

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

Alice Faisal, *Graduate Member, IEEE*, Ibrahim Al-Nahhal, *Senior Member, IEEE*, Octavia A. Dobre, *Fellow, IEEE* and Telex M. N. Ngatched, *Senior Member, IEEE*

## Abstract

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

## Index Terms

Reconfigurable intelligent surface (RIS), full-duplex, deep reinforcement learning, single and distributed RIS.

## I. INTRODUCTION

Recently, the reconfigurable intelligent surface (RIS) technology has been proposed as a key enabler to meet the demands of future technologies [1], [2]. RIS is a meta-surface consisting of

The authors are with the Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL, Canada, (e-mail: afaisal@mun.ca; ioalnahhal@mun.ca; odobre@mun.ca; tngatched@grenfell.mun.ca).

low-cost passive elements that can be programmed to turn the random nature of wireless channels into a partially deterministic space to improve the propagation of wireless signals [3]. In addition to the RIS technology, full-duplex (FD) transmission has been regarded as a potential approach to increase the spectral efficiency of wireless systems by enabling simultaneous transmission and reception [4], [5].

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

The above works that used AO techniques exhibit both loss of optimality and high computational complexity. Deep reinforcement learning (DRL) has emerged as a powerful approach to optimize the RIS phase shifts by overcoming the practical implementation problems of AO techniques. Furthermore, DRL approaches enable addressing mathematically intractable nonlinear problems directly, without the need of prior relaxations requirements. The work in [10] proposed a DRL algorithm to maximize the rate, where both half-duplex and FD operating modes are considered together. However, only a single RIS deployment was considered. The rapid changes in dynamic environments can obliterate/annihilate the RIS deployment benefits when the corresponding link is blocked/weak. In such cases, deploying distributed power-efficient RISs can cooperatively enhance the coverage of the system by providing multiple paths of received signals. Moreover, the computational complexity can be further reduced. To the best of the authors' knowledge, utilizing DRL for investigating the performance of single and distributed RIS deployment schemes in FD multiple-input single-output (MISO) systems has not yet been considered in the literature. Our contributions are summarized as follows:

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

optimized scenarios.

- The proposed DRL algorithm provides a considerable reduction in the computational complexity compared to the DRL algorithm in [10].
- The complexity analysis and Monte Carlo simulations support the findings.

The rest of this letter is organized as follows: Section II presents the system model and problem formulation. The proposed DRL algorithm is introduced in Section III. Simulation results and conclusions are presented in Sections IV and V, respectively.

## II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider an RIS-assisted FD MISO system, where single and distributed RIS deployment schemes are investigated.  $S_1$  and  $S_2$  represent the base station (BS) and user equipment (UE), respectively. Both the BS and UE are equipped with  $M$  transmit antennas and one receive antenna. The  $r$ -th RIS,  $R_r$ , consists of  $N_r$  programmable reflecting elements. Note that the total number of elements for both deployment schemes is defined as  $N = N_r\Lambda$  to ensure the same number of RIS elements for all scenarios, where  $\Lambda$  is the number of RISs. As illustrated in Fig. 1, three scenarios are investigated based on the links' quality. In the first scenario, the single and distributed RIS deployment schemes have strong line-of-sight (LoS) components in all links. Scenarios 2 and 3 assume that the links of  $R_1$ - $S_2$  and  $S_1$ - $R_2$  are weak due to obstacles, respectively. It is worth noting that from a practical point of view, it is more probable that the longer distance links (i.e.,  $R_1$ - $S_2$  and  $S_1$ - $R_2$ ) may experience blockage since the short-distance links are planned deployment links. It also ensures a fair comparison between the two deployment schemes as the RIS benefits are embraced in all scenarios.

