Partially Observable Mean Field Learning for Opportunistic FAMA

Huixian Gu, *Graduate Student Member, IEEE*, Kai Liang, *Member, IEEE*, Gan Zheng, *Fellow, IEEE*, Kai-Kit Wong, *Fellow, IEEE* and Chan-Byoung Chae, *Fellow, IEEE*

***Abstract*—In this letter, we investigate an opportunistic fluid antenna multiple access (O-FAMA) system in a multiuser downlink channel, where a multi-antenna base station (BS) communicates to many user equipments (UEs) on the same time-frequency resource block without precoding and each UE decides on its own whether it should access the channel at any given time, and if it does, utilizes solely its fluid antenna to overcome the inter-user interference. This setup is designed to combine decentralized opportunistic scheduling and FAMA for sum-rate maximization. Specifically, the joint optimization of opportunistic UE scheduling and port selection is prohibitively complex even if it is performed in a centralized fashion. To address this challenge, we propose a low-complexity mean field reinforcement learning approach, for each UE to select its best port and make scheduling decision based on the mean field behavior of other UEs and its own partially observable signal-to-interference-plus-noise ratio (SINR) at the ports. Numerical results show that the proposed method not only reduces training time but also outperforms benchmark solutions in terms of sum-rate performance in O-FAMA systems.**

***Index Terms*—6G, Fluid antenna system (FAS), fluid antenna multiple access (FAMA), mean field, reinforcement learning.**

I. Introduction

THE fluid antenna system (FAS) introduces a novel communication paradigm by enabling dynamic repositioning of radiating elements, offering a new degree of freedom (DoF) through position flexibility [1], [2], [3]. This enhanced spatial diversity and multiplexing capability make FAS an appealing technology for achieving the high reliability, hyper-low latency and massive connectivity that are crucial for future wireless systems such as the sixth generation (6G) [4]. The concept of FAS was explained through the lens of electromagnetic theory in [5]. Recently, FAS prototypes using different technologies and experiments have also been reported [6], [7], [8].

One of the main applications in FAS is multiple access. In [9], Wong *et al.* proposed that inter-user interference can be overcome by simply adjusting the receiver's antenna position with FAS, without the need of any interference management such as precoding or power control. This is possible because radio signal *fades* in scattering environments. While [9] introduced the concept of fluid antenna multiple access (FAMA), the version of FAMA in [9] demanded the knowledge of a symbol-level signal-to-interference plus noise energy ratio that is not known possible. This has motivated the development of a more practical FAMA scheme, called slow FAMA, in [10], which only requires to know the average signal-to-interference plus noise ratio (SINR) at the ports [11] and can easily handle several users on the same physical channel. Recent research has also shown the amazing impact of channel coding on slow FAMA [12] and in [13], it was demonstrated that slow FAMA with the codec in the fifth generation (5G) New Radio (NR) can accommodate a much greater number of user equipments (UEs) than the fixed-position antenna (FPA) counterpart.

Despite the encouraging results thus far, an uncoded slow FAMA system has limited interference rejection capability. To tackle this problem, [14] contemplated the idea of combining opportunistic scheduling and FAMA and revealed tremendous capacity benefits of an opportunistic FAMA (O-FAMA) system. Opportunistic scheduling, however, requires a centralized scheduler with channel state information (CSI) (including all the crosstalk channels) of all UEs, to perform an exhaustive search, which is infeasible. Motivated by the successes in [15], [16], the authors of [17] proposed a decentralized reinforcement learning (RL) framework for O-FAMA to autonomously select favorable UEs and their best FAS ports, for maximizing the network sum-rate. The technique in [17] further enhanced policy learning through derivative-based reward shaping and a dual actor-critic structure, but suffers from very high training complexity and lacks scalability for large networks.

