AI-Based Fluid Antenna Design for Client Selection in Over-the-Air Federated Learning

Mohsen Ahmadzadeh, Saeid Pakravan, *Graduate Student Member, IEEE*, Ghosheh Abed Hodtani, Ming Zeng, *Member, IEEE*, Qiang Ye, *Senior Member, IEEE*, Jean-Yves Chouinard, *Life Senior Member, IEEE*, and Leslie A. Rusch, *Life Fellow, IEEE*

***Abstract*—This article proposes an innovative approach to improve over-the-air federated learning (OTA-FL) systems by integrating fluid antennas (FAs) at the access point. By exploiting the mobility of FAs, we aim to increase the correlation among the users' channels, thereby improving the learning performance. We analyze the performance of over-the-air computation and the convergence behavior of the OTA-FL system, highlighting the benefits of FAs. Since the learning performance improves as more devices participate in the federated learning (FL) aggregation, we formulate a nonconvex optimization problem that maximizes the number of selected users by jointly optimizing FA positions and the beamforming vector, coupled with a user selection policy subject to a mean-squared error constraint. To address environmental dynamics, we describe the problem as a Markov decision process and develop a long–short-term memory (LSTM)-based algorithm for efficient decision-making. Simulation results demonstrate that the proposed FA-assisted OTA-FL framework significantly outperforms conventional setups, achieving higher user selection rates and improved learning performance compared to existing benchmarks.**

***Index Terms*—Deep reinforcement learning (DRL), fluid antenna (FA), optimality gap, over-the-air federated learning (OTA-FL).**

I. Introduction

FEDERATED learning (FL) has emerged as a promising approach in communication systems due to its decentralized structure and robust privacy preservation [1], [2]. By leveraging the computational capabilities of edge devices, FL facilitates collaborative training of a global model without the need to share local data. This approach is particularly advantageous for mobile Internet of Things (IoT) applications [3], [4], [5]. However, the practical deployment of FL faces significant challenges, notably high communication latency and associated costs, which can impede its efficiency and scalability in real-world implementations.

To mitigate these issues, over-the-air (OTA) computation has emerged as an effective solution for model aggregation in FL. By exploiting the inherent superposition property of wireless channels, OTA computation enables simultaneous data transmission, thereby significantly reducing aggregation overhead [6]. However, the performance of model aggregation in over-the-air FL (OTA-FL) faces challenges due to unfavorable wireless conditions, particularly in large-scale mobile IoT environments.

Recent studies have explored reconfigurable intelligent surfaces (RISs) as a means to improve OTA-FL reliability by manipulating wireless propagation environments [7], [8], [9], [10]. RIS employ passive reflecting elements with adjustable reflection coefficients to steer signals toward desired receivers, enhancing channel conditions and transmission performance [11], [12]. However, their static deployment and reliance on environmental factors limit adaptability in highly dynamic and mobile IoT scenarios, motivating the need for more flexible solutions. Furthermore, advanced beamforming techniques have been investigated to exploit spatial degrees of freedom (DoF) for improved signal reception [13]. However, these techniques are also constrained by the fixed positions of the receiver antennas, which limits the adaptability in dynamic environments.

To address these limitations, fluid antennas (FAs) have emerged as a promising enhancement for OTA-FL systems. In contrast to conventional fixed position antennas (FPAs), FAs possess the distinctive ability to dynamically adjust their spatial locations, thereby enabling real-time manipulation of wireless channel characteristics [14], [15], [16]. This inherent adaptability is particularly beneficial in mobile IoT environments, where rapid fluctuations in channel conditions often arise due to user mobility and environmental dynamics [17]. Existing studies have demonstrated the superior performance of FAs over FPAs in various wireless communication paradigms, including multiuser uplink transmissions [18], [19], mobile edge computing [20], and covert communications [21]. By introducing additional DoFs through position adaptability, FAs unlock new opportunities to enhance the performance of both OTA computation and FL systems [22], [23], [24], [25], [26], [27], [28]. For

Received 15 July 2025; accepted 26 July 2025. Date of publication 4 August 2025; date of current version 9 October 2025. This work was supported in part by the Canada Natural Science and Engineering Research Council under Grant RGPIN-2021-02636 and Grant CRC-2022-00115, and in part by FRQNT under Grant 341270. *(Corresponding author: Ming Zeng.)*

Mohsen Ahmadzadeh and Ghosheh Abed Hodtani are with the Department of Electric and Computer Engineering, Ferdowsi University, Mashhad 9177948974, Iran (e-mail: m.ahmadzadehbolghan@mail.um.ac.ir; hodtani@um.ac.ir).

Saeid Pakravan, Ming Zeng, Jean-Yves Chouinard, and Leslie A. Rusch are with the Department of Electric and Computer Engineering, Laval University, Quebec City, QC G1V 0A6, Canada (e-mail: saeid.pakravan.1@ulaval.ca; ming.zeng@gel.ulaval.ca; jean-yves.chouinard@gel.ulaval.ca; leslie.rusch@gel.ulaval.ca).

Qiang Ye is with the Department of Electrical and Software Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada (e-mail: qiang.ye@ucalgary.ca).

Digital Object Identifier 10.1109/JIOT.2025.3594141

example, [22] investigates joint transceiver design and antenna positioning to minimize mean-squared error (MSE), while [23] addresses robust resource allocation strategies under channel uncertainty. In [24], 2-D movable antennas (MAs) are deployed at the access point (AP) to reduce aggregation errors during wireless data collection. Similarly, [25] investigates the minimization of MSE under constrained communication rates in FA-aided networks considering dynamic operational scenarios. Reference [26] considers unmanned aerial vehicle-assisted MAs at the AP to achieve MSE minimization, whereas [27] focuses on FA-enhanced OTA-FL systems with fixed client participation to reduce the optimality gap in federated training. Moreover, [28] proposes a port selection mechanism for FA-equipped devices to accelerate the convergence of distributed learning processes. Despite these advancements, the comprehensive integration of FAs into OTA-FL systems, particularly for optimizing user selection policies, requires further investigation.

In this article, we integrate FAs with OTA-FL to modify channel gains to maximize the number of selected devices while ensuring that the MSE remains within a predefined threshold. First, we evaluate the impact of FAs on OTA-FL through convergence analysis, focusing on the MSE of model aggregation and optimizing power allocation to minimize errors. Our analysis highlights the critical role of the correlation between the beamforming vector and each device's channel gain in reducing the MSE. Then, we formulate a nonconvex optimization problem to design FA systems in dynamic environments. The objective is to jointly optimize the beamforming vector, antenna positioning, and user selection policy to minimize the optimality gap by maximizing device participation, while adhering to the MSE constraint. To tackle the complexity of this problem, we reformulate it as a Markov decision process (MDP) and employ deep reinforcement learning (DRL) techniques that are well-suited for dynamic environments. Specifically, we utilize the long–short-term memory deep deterministic policy gradient (LSTM-DDPG) algorithm to capture temporal correlations and enable real-time decision-making [29]. Finally, we conduct extensive simulations using both synthetic and real-world datasets to evaluate the proposed framework. The performance of the FA-assisted OTA-FL system is benchmarked against FPA systems, while the LSTM-DDPG algorithm is compared with other DRL approaches, such as soft actor–critic (SAC) and DDPG. Simulation results demonstrate that the proposed FA-assisted OTA-FL system significantly outperforms FPA-based implementations, and the LSTM-DDPG algorithm achieves superior performance compared to other DRL algorithms.

Notations: Italicized letters denote scalars, while boldface letters represent vectors. The transpose and conjugate transpose operations are denoted by $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{H}$, respectively. $\mathbb{E}[\cdot]$ represents the expectation operation, $|\cdot|$ denotes the magnitude of a scalar or the cardinality of a set, and $\|\cdot\|$ signifies the Euclidean norm of a vector.

II. System Model

We consider an OTA-FL framework comprising K single-antenna user equipment (UE) devices, each labeled as UE_k, where $k \in \mathcal{K} \triangleq \{1, 2, \ldots, K\}$. These devices are uniformly distributed within a designated area and dynamically collect local data samples to collaboratively train a global model at an AP equipped with FAs.

