Fluid XL-MIMO: Multiuser Antenna Allocation Exploiting the Visibility Region Information

Chao Zhang, Jue Wang, *Member, IEEE*, Li You, *Senior Member, IEEE*, Chan-Byoung Chae, *Fellow, IEEE*, and Kai-Kit Wong, *Fellow, IEEE*

***Abstract*—Fluid antenna (FA) technology can fully exploit spatial degrees of freedom by adapting antenna positions to reshape multiple-input multiple-output (MIMO) wireless channels. Most existing studies on FA assume consistent channel statistics across the antenna array, whereas this might not be true in spatially nonstationary channels. In this letter, we consider an FA-assisted extra-large MIMO (XL-MIMO) system, in which different users have different visibility regions (VRs) over the entire array, and the channel characteristics differ across the VRs. With the goal of maximizing the ergodic downlink sum-rate, we propose a transmission framework consisting of long-term antenna allocation, antenna position optimization, along with short-term precoding. Under this framework, we design multiuser antenna allocation algorithms exploiting the VR information, for practical scenarios that the users have either non-overlapping or overlapping VRs. Simulation results demonstrate the superiority of exploiting VR information in FA-aided XL-MIMO systems.**

***Index Terms*—Fluid antenna system (FAS), XL-MIMO, visibility region, antenna allocation, ergodic sum-rate.**

I. Introduction

MULTIPLE-input multiple-output (MIMO) has long been a key enabling technology for wireless communications [1]. Conventional MIMO systems usually employ fixed-position antenna (FPA) arrays. Recently, the fluid antenna (FA) technique, first proposed in [2], [3], has attracted considerable research attention due to its unique reconfigurability [4], [5], [6], [7]. Compared to FPA MIMO, the array structure in FA MIMO systems can be optimized by flexibly altering antenna positions, so as to fully exploit the spatial degree of freedom and improve the system capacity [8]. FAs can be realized by a variety of antenna technologies such as liquid antennas [9], reconfigurable pixels [10] and metamaterials [11].

The performance of a single-antenna FA system was studied in the early work of [3]. This line of research has progressed under different terminologies, such as movable antennas (MA) [12]. The capacity of MA-enabled MIMO systems based on instantaneous channel state information (CSI) was characterized in [13]. Further considering the difficulty of acquiring instantaneous CSI, ergodic rates with statistical CSI have been analyzed in [14], [15]. Note that existing studies on FA/MA MIMO typically assume that all users see the same propagation environment. However, with the evolution of massive MIMO to extra-large MIMO (XL-MIMO) towards 6G, such a channel assumption might become invalid.

In XL-MIMO systems where the array aperture becomes very large, spatial non-stationarity in the channels has to be considered. Specifically, each user, depending on its location, may only see a subregion of the entire array. This is called the user's visibility region (VR) [16], [17]. Having different VRs, users exhibit distinct channel statistics. The location-dependent nature of VR enables simultaneous user localization and VR identification [18]. This fact facilitates the incorporation of VR information into FA-assisted XL-MIMO transmission design. Specifically, multiuser antenna allocation becomes possible by switching FA elements to different VRs of different users.

In this context, we investigate VR-aware FA-assisted XL-MIMO transmission. To maximize the ergodic sum-rate, we propose a transmission framework which combines long-term antenna allocation and short-term precoding. We consider two practical scenarios, in which the VRs of different users might be either non-overlapping or overlapping. For non-overlapping VRs, we derive an analytical upper bound of the ergodic sum-rate under maximum ratio transmission (MRT), enabling efficient antenna allocation via standard mixed-integer nonlinear programming (MINLP). For overlapping VRs, we develop a dual-loop optimization framework, in which antenna allocation is conducted by employing genetic algorithm (GA). Afterwards antenna placement and precoding are alternatingly updated using particle swarm optimization (PSO) and zero-forcing (ZF), respectively. Numerical results indicate that the use of VR information significantly enhances FA-assisted XL-MIMO transmission performance over an FPA array.

II. System Model and Problem Formulation

A. Signal Model

Consider a multiuser MIMO system as shown in Fig. 1, in which a base station (BS) equipped with N FAs, forming an XL-MIMO array, serves K single-FPA users in the downlink. The XL-MIMO array is implemented in a rectangular area of size $A_x \times A_y$, which is divided into S subregions of size $a_x \times a_y$, where $S = n_{\text{row}} \times n_{\text{col}}$, $n_{\text{col}} = A_x/a_x$ and $n_{\text{row}} = A_y/a_y$. The s-th subregion is denoted as $\mathcal{A}_s$ for $s \in \{1, 2, \ldots, S\}$. The k-th user' VR is described by the union of its visible subregions, represented by the index set $\mathcal{I}_k \subseteq \{1, 2, \ldots, S\}$. Also, $\mathcal{A}_s$ may be visible to one or more users, for which we define the corresponding user set as $\mathcal{K}_s \subseteq \{1, 2, \ldots, K\}$.

