DDPG Performance in THz Communications over Cascaded RISs: A Machine Learning Solution to the Over-Determined System

Original
DDPG Performance in THz Communications over Cascaded RISs: A Machine Learning Solution to the Over-Determined System / Shehab, Muhammad; Badawy, Ahmed; Elsayed, Mohamed; Khattab, Tamer; Trinchero, Daniele. - ELETTRONICO. - (2023), pp. 210-215. (Intervento presentato al convegno 2023 International Wireless Communications and Mobile Computing (IWCMC) tenutosi a Marrakesh, Morocco nel 19-23 June 2023) [10.1109/IWCMC58020.2023.10182861].

Availability:
This version is available at: 11583/2980644 since: 2023-10-06T09:23:20Z

Publisher:
IEEE

Published
DOI:10.1109/IWCMC58020.2023.10182861

(Article begins on next page)

DDPG Performance in THz Communications over Cascaded RISs: A Machine Learning Solution to the Over-Determined System

Muhammad Shehab*, Ahmed Badawy†, Mohamed Elsayed‡, Tamer Khattab‡, and Daniele Trinchero*
*Politecnico di Torino (DET), (MuhammadShehab@ieee.org, daniele.trinchero@polito.it)
†Computer Science and Engineering, Qatar University, badawy@qu.edu.qa
‡Electrical Engineering, Qatar University, hamid@qu.edu.qa, tkhattab@ieee.org

***Abstract*—THz technology is considered a key element in 6G wireless communication because it provides ultra-high bandwidths, considerable capacities, and significant gains. However, wireless systems operating at high frequencies are faced with uncertainty and highly dynamic channels. Reflecting intelligent surfaces (RISs) can increase the range of the THz communication links and boost the rate at the receiver. In contrast to the existing literature, we investigate the scenario of multiple access multi-hop (cascaded) RISs uplink THz networks in a correlated channel environment. We show that our inspected cascaded RIS system is over-determined and that the rate maximization optimization problem is non-convex. To this end, we derive a closed-form expression of the received power and derive an analytical solution based on pseudo-inverse to obtain optimum RISs' phase shifts that maximize the received signal power and hence increase the rate. In addition, we utilize deep reinforcement learning (DRL), which is capable of solving non-convex optimization problems, to obtain the optimum cascaded RISs' phase shifts at the receiver taking into account the situation of the spatially correlated channels. Simulation results demonstrate that the DRL algorithm achieves higher rates than the mathematical sub-optimal method and the case of randomized phases.**

***Index Terms*—Wireless Communication, 5G, 6G, RIS, THz, Machine Learning, Deep Reinforcement learning.**

I. INTRODUCTION

THZ frequency bands (100 GHz - 10 THz) are considered cornerstones in the 6G communication networks. THz frequencies are favorable to support ultra-high bandwidths and significant data rates. These frequency bands can potentially provide considerable performance gains and significant capacities. Nonetheless, the transition towards the real and practical implementation of THz networks suffers from molecular losses, highly dynamic and varying channels, short-range links and communication distances, and the reliance on line-of-sight (LOS) or narrow-beam links [1], [2]. To optimize the achievable data rate at the receiver (R_x), this research paper examines the reflecting intelligent surface (RIS) as a modern technology and promising solution. The RIS is a two-dimensional (2D) electromagnetic surface, precisely metasurface, that constitutes a large number of semi-passive scattering elements. Every element can be controlled via a software-defined behavior to adjust the electromagnetic properties (i.e. phase-shift) of the reflection of the incident radio frequency (RF) signals upon the RIS elements [3], [4]. Thus, the RIS can instantly amend the wireless propagation channel to improve the signal transmission, boost the received signal power, and suppress the interference at the R_x. Therefore, it improves the data rate in a cost-effective and energy-efficient behavior and provides an innovative means to attain the 6G Key performance indicators (KPIs) [5], [6].

Several research papers inspected the deployment of RIS in THz networks to investigate its power in boosting the coverage and improving the achievable data rate at the R_x [7]–[16]. However, the following research studies [7]–[14] utilized mathematical methods to solve their optimization problems, whereas the research papers in [15], [16] employed the deep reinforcement learning (DRL) method to solve the joint design of the digital beamforming at the base station (BS) and the analog beamforming at the RISs to combat the propagation attenuations and molecular absorptions in downlink broadcast THz system, which is a Base station (i.e. single source) to multi-destination (i.e. users) scenario. To the best of our knowledge, none of the studies in the literature utilized the DRL technique to determine the solution to the over-determined system of equations for the scenario of multiple access multi-hop RIS uplink THz system, which is multiple sources(i.e. users) to the same destination (i.e. R_x).