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

$$y_i = \left( \sum_{r \in \Lambda} \mathbf{h}_{R_r S_i}^H \mathbf{\Theta}_r \mathbf{H}_{S_i R_r} + \mathbf{h}_{S_i S_i}^H \right) \mathbf{w}_i^H \mathbf{x}_{\bar{i}} + \mathbf{h}_{S_i S_i}^H \mathbf{w}_i x_i + n, \quad i = 1, 2, \quad \Lambda = \begin{cases} 1 & \text{Single RIS} \\ 2 & \text{Distributed RIS,} \end{cases} \quad (1)$$

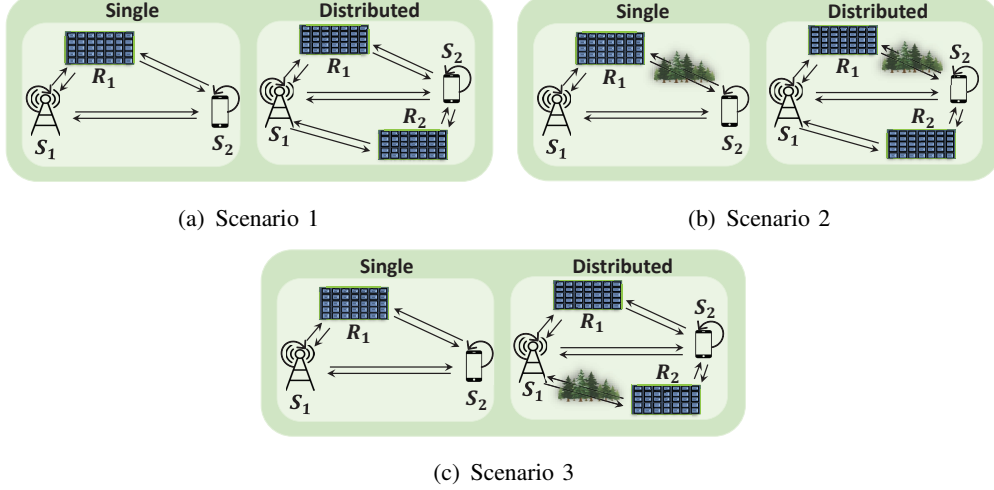


Fig. 1: RIS-assisted FD MISO system.

where  $n \sim \mathcal{CN}(0, \sigma^2)$  denotes the additive white complex Gaussian noise with zero-mean and variance  $\sigma^2$ . The diagonal matrix  $\mathbf{\Theta}_r = \text{diag}(e^{j\varphi_{r1}}, \dots, e^{j\varphi_{rn}}, \dots, e^{j\varphi_{rN_r}}) \in \mathbb{C}^{N_r \times N_r}$  represents the phase shifts of  $R_r$ , where  $\varphi_{rn} \in [-\pi, \pi)$  is the phase shift introduced by the  $n$ -th reflecting element. The source node,  $S_i$ , employs an active beamforming  $\mathbf{w}_i \in \mathbb{C}^{M \times 1}$  to transmit the information signal,  $x_i$ , with  $\mathbb{E}\{|x_i|^2\} = 1$ , where  $\mathbb{E}\{\cdot\}$  denotes the expectation operation.

Based on (1), the received signal-to-interference plus-noise ratio,  $\gamma_i$ , and achievable rate,  $\mathcal{R}_i$ , measured in bit per second per Hertz (bps/Hz), are respectively given as

$$\gamma_i = \frac{\left| \left( \sum_{r \in \Lambda} \mathbf{h}_{R_r S_i}^H \mathbf{\Theta}_r \mathbf{H}_{S_i R_r} + \mathbf{h}_{S_i S_i}^H \right) \mathbf{w}_i \right|^2}{|\mathbf{h}_{S_i S_i}^H \mathbf{w}_i|^2 + \sigma^2},$$

$$i = 1, 2, \quad \Lambda = \begin{cases} 1 & \text{Single RIS} \\ 2 & \text{Distributed RIS,} \end{cases} \quad (2)$$

and

$$\mathcal{R}_i = \log_2(1 + \gamma_i). \quad (3)$$

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

and is formulated as

$$(P1) \quad \max_{\mathbf{w}_i, \bar{\Theta}} \sum_{i=1}^2 \mathcal{R}_i \quad (4a)$$

