Reflective Intelligence Surface Technology for Future Wireless Networks

by

A dissertation submitted to the School of Graduate Studies
in partial fulfillment of the requirements for the degree of

Master of Electrical Engineering

Faculty of Engineering and Applied Science
Memorial University of Newfoundland

August 2022

St. John's, Newfoundland

Abstract

Reconfigurable intelligent surfaces (RISs) have witnessed significant attention due to their potential to improve the efficiency and coverage of wireless networks. RIS acts as a smart mirror, which reconfigures the wireless propagation environment by tuning the incoming waveform's phase shift, amplitude, and polarization. To fully realize the capabilities of RIS, the phase shifts should be efficiently optimized. Researchers have considered optimization-based techniques to tackle the phase shift optimization problem. However, such methods are complex in nature and are difficult to realize for large-scale systems. To this end, deep reinforcement learning (DRL) has emerged as a robust and powerful approach for optimizing wireless communication systems. DRL learns from interacting with the environment without needing a labeled dataset, enabling adapting to the dynamic changes in the communication environment. In this work, we develop DRL frameworks to optimize full-duplex (FD) RIS-assisted communication systems. FD communications are envisioned as one of the essential technologies for future wireless communications. Incorporating RIS into FD systems can efficiently establish a reliable communication system and resolve the co-channel interference issue of FD systems.

To this end, this work first proposes a low-complexity DRL algorithm to optimize the RIS phase shifts of a half-duplex (HD)-FD RIS-assisted communication system. The proposed algorithm is the first of its kind, which tackles the optimization problem

in the FD operating mode. It was shown that the proposed algorithm significantly improves the rate compared to the non-optimized case in both operating modes and reduces the computational complexity compared to the state-of-the-art algorithm in the HD operating mode. Furthermore, the deployment of distributed RISs is also investigated in this thesis. In particular, the preference of deploying single or distributed RIS schemes is studied based on the links' quality considering three practical scenarios. The sum-rate maximization problem is considered subject to transmit beamformers and RIS phase shifts of a FD RIS-assisted communication system. To address the optimization problem, a two-step solution is proposed. First, a closed-form solution is derived to optimize the beamformers. Second, a DRL algorithm is proposed to optimize the RIS phase shifts. The proposed solution was shown to efficiently outperform the conventional beamformers approximation and improve the sum rate compared to the non-optimized RIS phase shifts. Finally, this work considers a DRL approach for optimizing the discrete phase shifts of FD distributed RIS-assisted system. The discrete phase shifts are considered to offer a feasible solution, since the continuous phase shifts are infeasible to implement due to hardware limitations. A deep Q-learning algorithm is developed to optimize the RIS phase shifts, along with two mathematical beamformers derivations (i.e., closed-form and approximate). The performance of the proposed algorithm is further assessed through extensive simulations by considering two scenarios: the presence of the line-of-sight (LoS) link and when it is blocked. It was shown that the proposed algorithm achieves promising results compared to the ideal approach (the continuous baseline), which guarantees a near-optimal performance. The complexity analysis for all proposed algorithms and simulation results are provided to support these findings.

Acknowledgments

I wish to express my most profound appreciation to all who helped me throughout this journey. Special gratitude goes to my supervisors who provided me the opportunity to exploit my academic capabilities; Prof. Octavia A. Dobre and Prof. Telex M. N. Ngatched. They were beyond academic supervisors; they helped from their heart as if I was the only student to mentor. Thank you for your valuable supervision, your support, and positive energy. Thank you for being there, even in the most unfortunate moments of this period. I also wish to express my sincerest thanks to my mentor, Dr. Ibrahim Al-Nahhal, for his endless support, guidance, and valuable efforts throughout my master's journey. Thank you, not only for the academic help, but also for genuine guidance beyond academics. I was lucky to be in Prof. Octavia's group and to work with you all. I would also like to thank all the group members for being a valuable part of this experience.

My most profound gratitude goes to my parents for pushing me to exceed my limits, their unconditional love, and support. I would also like to thank my sister for being proud of my strengths and for making me smile no matter what the situation is. I would further like to express my appreciation for my friends for being there all along. My deepest gratitude goes to Mohsen for the emotional support, encouragement, and empowerment he gave me when I needed it. Finally, I would like to acknowledge the financial support provided by my supervisors, the Faculty of Engineering and Applied

Science, the School of Graduate Studies, and the Natural Science and Engineering Research Council of Canada (NSERC). *Thank you all for everything.*

Co-Authorship Statement

I, Alice Faisal, hold the principal author status for all the manuscripts included in this thesis. However, all manuscripts in this work are coauthored by my supervisors Prof. Octavia A. Dobre and Prof. Telex M. N. Ngatched, in addition to my mentor Dr. Ibrahim Al-Nahhal. The list of the manuscripts included in this thesis are described as below.

- Paper 1 in Chapter 2: A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Deep reinforcement learning for optimizing RIS-assisted HD-FD wireless systems," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3893-3897, Dec. 2021.
- Paper 2 in Chapter 3: A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Deep reinforcement learning for RIS-assisted FD systems: Single or distributed RIS?" *IEEE Commun. Lett.*, Early Access, Apr. 2022.
- Paper 3 in Chapter 4: A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Distributed RIS-assisted FD systems with discrete phase shifts: A reinforcement learning approach," Accepted for presentation at *IEEE Global Commun. Conf. (GLOBECOM)*, May 2022.

I was the primary author of the above papers, with authors 2-4 contributing to the idea, its formulation and development, and refinement of the presentation.

Alice Faisal	Date

Table of Contents

List of Tables

List of Figures

List of Abbreviations

5G	Fifth-generation
6G	Sixth-generation
AO	Alternating optimization
bps/Hz	Bit per second per Hertz
BS	Base station
CCI	Co-channel interference
D2D	Device-to-device
DDPG	Deep deterministic policy gradient
DNN	Deep neural network
DQL	Deep Q-learning
DRL	Deep reinforcement learning
EM	Electromagnetic
FD	Full-duplex
FDD	Frequency division duplex

FF	Feed-forward
GRU	Gated recurrent unit
HAPS	High-altitude platform station
HD	Half-duplex
LoS	Line-of-sight
MIMO	Multiple-input multiple-output
MISO	Multiple-input single-output
PL	Path loss
RIS	Reconfigurable intelligent surface
SGD	Stochastic gradient descent
SI	Self-interference
SWIPT	Simultaneous wireless information and power transfer
Tbps	Tera bit per second
THz	Terahertz band
UAV	Unmanned aerial vehicle
UE	User equipment
UM-MIMO	Ultra massive MIMO

Chapter 1

Introduction

1.1 Motivation and Background

As the sixth-generation (6G) mobile communications are advancing towards universal standards, researchers are investigating the feasibility of deploying several future services to fulfill the demands of wireless communications. The rapid growth of connected devices has brought significant challenges to the capabilities of fifth-generation (5G) wireless systems. 6G wireless networks are envisioned to provide extreme data rates (peak data rates up to 1 tera bit per second (Tbps)), enhanced spectral efficiency and coverage (the peak spectral efficiency can be increased up to 60 bps per Hertz (bps/Hz)), wide bandwidths (100 times the bandwidths in 5G networks), ultra-low latency, and extremely high reliability to enable mission and safety-critical applications [1, 2, 3]. To support such demands, new evolutionary technologies are developed to support the next-generation of wireless communications. The main potential technologies include reconfigurable intelligent surfaces (RISs), massive multiple-input multiple-output (MIMO), cell-free massive MIMO, high-frequency based technologies (i.e., terahertz band (THz) and visible light communications), full-duplex

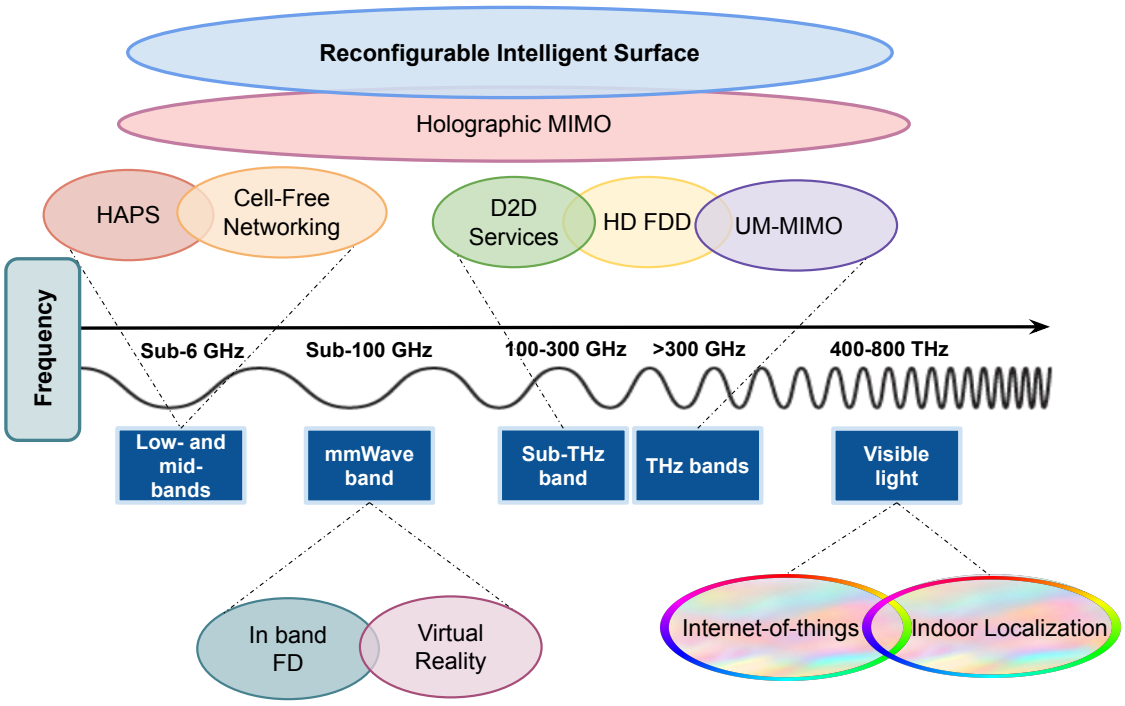


Fig. 1.1: 6G Spectrum decomposition and candidate applications. UM-MIMO and D2D denote ultra massive MIMO and device-to-device technologies, respectively.

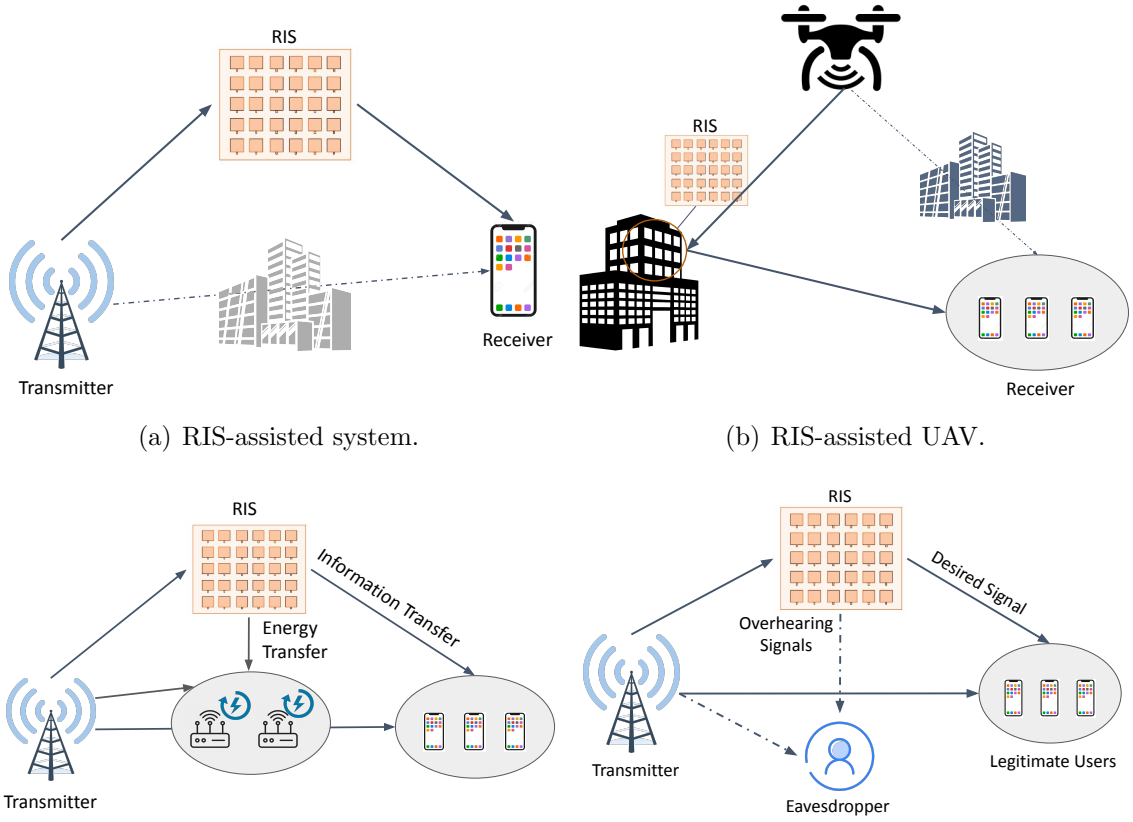
(FD) communication, and artificial intelligence [4, 1]. The spectrum decomposition, its corresponding candidate applications for 6G networks are illustrated in Fig. 1.1.

Recently, RIS has emerged as a promising future transmission technique in wireless communication systems. It consists of a two-dimensional array made up of low-cost, nearly passive electromagnetic (EM) elements. It overcomes the probabilistic nature of EM wave transmission. In particular, RISs provide an adaptive propagation environment by tuning phase shift, amplitude, and polarization of the incoming waveform. It enables controlling different characteristics of radio waves, such as scattering, reflection, and refraction, which effectively enhances the signal quality and boosts the wireless spectral efficiency by realizing a controllable environment. RISs

have other unique features, such as providing an inherently full-duplex transmission (not affected by receiver noise), having a full-band response, and operating with low power consumption while providing relatively high energy efficiency. Moreover, RISs are easy to deploy in diverse environments at a low-cost, which enables integrating them into different application scenarios. For their above unique features, RISs are envisioned as an important technology that plays a crucial role in 6G and beyond wireless communications.

Figure 1.2 illustrates prospective use cases of RIS in future wireless networks. In particular, RIS can be deployed to enhance the coverage and establish an improved connection between the transmitter and the receiver when the direct link is blocked. The RIS can also be deployed to enhance the physical layer security, where the reflected signals can be added constructively at the legitimate receiver to improve its reception power. The reflected signals can also be added destructively at the eavesdropper to degrade the quality of its reception, which can potentially yield a secure transmission. Furthermore, RIS can assist unmanned aerial vehicle (UAV) communications, where it can be deployed on the ground or attached to UAVs to assist terrestrial communications through exploiting the RIS reflection from the sky. RIS can be further deployed for simultaneous wireless information and power transfer (SWIPT) applications, where it tackles the low efficiency problem of the far-field power transfer and improves the energy harvesting performance.

The current wireless systems use half-duplex (HD) communications through time division duplex or frequency division duplex (FDD), in which the transmission and reception are not performed simultaneously. On the other hand, FD communications enable receiving a signal while also transmitting in the same frequency band. Due to this fact, FD technology has the potential to double the spectral efficiency and significantly increase the throughput of wireless communication systems [5]. FD



(a) RIS-assisted system.

(b) RIS-assisted UAV.

(c) RIS-assisted SWIPT.

(d) RIS-assisted physical layer security.

Fig. 1.2: RIS sample use cases.

communications have a wide range of benefits in many use-cases such as bidirectional communication, cooperative transmission, secure and cognitive radio applications [6]. The major challenge in FD is the residual self-interference (SI) from the transmit antennas to the receive antennas. However, a wide range of SI mitigation techniques were investigated and the FD feasibility has been experimentally demonstrated in small-scale wireless communications environments. Another challenge that needs to be addressed to successfully employ FD communications in 6G wireless systems is the co-channel interference (CCI) [7].

With the aid of RISs, the FD communication systems can be realized as RISs achieve high beamforming gain, suppress the CCI, and enhance multicasting performance.

Recent studies have proven that FD-aided RIS systems always outperform the half-duplex (HD) system despite the CCI. This is in sharp contrast with the conventional system without RIS, where the FD operation is not always beneficial, especially when the CCI is severe [8].

1.2 Literature Review

Due to the unique features of RIS, it has been thoroughly investigated in the recent years from an optimization and performance analysis perspectives. The deployment of RIS in various wireless applications has been considered including cognitive radio, secure communications, sparse code multiple access, cell-free systems, and high-altitude platform station (HAPS) [9, 10, 11, 12, 13, 14]. Most of the state-of-the-art works focused on investigating the RIS capabilities in a HD operating mode [15, 16, 17]. However, incorporating RIS into FD communications is not well investigated, yet, and is needed to fulfill the requirements of 6G applications. A number of existing works have studied RIS-assisted FD wireless networks [18, 11, 19, 8]. In [18], the authors developed a hybrid communication network which uses a FD decode-and-forward relay and an RIS to assist the communication over wireless channels. The aim was to maximize the minimum rate by optimizing the RIS phase shifts. To address the optimization problem, a semi-definite relaxation approach was used. The results proved that the SI at the relay is sufficiently suppressed due to the RIS deployment. The authors in [11] investigated the resource allocation design for RIS-assisted FD cognitive radio systems. To tackle the non-convex optimization problem, an iterative block coordinate descent is used.

Furthermore, the work in [19] studied the beamforming optimization problem of an RIS-assisted FD communication system. In particular, the sum-rate was maximized

by jointly optimizing the transmit beamforming and the RIS phase shifts using a fast converging alternating optimization (AO) technique. The investigation of the deployment of multi-RISs is studied in [8], where the authors considered the weighted sum-rate maximization for multi-RIS-assisted FD system with hardware impairments. An AO approach was proposed to obtain a sub-optimal solution, and the numerical results clarified that multiple RISs can significantly improve the performance metric under hardware impairments. The work in [20] further investigated the weighted sum transmit power consumption minimization problem of an RIS-assisted FD system, where an AO was proposed to solve the problem.

However, all the previous works on RIS-assisted FD communication systems used AO-based approaches. Such approaches are generally complex and difficult to realize in practical large scale systems. Moreover, AO approaches are considered sensitive to the system parameters and it poses the need of prior relaxations requirements. To this end, deep reinforcement learning (DRL) has emerged as a powerful and reliable approach to optimize the RIS phase shifts by overcoming the practical implementation problems of AO techniques. DRL algorithms enable addressing mathematically intractable nonlinear problems directly based on learning via interaction with the environment. In particular, the agent in DRL learns due to the feedback mechanism, where it gets rewarded for the optimized actions and punished otherwise [21]. To this end, this thesis focus on RIS-assisted FD communication systems, where DRL approaches are proposed to tackle the formulated optimization problems.

1.3 Thesis Contributions and Outline

Motivated by the aforementioned discussions, this thesis contributes to the literature by the following:

- A low-complexity DRL algorithm is designed which optimizes the RIS phase shifts, while significantly reducing the computational complexity for RIS-assisted HD communication system [22].

- For the first time in the literature, DRL is considered to efficiently optimize the RIS phase shifts in FD communication system. The proposed algorithm provides a significant improvement in the rate compared to the non-optimized RIS phase shifts [22].

- A closed-form solution is derived to optimize the transmit beamformers for single and distributed RIS schemes, which provides a remarkable improvement in the sum rate compared to state-of-the-art results [23].

- The RIS deployment problem is further studied in FD communication system which answers the question of when the single RIS deployment scheme outperforms the distributed RIS scheme, and vise versa [23].

- A practical DRL algorithm is proposed to optimize the discrete phase shifts of a distributed RIS-assisted FD network, for the first time in the literature. The proposed algorithm is shown to achieve promising results compared to the continuous-baseline [24].

- The performance of the proposed algorithm is assessed through extensive simulations, by considering two scenarios: the presence of the line-of-sight (LoS) link and when it is blocked [24].

The rest of this thesis is organized as follows: Chapter 2 proposes a low-complexity DRL algorithm for RIS-assisted wireless communication systems, where HD and FD operating modes are considered. Chapter 3 investigates the RIS deployment schemes (single versus distributed) in three practical FD scenarios and proposes a two-step

solution to solve the formulated optimization problem. Chapter 4 proposes a practical DRL algorithm to optimize the RIS discrete phase shifts in distributed RIS-assisted FD system. Finally, thesis conclusion and potential future developments are presented in Chapter 5.

