Multi-agent Reinforcement Learning-based Joint Precoding and Phase Shift Optimization for RIS-aided Cell-Free Massive MIMO Systems

Yiyang Zhu, Enyu Shi, Ziheng Liu, Jiayi Zhang, *Senior Member, IEEE*, Bo Ai, *Fellow, IEEE*

***Abstract*—Cell-free (CF) massive multiple-input multiple-output (mMIMO) is a promising technique for achieving high spectral efficiency (SE) using multiple distributed access points (APs). However, harsh propagation environments often lead to significant communication performance degradation due to high penetration loss. To overcome this issue, we introduce the reconfigurable intelligent surface (RIS) into the CF mMIMO system as a low-cost and power-efficient solution. In this paper, we focus on optimizing the joint precoding design of the RIS-aided CF mMIMO system to maximize the sum SE. This involves optimizing the precoding matrix at the APs and the reflection coefficients at the RIS. To tackle this problem, we propose a fully distributed multi-agent reinforcement learning (MARL) algorithm that incorporates fuzzy logic (FL). Unlike conventional approaches that rely on alternating optimization techniques, our FL-based MARL algorithm only requires local channel state information, which reduces the need for high backhaul capacity. Simulation results demonstrate that our proposed FL-MARL algorithm effectively reduces computational complexity while achieving similar performance as conventional MARL methods.**

***Index Terms*—Reconfigurable intelligent surface, cell-free massive MIMO, precoding, spectral efficiency, multi-agent reinforcement learning.**

I. Introduction

The sixth-generation (6G) network will be a vital component in all parts of future society, industry, and life, given its primary mission to fulfill the communication needs of humans and intelligent machines [1]. The integration of distributed networks and massive MIMO confers notable advantages upon an ultra-dense network known as cell-free (CF) massive multiple-input multiple-output (mMIMO). In the context of CF mMIMO networks, a substantial array of distributed access points (APs) collectively cater to a limited user base using concurrent time-frequency resources, while all base stations (BSs) are linked to a central processing unit (CPU) through backhaul wireless connections [2], [3].

To enhance the network capacity, the deployment of a large number of distributed APs in the cell-free network is necessary. However, this approach entails significant costs and power consumption. Fortunately, the authors of [4] present a promising solution for enhancing network capacity in a cost-effective and energy-efficient manner, which involves using reconfigurable intelligent surfaces (RIS) to assist CF mMIMO systems. RIS is increasingly recognized as a forward-looking smart radio technology for advancing future 6G communications [5]. Consequently, the utilization of RIS-aided CF mMIMO systems can lead to improvements in channel capacity, reduced transmission power, enhanced transmission reliability, and expanded wireless coverage [6]–[10].

The utilization of joint precoding in RIS-aided CF mMIMO systems, as opposed to conventional precoding at the APs alone, involves the coordinated design of the beamforming matrix at the AP and the phase shifts of the RIS elements. This approach has been explored in recent research, such as the joint active-and-passive precoding framework proposed in [11] and the partially connected CF mMIMO framework presented in [12]. Additionally, efforts towards addressing practical implementation challenges of these techniques, such as the creation of less complex iterative algorithms, have been undertaken, as explored in [13]. Despite these advancements, several challenges persist, particularly in the application of computationally intensive algorithms and joint learning in practical scenarios, necessitating further resolution for real-world deployment.

As a crucial technology for future 6G-and-beyond wireless communication systems, machine learning/artificial intelligence holds the promise of resolving non-convex optimization problems that are mathematically insoluble [14]. Specifically, the authors in [15] proposed a meta reinforcement learning (meta-RL)-based computation offloading policy to optimize RIS phase shift. In [16], the authors introduced a distributed machine learning-based approach to optimize the transmit beamforming at the AP. The aforementioned methods solve the computation complexity problems, however, the acquisition of instantaneous global CSI still entails substantial front-haul overhead [17].

To address the challenges in the RIS-aided CF mMIMO system as mentioned above, inspired by multi-agent reinforcement learning (MARL), in this paper, we introduce an innovative MARL-based downlink design for joint precoding and phase shift to mitigate these challenges. The principal contributions of this paper are delineated as follows:

- We investigate an RIS-aided CF mMIMO network and formulate the optimization problem for joint precoding and phase shift to maximize the sum-SE. In contrast

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grants 2023YJS015 and 2022JBQY004, in part by National Natural Science Foundation of China under Grants 62221001, in part by Natural Science Foundation of Jiangsu Province, Major Project under Grant BK20212002, and in part by ZTE Industry-University-Institute Cooperation Funds under Grant No. IA20240319002. (*Corresponding author: Jiayi Zhang*).

Y. Zhu, E. Shi, Z. Liu, J. Zhang, and B. Ai are with the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, P. R. China. (e-mail: jiayizhang@bjtu.edu.cn).

