Deep Unfolding Neural Networks for Fluid Antenna-Enhanced Vehicular Communication

Biqian Feng, Chenyuan Feng, *Member, IEEE*, Kai-Kit Wong, *Fellow, IEEE*, and Tony Q. S. Quek, *Fellow, IEEE*

***Abstract*—Fluid antenna (FA) technology has emerged as a promising technology to achieve higher spectral and energy efficiency by introducing a new dimension. However, the antenna position configuration inevitably increases computational complexity, presenting challenges under real-time configuration requirements, especially in vehicular communication systems characterized by rapidly time-varying channels. To address these issues, this paper investigates the classical weighted sum rate maximization problem and proposes an optimization-empowered neural network framework designed to accelerate convergence without compromising accuracy. Extensive simulations demonstrate that the proposed approach effectively mitigates the computational burdens associated with FAs, delivering superior performance in terms of convergence rate and system performance, thus paving the way for the deployment of next-generation FA-enabled communication systems.**

***Index Terms*—Fluid antenna, deep unfolding neural networks, weighted sum rate maximization, vehicular communication.**

I. Introduction

With the growing demand for higher capacity and efficiency in vehicular networks, technologies like multiple-input multiple-output (MIMO) have become increasingly prevalent, as they utilize spatial degrees of freedom to enhance system performance [1], [2]. However, traditional MIMO systems based on fixed-position antennas (FPA) face inherent limitations, as the static nature of these antennas hinders further optimization of communication performance. To overcome these challenges, fluid antenna (FA) technology has emerged as a promising solution to dynamically reconfigure their structure in response to varying environmental conditions, such as changes in the wireless channel, user mobility, and interference [3]–[5]. This adaptability is made possible by the use of advanced materials and technologies, such as reconfigurable pixels [6], which facilitate physical reconfiguration of the antenna's shape and position, allowing it to continuously adjust to the varying conditions of the wireless channel in vehicular networks.

Recent years have witnessed the research efforts dedicated to the analysis and enhancement of FA-enhanced systems [7]–[10]. Chen *et al* [7] investigated the deployment of FAs in transceivers to improve the performance of point-to-point MIMO systems. Hu *et al* [8] focused on the multiuser uplink communication system, which consists of a FA-equipped base station (BS) and multiple single-antenna users. Zhang *et al* [9] explored reconfigurable intelligent surfaces (RIS)-assisted FA systems and proposed a genetic algorithm to jointly optimize channel conditions and improve communication performance by locally moving the antenna solely in the receiving area. Furthermore, recent work has investigated optimizing FA performance across a range of configurations, including Unmanned Aerial Vehicle (UAV) network [11], integrated sensing and communication (ISAC) [12], over-the-air computation [13], and mobile edge computing [14]. Despite these advancements, much of the existing literature relies on traditional optimization methods, which, while offering strong interpretability, are often criticized for their slow convergence rates.

In this paper, we investigate an optimization-empowered neural network architecture aimed at balancing interpretability with computational efficiency, facilitating faster convergence without sacrificing optimization accuracy. The main contributions are summarized as follows: i) To provide better capacity in vehicular networks, we formulate a weighted sum rate (WSR) maximization problem within FA-enhanced communication systems and transform it into a more tractable weighted sum mean-square error minimization (WMMSE) problem. ii) To solve this optimization problem, we first employ the block successive upper bound minimization (BSUM) technique to iteratively optimize all variables. Building on this method, we propose a deep unfolding neural network to accelerate the convergence process. iii) Numerical results demonstrate that the FA-enhanced systems ensure superior performance over the conventional systems. Additionally, the proposed deep unfolding neural networks significantly accelerate convergence, thereby reducing computational delay.

Notations: Superscripts $*, T$, and H stand for the conjugate, transpose, and conjugate transpose, respectively. $\Pi_{\mathcal{C}}(\mathbf{x})$ denotes the projection of $\mathbf{x}$ onto the set $\mathcal{C}$.

II. System Model

In this section, we describe the signal model for the considered FA-enhanced vehicular networks, and then formulate

This work was supported in part by the National Natural Science Foundation of China under Grant 62301328, and in part by the SNS JU project 6G-GOALS under the EU's Horizon program Grant Agreement No. 101139232. *(Corresponding author: Chenyuan Feng.)*

B. Feng is with the State Key Laboratory of Internet of Things for Smart City and the Department of Computer and Information Science, University of Macau, Macao. (e-mail: fengbiqian@gmail.com).

C. Feng is with EURECOM, Sophia Antipolis 06410, France (e-mail: Chenyuan.Feng@eurecom.fr).

K. K. Wong is affiliated with the Department of Electronic and Electrical Engineering, University College London, Torrington Place, WC1E 7JE, United Kingdom and he is also affiliated with Yonsei Frontier Lab, Yonsei University, Seoul, Korea. (e-mail: kai-kit.wong@ucl.ac.uk).

T. Q. S. Quek is with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore 487372 (e-mail: tonyquek@sutd.edu.sg).

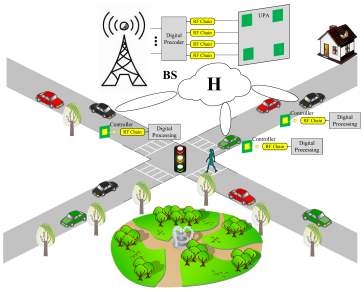

Fig. 1. Illustration of FA-enhanced vehicular communication systems.

a WSR maximization problem to jointly optimize the beamformer and the antenna positions mounted on the vehicles.