References

[1] N. Rajatheva, I. Atzeni, E. Bjornson, A. Bourdoux, S. Buzzi, J.-B. Dore, S. Erkucuk, M. Fuentes, K. Guan, Y. Hu, X. Huang, J. Hulkkonen, J. M. Jornet, M. Katz, R. Nilsson, E. Panayirci, K. Rabie, N. Rajapaksha, M. J. Salehi, H. Sarieddeen, S. Shahabuddin, T. Svensson, O. Tervo, A. Tolli, Q. Wu, and W. Xu, "White paper on broadband connectivity in 6g," *6G Research Visions*, vol. 10. [Online]. Available: https://par.nsf.gov/biblio/10223732

[2] R. Alghamdi, R. Alhadrami, D. Alhothali, H. Almorad, A. Faisal, S. Helal, R. Shalabi, R. Asfour, N. Hammad, A. Shams, N. Saeed, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Intelligent surfaces for 6G wireless networks: A survey of optimization and performance analysis techniques," *IEEE Access*, vol. 8, pp. 202795–202818, Oct. 2020.

[3] M. A. ElMossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, "Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities," *IEEE Trans. Cogn. Commun.*, vol. 6, no. 3, pp. 990–1002, Sep. 2020.

[4] B. C. Nguyen, T. M. Hoang, L. T. Dung, and T. Kim, "On performance of two-way full-duplex communication system with reconfigurable intelligent surface," *IEEE Access*, vol. 9, pp. 81 274–81 285, Jun. 2021.

[5] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," ser. SIGCOMM '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 375â386.

[6] A. H. Gazestani, S. A. Ghorashi, B. Mousavinasab, and M. Shikh-Bahaei, "A survey on implementation and applications of full duplex wireless communications," *Physical Communication*, vol. 34, pp. 121–134, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1874490718304440

[7] R. Askar, J. Chung, Z. Guo, H. Ko, W. Keusgen, and T. Haustein, "Interference handling challenges toward full duplex evolution in 5G and beyond cellular networks," *IEEE Wirel. Commun.*, vol. 28, no. 1, pp. 51–59, Feb. 2021.

[8] M. A. Saeidi, M. J. Emadi, H. Masoumi, M. R. Mili, D. W. K. Ng, and I. Krikidis, "Weighted sum-rate maximization for multi-IRS-assisted full-duplex systems with hardware impairments," *IEEE Trans. Cogn. Commun.*, vol. 7, no. 2, pp. 466–481, Jun. 2021.

[9] G. Zhou, C. Pan, H. Ren, K. Wang, and Z. Peng, "Secure wireless communication in RIS-aided MISO system with hardware impairments," *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1309–1313, Jun. 2021.

[10] Z. Yang and Y. Zhang, "Beamforming optimization for RIS-aided SWIPT in cell-free MIMO networks," *China Communications*, vol. 18, no. 9, pp. 175–191, Sep. 2021.

[11] D. Xu, X. Yu, Y. Sun, D. W. K. Ng, and R. Schober, "Resource allocation for IRS-assisted full-duplex cognitive radio systems," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7376–7394, Dec. 2020.

[12] I. Al-Nahhal, O. A. Dobre, and E. Basar, "Reconfigurable intelligent surface-

assisted uplink sparse code multiple access," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 2058–2062, Feb. 2021.

[13] K. O. Odeyemi, P. A. Owolawi, and O. O. Olakanmi, "Reconfigurable intelligent surface-assisted HAPS relaying communication networks for multiusers under af protocol: A performance analysis," *IEEE Access*, vol. 10, pp. 14 857–14 869, Jan 2022.

[14] I. Al-Nahhal, O. A. Dobre, E. Basar, T. M. N. Ngatched, and S. Ikki, "Reconfigurable intelligent surface optimization for uplink sparse code multiple access," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 133–137, Jan. 2022.

[15] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Dec. 2020.

[16] S. Zhou, W. Xu, K. Wang, M. Di Renzo, and M.-S. Alouini, "Spectral and energy efficiency of IRS-assisted miso communication with hardware impairments," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1366–1369, Sep. 2020.

[17] M. Jung, W. Saad, M. Debbah, and C. S. Hong, "On the optimality of reconfigurable intelligent surfaces (RISs): Passive beamforming, modulation, and resource allocation," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4347–4363, Jul. 2021.

[18] Z. Abdullah, G. Chen, S. Lambotharan, and J. A. Chambers, "Optimization of intelligent reflecting surface assisted full-duplex relay networks," *IEEE Wireless Commun. Lett.*, Feb. 2021.

[19] H. Shen, T. Ding, W. Xu, and C. Zhao, "Beamformig design with fast convergence

for IRS-aided full-duplex communication," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2849–2853, Aug. 2020.

[20] Y. Cai, M.-M. Zhao, K. Xu, and R. Zhang, "Intelligent reflecting surface aided full-duplex communication: Passive beamforming and deployment design," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 1, pp. 383–397, Jan. 2022.

[21] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, May 2016, pp. 1–14.

[22] A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Deep reinforcement learning for optimizing RIS-assisted HD-FD wireless systems," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3893–3897, Dec. 2021.

[23] ——, "Deep reinforcement learning for RIS-assisted FD systems: Single or distributed RIS?" *IEEE Commun. Lett., Early Access*, Apr. 2022.

[24] ——, "Distributed RIS-assisted FD systems with discrete phase shifts: A reinforcement learning approach," in *Submitted to Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2022.

Chapter 2

Deep Reinforcement Learning for Optimizing RIS-Assisted HD-FD Wireless Systems

2.1 Abstract

This chapter investigates the reconfigurable intelligent surface (RIS)-assisted multiple-input single-output (MISO) wireless system, where both half-duplex (HD) and full-duplex (FD) operating modes are considered together, for the first time in the literature. The goal is to maximize the rate by optimizing the RIS phase shifts. A novel deep reinforcement learning (DRL) algorithm is proposed to solve the formulated non-convex optimization problem. The complexity analysis and Monte Carlo simulations illustrate that the proposed DRL algorithm significantly improves the rate compared to the non-optimized scenario in both HD and FD operating modes using a single parameter setting. Besides, it significantly reduces the computational complexity of the downlink HD MISO system and improves the achievable rate with a reduced number of steps

per episode compared to the conventional DRL algorithm.

2.2 Introduction

Reconfigurable intelligent surfaces (RISs) have emerged as a promising paradigm to fulfill the need of a smart and programmable wireless environment, and meet the demands of future wireless networks [1, 2]. RIS consists of a two-dimensional array of low-cost passive electromagnetic (EM) elements [3]. By overcoming the random nature of EM wave propagation, RIS enables controlling different characteristics of radio waves, such as scattering, reflection, and refraction. Consequently, it effectively enhances the signal quality and boosts the wireless spectral efficiency by realizing a controllable environment [4].

RIS-assisted multiple-input multiple-output systems have recently drawn significant attention as a cost-effective solution to enhance the wireless transmission in both half-duplex (HD) and full-duplex (FD) operating modes [5, 6, 7, 8, 9, 10]. In the HD mode, systems require additional resources to receive and forward signals, which results in a decreased spectral efficiency. In contrast, the FD mode has the potential to significantly increase the throughput of wireless systems as it enables simultaneous transmission and reception of signals in the same frequency band. However, this comes at the cost of increased interference and implementation complexity. To this end, some researchers are considering HD-FD transmission schemes that combine the advantages of both HD and FD modes [11]. In [5] and [6], RIS-HD systems are optimized to minimize the total transmit power. In [7], a joint optimization problem is considered to maximize the achievable rate of an RIS-HD system. In [8] and [9], the sum-rate and spectral efficiency of an RIS-FD system is maximized, respectively. In [10], the weighted minimum rate is maximized for a multi-user RIS-FD system. Most of these

works decoupled the optimization variables using alternating optimization algorithms, which exhibit both loss of optimality and high computational complexity.

Deep learning has emerged as a powerful approach to optimize the RIS phase shifts by tackling the practical implementation problems of the optimization techniques [12, 13]. In particular, deep reinforcement learning (DRL) is a potential candidate to optimize the RIS phase shifts without the need for offline training with a labeled dataset. A few works have considered DRL approaches to optimize RIS-HD systems [14, 15, 16]. The authors in [14] proposed an optimization-driven deep deterministic policy gradient (DDPG) to minimize the access point's transmit power. The sum-rate maximization problem of a multi-user RIS-HD system was addressed in [15] using a DRL algorithm. Furthermore, a conventional DRL algorithm is introduced in [16] to maximize the received signal-to-noise ratio of the downlink RIS-HD multiple-input single-output (MISO) system. To the best of the authors' knowledge, utilizing DRL for RIS-FD systems has not yet been discussed in the literature.

In this chapter, a novel DRL algorithm is proposed to optimize the phase shifts of an RIS-assisted HD-FD MISO system. The contributions are summarized as follows:

- A DRL algorithm is proposed which achieves promising results in the HD and FD operating modes without the need of additional parameters tuning.
- The proposed DRL algorithm provides a significant improvement in the rate compared to the non-optimized RIS phase shifts in the HD and FD operating modes.
- It significantly reduces the computational complexity, while providing a considerable rate improvement with a reduced number of required steps for each episode, compared to the conventional DRL in [16] for the HD mode.
- The complexity analysis and Monte Carlo simulations support the findings.

The remainder of this chapter is organized as follows: Section 2.3 presents the system model and problem formulation for the RIS-assisted HD-FD MISO system. The proposed DRL algorithm is introduced in Section 2.4, and its computational complexity is analyzed in Section 2.5. Simulation results and conclusions are presented in Sections 2.6 and 2.7, respectively.

2.3 System Model and Problem Formulation

Consider an RIS-assisted HD-FD MISO system as illustrated in Fig. 2.1, where S_1 and S_2 represent the base station (BS) and user equipment (UE), respectively. Both the BS and UE are equipped with M transmit antennas and one receive antenna. The UE sometimes operates in a HD mode, where it only receives information from the BS (i.e., downlink HD mode), while other times the UE and BS transmit and receive information simultaneously in the same frequency band (i.e., FD mode). Henceforth, Ω denotes the operating mode, where $\Omega \in \{\text{HD}, \text{FD}\}$. The RIS is composed of N programmable reflecting elements, which assists the communication between S_1 and S_2 by optimizing the RIS phase shifts through an RIS controller. Given $\bar{i} = 3 - i$ $\forall\, i = 1, 2$, let $\mathbf{H}_{S_{\bar{i}}R} \in \mathbb{C}^{N\times M}$, $\mathbf{h}_{RS_i}^H \in \mathbb{C}^{1\times N}$, and $\mathbf{h}_{S_{\bar{i}}S_i}^H \in \mathbb{C}^{1\times M}$ denote the channel coefficients of the $S_{\bar{i}}$-RIS, RIS-S_i, and $S_{\bar{i}}$-S_i links, respectively. The self-interference (SI) channels, which are involved in the FD mode at the BS and UE are denoted by $\mathbf{h}_{S_iS_i}^H \in \mathbb{C}^{1\times M}$.

At the receiver-side, the signal is received from the direct and reflected links of the BS and RIS, respectively. Thus, the noisy received signals of the downlink HD and FD operating modes are respectively expressed as

$$y_i^{\Omega} = \Big(\underbrace{\mathbf{h}_{RS_i}^H \boldsymbol{\Theta} \mathbf{H}_{S_{\bar{i}}R}}_{\text{Reflected signal}} + \underbrace{\mathbf{h}_{S_{\bar{i}}S_i}^H}_{\text{Direct signal}}\Big)\mathbf{w}_{\bar{i}} x_{\bar{i}} + n,\, i = 2,\, \Omega = \text{HD}, \tag{2.1}$$

and

$$y_i^{\Omega} = \Big(\underbrace{\mathbf{h}_{RS_i}^H \boldsymbol{\Theta} \mathbf{H}_{S_{\bar{i}}R}}_{\text{Reflected signal}} + \underbrace{\mathbf{h}_{S_{\bar{i}}S_i}^H}_{\text{Direct signal}}\Big)\mathbf{w}_{\bar{i}} x_{\bar{i}} + \underbrace{\mathbf{h}_{S_iS_i}^H \mathbf{w}_i x_i}_{\text{Residual SI}} + n,\ i = 1, 2, \Omega = \text{FD}, \tag{2.2}$$

where $n \sim \mathcal{CN}(0, \sigma^2)$ denotes the additive white complex Gaussian noise with zero-mean and variance σ^2. The diagonal matrix $\boldsymbol{\Theta} = \text{diag}\big(e^{j\varphi_1}, \cdots, e^{j\varphi_n}, \cdots, e^{j\varphi_N}\big) \in \mathbb{C}^{N\times N}$ represents the phase shifts of the RIS, where $\varphi_n \in [-\pi, \pi)$ is the phase shift introduced by the n-th reflecting element. The source node, S_i, employs an active beamforming $\mathbf{w}_i \in \mathbb{C}^{M\times 1}$ to transmit the information signal, x_i, with $\mathbb{E}\{|x_i|^2\} = 1$, where $\mathbb{E}\{\cdot\}$ denotes the expectation operation. The third term in (2.2) represents the SI introduced by the FD mode operation.

The achievable rate and sum-rate of the downlink HD and FD operating modes, measured in bit per second per Hertz (bps/Hz), are respectively given as

$$\mathcal{R}^{\Omega} = \log_2\left(1 + \frac{\left|\left(\mathbf{h}_{RS_i}^H \boldsymbol{\Theta} \mathbf{H}_{S_{\bar{i}}R} + \mathbf{h}_{S_{\bar{i}}S_i}^H\right)\mathbf{w}_{\bar{i}}\right|^2}{\sigma^2}\right),\ i = 2, \Omega = \text{HD}, \tag{2.3}$$

and

$$\mathcal{R}^{\Omega} = \sum_{i=1}^{2} \log_2\left(1 + \frac{\left|\left(\mathbf{h}_{RS_i}^H \boldsymbol{\Theta} \mathbf{H}_{S_{\bar{i}}R} + \mathbf{h}_{S_{\bar{i}}S_i}^H\right)\mathbf{w}_{\bar{i}}\right|^2}{|\mathbf{h}_{S_iS_i}^H \mathbf{w}_i|^2 + \sigma^2}\right),\ \Omega = \text{FD}. \tag{2.4}$$

Here, the goal is to maximize the rate of the RIS-assisted HD-FD MISO system by optimizing the RIS phase shifts. Thus, the resulting optimization problem can be expressed as

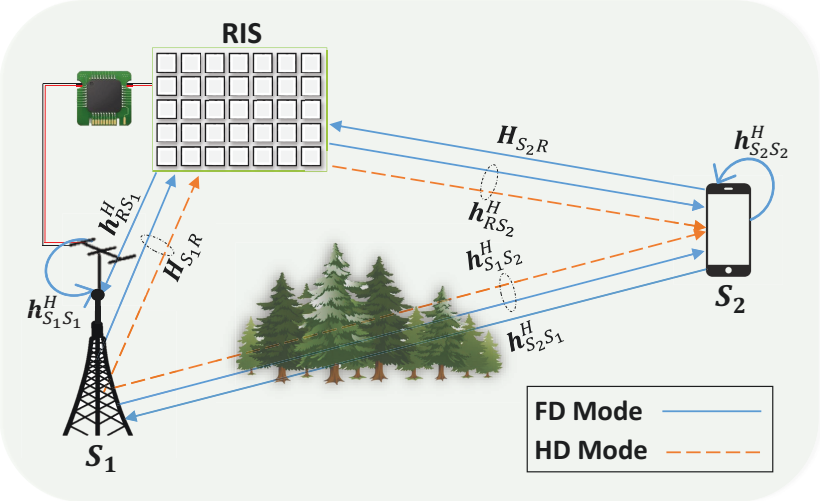


Fig. 2.1: RIS-assisted HD-FD MISO system.

$$\text{(P1)} \quad \max_{\boldsymbol{\varphi}} \quad \mathcal{R}^{\Omega},\ \Omega \in \{\text{HD}, \text{FD}\} \tag{2.5a}$$

$$\text{s.t.} \quad -\pi \leq \varphi_n \leq \pi,\ n = 1, \cdots, N. \tag{2.5b}$$

It is worth noting that the conventional DRL algorithm in [16] has been proposed to solve the non-convex problem (P1) only when $\Omega = \text{HD}$, and suffers from high computational complexity. Moreover, the DRL for the FD operating mode has not yet been investigated in the literature.

2.4 Proposed DRL Algorithm

This section proposes a novel DRL algorithm to solve (P1) for the RIS-assisted HD-FD MISO system. To deal with (P1), the RIS phase shifts are optimized using the proposed DRL algorithm. Then, for a given optimized $\boldsymbol{\Theta}$, the transmit beamformers, $\mathbf{w}_{\bar{i}}$, are optimized using a closed and semi-closed form solutions for the HD and FD operating

modes, respectively. The optimization problem is solved in an iterative fashion until the optimized $\mathbf{\Theta}$ and $\mathbf{w}_{\bar{i}}$ converge.

2.4.1 Beamforming Design for a Given $\mathbf{\Theta}$

The optimal beamforming vector for the HD operating mode is calculated using the maximum ratio transmission approach, whereas a semi-closed optimal solution of the FD beamforming vectors is given in [8]. Consequently, for a given optimized $\mathbf{\Theta}$, the optimal beamforming vectors of the HD and FD modes, $\mathbf{w}_{\bar{i}}$, are respectively given as

$$\mathbf{w}_{\bar{i}}^{\dagger} = \sqrt{P_{\max}} \frac{\left(\mathbf{h}_{RS_i}^H \mathbf{\Theta} \mathbf{H}_{S_{\bar{i}}R} + \mathbf{h}_{S_{\bar{i}}S_i}^H\right)^H}{\left|\left|\left(\mathbf{h}_{RS_i}^H \mathbf{\Theta} \mathbf{H}_{S_{\bar{i}}R} + \mathbf{h}_{S_{\bar{i}}S_i}^H\right)\right|\right|}, i = 2, \Omega = \text{HD}, \tag{2.6}$$

and

$$\mathbf{w}_{\bar{i}}^{\dagger} = (\delta \mathbf{h}_{S_{\bar{i}}S_{\bar{i}}} \mathbf{h}_{S_{\bar{i}}S_{\bar{i}}}^H + v^{\dagger}\mathbf{I})^{-1} \boldsymbol{\mathcal{B}}, i = 1, 2, \Omega = \text{FD}, \tag{2.7}$$

where $P_{\max}$ is the maximum transmitted power of $S_{\bar{i}}$, $\mathbf{I}$ is the identity matrix, and $v^{\dagger}$ is the optimal dual Lagrangian variable associated with the power constraint that is found by performing a bisection search over the interval $\left[0, \sqrt{\boldsymbol{\mathcal{B}}^T \boldsymbol{\mathcal{B}}}/\sqrt{P_{\max}}\right]$. Here, $\boldsymbol{\mathcal{B}}$ and δ are given as

$$\boldsymbol{\mathcal{B}} \triangleq \frac{1}{\tilde{b}_i} \left(1 + \frac{b_i}{|\mathbf{h}_{S_{\bar{i}}S_{\bar{i}}}^H \tilde{\mathbf{w}}_{\bar{i}}|^2 + \sigma^2}\right) \mathbf{h}_{\bar{i}} \mathbf{h}_{\bar{i}}^H \tilde{\mathbf{w}}_{\bar{i}}, \tag{2.8}$$

and

$$\delta \triangleq \frac{b_i \left(|\mathbf{h}_{\bar{i}}^H \tilde{\mathbf{w}}_{\bar{i}}|^2 + \tilde{b}_i\right)}{\tilde{b}_i \left(|\mathbf{h}_{S_{\bar{i}}S_{\bar{i}}}^H \tilde{\mathbf{w}}_{\bar{i}}|^2 + \sigma^2\right)^2}, \tag{2.9}$$

where $b_i \triangleq |\mathbf{h}_i^H \mathbf{w}_i|^2$, $\tilde{b}_i \triangleq |\mathbf{h}_{S_iS_i}^H \mathbf{w}_i|^2 + \sigma^2$, $\mathbf{h}_{\bar{i}} \triangleq \mathbf{H}_{S_{\bar{i}}R}^H \mathbf{\Theta}^H \mathbf{h}_{RS_i} + \mathbf{h}_{S_{\bar{i}}S_i}$, and $\tilde{\mathbf{w}}_{\bar{i}}$ is a given feasible point.