$$\text{s.t.} \quad -\pi \leq \varphi_{rn} \leq \pi, \quad n = 1, \dots, N_r, \quad (4b)$$

$$\|\mathbf{w}_i\|^2 \leq P_{\max}, \quad i = 1, 2. \quad (4c)$$

Here,  $\bar{\Theta} = \text{diag}(\Theta_1, \Theta_2)$  is a block matrix whose diagonal entries contain the phase shifts of the two RISs for the distributed RIS, and  $\bar{\Theta} = \text{diag}(\Theta_1)$  when a single RIS is considered.  $P_{\max}$  is the maximum transmitted power of  $S_i$ . It is worth noting that (P1) is challenging to solve due to the non-convexity of the objective function and constraints. Thus, an efficient solution is proposed which decouples the problem into two sub-problems.

### III. PROPOSED SOLUTION

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

#### A. Beamformers Optimization for a Given $\bar{\Theta}$

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

$$I(s; y) = \max_{q(s|y)} \mathbb{E}[\log(q(s|y)) - \log(p(s))], \quad (5)$$

where the optimal  $q^*(s|y)$  is the posterior probability [8], and is expressed as  $q^*(s|y) = \frac{p(s)p(y|s)}{p(y)} \triangleq p(s|y)$ . Based on (5), the achievable rate of  $S_i$  is

$$\mathcal{R}_i = \max_{q(s_i|y_i)} \mathbb{E}[\log(q(s_i|y_i)) - \log(p(s_i))], \quad (6)$$

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

$$f_i^* = \frac{b_i}{b_i^2 + |\mathbf{h}_{S_i S_i}^H \mathbf{w}_i|^2}, \quad (7)$$

$$b_i = \left| \left( \sum_{r \in \Lambda} \mathbf{h}_{R_r S_i}^H \mathbf{\Theta}_r \mathbf{H}_{S_i R_r} + \mathbf{h}_{S_i S_i}^H \right) \mathbf{w}_i \right|^2. \quad (8)$$

To this end, (4a) in (P1) can be re-expressed as

$$\max_{\mathbf{w}_i, \mathbf{\Theta}, f_i, \Sigma_i} \sum_{i=1}^2 \mathbb{E}[\log(p(s_i|y_i)) - \log(p(s_i))]. \quad (9)$$

Let  $\alpha_i = \sum_{r \in \Lambda} \mathbf{h}_{R_r S_i}^H \mathbf{\Theta}_r \mathbf{H}_{S_i R_r} + \mathbf{h}_{S_i S_i}^H$  and  $b_i = |\alpha_i \mathbf{w}_i|^2$ . The expectation term in (9) is calculated as

$$\begin{aligned} & \mathbb{E}[\log(\mathcal{CN}(f_i y_i, \Sigma_i)) - \log(\mathcal{CN}(0, 1))] \\ &= \exp\left(f_i y_i + \frac{\Sigma_i}{2}\right) - \exp\left(\frac{1}{2}\right) \\ &= -\frac{1}{2} f_i |\alpha_i \mathbf{w}_i|^2 + \mathbf{w}_i (f_i \alpha_i + f_i \mathbf{h}_{S_i S_i}^H). \end{aligned} \quad (10)$$

Furthermore, let  $\beta_i = f_i \alpha_i + f_i \mathbf{h}_{S_i S_i}^H$ . Thus, (10) can be defined as a convex quadratically constrained quadratic program:

$$-\frac{1}{2} f_i |\alpha_i \mathbf{w}_i|^2 + \mathbf{w}_i \beta_i, \quad (11)$$

where its solution can be derived as

$$\mathbf{w}_i^* = (v^* + f_i \alpha_i \alpha_i^H)^{-1} \beta_i. \quad (12)$$

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

### B. Phase Shift Optimization for a Given $\mathbf{w}_i$ and $\mathbf{w}_i$

Model-free RL can be employed to address a decision-making problem by learning the optimal solution in dynamic environments. Therefore, the RIS-assisted FD MISO system represents the DRL environment and the RIS controller represents the DRL agent. At each time step  $t$ , the

