j}, a_{2,j}, r_{\beta,j}, s'_j)$ from $\mathcal{D}$
        Compute target action:
          $a'_j = \mu^-(o'_j)$
        Compute target Q-value:
          $y_j = r_{\beta,j} + \gamma Q^-(s'_j, a'_{1,j}, a'_{2,j})$
        Update each critic by minimizing the loss:
          $L(\theta_i) = \frac{1}{B}\sum_{j=1}^{B}(y_j - Q^{\mu}_{\theta_i}(s_j, a_{i,1}, a_{i,2}))^2$
        Update each actor using the sampled policy gradient:
          $\nabla_{\phi_i} J(\phi_i) \approx \frac{1}{B}\sum_{j=1}^{B} \nabla_{a_i} Q^{\mu}_{\theta_i}(s_j, a_{1,j}, a_{2,j}) \nabla_{\phi_i}\mu_{\phi_i}(O_i)$
        Soft update target networks:
          $\theta_i^- \leftarrow \tau\theta_i + (1-\tau)\theta_i^-,\;\; \phi_i^- \leftarrow \tau\phi_i + (1-\tau)\phi_i^-$
    end for
end for
```

CHAPTER 4

BENCHMARK PRECODING SCHEMES

In this section, we will be explaining the benchmark precoding schemes MADDPG with no RS [3], maximum ratio transmission [40], zero-forcing precoding [40], and leakage based precoding [41]. We will also compare with the upper bounds [4, 5, 22] on interference channels.

4.1 MADDPG with no Rate-Splitting

If there is no rate-splitting, our scheme reduces to the one in [3]. Also, the system model reduces to Figure 4.1. There is no common message, $b_{ic} = \emptyset$, and $\boldsymbol{w}_{ic}$ is an all zero vector. the one in that for MADDPG without rate-splitting, (3.2) can be modified by using $r_1 = R_1$ and $r_2 = R_2$ that are given in (4.1) and (4.2). Also, since we only optimize precoders, but not α_{ic} or α_{ip}, we use actions only for precoder evaluation. Then, the rates achieved by this scheme for MIMO case become

$$R_1^{\pi} = \log\det\left(\boldsymbol{I}_{\boldsymbol{Q_1}} + (\boldsymbol{H}_1\boldsymbol{W}_1^{\pi}\boldsymbol{W}_1^{\pi,H}\boldsymbol{H}_1^{H})(\boldsymbol{G}_2\boldsymbol{W}_2^{\pi}\boldsymbol{W}_2^{\pi,H}\boldsymbol{G}_2^{H} + N_0\boldsymbol{I}_{\boldsymbol{Q_1}})^{-1}\right) \quad (4.1)$$

$$R_2^{\pi} = \log\det\left(\boldsymbol{I}_{\boldsymbol{Q_1}} + (\boldsymbol{H}_2\boldsymbol{W}_2^{\pi}\boldsymbol{W}_2^{\pi,H}\boldsymbol{H}_2^{H})(\boldsymbol{G}_1\boldsymbol{W}_1^{\pi}\boldsymbol{W}_1^{\pi,H}\boldsymbol{G}_1^{H} + N_0\boldsymbol{I}_{\boldsymbol{Q_1}})^{-1}\right) \quad (4.2)$$

where $\pi = \{drl\}$ indicates the precoders for MADDPG without rate-splitting.

For the case the expressions (4.1) and (4.2) become

$$R_1^{\pi} = \log\left(1 + \frac{|\boldsymbol{h}_1\boldsymbol{w}_1^{\pi}|^2}{|\boldsymbol{g}_2\boldsymbol{w}_2^{\pi}|^2 + N_0}\right) \quad (4.3)$$

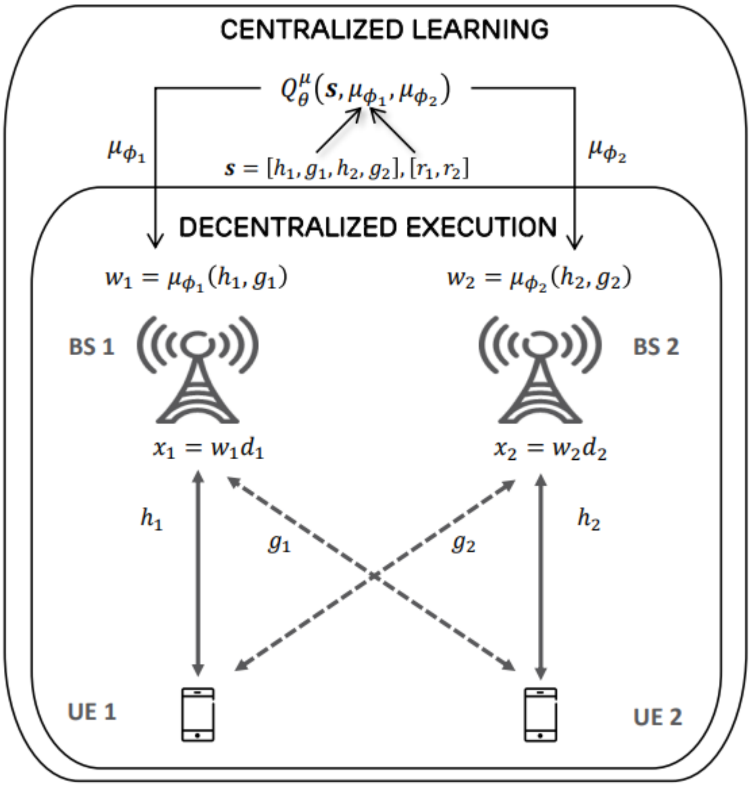
CENTRALIZED LEARNING
$Q_\theta^\mu(s, \mu_{\phi_1}, \mu_{\phi_2})$
μ_{ϕ_1}
μ_{ϕ_2}
$s = [h_1, g_1, h_2, g_2], [r_1, r_2]$
DECENTRALIZED EXECUTION
$w_1 = \mu_{\phi_1}(h_1, g_1)$
$w_2 = \mu_{\phi_2}(h_2, g_2)$
BS 1
BS 2
$x_1 = w_1 d_1$
$x_2 = w_2 d_2$
h_1
g_1
g_2
h_2
UE 1
UE 2

Figure 4.1: MADDPG with no RS system model [3]

$$R_2^\pi = \log\left(1 + \frac{|\boldsymbol{h}_2 \boldsymbol{w}_2^\pi|^2}{|\boldsymbol{g}_1 \boldsymbol{w}_1^\pi|^2 + N_0}\right), \tag{4.4}$$

Please note that the terms in the expressions 4.3 and 4.4 become scalar for SISO case. This scheme will then be used to justify the rate-splitting gain in the Chapter 5, when we present thesimulation results.

4.2 Maximum Ratio Transmission (MRT)

Maximum ratio transmission is employed at the transmitter side, where transmit antenna weights are matched to the channel [42], [40]. This way, the maximum received SNR is attained at the intended receivers. This process takes advantage of the spatial diversity offered by multiple antennas at both the transmitter and receiver ends. By adjusting the transmission weights in this manner, MRT aims to maximize the received signal power, effectively exploiting the available spatial dimensions and enhancing the system's overall performance in terms of reliability and data throughput. However, maximal ratio transmission does not take interference into consideration. In maximum ratio transmission there is no rate-splitting and there is no common message, $b_{ic} = \emptyset$, $\boldsymbol{w}_{ic}$ is an all zero vector, and for MIMO, MISO and SISO cases precoder expressions are as follows:

For a MIMO system $\boldsymbol{W}_i^{mrt} = \boldsymbol{W}_{ip}$, where

$$\boldsymbol{W}_i^{mrt} = \boldsymbol{H}_i^H. \tag{4.5}$$

For a MISO system $\boldsymbol{w}_i^{mrt} = \boldsymbol{w}_{ip}$, where

$$\boldsymbol{w}_i^{mrt} = \boldsymbol{h}_i^H. \tag{4.6}$$

For a SISO system $w_i^{mrt} = w_{ip}$, where

$$w_i^{mrt} = h_i^H. \tag{4.7}$$

4.3 Zero-Forcing (ZF)

As in maximum ratio transmission, there is no rate-splitting in zero-forcing transmission, and the transmitters aim to eliminate interferences among data streams by setting the transmission weights such that the signal transmitted from each antenna is orthogonal to the interference caused on the other receiving antennas. This is achieved by using the pseudo-inverse of the channel matrix to create a null space for the interference, i.e., by projecting input data symbols on the null space of $\boldsymbol{G_i}$. By nullifying the interference, ZF seeks to improve the reliability of data transmission without causing mutual interference among the multiple antennas, thereby enhancing the overall system performance in terms of throughput and signal quality. However, ZF might be sensitive to noise and can lead to amplication of noise in the process of eliminating interference. As a result, for ZF $b_{ic} = \emptyset$, $\boldsymbol{w}_{ic}$ is an all zero vector, and

for a MIMO system $\boldsymbol{W}_i^{zf} = \boldsymbol{W}_{ip}$, where

$$\boldsymbol{W}_i^{zf} = (\boldsymbol{G}_i^H \boldsymbol{G}_i)^{-1} \boldsymbol{H}_i^H, \tag{4.8}$$

for a MISO system $\boldsymbol{w}_i^{zf} = \boldsymbol{w}_{ip}$, where

$$\boldsymbol{w}_i^{zf} = (\boldsymbol{g}_i^H \boldsymbol{g}_i)^{-1} \boldsymbol{h}_i^H, \tag{4.9}$$

and for a SISO system $w_i^{zf} = w_{ip}$, where

$$w_i^{zf} = (g_i^H g_i)^{-1} h_i^H. \tag{4.10}$$

4.4 Leakage Based Precoding

SLNR precoding technique is designed to optimize signal transmission in multi-user communication systems by minimizing interference among users while considering the system's noise. Unlike traditional SNR-based approaches, SLNR focuses on minimizing both interference and noise to enhance the overall signal quality. In SLNR precoding, the precoding matrix is computed to maximize the desired signal power while minimizing the interference caused to other users. It aims to maintain a high signal-to-leakage-plus-noise ratio for the intended receiver, hence reducing interference while considering the system noise level. This precoding strategy is particularly useful in multi-user scenarios where reducing interference among users is crucial to improve overall system performance. In other words, leakage is a measure of how much signal power leaks into the other users. In this precoding scheme, the aim is to maximize the SLNR [41]. The leakage based precoder can be computed as follows:

For a MIMO systemk k'th column of $\boldsymbol{W}_i^{slnr}$, $\boldsymbol{W}_{ik}^{slnr}$, is equal to the eigenvector that corresponds to the largest eigenvalue of $\big((N_0\boldsymbol{I} + \boldsymbol{G}_{ik}^H \boldsymbol{G}_{ik})^{-1} \boldsymbol{H}_{ik}^H \boldsymbol{H}_{ik}\big)$ where $\boldsymbol{G}_{ik}$ and $\boldsymbol{H}_{ik}$ represent the k'th row of $\boldsymbol{G}_i$ and $\boldsymbol{H}_i$ respectively.

For a MISO system $\boldsymbol{w}_i^{slnr}$ is equal to the eigenvector that corresponds to the largest eigenvalue of $\big((N_0\boldsymbol{I} + \boldsymbol{g}_i^H \boldsymbol{g}_i)^{-1} \boldsymbol{h}_i^H \boldsymbol{h}_i\big)$.

For a SISO system w_i^{slnr} is equal to the eigenvector that corresponds to the largest eigenvalue of $\big((N_0 I + g_i^H g_i)^{-1} h_i^H h_i\big)$ which is just a scalar.

Similar to maximum ratio transmission and zero-forcing, there is no rate-splitting in leakage based precoding. The achievable rates for maximum ratio transmission, zero-forcing and leakage based precoding can all be written as in (4.1) and (4.2), where $\pi = \{mrt,\ zf,\ slnr\}$.

4.5 Interference Channel Upper Bounds

In [4] and [5], the authors suggest upper bounds for single antenna interference channels. In the next section, for the single antenna case, we will use the upper bound given in [5] for weak and mixed interference conditions. Although the definitions and explanations can be found in Appendix B, C, and D, the calculations and deductions for the weak and mixed interference channel are presented in [5], whereas [4] provides the derivations for the strong interference channel. Also, when some of the devices have multiple antennas, the above bounds are not directly applicable and we use the upper bound calculated in [22]. The derived expressions for computation are as follows :

4.5.1 SISO Weak Interference Channel

The weak interference channel describes a communication scenario where the interference imposed by one transmitter on the intended receiver of another transmitter is relatively weak compared to the signal power received at the intended receiver. In this context, the interference level is relatively lower than the received signal strength, allowing for more manageable interference mitigation techniques and potentially better communication performance compared to strong interference scenarios. For the weak interference condition, [5] is used for the upper bound evaluation. The computation of upper bound differs according to the interference terms.

4.5.1.1 $INR_1 \geq 1$ and $INR_2 \geq 1$

This condition aligns with the achievable region outlined in Appendix B.1, where $INR_{p_1} = 1$ and $INR_{p_2} = 1$. Consequently, adopting this criterion implies that we retain a considerable degree of optimality by essentially treating the entirety of user 2's signal as noise at receiver 1 and reciprocally treating the entirety of user 1's signal as noise at receiver 2. By doing so, the rates can be expressed as follows

$$
\begin{aligned}
R_1 &\leq \log\left(2+\mathrm{SNR}_1\right)-1\\
R_2 &\leq \log\left(2+\mathrm{SNR}_2\right)-1\\
R_1+R_2 &\leq \log\left(2\mathrm{INR}_2+\mathrm{SNR}_1\right)+\log\left(1+\frac{1+\mathrm{SNR}_2}{\mathrm{INR}_2}\right)-2\\
R_1+R_2 &\leq \log\left(2\mathrm{INR}_1+\mathrm{SNR}_2\right)+\log\left(1+\frac{1+\mathrm{SNR}_1}{\mathrm{INR}_1}\right)-2\\
R_1+R_2 &\leq \log\left(1+\mathrm{INR}_1+\frac{\mathrm{SNR}_1}{\mathrm{INR}_2}\right)+\log\left(1+\mathrm{INR}_2+\frac{\mathrm{SNR}_2}{\mathrm{INR}_1}\right)-2\\
2R_1+R_2 &\leq \log\left(1+\mathrm{SNR}_1+\mathrm{INR}_1\right)+\log\left(1+\mathrm{INR}_2+\frac{\mathrm{SNR}_2}{\mathrm{INR}_1}\right)\\
&\quad+\log\left(2+\frac{\mathrm{SNR}_1}{\mathrm{INR}_2}\right)-3\\
R_1+2R_2 &\leq \log\left(1+\mathrm{SNR}_2+\mathrm{INR}_2\right)+\log\left(1+\mathrm{INR}_1+\frac{\mathrm{SNR}_1}{\mathrm{INR}_2}\right)\\
&\quad+\log\left(2+\frac{\mathrm{SNR}_2}{\mathrm{INR}_1}\right)-3.