A. FA-Enhanced Vehicular Communication Systems

As illustrated in Fig. 1, we consider the downlink wireless communication in vehicular network, where the roadside unit (RSU) equipped with a uniform planar array (UPA) with $M = M_1 \times M_2$ antennas, serves K single-FA vehicles. Compared with the conventional communication system, the position of each FA in the receive region $\mathcal{C}_k^r, k = 1, \cdots, K$, can be adjusted via a controller (e.g., liquid metal) to harness spatial diversity gains. Let $\mathbf{r}_k = (x_k, y_k)^T$ denote the adjustable position of the FA on vehicle k, while $\mathbf{t}_m = (X_m, Y_m)^T$ represents the fixed position of the antenna m at the RSU. By aggregating the positions of all FAs, the antenna position vector is expressed as $\mathbf{r} = (\mathbf{r}_1^T, \cdots, \mathbf{r}_K^T)^T$.

Let L_k^t and L_k^r denote the total numbers of transmit and receive channel paths from the RSU to vehicle k. The signal propagation phase differences of the l-th transmit path for vehicle k between the m-th FA and the reference point $\mathbf{t}_0 = [0,0]^T$ at the RSU and the l-th receive path between this FA and reference point $\mathbf{r}_0 = [0,0]^T$ at vehicle k are, respectively, given by

$$\begin{aligned} \rho_{l,k}^t(\mathbf{t}_m) &= \mathbf{t}_m^T \mathbf{n}_{l,k}^t = X_m \sin\theta_{l,k}^t \cos\phi_{l,k}^t + Y_m \cos\theta_{l,k}^t, \\ \rho_{l,k}^r(\mathbf{r}_k) &= \mathbf{r}_k^T \mathbf{n}_{l,k}^r = x_k \sin\theta_{l,k}^r \cos\phi_{l,k}^r + y_k \cos\theta_{l,k}^r, \end{aligned} \tag{1}$$

where $\mathbf{n}_{l,k}^t \triangleq (\sin\theta_{l,k}^t \cos\phi_{l,k}^t, \cos\theta_{l,k}^t)^T$ and $\mathbf{n}_{l,k}^r \triangleq (\sin\theta_{l,k}^r \cos\phi_{l,k}^r, \cos\theta_{l,k}^r)^T$; $(\theta_{l,k}^t, \phi_{l,k}^t)$ and $(\theta_{l,k}^r, \phi_{l,k}^r)$ are the elevation and azimuth angles-of-arrival (AoAs) for the l-th transmit and receive paths between vehicle k and the RSU, respectively. Accordingly, the field-response vector of the receive channel paths between vehicle k and the m-th FA at the RSU is given by

$$\begin{aligned} \mathbf{f}_k(\mathbf{t}_m) &= \left[e^{j\frac{2\pi}{\lambda}\rho_{1,k}^t(\mathbf{t}_m)}, \cdots, e^{j\frac{2\pi}{\lambda}\rho_{L_k^t,k}^t(\mathbf{t}_m)} \right]^T, \\ \mathbf{g}_k(\mathbf{r}_k) &= \left[e^{j\frac{2\pi}{\lambda}\rho_{1,k}^r(\mathbf{r}_k)}, \cdots, e^{j\frac{2\pi}{\lambda}\rho_{L_k^r,k}^r(\mathbf{r}_k)} \right]^T. \end{aligned} \tag{2}$$

Thus, the channel vector between vehicle k and the RSU is given by

$$\mathbf{h}_k(\mathbf{r}_k) = \mathbf{F}_k^H \mathbf{\Sigma}_k \mathbf{g}_k(\mathbf{r}_k), \tag{3}$$

where $\mathbf{F}_k \triangleq (\mathbf{f}_k(\mathbf{t}_1), \cdots, \mathbf{f}_k(\mathbf{t}_M)) \in \mathbb{C}^{L_k^t \times M}$ denotes the field-response matrix at the RSU, and $\mathbf{\Sigma}_k \in \mathbb{C}^{L_k^t \times L_k^r}$ is the path-response matrix, capturing the multi-path response between all transmit and receive channel paths. Therefore, the discrete-time signal received by vehicle k can be expressed as

$$y_k = \mathbf{h}_k^H(\mathbf{r}_k) \sum_{k=1}^K \mathbf{w}_k s_k + n_k, \tag{4}$$

where $\mathbf{w}_k \in \mathbb{C}^{M\times 1}$ and $s_k \in \mathbb{C}$ represent the beamformer and the unit-power complex-valued information symbols intended for vehicle k, respectively; $n_k \sim \mathcal{CN}(0, \sigma^2)$ is the additive noise modeled as a complex Gaussian random variable with zero mean and variance σ^2. Thus, the signal-to-interference-plus-noise ratio (SINR) experienced at vehicle k is given by

$$\gamma_k \triangleq \frac{|\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_k|^2}{\sum\limits_{i\neq k} |\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_i|^2 + \sigma^2}. \tag{5}$$

