Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Electrical, Electronics and Communications Engineering
(36.th cycle)

On Applications of Machine Learning In IRS NOMA Architectures For 5G and Beyond

Muhammad Shehab

* * * * *

Supervisors
Prof. Tamer Khattab, Supervisor
Prof. Daniele Trinchero, Co-Supervisor

Politecnico di Torino
January 22, 2024

I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

. . . . . .

Muhammad Shehab

Turin, January 22, 2024

Summary

To foster the advancement of smart city components and the Internet of Things (IoT) applications, future wireless systems need to offer extensive system capacity, ultra-high data speeds, minimal latency, high dependability, low energy consumption, and large connection densities. These networks are expected to revolutionize various applications in a variety of contexts, including massive machine-type communications (mMTC), ultra-reliable low latency communication (uRLLC), and better mobile broadband services (EMBB).

Due to the widely varying and quickly growing need for mobile wireless networks, meeting the requirements of these applications has become challenging. To provide ubiquitous connections for billions of devices, efficiently manage the surging mobile data traffic, and support algorithm-driven applications, an effective solution is necessary. Machine learning (ML) is seen as a crucial tool to address these challenges. Moreover, the utilization of new frequencies and surge in number of users motivated new technologies such as non-orthogonal multiple access (NOMA) and intelligent reflecting surfaces (IRS).

The goal of this thesis is to establish ML-based methods for 5G and beyond (B5G) wireless communication systems that use NOMA and IRS. The study involves the utilization of deep reinforcement learning (DRL) to enhance IRS operations in downlink NOMA situations, as well as terahertz (THz) networks assisted by multi-hop IRS. The emphasis will be on maximizing the total data transmission rate. Furthermore, DRL is utilized for the secure cross-layer design of the IRS NOMA-SWIPT (Simultaneous Wireless Information and Power Transfer) IoT scenario, focusing on maximizing the secure sum-rate for legitimate IoT devices (IoTDs) operating under the IRS NOMA-SWIPT scheme, and meeting packet loss constraints within the system. It was shown that the utilization of the DRL in the aforementioned scenarios enabled significant performance enhancements compared to existing classical solutions, which are typically sub-optimal due to the non-convex nature of the formulated problems. Furthermore, DRL models had better complexity scaling properties with increased system complexity (e.g. number of IRS elements).

Acknowledgements

To my family, for their support and faith in me. ...

I would like to acknowledge and express my gratitude to my Father, Mother, grandfathers, and grandmothers, wife, children, and all my family members for their ongoing support, motivation, patience, encouragement, and enduring affection throughout my academic pursuits. Their constant motivation and encouragement have been a driving force that propels me forward in my research studies.

Furthermore, I take great pleasure in extending my sincere gratitude to my mentors, Professor Tamer and Dr.Daniele Trinchero, for their consistent guidance throughout my doctoral journey. I am genuinely appreciative of their unwavering support and constructive input, which was indispensable for the successful culmination of my dissertation. I also wish to convey my special appreciation to the members of the Ph.D. committee for generously dedicating their effort and time to evaluate my research work.

I would like to dedicate this thesis to my parents, Father, Mother, grandfathers, grandmothers, brother, sisters, my loving wife, children, and all my family members.

Contents

List of Tables

List of Figures

LIST OF PUBLICATIONS

- **M. Shehab** et al., "**Terahertz Multiple Access: A Deep Reinforcement Learning Controlled Multihop IRS Topology**," in IEEE Open Journal of the Communications Society, doi: 10.1109/OJCOMS.2024.3357701.
- **M. Shehab**, B. S. Ciftler, T. Khattab, M. M. Abdallah and D. Trinchero, "**Deep Reinforcement Learning Powered IRS-Assisted Downlink NOMA**," in IEEE Open Journal of the Communications Society, vol. 3, pp. 729-739, 2022, doi: 10.1109/OJCOMS.2022.3165590.
- **M. J. Shehab**, I. Kassem, A. A. Kutty, M. Kucukvar, N. Onat and T. Khattab, "**5G Networks Towards Smart and Sustainable Cities: A Review of Recent Developments, Applications and Future Perspectives**," in IEEE Access, vol. 10, pp. 2987-3006, 2022, doi: 10.1109/ACCESS.2021.3139436.
- **M. Shehab**, A. Badawy, M. Elsayed, T. Khattab and D. Trinchero, "**DDPG Performance in THz Communications over Cascaded RISs: A Machine Learning Solution to the Over-Determined System**,"2023 International Wireless Communications and Mobile Computing (IWCMC), Marrakesh, Morocco, 2023, pp. 210-215,doi:10.1109/IWCMC58020.2023.10182861.
- **M. Shehab**, T. Khattab, M. Kucukvar and D. Trinchero, "**The Role of 5G/6G Networks in Building Sustainable and Energy-Efficient Smart Cities**,"IEEE 7th ENERGYCON, Riga, Latvia, 2022, pp. 1-7, doi: 10.1109/ENERGYCON53164.2022.9830364.
- **M. Shehab**, M. Elsayed, A.Badawy, T. Khattab, M. Barhamgi, S. Salem, D. Trinchero, "**A DRL-Based Approach for Secure Cross-layer Design of IRS NOMA-SWIPT Systems in IoT Environments**," 2024 (Submitted to TCOM).

Chapter 1

Introduction

1.1 Motivation

In the 5G/6G era, the emergence of innovative use cases such as augmented reality (AR), virtual reality (VR), machine-type networking, online games, Voice over IP (VoIP), autonomous driving, smart applications, sensor networks, and the realm of interconnected devices including Internet of Things (IoT) has highlighted the necessity for reliable wireless networks that can overcome the limitations of current wireless technologies and bridge the critical performance gap [51], [50]. To meet these requirements, researchers are exploring modern various access and modulation schemes. This exploration is driven by the higher demand for bandwidth, necessitated by the need for faster communication, and the scarcity of available frequency bands. In response, 5G and 6G technologies are venturing into high frequencies like mmWave and THz. However, these higher frequencies bring challenges such as higher attenuation, reduced coverage, and Line of Sight (LOS) limitations. To address these issues, researchers are investigating technologies like Machine Learning (ML), Non-Orthogonal Multiple Access (NOMA), and Intelligent Reflecting Surfaces (IRS). Nevertheless, the incorporation of IRS NOMA introduces increased complexity to the system, a challenge that can be tackled using ML approaches [51], [50].

1.2 Problem Statement

In IRS NOMA scenarios, the problem of tuning the IRS phase shifts to optimize the sum-rate is non-convex as a result of the constant modulus restriction, which arises from the reflective nature of the IRS without signal amplification. Traditional methods struggle to find mathematical solutions for such large-scale communication schemes, and exhaustive search (ES) approaches are impractical due to their complexity. To overcome these challenges and solve the optimization problem, we

employ a DRL method, specifically deep deterministic policy gradient (DDPG), as an alternative to mathematical methods.

1.3 Deliverables

This dissertation yields the following key deliverables:

1. An algorithm for maximizing the total transmission data rate of an IRS NOMA network using DRL, thereby enhancing network performance and increasing user data rate attainment.

2. A method for overcoming non-convexity and unit modulus constraints in the NOMA IRS system using DDPG. This method addresses challenges in handling limited channel state information (CSI) and compares the efficiency of DDPG with the ES in IRS NOMA systems.

3. A framework for utilizing the DDPG scheme in the context of uplink multiple access cascaded IRS with THz communications. This framework solves overdetermined sets of equations in a multi-hop IRS scenario.

4. An approach for maximizing received power in multi-hop IRS scenarios using DDPG. It determines optimal phase shift reflecting matrices to maximize received power while considering interference from other users and compares the efficiency of DDPG with mathematical methods.

5. A method for optimizing the aggregate data transmission rate of users in multi-hop IRS scenarios using DDPG, demonstrating superior performance and reduced complexity compared to an ES.

6. A framework for secrecy sum-rate maximization for IoTDs using DDPG with cross-layer design. It optimizes the secrecy sum-rate by determining optimal phases of IRS, power allocation factors, and energy harvesting (EH) factors while addressing non-convex optimization problems.

1.4 Methodology

The research methodology of the dissertation involves analyzing and modeling the key research areas such as 5G, 6G, THz, IRS, cascaded IRS, NOMA, SWIPT, EH, DRL, etc. An optimization problem is developed based on the provided system model. considering the system's limitations, constraints, and non-convexity. Mathematical methods and the DRL algorithm are employed to solve this problem and find optimal solutions, which are compared to sub-optimal solutions generated from mathematical methods.

The methodology is divided into several work packages:

WP1: Conducting a background study and literature review on the main research areas, including IRS, NOMA, DRL, DDPG, multi-hop, THz, etc. This involves collecting relevant research topics, thoroughly reviewing research articles, and identifying the research focus and questions.

WP2: Building a comprehensive system and channel model that incorporates the aforementioned research areas. This step aims to develop a clear and robust model that facilitates the derivation and formulation of the optimization problem.

WP3: Formulating the optimization model to maximize overall system efficiency while taking into account the system model's constraints, limitations, and non-convex nature.

WP4: Proposing an innovative and effective strategy for solving the non-convex optimization problem and improving system performance.

WP5: Planning, designing, and simulating the system model using DRL, ES, and sub-optimal mathematical methods.

WP6: Conducting a comparison between the results obtained from each solution method, considering their respective complexities, and analyzing the advantages and disadvantages of each technique.

WP7: Publishing research papers and completing the dissertation write-up based on the findings and outcomes of the research.

By following this methodology, the dissertation aims to contribute to the existing knowledge in the field and provide insights into optimizing the efficiency of IRS NOMA and multi-hop IRS NOMA systems within the context of wireless systems.

1.5 Thesis Objectives

The overarching goal of this thesis is to propose optimal techniques for utilizing IRS and NOMA techniques in 5G and beyond communication technologies. In order to achieve the overall goal, we tackle three main technologies that can be used for future wireless networks; namely, THz physical layer, NOMA and IRS. The aforementioned technologies will be utilized to address specific gaps in the network design and performance namely; multi-user capacity enhancement, coverage extension, and security. More specifically, the following objectives are identified:

1. Propose a framework for using IRS in downlink NOMA systems to maximize sum rate under limited CSI knowledge.

2. Propose a framework for using cascaded IRS to extend the coverage of THz technology when used in uplink.

3. Propose a framework for using IRS in downlink NOMA with security and energy constraints.

1.6 Thesis Contributions

The following contributions are made by this dissertation to the field of wireless communications:

1. Proposing a DRL based method to optimize the sum rate of a downlink IRS NOMA under the assumption of limited CSI.

2. Proposing a DRL based method to optimize the sum rate of a multi-user uplink THz system supported by cascaded IRS backbone.

3. Proposing a DRL based method to optimize the secure sum rate of a downlink IRS NOMA system under energy harvesting and cross-layer packet-loss constraints.

These contributions provide valuable insights into the field of wireless communications by addressing optimization problems in IRS NOMA systems and cascaded IRS systems. Additionally, the dissertation explores the optimization of the secrecy sum-rate in the context of IoTDs, introducing a novel DDPG-based approach. The research findings provide valuable insights for establishing effective, secure, and safe wireless communication networks, facilitating advancements in future network technologies.

1.7 Thesis Layout

The dissertation is structured in the following manner:

Chapter 1: Introduction: This chapter serves as an introductory section to the dissertation, presenting the motivation behind the research, the methodology employed, the objectives of the study, and the contributions made to the field of wireless communications.

Chapter 2: Literature Review: The primary motifs, key concepts, and state-of-the-art schemes are illustrated in this chapter. An exhaustive analysis of related work in the literature is carried out, evaluating and comparing the proposed work with existing studies.

Chapter 3: Downlink Non-Orthogonal Multiple Access (NOMA) With Intelligent Reflecting Surfaces (IRS) Empowered by Deep Reinforcement Learning: Chapter 3 investigates the scenario of the IRS NOMA downlink network. The objective is to maximize the total transmission rate while considering limited CSI knowledge and the incorporation of imperfect SIC in practical NOMA systems. The chapter encompasses the presentation and development of the optimization problem and explores the suggested DDPG-based solution.

Chapter 4: A Deep Reinforcement Learning Technique for Terahertz Multiple Access in a Multi-hop IRS Network Topology: This chapter is

dedicated to exploring the uplink multiple access multi-hop IRS scenario in a THz system. The primary objective is to utilize the DDPG algorithm to maximize the total throughput of users. The chapter addresses the challenges posed by the IRSs' constant modulus restrictions and the over-determined set of equations.

Chapter 5: Deep Reinforcement Learning-Based Cross-Layer Design for Secure IRS NOMA-SWIPT Systems in IoT Environments: Chapter 5 examines the indoor scenario of the secure cross-layer design of IRS NOMA downlink system, focusing on maximizing the aggregated secrecy throughput of legitimate Internet of Things Devices (IoTDs) and conveying energy to near and far NOMA IoTDs. The chapter focuses on employing the IRS for adjusting the phase of the incoming signal, optimizing NOMA power allocation, and EH, and minimizing network layer packet loss. The problem of non-convex optimization is handled using DRL, specifically the DDPG algorithm, to obtain optimum parameters.

Chapter 6: Conclusion and Future Prospects: The dissertation concludes and summarizes the findings in this final chapter. The fundamental results of the research are reviewed, and suggestions for future research work in the field of IRS NOMA architecture are provided.

By following this layout, the dissertation presents a comprehensive exploration of IRS NOMA systems and their optimization in various scenarios, showcasing the proposed DDPG-based solutions and their contributions to the field of wireless communications.

Chapter 2

Background and literature survey

Multiple access schemes are crucial for efficiently utilizing constrained frequency resources and accommodating the diverse communication needs of various applications. Traditional multiple-access approaches including FDMA, TDMA, and CDMA have been extensively utilized in earlier generations of wireless networks. Nonetheless, these methods have their limitations concerning spectrum efficiency, latency, and the ability to support massive connectivity. To address these limitations, new multiple access methods are suggested for 5G and 6G networks. One example is NOMA, which allows multiple users to jointly utilize the same time-frequency resources by leveraging the power or code domain. NOMA improves spectrum efficiency and supports a large number of simultaneous connections. It is shown to be particularly effective in scenarios involving distinct service quality demands and fluctuating channel conditions [29].

The evolution towards 5G/6G networks is driven by the demand for reliable wireless communication to support a wide range of applications. To meet these requirements, modern multiple access and modulation techniques, such as NOMA, along with advanced technologies like IRS, SWIPT, and DRL, are being explored. These advancements aim to overcome the limitations of existing wireless networks and enable efficient and reliable connectivity in the era of 5G/6G [51], [50].

2.1 Background

2.1.1 Background on NOMA

In previous generations of wireless networks, such as 1G, 2G, 3G, and 4G, OMA techniques like CDMA, TDMA, FDMA, and OFDMA were used to allocate radio resources to users. These techniques ensured that multiple users were separated based on orthogonal properties, such as time, frequency, code, or space. Deviating from this orthogonality led to interference, resulting in degraded communication quality, inefficient resource utilization, and data loss. However, the

increasing demands for connectivity in 5G/6G networks have necessitated a new approach, giving rise to NOMA [6]. NOMA has garnered considerable interest as a promising solution for 5G and beyond because of its potential to boost spectrum efficiency, improve link connectivity, and substantially increase capacity compared to OMA schemes. NOMA empowers multiple users to simultaneously transmit within a common pool of shared resources through power domain multiplexing, without strict constraints on the orthogonality of frequency subcarriers. Although simultaneous transmission introduces interference, NOMA employs a technique known as Successive Interference Cancellation (SIC) to mitigate its impact.

2.1.2 Intelligent Reflecting Surfaces (IRSs)

IRS is considered a cutting-edge concept and a research hotspot in wireless communications, serving as an extension to M-MIMO technology. IRS has risen as a highly auspicious technology with significant potential for enabling smart and adaptable wireless communication environments. It consists of a digitally-controlled electromagnetic surface, known as a metasurface, comprising numerous low-cost semi-passive scattering elements. These elements consume minimal power and can be programmed to manipulate the electromagnetic properties, specifically altering the phase of the incoming RF signals. The unique capability of the IRS lies in its ability to coherently combine the reflected signals, optimizing the received power, or counter-actively combining them to alleviate interference. Unlike traditional wireless communication technologies implemented at the receiver, IRS actively controls the properties of the propagated signals, thereby enhancing the received signal strength. This leads to enhanced efficiency in contrast to wireless communication networks lacking IRS, effectively realizing the concept of a smart radio environment [51], [50], [75].

By reshaping the wireless propagation channel, IRS can amplify transmitted signals, improve received signal strength for intended users, and mitigate interference for unintended users. This not only improves the overall signal quality but also provides innovative and cost-effective methods for meeting the crucial benchmarks of 6G networks. In summary, IRS is an advanced technology that enables the development of intelligent and adaptable wireless communication environments. By actively manipulating signal properties, IRS holds immense potential for enhancing wireless communication systems, boosting performance, and realizing the goals of future 6G networks. [19].

2.1.3 Terahertz Technology (THz)

THz technology is anticipated to have a pivotal role in the advancement of 6G networks, which aim to support unprecedented services and bandwidth-intensive applications. These applications include Terabytes Per Second (Tbps) IoT, Tbps

wireless local area network (WLAN) systems, ultra-broadband THz space communication networks, wireless networks-on-chip communication networks, Tbps integrated access backhaul wireless communication systems, and more. Meeting the requirements of these applications necessitates a significant capacity boost, approximately 1000 times higher than that of 5G systems [4]. To address the challenge of spectrum scarcity and meet the increasing service demands, the migration towards THz bands with frequencies ranging from 0.1 up to 10 THz is seen as a promising solution. THz spectrum bands offer a vast amount of bandwidth when effectively utilized, making them a fundamental element of 6G communication networks. Operating in the THz frequencies enables substantial data rates and capacities. However, there are several challenges associated with THz communication, including high propagation losses, severe path attenuation, intermittent wireless links, and highly dynamic and uncertain channels. These limitations result in unreliable sporadic links, restricting THz communications to short-range and presenting challenges due to the extremely small wavelength (λ) values [51], [50], [20].

2.1.4 Energy Harvesting

EH involves capturing power from various sources such as radio frequency (RF) waves, thermal gradients, and light to provide energy for devices. It is particularly important in communication systems and IoT environments to extend battery life and promote sustainability. They offer the potential to capture energy from the environment and power IoT devices (IoTDs), reducing the reliance on batteries and contributing to a more eco-friendly and cost-effective IoT ecosystem. By harnessing energy from the surrounding environment, IoTDs can operate for longer periods without the need for frequent battery replacements. Challenges in EH models include power limitations, intermittent energy availability, complex energy management, and the cost and complexity of setting up EH infrastructure. These challenges need to be addressed to fully harness the benefits of EH in communication systems [51], [50], [37], [5].

2.1.5 Reinforcement Learning

Reinforcement Learning (RL) is a subset of ML that empowers an agent to interact with its environment, learning through experimentation and adjustment based on feedback received from the environment. The fundamental components of the RL model include the environment, action, state, reward, policy, and Q-value function. The environment represents the agent's surroundings, the state signifies its current condition, and the reward reflects the response obtained from the environment. The policy dictates how the agent maps its state to actions, while the Q-value function assesses the quality of actions taken in specific states.

During each temporal step, the agent acquires a state $s^{(\tau)}$ from the environment and selects an action $a^{(\tau)}$ according to a given policy π. Subsequently, the agent obtains a reward $r^{(\tau)}$ as a performance measure for the action undertaken within the state $s^{(\tau)}$, and the state transitions to $s^{(\tau+1)}$. The objective of the RL agent is to determine the optimal action $a^{(\tau)}$ within a provided state $s^{(\tau)}$ to maximize the total reward $r^{(\tau)}$ over an episode. The interaction between the agent and the environment is modeled as a Markov Decision Process (MDP). Notably, RL algorithms differ from dynamic programming as they do not assume knowledge of a particular MDP mathematical model, making them well-suited for large MDPs where the use of specific mathematical methods is impractical [14].

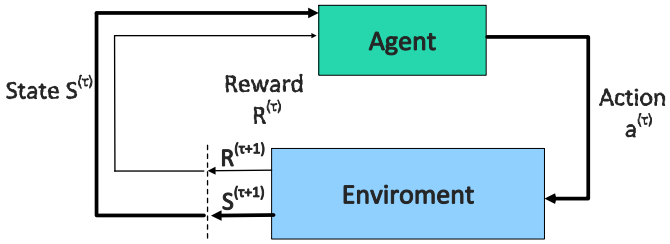


Figure 2.1: DRL-based Model

RL schemes such as deep Q-learning (DQL), policy gradient (PG), and DDPG share a common goal, which is to train an agent to acquire a strategy that maximizes rewards within an environment. However, DQL is limited to discrete-time spaces and cannot handle continuous action-spaces, making it a value-based learning method. The PG scheme has a slow convergence rate in the context of wireless communication. DDPG addresses these limitations by combining Q-networks and the PG techniques as revealed in Fig. 2.2. It is capable of handling continuous action-spaces dynamic, and non-convex wireless communication scenarios. Thus, it offers improved performance compared to the individual algorithms [14].

2.1.6 Deep Deterministic Policy Gradient

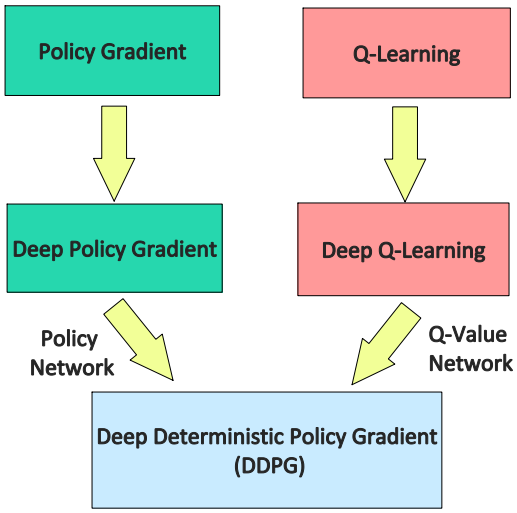


Figure 2.2: Hybrid approach: DDPG combining DPG and DQN

DDPG is an RL method designed for environments with continuous action spaces. It learns from experience, utilizing an off-policy approach with a replay buffer to manage a large store of past experiences. DDPG employs neural networks to estimate the Q-value function and policy, making it well-suited for complex, high-dimensional environments. Two sets of networks, online and target networks, are used to enhance stability. The online networks include the actor-network for selecting actions and the critic-network for evaluation, while the target networks mirror them. Soft network parameter updates gradually align target network parameters with online networks. The actor-network directly calculates actions based on the state, and the critic-network assesses state-action pairs by estimating the Q-value. Soft updates with small values further enhance learning stability and convergence, ensuring smoother updates and overall algorithm stability. The DDPG framework includes actor-networks for state-to-action mapping and critic-networks for action evaluation. To further stabilize learning, target actor and critic networks estimate future actions and Q-values, guiding training using the Q-target formula. This architecture enhances decision-making in RL tasks, promoting effective learning and performance improvement. The DDPG implementation process is illustrated in Fig. 2.3 and detailed in Algorithm 1.

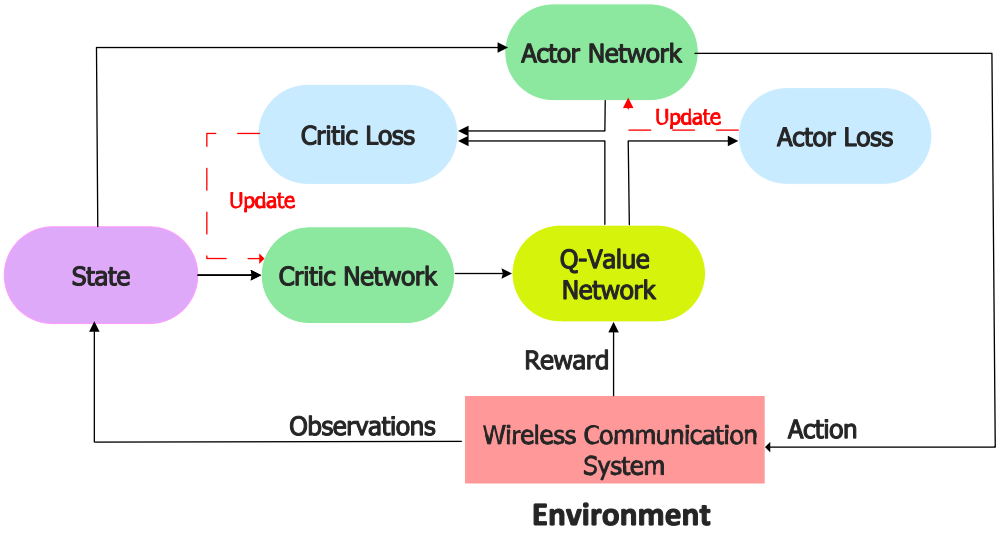


Figure 2.3: DDPG-based model

Algorithm 1 DDPG Implementation Process

Input: Learning rate $\digamma$, decay factor γ, batch quantity B, memory buffer D with capacity C, Maximum number of iterations T.
Output: Optimal actor-network weights signified as o^{μ} and critic-network weights designated as o^{Q}.
Initialization: Set $i = 0$ and initialize the experience memory buffer D with capacity C.
Randomly initialize the actor-network $\mu(s; o^{\mu})$ with weights o^{μ}, the critic-network $Q(s, a; o^{Q})$ with weights o^{Q}, the target actor-network $\mu^{\tau}(s; o^{\mu'})$ with weights $o^{\mu'}$, and the target critic-network $Q^{\tau}(s, a; o^{Q'})$ with weights $o^{Q'}$.
Set the initial target parameters equal to the main parameters, i.e., $o^{\mu'} \leftarrow o^{\mu}$ and $o^{Q'} \leftarrow o^{Q}$.
for episode $i = 1, 2, ..., B$ **do**
    Initiate random noise $\mathcal{N}$ for action exploration.
    Obtain initial information about the first state s_{τ}.
    for time step $\tau = 1, 2, ..., T$ **do**
        Determine the state s^{τ} and choose action $a^{\tau} = \mu(s^{\tau}|o^{\mu}) + \mathcal{N}$.
        Perform action a^{τ}, receive reward r^{τ}, and observe the subsequent state $s^{\tau+1}$.
        Store the tuple $(a^{\tau}, s^{\tau}, r^{\tau}, s^{\tau+1})$ in the replay memory D.
        Randomly sample a mini-batch of transitions (a_i, s_i, r_i, s_{i+1}) B from the memory buffer D.
        Compute the target $y_i = r_i + \gamma Q'(s^{i+1}, \mu'(s^{i+1}; o^{\mu'}); o^{Q'})$ if $i \leq B$, and $y_i = r_i$ if $i = B$.
        Update the critic-network o^{Q} parameters to minimize the loss: $L = \frac{1}{|B|} \sum_{i=1}^{|B|} \left(y_i - Q(\mathbf{s}^{(i)}, \mathbf{a}^{(i)}|o^{Q})\right)^2$
        Update the actor-network o^{μ} parameters using the extracted policy gradient: $\nabla_{o\mu} \boldsymbol{J} \approx \frac{1}{|B|} \sum_{i=1}^{|B|} \nabla_a Q(\mathbf{s}^{(i)}, \mathbf{a}^{(i)}|o^{Q}) \nabla_{o\mu} \mu(\mathbf{s}^{(i)}|o^{\mu})$
        Update the target actor and critic-network parameters: $o^{j'} = \upsilon o^{j} + (1 - \upsilon) o^{j'}$, where $j = Q$ or μ.

2.2 Relevant Literature Survey

2.2.1 IRS NOMA

Deploying IRS in wireless communication systems converts it into a smart radio environment capable of supporting diverse user requirements, such as extending coverage, enhancing data rates, minimizing power consumption, and improving transmission security [19]. Meanwhile, NOMA emerges as an innovative approach for optimizing performance in upcoming communication networks, offering higher capacity and increased spectrum efficiency in contrast to OMA approaches [28]. Several research studies have investigated schemes combining IRS with OMA [66, 25, 12, 54, 53]. However, the synergy between IRS and NOMA offers a potent remedy for enhancing spectral efficiency, energy efficiency, and communication coverage [11]. A typical IRS-aided NOMA model involves a BS transmitting signals to $\mathcal{K}$ NOMA users through the IRS in scenarios where there is no LOS communication link between the BS and the users.

Numerous research studies have investigated the use of IRS in NOMA systems, both in the downlink and uplink scenarios. These studies include [8, 72, 65, 78, 70, 62, 36, 69, 79, 77, 34, 15, 58, 16, 63, 46, 43, 44, 61, 68, 67, 33, 38, 76].

In the downlink scenario, [8] proposed a new NOMA transmission scheme using IRS to increase the number of served users. [72] focused on secure transmission in IRS NOMA networks. [65] optimized the rate performance and fairness in a downlink IRS-assisted NOMA system. [78] optimized power allocation and beamforming in an IRS mmWave NOMA scheme. [62] considered a single IRS aiding downlink transmission to maximize signal power. [36] investigated a multi-user downlink scenario using optical IRS for VLC systems. [69, 79, 77, 34, 15, 58] explored different objectives in STAR-IRS-aided downlink NOMA networks. [46] studied an IRS-assisted NOMA-HARQ system for IoT users situated in areas with no direct line of sight (NLoS) propagation. [43] analyzed error performance in a multi-user IRS NOMA scenario. [44] enhanced physical layer security in an IRS downlink MIMO NOMA system. [68] investigated IRS-assisted NOMA transmission to near and far users.

In the uplink scenario, [70] optimized the users' sum-rate in an IRS NOMA system. [16] proposed an uplink cognitive radio NOMA scheme with rate splitting and IRS. [63] utilized double IRSs to increase the sum-rate. [61] developed a symbiotic radio uplink NOMA IRS scheme. [67] addressed data extraction and detection in an IRS uplink NOMA system. [33] employed IRS in multi-IoTDs MEC systems to improve latency performance.

Unlike most studies using classical methods, [38] and [76] employed ML techniques. [38] utilized federated learning (FL) with IRS to reduce training latency in NOMA systems. [76] used distributed ML for passive beamforming in STAR-IRS downlink NOMA, adopting a partitioning approach and a DRL agent.

In summary, the existing literature extensively explores IRS NOMA scenarios in downlink and uplink configurations, addressing various objectives such as coverage extension, rate optimization, and spectral efficiency enhancement. The studies employ mathematical optimization methods [8, 72, 65, 78, 70, 62, 36, 69, 79, 77, 34, 15, 58, 16, 63, 46, 43, 44, 61, 68, 67, 33] to solve complex problems, with recent approaches integrating ML techniques like federated learning and distributed ML [38, 76]. However, a significant research gap exists as these studies assume accurate channel knowledge between the IRS and users, which contrasts real-world scenarios where limited channel knowledge is prevalent, especially with semi-passive IRS elements, which cannot be used to estimate channels accurately. Further, DRL is a distinct type of ML technique, different from federated learning, that offers significant potential for overcoming the limitations imposed by limited channel knowledge and improving the system's performance. By leveraging DRL techniques, we can effectively mitigate the challenges associated with limited channel knowledge and improve the overall system performance.

2.2.2 IRS in THz Communication

To enhance the achievable data rate and coverage in THz communication networks, researchers have recently explored the implementation of cascaded IRS. Several studies [41, 39, 52, 71, 2, 26, 27] have investigated different aspects of this technology. For example, in [41], the focus was on an IRS-aided multi-hop multi-pair unicast communication system, involving numerous transmitters and receivers engaging in mutual communication. The authors suggested distributed control mechanisms for the IRSs to optimize the attainable data rate in a multi-hop interference channel. In [39], a multi-IRS-assisted M-MIMO system was studied. The goal was to boost the minimum received signal power in a scenario where a BS with multiple antennas transmits independent signals to faraway users with single antennas. Cascaded LOS communication linkages were built between the BS and the users by exploiting collaborative signal reflections from several groups of IRSs. Another study [52] focused on a multi-IRS network with decode and forward relaying assistance. Here, a sole source (BS) establishes communication with a sole destination (user), aiming to determine the optimal arrangement of IRSs, the count of IRSs, and the count of reflecting units (RUs) that maximize the ergodic rate. The work in [71] investigated a communication network scenario that involves multiple users and employs multi-hop configurations for downlink communication using IRSs. The main goal was to maximize the total transmission rate by concurrently optimizing the beamforming at the BS and the phase adjustments of multiple IRSs while communicating with multiple users. Moreover, in [2], the focus shifted to an uplink cascaded IRS network. The main goals here were to increase the receiver power and expand the coverage distance in a THz communication system. To achieve this, the authors suggested a multi-hop system involving passive IRS operating in the THz frequency range to mitigate the substantial signal attenuation caused by air molecule absorption. In contrast to the research articles stated above [41, 39, 52, 71, 2] that used classical techniques to tackle their optimization challenges, the research articles [26] and [27] took a different approach by utilizing the DRL approach to tackle the issue of non-convex optimization. These studies introduced a hybrid beamforming strategy for networks with cascaded IRSs, to extend the coverage of THz networks. To reduce propagation losses in a THz downlink broadcast system, they inspected the joint design of the analog beamforming at the IRSs and digital beamforming at the BS. This scenario involves a single source transmitting to multiple destinations (multi-user scenario).

Notably, the existing literature on THz communication networks with cascaded IRS has extensively explored various aspects, including multi-hop unicast, M-MIMO systems, DF relaying, and uplink cascaded IRS scenarios. While classical optimization techniques have been employed in most studies, some have introduced DRL to tackle non-convex optimization challenges. However, a notable research gap exists as none of the existing works have utilized the DRL approach to solve the

over-determined system of equations in the uplink cascaded IRS multiple access scenario, specifically considering spatially correlated channels between IRS elements. This gap is addressed in this dissertation by employing DDPG to optimize the phase shifts of cascaded IRS for two alternative scenarios: maximizing the rate for a specific user and maximizing the aggregate rate for both users.

2.2.3 Security and Energy Harvesting in IRS NOMA

Several research studies have concentrated on the implementation of physical layer security IRS NOMA networks, recognizing the advantages offered by both IRS and NOMA technologies [44], [60, 59, 21, 18, 55, 73, 13, 3]. These studies aim to enhance network coverage, capacity, spectral efficiency, and connectivity on a large scale. For instance, [60] provided an overview of NOMA, addressing its limitations and highlighting the potential of the IRS to overcome these limitations. They also discussed the realization of Physical Layer Security (PLS) in IRS NOMA networks. Meanwhile, other studies such as [44, 59, 21, 18, 55, 73, 13] investigated secure NOMA IRS scenarios with diverse aims, including the maximization of sum-rate, improvement of secrecy performance, reduction of secrecy outage probability, enhancement of effective secrecy throughput, and reduction of eavesdropper's SINR. Notably, these prior works did not consider energy harvesting as a means of transferring energy to legitimate users in their scenarios. In contrast, [3] proposed a non-cooperative NOMA scenario with a PLS scheme that utilized artificial noise (AN) to secure the downlink and transfer energy to legitimate users but did not consider the incorporation of IRS. To the best of our knowledge, none of the existing studies in the literature have delved into cross-layer design, accounted for packet loss calculations at the network layer, or explored novel ML-based solutions instead of relying on conventional mathematical techniques to tackle optimization problems.

Chapter 3

Downlink Non-Orthogonal Multiple Access (NOMA) with Intelligent Reflecting Surfaces (IRS) Empowered by DRL

In this chapter, our primary focus is on the first major contribution of this thesis, which involves the development of a DRL approach for an IRS NOMA system. The primary aim is to maximize the total system sum-rate while addressing several critical challenges that arise in such a setup. The initial challenge we address pertains to the constraint of the constant amplitude of the IRS, which makes the optimization of sum-rate using IRS phase shifts a non-trivial task, requiring innovative solutions. The second challenge pertains to the limited CSI between the IRS and the users. In practice, obtaining accurate CSI can be difficult, which makes it necessary to develop strategies that work effectively under such uncertainty. Lastly, we consider the incorporation of the SIC technique. SIC is an essential aspect of NOMA networks, where users with stronger channel conditions help in decoding and canceling the signals of other users with weaker channels. This introduces additional complexity in the optimization process. To address these challenges, we suggest a novel DRL-based method that learns to optimize the phase shifts of the IRS and the distribution of power among the users. By employing DRL, the system can adapt and make decisions based on its environment and experience, even in the presence of limited CSI and constant modulus constraints. Overall, our focus in this chapter is on presenting the design and formulation of the DRL approach for the IRS-NOMA system, which is aimed at achieving significant gains in the sum-rate and addressing the practical limitations of IRS and NOMA technologies.

3.1 System Model

We investigate the downlink transmission of a NOMA system supported by an IRS, as illustrated in Fig. 3.1. The system consists of $\mathcal{K}$ users, each equipped with a single antenna. The BS is equipped with a single antenna as well. To simplify the analysis, we assume that the users are listed in ascending order based on their distances from the IRS, with user 1 being the farthest away and user $\mathcal{K}$ being the closest. This sequence is determined by the expected magnitudes of their channel gains, guaranteeing the condition $|\mathbf{h}_{r,1}|^2 < |\mathbf{h}_{r,2}|^2 < \ldots < |\mathbf{h}_{r,\mathcal{K}}|^2$. We examine a quasi-static scenario, where the ordering of users (based on their distances from the IRS) changes slowly over time. When such changes occur, the system initiates power allocation reconfiguration to adapt to the new user ordering effectively. In this setup, there exists no direct connection or LOS communication link that connects the BS to the users. Instead of communicating directly with each other, the BS and users communicate through an IRS. The IRS comprises N reflecting elements, organized in a grid structure with dimensions N_x and N_y.

Furthermore, our analysis incorporates several practical aspects of the system, including:

1. **Power Allocation with IRS using DRL**: The use of DRL is employed for average-based power allocation NOMA and equal power allocation OMA schemes. In the NOMA configuration, power allocation factors α_i are assigned to users according to their distances from the IRS, without relying on instantaneous CSI. The power allocation factors satisfy the condition $\alpha_1 > \alpha_2 > ... > \alpha_{\mathcal{K}}$, and the vector of power allocation coefficients is denoted as $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_{\mathcal{K}}]$. The power allocation is determined based on the long-term mean of the random channel gains.

2. **Successive Interference Cancellation (SIC)**: Users implement SIC using long-term channel statistics, considering the limited information of the instantaneous CSI between the IRS and users, the SIC process process relies on statistical channel information.

3. **End-to-end estimation**: The BS can estimate the complete channel, which combines the effects of both the transmit ($\mathbf{h}_t$) and receive ($\mathbf{h}_{r,\kappa}$) channels. Pilot transmission conducted by the BS and users is used to estimate the overall channel, as both are active nodes [64].