\end{aligned}
\tag{4.11}
$$

4.5.1.2 $INR_1 < 1$ and $INR_2 \geq 1$

This requirement coincides with the attainable area specified in Appendix B.1, specifically with $INR_{p_2} = 1$. As a result, adhering to this criterion implies maintaining a significant level of optimality, essentially treating the complete signal from user 1 as noise at receiver 2.

In such an instance, the rate expressions are as follows:

$$
\begin{aligned}
R_1 &\leq \log\left(1+\frac{\mathrm{SNR}_1}{1+\mathrm{INR}_1}\right)\\
R_2 &\leq \log\left(2+\mathrm{SNR}_2\right)-1\\
R_1+R_2 &\leq \log\left(\mathrm{INR}_2+\frac{\mathrm{SNR}_1}{1+\mathrm{INR}_1}\right)+\log\left(1+\frac{1+\mathrm{SNR}_2}{\mathrm{INR}_2}\right)-1\\
R_1+R_2 &\leq \left(1+\frac{\mathrm{SNR}_1}{1+\mathrm{INR}_1}\right)+\log\left(2+\mathrm{SNR}_2\right)-1\\
R_1+R_2 &\leq \log\left(\mathrm{INR}_2+\frac{\mathrm{SNR}_1}{1+\mathrm{INR}_1}\right)+\log\left(1+\frac{1+\mathrm{SNR}_2}{\mathrm{INR}_2}\right)-1
\end{aligned}
$$

$$\begin{aligned} 2R_1 + R_2 &\leq \log\left(1 + \mathrm{SNR}_1 + \mathrm{INR}_1\right) + \log\left(1 + \mathrm{INR}_2 + \mathrm{SNR}_2\right) \\ &\quad + \log\left(1 + \mathrm{INR}_1 + \frac{\mathrm{SNR}_1}{\mathrm{INR}_2}\right) - \log 2\left(1 + \mathrm{INR}_1\right)^2 \\ R_1 + 2R_2 &\leq \log\left(2 + \mathrm{SNR}_2\right) + \log\left(\mathrm{INR}_2 + \frac{\mathrm{SNR}_1}{1 + \mathrm{INR}_1}\right) \\ &\quad + \log\left(1 + \frac{1 + \mathrm{SNR}_2}{\mathrm{INR}_2}\right) - 2. \end{aligned} \tag{4.12}$$

4.5.1.3 $INR_1 \geq 1$ and $INR_2 < 1$

This condition aligns with the achievable region outlined in Appendix B.1, where $INR_{p_1} = 1$. Consequently, adopting this criterion implies that we retain a considerable degree of optimality by essentially treating the entirety of user 2's signal as noise at receiver 1. In this case, rate expression are as follows

$$\begin{aligned} R_1 &\leq \log\left(2 + \mathrm{SNR}_1\right) - 1 \\ R_2 &\leq \log\left(1 + \frac{\mathrm{SNR}_2}{1 + \mathrm{INR}_2}\right) \\ R_1 + R_2 &\leq \log\left(\mathrm{INR}_1 + \frac{\mathrm{SNR}_1}{1 + \mathrm{INR}_2}\right) + \log\left(1 + \frac{1 + \mathrm{SNR}_2}{\mathrm{INR}_1}\right) - 1 \\ R_1 + R_2 &\leq \left(1 + \frac{\mathrm{SNR}_1}{1 + \mathrm{INR}_2}\right) + \log\left(2 + \mathrm{SNR}_2\right) - 1 \\ R_1 + R_2 &\leq \log\left(\mathrm{INR}_1 + \frac{\mathrm{SNR}_1}{1 + \mathrm{INR}_2}\right) + \log\left(1 + \frac{1 + \mathrm{SNR}_2}{\mathrm{INR}_1}\right) - 1 \\ 2R_1 + R_2 &\leq \log\left(1 + \mathrm{SNR}_1 + \mathrm{INR}_2\right) + \log\left(1 + \mathrm{INR}_1 + \mathrm{SNR}_2\right) \\ &\quad + \log\left(1 + \mathrm{INR}_2 + \frac{\mathrm{SNR}_1}{\mathrm{INR}_1}\right) - \log 2\left(1 + \mathrm{INR}_2\right)^2 \\ R_1 + 2R_2 &\leq \log\left(2 + \mathrm{SNR}_2\right) + \log\left(\mathrm{INR}_1 + \frac{\mathrm{SNR}_1}{1 + \mathrm{INR}_2}\right) \\ &\quad + \log\left(1 + \frac{1 + \mathrm{SNR}_2}{\mathrm{INR}_1}\right) - 2. \end{aligned} \tag{4.13}$$

4.5.1.4 $INR_1 < 1$ and $INR_2 < 1$

Assessing the attainable region B.1 by setting $INR_{p_1} = INR_1$ and $INR_{p_2} = INR_2$, and eliminating unnecessary constraints, yields the subsequent region

$$\begin{aligned} R_1 &\leq \log\left(1 + \frac{\mathrm{SNR}_1}{1 + \mathrm{INR}_1}\right) \\ R_2 &\leq \log\left(1 + \frac{\mathrm{SNR}_2}{1 + \mathrm{INR}_2}\right). \end{aligned} \tag{4.14}$$

4.5.2 SISO Mixed Interference Channel

A mixed interference channel represents a communication scenario involving multiple transmitters and receivers, where the interference levels among the different communication links vary in strength. Unlike the weak or strong interference channels, the mixed interference channel encompasses a combination of interference scenarios, ranging from weak to strong, across the different transmitter-receiver pairs. This variability in interference levels poses additional challenges in managing interference and optimizing communication performance due to the diverse interference strengths present in the channel. For the mixed interference condition, [5] is used for the upper bound evaluation. The computation of upper bound differs according to user's interference. It is assume that $\mathrm{INR}_1 \geq \mathrm{SNR}_2$ and $\mathrm{INR}_2 < \mathrm{SNR}_1$ in the mixed interference channel. A remarkable feature of this channel is that user 2's message can be fully decoded at receiver 1. Using this fact, a natural scheme for user 2 is to use all of his power on the common message, i.e., set $INR_{p_1} = 0$. We also let INR_{p_2} to be as close to 1 as possible.

4.5.2.1 $\boldsymbol{INR_2 > 1}$

In this case we use the Han–Kobayashi scheme HK(1,0) which is defined in Appendix B.1. By evaluating that, we have the following rate expressions

$$\begin{aligned}
R_1 &\leq \log\left(1+\mathrm{SNR}_1\right) \\
R_2 &\leq \log\left(2+\mathrm{SNR}_2\right) - 1 \\
R_1 + R_2 &\leq \log\left(\mathrm{INR}_2+\mathrm{SNR}_1\right) + \log\left(1+\frac{1+\mathrm{SNR}_2}{\mathrm{INR}_2}\right) - 1 \\
R_1 + R_2 &\leq \log\left(1+\mathrm{INR}_1+\mathrm{SNR}_1\right) \\
R_1 + R_2 &\leq \log\left(1+\mathrm{INR}_1+\frac{\mathrm{SNR}_1}{\mathrm{INR}}\right) + \log\left(1+\mathrm{INR}_2\right) - 1 \\
2R_1 + R_2 &\leq \log\left(1+\mathrm{INR}_2\right) + \log\left(1+\mathrm{SNR}_1+\mathrm{INR}_1\right) \\
&\quad + \log\left(1+\frac{\mathrm{SNR}_1}{\mathrm{INR}_2}\right) - 1 \\
R_1 + 2R_2 &\leq \log\left(1+\mathrm{SNR}_2+\mathrm{INR}_2\right) \\
&\quad + \log\left(1+\mathrm{INR}_1+\frac{\mathrm{SNR}_1}{\mathrm{INR}_2}\right) - 1.
\end{aligned} \tag{4.15}$$

4.5.2.2 $INR_2 < 1$

In this case we use the Han–Kobayashi scheme HK(INR_2,0) which is defined in Appendix B.1. By evaluating that, we have the following rate expressions

$$\begin{aligned}
R_1 &\leq \log\left(1+\mathrm{SNR}_1\right)\\
R_2 &\leq \log\left(1+\frac{\mathrm{SNR}_2}{1+\mathrm{INR}_2}\right)\\
R_1+R_2 &\leq \log\left(1+\mathrm{SNR}_1\right)+\log\left(1+\frac{\mathrm{SNR}_2}{1+\mathrm{INR}_2}\right)\\
R_1+R_2 &\leq \log\left(1+\mathrm{SNR}_1+\mathrm{INR}_1\right)\\
2R_1+R_2 &\leq \left(1+\mathrm{SNR}_1\right)+\log\left(1+\mathrm{SNR}_1+\mathrm{INR}_1\right)\\
R_1+2R_2 &\leq \log\left(1+\frac{\mathrm{SNR}_2}{1+\mathrm{INR}_2}\right)+\log\left(1+\mathrm{SNR}_1+\mathrm{INR}_1\right). \qquad (4.16)
\end{aligned}$$

4.5.3 SISO Strong Interference Channel

In the context of communication systems, a strong interference channel denotes a scenario where interference significantly affects the transmission between multiple transmitter-receiver pairs. In such channels, the interference levels are notably higher compared to the signal strength, leading to substantial degradation in the quality of received signals. In situations with strong interference, the most effective approach is to use joint decoding. This means solving for every user simultaneously. This method is considered optimal and ensures better outcomes in handling strong interference, making it a suitable solution for such scenarios. For the strong interference condition, [4] is used for the upper bound evaluation. We have the following rate expressions under strong interference condition which is expressed in Appendix C.

$$\begin{aligned}
&0 \leq R_1 \leq \log\left(1+\frac{P_1}{N_1}\right),\\
&0 \leq R_2 \leq \log\left(1+\frac{P_2}{N_2}\right),\\
&0 \leq R_1+R_2 \leq \min\left[\log\left(1+\frac{P_1+\chi P_2}{N_1}\right),\log\left(1+\frac{\nu P_1+P_2}{N_2}\right)\right]. \qquad (4.17)
\end{aligned}$$

4.5.4 MIMO Channel Capacity

In multi-user MIMO systems applying rate-splitting, the channel capacity characterizes the maximum achievable rate at which information can be reliably transmitted over the channel. Rate-splitting in MIMO allows the transmission of multiple message signals with different priorities. It divides the transmitted signal into two parts: a common part, decoded by both receivers, and a private part, intended for a specific receiver. The channel capacity with rate-splitting accounts for this division, considering the optimal allocation of power and transmission strategies for both the common and private messages. It defines the maximum achievable rate for each user while ensuring successful reception and decoding at the receivers, leveraging the spatial degrees of freedom provided by multiple antennas at both ends of the communication link. The analysis involves optimizing the power allocation, precoding strategies, and decoding orders to maximize the overall achievable rates while considering the interference among different users. For the MIMO interference channel capacity [22] is used for the upper bound evaluation which is indicated in 4.18, and the necessary parameters (ρ_{11}, ρ_{12}, ρ_{22}, ρ_{21}, K_1 and K_2) for computation are provided in Appendix D.

$$\begin{aligned}
R_1 &\leq \log\det\left[I_{Q_1} + \rho_{11} H_1 H_1^H\right] \\
R_2 &\leq \log\det\left[I_{Q_2} + \rho_{22} H_2 H_2^H\right] \\
R_1 + R_2 &\leq \log\det\left[I_{Q_2} + \rho_{12} G_1 G_1^H + \rho_{22} H_2 H_2^H\right] + \log\det\left[I_{Q_1} + \rho_{11} H_1 K_1 H_1^H\right] \\
R_1 + R_2 &\leq \log\det\left[I_{Q_1} + \rho_{21} G_2 G_2^H + \rho_{11} H_1 H_1^H\right] + \log\det\left[I_{Q_2} + \rho_{22} H_2 K_2 H_2^H\right] \\
R_1 + R_2 &\leq \log\det\left[I_{Q_1} + \rho_{21} G_2 G_2^H + \rho_{11} H_1 K_1 H_1^H\right] \\
&\quad + \log\det\left[I_{Q_2} + \rho_{12} G_1 G_1^H + \rho_{22} H_2 K_2 H_2^H\right] \\
2R_1 + R_2 &\leq \log\det\left[I_{Q_1} + \rho_{21} G_2 G_2^H + \rho_{11} H_1 H_1^H\right] + \log\det\left[I_{Q_1} + \rho_{11} H_1 K_1 H_1^H\right] \\
&\quad + \log\det\left[I_{Q_2} + \rho_{12} G_1 G_1^H + \rho_{22} H_2 K_2 H_2^H\right] \\
R_1 + 2R_2 &\leq \log\det\left[I_{Q_2} + \rho_{12} G_1 G_1^H + \rho_{22} H_2 H_2^H\right] + \log\det\left[I_{Q_2} + \rho_{22} H_2 K_2 H_2^H\right] \\
&\quad + \log\det\left[I_{Q_1} + \rho_{21} G_2 G_2^H + \rho_{11} H_1 K_1 H_1^H\right] \qquad (4.18)
\end{aligned}$$

When we are averaging over different channel conditions in the next section, for each channel realization we check the interference condition (weak, mixed or strong), apply the appropriate bound and then take the average.