A. FL Model

FL aims to minimize the global loss function $F(\boldsymbol{w}_t)$, which is represented as

$$\min_{\boldsymbol{w}_t} F(\boldsymbol{w}_t) = \min_{\boldsymbol{w}_t} \frac{1}{K} \sum_{k=1}^{K} F_k(\boldsymbol{w}_t), \tag{1}$$

where $\boldsymbol{w}_t \in \mathbb{R}^d$ denotes the d-dimensional parameters of the learnable model. The local loss function associated with the kth device is expressed as

$$F_k(\boldsymbol{w}_t) = \frac{1}{|\mathcal{D}_k|} \sum_{j=1}^{|\mathcal{D}_k|} f\big(\boldsymbol{w}_t; d_{k,j}\big), \tag{2}$$

where $f(\cdot;\cdot)$ represents the loss function, $d_{k,j}$ indicates the jth data sample from the local dataset $\mathcal{D}_k$ of the kth device, and $|\mathcal{D}_k|$ denotes the number of samples in that dataset.

To solve the minimization problem outlined in (1), we employ the federated averaging (FedAvg) algorithm, which efficiently converges to the optimal solution while preserving device data privacy. The implementation of FedAvg involves a series of sequential steps as follows.

1) *User Selection:* The AP initiates the process by selecting a subset of UEs, denoted as $\mathcal{S}_t \subset \mathcal{K}$, and subsequently broadcasts the current global model $\boldsymbol{w}_t$ to these selected UEs.
2) *Local Training:* Each selected UE performs local model updates using stochastic gradient descent (SGD). The update is expressed as $\boldsymbol{w}_{k,t} = \boldsymbol{w}_t - \gamma \nabla \hat{F}_k(\boldsymbol{w}_t; \zeta_k)$, where $\gamma \in (0, 1)$ is the learning rate, and $\zeta_k \subset \mathcal{D}_k$ represents a randomly selected minibatch from the UE's local dataset. $\hat{F}_k(\boldsymbol{w}_t; \zeta_k)$ is an estimate of the local loss function, which is computed as
$$\hat{F}_k(\boldsymbol{w}_t; \zeta_k) = \frac{1}{|\zeta_k|} \sum_{j \in \zeta_k} f\big(\boldsymbol{w}_t; D_{k,j}\big) \tag{3}$$
where $f(\boldsymbol{w}_t, d_{k,j})$ indicates the sample-wise loss function associated with the data sample $D_{k,j}$.
3) *Model Aggregation:* Upon completing local training, each selected UE transmits its updated local model back to the AP. The AP then aggregates the local models by averaging
$$\boldsymbol{w}_{t+1} = \frac{1}{|\mathcal{S}_t|} \sum_{k \in \mathcal{S}_t} \boldsymbol{w}_{k,t}. \tag{4}$$

This process is iterated until the predefined maximum number of iterations is reached, ensuring continuous improvement of the global model.

B. Communication Model

We focus on the upload phase of the OTA-FL system, where each UE concurrently sends its locally trained model weights to the AP. The AP is equipped with N FAs arranged

in a linear array, which can dynamically reposition themselves along a 1-D line segment of length X. The position of each FA is constrained within the interval $[0, X]$, with a minimum distance X_0 maintained between adjacent FAs to avoid antenna coupling. The positions of the N FAs are represented by the vector $\mathbf{x} = [x_1, x_2, \ldots, x_N]^{\mathrm{T}}$, satisfying $x_1 < x_2 < \cdots < x_N$.

While the rayleigh fading model is widely adopted to characterize environments with rich scattering, it may fall short in scenarios representative of next-generation wireless systems, where strong line-of-sight (LoS) components often coexist with limited scattering [30], [31]. To accurately capture both deterministic LoS and non-LoS (NLoS) components, we adopt the rician fading model. Accordingly, the channel vector between the kth UE and the AP, denoted by $\mathbf{h}_k[x] \in \mathbb{C}^{N\times 1}$, is expressed as [30]

$$\begin{aligned}\mathbf{h}_k[\mathbf{x}] &= \sqrt{\frac{A_N d_k^{-\alpha_N}}{K_r+1}}\left[\sum_{i=1}^{N_p} e^{j\frac{2\pi}{\lambda}x_1\cos(\phi_{k,i})}, \ldots, \sum_{i=1}^{N_p} e^{j\frac{2\pi}{\lambda}x_N\cos(\phi_{k,i})}\right]^{\mathrm{T}} \\ &\quad + \sqrt{\frac{A_L d_k^{-\alpha_L} K_r}{K_r+1}}\left[e^{j\frac{2\pi}{\lambda}x_1\cos(\phi_k)}, \ldots, e^{j\frac{2\pi}{\lambda}x_N\cos(\phi_k)}\right]^{\mathrm{T}} \\ &= \sqrt{\frac{A_N d_k^{-\alpha_N}}{K_r+1}}\mathbf{h}_k^{\mathrm{NLoS}}[\mathbf{x}] + \sqrt{\frac{A_L d_k^{-\alpha_L} K_r}{K_r+1}}\mathbf{h}_k^{\mathrm{LoS}}[\mathbf{x}]. \end{aligned} \tag{5}$$

Here, K_r denotes the rician factor, and d_k is the distance between the FA system and the kth UE. The parameter N_p represents the number of scattered NLoS paths. The coefficients A_L and A_N indicate the reference path loss for the LoS and NLoS components, respectively, while α_L and α_N are the corresponding path loss exponents. The wavelength is denoted by λ, and ϕ_k and $\phi_{k,i}$ represent the angles of arrival (AoAs) for the specular LoS path and the ith scattered NLoS path, respectively.

For sufficiently large numbers of scattered paths (i.e., $N_p \to \infty$), the aggregate NLoS components can be approximated as zero-mean, circularly symmetric complex Gaussian random variables due to the central limit theorem, resulting in rayleigh-distributed channel amplitudes [30]. Moreover, due to the flexible placement of FA elements in close proximity, spatial correlation may arise in the NLoS channel components. This phenomenon, commonly observed in rich 2-D isotropic scattering environments, can be modeled using established statistical channel models, such as the Jakes' model [32], [33], which captures the spatial dependence of received signal fluctuations. It is further assumed that each UE transmits its model parameters from a stationary position within a predefined area [3]. Given the substantial difference between the signal path length and the movement range of the FAs, the far-field condition is assumed to hold between the AP and the UEs. Consequently, ϕ_k and d_k are treated as constant parameters during transmission in each time slot [18], [20].

During the tth training round, the AP receives the local model parameters transmitted by the selected UEs as

$$\mathbf{y}_t = \sum_{k\in\mathcal{S}_t} p_{k,t}\mathbf{h}_k[\mathbf{x}_t]\boldsymbol{w}_{k,t} + \mathbf{z}_t \tag{6}$$

where $p_{k,t}$ is the transmission power coefficient for the kth UE, and $\mathbf{z}_t \in \mathbb{C}^{N\times d}$ is an additive white Gaussian noise (AWGN) matrix with elements that are distributed as $\mathcal{CN}(0, \sigma^2)$. We assume that the transmission power allocated to each UE_k remains within the maximum limit $p_{\max}$, as specified in [13]. That is,

$$\frac{1}{d}p_{k,t}^2\mathbb{E}\left[\|\boldsymbol{w}_{k,t}\|^2\right] \le p_{\max} \quad \forall k \in \mathcal{K}. \tag{7}$$

The aggregated model parameter vector $\hat{\boldsymbol{w}}_{t+1}$ at the AP is computed as

$$\begin{aligned}\hat{\boldsymbol{w}}_{t+1} &= \frac{1}{|\mathcal{S}_t|}\left(\frac{1}{\sqrt{\eta_t}}\boldsymbol{m}_t^H\mathbf{y}_t\right) \\ &= \frac{1}{|\mathcal{S}_t|}\left(\sum_{k\in\mathcal{K}}\frac{1}{\sqrt{\eta_t}}\boldsymbol{m}_t^H p_{k,t}\mathbf{h}_k[\mathbf{x}_t]\boldsymbol{w}_{k,t} + \frac{\boldsymbol{m}_t^H\mathbf{z_t}}{\sqrt{\eta_t}}\right)\end{aligned} \tag{8}$$

where, $\boldsymbol{m}_t \in \mathbb{C}^{N\times 1}$ is the beamforming vector at the AP, and η_t is the amplitude scaling factor.