This letter aims to reduce the computational complexity for realizing O-FAMA systems for maximizing the sum-rate. The proposed approach is built upon mean field learning [18], in which the port selection decision of each UE is influenced by the average behavior (i.e., mean action) of all other UEs in the network. To the best of our knowledge, this is the first time to utilize a mean field learning approach to solve port selection and opportunistic scheduling in O-FAMA systems. Our main contributions are summarized as follows:

- First, we formulate a sum-rate maximization problem for the O-FAMA network as a stochastic game by jointly optimizing user scheduling and port selection at each UE, while taking into account partially observable SINR at the UEs. This formulation enables each agent (i.e., UE) to make sequential decisions under uncertainty and dynamic interactions, providing a principled framework to model long-term performance and multi-agent coordination.
- In large-scale systems, stochastic games suffer from the curse of dimensionality, since the joint action and state spaces grow exponentially with the number of UEs, making policy learning and coordination intractable. To address this problem, we propose a partially observable mean field learning approach, in which each UE makes decisions based on its own observable SINR and the mean field action of other UEs. As the number of UEs increases, the computational complexity remains nearly constant, since both policy and value updates rely on the mean field representation rather than individual actions. This makes the proposed approach highly scalable and

H. Gu and K. Liang are with are with the State Key Laboratory of Integrated Service Networks, the School of Telecommunications Engineering, Xidian University, Xi'an, 710071, China.

G. Zheng is with the School of Engineering, University of Warwick, Coventry, United Kingdom.

K. K. Wong is with the Department of Electronic and Electrical Engineering, University College London, United Kingdom, and also affiliated with Yonsei Frontier Laboratory, Yonsei University, Seoul, 03722, Korea.

C. B. Chae is with School of Integrated Technology, Yonsei University, Seoul, 03722, Korea.

well-suited for large-scale network scenarios.

II. System Model and Problem Formulation

We consider a downlink multiuser system in which a multi-FPA base station (BS) simultaneously transmits independent data stream to M users over a shared time-frequency resource block. The BS is equipped with a fixed antenna array, where each antenna is designated for communication with a specific UE. Each UE has a FAS spanning a linear region of $W\lambda$, where λ is the carrier wavelength, and K ports are uniformly distributed along this space. We assume that each FAS port can be considered as a point source and can be instantaneously switched amongst the K preset locations without delay. Thus, the received signal at the k-th port of UE u is given by

$$y_k^{(u)} = g_k^{(u,u)} v_u + \sum_{\substack{\tilde{u}=1 \\ \tilde{u}\neq u}}^{M} g_k^{(\tilde{u},u)} v_{\tilde{u}} + \eta_k^{(u)}, \tag{1}$$

where $g_k^{(\tilde{u},u)}$ is the fading coefficient between the $\tilde{u}$-th FPA of the BS and the k-th port of UE u, v_u is the data symbol transmitted to UE u, with $\mathbb{E}[|v_u|^2] = \sigma_v^2$, and $\eta_k^{(u)}$ is the additive white Gaussian noise (AWGN) at the k-th port of UE u, modeled as a circularly symmetric complex Gaussian variable with zero mean and variance σ_η^2 (i.e., $\eta_k^{(u)} \sim \mathcal{CN}(0, \sigma_\eta^2)$).

We assume that all channels follow a Rayleigh distribution with $\mathbb{E}[|g|^2] = \sigma^2$. Since the ports of a UE's FAS can be arbitrarily close to one another, the channel coefficients $g_k^{(\tilde{u},u)}$ corresponding to different ports of the u-th UE exhibit spatial correlation which can be modeled as

$$\mathbf{g} = \mathbf{A}\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1} & a_{K2} & \cdots & a_{KK} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{bmatrix}. \tag{2}$$

For notational brevity, the indices k, u, and $\tilde{u}$ are omitted. The variables $\{x_k\}_{k=1}^K$ are modeled as independent and identically distributed (i.i.d.) standard complex Gaussian random variables, i.e., $x_k \sim \mathcal{CN}(0,1)$, while $\mathbf{A}$ serves to impose the desired spatial correlation among the channel coefficients.

In our case of a linear fluid antenna with size $W\lambda$ operating in a rich two-dimensional isotropic scattering environment, the spatial correlation between any pair of ports follows the classical Jake's model, which gives

$$\mathbb{E}[g_k g_l^*] = \phi_{k,l} = \sigma^2 J_0\left(\frac{2\pi(k-l)W}{K-1}\right), \tag{3}$$

in which $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind. With $\mathbf{\Phi} \triangleq [\phi_{k,l}]$, we have $\mathbf{\Phi} = \mathbb{E}[\mathbf{g}\mathbf{g}^\dagger] = \mathbf{A}\mathbf{A}^\dagger$.