agent observes the current state,  $s_t$ , from the environment, takes an action,  $a_t$ , based on a policy,  $\tilde{\pi}$ , receives a reward,  $r_t$ , of executing  $a_t$ , and transitions to a new state  $s_{t+1}$ . The key elements of DRL are defined as follows: The *state space* at time step  $t$ , includes  $\varphi_{rn} \forall n = 1, \dots, N_r$  and the corresponding  $\sum_{i=1}^2 \mathcal{R}_i$  at time step  $t-1$ , i.e.,  $s_t = \left[ \sum_{i=1}^2 \mathcal{R}_i^{(t-1)}, \varphi_{r1}^{(t-1)}, \dots, \varphi_{rn}^{(t-1)}, \dots, \varphi_{rN_r}^{(t-1)} \right]$ . The *action space* at time step  $t$  is expressed as  $a_t = \left[ \varphi_{r1}^{(t)}, \dots, \varphi_{rn}^{(t)}, \dots, \varphi_{rN_r}^{(t)} \right]$ , and the *reward* at time step  $t$  is  $r_t = \sum_{i=1}^2 \mathcal{R}_i^{(t)}$ .

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

At the initialization stage, four networks are generated, i.e., target and evaluation DNN. The target networks are generated by making a copy of the actor and critic evaluation NNs,  $\mu'(s_t | \theta_{\mu'})$  and  $Q'(s_t, a_t | \theta_{q'})$ . The experience replay with memory  $D$  is built to reduce the correlation of the training samples. During each episode, all the channel state information is obtained. Then, the agent takes  $a_t$  generated by the actor network, calculates the  $r_t$ , and transitions to  $s_{t+1}$ . The experience is then stored  $(s_t, a_t, r_t, s_{t+1})$  into  $D$ , and the critic evaluation network randomly samples a minibatch transitions,  $N_B$ , to calculate the target value  $y_j$ , as

$$y_j = r_j + \rho Q'(s_{j+1}, \mu'(s_{j+1} | \theta_{\mu'}) | \theta_{q'}), \quad (13)$$

where  $\rho \in (0, 1]$  is the discount factor. The actor and critic NN parameters,  $\theta_\mu$  and  $\theta_q$ , are updated using the stochastic gradient descent and policy gradient, respectively, as

$$L = \frac{1}{N_B} \sum_j (y_j - Q(s_j, a_j | \theta_q))^2, \quad (14)$$

and

$$\nabla_{\theta_\mu} = \frac{1}{N_B} \sum_j \nabla_a Q(s, a | \theta_q) |_{s=s_j, a=\mu(s_j)} \nabla_{\theta_\mu} \mu(s | \theta_\mu) |_{s_j}. \quad (15)$$

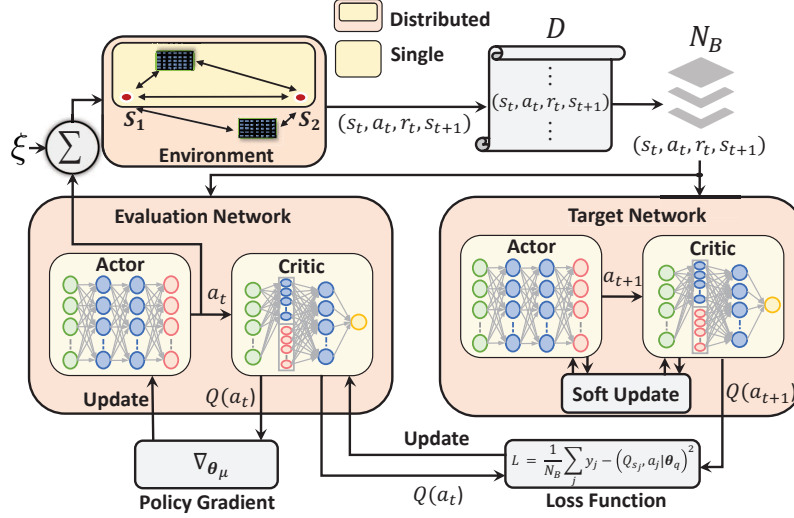


Fig. 2: The proposed DRL algorithm structure.