III. Performance Analysis

In this section, we evaluate the impact of FAs on OTA-FL by analyzing the MSE of model aggregation and the optimality gap. The primary objective of OTA computation is to minimize the distortion between the estimated and target received signal aggregations. This distortion is quantitatively assessed through the MSE at time step t, defined as

$$\begin{aligned}\mathrm{MSE}_t &= E\left\{\left|\boldsymbol{w}_{t+1} - \hat{\boldsymbol{w}}_{t+1}\right|^2\right\} \\ &= \frac{1}{|\mathcal{S}_t|^2}\sum_{k\in\mathcal{S}_t}\left|1 - \frac{1}{\sqrt{\eta_t}}\boldsymbol{m}_t^H p_{k,t}\mathbf{h}_k[\mathbf{x}_t]\right|^2 + \frac{d\|\boldsymbol{m}_t^H\|^2}{\eta_t|\mathcal{S}_t|^2}\sigma_n^2.\end{aligned} \tag{9}$$

Following [11], [13], the transmit scalars in (9) can be designed using a zero-forcing structure to suppress interclient interference while approximating the minimum MSE, thereby reducing aggregated signal distortion at the AP and enabling efficient OTA model aggregation. Based on the zero-forcing structure, the transmit scalars are expressed as follows:

$$p_{k,t} = \frac{\sqrt{\eta_t}\left(\boldsymbol{m}_t^H\boldsymbol{h}_k[\mathbf{x}_t]\right)^H}{\left|\boldsymbol{m}_t^H\boldsymbol{h}_k[\mathbf{x}_t]\right|^2}. \tag{10}$$

To meet each client's maximum power constraints, the upper bound of η_t must satisfy

$$\eta_t \le \frac{dp_{\max}\left|\boldsymbol{m}_t^H\boldsymbol{h}_k[\mathbf{x}_t]\right|^2}{\mathbb{E}\left[\|\boldsymbol{w}_{k,t}\|^2\right]} \quad \forall k \in \mathcal{S}_t. \tag{11}$$

Substituting (11) and (10) into (9), and assuming $\mathbb{E}[\|\boldsymbol{w}_{k,t}\|^2] \ge \Gamma$, the MSE at time step t is given by

$$\mathrm{MSE}_t = \frac{\Gamma\sigma_n^2}{|\mathcal{S}_t|^2 p_{\max}}\max_{k\in\mathcal{S}_t}\frac{\|\boldsymbol{m}_t^H\|^2}{\left|\boldsymbol{m}_t^H\boldsymbol{h}_k[\mathbf{x}_t]\right|^2}. \tag{12}$$

Remark 1: As highlighted in (12), the MSE remains invariant under any nonzero scaling of the aggregation beamforming vector. Specifically, both the original beamforming vector $\boldsymbol{m}$ and its scaled counterpart $\boldsymbol{m} = c\boldsymbol{m}$ (where $c \neq 0$) result in identical MSE values. This demonstrates that the correlation between $\boldsymbol{m}$ and the channel gain $\boldsymbol{h}_k[\mathbf{x}_t]$ is the primary factor influencing MSE, while the magnitude of $\boldsymbol{m}$

does not play a significant role. Furthermore, in the FA systems, the additional DoF provided by antenna mobility can facilitate the establishment of varying phases for channel gain. This flexibility can lead to improved alignment, thereby augmenting the term $|\boldsymbol{m}_t^H \boldsymbol{h}_k[\mathbf{x}_t]|^2$. Consequently, FA systems are well-suited for applications involving OTA computation.

To assess the impact of FAs on learning performance, we consider the user selection and the model aggregation error and analyze the convergence of OTA-FL, based on the widely accepted assumptions for convergence analysis [34].

Assumption 1: The loss function $F(\boldsymbol{w})$ is ℓ-smooth, meaning there exists a constant $\ell \geq 0$ such that for all $\boldsymbol{w}, \boldsymbol{v} \in \mathbb{R}^d$

$$F(\boldsymbol{w}) - F(\boldsymbol{v}) \leq (\boldsymbol{w} - \boldsymbol{v})^{\mathrm{T}} \nabla F(\boldsymbol{v}) + \frac{\ell}{2} \|\boldsymbol{w} - \boldsymbol{v}\|^2. \tag{13}$$

Assumption 2: The loss function satisfies the Polyak–Lojasiewicz (PL) inequality, characterized by the optimal loss function $F(\boldsymbol{w}^*)$ under the condition $\mu > 0$, given by

$$\|\nabla F(\boldsymbol{w})\|^2 \geq 2\mu \big[F(\boldsymbol{w}) - F\big(\boldsymbol{w}^*\big)\big]. \tag{14}$$

Assumption 3: Each FL client obtains an unbiased estimate of the stochastic gradient with bounded variance, given by

$$\begin{aligned} \mathbb{E}\Big[\nabla \hat{F}_k(\boldsymbol{w})\Big] &= \nabla F_k(\boldsymbol{w}) \\ \mathbb{E}\Big[|\nabla \hat{F}_k(\boldsymbol{w}) - \nabla F_k(\boldsymbol{w})|^2\Big] &\leq \frac{\sigma_g^2}{|\zeta_k|} \end{aligned} \tag{15}$$

where σ_g^2 denotes the variance introduced by SGD.

Assumption 4: For any labeled data samples $d \in \mathbb{R}^d \times \mathbb{R}$ and model parameter $\boldsymbol{w} \in \mathbb{R}^d$, the gradient norm is bounded as

$$\|\nabla f(\boldsymbol{w}, d)\|^2 \leq \kappa. \tag{16}$$

Theorem 1: Given Assumptions 1–4 and a learning rate $0 \leq \gamma \leq (1/2\ell)$, the optimality gap after T communication rounds is bounded by

$$\begin{aligned} \mathbb{E}\big[F(\boldsymbol{w}_{T+1})\big] - F\big(\boldsymbol{w}^*\big) \leq \psi^{\mathrm{T}} \big(\mathbb{E}[F(\boldsymbol{w}_1)] - F\big(\boldsymbol{w}^*\big)\big) \\ + \sum_{t=1}^{\mathrm{T}} \psi^{T-t} \Theta_t \end{aligned} \tag{17}$$

where $\Theta_t = 2\gamma\kappa(1 - [|\mathcal{S}_t|/K])^2 + (\gamma^2 \ell / |\mathcal{S}_t|^2) \sum_{k \in \mathcal{S}_t} (\sigma_g^2 / |\zeta_k|) + (\ell/2)\mathrm{MSE}_t$, and $\psi = 1 - \gamma\mu$.

Proof: See the Appendix. ∎

Remark 2: It is evident that selecting an appropriate learning rate γ such that $|\psi| < 1$ ensures the first term in (17) diminishes as $T \to \infty$, leaving the optimality gap predominantly influenced by the second term. Since ψ remains constant across all time slots, reducing the optimality gap depends on minimizing Θ_t in each communication round. Notably, Θ_t is inversely proportional to $|S_t|$ and directly proportional to MSE_t. Therefore, increasing $|S_t|$ while maintaining MSE_t within acceptable limits is crucial for achieving improved convergence.

IV. Problem Formulation

We design FA systems by jointly optimizing the parameters $\boldsymbol{m} = [m_1, \ldots, m_N]^{\mathrm{T}}$, $\boldsymbol{x} = [x_1, \ldots, x_N]^{\mathrm{T}}$, and the device set $\mathcal{S}_t$ to minimize the total optimality gap by maximizing the number of selected FL devices ($|\mathcal{S}_t|$) while ensuring that MSE_t remains within a predetermined limit, given by

$$\begin{aligned} \mathcal{P}_1 : \max_{\mathcal{S}_t \subset \mathcal{K}, \boldsymbol{m}, \boldsymbol{x}} \quad & |\mathcal{S}_t| \\ \text{s.t.} \quad & C_1 : \mathrm{MSE}_t \leq \tau \\ & C_2 : 0 \leq x_n \leq X \quad \forall n \in \{1, \ldots, N\} \\ & C_3 : x_1 < x_2 < \cdots < x_N \\ & C_4 : x_n - x_{n-1} > X_0 \quad \forall n \in \{2, \ldots, N\}. \end{aligned} \tag{18}$$

Here, C_1 ensures that the MSE remains below a predefined threshold in each training round, C_2 defines the allowable range for FA positions, C_3 establishes the ordering of FA placements, and C_4 imposes a minimum distance constraint between neighboring FAs.