Let $\mathbf{g} \triangleq [g_1, \ldots, g_K]^T$ denote the vector of channel coefficients. Applying eigenvalue decomposition to the correlation matrix $\mathbf{\Phi}$ yields $\mathbf{\Phi} = \mathbf{V}\mathbf{D}\mathbf{V}^\dagger$, where $\mathbf{D}$ denotes a diagonal matrix containing the eigenvalues, and $\mathbf{V}$ is a unitary matrix composed of the corresponding eigenvectors. Accordingly, the matrix $\mathbf{A}$ can be constructed as $\mathbf{A} = \mathbf{V}\mathbf{D}^{1/2}$ to ensure that the required correlation structure is satisfied. Followed by [17] the distances between the BS and UEs, as well as the effects of path loss, are intentionally omitted. This abstraction allows us to focus on evaluating the performance of O-FAMA.

In O-FAMA, $\tilde{M}$ favorable UEs are selected from a total of M UEs to communicate over a shared time-frequency resource block. Ideally, the SINRs of all UEs are evaluated and ranked, and the top $\tilde{M}$ users are chosen for transmission. For a given UE u, the SINR observed at its k-th port is expressed as

$$\Gamma_k^{(u)} = \frac{\mu_u \sigma_s^2 |g_k^{(u,u)}|^2}{\sigma_s^2 \sum_{\substack{\tilde{u}=1 \\ \tilde{u}\neq u}}^{M} \mu_{\tilde{u}} |g_k^{(\tilde{u},u)}|^2 + \sigma_\eta^2} \text{ s.t. } \sum_{\tilde{u}=1}^{M} \mu_{\tilde{u}} = \tilde{M}, \tag{4}$$

where $\mu_u = \{0,1\}$ indicates if UE u is selected.

Our aim is to solve:

$$\max_{\mathcal{K},\boldsymbol{\mu}} \quad \sum_{u=1}^{M} \log_2(1 + \Gamma_{k_u}^{(u)}) \tag{5a}$$

$$\text{s.t.} \quad \mathrm{C}_1 : \mu_u = \{0,1\}, \forall u, \tag{5b}$$

$$\mathrm{C}_2 : 2 \leq \sum_{u=1}^{M} \mu_u \leq M, \tag{5c}$$

$$\mathrm{C}_3 : k_u = [1, \ldots, K], \forall u, \tag{5d}$$

$$\mathrm{C}_4 : k_u^* = \arg\max_k \{\Gamma_k^{(u)}, k \in \mathcal{P}\}, \tag{5e}$$

where $\mathcal{K} = \{k_1, \ldots, k_u, \ldots, k_M\}$ is the port selection decision of the favourable UEs, and $\boldsymbol{\mu} = \{\mu_1, \ldots, \mu_u, \ldots, \mu_M\}$ is the scheduling decision of all UEs, (5b) specifies the binary scheduling decision for all UEs, (5c) indicates that the total number of selected UEs should not exceed the permissible range of the network, (5d) specifies the port indices, and (5e) is the availability of SINR information at the ports. Note that after solving (5), $\tilde{M}$ will be found, which is a random variable according to the CSI of all the UEs.

Solving (5) faces major challenges: (i) the port selection of UE k is affected by the other UEs' decisions; (ii) in massive access scenarios, it is impossible to compute the optimal port selection when considering the port selection of other UEs.

III. Mean Field Learning

To address the above challenges, we reformulate the problem as a stochastic game, which provides a principled framework for sequential decision-making. However, as the number of agents increases, traditional methods suffer from the curse of dimensionality, making joint optimization in large-scale systems intractable. Moreover, due to the limited observability of O-FAMA ports, it is difficult for each agent to obtain global information. To overcome these issues, we incorporate the mean field approximation and propose a partially observable mean field actor-critic algorithm, enabling scalable and decentralized decision-making under partial observability. The key elements in stochastic game can be represented using a five tuple: $(\mathcal{U}, \mathcal{S}, \mathcal{O}, \mathcal{A}, \mathcal{R})$. Each element is explained below.

*1) **Agent** $\mathcal{U}$:* Each EU acts as an agent to learn the optimal port selection policy. We have the UE set, $\mathcal{U} = \{1, \ldots, M\}$.

*2) **State space** $\mathcal{S}$:* The state space $\mathcal{S}$ includes all UEs' SINR information at each port at time slot t. The state space of agent u can be represented as $s_u = \{\Gamma_k^{(u)}\}_{k\in\mathcal{K}}$.