Finally, the target NN parameters are updated using a soft update coefficient,  $\tau$ , as

$$\theta_{q'} \leftarrow \tau \theta_q + (1 - \tau) \theta_{q'}, \quad (16)$$

$$\theta_{\mu'} \leftarrow \tau \theta_{\mu'} + (1 - \tau) \theta_{\mu'}. \quad (17)$$

This process is repeated for  $K$  and  $T$  until convergence is reached. The structure of the proposed DRL algorithm is illustrated in Fig. 2 and summarized in Algorithm 1.

### C. Proposed DNN Design

The proposed DNN models are designed as feedforward fully connected NNs. The proposed algorithm contains four NNs (actor and critic for each evaluation and target network). Each NN has an input layer, two hidden layers and output layer, as shown in Fig. 2. The input layer of the actor and critic networks contains  $N + 1$  neurons (i.e., size of  $s_t$ ). The input of the actor is passed to two hidden layers, each having  $\psi_i$  neurons, where  $\psi_i$  is the number of neurons of the  $i$ -th layer. On the other hand, the input of the critic network is passed to the first hidden layer that is concatenated with  $a_t$  (i.e., size of  $\psi_i + N$ ), and then passed to the second hidden layer. The two hidden layers for each of the actor and critic networks use the *ReLU* activation function whereas the output layer of the actor network uses the *tanh* activation function. The output layer of the actor and critic networks contains  $N$  neurons (i.e., size of  $a_t$ ) and one neuron (i.e., Q-value), respectively.



---

**Algorithm 1** Proposed DRL algorithm.

---

**Initialize:**  $\theta_\mu$  and  $\theta_q$  with random weights,  $D$ ,  $\rho$ ,  $\tau$ , learning rate  $\nu$ ,  $\theta_{\mu'} \leftarrow \theta_\mu$  and  $\theta_{q'} \leftarrow \theta_q$ ;

- 1: **repeat**
- 2:   Collect the channels of the  $k$ -th episode;
- 3:   Randomly initialize  $\varphi_{rn} \forall n = 1, \dots, N_r$ ;
- 4:   Calculate  $\mathbf{w}_{\bar{i}}$  using (12);
- 5:   Initialize  $\xi \sim \mathcal{CN}(0, 0.1)$ ;
- 6:   **repeat**
- 7:     Obtain  $a_t = \mu(s_t | \theta_\mu) + \xi$  from the actor network and reshape it;
- 8:     Repeat **Line #4**;
- 9:     Observe the new state,  $s_{t+1}$ , given  $a_t$ ;
- 10:    Store  $(s_t, a_t, r_t, s_{t+1})$  in  $D$ ;
- 11:    When  $D$  is full, sample a minibatch of  $N_B$  transitions  $(s_j, a_j, r_j, s_{j+1})$  randomly from  $D$ ;
- 12:    Compute the target value from (13);
- 13:    Update the critic using (14);
- 14:    Update the actor using (15);
- 15:    Update the target NNs using (16) and (17);
- 16:   **until**  $t = T$ ;
- 17: **until**  $k = K$ ;

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

---

#### D. Complexity Analysis

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

$$C_P = 2 \left( \sum_{i=1}^3 (\psi_i^A + 1) \psi_{i+1}^A + \sum_{i=1}^3 (\psi_i^C + 1) \psi_{i+1}^C \right), \quad (18)$$

$$C_M = 2 \left( \sum_{i=1}^3 \psi_i^A \psi_{i+1}^A + \sum_{i=1}^3 \psi_i^C \psi_{i+1}^C \right), \quad (19)$$

$$C_A = 2 \left( \sum_{i=1}^3 \psi_i^A \psi_{i+1}^A + \sum_{i=1}^3 \psi_{i+1}^A + \sum_{i=1}^3 \psi_i^C \psi_{i+1}^C + \sum_{i=1}^3 \psi_{i+1}^C \right), \quad (20)$$

where the actor and critic networks are expressed through the superscripts A and C, respectively. The complexity reduction of using the proposed DRL algorithm over the algorithm in [10] for the single RIS-assisted FD system is