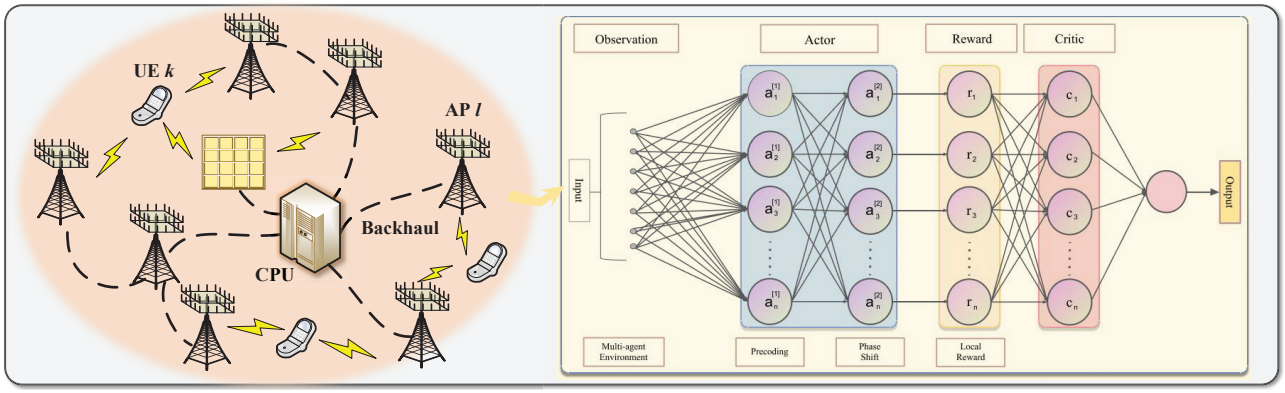

Fig. 1. The RIS-aided CF mMIMO system and the proposed MARL precoding network.

to centralized training centralized execution approach, such as alternating optimization (AO), we employ the method of centralized training and distributed execution of MARL to solve this problem.

Based on the proposed MARL algorithm, we design a two-layer network to address the joint optimization of AP precoding and RIS phase shift separately. Each AP in our approach only requires local CSI for precoding design, reducing the burden and overhead on the backhaul link. Additionally, we introduce fuzzy logic (FL) to enhance the convergence speed of MARL and further reduce computational complexity. The results demonstrate that our proposed algorithm outperforms AO-based precoding regarding SE performance in a limited training time.

Notation: The mathematical notation $(\cdot)^H$ is employed to denote the conjugate transpose operation. Boldface uppercase letters such as $\mathbf{X}$ are utilized to represent matrices, while boldface lowercase letters such as $\mathbf{x}$ are employed to denote column vectors. Furthermore, the complex Gaussian random variable x with variance σ^2 is represented by $x \sim \mathcal{CN}\left(0, \sigma^2\right)$.

II. System Model

In this section, we focus on an RIS-aided CF mMIMO system, as represented in Fig. 1, which utilizes several distributed APs and RISs to serve all UEs, simultaneously. For centralized control and training, a CPU is employed. All APs are connected to the CPU by optical cables or wireless fronthaul/backhaul links [18], [19]. This design enables the distributed APs to obtain important user-specific CSI and cooperatively service all UEs. The control of all RISs is overseen by the CPU, and facilitated by wired connections.

Specifically, we assume the network consists of L APs, K UEs, and R RISs. We assume that each AP and UE are equipped with M antennas, U antennas, respectively. Also, each RIS consists of N elements. We use the sets $\mathcal{N} = \{1, 2, \ldots, N\}$, $\mathcal{L} = \{1, 2, \ldots, L\}$, $\mathcal{K} = \{1, 2, \ldots, K\}$, and $\mathcal{R} = \{1, 2, \ldots, R\}$ to represent the index sets for RIS elements, APs, UEs, and RISs, respectively.

A. Channel Model

The utilization of RISs enables directional reflection, thereby structuring the channel between each AP and UE into distinct constituents. Specifically, the channel comprises an AP-UE link and R AP-RIS-UE links, with each AP-RIS-UE link further divisible into an AP-RIS link and a RIS-UE link. The architectural framework, which is facilitated by RISs, delineates the communication channels within wireless systems, offering enhanced control and optimization of signal propagation.

A phase shift matrix to the incident signal, followed by the transmission of the phase-shifted signal to the user, is applied to represent on the RISs. Consequently, the resulting equivalent channel, denoted as $\mathbf{h}_{l,k}^H$, originating from the l-th AP to the k-th UE, is represented as

$$\hat{\mathbf{h}}_{l,k}^H = \mathbf{H}_{l,k}^H + \sum_{r=1}^{R} \mathbf{F}_{r,k}^H \mathbf{\Theta}_r^H \mathbf{G}_{l,r}, \tag{1}$$

where $\mathbf{G}_{l,r} \in \mathbb{C}^{N \times M}$, $\mathbf{F}_{r,k}^H \in \mathbb{C}^{U \times N}$ denote the frequency-domain channel from the AP l to RIS r, and from RIS r to UE k,respectively; $\mathbf{\Theta}_r^H \in \mathbb{C}^{N \times N}$ denotes the phase shift matrix as the RIS r, which is written as

$$\mathbf{\Theta}_r^H \triangleq \operatorname{diag}(\theta_{r,1}, \ldots, \theta_{r,N}), \forall r \in \mathcal{R}, \tag{2}$$

where $\theta_{r,n} \in \mathcal{F}$. Note that $\mathcal{F}$ is the feasible set of the reflection coefficient at RIS. For simplicity but without loss of generality, here we assume $\mathcal{F}$ is the ideal case, i.e.,

$$\mathcal{F} \triangleq \{\theta_{r,n} | \theta_{r,n}| \leq 1\}, \forall r \in \mathcal{R}, \forall n \in \mathcal{N}. \tag{3}$$