*3) **Observable space** $\mathcal{O}$:* Each UE can only access its own observation at the observable ports, which consists of the available SINR information, $o_u = \{\Gamma_k^{(u)}\}_{k\in\mathcal{P}}$, $|\mathcal{P}| \ll |\mathcal{K}|$.

*4) **Action space** $\mathcal{A}$:* The action space of each UE is the port selection decision and the opportunistic scheduling decision, which can be represented as $a_u = \{k_u, \mu_u\}$. Due to the fact that $k_u = [0, \ldots, k_u, \ldots, 0]_{[1\times K]}$ and $\mu_u \in \{0, 1\}$ are discrete values, at any time step, only one port can be activated, so one-hot encoding can be adopted for action space.

*5) **Reward** $\mathcal{R}$:* Our objection is to maximize the sum rate. Thus, the agent receives a reward of $r = \sum_{u=1}^{M} \log_2(1+\Gamma_{k_u^*}^{(u)})$ if all the constraints are met; otherwise, $r = 0$.

The Nash equilibrium is an important concept in stochastic games. It represents a joint policy $\boldsymbol{\pi}^* = [\pi_1^*, \ldots, \pi_M^*]$, such that for each $u \in [1, \ldots, M]$, $v_u(o_u, \boldsymbol{\pi}^*) = v_u(o_u, \pi_u^*, \boldsymbol{\pi}_{-u}^*)$, where $\boldsymbol{\pi}_{-u}^* = [\pi_1^*, \ldots, \pi_{u-1}^*, \pi_{u+1}^*, \ldots, \pi_M^*]$. The best response of UE u is defined as $v_u(o_u, \pi_u^*, \boldsymbol{\pi}_{-u}^*) \geq v_u(o_u, \pi_u^*, \boldsymbol{\pi}_{-u})$, where π_u^* denotes the optimal policy of UE u, and $\boldsymbol{\pi}_{-u}$ represents the policies of all other UEs. If every UE's policy, π_u^*, satisfies this best response condition with respect to the others, the joint policy profile $\boldsymbol{\pi}^* = (\pi_u^*)_{u=1}^M$ constitutes a Nash equilibrium. In multi-agent RL, Nash Q-learning computes the Nash equilibrium through two iterative steps [19]. However, as the number of agents increases, the joint action space grows exponentially, making it unsuitable for large-scale multi-agent interaction scenarios.

Mean field approximation has been investigated in [18] to reduce the computational complexity of Nash Q-learning. In this case, we have $a_u = \bar{a}_{u'} + \delta a_{u,u'}$, in which $\bar{a}_{u'} = \frac{1}{M-1}\sum_u \bar{a}_u$ and $\delta a_{u,u'}$ is a small fluctuation. Additionally, $\bar{a}_{u'}$ is the action distribution of the neighbors of agent u'. Therefore, the Q-function can be expressed as $Q_u(o_u, \boldsymbol{a}) = \frac{1}{M-1}\sum_{u'} Q_u(o_u, a_u, a_{u'}) = Q_u(o_u, a_u, \bar{a}_u)$. During the learning phase, the action is selected by the Boltemann policy:

$$\pi_{\theta_u}(a_u|o_u, \bar{a}_u) = \frac{\exp(\beta Q_u(o_u, a_u, \bar{a}_u))}{\sum_{a_{u'}\in\mathcal{A}_u}\exp(\beta Q_u(o_u, a_u, \bar{a}_u))}, \quad (6)$$

where

$$Q_u^{t+1}(s_u, a_u, \bar{a}_u) = (1-\alpha_t)Q_u^t + \alpha[r_u + \gamma v_u^t(s')], \quad (7)$$

where α_t is the learning rate. The mean field value function $v_u^t(s')$ is found by

$$v_u^t(s') = \sum_{a_u} \pi_u^t(a_u|s', \bar{a}_u)\mathbb{E}_{\bar{a}_u(a_{-u})\sim\pi_{-u}^t} Q_u^t(s'). \quad (8)$$

With the mean field approximation, the best response of agent u can be calculated based on the mean field action of agent u. Therefore, we can implement the mean field learning by function approximation such as neural networks, where $Q_u(o_u, a_u, \bar{a}_u)$ is parameterized by weights ϕ. We can also use a neural network to model the policy selection process with weights θ. The update rules for Q-function and policy selection lead to an actor-critic approach.