2.4.2 Phase Shift Design Based on the Proposed DRL Algorithm

2.4.2.1 Problem Transformation

The RIS controller represents the DRL *agent*, while the RIS-assisted HD-FD MISO communication system represents the DRL *environment*. Thus, the *state space*, *action space*, and *reward* for the proposed DRL algorithm are defined as follows:

- State space: The state space at time step t, $s_t \in \mathbb{R}^{1\times(N+1)}$, includes $\varphi_n \forall\, n = 1, \cdots, N$ and the corresponding $\mathcal{R}^{\Omega}$ at time step $t-1$, and is defined as

$$s_t = \left[\mathcal{R}^{\Omega,(t-1)}, \varphi_1^{(t-1)}, \cdots, \varphi_n^{(t-1)}, \cdots, \varphi_N^{(t-1)}\right]. \tag{2.10}$$

- Action space: Since (P1) aims to optimize the RIS phase shifts, the action space at time step t, $a_t \in \mathbb{R}^{1\times N}$, is expressed as

$$a_t = \left[\varphi_1^{(t)}, \cdots, \varphi_n^{(t)}, \cdots, \varphi_N^{(t)}\right]. \tag{2.11}$$

- Reward: As the target of (P1) is to maximize $\mathcal{R}^{\Omega}$, the reward is expressed as

$$r_t = \mathcal{R}^{\Omega,(t)},\ \Omega \in \{\mathrm{HD}, \mathrm{FD}\}. \tag{2.12}$$

At each time step t, the agent receives the current state s_t from the environment, takes an action a_t based on a *policy* $\tilde{\pi}$, and receives a scalar reward r_t. Then, a new state s_{t+1} is obtained. The return of a state is defined as the *total discounted reward* from time step t onwards, and is given by $R_t = \sum_{k=t}^{\infty} \gamma^{k-t} r(s_k, a_k)$, where $\gamma \in (0, 1]$ is the DRL discount factor. The goal is to learn a policy that maximizes the expected cumulative discounted reward from the start state, as: $J(\tilde{\pi}) = \mathbb{E}\left[R_1|\tilde{\pi}\right]$. The DDPG, which combines the benefits of value-based and policy-based approaches [17], is used to

learn the optimal policy for a continuous a_t. In particular, the DDPG algorithm aims at maximizing the Q-value of (s, a) pair by training a deep neural network (DNN), defined as

$$Q^{\tilde{\pi}_{\boldsymbol{\theta}}}(s, a) = \mathbb{E}_{\tilde{\pi}_{\boldsymbol{\theta}}}\Big[R_1 | s_1 = s, a_1 = a\Big], \tag{2.13}$$

where $\boldsymbol{\theta}$ represents the DNN parameters, as well as finding the optimal policy by performing the gradient ascent of

$$\nabla_{\boldsymbol{\theta}} J(\tilde{\pi}_{\boldsymbol{\theta}}) = \mathbb{E}_{\tilde{\pi}_{\boldsymbol{\theta}}}\left[Q^{\tilde{\pi}_{\boldsymbol{\theta}}}(s, a) \nabla_{\boldsymbol{\theta}} \log \pi_{\boldsymbol{\theta}}(a|s)\right]. \tag{2.14}$$

The DDPG algorithm is based on the actor-critic technique, which consists of two DNN models: actor and critic. The actor, $\mu(s_t|\boldsymbol{\theta}_\mu)$, represents the policy network that takes the state as an input for a given $\boldsymbol{\theta}_\mu$ and outputs $a_t = \mu(s_t|\boldsymbol{\theta}_\mu) + \xi$, where ξ is a random process that is added to the actions for exploration. ξ is modeled as complex Gaussian process with zero mean and variance 0.1. The critic, $Q(s_t, a_t|\boldsymbol{\theta}_q)$, represents the network that evaluates the actions. It takes s_t and a_t as an input for a given $\boldsymbol{\theta}_q$, and outputs the Q-value. The DDPG algorithm utilizes the concept of experience replay with memory D to reduce the correlation of the training samples by randomly sampling minibatch transitions, N_B. Moreover, target networks are introduced to stabilize the learning process. The target networks are generated by making a copy of the actor and critic evaluation NNs, $\mu'(s_t|\boldsymbol{\theta}_{\mu'})$ and $Q'(s_t, a_t|\boldsymbol{\theta}_{q'})$, and are used to calculate the corresponding target values, y_t in (2.15). The actor and critic NN parameters, $\boldsymbol{\theta}_\mu$ and $\boldsymbol{\theta}_q$, are updated using the stochastic gradient descent (SGD) from (2.16) and policy gradient from (2.17), respectively. Finally, the target NN parameters are updated using a soft update coefficient, τ, based on (2.18) and (2.19). After T steps of each episode, the agent's performance saturates and it outputs

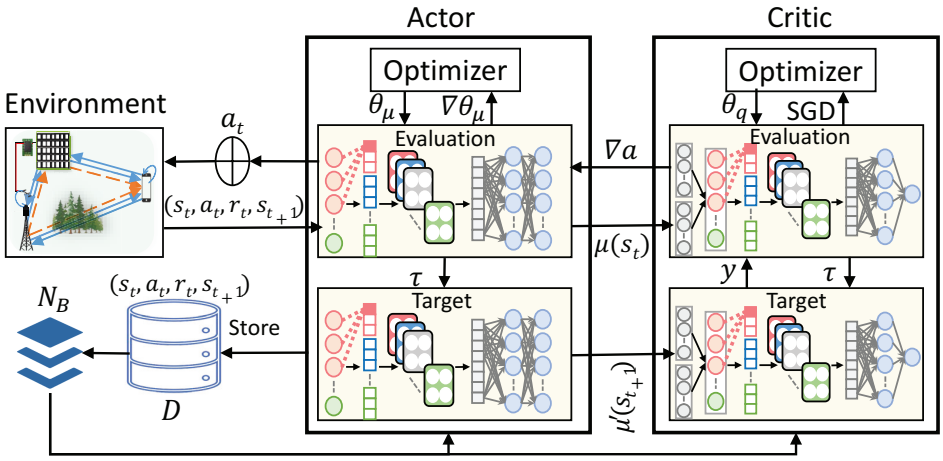


Fig. 2.2: The proposed DRL algorithm structure.

the optimized Θ. The structure of the proposed DRL algorithm is illustrated in Fig. 2.2 and summarized in Algorithm 1.

2.4.2.2 Proposed DNN Design

As can be seen from Fig. 2.2, the proposed DRL algorithm contains four NNs (i.e., two NNs for the actor and two NNs for the critic). A novel design is proposed for the four NNs, which consists of the input layer, two hidden layers and the output layer. The two hidden layers are a combination of one convolutional layer and one feed-forward (FF) layer with a flatten layer between them. The input layer of the actor and critic networks contains $N+1$ neurons (i.e., size of s_t) and $2N+1$ neurons (i.e., concatenation of s_t and a_t), respectively. The output layer of the actor and critic networks contains N neurons (i.e., size of a_t) and one neuron (i.e., scalar Q-value), respectively. The convolutional hidden layer for each of the actor and critic networks uses the $ReLU$ activation function since it does not suffer from vanishing or exploding gradient problems. In contrast, the FF hidden layer uses the *softmax* activation function to obtain probabilistic values for all inputs.

Algorithm 1 Proposed DRL algorithm.

Initialize: $\boldsymbol{\theta}_\mu$ and $\boldsymbol{\theta}_q$ with random weights, D, γ, τ, and learning rate α;
Set: $\boldsymbol{\theta}_{\mu'} \leftarrow \boldsymbol{\theta}_\mu$ and $\boldsymbol{\theta}_{q'} \leftarrow \boldsymbol{\theta}_q$;

1: **repeat** for K episodes:
2: Collect the channels of the k-th episode based on Ω;
3: Randomly initialize $\varphi_n \, \forall \, n = 1, \cdots, N$ to obtain the initial state;
4: **if** $\Omega =$ HD **then**
5: Calculate $\mathbf{w}_{\bar{i}}$ using (2.6);
6: **else**
7: Calculate $\mathbf{w}_{\bar{i}}$ using (2.7);
8: **end if**
9: Initialize $\xi \sim \mathcal{CN}(0, 0.1)$;
10: **repeat** for T steps:
11: Obtain $a_t = \mu(s_t|\boldsymbol{\theta}_\mu) + \xi$ from the actor network and reshape it;
12: Repeat **Lines** #4-8;
13: Observe the new state, s_{t+1}, given a_t;
14: Store (s_t, a_t, r_t, s_{t+1}) in D;
15: When D is full, sample a minibatch of N_B transitions randomly (s_j, a_j, r_j, s_{j+1}) from D;
16: Compute the target value using target networks:

$$y_j = r_j + \gamma Q'(s_{j+1}, \mu'(s_{j+1}|\boldsymbol{\theta}_{\mu'})|\boldsymbol{\theta}_{q'}); \tag{2.15}$$

17: Update the critic by minimizing the loss using SGD:

$$L = \frac{1}{N_B} \sum_j \left(y_j - Q(s_j, a_j|\boldsymbol{\theta}_q)\right)^2; \tag{2.16}$$

18: Update the actor using the policy gradient:

$$\nabla_{\boldsymbol{\theta}_\mu} = \frac{1}{N_B} \sum_j \nabla_a Q(s, a|\boldsymbol{\theta}_q)|_{s=s_j, a=\mu(s_j)} \nabla_{\boldsymbol{\theta}_\mu} \mu(s|\boldsymbol{\theta}_\mu)|_{s_j}; \tag{2.17}$$

19: Update the target NNs through soft update:

$$\boldsymbol{\theta}_{q'} \longleftarrow \tau \boldsymbol{\theta}_q + (1 - \tau)\boldsymbol{\theta}_{q'}, \tag{2.18}$$

$$\boldsymbol{\theta}_{\mu'} \longleftarrow \tau \boldsymbol{\theta}_\mu + (1 - \tau)\boldsymbol{\theta}_{\mu'}. \tag{2.19}$$

Output: Optimal action that corresponds to the optimal $\boldsymbol{\Theta}$.

2.5 Complexity Analysis

The computational complexity of the conventional DRL algorithm in [16] and the proposed DRL algorithm for Ω = HD is derived in terms of the number of NN parameters $C_{\mathcal{P}}$ required to be stored, real additions $C_{\mathcal{A}}$, and real multiplications $C_{\mathcal{M}}$. The conventional DRL algorithm uses two hidden FF layers, and its computational complexity is given as

$$C_{\mathcal{P}} = \sum_{i=1}^{3}(\eta_i + 1)\eta_{i+1}, \tag{2.20}$$

$$C_{\mathcal{M}} = \sum_{i=1}^{3}\eta_i\eta_{i+1}, \tag{2.21}$$

$$C_{\mathcal{A}} = \sum_{i=1}^{3}\eta_i\eta_{i+1} + \sum_{i=1}^{3}\eta_{i+1}, \tag{2.22}$$

where η_i is the number of neurons of the i-th layer. For simplicity, each activation function is considered to cost one real addition.

Based on the NNs design in Section 2.4.2.2, the complexity for the proposed DRL algorithm is given as

$$C_{\mathcal{P}} = (\eta_F\eta_3 + F_z + 1)F_n + (\eta_4 + 1)\eta_3 + \eta_4, \tag{2.23}$$

$$C_{\mathcal{M}} = (F_z + \eta_3)\eta_F F_n + \eta_3\eta_4, \tag{2.24}$$

$$C_{\mathcal{A}} = (F_z + \eta_3 + 1)\eta_F F_n + (\eta_4 + 1)\eta_3 + \eta_4, \tag{2.25}$$

where $\eta_F = \lfloor \frac{\eta_1 - F_z}{F_s} + 1 \rfloor$, with $\lfloor \cdot \rfloor$ as the floor operation, F_z is the filter size, F_n is the number of filters, and F_s is the stride. The complexity reduction of using the proposed

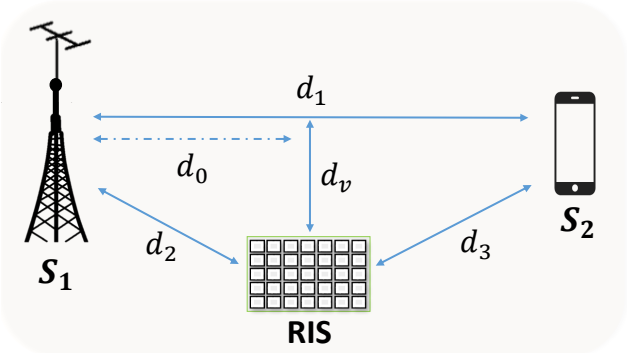


Fig. 2.3: Simulation setup.

DRL algorithm over the conventional one for $\Omega = \text{HD}$ is

$$\text{Reduction} = 1 - \frac{\left\{C_{\chi}^{\text{Actor}} + C_{\chi}^{\text{Critic}}\right\}_{\text{Proposed}}}{\left\{C_{\chi}^{\text{Actor}} + C_{\chi}^{\text{Critic}}\right\}_{\text{Conventional}}}, \ \chi \in \{\mathcal{P}, \mathcal{A}, \mathcal{M}\}. \tag{2.26}$$

2.6 Simulation Results

This section evaluates the performance of the proposed DRL algorithm for the RIS-assisted HD-FD MISO system. The simulation setup is shown in Fig. 2.3, where the considered parameters are $d_v = 2\,\text{m}$ and $d_1 = 50\,\text{m}$. The distances of the BS-RIS and UE-RIS links are calculated as $d_2 = \sqrt{d_0^2 + d_v^2}\,\text{m}$ and $d_3 = \sqrt{(d_1 - d_0)^2 + d_v^2}\,\text{m}$, respectively. The path loss (PL) at distance d_j, $\forall j \in \{1, 2, 3\}$ is modeled as $\text{PL} = PL_0 - 10\zeta \log_{10}\left(\frac{d_j}{D_r}\right)$ [16], where PL_0 is the PL at a reference distance D_r and ζ is the PL exponent, in which $PL_0 = -30\,\text{dB}$ and $D_r = 1\,\text{m}$. As in [16], the BS-UE channels are modeled as Rayleigh fading (assuming a blocking element between S_1 and S_2), while the rest of the channels are Rician with a factor of 10. The PL exponents of the BS-UE, BS-RIS, and UE-RIS channels are set to $\zeta_{\text{BU}} = 3$ and $\zeta_{\text{BR}} = \zeta_{\text{UR}} = 2$, respectively. The PL of the SI channels for the FD mode is $-95\,\text{dB}$. The total transmit power is $P = 5\,\text{dBm}$, while the noise power is $\sigma^2 = -80\,\text{dBm}$ [16]. The antenna gain at the BS and UE is $0\,\text{dBi}$, while the RIS gain is $5\,\text{dBi}$. The penetration loss in the

TABLE 2.1: DDPG Parameters.

Parameter	Value
(F_n, F_z, F_s)	(4, 3, 2)
$\eta_{\text{FF hidden layer}}$	60
(K, T)	(500, 800)
N_B	16
α	10^{-3}
γ	0.95
τ	0.005
D	50000

BS-UE and RIS-UE links is 10 dB.

The parameters of the proposed DRL algorithm are summarized in Table. 2.1. Furthermore, the design of the NNs is explained in Section 2.4.2.2, and its parameters are provided in Table. 2.1. The Adam optimizer is used to update the parameters of the NNs. To assess the performance of the proposed algorithm, it is compared with the non-optimized scenario, referred to as random phase shifts. The conventional DRL algorithm in [16] with $T = 1000$ is also included to show the superiority of the proposed DRL algorithm in the HD mode. It is worth noting that the current form of the conventional DRL algorithm can not be used to optimize the RIS phase shifts in the FD mode.

Figure. 2.4 studies the impact of the RIS location on the system performance. It is shown that the proposed DRL algorithm significantly improves the rate for both operating modes, compared to the random phase shifts and without-RIS scenarios, especially when the RIS is located closer to either the BS or the UE. On the other hand, the random phase shifts scenario does not improve the rate when the RIS is located relatively far from both BS and UE, compared to the scenario without-RIS. Consequently, a proper optimization for the RIS phase shifts is needed to achieve a satisfactory performance. Since the RIS should be deployed near the BS or UE to best

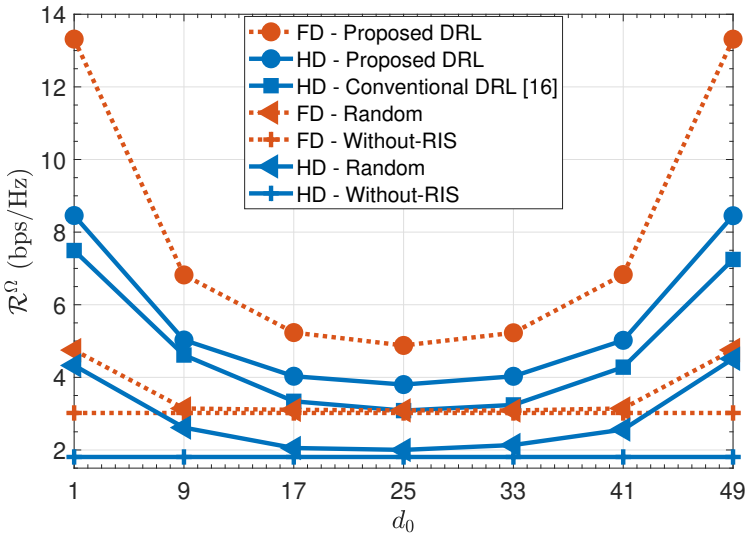


Fig. 2.4: RIS deployment investigation.

benefit from the RIS reflection links, it is considered that $d_0 = 1\,\text{m}$ for the rest of the chapter.

Figure. 2.5 illustrates the effect of increasing N on the system performance. As can be observed, R^{Ω} increases as N increases for all algorithms. The proposed DRL algorithm provides an improvement of 4.6 bps/Hz and 8.5 bps/Hz in the achievable rate and sum-rate of the HD and FD modes, respectively, compared to the random phase shifts scenario at $N = 40$. It is worth noting that the gain gap increases as N increases for the proposed DRL algorithm.

In the HD operating mode, the proposed DRL algorithm improves the achievable rate performance by 1.4 bps/Hz and 0.6 bps/Hz at $N = 20$ and $N = 40$, respectively, when compared to [16], as depicted in Fig. 2.5. Moreover, as shown in Fig. 2.6, the proposed DRL algorithm in the HD mode (with $T = 800$ steps) significantly reduces the computational complexity of each NN in the range of 94% to 86% for the practical case of $N = 20$ to 60, respectively, compared to conventional DRL in [16] (with $T = 1000$ steps). The proposed algorithm achieves a significant complexity reduction

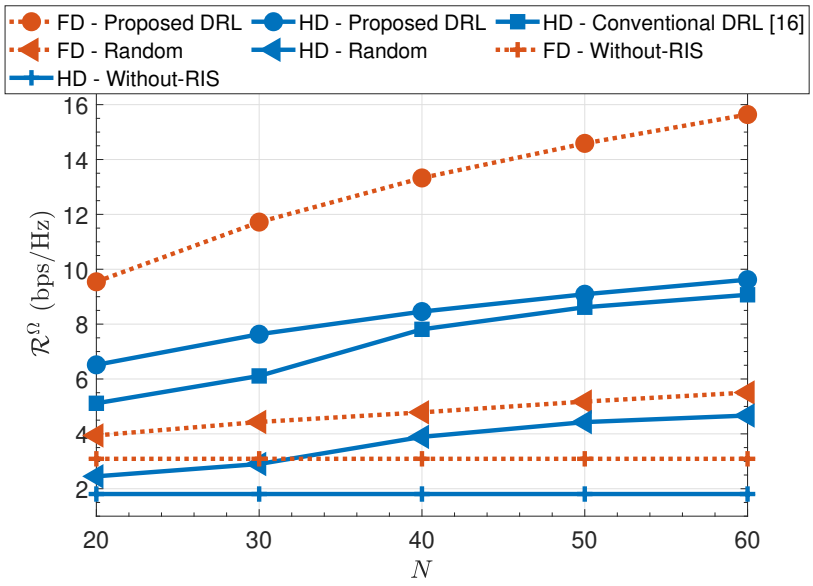


Fig. 2.5: The impact of varying N on the system performance.

percentage through proposing a novel and simpler NN along with hyperparameter tuning and optimized DRL configuration. Although the complexity reduction seems to decrease as N increases, it saturates at 63% for a certain large value of N, as seen from the asymptotic complexity bound in Fig. 2.7.

Finally, the proposed DRL algorithm provides a significant improvement in the rate for both operating modes, compared with the random phase shifts scenario. Besides, with a 20% reduction in the number of required steps when compared with the conventional DRL algorithm, the proposed DRL algorithm guarantees a faster convergence and improves the rate with lower computational complexity for each of the four NNs.

2.7 Conclusion

This chapter considered DRL for the rate maximization problem of the RIS-assisted HD-FD MISO system, for the first time in the literature. With a single parameter

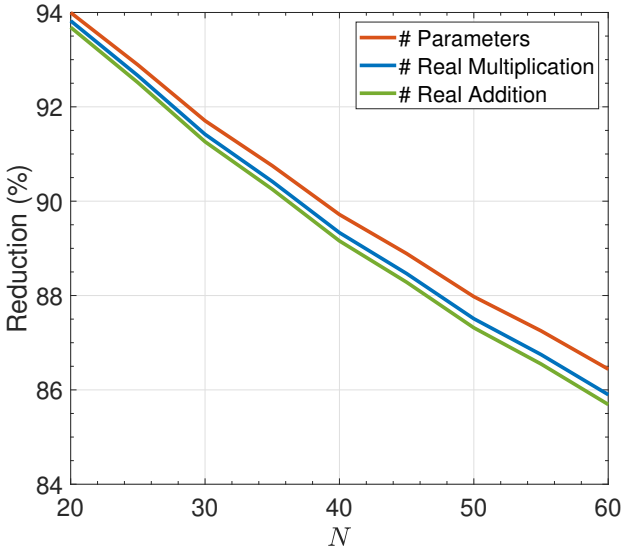


Fig. 2.6: Practical range of N.