4.6 No Interference

For this case, interference terms are assumed to be 0 to obtain a trivial upper-bound. The rates achieved by each user for a MIMO system written as

$$R_1 = \log\det\left(\boldsymbol{I}_{\boldsymbol{Q_1}} + (\boldsymbol{H}_1\boldsymbol{W}_1\boldsymbol{W}_1^H\boldsymbol{H}_1^H)(N_0\boldsymbol{I}_{\boldsymbol{Q_1}})^{-1}\right) \tag{4.19}$$

$$R_2 = \log\det\left(\boldsymbol{I}_{\boldsymbol{Q_1}} + (\boldsymbol{H}_2\boldsymbol{W}_2\boldsymbol{W}_2^H\boldsymbol{H}_2^H)(N_0\boldsymbol{I}_{\boldsymbol{Q_1}})^{-1}\right), \tag{4.20}$$

where $\boldsymbol{W}_i = \boldsymbol{W}_i^{mrt}$, for a MISO system

$$R_1 = \log\left(1 + \frac{|\boldsymbol{h}_1\boldsymbol{w}_1|^2}{N_0}\right) \tag{4.21}$$

$$R_2 = \log\left(1 + \frac{|\boldsymbol{h}_2\boldsymbol{w}_2|^2}{N_0}\right), \tag{4.22}$$

where $\boldsymbol{w}_i = \boldsymbol{w}_i^{mrt}$, and for a SISO system

$$R_1 = \log\left(1 + \frac{|h_1w_1|^2}{N_0}\right) \tag{4.23}$$

$$R_2 = \log\left(1 + \frac{|h_2w_2|^2}{N_0}\right), \tag{4.24}$$

where $w_i = w_i^{mrt}$.

CHAPTER 5

SIMULATION RESULTS

In this Chapter, we delve into an exhaustive examination of simulation results for the transmission scheme introduced in Chapter 3. This method is meticulously compared against its subset, the MADDPG with no rate-splitting [3], offering valuable insights into the role of rate-splitting in shaping the overall performance. Moreover, we extend our investigation to encompass a comprehensive set of benchmark schemes outlined in Chapter 4, covering a spectrum of scenarios including SISO, MISO and MIMO configurations. This extensive analysis allows us to examine the proposed scheme's adaptability and efficacy across diverse communication scenarios.

The benchmark schemes considered in our simulations span well-established methodologies for interference channels with multiple antennas. The comparative study systematically evaluates the proposed MADDPG with rate-splitting against these benchmarks, unraveling the nuanced intricacies of each scheme in SISO, MISO and MIMO cases. By conducting such a comprehensive set of simulations, our aim is to provide a holistic understanding of the strengths, weaknesses, and applicability of each scheme across a range of communication scenarios. This thorough examination serves as a critical step in elucidating the effectiveness and robustness of the proposed MADDPG with rate-splitting, showcasing its relative performance against alternative methodologies in diverse and challenging communication environments.

5.1 Performance Measures

Our primary objective is to maximize the sum-rate in a communication system by optimizing both the power allocation coefficient and the precoder. Maximizing the sum-rate in RSMA is crucial as it directly correlates with the overall efficiency and throughput of the communication system. A higher sum-rate implies the ability to transmit more information per unit of time, leading to improved data transfer capabilities and enhanced network performance. Utilizing the MADDPG algorithm becomes essential in this context, as it provides a framework for the coordinated learning of multiple agents to optimize the sum-rate. MADDPG enables these agents to adapt their strategies collaboratively, ensuring a balanced and effective allocation of resources, such as power and decoding orders, to maximize the sum-rate in RSMA-based communication systems. This optimization process, particularly in the context of RSMA, introduces inherent complexities. The nature of RSMA, with its simultaneous consideration of common and private streams, renders traditional optimization approaches less effective, making the utilization of RL approaches inevitable. In addition to maximizing the sum-rate, our study will delve into the performance implications of decoding order estimation, investigating how the system adapts to varying orders of decoding for transmitted messages. Furthermore, we will rigorously examine the system's resilience in the face of channel estimation errors, a crucial consideration in practical communication scenarios. This multifaceted exploration aims to provide a nuanced understanding of the proposed model's capabilities, particularly in the intricate landscape of RSMA, reinforcing the necessity of RL in addressing its inherent complexities.

5.2 Numerical Settings

The instantiation of the MADDPG algorithm was conducted within the PyTorch 1.9.1 framework. The training architecture was meticulously designed, employing four fully connected layers for both the critic and actors, thereby affording adaptability and sophistication in the learning process. The algorithm's convergence was vigilantly monitored across numerous episodes, each comprising 200 time steps, culminating in the acquisition of rates across varied SNR scenarios. Detailed specifications of pertinent values pertaining to episodes and network parameters are documented in Tables 5.1, 5.2, and 5.3, ensuring transparency and reproducibility.

Within the simulated environment, the channel coefficients h_i, $\boldsymbol{h}_i$, $\boldsymbol{H}_i$, and g_i, $\boldsymbol{g}_i$,

$\boldsymbol{G}_i$ for $i = 1, 2$ are presumed to follow an independent and identically distributed (i.i.d.) circularly symmetric complex Gaussian distribution with zero mean and unit variance. This modeling paradigm aligns with established conventions in the representation of wireless communication channels. Moreover, the introduction of channel estimation errors, a practical consideration in real-world scenarios, is systematically explored. The examination of these errors encompasses two distinct modalities: the first, wherein the estimation error dynamically fluctuates with the SNR value, and the second, where it remains invariant. The estimated channel can be represented as

$$\tilde{\boldsymbol{H}}_{\boldsymbol{i}} = \boldsymbol{H}_{\boldsymbol{i}} + \boldsymbol{E}_{\boldsymbol{i}} \tag{5.1}$$

where the resulting estimation error, denoted by $\boldsymbol{E}_{\boldsymbol{i}}$, arises from the disparity between the estimated channel coefficients, represented by $\tilde{\boldsymbol{H}}_{\boldsymbol{i}}$, and the actual channel coefficients denoted as $\boldsymbol{H}_{\boldsymbol{i}}$. It is important to note that these matrices $\boldsymbol{E}_{\boldsymbol{i}}$, $\boldsymbol{H}_{\boldsymbol{i}}$, and $\tilde{\boldsymbol{H}}_{\boldsymbol{i}}$ all belong to the complex field and have dimensions $Q_i \times M_i$, i.e. $\mathbb{C}^{Q_i \times M_i}$. In the case of varying imperfections, each entry $[\boldsymbol{E}_{\boldsymbol{i}}](i, j)$ is determined as $\frac{(SNR^{-0.6})}{5}\mathcal{CN}(0, \sigma^2)$, where $\sigma^2 = 1$. It is crucial to highlight that the estimation error solely influences the computation of precoders, while the rates are computed employing the exact channel coefficients. For clarity, this means that the actor network in each agent incorporates estimated channel coefficients, whereas the reward function is formulated based on the exact channel coefficients.

Irrespective of the antenna configuration, be it SISO, MISO, or MIMO the total power P_i in the power allocation equation (2.14c) is consistently set to 1. Additionally, for the purpose of decoding order estimation, an additional actor network is introduced for each agent across all antenna configurations, as elucidated in Chapter 3.

Our investigation will encompass a comprehensive analysis of our communication scheme across diverse SNR regimes, specifically considering the case where SNR is defined as the reciprocal of the noise power N_0 (i.e., SNR = 1/N_0), given that the signal power is fixed at 1. This SNR parameterization is consistent with the power constraint considerations outlined in Chapter 2, where N_0 represents the noise power. To maintain uniformity and adhere to power constraints, the precoders employed in each scheme will undergo normalization by their respective magnitudes. This normalization process ensures that the power constraints are consistently satisfied across various SNR scenarios, facilitating a systematic evaluation of the scheme's performance under different signal-to-noise conditions.

5.3 Simulation Results

The careful adjustment of hyperparameters is paramount in RL as it significantly influences the algorithm's performance and convergence. Hyperparameters, such as the learning rate, discount factor, exploration noise, and network architecture parameters, play a pivotal role in shaping the RL agent's behavior during training. Suboptimal hyperparameter settings can impede convergence, hinder learning, or lead to instability in the training process. Conversely, well-tuned hyperparameters are crucial for achieving a balance between exploration and exploitation, ensuring effective learning and adaptation to complex environments. Fine-tuning hyperparameters is a delicate process that requires a comprehensive understanding of the specific characteristics of the problem domain. Experimentation and iterative adjustments are essential to finding an optimal set of hyperparameters that facilitates efficient learning and robust performance of RL algorithms.

In that context, the MADDPG algorithm is strategically configured to accommodate the distinctive characteristics of SISO, MISO, and MIMO interference channel scenarios. In the SISO case in Table 5.1, where there is a single-antenna transmitters and receivers, the algorithm adopts a streamlined approach with a moderate minibatch size, hidden size, and episode length. This design prioritizes efficiency within the simplicity of a singular communication link.

As we move to the MISO scenario in Table 5.2, involving multiple-antenna transmitters and a single-antenna receivers, the algorithm undergoes targeted adjustments. These adaptations include an increased replay memory, additional training episodes, and a slight tuning of the learning rate. These modifications cater to the heightened complexity introduced by multiple transmitters feeding into a shared receiver.

The MIMO configuration in Table 5.3, characterized by multiple transmitters and receivers, demands a more intricate strategy. To address this complexity, the algorithm

Table 5.1: Hyperparameters of MADDPG algorithm for SISO

Parameter	Value	Parameter	Value
discount factor γ	0.99	optimizer	Adam
minibatch size B	64	loss function	MSE loss
replay memory length D	7000	no. of connected layers	4
activation function	ReLU	learning rate	10^{-4}
hidden size	36	episode length T	200
update rate τ	0.01	exploration noise σ_N^2	0.1
number of episodes E	2400	weight of user rates β	0.5

incorporates a larger minibatch size, an additional connected layer, and an augmented hidden size. The learning rate is further refined to strike a balance that accommodates the intricacies introduced by multiple antennas. Training is extended over a larger number of episodes, ensuring the algorithm captures the dynamics of the MIMO channel.

In essence, the tailored adjustments across SISO, MISO, and MIMO scenarios underscore the algorithm's adaptability and its ability to flexibly navigate the diverse challenges posed by different interference channel configurations. These adjustments, reflecting a understanding of each scenario, emphasize the MADDPG algorithm's robustness and efficacy in learning optimal policies suited to the intricacies of SISO, MISO, and MIMO communication setups.

In this specific setup, our system formulation aligns with the specifications detailed Section 2.1 in Chapter 2 under the SISO system characteristics. The solution methodology closely adheres to the outlined structure presented in Chapter 3, focusing specif-

Table 5.2: Hyperparameters of MADDPG algorithm for MISO

Parameter	Value	Parameter	Value
discount factor γ	0.99	optimizer	Adam
minibatch size B	64	loss function	MSE loss
replay memory length D	10000	no. of connected layers	4
activation function	ReLU	learning rate	10^{-4}
hidden size	36	episode length T	200
update rate τ	0.01	exploration noise σ_N^2	0.1
number of episodes E	4000	weight of user rates β	0.5

ically on the SISO scenario (refer to section 3.2). To evaluate the upper bound, we rely on methodologies presented in [4] and [5]. The details and specific parameters needed for this computation are extensively explained in Chapter 4. In the subsequent parts, namely parts 4.5.2, 4.5.1, and 4.5.3, these parameters are applied to the same channel coefficients utilized in the MADDPG framework during the testing phase. This structured approach ensures consistency in the assessment of the upper bound, as the channel conditions for both the upper bound evaluation and MADDPG testing are aligned, facilitating a meaningful and reliable comparison. For this particular SISO analysis, we assume that the base station is equipped with a single antenna, while each user possesses a single antenna as well. This antenna configuration enables us to thoroughly examine and assess the system performance within the defined SISO framework.

In Figure 5.1, a graphical representation illustrates the average sum-rate as a function of SNR in a scenario involving single-antenna base stations (M = 1) and two users.

Table 5.3: Hyperparameters of MADDPG algorithm for MIMO

Parameter	Value	Parameter	Value
discount factor γ	0.99	optimizer	Adam
minibatch size B	128	loss function	MSE loss
replay memory length D	15000	no. of connected layers	5
activation function	ReLU	learning rate	5×10^{-5}
hidden size	64	episode length T	200
update rate τ	0.01	exploration noise σ_N^2	0.1
number of episodes E	12000	weight of user rates β	0.5

In this setup, the system formulation is based on the configuration described in Section 2.1, while the solution methodology aligns with the structure outlined in Section 3.1. The objective is to discern the impact of rate-splitting on performance. The plot includes curves for the MADDPG algorithm both with and without rate-splitting, alongside the upper bound elucidated in Chapter 4. Notably, the results indicate that MADDPG with rate-splitting can attain the average upper bound, signifying its efficacy in approaching theoretical limits. Furthermore, as SNR increases, the disparity between the performances of MADDPG with and without rate-splitting, illustrating the rate-splitting gain, becomes more pronounced. This observation underscores the significance of rate-splitting as a strategic mechanism for augmenting the sum-rate, particularly in scenarios characterized by higher Signal-to-Noise Ratios. The MADDPG curves are derived from averaging 25 runs, each consisting of 200 time steps, post-convergence of the algorithm.

In this particular configuration, our system formulation adheres to the specifications detailed in Chapter 2 under the heading of MISO system characteristics (see section

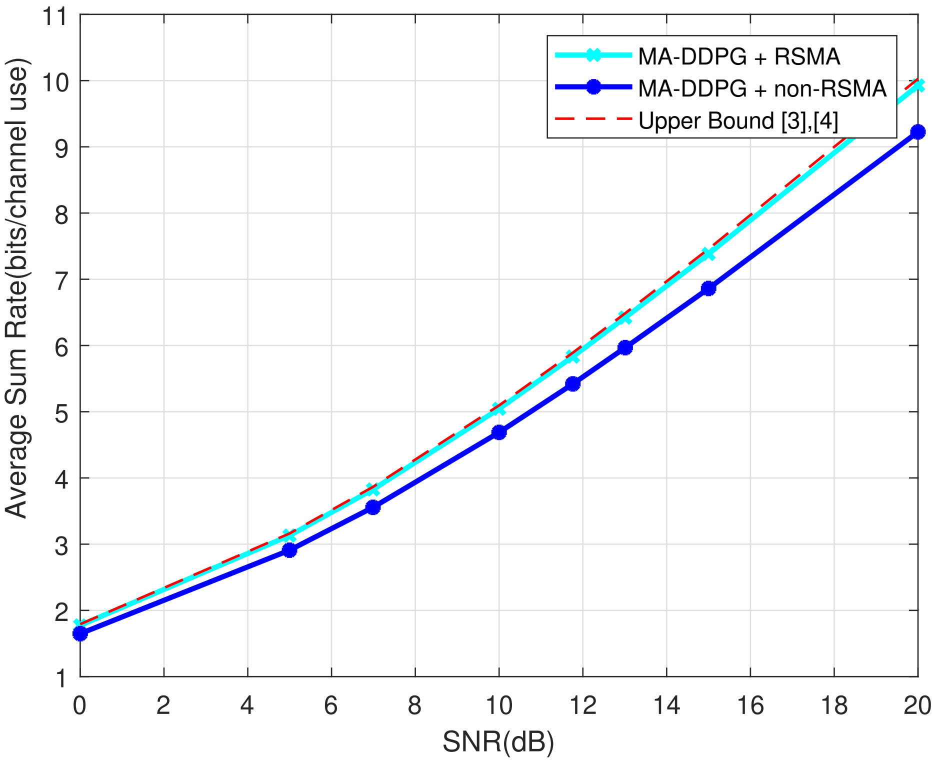

Figure 5.1: Average sum-rate achieved by MADDPG and the upper bound due to [4] and [5] for single-antenna base stations, $M = 1$, and two users each equipped with single antenna ($Q = 1$). The MADDPG curves are obtained by averaging 25 runs, each having 200 time steps after the algorithm achieves convergence.