   These practical aspects are taken into account to ensure the feasibility and effectiveness of the suggested DRL-based power allocation algorithm in IRS-aided NOMA systems.

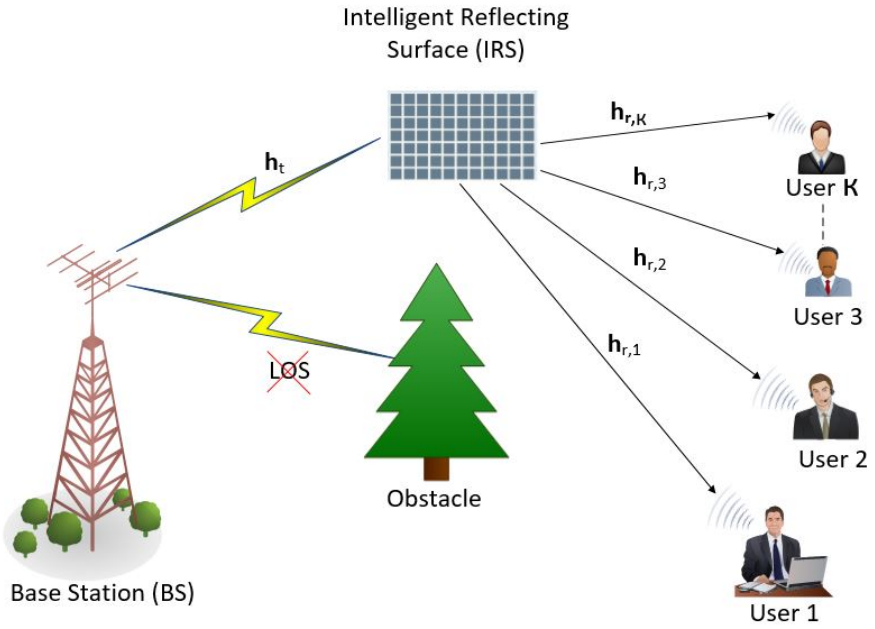


Figure 3.1: Enhancing wireless communication with IRS.

Our system model operates by utilizing static power allocation while simultaneously optimizing the proficiency of the NOMA system. This optimization is achieved by adapting the channel using phase adjustments implemented by the IRS. These adjustments compensate for random channel variations and account for the use of long-term SIC ordering, which remains fixed across multiple channel realizations, as opposed to the conventional approach where the SIC order changes for each realization.

The signal that is transmitted by the BS, denoted as x, is the sum of signals intended for each user, and it is expressed below:

$$x = \sum_{\kappa=1}^{\mathcal{K}} \sqrt{P_\kappa} S_\kappa, \tag{3.1}$$

In this context, the power allocated to each user κ, represented as $P_\kappa = \alpha_\kappa P_t$, adheres to the condition $P_1 > P_2 > ... > P_{\mathcal{K}}$, P_t is the BS transmit power. Each user's signal, denoted as S_κ, has a normalized power (i.e., $\mathbb{E}[|S_\kappa|^2] = 1$ for $\kappa \in 1, ..., \mathcal{K}$, and $\mathbb{E}[.]$ represents the mathematical expectation).

The signal received at user $\mathcal{K}$, designated as y_κ, is composed of several components, including the path loss between the BS and IRS, represented by c_t; the path loss between the IRS and user $\mathcal{K}$ denoted as $c_{r,\kappa}$; the communication channel link from the BS to the IRS, expressed as $\mathbf{h}_t \in \mathbb{C}^{1\times N}$; the matrix of phase shifts denoted as $\mathbf{\Phi} \in \mathbb{C}^{N\times N}$; the channel link from the IRS to user $\mathcal{K}$, symbolized by

$\mathbf{h}_{r,\kappa} \in \mathbb{C}^{N\times 1}$; and finally, the presence of additive white Gaussian noise (AWGN) represented as n. Accordingly,

$$y_\kappa = c_t c_{r,\kappa} \mathbf{h}_{r,\kappa}^H \mathbf{\Phi} \mathbf{h}_t x + n, \tag{3.2}$$
$$y_\kappa = c_t c_{r,\kappa} \mathbf{h}_{r,\kappa}^H \mathbf{\Phi} \mathbf{h}_t \sum_{\kappa=1}^{\mathcal{K}} \sqrt{P_\kappa} s_\kappa + n,$$

In the presented model, we have various parameters and notations. Let $c_t = (1 + d_t^{\Xi_t})^{-\frac{1}{2}}$ represent a certain coefficient related to the distance d_t separating the BS and the IRS. Similarly, $c_{r,\kappa} = (1 + d_{r,\kappa}^{\Xi_r})^{-\frac{1}{2}}$ is another coefficient associated with the distance $d_{r,\kappa}$ separating the IRS and user $\mathcal{K}$. Here, Ξ_t and Ξ_r represent the path loss exponents associated with the link between the BS and the IRS, and between the IRS and user, $\mathcal{K}$ [35, 9].

The matrix $\mathbf{\Phi}$ is a diagonal matrix that follows the unit modulus constraint; thereby, ensuring that the absolute value of each phase shift element equals 1. In other words, $|\phi_i|^2 = |e^{j\theta_i}|^2 = 1$ for all elements i in the range 1,2, ..., N, where N is the total number of passive reflecting elements. The phase shift associated with the i^{th} passive reflecting unit is denoted as θ_i, and its value ranges from 0 to 2π. Additionally, the channel links from the BS to the IRS and from the IRS to users are modeled using the rician fading model. The BS-IRS channel includes LoS and NLoS components, while the IRS-user channel encompasses both LoS and NLoS components.

$$\mathbf{h_t} = \sqrt{\frac{K_1}{K_1+1}} \bar{\mathbf{h}}_\mathbf{t} + \sqrt{\frac{1}{K_1+1}} \tilde{\mathbf{h}}_\mathbf{t}, \tag{3.3}$$

$$\mathbf{h_{r\kappa}} = \sqrt{\frac{K_2}{K_2+1}} \bar{\mathbf{h}}_{\mathbf{r}\kappa} + \sqrt{\frac{1}{K_2+1}} \tilde{\mathbf{h}}_{\mathbf{r}\kappa}, \tag{3.4}$$

The parameter K_1 characterizes the Rician factor of the channel $\mathbf{h_t}$. This channel is composed of two components: the LoS component, represented by $\bar{\mathbf{h}}_\mathbf{t} \in \mathbb{C}^{1\times N}$, and the NLoS component, represented by $\tilde{\mathbf{h}}_\mathbf{t} \in \mathbb{C}^{1\times N}$. Similarly, K_2 denotes the Rician factor associated with the channel $\mathbf{h}_{\mathbf{r},\kappa}$. This channel consists of two components as well: the LoS component, denoted by $\bar{\mathbf{h}}_{\mathbf{r},\kappa} \in \mathbb{C}^{N\times 1}$, and the NLoS component, designated as $\tilde{\mathbf{h}}_{\mathbf{r},\kappa} \in \mathbb{C}^{N\times 1}$.

Therefore, the received SINR for user κ is expressed as follows:

$$\gamma_k = \frac{c_t^2 c_{r,\kappa}^2 |\mathbf{h}_{r,\kappa}^H \mathbf{\Phi} \mathbf{h}_t|^2 P_\kappa}{c_t^2 c_{r,\kappa}^2 |\mathbf{h}_{r,\kappa}^H \mathbf{\Phi} \mathbf{h}_t|^2 \sum_{i=\kappa+1}^{\mathcal{K}} P_i + \sigma^2}. \tag{3.5}$$

Indeed, it is important to highlight that when $\kappa = \mathcal{K}$, the following expression $c_t^2 c_{r,\kappa}^2 |\mathbf{h}_{r,\kappa}^H \mathbf{\Phi} \mathbf{h}_t|^2 \sum_{i=\kappa+1}^{\mathcal{K}} P_i$ becomes equivalent to zero.

Moreover, the data rate of user κ is expressed as:

$$R_\kappa = \log_2(1 + \gamma_\kappa). \tag{3.6}$$

The illustration of the signaling and feedback model for our IRS NOMA system can be seen in Fig. 3.2. In this representation, the BS assigns power $P_k = [P_1, P_2, P_3, ..., P_{\mathcal{K}}]$ to the users. Following this, the users calculate their respective rates and transmit this information to the IRS, which plays the role of the DRL agent within the system. The agent calculates the reward based on the received information and then proceeds to adjust the phases of the IRS accordingly. Our DRL model tracks the changes in channel magnitude by continuously adapting the IRS phase to maximize the total transmission rate. Therefore, the system utilizes DRL to optimize the IRS phase according to the total transmission rate feedback from the users, enabling learning and achieving ideal sum-rate fine-tuning with the existence of partial CSI.

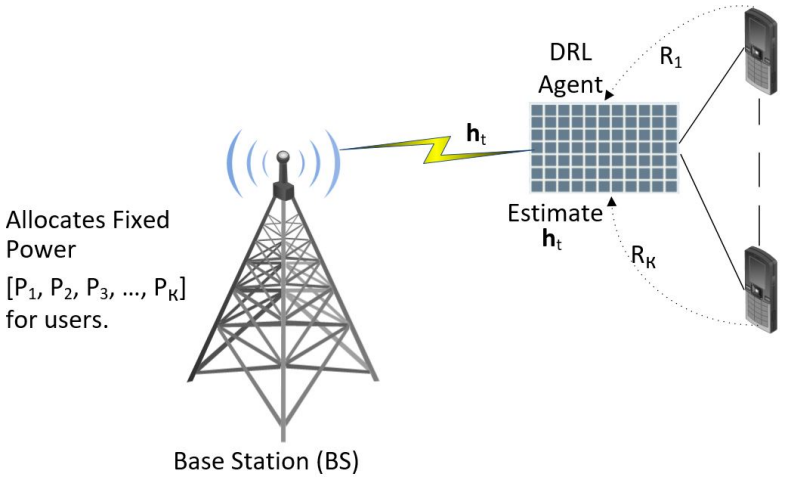


Figure 3.2: Feedback and signaling scheme for IRS-aided NOMA

3.2 Problem Formulation for IRS-NOMA System

The primary objective of this study is to determine the ideal phase shift values for the IRS units, aiming to maximize the total transmission rate of the users, as indicated by the following equation:

$$R_{sum} = \sum_{\kappa=1}^{\mathcal{K}} \log_2 (1 + \gamma_\kappa) . \tag{3.7}$$

Nevertheless, the assumption made in (3.7) of perfect SIC represents an ideal scenario. Achieving perfect SIC relies on two impractical assumptions: (i) having perfect knowledge of CSI and (ii) attaining perfect information decoding. In reality, these suppositions are not feasible. To present a more realistic approach, we consider that at every stage of interference cancellation, a residual fraction is introduced denoted as $0 \leq \epsilon \ll 1$, which represents the remaining interfering signal after the cancellation process [1]. Within our system, the power allocation and SIC for NOMA rely on the long-term averages of channel characteristics. Moreover, the adjustment of IRSs using DRL is employed to reconfigure the channel conditions and achieve optimal performance for NOMA. Consequently, the SINR received in (3.5) at user κ is given as:

$$\tilde{\gamma}_\kappa = \frac{c_t^2 c_{r,\kappa}^2 |\mathbf{h}_{r,\kappa}^H \mathbf{\Phi} \mathbf{h}_t|^2 P_\kappa}{c_t^2 c_{r,\kappa}^2 |\mathbf{h}_{r,\kappa}^H \mathbf{\Phi} \mathbf{h}_t|^2 \left(\epsilon \sum_{j=1}^{\mathcal{K}-1} P_j + \sum_{i=\kappa+1}^{\mathcal{K}} P_i \right) + \sigma^2}. \tag{3.8}$$

In the case where $\kappa = 1$, the expression $c_t^2 c_{r,\kappa}^2 |\mathbf{h}_{r,\kappa}^H \mathbf{\Phi} \mathbf{h}_t|^2 \sum_{j=1}^{\mathcal{K}-1} P_j$ becomes equal to 0. This implies that for the first user, there is no residual interference from the imperfect SIC caused by other users in the system. Furthermore, when $\epsilon = 0$, $\tilde{\gamma}\kappa$ is equal to $\gamma\kappa$, representing the ideal case. In this scenario, there is no residual interference after SIC, and the SINR received for user κ is equivalent to the SINR in an ideal, interference-free environment.

Thus, the problem formulated at the IRS is to obtain the matrix of the ideal phase shifts $\mathbf{\Phi}$ that maximizes R_{sum}, which is expressed as:

$$\max_{\Phi} \sum_{\kappa=1}^{\mathcal{K}} \log_2 (1 + \tilde{\gamma}_\kappa) , \tag{3.9}$$

$$\begin{aligned} & s.t. \\ & \text{K1} : |\phi_i|^2 = 1, \forall i \in \{1,2,...,N\}, \\ & \text{K2} : \alpha_\kappa \geq 0, , \\ & \text{K3} : \sum_{\kappa=1}^{\mathcal{K}} \alpha_\kappa = 1 \end{aligned} \tag{3.10}$$

Here, constraint K1 ensures that the IRS elements' phase shifts satisfy the constant amplitude constraint, ensuring that the magnitude of each phase shift element remains equal to 1. However, constraints K2 and K3, which represent the power allocation constraints, are redundant in the optimization problem. The

rationale behind this is that the system model relies on fixed power allocation coefficients determined from the statistics of the long-term channel, and thus, the power constraints are already satisfied. Hence, constraints K2 and K3 are no longer necessary for the optimization problem, and they can be eliminated.

Alternatively, the problem of optimization presented in (3.9) is reformulated as maximizing the SINR $\tilde{\gamma}_\kappa$ by obtaining the optimal phase shift $\mathbf{\Phi}$, as shown in (3.11). However, solving this problem is challenging because of the objective function and non-convex nature of the constant amplitude restriction [70] and [62]. Previous works have utilized the Semi-Definite Relaxation (SDR) method to achieve near-ideal solutions. Nevertheless, it is crucial to highlight that the computational complexity of the SDR method is excessively high, reaching $\mathcal{O}((N+1)^6)$ [49].

$$\max_{\Phi} \sum_{\kappa=1}^{\mathcal{K}} c_t^2 c_{r,\kappa}^2 |\mathbf{h}_{r,\kappa}^H \mathbf{\Phi} \mathbf{h}_t|^2 P_\kappa, \qquad (3.11)$$
$$s.t.\ |\phi_i|^2 = 1, \forall i \in \{1,2,...,N\}.$$

Moreover, considering the practical limitations of having incomplete information about the CSI for the IRS-user channels $\mathbf{h}_{r,\kappa}$ and the presence of imperfect interference cancellation, attempting to solve the optimization problem analytically becomes infeasible. Due to the system's dynamic nature and the complexity of the problem, we propose the adoption of RL techniques to find the ideal phase shifts of the IRS. By leveraging the power of RL, we can train an agent to acquire the optimal phase configurations that boost the total transmission rate, considering the system's dynamics and the limited CSI information available.

3.3 DDPG Approach for IRS Phase Control

This section outlines a method for controlling the phase shift of an IRS using the DDPG scheme (see Fig. 2.3). We address the problem of optimization described in (3.9) which aims to maximize the sum-rate considering SIC. DDPG is an RL technique that combines the benefits of PG and QL, and it does not require a model of the system. Given that the system's states mainly rely on the channel gains and the resultant sum-rate, and the actions are associated with the IRS phase shifts, we are dealing with a system characterized by continuous states and continuous actions. DDPG is advantageous in this situation as it can effectively manage systems with both continuous state and action spaces [14]. DDPG utilizes four NNs: actor, critic, target actor, and target critic networks. This architecture ensures stability during training. By learning the policy, DDPG can effectively solve the optimization problem presented in (3.9). This architecture is designed to ensure stability during training. By learning the policy, DDPG effectively addresses the optimization problem presented in (3.9).

3.3.1 System Mapping to DDPG

The initial step in using RL to solve a problem is to define the core elements of an RL scheme, which include the state space, action space, and reward function for the problem. In the subsequent discussion, we delve into this mapping process and provide an overview of the typical behavior exhibited by the RL scheme using the DDPG algorithm. During the mapping process, the problem at hand is translated into a suitable representation of actions, states, and rewards. The state space encompasses the relevant variables or information that characterizes the current state of the problem. The action-space represents the set of feasible actions that can be taken within the problem domain. The reward function quantifies the desirability or quality of the outcomes associated with different states and actions, guiding the RL algorithm during the learning process.

Once the problem is appropriately mapped, the RL method utilizing DDPG is employed. DDPG is a specific RL algorithm that combines the strengths of PG and QL. By utilizing NNs, DDPG can handle continuous state and action-spaces. It includes training for four various NNs. This architecture helps maintain stability during training.

In terms of behavior, the RL method using DDPG aims to learn an ideal decision-making strategy. It engages with the environment iteratively by observing states, taking actions, getting rewards, and adjusting the policy as necessary. Through this iterative process, the RL algorithm progressively improves its decision-making abilities, seeking to maximize the cumulative rewards obtained over time.

To summarize, the first step in utilizing RL to solve a problem involves mapping the problem onto the core elements of an RL system, encompassing the state-space, action-space, and reward function. Following this, the RL method, particularly DDPG, is utilized to acquire an optimal policy through iterative interactions with the environment, aiming to maximize cumulative rewards.

State-space

The DDPG agent's state-space at timestep τ is given below:

$$\mathbf{s}^{(\tau)} = [\mathbf{h}_t^{(\tau)}, \mathbf{\Phi}^{(\tau-1)}, \hat{\boldsymbol{\gamma}}^{(\tau-1)}], \tag{3.12}$$

Here, the communication link between the IRS serving as the agent and the BS acting as the source is represented by $\mathbf{h}_t$. The term $\mathbf{\Phi}^{(\tau-1)}$ denotes the previous action taken by the IRS. The phase action represents the specific adjustment or configuration applied to the IRS to manipulate the wireless signal reflections.

$\hat{\boldsymbol{\gamma}}^{(\tau-1)}$ refers to the estimated SINR values of the users in the previous state. These values depend on the rates of the users associated with the particular action taken by the IRS. The SINR reflects the ratio of the desired signal power to

the combined interference and noise power, indicating the quality of the wireless transmission for each user. It is represented as $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_\kappa, \ldots, \gamma_\mathcal{K}]$.

By incorporating these elements into the state-space, the DDPG agent can consider the channel gain, the previous IRS phase action, and the previous estimated SINR values for users as part of its decision-making process.

Action-space

The actions in the DDPG agent correspond to the IRS phase shift values. These phase shifts are real numbers obtained from the NN and utilized as the input of a complex exponential function to represent the actual phase. The resulting array determines the phase shift for each IRS unit. Consequently, the set of possible actions can be described using the subsequent policy function, formulated as follows:

$$\mathbf{a}^{(\tau)} = \mu(\mathbf{s}^{(\tau)}|o^\mu) + \mathbf{n}^{(\tau)} \tag{3.13}$$

Here, the policy function is represented by μ, while the parameters (i.e., weights of the NN) are represented by o^μ. The policy function accepts the state (τ) as input and returns the desired action $\mathbf{a}^{(\tau)}$.

Additionally, $\mathbf{n}^{(\tau)}$ represents the action noise based on the OU approach. This noise helps add exploration to the agent's actions, promoting diversity and avoiding the agent's confinement to local optima during the learning process.

To summarize, the DDPG agent's set of actions is defined by the policy function, taking the current state $\mathbf{s}^{(\tau)}$ as input and generating the action $\mathbf{a}^{(\tau)}$. The policy function is defined by the parameters o^μ, representing the NN weights. The action is further influenced by the OU process-based action noise $\mathbf{n}^{(\tau)}$, which encourages exploration during the learning process [56].

Reward function

The reward function is a function that rewards the agent for reaching the maximum capacity ever attained. It is expressed as below:

$$r^{(\tau)} = R_{\text{sum}}^{(\tau)} - R_{\text{sum,max}} \tag{3.14}$$

where $R_{\text{sum}}^{(\tau)}$ symbolizes the actual total users' transmission rate at time-step τ, which represents the aggregated data rate achieved by all users in the system. Conversely, $R_{\text{sum,max}}$ denotes the maximum attained sum-rate during the learning phase.

The reward $r^{(\tau)}$ is determined as the difference between the current aggregated rate $R_{\text{sum}}^{(\tau)}$ and the highest aggregated rate ever reached $R_{\text{sum,max}}$. This formulation allows the DDPG agent to strive for higher sum-rates by incentivizing improvements over the best performance achieved thus far.

In conclusion, the reward function bases its evaluation of the agent's performance on the total users' rate. Therefore, the agent seeks out improvements to optimize system capacity by comparing the present sum-rate to the total transmission rate achieved.

3.3.2 Balancing Exploration and Exploitation in RL

To explore the continuous action-space in DDPG, the exploration is facilitated using noise produced by the Ornstein-Uhlenbeck (OU). The OU method generates noise by drawing samples from a correlated Gaussian distribution. It is specifically designed to produce temporally correlated noise, which helps the agent explore the action-space effectively. The correlated nature of the noise allows the agent to have smoother transitions between consecutive actions, promoting stability during the learning process.

By incorporating the OU process-based action noise, the DDPG agent can explore a wide range of actions, preventing it from getting stuck in local optima and encouraging the discovery of better policies. This exploration is crucial for the agent to learn and adapt in complex environments with continuous action-spaces.

In summary, the OU process is employed to generate noise that enables exploration in the agent's action-space. This noise is sampled from a correlated normal distribution, providing smooth transitions between actions and facilitating effective learning and adaptation.

3.3.3 DDPG Scheme

The core purpose of the DDPG scheme is to facilitate the training of the IRS agent to make optimal decisions in response to variations in unknown environments. Specifically, the agent's specific objective is to iteratively determine the ideal phase shifts that maximize the total rate through successive iterations of the DDPG scheme. Notably, the agent has limited CSI concerning the link between the IRS and users. To address this, the agent relies on feedback from the BS regarding estimated channels to obtain the CSI of the channel between the BS and IRS. However, the agent faces the limitation of not having direct access to CSI for the communication link between the IRS and the receiver. To overcome this constraint, the agent employs an indirect indicator by utilizing the feedback SNR. This SNR feedback provides an alternative measure that the agent can use to find out the quality of the IRS-to-receiver channel, enabling it to make informed decisions in the absence of direct CSI. The tuning of the IRS phases in this setup is effectively tuning the channel effect.

The ultimate goal of the DDPG scheme is to train the IRS agent in such a way that it can choose actions leading to the maximization of the long-term mean reward, which, in this situation, is equivalent to maximizing the total transmission

rate while adapting to the changes in unknown environments. The agent achieves this objective by learning to adapt its randomized policy effectively, considering the stochastic behavior of the environment. This adaptive learning allows the agent to cope with uncertainties in the environment and ultimately results in the optimization of long-term rewards. It is crucial to emphasize that the prime objective of the DDPG scheme is not to provide the best possible immediate reaction to the channel's stochastic fluctuations. Instead, the algorithm focuses on learning the probabilistic representation of the uncertain environment and adapting its unique policy for responding to statistical patterns accordingly to maximize long-term rewards. This long-term optimization ensures that the agent can make informed decisions in the face of uncertainties and achieve overall better performance over time.

During each iteration, the IRS agent watches the current state, which consists of the transmitting channel ($\mathbf{h}_t^{(t)}$), the previous action taken ($\mathbf{\Phi}^{(t-1)}$), and the previously estimated SINR ($\hat{\boldsymbol{\gamma}}^{(t-1)}$). Subsequently, it employs the actor-network to compute the action ($\mathbf{\Phi}^{(t)}$) that boosts the reward, as specified in (3.14), which corresponds to the total sum-rate.

Furthermore, the critic-network uses both the state and action inputs to obtain the aggregated rate. When the aggregate rate from the users is collected, a new state is observed, prompting the IRS to adjust the phase shifts accordingly. The policy parameters (o^{μ}) used to calculate the action are modified according to the total rate feedback received from the user devices. This iterative process continues until the system acquires the ability to attain optimal sum-rate tuning despite the limitations posed by limited CSI. To gain a more comprehensive understanding of the DDPG algorithm, kindly refer to Algorithm 1.

3.3.4 Neural Network Architecture

The DDPG agent employs a four-NN architecture that includes actor, critic, target actor, and target critic networks. This design provides stability when training. Each actor and critic-network is made up of 2 hidden layers, each with 256 nodes. The actor-network accepts an input of size $2N + \mathcal{K}$ and produces an output of size N, aligning with the continuous nature of DDPG. The DDPG agent's structure enables significant scalability with progressively growing complexity.

3.4 Complexity Analysis

To assess the effectiveness of the DDPG algorithm, we compare its complexity to that of the ES algorithm. The complexities of both algorithms can be inferred from the provided descriptions in Algorithm 1 and Algorithm 2. In our analysis, we focused on the computations performed during the exploitation stage of the DDPG

algorithm to evaluate its complexity, which heavily relies on the actor-network's architecture. Deep NNs consist of input, output, and concealed layers (i.e. the layers that lie between the input and output strata of a DNN). We examine several factors in our analysis, including the count of states ($\mathcal{S}$), the number of neurons within the input of each layer ($\mathcal{I}$), the number of concealed layers ($\mathcal{H}$), the number of neurons in the output of each layer ($\mathcal{O}$), and the count of actions ($\mathcal{A}$). The complexity of the input layer relates to $\mathcal{S} \times \mathcal{I}$, and the complexity of the concealed layers correlates to $\mathcal{H} \times \mathcal{I} \times \mathcal{O}$, and the complexity of the output layer is connected to $\mathcal{O} \times \mathcal{A}$. Consequently, the cumulative complexity of the DRL algorithm can be expressed as follows:

$C_{DDPG} = \mathcal{S} \times \mathcal{I} + \mathcal{H} \times \mathcal{I} \times \mathcal{O} + \mathcal{O} \times \mathcal{A}$.

Furthermore, in the DDPG scheme, the action yielding the highest reward is always selected, and a linear search is performed on the output. Consequently, the overall computational complexity of a NN forward pass can be represented as:

$C_{DDPG} = \mathcal{S} \times \mathcal{I} + \mathcal{H} \times \mathcal{I} \times \mathcal{O} + \mathcal{O} \times \mathcal{A} + \mathcal{A}$.

Please note that the above analysis assumes a simplified perspective and does not consider additional factors such as activation functions, regularization techniques, or the training process itself.

The complexity of the ES approach is described as follows when $\mathcal{K}$ users, N IRS elements, and $\lfloor \frac{2\pi}{\Delta\Phi} \rfloor$ phase search steps are taken into account:

$$C_{ES} = O\left(\mathcal{K} \times \left(\lfloor \frac{2\pi}{\Delta\Phi} \rfloor + 1\right)^{N}\right). \tag{3.15}$$

Here, C_{ES} represents the complexity of the ES scheme. The computational effort increases exponentially as the count of IRS elements N and phase search steps $\lfloor \frac{2\pi}{\Delta\Phi} \rfloor$ increases. Hence, the search space expands significantly, making the ES approach computationally demanding and impractical for large-scale scenarios.

Consequently, with an increase in the number of users or IRS units, the complexity of the DDPG algorithm experiences a notably more pronounced reduction compared to the ES algorithm. Hence, DDPG demonstrates substantially reduced complexity compared to ES as the count of users or IRS units rises.

3.5 Bench-marking Assessment Schemes

To evaluate the effectiveness of the suggested algorithm, it is essential to have reference systems for comparison. However, the current body of literature lacks a definitive upper boundary or theoretical constraint for the examined system, and establishing such an upper limit poses intricate and demanding challenges. Hence, we introduce two benchmark reference models to assess the effectiveness of our approach. The first reference model is based on a discretized ES approach of the

IRS phases. While this method serves as an approximate upper bound, it is not a precise theoretical limit. The ES technique explores various combinations of IRS phases to seek the optimal solution. The second reference model employs OMA and serves as a performance lower bound. This benchmarking scheme provides a baseline comparison for assessing the efficiency of our suggested approach. OMA assumes that the available resources are divided among the users in an orthogonal manner, thus eliminating interference between users. By evaluating our suggested scheme's performance to that of existing reference models, we can gain insights into its effectiveness and understand how it performs relative to the upper and lower bounds.

3.5.1 Upperbound on Performance

Algorithm 2 ES Approach for IRS Phase Shifts' Matrix

Set $\Delta\Phi = \frac{2\pi}{30}$, N = 4.
for $\phi_1 = 0 : \frac{2\pi}{30} : 2\pi$; **do**
    for $\phi_2 = 0 : \frac{2\pi}{30} : 2\pi$; **do**
        for $\phi_3 = 0 : \frac{2\pi}{30} : 2\pi$; **do**
            for $\phi_4 = 0 : \frac{2\pi}{30} : 2\pi$; **do**
                Compute and save $R_{sum}(\phi_1, \phi_2, \phi_3, \phi_4)$
            Endfor
        Endfor
    Endfor
Endfor
Obtain the optimum $\Phi^* = \text{argmax}_{\phi_1,\phi_2,\phi_3,\phi_4} R_{sum}$

To evaluate the efficiency of the DDPG scheme compared to the upper bound, we employ a discretized ES method. This approach allows us to approximate the total transmission rate by searching for the matrix of optimal phase shifts. The algorithm used for this purpose is outlined in Algorithm 2. To mitigate the computational complexity associated with the ES, we limit the number of IRS RUs. This serves as evidence that the DDPG scheme can effectively approximate the upper limit. Specifically, for every IRS unit, we consider phase values ranging from 0 to 2π with an incremental step of $2\pi/30$. This yields 30^M phase shift matrix combinations. Subsequently, we compute the total transmission rate for $\mathcal{K}$ end-users based on these combinations.

3.5.2 Baseline Scheme: Orthogonal Multiple Access (OMA)

The OMA signal model assumes that the available resources, whether in terms of frequency or time, are equally divided among the $\mathcal{K}$ users. OMA scheme allows the

users to receive the signal without experiencing interference. In contrast, NOMA enables simultaneous transmission and offers control over interference. However, serving $\mathcal{K}$ OMA end-users using TDMA or FDMA demands $\mathcal{K}$ time slots or frequency channels, where each user occupies a specific frequency or time slot.

The BS transmitted signal in the OMA scheme is expressed as:

$$x_{\kappa}^{OMA} = \sqrt{P_t} s_{\kappa}, \tag{3.16}$$

From the user's perspective, the received signal can be expressed as:

$$y_{\kappa}^{OMA} = c_t c_{r,\kappa} \sqrt{P_t} s_{\kappa} \mathbf{h}_{r,\kappa}^{H} \mathbf{\Phi} \mathbf{h}_t + n, \tag{3.17}$$

Therefore, the SNR for user κ is given as:

$$\gamma_{\kappa} = \frac{c_t^2 c_{r,\kappa}^2 |\mathbf{h}_{r,\kappa}^{H} \mathbf{\Phi} \mathbf{h}_t|^2 P_t}{\sigma^2}, \tag{3.18}$$

The rate for user κ can be calculated as:

$$R_{\kappa}^{OMA} = \frac{1}{\mathcal{K}} \log_2(1 + \gamma_{\kappa}^{OMA}), \tag{3.19}$$

Hence, the total OMA rate is expressed as:

$$R_{sum}^{OMA} = \frac{1}{\mathcal{K}} \sum_{\kappa=1}^{\mathcal{K}} \log_2 \left(1 + \frac{c_t^2 c_{r,\kappa}^2 |\mathbf{h}_{r,\kappa}^{H} \mathbf{\Phi} \mathbf{h}_t|^2 P_t}{\sigma^2} \right). \tag{3.20}$$

where R_{sum}^{OMA} symbolizes the sum of the individual data rates for each user, normalized by $\mathcal{K}$.

3.6 Performance Evaluation and Analysis

The numerical results presented showcase the proficiency of the DRL-based IRS NOMA communication network, considering an IRS with $N = 16$ RUs. The default values used for the evaluation of the DRL scheme are outlined in Table 3.1. Specifically, the system configuration includes one BS antenna ($N_t = 1$) and one antenna per end-user ($N_r = 1$). The distance separating the BS and the IRS is fixed at 50 m, but the distances separating the IRS and the users are produced at random between 200 and 1500 m. The communication channel link separating the BS and IRS, as well as the links separating the IRS and users, are modeled using the rician fading model with rician factors $K_1 = K_2 = 10$. The communication link between the BS and the IRS is assumed to be accurately determined, whereas the link from the IRS to users has limited CSI. The system operates within a 10 MHz bandwidth, with a BS transmit power (P_t) of 40 dBm and a noise power spectral density

Table 3.1: Parameters Used in Simulation

Simulation Parameters	**Values**
IRS RUs (N)	16
User $\mathcal{K}$ antennas N_r	1
BS antennas N_t	1
d_t	50
$d_{r,\mathcal{K}}$	200 - 1500
P_t	40 dBm
Network bandwidth	10 MHz
Noise PSD	-174 dBm/Hz
Ξ_t	2
Ξ_r	2.8
K_1, K_2	10
Critic-network's $\digamma_C$	0.001
Actor-network's $\digamma_A$	0.0005
Discount Coefficient Γ	0.99
Soft Updates Factor υ	0.05
Batch size B	64
Capacity of the Memory Buffer C	100000

(PSD) of -174 dBm/Hz. The outcomes of the numerical simulations are obtained for 1000 Monte Carlo runs. Overall, these simulation results provide insights into the efficiency of the DRL-based IRS NOMA scheme, and they demonstrate the performance of the algorithm in achieving improved total rate performance through iterative learning and optimization.

The DDPG algorithm employs DNNs for both the actor and critic-networks. A total of 128 neurons are used as input by the actor-network, while 16 neurons are produced as output. It consists of two hidden layers, each with 256 neurons, succeeded by the ReLU activation unit. To ensure sufficient gradient, the actor-network output layer employs the tanh() function. The critic-network takes inputs for both states and actions. The critic-network takes inputs of both actions and states. The state is routed through 2 dense layers consisting of 128 and 256 neurons respectively, succeeded by ReLU activation units. The action passes through a single dense layer with 128 neurons. The outputs of both networks are then passed through separate layers before being concatenated to form the critic-network's input. Two further hidden strata each consisting of 256 neurons are added, succeeded by the ReLU activation unit. The critic-network concludes with an output layer containing 16 neurons. Adam optimizer is used by the actor and critic-networks to update their settings. Over 1000 channel iterations, the average sum-rate is computed. The learning rate for the actor-network, denoted as $\digamma_A$, is assigned a

value of 0.0005, the learning rate for the critic-network, denoted as $\digamma_C$, is assigned a value of 0.001, the soft update coefficient υ is set to 0.05, the discount factor Γ is adjusted to 0.99, and the buffer capacity C is configured to hold 100,000 entries. The noise is considered to be complex AWGN with a 0.1 variance and a zero mean.

Our primary focus in this section is to assess the DDPG algorithm's performance to ensure that the calculated total transmission rate values closely approximate the upper bound. To determine the maximum achievable sum-rate, we employ the ES approach, assuming perfect knowledge of the channel. Considering the high complexity of the ES scheme, we limit the count of IRS RUs to $N = 4$ instead of $N = 16$. This allows us to verify the capability of the DDPG algorithm to closely approximate the upper limit. For each reflecting unit, we examine phase shifts ranging from 0 to 2π with an incremental step of $\frac{2\pi}{30}$. Consequently, there are a total of 30^4 potential permutations of phase shift matrices. We compute the total rates for a system with 16 users and perform 1000 Monte Carlo iterations. By conducting these evaluations, we can determine the extent to which our suggested DDPG scheme can approximate the upper bound, thus validating its effectiveness in enhancing the total rate performance of the IRS NOMA network.

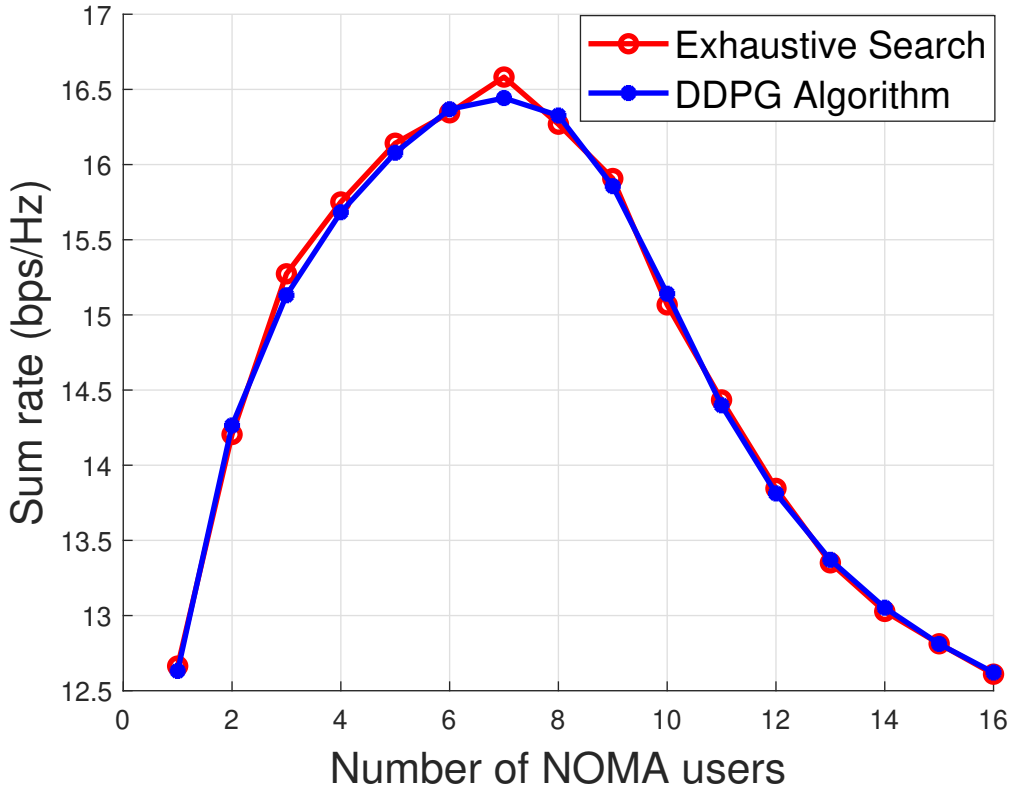


Figure 3.3: Comparing upper bound performance with the proposed DDPG scheme. $\mathcal{K} = 16$, $\mathrm{N} = 4$, and $\Delta\Phi = \frac{2\pi}{30}$.