Besides, $\mathbf{H}_{l,k}^H \in \mathbb{C}^{U \times M}$ denote the frequency-domain channel from the AP l to the UE k, which can be written as

$$\mathbf{H}_{l,k}^H = \beta_{l,k}^H |\mathbf{h}_{l,k}^H|^2, \tag{4}$$

where $\beta_{l,k}^H$ denotes the large-scale factor, $\mathbf{h}_{l,k}^H$ is the Rayleigh fading vector composed of the small-scale fading coefficients between AP l and UE k.

B. Transmitters and Receivers

Our proposed RIS-aided CF mMIMO system establishes synchronization among all APs, a prerequisite for facilitating coherent joint transmission to cater to all users. Let $\mathbf{s} \triangleq [s_1, s_2, \ldots, s_K]^T \in \mathbb{C}^K$ denote the vector of symbols, where each s_k corresponds to the symbol transmitted to the k-th user. It is ensured that the transmitted symbols adhere to power

normalization, implying that $\mathbb{E}\{ss^H\} = \mathbf{I}$, with $\mathbf{I}$ denoting the identity matrix.

In the downlink, the frequency-domain symbol s_k undergoes precoding using the precoding matrix $\mathbf{w}_{l,k} \in \mathbb{C}^M$ at the *l*-th AP. The initial precoding operation yields the precoded symbol $\mathbf{x}_l$ and can be mathematically represented as

$$\mathbf{x}_l = \sum_{k=1}^{K} \mathbf{w}_{l,k}\mathbf{s}_k. \tag{5}$$

Let's represent the baseband frequency-domain signal received by UE *k* as $\mathbf{y}_k \in \mathbb{C}^U$

$$\mathbf{y}_k = \sum_{l=1}^{L} \hat{\mathbf{h}}_{l,k}^H \mathbf{x}_l + \mathbf{z}_k^H. \tag{6}$$

C. Problem Formulation

Based on the system model above, the aim in this subsection is to enhance the overall SE gain realized over the network's operational duration. At first, the signal-to-interference-and-noise ratio (SINR) for the transmitted symbol s_k at UE *k* is calculated as

$$\gamma_k = \frac{|\sum_{l=1}^{L} \hat{\mathbf{h}}_{l,k}^H \mathbf{w}_l|^2}{\sum_{j\neq k} |\sum_{l=1}^{L} \hat{\mathbf{h}}_{l,k}^H \mathbf{w}_j|^2 + \sigma^2}. \tag{7}$$

Thereby, the SE of UE *k* R_k is given by

$$R_k = \log_2(1+\gamma_k). \tag{8}$$

Finally, the optimization problem of maximizing SE gain can be originally formulated as

$$\begin{aligned} \mathcal{P}^0 \max_{\mathbf{w}_k,\Theta} \text{sum-SE} &= \sum_{k=1}^{K} \log_2(1+\gamma_k), \\ s.t. \quad & \sum_{k=1}^{K} \|\mathbf{w}_k\|^2 \le P_{l,\max}, \quad \forall k \in K, l \in L, \\ & \theta_{r,n} \in [0, 2\pi], \quad \forall r \in R, n \in N, \end{aligned} \tag{9}$$

where the objective function pertains to an optimal problem with a multi-timescale horizon, $P_{l,max}$ denotes the maximum transmit power of the AP *l*, and θ_{rn} denotes the reflection coefficient at the RISs, respectively.

Given the complex characteristics of the non-convex objective function (9), the concurrent optimization of both the phase shift matrix and the precoding matrix presents a significant challenge. Nevertheless, drawing inspiration from MARL, we have proposed an innovative joint precoding network to tackle the optimization problem $\mathcal{P}^\circ$, as detailed in Section III.