In mean field actor-critic, the agent is trained by minimizing the loss function

$$\mathcal{L}(\phi_u) = (y_u - Q_{\phi_u}(o_u, a_u, \bar{a}_u))^2, \quad (9)$$

where $y_u = r_u + \gamma v_{\phi_u'}(s')$ is the target mean field value with the weight ϕ_u'. Differentiating $\mathcal{L}(\phi_u)$ gives

$$\nabla_{\phi_u}\mathcal{L}(\phi_u) = (y_u - Q_{\phi_u}(o_u, a_u, \bar{a}_u))\nabla_{\phi_u} Q_{\phi_u}(o_u, a_u, \bar{a}_u). \quad (10)$$

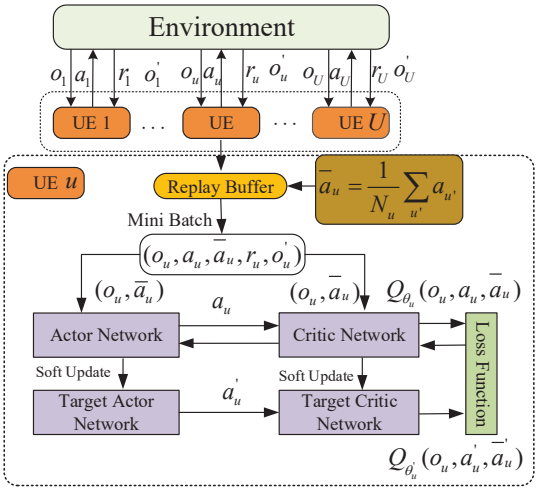

Fig. 1. The overall mean field learning framework.

Additionally, the actor network π_{θ_u} is trained by the sampled policy gradient

$$\nabla_{\theta_u}\mathcal{J}(\theta_u) \approx \nabla_{\theta_u}\log\pi_{\theta_u}(o_u)Q_{\phi_u}(o_u, a_u, \bar{a}_u). \quad (11)$$

The overall mean field learning framework is shown in Fig. 1, and the proposed algorithm is presented as Algorithm 1.

Algorithm 1 Mean Field Actor-Critic (MFAC) Algorithm for Port Selection and Opportunistic Scheduling in O-FAMA

Initialize parameters: ϕ^u and $\phi^u_{'}$, θ^u and $\theta^u_{'}$, initialize the mean action $\bar{a}_u$ for all $u \in \{1, \ldots, M\}$, initialize the total number of episode E and total number of step T.
while Episode < E **do**
  while Step < T **do**
    For each agent u choose action a_u from π_{θ_u}, and compute new mean action $\bar{\mathbf{a}} = [\bar{a}_1, \ldots, \bar{a}_M]$.
    Execute the joint action $\mathbf{a}$, observe the reward $\mathbf{r}$, and the next state $\mathbf{s}'$.
    Store $(o, \mathbf{a}, \mathbf{r}, \mathbf{s}', \bar{\mathbf{a}})$ in replay buffer $\mathcal{B}$.
  end while
  while $j = 1, \ldots, U$ **do**
    Sample a minibatch of Z samples $(o, \mathbf{a}, \mathbf{r}, \mathbf{s}', \bar{\mathbf{a}})$ from replay buffer $\mathcal{B}$.
    Set $y_u = r_u + \gamma v_{\phi_u'}(s')$.
    Update the critic network by minimizing the loss: $L(\phi_u) = \frac{1}{Z}\sum(y_u - Q_{\phi_u}(o_u, a_u, \bar{a}_u))^2$.
    Update the actot network by (11).
  end while
  Update the parameters of the target network for each agent u with learning rate: τ; $\phi_u' \leftarrow \tau\phi_u + (1-\tau)\phi_u'$, $\theta_u' \leftarrow \tau\theta_u + (1-\tau)\theta_u'$.
end while

The computational complexity of the proposed algorithm depends on the architecture of the critic network. Let h_i^{critic} represent the number of neurons in the i-th hidden layer. Then the complexity of our approach can be expressed as $\mathcal{O}\left((|S_u| + 2|A_u|)h_1^{\text{critic}} + \sum_{i=2}^{I-1} h_i^{\text{critic}} h_{i+1}^{\text{critic}} + h_I^{\text{critic}}\right)$, where only the local state and mean action are used as inputs. In contrast, the traditional actor-critic approach takes the actions of all M agents as input, resulting in a complexity of $\mathcal{O}\left((|S_u| + M|A_u|)h_1^{\text{critic}} + \sum_{i=2}^{I-1} h_i^{\text{critic}} h_{i+1}^{\text{critic}} + h_I^{\text{critic}}\right)$, which increases linearly with the number of agents [17]. Therefore,