In this study, we examine the above-mentioned gap in the literature by employing a DRL method, namely DDPG (deep deterministic policy gradient (DDPG), to jointly obtain the optimum RISs phase-shifts in the multi-hop RIS scenario to maximize the received power for user 1 at the R_x. The main challenge in our design while finding the optimum phases at RIS_1 and RIS_2 jointly lies in the non-convexity because of the constant modulus constraints of the RIS elements, and computationally intractable cascaded RIS links. Thus, the optimal solution to this optimization problem is not known, and it is not feasible to find an analytical solution using mathematical techniques. Further, employing the exhaustive search isn't practical for large-scale networks because of

M. Shehab and D. Trinchero are with the Dipartimento di Elettronica, Politecnico di Torino, Torino, Italy. M. Elsayed, and T. Khattab, are with the department of Electrical Engineering, Qatar University, Doha, Qatar. A. Badawy is with the department of Computer Science and Engineering, Qatar University, Doha, Qatar. This research work was made possible by grant no. AICC03-0530-200033 from Qatar National Research Fund (QNRF). The statements made herein are the sole responsibility of the authors.

it is high complexity [15], [16]. Therefore, to solve this optimization problem, we leverage a DRL method, namely the DDPG algorithm, to find out the possible solutions.

A. Contributions

Our contributions to this research paper can be summarized as follows:

- We formulate the multi-hop RIS correlated channel system that operates in the THz frequency range under the scenario of two transmitting users.
- We formulate user 1's rate optimization problem by jointly optimizing the phase shifts at RIS_1 and RIS_2 while the second user is considered an interferer and we show that the problem is non-convex.
- We derive a closed-form expression of the received power of the first user under the cascaded RIS correlated channel scenario and we show that the system is overdetermined.
- We find a suboptimal solution to the over-determined system to find cascaded RIS phases that maximize the received power.
- We leverage DDPG to solve the optimization problem which is non-convex and computationally intractable. Moreover, we reveal the performance of DDPG by solving the same problem with a sub-optimal mathematical method such as pseudo-inverse and comparing the performance of DDPG with the mathematical technique.
- We simulate our suboptimal and DDPG solutions and demonstrate that the DDPG algorithm is superior to the sub-optimal mathematical method, and the random generation case.

II. Communication System Model

In our system model, we consider a static topology in Fig. 1 where there are two users, transmitter 1 (T_{x_1}) and transmitter 2 (T_{x_2}) communicating with the R_x, via two multi-hop RISs. Both users and the R_x are furnished with high parabolic directional antennas, and they are transmitting to the center of the RIS_1, with diameter D_t for the two transmitters and D_r for the R_x. The distances between user 1 and RIS_1, user 2 and RIS_1, RIS_1 and RIS_2, and RIS_1 and R_x are represented as r_{11}, r_{12}, r_2, and r_3 respectively, whereas the horizontal distances between user 1 and RIS_1, user 2 and RIS_1, RIS_1 and RIS_2, and RIS_2 and R_x are denoted as $r_{11,h}$, $r_{21,h}$, $r_{2,h}$, and $r_{3,h}$ respectively. The heights of the two users T_{x_1}, and T_{x_2}, RIS_1, RIS_2, and R_x are represented as $h_{T_{x_1}}$, $h_{T_{x_2}}$, h_{s_1}, h_{s_2}, and h_r respectively. The number of elements for RIS_1 and RIS_2 is denoted as $V = V_x \times V_y$ and $W = W_x \times W_y$ elements respectively. Every RIS is furnished with a controller that determines the perfect knowledge of the channel state information (CSI). The channel model between each transmitter u and RIS_1, and between RIS_2 and the R_x, follow the rician fading model which is represented as:

$$\mathbf{h_{t,u}} = \sqrt{\frac{K1}{K1+1}}\bar{\mathbf{h}}_{\mathbf{t,u}} + \sqrt{\frac{1}{K1+1}}\tilde{\mathbf{h}}_{\mathbf{t,u}}, \tag{1}$$

$$\mathbf{h_r} = \sqrt{\frac{K2}{K2+1}}\bar{\mathbf{h}}_{\mathbf{r}} + \sqrt{\frac{1}{K2+1}}\tilde{\mathbf{h}}_{\mathbf{r}}, \tag{2}$$

where $\mathbf{h_{t,u}}$ is the channel between each user u and RIS_1, $K1$ is the rician factor of $\mathbf{h_{t,u}}$. Further, the LOS component for the channel $\mathbf{h_{t,u}}$ is represented as $\bar{\mathbf{h}}_{\mathbf{t,u}} \in C^{1\times V}$, and the non-LOS (NLOS) component is represented as $\tilde{\mathbf{h}}_{\mathbf{t,u}} \in C^{1\times V}$. Similarly, for the R_x channel $\mathbf{h_r}$, K_2 is the rician factor, the LOS component is represented as $\bar{\mathbf{h}}_{\mathbf{r}} \in C^{W\times 1}$, and the NLOS component is represented as $\tilde{\mathbf{h}}_{\mathbf{r}} \in C^{W\times 1}$.