III. Proposed Joint Precoding and Phase Shift Optimization Framework

A. Overview of the Framework

Within the context of a multi-agent architecture, every individual agent is constituted by two discrete components: an actor, which is responsible for the execution of actions, and a critic, which plays a pivotal role in the evaluation and refinement of the policy, respectively. MARL with centralized training decentralized execution (CTDE) has emerged as a viable alternative, streamlining centralized learning to a more computationally manageable extent. Consequently, the optimization problem denoted as $\mathcal{P}^0$ in equation (9) can be reformulated within the framework of MARL-CTDE as follows:

$$\begin{aligned} \mathcal{P}^1 \max_{\mathbf{w}_{lk},\Theta} \text{sum-SE} &= \sum_{k=1}^{K} \log_2\left(1+\frac{\left|\sum_{l=1}^{L} \hat{\mathbf{h}}_{lk,d}^{\mathrm{H}} \mathbf{w}_{lk}\right|^2}{\sum_{i=1,i\neq k}^{K}\left|\sum_{l=1}^{L} \hat{\mathbf{h}}_{lk}^{\mathrm{H}} \mathbf{w}_{li}\right|^2+\sigma^2}\right), \\ s.t. \quad & \sum_{k=1}^{K} \|\mathbf{w}_{lk}\|^2 \le P_{l,\max}, \quad \forall k \in K, l \in L, \\ & \theta_{r,n} \in [0, 2\pi], \quad \forall r \in R, n \in N. \end{aligned} \tag{10}$$

We transfer $\mathcal{P}^0$ to $\mathcal{P}^1$ for convenience in a MARL scenario. Each AP *l*, instead of UE *k*, is considered an agent, which is discussed in the following subsection.

B. Fuzzy Logic

Given the MARL policy in our network, it is evident that the RIS-aided CF mMIMO system, with a substantial number of APs and users, results in a large matrix calculation dimension and heightened complexity. Consequently, the conventional MARL necessitates simplification to ensure real-time interaction capability and scalability of the algorithms developed. Motivated by the integration of FL in a seminal work [20], we propose an innovative two-layer MARL-based downlink joint precoding method. This method strategically employs FL to formulate a correlation between fuzzy agents and entities, whose network is shown in Fig.1.

The precoding problem $\mathcal{P}^1$ and FL are described in this case, along with a MARL tuple $< \mathcal{S}^{(t)}, \mathcal{A}^{(t)}, r^{(t)} >$ at slot *t*. The state space $\mathcal{S}^{(t)} = \left(\mathcal{S}_1^{(t)}, \ldots, \mathcal{S}_n^{(t)}\right)$, action space $\mathcal{A}^{(t)} = \left(\mathcal{A}_1^{(t)}, \ldots, \mathcal{A}_n^{(t)}\right)$, and reward $r^{(t)} = \left(r_1^{(t)}, \ldots, r_n^{(t)}\right)$ are designed as follows.

1) Agent: We consider each AP as an agent.

2) State space: States are characterized as comprehensive representations of the entire system. To encompass the states of UEs dispersed across diverse locations, we employ an observational approach comprising both partial state variables and global state variables. In this framework, the agent is equipped to observe the relative positions of all UEs concerning all APs, denoted as $\boldsymbol{D}$. Specifically, local state is contemplated for AP *l* at slot *t*.

$$\mathcal{S}_l^{(t)} = \left(\mathbf{H}_{l,k}^{(t)}, \mathbf{F}_{r,k}^{(t)}, \mathbf{\Theta}_r^{(t)}, \mathbf{G}_{l,r}^{(t)}, \boldsymbol{D}_l, \mathbf{w}_l^{(t)}, \Theta^{(t)}, \gamma_{l,k}^{(t)}\right). \tag{11}$$

3) Initialization: Initially, every fuzzy agent's fuzzy state is denoted as $\hat{\mathcal{S}}^{(t)} = \left(\hat{\mathcal{S}}_1^{(t)}, \ldots, \hat{\mathcal{S}}_n^{(t)}\right)$, with each $\mathcal{S}_i^{(t)}$ being a random selection from the observed state, and n representing the total number of fuzzy agents. Subsequently, we divide each dimension of the state space into n unique fuzzy sets. For any given j-th dimension, the fuzzy state set is expressed as $\left(\hat{s}_{1,j}^{(t)}, \hat{s}_{2,j}^{(t)}, \ldots, \hat{s}_{K,j}^{(t)}\right)$. The corresponding membership func-

tion is denoted as $\xi^{(t)}_{\hat{s}_{i,j}(s)} = \mathbf{exp}(-\frac{1}{d_a * n}|s^{(t)} - \hat{s}^{(t)}_{i,j}|)$,where d_a represents the dimension of the action space [21].

4) Fuzzy Action space: Each fuzzy agent in a fuzzy system is assigned a policy based on the perceived fuzzy state, which is represented as $\mathcal{S}^{(t)}$. Subsequently, defuzzification is employed to establish a mapping from the fuzzy action $\hat{\mathcal{A}}^{(t)} = \left(\hat{\mathcal{A}}^{(t)}_1, \ldots, \hat{\mathcal{A}}^{(t)}_n\right)$ to the specific action $\mathcal{A}^{(t)}$. The mapping relationship between the p-th agent and the i-th fuzzy agent is denoted as $\Xi^{(t)}_{i,p} = \prod_{i=j}^{d_a} \xi^{(t)}_{\hat{s}_{i,j}(s_{k,j})}$. Here we define the relationship as $\mathcal{A}^{(t)}_p = \sum_{i=1}^{m} \bar{\Xi}^{(t)}_{i,p} \times \hat{\mathcal{A}}^{(t)}_i$, where $\bar{\Xi}_{i,p}$ represents the normalized mapping relationship.