2.2). The solution methodology, on the other hand, closely follows the outlined structure presented in Chapter 3, specifically addressing the MISO scenario (refer to Section 3.2). To assess the upper bound, we employ methodologies outlined in [22]. The pertinent parameters for computation are extensively discussed in Chapter 4. Subsequently, in the following part, namely part 4.5.4, these parameters are applied to the identical channel coefficients used in the MADDPG framework during the testing phase. This systematic approach ensures uniformity in the upper bound assessment, as the channel conditions for both the upper bound evaluation and MADDPG testing are harmonized, allowing for a meaningful and reliable comparison. For this specific MISO analysis, we assume that the base station is equipped with three antennas, while each user possesses a single antenna. This antenna configuration allows us to explore and evaluate the system performance under the outlined MISO setup, providing insights into the behavior and efficacy of the proposed solution in this specific antenna configuration.

In examining the convergence behavior of the proposed algorithm in the MISO case, a comprehensive analysis is conducted by plotting the convergence curves. In Figure 5.2, we present the average sum-rate versus the number of training episodes for the MISO scenario, and for a more insightful comparison, we include the upper bound as well. The convergence curves depict the performance evolution of the MADDPG algorithm, providing a visual representation of its learning trajectory over the course of training episodes. The incorporation of the upper bound serves as a benchmark, elucidating the algorithm's proximity to the theoretical limits. This visual analysis not only facilitates an assessment of the algorithm's convergence speed but also provides insights into how well the proposed solution converges towards the upper bound, shedding light on the effectiveness of the MADDPG algorithm in addressing the challenges posed by the MISO configuration. The convergence curve was drawn only for the MISO case, as a deliberate decision to avoid overwhelming the context with an excess of convergence curves and to maintain clarity and focus in the analysis.

In Figure 5.3, we illustrate the learning curve depicting the evolution of the weighted sum-rate under a 10 dB SNR. The learning curve analysis allows us to explore the rate region by varying the weight parameter (β). By observing how the system's performance changes with different beta values, we gain insights into the impact of user rate weights on the overall learning process. This provides a comprehensive view of the algorithm's behavior across a range of rate configurations, aiding in the assessment of its adaptability and responsiveness to changes in user rate priorities. The training process involves averaging our model over 100 runs, each comprising 1000 time steps realizations. Notably, this learning curve considers scenarios with no interference and includes an upper bound evaluation. In this context, the weighting parameter (β) is set to 0.5, signifying an equal weight distribution for each user. The choice of β plays a crucial role in determining the emphasis placed on individual user rates within the sum-rate optimization process. The learning curve provides insights into the convergence and performance of the proposed approach under these specified conditions, offering a comprehensive view of the training dynamics and the achieved sum-rate outcomes.

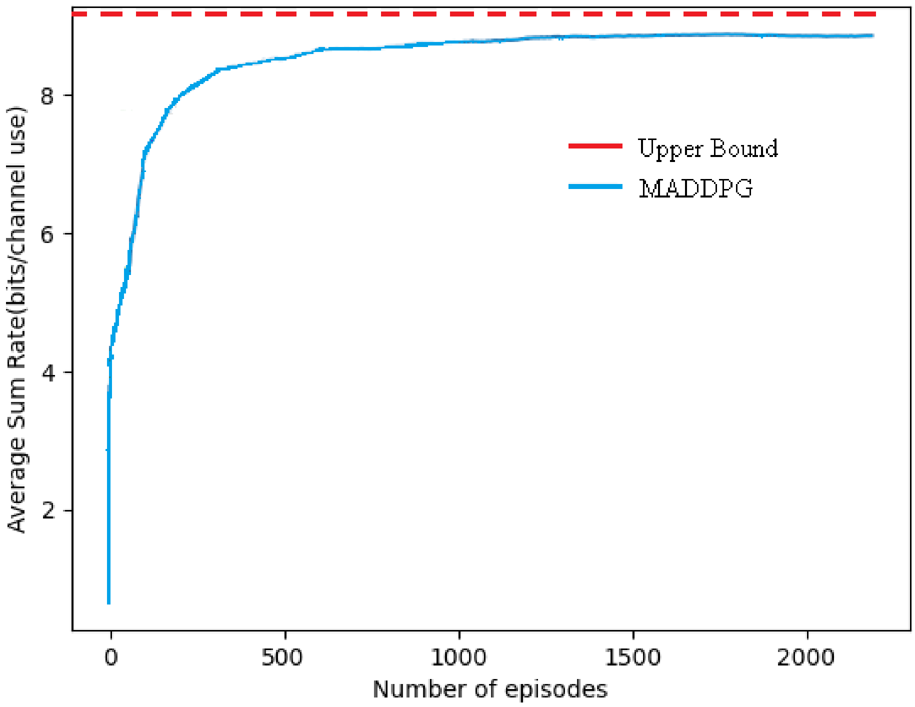

Figure 5.2: Convergence curve achieved by MADDPG for multiple-antenna base stations, ($M = 3$), and two users each equipped with single antenna ($Q = 1$) when $SNR = 10\,dB$.

In Figure 5.4, we present a detailed analysis of the average sum-rate results for $M = 3$. The comparison involves multiple schemes, including MADDPG with and without rate-splitting, maximum ratio transmission (MRT), zero-forcing (ZF), and leakage-based precoding. Additionally, the upper bound, as defined in [22] and provided in 4, serves as a reference for assessing the performance achieved by the proposed approach. Analyzing the results, it is evident that MADDPG with rate-splitting outperforms its no rate-splitting counterpart, which illustrates the rate-splitting gain and exhibits superior performance compared to MRT, ZF, and leakage-based precoding. MRT experiences challenges in the presence of severe interference, leading to a convergent behavior rather than maintaining an increasing average sum-rate curve. On the other hand, ZF, while immune to interference, struggles to achieve sufficiently high signal power, resulting in limited performance. Leakage-based precoding strikes a balance between desired signal power and leakage power, leading to higher rates

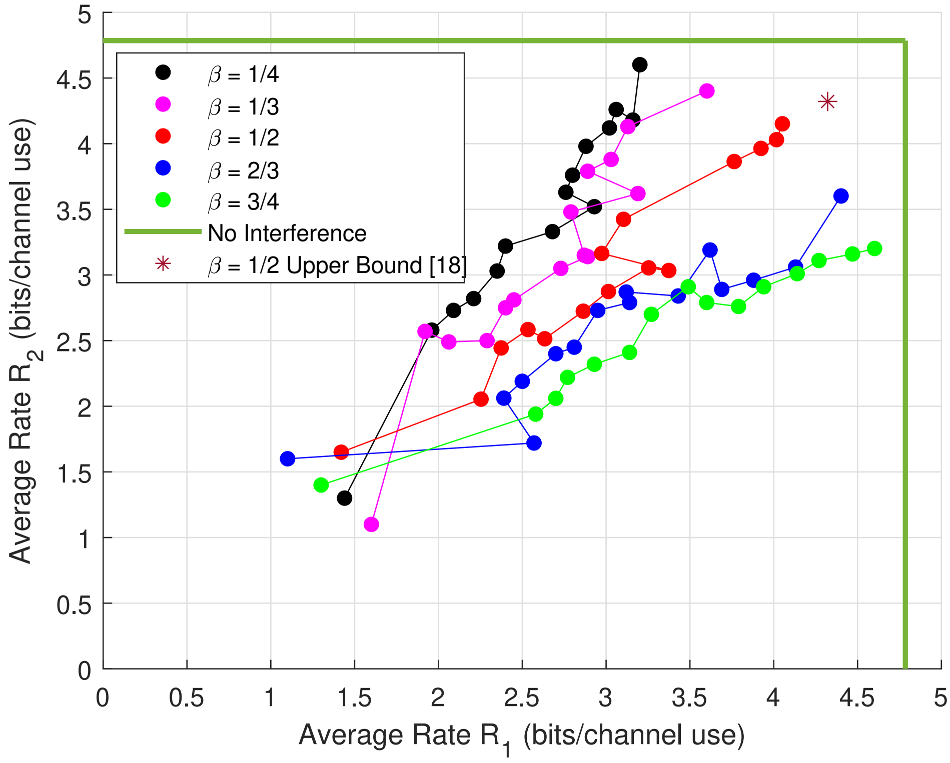

Figure 5.3: Evolution of the weighted sum-rate for MADDPG with rate-splitting for SNR = 10 dB and for varying β defined in (2.14a). The skip from one dot to the next represents 100 episodes of training, with the dots appearing after a delay of 500 episodes.

than both MRT and ZF. The unique advantages of MADDPG with rate-splitting become apparent in this comparison. Firstly, MADDPG utilizes the SINR as a metric, which is more relevant for evaluating system performance. Secondly, the incorporation of rate-splitting enables MADDPG to intelligently manage interference. In scenarios with weak interference, more power is allocated to private messages, while in the presence of strong interference, common messages are transmitted with higher power. The consideration of all possible decoding orders further enhances the adaptability of MADDPG with rate-splitting, resulting in superior performance compared to benchmark schemes. This comprehensive analysis sheds light on the nuanced benefits and capabilities of the proposed approach in addressing interference challenges in multi-antenna scenarios.

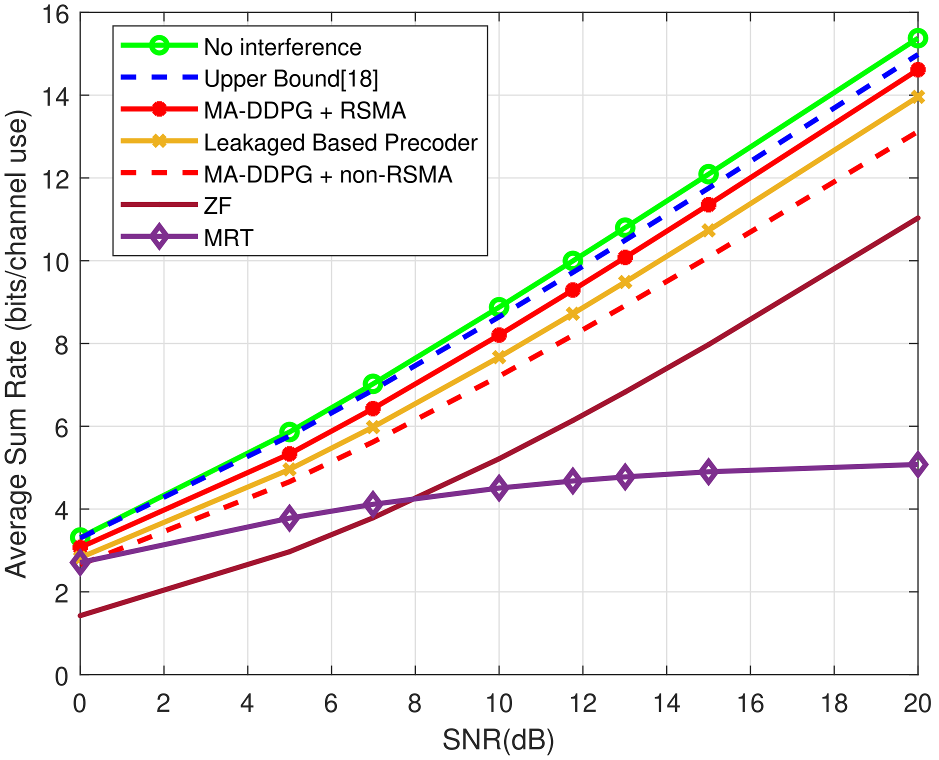

Figure 5.4: Average sum-rate achieved by MADDPG and the benchmark schemes for three-antenna base stations, $M = 3$, and two users each equipped with single antenna ($Q = 1$). The MADDPG curves are obtained by averaging 50 runs, each having 1000 time steps after the algorithm achieves convergence.

The confidence bounds of the RL results, specifically obtained through the MADDPG algorithm, showcase the robustness and reliability of the proposed approach. In assessing the performance of the system, the inclusion of confidence bounds adds a layer of statistical significance to the obtained results. The confidence bounds were derived by leveraging the information gathered from multiple runs of the MADDPG algorithm. To calculate the confidence bounds, the results were averaged over a significant number of runs, providing a representative performance metric. Subsequently, the standard deviation of each run was computed. By adding and subtracting this standard deviation from the averaged line, a confidence bound was established, offering insights into the variability and consistency of the RL algorithm's performance across different runs. This approach to confidence bounds is particularly valuable in reinforcing the reliability of the observed trends and outcomes. The incorporation of upper bounds in the comparison further contextualizes the RL results, allowing for a nuanced understanding of the algorithm's proximity to the theoretical performance limits. The robustness of the confidence bounds, derived through careful statistical analysis, enhances the credibility of the RL results, providing stakeholders and researchers with a comprehensive perspective on the algorithm's performance under varying conditions. The comprehensive assessment of the confidence bounds, as illustrated in Figure 5.5, augments the overall reliability and robustness of the acquired results derived from the MADDPG algorithm. The promising aspect of these confidence bounds becomes more pronounced when considering the extensive nature of the evaluation, encompassing a total of 1000 runs, each comprising 1000 time steps. After 2500 episodes, the algorithm consistently produces similar results. This stability and consistency in performance indicate the robustness of our proposed solution. The algorithm's ability to maintain comparable outcomes across multiple episodes highlights its reliability in handling the complexities of the MISO scenario. This robust behavior is crucial for ensuring the dependable and consistent performance of the algorithm under varying conditions, contributing to its practical applicability and effectiveness in real-world settings.