In problem $\mathcal{P}_1$, the set of selected devices participating in the aggregation process is meticulously designed to ensure that the MSE does not exceed a specified threshold. This can be mathematically formulated as

$$\mathcal{S}_t = \{k \in \mathcal{K} | \mathrm{MSE}_t \leq \tau\}. \tag{19}$$

Without loss of generality, we impose the constraint $\|\boldsymbol{m}_t\|^2 \leq 1$, thereby leveraging (12), we obtain

$$\frac{\Gamma \sigma_n^2}{|\mathcal{S}_t|^2 p_{\max}} \max_{k \in \mathcal{S}_t} \frac{\|\boldsymbol{m}_t^H\|^2}{\big|\boldsymbol{m}_t^H \boldsymbol{h}_k[\mathbf{x}_t]\big|^2} \leq \tau \quad \forall k \in \mathcal{S}_t. \tag{20}$$

Upon simplification, this leads to the condition, given by

$$\big|\boldsymbol{m}_t^H \boldsymbol{h}_k[\mathbf{x}_t]\big|^2 \geq \frac{\Gamma \sigma_n^2}{\tau |\mathcal{S}_t|^2 p_{\max}} \quad \forall k \in \mathcal{S}_t. \tag{21}$$

Introducing a scaling parameter $\delta = (\Gamma \sigma_n^2 / \tau |\mathcal{S}_t|^2 p_{\max})$, the user selection policy can be equivalently reformulated as

$$\mathcal{S}_t = \{k \in \mathcal{K} \mid \big|\boldsymbol{m}_t^H \boldsymbol{h}_k[\mathbf{x}_t]\big|^2 \geq \delta\}. \tag{22}$$

As expressed in (22), an increase in the correlation of channel gains results in a higher number of selected users. Therefore, FA systems can enhance user selection by dynamically adjusting channel vectors $\boldsymbol{h}_k[\mathbf{x}_t]$, unlike FPA systems, which optimize only the beamforming vector $\boldsymbol{m}$. Consequently, the optimization problem $\mathcal{P}_1$ can be reformulated as

$$\begin{aligned} \mathcal{P}_2 : \max_{\boldsymbol{m}, \mathbf{x}} \quad & |\mathcal{S}_t| \\ \text{s.t.} \quad & C_1 : |\boldsymbol{m}_t^H \boldsymbol{h}_k[\mathbf{x}_t]|^2 \geq \delta \\ & C_5 : \|\boldsymbol{m}_t\|^2 \leq 1 \\ & C_2, C_3, \text{and } C_4. \end{aligned} \tag{23}$$

The nonconvex nature of the objective function, coupled with the stochastic characteristics inherent in large-scale IoT environments, poses significant challenges to traditional optimization techniques in solving problem $\mathcal{P}_2$. The problem requires real-time decision-making based on current environmental conditions. To overcome these challenges, we reformulate the problem $\mathcal{P}_2$ as an MDP and propose an

online DRL algorithm, enabling dynamic and efficient adaptation to changing environmental conditions. Unlike traditional optimization methods that require explicit models of environmental dynamics, the DRL-based approach learns directly from interactions with the environment, making it particularly suitable for high-dimensional and nonconvex problems, such as $\mathcal{P}_2$. Furthermore, the proposed method embeds the physical and system-level constraints of $\mathcal{P}_2$ within the action space and reward function, ensuring feasibility and convergence during training and deployment.

V. Proposed DRL-Based Approach

We deploy a DRL agent at the AP to learn an optimal policy for jointly optimizing the beamforming vector $\boldsymbol{m}$ and FA locations $\boldsymbol{x}$ in each round, aiming to maximize the number of selected users $|\mathcal{S}_t|$. The details of the MDP framework are as follows.

1) *State Space:* The system state includes all possible values of the distance d_k and the AoAs of the LoS paths ϕ_k for all $k \in \mathcal{K}$, represented as $s_t = \{[d_{1,t}, \ldots, d_{K,t}], [\phi_{1,t}, \ldots, \phi_{K,t}]\}$.
2) *Action Space:* The action space consists of the beamforming vectors and the positions of the FAs, represented as $a_t = \{[m_{1,t}, \ldots, m_{N,t}], [x_{1,t}, \ldots, x_{N,t}]\}$.
3) *Reward Function:* According to Theorem 1, in order to maximize the $|\mathcal{S}_t|$, the reward function is formulated as

$$r(s_t, a_t) = \begin{cases} r_p, & \text{if } \left|\boldsymbol{m}_t^H \boldsymbol{h}_k[\mathbf{x}_t]\right| = 0 \\ |\mathcal{S}_t|, & \text{otherwise} \end{cases} \tag{24}$$

where r_p is a penalty term that needs to be tuned during the simulation to achieve optimal convergence behavior. The reward function (24) provides quantifiable feedback to guide policy improvement. It ensures that actions increasing the number of users satisfying the MSE threshold are positively reinforced.

Given the continuous nature of the action space in our problem, conventional model-free value-based DRL algorithms, such as deep Q-networks (DQNs), are not suitable to handle continuous actions. We address this by leveraging policy gradient-based methods, specifically the DDPG algorithm, which handles continuous actions effectively [35]. However, traditional DDPG with fully connected networks struggles to capture complex temporal patterns in user mobility. To improve adaptability, we integrate an long–short-term memory (LSTM) layer into the DDPG framework, to leverage temporal state patterns and adapt to dynamic environments. It is worth noting that while the state vector at time t contains current information on AoAs and distances, it does not explicitly encode how these features evolve over time. In dynamic wireless environments, such temporal information can be critical. The LSTM layer in our actor network captures this time evolution, enabling the policy to learn temporal correlations and anticipate environment changes.

Our LSTM-DDPG implementation comprises four neural networks. The actor network π_ϕ, with parameters ϕ, generates actions a_t from the current state s_t, incorporating exploration noise as $a_t = \pi_\phi(s_t) + \xi$. The critic network, parameterized by θ, evaluates state-action pairs to compute Q-values $Q_\theta(s_t, a_t)$, representing expected rewards. The target actor and critic networks, as delayed versions of the main networks, ensure stable learning. The framework aims to learn an optimal policy π^* to maximize cumulative expected rewards, defined as

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{s_t, a_t}\left[\sum_{t=0}^{\infty} r(s_t, a_t)\right]. \tag{25}$$

The actor network is updated by the gradient of the objective function $J(\phi)$, given by

$$\nabla_\phi J(\phi) = \mathbb{E}\left[\nabla_{a_t} Q_{\theta_1}(s_t, a_t)\Big|_{a_t=\pi_\phi(s_t)} \nabla_\phi \pi_\phi(s_t)\right]. \tag{26}$$

Simultaneously, the critic network is updated to minimize the error between its predictions and the target values Y_t, as specified by the loss function, expressed as

$$Y_t = r_t + \gamma Q_{\theta_i'}\big(s_{t+1}, \pi_{\phi'}(s_{t+1}) + \xi\big). \tag{27}$$

The proposed LSTM-DDPG method is described in Algorithm 1.