For every FA element, its position can be adjusted either within a subregion, or across different subregions. In order to describe the precise position of an FA element, we establish a coordinate system: The origin $\mathbf{o} = (0, 0)^{\mathrm{T}}$ is set at the

C. Zhang and J. Wang are with the School of Information Science and Technology, Nantong University, Nantong 226019, China, and also with the Nantong Research Institute for Advanced Communication Technologies, Nantong 226019, China (e-mail: zhangchao@stmail.ntu.edu.cn; wangjue@ntu.edu.cn).

L. You is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China, and also with the Purple Mountain Laboratories, Nanjing 211100, China (e-mail: lyou@seu.edu.cn).

C.-B. Chae is with the School of Integrated Technology, Yonsei University, Seoul, 03722, Korea (e-mail: cbchae@yonsei.ac.kr).

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

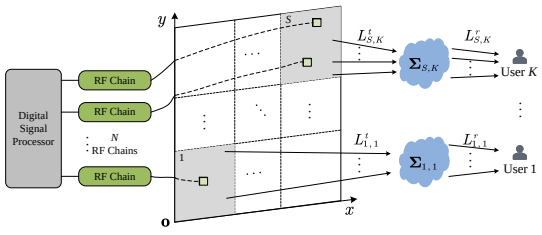


Fig. 1. An FA-assisted XL-MIMO communication system.

left-bottom corner of the entire array, with the x- and y-axes defined along its two sides. The entire array area is described by the coordinate region of $[0, A_x] \times [0, A_y]$. The coordinate of the n-th FA in the s-th subregion $\mathcal{A}_s$ is denoted by $\mathbf{t}_{s,n} = [x_{s,n}, y_{s,n}]^{\mathrm{T}}, 1 \le n \le N_s$, where N_s is the number of FA elements allocated to $\mathcal{A}_s$. For convenience, we stack the coordinates of all FA elements in $\mathcal{A}_s$ in a matrix form, described as $\tilde{\mathbf{t}}_s \triangleq [\mathbf{t}_{s,1}, \ldots, \mathbf{t}_{s,N_s}] \in \mathbb{R}^{2\times N_s}$.

Let $x_k \in \mathbb{C}$ denote the signal intended for the k-th user, where $\mathbb{E}\{x_k x_k^H\} = 1$ and $\mathbb{E}\{x_k x_{k'}^H\} = 0, \forall k' \neq k$. The received signal at the k-th user can be expressed as

$$y_k = \sum_{s\in\mathcal{I}_k} \mathbf{h}_{s,k}(\tilde{\mathbf{t}}_s)\mathbf{w}_{s,k}x_k + \sum_{s\in\mathcal{I}_k}\left(\mathbf{h}_{s,k}(\tilde{\mathbf{t}}_s)\sum_{\substack{k'\in\mathcal{K}_s\\ k'\neq k}}\mathbf{w}_{s,k'}^H x_{k'}\right) + n_k, \quad (1)$$

where $\mathbf{h}_{s,k}(\tilde{\mathbf{t}}_s) \in \mathbb{C}^{1\times N_s}$ denotes the channel vector between the FAs in $\mathcal{A}_s$ and user k (which depends on the FA positions $\tilde{\mathbf{t}}_s$), and $\mathbf{w}_{s,k} \in \mathbb{C}^{N_s\times 1}$ is the corresponding precoder, and $n_k \sim \mathcal{CN}(0, \sigma_k^2)$ is complex additive Gaussian noise.

B. Channel Model

We make the following assumptions on the wireless channels in (1): 1) Far-field propagation condition holds for every subregion. 2) Spatial non-stationarity appears across subregions. Specifically, denote the number of transmit and receive paths between subregion $\mathcal{A}_s$ and user k by $L_{s,k}^t$ and $L_{s,k}^r$, respectively; besides, denote the elevation and azimuth angles of departure (AoD), for the l-th transmit path from $\mathcal{A}_s$ to user k, as $\theta_{s,k,l}^t \in [-\pi/2, \pi/2]$ and $\phi_{s,k,l}^t \in [-\pi/2, \pi/2]$ for $l = 1, \ldots, L_{s,k}^t$, respectively. The number of paths and their AoDs vary across subregions due to spatial non-stationarity.

Then taking $\mathbf{o}$ as a reference point, the l-th path length difference from the n-th FA in $\mathcal{A}_s$ to user k is $\rho_{s,k,l}^t(\mathbf{t}_{s,n}) = x_{s,n}\sin\theta_{s,k,l}^t\cos\phi_{s,k,l}^t + y_{s,n}\cos\theta_{s,k,l}^t$ [13]. Consequently, the channel phase difference is $2\pi\rho_{s,k,l}^t(\mathbf{t}_{s,n})/\lambda$, where λ is the carrier wavelength. Hence, the transmit field response vector $\mathbf{g}_{s,k}(\mathbf{t}_{s,n}) \in \mathbb{C}^{L_{s,k}^t\times 1}$ is given as

$$\mathbf{g}_{s,k}(\mathbf{t}_{s,n}) = \left(e^{j\frac{2\pi}{\lambda}\rho_{s,k,1}^t(\mathbf{t}_{s,n})}, \ldots, e^{j\frac{2\pi}{\lambda}\rho_{s,k,L_{s,k}^t}^t(\mathbf{t}_{s,n})}\right)^T. \quad (2)$$