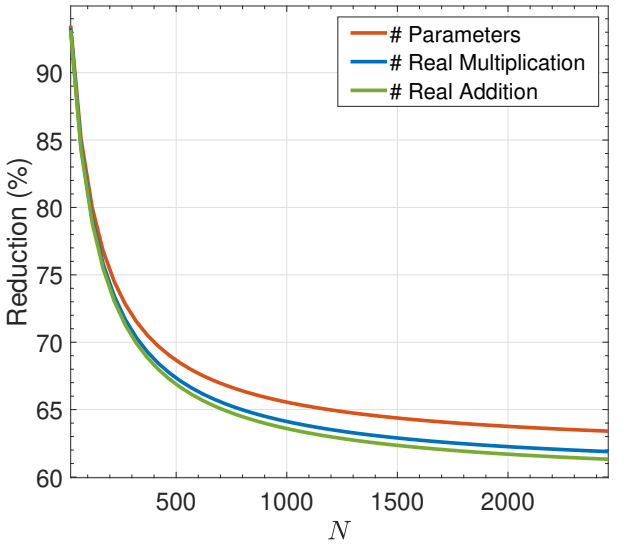


Fig. 2.7: Asymptotic range of N.

setting, the proposed DRL algorithm optimized the RIS phase shifts for both HD and FD operating modes. A novel DNN structure was proposed to learn the optimal policy of the proposed DRL algorithm. Compared to the non-optimized scenario, the proposed DRL algorithm significantly improved the achievable rate and sum-rate for the HD and FD operating modes, respectively. Compared to the conventional DRL algorithm in HD mode, the proposed DRL algorithm saved 20% of the required steps per episode and achieved up to 1.4 bps/Hz rate improvement with up to 94% reduction in the computational complexity. Future works can consider extending the proposed DRL algorithm to optimize the multi-user scenario.

References

[1] I. Al-Nahhal, O. A. Dobre, and E. Basar, "Reconfigurable intelligent surface-assisted uplink sparse code multiple access," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 2058–2062, Feb. 2021.

[2] L. Bariah, L. Mohjazi, S. Muhaidat, P. C. Sofotasios, G. K. Kurt, H. Yanikomeroglu, and O. A. Dobre, "A prospective look: Key enabling technologies, applications and open research topics in 6G networks," *IEEE Access*, vol. 8, pp. 174792–174820, Aug. 2020.

[3] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, Aug. 2019.

[4] R. Alghamdi, R. Alhadrami, D. Alhothali, H. Almorad, A. Faisal, S. Helal, R. Shalabi, R. Asfour, N. Hammad, A. Shams, N. Saeed, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Intelligent surfaces for 6G wireless networks: A

survey of optimization and performance analysis techniques," *IEEE Access*, vol. 8, pp. 202795–202818, Oct. 2020.

[5] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Dec. 2020.

[6] G. Zhou, C. Pan, H. Ren, K. Wang, M. D. Renzo, and A. Nallanathan, "Robust beamforming design for intelligent reflecting surface aided MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1658–1662, Jun. 2020.

[7] N. S. Perovic, L.-N. Tran, M. Di Renzo, and M. F. Flanagan, "Achievable rate optimization for MIMO systems with reconfigurable intelligent surfaces," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3865–3882, Feb. 2021.

[8] H. Shen, T. Ding, W. Xu, and C. Zhao, "Beamformig design with fast convergence for IRS-aided full-duplex communication," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2849–2853, Aug. 2020.

[9] J. Zhao, M. Chen, M. Chen, Z. Yang, Y. Wang, B. Cao, and M. Shikh-Bahaei, "Energy efficient full-duplex communication systems with reconfigurable intelligent surface," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, Feb. 2020, pp. 1–5.

[10] Z. Peng, Z. Zhang, C. Pan, L. Li, and A. L. Swindlehurst, "Multiuser full-duplex two-way communications via intelligent reflecting surface," *IEEE Trans. Signal Process.*, vol. 69, pp. 837–851, Jan. 2021.

[11] M. Elhattab, M. A. Arfaoui, C. Assi, and A. Ghrayeb, "Reconfigurable intelligent surface enabled full-duplex/half-duplex cooperative non-orthogonal multiple access," Jan. 2021. [Online]. Available: https://arxiv.org/abs/2101.01307

[12] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?" *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 7331–7376, Jun. 2019.

[13] Y. Chen, Y. Liu, M. Zeng, U. Saleem, Z. Lu, X. Wen, D. Jin, Z. Han, T. Jiang, and Y. Li, "Reinforcement learning meets wireless networks: A layering perspective," *IEEE Internet Things J.*, vol. 8, no. 1, pp. 85–111, Jan. 2021.

[14] J. Lin, Y. Zout, X. Dong, S. Gong, D. T. Hoang, and D. Niyato, "Deep reinforcement learning for robust beamforming in IRS-assisted wireless communications," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Jan. 2020, pp. 1–6.

[15] C. Huang, R. Mo, and C. Yuen, "Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1839–1850, Jun. 2020.

[16] K. Feng, Q. Wang, X. Li, and C.-K. Wen, "Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 745–749, Jan. 2020.

[17] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, May 2016, pp. 1–14.

Chapter 3

Deep Reinforcement Learning for RIS-Assisted FD Systems: Single or Distributed RIS?

3.1 Abstract

This chapter investigates reconfigurable intelligent surface (RIS)-assisted full-duplex multiple-input single-output wireless system, where the beamforming and RIS phase shifts are optimized to maximize the sum-rate for both single and distributed RIS deployment schemes. The preference of using the single or distributed RIS deployment scheme is investigated through three practical scenarios based on the links' quality. The closed-form solution is derived to optimize the beamforming vectors and a novel deep reinforcement learning (DRL) algorithm is proposed to optimize the RIS phase shifts. Simulation results illustrate that the choice of the deployment scheme depends on the scenario and the links' quality. It is further shown that the proposed algorithm significantly improves the sum-rate compared to the non-optimized scenario in both

single and distributed RIS deployment schemes. Besides, the proposed beamforming derivation achieves a remarkable improvement compared to the approximated derivation in previous works. Finally, the complexity analysis confirms that the proposed DRL algorithm reduces the computation complexity compared to the DRL algorithm in the literature.

3.2 Introduction

Recently, the reconfigurable intelligent surfaces (RISs) technology has been proposed as a key enabler to meet the demands of future technologies [1, 2]. RIS is a meta-surface consisting of low-cost passive elements that can be programmed to turn the random nature of wireless channels into a partially deterministic space to improve the propagation of wireless signals [3]. In addition to the RIS technology, full-duplex (FD) transmission has been regarded as a potential approach to increase the spectral efficiency of wireless systems by enabling simultaneous transmission and reception [4, 5].

Incorporating RIS into FD communications can provide new degrees of freedom, facilitating ultra spectrum-efficient communication systems [6]. A number of existing works have studied RIS-assisted FD wireless networks [7, 8, 9]. The works in [7, 8] considered alternating optimization (AO) techniques to optimize the RIS phase shifts in FD systems. The authors in [9] considered a multi RIS-assisted FD system to maximize the weighted system sum-rate, where the non-convex problem was addressed using the AO approach.

The above works that used AO techniques exhibit both loss of optimality and high computational complexity. Deep reinforcement learning (DRL) has emerged as a powerful approach to optimize the RIS phase shifts by overcoming the practical

implementation problems of AO techniques. Furthermore, DRL approaches enable addressing mathematically intractable nonlinear problems directly, without the need for prior relaxations requirements. The work in [10] proposed a DRL algorithm to maximize the rate, where both half-duplex and FD operating modes are considered together. However, only a single RIS deployment was investigated. The rapid changes in dynamic environments can obliterate/annihilate the RIS deployment benefits when the corresponding link is blocked/weak. In such cases, deploying distributed power-efficient RISs can cooperatively enhance the coverage of the system by providing multiple paths of received signals. Moreover, the computational complexity can be further reduced. To the best of the authors' knowledge, this work is the first of its kind, which utilizes DRL for investigating the performance of single and distributed RIS deployment schemes in FD multiple-input single-output (MISO) systems in the literature. Our contributions are summarized as follows:

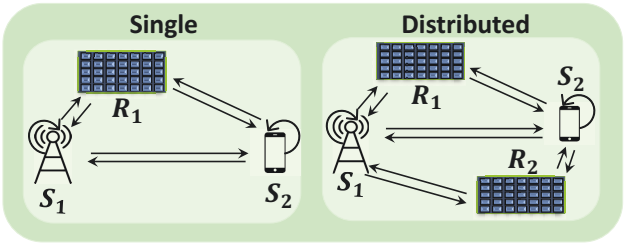
- Three practical scenarios are considered to investigate the sum-rate performance of deploying a single or distributed RIS in an FD-MISO system.
- A closed-form solution is derived to optimize the transmit beamformers, which provides a remarkable improvement in the sum-rate compared to the state-of-the-art approximated derivation in [10].
- An improved DRL algorithm is proposed to optimize the RIS phase shifts for both deployment schemes, which achieves a significant improvement in the sum-rate compared to the non-optimized scenarios.
- The proposed DRL algorithm provides a considerable reduction in the computational complexity compared to the DRL algorithm in [10].
- The complexity analysis and Monte Carlo simulations support the findings.

The rest of this chapter is organized as follows: Section 3.3 presents the system model and problem formulation. The proposed DRL algorithm is introduced in Section 3.4. Simulation results and conclusions are presented in Sections 3.5 and 3.6, respectively.

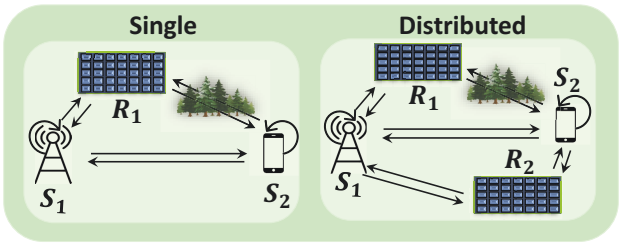
3.3 System Model and Problem Formulation

Consider an RIS-assisted FD MISO system, where single and distributed RIS deployment schemes are investigated. S_1 and S_2 represent the base station (BS) and user equipment (UE), respectively. Both the BS and UE are equipped with M transmit antennas and one receive antenna. The r-th RIS, R_r, consists of N_r programmable reflecting elements. Note that the total number of elements for both deployment schemes is defined as $N = N_r\Lambda$ to ensure the same number of RIS elements for all scenarios, where Λ is the number of RISs. As illustrated in Fig. 3.1, three scenarios are investigated based on the links' quality. In the first scenario, the single and distributed RIS deployment schemes have strong line-of-sight (LoS) components in all links. Scenarios 2 and 3 assume that the links R_1-S_2 and S_1-R_2 are weak due to obstacles, respectively. It is worth noting that from a practical point of view, it is more probable that the longer distance links (i.e., R_1-S_2 and S_1-R_2) may experience blockage since the short-distance links are planned deployment links. It also ensures a fair comparison between the two deployment schemes as the RIS benefits are embraced in all scenarios.

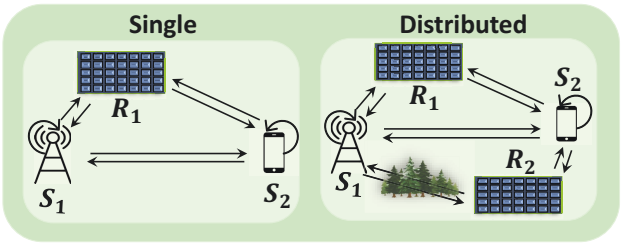
Given $\bar{i} = 3 - i \; \forall \, i \in \{1, 2\}$, let $\mathbf{H}_{S_{\bar{i}}R_r} \in \mathbb{C}^{N_r \times M}$, $\mathbf{h}^H_{R_r S_i} \in \mathbb{C}^{1 \times N_r}$, and $\mathbf{h}^H_{S_{\bar{i}}S_i} \in \mathbb{C}^{1 \times M}$ denote the channel coefficients of the $S_{\bar{i}}$-R_r, R_r-S_i, and $S_{\bar{i}}$-S_i links, respectively. The self-interference (SI) channels of both the BS and UE are denoted by $\mathbf{h}^H_{S_iS_i} \in \mathbb{C}^{1 \times M}$. Hence, the noisy received signal, y_i, is



(a) Scenario 1



(b) Scenario 2



(c) Scenario 3

Fig. 3.1: RIS-assisted FD MISO system.

$$y_i = \Big(\underbrace{\sum_{r\in\Lambda} \mathbf{h}_{R_r S_i}^H \boldsymbol{\Theta}_r \mathbf{H}_{S_{\bar{i}} R_r}}_{\text{Reflected signal}} + \underbrace{\mathbf{h}_{S_{\bar{i}} S_i}^H}_{\text{Direct signal}}\Big)\mathbf{w}_{\bar{i}} x_{\bar{i}} + \underbrace{\mathbf{h}_{S_i S_i}^H \mathbf{w}_i x_i}_{\text{Residual SI}} + n,$$

$$i = 1, 2, \;\; \Lambda = \begin{cases} 1 & \text{Single RIS} \\ 2 & \text{Distributed RIS,} \end{cases} \tag{3.1}$$

where $n \sim \mathcal{CN}(0, \sigma^2)$ denotes the additive white complex Gaussian noise with zero-mean and variance σ^2. The diagonal matrix $\boldsymbol{\Theta}_r = \text{diag}\big(e^{j\varphi_{r1}}, \cdots, e^{j\varphi_{rn}}, \cdots, e^{j\varphi_{rN_r}}\big) \in$

$\mathbb{C}^{N_r \times N_r}$ represents the phase shifts of R_r, where $\varphi_{rn} \in [-\pi, \pi)$ is the phase shift introduced by the n-th reflecting element. The source node, S_i, employs an active beamforming $\mathbf{w}_i \in \mathbb{C}^{M \times 1}$ to transmit the information signal, x_i, with $\mathbb{E}\{|x_i|^2\} = 1$, where $\mathbb{E}\{\cdot\}$ denotes the expectation operation.

Based on (3.1), the received signal-to-interference plus-noise ratio, γ_i, and achievable rate, $\mathcal{R}_i$, measured in bit per second per Hertz (bps/Hz), are respectively given as

$$\gamma_i = \frac{\left|\left(\sum_{r\in\Lambda} \mathbf{h}_{R_r S_i}^H \mathbf{\Theta}_r \mathbf{H}_{S_{\bar{i}} R_r} + \mathbf{h}_{S_{\bar{i}} S_i}^H\right) \mathbf{w}_{\bar{i}}\right|^2}{|\mathbf{h}_{S_i S_i}^H \mathbf{w}_i|^2 + \sigma^2}, i = 1, 2, \ \Lambda = \begin{cases} 1 & \text{Single RIS} \\ 2 & \text{Distributed RIS,} \end{cases} \tag{3.2}$$

and

$$\mathcal{R}_i = \log_2 (1 + \gamma_i) . \tag{3.3}$$

The objective is to maximize the sum-rate by optimizing the beamformers and RIS phase shifts, and is formulated as

$$\text{(P1)} \quad \max_{\mathbf{w}_i, \bar{\mathbf{\Theta}}} \quad \sum_{i=1}^{2} \mathcal{R}_i \tag{3.4a}$$

$$\text{s.t.} \quad -\pi \leq \varphi_{rn} \leq \pi, \ n = 1, \cdots, N_r, \tag{3.4b}$$

$$||\mathbf{w}_i||^2 \leq P_{\max}, \ i = 1, 2. \tag{3.4c}$$

Here, $\bar{\mathbf{\Theta}} = \text{diag}\big(\mathbf{\Theta}_1, \mathbf{\Theta}_2\big)$ is a block matrix whose diagonal entries contain the phase shifts of the two RISs for the distributed RIS, and $\bar{\mathbf{\Theta}} = \text{diag}\big(\mathbf{\Theta}_1\big)$ when a single RIS is considered. $P_{\max}$ is the maximum transmitted power of S_i.

It is worth noting that (P1) is challenging to solve due to the non-convexity of the objective function and constraints. Thus, an efficient solution is proposed which decouples the problem into two sub-problems.

3.4 Proposed Solution

This section proposes a novel algorithm to solve (P1). First, a closed-form solution is derived to optimize the transmit beamformers, $\mathbf{w}_i^*$, for a fixed $\bar{\mathbf{\Theta}}$. Then, the RIS phase shifts, $\bar{\mathbf{\Theta}}$, are obtained using the proposed DRL algorithm. This process is repeated until $\bar{\mathbf{\Theta}}^*$ and $\mathbf{w}_i^*$ converge. In what follows, more details about the two-step solution are provided.

3.4.1 Beamformers Optimization for a Given $\bar{\mathbf{\Theta}}$

The mutual information $I(s;y)$ with an arbitrary input probability distribution $p(s)$ for a channel with input s, output y, and a transition probability of $p(y|s)$ is given by

$$I(s;y) = \max_{q(s|y)} \mathbb{E}\big[\log\left(q(s|y)\right) - \log\left(p(s)\right)\big], \tag{3.5}$$

where the optimal $q^*(s|y)$ is the posterior probability [8], and is expressed as $q^*(s|y) = \frac{p(s)p(y|s)}{p(y)} \triangleq p(s|y)$. Based on (3.5), the achievable rate of S_i is

$$\mathcal{R}_i = \max_{q(s_i|y_{\bar{i}})} \mathbb{E}\big[\log\left(q(s_i|y_{\bar{i}})\right) - \log\left(p(s_i)\right)\big], \tag{3.6}$$

where the input probability distribution $p(s_i)$ is $\mathcal{CN}(0,1)$ and the channel transition probability $p(y_{\bar{i}}|s_i)$ is obtained from (3.1). According to [11], $p(s_i|y_{\bar{i}})$ follows the complex Gaussian distribution of $\mathcal{CN}(f_{\bar{i}}^* y_{\bar{i}}, \Sigma_{\bar{i}}^*)$. $\Sigma_{\bar{i}}^*$ is defined as $\Sigma_{\bar{i}}^* = 1 - f_{\bar{i}}^* b_{\bar{i}}$, where $f_{\bar{i}}^*$ and $b_{\bar{i}}$ are respectively expressed as

$$f_{\bar{i}}^{*} = \frac{b_{\bar{i}}}{b_{\bar{i}}^{2} + |\mathbf{h}_{S_{\bar{i}}S_{\bar{i}}}^{H}\mathbf{w}_{\bar{i}}|^{2}}, \tag{3.7}$$

and

$$b_{\bar{i}} = \left| \left(\sum_{r\in\Lambda} \mathbf{h}_{R_r S_{\bar{i}}}^{H} \boldsymbol{\Theta}_r \mathbf{H}_{S_i R_r} + \mathbf{h}_{S_i S_{\bar{i}}}^{H} \right) \mathbf{w}_i \right|^{2}. \tag{3.8}$$

To this end, (3.4a) in (P1) can be re-expressed as

$$\max_{\mathbf{w}_i,\, \bar{\boldsymbol{\Theta}},\, f_i,\, \Sigma_i} \quad \sum_{i=1}^{2} \mathbb{E}\big[\log(p(s_i|y_{\bar{i}})) - \log(p(s_i))\big]. \tag{3.9}$$

Let $\boldsymbol{\alpha}_{\bar{i}} = \sum_{r\in\Lambda} \mathbf{h}_{R_r S_{\bar{i}}}^{H} \boldsymbol{\Theta}_r \mathbf{H}_{S_i R_r} + \mathbf{h}_{S_i S_{\bar{i}}}^{H}$ and $b_{\bar{i}} = |\boldsymbol{\alpha}_{\bar{i}}\mathbf{w}_i|^2$. The expectation term in (3.9) is calculated as

$$\begin{aligned} \mathbb{E}\big[\log(\mathcal{CN}(f_{\bar{i}}y_{\bar{i}}, \Sigma_{\bar{i}})) - \log(\mathcal{CN}(0,1))\big] &= \exp\left(f_i y_{\bar{i}} + \frac{\Sigma_{\bar{i}}}{2}\right) - \exp\left(\frac{1}{2}\right) \\ &= -\frac{1}{2} f_{\bar{i}} |\boldsymbol{\alpha}_{\bar{i}}\mathbf{w}_i|^2 + \mathbf{w}_i \left(f_{\bar{i}}\boldsymbol{\alpha}_{\bar{i}} + f_i \mathbf{h}_{S_i S_i}^{H}\right). \end{aligned} \tag{3.10}$$

Furthermore, let $\boldsymbol{\beta}_{\bar{i}} = f_{\bar{i}}\boldsymbol{\alpha}_{\bar{i}} + f_i \mathbf{h}_{S_i S_i}^{H}$. Thus, (3.10) can be defined as a convex quadratically constrained quadratic program:

$$-\frac{1}{2} f_{\bar{i}} |\boldsymbol{\alpha}_{\bar{i}}\mathbf{w}_i|^2 + \mathbf{w}_i \boldsymbol{\beta}_{\bar{i}}, \tag{3.11}$$

where its solution can be derived as [12]

$$\mathbf{w}_i^{*} = (v^{*} + f_{\bar{i}}\boldsymbol{\alpha}_{\bar{i}}\boldsymbol{\alpha}_{\bar{i}}^{H})^{-1}\boldsymbol{\beta}_{\bar{i}}. \tag{3.12}$$

Here, v^{*} is the optimal dual Lagrangian variable associated with the power constraint and is found by performing a bisection search over the interval $\left[0, \sqrt{\boldsymbol{\beta}_{\bar{i}}^{T}\boldsymbol{\beta}_{\bar{i}}}/\sqrt{P_{\max}}\right]$ [12].

