

Terahertz Multiple Access: A Deep Reinforcement Learning Controlled Multihop IRS Topology

*Original*

Terahertz Multiple Access: A Deep Reinforcement Learning Controlled Multihop IRS Topology / Shehab, Muhammad; Elsayed, Mohamed; Almohamad, Abdullateef; Badawy, Ahmed; Khattab, Tamer; Zorba, Nizar; Hasna, Mazen; Trincherio, Daniele. - In: IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY. - ISSN 2644-125X. - ELETTRONICO. - 5:(2024), pp. 1072-1087. [10.1109/OJCOMS.2024.3357701]

*Availability:*

This version is available at: 11583/2985361 since: 2024-01-25T04:52:45Z

*Publisher:*

IEEE

*Published*

DOI:10.1109/OJCOMS.2024.3357701

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

IEEE postprint/Author's Accepted Manuscript

©2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

# Terahertz Multiple Access: A Deep Reinforcement Learning Controlled Multihop IRS Topology

Muhammad Shehab, *Member, IEEE*, Mohamed Elsayed, *Member, IEEE*, Abdullateef Almohamad, *Member, IEEE*, Ahmed Badawy, *Member, IEEE*, Tamer Khattab, *Senior Member, IEEE*, Nizar Zorba, *Senior Member, IEEE*, Mazen Hasna, *Senior Member, IEEE*, Daniele Trinchero.

We explore THz communication uplink multi-access with multi-hop Intelligent reflecting surfaces (IRSs) under correlated channels. Our aims are twofold: 1) enhancing the data rate of a desired user while dealing with interference from another user and 2) maximizing the combined data rate. Both tasks involve non-convex optimization challenges. For the first aim, we devise a sub-optimal analytical approach that focuses on maximizing the desired user's received power, leading to an over-determined system. We also attempt to use approximate solutions utilizing pseudo-inverse ( $P_{inv}$ ) and block solution (BLS) based methods. For the second aim, we establish a loose upper bound and employ an exhaustive search (ES). We employ deep reinforcement learning (DRL) to address both aims, demonstrating its effectiveness in complex scenarios. DRL outperforms mathematical approaches for the first aim, with the performance improvement of DDPG over the block solution ranging from 8% to 57.12%, and over the pseudo-inverse ranging from 41% to 190% for a correlation-factor equal to 1. Moreover, DRL closely approximates the ES for the second aim. Furthermore, our findings show that as channel correlation increases, DRL's performance improves, capitalizing on the correlation for enhanced statistical learning.

**Index Terms**—Artificial intelligence, multi-access communication, sub-millimeter wave communication, communication system performance.

## I. INTRODUCTION

IN the field of wireless communication, 6G is expected to cater to significantly advanced services and data-intensive applications compared to its predecessor, 5G. These applications, such as immersive remote presence, connected robotics (CRAS), digital twins, and immersive extended reality (XR), necessitate a colossal 1000-fold increase in capacity compared to 5G mobile systems [1]. To address these demands and reconcile the tension between service needs and limited spectrum resources [2], there is a call to extend existing wireless spectrum bands and venture into the higher terahertz (THz) frequency range, spanning from 0.1 THz to 10 THz. These frequencies are poised to play a pivotal role

in 6G wireless communications due to their potential for substantially enhanced capacities and data rates. However, these higher Radio Frequencies (RF) bring challenges such as substantial path attenuation, elevated propagation losses, and intermittent wireless connections. Moreover, RF waves with extremely short wavelengths result in limited communication distances and increased vulnerability to molecular absorption and obstruction [1]. To enhance received signal power and achievable data rates, this paper explores the Intelligent reflecting surfaces (IRS) as a burgeoning and highly promising technological advancement. [3]. IRS operates by manipulating incident electromagnetic waves, programmatically adjusting phase shifts of semi-passive reflecting elements to enable a smart radio environment, enhancing data rates in a cost-effective and energy-efficient manner [4].

## II. LITERATURE REVIEW

Numerous recent studies have scrutinized the deployment of IRS in THz communications to evaluate its potential in improving coverage and achievable data rates [5] - [11]. For example, one study [5] focused on an IRS-assisted multi-hop multi-pair unicast network, where multiple sources communicate with multiple destinations, proposing distributed control of multiple IRSs to optimize achievable rates. Another study [6] explored a multi-IRS-assisted massive multiple-input multiple-output system, aiming to boost minimum received signal power. It involved a base station (BS) equipped with multi-antennas transmitting independent signals to remote users with single antennas, enabling cascaded line-of-sight (LOS) communication links through cooperative signal reflections from various IRS groups. In [7], the authors assessed the effectiveness of a decode-and-forward (DF) relaying assisted multi-IRS system in a scenario where a single source communicates with a single destination, seeking to determine the optimal IRS configuration, the number of IRSs, and the number of IRS reflecting elements that maximize the ergodic rate. Additionally, in [8], the authors considered a multi-hop IRS-assisted multi-user downlink communication scenario where the BS communicates with  $K$  users, optimizing the beamforming at the BS and multiple IRS phase shift reflection matrices to maximize the sum rate. Furthermore, in [9], the authors explored an uplink multi-hop IRS communication system where multiple users communicate with a single destination, to extend the link range in THz communications and maximize

M. Shehab, and D. Trinchero are with Dipartimento di Elettronica, Politecnico di Torino, Torino, Italy. M. Elsayed, A. Almohamad, T. Khattab, N. Zorba, and M. Hasna are with Electrical Engineering, Qatar University, Doha, Qatar. A. Badawy is with Computer Science and Engineering, Qatar University, Doha, Qatar. E-mail: MuhammadShehab@ieee.org, hamid@qu.edu.qa, abdullateef@ieee.org, badawy@qu.edu.qa, tkhattab@ieee.org, nizarz@qu.edu.qa, hasna@qu.edu.qa, and daniele.trinchero@polito.it. This research work was made possible by grant number AICCO3-0530-200033 from the Qatar National Research Fund (QNRF). Statements made herein are the sole responsibility of authors.

power at the receiver ( $R_x$ ). They introduced a cascaded passive IRS THz system to overcome the substantial propagation losses due to air molecule absorption.

Notably, the studies from [5] to [9] employed mathematical techniques to address their optimization problems. In contrast, research papers [10] and [11] utilized the DRL algorithm to tackle non-convex optimization challenges. The authors suggested a hybrid beamforming scheme for multi-hop IRS-aided networks to enhance the coverage of THz communication links. They investigated the joint design of analog beamforming at the IRSs and digital beamforming at the BS to mitigate propagation losses in THz downlink broadcast systems, which involve a single source and multiple destinations.

### A. Contributions

To enhance the range of THz links and compensate for losses at such high frequencies, we employ multi-hop IRS, often referred to as cascaded IRSs, as a fundamental element in our system model. Given the limited coverage typically associated with THz links, we focus on small areas where multiple users are not expected. Our optimization problem centers around a two-user system, to find the optimal phase shifts for the multi-hop IRS elements. This optimization aims to maximize the received rate for a specific user and the combined rate for both users. The primary challenge in our approach arises from the non-convex nature of the objective function due to constraints related to the constant amplitude of the IRS reflecting elements, as well as non-linear constraints and complex multi-hop links. Solving this NP-hard problem optimally remains elusive, and conventional mathematical techniques struggle to provide an analytical solution. Moreover, an exhaustive search ( $ES$ ) approach is impractical for large-scale communication systems [10], [11]. Additionally, optimizing  $R_x$  rates leads to an over-determined system. To address these combined challenges, we employ DRL, specifically utilizing a Deep Deterministic Policy Gradient (DDPG) based scheme. This approach allows us to obtain efficient and feasible solutions. To the best of our knowledge, no prior studies in the existing literature have utilized the DRL approach to tackle the challenge of solving over-determined systems of equations in uplink cascaded IRS multiple access scenario. In this research, our objective is to address the existing gap in the literature by utilizing DDPG to simultaneously tune the phases of every individual IRS within the cascaded IRS system. We specifically consider the scenario of spatially correlated channels [12] between IRS<sub>1</sub> and IRS<sub>2</sub>. We aim to approach two alternative scenarios, each with a different objective: a) maximizing the rate for any specific user, or b) maximizing the sum rate for both users.

We detail our contributions below against our two main objectives (two alternative scenarios), where both scenarios consider multi-hop IRSs and multiple access systems operating in the THz range:

- **Scenario 1:** Maximize the data rate for a single user, considering the second user as an interferer.
  - We formulate the optimization problem for the cascaded IRS phase shifts, encompassing IRS<sub>1</sub> and

IRS<sub>2</sub>, and establish that it is non-convex and computationally challenging.

- Noticing the over-determined nature of this problem, we present two sub-optimal solutions to determine the ideal phase shift matrices of IRS<sub>1</sub> and IRS<sub>2</sub> that optimize the received power for the desired user, utilizing pseudo-inverse  $P_{inv}$  and block solutions ( $BLS$ ).
- We develop a DDPG scheme to obtain the ideal phase shifts for the desired user and compare its performance against the sub-optimal state-of-art schemes, as shown in Table 1.
- **Scenario 2:** Maximize the combined data rate for two users.
  - We provide analytical insights into this problem, specifically for cascaded IRSs.
  - We design a DDPG algorithm to calculate the ideal phase shifts for the cascaded IRSs that optimize the total data rate
- We then simulate and compare the results obtained for both scenarios. We benchmark these results against an upper bound derived from  $ES$  and a lower bound generated from randomly assigned phase shifts.

The subsequent sections of this research paper are organized as follows: In section II, we delve into the system and channel model concerning the multi-hop IRS scenario. Section III delves into the problem formulation, focusing on maximizing the data rate for the desired user while dealing with interference, while in section IV, we explore the problem of maximizing the total data rate for both users. Here, we derive the end-to-end sum rate within the cascaded IRS scenario. Moving on to section V, we introduce our proposed solution that employs the DDPG algorithm for controlling the phase shifts of the cascaded IRS. Numerical simulations, which shed light on our findings, are the subject of discussion in section VI. Lastly, in section VII, we draw our conclusion.

Notation: For more convenience, frequent symbols and parameters along with their description are illustrated in Table 2.

## III. SYSTEM MODEL

In our system model, we consider an uplink multi-hop IRS communication system operating in the THz frequency range, depicted in Figure 1. Within this system, we have two users, each equipped with a single antenna. Notably, the communication process is facilitated by highly directional parabolic antennas employed by both users to transmit signals precisely focused at the center of IRS<sub>1</sub>. Subsequently, these signals undergo reflection by the IRS<sub>1</sub> elements, followed by a secondary reflection from IRS<sub>2</sub> before reaching the final destination at  $R_x$ . The choice of THz frequencies is deliberate, primarily due to their suitability for scenarios with limited coverage areas—perfect for users 1 and 2. To mitigate the significant THz propagation losses caused by absorption in the air molecules we implemented a cascaded IRS system. Consequently, this setup effectively amplifies signal strength, particularly for signals facing challenges in traversing long

TABLE 1: Comparing Pseudo-Inverse with Block Solution and DDPG.