We further define the path response matrix from $\mathcal{A}_s$ to user k as $\mathbf{\Sigma}_{s,k} \in \mathbb{C}^{L_{s,k}^r\times L_{s,k}^t}$, in which the entry $[\mathbf{\Sigma}_{s,k}]_{q,p}$ characterizes the response coefficient between the p-th transmit path and the q-th receive path. Each entry is modeled as an independent identically distributed (i.i.d.) circularly symmetric complex Gaussian random variable with zero mean and variance $\alpha_{s,k}^2$. At last, the channel vector $\mathbf{h}_{s,k}(\tilde{\mathbf{t}}_s)$ is [13]

$$\mathbf{h}_{s,k}(\tilde{\mathbf{t}}_s) = \mathbf{f}_{s,k}^H\mathbf{\Sigma}_{s,k}\mathbf{G}_k(\tilde{\mathbf{t}}_s), \quad (3)$$

where $\mathbf{f}_{s,k} = \mathbf{1}_{L_{s,k}^r\times 1}$ is an all-one vector, and $\mathbf{G}_k(\tilde{\mathbf{t}}_s) = [\mathbf{g}_{s,k}(\mathbf{t}_{s,1}), \ldots, \mathbf{g}_{s,k}(\mathbf{t}_{s,N_s})] \in \mathbb{C}^{L_{s,k}^t\times N_s}$ is the field-response matrix of all the FA elements in $\mathcal{A}_s$ for the k-th user.

C. Problem Formulation

With the signal model in (1), the achievable ergodic sum-rate can be found as [14]

$$R = \mathbb{E}_{\mathbf{\Sigma}_{s,k}}\left[\sum_{k=1}^K \log\left(1 + D_k^{-1}\left|\sum_{s\in\mathcal{I}_k}\mathbf{h}_{s,k}(\tilde{\mathbf{t}}_s)\mathbf{w}_{s,k}\right|^2\right)\right], \quad (4)$$

where $D_k = \sigma_k^2 + \sum_{k'\neq k}|\sum_{s\in\mathcal{I}_k\cap\mathcal{I}_{k'}'}\mathbf{h}_{s,k}(\tilde{\mathbf{t}}_s)\mathbf{w}_{s,k'}|^2$ denotes the noise plus interference term for user k. Our objective is to maximize this ergodic sum-rate by jointly optimizing the antenna allocation ($\{N_s\}$), FA positions ($\{\tilde{\mathbf{t}}_s\}$), and downlink precoding ($\{\mathbf{w}_{s,k}\}$). The problem is formulated as

$$\begin{aligned}
(\mathrm{P1}) \max_{\{N_s\},\{\mathbf{w}_{s,k}\},\{\tilde{\mathbf{t}}_s\}} \quad & R && (5a)\\
\text{s.t.} \quad & N_s = 0 \text{ or } |\mathcal{K}_s| \le N_s \le N_{s,\max}, && (5b)\\
& \sum_{s=1}^S N_s \le N, && (5c)\\
& \sum_{k\in\mathcal{K}_s}\mathrm{tr}(\mathbf{w}_{s,k}\mathbf{w}_{s,k}^H) \le N_s p, && (5d)\\
& \|\mathbf{t}_{s,i} - \mathbf{t}_{s,j}\| \ge D,\ \forall s,\ i,\ j, && (5e)\\
& \tilde{\mathbf{t}}_s \in \mathcal{A}_s,\ \forall s. && (5f)
\end{aligned}$$

Note that $N_s = 0$ in (5b) allows a subregion $\mathcal{A}_s$ to be inactive. Otherwise, once activated, N_s must be no less than $|\mathcal{K}_s|$, which is the number of users covered by $\mathcal{A}_s$. Furthermore, N_s cannot exceed the maximum allowable antenna count $N_{s,\max} = \lfloor a_x/D + 1\rfloor \times \lfloor a_y/D + 1\rfloor$. The constraints on the total number of antennas, transmit power (where $p = P/N$ in (5d) denotes the average power constraint per antenna, with P denoting the total transmit power), and antenna positions are specified in (5c)–(5f), respectively.

III. Non-Overlapping VRs

We start solving (P1) by considering an ideal case, that the users have non-overlapping VRs on the XL-MIMO array, i.e., $\mathcal{I}_k \cap \mathcal{I}_{k'} = \emptyset$ for $k' \neq k$. In practice, this can be guaranteed by proper user scheduling, or simply by discarding the overlapping VRs in transmission design. In this case, the interference term in (4) is eliminated, and hence MRT becomes optimal in every subregion $\mathcal{A}_s$. That is, $\mathbf{w}_{s,k} = \sqrt{N_s p}\mathbf{h}_{s,k}^H/\|\mathbf{h}_{s,k}\|$. As such, the ergodic sum-rate in (4) is simplified to

$$R = \mathbb{E}_{\mathbf{\Sigma}_{s,k}}\left[\sum_{k=1}^K \log\left(1 + \frac{1}{\sigma_k^2}\left(\sum_{s\in\mathcal{I}_k}\sqrt{N_s p}\|\mathbf{h}_{s,k}(\tilde{\mathbf{t}}_s)\|\right)^2\right)\right]. \quad (6)$$