In Figure 5.6, we systematically investigate the impact of a fixed value of imperfection over SNR, as defined in equation 5.1 in section 5.2. The imperfection is set as $\frac{(10^{-0.6})}{5}\mathcal{CN}(0,\sigma^2)$ with $\sigma^2 = 1$. Additionally, Figure 5.7 explores the scenario where imperfection varies according to $\frac{(SNR^{-0.6})}{5}\mathcal{CN}(0,\sigma^2)$, with the same $\sigma^2 = 1$. In these analyses, the resilience and adaptability of the proposed RSMA framework, especially MADDPG with rate-splitting, are highlighted. This quality becomes particularly pronounced when compared against alternative schemes, including ZF, MRT, leakage-based precoding, and non-rate-splitting MADDPG. The intricate architecture of the RSMA system, with its unique rate-splitting mechanism, exhibits exceptional proficiency in alleviating the deleterious effects of channel estimation errors. The inherent resilience observed in RSMA translates into a robust and dependable performance, showcasing the superiority of the proposed framework under diverse SNR conditions. This comparative analysis reinforces the adaptability and reliability of RSMA, positioning it as a promising solution in realistic communication scenarios where channel imperfections are prevalent and challenging to address effectively.

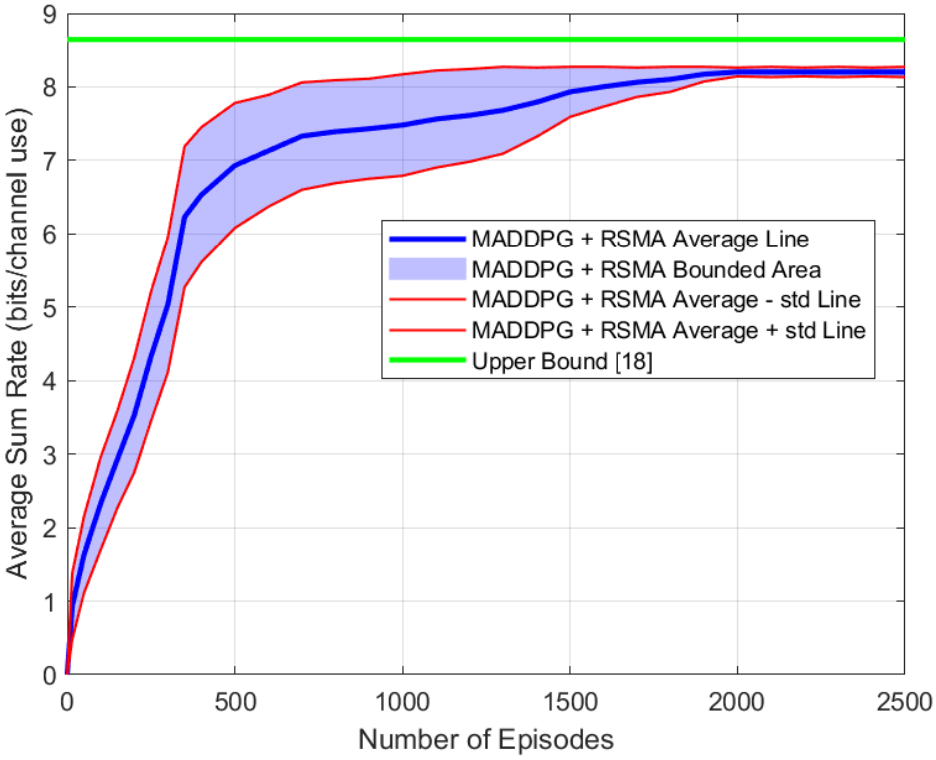

Figure 5.5: Confidence bound achieved by MADDPG for three-antenna base stations, $M = 3$, and two users each equipped with single antenna ($Q = 1$). The confidence interval are obtained by averaging 1000 runs, each having 1000 time steps after the algorithm achieves convergence.

In scenarios with imperfect channel state information at the transmitter (CSIT), the conventional schemes—ZF, MRT, leakage-based precoding, and non-rate-splitting MADDPG—encounter challenges in maintaining their performance due to their limited adaptability to variations in channel conditions. On the contrary, MADDPG with rate-splitting, displays a remarkable capacity to navigate through the complexities introduced by imperfect CSIT. When there are channel estimation errors, the actual channel conditions may differ from the estimated ones. SIC helps in handling these errors by iteratively canceling the interference, which becomes especially valuable in scenarios where accurate channel information is challenging to obtain. As a result, RSMA with SIC demonstrates resilience to channel estimation errors, making it a robust choice for communication systems, particularly in the presence of imperfect channel knowledge. The ability to effectively manage and mitigate the impact

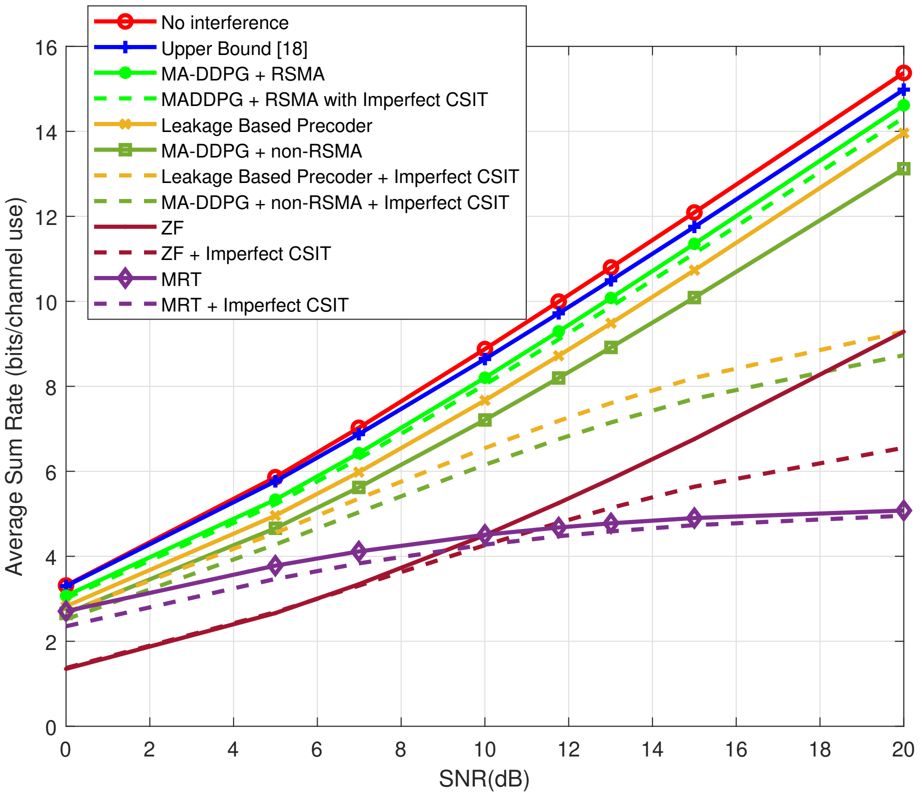

Figure 5.6: Channel estimation under fixed imperfection over SNR achieved by MADDPG for three-antenna base stations, $M = 3$, and two users each equipped with single antenna ($Q = 1$).

of fixed channel estimation errors marks a pivotal distinction for RSMA, reinforcing its standing as a robust and reliable solution in real-world and demanding communication environments. This robustness in the face of imperfect CSIT not only underscores the adaptability of MADDPG with rate-splitting but also positions it as a preferred choice for communication systems, where accurate channel information is challenging to obtain. The enhanced performance and reliability exhibited by RSMA in the presence of imperfect CSIT contribute to its broader applicability, especially in practical communication scenarios where environmental conditions and system uncertainties necessitate a robust and adaptable solution.

In the context of decoding order estimation, an additional actor network is introduced for each agent. Our investigation into the performance of this decoding order estimation involves two key scenarios. Firstly, we examine the case where there is a

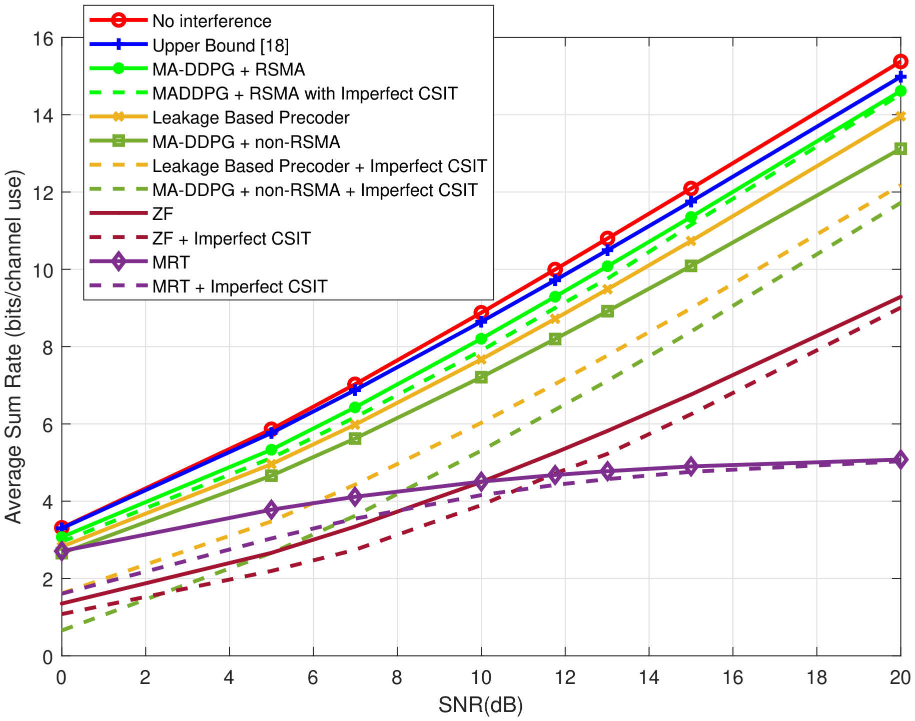

Figure 5.7: Channel estimation under varying imperfection over SNR achieved by MADDPG for three-antenna base stations, $M = 3$, and two users.

fixed value of channel estimation error over SNR, as defined earlier in this Chapter. The second scenario considers a situation with no channel estimation error. Remarkably, our findings reveal that decoding order estimation exhibits superior performance across different SNR regimes. This observation is particularly noteworthy as it underscores the robustness and efficiency of the proposed MADDPG framework under decoding order estimation. Decoding order estimation is a critical aspect, especially considering the complexity introduced by the RSMA structure, particularly in the SIC part. The significance of this result lies in the ability of MADDPG to maintain superior performance even with decoding order estimation, addressing a crucial concern related to the complexity of RSMA, particularly in the SIC component. This suggests that by incorporating decoding order estimation, the system can achieve competitive performance while mitigating the associated complexity. This nuanced exploration of decoding order estimation adds valuable insights to the understanding of RSMA's adaptability and efficacy in addressing practical challenges in communication systems.

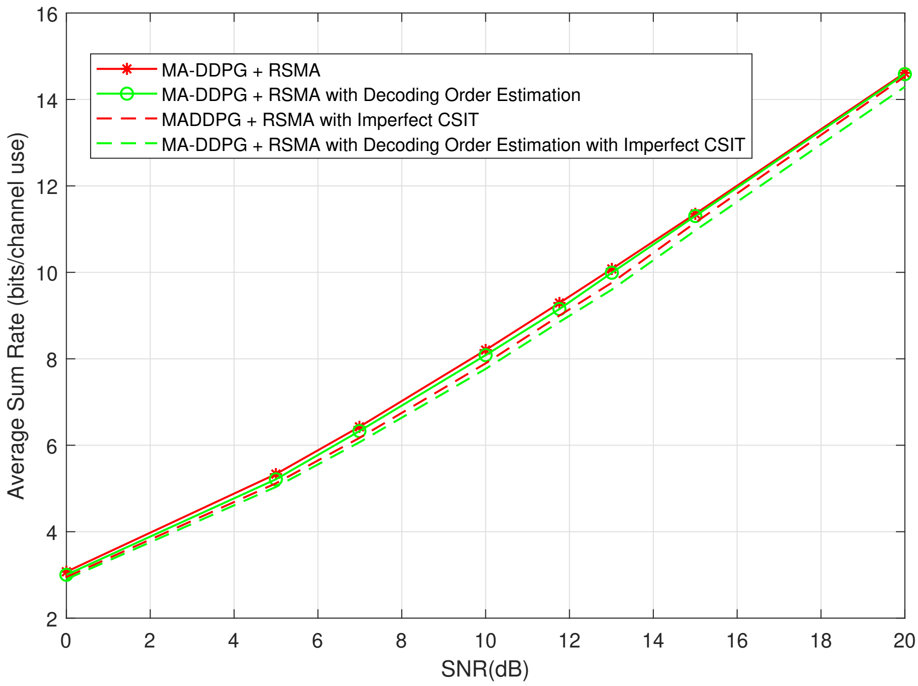

Figure 5.8: Decoding order estimation achieved by MADDPG for three-antenna base stations, $M = 3$, and two users each equipped with single antenna ($Q = 1$).

These insightful findings regarding the superior performance of decoding order estimation, particularly in the presence of fixed channel estimation errors and across various SNR regimes, are visually depicted in Figure 5.8. The depicted results provide a clear and illustrative representation of the robustness and effectiveness of the proposed MADDPG framework in addressing the complexities associated with decoding order estimation in the context of the Rate-Splitting Multiple Access (RSMA) structure.

In our investigation, we observed a widening gap between the training and test performance as we progressed from SISO to MISO and MIMO scenarios. This increasing gap can be attributed to the growing number of parameters that need to be estimated as we move to more complex antenna configurations. The inherent challenges of training a model with an expanding parameter space contribute to this performance gap.