Feature	Pseudo-Inverse	Block Solution	DDPG
Complexity	High	Medium	Low
Adaptability	Static Configuration	Static Configuration	Dynamic Learning
Objective 1 (Rate of Desired User)	Sub-optimal	Decent Performance	Superior
Objective 2 (Sum Rate)	N/A	N/A	Near-optimal (Close to $ES$ )
Handling Non-Convexity	Not suitable	Not suitable	Effective
Learning from Environment	No	No	Yes
Channel Correlation Impact	Low	Moderate	Significant

distances due to their short wavelength and high frequency. The cascaded IRS configuration emerges as a strategic solution to enhance the overall robustness and efficiency of the communication process within the challenging THz spectrum. Further, each transmitter ( $T_x$ ) and  $R_x$  is equipped with antennas having diameters of  $D_t$  and  $D_r$  respectively. The distances between various points in the system are defined as follows:  $r_{tk}$ , for the distance between the users and IRS<sub>1</sub>,  $r_2$  for the distance between IRS<sub>1</sub> and IRS<sub>2</sub>, and  $r_3$  for the distance between IRS<sub>2</sub> and the  $R_x$ . The horizontal distances between the transmitters and the center of IRS<sub>1</sub>, IRS<sub>1</sub> and IRS<sub>2</sub>, the center of IRS<sub>2</sub> and the  $R_x$  are denoted as  $r_{k,1,h}$ ,  $r_{2,h}$ , and  $r_{3,h}$ , respectively. The angles of incidence and reflection with respect to the center of the illuminated areas at IRS<sub>1</sub> and IRS<sub>2</sub> are represented by  $\theta_{i,1,1}$ ,  $\theta_{i,2,1}$ ,  $\theta_{r,1}$ ,  $\theta_{i,2}$ , and  $\theta_{r,2}$ . The heights of the two transmitters, IRS<sub>1</sub>, IRS<sub>2</sub>, and the  $R_x$  are indicated as  $\ell_{Tx,1}$ ,  $\ell_{Tx,2}$ ,  $\ell_{s1}$ ,  $\ell_{s2}$ , and  $\ell_{Rx}$  respectively. The IRSs act as beamformers that focus the incoming signal at a particular reflection direction by modifying the phases of the reflecting units (RUs). The number of RUs in IRS<sub>1</sub> and IRS<sub>2</sub> are  $M$  and  $N$  respectively. The transmitted signal for each user  $k$ , where  $k \in 1, 2$ , is represented by

$$x_k = \sqrt{P_t} z_k, \quad (1)$$

Here,  $z_k$  denotes the signal corresponding to user  $k$  and has a unit power (i.e.,  $\mathbb{E}[|z_k|^2] = 1$ ,  $\mathbb{E}[\cdot]$  designates the expectation), and  $P_t$  represents the transmit power for each user. Thus, the received signal for each user  $k$  is

$$y_k = \mathbf{h}_r^H \Phi_N \mathbf{H}_{m,n}^H \Phi_M \mathbf{h}_{t,k}^H x_k + n_0, \\ y_k = \mathbf{h}_r^H \Phi_N \mathbf{H}_{m,n}^H \Phi_M \mathbf{h}_{t,k}^H \sqrt{P_t} z_k + n_0, \quad (2)$$

where  $\mathbf{h}_{t,k} \in \mathbb{C}^{1 \times M}$  is the communication link between each user  $k$  and IRS<sub>1</sub>;  $\mathbf{H}_{m,n} \in \mathbb{C}^{M \times N}$  is the communication link between IRS<sub>1</sub> and IRS<sub>2</sub>;  $\mathbf{h}_r \in \mathbb{C}^{N \times 1}$  is the communication link between IRS<sub>2</sub> and the  $R_x$ ;  $\Phi_M = \text{diag}(e^{-j\eta_1}, e^{-j\eta_2}, \dots, e^{-j\eta_M})$  and  $\Phi_N = \text{diag}(e^{-j\psi_1}, e^{-j\psi_2}, \dots, e^{-j\psi_N})$  are the phase shift reflection matrices for IRS<sub>1</sub> and IRS<sub>2</sub>, respectively, that satisfy the constant modulus constraint on each diagonal element of the matrix,  $|\phi_m|^2 = |e^{-j\eta_m}|^2 = 1$ ,  $\forall m \in \{1, 2, \dots, M\}$ ,  $|\phi_n|^2 = |e^{-j\psi_n}|^2 = 1$ ,  $\forall n \in \{1, 2, \dots, N\}$ ; and  $\text{diag}(\cdot)$  symbolizes the diagonal matrix. Moreover, the phase shifts of the  $m^{\text{th}}$  and  $n^{\text{th}}$  reflecting elements are represented by  $\eta_m$  and  $\psi_n$ , where the values of  $\eta_m$  and  $\psi_n$  range between 0 and  $2\pi$ , and the noise  $n_0 \sim \mathcal{CN}(0, \sigma^2)$  denotes the AWGN for each user in linear scale. The deterministic phase shifts are associated with the distances that the signals from each user

$k$  travel during the first hop, as well as the link between IRS<sub>2</sub> and  $R_x$  during the third hop.

$$\Omega_k = 2\pi r_{tk}/\lambda, \quad \text{and} \quad \Omega_3 = 2\pi r_3/\lambda, \quad (3)$$

where  $\lambda$  is the wavelength.

#### A. Communication Channel Model

The  $T_x$  and  $R_x$  channels  $\mathbf{h}_{t,k}$ , and  $\mathbf{h}_r$  follow the Rician fading model [11], [13]

$$\mathbf{h}_{t,k} = \sqrt{\frac{K_1}{K_1 + 1}} \bar{\mathbf{h}}_{t,k} + \sqrt{\frac{1}{K_1 + 1}} \tilde{\mathbf{h}}_{t,k}, \quad (4)$$

$$\mathbf{h}_r = \sqrt{\frac{K_2}{K_2 + 1}} \bar{\mathbf{h}}_r + \sqrt{\frac{1}{K_2 + 1}} \tilde{\mathbf{h}}_r, \quad (5)$$

Here,  $K_1$  signifies the Rician factor for  $\mathbf{h}_{t,k}$ , with  $\bar{\mathbf{h}}_{t,k}$  in  $\mathbb{C}^{1 \times M}$  representing the Line-of-Sight (LOS) component and  $\tilde{\mathbf{h}}_{t,k}$  in  $\mathbb{C}^{1 \times M}$  as the non-Line-of-Sight (NLOS) component. Similarly,  $K_2$  is the Rician factor for  $\mathbf{h}_r$ , where  $\bar{\mathbf{h}}_r$  in  $\mathbb{C}^{N \times 1}$  denotes the LOS component, and  $\tilde{\mathbf{h}}_r$  in  $\mathbb{C}^{N \times 1}$  represents the NLOS component. The channel between IRS<sub>1</sub> and IRS<sub>2</sub>,  $\mathbf{H}_{m,n} \sim \mathcal{CN}(0, \mathbf{R})$ , follows the spatially correlated Rayleigh fading channel model, where  $\mathbf{R}$  represents the covariance matrix that is derived according to the exponential spatial correlation model. It is controlled by the parameter  $\rho$  within the interval  $[0, 1]$ , denoting the correlation coefficient among neighboring RUs, and it is given as below:

$$[\mathbf{R}]_{m,n} = \rho^{|m-n|} e^{j|m-n|\theta_{i,2}}, \quad (6)$$

where  $\theta_{i,2}$  is the angle of arrival between IRS<sub>1</sub> and IRS<sub>2</sub>. High values of  $\rho$ , result in high correlation among  $\mathbf{H}_{mn}$  elements, and in cases where  $\rho$  is less than 1 (i.e. not equal to 1), the significant correlations are between adjacent RUs only, with considerably low correlation at large distances. Further, we assume that the channels  $\mathbf{h}_{t,k}$ , and  $\mathbf{h}_r$  are perfectly known for all the transmitters and the  $R_x$ .

#### B. Transmitters and Receiver Antenna Gains

The gains for the users' and  $R_x$  antennas  $G_t(o)$  and  $G_r(o)$  are expressed as

$$G_{t,k}(o) = 4e_t \frac{J_1\left(\frac{\pi D_t \sin(o)}{\lambda}\right)}{\sin(o)}. \quad (7)$$

$$G_r(o) = 4e_r \frac{J_1\left(\frac{\pi D_r \sin(o)}{\lambda}\right)}{\sin(o)}. \quad (8)$$

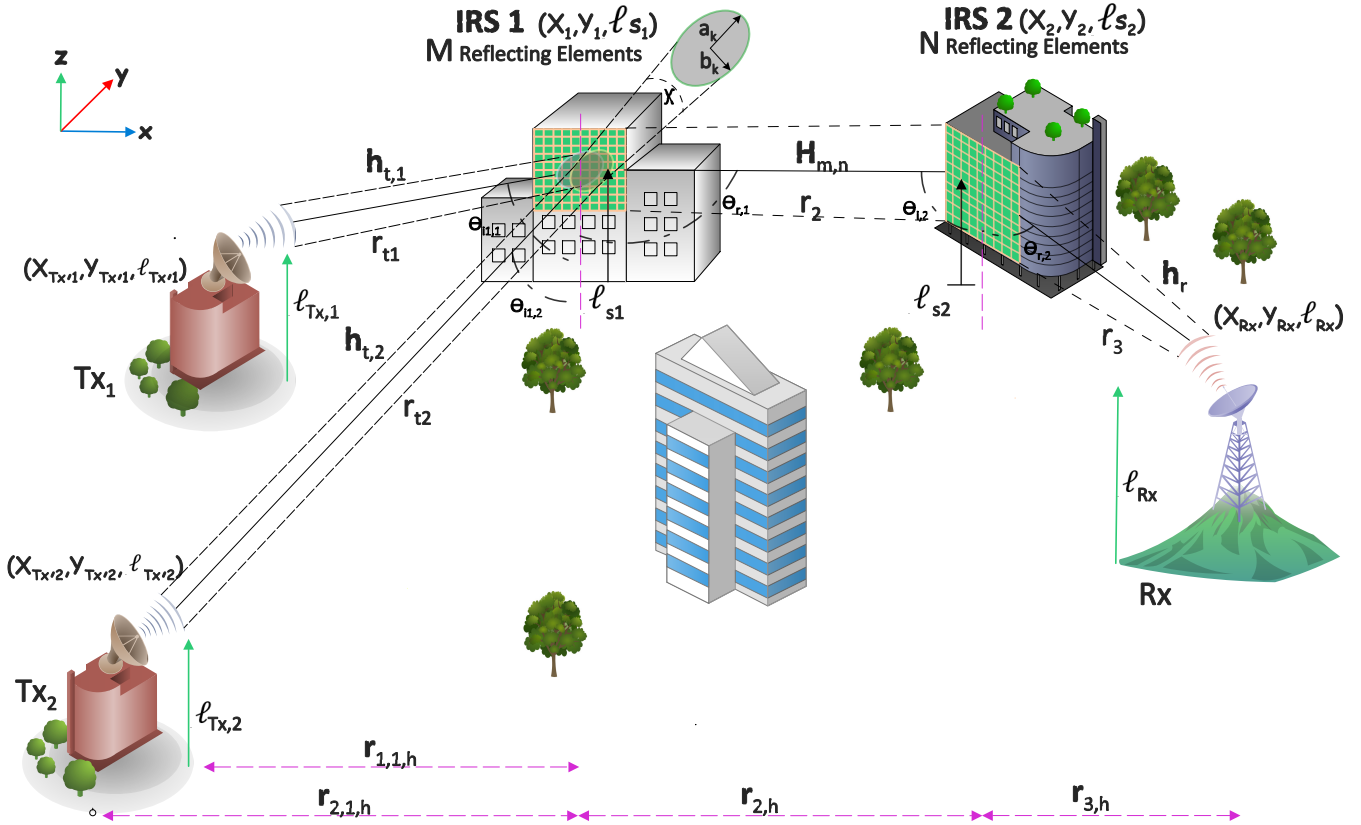


Fig. 1: Cascaded IRS system model.

Here,  $J_1(\cdot)$  denotes the first-order Bessel function of the first kind,  $D_t$  stands for the antenna diameter of the  $T_x$ , and  $D_r$  stands for the antenna diameter of the  $R_x$ . The angle, measured from the broadside of the antenna, is denoted as  $o$  [14].

Thus, the maximum gain is for  $o = 0$  and is denoted as

$$G_{t,k}(o) = e_t \left( \frac{\pi D_t}{\lambda} \right)^2. \quad (9)$$

$$G_r(o) = e_r \left( \frac{\pi D_r}{\lambda} \right)^2. \quad (10)$$

Here,  $e_t$  and  $e_r$  represents the aperture efficiencies for the  $T_x$  and the  $R_x$  respectively. Additionally, the gain of each RU is expressed as detailed in [14].

$$G(\theta_{i1,k}) = 4 \cos(\theta_{i1,k}), \quad 0 \leq \theta_{i1,k} \leq \pi/2, \quad (11)$$

where  $\theta_{i1,k}$  is the angle of incidence from user  $k$  to IRS<sub>1</sub> [14].