Algorithm 1: LSTM-DDPG Algorithm

Initialize: experience replay buffer M, mini-batch size H, the actor network π_ϕ, the critic network Q_θ with random values, and create the target networks by setting $\theta' \leftarrow \theta$ and $\phi' \leftarrow \phi$.
Set: Set E and T to be the maximum number of episodes and episode length, respectively.
for *each episode* $e : E$ **do**
  Initialize the environment state s_0, and the exploration noise ξ;
  for $t = 1 : T$ **do**
    Receive s_t from the environment;
    Obtain $a_t = \pi_\phi(s_t) + \xi$ from the actor network and re-shape it;
    Compute r_t using (25);
    Observe the new state s_{t+1};
    Store transition (s_t, a_t, r_t, s_{t+1}) into M;
  end
  Randomly sample a H mini-batch of transitions from M;
  Compute the target function Y_t according to (27);
  Update the actor and critic networks using the Adam optimizer.
  Perform a soft update of the target actor and critic networks using the update coefficient $\tau_N \in [0, 1]$:

$$\phi' \leftarrow \tau_N \phi + (1 - \tau_N)\phi', \quad \theta' \leftarrow \tau_N \theta + (1 - \tau_N)\theta'$$

end

A. Computational Complexity Analysis

The computational complexity of the proposed algorithm consists of two main components: action selection and training [29]. The network architecture includes an actor and a critic network, each with $\mathcal{U}$ hidden layers containing $\mathcal{L}$ neurons per layer. The action selection complexity, which refers to

generating the network output for a given input, is determined by the size of the consecutive layers. For the actor network, this is expressed as: $\mathcal{J} \times (|S| + |A|) \times \mathcal{L}$ for the input and first layer, $\mathcal{L}^2$ for the successive hidden layers, and $\mathcal{L} \times |A|$ for the output layer, where $\mathcal{J}$ represents the previous trajectory length, $|S|$ denotes the state dimension, and $|A|$ represents the action dimension.

Similarly, for the critic network, the computation of the subsequent layers is $\mathcal{J} \times (|S| + 2 \times |A|) \times \mathcal{L}$ for the input and first layer, $\mathcal{L}^2$ for the successive hidden layers, and $\mathcal{L} \times |A|$ for the final connection. Here, $|S|$ and $|A|$ denote the dimensions of the agent state and action spaces, respectively. Thus, the computational complexity of action selection for the proposed method is $\mathcal{O}(\mathcal{L}^2)$.

During the training process, the computational complexity of LSTM-DDPG is primarily determined by the number of network edges, calculated as $I \times C + C^2 + C \times O$, where I is the input size, C is the number of neurons, and O is the output size [29]. The complexity for the actor and critic networks can be further refined as: $(H|S|\mathcal{L} + H\mathcal{L}^2 + H\mathcal{L}|A|)$ and $(H(|S| + |A|)\mathcal{L} + H\mathcal{L}^2 + H\mathcal{L})$, respectively, where H denotes the batch size. Hence, the overall training complexity of the LSTM-DDPG method is $\mathcal{O}(H\mathcal{L}^2)$. For comparison, the computational complexity of other DRL algorithms for one episode, such as SAC and DDPG, is given by $\mathcal{O}((\sum_{\mathcal{N}=1}^{\mathcal{N}_l} C_{\mathcal{N}} C_{\mathcal{N}-1})H)$, where $\mathcal{N}_l$ is the number of layers, and $C_{\mathcal{N}}$ denotes the number of neurons per layer [27], [36]. To further illustrate the practical implications of this complexity, Fig. 1 presents the number of floating-point operations (FLOPs) required by the proposed algorithm under different mini-batch sizes H and varying numbers of FAs N. The results indicate that increasing the number of FA elements leads to higher computational complexity across all batch sizes, which is expected due to the expanded dimensionality of both state and action spaces. For instance, when $H = 256$, doubling the number of FA elements from $N = 10$ to $N = 20$ results in an 8.32% increase in computational load. This trend highlights the tradeoff between performance gains enabled by FA diversity and the resulting computational overhead.

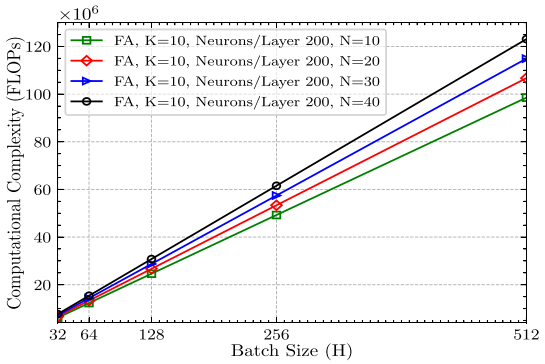

Fig. 1. Computational complexity of the proposed algorithm in terms of FLOPs versus batch size H for different numbers of FAs (N).

TABLE I
SYSTEM PARAMETERS [15], [23], [29], [37]

Par.	Description	Value
K	Total users	10
N	Number of FAs	5
σ^2	Noise power density	-170 dBm/Hz
–	User-to-AP distance range	$[20, 100]$ m
–	AoA range	$[-\pi/2, \pi/2]$ rad
X_0	Minimum FA separation	0.5λ
$[0, X]$	FA interval distance	$[0, 8\lambda]$
$P_{\max}$	Max transmit power	36 dBm
$\mathbf{h}_k[\mathbf{x}]$	Channel gain	Rician
α_L, α_N	Path loss exponent (LOS/NLOS)	3.5
K_r	Rician factor	8
A_L, A_N	Path loss at reference distance	30 dBm
τ	MSE requirement	2 dBm
$\mathcal{N}_e$	Episodes	6000
H	Batch size	64
τ_N	Target network update period	0.001
$\mathcal{U}$	Hidden layers	2
–	Actor learning rate	0.0005
–	Critic learning rate	0.0001
γ	Discount factor	0.8
M	Replay buffer size	600000
–	Input/Hidden layer neurons	200
–	Output layer neurons	10
–	Input/Hidden activation	ReLU
–	Output activation	Softmax
–	Communication rounds	100

VI. SIMULATION RESULTS

This section presents numerical results to demonstrate the effectiveness of our approach, which integrates adaptive antenna positioning in the FA system with real-time decision-making using the proposed LSTM-DDPG algorithm for optimizing OTA-FL convergence. This algorithm is initialized with the following parameters [15], [23], [37]: the user-to-AP distance is uniformly distributed within the range of [20, 100] m, and the AoAs is uniformly distributed within $[-\pi/2, \pi/2]$ radians. For the FA configuration, we set $X_0 = 0.5\lambda$ and $X = 8\lambda$. The simulation involves a total of 100 devices, with 5 FAs deployed. The simulations are conducted using TensorFlow 2.6.1 and Torch 1.4.0, along with the Adam optimizer. The key hyperparameters include a learning rate of 0.0005, a batch size of 64, a replay buffer of 10^4, a soft update rate of 0.001, and a discount factor of 0.9. Performance is compared between the FA system and the FPA system [23], [38] using a uniform location vector $\boldsymbol{x} = [(X/[N+1]), \ldots, (NX/[N+1])]^{\mathrm{T}}$. Furthermore, we assess the LSTM-DDPG algorithm against the SAC algorithm [26], [39], [40] and the standard DDPG algorithm [29] based on average rewards, calculated at each episode e as $R_{\text{avg}}(e) = (1/100)\sum_{i=e-100}^{e} R_i$, where R_i denotes the mean reward obtained in episode i. Table I provides a summary of the simulation parameters.

Fig. 2 depicts the convergence behavior of various DRL algorithms across training episodes. The solid curves represent the average rewards, while the shaded regions correspond to the standard deviations, offering insights into the stability of each algorithm. The results demonstrate that the proposed LSTM-DDPG approach exhibits higher average rewards and lower variance compared to standard DRLs, demonstrating superior performance and improved stability in dynamic environments. The narrower variance band of LSTM-DDPG further confirms its enhanced training stability compared to

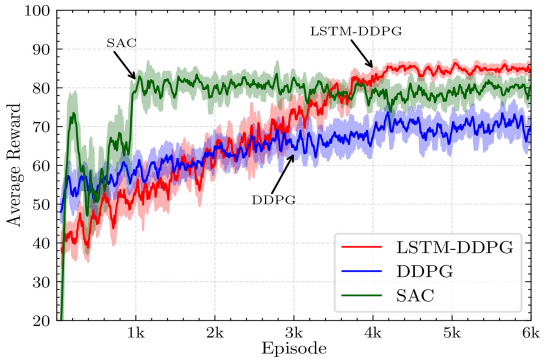

Fig. 2. Convergence analysis of DRL algorithms based on average reward per training episode.