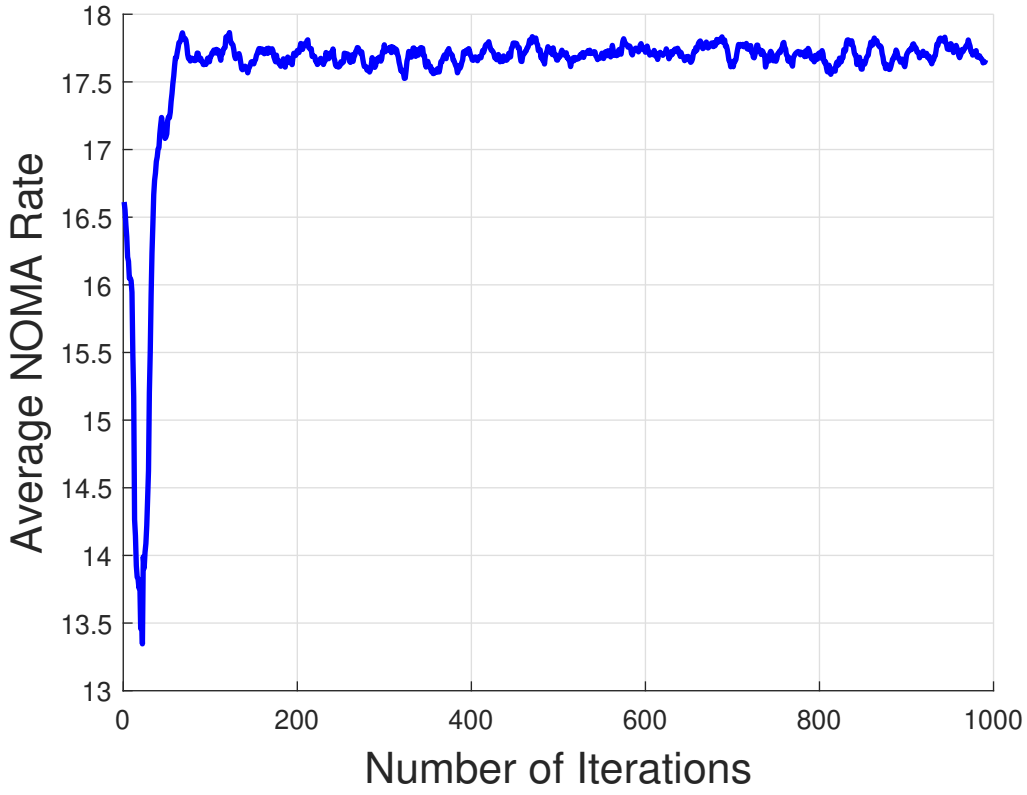


Figure 3.4: Comparing NOMA sum-rate with iteration plots.

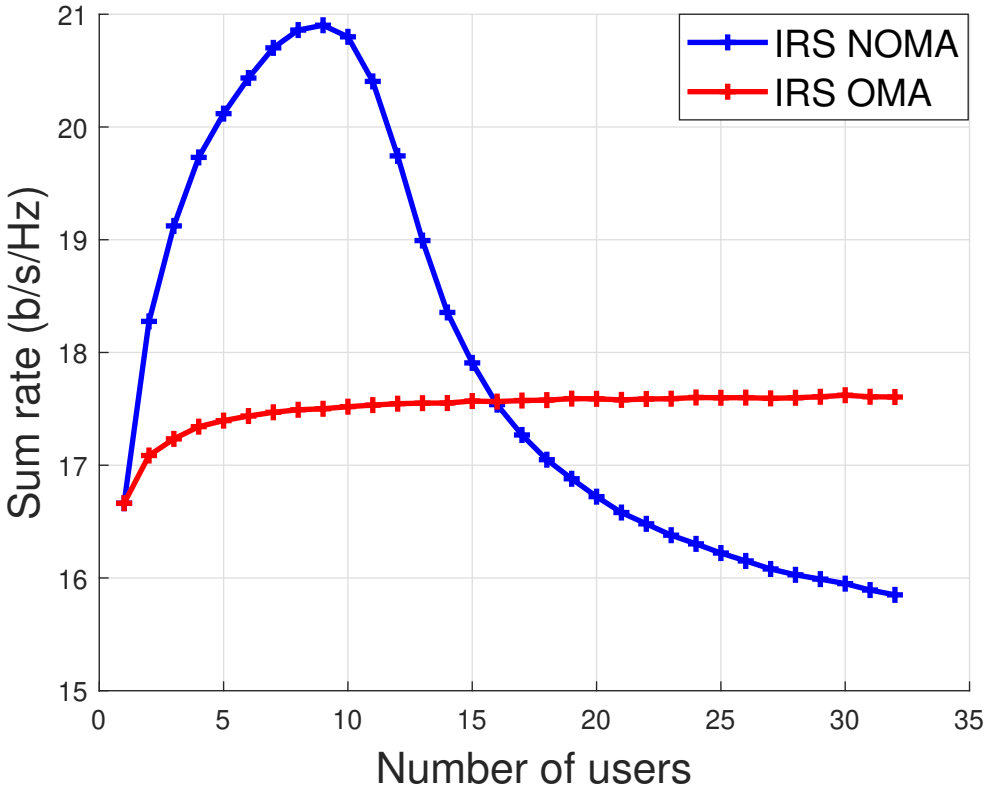


Figure 3.5: Comparing sum-rates of IRS NOMA and IRS OMA for various number of users. Pt = 40 dbm, and N = 16.

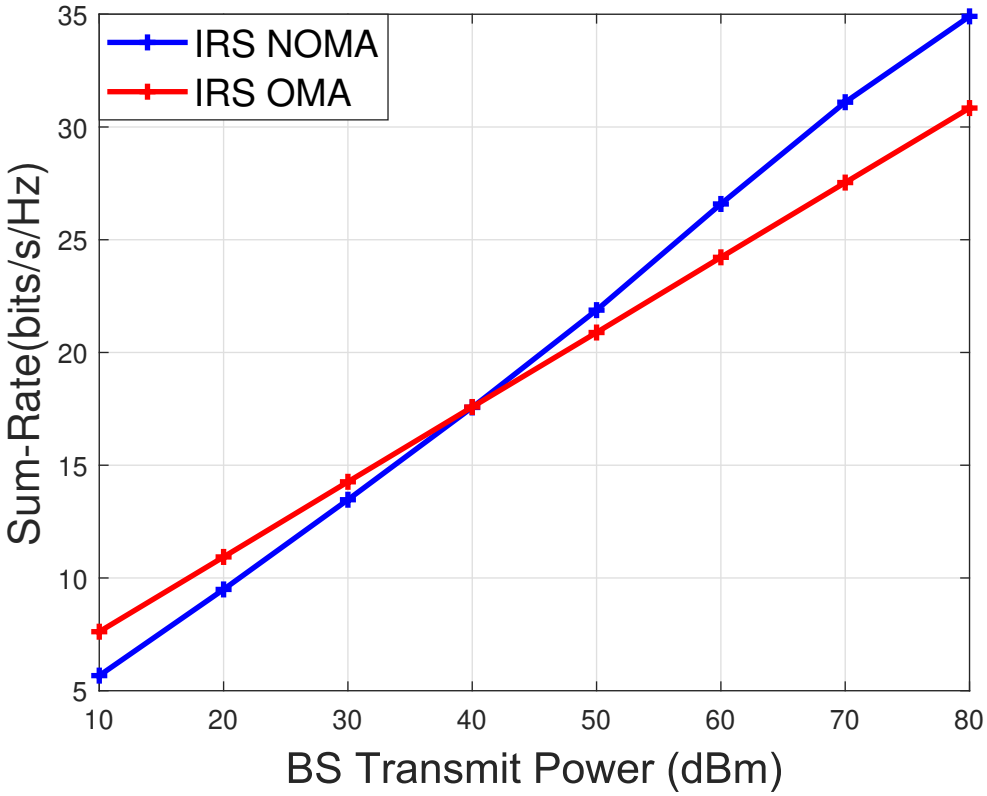


Figure 3.6: Comparison between IRS NOMA and IRS OMA sum-rates vs different power levels. $\mathcal{K} = 16$, and $N = 16$.

Comparatively, the complexity of the ES algorithm can be computed as the product of the number of Monte Carlo simulations 1000, the total users' count 16, and the total count of phase shift matrices (30^4), yielding a total of 1.2960×10^{10} iterations. The time necessary to complete this ES is roughly 124.86 hours. The total rate obtained by the DDPG scheme closely approximates the upper bound and is near to ideal, as revealed in Fig. 3.3. These findings reveal the computational efficiency and effectiveness of the DDPG algorithm in achieving near-optimal results for the problem of sum-rate optimization in the IRS NOMA communication system, as it achieves comparable performance to the ES algorithm with significantly reduced computational burden.

In Fig. 3.4, the convergence of the DRL method is demonstrated through the average NOMA rate versus iteration plots. It can be observed that the average rate increases with time until reaching an almost stable value, indicating the successful training process of the DRL algorithm. This reaffirms the effectiveness of our approach in optimizing the efficiency of the IRS NOMA network.

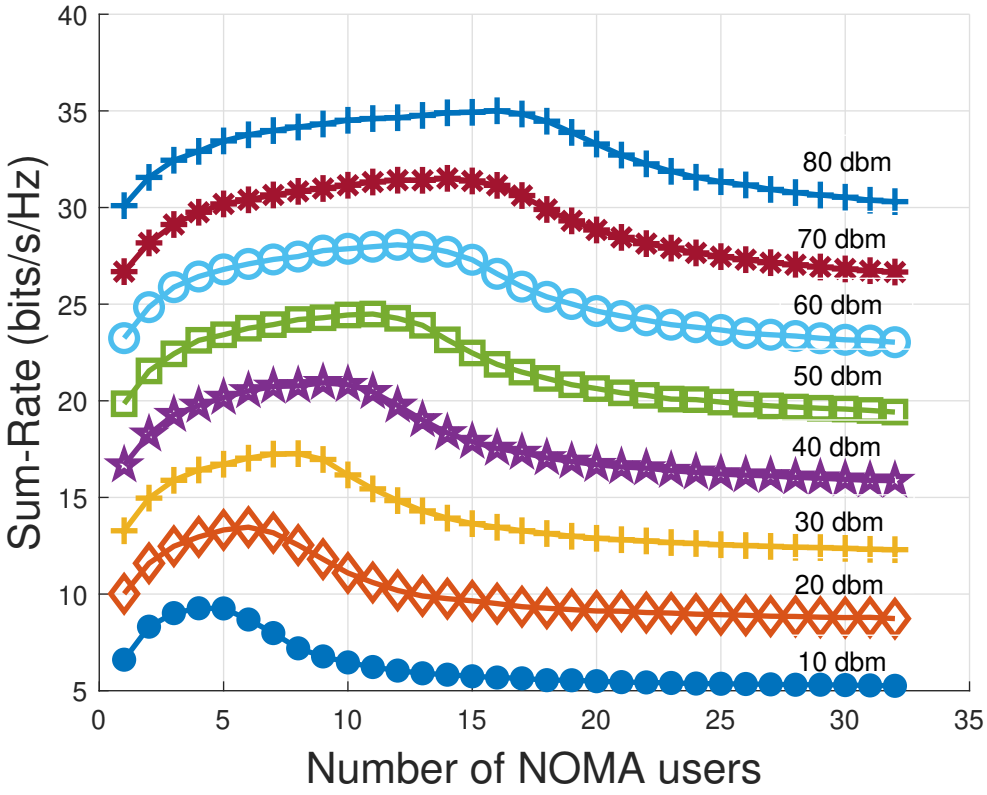


Figure 3.7: NOMA sum-rate analysis for different power levels N = 16.

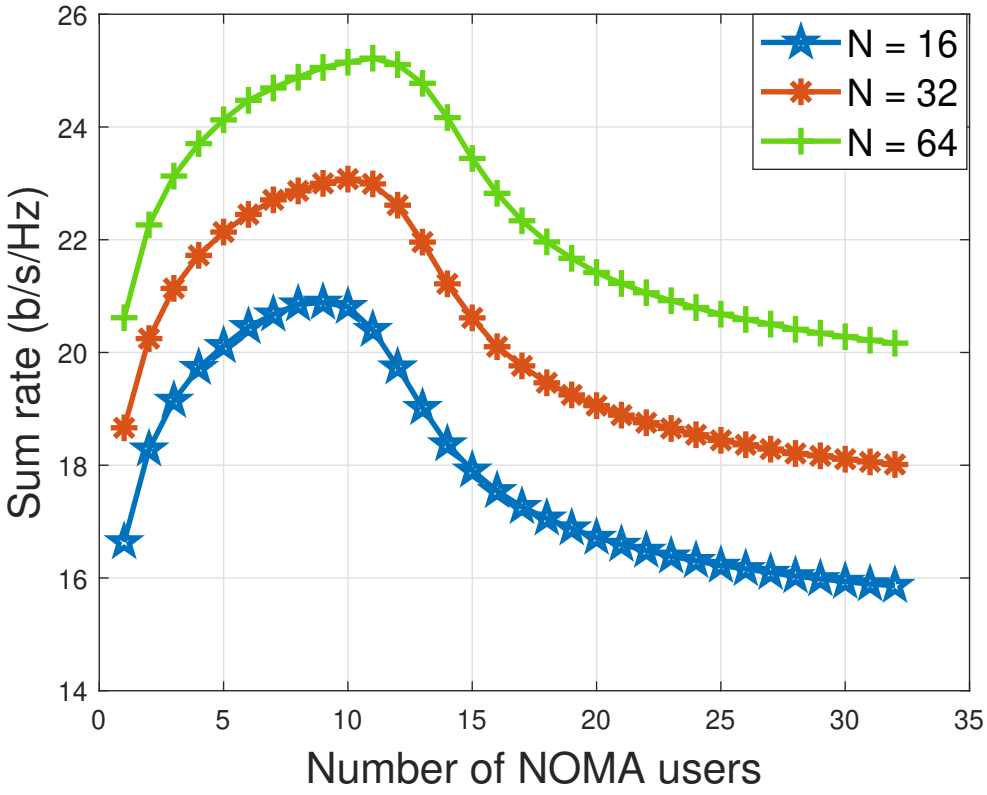


Figure 3.8: NOMA sum-rate analysis for different numbers of reflecting units. $\mathcal{K}$ = 32, and Pt = 40 dbm.

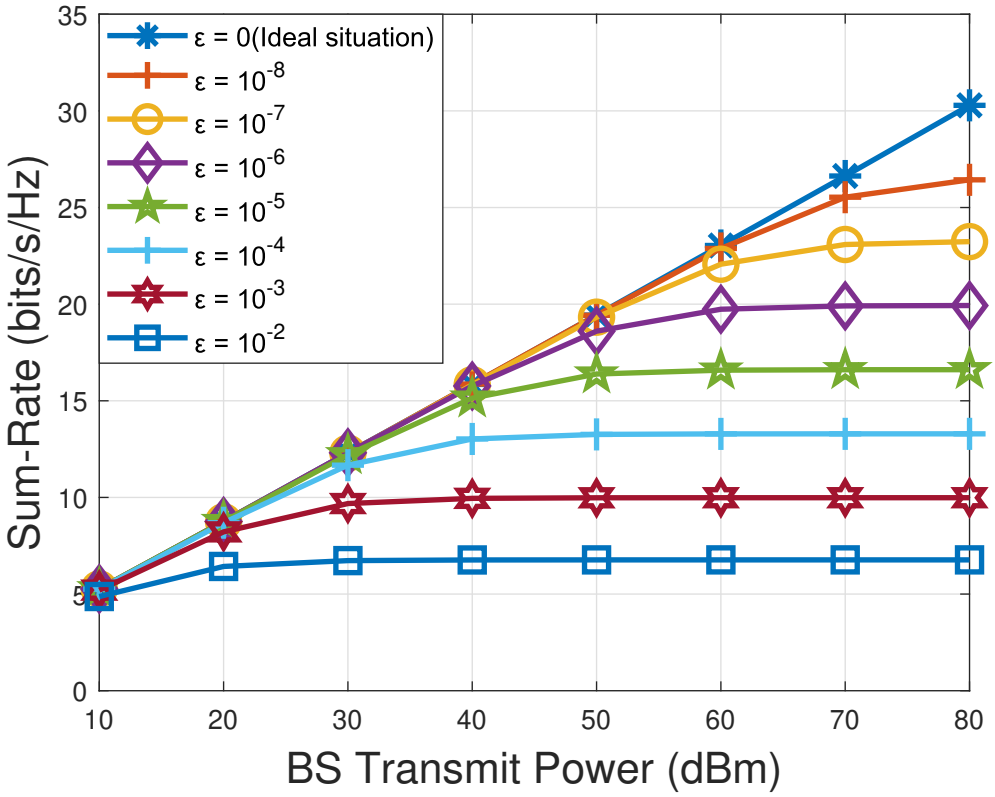


Figure 3.9: Total achievable data rate at the closest user, accounting for imperfect SIC. $\mathcal{K} = 32$, and $N = 16$.

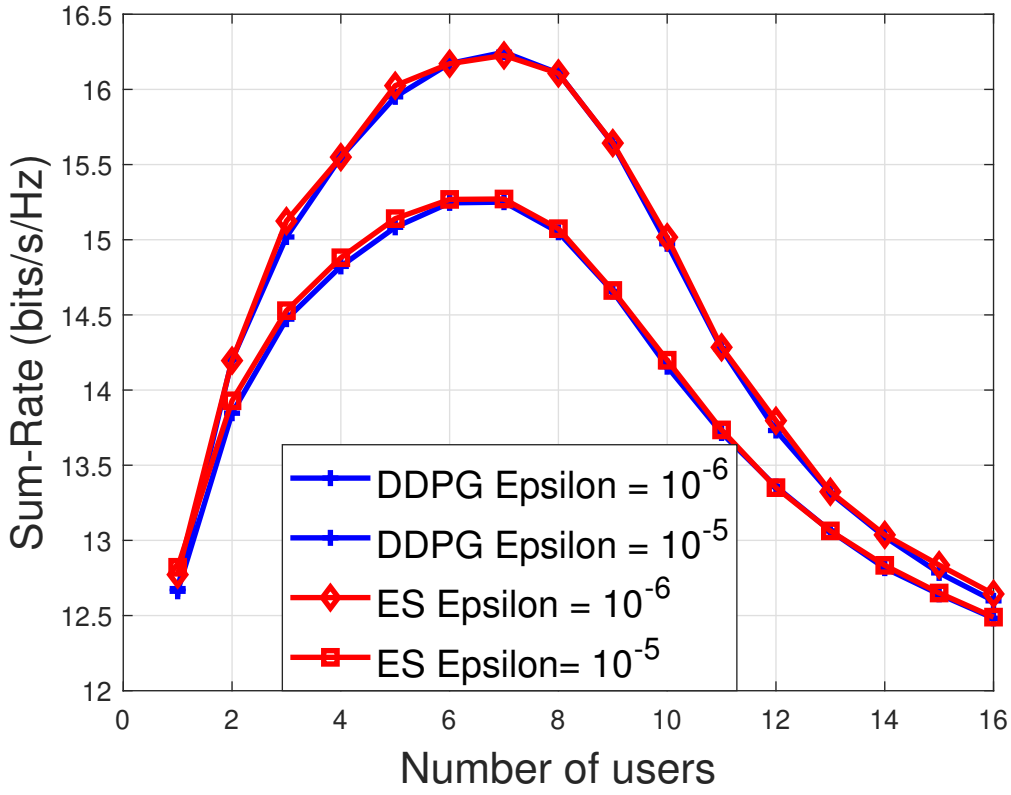


Figure 3.10: Comparing upper bound performance with the proposed DDPG algorithm incorporating imperfect SIC. $\mathcal{K} = 16$, N = 4,and $\Delta\Phi = \frac{2\pi}{30}$.

Figure 3.5 illustrates a comparison between the sum-rates of IRS NOMA and IRS OMA across varying user counts, with a transmit power of 40 dBm. NOMA outperforms OMA by offering a greater sum-rate value when the number of users is limited (less than 16). This is because NOMA allows users in the power domain to share resources, removing the requirement for bandwidth division. As a result, both the rate and the spectral efficiency have increased. OMA, on the other hand, does not entail resource sharing, resulting in user bandwidth partition. As the user count surpasses 16, the interference among users increases, and OMA outperforms NOMA by giving a greater sum-rate. Fig.3.6 shows that IRS OMA performs slightly superior to IRS NOMA at low transmission power levels and in scenarios involving 16 users. This is due to interference experienced by IRS NOMA users during simultaneous transmission at low SINR, but OMA users do not. At low SINR, the NOMA system lacks enough power to provide significant channel disparity across users, reducing the potential advantage afforded by NOMA. However, under conditions of higher power levels, the sum-rate of the IRS NOMA system surpasses that of the IRS OMA system.

Further, Fig. 3.7 illustrates the total rate of NOMA across varying power levels ranging from 10 dBm to 80 dBm. The uppermost curve signifies the aggregate rate at 80 dBm, while the lower curve corresponds to the total rate at 10 dBm. Notably, as the power level rises, the total data rate also demonstrates an increase, enabling the IRS NOMA system to accommodate more users effectively. Additionally, Figure 3.8 demonstrates the capacity of the DDPG algorithm to effectively handle a greater amount of IRS units, showcasing its scalability. However, due to the computationally prohibitive complexity, the ES approximate upper bound is not displayed in these cases. Importantly, it's crucial to highlight that the overall data rate performance sees enhancement when the number of IRS units increases. This improvement is attributed to the added degrees of freedom and the enhanced capability to focus the signal more effectively at the intended destination.

Moreover, Fig.3.9 and Fig.3.10 provide visual representations of the users' total data rate in scenarios involving imperfect SIC. In this context, the residual interference stemming from the power of all users remains within the denominator of the rate equation. Moreover, Fig. 3.9 displays the rate attributed to user κ, who is the nearest user to the BS, across varying power levels and ϵ values. User 1 positioned farthest from the BS, does not perform SIC, thereby directing the analysis toward examining the data rate specifically for user $\mathcal{K}$. It is evident that with increasing imperfection, the data rate for user $\mathcal{K}$ demonstrates a decline. The depicted curves in the plots correspond to various ϵ values, which signifies the proportion of residual interference. When ϵ is equal to 0, perfect SIC is achieved, resulting in the highest data rate for user $\mathcal{K}$. In comparison, when ϵ holds a value greater than 0, the achieved rates for user $\mathcal{K}$ are lower.

As the value of ϵ increases, the rate decreases due to the growing fraction of imperfection. Consequently, imperfect SIC has a negative impact on the rates of users employing SIC. Fig.3.10 depicts the performance of the DDPG algorithm versus the ES scheme during imperfect SIC. With the increase of ϵ values, there is a decrease in the data rate owing to the increased proportion of imperfection. Consequently, the rates of users employing SIC are adversely affected by the imperfection. This phenomenon is illustrated in Fig. 3.10, which showcases the comparative performance of the DDPG scheme against the ES algorithm under conditions of imperfect SIC. The calculation of the optimal phase shifts is performed to optimize the total data rate, considering the constraints of limited channel knowledge and ϵ values greater than zero. Remarkably, despite the presence of imperfect SIC, it's evident that the DDPG method consistently approaches the upper-performance limit.

Chapter 4

A Deep Reinforcement Learning Approach for THz Multiple Access in a Multi-hop IRS Network Topology

This chapter focuses primarily on the thesis's second main contribution, which involves developing a DRL approach for THz multiple access in a multi-hop IRS Network Topology. The objectives are twofold: maximizing the targeted user's rate while treating the second user as an interfering factor, and maximizing the total rate of the system.

For the first objective, one challenge involves handling the constant modulus constraints of the multi-hop IRSs, which makes optimizing the phase shifts at IRS_1 and IRS_2 complex and requires innovative solutions. Another potential problem in the uplink multi-hop IRS multiple access scheme is dealing with an over-determined set of equations, considering the different paths established through combinations of reflective elements in IRS_1 and IRS_2. To address these challenges, sub-optimal mathematical methods are utilized, and the results are compared with the proposed DRL scheme. The second objective is a more difficult problem to solve, as it is non-convex, NP-hard, and more complicated. It is challenging to find an analytical solution utilizing typical mathematical approaches, and ES becomes impractical for large-scale communication systems. To overcome these challenges, a novel DRL-based approach is introduced, employing DDPG to determine the optimum IRS phases in the multi-hop communication network, and the obtained results are compared to that of the ES scheme considering the limited number of RUs. The approach takes into consideration the spatial correlation of the channel between the two IRSs, and aims to achieve two scenarios: enhancing the rate for the desired user and increasing the overall rate for all users in our system.

4.1 System Model

We assume in the system model that our setup comprises two users, each equipped with a single antenna, operating in a wireless communication system where these users transmit data to a BS with the help of a multi-hop IRS. This configuration is depicted in Fig. 4.1. It is valuable to acknowledge that increasing the number of antennas and users does not provide additional insights but introduces complexity to the system without clear benefits. Additionally, both users have parabolic antennas that are highly directional, meaning that they can focus the transmitted signal toward the center of IRS_1. Further, the communication system operates at a frequency within the THz range. The decision to utilize THz frequencies is intentional, mainly because of their appropriateness in scenarios with constrained coverage areas, which is ideal for users 1 and 2. To address the substantial propagation losses experienced by THz signals due to absorption in air molecules, we have integrated a cascaded IRS system. As a result, this configuration significantly boosts signal strength, especially for signals encountering difficulties in traveling long distances due to their short wavelength and high frequency. The cascaded IRS setup emerges as a strategic solution to augment the overall resilience and efficiency of the communication process within the challenging THz spectrum. Moreover, the transmitter antennas have a diameter of D_t, while the receiver antenna has a diameter of D_r. The distances from each user to IRS_1, from IRS_1 to IRS_2, and from IRS_2 to the receiver are designated as $r_{t\kappa}$, r_2, and r_3, respectively. Similarly, the horizontal distances between each user and the center of IRS_1, IRS_1 and IRS_2, and IRS_2 and the receiver are symbolized as $r_{\kappa,1,h}$, $r_{2,h}$, and $r_{3,h}$, respectively. The angles of the incoming signal w.r.t. the center of IRS_1 and IRS_2 are denoted as $\psi_{i1,1}$, and $\psi_{i2,1}$, respectively. Similarly, the reflected angles from IRS_1 and IRS_2 w.r.t. their center are designated as $\psi_{i,2}$, and $\psi_{r,2}$, respectively. The two users' heights, IRS_1 height, IRS_2 height, as well as the receiver's height, are represented as $\ell_{Tx,1}, \ell_{Tx,2}$, ℓ_{s1}, ℓ_{s2}, and ℓ_{Rx}, respectively. The phase of each RU is adjusted by the IRS to steer the incoming signal toward a particular reflection path. E indicates the quantity of RUs within IRS_1, while F represents the number of RUs in IRS_2.

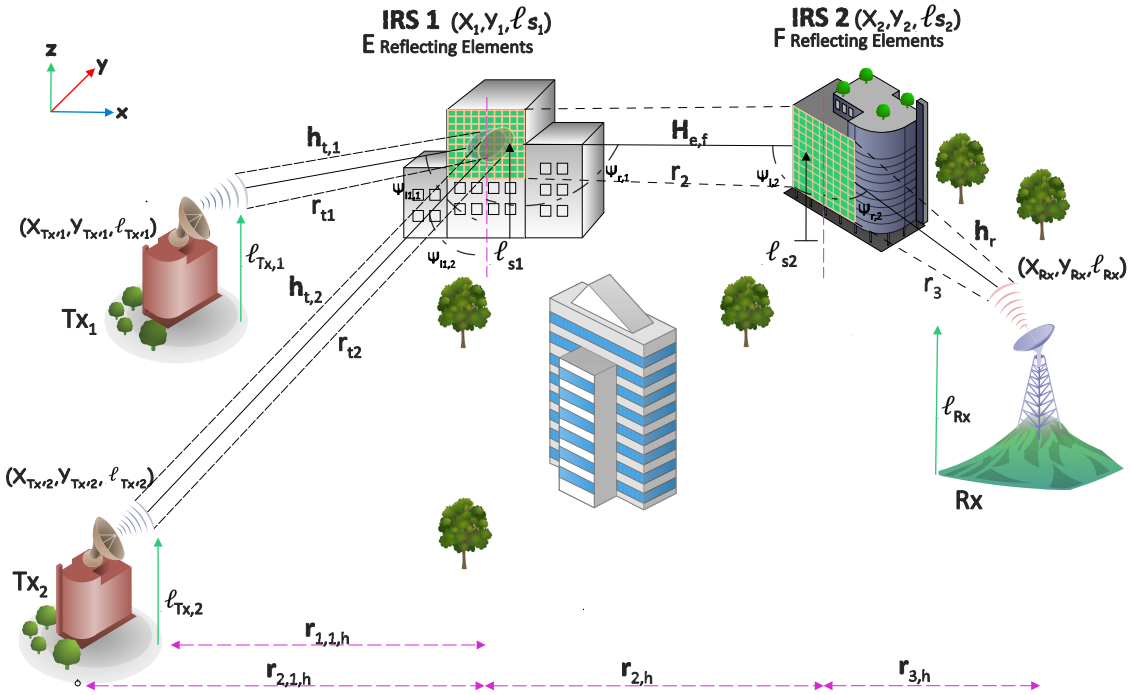


Figure 4.1: Multi-hop IRS network model.

The signal transmitted by each user κ is given as:

$$x_\kappa = \sqrt{P_t} z_\kappa, \tag{4.1}$$

where $\kappa \in [1,2]$, P_t designates the transmitting power for every transmitter, z_κ denotes user κ's signal with normalized power ($\mathbb{E}[|z_\kappa|^2] = 1$), and $\mathbb{E}[.]$ represents the expected value.

Hence, user κ's received signal is given as below:

$$y_\kappa = \mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{e,f}^H \mathbf{\Phi}_E \mathbf{h}_{t,\kappa}^H x_\kappa + n_0 \tag{4.2}$$

$$y_\kappa = \mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{e,f}^H \mathbf{\Phi}_E \mathbf{h}_{t,\kappa}^H \sqrt{P_t} z_\kappa + n_0, \tag{4.3}$$

In the system model, several parameters and variables are defined as follows: The communication channel link connecting user κ and IRS_1 is represented by the vector $\mathbf{h}_{t,\kappa}$, which has the dimension $1 \times E$. The communication link connecting IRS_1 and IRS_2 is depicted by the matrix $\mathbf{H}_{e,f}$, which possesses dimensions of $E \times F$. The communication link connecting IRS_2 and the receiver is illustrated by the vector $\mathbf{h}_r$, which is of size $F \times 1$. The matrices representing the phase adjustments for IRS_1 and IRS_2 are expressed as $\mathbf{\Phi}_E = \text{diag}(e^{-j\theta_1}, e^{-j\theta_2}, ..., e^{-j\theta_E})$ and $\mathbf{\Phi}_F = \text{diag}(e^{-j\omega_1}, e^{-j\omega_2}, ..., e^{-j\omega_F})$, correspondingly. These matrices ensure that the

constraint of the constant modulus is satisfied, such that $|\phi_e|^2 = |\Lambda e^{-j\theta_e}|^2 = 1$ for all $e \in 1,2,...,E$, and $|\phi_f|^2 = |\Lambda e^{-j\omega_f}|^2 = 1$ for all $f \in 1,2,...,F$, where Λ symbolizes the reflection amplitude, assumed to be uniform across all elements. The function diag(.) represents the diagonal matrix. The phase shifts of the $e^{th}$and f^{th} RUs are denoted by θ_e and ω_f, respectively. The superscripts e and f refer to the $e^{th}$and f^{th} RUs, respectively. These phase adjustments are used to manipulate the channel between the transmitter and the receiver. The objective is to tune the phase shifts of the RUs to maximize the received signal strength at the receiver. The values of θ_e and ω_f range from 0 to 2π. The AWGN at the receiver is represented by the random variable n_0, which takes the form of a complex Gaussian random variable with zero mean and a variance of σ^2, where σ^2 represents the noise power in linear scale. The signal phase originating from each user to IRS_1, as well as from IRS_2 to the receiver, can be calculated using the following equations.

$$\eta_\kappa = \frac{2\pi r_{t_\kappa}}{\lambda}, \tag{4.4}$$

$$\eta_3 = \frac{2\pi r_3}{\lambda}, \tag{4.5}$$

Here, the parameter λ denotes the wavelength, which is calculated as the speed of light (c), approximately 3×10^8 meters per second (m/s), divided by the frequency (f) measured in Hertz (Hz). By applying these equations, the deterministic phase shifts can be obtained based on the distances traveled by the signals. The distances (r_{t_κ} and r_3) are multiplied by 2π and divided by the wavelength to convert them into phase shifts in terms of cycles or radians. This allows precise control and manipulation of signal propagation in systems utilizing IRSs.

The communication channel links between the transmitters and IRS_1, $\mathbf{h}_{t,\kappa}$, and the communication link between the IRS_2 and the receiver, $\mathbf{h}_r$, are represented to follow the Rician fading model [27], [12]. This model is described as follows:

$$\mathbf{h}_{t,\kappa} = \sqrt{\frac{K_1}{K_1+1}}\bar{\mathbf{h}}_{t,\kappa} + \sqrt{\frac{1}{K_1+1}}\tilde{\mathbf{h}}_{t,\kappa}, \tag{4.6}$$

$$\mathbf{h}_r = \sqrt{\frac{K_2}{K_2+1}}\bar{\mathbf{h}}_r + \sqrt{\frac{1}{K_2+1}}\tilde{\mathbf{h}}_r, \tag{4.7}$$

Here, K_1 represents the Rician factor for $\mathbf{h}_{t,\kappa}$, while $\bar{\mathbf{h}}_{t,\kappa}$ corresponds to the LOS constituent of the channel, and $\tilde{\mathbf{h}}_{t,\kappa}$ represents the NLOS constituent of the channel. Similarly, K_2 signifies the Rician factor for $\mathbf{h}_r$, where $\bar{\mathbf{h}}_r$ denotes the LOS portion of the channel and $\tilde{\mathbf{h}}_r$ signifies the NLOS portion of the channel.

The channel between the two IRSs denoted as $\mathbf{H}_{e,f}$, follows a spatially correlated Rayleigh fading channel model. The choice of the Rayleigh fading model for the channels between the two IRSs is based on their proximity in location and operation

on the same frequency. In such conditions, there is no prevalent Line-of-Sight (LOS) link, hence the applicability of the Rayleigh fading model to characterize the channel behavior. Further, the exponential spatially correlated model is used to calculate the covariance matrix $\mathbf{R}$, where the correlation coefficient ρ acts as the controlling parameter and takes values within the range of [0,1]. The elements of the matrix $\mathbf{R}$ are expressed as follows:

$$[\mathbf{R}]_{e,f} = \rho^{|e-f|} e^{|e-f|\psi_{i,2}}, \tag{4.8}$$

Here, $\psi_{i,2}$ represents the angle of arrival between IRS_1 and IRS_2. Higher ρ values indicate a stronger correlation among the components of $\mathbf{H}_{e,f}$, with a significant correlation observed between adjacent RUs. The correlation decreases as the distance between RUs increases[49], [47].

It is assumed that the first hop channels $\mathbf{h}_{t,\kappa}$ and the third hop channel $\mathbf{h}_r$ are precisely known to both the users and the receiver. While channel estimation and obtaining CSI is a challenging task in a communication network including the IRS, several methods have been proposed to address this issue. For example, in [74], an effective channel estimation approach for a double multi-user MIMO IRS communication scheme was proposed. In [75], an extensive survey was conducted regarding channel estimation for wireless communications aided by the IRS, focusing on practical design solutions. Additionally, in [24], a system for channel estimation involving the IRS was introduced. This system employs a limited number of IRS units to process received pilot signals through compressed sensing techniques, thereby enabling effective channel estimation. These approaches empower the IRS to enhance the communication links between the IRS and the BS, as well as between the IRS and the users.

The gains for the users' and receiver antennas, denoted as $G_t(O)$ and $G_r(O)$, respectively, can be given as follows:

$$G_{t,\kappa}(O) = 4e_t \frac{J_1\left(\frac{\pi D_t \sin(O)}{\lambda}\right)}{\sin(O)}. \tag{4.9}$$

$$G_r(O) = 4e_r \frac{J_1\left(\frac{\pi D_r \sin(O)}{\lambda}\right)}{\sin(O)}. \tag{4.10}$$

Here, $J_1(.)$ represents the Bessel function of the first kind of order 1, D_t and D_r are the diameter of the transmitter and receiver antennas. The parameter O represents the angle measured from the broadside of the antenna [42], where the highest gain occurs when $O = 0$ and is written as:

$$G_{t,\kappa}(O) = e_t \left(\frac{\pi D_t}{\lambda}\right)^2. \tag{4.11}$$

$$G_{r,\kappa}(O) = e_r \left(\frac{\pi D_r}{\lambda}\right)^2. \tag{4.12}$$

The aperture efficiencies for the transmitter and receiver antennas are represented by e_t and e_r, respectively. The gain of each RU can also be given as [42].

Each RU's power radiation pattern can be represented as follows:

$$G(\psi_{i1,\kappa}) = 4\cos(\psi_{i1,\kappa}), \quad 0 \leq \psi_{i1,\kappa} \leq \frac{\pi}{2}, \tag{4.13}$$

Here, the angle at which the signal from user κ arrives at IRS_1 is represented as $\psi_{i1,\kappa}$. [42].

The cumulative gains and losses between each user and the receiver are expressed as $L_{\tau,\kappa}$. These include free space path losses (FSPL), antenna gains, and THz absorption losses [49], [47]. Mathematically, we have:

$$L_{\tau,\kappa} = L_{FSPL,\tau,\kappa} L_{abs,\kappa} \tag{4.14}$$

The entire THz absorption losses for each transmitter T_{x_κ} are represented by $L_{abs,\kappa}$. These losses are calculated using a simplified model presented in [31] at typical atmospheric conditions. On the other hand, the entire FSPL for each transmitter T_{x_κ} is represented by $L_{FSPL,\kappa}$ and can be expressed as[49], [47]:

$$L_{FSPL,\tau,\kappa} = L_{FSPL,\kappa} L_{FSPL,r}, \tag{4.15}$$

The term $L_{FSPL,\kappa}$ represents the FSPL for the signal reflected from IRS_1 to IRS_2 and is given by:

$$L_{FSPL,\kappa} = \frac{\left(\frac{\lambda}{4\pi}\right)^2 G_{t,\kappa} G_{\psi_{i1,\kappa}} G_{\psi_{r,1}}}{r_{t_\kappa}^2}, \tag{4.16}$$

Similarly, $L_{FSPL,r}$ corresponds to the FSPL for the third hop, linking IRS_2 and the receiver, is formulated as:

$$L_{FSPL,r} = \frac{\left(\frac{\lambda}{4\pi}\right)^4 G_{\psi_{i,2}} G_{\psi_{r,2}} G_r}{r_2^2 r_3^2}. \tag{4.17}$$

The total FSPL for each transmitter T_{x_κ} can be obtained as follows:

$$L_{FSPL,\tau,\kappa} = \left(\frac{\lambda}{4\pi}\right)^6 \frac{G_{t,\kappa} G(\psi_{i1,\kappa}) G(\psi_{r,1}) G(\psi_{i,2}) G(\psi_{r,2}) G_r}{r_{t\kappa}^2 r_2^2 r_3^2}. \tag{4.18}$$

In (4.18), $G_{t,\kappa}$ designates the transmitter T_{x_κ} gain, $G(\psi_{i1,\kappa})$ and $G(\psi_{r,1})$ denote the power radiation patterns of the incident angle $\psi_{i1,\kappa}$ towards IRS_1 and the reflected angle $\psi_{r,1}$ towards IRS_2, respectively. Similarly, $G(\psi_{i,2})$ and $G(\psi_{r,2})$ represent the power radiation patterns of the incident angle $\psi_{i,2}$ towards IRS_2 and the reflected angle $\psi_{r,2}$ towards the receiver, respectively. Finally, G_r denotes the gain of the receiver antenna [49], [47].