3.4.2 Phase Shift Optimization for a Given $\mathbf{w}_i$ and $\mathbf{w}_{\bar{i}}$

Model-free RL can be employed to address a decision-making problem by learning the optimal solution in dynamic environments. Therefore, the RIS-assisted FD MISO system represents the DRL environment and the RIS controller represents the DRL agent. At each time step t, the agent observes the current state, s_t, from the environment, takes an action, a_t, based on a policy, $\tilde{\pi}$, receives a reward, r_t, of executing a_t, and transitions to a new state s_{t+1}. The key elements of DRL are defined as follows: The *state space* at time step t, includes $\varphi_{rn} \forall\, n = 1, \cdots, N_r$ and the corresponding $\sum_{i=1}^{2} \mathcal{R}_i$ at time step $t-1$, i.e., $s_t = \left[\sum_{i=1}^{2} \mathcal{R}_i^{(t-1)}, \varphi_{r1}^{(t-1)}, \cdots, \varphi_{rn}^{(t-1)}, \cdots, \varphi_{rN_r}^{(t-1)}\right]$. The *action space* at time step t is expressed as $a_t = \left[\varphi_{r1}^{(t)}, \cdots, \varphi_{rn}^{(t)}, \cdots, \varphi_{rN_r}^{(t)}\right]$, and the *reward* at time step t is $r_t = \sum_{i=1}^{2} \mathcal{R}_i^{(t)}$.

The goal of a RL agent is to learn a policy that maximizes the expected cumulative discounted reward from the start state, as: $J(\tilde{\pi}) = \mathbb{E}\left[R_1|\tilde{\pi}\right]$. The policy gradient based algorithms can be used to learn the optimal policy for continuous a_t. In particular, the proposed algorithm aims at maximizing the return by training deep neural networks (DNN) to approximate the Q-value function. It is based on the *actor-critic* approach, which consists of two DNN models: *actor*, $\mu(s_t|\boldsymbol{\theta}_\mu)$, and *critic*, $Q(s_t, a_t|\boldsymbol{\theta}_q)$, where $\boldsymbol{\theta}$ represents the DNN parameters. The actor takes the state as an input and outputs $a_t = \mu(s_t|\boldsymbol{\theta}_\mu) + \xi$, where ξ is a random process that is added to the actions for exploration, representing the policy network. The critic takes s_t and a_t as an input and outputs the Q-value, representing the evaluation network [13].

At the initialization stage, four networks are generated, i.e., target and evaluation DNN. The target networks are generated by making a copy of the actor and critic evaluation NNs, $\mu'(s_t|\boldsymbol{\theta}_{\mu'})$ and $Q'(s_t, a_t|\boldsymbol{\theta}_{q'})$. The experience replay with memory D is built to reduce the correlation of the training samples. During each episode, all the channel state information is obtained. Then, the agent takes a_t generated by the

actor network, calculates the r_t, and transitions to s_{t+1}. The experience is then stored (s_t, a_t, r_t, s_{t+1}) into D, and the critic evaluation network randomly samples a minibatch transitions, N_B, to calculate the target value y_j, as

$$y_j = r_j + \rho Q'(s_{j+1}, \mu'(s_{j+1}|\boldsymbol{\theta}_{\mu'})|\boldsymbol{\theta}_{q'}), \tag{3.13}$$

where $\rho \in (0, 1]$ is the discount factor. The actor and critic NN parameters, $\boldsymbol{\theta}_\mu$ and $\boldsymbol{\theta}_q$, are updated using the stochastic gradient descent and policy gradient, respectively, as

$$L = \frac{1}{N_B} \sum_j \left(y_j - Q(s_j, a_j|\boldsymbol{\theta}_q)\right)^2, \tag{3.14}$$

and

$$\nabla_{\boldsymbol{\theta}_\mu} = \frac{1}{N_B} \sum_j \nabla_a Q(s, a|\boldsymbol{\theta}_q)|_{s=s_j, a=\mu(s_j)} \nabla_{\boldsymbol{\theta}_\mu} \mu(s|\boldsymbol{\theta}_\mu)|_{s_j}. \tag{3.15}$$

Finally, the target NN parameters are updated using a soft update coefficient, τ, as

$$\boldsymbol{\theta}_{q'} \longleftarrow \tau \boldsymbol{\theta}_q + (1 - \tau)\boldsymbol{\theta}_{q'}, \tag{3.16}$$

and

$$\boldsymbol{\theta}_{\mu'} \longleftarrow \tau \boldsymbol{\theta}_\mu + (1 - \tau)\boldsymbol{\theta}_{\mu'}. \tag{3.17}$$

This process is repeated for K and T until convergence is reached. The structure of the proposed DRL algorithm is illustrated in Fig. 3.2 and summarized in Algorithm 1.

3.4.3 Proposed DNN Design

The proposed DNN models are designed as feedforward fully connected NNs. The proposed algorithm contains four NNs (actor and critic for each evaluation and target network). Each NN has an input layer, two hidden layers and output layer, as shown

Algorithm 1 Proposed DRL algorithm.

Initialize: $\boldsymbol{\theta}_{\mu}$ and $\boldsymbol{\theta}_{q}$ with random weights, D, ρ, τ, learning rate ν, $\boldsymbol{\theta}_{\mu'} \leftarrow \boldsymbol{\theta}_{\mu}$ and $\boldsymbol{\theta}_{q'} \leftarrow \boldsymbol{\theta}_{q}$;

1: **repeat** for K episodes:
2: Collect the channels of the k-th episode;
3: Randomly initialize $\varphi_{rn} \, \forall \, n = 1, \cdots, N_r$;
4: Calculate $\mathbf{w}_{\bar{i}}$ using (3.12);
5: Initialize $\xi \sim \mathcal{CN}(0, 0.1)$;
6: **repeat** for T steps:
7: Obtain $a_t = \mu(s_t|\boldsymbol{\theta}_{\mu}) + \xi$ from the actor network and reshape it;
8: Repeat **Line** #4;
9: Observe the new state, s_{t+1}, given a_t;
10: Store (s_t, a_t, r_t, s_{t+1}) in D;
11: When D is full, sample a minibatch of N_B transitions (s_j, a_j, r_j, s_{j+1}) randomly from D;
12: Compute the target value from (3.13);
13: Update the critic using (3.14);
14: Update the actor using (3.15);
15: Update the target NNs using (3.16) and (3.17);

Output: Optimal action that corresponds to the optimal $\bar{\boldsymbol{\Theta}}^{*}$.

in Fig. 3.2. The input layer of the actor and critic networks contains $N + 1$ neurons (i.e., size of s_t). The input of the actor is passed to two hidden layers, each having ψ_i neurons, where ψ_i is the number of neurons of the i-th layer. On the other hand, the input of the critic network is passed to the first hidden layer that is concatenated with a_t (i.e., size of $\psi_i + N$), and then passed to the second hidden layer. The two hidden layers for each of the actor and critic networks use the $ReLU$ activation function whereas the output layer of the actor network uses the *tanh* activation function. The output layer of the actor and critic networks contains N neurons (i.e., size of a_t) and one neuron (i.e., Q-value), respectively.

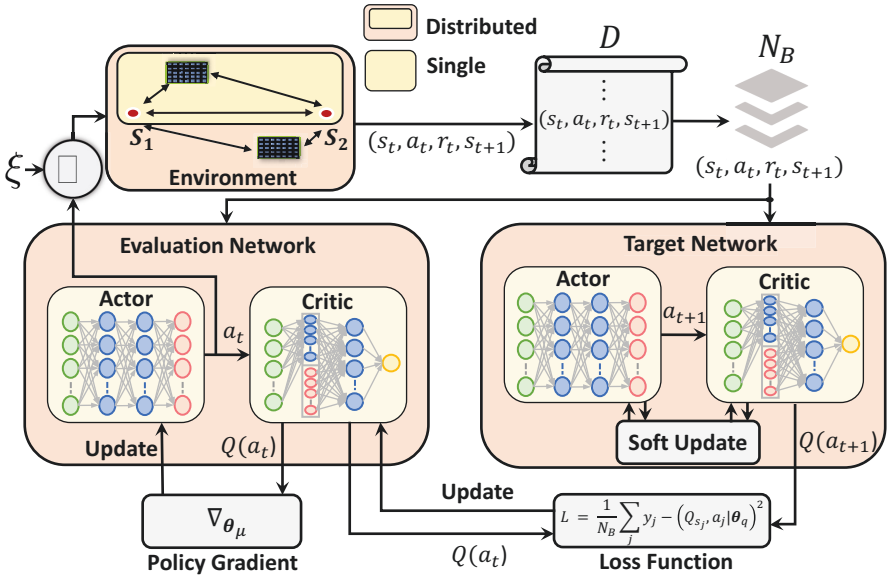


Fig. 3.2: The proposed DRL algorithm structure.

3.4.4 Complexity Analysis

The computational complexity of the proposed DRL algorithm is analyzed in terms of the number of NN parameters $C_{\mathcal{P}}$ required to be stored, real additions $C_{\mathcal{A}}$, and real multiplications $C_{\mathcal{M}}$. It is worth noting that, for simplicity, each activation function is considered to cost one real addition. Henceforth, the complexity for the proposed DRL algorithm based on the NNs design is given as

$$C_{\mathcal{P}} = 2\left(\sum_{i=1}^{3}(\psi_i^{\mathrm{A}}+1)\psi_{i+1}^{\mathrm{A}} + \sum_{i=1}^{3}(\psi_i^{\mathrm{C}}+1)\psi_{i+1}^{\mathrm{C}}\right), \tag{3.18}$$

$$C_{\mathcal{M}} = 2\left(\sum_{i=1}^{3}\psi_i^{\mathrm{A}}\psi_{i+1}^{\mathrm{A}} + \sum_{i=1}^{3}\psi_i^{\mathrm{C}}\psi_{i+1}^{\mathrm{C}}\right), \tag{3.19}$$

$$C_{\mathcal{A}} = 2\left(\sum_{i=1}^{3}\psi_i^{\mathrm{A}}\psi_{i+1}^{\mathrm{A}} + \sum_{i=1}^{3}\psi_{i+1}^{\mathrm{A}} + \sum_{i=1}^{3}\psi_i^{\mathrm{C}}\psi_{i+1}^{\mathrm{C}} + \sum_{i=1}^{3}\psi_{i+1}^{\mathrm{C}}\right), \tag{3.20}$$

where the actor and critic networks are expressed through the superscripts A and C, respectively. The complexity reduction of using the proposed DRL algorithm over the

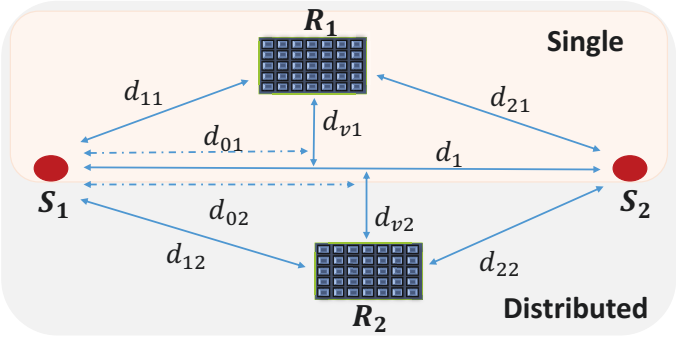


Fig. 3.3: Simulation setup.

algorithm in [10], where the computational complexity is presented in (2.24-2.25), for the single RIS-assisted FD system is

$$\text{Reduction} = 1 - \frac{\{C_{\chi}^{\text{A}} + C_{\chi}^{\text{C}}\}_{\text{Proposed}}}{\{C_{\chi}^{\text{A}} + C_{\chi}^{\text{C}}\}_{[10]}}, \ \chi \in \{\mathcal{P}, \mathcal{A}, \mathcal{M}\}. \tag{3.21}$$

3.5 Simulation Results

Figure 3.3 illustrates the simulation setup, where the considered parameters are: $d_{v1} = d_{v2} = 2\,\text{m}$ and $d_1 = 50\,\text{m}$. The distances between the links are: $d_{11} = \sqrt{d_{01}^2 + d_{v1}^2}\,\text{m}$, $d_{12} = \sqrt{d_{02}^2 + d_{v2}^2}$ m, $d_{21} = \sqrt{(d_1 - d_{01})^2 + d_{v1}^2}$ m, and $d_{22} = \sqrt{(d_1 - d_{02})^2 + d_{v2}^2}$ m. The path loss (PL) at distance d_{ir} is modeled as $\text{PL} = PL_0 - 10\zeta \log_{10}\left(\frac{d_{ir}}{D_r}\right)$ [14], where PL_0 is the PL at a reference distance D_r and ζ is the PL exponent, in which $PL_0 = -35.6\,\text{dB}$ and $D_r = 1\,\text{m}$. The channels are modeled as Rayleigh fading whenever a blocking element exists. Otherwise, the channels are modeled as Rician with a factor of 10. The PL exponents of the S_1-S_2, S_1-R_r, and S_2-R_r channels are set to $\zeta_{\text{BU}} = 4$, $\zeta_{\text{BR}} = 2.1$, and $\zeta_{\text{UR}} = 2.2$, respectively [9]. The PL of the SI channels is $-95\,\text{dB}$. The total transmit power is $P = 15\,\text{dBm}$, while the noise power is $\sigma^2 = -80\,\text{dBm}$ [7].

The parameters of the proposed DRL are as follows: $T = 800$, $K = 500$, $N_B = 16$,

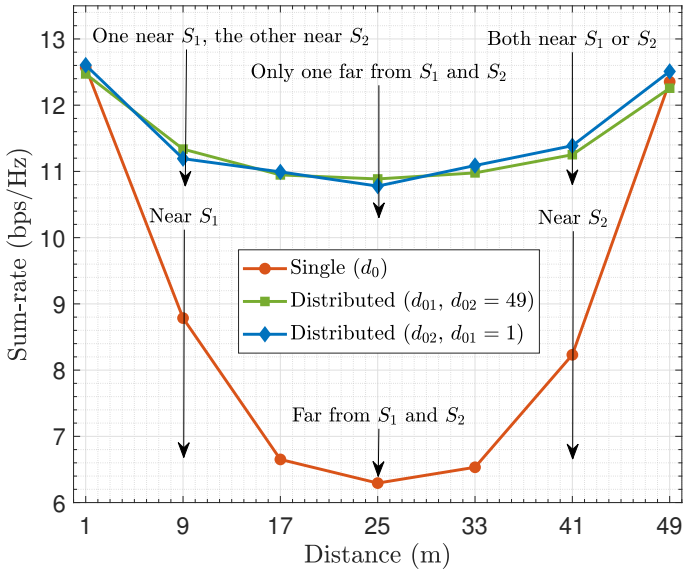


Fig. 3.4: RIS deployment investigation.

$\nu_{\mathrm{A}} = 0.0001$, $\nu_{\mathrm{C}} = 0.0002$, decaying rate $= 0.0001$, $\rho = 0.99$, $\tau = 0.001$, and $D = 50000$. Both actor and critic networks use the Adam optimizer for updating the parameters. The number of neurons of the hidden layers are, $\psi_1 = 100$ and $\psi_2 = 45$.

To validate the performance of the proposed algorithm, it is compared with the non-optimized scheme, referred to as random phase shifts. Furthermore, it is compared with the algorithm in [10] for the single RIS-assisted FD system to show the superiority of the proposed beamforming derivations over the approximated derivations in [10]. To ensure a fair comparison, it is assumed that N is the same for both deployment schemes. Hence, each RIS in the distributed scheme has half the number of elements of the single scheme.

Figure 3.4 studies the RIS deployment problem in both single and distributed RIS-assisted FD system. In the single RIS scheme, the sum-rate gradually increases when the RIS gets closer to S_1 or S_2. Generally, the RIS is deployed near the BS to best benefit from the RIS reflection links, ensuring a strong LoS link between the RIS

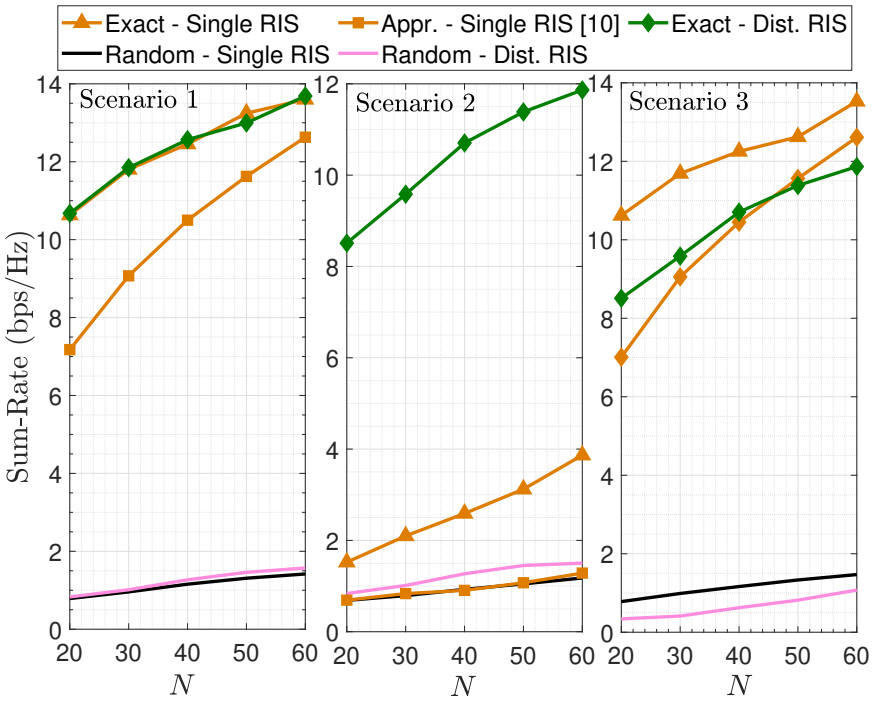


Fig. 3.5: The impact of varying N on the system performance.

and BS. Therefore, in the distributed RIS scheme, two cases are considered: varying d_{01} when $d_{02} = 49\,\text{m}$ and varying d_{02} when $d_{01} = 1\,\text{m}$. As both RISs get near the ends or if one is fixed near S_1 and the other is near S_2, the sum-rate increases. It is shown that when the RIS is located relatively far from both S_1 and S_2 in the single RIS scheme, the distributed RIS scheme significantly improves the sum-rate. This is because deploying distributed RISs enables providing alternative paths when the other RIS experiences a poor quality link. For the rest of the chapter, it is considered that $d_{01} = 1\,\text{m}$ and $d_{02} = 49\,\text{m}$.

Figure 3.5 illustrates the effect of increasing N on the system performance. Three practical scenarios are considered to investigate the preference of using single or distributed RIS schemes. In Scenario 1, the distributed and single RIS schemes achieve a similar performance due to the strong LoS components (i.e., good quality links), and N is the same in both schemes. In Scenario 2, the results illustrate that the

distributed RIS system significantly outperforms the single RIS system when the R_1-S_2 link is blocked/weak. In this case, the distributed RIS scheme has a higher sum-rate since it compensates for the poor quality link by providing an alternative path. On the other hand, if the link between R_2-S_2 is blocked/weak, as in Scenario 3, the single RIS scheme outperforms the distributed RIS since the former has double the number of elements compared to the latter. It is also worth noting that the proposed DRL algorithm provides a significant improvement in the sum-rate for the single and distributed RIS schemes compared to the random RIS phase shifts in all scenarios. The performance of the studied scenarios provides important insights into the preference of each deployment scheme based on the link conditions. Scenario 1 further points that the deployment cost should be considered if both schemes yield similar performance, as the required channel state information of the single RIS scheme is less than that of the distributed RIS scheme.

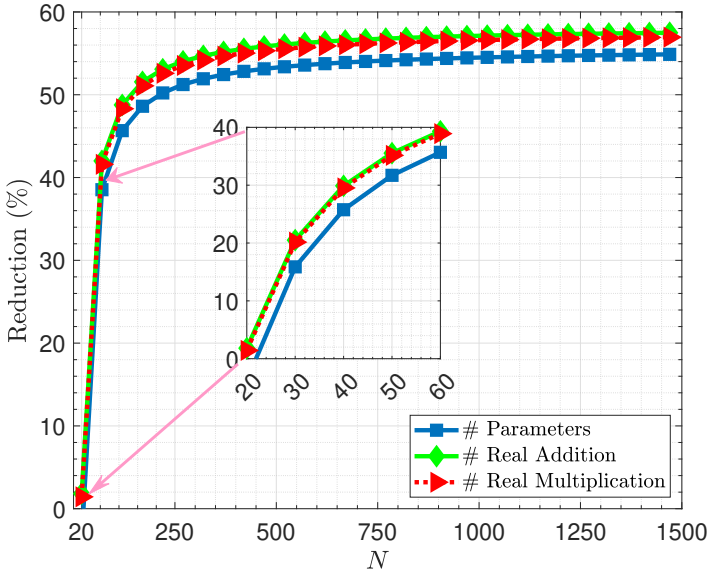


Fig. 3.6: Complexity reduction percentage versus N.