Proposition 1. *With non-overlapping VRs and MRT precoding, an upper bound of R in (6) can be obtained as*

$$\bar{R} = \sum_{k=1}^{K} \log\left(1 + \frac{1}{\sigma_k^2}\left[\sum_{s\in\mathcal{I}_k} pN_s^2\alpha_{s,k}^2 L_{s,k}^r L_{s,k}^t + \sum_{\substack{s,s'\in\mathcal{I}_k \\ s\neq s'}} p\alpha_{s,k}\alpha_{s',k}\sqrt{N_s N_{s'} L_{s,k}^r L_{s,k}^t L_{s',k}^r L_{s',k}^t}\,\frac{\Gamma(N_s+1/2)}{\Gamma(N_s)}\frac{\Gamma(N_{s'}+1/2)}{\Gamma(N_{s'})}\right]\right), \quad (7)$$

where $\Gamma(\cdot)$ denotes the Gamma function. Note that (7) is a function of the channel path parameters ($\alpha_{s,k}$, $L_{s,k}^r$ and $L_{s,k}^t$), the VR sets ($\mathcal{I}_k$), and the FA allocation parameters (N_s).

Proof. With Jensen's inequality, R in (6) is upper bounded by

$$\begin{aligned} R \le \bar{R} &\triangleq \sum_{k=1}^{K} \log\left(1 + \frac{1}{\sigma_k^2}\mathbb{E}_{\mathbf{\Sigma}_{s,k}}\left[\left(\sum_{s\in\mathcal{I}_k} X_s\right)^2\right]\right) \\ &= \sum_{k=1}^{K} \log\left(1 + \frac{1}{\sigma_k^2}\left(\sum_{s\in\mathcal{I}_k}\mathbb{E}_{\mathbf{\Sigma}_{s,k}}\left[X_s^2\right] + \sum_{\substack{s,s'\in\mathcal{I}_k \\ s\neq s'}}\mathbb{E}_{\mathbf{\Sigma}_{s,k}}\left[X_s\right]\mathbb{E}_{\mathbf{\Sigma}_{s,k}}\left[X_{s'}\right]\right)\right), \end{aligned} \quad (8)$$

where $X_s = \sqrt{N_s p}\|\mathbf{h}_{s,k}(\tilde{\mathbf{t}}_s)\|$. Since the elements of $\mathbf{h}_{s,k}(\tilde{\mathbf{t}}_s)$ follow complex Gaussian distribution of $\mathcal{CN}(0, \alpha_{s,k}^2 L_{s,k}^r L_{s,k}^t)$, we have $\|\mathbf{h}_{s,k}(\tilde{\mathbf{t}}_s)\|^2 \sim \text{Gamma}(N_s, \alpha_{s,k}^2 L_{s,k}^r L_{s,k}^t)$ and $\|\mathbf{h}_{s,k}(\tilde{\mathbf{t}}_s)\| \sim \text{Nakagami}(N_s, N_s\alpha_{s,k}^2 L_{s,k}^r L_{s,k}^t)$. As a result, the expectation terms in (8) can be respectively derived as

$$\mathbb{E}_{\mathbf{\Sigma}_{s,k}}\left[X_s^2\right] = pN_s^2\alpha_{s,k}^2 L_{s,k}^r L_{s,k}^t, \quad (9)$$

$$\mathbb{E}_{\mathbf{\Sigma}_{s,k}}\left[X_s\right] = \sqrt{pN_s L_{s,k}^r L_{s,k}^t}\,\alpha_{s,k}\frac{\Gamma(N_s+1/2)}{\Gamma(N_s)}. \quad (10)$$

This leads to the upper bound in (7). □

Although $\bar{R}$ in (7) is an upper bound, it is tight under different parameter settings, as will be confirmed by Monte-Carlo simulations. Notably, with non-overlapping VRs and MRT, $\bar{R}$ solely depends on N_s while it is not affected by particular FA positions. In this case, Problem (P1) is simplified to an antenna allocation problem only, described as

$$\text{(P2)} \quad \max_{\{N_s\}} \bar{R} \quad \text{s.t.} \quad (5b), (5c), \quad (11)$$

which is a nonlinear integer programming problem, which can be efficiently solved using a standard MINLP solver.

IV. Overlapping VRs

We extend the discussion to a more general scenario, in which different users may have overlapping VRs on the XL-MIMO array. That is, for two users $k' \neq k$, $\mathcal{I}_k \cap \mathcal{I}_{k'} \neq \emptyset$ may hold. In this case, interference-cancelling precoding has to be considered instead of simply using MRT. With $N_s \geq |\mathcal{K}_s|$, we adopt ZF precoders for the design of $\mathbf{w}_{s,k}$ in (5a).