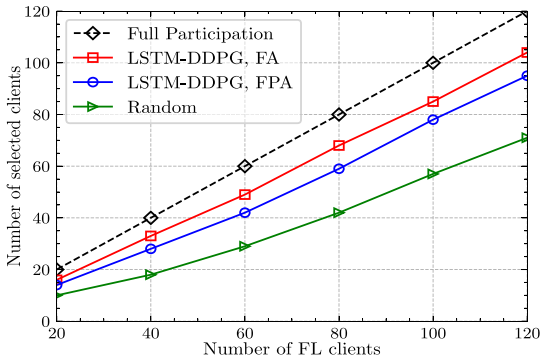

Fig. 3. Number of selected devices versus number of FL clients.

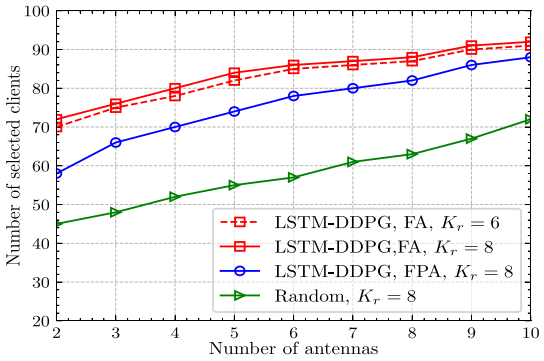

Fig. 4. Number of selected devices versus number of antenna.

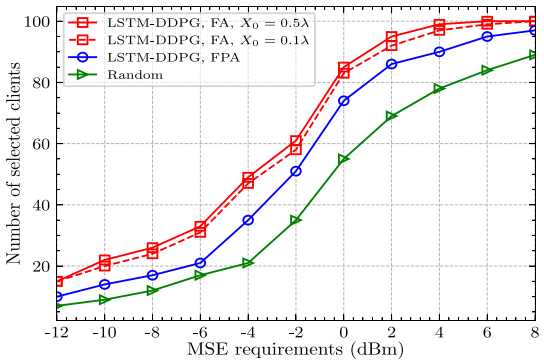

Fig. 5. Number of selected devices versus MSE requirement.

conventional DRL methods, making it particularly suitable for real-world applications requiring reliable convergence. Furthermore, experimental results demonstrate that the SAC algorithm achieves superior performance compared to DDPG, exhibiting both higher average rewards and lower variance across training episodes.

Fig. 3 illustrates the relationship between the total number of devices, K, and the selected devices, $|\mathcal{S}_t|$, with full participation ($|\mathcal{S}_t| = K$) used as a benchmark. The number of selected devices increases nearly linearly with K, aligning with the analysis that random beamforming ensures proportionality between selected and total devices, resulting in consistent system performance. Furthermore, the proposed FA method consistently outperforms both FPA and the random algorithm across varying total FL client populations. This demonstrates the efficacy of adaptive antenna positioning and robust client selection capabilities. These results indicate that the proposed framework is particularly well-suited for large-scale FL systems, where dynamic client participation poses significant challenges.

Fig. 4 shows the performance of the proposed algorithms with varying antenna numbers. As N increases, the number of selected devices improves in all scenarios, though the rate of improvement slows as N continues to grow, attributed to the fixed length of X. FA systems consistently outperform FPA systems for all N values, owing to the enhanced DoF provided by the adjustable antenna configurations. Additionally, to further validate the generalizability of our approach under different channel conditions, we examine the impact of the rician factor K_r, which determines the dominance of the LoS component relative to scattered multipath components in the wireless channel. We conducted additional simulations with $K_r = 6$ and observed that the proposed algorithm maintains highly robust performance, achieving results comparable to those at $K_r = 8$ while preserving the same key trends in device selection and MSE control. These results confirm that our method exhibits strong consistency across similar LoS dominant channel conditions.

Fig. 5 illustrates the relationship between the number of selected devices $|\mathcal{S}|$ and the MSE requirement τ. As τ increases, more devices are selected due to the lower selection threshold δ, allowing more devices to meet the MSE constraint. Notably, to assess the robustness of the proposed method under varying antenna spacing values X_0, the performance of the LSTM-DDPG approach with FA under different X_0 values is also evaluated. The results indicate that the proposed method exhibits slightly improved performance at higher values of X_0. This marginal variation is primarily attributed to the presence of a strong LoS component and the dominant influence of the rician channel gain, which render the performance relatively insensitive to the changes in X_0. In addition, the proposed LSTM-DDPG approach with FA consistently outperforms other methods across different values of τ. Specifically, when τ exceeds 4 dBm, the proposed scheme selects all 100 devices, demonstrating its superior efficiency in accommodating a larger number of devices under the higher MSE thresholds.

To further investigate the relationship between OTA-FL performance, the number of selected devices, and the MSE

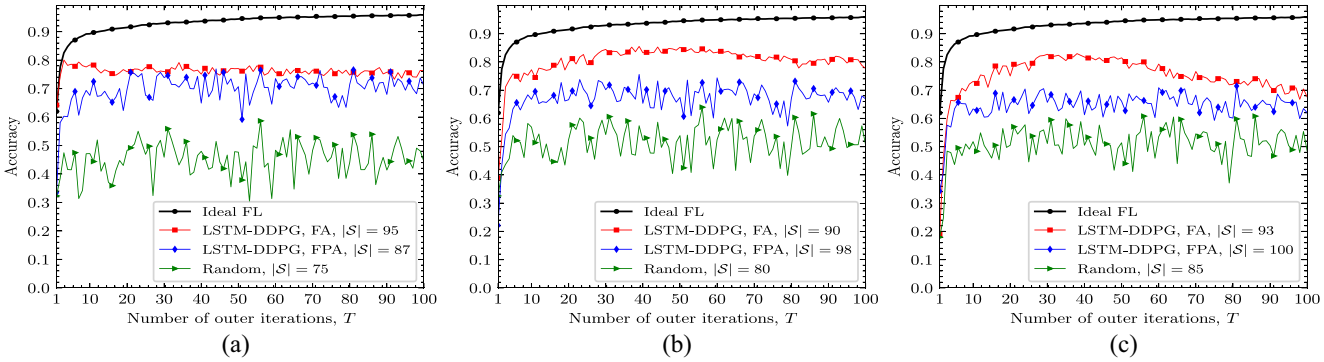

Fig. 6. Learning performance over different communication rounds with (a) $\tau = 2$ dBm; (b) $\tau = 4$ dBm; and (c) $\tau = 6$ dBm.

threshold, we conducted extensive experiments involving collaborative training of image classification models using the MNIST dataset. The experimental setup consisted of 100 users, each assigned a unique data subset, and training was performed over 100 communication rounds. The MNIST dataset was divided into 90% training and 10% testing sets, with the training set further partitioned into 100 nonoverlapping subsets, each assigned to an individual client. Each client utilized a feedforward neural network for local model training. The architecture comprised an input layer with 200 neurons, followed by a hidden layer with 200 neurons, both employing ReLU activation functions. The output layer contained a number of neurons equal to the number of classes in the classification task, utilizing a softmax activation function for final predictions. Model updates from all clients were aggregated using OTA-FL, leveraging the wireless medium for OTA computation. To establish a performance baseline, we also considered an ideal FL scenario, assuming full device participation and a noise-free communication environment. This benchmark setting serves as a reference to quantify the impact of communication constraints and user selection policies on model convergence and accuracy.

Fig. 6 depicts the accuracy of OTA-FL over 100 communication rounds for different values of τ. The results indicate that, for each τ, the proposed method consistently outperforms existing approaches. This performance gain is primarily attributed to the higher number of selected users under the same τ, which leads to improved accuracy, reduced fluctuations across communication rounds, and faster convergence. However, while increasing τ directly increases the number of selected users, its impact on accuracy and convergence speed is not strictly monotonic. To further investigate this trend, Table II presents the average accuracy over the last 10 communication rounds for different τ values. Specifically, as τ increases from 2 to 4, the average accuracy improves and convergence accelerates due to the larger set of selected users, which enhances learning performance. In contrast, when τ is further increased to 6, performance degrades due to the adverse effects of a higher τ. These results highlight the critical role of selecting an optimal τ to balance user participation and model convergence, ultimately achieving the best performance in OTA-FL systems.