### C. Analysis of Loss Factors

The total losses and gains on the path between each  $T_{x_k}$  and the  $R_x$  are denoted by  $L_{\tau,k}$ . This includes the antenna gains, free space path loss (FSPL), and THz absorption loss (AS).

$$L_{\tau,k} = L_{FSPL,\tau,k} L_{abs,k}, \quad (12)$$

where  $L_{abs,k}$  represents the total THz absorption losses for each  $T_{x_k}$ . The losses in the THz range are calculated based on

atmospheric conditions using the simplified model suggested in [16].  $L_{FSPL,\tau,k}$  represents the total FSPL for each  $T_{x_k}$  and is denoted as

$$L_{FSPL,\tau,k} = L_{FSPL,k} L_{FSPL,r}. \quad (13)$$

$L_{FSPL,k}$  for the signal reflected from IRS<sub>1</sub> towards IRS<sub>2</sub> is represented as

$$L_{FSPL,k} = \frac{\left(\frac{\lambda}{4\pi}\right)^2 G_{t,k} G_{\theta_{i1,k}} G_{\theta_{r,1}}}{r_{t_k}^2}, \quad (14)$$

and  $L_{FSPL,r}$  between IRS<sub>1</sub> and the  $R_x$  is expressed as

$$L_{FSPL,r} = \frac{\left(\frac{\lambda}{4\pi}\right)^4 G_{\theta_{i,2}} G_{\theta_{r,2}} G_r}{r_2^2 r_3^2}. \quad (15)$$

The total FSPL for each  $T_{x_k}$  is expressed as

$$L_{FSPL,\tau,k} = \left(\frac{\lambda}{4\pi}\right)^6 \frac{G_{t,k} G(\theta_{i1,k}) G(\theta_{r,1}) G(\theta_{i,2}) G(\theta_{r,2}) G_r}{r_{t_k}^2 r_2^2 r_3^2}. \quad (16)$$

### D. Rate of the Desired User Under Interference

The rate of the desired user under an interference scenario is expressed as

$$R_k = \log_2(1 + \gamma_k), \quad (17)$$

Here,  $\gamma_k$  represents the Signal-to-Interference-plus-Noise-Ratio (SINR) for user  $k$ :

$$\gamma_k = \frac{P_{Rx}^{(k)}}{\sum_{\substack{i=1 \\ i \neq k}}^K P_{Rx}^{(i)} + \sigma^2}, \quad (18)$$

In this equation,  $P_{Rx}^{(k)}$  denotes the power received for user  $k$ , and  $\sigma^2$  represents the noise variance.

#### 1) Derivation of the User's Received Power ( $P_{Rx}^{(k)}$ )

In our setup, both users transmit at IRS<sub>1</sub>, covering all elements of IRS<sub>1</sub> from different angles and distances. The power reflected from the  $m^{th}$  RU of IRS<sub>1</sub> can be expressed as in [14], excluding the calculation of absorption losses.<sup>1</sup>

$$P_{r,m}^{(k)} = \left(\frac{\lambda}{4\pi}\right)^2 \frac{G_{t,k}G(\theta_{i1,k})G(\theta_{r,1})}{r_{tk}^2} \times |h_{t,km}|^2 |\phi_m|^2 P_t, \quad (19)$$

Here,  $\phi_m = e^{-j\eta_m}$  represents the reflection coefficient of the  $m^{th}$  RU of IRS<sub>1</sub>;  $G_{t,k}$  denotes the antenna gain of user  $k$  at the  $T_x$ ;  $G(\theta_{i1,k})$  is the gain of the RU from the incident angle,  $G(\theta_{r,1})$  is the gain of the RU from the reflection angle, and  $r_{t,k}$  is the distance between  $T_{xk}$  and RU  $m$ . Similarly, the power reflected from the  $n^{th}$  RU of IRS<sub>2</sub> when it is illuminated by the signal reflected by the  $m^{th}$  RU of IRS<sub>1</sub> can be expressed as follows:

$$P_{r,mn}^{(k)} = \left(\frac{\lambda}{4\pi}\right)^4 \frac{G_{t,k}G(\theta_{i1,k})G(\theta_{r,1})G(\theta_{i,2})G(\theta_{r,2})}{r_{tk}^2 r_{tn}^2} \times |h_{t,km}|^2 |\phi_m|^2 |H_{mn}|^2 |\phi_n|^2 P_t, \quad (20)$$

In this equation,  $\phi_n = e^{-j\psi_n}$  represents the reflection coefficient of the  $n^{th}$  RU of IRS<sub>2</sub>, and  $H_{mn}$  denotes the  $(m, n)$  element of the channel matrix  $\mathbf{H}$  connecting IRS<sub>1</sub> to IRS<sub>2</sub>. The received power captured at the receiver  $R_x$  through the channel  $H_{mn}$  can be expressed as follows:

$$P_{rx,mn}^{(k)} = \left(\frac{\lambda}{4\pi}\right)^2 \frac{P_{r,mn}^{(k)}}{r_3^2} G_r |h_{rn}|^2, \quad (21)$$

and the total power received for user  $k$  at the  $R_x$  is expressed as [14]

$$P_{Rx}^{(k)} = \left| \sqrt{L_{\tau,k}} \sum_{m=1}^M \sum_{n=1}^N |h_{t,km}| |H_{mn}| |h_{rn}| \times e^{-j(\varphi_{t_{km}} + \eta_m + \varphi_{mn} + \psi_n + \varphi_{rn} + \Omega_k + \Omega_3)} \right|^2 P_t, \quad (22)$$

where  $\varphi_{t_{km}}$  is the phase for the  $T_x$  channel for user  $k$  and  $m^{th}$  RU,  $\varphi_{mn}$  is the phase for the  $\mathbf{h}_{m,n}$  channel for  $m^{th}$  and  $n^{th}$  RU, and  $\varphi_{rn}$  is the phase for the  $R_x$  channel for  $n^{th}$  RU. Therefore, (22) can be re-written as

$$P_{Rx}^{(k)} = \left| \sqrt{L_{\tau,k}} \sum_{m=1}^M \sum_{n=1}^N |h_{t,km}| |H_{mn}| |h_{rn}| \times e^{-j(\varphi_{t_{km}} + \eta_m + \varphi_{mn} + \psi_n + \varphi_{rn} + \Omega_k + \Omega_3)} \right|^2 P_t. \quad (23)$$

<sup>1</sup>The absorption loss,  $L_{abs,k}$ , is included in  $L_{\tau,k}$  which will show later in the expression for the total received power.

**Proposition 1.** It can be shown that the total received power for each  $T_{xk}$  at the  $R_x$  can be given as follows:

$$P_{Rx}^{(k)} = \left| \sqrt{L_{\tau,k}} e^{-j\Omega_3} \mathbf{h}_r^H \mathbf{\Phi}_N \mathbf{H}_{mn}^H \mathbf{\Phi}_M \mathbf{h}_{t,k}^H e^{-j\Omega_k} \right|^2 P_t. \quad (24)$$

*Proof.* The proof is given in Appendix A.  $\square$

#### 2) Derivation of the User's SINR

To this end, substituting (24) in (18) leads to the below signal-to-interference-plus-noise ratio (SINR) for user 1

$$\gamma_k = \frac{\left| \sqrt{L_{\tau,k}} e^{-j\Omega_3} \mathbf{h}_r^H \mathbf{\Phi}_N \mathbf{H}_{m,n}^H \mathbf{\Phi}_M \mathbf{h}_{t,k}^H e^{-j\Omega_k} \right|^2 P_t}{\sum_{\substack{i=1 \\ i \neq k}}^K \left| \sqrt{L_{\tau,i}} e^{-j\Omega_3} \mathbf{h}_r^H \mathbf{\Phi}_N \mathbf{H}_{m,n}^H \mathbf{\Phi}_M \mathbf{h}_{t,i}^H e^{-j\Omega_i} \right|^2 P_t + \sigma^2}. \quad (25)$$

### IV. MAXIMIZING THE RATE OF A DESIRED USER UNDER INTERFERENCE

In this section, we provide the analytical derivations of the first objective, which is maximizing the rate of a desired user, while the other user is considered an interferer. Our objective is to find the optimum phases for the multi-hop IRSs that maximize the received rate of the desired user. We will show that the rate maximization problem is non-convex and finding a closed-form expression of the IRS phases is mathematically intractable. Then we propose a sub-optimal solution to the problem through maximizing the received power of the desired user. In addition, we propose a DDPG algorithm that maximizes the rate of the desired user.

To maximize the rate of the desired user under interference, we need to solve:

$$\begin{aligned} & \max_{\mathbf{\Phi}_N, \mathbf{\Phi}_M} \log_2(1 + \gamma_k), \\ & \text{s.t. } C1 : |\phi_m|^2 = 1, \forall m \in \{1, 2, \dots, M\}, \\ & \quad C2 : |\phi_n|^2 = 1, \forall n \in \{1, 2, \dots, N\}, \end{aligned} \quad (26)$$

The optimization problem presented in (26) is considered NP-hard, making the solution non-trivial due to the non-convexity arising from the constant amplitude constraints of IRS<sub>1</sub> and IRS<sub>2</sub> reflecting elements. The constant modulus constraint is a mathematical condition that restricts the square of the magnitude of a complex variable to a fixed value. This non-convexity introduces challenges in finding the global optimum, resulting in multiple local optima. Consequently, obtaining an analytical closed-form expression for the optimal phase shifts of both IRSs is mathematically intractable. The optimal solution needs to strike a balance between enhancing the received Signal-to-Noise Ratio (SNR) for the desired user and mitigating interference from the other user, although these sub-objectives may not necessarily align. Therefore, we employ a sub-optimal approach for (26) by focusing on maximizing the power received for the desired user [15].

#### 1) Sub-optimal Solutions: Maximizing the Received Power of the Desired User

To maximize the total received power of the desired user (e.g. user 1), we will solve the following system of equations:

$$\eta_m + \psi_n + \varphi_{t_{1m}} + \varphi_{mn} + \varphi_{r_n} + \Omega_1 + \Omega_3 = \nu, \quad \forall m, n. \quad (27)$$

TABLE 2: List of frequently used parameters and symbols.