B. Problem Formulation

This work focuses on maximizing the WSR of all vehicles by jointly optimizing the transmit beamformer at the RSU and the antenna position vectors at K vehicles, subject to the minimum inter-FA distance at the RSU and the transmit power constraints. Let $\mathbf{W} \triangleq (\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_K) \in \mathbb{C}^{M\times K}$, the WSR maximization problem is formulated as

$$\underset{\mathbf{W},\mathbf{r}}{\text{maximize}} \quad \sum_{k=1}^K \alpha_k \log(1+\gamma_k) \tag{6a}$$

$$\text{subject to} \quad \sum_{k=1}^K \|\mathbf{w}_k\|^2 \le P, \tag{6b}$$

$$\mathbf{r}_k \in \mathcal{C}_k^r, \tag{6c}$$

where the weight $\alpha_k \ge 0$ is the priority weight for vehicle k, P is the maximum transmit power. The non-convexity of (6a) is a huge obstacle for the development of an optimal solution. For the convenience of the design, we integrate the power constraints (6b) into the SINR and reformulate it as

$$\underset{\mathbf{W},\mathbf{r}}{\text{maximize}} \quad \sum_{k=1}^K \alpha_k \log(1+\bar{\gamma}_k) \tag{7a}$$

$$\text{subject to } \mathbf{r}_k \in \mathcal{C}_k^r, \tag{7b}$$

where

$$\bar{\gamma}_k \triangleq \frac{|\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_k|^2}{\sum_{i\neq k} |\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_i|^2 + \frac{\sigma^2}{P}\sum_{i=1}^K \|\mathbf{w}_i\|^2}. \tag{8}$$

Proposition 1. *The problems (6) and (7) are equivalent, as they yield an identical global optimal solution. Moreover, for any given* $\mathbf{r}$*, the optimal solutions* $\mathbf{W}(\mathbf{r})^\#$ *in problem (6) and* $\mathbf{W}(\mathbf{r})^*$ *in problem (7) satisfy*

$$\mathbf{W}(\mathbf{r})^* = \sqrt{P}\frac{\mathbf{W}(\mathbf{r})^\#}{\|\mathbf{W}(\mathbf{r})^\#\|}. \tag{9}$$

Proof. Please refer to the Appendix. □

To reduce the complexity and transform it into a more tractable form, we adopt a linear receive beamforming strategy so that the estimated signal at vehicle k is given by

$$\hat{s}_k = u_k^* y_k, \tag{10}$$

where $\mathbf{u} \triangleq (u_1, u_2, \cdots, u_K)^T \in \mathbb{C}^{K\times 1}$ represents the receive beamformer. Assuming the independence of s_k and n_k, the expected mean-square error (MSE) can be written as

$$\begin{aligned} e_k &\triangleq \mathbb{E}\left[|\hat{s}_k - s_k|^2\right] \\ &= 1 + |u_k|^2(\sigma^2 + \sum_{j=1}^{K} |\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_j|^2) \\ &\quad - 2\mathrm{Re}(u_k^*\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_k) \\ &= 1 + |u_k|^2\left(\frac{\sigma^2}{P}\sum_{i=1}^{K}\|\mathbf{w}_i\|^2 + \sum_{i=1}^{K}|\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_i|^2\right) \\ &\quad - 2\mathrm{Re}(u_k^*\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_k), \end{aligned} \tag{11}$$

where the last equality follows from the optimal solution satisfying $P = \|\mathbf{W}\|^2$. Inspired by the WMMSE algorithm proposed for conventional MIMO systems [15], we introduce an auxiliary optimization variable vector $\mathbf{v} = (v_1, v_2, \cdots, v_K)^T$ and establish a more tractable FA-enabled WMMSE problem as follows

$$\underset{\mathbf{W},\mathbf{r},\mathbf{u},\mathbf{v}}{\text{minimize}} \quad \sum_{k=1}^{K}\alpha_k(v_k e_k - \log(v_k)) \tag{12a}$$

$$\text{subject to } \mathbf{r}_k \in \mathcal{C}_k^r, \tag{12b}$$

$$\mathbf{v} \geq \mathbf{0}. \tag{12c}$$

III. Block Successive Upper Bound Minimization

The BSUM method decomposes the optimization variables into multiple blocks and iteratively optimizes surrogate or upper-bound functions of the original objective in a block-by-block fashion [16]. To ensure systematic updates, we adopt an essentially cyclic (E-C) coordinate selection strategy, which updates each block alternately over a predefined cycle. Particularly, during a period of $N+3$ updates: i) the blocks $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{W}$ are each updated once; ii) the remaining N updates are allocated to the block $\mathbf{r}_k, k = 1, 2\cdots, K$,ensuring each vehicle's antenna position vector receives dedicated attention. This update strategy balances computational efficiency and convergence, leveraging the problem structure to achieve iterative refinement of all variables. Specifically:

- For fixed $\mathbf{r}, \mathbf{W}, \mathbf{v}$, the optimal $\mathbf{u}$ satisfy

$$u_k = \left(\frac{\sigma^2}{P}\sum_{i=1}^{K}\|\mathbf{w}_i\|^2 + \sum_{i=1}^{K}|\mathbf{h}_k(\mathbf{r}_k)^H\mathbf{w}_i|^2\right)^{-1} \mathbf{h}_k(\mathbf{r}_k)^H\mathbf{w}_k, \quad \forall k. \tag{13}$$

- For fixed $\mathbf{r}, \mathbf{W}, \mathbf{u}$, the optimal $\mathbf{v}$ satisfy

$$v_k = \left(1 - u_k^*\mathbf{h}_k(\mathbf{r}_k)^H\mathbf{w}_k\right)^{-1}, \forall k. \tag{14}$$

- For fixed $\mathbf{r}, \mathbf{u}, \mathbf{v}$, the optimal beamformer $\mathbf{W}$ satisfy

$$\mathbf{w}_k = \alpha_k u_k v_k \left(\frac{\sigma^2}{P}\alpha_k|u_k|^2 v_k \mathbf{I}_M + \sum_{i=1}^{K}\alpha_i|u_i|^2 v_i \mathbf{h}_i(\mathbf{r}_i)\mathbf{h}_i(\mathbf{r}_i)^H\right)^{-1} \mathbf{h}_k(\mathbf{r}_k). \tag{15}$$