Unlike $\bar{R}$ in (7), the ergodic sum-rate achieved with ZF precoders now depends on specific FA positions. This renders the optimization of $\{N_s\}$, $\{\mathbf{w}_{s,k}\}$, and $\{\tilde{\mathbf{t}}_s\}$ all coupled together when solving (P1). To address this challenge, we propose a dual-loop algorithm to obtain a solution, as described by the flowchart in Fig. 2. In the inner loop, the antenna allocation set $\{N_s\}$ is given and fixed, while $\{\tilde{\mathbf{t}}_s\}$ and $\{\mathbf{w}_{s,k}\}$ are alternately updated using a PSO algorithm. In the outer loop, a GA-based solution of $\{N_s\}$ is obtained. These inner- and outer-loops are iteratively updated until an end-loop condition is satisfied. Details of the algorithm are explained in the following.

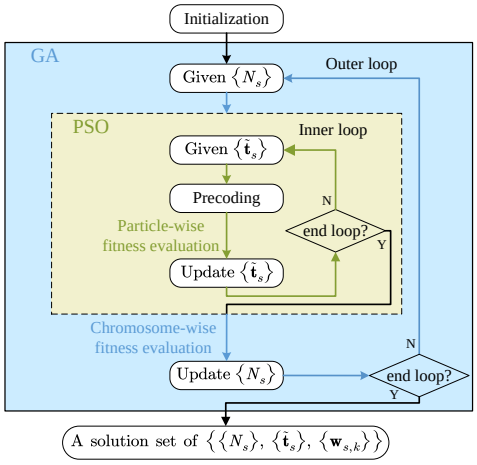


Fig. 2. The flowchart of the proposed dual-loop algorithm.

A. Inner Loop: Precoding and FA Position Optimization

In the inner loop, ZF precoders are determined as follows. Define $\bar{\mathbf{H}}_{s,k} \triangleq [\mathbf{h}_{s,1}^T, \ldots, \mathbf{h}_{s,k-1}^T, \mathbf{h}_{s,k+1}^T, \ldots, \mathbf{h}_{s,|\mathcal{K}_s|}^T]^T \in \mathbb{C}^{(|\mathcal{K}_s|-1)\times N_s}$, ZF precoding is performed on each subregion, with the normalized precoder for user k, given by [19]

$$\mathbf{w}_{s,k} = \sqrt{\frac{N_s p}{|\mathcal{K}_s|}} \times \frac{\mathbf{P}_{s,k}\mathbf{h}_{s,k}(\tilde{\mathbf{t}}_s)^H}{\|\mathbf{P}_{s,k}\mathbf{h}_{s,k}(\tilde{\mathbf{t}}_s)^H\|}, \quad (12)$$

where $\mathbf{P}_{s,k} = \mathbf{I}_{N_s} - \bar{\mathbf{H}}_{s,k}^H(\bar{\mathbf{H}}_{s,k}\bar{\mathbf{H}}_{s,k}^H)^{-1}\bar{\mathbf{H}}_{s,k} \in \mathbb{C}^{N_s\times N_s}$ is the orthogonal projection matrix onto the null space of $\bar{\mathbf{H}}_{s,k}$. Specifically, when $|\mathcal{K}_s| = 1$, $\mathbf{P}_{s,k}$ degenerates into the identity matrix, and $\mathbf{w}_{s,k}$ is simplified to the MRT precoder.

Given sets of $\{N_s\}$ and $\{\tilde{\mathbf{t}}_s\}$, the ZF precoders completely eliminate the inter-user interference term in (4). However, it is still very challenging to derive an analytical expression for R in (4). Therefore, we use the Monte Carlo average as an approximation of the ergodic sum-rate, defined as

$$R_{\text{ZF}} \triangleq \frac{1}{M}\sum_{m=1}^{M}\sum_{k=1}^{K}\log\left(1 + \frac{1}{\sigma_k^2}\left|\sum_{s\in\mathcal{I}_k}\mathbf{h}_{s,k}^{(m)}(\tilde{\mathbf{t}}_s)\mathbf{w}_{s,k}^{(m)}\right|^2\right), \quad (13)$$

where $\mathbf{h}_{s,k}^{(m)}(\tilde{\mathbf{t}}_s)$ denotes the m-th random channel realization (generated with fixed FA positions $\tilde{\mathbf{t}}_s$), and $\mathbf{w}_{s,k}^{(m)}$ is the corresponding precoder obtained with (12). Now, the FA position optimization problem is described as

$$\max_{\{\tilde{\mathbf{t}}_s\}} R_{\text{ZF}} \quad \text{s.t.} \quad (5e), (5f). \quad (14)$$

This problem is tackled using the PSO algorithm [20], which yields a solution $\{\tilde{\mathbf{t}}_s\}$. Steps (12)–(14) are iteratively executed until an end-loop condition is met, e.g., the difference of R_{ZF}

obtained in two adjacent iterations falls below a predefined threshold. Once the inner loop is ended, the resulting $\{\mathbf{w}_{s,k}\}$ and $\{\tilde{\mathbf{t}}_s\}$ are passed to the outer loop to optimize $\{N_s\}$.