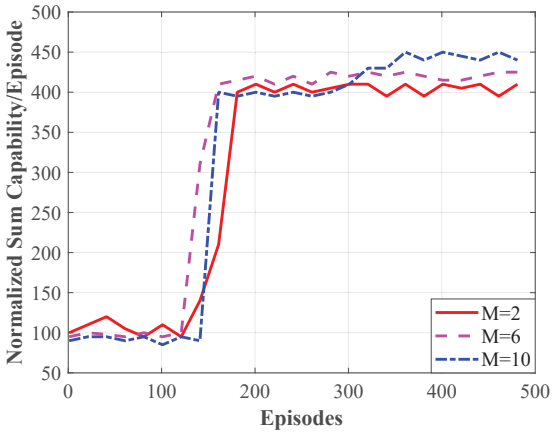

Fig. 2. Convergence under different number of M.

by leveraging the mean field approximation, the proposed approach ensures that the computational complexity does not increase with the number of agents. This design effectively compresses the input dimension of the critic network, thereby significantly reducing the overall computational complexity.

IV. Simulation Results

In the simulations, the actor network at each UE takes as input the agent's local observation and the mean action of other agents. It consists of five fully connected layers with hidden sizes 1024, 512, 256, and 128, followed by a softmax output layer to produce action probabilities. The output of critic part is the Q-value through a similar architecture. ReLU activation is applied to all hidden layers. Both networks are optimized using Adam with a learning rate of 0.0005, and soft target updates with $\tau = 0.05$ are applied. A learning rate scheduler is used to decay the learning rate by 0.95 every 50 steps, and the total number of episodes is 400. The system parameters are set as: $K = 100$, with $M = 6$, $W = 2$, and $|\mathcal{P}| = 30$.

We compare the proposed approach with (1) the **MA SD3-DerivNet** scheme proposed in [17] which employs dual policy networks and high-order derivative estimations to enhance the reward signal and (2) **exhaustive search** which evaluates all possible action combinations to obtain the optimal solution, is adopted as the performance upper bound.

As can be observed from Fig. 2, the proposed algorithm consistently achieves relatively stable convergence within approximately 50 episodes. This is primarily attributed to the incorporation of the mean field action, which abstracts the interactions among agents (i.e., UEs) into a low-dimensional representation. As a result, the input size of the value network remains invariant with respect to the number of agents, thereby enabling the algorithm to maintain stable learning performance regardless of the scale of the multi-agent system.

Table I shows the training time of the proposed approach and MA SD3-DerivNet under different number of UEs. As the number of UEs increases from 3 to 5, the training time of MA SD3-DerivNet grows rapidly, reaching over $12,000$ seconds, due to the exponential increase in joint action space. In contrast, our proposed approach maintains a significantly lower and more stable training time (405s to 662s), benefiting from the mean field approximation. This highlights our superior scalability and efficiency in large-scale O-FAMA.

TABLE I
Training Time vs. Different Number of UEs

Algorithm \ Number of UEs, M	3	4	5
The proposed method	405s	542s	662s
MA SD3-DerivNet [17]	9600s	11200s	12800s

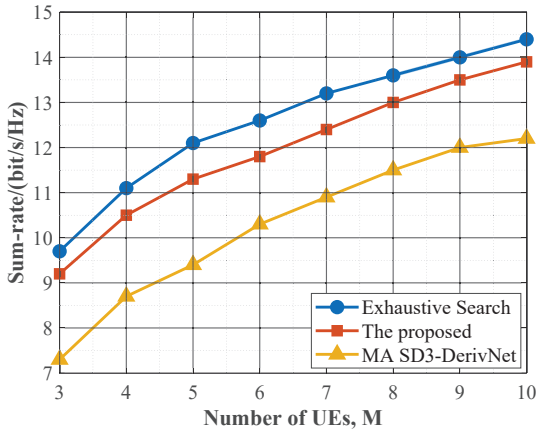

Fig. 3. Sum-rate vs. different number of UEs.