Parameters and Symbols	Description
$x_k$	Transmitted signal for each user $k$
$y_k$	Received signal for each user $k$
$z_k$	Signal for each user $k$
$P_t$	Transmit power for each user
$\lambda$	Wavelength
$\mathbf{h}_r$	Channel between IRS <sub>2</sub> and the receiver
$\mathbf{H}_{m,n}$	Channel between IRS <sub>1</sub> and IRS <sub>2</sub>
$\mathbf{h}_{t,k}$	Channel between each user $k$ and IRS <sub>1</sub>
$n_0$	AWGN in linear scale
$K_1, K_2$	Rician Factor for the transmitter channel and receiver channel
$\mathbf{R}$	Covariance Matrix
$\Phi_M, \Phi_N$	Phase shift reflection matrix for IRS <sub>1</sub> and IRS <sub>2</sub> , respectively
$\phi_m, \phi_n$	Phase shift of IRS <sub>1</sub> reflecting element $m$ and IRS <sub>2</sub> reflecting element $n$ , respectively
$\eta_m$	Phase shift of $m^{th}$ IRS <sub>1</sub> reflecting element
$\psi_n$	Phase shift of $n^{th}$ IRS <sub>2</sub> reflecting element
$D_t, D_r$	Antenna diameters for each $T_{x_k}$ and $R_x$ , respectively
$r_{tk}, r_2, r_3$	Distance between: each user $k$ and IRS <sub>1</sub> , IRS <sub>1</sub> and IRS <sub>2</sub> , and IRS <sub>2</sub> and $R_x$ , respectively
$r_{k,1,h}, r_{2,h}, r_{3,h}$	Horizontal distance between: user $k$ and center of IRS <sub>1</sub> , centers of IRS <sub>1</sub> and IRS <sub>2</sub> , and center of IRS <sub>2</sub> and $R_x$ , respectively
$\theta_{ik,1}$	Incident angle from user $k$ w.r.t. the center of the illuminated area at IRS <sub>1</sub>
$\theta_{r,1}$	Reflected angle w.r.t. the center of the illuminated area at IRS <sub>1</sub>
$\theta_{i,2}$	Incident angle from IRS <sub>1</sub> w.r.t. the center of the illuminated area at IRS <sub>2</sub>
$\theta_{r,2}$	Reflected angle w.r.t. the center of the illuminated area at IRS <sub>2</sub>
$\ell_{Tx,k}, \ell_{Rx}, \ell_{s1}, \ell_{s2}$	The height of the $T_{x_k}$ , $R_x$ , IRS <sub>1</sub> and IRS <sub>2</sub> , respectively
$M, N$	Number of reflecting elements for IRS <sub>1</sub> and IRS <sub>2</sub> respectively
$[\mathbf{R}]_{m,n}$	Covariance matrix obtained based on the exponential spatial correlation model
$\rho^{ m-n }$	Correlation-coefficient among the adjacent RUs
$\Omega_k$	Phase shifts corresponding to signal traveled from user $k$ to IRS <sub>1</sub>
$\Omega_3$	Phase shift corresponding to signal traveled from IRS <sub>2</sub> to $R_x$
$o$	Angle measured from the broadside of the antenna
$G_t(o), G_r(o)$	Gains for the users' and receiver antennas respectively
$L_{\tau,k}$	Total losses and gains on the path between each $T_{x_k}$ and the $R_x$
$L_{FSPL,\tau,k}$	Total FSPL for $T_{x_k}$
$L_{abs,\tau,k}$	Total absorption loss for $T_{x_k}$
$G_{t,k}, G_r$	$T_{x_k}$ , and $R_x$ antenna gains
$G(\theta_{i1,k}), (\theta_{r,1})$	Gain of IRS <sub>1</sub> RU from the incident and reflection angles
$P_{r,m}$	The power reflected from the $m^{th}$ RU of IRS <sub>1</sub>
$P_{r,mn}$	The power reflected from the $n^{th}$ RU of IRS <sub>2</sub> because of being illuminated by the signal reflected by the $m^{th}$ RU of IRS <sub>1</sub>
$P_{rx,mn}$	The received captured power at the $R_x$
$P_{Rx}$	The total received power for user $k$ at the receiver ( $R_x$ )
$\varphi_{t,k_m}$	The phase for the transmitter channel for user $k$ and $m^{th}$ RU
$\varphi_{mn}$	The phase for the $\mathbf{h}_{m,n}$ channel for $m^{th}$ and $n^{th}$ RU
$\varphi_{r_n}$	The phase for the receiver channel for $n^{th}$ RU
$\gamma_k$	The received SINR for the $T_{x_k}$ at the $R_x$
$R_k$	The data rate for user $k$
$R_{sum}$	The sum rate for both users
$\Delta\Phi$	Phase Search Step

Here,  $\nu$  is an arbitrary constant. Equation (27) can be interpreted as follows:  $P_{Rx}^{(k)}$  will be maximized when the phases across all the paths established by the different IRS elements of both IRSs are constant for all  $m$  and  $n$ . Without loss of generality, we can set  $\nu = 0$ . Equation (27) represents an over-determined system of equations with  $M + N$  unknowns (the phase shifts of the elements of IRS<sub>1</sub> and IRS<sub>2</sub>) and  $M \times N$  equations (corresponding to the different paths established through all combinations of the  $M$  reflective elements of IRS<sub>1</sub> and the  $N$  reflective elements of IRS<sub>2</sub>).

The set of equations in (27) can be represented as:

$$\mathbf{A}\Theta = \mathbf{C}, \quad (28)$$

Here,  $\Theta$  is a column vector with dimensions  $(M + N) \times 1$  representing the phase shifts of IRS<sub>1</sub> and IRS<sub>2</sub>, denoted as  $\eta_1, \eta_2, \dots, \eta_M, \psi_1, \psi_2, \dots, \psi_N$ .  $\mathbf{A}$  is a binary matrix with

dimensions  $(M \times N) \times (M + N)$ , and  $\mathbf{C}$  is a column vector with dimensions  $(M \times N) \times 1$ .  $\mathbf{C}$  contains known constant values, including the phase shifts of the  $T_x$  channel,  $\mathbf{h}_{t,k}$ , the phase shifts of the channel between IRS<sub>1</sub> and IRS<sub>2</sub>,  $\mathbf{H}_{mn}$ , and the phase shifts of the  $R_x$  channel,  $\mathbf{h}_r$ . We address this problem by employing two sub-optimal mathematical techniques, namely, the  $P_{inv}$  method and the  $BLS$ . These methods allow us to determine the unknown phase shifts,  $\eta_m$  and  $\psi_n$ , and calculate the received power for a selected user (in this case, user 1, chosen arbitrarily as the desired user).

*Pseudo-Inverse Solution:* The  $P_{inv}$  solution for the over-determined system in (28) is expressed as

$$\Theta = \mathbf{A}^+ \mathbf{C}. \quad (29)$$

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

$$\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T, \quad (30)$$

**Block Solution:** Based on the spatially correlated channel assumption, a low-complexity approximate solution built on the exponential correlation model can be developed for higher values of the correlation-coefficient,  $\rho$ , between the adjacent RUs. In the case where the channel correlation is very high, the channel,  $\mathbf{H}_{m,n}$ , can be assumed to have a block structure, where the elements in the channel matrix are organized into groups, where each group shares the same phase and exhibits no correlation between the contiguous blocks [9]. Since the number of unknowns  $M + N$ , the total number of IRS<sub>1</sub> and IRS<sub>2</sub> elements, is smaller than the number of equations  $M \times N$ , the total number of elements in the channel matrix between IRS<sub>1</sub> and IRS<sub>2</sub>, then we need to reduce the number of equations. The primary concept behind the *BLS* is to minimize the number of equations by treating each group of channel elements as one block with the same value. This will result in redundant equations in the channel matrix  $\mathbf{H}_{m,n}$  that can be eliminated using row reduction methods such as Gaussian elimination reducing the rank of  $\mathbf{H}_{m,n}$  to the number of blocks  $\frac{M \times N}{N_{\text{blk}}}$ . This reduction in the rows of the channel matrix will be reflected in the equation (27) which will no longer be over-determined because the number of equations is reduced. An important point to emphasize is that the count of independent IRS<sub>1</sub> and IRS<sub>2</sub> elements remains consistent. However, through this approach, the number of equations has been streamlined from  $M \times N$  to  $M + N$ . By employing this method, the previously over-determined system can be resolved whenever the number of blocks  $\frac{M \times N}{N_{\text{blk}}} \leq M + N$ .

As described in Algorithm 1, we will first check if the

---

**Algorithm 1** Block Solution-based Framework

---

```

1: Input:  $h_t, M, H_{m,n}, N, h_r, N_{\text{blk}}$ 
2: Output:  $\eta_m, \psi_n, P_{R_x k}$ 
3: if  $M + N \leq M \times N$  then
4:   Divide the elements of channel  $\mathbf{H}_{m,n}$  into blocks.
5:   if  $M + N \geq \frac{M \times N}{N_{\text{blk}}}$  then
6:     Calculate the total received power using (23) by treating
       each group of channel elements as one block with the same value.
7:   else
8:     if  $M + N < \frac{M \times N}{N_{\text{blk}}}$  then
9:       Solve (27) using the  $P_{inv}$  solution, and calculate the
       total received power using (23).
10:   endif
11: endif
12: endif

```

---

number of unknowns  $M + N$  is less than or equal to the number of equations  $M \times N$ . If this condition is met, it implies that the system is over-determined. In this case, we will divide the elements of the channel matrix  $\mathbf{H}_{m,n}$  into blocks. After dividing it into blocks, if the number of unknowns  $M + N$  is greater or equal to the number of equations  $\frac{M \times N}{N_{\text{blk}}}$ , then we will calculate the total received power using the formula (23) treating each group of channel

elements as one block with the same value. Else if the number of unknowns  $M + N$  is still smaller than the number of equations  $\frac{M \times N}{N_{\text{blk}}}$ , then we will solve (27) using the  $P_{inv}$  solution, and calculate the total received power using (23).

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

## V. MAXIMIZING THE SUM RATE FOR BOTH USERS

In our second scenario, we aim to find the optimal phase shifts for the elements of both IRS<sub>1</sub> and IRS<sub>2</sub> to maximize the total data rate for all users at the receiver  $R_x$ , which can be formulated as follows:

$$R_{sum} = \sum_{k=1}^K \log_2(1 + \gamma_k). \quad (31)$$

Therefore, the problem formulated for IRS<sub>1</sub> and IRS<sub>2</sub> is to determine the phase shift matrices  $\Phi_N$  and  $\Phi_M$  that maximize the total sum rate  $R_{sum}$ , and it can be formulated as

$$\begin{aligned}
& \max_{\Phi_N, \Phi_M} \sum_{k=1}^K \log_2(1 + \gamma_k), \\
& \text{s.t. } C1 : |\phi_m|^2 = 1, \forall m \in \{1, 2, \dots, M\}, \\
& \quad C2 : |\phi_n|^2 = 1, \forall n \in \{1, 2, \dots, N\},
\end{aligned} \quad (32)$$

Similar to the optimization problem presented in (26), this optimization problem is NP-hard, and finding a solution is non-trivial because of its non-convex nature caused by the constant amplitude constraints of IRS<sub>1</sub> and IRS<sub>2</sub> reflecting elements. As a result, obtaining an analytical solution is not feasible. To address this issue, we employ the Deep Reinforcement Learning (DRL) technique, specifically the DDPG algorithm, instead of attempting to solve the challenging problem through mathematical methods. Moreover, we provide two limiting cases for the solution of this problem; an upper bound case where there is no interference and full channel phase compensation can be achieved as well as a lower bound case where phases of the elements of the IRSs are randomly chosen [15].

### A. Upper bound on Performance

Assuming the case of null interference and that the IRSs can be used to fully cancel phase shifts between different reflection paths, a relaxed upper bound on the sum rate can be established using (25), where the SINR on user  $k$  becomes

$$\gamma_k^U = \frac{\left| \sqrt{L_{\tau,k}} \mathbf{h}_r^H \mathbf{H}_{m,n}^H \mathbf{h}_{t,k} \right|^2 P_t}{\sigma^2}, \quad (33)$$

and the upper bound on the sum rate becomes

$$R_{sum}^U = \sum_{k=1}^K \log_2(1 + \gamma_k^U). \quad (34)$$



## VI. DDPG FOR CASCADED IRS PHASE CONTROL

### A. Introduction to DDPG

In the DDPG scheme (see Fig. 2), the agents (represented by the IRSs) interacts with the environment (the communication system model) to learn an optimal policy. The state  $\mathbf{s}^{(T)}$  of the agent at time step  $T$  is determined by various factors such as the received SINRs for  $T_{x_1}$  and  $T_{x_2}$  at the  $R_x$ , and the previous sum rate for both users at time step  $(T - 1)$ . The action  $\mathbf{a}^{(T)}$  taken by the agent includes the  $\text{IRS}_1$  and  $\text{IRS}_2$  phase shift matrices. The agent's objective is to maximize the reward  $\mathbf{r}^{(T)}$ , which stands for the highest sum-rate achieved. The DDPG scheme is a powerful and effective approach for solving the non-convex optimization problems in our system. It enables the agent to learn and adapt to the dynamic environment, making it well-suited for the optimizations problems in the THz cascaded IRS system [19].