4.1.1 Rate of the Desired User in the Presence of Interference

The desired user's rate in the presence of interference is given by (4.19). Here, R_κ, and γ_κ represent the data rate and the SINR of user κ, respectively.

$$R_\kappa = \log_2(1 + \gamma_\kappa), \tag{4.19}$$

The SINR γ_κ is defined as:

$$\gamma_\kappa = \frac{P_{Rx}^{(\kappa)}}{\sum_{i=1\ i\neq\kappa}^{\mathcal{K}} P_{Rx}^{(i)} + \sigma^2}, \tag{4.20}$$

Here, the received power of user κ is represented by $P_{Rx}^{(\kappa)}$, and σ^2 denotes the noise variance at the receiver. The denominator in (4.20) represents the aggregate interference power stemming from all remaining users. The objective of this section is to derive the expression for the received power given by (4.24) and the E2E total rate (4.32). Additionally, we provide a detailed analysis of two objectives: 1) Optimize the received signal power for user 1. 2) Maximize the sum data rate for both users.

However, it is important to note that maximizing the rate, (4.19), for individual users or the sum rate, (4.32), for both users using analytical solutions is mathematically intractable and not feasible. For the first goal, we utilize sub-optimal optimization solutions available in the literature. We address the over-determined system of equations that has more equations than unknowns by considering the assumption of a spatially correlated channel. Mathematical solutions for this system are typically challenging to obtain, and thus, sub-optimal optimization methods are employed. Additionally, we suggest employing DRL techniques, specifically DDPG, to solve the same problem. We assess the performance of the DDPG algorithm in comparison to sub-optimal methods. For the second goal, we use the DDPG scheme to solve a non-convex optimization problem. To assess the DDPG algorithm's performance for this objective, we compare its results with the results obtained using a discretized ES method.

Deriving the Received Power for user κ ($P_{Rx}^{(\kappa)}$)

In our setup, the two users are transmitting at IRS_1, and their signals cover all IRS_1 units from different distances and angles.

The power that is reflected from the e^{th} RU of IRS_1 is represented as follows:

$$P_{r,e}^{(\kappa)} = \left(\frac{\lambda}{4\pi}\right)^2 \frac{G_{t,\kappa} G(\psi_{i1,\kappa}) G(\psi_{r,1})}{r_{t\kappa}^2} |h_{t,\kappa e}|^2 |\phi_e|^2 P_t, \tag{4.21}$$

Here, $\phi_e = \Lambda_e e^{-j\theta_e}$ denotes the exponential of the phase shift of the e^{th} reflecting element in IRS_1, and Λ_e stands for the reflection coefficient of the e^{th} RU in IRS_1. The antenna gain for user κ is designated by $G_{t,\kappa}$, the RU gain from the incident angle is symbolized as $G(\psi_{i1,\kappa})$, the gain of the RU from the reflected angle is represented as $G(\psi_{r,1})$, and $r_{t\kappa}$ signifies the distance separation between user κ and the e^{th} IRS_1's RU. $|h_{t,\kappa e}|^2$ represents the squared magnitude of the channel coefficient between user κ and the eth RU. P_t denotes the transmitted power. Similarly, the power that is reflected from the f^{th} RU of IRS_2's as a result of being reflected by the e^{th} RU of IRS_1 can be expressed as:

$$P_{r,ef}^{(\kappa)} = \left(\frac{\lambda}{4\pi}\right)^4 \frac{G_{t,\kappa}G(\psi_{i1,\kappa})G(\psi_{r,1})G(\psi_{i,2})G(\psi_{r,2})}{r_{tk}^2 r_2^2} |h_{t,\kappa e}|^2 |\phi_e|^2 |H_{ef}|^2 |\phi_n|^2 P_t, \quad (4.22)$$

Here, $\phi_f = \Lambda_f e^{-j\omega_f}$ denotes the exponential of the phase shift of IRS_2 f^{th} reflecting unit, and its reflection coefficient is designated as Λ_f. $G(\psi_{i,2})$ signifies the gain of the RU based on the incident angle $\psi_{i,2}$, and $G(\psi_{r,2})$ denotes the gain of the RU corresponding to the reflected angle. H_{ef} signifies the channel coefficient connecting the e^{th} RU of IRS_1 and the f^{th} RU of IRS_2. Finally, the power received at the receiver across the channel H_{ef} is expressed as:

$$P_{rx,ef}^{(\kappa)} = \left(\frac{\lambda}{4\pi}\right)^2 \frac{P_{r,ef}^{(\kappa)}}{r_3^2} G_r |h_{rf}|^2, \quad (4.23)$$

Here, r_3 designates the distance between IRS_2 and the receiver. G_r represents the antenna gain of the receiver, and $|h_{rf}|^2$ represents the squared magnitude of the channel coefficient between IRS_2's f^{th} RU and the receiver.

The overall received power for user κ can be expressed as:

$$P_{Rx}^{(\kappa)} = \left| \sqrt{L_{\tau,\kappa}} \sum_{e=1}^{E} \sum_{f=1}^{F} |h_{t,\kappa e}||H_{ef}||h_{rf}| \times e^{-j\left(\varphi_{t_{\kappa e}} + \theta_e + \varphi_{ef} + \omega_f + \varphi_{r_f} + \eta_\kappa + \eta_3\right)} \right|^2 P_t, \quad (4.24)$$

Here, $L_{abs,\kappa}$ represents the absorption loss, which is included in $L_{\tau,\kappa}$. The phase terms $\varphi_{t_{\kappa e}}$, φ_{ef}, φ_{r_f} denote the phases for the user channel $\mathbf{h}_{t,\kappa e}$, $\mathbf{H}_{e,f}$ channel, and receiver channel $\mathbf{h}_{rf}$, respectively. η_κ and η_3 represent additional phase terms.

It is essential to notice that the absorption loss term, $L_{\tau,\kappa}$, is included in the expression for the total received power given by (4.24).

Proposition 1. *The overall received power for user κ is represented as follows:*

$$P_{Rx}^{(\kappa)} = \left| \sqrt{L_{\tau,\kappa}} e^{-j\eta_3} \mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{ef}^H \mathbf{\Phi}_F \mathbf{h}_{t,\kappa}^H e^{-j\eta_\kappa} \right|^2 P_t. \quad (4.25)$$

Proof. The proof of Proposition 1 is given in Appendix 6.1. □

By substituting (4.25) into (4.20), the SINR γ_κ for user κ can be obtained as follows:

$$\gamma_\kappa = \frac{\left|\sqrt{L_{\tau,\kappa}} e^{-j\eta_3} \mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{ef}^H \mathbf{\Phi}_E \mathbf{h}_{t,\kappa}^H e^{-j\eta_\kappa}\right|^2 P_t}{\sum_{i=1\ i\neq\kappa}^{\mathcal{K}} \left|\sqrt{L_{\tau,\mathrm{i}}} e^{-j\eta_3} \mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{ef}^H \mathbf{\Phi}_E \mathbf{h}_{t,i}^H e^{-j\eta_i}\right|^2 P_t + \sigma^2}. \tag{4.26}$$

4.1.2 Maximizing Desired User's Data Rate in the Presence of Interference

In this section, our primary focus is on analytically deriving the first goal, which is centered around enhancing the data rate for the intended user while accounting for the second user's presence as an interfering factor. Our aim revolves around identifying the optimal phases of the interconnected IRSs that lead to the highest achievable rate for the desired user. This entails determining the ideal phase settings of the multi-hop IRSs to maximize the data rate received by the intended user. Nevertheless, we face two significant challenges. Firstly, the problem is non-convex, and secondly, obtaining a closed-form solution proves to be mathematically complex. To tackle these obstacles, we propose two sub-optimal methods to maximize the received power for the intended user. Additionally, we introduce a DDPG approach aimed at optimizing the data rate for the intended user.

Desired User's Rate Maximization Problem

To maximize the rate for the desired user while accounting for interference from the other user, the optimization is given by (4.27).

$$\begin{aligned} \max_{\mathbf{\Phi}_F, \mathbf{\Phi}_E} \ & \log_2\left(1+\gamma_\kappa\right), \\ \text{s.t.} \ & |\phi_e|^2 = 1, \forall e \in \{1,2,...,E\}, \\ & |\phi_f|^2 = 1, \forall f \in \{1,2,...,F\}, \end{aligned} \tag{4.27}$$

However, because of the restrictions imposed by the constant modulus on the reflective units of IRS_1 and IRS_2, the problem is non-convex and falls into the category of NP-hard problems [48]. Consequently, obtaining a closed-form analytical expression for the IRSs' ideal phase shifts is mathematically intractable. The ideal approach necessitates striking a balance between enhancing the desired user's received SNR and reducing interference from the second user. These sub-objectives may not necessarily align with each other. As a result, we propose a sub-optimal greedy method by focusing on maximizing the desired user's received power. Although this solution does not guarantee optimal performance according to the rate maximization objective, our simulation results demonstrate that it provides a good compromise between complexity and achieved results. This approach offers a practical alternative to mitigate the challenges posed by the non-convexity of the problem and the complexity of optimizing the rate directly through ES.

Sub-optimal Methods: Maximizing the Desired User's Received Power

Our goal can be recast as finding a solution for (4.28) to maximize the total received power $P_{Rx}^{(\kappa)}$ for the intended user.

$$\theta_e + \omega_f + \varphi_{t_{1_e}} + \varphi_{ef} + \varphi_{r_f} + \eta_1 + \eta_3 = \nu, \quad \forall e, f. \tag{4.28}$$

This system of equations can be understood as finding a set of phase shifts across all paths established by the reflective elements of IRS_1 and IRS_2 that result in a maximum received power. Without loss of generality, we can choose a constant value $\nu = 0$.

Equation (4.28) represents an over-determined set of equations containing $E+F$ unknowns, corresponding to the phase shifts of IRS_1 and IRS_2 RUs, and $E \times F$ equations, corresponding to the different paths formed by the combinations of the reflective elements. The goal is to obtain the phase shifts that satisfy this system of equations and maximize the received power.

We can express (4.28) as:

$$\mathbf{A\Theta} = \mathbf{C} \tag{4.29}$$

where $\mathbf{\Theta}$ is $(E + F) \times 1$ vector representing the phase shifts of IRS_1 and IRS_2 (i.e., $\theta_1, \theta_2, ..., \theta_E, \omega_1, \omega_2, ..., \omega_F$), $\mathbf{A}$ is a binary matrix with dimensions $(E \times F) \times (E+F)$, and $\mathbf{C}$ is a $(E \times F) \times 1$ vector with known constant values including the transmitter channel's ($\mathbf{h}_{t,\kappa}$) phase shifts, $\mathbf{H_{ef}}$ channel's phase shifts, and the receiver channel's ($\mathbf{h}_r$) phase shifts.

By solving this equation, we determine the optimal phase shifts that will result in maximizing the received power. It is important to note that finding an analytical closed-form solution for this problem is very hard due to the over-determined situation.

We provide a detailed explanation of (4.28) below:

$$\underbrace{\begin{bmatrix} 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 & 1 \end{bmatrix}}_{E+F} \times \underbrace{\begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_e \\ \vdots \\ \theta_{E-1} \\ \theta_E \\ \omega_{E+1} \\ \vdots \\ \omega_{E+f} \\ \vdots \\ \omega_{E+F-1} \\ \omega_{E+F} \end{bmatrix}}_{(E+F)\times 1} = \underbrace{\begin{bmatrix} -\varphi_{t_\kappa,1} - \varphi_{11} - \varphi_{r,1} - \eta_{\kappa,1} + \eta_{3,1} \\ \vdots \\ -\varphi_{t_\kappa,E} - \varphi_{E1} - \varphi_{r,1} - \eta_{\kappa,E} + \eta_{3,1} \\ \vdots \\ -\varphi_{t_\kappa,1} - \varphi_{12} - \varphi_{r,2} - \eta_{\kappa,1} + \eta_{3,2} \\ \vdots \\ -\varphi_{t_\kappa,E} - \varphi_{E2} - \varphi_{r,2} - \eta_{\kappa,E} + \eta_{3,1} \\ \vdots \\ -\varphi_{t_\kappa,e} - \varphi_{ef} - \varphi_{r,f} - \eta_{\kappa,e} + \eta_{3,f} \\ \vdots \\ -\varphi_{t_\kappa,1} - \varphi_{1F} - \varphi_{r,F} - \eta_{\kappa,1} + \eta_{3,F} \\ \vdots \\ -\varphi_{t_\kappa,E} - \varphi_{EF} - \varphi_{r,F} - \eta_{\kappa,E} + \eta_{3,F} \end{bmatrix}}_{(E\times F)\times 1}$$

To address the inconsistency of the system described in (4.28), we employ two approximate mathematical techniques: the P_{inv} and the BLS. These techniques allow us to find approximate solutions for the unknown variables θ_e and ω_f, and subsequently calculate the received power for a selected desired user (i.e. user 1). We compare the results obtained from the mathematical methods with those produced a the DRL solution, utilizing a DDPG architecture.

Pseudo-Inverse Solution: The P_{inv} method is employed to solve the over-determined set of equations in (4.28):

$$\mathbf{\Theta} = \mathbf{A}^{+}\mathbf{C}, \tag{4.30}$$

where $\mathbf{A}^{+}$ is the P_{inv} of the matrix $\mathbf{A}$ defined as

$$\mathbf{A}^{+} = (\mathbf{A}^{\intercal}\mathbf{A})^{-1}\mathbf{A}^{\intercal}. \tag{4.31}$$

This equation represents the system of equations where the matrix $\mathbf{A}$ of size $(EF)\times(E+F)$ is multiplied by the matrix $\mathbf{\Theta}$ of size $(E+F)\times 1$ and yields the matrix $\mathbf{C}$ of size $(EF)\times 1$.

Block Solution: By assuming spatially correlated channels, it is possible to develop a low-complexity approximate solution for cases with higher ρ values between adjacent RUs using the exponential correlation model [2]. In scenarios with high

channel correlation, the channel matrix $\mathbf{H_{ef}}$ can be assumed to have a block structure, where elements within each block share the same phase, while there is no interdependence or correlation observed between adjacent blocks.

Since the number of unknowns, $E+F$, representing the total number of elements in IRS_1 and IRS_2, is smaller than the number of equations, $E \times F$, representing the entire elements in the channel matrix between IRS_1 and IRS_2, we need to reduce the number of equations. The key idea of the block solution is to treat each group of channel elements as one block with the same value. By doing so, redundant equations in the channel matrix $\mathbf{H_{e,f}}$ can be eliminated using row reduction methods such as Gaussian Elimination, reducing the rank of $\mathbf{H_{e,f}}$ to the number of blocks $\frac{E\times F}{N_{\text{blk}}}$, where N_{blk} is the number of channel elements within a single block. This reduction in the number of rows within the channel matrix will also be reflected in (4.28), which will no longer be over-determined since the number of equations is reduced. It is important to mention that the count of independent elements in IRS_1 and IRS_2 remains the same, but the number of equations is reduced from $E \times F$ to less than $E + F$.

Using this approach, the over-determined set of equations becomes solvable if the number of blocks $\frac{E\times F}{N_{\text{blk}}} \leq E + F$. Thus, we can employ the block solution to reformulate the overall received power for user 1 in (4.24).

Algorithm 3 Block solution (BLS) Based Scheme

Input: E, F, h_t, $H_{e,f}$, h_r, N_{blk}
Output: P_{Rx_κ}, θ_e, ω_f
if $E + F \leq E \times F$ **then**
    Separate the channel $\mathbf{H_{e,f}}$ elements into blocks.
    if $E + F \geq \frac{E\times F}{N_{\text{blk}}}$ **then**
        Compute the overall received power by applying equation (4.24), considering each set of channel elements as a single block with identical values.
    else
        if $E + F < \frac{E\times F}{N_{\text{blk}}}$ **then**
            Obtain the solution to (4.28) using the P_{inv} method and compute the overall received power using (4.24).
        endif
    endif
endif

Since the problem of maximizing the rate of one of the users under interference from the other is a sub-problem from the sum rate maximization problem, we directly move to the latter and address the former within after establishing DRL setup.

4.1.3 Maximizing the Total Rate for Two Users

The objective in the second scenario is to obtain the ideal phase shifts for IRS_1 and IRS_2. This optimization aims to maximize the overall data transmission rate for both users. The sum-rate equation is expressed as follows:

$$R_{sum} = \sum_{\kappa=1}^{\mathcal{K}} \log_2 (1 + \gamma_\kappa) . \tag{4.32}$$

The challenge of determining the optimal phase shift matrices $\mathbf{\Phi}_F$ and $\mathbf{\Phi}_E$ that maximize the total rate, denoted as R_{sum}, is formulated as follows:

$$\begin{aligned} \max_{\mathbf{\Phi}_F, \mathbf{\Phi}_E} & \sum_{\kappa=1}^{\mathcal{K}} \log_2 (1 + \gamma_\kappa) , \\ \text{s.t.} \quad & |\phi_e|^2 = 1, \quad \forall e \in 1,2,...,E, \\ & |\phi_f|^2 = 1, \quad \forall f \in 1,2,...,F, \end{aligned} \tag{4.33}$$

The problem formulated in (4.33) is a problem with an NP-hardness that has a non-tractable solution owing to the non-convex nature arising from the constraints imposed by the constant modulus properties of the reflecting elements in IRS_1 and IRS_2 [48]. Analytically solving this problem is very hard, and traditional mathematical methods are not applicable. To address this challenge, we employ a DRL technique, specifically the DDPG algorithm, as an alternative approach to solve this complex problem. Additionally, we consider two limiting cases: an upper bound case where there is no interference and full channel phase compensation can be achieved, and a lower bound case where the phases of the IRS elements are randomly chosen.

4.1.4 Upper bound on Performance

Given the condition of zero interference and the ability of IRSs to completely nullify phase shifts among various paths of reflection, we can approximate an upper limit for the total transmission rate based on the equation (4.26) representing the SINR. Consequently, the SINR for user κ is represented as below:

$$\gamma_\kappa^U = \frac{\left| \sqrt{L_{\tau,\kappa}} |\mathbf{h}_r^H| |\mathbf{H}_{e,f}^H| |\mathbf{h}_{t,\kappa}^H| \right|^2 P_t}{\sigma^2}, \tag{4.34}$$

The upper limit on the total rate is obtained as a result of this condition, which is formulated as:

$$R_{sum}^U = \sum_{\kappa=1}^{\mathcal{K}} \log_2 \left(1 + \gamma_\kappa^U\right) . \tag{4.35}$$

4.2 Controlling Multi-hop IRS Phases utilizing DDPG

As previously stated, the optimization challenges presented in (4.27) and (4.33) are characterized as non-convex, and employing ES to identify the optimal phases that achieve the maximum rate is computationally not feasible. As a result, we develop a DDPG solution to determine the ideal phase shifts of the multi-hop IRS.

In this part, we explain the framework of the suggested DDPG method as shown in Fig. 2.3 to address the problem of optimization for the multi-hop IRS system in (4.27) and (4.33). Deep Q-Networks are unsuitable for this scenario due to their limitation in handling only discrete temporal spaces. Furthermore, in the realm of wireless communication, the convergence of the PG method is insufficient [14]. DDPG represents a model-free RL method that amalgamates the advantages of Q-networks and the PG algorithm, effectively mitigating the shortcomings inherent in both techniques. This method leverages both continuous state and action spaces, as mentioned in [14]. Thus, knowing the policy leads to the resolution of the optimization problem in (4.27) and (4.33).

The DDPG scheme is made up of various essential parts, including the agent, the action designated as $\mathbf{a}^{(\tau)}$, the state denoted as $\mathbf{s}^{(\tau)}$, the reward symbolized as $\mathbf{r}^{(\tau)}$, the Q-value function represented as $Q(\mathbf{s}, \mathbf{a}|o^Q)$, and the policy function indicated as μ. The IRS_1 and IRS_2 are the agents in our system. The states denoted as $\mathbf{s}^{(\tau)}$ encompass the SINR for user 1 at the receiver, the SINR received for user 2 at the receiver, and the total rate for all users at the previous time step $(\tau - 1)$. The actions $\mathbf{a}^{(\tau)}$ refer to the phases of IRS_1 and IRS_2, while the reward $\mathbf{r}^{(\tau)}$ corresponds to the received power for user 1, serving as our first objective, and the collective rate of all the users, serving as our second objective. To ensure stability, the DDPG approach consists of the following NNs: the critic-network, the target critic-network, the actor-network, and the target actor-network.

4.2.1 System Mapping to DDPG

The first step in addressing the system model's optimization challenge is to transform it into the essential components of the DDPG scheme. These components include the action-space, the state-space, and the reward function. In the subsequent paragraphs, we will elaborate on the details of this mapping process and offer an overview of the DDPG algorithm's overall behavior. To begin, the system model needs to be represented in a way that can be interpreted by the DDPG algorithm. This involves defining the state space, which encapsulates the relevant variables and information about the system at a given time. Additionally, the action-space needs to be defined to encompass the available actions that can be taken by the algorithm to interact with the system.

In terms of the general behavior of the DDPG algorithm, it follows a process that involves exploration and exploitation. Throughout the exploration phase, the algorithm explores different actions and their outcomes to gather information about the system. This exploration helps in learning the underlying dynamics and uncertainties of the system. Once the exploration phase is complete, the exploitation phase begins, where the algorithm leverages the knowledge gained to make informed decisions and select actions that maximize the expected rewards.

State space

The state space of the DDPG agent at time step τ is delineated as follows:
In pursuit of the first aim: (Maximization of the user rate under interference)

$$\mathbf{s}^{(\tau)} = [\gamma_1^{(\tau-1)}, \gamma_2^{(\tau-1)}, P_{Rx_\kappa}^{(\tau-1)}], \tag{4.36}$$

To fulfill the second objective: (Maximization of the sum rate)

$$\mathbf{s}^{(\tau)} = [\gamma_1^{(\tau-1)}, \gamma_2^{(\tau-1)}, R_{sum}^{(\tau-1)}], \tag{4.37}$$

Here, $\gamma_1^{(\tau-1)}$ and $\gamma_2^{(\tau-1)}$ represent the received SINR for user 1 and user 2 respectively, at receiver R_x. $P_{Rx_\kappa}^{(\tau-1)}$ corresponds to user x's received power at receiver. $R_{sum}^{(\tau-1)}$ represents the total rate of the users at timestep $\tau - 1$.

By considering these state representations, the DDPG agent can effectively capture and utilize the relevant information needed to make informed decisions and optimize the system's performance.

Action-space

The actions in the DDPG algorithm correspond to the phase shift values of IRS_1 and IRS_2. The outcome is an array that encompasses the phase shift for each IRS unit. Therefore, the action-space can be described using the below policy function:

$$\mathbf{a}^{(\tau)} = \mu(\mathbf{s}^{(\tau)}|o^\mu) + \mathbf{n}^{(\tau)} \tag{4.38}$$

Here, μ represents the policy function, whereas o^μ denotes the parameters of the NN. $\mathbf{n}^{(\tau)}$ refers to the action noise, that is produced using the OU noise [56].

Since the action-space is continuous, it is explored by incorporating noise produced by the OU process. This process draws noise from a correlated normal distribution and samples it, allowing for exploration while maintaining some correlation between consecutive actions. The DDPG algorithm employs a policy function to determine the actions, which are the IRS_1 and IRS_2 phase shift values.

Reward function

The reward function is crucial in guiding the DDPG algorithm towards the desired behavior. It evaluates the outcome of each action taken by the algorithm and provides a numerical reward or penalty based on predefined criteria. The aim is to shape the behavior of the algorithm to maximize the cumulative rewards over time.

The reward function for the first aim is determined by the intended user's highest obtained power:

$$r^{(\tau)} = P_{Rx_\kappa}^{(\tau)}. \tag{4.39}$$

Here, $P_{Rx_\kappa}^{(\tau)}$ represents the highest received power at timestep τ. The reward is directly proportional to the received power, indicating that the agent is incentivized to enhance the received power for the desired user.

For the second aim, the reward function is the highest total rate across both users:

$$r^{(\tau)} = R_{sum}^{(\tau)}. \tag{4.40}$$

In this case, $R_{sum}^{(\tau)}$ represents the total rate across both users at timestep τ. The reward relies on the total rate, which motivates the agent to optimize the overall rate achieved by the users.

4.2.2 DDPG Scheme Operation

The main purpose of the DDPG scheme, as described in Algorithm 1, is to train the agents IRS_1 and IRS_2 to make decisions or carry out actions that achieve maximum long-term mean reward. In this case, the reward is defined based on the received power of user 1 and the sum-rate of both users. The focus is on maintaining a high average reward over time, rather than responding instantaneously to changes in the environment. In each iteration of the algorithm, the agents IRS_1 and IRS_2 observe the current state, that includes the SINRs $\gamma_1^{(\tau-1)}$, and $\gamma_2^{(\tau-1)}$ at state $(\tau - 1)$. They also observe the reward obtained in the previous state. Based on this information, the agents calculate the actions $\mathbf{\Phi}_E$ and $\mathbf{\Phi}_F$ that maximize the cumulative or extended-term reward. This action selection process is performed by the actor-network, which maps the observed state to the optimal action.

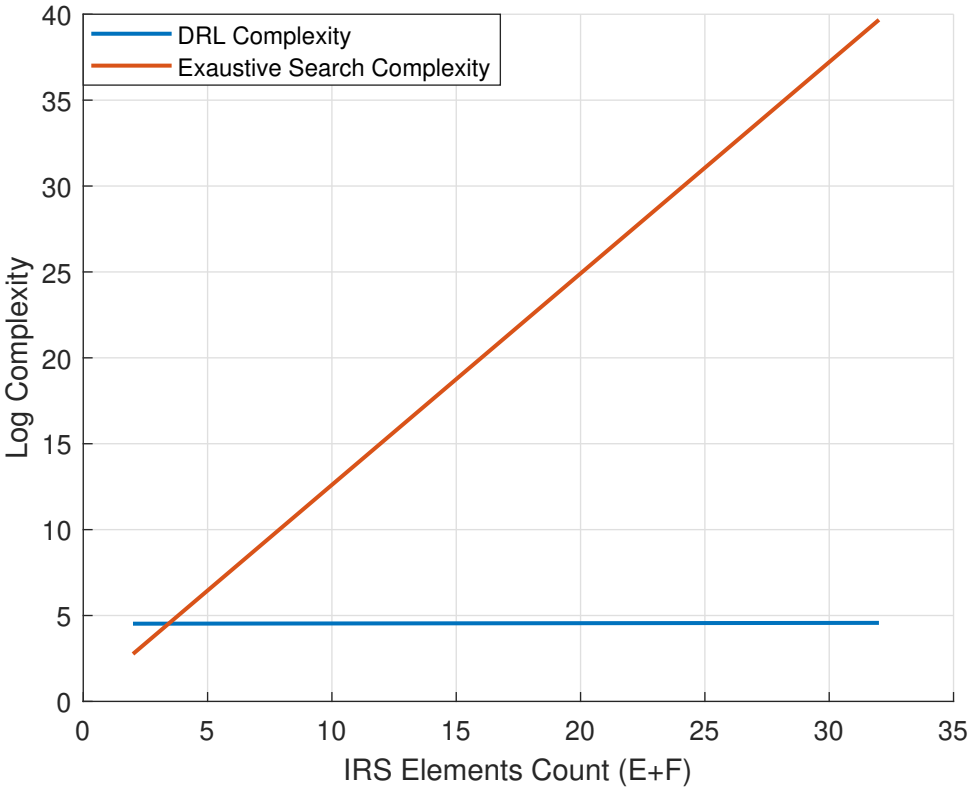
The critic-network works in tandem with the actor-network. It accepts the action and observed state as inputs and estimates the expected reward. In this case, the expected reward corresponds to the received power of the primary user and the combined rate of both users. The critic-network helps in evaluating the quality of the chosen actions and guides the learning process. After calculating the reward, the agents observe the new state and adjust the phases of IRS_1 and IRS_2

accordingly. This procedure is iteratively repeated until the system learns to attain the optimum reward by adapting the actions and phases.

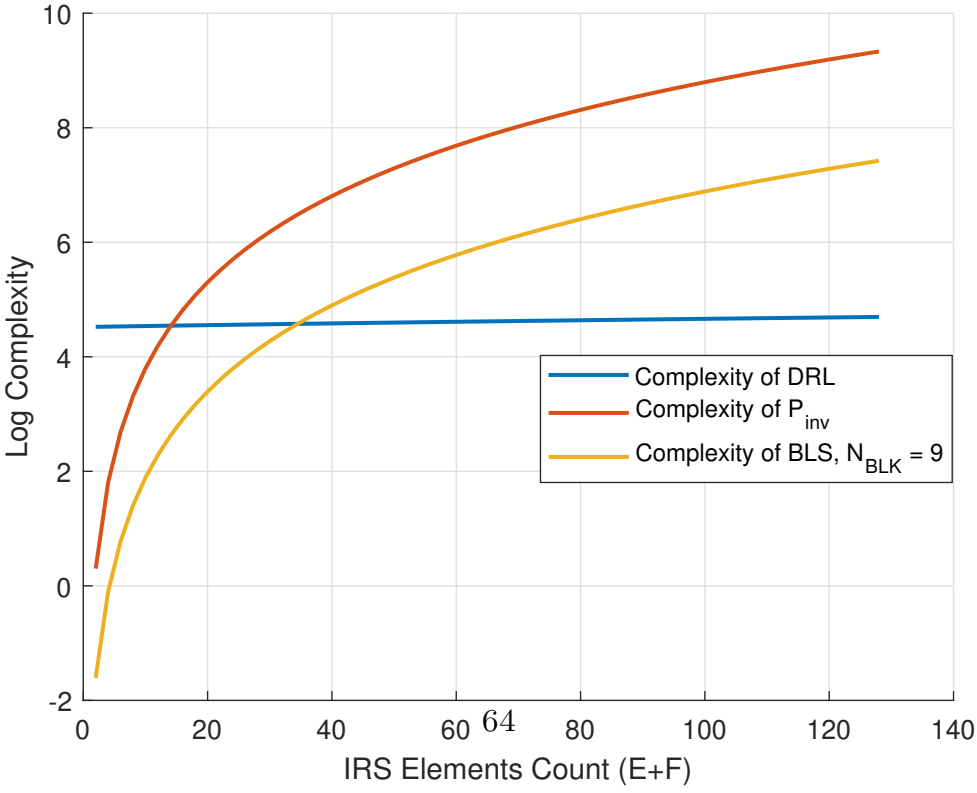
To enhance stability during the learning process, the target networks are periodically updated based on the latest actor and critic parameter values. This helps in reducing the effects of parameter fluctuations and improves the convergence of the algorithm. In essence, the DDPG algorithm strives to train the agents to make choices that optimize the average reward over the long term, taking into account the previous states, actions, rewards, and observations of the environment. By iteratively adjusting the actions and phases based on the feedback received, the agents learn to optimize the system performance over time.

4.3 Complexity Analysis

To demonstrate the DRL algorithm's advantages, we compare its complexity to that of the P_{inv} solution, the BLS method, and the ES scheme.



(a) Comparing complexity: DRL vs. ES.



(b) Comparing complexity: DRL vs. P_{inv} vs. BLS.

Figure 4.2: Comparing complexity:DRL vs. P_{inv} vs. BLS vs. ES. E=F=18, NBLK = 9.

The DRL algorithm's complexity is explained in section 3.4 as follows:

$$C_{DRL} = \mathcal{S} \times \mathcal{I} + \mathcal{H} \times \mathcal{I} \times \mathcal{O} + \mathcal{O} \times \mathcal{A} + \mathcal{A} \tag{4.41}$$

In contrast, the P_{inv} solution of a matrix A with dimensions $EF \times (E+F)$ is computed using its singular value decomposition (SVD), a useful computational technique for dimensionality reduction in over-determined systems, with a complexity of $O((EF)^2 \times (E+F))$, where $EF > (E+F)$, and E and F represent IRS_1 and IRS_2 number of elements, respectively [30, 57]. On the other hand, for the BLS method, the complexity of inverting a matrix A with dimension of $\frac{EF}{N\text{blk}} \times (E+F)$ is reduced to $O((\frac{EF}{N_{\text{blk}}})^2 \times (E+F)$, where $\frac{E \times F}{N_{\text{blk}}} \leq (E+F)$.

Furthermore, the complexity of the ES scheme (C_{ES}) can be inferred from Algorithm 4 and is expressed as:

$$C_{ES} = O\left(\mathcal{K} \times \left(\lfloor \frac{2\pi}{\Delta\Phi} \rfloor + 1\right)^{(E+F)}\right), \tag{4.42}$$

where $\mathcal{K}$ designates the number of users and $\lfloor \frac{2\pi}{\Delta\Phi} \rfloor$ denotes the number of phase search steps.

Consequently, the DRL algorithm's shows lower complexity than that of the P_{inv}, BLS, and ES techniques as the IRS elements count increases. This relationship is depicted in figures: Fig.4.2a and Fig.4.2b.

4.4 Numerical Results

We examine the proficiency of the proposed DDPG-based cascaded IRS-assisted wireless THz system in this section. To evaluate the efficiency of the DDPG scheme, we compare it to benchmark schemes in two scenarios: maximizing the rate for the intended user (user 1) and maximizing the total transmission rate for two users.

We present two evaluation schemes as benchmark models for our communication system in the rate maximization scenario. The first scheme is based on the P_{inv} method, while the second scheme uses the BLS approach. Both schemes consider IRS_1 with $E = 18$ reflecting elements and IRS_2 with $F = 18$ reflecting elements.

In the sum rate maximization scenario, we compare the total rates achieved through the utilization of the DDPG approach with those resulting from employing a discretized ES approximation as an approximate benchmark. To minimize the complexity of the ES scheme, we limit the count of reflecting elements to $E = 4$ for IRS_1 and $F = 4$ for IRS_2. We examine phase shifts ranging from 0 to 2π with an incremental step of $2\pi/72$ for each IRS element, resulting in $(72+1)^{E+F}$ possible combinations of matrices for phase shifts. The choice of search steps significantly influences the computational complexity and level of detail in exploring the angle

Algorithm 4 Exhaustive Search Scheme

Initialize $E = 4$, $F = 4$, $\Delta\theta = \frac{2\pi}{72}$, $\Delta\omega = \frac{2\pi}{72}$,
for $\theta_1 = 0 : \frac{2\pi}{72} : 2\pi$; **do**
    for $\theta_2 = 0 : \frac{2\pi}{72} : 2\pi$; **do**
        for $\theta_3 = 0 : \frac{2\pi}{72} : 2\pi$; **do**
            for $\theta_4 = 0 : \frac{2\pi}{72} : 2\pi$; **do**
                for $\omega_1 = 0 : \frac{2\pi}{72} : 2\pi$; **do**
                    for $\omega_2 = 0 : \frac{2\pi}{72} : 2\pi$; **do**
                        for $\omega_3 = 0 : \frac{2\pi}{72} : 2\pi$; **do**
                            for $\omega_4 = 0 : \frac{2\pi}{72} : 2\pi$; **do**
                                Compute and save $R_{sum}(\theta_1, \theta_2, \theta_3, \theta_4, \omega_1, \omega_2, \omega_3, \omega_4)$
                            end for
                        end for
                    end for
                end for
            end for
        end for
    end for
end for
Find $\mathbf{\Theta}^* = \text{argmax}_{\theta_1,\theta_2,\theta_3,\theta_4,\omega_1,\omega_2,\omega_3,\omega_4} R_{sum}$

Table 4.1: Simulation Parameters

Simulation Parameters	**Values**
Users' count ($\mathcal{K}$)	2
Users' antenna count (N_t)	1
Receiver's antenna count (N_r)	1
Speed of wireless waves free space (c)	3×10^8
Transmission Frequency (f)	300×10^9
Wave Length λ	1×10^{-3}
IRS_1 Reflecting Elements Count (E)	18
IRS_2 Reflecting Elements Count (F)	18
IRS_1 Coordinates (x_{r1}, y_{r1}, h_{r1})	(5,10,12)
IRS_2 Coordinates (x_{r2}, y_{r2}, h_{r2})	(10,10,12)
r_{t_k}	3 to 15
r_2	15
Reflection Coefficient of IRS_1 and IRS_2 (Λ)	1
Antenna Diameter (D_t)	0.12
R_x Coordinates (x_{rx}, y_{rx}, h_r)	(20,0,5)
Channel Width	2×10^9 MHz
Noise PSD (N_{PSD})	-174 dB/Hz
R_x Noise figure (F_{dB})	10
Average noise power represented in dB as ($N0$)	-174 dB/Hz
Noise power expressed in a linear scale (no)	7.9621×10^{-11}
Path attenuation exponent (user κ to IRS_1)	2
Path attenuation exponent (IRS_2 to receiver R_x)	2
Rician factors (K_1 and K_2)	10
F_C	3×10^{-4}
F_A	1×10^{-4}
Discount factor (Γ)	0.99
Soft updates weight(υ)	1×10^{-3}
Batch Quantity (B)	128
Episodes Count	10000
Capacity of the Replay Memory (C)	10^5

space during an exhaustive search. A search step of $2\pi/30$, with a smaller denominator used in chapter 3, reduces the computational complexity by employing a smaller search space, enabling a less detailed exploration. Conversely, a step of $2\pi/72$, with a larger denominator, strikes a balance between computational complexity and detail, providing a moderate resolution for angle exploration. This step is more complex than the finer resolution used in chapter 3, allowing for a more detailed exploration of angles during the exhaustive search. Once the optimal matrices for phase shifts are obtained through the DDPG algorithm, the corresponding sum-rates for the users are calculated accordingly.

In addition to the benchmark schemes, we also compare sum-rates produced by the DDPG algorithm with those computed according to the random phase shift generation as another lower benchmark scheme to our system.