TABLE II
AVERAGE ACCURACY FOR DIFFERENT MSE REQUIREMENTS τ

τ	Algorithms	$\lvert\mathcal{S}\rvert$	Average Accuracy
2 dBm	LSTM-DDPG, FA	100	0.751
	LSTM-DDPG, FPA	87	0.717
4 dBm	LSTM-DDPG, FA	98	0.804
	LSTM-DDPG, FPA	90	0.680
6 dBm	LSTM-DDPG, FA	100	0.697
	LSTM-DDPG, FPA	93	0.646

VII. Conclusion

In this article, we have studied integrating FAs into AP to enhance OTA-FL systems. Our convergence analysis has shown that the optimality gap is significantly affected by the number of selected devices and MSE, both of which are influenced by the FA positions and beamforming vectors. To maximize user participation, we have formulated a nonconvex optimization problem that jointly optimizes FA positions, beamforming vectors, and user selection policies under MSE constraints. We have further transformed the problem as an MDP formulation to capture the environment dynamic nature and used an LSTM-based algorithm for real-time optimization. The LSTM-based approach effectively learns long-term correlations in the optimization process, allowing for adaptive decision-making under varying channel conditions. Simulation results have shown that the FA-assisted system significantly outperforms FPA-based systems, achieving higher user selection rates, and leading to improved performance of OTA-FL systems.

Appendix

Proof of Theorem 1

Let $\nabla\hat{F}_s(\boldsymbol{w}_t) = (1/|\mathcal{S}_t|)\sum_{k\in\mathcal{S}_t}\nabla\hat{F}_k(\boldsymbol{w}_t)$ and $\nabla F_s(\boldsymbol{w}_t) = (1/|\mathcal{S}_t|)\sum_{k\in\mathcal{S}_t}\nabla F_k(\boldsymbol{w}_t)$. In the tth communication round, the global model update is formulated as follows:

$$\hat{\boldsymbol{w}}_{t+1} = \frac{1}{|\mathcal{S}_t|}\left(\sum_{k\in\mathcal{S}_t}\left(\boldsymbol{w}_t - \gamma\nabla\hat{F}_k(\boldsymbol{w}_t)\right) + \frac{\boldsymbol{m}_t^H\mathbf{z_t}}{\sqrt{\eta_t}}\right)$$
$$= \boldsymbol{w}_t - \gamma\nabla\hat{F}_s(\boldsymbol{w}_t) + \frac{\boldsymbol{m}_t^H\mathbf{z}_t}{|\mathcal{S}_t|\sqrt{\eta_t}}. \quad (28)$$

Based on Assumptions 1 and 3, we have

$$\Lambda_t = \mathbb{E}\big[F(\boldsymbol{w}_{t+1})\big] - \mathbb{E}[F(\boldsymbol{w}_t)] \leq -\gamma \langle \nabla F_s(\boldsymbol{w}_t)^{\mathrm{T}}, \nabla F(\boldsymbol{w}_t)\rangle + \frac{\gamma^2 \ell}{2}\mathbb{E}\Big[\|\nabla \hat{F}_s(\boldsymbol{w}_t)\|^2\Big] + \frac{\ell}{2}\mathrm{MSE}_t. \tag{29}$$

The upper bound of $\mathbb{E}[\|\nabla \hat{F}_s(\boldsymbol{w}_t)\|^2]$ is derived as

$$\begin{aligned}\mathbb{E}\Big[\Big\|\nabla \hat{F}(\boldsymbol{w}_t)\Big\|^2\Big] &= \mathbb{E}\Big[\Big\|\Big(\nabla \hat{F}_s(\boldsymbol{w}_t) - \nabla F_s(\boldsymbol{w}^{[t]}) + \nabla F_s(\boldsymbol{w}^{[t]})\Big)\Big\|^2\Big] \\ &\overset{(a_1)}{\leq} 2\mathbb{E}\Big[\Big\|\nabla \hat{F}_s(\boldsymbol{w}_t) - \nabla F_s\big(\boldsymbol{w}^{[t]}\big)\Big\|^2\Big] + 2\|\nabla F_s\big(\boldsymbol{w}^{[t]}\big)\|^2 \\ &\leq \frac{2}{|\mathcal{S}_t|^2}\sum_{k\in\mathcal{S}_t}\frac{\sigma_g^2}{|\zeta_k|} + 2\big\|\nabla F_s\big(\boldsymbol{w}^{[t]}\big)\big\|^2 \end{aligned} \tag{30}$$

where a_1 follows from the inequality $\|\mathbf{a}+\mathbf{b}\|^2 \leq 2\|\mathbf{a}\|^2 + 2\|\mathbf{b}\|^2$. This leads to the conclusion that

$$\Lambda_t \leq -\gamma \langle \nabla F_s(\boldsymbol{w}_t)^{\mathrm{T}}, \nabla F(\boldsymbol{w}_t)\rangle + \gamma^2 \ell \big\|\nabla F_s\big(\boldsymbol{w}^{[t]}\big)\big\|^2 + \frac{\gamma^2 \ell}{|\mathcal{S}_t|^2}\sum_{k\in\mathcal{S}_t}\frac{\sigma_g^2}{|\zeta_k|} + \frac{\ell}{2}\mathrm{MSE}_t. \tag{31}$$

To evaluate the impact of user selection on the optimality gap, we define $\nabla F_s(\boldsymbol{w}_t) = \nabla F(\boldsymbol{w}_t) - o_t$. Substituting this into (31), we obtain

$$\Lambda_t \leq \gamma(\gamma\ell - 1)\|\nabla F(\boldsymbol{w}_t)\|^2 - \gamma(2\gamma\ell - 1)\big\langle o_t^{\mathrm{T}}, \nabla F(\boldsymbol{w}_t)\big\rangle + \gamma^2 \ell \|o_t\|^2 + \frac{\gamma^2 \ell}{|\mathcal{S}_t|^2}\sum_{k\in\mathcal{S}_t}\frac{\sigma_g^2}{|\zeta_k|} + \frac{\ell}{2}\mathrm{MSE}_t. \tag{32}$$

Then, by setting $0 \leq \gamma \leq (1/2\ell)$ and applying the Cauchy-Schwarz inequality, $\langle o_t^{\mathrm{T}}, \nabla F(\boldsymbol{w}_t)\rangle \leq 0.5\|\nabla F(\boldsymbol{w}_t)\|^2 + 0.5\|o_t\|^2$, we can conclude

$$\Lambda_t \leq \frac{\gamma}{2}\Big(\|o_t\|^2 - \|\nabla F(\boldsymbol{w}_t)\|^2\Big) + \frac{\gamma^2 \ell}{|\mathcal{S}_t|^2}\sum_{k\in\mathcal{S}_t}\frac{\sigma_g^2}{|\zeta_k|} + \frac{\ell \mathrm{MSE}_t}{2}. \tag{33}$$

According to [11], [13], and Assumption 4, we obtain

$$|o_t|^2 \leq 4\kappa\left(1 - \frac{|\mathcal{S}_t|}{K}\right)^2. \tag{34}$$

Thus, based on Assumption 2, we derive the following inequality:

$$\mathbb{E}\big[F(\boldsymbol{w}_{t+1})\big] - \mathbb{E}\big[F\big(\boldsymbol{w}^*\big)\big] \leq \psi_t\big(\mathbb{E}[F(\boldsymbol{w}_t)] - \mathbb{E}\big[F\big(\boldsymbol{w}^*\big)\big]\big) + \Theta_t \tag{35}$$

where ψ_t and Θ_t are defined in (20) and (21). By recursively operating on (35), the cumulative optimality gap in Theorem 1 is calculated.

REFERENCES

[1] Q.-V. Pham, M. Zeng, T. Huynh-The, Z. Han, and W.-J. Hwang, "Aerial access networks for federated learning: Applications and challenges," *IEEE Netw.*, vol. 36, no. 3, pp. 159–166, Jul. 2022.

[2] X. Wang, Y. Chen, Q. Ye, and O. A. Dobre, "Connectivity enrichment for decentralized federated learning networks with teleportation," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM WKSHPS)*, May 2025, pp. 19–22.

[3] Y. Wang, Z. Su, N. Zhang, and A. Benslimane, "Learning in the air: Secure federated learning for UAV-assisted crowdsensing," *IEEE Trans. Netw. Service Manag.*, vol. 8, no. 2, pp. 1055–1069, Apr.–Jun. 2020.

[4] D. Yang et al., "DetFed: Dynamic resource scheduling for deterministic federated learning over time-sensitive networks," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 5162–5178, Aug. 2024.