5) Reward space: Subsequent to the reception of the specific action $\mathcal{A}^{(t)}$ by the agents, the corresponding reward r is ascertained in accordance with the predefined reward function. In the context of interacting with the environment, it is essential to employ fuzzy agents rather than entities. This necessitates the process of fuzzification to derive the fuzzy reward $\hat{r}^{(t)} = \left(\hat{r}^{(t)}_1, \ldots, \hat{r}^{(t)}_n\right)$ within the framework of reinforcement learning. The fuzzy reward can be expressed as $\hat{r}^{(t)}_i = \sum_{k=1}^{K} \bar{\Xi}^{(t)}_{i,k} \times r^{(t)}_k$. Therefore, the incorporation of fuzzification is imperative for the successful completion of the reinforcement learning model.

C. FL-MARL Algorithm

In the context of FL-MARL, individual fuzzy agents independently calculate the policy gradient for their respective local actor networks, utilizing the collective abstract state and action as a basis. Additionally, the objective function for the i-th policy, denoted as π_i, can be formulated as $L(\pi_i) = \sum_{\hat{\mathcal{S}}^{(t)}_i} p_\pi\left(\hat{\mathcal{S}}^{(t)}_i\right) \sum_{\hat{\mathcal{A}}^{(t)}_i} \pi\left(\hat{\mathcal{A}}^{(t)}_i \mid \hat{\mathcal{S}}^{(t)}_i\right) \hat{r}^{(t)}_i$.

The i-th fuzzy reward, symbolized as r_i, is correspondingly linked with a universal fuzzy action, denoted as $\mathcal{A}$, and a state, represented as $\mathcal{S}$. This association results in a consolidated global behavior value, $Q_\pi\left(\hat{\mathcal{S}}^{(t)}, \hat{\mathcal{A}}^{(t)}\right)$, which is the output of the i-th critic network's computation. In the context of π_i, the policy gradient of the local actor network can be articulated as follows:

$$\Delta J(\theta_{\pi_i}) = \sum^{\hat{\mathcal{A}}^{(t)}_i} Q_\pi\left(\hat{\mathcal{S}}^{(t)}, \hat{\mathcal{A}}^{(t)}\right) \Delta\pi_i\left(\hat{\mathcal{A}}^{(t)}_i | \hat{\mathcal{S}}^{(t)}_i; \theta_{\pi_i}\right), \tag{12}$$

where $\Delta\pi_i\left(\hat{\mathcal{A}}^{(t)}_i \mid \hat{\mathcal{S}}^{(t)}_i; \theta_{\pi_i}\right)$ is the output by the local policy network.

In accordance with Algorithm 1, the soft update procedure is implemented concurrently with the existing network configuration. The target actor network undergoes modification according to the expression $\theta_{\pi'_i} \leftarrow \tau\theta_{\pi'_i} + (1-\tau)\theta_{\pi_i}$, whereas the target critic network is adjusted based on the formula $\theta_{Q_{\pi'}} \leftarrow \tau\theta_{Q_{\pi'}} + (1-\tau)\theta_{Q_\pi}$.

D. Complexity comparison

We compare the computational complexity of different algorithms. We assume that Q_a and Q_c denote the output size of the a-th and c-th layer or the input size of the next layer, and Q_{SE} represents the computational complexity of calculating SE expressions, respectively. Note that as observed from Table I, the computational complexity of MADDPG exhibits an exponential increase with the number of APs L increasing. In contrast, upon integrating FL, the computational complexity increases linearly with the number of APs L. Correspondingly, the introduction of FL makes the computational complexity of FL-MADDPG linearly related to the number of fuzzy agents N_F, thus reducing the need for high backhaul capacity for $\left(\frac{N_F}{L}\right)^2$.

Algorithm 1 FL-MARL Algorithm for Maximizing sum-SE

Initialize AP agent states $\mathcal{S}^{(t)}_1, \ldots, \mathcal{S}^{(t)}_n$ by randomly sampling fuzzy agent states: $\hat{\mathcal{S}}^{(t)}_1, \ldots, \hat{\mathcal{S}}^{(t)}_n$
count = 0
while count $\leq$ N **do**
  Evaluate the network actor to decide the downlink precoding and phase shift design: $\hat{\mathcal{A}}^{(t)}_i = \pi_i\left(\mathcal{S}^{(t)}_i\right)$
  Calculate actual actions $\mathcal{A}^{(t)}_i = \sum_{i=1}^{m} \bar{\Xi}^{(t)}_{i,k} \times \hat{\mathcal{A}}^{(t)}_i$
  Get actual rewards r_i with reward function
  Use fuzz function: $\hat{r}^{(t)}_i = \sum_{k=1}^{K} \bar{\Xi}^{(t)}_{i,k} \times r^{(t)}_k$ to calculate fuzzy rewards $\hat{r}_i$
  Update environment
  Get next actual states $\mathcal{S}^{(t)}_i$
  Get next fuzzy states $\hat{\mathcal{S}}^{(t)}_1$ by fuzz function: $\hat{\mathcal{S}}^{(t)}_i = \sum_{k=1}^{K} \bar{\Xi}^{(t)}_{i,k} \times \hat{\mathcal{S}}^{(t)}_k$
  Update membership function with $\xi^{(t+1)}_{\hat{s}_{i,j}(\hat{s}_{k,j})}$
  Store fuzzy experience $< \mathcal{S}^{(t)}, \mathcal{A}^{(t)}, r^{(t)} >$ in replay buffer $\mathcal{D}_i$
  for each training step **do**
    Randomly sample a mini-batch of $\mathcal{B}_i$ transitions uniformly from $\mathcal{D}_i$
    Update weights of joint precoding and phase shift critic network
    Calculate policy gradient of the two layer actor network $\Delta J(\theta_{Q_\pi})$ and update target network
  count += 1