The target networks and the online networks are the two sets of networks that the DDPG algorithm uses. An actor network and a critic network are part of the online networks. At each time step, the actor network chooses an action for each state based on the input state. The actor network's choice of action is evaluated by the critic network for quality. On the other side, the target networks are utilized to increase the learning process stability. The specifications of the online networks are used to update the target networks regularly. Only a small portion of the parameters from the online networks are communicated to the target networks during the soft updates. The agent interacts with the environment by taking actions based on the current situation and receives feedback in the form of rewards during the learning process. The agent stores these experiences in a replay memory, which is a collection of past experiences. This replay memory is then used to randomly sample experiences to train the networks. The DDPG algorithm employs off-policy learning, meaning that the agent learns from past experiences stored in the replay memory rather than relying solely on the most recent experience. The weights of the networks are updated using the DDPG algorithm based on the loss function using an optimization technique, such as the Adam optimizer. The loss function is intended to maximize the expected cumulative reward (Q-value) estimated by the critic network and reduce the difference between the predicted actions and the target actions. The DDPG algorithm seeks to identify the best possible policy that maximizes the sum rate in the current environment by repeatedly updating the weights of the networks based on the accumulated experiences. The agent's learning process doesn't stop until it reaches an acceptable level of performance or convergence. The DDPG scheme is a powerful and effective approach for solving the non-convex optimization problems in (26) and (32) in our system. It enables the agent to learn and adapt to the dynamic environment, making it well-suited for the optimization of the THz cascaded IRS uplink scheme [19].

As stated earlier, the optimization problems in (26) and (32) are non-convex, and finding the optimum phases that maximize the rate through  $ES$  is computationally infeasible. Hence, we design DDPG solutions to find the optimum

phases of the cascaded IRS. In this section, we introduce our approach using the DDPG algorithm to tackle the optimization problems outlined in equations (26) and (32) for the cascaded IRS system. Deep learning isn't a suitable fit for such a dynamic wireless communication context because it relies on the availability of training data, and deep Q-networks are also unsuitable since they are designed for discrete-time spaces exclusively. Furthermore, the convergence of the policy gradient (PG) algorithm isn't sufficient within the framework of wireless communication. The DDPG model we've chosen is well-suited for our dynamic, continuous, and non-convex wireless communication scenario. It is designed for continuous action spaces and is well-suited for problems where the action space is not discrete. It has been applied to various continuous dynamic tasks, making it suitable for scenarios where wireless communication parameters need to be adjusted continuously [17]-[18].

The primary aim of the DDPG model is to learn the policy that solves the optimization problems outlined in (26) and (32). The DDPG scheme consists of essential components: agents, states  $\mathbf{s}^{(T)}$ , actions  $\mathbf{a}^{(T)}$ , rewards  $\mathbf{r}^{(T)}$ , the policy function  $\mu$ , and the Q-value function  $Q(\mathbf{s}, \mathbf{a}|\theta^Q)$ . Our agents,  $\text{IRS}_1$  and  $\text{IRS}_2$ , operate in the environment which is the communication system, and the states  $\mathbf{s}^{(T)}$  represent the received SINR for  $T_{x_1}$  at  $R_x$ , the SINR for  $T_{x_2}$  at  $R_x$ , and the sum rate for users at time step  $(T - 1)$ . The actions  $\mathbf{a}^{(T)}$  correspond to the phase shifts of  $\text{IRS}_1$  and  $\text{IRS}_2$ , and the reward  $\mathbf{r}^{(T)}$  is based on the received power for user 1 for our first objective and the sum rate for the users for the second objective. Our goal is to optimize the average rewards, which involve both immediate and future rewards. The DDPG scheme incorporates four  $NN$ s: the actor, the critic, the target actor, and the target critic networks, ensuring stability in the learning process [19].

### B. DDPG System Mapping

The initial stage in addressing the optimization problem in our system model involves mapping it to the fundamental components of the DDPG algorithm. These components include defining the state, the action, and the reward functions. We will explore this mapping process and provide an overview of how the DDPG algorithm behaves.

#### 1) State space

The state space of the DDPG agent at timestep  $T$  is specified as follows:

For the first objective:

$$\mathbf{s}^{(T)} = [\gamma_1^{(T-1)}, \gamma_2^{(T-1)}, P_{R_{x_k}}^{(T-1)}], \quad (35)$$

For the second objective:

$$\mathbf{s}^{(T)} = [\gamma_1^{(T-1)}, \gamma_2^{(T-1)}, R_{sum}^{(T-1)}], \quad (36)$$

Here,  $\gamma_1^{(T-1)}$ ,  $\gamma_2^{(T-1)}$ ,  $P_{R_{x_k}}^{(T-1)}$ , and  $R_{sum}^{(T-1)}$  represent the received SINR for  $T_{x_1}$  at the  $R_x$ , the received SINR for  $T_{x_2}$  at the  $R_x$ , and the sum rate for users at time step  $(T - 1)$  respectively.



to estimate the value of future states, which is used to train the current networks [19].

### Algorithm 2 DDPG-based Framework

- 1: **Initialization:** Begin by setting  $T$  to 0 and initializing the replay buffer of the DDPG agent, denoted as  $\mathcal{D}$ , with a capacity of  $M$ .
- 2: Randomly initialize the actor network weights as  $\theta^\mu$  and the critic network weights as  $\theta^Q$ .
- 3: Initialize the target networks: Set  $\theta^{\mu'}$  to  $\theta^\mu$  and  $\theta^{Q'}$  to  $\theta^Q$ .
- 4: **for**  $T = 1$  to  $\infty$  **do**
- 5: Observe the current state  $\mathbf{s}^{(T)}$  and choose an action while taking into account the exploration noise generated by the OU process.  $\mathbf{a}^{(T)} = \mu(\mathbf{s}^{(T)}|\theta^\mu) + \mathbf{n}_T$
- 6: Execute action  $\mathbf{a}^{(T)}$  at IRS<sub>1</sub> and IRS<sub>2</sub>.
- 7: Obtain the immediate reward  $r^{(T)}$ , observe the subsequent state  $\mathbf{s}^{(T+1)}$ , and then record this transition as  $(\mathbf{s}^{(T)}, \mathbf{a}^{(T)}, r^{(T)}, \mathbf{s}^{(T+1)})$  within the replay buffer  $\mathcal{D}$ .
- 8: Select a random mini-batch of transitions from the replay buffer  $\mathcal{D}$   $B \leftarrow \{(\mathbf{s}^{(i)}, \mathbf{a}^{(i)}, r^{(i)}, \mathbf{s}^{(i+1)})\} \in \mathcal{D}$ .
- 9: Compute the targets in the following manner:  $\tilde{Q}(\mathbf{s}^{(i)}, \mathbf{a}^{(i)}|\theta^Q) = r^{(i)} + \gamma Q(\mathbf{s}^{(i+1)}, \mu(\mathbf{s}^{(i+1)}|\theta^{\mu'})|\theta^Q)$
- 10: Update the parameters  $\theta^Q$  in the critic network by minimizing the loss function:  $L = \frac{1}{|B|} \sum_{i=1}^{|B|} (\tilde{Q}(\mathbf{s}^{(i)}, \mathbf{a}^{(i)}|\theta^Q) - Q(\mathbf{s}^{(i)}, \mathbf{a}^{(i)}|\theta^Q))^2$
- 11: Adjust the parameters  $\theta^\mu$  in the actor network using the sampled policy gradient:  $\nabla_{\theta^\mu} J \approx \frac{1}{|B|} \sum_{i=1}^{|B|} \nabla_a Q(\mathbf{s}^{(i)}, \mathbf{a}^{(i)}|\theta^Q) \nabla_{\theta^\mu} \mu(\mathbf{s}^{(i)}|\theta^\mu)$
- 12: Update the target actor and target critic networks:  $\theta^{Q'} \leftarrow \tau \theta^Q + (1 - \tau) \theta^{Q'}$   $\theta^{\mu'} \leftarrow \tau \theta^\mu + (1 - \tau) \theta^{\mu'}$
- 13: **end for**

### C. Complexity Analysis

To disclose the complexity of the DDPG algorithm, we demonstrate a quantitative analysis of the proposed DDPG-based scheme  $C_{DDPG}$  versus the complexity of  $P_{inv}$ , BLS, and ES methods. In our analysis, we focused on the computations performed during the exploitation stage of the DDPG algorithm to evaluate its complexity, which heavily relies on the actor network's architecture. Deep NNs (DNNs) consist of input, output, and concealed layers (i.e. the layers that lie between the input and output strata of a DNN). We examine several factors in our analysis, including the count of states ( $\mathcal{S}$ ), the count of neurons within the input of each layer ( $\mathcal{J}$ ), the count of concealed layers ( $\mathcal{H}$ ), the count of neurons in the output of each layer ( $\mathcal{O}$ ), and the count of actions ( $\mathcal{A}$ ). The complexity of the input layer relates to  $\mathcal{S} \times \mathcal{J}$ , and the complexity of the concealed layers correlates to  $\mathcal{H} \times \mathcal{J} \times \mathcal{O}$ , and the complexity of the output layer is connected to  $\mathcal{O} \times \mathcal{A}$ . Consequently, the cumulative complexity of the DDPG scheme is expressed as follows:

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

Furthermore, in the DDPG scheme, the action yielding the highest reward is always selected, and a linear search is performed on the output. Consequently, the overall computational complexity of a NN forward pass can be represented as [21]:

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

Please note that the above analysis assumes a simplified perspective and doesn't consider additional factors such as

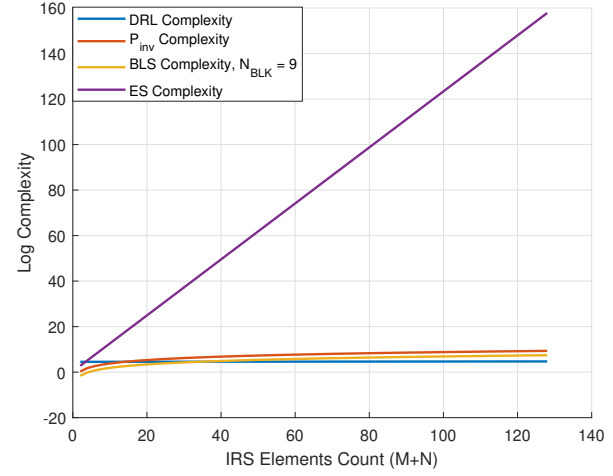


Fig. 3: Complexity of DRL vs.  $P_{inv}$  vs. BLS vs. ES.  $M = N = 64$ ,  $NBLK = 9$ .

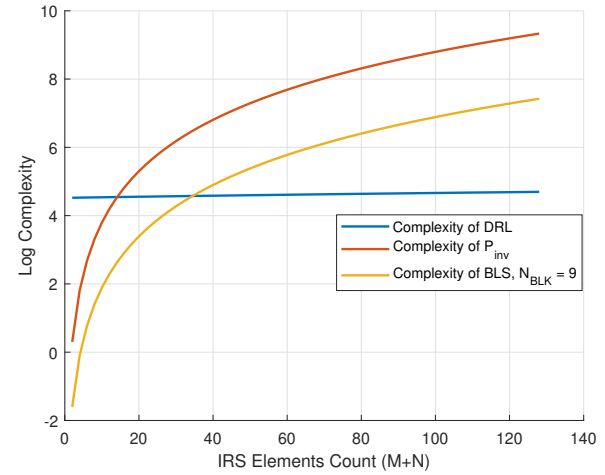


Fig. 4: Complexity of DRL vs.  $P_{inv}$  vs. BLS.  $M = N = 64$ ,  $NBLK = 9$ .

activation functions, regularization techniques, or the training process itself.

In contrast, the  $P_{inv}$  solution of a matrix  $A$  with dimensions  $MN \times (M + N)$  is computed using its singular value decomposition (SVD), a useful computational technique for dimensionality reduction in over-determined systems, with a complexity of  $O((MN)^2 \times (M + N))$ , where  $MN > (M + N)$ , and  $M$  and  $N$  represent IRS<sub>1</sub> and IRS<sub>2</sub> number of elements, respectively [22], [23]. On the other hand, for the BLS method, the complexity of inverting a matrix  $A$  with dimension of  $\frac{MN}{N_{blk}} \times (M + N)$  is reduced to  $O((\frac{MN}{N_{blk}})^2 \times (M + N))$ , where  $\frac{MN}{N_{blk}} \leq (M + N)$ .