Moreover, the channel between RIS_1 and RIS_2, $\mathbf{h}_{v,w} \sim \mathcal{CN}(0, \mathbf{R})$, is spatially correlated and it follows the Rayleigh fading model, where $\mathbf{R}$ represents the covariance matrix, which is obtained based on the model of the exponential spatial correlation. This is controlled by the correlation factor among the adjacent reflecting units $\rho \in [0, 1]$, which is represented as:

$$[\mathbf{R}]_{v,w} = \rho^{|v-w|}e^{|v-w|\xi} \tag{3}$$

where ξ denotes the angle of arrival. Increasing values of ρ enhances the correlation among the elements of the channel $\mathbf{h_{w,v}}$. In practical cases, the value of the correlation factor ρ is less than 1, and thus, the correlations happen among the adjacent elements with considerably reduced correlation at huge distances.

III. End-To-End Rate Derivation

In our scenario, the two users are sending their signals at RIS_1 from different distances and angles covering all the elements of RIS_1. The reflected power from the v_{th} reflecting element of the at RIS_1 can be expressed as [17].

$$P^u_{r,v} = \left(\frac{\lambda}{4\pi}\right)^2 \frac{G_{t,u}G(\theta_{i1,u})G(\theta_{r,1})}{r^2_{tu}} \times |h_{t,uv}|^2|\gamma_v|^2P_t, \tag{4}$$

where $G_{t,u}$ represents the Tx antenna gain of transmitter T_{x_u}; $G(\theta_{i1,u})$ and $G(\theta_{r,1})$ denote the gain of the reflecting element from the incident and reflection angles respectively; $\gamma_v = \gamma e^{-j\alpha_v}$ symbolizes the reflection coefficient of the v^{th}reflecting element of RIS_1; and $r_{t,u}$ denotes the distance between user u and the RIS_1 reflecting element v. Similarly, the reflected power from the w^{th} reflecting element of RIS_2 due to being illuminated by the signal reflected by the v^{th} reflecting element of RIS_1 is

$$P^u_{r,vw} = \left(\frac{\lambda}{4\pi}\right)^4 \frac{G_{t,u}G(\theta_{i1,u})G(\theta_{r,1})G(\theta_{i,2})G(\theta_{r,2})}{r^2_{tu}r^2_2} \times|h_{t,uv}|^2|\gamma_v|^2|h_{vw}|^2|\gamma_w|^2P_t, \tag{5}$$

where h_{vw} represents the (v, w) element of the RIS_1-RIS_2 channel matrix $\mathbf{H}$, $\gamma_w = \gamma e^{-j\varphi_w}$ symbolizes the reflection coefficient of the w^{th} reflecting element RIS_2. Finally, the