In Fig. 3, we show the sum-rate performance by varying the number of UEs from 3 to 10. We see that the proposed approach consistently outperforms the benchmark scheme. For instance, when $M = 6$, the proposed approach achieves a sum-rate improvement of approximately 13.0% over MA SD3-DerivNet. As M increases to 10, the performance gap widens further, with the proposed approach achieving a sum-rate that is about 13.9% higher than MA SD3-DerivNet. This performance gain stems from the design of the proposed MFAC approach. Unlike MA SD3-DerivNet, which introduces dual policy networks and high-order derivative networks to enhance the reward but incurs significant computational overhead as the number of agents grows, the proposed approach leverages the mean field approximation. This abstraction replaces the modeling of joint actions with a representative mean action, thereby avoiding the complexity of the high-dimensional action space. Consequently, each agent optimizes its policy based on local information and the mean field action, leading to improved scalability and a significant increase in system sum-rate.

Fig. 4 illustrates the sum-rate performance versus the number of ports K at the FAS of each UE. Compared with the MA SD3-DerivNet algorithm, the proposed approach consistently achieves better performance, especially as the number of ports increases. For example, when $K = 60$, the proposed approach achieves a sum-rate of 11.5 bps/Hz, which is 11.7% higher than that of MA SD3-DerivNet (10.3 bps/Hz). As K increases to 100, the performance gap becomes more pronounced, with the proposed approach reaching 11.8 bps/Hz, while MA SD3-DerivNet remains saturated at 10.3 bps/Hz, resulting in a gain of 14.6%. This gain stems from the efficiency of the MFAC

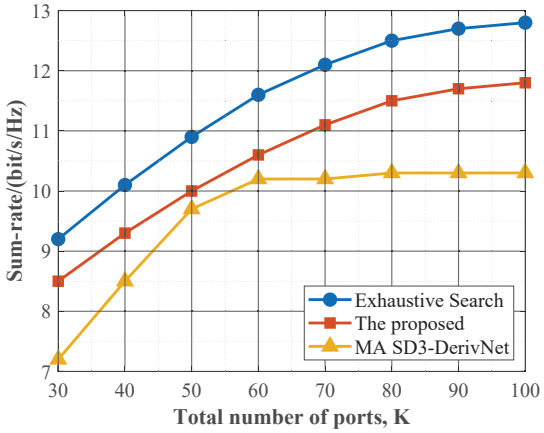

Fig. 4. Sum-rate vs. different number of ports.

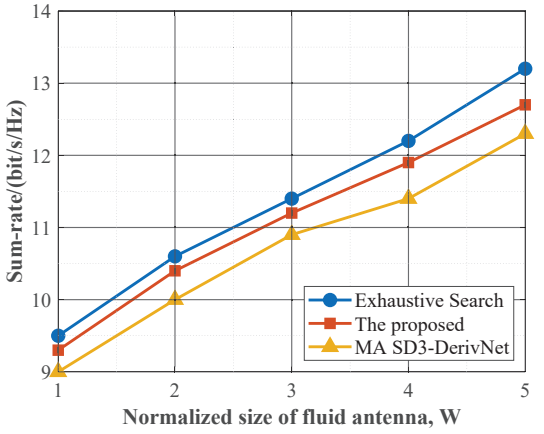

Fig. 5. Sum-rate vs. different normalized size of fluid antenna.

algorithm. Each agent can make more informed decisions by considering the aggregated behavior of others, rather than modeling the full joint action space explicitly.

Finally, the results in Fig. 5 are provided to examine the impact of the normalized size of the FAS at each UE, denoted by W. The results demonstrate that the proposed approach consistently outperforms the benchmark scheme. For instance, when $W = 3$, the proposed approach achieves a sum-rate of 11.2 bps/Hz, compared to 10.9 bps/Hz for MA SD3-DerivNet. As W increases to 5, the proposed approach reaches 12.7 bps/Hz, while MA SD3-DerivNet only achieves 12.3 bps/Hz, yielding a relative gain of about 3.3%. This gain arises from the MFAC algorithm's ability to capture average interaction effects through mean field approximation. Unlike MA SD3-DerivNet, which again relies on higher-order derivative networks and suffers from increased computational complexity in more dynamic environments, MFAC maintains both scalability and decision quality as W increases. By abstracting the joint actions of other agents into a representative mean action, our approach allows each agent to make efficient decisions even as the system becomes more complex.