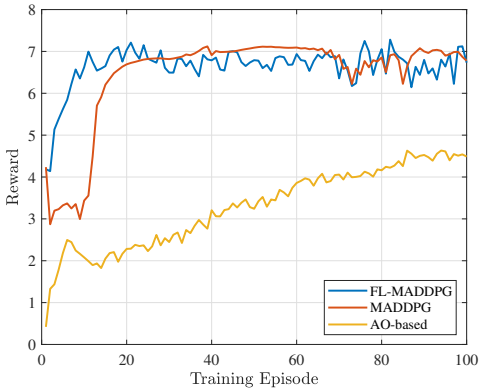

Fig. 2. Average reward against the training step with $step$ = 100, L = 4, K = 4, R = 4, M = 8, U = 1, and N = 16.

TABLE I
COMPARISON OF COMPUTATIONAL COMPLEXITY.

Parameters	Computational Complexity
MADDPG	$\left\{\mathcal{O}(L^2MKN^2\sum_{a=1}^{A_L}Q_a^2+L^2MK\sum_{c=1}^{C_L}Q_c^2)\right\}\times\left\{\mathcal{O}(L^2MKN^2\sum_{a=1}^{A_H}Q_a^2+LN\sum_{c=1}^{C_H}Q_c^2+L^3Q_{SE})\right\}$
FL-MADDPG	$\left\{\mathcal{O}(LN_FMKN^2\sum_{a=1}^{A_L}Q_a^2+LN_FMK\sum_{c=1}^{C_L}Q_c^2)\right\}\times\left\{\mathcal{O}(LN_FMKN^2\sum_{a=1}^{A_H}Q_a^2+LN\sum_{c=1}^{C_H}Q_c^2+L^2N_FQ_{SE})\right\}$

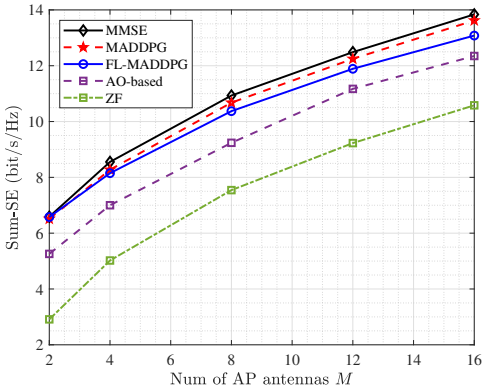

Fig. 3. Sum-SE against the number of AP antennas with L = 4, K = 4, R = 4, U = 1, and N = 16.

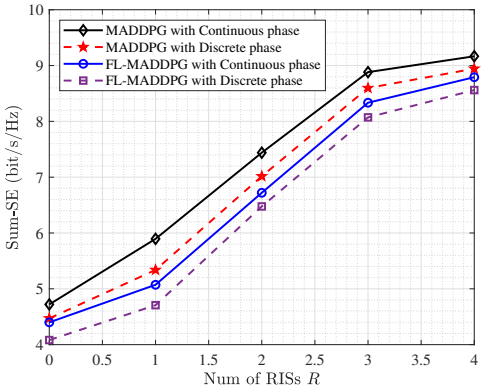

Fig. 4. Sum-SE against the number of RISs with L = 4, K = 4, M = 8, U = 1 and N = 16.

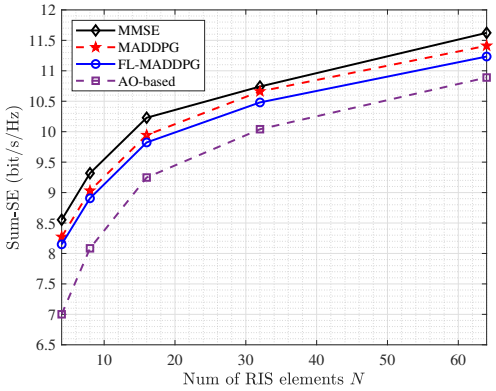

Fig. 5. Sum-SE against the number of RIS elements with L = 4, K = 4, R = 4, M = 8, and U = 1.

TABLE II
THE MODEL STRUCTURE AND EXPERIMENTAL DETAILS.