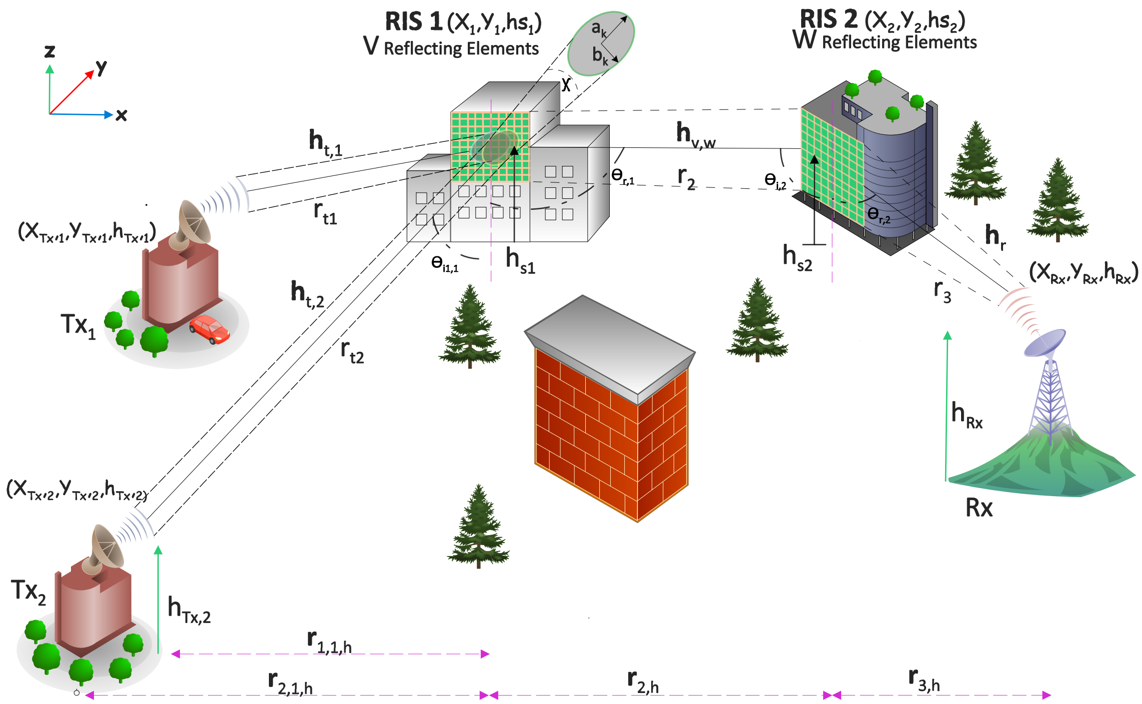

Figure 1: Communication System Model.

received captured power at the R_x through channel h_{vw} can be written as follows:

$$P^u_{rx,vw} = \left(\frac{\lambda}{4\pi}\right)^2 \frac{P^u_{r,vw}}{r_3^2} G_r |h_{rw}|^2. \quad (6)$$

$$P^u_{rx,vw} = \left(\frac{\lambda}{4\pi}\right)^6 \frac{G_{t,u} G(\theta_{i1,u}) G(\theta_{r,1}) G(\theta_{i,2}) G(\theta_{r,2}) G_r}{r_{tu}^2 r_2^2 r_3^2} \quad (7)$$
$$\times |h_{t,uv}|^2 |\gamma_v|^2 |h_{vw}|^2 |\gamma_w|^2 |h_{rw}|^2 P_t,$$

Further, user u's total received power at the R_x is expressed as [17]

$$P^u_{Rx} = |\sqrt{L_{\tau,\mathrm{u}}} \sum_{v=1}^{V} \sum_{w=1}^{W} |h_{t,uv}||\gamma||h_{vw}||\gamma||h_{rw}| \quad (8)$$
$$e^{-j\left(\psi_{t_{u_v}} + \alpha_v + \psi_{vw} + \varphi_w + \psi_{r_w} + \beta_U + \beta_3\right)}|^2 P_t,$$

where $\psi_{t_{u_v}}$ is the phase for the transmitter channel for user u and v^{th} reflecting element, ψ_{vw} is the phase for the $\mathbf{h}_{v,w}$ channel for v^{th} and w^{th} reflecting element, ψ_{r_w} is the phase for the R_x channel for w^{th} reflecting element, $|\gamma|$ is assumed to equal 1.

β_U and β_3 designate the deterministic phases corresponding to the traveled distance between each user u to RIS_1 over the first hop, and the traveled distance between RIS_2 and the R_x over the third hop. These phases are designated as follows:

$$\beta_u = 2\pi \times r_{t_u}/\lambda.$$

$$\beta_3 = 2\pi \times r_3/\lambda. \quad (9)$$

$$L_{\tau,\mathrm{u}} = L_{\text{FSPL,u}} \times L_{\text{absorption}} \quad (10)$$

$L_{\text{absorption}}$ designates the absorption losses according to [18], and $L_{\text{FSPL,u}}$ represents the free space path loss for each user u, and it is written as:

$$L_{FSPL,\tau,u} = \left(\frac{\lambda}{4\pi}\right)^6 \frac{G_{t,u} G(\theta_{i1,u}) G(\theta_{r,1}) G(\theta_{i,2}) G(\theta_{r,2}) G_r}{r_{tu}^2 r_2^2 r_3^2} \quad (11)$$

Therefore, equation (8) can be expressed as below:

$$P^u_{Rx} = |\sqrt{L_{\tau,\mathrm{u}}} \sum_{v=1}^{V} \sum_{w=1}^{W} |h_{t,uv}||h_{vw}||h_{rw}| \quad (12)$$
$$e^{-j\left(\psi_{t_{u_v}} + \alpha_v + \psi_{vw} + \varphi_w + \psi_{r_w} + \beta_U + \beta_3\right)}|^2 P_t,$$