B. Outer Loop: Antenna Allocation

The outer loop treats the ergodic sum-rate as a function of the antenna allocation $\{N_s\}$. With $\{\mathbf{w}_{s,k}\}$ and $\{\tilde{\mathbf{t}}_s\}$ obtained in the inner loop, Problem (P1) is reduced to:

$$\max_{\{N_s\}} \; R \;\; \text{s.t.} \;\; (5b), (5c). \tag{15}$$

Given it being NP-hard, we propose Algorithm 1 for an effective solution with the following key steps:

- *Initialization (Step 1):* First, the GA initializes a population $\mathcal{P}^{(0)} = \{\mathbf{N}_1^{(0)}, \mathbf{N}_2^{(0)}, \ldots, \mathbf{N}_I^{(0)}\}$ of size I, where each chromosome encodes a feasible antenna allocation solution set in the form of

$$\mathbf{N}_i^{(0)} = \left\{N_{i,1}^{(0)}, N_{i,2}^{(0)}, \ldots, N_{i,S}^{(0)}\right\}, \tag{16}$$

  where $N_{i,s}^{(0)}$ denotes the number of FAs allocated to $\mathcal{A}_s$ in the i-th chromosome, for $1 \le i \le I$.
- *Evaluation (Step 3):* Through the fitness evaluation, the sets of $\{N_s\}$, $\{\tilde{\mathbf{t}}_s\}$ and $\{\mathbf{w}_{s,k}\}$ that achieve the highest ergodic sum-rate are identified in the g-th iteration.
- *Reproduction and Evolution (from Step 4 to Step 14):* Use crossover operation to generate two offspring chromosomes $\mathbf{N}_1^{(\mathrm{o})}$ and $\mathbf{N}_2^{(\mathrm{o})}$ from a pair of parent solutions:

$$(\mathbf{N}_1^{(\mathrm{o})}, \mathbf{N}_2^{(\mathrm{o})}) = \texttt{Crossover}(\mathbf{N}_1^{(\mathrm{p})}, \mathbf{N}_2^{(\mathrm{p})}, \mathbb{P}_c). \tag{17}$$

  With probability $\mathbb{P}_c$, the single-point crossover is applied at a random point $s \in \{1, \ldots, S-1\}$; that is, the elements (of $\mathbf{N}_1^{(\mathrm{p})}$ and $\mathbf{N}_2^{(\mathrm{p})}$) beyond s are exchanged. Otherwise, $\mathbf{N}_1^{(\mathrm{o})}$ and $\mathbf{N}_2^{(\mathrm{o})}$ are exact copies of the parents.
- *Repair (Step 10):* As $\mathbf{N}_1^{(\mathrm{o})}$ and $\mathbf{N}_2^{(\mathrm{o})}$ may violate constraint (5c), a repair operation is applied such that:

$$\mathbf{N}_j^{(\mathrm{o})} = \texttt{Repair}(\mathbf{N}_j^{(\mathrm{o})}, N, \{N_{s,\max}\}, \{|\mathcal{K}_s|\}), \tag{18}$$

  This operation iteratively adjusts the allocation $N_{j,s}^{(\mathrm{o})}$ (the s-th element of $\mathbf{N}_j^{(\mathrm{o})}$) to satisfy constraint (5c), while ensuring that constraint (5b) is maintained.
- *Mutate (Step 11):* To maintain population diversity, mutation operation is applied to each repaired offspring:

$$\mathbf{N}_j^{(\mathrm{o})} = \texttt{Mutate}(\mathbf{N}_j^{(\mathrm{o})}, \mathbb{P}_m, \{N_{s,\max}\}, \{|\mathcal{K}_s|\}). \tag{19}$$

  With probability $\mathbb{P}_m$, a subregion pair $(\mathcal{A}_s, \mathcal{A}_{s'})$ is selected, and n antennas are re-allocated from $\mathcal{A}_s$ to $\mathcal{A}_{s'}$ according to the following rule:

$$n = \begin{cases} 1, & \text{if } N_{j,s}^{(\mathrm{o})} - 1 \ge |\mathcal{K}_s|, \\ N_{j,s}^{(\mathrm{o})}, & \text{otherwise.} \end{cases} \tag{20}$$

  An elitism strategy is further employed to retain the best-performing individuals across generations. The algorithm then proceeds to the evaluation step (Step 3) until a termination condition is satisfied.