In the single RIS scheme, the proposed beamforming derivation improves the

sum-rate performance in all scenarios, when compared to [10], as depicted in Fig. 3.5. Moreover, as shown in Fig. 3.6, the proposed DRL algorithm provides a complexity reduction percentage up to 40% for the range of N from 20 to 60 compared to the DRL presented in [10], and it saturates at 57% when N is very large.

3.6 Conclusion

This chapter optimized the beamformers and RIS phase shifts to maximize the sum-rate for both single and distributed RIS deployment schemes. Three practical scenarios were considered to investigate the preference of using single or distributed RIS deployment schemes. A closed-form solution is derived to obtain the optimal beamformers, and a novel DRL algorithm is considered for the RIS phase shifts optimization. It was shown that the superiority of a deployment scheme depends on the links' quality. Compared to the non-optimized scenarios, the proposed algorithm significantly improved the sum-rate for both deployment schemes. The proposed DRL algorithm achieved up to 57% complexity reduction compared to the DRL algorithm in the literature. Future works may consider generalizing the proposed DRL by jointly optimizing the beamformers and RIS phase shifts for multi-user systems.

References

[1] I. Al-Nahhal, O. A. Dobre, and E. Basar, "Reconfigurable intelligent surface-assisted uplink sparse code multiple access," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 2058–2062, Feb. 2021.

[2] I. Al-Nahhal, O. A. Dobre, E. Basar, T. M. N. Ngatched, and S. Ikki,

"Reconfigurable intelligent surface optimization for uplink sparse code multiple access," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 133–137, Jan. 2022.

[3] M. A. ElMossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, "Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 3, pp. 990–1002, Sep. 2020.

[4] R. Askar, J. Chung, Z. Guo, H. Ko, W. Keusgen, and T. Haustein, "Interference handling challenges toward full duplex evolution in 5G and beyond cellular networks," *IEEE Wirel. Commun.*, vol. 28, no. 1, pp. 51–59, Feb. 2021.

[5] A. Yadav and O. A. Dobre, "All technologies work together for good: A glance at future mobile networks," *IEEE Wirel. Commun.*, vol. 25, no. 4, pp. 10–16, Aug. 2018.

[6] R. Alghamdi, R. Alhadrami, D. Alhothali, H. Almorad, A. Faisal, S. Helal, R. Shalabi, R. Asfour, N. Hammad, A. Shams, N. Saeed, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Intelligent surfaces for 6G wireless networks: A survey of optimization and performance analysis techniques," *IEEE Access*, vol. 8, pp. 202795–202818, Oct. 2020.

[7] H. Shen, T. Ding, W. Xu, and C. Zhao, "Beamforming design with fast convergence for IRS-aided full-duplex communication," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2849–2853, Dec. 2020.

[8] Y. Zhang, C. Zhong, Z. Zhang, and W. Lu, "Sum rate optimization for two way communications with intelligent reflecting surface," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1090–1094, May 2020.

[9] M. A. Saeidi, M. J. Emadi, H. Masoumi, M. R. Mili, D. W. K. Ng, and I. Krikidis, "Weighted sum-rate maximization for multi-IRS-assisted full-duplex systems with hardware impairments," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 466–481, Jun. 2021.

[10] A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Deep reinforcement learning for optimizing RIS-assisted HD-FD wireless systems," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3893–3897, Dec. 2021.

[11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1997.

[12] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[13] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, May 2016, pp. 1–14.

[14] K. Feng, Q. Wang, X. Li, and C.-K. Wen, "Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 745–749, May 2020.

Chapter 4

Distributed RIS-Assisted FD Systems with Discrete Phase Shifts: A Reinforcement Learning Approach

4.1 Abstract

This chapter studies the sum-rate maximization problem of a distributed reconfigurable intelligent surface (RIS)-assisted full-duplex wireless system, where the availability of finite-resolution phase shifts at the RIS is considered. The aim is to optimize the transmit beamformers and RIS phase shifts, subject to the practical discrete phase shift and power constraints. The optimization problem is decoupled into two sub-problems; transmit beamforming and RIS phase shifts optimization. The transmit beamforming problem is mathematically addressed using approximate and closed-form solutions, while the discrete RIS phase shifts are optimized using a reinforcement learning (RL)

approach. The existence and absence of a strong direct line-of-sight is investigated to show the effect of the phase shift optimization on the sum-rate. Simulation results illustrate that the proposed RL for the discrete phase shifts optimization provides a near-optimal performance with a small number of bits (i.e., $b = 6$) even for a large number of RIS elements, while improving the sum-rate compared to the random phase shift scenario and reducing the computational complexity compared to the state-of-the-art works.

4.2 Introduction

Driven by the high data rate demands and rapid advancements of wireless communications, next-generation wireless networks should support massive network capacity and reliability [1, 2]. The reconfigurable intelligent surfaces (RISs) have recently attracted significant attention not only for controlling the propagation environment but also for offering a competitive low-cost solution [3, 4]. RIS is composed of a low-cost nearly passive two-dimensional electromagnetic elements that enable operating with low power consumption while providing relatively high energy efficiency. Therefore, RIS is envisioned as an important technology that will play a crucial role in 6G and beyond wireless communications.

With the aid of RIS, full-duplex (FD) communication systems have the potential to effectively double the spectral efficiency [5]. Several works have exploited alternating optimization (AO) algorithms for RIS-assisted FD systems [5, 6, 7, 8]. The work in [6] proposed an AO algorithm for the discrete RIS phase shift optimization of two-way device-to-device multi-pair orthogonal frequency division multiplexing systems. Moreover, the authors in [7] studied the weighted system sum-rate of a multi RIS-assisted FD system while considering an ideal phase shift model. The authors in

[8] investigated the weighted minimum rate maximization problem for a multi-user RIS-assisted FD system. However, utilizing AO techniques results in an increased computational complexity and sub-optimal solution. Furthermore, as the complexity of phase shift optimization increases (in large scale systems), such approaches become less efficient [9].

In contrast, deep reinforcement learning (DRL) has emerged as an efficient and stable framework in wireless communications [10, 11, 12]. In particular, the authors in [11] considered the rate maximization problem of a half-duplex-FD RIS-assisted system and proposed a DRL algorithm to solve it. Furthermore, [12] proposed a DRL algorithm for single and distributed RIS deployment schemes. However, the existing works on RIS-assisted FD systems assume an ideal phase shift model (i.e., continuous values), which is infeasible to implement due to hardware limitations [13, 14, 15]. Therefore, this work considers optimizing the discrete RIS phase shifts utilizing an efficient DRL approach.

It is worth mentioning that using DRL for optimizing the practical phase shifts in distributed RIS-assisted FD multiple-input single-output (MISO) systems has not been studied in the state-of-the-art works. The main contributions of our work are summarized as follows:

- A DRL algorithm for optimizing the discrete phase shifts of a distributed RIS-assisted FD network is proposed. The proposed algorithm is shown to achieve promising results compared to the continuous-baseline in [12].
- A novel NN design to learn the Q-value of the proposed DRL algorithm is designed and its complexity is investigated.
- Two mathematical solutions for optimizing the transmit beamformers are considered.
- The performance of the proposed DRL algorithm is assessed through extensive

simulations, by considering two scenarios: the presence of the line-of-sight (LoS) link and when it is blocked.

The rest of this chapter is organized as: Section 4.3 describes the system model and problem formulation. Section 4.4 introduces the proposed algorithm. Sections 4.5 and 4.6 present simulation results and conclusions, respectively.

4.3 System Model and Problem Formulation

This chapter considers a distributed RIS-assisted FD MISO system. The RISs are used to enhance the communication between S_1 and S_2 by optimizing their phase shifts via an RIS controller, as illustrated in Fig. 4.1. S_1 and S_2 denote the base station (BS) and user equipment (UE), respectively, and are equipped with M transmit antennas and one receive antenna. Each RIS, R_r, contains N_r elements and the total number of reflecting elements is denoted by N. For simplicity, it is assumed that we have only two RISs, and each RIS contains the same number of elements (i.e., $N_r = N/2$, $r \in \{1, 2\}$). Let $\bar{k} = 3 - k \; \forall \, k \in \{1, 2\}$. The channel coefficients of $S_{\bar{k}}$-R_r, R_r-S_k, and $S_{\bar{k}}$-S_k links are represented as $\mathbf{H}_{S_{\bar{k}}R_r} \in \mathbb{C}^{N_r \times M}$, $\mathbf{h}^H_{R_r S_k} \in \mathbb{C}^{1 \times N_r}$, and $\mathbf{h}^H_{S_{\bar{k}} S_k} \in \mathbb{C}^{1 \times M}$, respectively. Furthermore, $\mathbf{h}^H_{S_k S_k} \in \mathbb{C}^{1 \times M}$ represents the self-interference (SI) channels of S_1 and S_2. Hence, the signal is received from both the reflected and the LoS links, and is expressed as

$$y_i = \Big(\underbrace{\sum_{r=1}^{2} \mathbf{h}_{R_r S_k}^H \boldsymbol{\Theta}_r \mathbf{H}_{S_{\bar{k}} R_r}}_{\text{Reflected signal}} + \underbrace{\xi \mathbf{h}_{S_{\bar{k}} S_k}^H}_{\text{Direct signal}}\Big) \mathbf{w}_{\bar{k}} x_{\bar{k}} + \underbrace{\mathbf{h}_{S_k S_k}^H \mathbf{w}_i x_i}_{\text{Residual SI}} + n,$$

$$\bar{k} = 3 - k \,\forall\, k \in \{1, 2\},\ \xi = \begin{cases} 0 & \text{No LoS} \\ 1 & \text{LoS}, \end{cases} \tag{4.1}$$

where ξ is a factor that represents the LoS link condition, i.e., 0 means that the LoS link does not exists and 1 otherwise, and n is the additive white complex Gaussian noise with zero-mean and variance σ^2, $n \sim \mathcal{CN}(0, \sigma^2)$. Let $\boldsymbol{\Theta}_r = \text{diag}\big(e^{j\varphi_{r1}}, \cdots, e^{j\varphi_{rn}}, \cdots, e^{j\varphi_{rN_r}}\big) \in \mathbb{C}^{N_r \times N_r}$ denote the diagonal matrix whose elements are the phase shifts of R_r. The phase shift of each reflecting element is $\varphi_{rn} \in [-\pi, \pi)$. This chapter considers a practical phase shift model, where the discrete RIS phase shifts are chosen from the following set

$$\Upsilon = \{0, \Delta\varphi, \cdots, \Delta\varphi(K-1)\}, \Delta\varphi = 2\pi/K, K = 2^b, \tag{4.2}$$

where b is the number of quantization bits for each RIS phase shift. The sum-rate, $\mathcal{R}_k$, in bit per second per Hertz (bps/Hz) is expressed as

$$\mathcal{R}_k = \log_2 \left(1 + \frac{\left|\left(\sum_{r=1}^{2} \mathbf{h}_{R_r S_k}^H \boldsymbol{\Theta}_r \mathbf{H}_{S_{\bar{k}} R_r} + \xi \mathbf{h}_{S_{\bar{k}} S_k}^H\right) \mathbf{w}_{\bar{k}}\right|^2}{|\mathbf{h}_{S_k S_k}^H \mathbf{w}_k|^2 + \sigma^2}\right),$$

$$\bar{k} = 3 - k \,\forall\, k \in \{1, 2\},\ \xi = \begin{cases} 0 & \text{No LoS} \\ 1 & \text{LoS}, \end{cases} \tag{4.3}$$

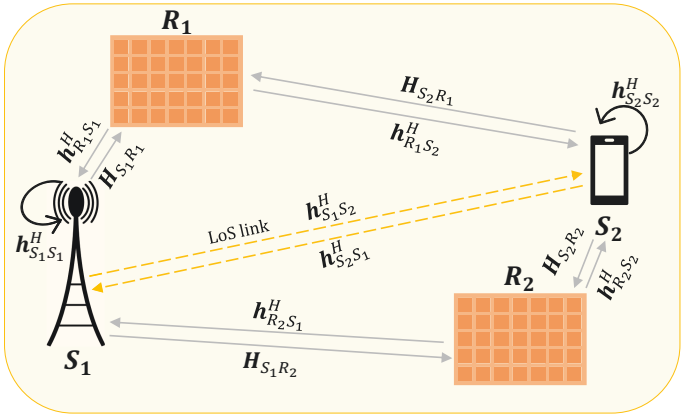


Fig. 4.1: Distributed RIS-assisted FD MISO system.

where $\mathbf{w}_k \in \mathbb{C}^{M\times 1}$ denotes the beamforming for the information signal transmission, x_i. It is considered that $\mathbb{E}\{|x_k|^2\} = 1$, where $\mathbb{E}\{\cdot\}$ is the expectation operation. In this chapter, the aim is to optimize the beamformers and practical RIS phase shifts to maximize the sum-rate. The problem formulation is given as

$$\text{(P1)} \quad \max_{\mathbf{w}_k, \bar{\mathbf{\Theta}}} \quad \sum_{k=1}^{2} \mathcal{R}_k \tag{4.4a}$$

$$\text{s.t.} \quad \varphi_{rn} \in \Upsilon,\ r = 1, 2,\ n = 1, \cdots, N_r, \tag{4.4b}$$

$$||\mathbf{w}_k||^2 \leq P_{\max},\ k = 1, 2. \tag{4.4c}$$

$\bar{\mathbf{\Theta}}$ is a matrix that contains the phase shifts of the two RISs ($\text{diag}\big(\mathbf{\Theta}_1, \mathbf{\Theta}_2\big)$), and $P_{\max}$ is the maximum transmitted power of the source node. Due to the non-convexity of the objective function subjected to a discrete phase shift constraint, (P1) is challenging to solve. Generally, there is no conventional algorithm to efficiently find the optimal solution to (P1). The optimal solution can be found by performing a search over all possible combinations of discrete phase shifts for all the elements. However, this results in a significant computational complexity of the order of $\mathcal{O}(2^{bN})$, which is practically

infeasible for large scale systems [16]. Thus, a two-step efficient solution is proposed to solve such challenging problems.

4.4 Proposed Solution

A practical algorithm to solve (P1) is proposed in this section. A DRL algorithm is proposed to efficiently optimize the discrete RIS phase shifts. For the transmit beamformer, $\mathbf{w}_k^*$, the solutions presented in [17] and [12] are used. The problem is solved iteratively until convergence is reached. The details about the proposed algorithm are provided below.

4.4.1 Beamformers Optimization

4.4.1.1 Approximate Solution

For a fixed $\bar{\mathbf{\Theta}}$, the corresponding beamforming vector $\mathbf{w}_{\bar{k}}$ can be obtained using an approximate solution as detailed in [17]

$$\mathbf{w}_{\bar{k}}^* = (\delta \mathbf{h}_{S_{\bar{k}}S_{\bar{k}}} \mathbf{h}_{S_{\bar{k}}S_{\bar{k}}}^H + v^* \mathbf{k})^{-1} \boldsymbol{\mathcal{B}}, \, k \in \{1, 2\}. \tag{4.5}$$

Here, $\boldsymbol{\mathcal{B}}$ and δ are given as

$$\boldsymbol{\mathcal{B}} \triangleq \frac{1}{|\mathbf{h}_{S_k S_k}^H \mathbf{w}_k|^2 + \sigma^2} \left(1 + \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{|\mathbf{h}_{S_{\bar{k}}S_{\bar{k}}}^H \tilde{\mathbf{w}}_{\bar{k}}|^2 + \sigma^2}\right) \mathbf{h}_{\bar{k}} \mathbf{h}_{\bar{k}}^H \tilde{\mathbf{w}}_{\bar{k}}, \tag{4.6}$$

and

$$\delta \triangleq \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2 \left(|\mathbf{h}_{\bar{k}}^H \tilde{\mathbf{w}}_{\bar{k}}|^2 + |\mathbf{h}_{S_k S_k}^H \mathbf{w}_k|^2 + \sigma^2\right)}{|\mathbf{h}_{S_k S_k}^H \mathbf{w}_k|^2 + \sigma^2 \left(|\mathbf{h}_{S_{\bar{k}}S_{\bar{k}}}^H \tilde{\mathbf{w}}_{\bar{k}}|^2 + \sigma^2\right)^2}, \tag{4.7}$$

where $\tilde{\mathbf{w}}_{\bar{k}}$ is a given feasible point and $\mathbf{h}_{\bar{k}}$ is expressed as

$$\mathbf{h}_{\bar{k}} \triangleq \mathbf{H}^{H}_{S_{\bar{k}}R_r}\boldsymbol{\Theta}^{H}_{r}\mathbf{h}_{R_rS_k} + \mathbf{h}_{S_{\bar{k}}S_k}. \tag{4.8}$$

4.4.1.2 Closed-form Solution

Based on probability theory, the optimal beamformers are derived in [12], where the closed form solution is given as

$$\mathbf{w}^{*}_{k} = (v^{*} + f_{\bar{k}}\boldsymbol{\alpha}_{\bar{k}}\boldsymbol{\alpha}^{H}_{\bar{k}})^{-1}\boldsymbol{\beta}_{\bar{k}},\, k \in \{1,2\}, \tag{4.9}$$

where $\boldsymbol{\alpha}_{\bar{k}}$, $f^{*}_{\bar{k}}$, and $\boldsymbol{\beta}_{\bar{k}}$ are respectively expressed as

$$\boldsymbol{\alpha}_{\bar{k}} = \sum_{r=1}^{2}\mathbf{h}^{H}_{R_rS_{\bar{k}}}\boldsymbol{\Theta}_{r}\mathbf{H}_{S_kR_r} + \xi\mathbf{h}^{H}_{S_kS_{\bar{k}}}, \tag{4.10}$$

$$f_{\bar{k}} = \frac{|\boldsymbol{\alpha}_{\bar{k}}\mathbf{w}_k|^2}{(|\boldsymbol{\alpha}_{\bar{k}}\mathbf{w}_k|^2)^2 + |\mathbf{h}^{H}_{S_{\bar{k}}S_{\bar{k}}}\mathbf{w}_{\bar{k}}|^2}, \tag{4.11}$$

and

$$\boldsymbol{\beta}_{\bar{k}} = f_{\bar{k}}\boldsymbol{\alpha}_{\bar{k}} + f_k\mathbf{h}^{H}_{S_kS_i}. \tag{4.12}$$

In the above formulations, v^{*} denotes the optimal dual Lagrangian coefficient. It is found using the bisection search algorithm, where the search interval for the approximate and closed form solutions are respectively given as $\left[0, ||\boldsymbol{\mathcal{B}}||/\sqrt{P_{\max}}\right]$ and $\left[0, ||\boldsymbol{\beta}_{\bar{k}}||/\sqrt{P_{\max}}\right]$.

4.4.2 Discrete Phase Shift Optimization

4.4.2.1 Overview and DRL Problem Transformation

DRL approaches enable efficient learning of dynamic environments. In particular, the RL agent learns the optimal solution based on experience, through a trial and error approach. After gaining enough experience, the agent chooses the optimized action that maximizes its reward in the environment. In this chapter, the distributed RIS-assisted FD MISO system is regarded as the environment and the RIS controller is the DRL agent. To this end, the problem transformation is given as follows

- State space: The state space at time step t, is the current configuration of the environment. It includes $\varphi_{rn} \forall\, n = 1, \cdots, N_r$ and the corresponding $\sum_{k=1}^{2} \mathcal{R}_k$ at time step $t-1$, and is expressed as

$$s_t = \left[\sum_{k=1}^{2} \mathcal{R}_k^{(t-1)}, \varphi_{r1}^{(t-1)}, \cdots, \varphi_{rn}^{(t-1)}, \cdots, \varphi_{rN_r}^{(t-1)} \right]. \tag{4.13}$$

- Action space: The action space at time step t includes all the available discrete RISs phase shifts, where $\varphi_{rn} \in \Upsilon$. It is defined as

$$a_t = \left[\varphi_{r1}^{(t)}, \cdots, \varphi_{rn}^{(t)}, \cdots, \varphi_{rN_r}^{(t)} \right]. \tag{4.14}$$

- Reward: The goal is to train the agent to maximize the sum-rate, $\sum_{k=1}^{2} \mathcal{R}_k$. Hence, the reward is expressed as

$$r_t = \sum_{k=1}^{2} \mathcal{R}_k^{(t)}. \tag{4.15}$$

4.4.2.2 Deep Q-learning Algorithm

The deep Q-learning (DQL) algorithm is a powerful algorithm that is based on the concept of Q-learning. The aim of the Q-learning agent is to learn the Q-values such that it can always choose the action with the highest value (i.e., highest reward). In particular, the Q-value defines the expected reward of each action at every step as [18]

$$Q(s_t, a_t) = r(s_t, a_t) + \gamma \max_a Q(s_{t+1}, a_t). \tag{4.16}$$

Here, $\gamma \in (0, 1]$ is the discount factor which controls the contribution of future rewards. With experience, the Q-values converge to the optimal policy. In practical situations, it can be updated iteratively as follows

$$Q(s_t, a_t) \quad \leftarrow \quad Q(s_t, a_t) \quad + \quad \mu \left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a_t) - Q(s_t, a_t) \right], \tag{4.17}$$

where μ is the learning rate. To this end, the optimal Q-values allow the agent to choose the best action based on the current state. However, the main problem of Q-learning is that it is only suitable for finite state environments and action spaces. As the size of state and action spaces increases, the time required to explore each state becomes unrealistic (i.e., similar to exhaustive search). Therefore, the DQL algorithm is developed to solve this problem by approximating the Q-value of any state action pair (s, a) using a deep neural network (DNN).