- For fixed $\mathbf{W}$, $\mathbf{u}$, and $\mathbf{v}$, each $\mathbf{r}_k$ can be updated in parallel, the problem of designing FA position at vehicle k accordingly reduces to

$$\underset{\mathbf{r}_k\in\mathcal{C}_k^r}{\text{minimize}} \ \mathbf{g}_k^H(\mathbf{r}_k)\mathbf{C}_k\mathbf{g}_k(\mathbf{r}_k) + \mathrm{Re}(\mathbf{d}_k^H\mathbf{g}_k(\mathbf{r}_k)), \tag{16}$$

where

$$\begin{aligned} \mathbf{C}_k &\triangleq \sum_{j=1}^{K}|u_k|^2\boldsymbol{\Sigma}_k^H\mathbf{F}_k\mathbf{w}_j\mathbf{w}_j^H\mathbf{F}_k^H\boldsymbol{\Sigma}_k, \\ \mathbf{d}_k &\triangleq -2u_k^*\boldsymbol{\Sigma}_k^H\mathbf{F}_k\mathbf{w}_k. \end{aligned} \tag{17}$$

A majorizing function for the objective function is constructed at $\mathbf{r}_k = \mathbf{r}_{k,0}$ as follows:

$$\begin{aligned} &\mathbf{g}_k^H(\mathbf{r}_k)\mathbf{C}_k\mathbf{g}_k(\mathbf{r}_k) + \mathrm{Re}(\mathbf{d}_k^H\mathbf{g}_k(\mathbf{r}_k)) \\ &\leq \frac{4\pi^2}{\lambda^2}\|\hat{\mathbf{d}}_k\|_1\|\mathbf{r}_k\|^2 \\ &\quad + (\nabla z_k(\mathbf{r}_{k,0}) - \frac{8\pi^2}{\lambda^2}\|\hat{\mathbf{d}}_k\|_1\mathbf{r}_{k,0})^T\mathbf{r}_k + \mathrm{const.}, \end{aligned} \tag{18}$$

where the inequality follows from [17, Lemmas 1-2]

$$\begin{aligned} \hat{\mathbf{d}}_k &\triangleq 2(\mathbf{C}_k - \sum_{j=1}^{K}|u_k|^2\|\boldsymbol{\Sigma}_k^H\mathbf{F}_k\mathbf{w}_j\|^2\mathbf{I})\mathbf{g}_k(\mathbf{r}_{k,0}) + \mathbf{d}_k, \\ z_k(\mathbf{r}_k) &\triangleq \mathrm{Re}(\hat{\mathbf{d}}_k^H\mathbf{g}_k(\mathbf{r}_k)). \end{aligned} \tag{19}$$

Therefore, the FA position at vehicle k is updated as

$$\begin{aligned} &\underset{\mathbf{r}_k\in\mathcal{C}_k^r}{\mathrm{argmin}}\frac{4\pi^2}{\lambda^2}\|\hat{\mathbf{d}}_k\|_1\|\mathbf{r}_k\|^2 \\ &\qquad + \left(\nabla z_k(\mathbf{r}_{k,0}) - \frac{8\pi^2}{\lambda^2}\|\hat{\mathbf{d}}_k\|_1\mathbf{r}_{k,0}\right)^T\mathbf{r}_k \\ &= \Pi_{\mathcal{C}_k^r}\left(\mathbf{r}_{k,0} - \frac{\nabla z_k(\mathbf{r}_{k,0})}{\frac{8\pi^2}{\lambda^2}\|\hat{\mathbf{d}}_k\|_1}\right). \end{aligned} \tag{20}$$

Let $\mathbf{r}^{(t-1,0)} = \mathbf{r}^{(t-1)}$ and $\mathbf{r}^{(t-1,N)} = \mathbf{r}^{(t)}$, the iterative BSUM algorithm can be summarized as follows

$$\begin{aligned} \mathbf{u}^{(t)} &= \Phi_t\left(\mathbf{r}^{(t-1)}, \mathbf{W}^{(t-1)}\right), \\ \mathbf{v}^{(t)} &= \Psi_t\left(\mathbf{r}^{(t-1)}, \mathbf{W}^{(t-1)}, \mathbf{u}^{(t)}\right), \\ \mathbf{W}^{(t)} &= \Omega_t\left(\mathbf{r}^{(t-1)}, \mathbf{u}^{(t)}, \mathbf{v}^{(t)}\right), \\ \mathbf{r}^{(t-1,n)} &= J_t\left(\mathbf{r}^{(t-1,n-1)}, \mathbf{W}^{(t)}, \mathbf{u}^{(t)}\right), \end{aligned} \tag{21}$$

where Φ_t, Ψ_t, Ω_t, and J_t denote the iterative mapping functions at the t-th iteration. The flowchart of the iterative process is presented in Fig. 2.

Remark: We adopt the BSUM framework over the AO algorithm due to its strong theoretical guarantees. Unlike AO, which generally necessitates convergence to a stationary point at each iteration, BSUM has been rigorously shown in the literature to converge to KKT points. Moreover, the E-C coordinate selection strategy within BSUM provides two key benefits: it guarantees sufficient descent across all blocks and dynamically allocates additional iterations to slower-converging modules. This combination of convergence assurance and adaptability is not inherently available in the AO algorithm.