The total received power for the first user can be maximized by solving the following system of equations:

$$\alpha_v + \varphi_w + \psi_{t_{1_v}} + \psi_{vw} + \psi_{r_w} + \beta_1 + \beta_3 = \mathbf{C}, \quad \forall v, w. \quad (13)$$

where $\mathbf{C}$ designates any constant value. This means that the power P_{Rx}^{u} is maximum when $\mathbf{C}$ is equal to constant $\forall v, w$. In this research paper, we will select $\mathbf{C} = 0$. Hence, equation (13) represents an over-determined system of equations with $V + W$ unknowns, and $V \times W$ equations. This problem is almost always inconsistent and it has no solution. Nonetheless, we will solve this problem using machine learning (ML) and specifically the DRL method to obtain the unknowns α_v and φ_w, calculate the received power for the first user, and then compare the results to the sub-optimal solution obtained from the mathematical technique pseudo inverse.

1) Moore-Penrose Pseudo Inverse Method: The Moore-Penrose pseudo-inverse solution for the over-determined system is expressed as follows:

$$\mathbf{M\Phi} = \mathbf{K}, \tag{14}$$

where $\mathbf{\Phi}$ with dimensions $(V + W) \times 1$ is the matrix that represents RIS_1 and RIS_2 phase shifts, $\mathbf{M}$ with dimensions $(V \times W) \times (V + W)$ represents the binary matrix, and $\mathbf{K}$ with dimensions $(V \times W) \times 1$ represents the matrix containing constant values such as the phase shifts of the transmitter channel $\mathbf{h}_{t,u}$, the phase shifts of the channel between RIS_1 and RIS_2 $\mathbf{h_{v,w}}$, and the phase shifts of the R_x channel $\mathbf{h}_r$.

$$\mathbf{M}^{+} = (\mathbf{M^T M})^{-1}\mathbf{M^T}, \tag{15}$$

where $\mathbf{M}^{+}$ is pseudo-inverse of a matrix $\mathbf{M}$. Thus, the pseudo-inverse solution $\mathbf{\Phi}$ is expressed as

$$\mathbf{\Phi} = \mathbf{M}^{+}\mathbf{K} \tag{16}$$

.

Moreover, the received signal-to-noise ratio (SINR for) T_{x_u} at the R_x can be written as::

$$\Omega_u = \frac{P_{Rx}^{u}}{\sum_{\substack{i=1 \\ i \neq u}}^{U} P_{Rx}^{i} + \sigma^2}. \tag{17}$$

Therefore, the data rate of user u is designated as:

$$R_u = \log_2(1 + \Omega_u). \tag{18}$$

Our goal in this work is to obtain the values of the RIS_1 and RIS_2 phase shifts that maximize the rate for user 1. Accordingly, the formulated problem at RIS_1 and RIS_2 is to obtain the phase shift matrices α_v and φ_w $\forall v, w$ that maximizes R_1, and it is written as:

$$\begin{aligned} &\max_{\alpha,\varphi} \sum_{u=1}^{U} \log_2\left(1 + \Omega_1\right), \\ &s.t. \\ &\text{C1}: |\gamma_v|^2 = 1, \forall v \in \{1, 2, ..., V\}, \\ &\text{C2}: |\gamma_w|^2 = 1, \forall w \in \{1, 2, ..., W\}, \end{aligned} \tag{19}$$

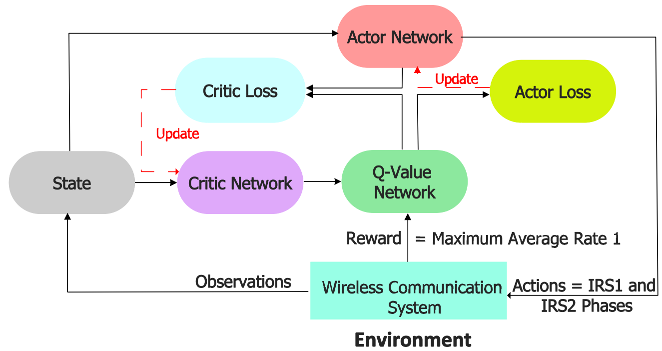

Figure 2: DDPG Model.

The solution to the optimization problem is non-trivial, it is an NP-hard problem, because of constant modulus constraints of RIS_1 and RIS_2 reflecting elements that make the problem non-convex. Thus, it is nearly not possible to find out a mathematical solution for the cascaded RIS optimization problem. To solve it, we utilize the DRL method, namely DDPG, rather than solving the challenging problem mathematically. The proposed DDPG scheme details are given in section IV.