Furthermore, the complexity of the ES method ( $C_{ES}$ ) is given as:

$$C_{ES} = O\left(\mathcal{K} \times \left(\left\lfloor \frac{2\pi}{\Delta\Phi} \right\rfloor + 1\right)^{(M+N)}\right), \quad (40)$$

where  $\mathcal{K}$  designates the users' count and  $\left\lfloor \frac{2\pi}{\Delta\Phi} \right\rfloor$  stands for the number of phase search steps. Consequently, the DDPG algorithm shows lower complexity than that of the  $P_{inv}$ , BLS, and ES techniques as the IRS elements count increases. This relationship is illustrated in figures: Fig.3 and Fig.4.

## VII. NUMERICAL RESULTS

In this section, we assess the performance of the suggested DDPG-based cascaded IRS-aided wireless THz communication scheme. To evaluate the DDPG system's effectiveness, we conduct a comparative analysis, focusing on two key scenarios: one aimed at maximizing the rate for the desired user (user 1) and the other at maximizing the total rate for both users.

When optimizing the rate for the desired user, we establish two reference benchmarking schemes for our system, both utilizing 18 reflecting elements for IRS<sub>1</sub> and IRS<sub>2</sub>. The first scheme relies on a  $P_{inv}$  approach, while the second employs the BLS method. Furthermore, for the sum rate maximization scenario, we compare the sum rates achieved using the DDPG algorithm with a discretized ES approximation. This comparison serves to demonstrate that the DDPG system performs closely to the ES method. To mitigate the computational complexity associated with the ES, we limit the number of reflecting elements to 4 for both IRS<sub>1</sub> and IRS<sub>2</sub>. We discretize the phase shifts within the range of 0 to  $2\pi$  with a search step of  $2\pi/72$ , resulting in  $(72 + 1)^{(M+N)}$  possible combinations of phase shift reflection matrices. After obtaining the optimal phase shift matrices, we compute the corresponding total rate for users. Additionally, we benchmark the DDPG-derived sum rates against those calculated using random phase generation (i.e., without optimization) as an additional reference point.

### A. Simulation Parameters

The simulation parameters, as shown in Table 3, provide the details of the experimental setup, which include the use of 18 reflecting elements for both IRS<sub>1</sub> and IRS<sub>2</sub>, two users ( $K = 2$ ), each equipped with a single antenna ( $N_t = 1$ ), and the  $R_x$  also had one antenna ( $N_r = 1$ ). The wavelength was set at  $\lambda = 10^{-3}$ . The channels, encompassing the link between user 1 and IRS<sub>1</sub>, user 2 and IRS<sub>1</sub>, IRS<sub>2</sub>, and the  $R_x$ , are characterized by a Rician fading model involving Rician factors of  $K_1 = 10$  and  $K_2 = 10$ . The path loss exponent for the  $T_x$  to IRS<sub>1</sub> channel is 2, and the same exponent is applied to the channel between IRS<sub>2</sub> and the  $R_x$ . We selected a carrier frequency of  $f = 300 \times 10^9$  and a bandwidth of  $BW = 2 \times 10^9$ . The noise spectral density was set to  $N_{PSD} = -174 \text{ dB/Hz}$ , with a noise figure at the  $R_x$  of  $F_{dB} = 10 \text{ dB}$ . The coordinates for IRS<sub>1</sub> were  $(x_{r1} = 5, y_{r1} = 10, h_{r1} = 12)$ , and for IRS<sub>2</sub>, the coordinates were  $(x_{r2} = 10, y_{r2} = 10, h_{r2} = 12)$ . The distance between user 1 and IRS<sub>1</sub> varied between  $r_{t1} = 3 \text{ m}$  and  $15 \text{ m}$ , while user 2 maintained a fixed distance of  $15 \text{ m}$  from IRS<sub>1</sub>. The coordinates of the  $R_x$  were  $(x_{rx} = 20, y_{rx} = 0, h_r = 5)$ . The antenna diameter was set at  $D_t = 0.12 \text{ m}$ ,

TABLE 3: Parameters Used in Simulation

Simulation Parameters	Values
Number of Users ( $K$ )	2
Number of antennas per user $N_t$	1
Number of antennas at the receiver $N_r$	1
Speed of the light $c$	$3 \times 10^8$
Carrier Frequency $f$	$300 \times 10^9$
Wavelength $\lambda$	$1 \times 10^{-3}$
Number of IRS <sub>1</sub> (M) and IRS <sub>2</sub> (N) Reflecting Elements	18, 18
Coordinates of IRS <sub>1</sub> ( $x_{r1}, y_{r1}, h_{r1}$ )	(5,10,12)
Coordinates of IRS <sub>2</sub> ( $x_{r2}, y_{r2}, h_{r2}$ )	(10,10,12)
Distance between User 1 and IRS <sub>1</sub>	3 to 15
Distance between User 2 and IRS <sub>1</sub>	15
IRS <sub>1</sub> and IRS <sub>2</sub> Reflection Amplitudes $\alpha$	1
IRS <sub>1</sub> and IRS <sub>2</sub> half-power Spacing $d_x$	$\lambda/2$
IRS <sub>1</sub> and IRS <sub>2</sub> Element Spacing $d_y$	$\lambda/2$
Antenna diameter in meters $D_t$	0.12
Coordinates of Rx ( $x_{rx}, y_{rx}, h_r$ )	(20,0,5)
Bandwidth	$2 \times 10^9 \text{ MHz}$
Noise power spectral density $N_{PSD}$	$-174 \text{ dB/Hz}$
Noise figure at the receiver $F_{dB}$	10
Average Noise power in dB $N_0$	$-174 \text{ dB/Hz}$
Noise power in linear scale $n_0$	$7.9621 \times 10^{-11}$
Transmitters to IRS <sub>1</sub> Path loss exponent	2
IRS <sub>2</sub> to receiver $R_x$ Path loss exponent	2
Rician Factor	10
Critic Network learning rate	$3 \times 10^{-4}$
Actor Network learning rate	$1 \times 10^{-4}$
Target Critic Network learning rate	$3 \times 10^{-4}$
Target Actor Network learning rate	$1 \times 10^{-4}$
Discount factor of the future reward $\Gamma$	0.99
Coefficient of Soft Updates $\tau$	$1 \times 10^{-3}$
Batch size	128
Replay Buffer Capacity $\mathcal{C}$	$10^5$
Number of episodes	10000

and the heights of user 1 and user 2 were both  $h_t = 5$ . We defined the distance ratio (DR) as  $r_{t1}$ , the distance from user 1 to IRS<sub>1</sub>, divided by  $r_{t2}$ , the distance from user 2 to IRS<sub>1</sub>. Our results were obtained through  $10^3$  Monte-Carlo simulations.

The suggested DDPG scheme involves two NNs: the actor and critic networks. Both of these networks are constructed as dense NNs, each comprising four layers. In these networks, each layer is a linear module with two parameters: the input size and the output size. For the actor network, the input consists of the states, with a size of 3 neurons, while the output represents the action and comprises 36 neurons. Between the input and output layers, there are two hidden layers, each with 128 neurons. These hidden layers employ the Rectified Linear Unit (ReLU) activation function. To ensure sufficient gradient information, the output layer of the actor network uses the  $\tanh(\cdot)$  activation function. As for the critic network, the input includes both the count of states and actions, which are concatenated to form the input of the critic network. Two hidden layers exist between the input layer and the output layer, with 128 neurons in each hidden layer. Again, ReLU activation functions are applied in the hidden layers. The output layer of the critic network produces the Q-value and consists of 36 neurons. To update the network parameters, both the actor and critic networks use the Adam optimizer. The results are obtained by considering the average rate over 1000 iterations. The actor network is configured with a learning rate of  $3 \times 10^{-4}$ , while the critic network uses a learning rate of  $1 \times 10^{-4}$ . The discount factor for future rewards  $\Gamma$  is 0.99, and the batch size is set at 128. The replay buffer (denoted as  $\mathcal{C}$ ) has a capacity of  $10^5$ , and the number of episodes for

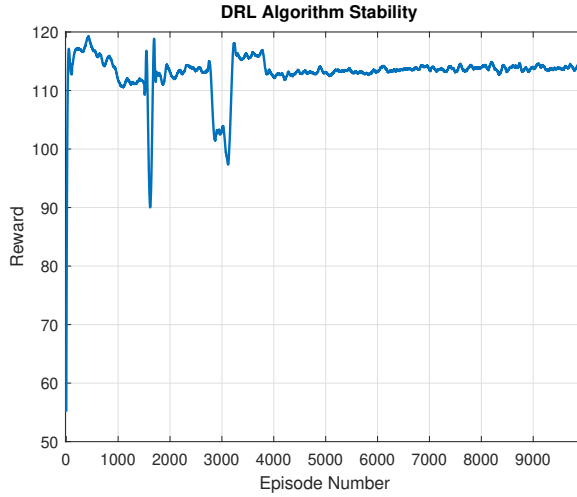


Fig. 5: DDPG algorithm convergence.

training the model is set to 10,000.

### B. DDPG Convergence

The results depicted in Fig. 5 demonstrate the convergence of the DDPG scheme. In the figure, rewards are plotted against episodes, and it's evident that rewards increase over time, indicating a successful learning process. The DDPG algorithm learning time requires 11.53 hours for 10,000 episodes for each distance ratio value when running a single program at a time. Once trained, the implementation time for 1000 iterations is 16.421875 seconds for the DDPG, 30.6875 seconds for the Block Solution, and 947.546875 seconds for the Pinv solution. We conducted these computations on the HP EliteBook x360 830 G7 Notebook PC, featuring an Intel(R) Core(TM) i7-10510U CPU running at 1.80GHz (2.30 GHz) and 32.0 GB of RAM.

### C. Maximizing the Rate of a Desired User under Interference

#### 1) Distance Ratio Impact:

In the scenarios aiming to maximize the rate for user 1 at the  $R_x$ , we consider different correlation factors ( $\rho$ ) and DRs between user 1 and user 2. We employ various methods, including DDPG,  $BLS$ , and  $P_{inv}$  solution, to showcase the performance differences among them. Figures 6, 7, and 8 depict the rate of user 1 achieved with the DDPG scheme across DRs ranging from 0.2 to 1. As the DR increases, user 1's rate decreases due to the increased interference from user 2, who is getting closer to user 1.

#### 2) Correlation-Factor Impact:

Fig. 6, Fig. 7, and Fig. 8 illustrate the behavior of user 1's data rate concerning  $\rho$ . As  $\rho$  decreases, the data rates for user 1 decrease for all methods, including DDPG,  $BLS$ , and  $P_{inv}$ . Conversely, when  $\rho$  increases, the data rates achieved by the DDPG technique increase as well. Thus, the enhancement in performance achieved by the DDPG algorithm over the  $BLS$  ranges from 8% to 57.12%, while its improvement over

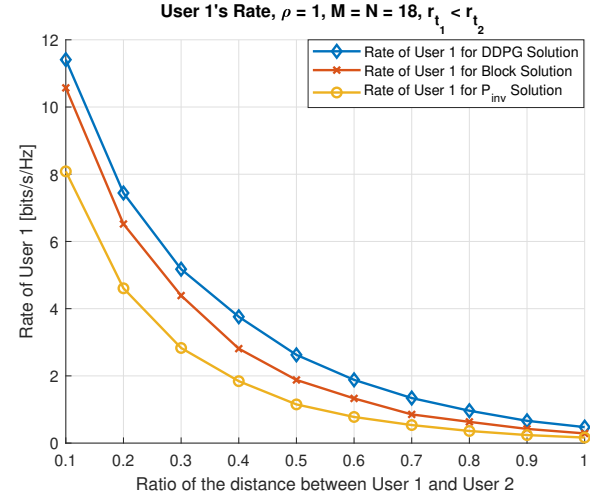


Fig. 6: Correlation-factor  $\rho$  equal to 1.0.