Table 4.1 shows the default simulation settings used in the DDPG-based multi-hop IRS method. The total rates are determined through numerical calculations involving 10^3 Monte Carlo iterations.

The suggested DDPG method consists of both critic and actor-networks, each

of which is implemented as DNNs having 4 layers. Every layer is represented by a linear module with defined input and output sizes. The actor-network takes states as input and generates actions as output, having a total of 36 neurons. It consists of two hidden layers with 128 neurons each, preceded by the rectified linear activation unit. To ensure an appropriate gradient, the output layer of the actor-network employs the tanh() function. The actions and states are merged to form the input that is fed into the critic-network. It also has two hidden layers of 128 neurons apiece, which are succeeded by the ReLU activation unit. The critic-network's output layer represents the Q-value having 36 neurons. The Adam optimizer is employed for updating parameters in both the actor and critic-networks. The simulation outcomes are an average of 1000 iterations. The actor-network's learning step size, indicated as $\digamma_A$, is equal to 3×10^{-4}, while the critic-network's learning step size, denoted as $\digamma_C$, is equal to 1×10^{-4}. The discount factor Γ is set to 0.99, the batch size is set to 128, the replay buffer C is of size 10^5, and the total episodes are set to 10,000.

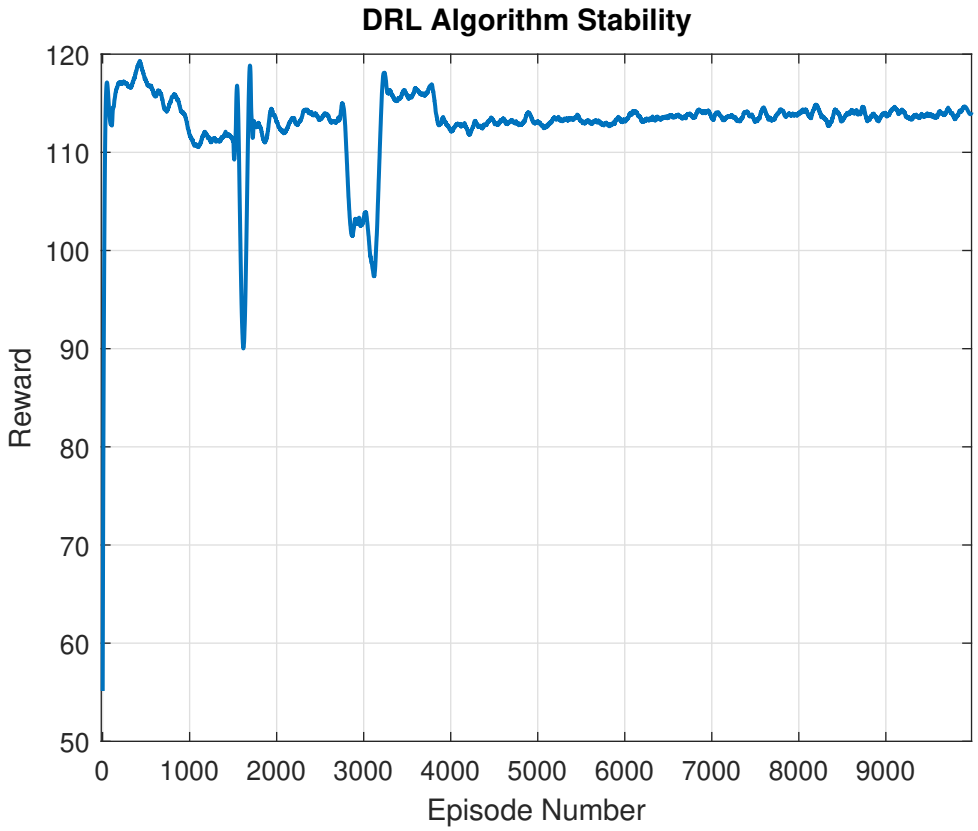


Figure 4.3: Convergence analysis of the DDPG algorithm

The obtained results, as depicted in Fig. 4.3, demonstrate the convergence of the DRL scheme. The plot illustrates the rewards plotted against the episodes, and

it is evident that the rewards consistently increase over time until reaching a steady state value. This indicates the successful learning process of the algorithm.

4.4.1 Maximizing Desired User's Data Rate in the Presence of Interference

In the forthcoming numerical simulations, we present the outcomes of maximizing the rate for the intended user at the receiver by plotting the rate of the first user against the ratio of distances between the two users. We consider several methods, including DDPG, BLS, and P_{inv} techniques, to compare their performances.

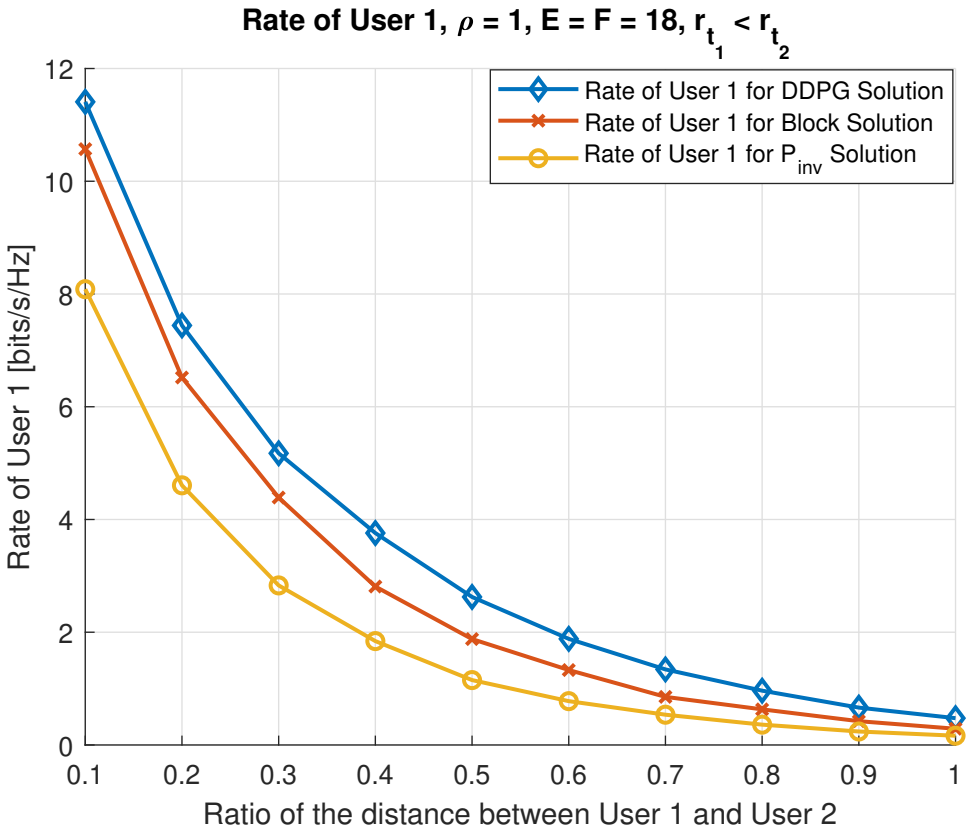


Figure 4.4: Data rate of user 1 as a function of distance ratio for $\rho = 1.0$.

In Fig. 4.4, the rate of user 1 is depicted for the DDPG scheme as a function of the distance ratio ranging from 0.2 to 1. It is shown that as the ratio of the distance between the two users increases, the rate of the first user decreases. This is because the second user becomes closer to the first user, resulting in increased interference from user 2 to user 1. Furthermore, Fig. 4.4 illustrates the rate of user 1 for the DDPG scheme compared to the P_{inv} and BLS methods, assuming $\rho = 1.0$.

It is shown that the rate achieved by the DDPG approach surpasses that of the P_{inv} and BLS methods. This highlights the effectiveness and superiority of the DDPG algorithm in maximizing the rate for user 1 in comparison to the P_{inv} and BLS methods.

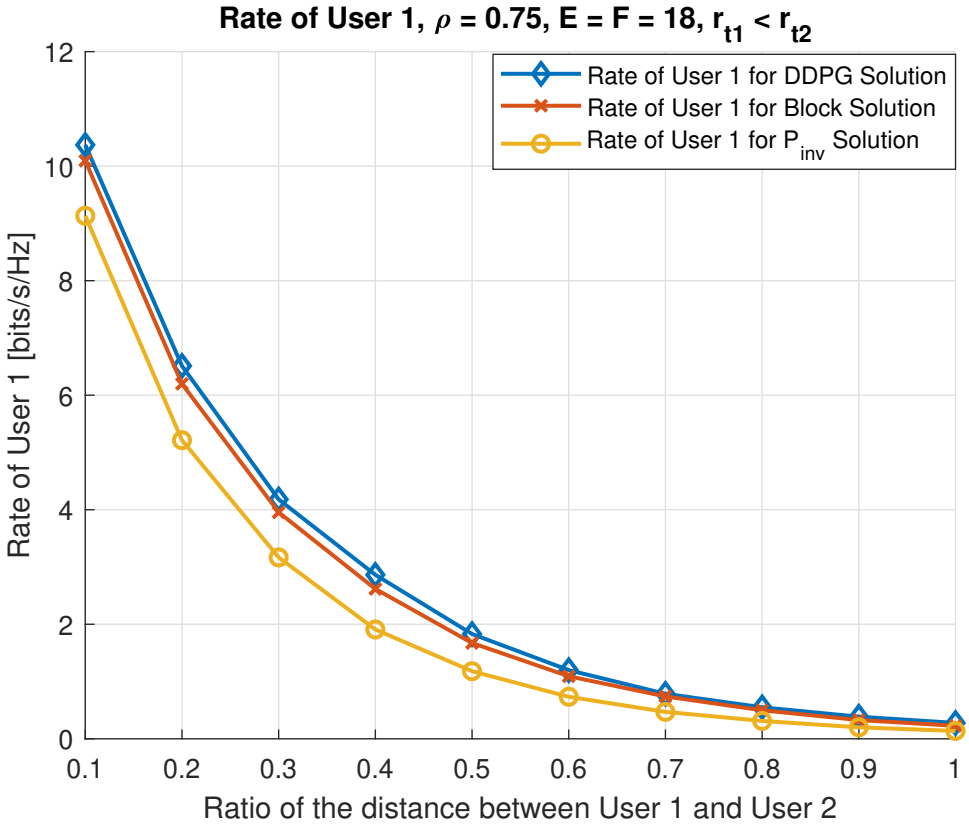


Figure 4.5: Data rate of User 1 as a function of distance ratio with ρ value equal to 0.75.

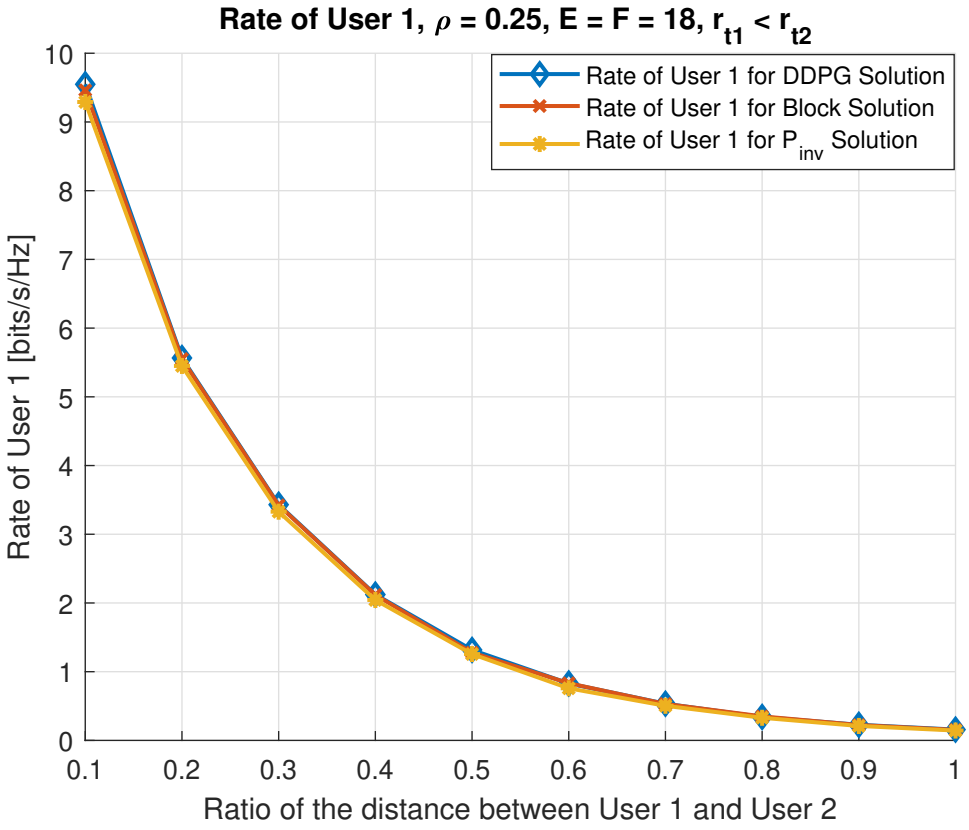


Figure 4.6: Data rate of User 1 as a function of distance ratio with for $\rho = 0.25$.

In the following figures Fig.4.4, Fig.4.5, and Fig. 4.6, the rate of user 1 is plotted for the DDPG algorithm, BLS, and P_{inv} solution as a function of the distance ratio, considering different ρ values. It is shown that as ρ decreases, the rate of user 1 decreases for all three methods. This is because a lower ρ implies less correlation between the channels, leading to increased interference and a reduced rate for user 1. Furthermore, as ρ increases, the data rates produced by the DDPG scheme exhibit higher values compared to the BLS and P_{inv} methods. This is because increasing ρ enhances the DDPG's learning efficiency, allowing it to exploit the benefits of correlated channels and achieve higher data rates. However, It's crucial to emphasize that when ρ values are low, the difference in data rates between the DDPG scheme and the other methods diminishes.

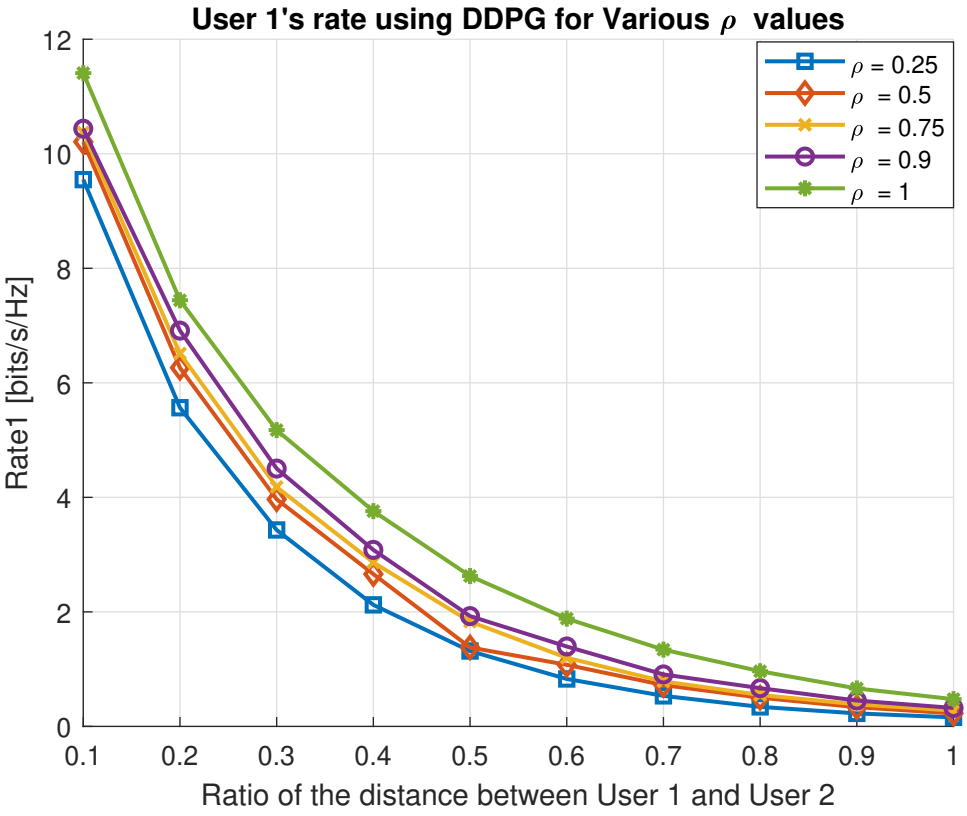


Figure 4.7: Data rate of user 1 as a function of distance ratio for different ρ values. E=F=18.

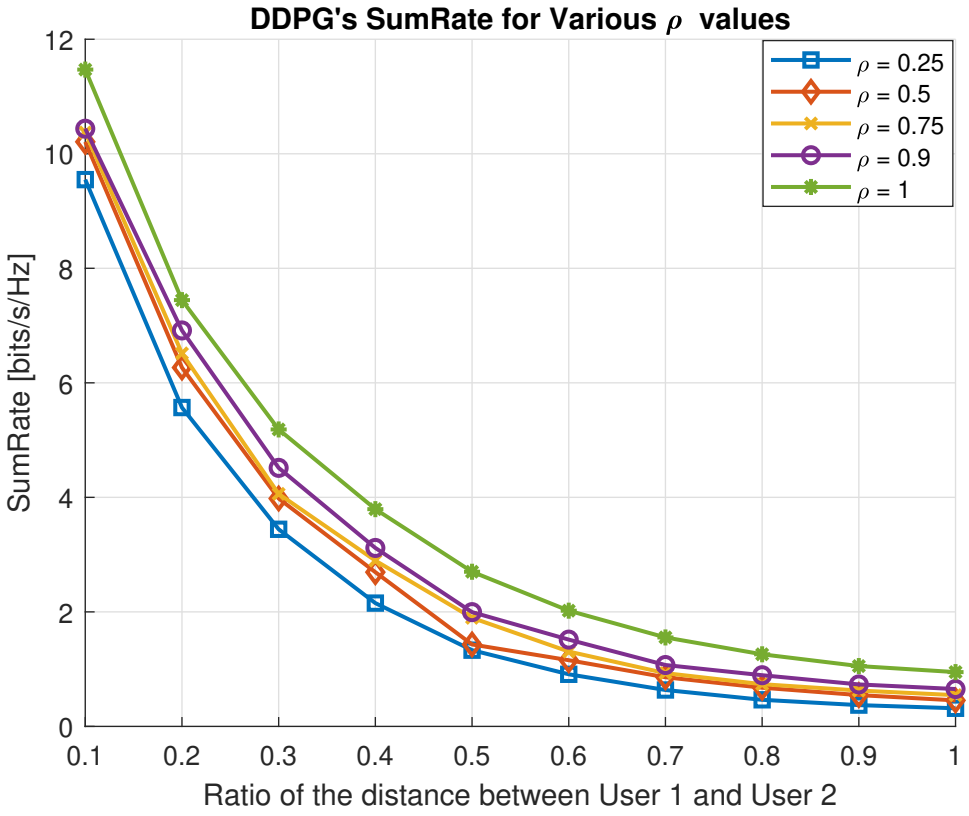
In Fig. 4.7, the rates generated by the DDPG algorithm are plotted against the distance ratio for different ρ values. It is observed that as ρ increases, the data rates for the DDPG scheme also increase. This is because a higher ρ indicates a stronger relationship between the channels, allowing the DDPG algorithm to leverage this correlation and improve its learning efficiency. As a result, the DDPG approach attains superior data rates compared to other techniques, particularly for high values of ρ. This demonstrates the importance of correlated channels in the DDPG scheme and their beneficial impact on data rates. By exploiting the correlations, the DDPG algorithm can more efficiently optimize the phase shifts of the cascaded IRSs, leading to improved performance in data rates.

4.4.2 Maximizing the Aggregated Rate for Both Users

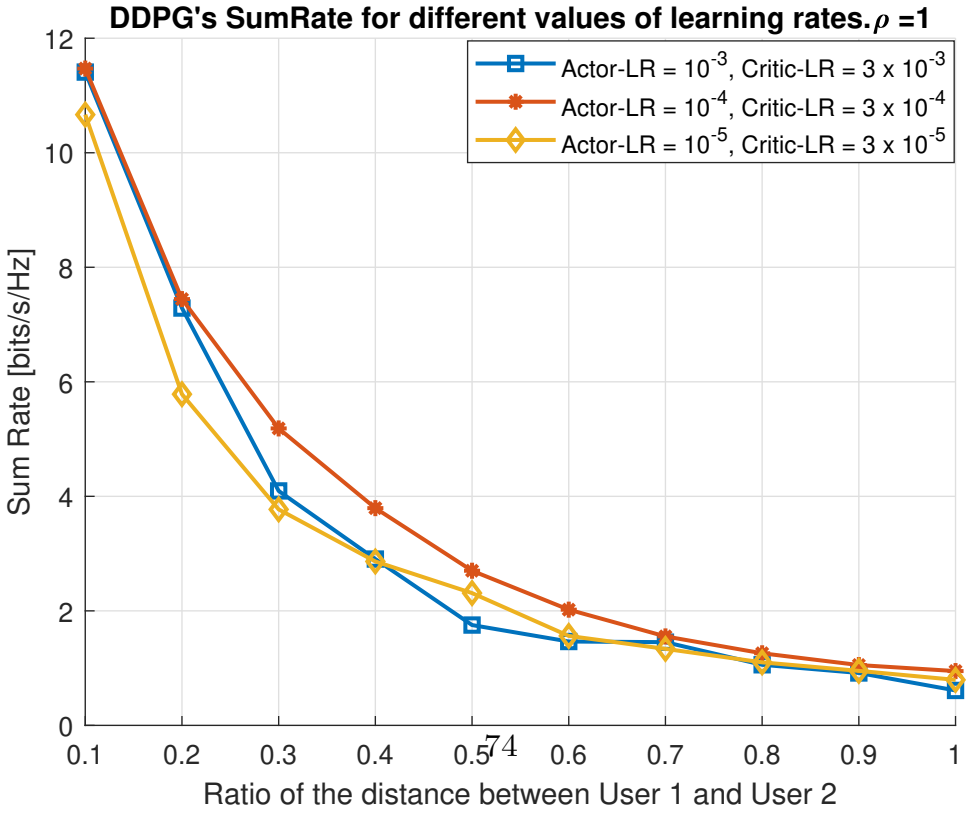
In the subsequent simulations, we demonstrate the outcomes of employing the DDPG method to optimize the total rate for both users. We graphically illustrate the sum-rate as a function of the ratio of distances between user 1 and user 2 for

various ρ values, as revealed in Fig. 4.8a. From the results, it is evident that the total rate achieved through the DDPG approach increases as the ρ value increases. This indicates that the DDPG algorithm leverages the higher correlation between channels to tune the phase shift values of the multi-hop IRSs and achieve higher overall data rates for the two users.

In the below numerical simulations, we employed fixed learning rates for the DDPG algorithm. Specifically, the actor-network's learning rate (F_A) equals 10^{-4}, and the critic-network's learning rate (F_C) equals 3×10^{-4}. The impact of different learning rates on DDPG data rates is illustrated in Fig. 4.8b, where we compare learning rates of 10^{-3}, 10^{-4}, and 10^{-5} for the actor-network, and 3×10^{-3}, 3×10^{-4}, and 3×10^{-5} for the critic-network. Results show that the peak DDPG data rate is attained when the actor-network's learning rate is set to 10^{-4} and the critic-network's learning rate is set to 3×10^{-4}. This demonstrates the importance of selecting appropriate learning rates for optimal performance. Learning rates that are too small (10^{-5}) or too large (10^{-3}) result in lower average rewards, while the learning rate of 10^{-4} achieves better rewards and higher data rates.



(a) DDPG total transmission rate vs ratio of distances for different ρs. E=F=18.



(b) DDPG total transmission rate vs ratio of distances for $\rho = 1.0$. E=F=18.

Figure 4.8: DDPG total transmission rate vs ratio of distances for different ρ and learning rate values.

To evaluate the proficiency of our DDPG scheme, we conduct a comparison between the total rates achieved by the DDPG scheme and the upper bound on the system performance, as depicted in Fig. 4.9. The findings reveal that the sum-rates produced by the DDPG closely approximate the upper bound. Although the upper bound sum-rate calculation does not consider interference between the two users, the difference between the upper bound sum-rates and those obtained from the DDPG scheme decreases with the distance ratio. This is due to the decrease in received power P_{Rx}^{k} as user 1 moves farther away from IRS_1 and closer to user 2. While there is no interference between the users in the upper bound calculation, the decrease in received power diminishes the disparity between the upper bound sum-rates and those achieved by the DDPG scheme. These results confirm that the DDPG scheme performs adequately well, with sum-rates closely approximating the upper bound. The decreasing difference between the upper bound and DDPG sum-rates highlights the impact of the received power and the proximity of the users on system performance, even in the absence of direct interference.

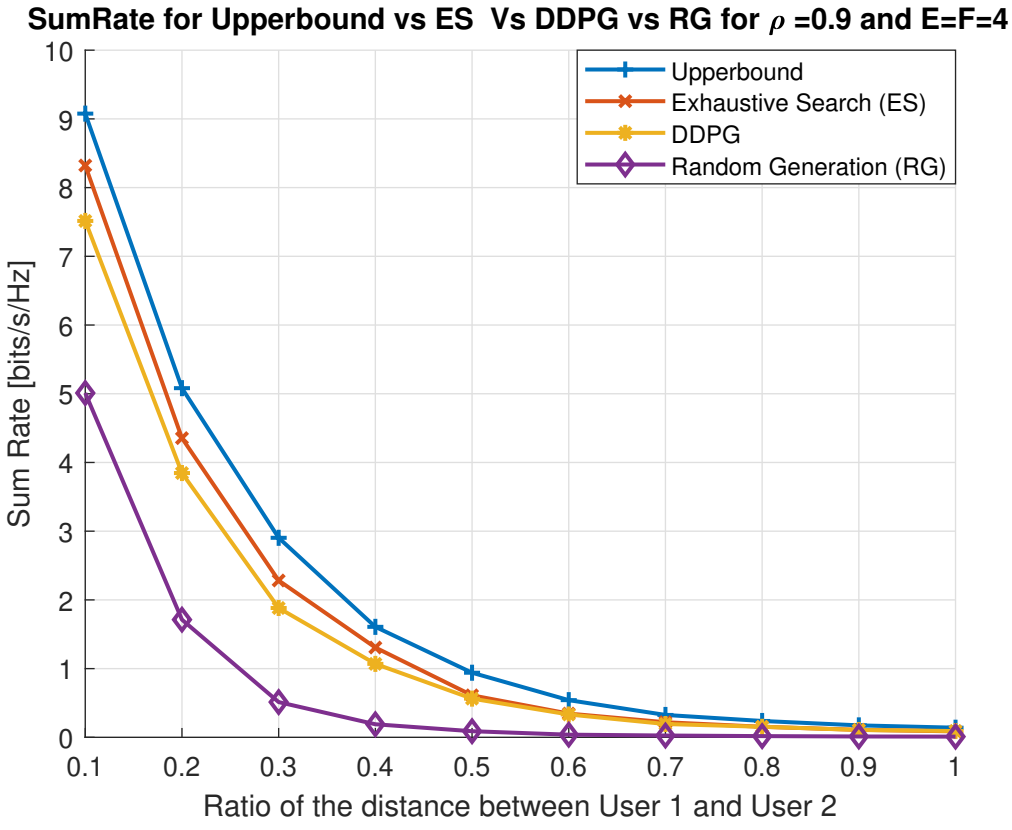


Figure 4.9: Comparing DDPG total transmission rate, upper bound, ES vs distance ratio. $\rho = 0.9$. E=F=4.

To further validate the effectiveness of the DDPG algorithm, We contrast the

total rates obtained from the DDPG scheme with those achieved using the discretized ES method. The ES algorithm is utilized to compute the highest sum-rate by finding the matrix of the optimal phase shifts. The total rate is computed for the two users based on 1000 Monte Carlo iterations. Numerical results depicted in Fig. 4.9 demonstrate that the DDPG algorithm's sum-rates closely match those obtained from the ES algorithm with the specified granularity. This confirms the effectiveness of our DDPG scheme in achieving high-performance results that are comparable to the optimal solution obtained from the ES algorithm, even with a reduced number of reflecting elements.

Chapter 5

A DRL-Based Approach for Secure Cross-layer Design of IRS NOMA-SWIPT Systems in IoT Environments

In this chapter, the focus is on the third main contribution of the thesis, which involves developing a DRL approach, specifically DDPG, to enhance the security of an indoor IoT scenario. The objective is to achieve a secure cross-layer design for IRS NOMA-SWIPT systems. Compared to existing literature, the main emphasis of this chapter is on employing DDPG to find a solution for the secrecy sum-rate maximization problem, considering the cross-layer design aspect. The primary goal is to increase the secure sum-rate for IoTDs by determining the optimal phases for the IRS, power allocation factors, and EH factors. These optimizations are subject to various constraints such as the IRS unit modulus, EH factors, NOMA dynamic power allocation factors, and packet loss factors.

To tackle this optimization problem, our work utilizes the DDPG approach, which proves effective in finding viable solutions despite the non-convex nature of the problem caused by the constraints of the IRS unit modulus. By employing DDPG, the work in this chapter aims to enhance the security and performance of the IoT system by optimizing various parameters across different layers. The use of DRL enables intelligent decision-making and adaptation based on specific system conditions and constraints.

5.1 System Model

An IRS-assisted NOMA-SWIPT scheme is represented in Fig. 5.1, where the access point (AP) with directional antennas is transmitting to two legitimate IoTDs

using a non-cooperative downlink NOMA transmission scheme. The system also includes a passive eavesdropper, denoted as e, who attempts to intercept the transmitted signal. The signal from the AP is transmitted to the IRS, which consists of M reflecting semi-passive elements. The IRS adjusts the phase of the incoming signal and reflects it toward the legitimate IoTDs. There is no direct physical connection or link between the AP and the legitimate IoTDs. Instead, the IRS acts as an intermediary.[1] The legitimate IoTDs are divided into two clusters: the near cluster and the far cluster, based on their respective distances from the IRS. In this system, IoTD n is selected from the near cluster, while IoTD f is selected from the far cluster. To ensure that the channels between the IRS and the eavesdropper, as well as the channels between the IRS and the legitimate IoTDs, are uncorrelated, the distances between IoTD n and e and between IoTD f and e are kept at least half a wavelength. The distances between the IRS and IoTD n and between the IRS and IoTD f are denoted as d_n and d_f, respectively. The wireless channels in the system are modeled using Rician fading [45], [40]. Specifically, $\mathbf{h_t}$ represents the transmitter channel between the AP and the IRS, $\mathbf{h_n}$ represents the receiver channel between the IRS and IoTD n, $\mathbf{h_f}$ represents the receiver channel between the IRS and IoTD f, and $\mathbf{h_e}$ represents the channel between the IRS and the eavesdropper. It's important to note that the channel between the IRS and the eavesdropper follows a Rician distribution, indicating a worst-case scenario with a dominant Line-of-Sight (LOS) link in this particular channel.

$$\mathbf{h_t} = \sqrt{\frac{\kappa_1}{\kappa_1+1}}\bar{\mathbf{h}}_\mathbf{t} + \sqrt{\frac{1}{\kappa_1+1}}\tilde{\mathbf{h}}_\mathbf{t}, \tag{5.1}$$

$$\mathbf{h_n} = \sqrt{\frac{\kappa_2}{\kappa_2+1}}\bar{\mathbf{h}}_\mathbf{n} + \sqrt{\frac{1}{\kappa_2+1}}\tilde{\mathbf{h}}_\mathbf{n}, \tag{5.2}$$

$$\mathbf{h_e} = \sqrt{\frac{\kappa_2}{\kappa_2+1}}\bar{\mathbf{h}}_\mathbf{e} + \sqrt{\frac{1}{\kappa_2+1}}\tilde{\mathbf{h}}_\mathbf{e}, \tag{5.3}$$

$$\mathbf{h_f} = \sqrt{\frac{\kappa_2}{\kappa_2+1}}\bar{\mathbf{h}}_\mathbf{f} + \sqrt{\frac{1}{\kappa_2+1}}\tilde{\mathbf{h}}_\mathbf{f}. \tag{5.4}$$

[1] The considered scenario utilizes IRS for coverage extension.

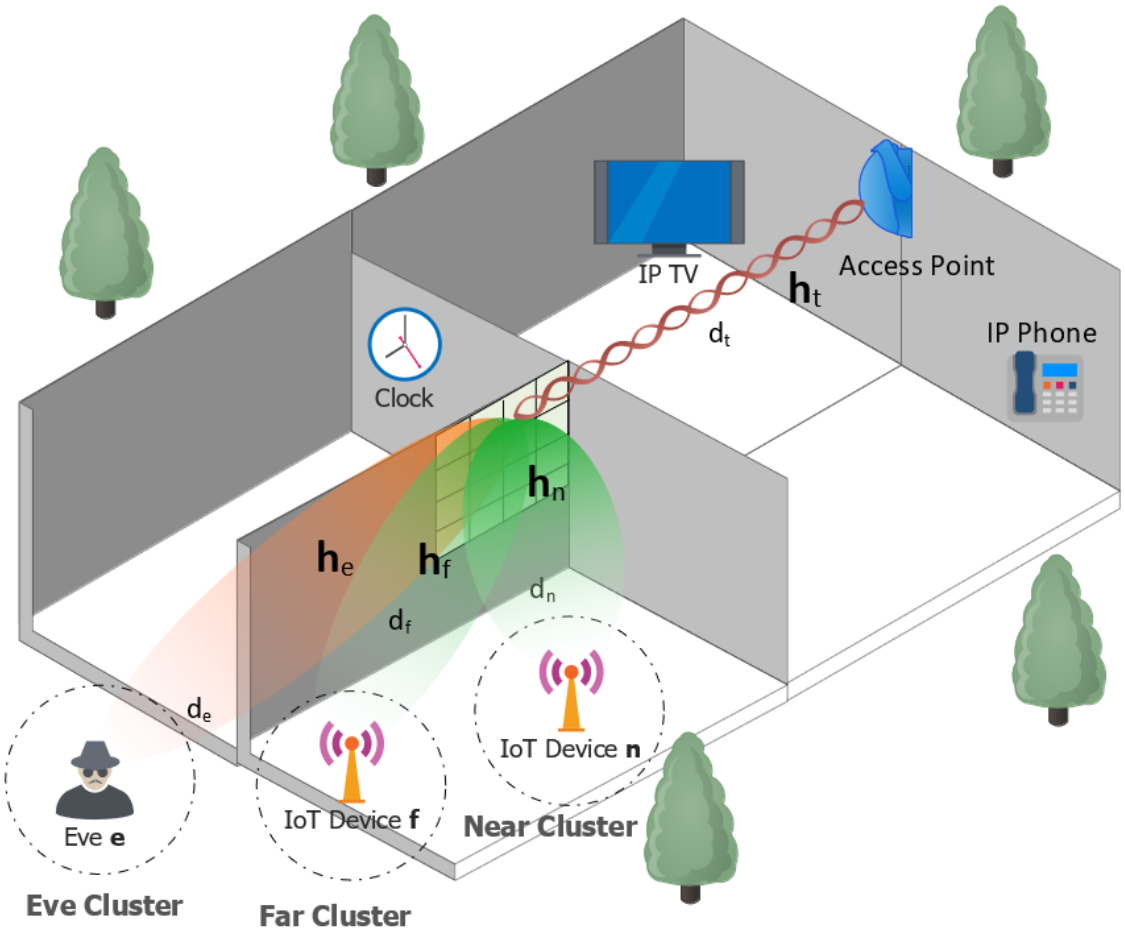


Figure 5.1: Downlink indoor extended coverage using IRS secure NOMA-SWIPT network topology.

The signal transmitted from the AP can be represented as follows:

$$x = \sqrt{\alpha P_t} s_n + \sqrt{\bar{\alpha} P_t} s_f \tag{5.5}$$

In this equation, x is a scalar representing the transmitted signal for one sample. s_n and s_f denote the signals intended for IoTDs n and f, respectively, with unit power (i.e., $\mathbb{E}[|s_n|^2] = 1$ and $\mathbb{E}[|s_f|^2] = 1$, where $\mathbb{E}[\cdot]$ denotes the expectation function).

The power allocated to IoTD n is given by $P_n = \alpha P_t$, where α is the NOMA power allocation factor for IoTD n. Similarly, the power allocated to IoTD f is $P_f = \bar{\alpha} P_t$, where $\bar{\alpha} = 1 - \alpha$ is the power allocation factor for IoTD f. Here, $0 \leq \alpha \leq 1$ represents the power allocation factor, and P_t corresponds to the total transmit power at the AP.

The received signals for IoTDs n and f in the frequency domain can be expressed

as:

$$y_1 = c_n c_t \mathbf{h}_n^H \mathbf{\Phi_M} \mathbf{h}_t^H x + n_n, \tag{5.6}$$

$$y_2 = c_f c_t \mathbf{h}_f^H \mathbf{\Phi_M} \mathbf{h}_t^H x + n_f, \tag{5.7}$$

In these equations, $c_t = (1 + d_t^{\Xi_t})^{-\frac{1}{2}}$, $c_n = (1 + d_n^{\Xi_n})^{-\frac{1}{2}}$, and $c_f = (1 + d_f^{\Xi_f})^{-\frac{1}{2}}$ are channel coefficients that account for the path loss between the AP, IRS, and the IoTDs [35], [9]. The variables d_t, d_n, and d_f represent the distances between the AP and the IRS, the IRS and IoTD n, and the IRS and IoTD f, respectively. The symbols Ξ_t, Ξ_n, and Ξ_f represent the path loss exponents. The channel between the AP and the IRS is represented by $\mathbf{h_t} \in \mathbb{C}^{1\times M}$, while the channels between the IRS and IoTDs n and f are denoted by $\mathbf{h}_n \in \mathbb{C}^{M\times 1}$ and $\mathbf{h}_f \in \mathbb{C}^{M\times 1}$, respectively. The phase shift matrix $\mathbf{\Phi_M} = \text{diag}(e^{j\theta_1}, e^{j\theta_2}, ..., e^{j\theta_M})$ represents the phase shifts applied by the IRS elements, where θ_m is the phase shift of the m^{th} element. The phase shift matrix satisfies the unit modulus constraint $|\phi_m|^2 = |e^{j\theta_m}|^2 = 1$ for all $m \in 1,2,...,M$, as the IRS reflects the signal without amplification. The terms $n_n \sim \mathcal{CN}(0, \sigma_n^2)$ and $n_f \sim \mathcal{CN}(0, \sigma_f^2)$ represent the AWGN for IoTDs n and f, respectively.