At the initialization stage, the Q-value DNN is generated. An experience replay with capacity D is initialized to minimize the correlation of consecutive training samples by storing/sampling the past experiences. In the beginning of each episode, the entire channel state information is obtained. After that, the agent takes a_t via exploration or exploitation. In particular, in the early stage of training, the agent

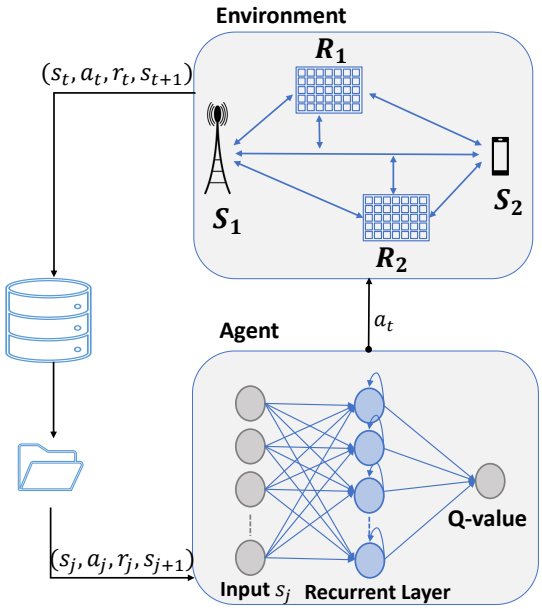


Fig. 4.2: The proposed DQL algorithm structure.

chooses actions randomly more often due to the lack of experience. As the agent becomes more experienced, the exploration rate, ϵ, is decreased to exploit the agent knowledge (i.e., choosing a_t with the maximum Q-value from the DNN). Based on the chosen action, the agent receives a reward, r_t, and passes to the next state, s_{t+1}. The experience (s_t, a_t, r_t, s_{t+1}) is then stored into D, and the agent randomly samples a minibatch transitions, N_B, to calculate the target value y_j, as

$$y_j = r_j + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}; \boldsymbol{\theta}). \tag{4.18}$$

The NN parameters are updated using the gradient descent algorithm. The DQL algorithm trains the agent for K episodes and T steps until achieving the convergence. It is worth noting that γ, μ, ϵ, the exploration threshold, η, and exploration decay, ρ, are all hyperparameters that are tuned for efficient convergence. The proposed DQL algorithm is explained in Algorithm 1 and its structure is shown in Fig. 4.2.

Algorithm 1 Proposed DQL algorithm.

Initialize: $\boldsymbol{\theta}$ with random weights, D, μ, γ, ϵ, η, and ρ;

1: **repeat** for K episodes:
2: Collect the channel state information of the current episode;
3: Initialize $\varphi_{rn} = 0 \, \forall \, n = 1, \cdots, N_r$;
4: Calculate $\mathbf{w}_{\bar{k}}$ using (4.5) or (4.9);
5: **repeat** for T steps:
6: **if** random $\leq \epsilon$ **then**
7: Select a random action $a_t \in \Upsilon$;
8: **else**
9:

$$a_t = \max_a Q(s_t, a; \boldsymbol{\theta}); \tag{4.19}$$

10: **end if**
11: Repeat **Line** #4;
12: Calculate the reward r_t and observe the new state, s_{t+1}, given a_t;
13: Store (s_t, a_t, r_t, s_{t+1}) in D;
14: Sample a minibatch of N_B transitions (s_j, a_j, r_j, s_{j+1}) randomly from D when it is full;
15: **if** $\epsilon > \eta$ **then**
16: $\epsilon \leftarrow \epsilon\rho$;
17: **end if**
18: Compute the target value from (4.18);
19: Perform a gradient descent algorithm step on $(y_j - Q(s_j, a_j; \boldsymbol{\theta}))^2$;

Output: Optimized $\bar{\boldsymbol{\Theta}}^*$.

4.4.3 Proposed DNN Structure and Complexity Analysis

The proposed NN model contains three layers: the input layer, a hidden layer and the output layer. The input layer has the size of s_t, and contains Γ_i neurons. The hidden layer is modeled as gated recurrent unit (GRU). It contains Γ_h neurons and it uses the *ReLU* activation function. The internal recurrent squashing function is the *Sigmoid* function. The GRU uses two gates to control the flow of information: the reset and update gates. These gates decide how much of the past information should be passed to the output. In particular, the reset gate is responsible for short-term memory, whereas the update gate is responsible for long-term memory. The output layer, which contains Γ_0 neurons, gives the approximated Q-value. The structure of

the NN model is shown in Fig. 4.2.

The complexity of the proposed NN model is analyzed based on the number of real additions, C_A, and real multiplications, C_M. Therefore, the complexity of the proposed NN design is given as

$$C_A = 3\left(\Gamma_i + \Gamma_h + \frac{5}{3}\right)\Gamma_h + \Gamma_h\Gamma_o, \tag{4.20}$$

and

$$C_M = 3\left(\Gamma_i + \Gamma_h + 1\right)\Gamma_h + \Gamma_h\Gamma_o. \tag{4.21}$$

4.5 Simulation Results

Figure 4.3 shows the simulation setup and the distances between the R_r-S_i links. We assume that $d_{v1} = d_{v2} = 2\,\text{m}$, $d_1 = 50\,\text{m}$. Generally, the RIS is deployed near the BS to best benefit from the RIS reflection links, ensuring a strong LoS link between the RIS and BS. Therefore, d_{01} is considered as 1 m, and d_{02} as 49 m. The path loss is given as $\text{PL} = PL_r - 10\zeta\log_{10}\left(\frac{d_{ir}}{D_r}\right)$ [10], where d_{ir} is the distance between the R_r-S_i links, PL_r is the path loss at a reference distance D_r, and ζ is the path loss exponent. We set $PL_r = -35.6\,\text{dB}$ and $D_r = 1\,\text{m}$. The path loss exponents of the S_1-S_2, S_1-R_r, and S_2-R_r links are $\zeta_{\text{BU}} = 4$, $\zeta_{\text{BR}} = 2.1$, and $\zeta_{\text{UR}} = 2.2$, respectively, while the path loss of the SI channels is $-95\,\text{dB}$. The total maximum transmit power is $P_{\max} = 15\,\text{dBm}$ and the noise power is $\sigma^2 = -80\,\text{dBm}$ [12, 7]. The channels are modeled as Rician [12], and are expressed as

$$\mathbf{h} = \sqrt{PL}\left(\frac{K_1}{K_1+1}\bar{\mathbf{h}} + \frac{1}{K_1+1}\tilde{\mathbf{h}}\right), \tag{4.22}$$

where $\mathbf{h}$ represents the channel and K_1 is the Rician factor which is set to 10. $\tilde{\mathbf{h}}$ is

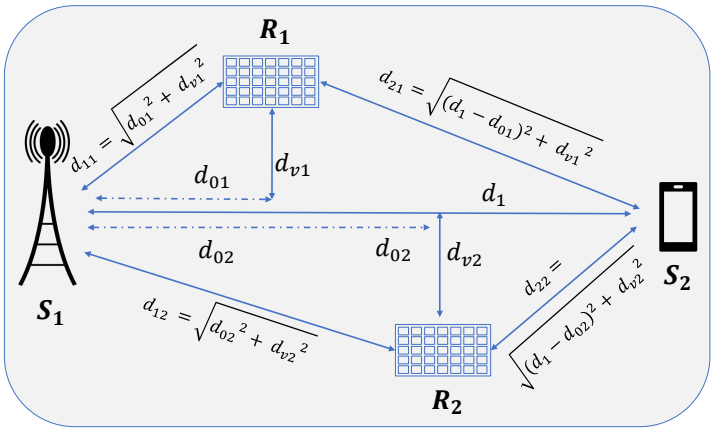


Fig. 4.3: Simulation setup.

the random component that contains independent and identical distributed $\mathcal{CN}(0,1)$ elements, while $\bar{\mathbf{h}}$ is the deterministic component.

The parameters of the DQL algorithm are set as follows: $T = 1000$, $K = 500$, $N_B = 32$, $\mu = 0.0001$, decaying rate of 0.0001, $\gamma = 0.95$, $\epsilon = 1$, $\eta = 0.01$, $\rho = 0.995$, and $D = 10000$. The Adam optimizer is used to update the NN parameters. Finally, the DNN model structure is defined as $\Gamma_i = N + 1$, $\Gamma_h = 300$, and $\Gamma_o = 1$.

To assess the performance of the discrete phase shift optimization, it is compared with the optimization of the continuous phase shift model presented in [12], referred to as *cont.* in the simulation results. Two beamforming mathematical derivations are used to optimize the transmit beamformers, as discussed in Sec. 4.4.1, where the approximate and closed-form solutions are referred to as *appr.* and *exact*, respectively. The performance of the discrete phase shift optimization is also compared with the non-optimized case, referred to as *random*. Furthermore, the results are shown for two scenarios: No LoS ($\xi = 0$) and LoS ($\xi = 1$).

Figure 4.4 shows the average system reward versus training episodes when $N = 40$. It can be seen that the average reward increases over time, indicating that the agent is learning. In the early stages of the learning process, more fluctuations can be seen

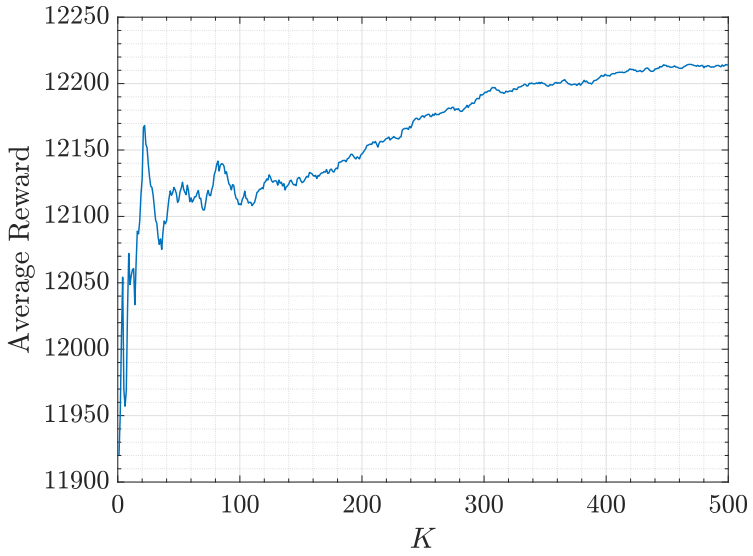


Fig. 4.4: Average reward performance versus number of episodes for $N = 40$ and $b = 6$.

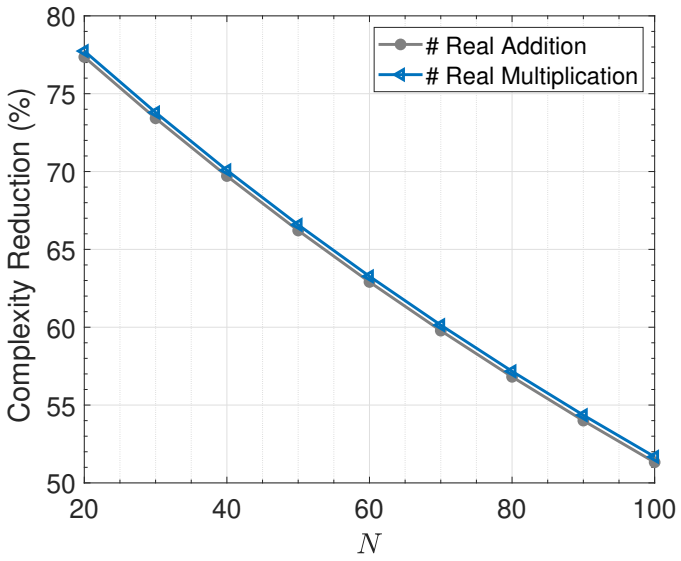


Fig. 4.5: Complexity reduction percentage versus N.

due to the higher probability of randomness in choosing actions. However, as the number of episodes increases, the learning of the agent is more stable, and it chooses actions based on exploiting its own knowledge (i.e., choosing the action with the highest Q-value). It can be observed that the average reward reaches convergence

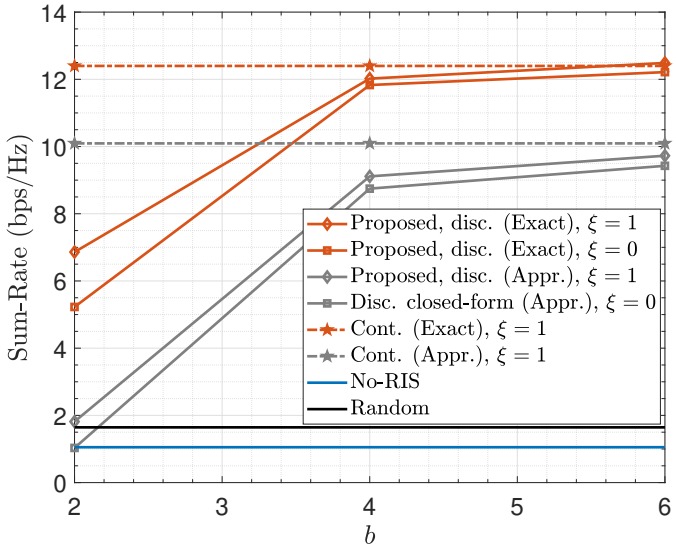


Fig. 4.6: The effect of the number of bits on the system performance for $N = 40$.

when $K > 400$.

Figure 4.5 illustrates the complexity reduction percentage of using the proposed NN over the NN presented in [12], where its computational complexity is presented in (3.19- 3.20), which is given as

$$\text{Reduction} = 1 - \frac{\{C_\chi\}_{\text{Proposed}}}{\{C_\chi^{\text{A}} + C_\chi^{\text{C}}\}_{[12]}}, \ \chi \in \{A, M\}. \tag{4.23}$$

where A and C denote the actor and critic networks used in [12]. It can be seen that the proposed NN significantly reduces the computational complexity in the range of 78% to 52% for $N = 20$ to $N = 100$.

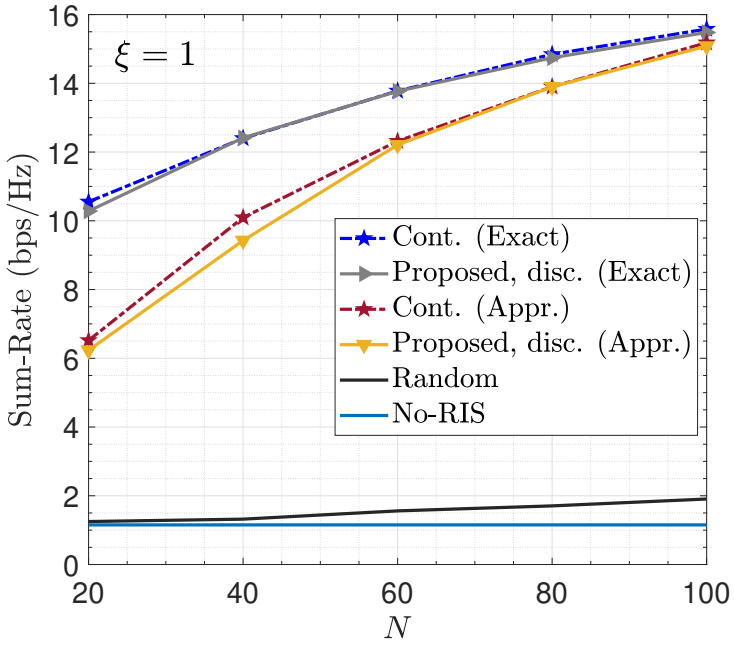
Figure 4.6 illustrates the sum-rate performance of the proposed algorithm versus the number of quantization bits, b. The results are compared with the optimization of the continuous phase shift model [12] (using the two beamforming mathematical solutions) and the scenario where no RIS is deployed, referred to as no-RIS. The results shows that as b increases, the proposed DRL algorithm almost achieves the

upper bound (i.e., the continuous phase shift model), which testifies its practicality. This is applicable for both beamforming solutions, i.e., approximate and closed-form derivations. Besides, the simulation results show that there is a small gap in the performance when $\xi = 0$ and $\xi = 1$, which emphasizes that the sum-rate improvement is due to the RIS deployment. It is worth noting that when $b = 2$, only four phase shift values are available for each RIS element, $\left(\text{i.e.,}\{-\pi, -\frac{\pi}{2}, 0, \frac{\pi}{2}\}\right)$. In this case, the sum-rate performance would be far from the optimized solution as shown in the figure. It can also be observed that the near-optimal solutions are obtained with only 6 bits.

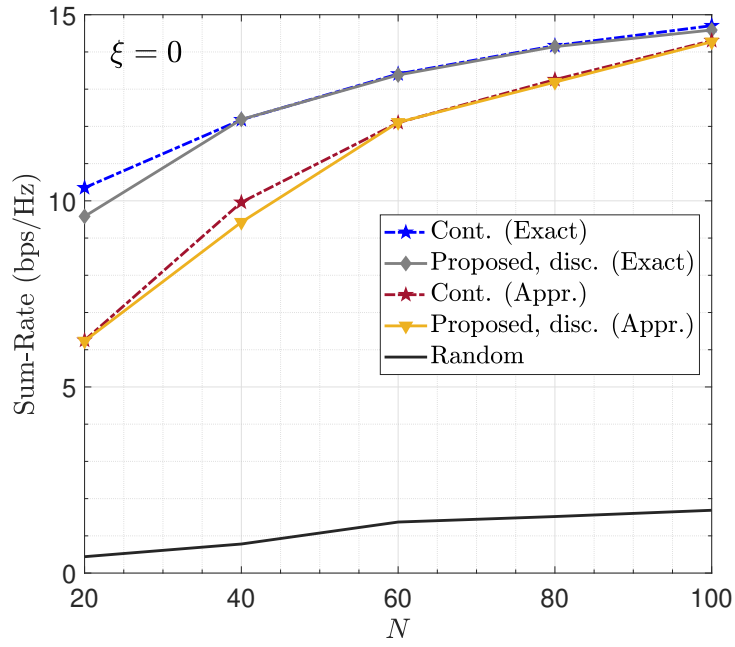
Figure 4.7 illustrates the effect of varying N on the system performance when $\xi = 1$ and $\xi = 0$ for $b = 6$. It can be noted that for all algorithms, the sum-rate increases as N increases. The results depict that the proposed algorithm achieves a near-optimal performance even when the number of reflecting elements is large. In particular, the gap in the sum-rate performance between the continuous and discrete phase shift models decreases as N increases. It can also be observed that the proposed DQL algorithm provides a remarkable sum-rate improvement compared to the random RIS phase shifts scenario. Lastly, it can be seen that when $\xi = 0$, the sum-rate performance is lower than that of $\xi = 1$. However, the results are almost similar due to the fact that the improvement is mainly coming from the RIS deployment.

4.6 Conclusion

This chapter maximized the sum-rate of a distributed RIS-assisted FD MISO system, considering a practical phase shift model (i.e., discrete values). The optimization problem was addressed using a two-step solution: mathematical derivations to optimize the transmit beamformers and a DRL approach to optimize the discrete phase shifts. Two scenarios, namely the presence and the absence of the LoS link, were investigated



(a) $\xi = 1$.



(b) $\xi = 0$.

Fig. 4.7: The effect of increasing N on the system performance for $b = 6$.

to validate the performance of the proposed algorithm. The proposed DRL algorithm achieved a near-optimal performance compared to the ideal phase shift model considered in the literature (i.e., continuous values) in the two scenarios. It was further shown that the algorithm provides a remarkable improvement over the non-optimized cases while reducing the complexity in the range of 78% to 52% for $N = 20$ to $N = 100$.

References

[1] R. Alghamdi, R. Alhadrami, D. Alhothali, H. Almorad, A. Faisal, S. Helal, R. Shalabi, R. Asfour, N. Hammad, A. Shams, N. Saeed, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Intelligent surfaces for 6G wireless networks: A survey of optimization and performance analysis techniques," *IEEE Access*, vol. 8, pp. 202795–202818, Oct. 2020.