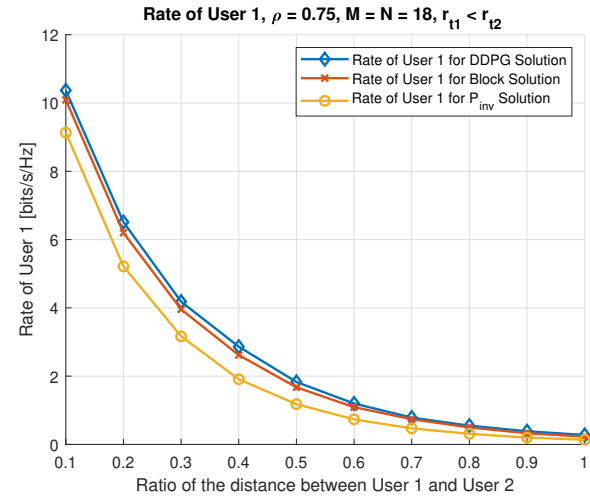


Fig. 7: Correlation-factor  $\rho$  equal to 0.75.

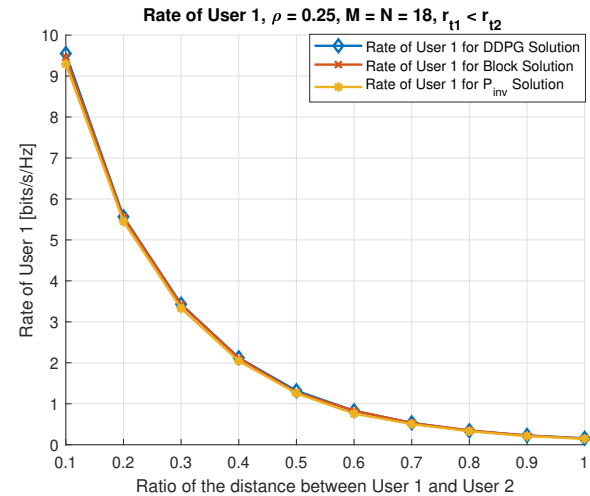


Fig. 8: Correlation-factor  $\rho$  equal to 0.25.

the  $P_{inv}$  spans from 41% to 190% for  $\rho = 1$ . This is due to the DDPG algorithm's improved learning efficiency with higher correlation values. Consequently, the DDPG scheme outperforms other methods, showcasing superior performance in scenarios with high correlation. This aligns with our observations, indicating that as channel correlation increases, the performance of DRL exhibits improvement, effectively leveraging the correlation for enhanced learning. These results underscore the significance of correlated channels in our scenario and their positive impact on data rates. It's important to note, however, that for low values of  $\rho$ , the gap in data rates between DDPG and other methods retracts. Thus, for  $\rho = 0.75$ , the enhancement in performance achieved by the DDPG algorithm over  $BLS$  ranges from 2.75% to 21.6%, while its improvement over  $P_{inv}$  spans from 13.5% to 102.2%. Whereas for  $\rho = 0.25$ , the DDPG algorithm demonstrates a performance improvement over  $BLS$  ranging from 0.25% to 3%, and over  $P_{inv}$ , it spans from 2.1% to 9.7%.

#### D. Maximizing the Sum Rate for both Users

In the scenario of maximizing the sum rate for both users at the  $R_x$ , Fig.9 displays the sum rates as a function of the DR between user 1 and user 2, using the DDPG method with various correlation coefficients  $\rho$ .

##### 1) Learning Rate Impact:

Notably, the results show that as  $\rho$  increases, the sum rates achieved with the DDPG solution also increase. It's important to highlight that constant learning rates were maintained in all our simulations for the DDPG scheme, with a rate of  $10^{-4}$  for the actor network and  $3 \times 10^{-4}$  for the critic network. The impact of these learning rates on the DDPG data rates is demonstrated in Fig.10, where we compare different learning rates, including  $10^{-3}$ ,  $10^{-4}$ , and  $10^{-5}$  for the actor network, and  $3 \times 10^{-3}$ ,  $3 \times 10^{-4}$ , and  $3 \times 10^{-5}$  for the critic network. The results show that the highest DDPG data rate is achieved when the learning rate for the actor networks is  $10^{-4}$ , and for the critic networks, it's  $3 \times 10^{-4}$ . Therefore, these learning rates produce the best average rewards, while too small ( $10^{-5}$ ) or too large ( $10^{-3}$ ) learning rates result in lower average rewards. The optimal learning rate of  $10^{-4}$  is a key factor for achieving better rewards in the DDPG scheme.

##### 2) DDPG vs Exhaustive Search:

Moreover, to assess the effectiveness of our DDPG scheme, we conducted a comparison between the sum rates generated by the DDPG algorithm and those obtained through a discretized  $ES$  approach used to find the maximum sum rate by determining the optimal phase shift matrix. The  $ES$  method, while highly accurate, involves significant computational complexity. To mitigate this, we reduced the number of reflecting elements for  $IRS_1$  to  $M = 4$  and  $IRS_2$  to  $N = 4$ , instead of the original  $M = N = 18$ . For each IRS element, we considered phase shifts ranging from 0 to  $2\pi$  with a search step size of  $\frac{2\pi}{72}$ . Consequently, the total number of phase shift matrix combinations amounts to  $(72 + 1)^4$ . The sum rates were computed for two users across 100 Monte-Carlo simulations. The results are presented in Fig. 11, where it's evident that the DDPG algorithm's sum

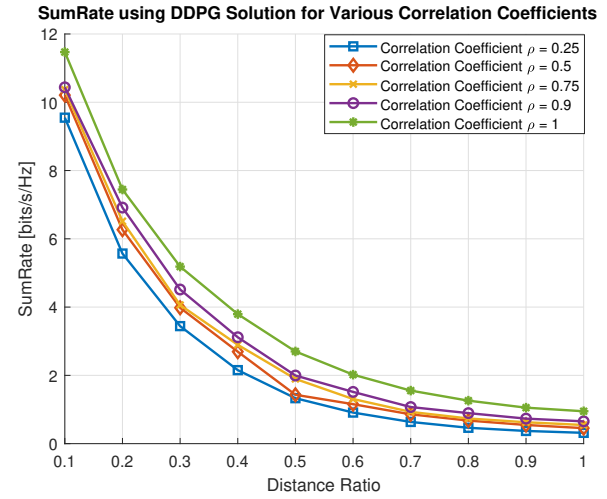


Fig. 9: DDPG sum rate vs distance ratio for various correlation-coefficients  $\rho$ .  $M=N=18$ .

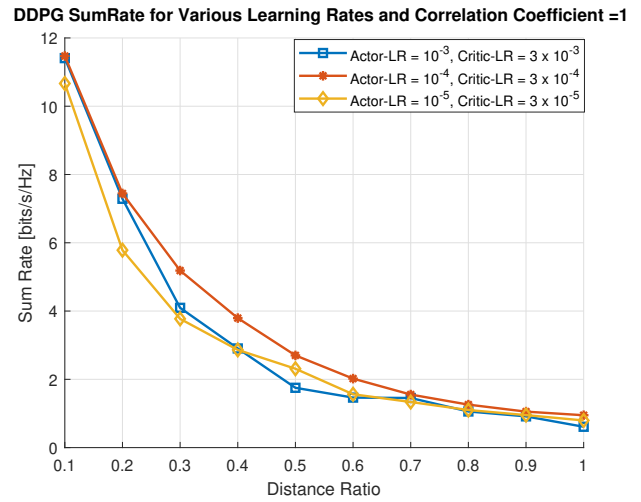


Fig. 10: DDPG sum rate vs distance ratio for various learning rates and correlation-coefficient  $\rho = 1.0$ .  $M = N = 18$ .

rates closely align with the  $ES$  outcomes with the specified granularity.

## VIII. CONCLUSION

In this research paper, we focused on the uplink multiple access scenario within a cascaded IRS system, to address the challenges associated with short-range communications in THz networks. Our primary goals were twofold. The first objective was to maximize the data rate for the intended user. We determined that this problem was not amenable to closed-form mathematical solutions due to its non-convex nature. As a result, we proposed two sub-optimal approaches to enhance the received power for the desired user. The second objective was to maximize the total data rate for both users, which presented a more intricate, non-convex problem. To tackle these optimization challenges, we employed DDPG algorithms, known for their effectiveness in handling non-



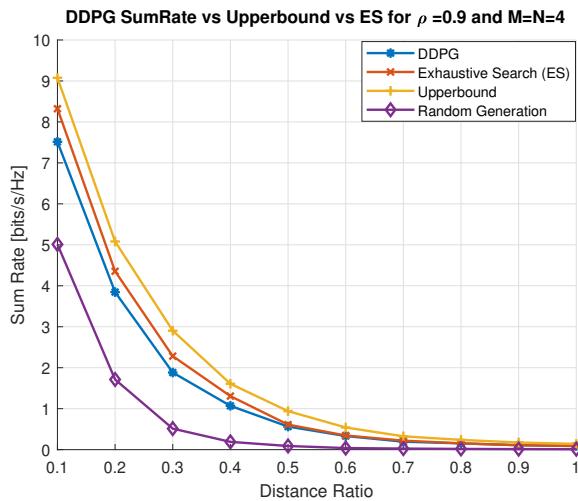


Fig. 11: Comparison between DDPG sum rate,  $ES$ , and upper bound vs distance ratio for correlation-coefficient  $\rho = 0.9$ .  $M = N = 4$ .

convex optimization problems, to optimize the cascaded IRS system. The DDPG algorithm allowed us to approximate optimal IRS phase configurations that boosted the received data rate for the intended user as well as the overall data rate for both users.

Our simulation results demonstrated that, for the first objective, DDPG consistently outperformed sub-optimal methods such as  $P_{inv}$  and  $BLS$ , achieving higher data rates. Regarding the second objective, DDPG's achieved data rates closely matched those of a discretized  $ES$ , with a search step equal to  $2\pi/72$ . Furthermore, DDPG highlighted the importance of channel correlation in improving the learning process and achieving enhanced data rates. In summary, our paper presents a pioneering use of DDPG for optimizing cascaded IRS phase shifts, addressing an unexplored research area, and achieving superior performance.

## REFERENCES

- [1] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, P. Popovski and M. Debbah, "Seven Defining Features of Terahertz (THz) Wireless Systems: A Fellowship of Communication and Sensing," in IEEE Communications Surveys and Tutorials, DOI: 10.1109/COMST.2022.3143454.
- [2] R. Imran, M. Odeh, N. Zorba and C. Verikoukis, "Quality of Experience for Spatial Cognitive Systems within Multiple Antenna Scenarios," in IEEE Transactions on Wireless Communications, vol. 12, no. 8, pp. 4153-4161, August 2013, doi: 10.1109/TWC.2013.071113.122037.
- [3] I. Yildirim, A. Uyrus and E. Basar, "Modeling and Analysis of Reconfigurable Intelligent Surfaces for Indoor and Outdoor Applications in Future Wireless Networks," in IEEE Transactions on Communications, vol. 69, no. 2, pp. 1290-1301, Feb. 2021, doi: 10.1109/TCOMM.2020.3035391.
- [4] Z. Chen, X. Ma, C. Han and Q. Wen, "Towards intelligent reflecting surface empowered 6G terahertz communications: A survey," in China Communications, vol. 18, no. 5, pp. 93-119, May 2021, DOI: 10.23919/JCC.2021.05.007.
- [5] T. V. Nguyen, T. P. Truong, T. M. T. Nguyen, W. Noh, and S. Cho, "Achievable Rate Analysis of Two-Hop Interference Channel with Coordinated IRS Relay," in IEEE Transactions on Wireless Communications, DOI: 10.1109/TWC.2022.3154372.
- [6] W. Mei and R. Zhang, "Multi-Beam Multi-Hop Routing for Intelligent Reflecting Surfaces Aided Massive MIMO," in IEEE Transactions on Wireless Communications, vol. 21, no. 3, pp. 1897-1912, March 2022, DOI: 10.1109/TWC.2021.3108020.