Furthermore, we noted a noteworthy trend in the decreasing performance gap between

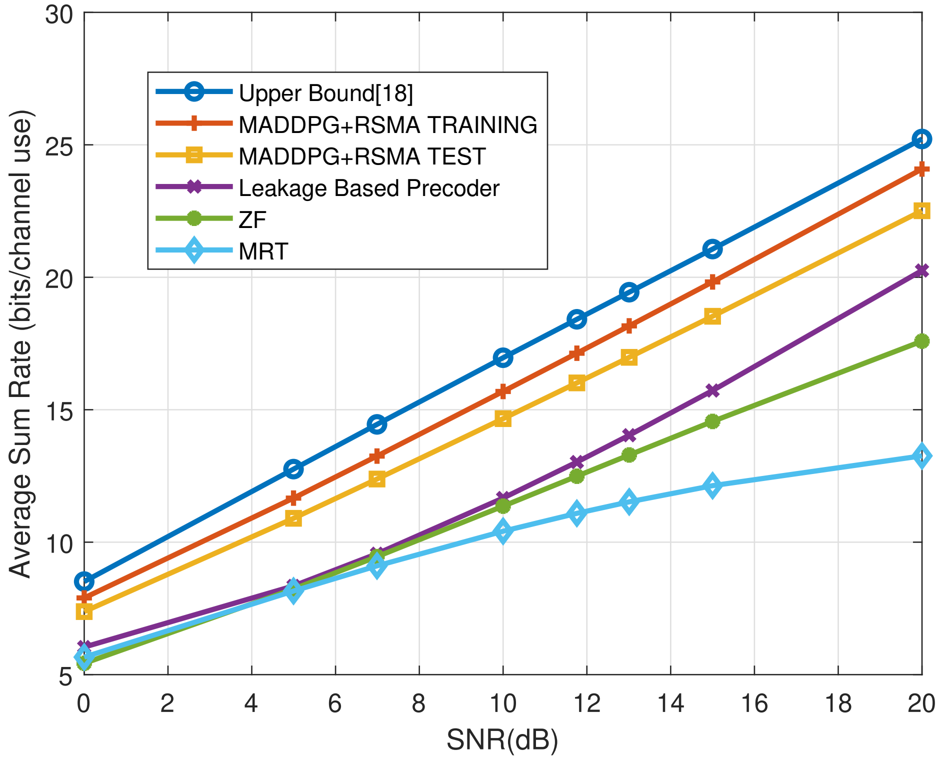

Figure 5.9: The average sum-rate achieved by MADDPG algorithm, along with benchmark schemes, is evaluated in the context of a communication system with three-antenna base stations ($M = 3$) and two users each equipped with three antennas ($Q = 3$).

Zero Forcing (ZF) and the proposed MADDPG+RSMA method as we transitioned from MISO to MIMO scenarios. This phenomenon can be explained by the inherent characteristics of ZF and its adaptability to the increased complexity of MIMO systems. ZF is known for its ability to mitigate interference in multi-antenna scenarios, and as the number of antennas increases in MIMO configurations, ZF may be better suited to exploit spatial diversity and handle interference effectively. The observed improvement in ZF's performance through MISO to MIMO scenarios could be attributed to its inherent capabilities in these complex setups.

Examining the number of training episodes across SISO, MISO, and MIMO configurations, a consistent upward trend was observed. This escalation in training episodes is a consequence of the heightened complexity inherent in transitioning from SISO to MIMO configurations. The increase in the number of antennas and parameters in the communication system demands more extensive training to optimize the reinforcement learning model effectively. The complexity of MIMO scenarios introduces additional intricacies, requiring the learning agent to navigate a larger action space and gain a more nuanced understanding of the environment. Consequently, the iterative learning process is extended to accommodate the increased complexity and ensure the convergence of the reinforcement learning algorithm to an optimal policy.

In the context of a multiple antenna scenario, the challenges associated with estimating parameters become more pronounced, leading to a potential decrease in estimation performance. As the number of antennas increases, the complexity of the communication system grows, and a higher number of parameters need to be estimated. This heightened complexity introduces additional intricacies in capturing and modeling the nuances of the channel, making parameter estimation a more challenging task. Consequently, the performance of parameter estimation may exhibit a decline in accuracy due to the increased dimensionality and complexity of the estimation problem.

Moreover, the observed differences between the results obtained during training and testing phases can be attributed to the challenges posed by the growing number of parameters. During the training phase, the algorithm adapts to the training data, attempting to learn the underlying patterns and relationships within the given parameter space. However, when the model encounters the testing phase with a potentially different set of conditions, including different channel realizations or environmental factors, the generalization performance may vary. The discrepancies between training and testing results highlight the impact of the escalating parameter count in massive MIMO scenarios and underscore the need for robust algorithms capable of handling the inherent complexities introduced by massive MIMO configurations.

In Figure 5.9 we present a detailed analysis of the average sum-rate results for $M = 3$ and $Q = 3$, our proposed MADDPG+RSMA method consistently outperforms benchmark schemes such as ZF, MRT, and Leakage-Based Precoder. This superiority can be attributed to the unique advantages offered by RSMA. By intelligently allocating power and managing interference, RSMA ensures more efficient resource utilization, resulting in enhanced overall system performance. The robustness and adaptability of our proposed method in the face of increasing complexity and interference make it a promising solution for advanced communication scenarios.

CHAPTER 6

CONCLUSION

6.1 Conclusions

RSMA has become a remarkable transmission strategy in the development of communication systems, especially in the transitions from 5G to the anticipated advancements in 6G. This research delves into the complexities of RSMA, specifically focusing on precoding within a multiple-antenna interference channel, and employs DRL techniques. The primary objective is to optimize precoders and manage transmit power for both common and private data streams. Successfully addressing this challenge, particularly in the continuous action space, necessitates the collaboration of multiple decision-makers. The choice of utilizing DRL, as opposed to other machine learning methods, is crucial. DRL proves critical in handling the intricate nature of RSMA tasks, offering a more adaptive and efficient approach to mitigating interference in modern communication systems.

Our approach tackles the complex optimization landscape of multiple antenna interference channels employing rate-splitting strategies by harnessing a MADDPG framework. The MADDPG algorithm, specifically tailored for the optimization of precoding and power allocation coefficients, exhibits a distinctive framework that allows for centralized learning while enabling decentralized execution. This dual capability makes a significant contribution, providing a decentralized and scalable solution for interference management without the need for constant coordination from a central entity. This adaptability and robustness enhance the algorithm's effectiveness in real-world scenarios, making it a valuable tool for addressing the challenges posed by interference in multiple antenna environments. The results from our simulations highlight the effectiveness of the rate-splitting method we proposed. In cases with a single antenna, our approach achieves the information-theoritical sum-rate upper bound. Even in scenarios with multiple antennas, it demonstrates impressive proximity to the upper bound. Additionally, our method outperforms alternative approaches such as MADDPG without rate-splitting, maximal ratio transmission, zero-forcing, and leakage-based precoding in both single and multiple antenna cases, showcasing its superior performance in terms of sum-rate.

In addition to these, our work delves into the intricate aspects of channel estimation errors and optimal decoding order selection within the learning algorithm. These deliberate considerations contribute significantly to the algorithm's robustness and expand its applicability to various scenarios. Our comprehensive analysis demonstrates the superior performance of our proposed approach compared to other baseline schemes, particularly in the contexts of decoding order and channel estimation cases. In terms of channel estimation, RSMA proves to be robust, showcasing resilience against errors in estimating the channel conditions. Moreover, the incorporation of optimal decoding order selection eliminates considerable complexity, further enhancing the efficiency and practicality of the proposed algorithm in managing interference and decoding strategies for multiple messages.

We have also investigated the learning curve, accompanied by the variation in the weight of the user, serves as a powerful tool to explore the rate region systematically, providing insights into the algorithm's adaptability across different scenarios. This dynamic approach enables us to understand how the system performance varies with changes in the weight assigned to users, shedding light on the algorithm's flexibility in different user-centric scenarios. Additionally, the inclusion of a confidence bound in our results signifies the reliability and robustness of our proposed algorithm. The confidence bound acts as a measure of certainty in the algorithm's performance, bolstering confidence in its consistency across various conditions. In conclusion, the learning curve and confidence bound collectively underscore the versatility, reliability, and robustness of our proposed MADDPG algorithm, making it a promising solution for optimizing precoding and power allocation coefficients in multiple antenna interference channels employing rate-splitting strategies.

REFERENCES

[1] Y. Mao, B. Clerckx, and V. Li, "Rate-splitting multiple access for cooperative multi-cell networks," *arXiv preprint arXiv:1804.10516*, 2018.

[2] R. Amiri, H. Mehrpouyan, L. Fridman, R. K. Mallik, A. Nallanathan, and D. Matolak, "A machine learning approach for power allocation in HetNets considering QoS," in *2018 IEEE international Conference on Communications (ICC)*, pp. 1–7, IEEE, 2018.

[3] H. Lee and J. Jeong, "Multi-agent deep reinforcement learning (MADRL) meets multi-user MIMO systems," *2021 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2021.

[4] H. Sato, "The capacity of the Gaussian interference channel under strong interference," *IEEE Transactions on Information Theory*, vol. 27, no. 6, pp. 786–788, 1981.

[5] R. H. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5534–5562, 2008.

[6] R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. Pieter Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[7] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021.

[8] Y. Lu and X. Zheng, "6G: A survey on technologies, scenarios, challenges, and the related issues," *Journal of Industrial Information Integration*, vol. 19, p. 100158, 2020.

[9] M. U. A. Siddiqui, H. Abumarshoud, L. Bariah, S. Muhaidat, M. A. Imran, and L. Mohjazi, "URLLC in beyond 5G and 6G networks: An interference management perspective," *IEEE Access*, 2023.

[10] Z. Wei, H. Qu, Y. Wang, X. Yuan, H. Wu, Y. Du, K. Han, N. Zhang, and Z. Feng, "Integrated sensing and communication signals towards 5G-A and 6G: A survey," *IEEE Internet of Things Journal*, 2023.

[11] F. Salahdine, T. Han, and N. Zhang, "5G, 6G, and beyond: Recent advances and future challenges," *Annals of Telecommunications*, pp. 1–25, 2023.

[12] B. Clerckx, Y. Mao, E. A. Jorswieck, J. Yuan, D. J. Love, E. Erkip, and D. Niyato, "A primer on rate-splitting multiple access: Tutorial, myths, and frequently asked questions," *IEEE Journal on Selected Areas in Communications*, 2023.

[13] Z. Lin, M. Lin, T. De Cola, J.-B. Wang, W.-P. Zhu, and J. Cheng, "Supporting IoT with rate-splitting multiple access in satellite and aerial-integrated networks," *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11123–11134, 2021.

[14] Y. Mao, O. Dizdar, B. Clerckx, R. Schober, P. Popovski, and H. V. Poor, "Rate-splitting multiple access: Fundamentals, survey, and future research trends," *IEEE Communications Surveys & Tutorials*, 2022.

[15] X. Li, T. Wang, H. Tong, Z. Yang, Y. Mao, and C. Yin, "Sum-rate maximization for active RIS-aided downlink RSMA system," *arXiv preprint arXiv:2301.12833*, 2023.

[16] A. Z. Yalcin, M. K. Cetin, and M. Yuksel, "Max-min fair precoder design and power allocation for MU-MIMO NOMA," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 6217–6221, 2021.

[17] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[18] H. Dahrouj, R. Alghamdi, H. Alwazani, S. Bahanshal, A. A. Ahmad, A. Faisal, R. Shalabi, R. Alhadrami, A. Subasi, M. T. Al-Nory, *et al.*, "An overview of machine learning-based techniques for solving optimization problems in communications and signal processing," *IEEE Access*, vol. 9, pp. 74908–74938, 2021.

[19] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[20] A. El Gamal and T. M. Cover, “Multiple user information theory,” *Proceedings of the IEEE*, vol. 68, no. 12, pp. 1466–1483, 1980.

[21] T. Han and K. Kobayashi, “A new achievable rate region for the interference channel,” *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 49–60, 1981.

[22] S. Karmakar and M. K. Varanasi, “The capacity region of the MIMO interference channel and its reciprocity to within a constant gap,” *IEEE Transactions on Information Theory*, vol. 59, no. 8, pp. 4781–4797, 2013.

[23] M. Wu, Z. Gao, Y. Huang, Z. Xiao, D. W. K. Ng, and Z. Zhang, “Deep learning-based rate-splitting multiple access for reconfigurable intelligent surface-aided tera-hertz massive mimo,” *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 5, pp. 1431–1451, 2023.

[24] O. Dizdar, Y. Mao, Y. Xu, P. Zhu, and B. Clerckx, “Rate-splitting multiple access for enhanced URLLC and eMBB in 6G,” in *2021 17th International Symposium on Wireless Communication Systems (ISWCS)*, pp. 1–6, IEEE, 2021.

[25] M. Rahmani, M. J. Dehghani, P. Xiao, M. Bashar, and M. Debbah, “Multi-agent reinforcement learning-based pilot assignment for cell-free massive MIMO systems,” *IEEE Access*, vol. 10, pp. 120492–120502, 2022.

[26] H. Lee, M. Girnyk, and J. Jeong, “Deep reinforcement learning approach to MIMO precoding problem: Optimality and robustness,” *arXiv preprint arXiv:2006.16646*, 2020.

[27] N. Q. Hieu, D. T. Hoang, D. Niyato, and D. I. Kim, “Optimal power allocation for rate splitting communications with deep reinforcement learning,” *IEEE Wireless Communications Letters*, vol. 10, no. 12, pp. 2820–2823, 2021.

[28] J. Huang, Y. Yang, L. Yin, D. He, and Q. Yan, “Deep reinforcement learning-based power allocation for rate-splitting multiple access in 6G LEO satellite communication system,” *IEEE Wireless Communications Letters*, vol. 11, no. 10, pp. 2185–2189, 2022.

[29] S. Naser, A. A. Sani, and S. Muhaidat, “Deep reinforcement learning for RSMA-based multi-functional wireless networks,” *Authorea Preprints*, 2023.

[30] F. B. Mismar, B. L. Evans, and A. Alkhateeb, “Deep reinforcement learning for 5G networks: Joint beamforming, power control, and interference coordination,” *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1581–1592, 2019.

[31] M. Vaezi, X. Lin, H. Zhang, W. Saad, and H. V. Poor, “Deep reinforcement learning for interference management in UAV-based 3D networks: Potentials and challenges,” *arXiv preprint arXiv:2305.07069*, 2023.

[32] M. Diamanti, G. Kapsalis, E. E. Tsiropoulou, and S. Papavassiliou, “Energy-efficient rate-splitting multiple access: A deep reinforcement learning-based framework,” *IEEE Open Journal of the Communications Society*, 2023.