IV. Proposed DDPG Scheme Based RISs' Phases Control

DDPG is a model-free RL method that merges the merits of both the policy gradients and Q-learning algorithms. Taking into consideration that the states in our scenario are dependent on the output sum rate and channel gains, while the actions are the RIS phase shifts, we are considering a continuous action and a continuous state system. Thus, we chose to use DDPG since its main advantage arises from the fact that it utilizes both the continuous state and action spaces.

The DDPG algorithm consists of several vital elements. These elements include the agents operating in our communication system model which are the RIS_1 and RIS_2, the states $\mathbf{s}^{(T)}$ represented by $\mathbf{h}_{t,u}^{(T)}$, $\mathbf{h}_{v,w}^{(T)}$, $\mathbf{h}_{r}^{(T)}$, and the rate for user 1 of the previous state $R_1^{(T-1)}$, the reward $\mathbf{r}^{(t)}$ which is the rate for user 1 $R_1^{(T)}$ at the current state, the action $\mathbf{a}^{(t)}$ which is represented by the phases of the RIS_1 and RIS_2, the policy function μ, and the Q-value function $Q(\mathbf{s}, \mathbf{a}|\theta^{Q})$ which measures how good is the action. The goal of this research paper is to improve the average rewards including the instant and future rewards. The DDPG algorithm consists of four neural networks that include the actor-network, the critic network, the target critic network, and the target actor-network to increase stability in the system.

A. DDPG Algorithm Framework

The goal of the DDPG scheme is to train the agents RIS_1 and RIS_2 to take actions that maximize the long-term average reward, which is in our system the rate of user 1 via the environment changes as shown in Fig. 2 taking into consideration the constant modulus constraints mentioned

Algorithm 1 DDPG Algorithm-based Framework

1: **Algorithm Initialization:** Set the timestep $T = 0$ and initialize the reply buffer of the DDPG agent $\mathcal{D}$ with capacity M.
2: Initialize the weights of the critic networks θ^{Q}, and actor networks θ^{μ} randomly.
3: Initialize target network parameters: $\theta^{Q'} \leftarrow \theta^{Q}$, and $\theta^{\mu'} \leftarrow \theta^{\mu}$.
4: **for** $T = 1$ to ∞ **do**
5: Observe the state $\mathbf{s}^{(T)}$ and select an action with exploration OU noise $\mathbf{a}^{(T)} = \mu(\mathbf{s}^{(T)}|\theta^{\mu}) + \mathbf{n}_T$.
6: Execute the action $\mathbf{a}^{(T)}$ at RIS_1 and RIS_2.
7: Receive the immediate reward $r^{(T)}$, observe the next state $s^{(T+1)}$, and store the transition $(\mathbf{a}^{(T)}, \mathbf{s}^{(T)}, r^{(T)}, \mathbf{s}^{(T+1)})$ in the replay memory D.
8: Sample the mini-batch transitions from $\mathcal{D}$ randomly: $B \leftarrow \{(\mathbf{s}^{(i)}, \mathbf{a}^{(i)}, r^{(i)}, \mathbf{s}^{(i+1)})\} \in \mathcal{D}$.
9: Calculate the target Q-value: $\tilde{Q}(\mathbf{s}^{(i)}, \mathbf{a}^{(i)}|\theta^{Q'}) = r^{(i)} + \Gamma Q(\mathbf{s}^{(i+1)}, \mu(\mathbf{s}^{(i)}|\theta^{\mu'})|\theta^{Q'})$ where Γ is the discount factor.
10: Update the parameters θ^{Q} in the critic network by minimizing the loss using the obtained target Q-value. $L = \frac{1}{|B|}\sum_{i=1}^{|B|}\left(\tilde{Q}(\mathbf{s}^{(i)}, \mathbf{a}^{(i)}|\theta^{Q'}) - Q(\mathbf{s}^{(i)}, \mathbf{a}^{(i)}|\theta^{Q})\right)^2$
11: Update the parameters θ^{μ} in actor-network based on the sampled policy gradient:
$$\nabla_{\theta^{\mu}}\boldsymbol{J} \approx \frac{1}{|B|}\sum_{i=1}^{|B|}\nabla_a Q(\mathbf{s}^{(i)}, \mathbf{a}^{(i)}|\theta^{Q})\nabla_{\theta^{\mu}}\mu(\mathbf{s}^{(i)}|\theta^{\mu})$$
12: Update the target actor and critic networks using the soft updates τ to increase the learning stability:
$$\theta^{Q'} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'}$$
$$\theta^{\mu'} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'}$$
13: **end for**

in eq 19. The agents modify the randomized phase shift matrices, and the policy in a way that copes with the random environmental statistical behavior to maintain a long-term average reward, instead of an immediate response to the channel random changes. The DDPG algorithm-based framework is shown in algorithm 1.