[2] I. Al-Nahhal, O. A. Dobre, E. Basar, T. M. N. Ngatched, and S. Ikki, "Reconfigurable intelligent surface optimization for uplink sparse code multiple access," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 133–137, Jan. 2022.

[3] K. M. Faisal and W. Choi, "Machine learning approaches for reconfigurable intelligent surfaces: A survey," *IEEE Access*, vol. 10, pp. 27 343–27 367, Mar. 2022.

[4] I. Al-Nahhal, O. A. Dobre, and E. Basar, "Reconfigurable intelligent surface-assisted uplink sparse code multiple access," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 2058–2062, Feb. 2021.

[5] Y. Cai, M.-M. Zhao, K. Xu, and R. Zhang, "Intelligent reflecting surface aided full-duplex communication: Passive beamforming and deployment design," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 1, pp. 383–397, Jan. 2022.

[6] C. Pradhan, A. Li, L. Song, J. Li, B. Vucetic, and Y. Li, "Reconfigurable intelligent surface (RIS)-enhanced two-way OFDM communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16 270–16 275, Dec. 2020.

[7] M. A. Saeidi, M. J. Emadi, H. Masoumi, M. R. Mili, D. W. K. Ng, and I. Krikidis, "Weighted sum-rate maximization for multi-IRS-assisted full-duplex systems with hardware impairments," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 466–481, Jun. 2021.

[8] Z. Peng, Z. Zhang, C. Pan, L. Li, and A. L. Swindlehurst, "Multiuser full-duplex two-way communications via intelligent reflecting surface," *IEEE Trans. Signal Process.*, vol. 69, pp. 837–851, Jan. 2021.

[9] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 1, pp. 375–388, Jan. 2021.

[10] K. Feng, Q. Wang, X. Li, and C.-K. Wen, "Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 745–749, May 2020.

[11] A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Deep reinforcement learning for optimizing RIS-assisted HD-FD wireless systems," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3893–3897, Dec. 2021.

[12] ——, "Deep reinforcement learning for RIS-assisted FD systems: Single or distributed RIS?" *IEEE Commun. Lett., Early Access*, Apr. 2022.

[13] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Dec. 2020.

[14] Y. Zhang, J. Zhang, M. D. Renzo, H. Xiao, and B. Ai, "Performance analysis of RIS-aided systems with practical phase shift and amplitude response," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4501–4511, May 2021.

[15] Q. Wu and R. Zhang, "Beamforming optimization for intelligent reflecting surface with discrete phase shifts," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 7830–7833.

[16] B. Zheng, Q. Wu, and R. Zhang, "Intelligent reflecting surface-assisted multiple access with user pairing: NOMA or OMA?" *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 753–757, Apr. 2020.

[17] H. Shen, T. Ding, W. Xu, and C. Zhao, "Beamforming design with fast convergence for IRS-aided full-duplex communication," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2849–2853, Dec. 2020.

[18] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

Chapter 5

Conclusions and Future Work

This chapter concludes the thesis by summarizing the main contributions. It further discusses some of the promising upcoming research challenges that should be considered in the future work.

5.1 Conclusions

RISs have recently emerged as a key enabler for beyond 5G communications to fulfill the rapid increasing demands for wireless network capabilities. RISs are not only considered to enhance the communication quality, but also to reduce the power consumption as compared to the conventional wireless networks. This work considered incorporating RIS into FD communication systems to offer new degrees of freedom, facilitating ultra spectrum-efficient systems. In particular, FD communications allow transmitting and receiving simultaneously in the same frequency band, which offer the potential to double the spectral efficiency and increase the sum-rate of wireless communication systems significantly. To fully exploit the RIS capabilities in FD communication systems, the phase shifts should be efficiently optimized. Therefore, this work considers very powerful DRL approaches, to solve the challenging optimization problems.

To this end, Chapter 1 introduced the motivation, contributions and the outline of this thesis. In Chapter 2, a low-complexity DRL is proposed to optimize the RIS phase shifts in HD-FD RIS-assisted communication systems. It is worth noting that DRL is exploited for FD communication systems for the first time in the literature. The proposed algorithm was shown to significantly improve the rate compared to the non-optimized RIS phase shifts in both operating modes (i.e., HD and FD). It further significantly reduces the computational complexity as compared to the conventional DRL algorithms of the HD mode. In Chapter 3, the single and distributed RIS deployment schemes are investigated to answer the question of which deployment scheme is preferred. Three practical scenarios based on the links' quality are considered to study the sum-rate performance of deploying a single or distributed RIS in an FD-MISO system. A two-step solution was proposed, where a closed-form solution is derived to optimize the transmit beamformers and an efficient DRL algorithm is proposed to optimize the RIS phase shifts for both deployment schemes. The proposed solution provides a remarkable improvement in the sum-rate compared to the considered benchmark. Finally, the discrete phase shift optimization problem of distributed RIS-assisted FD system is considered in Chapter 4. In this work, two mathematical solutions (approximate and closed-form) for optimizing the transmit beamformers are considered. The proposed algorithm is shown to achieve promising results compared to the continuous-baseline (the DDPG algorithm). The proposed algorithms can fit a wide range of practical applications as DRL approaches are seen as a key enabler for future wireless application. The mathematical formulation, complexity analysis for all proposed algorithms, and simulation results were provided to support these findings.

5.2 Future Research Directions

The work in this thesis identifies numerous exciting open challenges for FD RIS-aided wireless networks, which include:

- Studying the sum-rate maximization problem for a multi-user FD RIS-assisted communication system using DRL. Since DRL allows the user to learn from the interactions with the environment, it allows extending the problem to cover large scale practical systems efficiently compared with the traditional optimization techniques.

- Developing a DRL approach for FD RIS-assisted cellular communications using energy harvesting is a promising research direction. It allows the BS to dynamically adapt to wireless environment by deciding the RIS phase shift configuration using a neural network. In particular, besides optimizing the RIS phase shifts, the RIS elements can be turned on or off, depending on the network performance to further enhance the overall energy efficiency of the FD communication system.

- Adopting DRL for FD RIS-assisted system, where the integration of sensing and communications is considered to offer efficient spectrum utilization.

- Investigating the sum-rate performance of different DRL approaches for a FD RIS-assisted system, while assuming imperfect transceivers to study the robustness and reliability of DRL in this context.

- Investigating the DRL approaches in the physical layer security and data transmission for the underlay device-to-device networks, while considering the RIS deployment and a FD jamming receiver for the robustness and security enhancements of the system.

In conclusion, investigating the performance of DRL approaches of various FD RIS-assisted applications remains an open problem that needs to be investigated. DRL can significantly improve the system throughput and enable the full exploitation of the capabilities of the RIS technology in different application scenarios.

References

Chapter 1

[1] N. Rajatheva, I. Atzeni, E. Bjornson, A. Bourdoux, S. Buzzi, J.-B. Dore, S. Erkucuk, M. Fuentes, K. Guan, Y. Hu, X. Huang, J. Hulkkonen, J. M. Jornet, M. Katz, R. Nilsson, E. Panayirci, K. Rabie, N. Rajapaksha, M. J. Salehi, H. Sarieddeen, S. Shahabuddin, T. Svensson, O. Tervo, A. Tolli, Q. Wu, and W. Xu, "White paper on broadband connectivity in 6g," *6G Research Visions*, vol. 10. [Online]. Available: https://par.nsf.gov/biblio/10223732

[2] R. Alghamdi, R. Alhadrami, D. Alhothali, H. Almorad, A. Faisal, S. Helal, R. Shalabi, R. Asfour, N. Hammad, A. Shams, N. Saeed, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Intelligent surfaces for 6G wireless networks: A survey of optimization and performance analysis techniques," *IEEE Access*, vol. 8, pp. 202795-202818, Oct. 2020.

[3] M. A. ElMossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, "Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 3, pp. 990-1002, Sep. 2020.

[4] B. C. Nguyen, T. M. Hoang, L. T. Dung, and T. Kim, "On performance of two-way full-duplex communication system with reconfigurable intelligent surface," *IEEE*

Access, vol. 9, pp. 81 274-81 285, Jun. 2021.

[5] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," ser. SIGCOMM '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 375-386.

[6] A. H. Gazestani, S. A. Ghorashi, B. Mousavinasab, and M. Shikh-Bahaei, "A survey on implementation and applications of full duplex wireless communications," *Physical Communication*, vol. 34, pp. 121-134, 2019.

[7] R. Askar, J. Chung, Z. Guo, H. Ko, W. Keusgen, and T. Haustein, "Interference handling challenges toward full duplex evolution in 5G and beyond cellular networks," *IEEE Wirel. Commun.*, vol. 28, no. 1, pp. 51-59, Feb. 2021.

[8] M. A. Saeidi, M. J. Emadi, H. Masoumi, M. R. Mili, D. W. K. Ng, and I. Krikidis, "Weighted sum-rate maximization for multi-IRS-assisted full-duplex systems with hardware impairments," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 466-481, Jun. 2021.

[9] G. Zhou, C. Pan, H. Ren, K. Wang, and Z. Peng, "Secure wireless communication in RIS-aided miso system with hardware impairments," *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1309-1313, Jun. 2021.

[10] Z. Yang and Y. Zhang, "Beamforming optimization for RIS-aided SWIPT in cell-free mimo networks," *China Communications*, vol. 18, no. 9, pp. 175-191, Sep. 2021.

[11] D. Xu, X. Yu, Y. Sun, D. W. K. Ng, and R. Schober, "Resource allocation for IRS-assisted full-duplex cognitive radio systems," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7376-7394, Sep. 2020.

[12] I. Al-Nahhal, O. A. Dobre, and E. Basar, "Reconfigurable intelligent surface-assisted uplink sparse code multiple access," *IEEE Commun. Lett.*, vol. 25, no. 6, pp.

2058-2062, Feb. 2021.

[13] K. O. Odeyemi, P. A. Owolawi, and O. O. Olakanmi, "Reconfigurable intelligent surface-assisted haps relaying communication networks for multiusers under AF protocol: A performance analysis," *IEEE Access*, vol. 10, pp. 14 857-14 869, Jan. 2022.

[14] I. Al-Nahhal, O. A. Dobre, E. Basar, T. M. N. Ngatched, and S. Ikki, "Reconfigurable intelligent surface optimization for uplink sparse code multiple access," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 133-137, Jan. 2022.

[15] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838-1851, Dec. 2020.

[16] S. Zhou, W. Xu, K. Wang, M. Di Renzo, and M.-S. Alouini, "Spectral and energy efficiency of IRS-assisted miso communication with hardware impairments," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1366-1369, Sep. 2020.

[17] M. Jung, W. Saad, M. Debbah, and C. S. Hong, "On the optimality of reconfigurable intelligent surfaces (RISs): Passive beamforming, modulation, and resource allocation," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4347-4363, Jul. 2021.

[18] Z. Abdullah, G. Chen, S. Lambotharan, and J. A. Chambers, "Optimization of intelligent reflecting surface assisted full-duplex relay networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 363-367, Feb. 2021.

[19] H. Shen, T. Ding, W. Xu, and C. Zhao, "Beamformig design with fast convergence for IRS-aided full-duplex communication," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2849-2853, Aug. 2020.

[20] Y. Cai, M.-M. Zhao, K. Xu, and R. Zhang, "Intelligent reflecting surface aided

full-duplex communication: Passive beamforming and deployment design," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 1, pp. 383-397, Jan. 2022.

[21] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, May 2016, pp. 1-14.

[22] A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Deep reinforcement learning for optimizing RIS-assisted HD-FD wireless systems," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3893-3897, Dec. 2021.

[23] ——, "Deep reinforcement learning for RIS-assisted FD systems: Single or distributed RIS?" *IEEE Commun. Lett.*, Early Access, Apr. 2022.

[24] ——, "Distributed RIS-Assisted FD Systems with Discrete Phase Shifts: A Reinforcement Learning Approach," Accepted for presentation at *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, May 2022.

Chapter 2

[1] I. Al-Nahhal, O. A. Dobre, and E. Basar, "Reconfigurable intelligent surface-assisted uplink sparse code multiple access," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 2058-2062, Feb. 2021.

[2] L. Bariah, L. Mohjazi, S. Muhaidat, P. C. Sofotasios, G. K. Kurt, H. Yanikomeroglu, and O. A. Dobre, "A prospective look: Key enabling technologies, applications and open research topics in 6G networks," *IEEE Access*, vol. 8, pp. 174792-174820, Aug. 2020.

[3] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*,

vol. 7, pp. 116753-116773, Aug. 2019.

[4] R. Alghamdi, R. Alhadrami, D. Alhothali, H. Almorad, A. Faisal, S. Helal, R. Shalabi, R. Asfour, N. Hammad, A. Shams, N. Saeed, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Intelligent surfaces for 6G wireless networks: A survey of optimization and performance analysis techniques," *IEEE Access*, vol. 8, pp. 202795-202818, Oct. 2020.

[5] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838-1851, Dec. 2020.

[6] G. Zhou, C. Pan, H. Ren, K. Wang, M. D. Renzo, and A. Nallanathan, "Robust beamforming design for intelligent reflecting surface aided MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1658-1662, Jun. 2020.

[7] N. S. Perovic, L.-N. Tran, M. Di Renzo, and M. F. Flanagan, "Achievable rate optimization for MIMO systems with reconfigurable intelligent surfaces," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3865-3882, Feb. 2021.

[8] H. Shen, T. Ding, W. Xu, and C. Zhao, "Beamformig design with fast convergence for IRS-aided full-duplex communication," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2849-2853, Aug. 2020.

[9] J. Zhao, M. Chen, M. Chen, Z. Yang, Y. Wang, B. Cao, and M. Shikh-Bahaei, "Energy efficient full-duplex communication systems with reconfigurable intelligent surface," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, Feb. 2020, pp. 1-5.

[10] Z. Peng, Z. Zhang, C. Pan, L. Li, and A. L. Swindlehurst, "Multiuser full-duplex two-way communications via intelligent reflecting surface," *IEEE Trans. Signal Process.*, vol. 69, pp. 837-851, Jan. 2021.

[11] M. Elhattab, M. A. Arfaoui, C. Assi, and A. Ghrayeb, "Reconfigurable intelligent surface enabled full-duplex/half-duplex cooperative non-orthogonal multiple access," Jan. 2021. [Online]. Available: https://arxiv.org/abs/2101.01307 20

[12] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?" *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 7331-7376, Jun. 2019.

[13] Y. Chen, Y. Liu, M. Zeng, U. Saleem, Z. Lu, X. Wen, D. Jin, Z. Han, T. Jiang, and Y. Li, "Reinforcement learning meets wireless networks: A layering perspective," *IEEE Internet Things J.*, vol. 8, no. 1, pp. 85-111, Jan. 2021.

[14] J. Lin, Y. Zout, X. Dong, S. Gong, D. T. Hoang, and D. Niyato, "Deep reinforcement learning for robust beamforming in IRS-assisted wireless communications," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Jan. 2020, pp. 1-6.

[15] C. Huang, R. Mo, and C. Yuen, "Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1839-1850, Jun. 2020.

[16] K. Feng, Q. Wang, X. Li, and C.-K. Wen, "Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 745-749, May. 2020.

[17] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, May 2016, pp. 1-14.

Chapter 3

[1] I. Al-Nahhal, O. A. Dobre, and E. Basar, "Reconfigurable intelligent surface-assisted uplink sparse code multiple access," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 2058-2062, Feb. 2021.

[2] I. Al-Nahhal, O. A. Dobre, E. Basar, T. M. N. Ngatched, and S. Ikki, "Reconfigurable intelligent surface optimization for uplink sparse code multiple access," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 133-137, Jan. 2022.

[3] M. A. ElMossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, "Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 3, pp. 990-1002, Sep. 2020.

[4] R. Askar, J. Chung, Z. Guo, H. Ko, W. Keusgen, and T. Haustein, "Interference handling challenges toward full duplex evolution in 5G and beyond cellular networks," *IEEE Wirel. Commun.*, vol. 28, no. 1, pp. 51-59, Feb. 2021.

[5] A. Yadav and O. A. Dobre, "All technologies work together for good: A glance at future mobile networks," *IEEE Wirel. Commun.*, vol. 25, no. 4, pp. 10-16, Aug. 2018.

[6] R. Alghamdi, R. Alhadrami, D. Alhothali, H. Almorad, A. Faisal, S. Helal, R. Shalabi, R. Asfour, N. Hammad, A. Shams, N. Saeed, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Intelligent surfaces for 6G wireless networks: A survey of optimization and performance analysis techniques," *IEEE Access*, vol. 8, pp. 202795-202818, Oct. 2020.

[7] H. Shen, T. Ding, W. Xu, and C. Zhao, "Beamforming design with fast convergence for IRS-aided full-duplex communication," *IEEE Commun. Lett.*, vol. 24, no. 12, pp.

2849-2853, Dec. 2020.

[8] Y. Zhang, C. Zhong, Z. Zhang, and W. Lu, "Sum rate optimization for two way communications with intelligent reflecting surface," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1090-1094, May 2020.

[9] M. A. Saeidi, M. J. Emadi, H. Masoumi, M. R. Mili, D. W. K. Ng, and I. Krikidis, "Weighted sum-rate maximization for multi-IRS-assisted full-duplex systems with hardware impairments," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 466-481, Jun. 2021.

[10] A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Deep reinforcement learning for optimizing RIS-assisted HD-FD wireless systems," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3893-3897, Dec. 2021.

[11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1997.

[12] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[13] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, May 2016, pp. 1-14.

[14] K. Feng, Q. Wang, X. Li, and C.-K. Wen, "Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 745-749, May 2020.

Chapter 4

[1] R. Alghamdi, R. Alhadrami, D. Alhothali, H. Almorad, A. Faisal, S. Helal, R. Shalabi, R. Asfour, N. Hammad, A. Shams, N. Saeed, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Intelligent surfaces for 6G wireless networks: A survey of optimization and performance analysis techniques," *IEEE Access*, vol. 8, pp. 202795-202818, Oct. 2020.

[2] I. Al-Nahhal, O. A. Dobre, E. Basar, T. M. N. Ngatched, and S. Ikki, "Reconfigurable intelligent surface optimization for uplink sparse code multiple access," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 133-137, Jan. 2022.

[3] K. M. Faisal and W. Choi, "Machine learning approaches for reconfigurable intelligent surfaces: A survey," *IEEE Access*, vol. 10, pp. 27343-27367, Mar. 2022.

[4] I. Al-Nahhal, O. A. Dobre, and E. Basar, "Reconfigurable intelligent surface-assisted uplink sparse code multiple access," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 2058-2062, Feb. 2021.

[5] Y. Cai, M.-M. Zhao, K. Xu, and R. Zhang, "Intelligent reflecting surface aided full-duplex communication: Passive beamforming and deployment design," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 1, pp. 383-397, Jan. 2022.

[6] C. Pradhan, A. Li, L. Song, J. Li, B. Vucetic, and Y. Li, "Reconfigurable intelligent surface (RIS)-enhanced two-way OFDM communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16 270-16 275, Dec. 2020.

[7] M. A. Saeidi, M. J. Emadi, H. Masoumi, M. R. Mili, D. W. K. Ng, and I. Krikidis, "Weighted sum-rate maximization for multi-IRS-assisted full-duplex systems with hardware impairments," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 466-481, Jun. 2021.

[8] Z. Peng, Z. Zhang, C. Pan, L. Li, and A. L. Swindlehurst, "Multiuser full-duplex two-way communications via intelligent reflecting surface," *IEEE Trans. Signal Process.*, vol. 69, pp. 837-851, Jan. 2021.

[9] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 1, pp. 375-388, Jan. 2021.

[10] K. Feng, Q. Wang, X. Li, and C.-K. Wen, "Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 745-749, May. 2020.

[11] A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Deep reinforcement learning for optimizing RIS-assisted HD-FD wireless systems," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3893-3897, Dec. 2021.

[12] ——, "Deep reinforcement learning for RIS-assisted FD systems: Single or distributed RIS?" *IEEE Commun. Lett.*, Early Access, Apr. 2022.

[13] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838-1851, Dec. 2020.

[14] Y. Zhang, J. Zhang, M. D. Renzo, H. Xiao, and B. Ai, "Performance analysis of RIS-aided systems with practical phase shift and amplitude response," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4501-4511, May 2021.

[15] Q. Wu and R. Zhang, "Beamforming optimization for intelligent reflecting surface with discrete phase shifts," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 7830-7833.

[16] B. Zheng, Q. Wu, and R. Zhang, "Intelligent reflecting surface-assisted multiple

access with user pairing: NOMA or OMA?" *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 753-757, Apr. 2020.

[17] H. Shen, T. Ding, W. Xu, and C. Zhao, "Beamformig design with fast convergence for IRS-aided full-duplex communication," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2849-2853, Dec. 2020.

[18] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529-533, Feb. 2015.