Parameters	Size
Hidden layer of AP Precoding	512, Leaky Relu (0.01)
Hidden layer of RIS Phase Shift	256, Leaky Relu (0.01)
Mini-batch	32
Discounted factor γ^{p} and γ^{f}	0.99
Experience pool size $\mathcal{D}^{\mathrm{p}}$ and $\mathcal{D}^{\mathrm{f}}$	4096 and 2048
Soft update rate τ^{p} and τ^{f}	0.0001 and 0.001

IV. SIMULATION RESULTS

A. Simulation Setup

In the proposed RIS-aided cell-free network simulation, we consider a 50 m $\times$ 50 m region served by a cell-free network with all APs simultaneously serving all UEs. Each AP divides the area into four equal squares, and UEs are randomly deployed within these squares. To enhance network capacity, RISs are strategically placed at the center of each of the four equal squares that the APs divide the area into. We assume that the maximum transmit power for APs is $P_{l,\max} = 0$ dBm, and the initial number of antennas per AP is $M = 4$, with a noise power of $\delta^2 = -96$ dBm. Considering the limited antennas and low transmit power of APs in cell-free networks, we adopt the channel model from [22]. For the performance enhancement of sum SE, we consider a proper experience pool size in the simulation to improve the generalization ability of the model. The experimental details of MARL are given in Table II.

B. Convergence of the proposed algorithm

To demonstrate the convergence of the proposed algorithms, we present a plot of the reward as a function of training time in Fig. 2. The outcomes depicted in Fig. 2 show that the MARL method can converge faster in the same number of training steps than the traditional AO-based method [23]. It is worth noting that in the initial stage of algorithm training, MADDPG showed a more stable upward trend compared with FL-MADDPG. However, FL-MADDPG has a faster convergence rate, about 30% faster than MADDPG, allowing FL-MADDPG to reach a stable state earlier due to its combination of fuzzy logic mechanisms to complete the agent-to-fuzzy agent mapping. Therefore, FL-MADDPG is more stable in comparison, demonstrating that the integration of FL into MADDPG yields substantial savings in computing resources.

C. Impact of key system parameters

We evaluate the sum-SE of the proposed RIS-aided cell-free network in this subsection.

1) SE against the number of antennas per AP: We illustrate the average sum-SE in relation to the number of AP antennas in Fig. 3. The figure reveals a notable increase in the sum-SE across all instances as the number of AP antennas increases. Notably, FL-MADDPG demonstrates a performance closely aligned with MADDPG, which is near MMSE. SE has a 42% gain compared to ZF and 18% over the AO-based algorithm, signifying that our proposed FL-MADDPG framework can approximate the performance of MADDPG with a relatively fast convergence.

2) SE against the number of RISs: We depict the sum-SE relative to the number of RISs in Fig. 4. It is easy to find that with the increase in RIS, the sum-SE shows an increasing trend. The gap between the continuous phase and the discrete phase gradually decreases with the increase in the number of RISs, which is because there are blind areas in CF-mMIMO. These blind areas need RIS to provide and enhance communication services. It is worth noting that the percentage gap between FL-MADDPG and MADDPG under continuous is smaller than the gap under discrete, indicating that FL-MADDPG is more suitable for a broader continuous scenario.

3) SE against the number of RIS elements: We depict the sum-SE relative to the number of RIS elements in Fig. 5. As the number of RIS elements increases, the sum SE of the RIS-aided cell-free network shows a notable improvement. Compared with alternate optimization, the MARL we adopted has more than 10% performance improvement, which can quickly approach the theoretical value in a limited training time. However, this enhancement is accompanied by heightened implementation complexity of the RIS. Furthermore, this state underscores the robustness of our proposed algorithm across a broader spectrum of application scenarios, approaching performance levels indicative of optimality.

V. Conclusions

In this paper, we investigated the maximization of downlink SE in a RIS-aided CF mMIMO system through joint precoding and phase shift design. To achieve this goal, we proposed a MARL-based method incorporating fuzzy logic. The method presented leverages parallel computing to diminish computational time, rendering it highly suitable for deployment in expansive networks. Our simulation findings substantiate that the MARL method, underpinned by fuzzy logic, significantly curtails computational complexity by approximately 42%, thereby enhancing reliability in practical settings relative to traditional AO-based algorithms in a limited training time. In future work, it is interesting to investigate the simultaneous transmitting and reflecting (STAR)-RIS-aided CF mMIMO systems and design joint precoding with imperfect CSI to enhance the network.

References

[1] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1637–1660, Aug. 2020.

[2] N. T. Nguyen, V.-D. Nguyen, H. V. Nguyen, H. Q. Ngo, S. Chatzinotas, and M. Juntti, "Spectral efficiency analysis of hybrid relay-reflecting intelligent surface-assisted cell-free massive MIMO systems," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 5, pp. 3397–3416, May. 2023.

[3] S. Chen, J. Zhang, E. Björnson, J. Zhang, and B. Ai, "Structured massive access for scalable cell-free massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1086–1100, Apr. 2021.