V. Numerical Results

In this section, we assess the performance of the DRL-based multi-hop RIS in THz networks. To demonstrate the performance of the DDPG algorithm in our system, we compare the results generated by DDPG with the pseudo inverse solution, and with those generated by using random phases. The default simulation parameters in our system are revealed in Table I. We define the term distance ratio as the ratio between the distance between user 1 and RIS_1 r_{t1}, and the distance between user 2 and RIS_1 r_{t2}. Numerical results for the data rates are calculated for 2 users and 1000 Monte-Carlo simulations.

We reveal the results for maximizing the received power for the first user at the R_x by plotting the rate for the first user versus the distance ratio between the two users using DDPG, and pseudo-inverse methods. Fig. 3 shows a comparison between the results generated by DDPG, pseudo-inverse, and the case of randomized phases (i.e. without optimization) for a correlation factor equal to 0.9. It is clear from the figure that as the distance ratio increases the rate decreases because the interference between user 1 and user 2 increases. Further, DDPG achieves higher rates than that of the pseudo-inverse, and randomized phases case. Moreover, Fig. 4 reveals the rates for the DDPG scheme versus the distance ratio for different correlation factors. It is obvious that when the correlation factor ρ increases, the data rates for the DDPG scheme increase. The reason for that, increasing the correlation factor value, will increase the learning efficiency of the DDPG scheme. Therefore, the DDPG algorithm attains higher data rates than other methods especially when the correlation factor ρ is high.

Table I: Parameters Used in Simulation

Simulation Parameters	**Values**
Number of Users (U)	2
The Speed of light c	3×10^8
Carrier Frequency f	300×10^9
Number of antennas per transmitterN_t	1
Number of antennas at the R_x N_r	1
Wavelength (λ)	1×10^{-3}
Number of RIS $_1$ Reflecting Elements (V)	18
Number of RIS_2 Reflecting Elements (W)	18
X-axis of RIS_1 x_{r1}	5
Y-axis of RIS_1 y_{r1}	10
height of RIS_1 h_{r1}	12
X-axis of RIS_2 x_{r2}	10
Y-axis of RIS_2 y_{r2}	10
height of RIS_2 h_{r2}	12
Distance between User 1 and RIS_1	3 to 15
Distance between User 2 and RIS_1	15
Heights of User 1 and User 2 h_t	5
RIS_1 and RIS_2 Reflection Coefficients α	1
RIS_1 and RIS_2 half-power Spacing d_x	$\lambda/2$
RIS_1 and RIS_2 Element Spacing d_y	$\lambda/2$
Antenna diameter in meters D_t	0.12
X-axis of Rx x_{rx}	20
Y-axis of Rx y_{rx}	0
height of Rx h_r	5
Bandwidth	2×10^9 MHz
Noise power spectral density N_{PSD}	-174 dB/Hz
Noise figure at the R_X F_{dB}	10
Average Noise power in dB $N0$	-174 dB/Hz
Noise power in linear scale no	7.9621×10^{-11}
Transmitters to RIS_1 Path loss exponent	2
RIS_2 to R_x Path loss exponent	2
Rician Factor	10
Coefficient of Soft Updates τ	1×10^{-3}
Batch size	128
Replay Buffer Capacity $\mathcal{C}$	10^5
Number of episodes	10000
Critic-Network learning rate	3×10^{-4}
Actor-Network learning rate	1×10^{-4}
Target Critic-Network learning rate	3×10^{-4}
Target Actor-Network learning rate	1×10^{-4}
Discount factor of the future reward Γ	0.99

VI. Conclusion

In this research paper, a multiple access scenario with a multi-hop RIS uplink system is considered to overcome the short-range communications issues in THz networks, to maximize the received power for the first user. The maximization problem for the multi-hop RIS network is over-determined because it consists of a number of equations greater than the number of unknowns. Further, it is non-convex

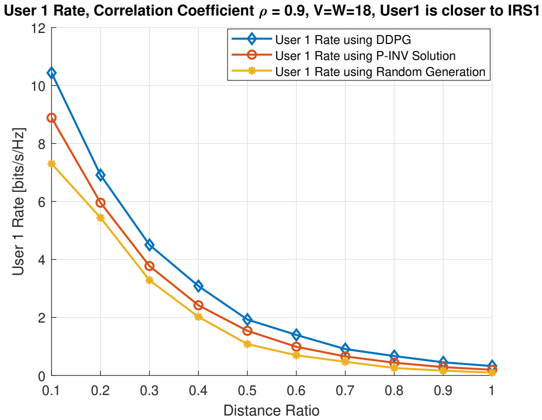

Figure 3: Rate of user 1 vs Distance ratio for correlation factor ρ equal to 0.9. V = W = 18.