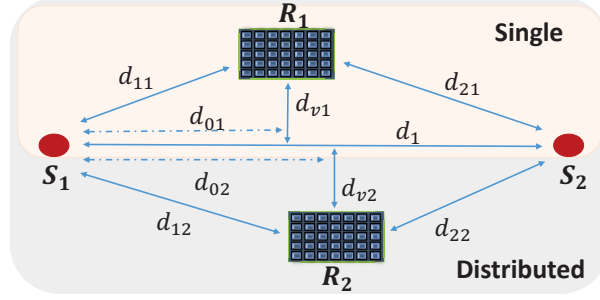


Fig. 3: Simulation setup.

$$\text{Reduction} = 1 - \frac{\{C_{\chi}^A + C_{\chi}^C\}_{\text{Proposed}}}{\{C_{\chi}^A + C_{\chi}^C\}_{[10]}}, \chi \in \{\mathcal{P}, \mathcal{A}, \mathcal{M}\}. \quad (21)$$

#### IV. SIMULATION RESULTS

Figure 3 illustrates the simulation setup, where the considered parameters are:  $d_{v1} = d_{v2} = 2$  m and  $d_1 = 50$  m. The distances between the links are:  $d_{11} = \sqrt{d_{01}^2 + d_{v1}^2}$  m,  $d_{12} = \sqrt{d_{02}^2 + d_{v2}^2}$  m,  $d_{21} = \sqrt{(d_1 - d_{01})^2 + d_{v1}^2}$  m, and  $d_{22} = \sqrt{(d_1 - d_{02})^2 + d_{v2}^2}$  m. The path loss (PL) at distance  $d_{ir}$  is modeled as  $\text{PL} = PL_0 - 10\zeta \log_{10} \left( \frac{d_{ir}}{D_r} \right)$  [14], where  $PL_0$  is the PL at a reference distance  $D_r$  and  $\zeta$  is the PL exponent, in which  $PL_0 = -35.6$  dB and  $D_r = 1$  m. The channels are modeled as Rayleigh fading whenever a blocking element exists. Otherwise, the channels are modeled as Rician with a factor of 10. The PL exponents of the  $S_1$ - $S_2$ ,  $S_1$ - $R_r$ , and  $S_2$ - $R_r$  channels are set to  $\zeta_{\text{BU}} = 4$ ,  $\zeta_{\text{BR}} = 2.1$ , and  $\zeta_{\text{UR}} = 2.2$ , respectively [9]. The PL of the SI channels is  $-95$  dB. The total transmit power is  $P = 15$  dBm, while the noise power is  $\sigma^2 = -80$  dBm [7].

The parameters of the proposed DRL are as follows:  $T = 800$ ,  $K = 500$ ,  $N_B = 16$ ,  $\nu_A = 0.0001$ ,  $\nu_C = 0.0002$ , decaying rate = 0.0001,  $\rho = 0.99$ ,  $\tau = 0.001$ , and  $D = 50000$ . Both actor and critic networks use the Adam optimizer for updating the parameters. The number of neurons of the hidden layers are,  $\psi_1 = 100$  and  $\psi_2 = 45$ . To validate the performance of the proposed algorithm, it is compared with the non-optimized scheme, referred to as random phase shifts. Furthermore, it is compared with the algorithm in [10] for the single RIS-assisted FD system to show the superiority of the proposed beamforming derivations over the approximated derivations in [10]. To ensure a fair comparison, it is assumed that  $N$  is the same for both deployment schemes. Hence, each RIS in the distributed scheme has half the number of elements of the single scheme.