[4] E. Shi, J. Zhang, D. W. K. Ng, and B. Ai, "Uplink performance of RIS-aided cell-free massive MIMO system with electromagnetic interference," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2431–2445, Aug. 2023.

[5] Y. Liu, X. Mu, J. Xu, R. Schober, Y. Hao, H. V. Poor, and L. Hanzo, "Star: Simultaneous transmission and reflection for 360° coverage by intelligent surfaces," *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 102–109, Dec. 2021.

[6] Y. Zhang, W. Xia, H. Zhao, G. Zheng, S. Lambotharan, and L. Yang, "Performance analysis of RIS-assisted cell-free massive MIMO systems with transceiver hardware impairments," *IEEE Trans. Commun.*, vol. 71, no. 12, pp. 7258–7272, Dec. 2023.

[7] J. Dai, J. Ge, K. Zhi, C. Pan, Z. Zhang, J. Wang, and X. You, "Two-timescale transmission design for RIS-aided cell-free massive MIMO systems," *IEEE Trans. Wirel. Commun.*, pp. 1–1, Early Access 2023.

[8] J. Zhang, H. Liu, Q. Wu, Y. Jin, Y. Chen, B. Ai, S. Jin, and T. J. Cui, "RIS-aided next-generation high-speed train communications: Challenges, solutions, and future directions," *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 145–151, Dec. 2021.

[9] X. Ma, X. Lei, P. T. Mathiopoulos, and D. B. d. Costa, "Active STAR-RIS aided cell-free massive MIMO: A performance study," *IEEE Trans. Veh. Technol.*, pp. 1–6, 2023.

[10] B. Li, Y. Hu, Z. Dong, E. Panayirci, H. Jiang, and Q. Wu, "Energy-efficient design for reconfigurable intelligent surface aided cell-free ultra dense hetnets," *IEEE Trans. Veh. Technol.*, pp. 1–17, 2023.

[11] Z. Zhang and L. Dai, "A joint precoding framework for wideband reconfigurable intelligent surface-aided cell-free network," *IEEE Trans. Signal Process.*, vol. 69, no. 6, pp. 4085–4101, Jun. 2021.

[12] X. Ma, D. Zhang, M. Xiao, C. Huang, and Z. Chen, "Cooperative beamforming for RIS-aided cell-free massive MIMO networks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 7243–7258, Mar. 2023.

[13] J. Yao, J. Xu, W. Xu, D. W. K. Ng, C. Yuen, and X. You, "Robust beamforming design for RIS-aided cell-free systems with CSI uncertainties and capacity-limited backhaul," *IEEE Trans. Commun.*, vol. 71, no. 8, pp. 4636–4649, Aug. 2023.

[14] L. Dai and X. Wei, "Distributed machine learning based downlink channel estimation for RIS assisted wireless communications," *IEEE Trans. Commun.*, vol. 70, no. 7, pp. 4900–4909, Jul. 2022.

[15] Y. Lu, Y. Jiang, L. Zhang, M. Bennis, D. Niyato, and X. You, "Meta reinforcement learning-based computation offloading in RIS-aided MEC-enabled cell-free RAN," in *Proc. IEEE ICC*, 2023, pp. 5370–5376.

[16] C. Chen, S. Xu, J. Zhang, and J. Zhang, "A distributed machine learning-based approach for IRS-enhanced cell-free MIMO networks," *IEEE Trans. Wireless Commun., Early Access*, 2023.

[17] Z. Liu, J. Zhang, Z. Liu, H. Du, Z. Wang, D. Niyato, M. Guizani, and B. Ai, "Cell-free XL-MIMO meets multi-agent reinforcement learning: Architectures, challenges, and future directions," *IEEE Wireless Commun.*, to appear, 2024.

[18] J. Zhang, J. Zhang, D. W. K. Ng, S. Jin, and B. Ai, "Improving sum-rate of cell-free massive MIMO with expanded compute-and-forward," *IEEE Trans. Signal Process.*, vol. 70, pp. 202–215, 2021.

[19] E. Shi, J. Zhang, H. Du, B. Ai, C. Yuen, D. Niyato, K. B. Letaief, *et al.*, "RIS-aided cell-free massive MIMO systems for 6G: Fundamentals, system design, and applications," *arXiv:2310.00263*, 2023.

[20] J. Li, H. Shi, and K.-S. Hwang, "Using fuzzy logic to learn abstract policies in large-scale multiagent reinforcement learning," *IEEE Trans. Fuzzy Systems*, vol. 30, no. 12, pp. 5211–5224, Apr. 2022.

[21] Z. Liu, J. Zhang, Z. Liu, H. Xiao, and B. Ai, "Double-layer power control for mobile cell-free XL-MIMO with multi-agent reinforcement learning," *IEEE Trans. Wireless Commun.*, to appear, 2023.

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

[23] X. Gan, C. Zhong, C. Huang, Z. Yang, and Z. Zhang, "Multiple riss assisted cell-free networks with two-timescale csi: Performance analysis and system design," *IEEE Trans. Commun.*, vol. 70, no. 11, pp. 7696–7710, Sep. 2022.