In this scenario, the legitimate IoTDs are considered as EH nodes. A power splitting EH scheme is applied, where a portion of the received signal is used for EH, and the remaining portion is used for information decoding. The received signals after power splitting can be written as:

$$y_n = c_n c_t \sqrt{\beta_n} \mathbf{h}_n^H \mathbf{\Phi_M} \mathbf{h}_t^H x + n_n, \tag{5.8}$$

$$y_f = c_f c_t \sqrt{\beta_f} \mathbf{h}_f^H \mathbf{\Phi_M} \mathbf{h}_t^H x + n_f, \tag{5.9}$$

Here, β_n and β_f represent the power splitting factors for IoTD n and IoTD f, respectively. These factors determine how the incoming signal power is divided between EH and information decoding. The terms $\sqrt{\beta_n}$ and $\sqrt{\beta_f}$ scale the received signals accordingly.

5.2 Secrecy Rate Derivation

In the non-cooperative downlink NOMA scheme, IoTD f, which is the device located farther from the IRS, decodes its signal in the presence of interference from the nearby IoTD n. IoTD f treats the signal from IoTD n as noise without removing it from the received signal. On the other hand, IoTD n performs SIC [2] by removing the message intended for IoTD f and subsequently decoding its signal. Consequently, the received SINR at IoTDs n and f can be expressed as follows:

[2] We assume perfect SIC.

$$\gamma_n = \left(\frac{c_n^2 c_t^2 \beta_n |\mathbf{h}_n^H \mathbf{\Phi_M} \mathbf{h}_t^H|^2 \alpha P_t}{\sigma_n^2} \right), \tag{5.10}$$
$$\gamma_f = \left(\frac{c_f^2 c_t^2 \beta_f |\mathbf{h}_f^H \mathbf{\Phi_M} \mathbf{h}_t^H|^2 \bar{\alpha} P_t}{c_f^2 c_t^2 |\mathbf{h}_f^H \mathbf{\Phi_M} \mathbf{h}_t^H|^2 \alpha P_t + \sigma_f^2} \right).$$

The data rates for IoTDs n and f are given by equations:

$$R_n = \log_2(1 + \gamma_n), \tag{5.11}$$
$$R_f = \log_2(1 + \gamma_f).$$

The achievable secrecy rate is the rate where the legitimate devices n and f can securely communicate in the presence of e. Therefore, the achievable secrecy rate at device n can be defined as the discrepancy between the rate of the main communication channel and the maximum rate achievable by the e. This metric quantifies the level of secure communication for the legitimate device by considering the potential information leakage to e. If e is interested in decoding either n's or f's message, then he will consider the signal that isn't interested in as interference. Hence, if e is interested in decoding n's message only, then e's rate will be expressed as:

$$R_{e,n} = \log_2(1 + \gamma_{e,n}), \tag{5.12}$$

where

$$\gamma_{e,n} = \left(\frac{c_e^2 c_t^2 |\mathbf{h}_e^H \mathbf{\Phi_M} \mathbf{h}_t^H|^2 \alpha P_t}{c_e^2 c_t^2 |\mathbf{h}_e^H \mathbf{\Phi_M} \mathbf{h}_t^H|^2 \bar{\alpha} P_t + \sigma_e^2} \right), \tag{5.13}$$

The variable c_e represents the path loss coefficient between the IRS and e, and it is computed as $c_e = (1 + d_e^{\Xi_e})^{-\frac{1}{2}}$. Here, d_e denotes the distance between the IRS and e, and it is required that $d_e > d_f$ to ensure that e is farther away from the IRS than IoTD f. Ξ_e represents the path loss exponent.[3] The term σ_e^2 represents the variance of the AWGN at e. If e is interested in decoding f's message only, then e's rate will be expressed as:

$$R_{e,f} = \log_2(1 + \gamma_{e,f}), \tag{5.14}$$

where

$$\gamma_{e,f} = \left(\frac{c_e^2 c_t^2 |\mathbf{h}_e^H \mathbf{\Phi_M} \mathbf{h}_t^H|^2 \bar{\alpha} P_t}{c_e^2 c_t^2 |\mathbf{h}_e^H \mathbf{\Phi_M} \mathbf{h}_t^H|^2 \alpha P_t + \sigma_e^2} \right), \tag{5.15}$$

[3]Practically for Eve to be hidden from legitimate users, it has to be outside the office structure.

The secrecy rates, $R_{s,n}$ and $R_{s,f}$, for IoTDs n and f, respectively are expressed as follows:

$$R_{s,n} \leq [R_n - R_{e,n}]^+, \tag{5.16}$$
$$R_{s,f} \leq [R_f - R_{e,f}]^+. \tag{5.17}$$

In (5.16), and (5.17) $[.]^+$ represent the positive part function, and it ensures that the secrecy sum-rate is non-negative.
The total secrecy rate, R_s, is calculated by adding the secrecy rates for IoTDs n and f, and it is upper bounded by the following equation: [3]

$$R_s = R_{s,n} + R_{s,f} \leq [R_n + R_f - (R_{e,n} + R_{e,f})]^+ . \tag{5.18}$$

Both rates ($R_{s,n}$, $R_{s,f}$) need to satisfy the constraints given in (5.16), (5.17), respectively. Additionally, the total secrecy rate, R_s, needs also to satisfy the upper bound constraint given in (5.18).

5.3 Energy Harvesting

The signals received through the energy harvesting circuitry at both IoTDs n and f can be expressed as follows:

$$\upsilon_n = c_n c_t \sqrt{\bar{\beta}_n} \mathbf{h}_n^H \mathbf{\Phi_M} \mathbf{h}_t^H x, \tag{5.19}$$
$$\upsilon_f = c_f c_t \sqrt{\bar{\beta}_f} \mathbf{h}_f^H \mathbf{\Phi_M} \mathbf{h}_t^H x,$$

Here, $\bar{\beta}_n = 1 - \beta_n$, and $\bar{\beta}_f = 1 - \beta_f$.

$$\upsilon_n = c_n c_t \sqrt{\bar{\beta}_n} \mathbf{h}_n^H \mathbf{\Phi_M} \mathbf{h}_t^H (\sqrt{\alpha P_t} s_n + \sqrt{\bar{\alpha} P_t} s_f), \tag{5.20}$$
$$\upsilon_f = c_f c_t \sqrt{\bar{\beta}_f} \mathbf{h}_f^H \mathbf{\Phi_M} \mathbf{h}_t^H (\sqrt{\alpha P_t} s_n + \sqrt{\bar{\alpha} P_t} s_f),$$

The harvested energy E_n and E_f at IoTDs n and f, respectively, can be obtained by taking the squared magnitude of the received signals and considering the efficiency factor η for RF energy conversion:

$$E_n = \eta c_n^2 c_t^2 \bar{\beta}_n \left| \mathbf{h}_n^H \mathbf{\Phi_M} \mathbf{h}_t^H (\sqrt{\alpha P_t} s_n + \sqrt{\bar{\alpha} P_t} s_f) \right|^2 \tag{5.21}$$
$$E_f = \eta c_f^2 c_t^2 \bar{\beta}_f \left| \mathbf{h}_f^H \mathbf{\Phi_M} \mathbf{h}_t^H (\sqrt{\alpha P_t} s_n + \sqrt{\bar{\alpha} P_t} s_f) \right|^2 ,$$

In these equations, $0 \leq \eta \leq 1$ represents the efficiency factor of the RF energy conversion operation at the energy harvester circuit. It indicates the efficiency of converting the received RF signal into usable energy.

5.4 Packet Loss

To address the cross-layer design aspect in our scenario, we aim to optimize the network layer packet loss in addition to improving the physical layer security (PLS), as discussed earlier. This involves maximizing the secure sum-rate while also enhancing the packet loss performance at different layers of the system. To model the packet loss in our wireless system, we will adopt the two-state Gilbert-Elliot Markov channel model, as illustrated in Fig. 5.2. This model defines two states: the bad state, representing high packet losses, and the good state, representing low packet losses. In addition, some parameters in the model characterize the behavior of the channel. These parameters include Ω, which is the probability of the channel remaining in the bad state, and P, which represents the probability of transitioning from the good state to the bad state. The steady-state probabilities of the channel being in the bad or good state can be expressed as follows:[22]:

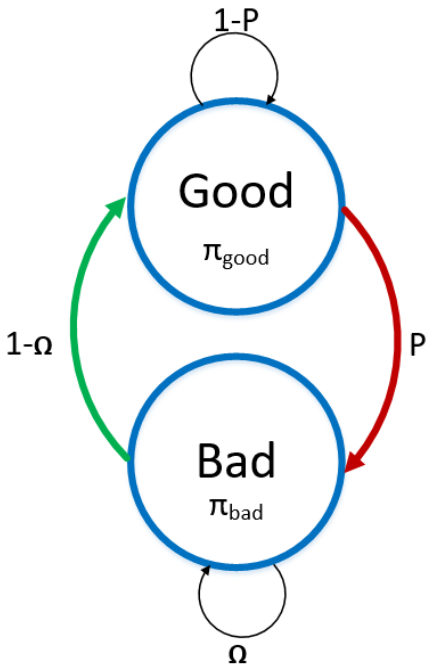


Figure 5.2: Gilbert-Elliot channel model.

$$\pi_{bad} = \frac{P}{1-\Omega+P}, \qquad \pi_{good} = \frac{1-\Omega}{1-\Omega+P}, \tag{5.22}$$

where $0 < P < 1$, and $0 < \Omega < 1$.

In the context of IoT, IEEE standard 802.15.4, wireless communications is characterized by packet loss. This standard is designed for low data rates and low-cost wireless personal area networks (LR-WPANs). It encompasses the physical (PHY) and medium access control (MAC) layers and serves as the foundation for many IoT protocols, and it focuses on low-speed pervasive communication among the devices. The network layer average packet loss probability is expressed as [23]:

$$P_l = e_B \pi_{bad} + e_G \pi_{good}. \tag{5.23}$$

Here, e_B denotes the probability of packet error when the channel is in a bad state, while e_G represents the probability of packet error when the channel is in a good state. To determine the values of e_B and e_G, we will employ the K-State Markov chain model described in the [10]. It is important to note that the original model was developed for the Rayleigh fading scenario, which assumes a small distance between the transmitter (T_x) and receiver (R_x), such as in Wireless Local Area Network (WLAN) setups. However, in our specific scenario, we extend the analysis to incorporate the cascaded Rician fading channel case. The bit error probabilities for a K-state channel model are expressed as [10]:

$$b_{k,n} = \frac{\Gamma_{k,n} - \Gamma_{k+1,n}}{\tilde{\gamma_n}}, \tag{5.24}$$

$$b_{k,f} = \frac{\Gamma_{k,f} - \Gamma_{k+1,f}}{\tilde{\gamma_f}}. \tag{5.25}$$

$$\Gamma_{k,n} = \exp\left(-\frac{A_k}{\tilde{\gamma_n}}\right)\left(1 - F(\sqrt{2A_k})\right) + \sqrt{\frac{\tilde{\gamma_n}}{\tilde{\gamma_n}+1}} F\left(\sqrt{\frac{2A_k\tilde{\gamma_n}+1}{\tilde{\gamma_n}}}\right), \tag{5.26}$$

$$\Gamma_{k,f} = \exp\left(-\frac{A_k}{\tilde{\gamma_f}}\right)\left(1 - F(\sqrt{2A_k})\right) + \sqrt{\frac{\tilde{\gamma_f}}{\tilde{\gamma_f}+1}} F\left(\sqrt{\frac{2A_k\tilde{\gamma_f}+1}{\tilde{\gamma_f}}}\right). \tag{5.27}$$

Hence, the packet error probabilities for IoTDs n and f can be written as follows [10]:

$$e_{k,n} = 1 - (1 - b_{k,n})^{\mathcal{L}}, \tag{5.28}$$

$$e_{k,f} = 1 - (1 - b_{k,f})^{\mathcal{L}}. \tag{5.29}$$

where k denotes the number of states from 0 to K, $A_0, \ldots, A_K$ is a partition of the receiver SNR (R-SNR) range, A_k represents the threshold for the R-SNR range, $\tilde{\gamma_n}$ is the mean of the R-SNR for IoTD n, and $\tilde{\gamma_f}$ is the mean of the R-SNR for IoTD f, and $\mathcal{L}$ is the packet length.

Considering the special case of Gilbert-Elliott channel model ($K = 2$) and assuming that state 'bad' is state 0 and state 'good' is state 1, taking into account that $A_0 = 0$, A_1 is the pre-selected SNR threshold that separates good and bad channel conditions, and $A_2 = \infty$, the bit and packet error probabilities can be re-written as:

$$e_{0,n} = 1 - (1 - b_{0,n})^{\mathcal{L}}, \tag{5.30}$$
$$e_{0,f} = 1 - (1 - b_{0,f})^{\mathcal{L}}. \tag{5.31}$$

$$e_{1,n} = 1 - (1 - b_{1,n})^{\mathcal{L}}, \tag{5.32}$$
$$e_{1,f} = 1 - (1 - b_{1,f})^{\mathcal{L}}. \tag{5.33}$$

$$b_{0,n} = \frac{\Gamma_{0,n} - \Gamma_{1,n}}{\tilde{\gamma_n}}, \tag{5.34}$$
$$b_{0,f} = \frac{\Gamma_{0,f} - \Gamma_{1,f}}{\tilde{\gamma_f}}. \tag{5.35}$$

$$b_{1,n} = \frac{\Gamma_{1,n} - \Gamma_{2,n}}{\tilde{\gamma_n}}, \tag{5.36}$$
$$b_{1,f} = \frac{\Gamma_{1,f} - \Gamma_{2,f}}{\tilde{\gamma_f}}. \tag{5.37}$$

$$\Gamma_{0,n} = 1 - F(0) + \sqrt{\frac{\tilde{\gamma_n}}{\tilde{\gamma_n} + 1}} F\left(\sqrt{\frac{1}{\tilde{\gamma_n}}}\right), \tag{5.38}$$

$$\Gamma_{1,n} = \exp\left(-\frac{A_1}{\tilde{\gamma_n}}\right)\left(1 - F(\sqrt{2A_1})\right) + \sqrt{\frac{\tilde{\gamma_n}}{\tilde{\gamma_n} + 1}} F\left(\sqrt{\frac{2A_1\tilde{\gamma_n} + 1}{\tilde{\gamma_n}}}\right), \tag{5.39}$$

$$\Gamma_{2,n} = \sqrt{\frac{\tilde{\gamma_n}}{\tilde{\gamma_n} + 1}}. \tag{5.40}$$

$$\Gamma_{0,f} = 1 - F(0) + \sqrt{\frac{\tilde{\gamma}_f}{\tilde{\gamma}_f + 1}} F\left(\sqrt{\frac{1}{\tilde{\gamma}_f}}\right), \tag{5.41}$$

$$\Gamma_{1,f} = \exp\left(-\frac{A_1}{\tilde{\gamma}_f}\right)\left(1 - F(\sqrt{2A_1})\right) + \sqrt{\frac{\tilde{\gamma}_f}{\tilde{\gamma}_f + 1}} F\left(\sqrt{\frac{2A_1\tilde{\gamma}_f + 1}{\tilde{\gamma}_f}}\right). \tag{5.42}$$

$$\Gamma_{2,f} = \sqrt{\frac{\tilde{\gamma}_f}{\tilde{\gamma}_f + 1}}. \tag{5.43}$$

Furthermore, $F(\gamma_n)$ corresponds to the cumulative distribution function (CDF) of the composite channel linking the access point to IoTD n, which is $\mathbf{h}_n^H \mathbf{\Phi_M} \mathbf{h}_t^H$. Similarly, $F(\gamma_f)$ signifies the CDF of the composite channel connecting the access point and IoTD f, which is $\mathbf{h}_f^H \mathbf{\Phi_M} \mathbf{h}_t^H$.

The CDF of q cascaded Rician channels can be expressed as [17]:

$$F(z) = \frac{1}{2^q} \sum_{\zeta_1=0}^{\infty} \sum_{\zeta_2=0}^{\infty} \cdots \sum_{\zeta_q=0}^{\infty} \mathcal{C}_q z^{2(\zeta_1+1)} \prod_{i=2}^{q} \left[\left(\frac{1}{2\delta_i^2}\right)^{\zeta_1-\zeta_i}\right] \times G_{1q+1}^{q1}\left(r^{-\zeta_1}\middle| \frac{z^2}{2^q \prod_{i=1}^{q} \delta_i^2}\right). \tag{5.44}$$

where $\mathcal{C}_q = \prod_{i=1}^{q}\left[\frac{1}{\zeta_i! \delta_i^2}\right] \prod_{i=1}^{q}\left[\frac{1}{\zeta_i! \delta_i^2} \exp(\frac{-v_i^2}{2\delta_i^2})(\frac{v_i^2}{2\delta_i^2})^{2\zeta_i}\right]$, the rice factor $\kappa_i = \frac{v_i^2}{2\delta_i^2}$, v_i^2 represents the average power in the LOS component, δ_i^2 represents the average power in the non-LOS component, G_{1q+1}^{q1} denotes the Meijer G-function, and $r = (\zeta_2 - \zeta_1, \zeta_3 - \zeta_1, \ldots, \zeta_q - \zeta_1, 0, -\zeta_1 - 1)$.

In our system model, only two cascaded Rician channels are used, so (5.44) is reduced to the below form:

$$F(z) = \frac{1}{2^2} \sum_{\zeta_1=0}^{\infty} \sum_{\zeta_2=0}^{\infty} \mathcal{C}_2 z^{2(\zeta_1+1)} \left[\left(\frac{1}{2\delta_2^2}\right)^{\zeta_1-\zeta_2}\right] \times G_{13}^{21}\left(r^{-\zeta_1}\middle| \frac{z^2}{2^2 \prod_{i=1}^{2} \delta_i^2}\right). \tag{5.45}$$

where $\mathcal{C}_2 = \frac{1}{\zeta_1!\delta_1^2}\frac{1}{\zeta_2!\delta_2^2}\exp(\frac{-v_1^2}{2\delta_1^2})(\frac{v_1^2}{2\delta_1^2})^{2\zeta_1}\exp(\frac{-v_2^2}{2\delta_2^2})(\frac{v_2^2}{2\delta_2^2})^{2\zeta_2}$, and $r = (\zeta_2 - \zeta_1, 0, -\zeta_1 - 1)$.

A linear combination of cascaded Rician random variables represents the aggregation of channel characteristics from the AP to the IoTDs, denoted as n and f. Thus, it's the summation of the cascaded Rician channels that encapsulate these channel properties. The central limit theorem states that the sum (or average) of a large number of independent and identically distributed random variables approaches a Gaussian (normal) distribution. In summary, if there is a large number of cascaded Rician channels with the same K-factor and they are independent and identically distributed, it can be expected that the distribution of their sum to be approximately Gaussian, in accordance with the central limit theorem [32].

The means μ_{g_n} of the resulting approximate gaussian distribution for IoTD n is represented as:

$$\begin{aligned} \mu_{g_n} &= E(\sum_{i=1}^{M}(\mathbf{H}_{n_i} \times \mathbf{H}_{t_i})) = \sum_{i=1}^{M}(E(\mathbf{H}_{n_i} \times \mathbf{H}_{t_i})) \\ \mu_{g_n} &= \sum_{i=1}^{M}(E(\mathbf{H}_{n_i}) \times E(\mathbf{H}_{t_i})) \end{aligned} \tag{5.46}$$

Similarly, for the mean μ_{g_f} of the resulting approximate Gaussian distribution for IoTD f:

$$\begin{aligned} \mu_{g_f} &= E(\sum_{i=1}^{M}(\mathbf{H}_{f_i} \times \mathbf{H}_{t_i})) = \sum_{i=1}^{M}(E(\mathbf{H}_{f_i} \times \mathbf{H}_{t_i})) \\ \mu_{g_f} &= \sum_{i=1}^{M}(E(\mathbf{H}_{f_i}) \times E(\mathbf{H}_{t_i})) \end{aligned} \tag{5.47}$$

Here, $\mathbf{H}_{t_i}$, $\mathbf{H}_{n_i}$, and, $\mathbf{H}_{f_i}$ represents the random variable for the channel response from the AP to the IRS, IRS to IoTD n, and IRS to IoTD f respectively for the i^{th} channel.

Further, the variance of the resulting approximate gaussian distribution for IoTD n can be expressed as follows:

$$\begin{aligned} \sigma_{g_n}^2 &= \mathrm{Var}(\sum_{i=1}^{M}(\mathbf{H}_{n_i} \times \mathbf{H}_{t_i})) \\ \sigma_{g_n}^2 &= \sum_{i=1}^{M}(E(\mathbf{H}_{n_i} \times \mathbf{H}_{t_i})^2 - (E(\mathbf{H}_{n_i} \times \mathbf{H}_{t_i}))^2) \\ \sigma_{g_n}^2 &= \sum_{i=1}^{M} E(\mathbf{H}_{n_i})^2 \times E(\mathbf{H}_{t_i})^2 - (E(\mathbf{H}_{n_i} \times \mathbf{H}_{t_i}))^2) \end{aligned} \tag{5.48}$$

Similarly, for the variance of the resulting approximate Gaussian distribution for IoTD f:

$$\sigma_{g_f}^2 = \mathrm{Var}(\sum_{i=1}^{M}(\mathbf{H}_{f_i} \times \mathbf{H}_{t_i})) \tag{5.49}$$
$$\sigma_{g_f}^2 = \sum_{i=1}^{M}(E(\mathbf{H}_{f_i} \times \mathbf{H}_{t_i})^2 - (E(\mathbf{H}_{f_i} \times \mathbf{H}_{t_i}))^2)$$
$$\sigma_{g_f}^2 = \sum_{i=1}^{M} E(\mathbf{H}_{f_i})^2 \times E(\mathbf{H}_{t_i})^2 - (E(\mathbf{H}_{f_i} \times \mathbf{H}_{t_i}))^2)$$

Therefore, the CDF of the composite channel linking the access point to IoTDs n and f can be expressed as [7]:

$$F(z) = \frac{1}{2} \times \left(1 + \mathrm{erf}\left(\frac{z - \mu_{g_n}}{\sqrt{2} \times \sigma_{g_n}}\right)\right) \tag{5.50}$$

$$F(z) = \frac{1}{2} \times \left(1 + \mathrm{erf}\left(\frac{z - \mu_{g_f}}{\sqrt{2} \times \sigma_{g_f}}\right)\right) \tag{5.51}$$

Through substitution, we get the following equations for IoTD n:

$$F\left(\sqrt{\frac{1}{\tilde{\gamma_n}}}\right) = \frac{1}{2} \times \left(1 + \mathrm{erf}\left(\frac{\sqrt{\frac{1}{\tilde{\gamma_n}}} - \mu_{g_n}}{\sqrt{2} \times \sigma_{g_n}}\right)\right) \tag{5.52}$$

$$F(\sqrt{2A_1}) = \frac{1}{2} \times \left(1 + \mathrm{erf}\left(\frac{\sqrt{2A_1} - \mu_{g_n}}{\sqrt{2} \times \sigma_{g_n}}\right)\right) \tag{5.53}$$

$$F\left(\sqrt{\frac{2A_1\tilde{\gamma_n} + 1}{\tilde{\gamma_n}}}\right) = \frac{1}{2} \times \left(1 + \mathrm{erf}\left(\frac{\sqrt{\frac{2A_1\tilde{\gamma_n}+1}{\tilde{\gamma_n}}} - \mu_{g_n}}{\sqrt{2} \times \sigma_{g_n}}\right)\right) \tag{5.54}$$

Similarly for IoTD f:

$$F\left(\sqrt{\frac{1}{\tilde{\gamma_f}}}\right) = \frac{1}{2} \times \left(1 + \mathrm{erf}\left(\frac{\sqrt{\frac{1}{\tilde{\gamma_f}}} - \mu_{g_f}}{\sqrt{2} \times \sigma_{g_f}}\right)\right) \tag{5.55}$$

$$F(\sqrt{2A_1}) = \frac{1}{2} \times \left(1 + \text{erf}\left(\frac{\sqrt{2A_1} - \mu_{g_f}}{\sqrt{2} \times \sigma_{g_f}}\right)\right) \tag{5.56}$$

$$F\left(\sqrt{\frac{2A_1\tilde{\gamma}_f + 1}{\tilde{\gamma}_f}}\right) = \frac{1}{2} \times \left(1 + \text{erf}\left(\frac{\sqrt{\frac{2A_1\tilde{\gamma}_f+1}{\tilde{\gamma}_f}} - \mu_{g_f}}{\sqrt{2} \times \sigma_{g_f}}\right)\right) \tag{5.57}$$

After calculating the packet error probabilities $e_{0,n}$, $e_{0,f}$, $e_{1,n}$, and $e_{1,f}$ in (5.30), (5.31), (5.32) and (5.33), our objective is to reduce these probabilities, thereby minimizing P_l as defined in (5.23).

5.5 Optimization Problem

The optimization problem in focus seeks to maximize the NOMA IoTDs n and f secrecy sum-rate. The variables to be optimized include the phase shift matrix Φ_M, the power splitting factors β_n and β_f, and the NOMA power allocation parameter α. The goal is to increase the secrecy sum-rate's expected value as much as possible. The optimization problem is subject to several constraints. First, the phase shift matrix should satisfy the unit modulus constraint, ensuring that $|\phi_m|^2 = 1$ for all elements of the matrix. The power splitting factors and the NOMA power allocation factor are bounded between 0 and 1. Additionally, the energy harvested at IoTDs n and f should be greater than or equal to their respective targeted average rates ψ_n and ψ_f. The packet loss at the network layer should not exceed the specified thresholds ϵ_n and ϵ_f.

$$\begin{aligned}
\max_{\Phi_M,\beta_n,\beta_f,\alpha} &: \{\mathbb{E}\{R_s\}\}, \\
\text{s.t.} |\phi_m|^2 = 1, \forall m \in \{1,2,...,M\}, \\
0 \leq \alpha \leq 1, \\
0 \leq \beta_n \leq 1, \\
0 \leq \beta_f \leq 1, \\
E_n \geq \psi_n, \\
E_f \geq \psi_f, \\
P_{l_n} \leq \epsilon_n \\
P_{l_f} \leq \epsilon_f
\end{aligned} \tag{5.58}$$

where ψ_n and ψ_f are the targeted average rates for the harvested energy at IoTDs n and f, respectively, and ϵ, is the threshold that the packet loss should not exceed.

The optimization problem in question is NP-hard due to the non-convexity caused by the constant modulus restrictions at the IRS, as stated in references [48], [49], [47]. To address this challenging problem, DRL is employed, specifically using the DDPG algorithm. In this approach, the DDPG algorithm is utilized to find the optimal values for the optimization parameters. An agent, represented by the IRS, is trained using DRL techniques and receives feedback from the IoTDs. Based on this feedback, the agent adjusts the phase shifts to maximize the secure sum-rate. In the proposed system, the transmit power is initially sent from the AP to the IoTDs. The IoTDs then estimate the rates and communicate them back to the IRS, which acts as the DRL agent. The IRS uses this information to optimize the phase shifts and maximize the overall secure sum-rate in the system. Through this learning process, the system aims to achieve the highest secure sum-rate by leveraging DRL to adapt the phase shift matrices based on the feedback received from the IoTDs.

5.6 Deep Deterministic Policy Gradient

5.6.1 System Mapping to DDPG

In the mapping between the communication system model and the DDPG scheme, we define the action-space, state-space, and reward to obtain a solution for the optimization problem. Further, we explain below this mapping between the system model and the DDPG scheme in detail:

State-space

The state-space of the DDPG at algorithm at a given timestep (τ) can be defined as follows:

$$\mathbf{s}^{(\tau)} = [E_n^{(\tau-1)}, E_f^{(\tau-1)}, P_{l_n}^{(\tau-1)}, P_{l_f}^{(\tau-1)}, R_s^{(\tau-1)}].$$

Here, $E_n^{(\tau-1)}$ and $E_f^{(\tau-1)}$ represent the energy harvested at IoTDs n and f, respectively, at the previous time step $(\tau - 1)$. $P_{l_n}^{(\tau)}$ and $P_{l_f}^{(\tau)}$ represent the packet loss for IoTDs n and f, $R_s^{(\tau-1)}$ represents the secure sum-rate at the previous time step $(\tau - 1)$.

By including these components in the state-space, the DDPG agent has access to the relevant information about the system, including channel conditions, previous actions, EH, packet loss, and previous secure sum-rate. This information empowers the agent to make well-informed decisions and acquire an optimal policy that maximizes the total transmission secure rate while simultaneously considering the constraints imposed by the system.

Action-space

The action-space of the DDPG algorithm at time step (τ) represents the set of actions taken by the DRL agent. It consists of four components: $\alpha^{(\tau)}$, $\beta_n^{(\tau)}$, $\beta_f^{(\tau)}$, and $\mathbf{\Phi}_M^{(\tau)}$, which correspond to the NOMA power allocation factor, the EH factor of IoTD n, the EH factor of IoTD f, and the IRS phase shift matrix, respectively.

The action-space $\mathbf{a}^{(\tau)}$ is defined as:

$$\mathbf{a}^{(\tau)} = [\alpha^{(\tau)}, \beta_n^{(\tau)}, \beta_f^{(\tau)}, \mathbf{\Phi}_M^{(\tau)}]. \tag{5.59}$$

These actions describe the DRL agent's decisions at each time step (τ). The agent determines the NOMA power allocation factor, EH factors, and IRS phase shift matrix values. The DRL agent's goal is to optimize the secure sum-rate while ensuring that the system's limitations are met.

Reward function

In the DDPG scheme, the reward function $r^{(\tau)}$ is a function of the current state and action and it is defined as the maximum secrecy rate $R_s^{(\tau)}$ achieved at timestep (τ) that satisfies the constraints mentioned in (5.58). Hence, it is necessary to include $R_s^{(\tau)}$ in the reward function, along with the other constraints. The secrecy rate represents the secure communication rate between legitimate devices in the presence of an eavesdropper. The reward is intended to motivate the DRL agent to maximize the secure sum-rate, which is the fundamental goal of the optimization problem.

The reward function is given by:

$$\mathbf{r}^{\tau} = \begin{cases} R_s^{(\tau)}, & \text{if (5.58) holds} \\ R_s^{(\tau)} + \chi_1 \min(0, (\psi_n - E_n^{(\tau)})) & \\ +\chi_2 \min(0, (\psi_f - E_f^{(\tau)})) & \\ +\chi_3 \min(0, (\epsilon_n - P_{l_n}^{(\tau)})) & \\ +\chi_4 \min(0, (\epsilon_f - P_{l_f}^{(\tau)})), & \text{otherwise} \end{cases} \tag{5.60}$$

where $\chi_1, \chi_2, \chi_3, \chi_4$ represent the weights assigned to the discount factors associated with each constraint if a particular constraint is violated. The reward condition requires that the constraints mentioned in 5.58 be satisfied to ensure that the parameters do not fall below a specified lower bound. Failure to meet the constraints will result in a penalty to the reward, proportionate to the relevant discount factor. These weights are crucial for optimizing the system's performance to achieve desired outcomes. By maximizing the reward, the DRL agent learns to make decisions that result in higher secrecy rates and improved system performance.

5.7 Complexity Analysis

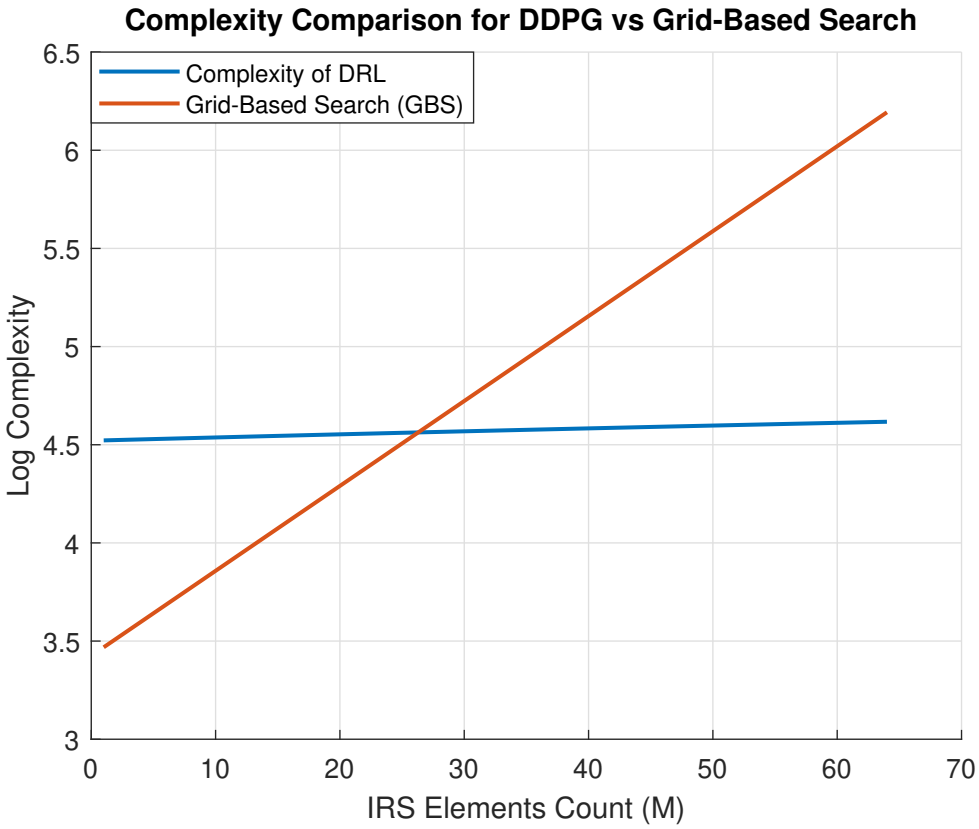


Figure 5.3: DDPG scheme's complexity vs. grid-based search.

To evaluate the complexity of the DDPG algorithm, we conducted a quantitative analysis comparing it to the complexity of the grid-based search method. The DDPG algorithm's complexity is explained in section 3.4 as follows:

$C_{DDPG} = \mathcal{S} \times \mathcal{I} + \mathcal{H} \times \mathcal{I} \times \mathcal{O} + \mathcal{O} \times \mathcal{A} + \mathcal{A}$

The complexity of the grid-based search scheme, denoted as C_{GBS}, can be determined based on the number of IoTDs $\mathcal{U}$, the number of elements for IRS as M, the IRS phase search steps denoted as $\lfloor \frac{2\pi}{\Delta\Phi} \rfloor$, the search step for the EH factors $\Delta\beta_n$ and $\Delta\beta_f$, and the power allocation factor search step $\Delta\alpha$. The complexity of the grid-based search scheme can then be expressed as:

$C_{GBS} = O(\mathcal{U} \times (1 + \frac{1}{\Delta\beta_n}) \times (1 + \frac{1}{\Delta\beta_f}) \times (1 + \frac{1}{\Delta\alpha}) \times (1 + \lfloor \frac{2\pi}{\Delta\Phi} \rfloor)^M)$

This formula indicates that the complexity of the grid-based search scheme is proportional to the number of IoTDs multiplied by the number of possible combinations for β_n multiplied by the number of possible combinations for β_f multiplied by the number of possible combinations for α multiplied by the number of possible phase combinations for the IRS raised to the power of the number of elements M.

Hence, the DDPG scheme's complexity is significantly lower than that of the gird-based search as the number of IRS elements increases as revealed in Fig. 5.3.

5.8 Simulation Results

Drawing from the analysis and derivations we previously discussed, we proceed to carry out simulations for our IRS NOMA-SWIPT system. Our objective is to explore how the optimization parameters influence the performance of crucial metrics. To facilitate a clearer comprehension of this influence, we generate plots depicting the secure sum-rate, harvested energy, and packet loss rate. Toward the conclusion of this section, we provide an overview of the optimized values obtained through both the grid-based search and the DRL methodology.