Algorithm 1 GA-based Antenna Allocation

Input: Array parameters: a_x, a_y, A_x, A_y,
System and channel parameters: N, $\{N_{s,\max}\}$, K, $\{\mathcal{I}_k\}$, $\{\mathcal{K}_s\}$, λ, D, P, σ_k^2, $\{L_{s,k}^t\}$, $\{L_{s,k}^r\}$, $\{\theta_{s,k,l}^t\}$, $\{\phi_{s,k,l}^t\}$, $\{\mathbf{\Sigma}_{s,k}\}$,
GA algorithm parameters: I, G, $\mathbb{P}_c$, $\mathbb{P}_m$.
Output: $R_{\max}$, $\mathbf{N}_{\max}$, $\{\tilde{\mathbf{t}}_s\}$, $\{\mathbf{w}_{s,k}\}$.

```
Initialize population P^(0) = {N_1^(0), N_2^(0), ..., N_I^(0)}.
for g = 1 to G do
    N_max ← arg max_{N_i^(g-1)} R; record the corresponding R as R_max.
    Initialize parent population P' = ∅, offspring population P'' = ∅.
    for i = 1 to I do
        Randomly select two individuals N_x^(g-1), N_y^(g-1) from P^(g-1).
        Add the one with better fitness (higher R) to P'.
    end for
    for i = 1 to I by 2 do
        Generate offspring individuals N_1^(o) and N_2^(o) according to (17).
        Repair each offspring according to (18).
        Mutate each offspring according to (19).
        Add N_1^(o), N_2^(o) to P''.
    end for
    Replace the first individual in P'' with the best individual in P^(g-1),
    then P^(g) ← P''.
end for
return R_max, N_max, {t̃_s}, {w_{s,k}}.
```

TABLE I
SIMULATION PARAMETERS.

Parameter	Description	Value
f	Carrier frequency	3.5 GHz
N	Number of antennas at the BS	$16, 32, 64, 128, 256$
A_x, A_y	Side length of the array region	2 m
a_x, a_y	Side length of the subregion	1 m
D	Minimum inter-FA distance	$\lambda/2$
$L_{s,k}^t$, $L_{s,k}^r$	Number of paths	$[3, 6]$
$\theta_{s,k,l}^t$, $\phi_{s,k,l}^t$	Elevation and azimuth angles	$\mathcal{U}\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$
β_0	Reference channel gain at 1 m	-40 dB
α	Path loss exponent	2.8
P	Transmit power	30 dBm
σ_k^2	Noise power	-80 dBm
I	Population size of the GA	100
G	Maximum iterations of the GA	20
$\mathbb{P}_c$	Crossover probability	0.8
$\mathbb{P}_m$	Mutate probability	0.1

V. SIMULATION RESULTS

We have conducted simulations to evaluate the performance of the proposed design. Parameters are set as described in Table I. The large-scale fading of the channels is given by $\beta_k = \beta_0 d_{s,k}^{-\alpha}$, where $d_{s,k}$ denotes the distance between $\mathcal{A}_s$ and user k. We consider an example scenario with two users, with $d_{s,1} = 50$ m and $d_{s,2} = 100$ m for each $\mathcal{A}_s$. The variance of $[\mathbf{\Sigma}_{s,k}]_{q,p}$ is set as $\alpha_{s,k}^2 = \beta_k/(L_{s,k}^t L_{s,k}^r)$. The PSO parameters follow [20], except that the population size is 50 and the number of iterations is 100.

Two VR scenarios are simulated: $\mathcal{I}_1 = \{2\}, \mathcal{I}_2 = \{1, 3\}$ (non-overlapping) and $\mathcal{I}_1 = \{1, 2\}, \mathcal{I}_2 = \{1, 3\}$ (overlapping). For these two VR conditions, the proposed algorithms described earlier are applied and respectively, denoted as "FA, non-overlapping VR" and "FA, overlapping VR" in the following. We compare the proposed FA designs with the following baseline schemes:

- **FPA, non-overlapping VR**: Conventional uniform planar array (UPA) is employed and MRT is adopted on each

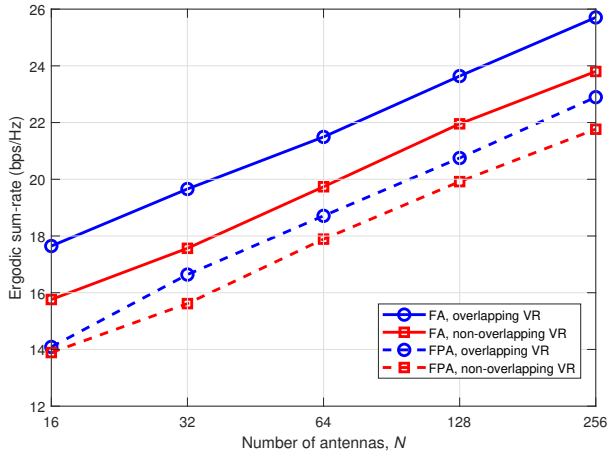
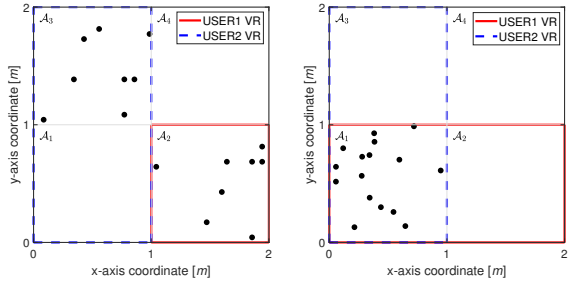


Fig. 3. Achievable ergodic sum-rate versus N.



(a) Non-overlapping VRs (b) Overlapping VRs

Fig. 4. Optimized FA layouts for different VR conditions.