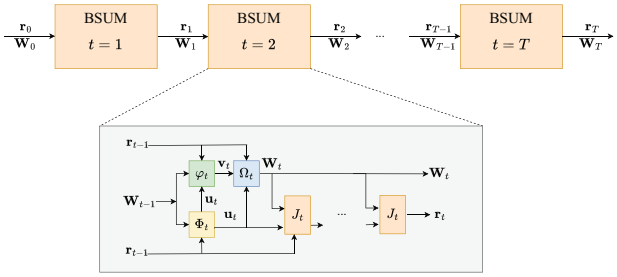

Fig. 2. Structure of the BSUM algorithm.

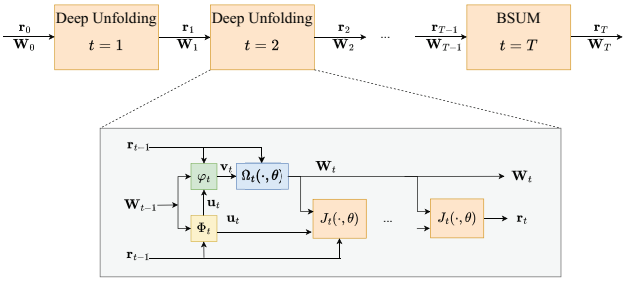

Fig. 3. Structure of the proposed deep unfolding neural network.

IV. Proposed Deep Unfolding Neural Networks

This section begins by discussing the limitations of the BSUM algorithm and then develops a deep unfolding neural network to alleviate them.

A. Limitations of BSUM Algorithm

The limitations of the BSUM algorithm outlined in the previous section can be summarized as follows:

1) Limitation 1–High computational complexity of matrix inversion: One significant drawback of the BSUM algorithm is its reliance on computing the inverse of the Hessian matrix in Eq. (15). This operation is computationally intensive, particularly in massive MIMO systems, where the complexity scales as $\mathcal{O}(n^3)$, with n representing the number of antennas. This heavy computational cost creates substantial challenges for practical deployment, especially in large-scale networks.

2) Limitation 2–Small step size: Another notable limitation arises from the upper bound constructed by second-order Taylor expansion for the objective function in Eq. (18). To ensure convergence, the step size must not exceed $1/L$, where L is the Lipschitz constant of the gradient. However, this constraint often results in an excessively small step size, leading to slow convergence. In practical applications, where the algorithm is limited to a finite number of iterations, this issue becomes even more pronounced, ultimately diminishing its efficiency and overall performance.

B. Structure of Deep Unfolding Neural Networks

To solve Limitation 1, let us define

$$\mathbf{A}_k \triangleq \frac{\sigma^2}{P}\alpha_k|u_k|^2 v_k \mathbf{I}_M + \sum_{i=1}^{K} \alpha_i |u_i|^2 v_i \mathbf{h}_i(\mathbf{r}_i)\mathbf{h}_i(\mathbf{r}_i)^H. \quad (22)$$

Then, the update of $\mathbf{w}_k$ in Eq. (15) can be rewritten as

$$\mathbf{w}_k = \alpha_k u_k v_k \mathbf{A}_k^{-1}\mathbf{h}_k(\mathbf{r}_k). \quad (23)$$

Given the high computational complexity of directly computing the matrix inverse $\mathbf{A}_k^{-1}$, we approximate it using a more efficient structure that combines an element-wise inverse with a first-order expansion. Specifically, we employ the formulation $\mathbf{A}_k^{\dagger}\mathbf{X}_k + \mathbf{A}_k\mathbf{Y}_k + \mathbf{Z}_k$, where $\mathbf{A}_k^{\dagger}$ is a diagonal matrix obtained by taking the reciprocal of each diagonal element of $\mathbf{A}_k$ while setting all off-diagonal elements to zero. The matrices $\mathbf{X}_k$, $\mathbf{Y}_k$, and $\mathbf{Z}_k$ are trainable parameters, enabling the approximation to adapt and improve efficiency while maintaining accuracy. Therefore the update of $\mathbf{w}_k$ becomes

$$\mathbf{w}_k = \alpha_k u_k v_k \left(\mathbf{A}_k^{\dagger}\mathbf{X}_k + \mathbf{A}_k\mathbf{Y}_k + \mathbf{Z}_k\right)\mathbf{h}_k(\mathbf{r}_k) + \mathbf{o}_k, \quad (24)$$

where $\mathbf{o}_k$ denotes the trainable offset.

To solve Limitation 2, we introduce a trainable parameter $\mu_k \geq 0$ to increase the stepsize of BSUM algorithm, i.e.,

$$\mathbf{r}_k = \Pi_{\mathcal{C}_k^r}\left(\mathbf{r}_{k,0} - \frac{\mu_k + 1}{\frac{8\pi^2}{\lambda^2}\|\hat{\mathbf{d}}_k\|_1}\nabla z_k(\mathbf{r}_{k,0})\right). \quad (25)$$

Moreover, since the updates of u_k and v_k involve only scalar inversion, which incurs negligible computational cost, we retain the original update Eqs. (13)-(14) without introducing any approximations.

Overall, we take Eq. (21) as the last layer in the proposed deep unfolding neural network, while the other layers can be summarized as follows

$$\begin{aligned}
\mathbf{u}^{(t)} &= \Phi_t\left(\mathbf{r}^{(t-1)}, \mathbf{W}^{(t-1)}\right), \\
\mathbf{v}^{(t)} &= \Psi_t\left(\mathbf{r}^{(t-1)}, \mathbf{W}^{(t-1)}, \mathbf{u}^{(t)}\right), \\
\mathbf{W}^{(t)} &= \Omega_t\left(\mathbf{r}^{(t-1)}, \mathbf{u}^{(t)}, \mathbf{v}^{(t)}; \boldsymbol{\theta}\right), \\
\mathbf{r}^{(t-1,n)} &= J_t\left(\mathbf{r}^{(t-1,n-1)}, \mathbf{W}^{(t)}, \mathbf{u}^{(t)}; \boldsymbol{\theta}\right),
\end{aligned} \quad (26)$$

where $\boldsymbol{\theta}$ contains all trainable parameters. The architecture of the proposed deep unfolding neural network is shown in Fig. 3.