5.8.1 Simulation Parameters

Table 5.1: Simulation Parameters

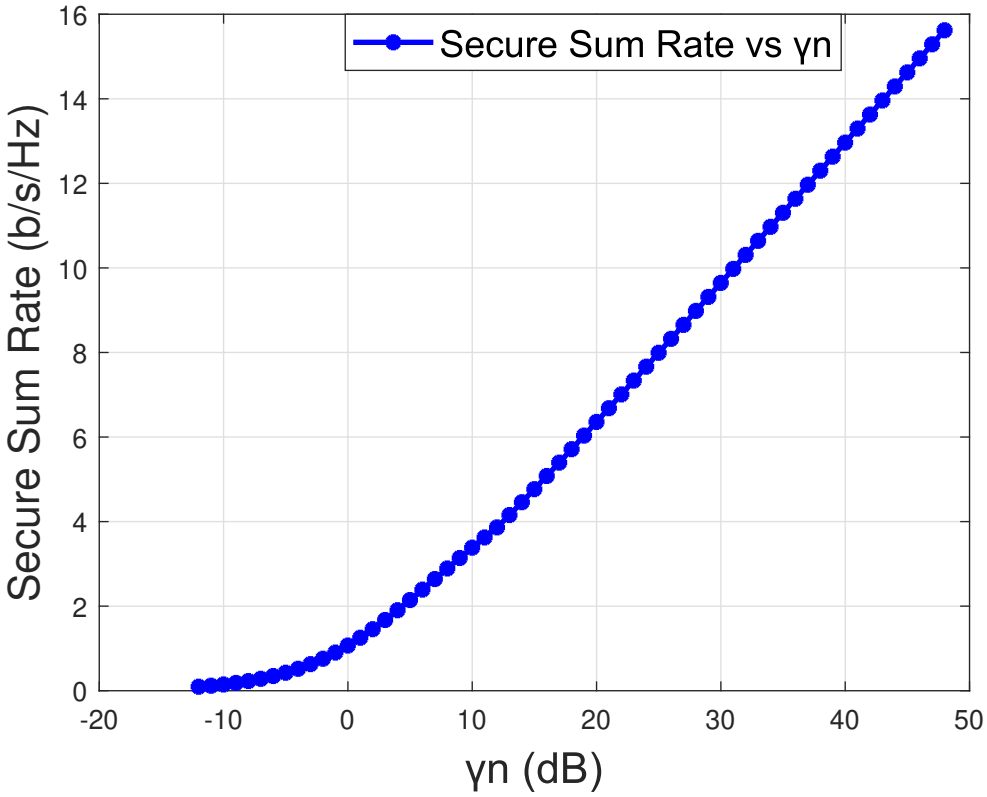
Simulation Parameters	**Values**
Pt	-30 dbm to 30 dbm
n_0	1×10^{-4} mW
mc	1000
M	12
α	$[0:1]$
β_n	$[0:1]$
β_f	$[0:1]$
η	0.75
M	12
K_1	10
K_2	10
Ξ_t	2
Ξ_n	2
Ξ_f	2
Ξ_e	2
P	0.9
Ω	0.5
$\mathcal{L}$	256
Total number of episodes	5000

To begin, we set the simulation parameters as follows. The number of IRS reflecting elements is 12, and the number of IoTDs is 2. The channels between

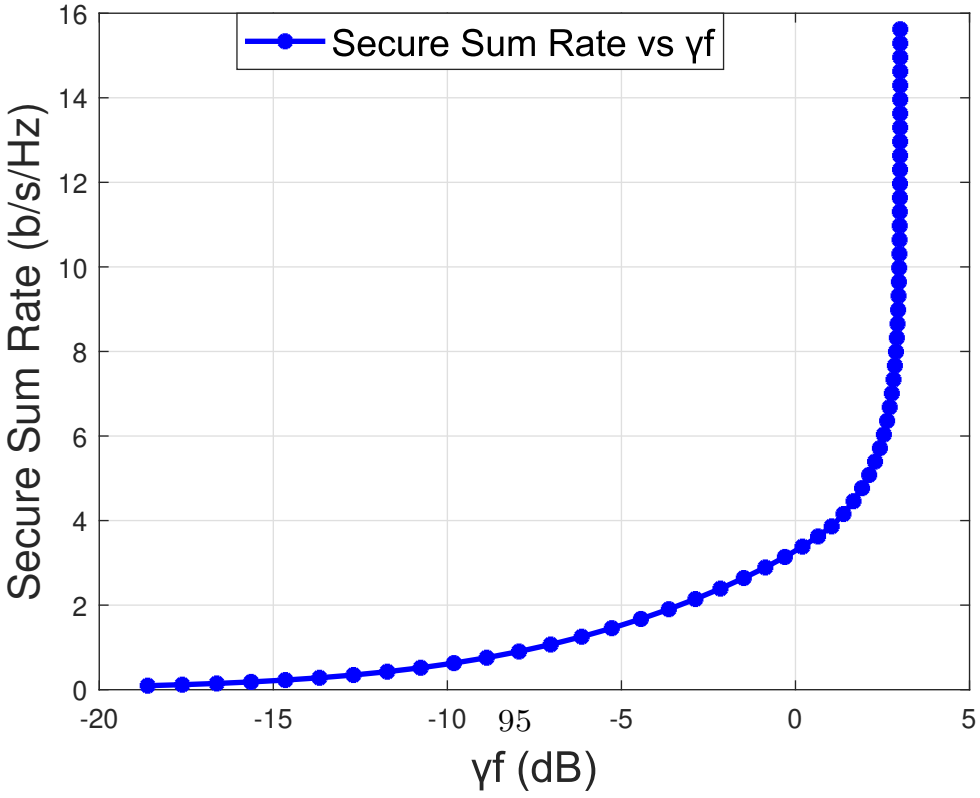
the AP and the IRS, IRS and IoTD n, IRS and IoTD f, and IRS and IoTD e are Rician channels with a factor of $K_1 = K_2 = 10$ and a path loss exponent of 2, the linear noise n_0 is set to 1×10^{-4} mW. The default distances between the AP and IRS, IRS and IoTD n, IRS and IoTD f, and IRS and IoTD e are 3 m, 2 m, 4 m, and 6 m, respectively. The probabilities P and Ω are both set to 0.9 and 0.5, respectively. The packet length, represented as $\mathcal{L}$, equals 256. The parameters α, β_n, and β_f, fall within the range of 0 to 1, and η is assigned a value of 0.75. We perform 1000 Monte Carlo simulations to obtain the numerical results. It is important to highlight that our solutions were implemented on a machine running a 64-bit Windows operating system. The machine is equipped with an Intel(R) Core(TM) i7-10510U CPU operating at 2.30GHz. Additionally, the machine has a memory capacity of 32 GB. A summary of our simulation parameters is provided in Table 5.1.

5.8.2 Performance Metrics Evaluation

To gain a deeper understanding of how different optimization parameters affect our system, we have presented the performance evaluation results in the following figures (Fig. 5.4 to Fig. 5.8). These figures illustrate the variations in our main performance metrics, namely secure sum rate, harvested energy, and packet loss rate, as we modify the optimization parameter.



(a) Secure sum-rate vs γ_n. $\beta_n = \beta_f = 0.5$, $\alpha = 0.2$.



(b) Secure sum-rate vs γ_f. $\beta_n = \beta_f = 0.5$, $\alpha = 0.2$.

Figure 5.4: Secure sum-rate for IoTDs n and f.

Secure Sum Rate

Figure 5.4 depicts the graphical representations of the secure sum-rate concerning IoTDs n and f. These plots showcase how the SINR for each IoTD affects the secure sum-rate, while maintaining constant values of $\beta_n = \beta_f = 0.5$ and $\alpha = 0.2$. In Fig. 5.4a, the secure sum-rate for IoTD n shows an increasing trend as the SINR rises. This indicates that as the quality of the channel and the received signal strength improve, IoTD n can achieve higher secure communication rates. On the other hand, in Fig. 5.4b, the secure sum-rate initially increases with an increasing SINR for IoTD f, but after reaching a certain point (around 3 dB), γ_f stops increasing with the increase of secure sum-rate. This behavior can be attributed to the interference caused by the nearby IoTD n in the downlink NOMA scheme. As the interference level from IoTD n increases, it occupies a significant portion of the available channel capacity, leaving less capacity for IoTD f to transmit its information. Consequently, the secure sum-rate for IoTD f starts to decrease due to the increased interference and noise. Therefore, while the secure sum-rate may increase overall due to the secure communication of IoTD n, the presence of interference limits the achievable secure rate for IoTD f. This highlights the trade-off and interference management challenges in downlink NOMA systems, where the performance of one IoTD can affect the performance of others in the network.

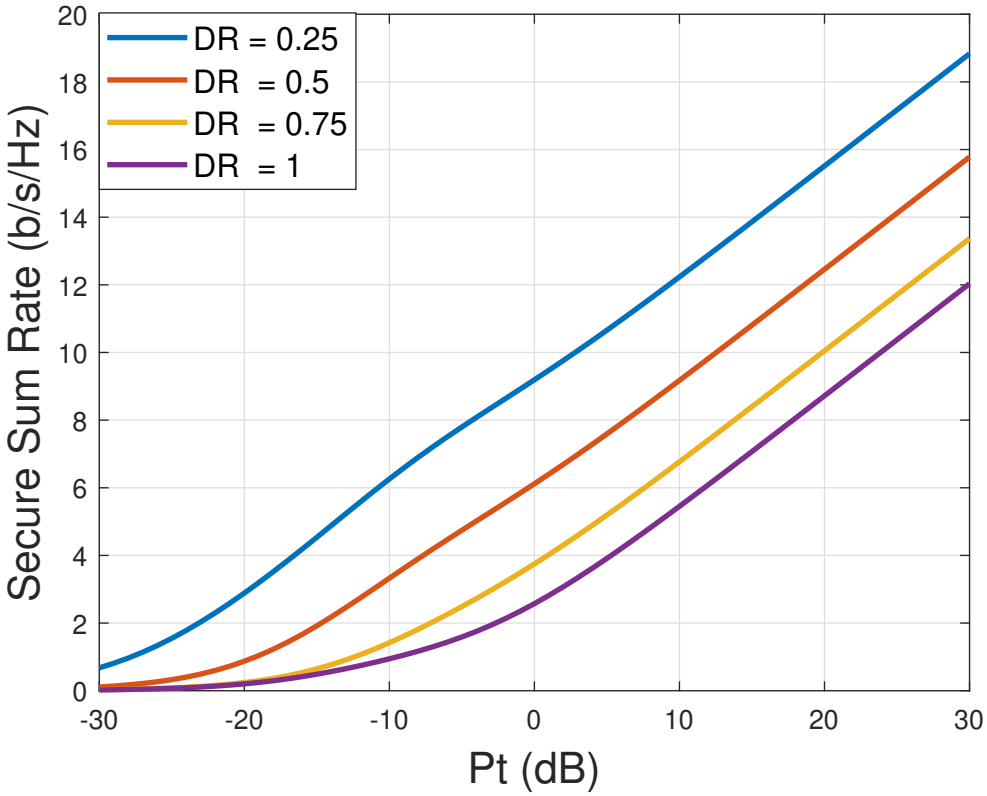
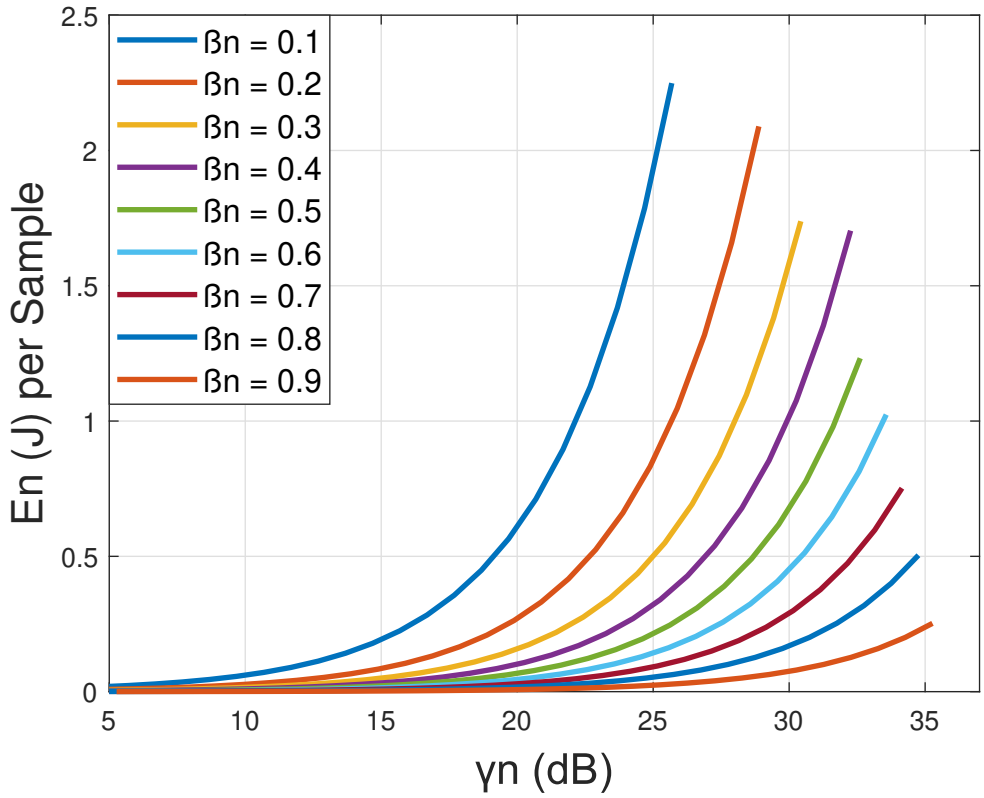


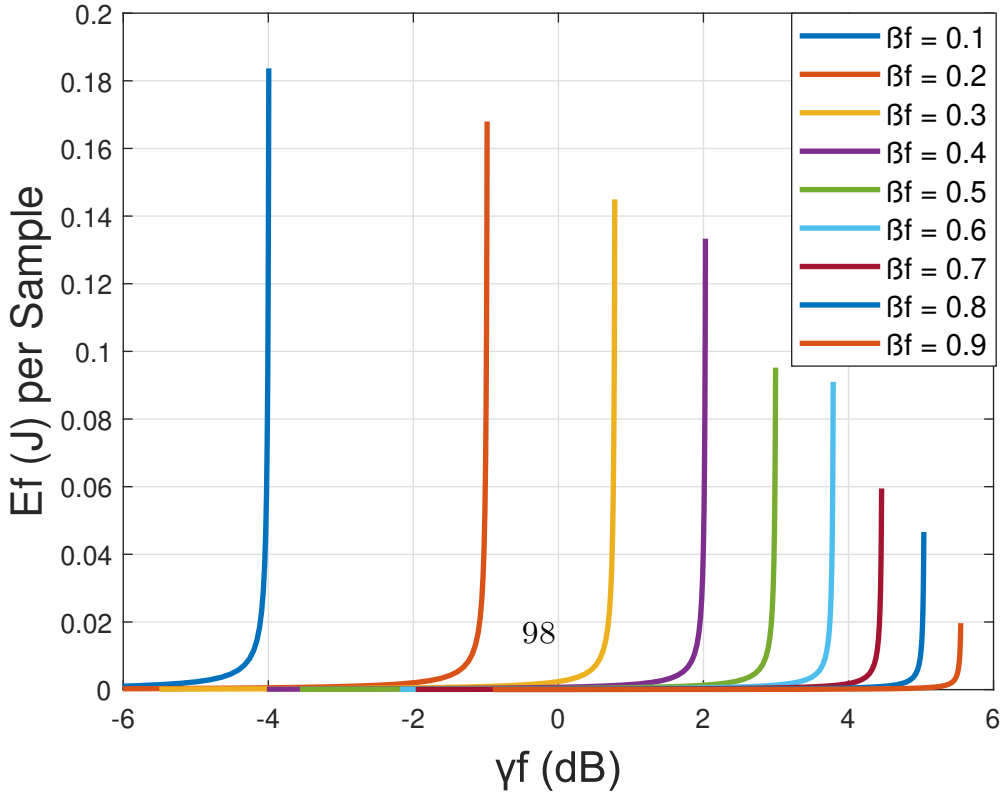
Figure 5.5: Secure sum-rate vs P_t (dB). $\beta_n = \beta_f = 0.5$, $\alpha = 0.2$, $d_t = 3$, $d_n = 1$ m to 4m, $d_f = 4$m, $d_e = 6$m.

In Fig. 5.5, we can observe the total transmission secure rate plotted against the transmit power P_t for various distance ratios between the IRS and IoTDs. The distance ratio represents the ratio of the distance between the IRS and IoTD n to the distance between the IRS and IoTD f. As the distance ratio increases, indicating that IoTD n is farther from the IRS compared to IoTD f, the total transmission secure rate decreases. This can be attributed to the fact that IoTD n experiences a weaker channel due to the increased distance from the IRS. Consequently, the secure rate of IoTD n decreases.

Harvested Energy



(a) Harvested energy for IoTD n vs γ_n. $\beta_f = 0.5$, $\alpha = 0.2$.$Pt = -15$ to 15 dbm.

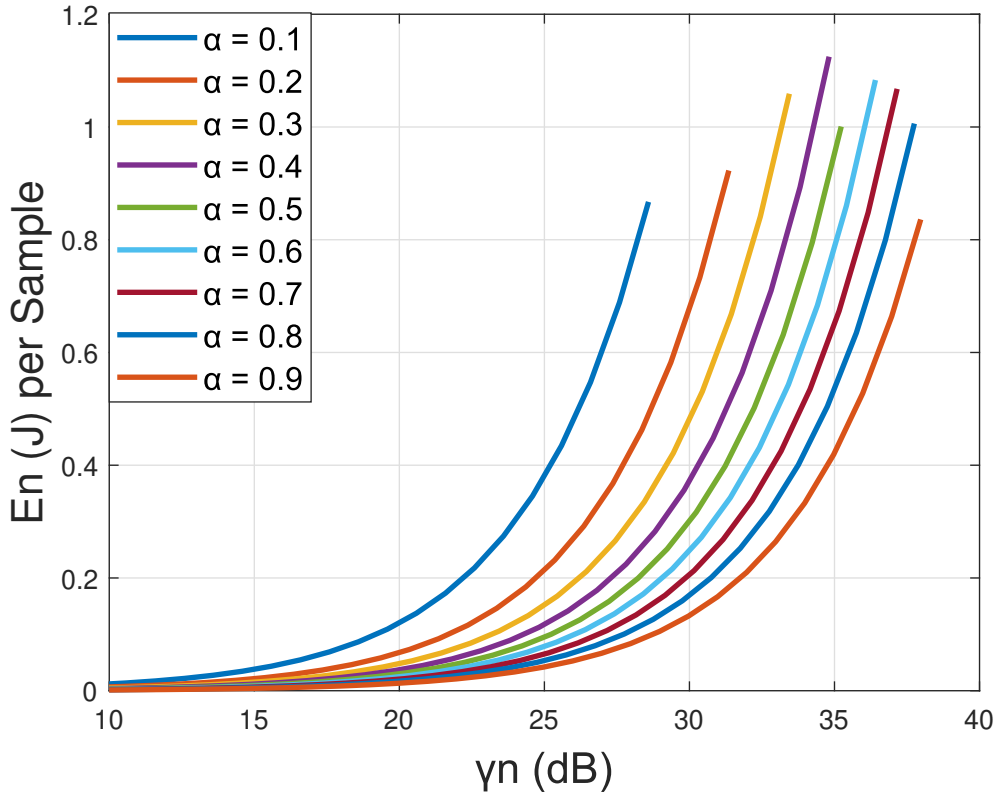


(b) Harvested energy for IoTD f vs γ_f. $\beta_n = 0.5$, $\alpha = 0.2$.$Pt = -15$ to 15 dbm.

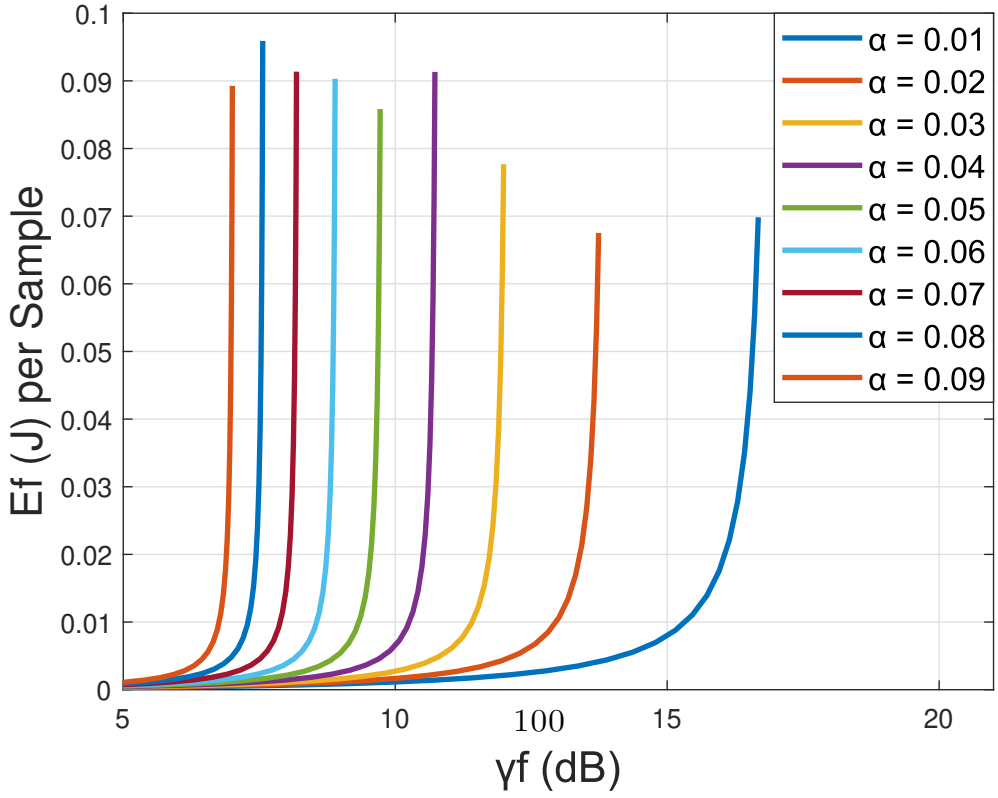
Figure 5.6: Harvested energy for IoTD devices n and f for various β_n and β_f.

In Fig. 5.6, we analyze the EH performance of the legitimate IoTDs for different EH factors β_n and β_f as a function of their respective SINRs, γ_n, and γ_f. The EH factors represent the power split ratios for IoTDs n and f, determining the portion of the received signal used for EH. As expected, as β_n and β_f decrease, the complementary power splitting factors $\bar{\beta}_n$ and $\bar{\beta}_f$ increase. This implies that a larger part of the received signal is allocated for EH whereas a smaller part is used for signal decoding. Consequently, the energy harvested by the IoTDs increases. The results in Fig. 5.6a demonstrate that the EH can increase by up to 10 orders of magnitude as β_n decreases from 0.9 to 0.1 (i.e., $\bar{\beta}_n$ increases from 0.1 to 0.9). This significant increase in EH is due to the higher power allocation for EH, allowing the IoTDs to capture more energy from the received signals. Furthermore, the results depicted in Figure 5.6b reveal that as the energy harvesting factor E_f increases, the SINR γ_f also increases until a certain point. Beyond that point, increasing E_f no longer leads to a further increase in γ_f. This phenomenon can be attributed to the interference caused by IoT device n on IoT device f. As E_f continues to increase, the interference from IoT device n becomes more significant, limiting the potential gains in SINR for IoT device f.

Moreover, the power allocation factor α plays a significant role in determining the energy harvested by legitimate IoTDs, as depicted in Figure 5.7, when α increases, the signal power allocated to IoTD n also increases, resulting in higher energy harvesting capabilities. It is important to note that at $\alpha = \bar{\alpha} = 0.5$, more energy is harvested from the nearby IoTD n compared to the distant IoTD f. This disparity is due to the higher path loss experienced by the farther device, leading to reduced received signal power and subsequently affecting energy harvesting performance. These observations underscore the trade-off between power splitting for energy harvesting and signal decoding, as well as the influence of path loss and power allocation on the energy harvesting capabilities of legitimate IoTDs. By optimizing the energy harvesting factors, along with the power allocation factor, the system can enhance energy harvesting efficiency and overall performance of the IoTDs.



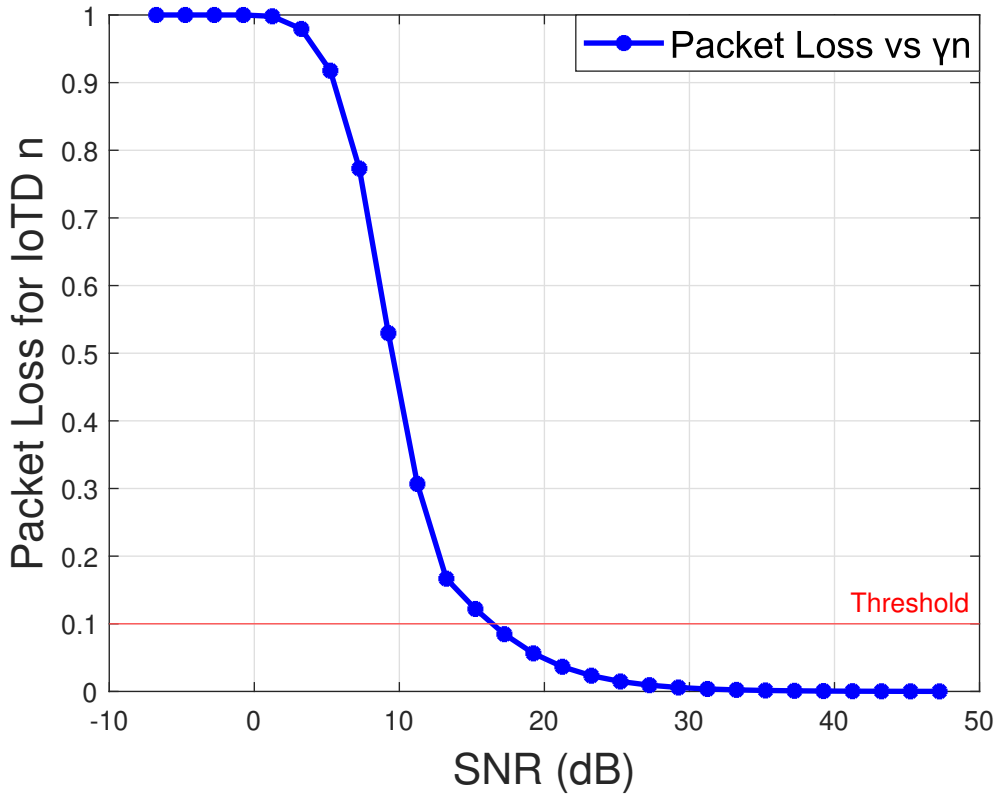
(a) Harvested energy for IoTD n vs γ_n. $\beta_n = \beta_f = 0.5$.



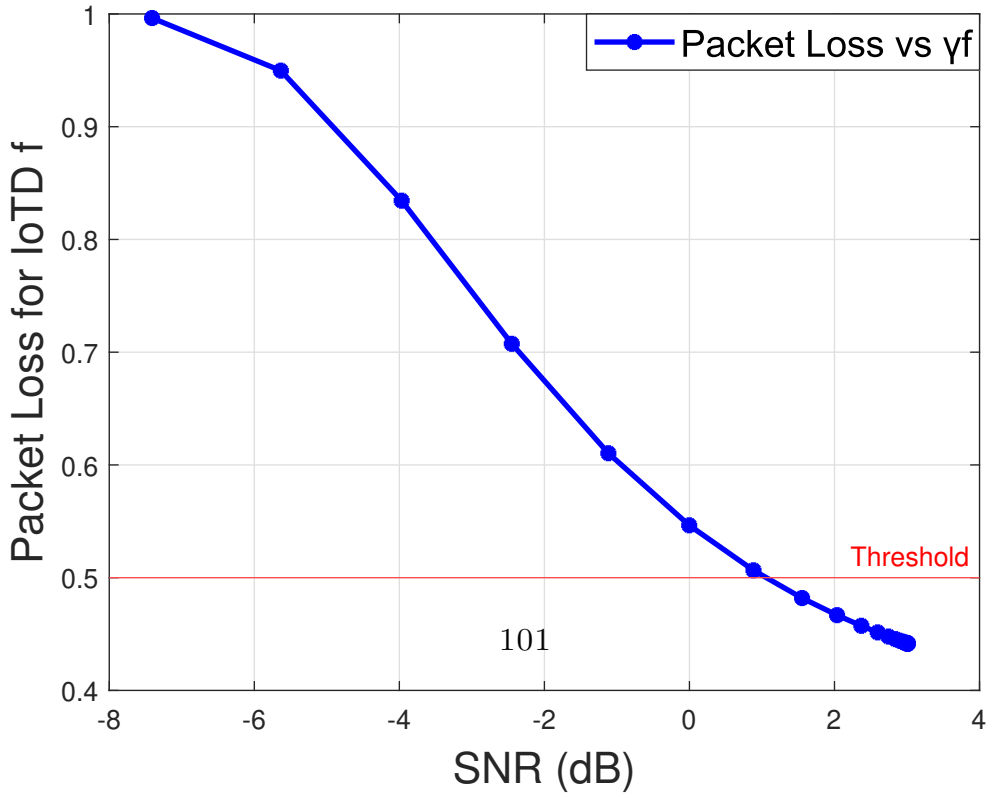
(b) Harvested energy for IoTD f vs γ_f. $\beta_n = \beta_f = 0.5$. $Pt = -15$ to 15 dbm.

Figure 5.7: Harvested energy for IoTD devices n and f for various α.

Packet Loss Rate



(a) Packet loss for IoTD n vs γ_n. $\alpha = 0.2, \beta_n = \beta_f = 0.5$. $Pt = -25$ dbm to 30 dbm.



(b) Packet loss for IoTD f vs γ_f. $\alpha = 0.2, \beta_n = \beta_f = 0.5$. $Pt = -20$ dbm to 30 dbm.

Figure 5.8: Packet loss for IoTDs n and f.

In Fig. 5.8, we analyze the network layer packet loss for the legitimate IoTDs in our system as a function of their respective SINRs, γ_n and γ_f. The packet loss represents the rate at which packets are lost or dropped during transmission, impacting the quality and reliability of IoT communication. The results demonstrate that as γ_n and γ_f increase, the packet loss in our system for the IoTDs (P_{l_n} and P_{l_f}) decreases accordingly. This indicates that higher SINRs result in lower packet loss rates, enhancing the reliability of communication for legitimate IoTDs.

In Fig. 5.8a, the packet loss P_{l_n} initially exceeds the threshold ϵ_n at $\gamma_n \leq 16.7344$ dB, indicating a high packet loss rate. However, as γ_n increases, the packet loss gradually decreases until it reaches 0 (i.e., no packet loss) at $\gamma_n = 28.7344$ dB. This implies that achieving higher SINRs significantly reduces the packet loss for IoTD n and improves the reliability of its communication. Similarly, in Fig. 5.8b, it is observed that there is packet loss P_{l_f} for $\gamma_f < 1.02312$ dB. However, for SINRs $\gamma_f > 1.02312$ dB, all the packet loss values are below the threshold ϵ_f, indicating that there is no packet loss in the system. This demonstrates that higher SINRs for IoTD f result in more reliable and robust communication with negligible packet loss.

5.8.3 Grid-based Search Results

Within this section, we employ a grid-based search methodology to serve as a benchmark for evaluating the outcomes produced by the DDPG algorithm. Utilizing the grid-based search approach depicted in Fig.5.9, we address the optimization problem in (5.58). The process begins by configuring the system parameters, followed by establishing a search space and incremental steps for each optimization parameter. Specifically, the search space for Φ_M spans from 0 to $2 \times \pi$ with an incremental step of $2 \times \pi/72$, for α it ranges from 0 to 1 with increments of 0.01, and for β_n and β_n, the range is 0 to 1 with increments of 0.1. Performance metrics including R_s, E_n, E_f, P_{l_n}, and P_{l_f} are computed for each set of discretized optimization parameters. Subsequently, these performance metric values are compared against predefined thresholds. Sets of optimization parameters that satisfy all constraints are recorded, forming a substantial pool of such sets. From this collection, the set generating the highest R_s is selected, accompanied by its corresponding parameters E_n, E_f, P_{l_n}, and P_{l_f}.

The grid-based search methodology systematically explores all feasible combinations of Φ_M, β_n, β_f, and α within the designated search space to identify the optimal solution. It is essential to acknowledge that although the grid-based search guarantees locating the optimal solution within the provided search space, its applicability to our system is limited due to its high complexity, especially as the number of IRS elements M or users increases. Consequently, the DDPG algorithm is adopted to efficiently identify satisfactory solutions in such intricate scenarios.

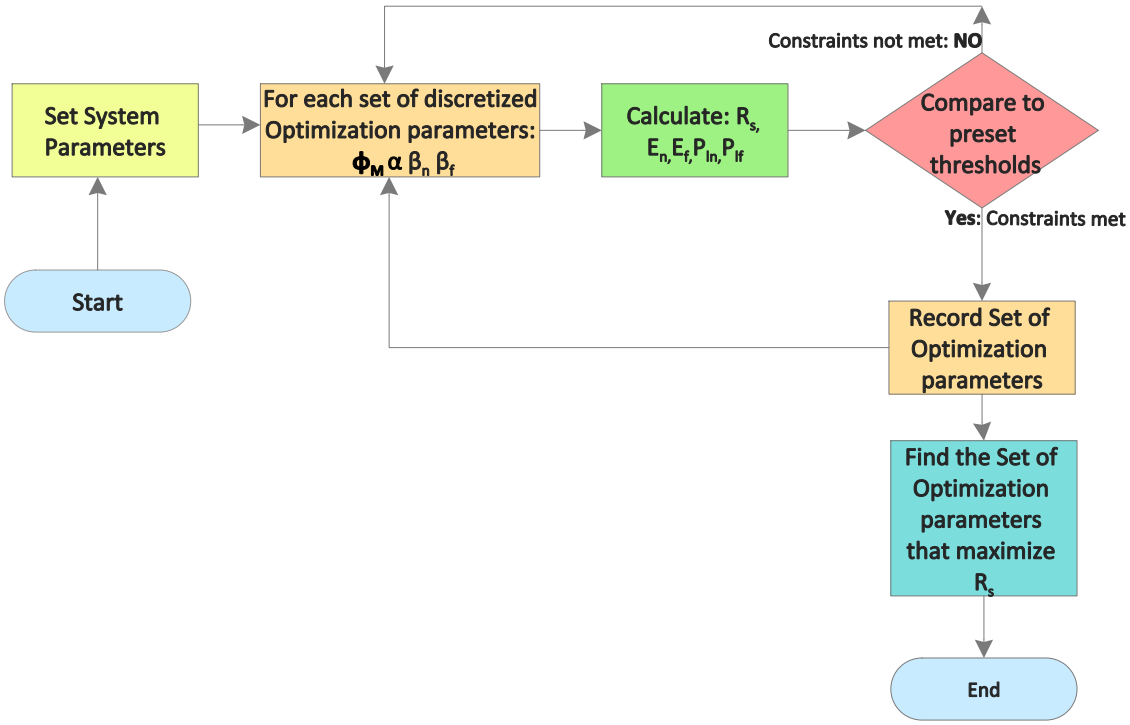


Figure 5.9: Flow chart of the grid-based simulation.

5.8.4 DRL Results

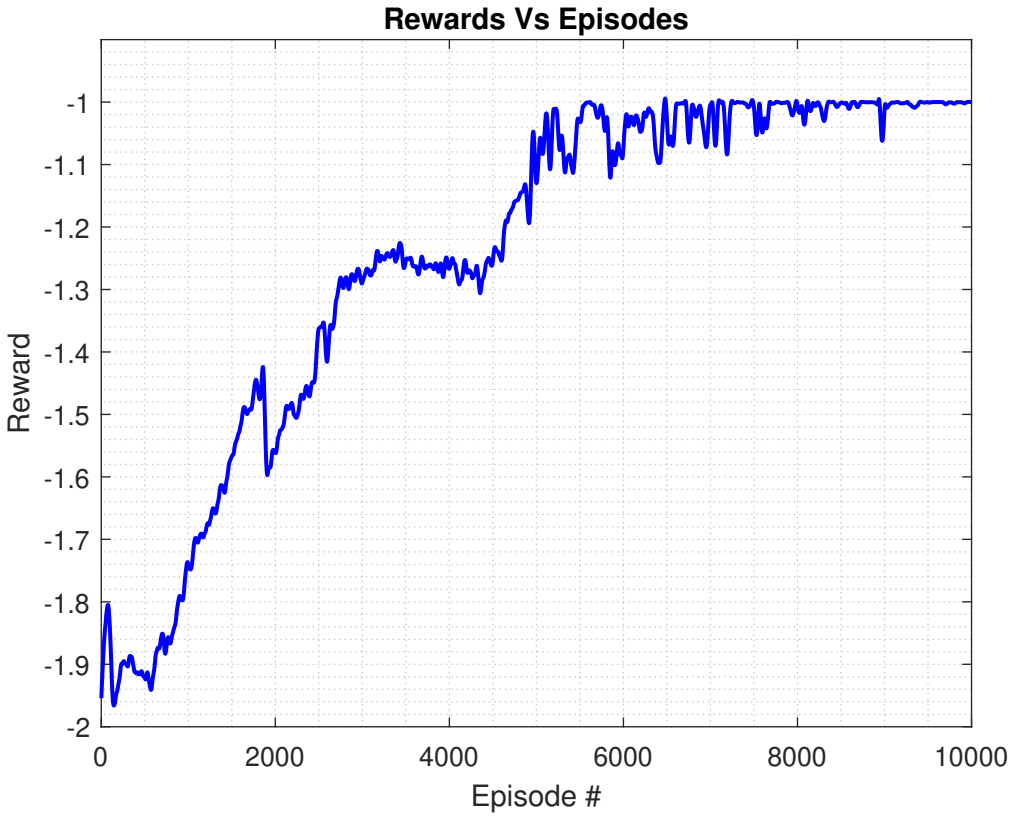


Figure 5.10: Convergence of the DDPG algorithm. Rewards vs. episodes.

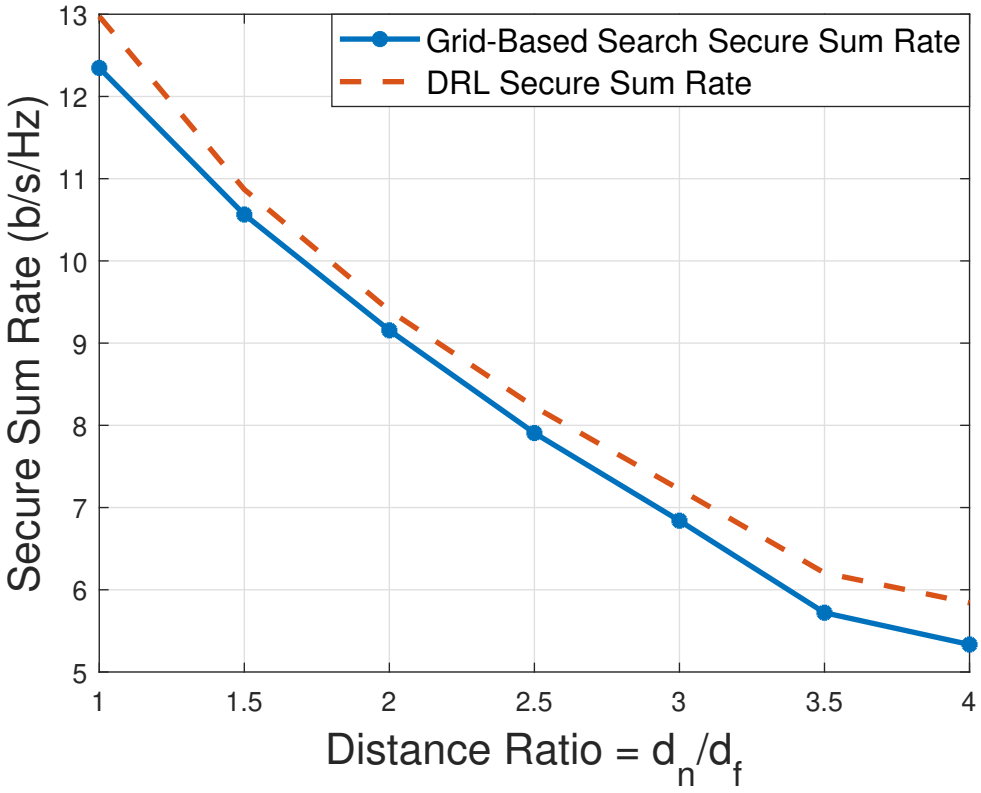


Figure 5.11: Secure sum rate for DRL versus grid-based search. $d_t = 2$, $d_n = 1$ m to 4m, $d_f = 4$m, $d_e = 6$m.

In our numerical assessment of the proposed DDPG scheme for IRS downlink NOMA-SWIPT systems, we aim to obtain the optimal solution for the non-convex optimization problem described in (5.58). To accomplish this, we employ the DDPG algorithm to find the optimum values of the IRS phase shift matrix $\boldsymbol{\Phi}_{\boldsymbol{M}}$, NOMA power allocation factor α, and EH factors β_n and β_f that maximize the secrecy sum-rate R_s while satisfying the constraints of IRS constant modulus, EH thresholds, and packet loss thresholds.