$\mathcal{A}_s$ with instantaneous CSI.

- **FPA, overlapping VR**: UPA is adopted and ZF precoding is used on each $\mathcal{A}_s$ with instantaneous CSI.

Fig. 3 demonstrates the rate performance versus the number of antennas. As shown, the proposed FA design outperforms the FPA schemes at different values of N, for both the non-overlapping and overlapping VR cases. Specifically, at $N = 16$, "FA, overlapping VR" achieves nearly 25% performance gains over "FPA, overlapping VR". Although the gains in percentage might become lower as N increases, an absolute gain larger than 2 bps/Hz is still guaranteed at large antenna numbers, e.g., $N = 256$. More importantly, the simulation results indicate that it is possible to use much less FA elements to achieve a comparable performance of FPA XL-MIMO.

The gains of FA XL-MIMO mainly stem from its capability of flexibly adapting antenna allocation to the VR condition. In Fig. 4, we take $N = 16$ as an example to intuitively illustrate the optimized FA layouts under different VR conditions. In Fig. 4(a), only $\mathcal{A}_3$ is activated for user 2, since this subregion has higher channel gain than $\mathcal{A}_1$. Besides, antenna allocation between $\mathcal{A}_3$ and $\mathcal{A}_2$ is optimized to maximize the sum-rate for two users. Fig. 4(b) illustrates the antenna allocation result for the overlapping VR case, where $\mathcal{A}_2$ is visible to both users. Allocating antennas to this shared subregion improves reuse efficiency under a limited antenna budget, therefore, all antennas are allocated to $\mathcal{A}_2$ for better performance.

VI. Conclusion

This paper investigated the ergodic sum-rate maximization problem for an FA-aided XL-MIMO system exploiting VR information. We proposed a framework that integrates long-term antenna allocation and short-term precoding, and conducted optimization for both the non-overlapping and overlapping VR scenarios. Simualtion results revealed that the proposed scheme that exploits VR information significantly outperforms conventional FPA systems. The results also illustrated that comparable performance can be achieved using fewer FAs than FPAs, reducing the number of required RF chains.

References

[1] C.-X. Wang *et al.*, "On the road to 6G: Visions, requirements, key technologies, and testbeds," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 905–974, Secondquarter 2023.

[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.*, early access, 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] C. Wang, K.-K. Wong, Z. Li, L. Jin, and C.-B. Chae, "Large language model empowered design of fluid antenna systems: Challenges, frameworks, and case studies for 6G," to appear in *IEEE Wireless Commun.*, arXiv:2505.09751, 2025.

[7] J. Zheng *et al.*, "Flexible-position MIMO for wireless communications: Fundamentals, challenges, and future directions," *IEEE Wireless Commun.*, vol. 31, no. 5, pp. 18–26, Oct. 2024.

[8] W. K. New, K.-K. Wong, H. Xu, K.-F. Tong and C.-B. Chae, "An information-theoretic characterization of MIMO-FAS: Optimization, diversity-multiplexing tradeoff and q-outage capacity," *IEEE Trans. Wireless Commun.*, vol. 23, no. 6, pp. 5541–5556, Jun. 2024.

[9] 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.

[10] 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.

[11] 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

[12] L. Zhu, and K. K. Wong, "Historical review of fluid antennas and movable antennas," *arXiv preprint*, arXiv:2401.02362v2, Jan. 2024.

[13] W. Ma, L. Zhu, and R. Zhang, "MIMO capacity characterization for movable antenna systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 4, pp. 3392–3407, Apr. 2024.

[14] Y. Ye *et al.*, "Fluid antenna-assisted MIMO transmission exploiting statistical CSI," *IEEE Commun. Lett.*, vol. 28, no. 1, pp. 223–227, Jan. 2024.

[15] G. Hu *et al.*, "Two-timescale design for movable antenna array-enabled multiuser uplink communications," *IEEE Trans. Veh. Technol.*, vol. 74, no. 3, pp. 5152–5157, Mar. 2025.

[16] E. D. Carvalho, A. Ali, A. Amiri, M. Angjelichinoski and R. W. Heath, "Non-stationarities in extra-large-scale massive MIMO," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 74–80, Aug. 2020.

[17] D. Liu *et al.*, "Location-based visible region recognition in extra-large massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 8186–8191, Jun. 2023.

[18] J. Tian, Y. Han, S. Jin and M. Matthaiou, "Low-overhead localization and VR identification for subarray-based ELAA systems," *IEEE Wireless Commun. Lett.*, vol. 12, no. 5, pp. 784–788, May 2023.

[19] S. Huang, H. Yin, J. Wu and V. C. M. Leung, "User selection for multiuser MIMO downlink with zero-forcing beamforming," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3084–3097, Sept. 2013.

[20] Z. Xiao, X. Pi, L. Zhu, X.-G. Xia and R. Zhang, "Multiuser communications with movable-antenna base station: Joint antenna positioning, receive combining, and power control," *IEEE Trans. Wireless Commun.*, vol. 23, no. 12, pp. 19744–19759, Dec. 2024.