Placing the BSUM module as the final layer is motivated by the following two considerations:

1) Integrating the BSUM module in the last layer ensures that the output $\mathbf{w}$ satisfies the power constraints of the original Problem (6), thereby guaranteeing both the feasibility of the solution and the convergence of the algorithm.

2) Since the dimensions of u_k and v_k are significantly smaller than those of $\mathbf{w}_k$, applying the iterative BSUM formulation in the final layer is more computationally efficient. This design choice not only reduces computational complexity but also preserves the overall performance of the algorithm.

C. PyTorch Platform

In this paper, we intend to leverage the PyTorch platform for the construction of neural networks and the execution of backpropagation. Nonetheless, the network architecture outlined in the previous subsection incorporates complex numbers, which

are not inherently accommodated by the PyTorch environment. To address this limitation, this subsection presents strategies for converting complex-valued data into a real-number format, enabling seamless integration within the PyTorch framework. The subsequent discussion provides a detailed explanation from two key perspectives: forward propagation and backward propagation.

1) Forward propagation: In the forward propagation stage, as outlined in Eqs. (13)-(14),(24)-(25), two main types of matrix operations are involved, namely matrix multiplication and matrix inversion.

Let $\mathbf{A} \triangleq \mathbf{A}_r + j\mathbf{A}_i \in \mathbb{C}^{N_1 \times N_2}$ and $\mathbf{B} \triangleq \mathbf{B}_r + j\mathbf{B}_i \in \mathbb{C}^{N_2 \times N_3}$ denote two complex matrices, where $\mathbf{A}_r, \mathbf{A}_i \in \mathbb{R}^{N_1 \times N_2}$ and $\mathbf{B}_r, \mathbf{B}_i \in \mathbb{R}^{N_2 \times N_3}$ represent their real and imaginary components, respectively. These matrices can be efficiently stored in real-valued tensors as $\tilde{\mathbf{A}} = [\mathbf{A}_r, \mathbf{A}_i] \in \mathbb{R}^{2 \times N_1 \times N_2}$ and $\tilde{\mathbf{B}} = [\mathbf{B}_r, \mathbf{B}_i] \in \mathbb{R}^{2 \times N_2 \times N_3}$. The product of $\mathbf{A}$ and $\mathbf{B}$ is computed as follows:

$$\begin{aligned} \mathbf{C} &\triangleq \mathbf{AB} \\ &= (\mathbf{A}_r + j\mathbf{A}_i)(\mathbf{B}_r + j\mathbf{B}_i) \\ &= (\mathbf{A}_r\mathbf{B}_r - \mathbf{A}_i\mathbf{B}_i) + j(\mathbf{A}_r\mathbf{B}_i + \mathbf{A}_i\mathbf{B}_r). \end{aligned} \tag{27}$$

Consequently, the resulting real tensor is given by $\tilde{\mathbf{C}} = [\mathbf{A}_r\mathbf{B}_r - \mathbf{A}_i\mathbf{B}_i, \mathbf{A}_r\mathbf{B}_i + \mathbf{A}_i\mathbf{B}_r]$.

Similarly, the inverse of a complex matrix $\mathbf{A}$ can be expressed as:

$$\begin{aligned} &\mathbf{A}^{-1} = (\mathbf{A}_r + j\mathbf{A}_i)^{-1} \\ &= (\mathbf{A}_r + \mathbf{A}_i\mathbf{A}_r^{-1}\mathbf{A}_i)^{-1} - j\mathbf{A}_r^{-1}\mathbf{A}_i(\mathbf{A}_r + \mathbf{A}_i\mathbf{A}_r^{-1}\mathbf{A}_i)^{-1}. \end{aligned} \tag{28}$$

The corresponding real tensor for the inverse is $\tilde{\mathbf{C}} = [(\mathbf{A}_r + \mathbf{A}_i\mathbf{A}_r^{-1}\mathbf{A}_i)^{-1}, -\mathbf{A}_r^{-1}\mathbf{A}_i(\mathbf{A}_r + \mathbf{A}_i\mathbf{A}_r^{-1}\mathbf{A}_i)^{-1}]$.

In the special case where $\mathbf{A}$ is a one-dimensional matrix (i.e., a scalar), the inverse simplifies to:

$$a^{-1} = (a_r + a_i^2 a_r^{-1})^{-1} - ja_r^{-1}a_i(a_r + a_i^2 a_r^{-1})^{-1}. \tag{29}$$

The corresponding real tensor for the inverse is $\tilde{c} = [(a_r + a_i^2 a_r^{-1})^{-1}, -a_r^{-1}a_i(a_r + a_i^2 a_r^{-1})^{-1}]$.

By leveraging Fig. 3 and Eq. (26), as well as the methodology of representing complex number operations through real number operations, we are able to architect a forward propagation neural network using the PyTorch platform.

2) Back propagation: During the backward propagation phase, we employ PyTorch's built-in automatic differentiation system alongside the Adam optimizer to facilitate efficient parameter updates. The integration of automatic differentiation with adaptive optimization techniques has become a standard approach in modern neural network implementations, as widely documented in the literature [19]. Due to space constraints, we omit the detailed matrix calculus derivations.