The suggested DDPG approach utilizes a two-network actor-critic architecture, incorporating an actor-network and a critic -network. Each network comprises four nn.Linear layers. The actor-network receives states as input and produces actions as output. It has two hidden layers, each of which has 128 neurons. ReLU (Rectified Linear Unit) activation function is employed in the hidden layers of both the actor-network and the critic-network. The output layer of the actor-network utilizes the $\tanh(\cdot)$ function, which provides higher gradients and facilitates more substantial weight updates. Additionally, the tanh function's output is symmetric, aiding in faster convergence. Conversely, the critic-network takes both states and

actions as input and estimates the expected future reward, commonly referred to as the Q-value. The Q-value represents the expected cumulative reward achievable by the agent when following a specific policy in a given state. The critic-network is composed of two hidden layers, each containing 128 neurons, and utilizes the ReLU activation function. The Adam optimizer is used to update the weights of both the actor and critic-networks. The Adam optimizer is a popular optimization algorithm that adapts the learning rate dynamically based on the gradients' characteristics. The reported results are obtained by averaging over 1000 iterations. The results obtained from the DDPG algorithm demonstrate the convergence of our scheme, as depicted in Fig. 5.10. The plot shows that the rewards increase over time, indicating the successful learning process of the DDPG algorithm. Moreover, Fig. 5.11 illustrates a comparison between the secure sum rates achieved using the DDPG algorithm and those obtained via grid-based search. The findings indicate a marginal superiority of the secure sum rate obtained through DDPG in comparison to that of a grid-based search. This outcome can be attributed to the dissimilar granularity between the two approaches, with the DDPG employing a continuous scheme, while the grid-based search involves steps of 0.1 (ranging from 0 to 1) in its search process.

Furthermore, Table 5.2 presents the optimized values for energy harvesting factors β_n and β_f, and power allocation factor α obtained using the DDPG algorithm. These values are derived for a specific combination of distances (d_t, d_n, d_f, and d_e) that maximizes the total secure rate for IoT devices while ensuring energy harvesting levels (E_n and E_f) above the thresholds for both near and far NOMA IoTDs,

Table 5.2: Optimized values for the optimization parameters obtained through DRL for $Pt = 1$ mw.

d_t	d_n	d_f	d_e	α	β_n	β_f	R_s	E_n	E_f	P_{l_n}	P_{l_f}
2	1	4	6	0.3399	0.971	0.936	12.9751	0.1504	0.0017	$1.1115e-05$	0.4531
2	1	6	7	0.1	0.37902	0.806	10.4798	3.0979	0.0015	$2.9357e-04$	0.3338
2	2	5	7	0.2677	0.4745	0.8854	8.6773	0.3181	0.0017	0.0050	0.4283
2	2	6	8	0.24989	0.3719	0.7844	7.8646	0.2824	0.0015	0.0140	0.4350
3	2	5	7	0.1963	0.4293	0.5845	6.1537	0.0733	0.0015	0.0239	0.4457

Table 5.3: Optimized values for the optimization parameters obtained through grid-based search for $Pt = 1$ mw.

d_t	d_n	d_f	d_e	α	β_n	β_f	R_s	E_n	E_f	P_{l_n}	P_{l_f}
2	1	4	6	0.28	0.9	0.9	12.3430	0.3996	0.0024	$1.6765e-05$	0.43
2	1	6	7	0.1	0.3	0.8	10.1718	3.5079	0.0013	$4.0114e-04$	0.3356
2	2	5	7	0.25	0.5	0.9	8.6945	0.2979	0.0014	0.0052	0.4202
2	2	6	8	0.27	0.3	0.8	7.8366	0.3555	0.0014	0.0155	0.4427
3	2	5	7	0.19	0.5	0.6	6.2590	0.0612	0.0013	0.0301	0.4417

while also minimizing packet loss (P_{l_n}, and P_{l_f}) for both IoTDs. These values are obtained through the DDPG algorithm while satisfying the constraints defined in the optimization problem (5.58). The DDPG algorithm allows us to find the optimal solution by iteratively improving the values of the optimization parameters. By utilizing the rewards and feedback from the IoTDs, the algorithm adapts and adjusts the parameters $\mathbf{\Phi_M}$, β_n, β_f, and α to achieve the highest secure sum-rate, provide energy to the IoTDs, and minimize their packet loss. The optimized values obtained from the DDPG algorithm provide insights into the optimal configuration for maximizing the performance of the system in terms of secure communication and energy transfer. Table 5.3 provides the optimized values of β_n, β_f, and α for various combinations of distances (d_t, d_n, d_f, and d_e) obtained through the discretized grid-based search scheme. These optimized values aim to maximize the total transmission sum rate for legitimate IoT devices. The main objective of this table is to compare the secure sum rates, energy harvesting levels, and packet loss rates achieved by the DDPG algorithm with those obtained through the ES method. This comparison serves to demonstrate the performance of our DDPG scheme, highlighting its close resemblance to the grid-based search method.

By utilizing both the DDPG algorithm and the grid-based search scheme, we determine the optimal parameters that yield an approximation of the maximum secure sum rate while adhering to the constraints outlined in the optimization problem (5.58). This comparative analysis allows us to evaluate the effectiveness and performance of our DDPG scheme, providing evidence of its ability to closely approximate the results obtained through the grid-based search method.

Chapter 6

Conclusion and Future Prospects

In this thesis, we explored the application of ML techniques in IRS and NOMA architectures for 5G/B5G wireless communication systems. The main objectives were to develop ML-based approaches for optimizing the performance of these systems, including maximizing the total data rate, achieving target secrecy rates, and enhancing energy transfer capabilities.

Firstly, we concentrated on the downlink scenario of an IRS NOMA system, to increase the total data rate of NOMA users under limited CSI availability. We formulated the problem as a non-convex optimization problem as a result of the constant modulus restriction and non-convex objective function. To address this complex problem, we employed DRL techniques, specifically the DDPG algorithm. We demonstrated that our proposed DDPG-based approach achieved high sum-rates, surpassing OMA in terms of user count and transmit power. We also observed the impact of imperfect SIC and found that our DDPG algorithm was robust to imperfections and approached the upper bound of the achievable sum-rate.

Secondly, we investigated the uplink multiple access scenario in cascaded IRS systems for short-range communications in THz networks. The main objective was to maximize the data transmission rate of the intended user and the sum-rate for both users. We faced non-convex optimization problems in both objectives, making closed-form expressions intractable. To overcome these challenges, we introduced sub-optimal solutions as bounding benchmarks and employed the DDPG algorithm known for its ability to handle non-convex problems. Our DDPG-based approach achieved higher data rates compared to the sub-optimal solutions for the first goal and achieved comparable sum-rates to the discretized ES for the second goal. Furthermore, we emphasized the importance of considering channel correlation to improve the learning process and achieve higher data transmission rates.

Lastly, we focused on maximizing the secrecy rate of legitimate IoT nodes while enabling energy transfer in an IRS-aided downlink secure NOMA-SWIPT scenario. The problem of optimization involved determining optimal values for NOMA power allocation parameters, IRS phase shift matrix, EH factors, and packet loss error

rate. Given the non-convex nature of the problem, we utilized the DDPG technique to search for the optimal solution. The obtained results demonstrated the effectiveness of our DDPG-based optimization approach in achieving improved secrecy rates and enhanced energy transfer capabilities in the IRS-assisted NOMA-SWIPT system. Moreover, the outcomes obtained using the DDPG algorithm closely align with those obtained through the ES method. This alignment demonstrates that the DDPG algorithm can effectively approximate the results achieved by the ES method, highlighting its capacity to handle the intricacies of the problem.

Overall, this thesis work highlighted the significance of ML, specifically DRL techniques, in tackling non-convex optimization problems and optimizing the performance of IRS and NOMA architectures in 5G and beyond wireless communication systems. By leveraging ML, we successfully achieved superior performance under more realistic system models. We showcase the added value of these technologies in the context of IRS NOMA architectures.

6.1 Future Directions

Future research on IRS NOMA networks has several promising directions. Among the most important potential areas are:

- **Advanced resource allocation algorithms**: An important area of research is the development of sophisticated resource allocation algorithms that optimize the performance of IRS NOMA networks. This includes looking into new power allocation schemes, phase shift optimization algorithms, and user grouping techniques to enhance both the system's sum-rate and energy efficiency.

- **Security and privacy within networks that incorporate IRS NOMA**: It is vital to investigate security and privacy elements within networks that incorporate IRS NOMA to ensure secure and private communication. Future studies could look at techniques for mitigating security threats in systems that utilize the IRS to assist NOMA, such as eavesdropping and jamming assaults. Furthermore, in such networks, privacy-preserving algorithms and authentication processes can be implemented to secure user information.

- **Interference management and beamforming:** Interference management in IRS NOMA networks is a key concern. Future research should concentrate on developing effective interference management techniques, like beamforming and interference alignment, to reduce interference generated by NOMA broadcasts and improve overall network performance.

- **Integration with various emerging technologies**: Investigating the integration of IRS NOMA networks with various emerging techniques can lead

to new research avenues. Investigating the use of IRS in conjunction with ML, artificial intelligence, or block-chain technology, for example, can lead to novel solutions for enhanced resource allocation, security, and scalability in wireless networks.

- **Real-world implementation and experimental validation**: Future research endeavors could focus on real-world implementations and experimental validations of IRS NOMA networks to gain practical insights into their viability and performance. The development of test-beds and prototypes is recommended to assess the effectiveness of IRS NOMA approaches across diverse environmental conditions and deployment scenarios. While implementing IRS in real-time NOMA systems poses challenges such as hardware complexity, cost, real-time signal processing, CSI acquisition, security and privacy concerns, and the need for dynamic adaptability, addressing these concerns through collaborative efforts is crucial for optimizing hardware, developing efficient algorithms, and ensuring regulatory compliance. Overcoming these challenges is essential for unlocking the full potential of IRS NOMA systems in enhancing wireless communication networks.

- **Dynamic Intersection Analysis: Identifying and Controlling the Crossing Point between Performance Curves in IRS NOMA and IRS OMA Systems**: Explore adaptive algorithms or optimization techniques to dynamically adjust system parameters, ensuring optimal performance at the intersection, examine the influence of various system configurations on the identified crossing point, investigate the impact of IRS placement, number of users, and transmit power levels on the dynamic behavior of NOMA and OMA systems, and explore the potential of machine learning or intelligent control mechanisms for real-time adjustment of system parameters.

- **Strategic Power Allocation Analysis in NOMA and OMA Scenarios with IRS**: Investigate the trade-offs between power allocation efficiency and spectral efficiency in both NOMA and OMA systems. Explore the impact of power control mechanisms on overall system performance, considering user fairness and energy efficiency.

- **Exploring Hybrid NOMA-OMA Systems: Performance Evaluation Compared to standalone NOMA and OMA Scenarios**: Investigate how a hybrid approach can leverage the benefits of both NOMA and OMA in different communication scenarios, and examine the potential for adaptive switching between NOMA and OMA based on real-time channel conditions or user requirements.

In summary, the future research agenda for IRS-assisted NOMA and OMA systems encompasses understanding their dynamic behavior, exploring adaptive

control mechanisms, and assessing trade-offs in power allocation and system configurations. Additionally, the exploration of hybrid NOMA-OMA systems presents a promising avenue for further investigation. Moreover, in the broader scope of IRS NOMA network research, addressing practical issues, enhancing system performance, improving security and privacy, and exploring integration with emerging technologies are crucial areas of focus. Investigating these topics will contribute to expanding our understanding and application of IRS NOMA networks in future wireless communication systems.

References

[1] Islam Abu Mahady et al. "Sum-Rate Maximization of NOMA Systems Under Imperfect Successive Interference Cancellation". In: *IEEE Communications Letters* 23.3 (2019), pp. 474–477. DOI: 10.1109/LCOMM.2019.2893195.

[2] Abdullateef Almohamad et al. "Performance of THz Communications Over Cascaded RISs: A Practical Solution to the Over-Determined Formulation". In: *IEEE Communications Letters* 26.2 (2022), pp. 291–295. DOI: 10.1109/LCOMM.2021.3132655.

[3] Ahmed Badawy and Ahmed El Shafie. "Securing OFDM-Based NOMA SWIPT Systems". In: *IEEE Transactions on Vehicular Technology* 69.10 (2020), pp. 12343–12347. DOI: 10.1109/TVT.2020.3017570.

[4] Christina Chaccour et al. "Seven Defining Features of Terahertz (THz) Wireless Systems: A Fellowship of Communication and Sensing". In: *IEEE Communications Surveys and Tutorials* 24.2 (2022), pp. 967–993. DOI: 10.1109/COMST.2022.3143454.

[5] Kae Won Choi et al. "Simultaneous Wireless Information and Power Transfer (SWIPT) for Internet of Things: Novel Receiver Design and Experimental Validation". In: *IEEE Internet of Things Journal* 7.4 (2020), pp. 2996–3012. DOI: 10.1109/JIOT.2020.2964302.

[6] Linglong Dai et al. "A survey of non-orthogonal multiple access for 5G". In: *IEEE communications surveys & Tutorials* 20.3 (2018), pp. 2294–2323.

[7] Jay L Devore. *Probability and Statistics for Engineering and the Sciences*. Cengage Learning, 2015.

[8] Zhiguo Ding and H. Vincent Poor. "A Simple Design of IRS-NOMA Transmission". In: *IEEE Communications Letters* 24.5 (2020), pp. 1119–1123. DOI: 10.1109/LCOMM.2020.2974196.

[9] Zhiguo Ding et al. "On the Performance of Non-Orthogonal Multiple Access in 5G Systems with Randomly Deployed Users". In: *IEEE Signal Processing Letters* 21.12 (2014), pp. 1501–1505. DOI: 10.1109/LSP.2014.2343971.

[10] Jean-Pierre Ebert and Andreas Willig. "A Gilbert-Elliot Bit Error Model and the Efficient Use in Packet Level Simulation". In: 1999.

[11] Fang Fang et al. "Energy-Efficient Design of IRS-NOMA Networks". In: *IEEE Transactions on Vehicular Technology* 69.11 (2020), pp. 14088–14092. DOI: 10.1109/TVT.2020.3024005.

[12] Keming Feng et al. "Deep Reinforcement Learning Based Intelligent Reflecting Surface Optimization for MISO Communication Systems". In: *IEEE Wireless Communications Letters* 9.5 (2020), pp. 745–749. DOI: 10.1109/LWC.2020.2969167.

[13] Yunpeng Feng et al. "Max-Min Fair Beamforming for IRS-Aided Secure NOMA Systems". In: *IEEE Communications Letters* 26.2 (2022), pp. 234–238. DOI: 10.1109/LCOMM.2021.3131350.

[14] Vincent Franois-Lavet et al. "An introduction to deep reinforcement learning". In: *Foundations and Trends® in Machine Learning* 11.3-4 (2018), pp. 219–354.

[15] Qiling Gao et al. "Joint Location and Beamforming Design for STAR-RIS Assisted NOMA Systems". In: *IEEE Transactions on Communications* (2023), pp. 1–1. DOI: 10.1109/TCOMM.2023.3247753.

[16] Ying Gao et al. "Rate-Splitting Multiple Access for Intelligent Reflecting Surface-Aided Secure Transmission". In: *IEEE Communications Letters* 27.2 (2023), pp. 482–486. DOI: 10.1109/LCOMM.2022.3224499.

[17] Ibrahim Ghareeb and Deemah Tashman. "Statistical Analysis of Cascaded Rician Fading Channels". In: *International Journal of Electronics Letters* 8 (Nov. 2018). DOI: 10.1080/21681724.2018.1545925.

[18] Caihong Gong et al. "Intelligent Reflecting Surface Aided Secure Communications for NOMA Networks". In: *IEEE Transactions on Vehicular Technology* 71.3 (2022), pp. 2761–2773. DOI: 10.1109/TVT.2021.3129075.

[19] Shimin Gong et al. "Toward Smart Wireless Communications via Intelligent Reflecting Surfaces: A Contemporary Survey". In: *IEEE Communications Surveys & Tutorials* 22.4 (2020), pp. 2283–2314. DOI: 10.1109/COMST.2020.3004197.

[20] Chong Han et al. "Terahertz Wireless Channels: A Holistic Survey on Measurement, Modeling, and Analysis". In: *IEEE Communications Surveys and Tutorials* 24.3 (2022), pp. 1670–1707. DOI: 10.1109/COMST.2022.3182539.

[21] Hui Han et al. "IRS-Aided Secure NOMA Networks Against Internal and External Eavesdropping". In: *IEEE Transactions on Communications* 70.11 (2022), pp. 7536–7548. DOI: 10.1109/TCOMM.2022.3208341.

[22] Gerhard Haßlinger and Oliver Hohlfeld. "The Gilbert-Elliott Model for Packet Loss in Real Time Services on the Internet." In: Jan. 2008, pp. 269–286.

[23] Rolando Herrero. "Analysis of IoT Mechanisms for Media Streaming". In: *Internet of Things* 9 (Mar. 2020), p. 100168. DOI: 10.1016/j.iot.2020.100168.

[24] Xiaoling Hu, Rui Zhang, and Caijun Zhong. "Semi-Passive Elements Assisted Channel Estimation for Intelligent Reflecting Surface-Aided Communications". In: *IEEE Transactions on Wireless Communications* 21.2 (2022), pp. 1132–1142. DOI: 10.1109/TWC.2021.3102446.

[25] Chongwen Huang, Ronghong Mo, and Chau Yuen. "Reconfigurable Intelligent Surface Assisted Multiuser MISO Systems Exploiting Deep Reinforcement Learning". In: *IEEE Journal on Selected Areas in Communications* 38.8 (2020), pp. 1839–1850. DOI: 10.1109/JSAC.2020.3000835.

[26] Chongwen Huang et al. "Hybrid Beamforming for RIS-Empowered Multi-hop Terahertz Communications: A DRL-based Method". In: *2020 IEEE Globecom Workshops (GC Wkshps*. 2020, pp. 1–6. DOI: 10.1109/GCWkshps50303.2020.9367503.

[27] Chongwen Huang et al. "Multi-Hop RIS-Empowered Terahertz Communications: A DRL-Based Hybrid Beamforming Design". In: *IEEE Journal on Selected Areas in Communications* 39.6 (2021), pp. 1663–1677. DOI: 10.1109/JSAC.2021.3071836.

[28] S. M. Riazul Islam et al. "Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges". In: *IEEE Communications Surveys & Tutorials* 19.2 (2017), pp. 721–742. DOI: 10.1109/COMST.2016.2621116.

[29] Tewelgn Kebede et al. "Multi-Carrier Waveforms and Multiple Access Strategies in Wireless Networks: Performance, Applications, and Challenges". In: *IEEE Access* 10 (2022), pp. 21120–21140. DOI: 10.1109/ACCESS.2022.3151360.

[30] Walter Keller-Gehrig. "Fast algorithms for the characteristics polynomial". In: *Theoretical computer science* 36 (1985), pp. 309–317.

[31] Joonas Kokkoniemi, Janne Lehtomäki, and Markku Juntti. "Simplified molecular absorption loss model for 275–400 gigahertz frequency band". In: *12th European Conference on Antennas and Propagation (EuCAP 2018)*. 2018, pp. 1–5. DOI: 10.1049/cp.2018.0446.

[32] Pierre-Simon Laplace. "Sur les approximations des formules qui sont fonctions de tres grands nombres et sur leur application aux probabilites". In: *Œuvres complètes* 12 (1810), pp. 301–345.

[33] Gen Li et al. "Latency Minimization for IRS-Aided NOMA MEC Systems with WPT-enabled IoT Devices". In: *IEEE Internet of Things Journal* (2023), pp. 1–1. DOI: 10.1109/JIOT.2023.3240395.

[34] Xingwang Li et al. “Enhancing Secrecy Performance for STAR-RIS NOMA Networks”. In: *IEEE Transactions on Vehicular Technology* 72.2 (2023), pp. 2684–2688. DOI: 10.1109/TVT.2022.3213334.

[35] Xingwang Li et al. “UAV-Aided Multi-Way NOMA Networks With Residual Hardware Impairments”. In: *IEEE Wireless Communications Letters* 9.9 (2020), pp. 1538–1542. DOI: 10.1109/LWC.2020.2996782.

[36] Zehao Liu et al. “Sum Rate Maximization for NOMA-based VLC with Optical Intelligent Reflecting Surface”. In: *IEEE Wireless Communications Letters* (2023), pp. 1–1. DOI: 10.1109/LWC.2023.3246064.

[37] Xiao Lu et al. “Wireless Networks With RF Energy Harvesting: A Contemporary Survey”. In: *IEEE Communications Surveys and Tutorials* 17.2 (2015), pp. 757–789. DOI: 10.1109/COMST.2014.2368999.

[38] Sun Mao et al. “Intelligent Reflecting Surface-Assisted Low-Latency Federated Learning Over Wireless Networks”. In: *IEEE Internet of Things Journal* 10.2 (2023), pp. 1223–1235. DOI: 10.1109/JIOT.2022.3204637.

[39] Weidong Mei and Rui Zhang. “Multi-Beam Multi-Hop Routing for Intelligent Reflecting Surfaces Aided Massive MIMO”. In: *IEEE Transactions on Wireless Communications* 21.3 (2022), pp. 1897–1912. DOI: 10.1109/TWC.2021.3108020.

[40] Sandeep Mukherjee et al. “Analytical calculation of Rician K-factor for indoor wireless channel models”. In: *IEEE Access* 5 (2017), pp. 19194–19212.

[41] The Vi Nguyen et al. “Achievable Rate Analysis of Two-Hop Interference Channel With Coordinated IRS Relay”. In: *IEEE Transactions on Wireless Communications* 21.9 (2022), pp. 7055–7071. DOI: 10.1109/TWC.2022.3154372.

[42] Konstantinos Ntontin et al. “Reconfigurable Intelligent Surface Optimal Placement in Millimeter-Wave Networks”. In: *IEEE Open Journal of the Communications Society* 2 (2021), pp. 704–718. DOI: 10.1109/OJCOMS.2021.3068790.

[43] Aditya Raosaheb Pawar, Salil Kashyap, and Sonali Chouhan. “Error Probability Analysis of Intelligent Reflecting Surface-Enabled Non-Orthogonal Multiple Access Systems”. In: *IEEE Transactions on Vehicular Technology* (2023), pp. 1–14. DOI: 10.1109/TVT.2023.3252083.

[44] Yue Qi and Mojtaba Vaezi. “IRS-Assisted Physical Layer Security in MIMO-NOMA Networks”. In: *IEEE Communications Letters* (2023), pp. 1–1. DOI: 10.1109/LCOMM.2023.3235722.

[45] J.A. Roberts and J.M. Bargallo. “DPSK performance for indoor wireless Rician fading channels”. In: *IEEE Transactions on Communications* 42.234 (1994), pp. 592–596. DOI: 10.1109/TCOMM.1994.577086.

[46] Debbarni Sarkar et al. “Intelligent Reflecting Surface Aided NOMA-HARQ Based IoT Framework for Future Wireless Networks”. In: *IEEE Transactions on Vehicular Technology* (2023), pp. 1–15. DOI: 10.1109/TVT.2022.3233090.

[47] Muhammad Shehab et al. “DDPG Performance in THz Communications over Cascaded RISs: A Machine Learning Solution to the Over-Determined System”. In: *2023 International Wireless Communications and Mobile Computing (IWCMC)*. 2023, pp. 210–215. DOI: 10.1109/IWCMC58020.2023.10182861.

[48] Muhammad Shehab et al. “Deep Reinforcement Learning Powered IRS-Assisted Downlink NOMA”. In: *IEEE Open Journal of the Communications Society* 3 (2022), pp. 729–739. DOI: 10.1109/OJCOMS.2022.3165590.

[49] Muhammad Shehab et al. “Terahertz Multiple Access: A Deep Reinforcement Learning Controlled Multihop IRS Topology”. In: *IEEE Open Journal of the Communications Society* (2024), pp. 1–1. DOI: 10.1109/OJCOMS.2024.3357701.

[50] Muhammad Shehab et al. “The Role of 5G/6G Networks in Building Sustainable and Energy-Efficient Smart Cities”. In: *2022 IEEE 7th International Energy Conference (ENERGYCON)*. 2022, pp. 1–7. DOI: 10.1109/ENERGYCON53164.2022.9830364.

[51] Muhammad J. Shehab et al. “5G Networks Towards Smart and Sustainable Cities: A Review of Recent Developments, Applications and Future Perspectives”. In: *IEEE Access* 10 (2022), pp. 2987–3006. DOI: 10.1109/ACCESS.2021.3139436.

[52] Qiang Sun et al. “Ergodic Rate Analysis and IRS Configuration for Multi-IRS Dual-Hop DF Relaying Systems”. In: *IEEE Communications Letters* 25.10 (2021), pp. 3224–3228. DOI: 10.1109/LCOMM.2021.3100347.

[53] Abdelrahman Taha, Muhammad Alrabeiah, and Ahmed Alkhateeb. “Deep Learning for Large Intelligent Surfaces in Millimeter Wave and Massive MIMO Systems”. In: *2019 IEEE Global Communications Conference (GLOBECOM)*. 2019, pp. 1–6. DOI: 10.1109/GLOBECOM38437.2019.9013256.

[54] Abdelrahman Taha et al. “Deep Reinforcement Learning for Intelligent Reflecting Surfaces: Towards Standalone Operation”. In: *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. 2020, pp. 1–5. DOI: 10.1109/SPAWC48557.2020.9154301.

[55] Wei Tan et al. “Secure Transmission via IUI Engineering for IRS-Assisted NOMA Systems”. In: *IEEE Wireless Communications Letters* 11.7 (2022), pp. 1369–1373. DOI: 10.1109/LWC.2022.3169350.

[56] George E Uhlenbeck and Leonard S Ornstein. “On the theory of the Brownian motion”. In: *Physical review* 36.5 (1930), p. 823.

[57] Vinita Vasudevan and M Ramakrishna. "A hierarchical singular value decomposition algorithm for low rank matrices". In: *arXiv preprint arXiv:1710.02812* (2017).

[58] Tianqi Wang, Fang Fang, and Zhiguo Ding. "Joint Phase Shift and Beamforming Design in a Multi-User MISO STAR-RIS Assisted Downlink NOMA Network". In: *IEEE Transactions on Vehicular Technology* (2023), pp. 1–12. DOI: 10.1109/TVT.2023.3248789.

[59] Wei Wang et al. "Beamforming and Jamming Optimization for IRS-Aided Secure NOMA Networks". In: *IEEE Transactions on Wireless Communications* 21.3 (2022), pp. 1557–1569. DOI: 10.1109/TWC.2021.3104856.

[60] Wei Wang et al. "Secure Beamforming for IRS-Enhanced NOMA Networks". In: *IEEE Wireless Communications* 30.1 (2023), pp. 134–140. DOI: 10.1109/MWC.012.2100639.

[61] Mingjiang Wu et al. "RIS-Assisted Energy- and Spectrum-Efficient Symbiotic Transmission in NOMA Systems". In: *IEEE Transactions on Communications* (2023), pp. 1–1. DOI: 10.1109/TCOMM.2023.3241355.

[62] Qingqing Wu and Rui Zhang. "Intelligent Reflecting Surface Enhanced Wireless Network: Joint Active and Passive Beamforming Design". In: *2018 IEEE Global Communications Conference (GLOBECOM)*. 2018, pp. 1–6. DOI: 10.1109/GLOCOM.2018.8647620.

[63] Zhen Xie et al. "Resource Allocation for Double IRSs Assisted Wireless Powered NOMA Networks". In: *IEEE Wireless Communications Letters* (2023), pp. 1–1. DOI: 10.1109/LWC.2023.3244997.

[64] Tingzhen Xu et al. "Pilot-based channel estimation design in covert wireless communication". In: *arXiv preprint arXiv:1908.00226* (2019).

[65] Gang Yang, Xinyue Xu, and Ying-Chang Liang. "Intelligent Reflecting Surface Assisted Non-Orthogonal Multiple Access". In: *2020 IEEE Wireless Communications and Networking Conference (WCNC)*. 2020, pp. 1–6. DOI: 10.1109/WCNC45663.2020.9120476.

[66] Helin Yang et al. "Deep Reinforcement Learning-Based Intelligent Reflecting Surface for Secure Wireless Communications". In: *IEEE Transactions on Wireless Communications* 20.1 (2021), pp. 375–388. DOI: 10.1109/TWC.2020.3024860.

[67] Zikun Yang, Feng Li, and Dou Zhang. "A Joint Model Extraction and Data Detection Framework for IRS-NOMA System". In: *IEEE Transactions on Signal Processing* 71 (2023), pp. 164–177. DOI: 10.1109/TSP.2023.3241781.

[68] Lei Yuan et al. "Performance Analysis of IRS-Aided Short-Packet NOMA Systems Over Nakagami-m Fading Channels". In: *IEEE Transactions on Vehicular Technology* (2023), pp. 1–6. DOI: 10.1109/TVT.2023.3243998.

[69] Xinwei Yue et al. "Simultaneously Transmitting and Reflecting Reconfigurable Intelligent Surface Assisted NOMA Networks". In: *IEEE Transactions on Wireless Communications* 22.1 (2023), pp. 189–204. DOI: 10.1109/TWC.2022.3192211.

[70] Ming Zeng et al. "Sum Rate Maximization for IRS-Assisted Uplink NOMA". In: *IEEE Communications Letters* 25.1 (2021), pp. 234–238. DOI: 10.1109/LCOMM.2020.3025978.

[71] Zepeng Zhang and Ziping Zhao. "Weighted Sum-Rate Maximization for Multi-Hop RIS-Aided Multi-User Communications: A Minorization-Maximization Approach". In: *2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. 2021, pp. 106–110. DOI: 10.1109/SPAWC51858.2021.9593114.

[72] Zheng Zhang et al. "Robust and Secure Communications in Intelligent Reflecting Surface Assisted NOMA Networks". In: *IEEE Communications Letters* 25.3 (2021), pp. 739–743. DOI: 10.1109/LCOMM.2020.3039811.

[73] Zheng Zhang et al. "Securing NOMA Networks by Exploiting Intelligent Reflecting Surface". In: *IEEE Transactions on Communications* 70.2 (2022), pp. 1096–1111. DOI: 10.1109/TCOMM.2021.3126636.

[74] Beixiong Zheng, Changsheng You, and Rui Zhang. "Efficient Channel Estimation for Double-IRS Aided Multi-User MIMO System". In: *IEEE Transactions on Communications* 69.6 (2021), pp. 3818–3832. DOI: 10.1109/TCOMM.2021.3064947.

[75] Beixiong Zheng et al. "A Survey on Channel Estimation and Practical Passive Beamforming Design for Intelligent Reflecting Surface Aided Wireless Communications". In: *IEEE Communications Surveys & Tutorials* 24.2 (2022), pp. 1035–1071. DOI: 10.1109/COMST.2022.3155305.

[76] Ruikang Zhong et al. "STAR-RISs Assisted NOMA Networks: A Distributed Learning Approach". In: *IEEE Journal of Selected Topics in Signal Processing* 17.1 (2023), pp. 264–278. DOI: 10.1109/JSTSP.2022.3229567.

[77] Jing Zhu et al. "Index Modulation for STAR-RIS Assisted NOMA System". In: *IEEE Communications Letters* 27.2 (2023), pp. 716–720. DOI: 10.1109/LCOMM.2022.3223968.

[78] Jiakuo Zuo et al. "Intelligent Reflecting Surface Enhanced Millimeter-Wave NOMA Systems". In: *IEEE Communications Letters* 24.11 (2020), pp. 2632–2636. DOI: 10.1109/LCOMM.2020.3009158.

[79] Jiakuo Zuo et al. "Joint Design for Simultaneously Transmitting and Reflecting (STAR) RIS Assisted NOMA Systems". In: *IEEE Transactions on Wireless Communications* 22.1 (2023), pp. 611–626. DOI: 10.1109/TWC.2022.3197079.

Proof of Proposition 1 (See page 54)

The aggregate signal power received by each T_{x_κ} at R_x in (4.24) can be formulated as follows:

$$P_{Rx}^{(\kappa)} = |\sqrt{L_{\tau,\kappa}} e^{-j\eta_3} \mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{e,f}^H \mathbf{\Phi}_E \mathbf{h}_{t,\kappa}^H e^{-j\eta_\kappa}|^2 P_t.$$

Proof.

- Perform matrix multiplication between the conjugate transpose of the receiver channel matrix $\mathbf{h}_r^H$ and the phase shift reflection matrix of IRS_2 denoted as $\mathbf{\Phi}_F$:

$$\left[\mathbf{h}_r^H \mathbf{\Phi}_F\right] = \underbrace{\left[h_{r1}^*, \ldots, h_{rf}^*, \ldots, h_{rF}^*\right]}_{1\times F} \times \underbrace{\begin{bmatrix} e^{-j\omega_1} & 0 & \ldots & \ldots & \ldots & 0 \\ 0 & e^{-j\omega_2} & 0 & \ddots & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & e^{-j\omega_f} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & \ldots & 0 & e^{-j\omega_F} \end{bmatrix}}_{F\times F}$$

$$\left[\mathbf{h}_r^H \mathbf{\Phi}_F\right] = \underbrace{\left[h_{r1}^* e^{-j\omega_1}, \ldots, h_{rf}^* e^{-j\omega_f}, \ldots, h_{rF}^* e^{-j\omega_F}\right]}_{1\times F}$$

- Subsequently, multiply the outcome of the preceding operation by the matrix $\mathbf{H}_{e,f}^H$:

$$\left[\mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{ef}^H\right] = \underbrace{\left[h_{r1}^* e^{-j\omega_1}, \ldots, h_{rf}^* e^{-j\omega_f}, \ldots, h_{rF}^* e^{-j\omega_F}\right]}_{1\times F} \times \underbrace{\begin{bmatrix} h_{11}^* & h_{21}^* & \ldots & \ldots & \ldots & h_{E1}^* \\ h_{12}^* & h_{22}^* & \ddots & \ddots & \ddots & h_{E2}^* \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & h_{ef}^* & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ h_{1F}^* & \ldots & \ldots & \ldots & \ldots & h_{EF}^* \end{bmatrix}}_{F\times E}$$

$$\left[\mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{ef}^H\right] = \underbrace{\left[\sum_{f=1}^{F} h_{rf}^* e^{-j\omega_f} h_{1f}^*, \sum_{f=1}^{F} h_{rf}^* e^{-j\omega_f} h_{2f}^*, \ldots, \sum_{f=1}^{F} h_{rf}^* e^{-j\omega_f} h_{EF}^*\right]}_{1\times E}$$

- Multiply the product obtained from the previous step by the phase shift reflection matrix of IRS_1, denoted as $\mathbf{\Phi}_E$:

$$\left[\mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{ef}^H \mathbf{\Phi}_E\right] = \underbrace{\left[\mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{ef}^H\right]}_{1\times E} \times \underbrace{\begin{bmatrix} e^{-j\theta_1} & 0 & \ldots & \ldots & \ldots & 0 \\ 0 & e^{-j\theta_2} & 0 & \ddots & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & e^{-j\theta_e} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & \ldots & 0 & e^{-j\theta_E} \end{bmatrix}}_{E\times E}$$

$$\left[\mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{ef}^H \mathbf{\Phi}_E\right] = \underbrace{\left[\sum_{e=1}^{E} \sum_{f=1}^{F} h_{rf}^* e^{-j\omega_f} H_{ef}^* e^{-j\theta_e}\right]}_{1\times E}$$

- Multiply the outcome from the previous step by the conjugate transpose of the transmitter channel matrix of user κ, represented as $\mathbf{h}_{t,\kappa}^H$:

$$\left[\mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{ef}^H \mathbf{\Phi}_E \mathbf{h}_{t,\kappa}^H\right] = \underbrace{\left[\sum_{e=1}^{E} \sum_{f=1}^{F} h_{rf}^* e^{-j\omega_f} H_{ef}^* e^{-j\theta_e}\right]}_{1\times E} \times \underbrace{\begin{bmatrix} h_{t,\kappa,1}^* \\ h_{t,\kappa,2}^* \\ \vdots \\ h_{t,\kappa,E}^* \end{bmatrix}}_{E\times 1}$$

$$\mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{ef}^H \mathbf{\Phi}_E \mathbf{h}_{t,\kappa}^H = \sum_{e=1}^{E} \sum_{f=1}^{F} h_{rf}^* e^{-j\omega_f} H_{ef}^* e^{-j\theta_e} h_{t,\kappa e}^*,$$

$$\mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{e,f}^H \mathbf{\Phi}_E \mathbf{h}_{t,\kappa}^H = \sum_{e=1}^{E} \sum_{f=1}^{F} |h_{rf}| e^{-j\phi_{rf}} e^{-j\omega_f} |H_{ef}| e^{-j\phi_{ef}} e^{-j\theta_e} |h_{t,\kappa e}| e^{-j\phi_{t,\kappa e}},$$

$$\mathbf{h}_r^H \mathbf{\Phi}_N \mathbf{H}_{ef}^H \mathbf{\Phi}_E \mathbf{h}_{t,\kappa}^H = \sum_{e=1}^{E} \sum_{f=1}^{F} |h_{t,\kappa e}||H_{ef}||h_{rf}| e^{-j(\phi_{t,\kappa e} + e^{-j\theta_e} + e^{-j\phi_{ef}} + e^{-j\omega_f} + \phi_{rf})},$$

$$\mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{ef}^H \mathbf{\Phi}_E \mathbf{h}_{t,\kappa}^H = \sum_{e=1}^{E} \sum_{f=1}^{F} |h_{t,\kappa e}||\Lambda_e||H_{ef}||\Lambda_f||h_{rf}| e^{-j(\phi_{t,\kappa e} + \theta_e + \phi_{ef} + \omega_f + \phi_{rf})},$$

- Based on the derived outcome, we can conclude that the overall received signal power for each T_{x_κ} at R_x can be expressed in the following manner:

$$P_{Rx}^{(\kappa)} = \left|\sqrt{L_{\tau,\kappa}} e^{-j\eta_3} \mathbf{h}_r^H \mathbf{\Phi}_F \mathbf{H}_{ef}^H \mathbf{\Phi}_E \mathbf{h}_{t,\kappa}^H e^{-j\eta_\kappa}\right|^2 P_t,$$
$$= \left|\sqrt{L_{\tau,\kappa}} \sum_{e=1}^{E} \sum_{f=1}^{F} |h_{t,\kappa e}||\Lambda_e||H_{ef}||\Lambda_f||h_{rf}| e^{-j\left(\varphi_{t_{\kappa e}} + \theta_e + \varphi_{ef} + \omega_f + \varphi_{r_f} + \eta_\kappa + \eta_3\right)}\right|^2 P_t.$$

□

This Ph.D. thesis has been typeset by means of the TEX-system facilities. The typesetting engine was pdfLATEX. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi=scudo`. This class is available in every up-to-date and complete TEX-system installation.