[33] S. Muy, D. Ron, and J.-R. Lee, “Energy efficiency optimization for SWIPT-based D2D-underlaid cellular networks using multiagent deep reinforcement learning,” *IEEE Systems Journal*, vol. 16, no. 2, pp. 3130–3138, 2021.

[34] D.-T. Hua, Q. T. Do, N.-N. Dao, T.-V. Nguyen, D. S. Lakew, and S. Cho, “Learning-based reconfigurable intelligent surface-aided rate-splitting multiple access networks,” *IEEE Internet of Things Journal*, 2023.

[35] C. Huang, R. Mo, and C. Yuen, “Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning,” *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1839–1850, 2020.

[36] Z. Tang, X. Zhu, H. Zhu, and H. Xu, “Energy efficient optimization algorithm based on reconfigurable intelligent surface and rate splitting multiple access for 6G multicell communication system,” *IEEE Internet of Things Journal*, 2023.

[37] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous control with deep reinforcement learning,” *arXiv preprint arXiv:1509.02971*, 2015.

[38] H. Albinsaid, K. Singh, S. Biswas, and C.-P. Li, “Multi-agent reinforcement learning-based distributed dynamic spectrum access,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 2, pp. 1174–1185, 2021.

[39] V. Mnih, K. Kavukcuoglu, D. Silver, *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[40] E. A. Jorswieck, E. G. Larsson, and D. Danev, "Complete characterization of the pareto boundary for the MISO interference channel," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5292–5296, 2008.

[41] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1711–1721, 2007.

[42] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[43] K. Zhang, Z. Yang, and T. Başar, "Multi-agent reinforcement learning: A selective overview of theories and algorithms," *Handbook of reinforcement learning and control*, pp. 321–384, 2021.

[44] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[45] H.-F. Chong, M. Motani, H. K. Garg, and H. El Gamal, "On the Han–Kobayashi region for the interference channel," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3188–3195, 2008.

[46] A. Raja, V. M. Prabhakaran, and P. Viswanath, "The two user Gaussian compound interference channel," in *2008 IEEE International Symposium on Information Theory*, pp. 569–573, IEEE, 2008.

[47] I. Sason, "On achievable rate regions for the Gaussian interference channel," *IEEE transactions on information theory*, vol. 50, no. 6, pp. 1345–1356, 2004.

[48] L. D. Davisson, "Rate-distortion theory and applications," *Proceedings of the IEEE*, vol. 60, no. 7, pp. 800–808, 1972.

[49] G. Bodenstein and H. M. Praetorius, "Feature extraction from the electroencephalogram by adaptive segmentation," *Proceedings of the IEEE*, vol. 65, no. 5, pp. 642–652, 1977.

[50] A. C. Sanderson and B. Kobler, “Sequential interval histogram analysis of non-stationary neuronal spike trains,” *Biological Cybernetics*, vol. 22, no. 2, pp. 61–71, 1976.

[51] J. Anderson and C.-W. Law, “Real-number convolutional codes for speech-like quasi-stationary sources,” *IEEE Transactions on Information Theory*, vol. 23, no. 6, pp. 778–782, 1977.

Appendix A

APPENDIX-1

In this section, we will break down the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm which is outlined in [6], exploring how it works and understanding its fundamental structure for learning policies in a multi-agent setting. We start our explanation by introducing DDPG, which serves as a foundation and single-agent version preceding the broader MADDPG framework. In recent years, there has been a surge in efforts to tackle challenges in RL, particularly in scenarios reflecting real-life complexities. The focus of many of these approaches has shifted towards frameworks involving multiple agents. When dealing with scenarios like robotics, it becomes apparent that considerations must extend beyond the individual agent to include interactions with other agents, such as robots or humans, sharing the same environment. This category of algorithms is commonly known as MADRL. For a comprehensive overview of MADRL, [43] serves as a valuable survey.

Most recent achievements in RL primarily pertain to single-agent scenarios, where an agent takes a series of sequential actions influencing the environment, receiving a new state and a corresponding reward. The environment is typically modeled as a MDP, aiming to find a policy mapping from the state space to an action space distribution for maximizing cumulative rewards. However, the assumption of a stationary environment doesn't extend well to multi-agent frameworks, where the actions of each agent impact the states and rewards of others. In MADRL, the problem is treated as a generalization of MDP called Markov Game (MG). Traditional RL methods like Q-Learning or policy gradient face challenges in adapting to multi-agent scenarios due to issues like changing environments affecting learning stability and high variance gradient estimates in coordination-dependent scenarios.

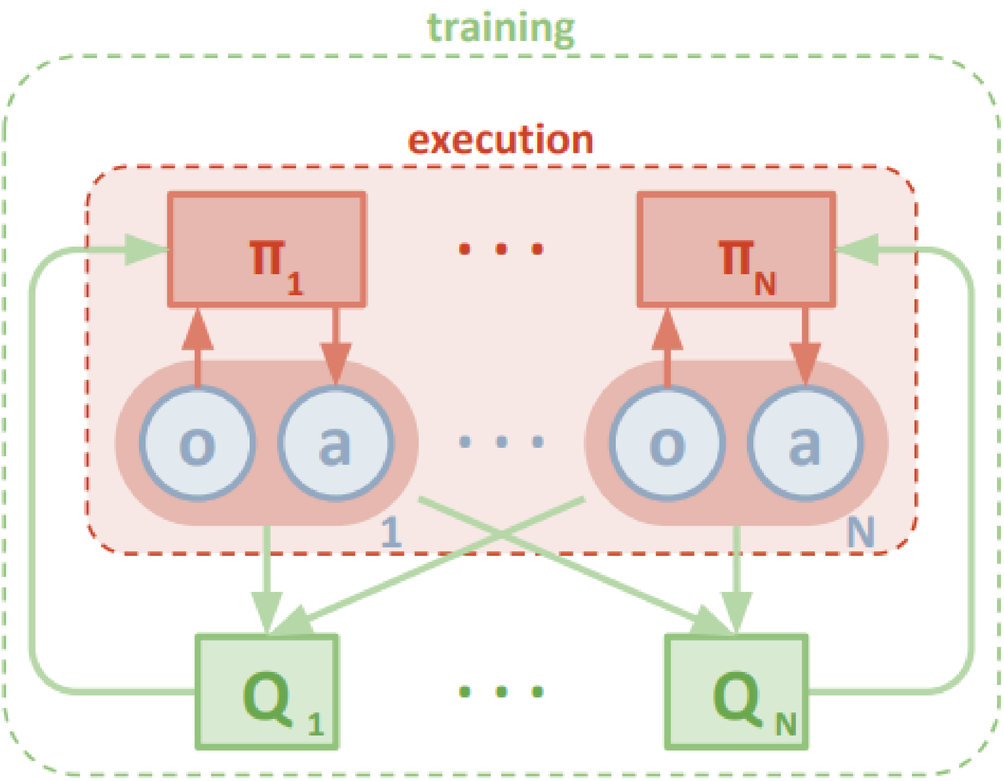

Figure A.1: Overview of multi-agent decentralized actor, centralized critic approach [6].

In policy gradient methods, we directly update the policy. This policy essentially maps the state space to a probability distribution over the action space. The updating process usually involves using a form of gradient ascent, specifically designed for a performance function, which is commonly associated with the state value function:

$$\nabla_\theta J(\theta) = \mathbb{E}_{s \sim p^\pi, a \sim \pi_\theta} \left[\nabla_\theta \log \boldsymbol{\pi}_\theta(a \mid s) Q^\pi(s, a) \right]$$

DDPG serves as a variation of Deterministic Policy Gradient, where we utilize deep neural networks to approximate both the deterministic policy and the critic. This algorithm operates off-policy and incorporates a replay buffer, which necessitates a continuous action space. It's worth noting that MADDPG can be viewed as a further development of DDPG. In DDPG, the gradient is expressed as follows:

$$\nabla_\theta J(\theta) = \mathbb{E}_{s\sim\mathcal{D}}\left[\nabla_\theta \boldsymbol{\mu}_\theta(a \mid s)\nabla_a Q^{\boldsymbol{\mu}}(s,a)\big|_{a=\boldsymbol{\mu}_\theta(s)}\right]$$

The proposed algorithm [6], MADDPG, addresses the mentioned limitations by relying solely on local information during execution, avoiding assumptions about a differentiable model for environmental dynamics or any specific communication method. Moreover, it is designed to be applicable in both cooperative and competitive frameworks.

The fundamental concept of MADDPG involves enriching the information utilized in actor-critic policy gradient methods. In the training phase, each agent's centralized critic accesses its own policy and the policies of all other agents. However, during execution, individual agents use only their own policy in a decentralized manner. Despite the centralization, employing a separate critic for each agent, as opposed to a single critic for all agents, provides the advantage of accommodating diverse reward functions, essential for addressing competitive frameworks as illustrated in A.1.

We examine N agents, each equipped with an individual continuous deterministic policy and a dedicated centralized action-value function. This function takes input from the actions of all agents and state information (typically corresponding to the observations of all agents) to generate the agent's Q-value as output. The update process for each agent's critic involves the use of a replay buffer, similar to the approach in DQN [44], accessing state transitions, actions, and associated rewards of all agents. Subsequently, the loss function employed for updating the critic is formulated in a manner similar to DQN:

$$\mathcal{L}(\theta_i) = \mathbb{E}_{\mathbf{x},a,r,\mathbf{x}'}\left[\left(Q_i^{\boldsymbol{\mu}}(\mathbf{x},a_1,\ldots,a_N)-y\right)^2\right], \quad y = r_i+\gamma Q_i^{\boldsymbol{\mu}'}(\mathrm{x}',a_1',\ldots,a_N')\Big|_{a_j'=\boldsymbol{\mu}_j'(o_j)}$$

,

This equation is applicable when every agent can access the policies of other agents, allowing the actions to be computed and utilized as input for the critic. In situations where this assumption is not feasible, agents can employ approximations of other agents' policies. In such cases, agent i refines the parameters corresponding to the policy of agent j in an online manner, incorporating an entropy regularizer that modifies the target used for updating the critic:

$$\mathcal{L}\left(\phi_i^j\right) = -\mathbb{E}_{o_j,a_j}\left[\log \hat{\boldsymbol{\mu}}_i^j\left(a_j \mid o_j\right) + \lambda H\left(\hat{\boldsymbol{\mu}}_i^j\right)\right]$$

,

$$\hat{y} = r_i + \gamma Q_i^{\boldsymbol{\mu}'}\left(\mathrm{x}', \hat{\boldsymbol{\mu}}_i'^1\left(o_1\right), \ldots, \boldsymbol{\mu}_i'\left(o_i\right), \ldots, \hat{\boldsymbol{\mu}}_i'^N\left(o_N\right)\right)$$

,

The primary objective of computing the critic is to update the actor, which is the component employed during execution. To achieve this, we can iteratively progress in the direction of the negative/positive gradient. This is particularly relevant for deterministic policies, where the input comprises the observations of the agents:

$$\nabla_{\theta_i} J\left(\boldsymbol{\mu}_i\right) = \mathbb{E}_{\mathbf{x},a\sim\mathcal{D}}\left[\nabla_{\theta_i}\boldsymbol{\mu}_i\left(a_i \mid o_i\right)\nabla_{a_i}Q_i^{\boldsymbol{\mu}}\left(\mathrm{x}, a_1, \ldots, a_N\right)\big|_{a_i=\boldsymbol{\mu}_i\left(o_i\right)}\right]$$

,

In a competitive environment, addressing non-stationarity poses a significant challenge, as agents might overly tailor their behaviors to mimic others. To tackle this issue, the authors suggest training a set of K sub-policies for each agent. During each episode, an agent randomly selects one of its K sub-policies and samples from a dedicated replay buffer associated with that sub-policy. In this scenario, the gradient update for the ensemble of policies is modified accordingly:

$$\nabla_{\theta_i^{(k)}} J_e\left(\boldsymbol{\mu}_i\right) = \frac{1}{K}\mathbb{E}_{\mathbf{x},a\sim\mathcal{D}_i^{(k)}}\left[\nabla_{\theta_i^{(k)}}\boldsymbol{\mu}_i^{(k)}\left(a_i \mid o_i\right)\nabla_{a_i}Q^{\boldsymbol{\mu}_i}\left(\mathbf{x}, a_1, \ldots, a_N\right)\Big|_{a_i=\boldsymbol{\mu}_i^{(k)}\left(o_i\right)}\right]$$

Appendix B

APPENDIX-2

In this section, we present the prerequisites, definitions, and calculations required for assessing upper bounds outlined in [5]. Consider a two-user complex Gaussian interference channel with two transmitter-receiver pairs with single antenna at both sides, each aiming to communicate exclusively within their respective pairs. These transmissions are represented by the following equations.

$$y_1 = h_1x_1 + g_2x_2 + z_1$$
$$y_2 = g_1x_1 + h_2x_2 + z_2$$

where for $i = 1, 2, x_i \in \mathbb{C}$ is subject to a power constraint P_i, i.e., $E\left[|x_i|^2\right] \leq P_i$, and the noise processes $z_i \sim \mathcal{CN}\left(0, N_0\right)$ are independent and identically distributed (i.i.d.) over time.

While the capacity region of the complex Gaussian interference channel may depend on the phases of the channel gains $\{h_{i,j}\}$, the inner and outer bounds that we present in this part only depend on the magnitudes $\{|h_{i,j}|\}$. As a result, we can use for our bounds a parameterization in terms of the signal-to-noise and interference-to-noise ratios. For $i = 1, 2$, let $\mathrm{SNR}_i = |h_{ii}|^2 P_i/N_0$ be the SNR of user i, and $\mathrm{INR}_1 = |g_2|^2 P_2/N_0$ $\left(\mathrm{INR}_2 = |g_1|^2 P_1/N_0\right)$ be the interference to noise ratio of user 1.

Definition 1: Weak interference channel is the channel that the parameters of the Gaussian interference channel satisfy $INR_1 < SNR_2$ and $INR_2 < SNR_1$.

Definition 2: Mixed interference channel is the channel that the parameters of the Gaussian interference channel satisfy $INR_1 \geq SNR_2$ and $INR_2 < SNR_1$, or $INR_1 < SNR_2$ and $INR_2 \geq SNR_1$.

Definition 3: Strong interference channel is the channel that the parameters of the

Gaussian interference channel satisfy $INR_1 \geq SNR_2$ and $INR_2 \geq SNR_1$.