- [7] Q. Sun, P. Qian, W. Duan, J. Zhang, J. Wang and K. -K. Wong, "Ergodic Rate Analysis and IRS Configuration for Multi-IRS Dual-Hop DF Relaying Systems," in IEEE Communications Letters, vol. 25, no. 10, pp. 3224-3228, Oct. 2021, DOI: 10.1109/LCOMM.2021.3100347.
- [8] Z. Zhang and Z. Zhao, "Weighted Sum-Rate Maximization for Multi-Hop RIS-Aided Multi-User Communications: A Minorization-Maximization Approach," 2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2021, pp. 106-110, DOI: 10.1109/SPAWC51858.2021.9593114.
- [9] A. Almohamad, M. Hasna, N. Zorba, and T. Khattab, "Performance of THz Communications Over Cascaded RISs: A Practical Solution to the Over-Determined Formulation," in IEEE Communications Letters, vol. 26, no. 2, pp. 291-295, Feb. 2022, DOI: 10.1109/LCOMM.2021.3132655.
- [10] C. Huang et al., "Hybrid Beamforming for RIS-Empowered Multi-hop Terahertz Communications: A DRL-based Method," 2020 IEEE Globecom Workshops (GC Wkshps, Taipei, Taiwan, 2020, pp. 1-6, doi: 10.1109/GCWkshps50303.2020.9367503.
- [11] C. Huang et al., "Multi-Hop RIS-Empowered Terahertz Communications: A DRL-Based Hybrid Beamforming Design," in IEEE Journal on Selected Areas in Communications, vol. 39, no. 6, pp. 1663-1677, June 2021, DOI: 10.1109/JSAC.2021.3071836.
- [12] C. Soni and N. Gupta, "Channel Estimation of Spatial Correlated Channel in Massive MIMO," 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2021, pp. 836-841.
- [13] K. Feng, Q. Wang, X. Li and C. -K. Wen, "Deep Reinforcement Learning Based Intelligent Reflecting Surface Optimization for MISO Communication Systems," in IEEE Wireless Communications Letters, vol. 9, no. 5, pp. 745-749, May 2020, doi: 10.1109/LWC.2020.2969167.
- [14] K. Ntontin et al., "Reconfigurable Intelligent Surface Optimal Placement in Millimeter-Wave Communications," 2021 15th European Conference on Antennas and Propagation (EuCAP), Dusseldorf, Germany, 2021, pp. 1-5, doi: 10.23919/EuCAP51087.2021.9411076.
- [15] K. Feng, Q. Wang, X. Li, and C. Wen, "Deep Reinforcement Learning Based Intelligent Reflecting Surface Optimization for MISO Communication Systems," in IEEE Wireless Communications Letters, vol. 9, no. 5, pp. 745-749, May 2020, DOI: 10.1109/LWC.2020.2969167.
- [16] J. Kokkonen, J. Lehtomäki and M. Juntti, "Simplified molecular absorption loss model for 275-400 gigahertz frequency band," 12th European Conference on Antennas and Propagation (EuCAP 2018), 2018, pp. 1-5, DOI: 10.1049/cp.2018.0446.
- [17] M. Shehab, B. S. Ciftler, T. Khattab, M. M. Abdallah and D. Trinchero, "Deep Reinforcement Learning Powered IRS-Assisted Downlink NOMA," in IEEE Open Journal of the Communications Society, vol. 3, pp. 729-739, 2022, doi: 10.1109/OJCOMS.2022.3165590.
- [18] C. Huang, R. Mo and C. Yuen, "Reconfigurable Intelligent Surface Assisted Multiuser MISO Systems Exploiting Deep Reinforcement Learning," in IEEE Journal on Selected Areas in Communications, vol. 38, no. 8, pp. 1839-1850, Aug. 2020, doi: 10.1109/JSAC.2020.3000835.
- [19] V. François-Lavet, P. Henderson, R. Islam, M. Bellemare, and J. Pineau (2018), "An Introduction to Deep Reinforcement Learning", Foundations and Trends in Machine Learning: Vol. 11, No. 3-4, 2018.
- [20] G. E. Uhlenbeck and L. S. Ornstein, "On the Theory of the Brownian Motion", Revista Latinoamericana De Microbiologia, 1930.
- [21] M. Elsayed, A. Badawy, A. E. Shafie, A. Mohamed and T. Khattab, "A Deep Reinforcement Learning Framework for Data Compression in Uplink NOMA-SWIPT Systems," in IEEE Internet of Things Journal, vol. 9, no. 14, pp. 11656-11674, 15 July 15, 2022, doi: 10.1109/JIOT.2021.3131524.
- [22] Keller-Gehrig, Walter: Fast algorithms for the characteristic polynomial. Theoretical Computer Science, 36(2-3):309-317, 1985, ISSN 0304-3975. [http://dx.doi.org/10.1016/0304-3975\(85\)90049-0](http://dx.doi.org/10.1016/0304-3975(85)90049-0). x.doi.org/10.1145/345542.345644.
- [23] V. Vasudevan, M. Ramakrishna, "A Hierarchical Singular Value Decomposition Algorithm for Low Rank Matrices,". 2017-10-08 — Preprint. ARXIV: arXiv:1710.02812v2.

## APPENDIX

The total received signal power for each  $T_{x_k}$  at the  $R_x$  in eq. (23) can be expressed as:

$$P_{Rx}^{(k)} = |\sqrt{L_{\tau,k}} e^{-j\Omega_3} \mathbf{h}_r^H \Phi_N \mathbf{H}_{m,n}^H \Phi_M \mathbf{h}_{t,k}^H e^{-j\Omega_k}|^2 P_t. \quad (41)$$

*Proof.* • Multiply the  $R_x$  channel  $\mathbf{h}_r^H$  by IRS<sub>2</sub> phase shift reflection matrix  $\Phi_N$ :

$$[\mathbf{h}_r^H \Phi_N] = \underbrace{[h_{r1}^*, \dots, h_{rn}^*, \dots, h_{rN}^*]}_{1 \times N} \times \underbrace{\begin{bmatrix} e^{-j\psi_1} & 0 & \dots & \dots & \dots & 0 \\ 0 & e^{-j\psi_2} & 0 & \ddots & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & e^{-j\psi_n} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 & e^{-j\psi_N} \end{bmatrix}}_{N \times N} \quad (42)$$

$$[\mathbf{h}_r^H \Phi_N] = \underbrace{[h_{r1}^* e^{-j\psi_1}, \dots, h_{rn}^* e^{-j\psi_n}, \dots, h_{rN}^* e^{-j\psi_N}]}_{1 \times N} \quad (43)$$

• Multiply the result of the previous operation by  $\mathbf{H}_{m,n}^H$ :

$$[\mathbf{h}_r^H \Phi_N \mathbf{H}_{mn}^H] = \underbrace{[h_{r1}^* e^{-j\psi_1}, \dots, h_{rn}^* e^{-j\psi_n}, \dots, h_{rN}^* e^{-j\psi_N}]}_{1 \times N} \times \underbrace{\begin{bmatrix} h_{11}^* & h_{21}^* & \dots & \dots & \dots & h_{M1}^* \\ h_{12}^* & h_{22}^* & \ddots & \ddots & \ddots & h_{M2}^* \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & H_{mn}^* & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ h_{1N}^* & \dots & \dots & \dots & \dots & h_{MN}^* \end{bmatrix}}_{N \times M} \quad (44)$$

$$[\mathbf{h}_r^H \Phi_N \mathbf{H}_{mn}^H] = \underbrace{\left[ \sum_{n=1}^N h_{rn}^* e^{-j\psi_n} h_{1n}^*, \sum_{n=1}^N h_{rn}^* e^{-j\psi_n} h_{2n}^*, \dots, \sum_{n=1}^N h_{rn}^* e^{-j\psi_n} h_{MN}^* \right]}_{1 \times M} \quad (45)$$

• Multiply the result of the previous operation by IRS<sub>1</sub> phase shift reflection matrix  $\Phi_M$ :

$$[\mathbf{h}_r^H \Phi_N \mathbf{H}_{mn}^H \Phi_M] = \underbrace{[\mathbf{h}_r^H \Phi_N \mathbf{H}_{mn}^H]}_{1 \times M} \times \underbrace{\begin{bmatrix} e^{-j\eta_1} & 0 & \dots & \dots & \dots & 0 \\ 0 & e^{-j\eta_2} & 0 & \ddots & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & e^{-j\eta_m} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 & e^{-j\eta_M} \end{bmatrix}}_{M \times M} \quad (46)$$

$$[\mathbf{h}_r^H \Phi_N \mathbf{H}_{mn}^H \Phi_M] =$$

$$\underbrace{\left[ \sum_{n=1}^N h_{rn}^* e^{-j\psi_n} h_{1n}^*, \sum_{n=1}^N h_{rn}^* e^{-j\psi_n} h_{2n}^*, \dots, \sum_{n=1}^N h_{rn}^* e^{-j\psi_n} H_{MN}^* \right]}_{1 \times M} \times \underbrace{\begin{bmatrix} e^{-j\eta_1} & 0 & \dots & \dots & \dots & 0 \\ 0 & e^{-j\eta_2} & 0 & \ddots & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & e^{-j\eta_m} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 & e^{-j\eta_M} \end{bmatrix}}_{M \times M} \quad (47)$$



$$[\mathbf{h}_r^H \Phi_N \mathbf{H}_{mn}^H \Phi_M] = \underbrace{\left[ \sum_{m=1}^M \sum_{n=1}^N h_{rn}^* e^{-j\psi_n} H_{mn}^* e^{-j\eta_m} \right]}_{1 \times M} \quad (48)$$

- Multiply the result of the previous operation by the  $T_x$  channel of user  $K$   $\mathbf{h}_{t,k}^H$ :

$$[\mathbf{h}_r^H \Phi_N \mathbf{H}_{mn}^H \Phi_M \mathbf{h}_{t,k}^H] = \underbrace{\left[ \sum_{m=1}^M \sum_{n=1}^N h_{rn}^* e^{-j\psi_n} H_{mn}^* e^{-j\eta_m} \right]}_{1 \times M} \times \underbrace{\begin{bmatrix} h_{t,k,1}^* \\ h_{t,k,2}^* \\ \vdots \\ h_{t,k,M}^* \end{bmatrix}}_{M \times 1} \quad (49)$$

$$\mathbf{h}_r^H \Phi_N \mathbf{H}_{mn}^H \Phi_M \mathbf{h}_{t,k}^H = \sum_{m=1}^M \sum_{n=1}^N h_{rn}^* e^{-j\psi_n} H_{mn}^* e^{-j\eta_m} h_{t,k,m}^* \quad (50)$$

$$\mathbf{h}_r^H \Phi_N \mathbf{H}_{m,n}^H \Phi_M \mathbf{h}_{t,k}^H = \sum_{m=1}^M \sum_{n=1}^N |h_{rn}| e^{-j\phi_{rn}} e^{-j\psi_n} |H_{mn}| e^{-j\phi_{mn}} e^{-j\eta_m} |h_{t,k,m}| e^{-j\phi_{t,k,m}} \quad (51)$$

$$\mathbf{h}_r^H \Phi_N \mathbf{H}_{mn}^H \Phi_M \mathbf{h}_{t,k}^H = \sum_{m=1}^M \sum_{n=1}^N |h_{t,k,m}| |H_{mn}| |h_{rn}| e^{-j(\phi_{t,k,m} + \eta_m + \phi_{mn} + \psi_n + \phi_{rn})} \quad (52)$$

- From the obtained result we can deduce that the total received signal power for each  $T_{x_k}$  at the  $R_x$  can be written as follows:

$$\begin{aligned} P_{Rx}^{(k)} &= \left| \sqrt{L_{\tau,k}} e^{-j\Omega_3} \mathbf{h}_r^H \Phi_N \mathbf{H}_{mn}^H \Phi_M \mathbf{h}_{t,k}^H e^{-j\Omega_k} \right|^2 P_t, \\ &= \left| \sqrt{L_{\tau,k}} \sum_{m=1}^M \sum_{n=1}^N |h_{t,k,m}| |H_{mn}| |h_{rn}| e^{-j(\varphi_{t,k,m} + \eta_m + \varphi_{mn} + \psi_n + \varphi_{rn} + \Omega_k + \Omega_3)} \right|^2 P_t \end{aligned} \quad (53)$$

□