Figure 4 studies the RIS deployment problem in both single and distributed RIS-assisted FD system. In the single RIS scheme, the sum-rate gradually increases when the RIS gets closer to

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

Figure 5 illustrates the effect of increasing  $N$  on the system performance. Three practical scenarios are considered to investigate the preference of using single or distributed RIS schemes. In Scenario 1, the distributed and single RIS schemes achieve a similar performance due to the strong LoS components (i.e., good quality links), and  $N$  is the same in both schemes. In Scenario 2, the results illustrate that the distributed RIS system significantly outperforms the single RIS system when the  $R_1$ - $S_2$  link is blocked/weak. In this case, the distributed RIS scheme has a higher sum-rate since it compensates for the poor quality link by providing an alternative path. On the other hand, if the link between  $R_2$ - $S_2$  is blocked/weak, as in Scenario 3, the single RIS scheme outperforms the distributed RIS since the former has double the number of elements compared to the latter. It is also worth noting that the proposed DRL algorithm provides a significant improvement in the sum-rate for the single and distributed RIS schemes compared to the random RIS phase shifts in all scenarios. The performance of the studied scenarios provides important insights into the preference of each deployment scheme based on the link conditions. Scenario 1 further points that the deployment cost should be considered if both schemes yield similar performance, as the required channel state information of the single RIS scheme is less than that of the distributed RIS scheme.

In the single RIS scheme, the proposed beamforming derivation improves the sum-rate performance in all scenarios, when compared to [10], as depicted in Fig. 5. Moreover, as shown in Fig. 6, the proposed DRL algorithm provides a complexity reduction percentage up to 40% for the range of  $N$  from 20 to 60 compared to the DRL presented in [10], and it saturates at 57% when  $N$  is very large.

## V. CONCLUSION

This letter optimized the beamformers and RIS phase shifts to maximize the sum-rate for both single and distributed RIS deployment schemes. Three practical scenarios were considered to

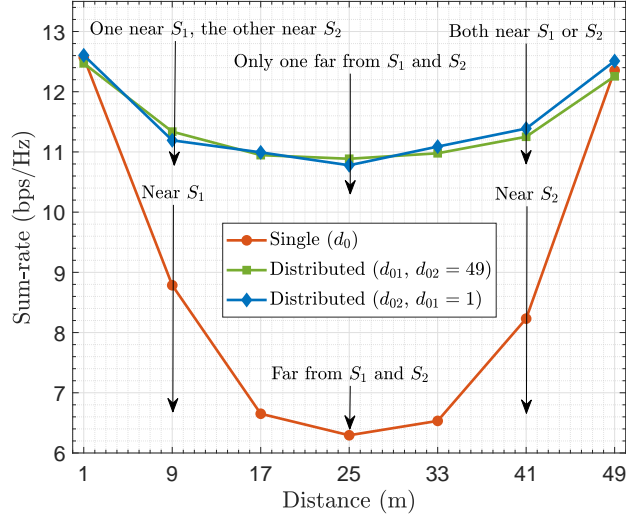


Fig. 4: RIS deployment investigation.