[5] M. Le, T. Huynh-The, T. Do-Duy, T.-H. Vu, W.-J. Hwang, and Q.-V. Pham, "Applications of distributed machine learning for the Internet-of-Things: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 27, no. 2, pp. 1053–1100, 2nd Quart., 2025.

[6] Z. Wang, Y. Zhao, Y. Zhou, Y. Shi, C. Jiang, and K. B. Letaief, "Over-the-air computation for 6G: Foundations, technologies, and applications," *IEEE Internet Things J.*, vol. 11, no. 14, pp. 24634–24658, Jul. 2024.

[7] H. Liu, X. Yuan, and Y.-J. A. Zhang, "CSIT-free model aggregation for federated edge learning via reconfigurable intelligent surface," *IEEE Wireless Commun. Lett.*, vol. 10, no. 11, pp. 2440–2444, Nov. 2021.

[8] M. Ahmadzadeh, S. Pakravan, G. A. Hodtani, M. Zeng, and J.-Y. Chouinard, "Deep reinforcement learning for robust RIS-aided over-the-air federated learning in cognitive radio," in *Proc. IEEE Middle East Conf. Commun. Netw. (MECOM)*, Nov. 2024, pp. 368–373.

[9] X. Yuan et al., "Low-cost federated broad learning for privacy-preserved knowledge sharing in the RIS-aided Internet of Vehicles," *Engineering*, vol. 33, pp. 178–189, Feb. 2023.

[10] B. Wei, P. Zhang, and Q. Zhang, "Active reconfigurable intelligent surface-aided over-the-air computation networks," *IEEE Wireless Commun. Lett.*, vol. 13, no. 4, pp. 1148–1152, Apr. 2024.

[11] Z. Wang et al., "Federated learning via intelligent reflecting surface," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 808–822, Feb. 2022.

[12] D. Zhang, M. Xiao, M. Skoglund, and H. V. Poor, "Beamforming design for active RIS-aided over-the-air computation," Nov. 2023, *arXiv:2311.18418*.

[13] C. Chen, Y.-H. Chiang, H. Lin, J. C. Lui, and Y. Ji, "Joint client selection and receive beamforming for over-the-air federated learning with energy harvesting," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 1127–1140, 2023.

[14] K.-K. Wong, A. Shojaeifard, K.-F. Tong, and Y. Zhang, "Fluid antenna systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1950–1962, Mar. 2021.

[15] L. Zhu, W. Ma, and R. Zhang, "Modeling and performance analysis for movable antenna enabled wireless communications," *IEEE Trans. Wireless Commun.*, vol. 23, no. 6, pp. 6234–6250, Jun. 2024.

[16] W. K. New et al., "A tutorial on fluid antenna system for 6G networks: Encompassing communication theory, optimization methods and hardware designs," *IEEE Commun. Surveys Tuts.*, vol. 27, no. 4, pp. 2325–2377, 4th Quart., 2025.

[17] L. Zhu and K.-K. Wong, "Historical review of fluid antenna and movable antenna," Jan. 2024, *arXiv:2401.02362*.

[18] H. Qin, W. Chen, Z. Li, Q. Wu, N. Cheng, and F. Chen, "Antenna positioning and beamforming design for fluid antenna-assisted multi-user downlink communications," *IEEE Wireless Commun. Lett.*, vol. 13, no. 4, pp. 1073–1077, Apr. 2024.

[19] G. Hu et al., "Fluid antennas-enabled multiuser uplink: A low-complexity gradient descent for total transmit power minimization," *IEEE Commun. Lett.*, vol. 28, no. 3, pp. 602–606, Mar. 2024.

[20] Y. Zuo, J. Guo, B. Sheng, C. Dai, F. Xiao, and S. Jin, "Fluid antenna for mobile edge computing," *IEEE Commun. Lett.*, vol. 28, no. 7, pp. 1728–1732, Jul. 2024.

[21] W. Xie et al., "Movable antenna-assisted covert communications with reconfigurable intelligent surfaces," *IEEE Internet Things J.*, vol. 12, no. 9, pp. 12369–12382, May 2025.

[22] D. Zhang, S. Ye, M. Xiao, K. Wang, M. Di Renzo, and M. Skoglund, "Fluid antenna array enhanced over-the-air computation," *IEEE Wireless Commun. Lett.*, vol. 13, no. 6, pp. 1541–1545, Jun. 2024.

[23] S. Pakravan, M. Ahmadzadeh, M. Zeng, G. Abed Hodtani, J.-Y. Chouinard, and L. A. Rusch, "Robust resource allocation for over-the-air computation networks with fluid antenna array," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Sep. 2024, pp. 1–6.

[24] N. Li, P. Wu, B. Ning, L. Zhu, and W. Mei, "Over-the-air computation via 2D movable antenna array," *IEEE Wireless Commun. Lett.*, vol. 14, no. 1, pp. 33–37, Jan. 2025.

[25] S. Ye, D. Zhang, M. Xiao, and M. Skoglund, "Integrated communication and computation empowered by fluid antenna array," in *Proc. IEEE 25th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, pp. 276–280, Sep. 2024.

[26] M. Ahmadzadeh, S. Pakravan, and G. A. Hodtani, "Movable antenna design for UAV-aided federated learning via deep reinforcement learning," in *Proc. IEEE 15th Int. Conf. Inf. Knowl. Technol. (IKT)*, Dec. 2024, pp. 91–95,

[27] M. Ahmadzadeh, S. Pakravan, G. Abed Hodtani, M. Zeng, J.-Y. Chouinard, and L. A. Rusch, "Enhanced over-the-air federated learning using AI-based fluid antenna system," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2025, pp. 1–6.

[28] S. Park and H. Seo, "Federated learning meets fluid antenna: Towards robust and scalable edge intelligence," Mar. 2025, *arXiv:2503.03054*.

[29] S. Sheikhzadeh, M. Pourghasemian, M. R. Javan, N. Mokari, and E. A. Jorswieck, "AI-based secure NOMA and cognitive radio-enabled green communications: Channel state information and battery value uncertainties," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 2, pp. 1037–1054, Jun. 2022.

[30] K.-K. Wong, K.-F. Tong, Y. Chen, and Y. Zhang, "Fast fluid antenna multiple access enabling massive connectivity," *IEEE Commun. Lett.*, vol. 27, no. 2, pp. 711–715, Feb. 2023.

[31] J. Huangfu et al., "Performance analysis of fluid antenna system under spatially-correlated Rician fading channels," May 2025, *arXiv:2505.15200*.

[32] C. Psomas, G. M. Kraidy, K.-K. Wong, and I. Krikidis, "On the diversity and coded modulation design of fluid antenna systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 3, pp. 2082–2096, Jul. 2024.

[33] K.-K. Wong and K.-F. Tong, "Fluid antenna multiple access," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 4801–4815, Jul. 2022.

[34] J. Yao and N. Ansari, "Secure federated learning by power control for Internet of Drones," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 4, pp. 1021–1031, Dec. 2021.

[35] D. Zhang et al., "Training beam sequence design for mmWave tracking systems with and without environmental knowledge," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10780–10795, Dec. 2022.

[36] A. Gharehgoli, A. Nouruzi, N. Mokari, P. Azmi, M. R. Javan, and E. A. Jorswieck, "AI-based resource allocation in end-to-end network slicing under demand and CSI uncertainties," *IEEE Trans. Netw. Service Manag.*, vol. 20, no. 3, pp. 3630–3651, Sep. 2023.

[37] L. Zhu, W. Ma, B. Ning, and R. Zhang, "Movable-antenna enhanced multiuser communication via antenna position optimization," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 7214–7229, Jul. 2024.

[38] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2022–2035, Mar. 2020.

[39] Y. Cui, T. Lv, W. Li, W. Ni, and E. Hossain, "Over-the-air federated learning in user-centric cell-free networks," *IEEE Wireless Commun. Lett.*, vol. 13, no. 12, pp. 3683–3687, Dec. 2024.

[40] T. Haarnoja, Z. Aurick, A. Pieter, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. IEEE Int. Conf. Mach. Learn.*, Jul. 2018, pp. 1861–1870.