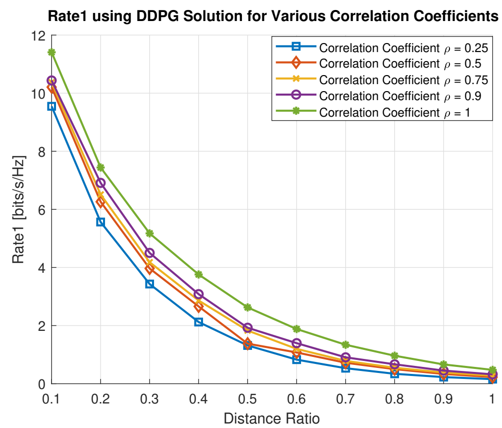

Figure 4: Rate of user 1 vs Distance ratio for various correlation factors ρ. V = W = 18.

due to the constant modulus constraints at RIS_1 and RIS_2. To solve this problem, we utilized the DDPG scheme which possesses the ability to deal with over-determined systems and non-convex optimization problems. DDPG finds out the optimal RIS phases which in turn maximizes the received power for the first user. Numerical results reveal that DDPG attains rates higher than that of the pseudo-inverse solution and randomized phases. Furthermore, DDPG shows the importance of the correlation in the channels to optimize the learning process and attain higher data rates.

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] M. Jian and R. Liu, "Baseband Signal Processing for Terahertz: Waveform Design, Modulation and Coding," 2021 International Wireless Communications and Mobile Computing (IWCMC), Harbin City, China, 2021, pp. 1710-1715, doi: 10.1109/IWCMC51323.2021.9498810.

[3] S. Gong et al., "Toward Smart Wireless Communications via Intelligent Reflecting Surfaces: A Contemporary Survey," in IEEE Communications Surveys and Tutorials, vol. 22, no. 4, pp. 2283-2314, Fourthquarter 2020, doi: 10.1109/COMST.2020.3004197.

[4] R. Liu, Q. Wu, M. Di Renzo and Y. Yuan, "A Path to Smart Radio Environments: An Industrial Viewpoint on Reconfigurable Intelligent Surfaces," in IEEE Wireless Communications, vol. 29, no. 1, pp. 202-208, February 2022, doi: 10.1109/MWC.111.2100258.

[5] B. Zheng, C. You, W. Mei and R. Zhang, "A Survey on Channel Estimation and Practical Passive Beamforming Design for Intelligent Reflecting Surface Aided Wireless Communications," in IEEE Communications Surveys and Tutorials, vol. 24, no. 2, pp. 1035-1071, Secondquarter 2022, doi: 10.1109/COMST.2022.3155305.

[6] Y. Liu et al., "Reconfigurable Intelligent Surfaces: Principles and Opportunities," in IEEE Communications Surveys and Tutorials, vol. 23, no. 3, pp. 1546-1577, thirdquarter 2021, doi: 10.1109/COMST.2021.3077737.

[7] 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.

[8] W. Mei and R. Zhang, "Cooperative Multi-Beam Routing for Multi-IRS Aided Massive MIMO," ICC 2021 - IEEE International Conference on Communications, 2021, pp. 1-6, DOI: 10.1109/ICC42927.2021.9500347.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] Y. Pan, K. Wang, C. Pan, H. Zhu and J. Wang, "Sum-Rate Maximization for Intelligent Reflecting Surface Assisted Terahertz Communications," in IEEE Transactions on Vehicular Technology, vol. 71, no. 3, pp. 3320-3325, March 2022, doi: 10.1109/TVT.2022.3140869.

[14] J. Qiao, C. Zhang, A. Dong, J. Bian and M. -S. Alouini, "Securing Intelligent Reflecting Surface Assisted Terahertz Systems," in IEEE Transactions on Vehicular Technology, vol. 71, no. 8, pp. 8519-8533, Aug. 2022, doi: 10.1109/TVT.2022.3172763.

[15] 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.

[16] 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.

[17] K. Ntontin, A. -A. A. Boulogeorgos, D. G. Selimis, F. I. Lazarakis, A. Alexiou and S. Chatzinotas, "Reconfigurable Intelligent Surface Optimal Placement in Millimeter-Wave Networks," in IEEE Open Journal of the Communications Society, vol. 2, pp. 704-718, 2021, DOI: 10.1109/OJCOMS.2021.3068790.

[18] J. Kokkoniemi, 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.