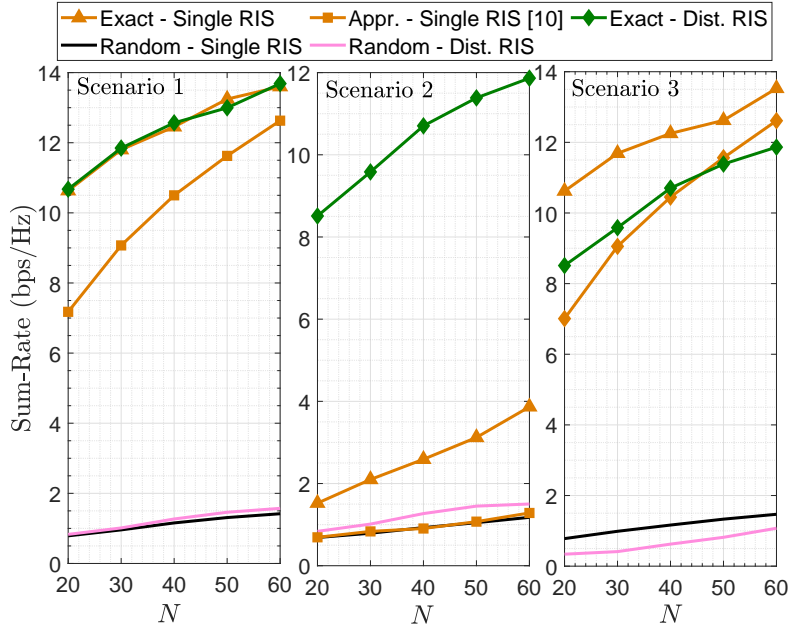


Fig. 5: The impact of varying  $N$  on the system performance.

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

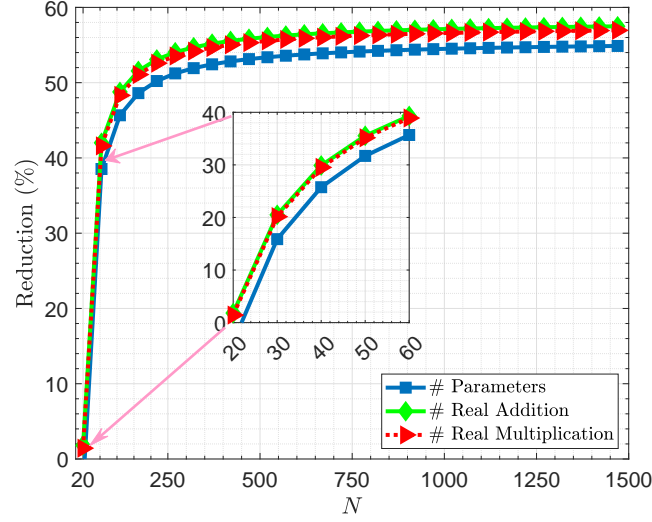


Fig. 6: Complexity reduction percentage versus  $N$ .

RIS phase shifts for multi-user systems.

#### REFERENCES

- [1] I. Al-Nahhal *et al.*, “Reconfigurable intelligent surface-assisted uplink sparse code multiple access,” *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 2058–2062, Feb. 2021.
- [2] —, “Reconfigurable intelligent surface optimization for uplink sparse code multiple access,” *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 133–137, Jan. 2022.
- [3] M. A. ElMossallamy *et al.*, “Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities,” *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 3, pp. 990–1002, Sep. 2020.
- [4] R. Askar *et al.*, “Interference handling challenges toward full duplex evolution in 5G and beyond cellular networks,” *IEEE Wirel. Commun.*, vol. 28, no. 1, pp. 51–59, Feb. 2021.
- [5] A. Yadav and O. A. Dobre, “All technologies work together for good: A glance at future mobile networks,” *IEEE Wirel. Commun.*, vol. 25, no. 4, pp. 10–16, Aug. 2018.
- [6] R. Alghamdi *et al.*, “Intelligent surfaces for 6G wireless networks: A survey of optimization and performance analysis techniques,” *IEEE Access*, vol. 8, pp. 202795–202818, Oct. 2020.
- [7] H. Shen *et al.*, “Beamforming design with fast convergence for IRS-aided full-duplex communication,” *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2849–2853, Dec. 2020.
- [8] Y. Zhang *et al.*, “Sum rate optimization for two way communications with intelligent reflecting surface,” *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1090–1094, May 2020.
- [9] M. A. Saeidi *et al.*, “Weighted sum-rate maximization for multi-IRS-assisted full-duplex systems with hardware impairments,” *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 466–481, Jun. 2021.
- [10] A. Faisal *et al.*, “Deep reinforcement learning for optimizing RIS-assisted HD-FD wireless systems,” *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3893–3897, Dec. 2021.
- [11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1997.
- [12] S. Boyd *et al.*, *Convex optimization*. Cambridge university press, 2004.
- [13] T. P. Lillicrap *et al.*, “Continuous control with deep reinforcement learning,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, May 2016, pp. 1–14.

- [14] K. Feng *et al.*, “Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems,” *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 745–749, May 2020.