Definition 4: An achievable region is said to be within one bit of the capacity region if for any rate pair (R_1, R_2) on the boundary of the achievable region, the rate pair (R_1+1, R_2+1) is not achievable. Equivalently, (R_1-1, R_2-1) is in the achievable region for any rate pair (R_1, R_2) in the capacity region.

Theorem 1: The achievable region for weak interference channel

$$\mathfrak{R}\left(\min\left(1, \mathrm{INR}_2\right), \min\left(1, \mathrm{INR}_1\right)\right)$$

is within one bit of the capacity region of the Gaussian weak interference channel.

The Han-Kobayashi scheme, introduced in [21], stands out as the most renowned achievable approach for handling interference channels. A more streamlined but equally effective version of the Han-Kobayashi achievable region was recently presented in [45], outlined in Lemma 1. This simplified region's achievability was subsequently established through a direct coding theorem in [[46], Theorem 1] (for a specific focus on a noncompound interference channel, refer to [[46], Sec. VI-A], noting that the region described in Lemma 1 is derived by applying the Fourier-Motzkin elimination method to the inequalities defining $\mathcal{R}_{\text{in}}^{(4)}$ in [[46]], alongside the expressions $R_1 = S_1 + T_1$ and $R_2 = S_2 + T_2$).

Lemma 1: Let $\mathcal{P}^*$ be the set of joint probability distributions $P^*(\cdot)$ that factor as

$$\begin{aligned} & P^*\left(q, w_1, w_2, x_1, x_2, y_1, y_2\right) \\ & \quad = P(q) \cdot P\left(w_1, x_1 \mid q\right) \cdot P\left(w_2, x_2 \mid q\right) \cdot P\left(y_1, y_2 \mid x_1, x_2\right). \end{aligned}$$

where $w_1\,(w_2)$ is the common information of user 1 (user 2) that can be decoded at both receivers, and q is the timesharing parameter. For the Gaussian interference channel, if we use Gaussian codebooks, and use u_1 and u_2 to denote the private information of user 1 and user 2 respectively.

For a fixed $P^* \in \mathcal{P}^*$, let $\mathcal{R}\left(P^*\right)$ be the set of (R_1, R_2) satisfying

$$\begin{aligned}
R_1 \leq & I\left(x_1 ; y_1 \mid w_2 q\right) \\
R_2 \leq & I\left(x_2 ; y_2 \mid w_1 q\right) \\
R_1+R_2 \leq & I\left(x_2 w_1 ; y_2 \mid q\right)+I\left(x_1 ; y_1 \mid w_1 w_2 q\right) \\
R_1+R_2 \leq & I\left(x_1 w_2 ; y_1 \mid q\right)+I\left(x_2 ; y_2 \mid w_1 w_2 q\right) \\
R_1+R_2 \leq & I\left(x_1 w_2 ; y_1 \mid w_1 q\right)+I\left(x_2 w_1 ; y_2 \mid w_2 q\right) \\
2 R_1+R_2 \leq & I\left(x_1 w_2 ; y_1 \mid q\right)+I\left(x_1 ; y_1 \mid w_1 w_2 q\right) \\
& +I\left(x_2 w_1 ; y_2 \mid w_2 q\right) \\
R_1+2 R_2 \leq & I\left(x_2 w_1 ; y_2 \mid q\right)+I\left(x_2 ; y_2 \mid w_1 w_2 q\right) \\
& +I\left(x_1 w_2 ; y_1 \mid w_1 q\right) .
\end{aligned} \tag{B.1}$$

Then the Han-Kobayashi achievable region is given by $\mathcal{R}=\bigcup_{P^* \in \mathcal{P}^*} \mathcal{R}\left(P^*\right)$.

$$\begin{aligned}
x_1 &= u_1+w_1 \\
x_2 &= u_2+w_2
\end{aligned}$$

where u_1, u_2, w_1, and w_2 are independent complex Gaussian random variables. Different $P^* \in \mathcal{P}^*$ correspond to different power splits between common and private messages, and different time-sharing strategies between the power splits.

Let's consider a scenario with a fixed power distribution—no time-sharing—between the private and common information of two users. Here, denote P_{u_1} as the power of user 1's private message and P_{u_2} as that of user 2's private message. We introduce INR_{p_2} as the interference-to-noise ratio of user 1's private message at receiver 2, and INR_{p_1} as the interference-to-noise ratio of user 2's private message at receiver 1.

$$\begin{aligned}
\mathrm{INR}_{p_2} &= \frac{\left|g_1\right|^2 P_{u1}}{N_0} \\
\mathrm{INR}_{p_1} &= \frac{\left|g_2\right|^2 P_{u2}}{N_0} .
\end{aligned}$$

It's clear that $0 \leq \mathrm{INR}_{p_2} \leq \mathrm{INR}_2$ and $0 \leq \mathrm{INR}_{p_1} \leq \mathrm{INR}_1$. With these definitions, the Signal-to-Noise Ratio (SNR) of user 1's private message at receiver 1 becomes $\mathrm{SNR}_{p_1}=\mathrm{INR} p_2 \mathrm{SNR}_1$, while the SNR of user 2's private message at receiver 2 is $\mathrm{SNR}_{p_2}=\mathrm{INR}_{p_1} \frac{\mathrm{SNR}_2}{\mathrm{INR}_1}$. We can specify a Han-Kobayashi achievable scheme with a fixed power allocation using INR_{p_2} and INR_{p_1}. This particular Han-Kobayashi scheme, characterized by parameters $\mathrm{INR}p_2$ and INR_{p_1}, is denoted as $\mathrm{HK}\left(\mathrm{INR}_{p_2}, \mathrm{INR}_{p_1}\right)$, with its associated achievable region labeled as $\mathfrak{R}\left(\mathrm{INR}_{p_2}, \mathrm{INR}_{p_1}\right)$

Notice that $\mathrm{HK}\left(\mathrm{INR}_{p_2}, \mathrm{INR}_{p_1}\right)$ and $\mathfrak{R}\left(\mathrm{INR}_{p_2}, \mathrm{INR}_{p_1}\right)$ pertain to a Han-Kobayashi scheme characterized by a consistent division of private and common message power, without employing time-sharing (i.e., the time-sharing variable q remains constant). Hence, $\mathfrak{R}\left(\mathrm{INR}_{p_2}, \mathrm{INR}_{p_1}\right) \subset \mathcal{R}$, wherein $\mathcal{R}$ represents the broader Han-Kobayashi achievable region outlined in Lemma 1. Generally, this inclusion is strict, signifying that by altering power allocations and employing time-sharing among various divisions of private and common message powers, a larger rate region can be achieved. However, it becomes evident that employing a strategic fixed allocation of private and common message power without time-sharing yields a rate region that closely approaches the channel's capacity region.

To evaluate the Han-Kobayashi region for the Gaussian interference channel, even if we restrict ourselves to use only Gaussian codebooks, we need to consider all possible power splits and different time-sharing strategies among them. This is in general very complicated and a calculation of a subset of the Han-Kobayashi achievable region using some special choices of power splitting and time sharing strategies can be found in [47]. We know in [5] that a good power splitting should have the property that $\mathrm{INR}_{p_2} = 1$ and $\mathrm{INR}_{p_1} = 1$, i.e., the interference-to-noise ratio of each user's private message at the other user's receiver is one. it is also shown that this power splitting can achieve to within one bit the symmetric rate capacity of the symmetric Gaussian interference channel. In the next section, it will shown that this is also a good splitting for the entire capacity region. More specifically, we will show that by choosing $\mathrm{INR}_{p_2}, \mathrm{INR}_{p_1}$ as close to 1 as possible, we can achieve rates within one bit of the whole capacity region.

To assess the Han-Kobayashi region concerning the Gaussian interference channel, even within the constraint of exclusively using Gaussian codebooks, it's necessary to explore all potential power distribution scenarios and various time-sharing strategies among them. This task is generally intricate, and a computation of a subset of the Han-Kobayashi achievable region, employing specific choices of power allocation and time-sharing strategies, is detailed in [47]. However, drawing from the insights gained in [5], we recognize that an optimal power distribution should exhibit $\mathrm{INR}_{p_2} = 1$ and $\mathrm{INR}_{p_1} = 1$. In other words, the interference-to-noise ratio of each user's private message at the opposite user's receiver equals one. It is also demonstrated that this particular power distribution nearly achieves the symmetric rate capacity of the symmetric Gaussian interference channel, differing by only one bit.

Appendix C

APPENDIX-3

In this section, we present the prerequisites, definitions, and calculations required for assessing upper bounds outlined in [4]. The capacity region for an interference channel when two distinct messages are transmitted has been established primarily for scenarios involving highly potent interference [48, 49]. However, these capacities have been computed solely for very strong interference scenarios [50]- [51], as well as for certain straightforward cases. Our focus is to delve into this issue across a broader spectrum of channels characterized by strong interference.

Once appropriately normalized, a Gaussian interference channel is represented by the following expression:

$$\begin{aligned} y_1 &= x_1 + \sqrt{\chi} x_2 + n_1, \\ y_2 &= \sqrt{\nu} x_1 + x_2 + n_2, \end{aligned}$$

In this context, x_i, y_i, and n_i (where $i = 1, 2$) denote sampled values of the input signal, output signal, and superposed noise, respectively. The noise is assumed to adhere to a Gaussian distribution and exhibit white characteristics within the specified frequency band, with a power of N_i for both noise sources. Importantly, these noise sources remain independent of the input signals. Additionally, the powers of the input signals are constrained to values lower than P_i ($i = 1, 2$). Within this framework, χ and ν represent the level of interference present in the first and second output signals y_1 and y_2, respectively.

Carleial [50] demonstrated that interference doesn't diminish the capacity in scenarios of very intense interference, namely when χ and ν satisfy the following conditions simultaneously:

$$\chi \geq (P_1 + N_1)/N_2,$$
$$\nu \geq (P_2 + N_2)/N_1.$$

Appendix D

APPENDIX-4

In this section, we outline the prerequisites, definitions, and computations necessary to evaluate the upper bounds as detailed in [22]. The scenario involves a two-user Multiple-Input Multiple-Output Interference Channel (MIMO IC) where transmitter i (denoted as Tx_i) possesses M_i antennas, and receiver i (denoted as Rx_i) has N_i antennas, respectively, for $i = 1, 2$. This specific MIMO IC is henceforth referenced as the $(M_1, N_1, M_2, N_2\ )$ MIMO IC.

Let $G_i \in \mathbb{C}^{N_j \times M_i}$ denote the channel matrix between Tx_i and Rx_j, with $|G_i|_F^2 = 1$. This normalization doesn't affect the generality as the Frobenius norm of an unnormalized channel matrix can be absorbed in Signal-to-Noise Ratio (SNR) or Interference-to-Noise Ratio (INR) considerations.

We focus on a time-invariant or fixed channel, where the channel matrices remain constant throughout the communication duration. At any given time t, Tx_i selects a vector $X_{it} \in \mathbb{C}^{M_i \times 1}$ and transmits $\sqrt{P_i} X_{it}$ across the channel. We assume the following average input power constraint at Tx_i:

$$\frac{1}{n} \sum_{t=1}^{n} \operatorname{Tr}\left(Q_{it}\right) \leq 1$$

for $i \in \{1, 2\}$, where $Q_{it} = \mathbb{E}\left(X_{it} X_{it}^H\right)$. Note that in the above power constraint, Q_{it} 's can depend on the channel matrices. The received signals at time t can be written as

$$\begin{aligned} Y_{1t} &= \sqrt{\rho_{11}} H_1 X_{1t} + \sqrt{\rho_{21}} G_2 X_{2t} + Z_{1t} \\ Y_{2t} &= \sqrt{\rho_{22}} H_2 X_{2t} + \sqrt{\rho_{12}} G_1 X_{1t} + Z_{2t} \end{aligned}$$

Here, $Z_{it} \in \mathbb{C}^{N_i \times 1}$ stands for independent and identically distributed complex Gaussian variables $\mathcal{CN}\left(\mathbf{0}, I_{N_i}\right)$ across both i and t. ρ_{ii} and ρ_{ij} denote the Signal-to-Noise

Ratio (SNR) at receiver i and Interference-to-Noise Ratio (INR) at receiver j, respectively, where $i \neq j \in 1, 2$.

In subsequent discussions, the MIMO IC, defined by channel matrices, SNRs, and INRs as described above, will be denoted as $\mathcal{IC}(\mathcal{H}, \bar{\rho})$, where $\mathcal{H} = \{H_1, G_1, G_2, H_2\}$ and $\bar{\rho} = [\rho_{11}, \rho_{12}, \rho_{21}, \rho_{22}]$.

If the normalized signal vector, X_i, possesses a covariance matrix Q_i, the covariance matrix of the received signal at Rx_i becomes $P_i H_i Q_i H_i^H$. Consequently, the total received signal power amounts to $\operatorname{Tr}\left(P_i H_i Q_i H_i^H\right)$, leading to the corresponding Signal-to-Noise Ratio (SNR) of $\rho_{ii} = \frac{P_i \operatorname{Tr}\left(H_i Q_i H_i^H\right)}{N_i}$. The Interference-to-Noise Ratios (INRs) of the channel, denoted as ρ_{ij}, can be computed in a similar manner.

$$n_i = \hat{m}_{ji} + \max\left\{\left(m_{ii} \log\left(M_i\right) + m_{ij} \log\left(M_i + 1\right)\right) \; \min\left\{N_i, M_s\right\} \log\left(M_x\right)\right\}$$

for $1 \leq i \neq j \leq 2$, with $M_x = \max\left\{M_1, M_2\right\}, M_s = \left(M_1 + \; M_2\right), m_{ij}$ representing the rank of the matrix G_i, and $\hat{m}_{ij} = m_{ij} \log\left(\frac{(M_i+1)}{M_i}\right)$. Note that $m_{ij} \leq \min\left\{M_i, N_j\right\}$.

The achievable region of this explicit HK coding scheme is within (n_1^*, n_2^*) bits to the capacity region, where

$$n_i^* = \min\left\{N_i, M_s\right\} \log\left(M_x\right) + \hat{m}_{ji}, \text{ for } 1 \leq i \neq j \leq 2.$$

To keep things concise, we introduce the matrices as follows:

$$K_i \triangleq \left(I_{M_i} + \rho_{ij} G_i^H G_i\right)^{-1} \quad 1 \leq i \neq j \leq 2.$$