D. Loss Function

Due to the short coherence time of wireless communication channels in vehicular networks, we assume that only T-layer neural network can be supported within the coherence time. Given the tractable weighted MSE, the loss function utilized in the deep unfolding neural network can be defined as

$$\mathcal{L} = \mathbb{E}\left\{\sum_{k=1}^{K} \alpha_k (v_k^{(T)} e_k^{(T)} - \log(v_k^{(T)}))\right\}, \tag{30}$$

where $v_k^{(T)}$ and $e_k^{(T)}$ are the values obtained from the forward inference of the final layer.

V. Numerical Results

We numerically evaluate the performance of the proposed algorithm in the FA-enhanced MIMO system. The RSU is equipped with $M = 16$ antennas, and each element of the path-response matrix is assumed to be independently and identically distributed (i.i.d.) Gaussian. The noise power is set to $\sigma^2 = 15$ dBm, and the signal wavelength is $\lambda = 1$ m. The minimum distance between adjacent antennas is set to $D = \lambda/2$. The size of the movement area at the user is $\lambda \times \lambda$.

We compare the proposed design with the following baselines: i) FPA: All antenna positions are fixed; ii) BSUM: Traditional optimization-based algorithm without any acceleration.

Fig. 4 illustrates the WSR of our proposed scheme, the FPA design, and the traditional BSUM algorithm. It is clear that our proposed scheme not only exhibits the fastest convergence rate but also achieves the best performance. This superior performance can be attributed to the ability of our proposed network to effectively approximate matrix inversion. Secondly, adopting either MSE or WSR as the loss function yields comparable performance metrics, thereby empirically validating the transformation proposed in Prop. (1). Additionally, the larger number of trainable parameters enables more precise step size selection, which accelerates the convergence of antenna position updates.

Fig. 5 shows the relationship between WSR performance and the transmission power budget of the RSU. As expected, all three schemes—the proposed method, FPA, and BSUM—achieve higher WSR as the transmission power budget increases. This improvement is due to the availability of more power, which enhances overall system performance. Similar to Fig. 4, our proposed scheme consistently demonstrates the fastest convergence rate and achieves the best performance.

VI. Conclusion

In this paper, we utilized FAs at the vehicles to enhance the WSR in vehicular networks. To reduce the computation cost introduced by FAs, we designed a deep unfolding neural network based on the BSUM algorithm to accelerate the optimization process. Simulation results demonstrated the huge potential of FAs in improving the performance of future communication systems and the effectiveness of the deep unfolding neural network in reducing computational complexity.

Appendix

Firstly, for any given $\mathbf{r}$, it is evident that the optimal $\mathbf{W}(\mathbf{r})^*$ satisfies $P = \|\mathbf{W}(\mathbf{r})^*\|^2$, as each $\mathbf{w}_k$ can be proportionally scaled to enhance the SINR and ensure the equality. Secondly,

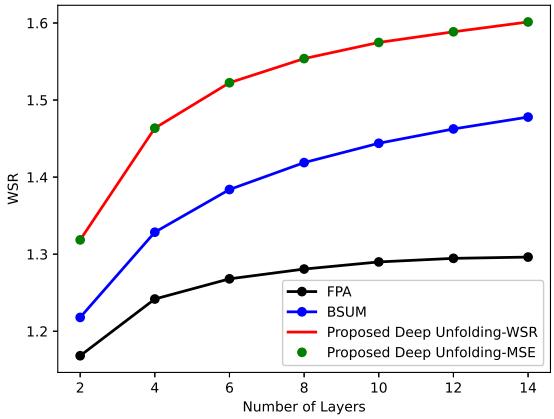

Fig. 4. WSR over the number of neural-network layers.

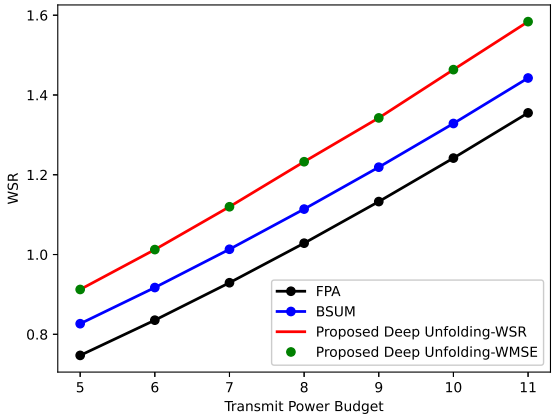

Fig. 5. WSR versus the transmit power budget.

let $\mathcal{W}$ denote the set of all optimal solutions to problem (7). For any given $\mathbf{r}$, we can proportionally scale the optimal solution to satisfy $P = \|\mathbf{W}(\mathbf{r})^*\|^2$, and the SINR for each vehicle remains unchanged. Therefore, the WSRs in (6) and (7) satisfy