V. Conclusion

In this letter, we proposed a mean field learning approach to address the joint port selection and opportunistic scheduling problem in O-FAMA systems, aiming to maximize the sum-rate while taking into account partial SINR observability at the UEs. Our approach revealed superior performance compared to the MA SD3-DerivNet baseline [17] in terms of training complexity, achievable sum-rate, and computational efficiency. Numerical results validated the effectiveness of our mean field learning framework for addressing the challenges in O-FAMA systems. In the future, it would be useful to accommodate heterogeneous FAS configurations across the UEs, incorporate beamforming at the BS, and explore integrated communication and sensing capabilities within the O-FAMA system.

References

[1] K. K. Wong, K.-F. Tong, Y. Zhang, and Z. Zheng, "Fluid antenna system for 6G: When Bruce Lee inspires wireless communications," *Elect. Lett.*, vol. 56, no. 24, pp. 1288–1290, Nov. 2020.

[2] K. K. Wong, A. Shojaeifard, K.-F. Tong, and Y. Zhang, "Performance limits of fluid antenna systems," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2469–2472, Nov. 2020.

[3] 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.

[4] W. K. New *et al.*, "A tutorial on fluid antenna system for 6G networks: Encompassing communication theory, optimization methods and hardware designs," *IEEE Commun. Surv. & Tuts.*, doi:10.1109/COMST.2024.3498855, 2024.

[5] W.-J. Lu *et al.*, "Fluid antennas: Reshaping intrinsic properties for flexible radiation characteristics in intelligent wireless networks," *IEEE Commun. Mag.*, vol. 63, no. 5, pp. 40–45, May 2025.

[6] Y. Shen *et al.*, "Design and implementation of mmWave surface wave enabled fluid antennas and experimental results for fluid antenna multiple access," *arXiv preprint*, arXiv:2405.09663, May 2024.

[7] B. Liu, K.-F. Tong, K. K. Wong, C.-B. Chae, and H. Wong, "Programmable meta-fluid antenna for spatial multiplexing in fast fluctuating radio channels," *Optics Express*, vol. 33, no. 13, pp. 28898–28915, 2025.

[8] J. Zhang *et al.*, "A novel pixel-based reconfigurable antenna applied in fluid antenna systems with high switching speed," *IEEE Open J. Antennas Propag.*, vol. 6, no. 1, pp. 212–228, Feb. 2025.

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

[10] K. K. Wong, D. Morales-Jimenez, K. F. Tong, and C. B. Chae, "Slow fluid antenna multiple access," *IEEE Trans. Commun.*, vol. 71, no. 5, pp. 2831–2846, May 2023.

[11] M. Eskandari, A. G. Burr, K. Cumanan and K. K. Wong, "cGAN-Based slow fluid antenna multiple access," *IEEE Wireless Commun. Lett.*, vol. 13, no. 10, pp. 2907–2911, Oct. 2024.

[12] H. Hong, K. K. Wong, K. F. Tong, H. Shin, and Y. Zhang, "Coded fluid antenna multiple access over fast fading channels," *IEEE Wireless Commun. Lett.*, vol. 14, no. 4, pp. 1249–1253, Apr. 2025.

[13] H. Hong *et al.*, "Downlink OFDM-FAMA in 5G-NR systems," *IEEE Trans. Wireless Commun.*, doi:10.1109/TWC.2025.3577771, 2025.

[14] K. K. Wong, K. F. Tong, Y. Chen, Y. Zhang, and C.-B. Chae, "Opportunistic fluid antenna multiple access," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 7819–7833, Nov. 2023.

[15] C. Wang *et al.*, "Fluid antenna system liberating multiuser MIMO for ISAC via deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 10 879–10 894, Sep. 2024.

[16] C. Wang *et al.*, "AI-empowered fluid antenna systems: Opportunities, challenges, and future directions," *IEEE Wireless Commun.*, vol. 31, no. 5, pp. 34–41, Oct. 2024.

[17] N. Waqar *et al.*, "Opportunistic fluid antenna multiple access via team-inspired reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 12068–12083, Sep. 2024.

[18] Y. Yang *et al.*, "Mean field multi-agent reinforcement learning," in *Proc. Mach. Learn. Res. (ICML)*, pp. 5567–5576, Jul. 2018.

[19] J. Hu and M. P. Wellman, "Nash Q-learning for general-sum stochastic games", *J. Machine Learning Research*, vol. 4, no. 11, pp. 1039–1069, Nov. 2003.