$$\begin{aligned}
&\sum_{k=1}^{K} \alpha_k \log\left(1+\frac{|\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_k(\mathbf{r})^{\#}|^2}{\sum_{i\neq k}|\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_i(\mathbf{r})^{\#}|^2+\sigma^2}\right)\\
&\overset{(a)}{=}\sum_{k=1}^{K} \alpha_k \log\left(1+\frac{|\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_k(\mathbf{r})^{\#}|^2}{\sum\limits_{i\neq k}|\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_i(\mathbf{r})^{\#}|^2+\frac{\sigma^2}{P}\|\mathbf{W}(\mathbf{r})^{\#}\|^2}\right)\\
&\overset{(b)}{\leq}\sum_{k=1}^{K} \alpha_k \log\left(1+\frac{|\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_k(\mathbf{r})^{*}|^2}{\sum\limits_{i\neq k}|\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_i(\mathbf{r})^{*}|^2+\frac{\sigma^2}{P}\|\mathbf{W}(\mathbf{r})^{*}\|^2}\right)\\
&\overset{(c)}{=}\sum_{k=1}^{K} \alpha_k \log\left(1+\frac{|\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_k(\mathbf{r})^{*}|^2}{\sum_{i\neq k}|\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_i(\mathbf{r})^{*}|^2+\sigma^2}\right)\\
&\overset{(d)}{\leq}\sum_{k=1}^{K} \alpha_k \log\left(1+\frac{|\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_k(\mathbf{r})^{\#}|^2}{\sum_{i\neq k}|\mathbf{h}_k^H(\mathbf{r}_k)\mathbf{w}_i(\mathbf{r})^{\#}|^2+\sigma^2}\right),
\end{aligned}$$

where the equalities (a) and (c) hold because the power of the optimal solution in both (6) and (7) is P. The inequalities (b) and (d) arise from the optimality of $\mathbf{W}(\mathbf{r})^{\#}$ and $\mathbf{W}(\mathbf{r})^{*}$ in (6) and (7), respectively. Therefore, we can obtain the desired $\mathbf{W}(\mathbf{r})^{\#}$ by proportionally scaling the optimal $\mathbf{W}(\mathbf{r})^{*}$.

REFERENCES

[1] B. Feng, C. Feng, G. Min, and T. Q. S. Quek, "Two-timescale adaptive live video streaming transmission mechanism for vehicular networks," *IEEE Trans. Veh. Technol.*, Early Access, 2024.

[2] S. Guo and K. Qu, "Beamspace modulation for near field capacity improvement in XL-MIMO communications," *IEEE Wireless Commun. Lett.*, vol. 12, no. 8, pp. 1434–1438, Aug. 2023.

[3] W. K. New, K. K. Wong, H. Xu, C. Wang, F. R. Ghadi, J. Zhang, J. Rao, R. Murch, P. Ramírez-Espinosa, D. Morales-Jimenez, C.-B. Chae, and K. F. Tong, "A tutorial on fluid antenna system for 6G networks: Encompassing communication theory, optimization methods and hardware designs," *IEEE Commun. Surv. & Tut.* , DOI: 10.1109/COMST.2024.3498855, 2024.

[4] 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.

[5] 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.

[6] J. Zhang, J. Rao, Z. Ming, Z. Li, C.-Y. Chiu, K. K. Wong, K.-F. Tong, and R. Murch, "A novel pixel-based reconfigurable antenna applied in fluid antenna systems with high switching speed," *IEEE Open J. Antennas Propag.*, Early Access, 2024.

[7] X. Chen, B. Feng, Y. Wu, D. W. K. Ng, and R. Schober, "Joint beamforming and antenna movement design for moveable antenna systems based on statistical CSI," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Kuala Lumpur, Malaysia, Dec. 2023, pp. 1–6.

[8] G. Hu *et al.*, "Two-timescale design for movable antenna array-enabled multiuser uplink communications," *IEEE Trans. Veh. Technol.*, doi: 10.1109/TVT.2024.3485647.

[9] B. Zhang *et al.*, "Sum-rate enhancement for RIS-assisted movable antenna systems: Joint transmit beamforming, reflecting design, and antenna positioning," *IEEE Trans. Veh. Technol.*, doi: 10.1109/TVT.2024.3493099.

[10] R. Wang, Y. Jing, C. Gu, S. He and J. Chen, "End-to-end multitarget flexible job shop scheduling with deep reinforcement learning," *IEEE Internet Things J.*, vol. 12, no. 4, pp. 4420-4434, 15 Feb.15, 2025.

[11] S. B. S. Abdou, W. K. New, C. Y. Leow, S. Won, K. -K. Wong and Z. Ding, "Sum-rate maximization for UAV relay-aided fluid antenna system with NOMA," *IEEE Int. Symp. Telecommun. Technol. (ISTT)*, Langkawi Island, Malaysia, 2024, pp. 53-58.

[12] C. Wang, G. Li, H. Zhang, K.-K. Wong, Z. Li, and D. W. K. Ng, "Fluid antenna system liberating multiuser MIMO for ISAC via deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 10879–10894, Sep. 2024.

[13] D. Zhang, S. Ye, M. Xiao, K. Wang, M. D. 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.

[14] Y. Zuo, J. Guo, B. Sheng, C. Dai, F. Xiao, and S. Jin, "Fluid antenna array for mobile edge computing," *IEEE Commun. Lett.*, vol. 28, no. 7, pp. 1728–1832, Jul. 2024.

[15] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[16] M. Hong, M. Razaviyayn, Z.-Q. Luo, and J.-S. Pang, "A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 57–77, Jan. 2016.

[17] B. Feng, Y. Wu, X.-G. Xia, and X. Xiao, "Weighted sum-rate maximization for movable antenna-enhanced wireless networks," *IEEE Wireless Commun. Lett.*, vol. 13, no. 6, pp. 1770–1774, Jun. 2024.

[18] Q. Hu, Y. Cai, Q. Shi, K. Xu, G. Yu, and Z. Ding, "Iterative algorithm induced deep-unfolding neural networks: Precoding design for multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1394–1410, Feb. 2021.

[19] D. P. Kingma, J. Ba, "Adam: A Method for Stochastic Optimization," Arxiv:1412.6980, Jan. 2017.