IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 January 2026, accepted 28 January 2026, date of publication 6 February 2026, date of current version 11 February 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3661737

RESEARCH ARTICLE

On Fairness Degradation in Cell-Free Multi-Stream Massive MIMO With Mobile-Grade Antennas

MUTEEN MUNAWAR, (Student Member, IEEE), MAMOUN GUENACH, (Senior Member, IEEE), AND INGRID MOERMAN, (Senior Member, IEEE)
Interuniversity Microelectronics Centre (IMEC), 3001 Leuven, Belgium
IDLab, Ghent University, 9000 Ghent, Belgium

Corresponding author: Muteen Munawar (muteen.munawar@imec.be)

This work was supported by European Community's Research Foundation Flanders (FWO) through the Project Strip-Link multiple-input multiple-output (MIMO) FWO—97707 under Grant A2582000101.

ABSTRACT For the commercial deployment of cell-free (CF) massive MIMO, mobile-grade antennas are envisioned for use in low-cost, low-complexity access points. It is essential to consider the practical constraints of these mobile-grade components when comparing the performance of CF and cellular networks. In this context, this paper focuses on the following main aspects: 1) we consider a generalized CF system with multiple APs and UEs, each equipped with multiple antennas and supporting multiple data streams per user. We exploit mobile-grade antennas with lower per-antenna transmit power constraints and compare their performance with conventional cellular networks, which are typically less constrained in this regard; and 2) we study a novel optimization problem aimed at achieving performance fairness in the presence of multiple data streams per user. To solve this problem, we derive convex formulations for the number of data streams, power allocations, and fairness weights. Based on these formulations, we propose a multistep alternating optimization (mAO) algorithm. To reduce the computational complexity of the mAO algorithm, we also propose a second algorithm based on closed-form solutions, which does not require convex solvers. Given the same total power budget in CF and cellular systems, the numerical analysis highlights the impact of mobile-grade antennas and shows that the fairness difference between cellular and CF is not as significant as portrayed in idealized scenarios. Specifically, we show that to achieve a sufficient fairness gap, CF requires up to 20 times more print-quality transmit antennas, i.e., radio stripes, than cellular systems. To assist future comparative studies, we also provide a computational complexity analysis of the proposed solutions.

INDEX TERMS Cell-free systems, mobile-grade antennas, fairness degradation, multi-stream transmission, alternating optimization, mFPI-AO algorithm, computational complexity, power allocation, fronthaul constraints, generalized cell-free architecture.

I. INTRODUCTION

The evolution of wireless communication systems has seen a major shift with the introduction of new network topologies. One such concept is the cell-free (CF) massive multiple-input multiple-output (mMIMO) wireless network [1], [2]. This design moves away from traditional cellular networks by incorporating macro-diversity and improved interference management as key principles [3], [4]. The motivation for CF mMIMO comes from the limitations found in conventional cellular systems, including issues with non-uniform performance, interference, and limited scalability to support the rising demand for high data rates and consistent connectivity [5]. To address these challenges, the

The associate editor coordinating the review of this manuscript and approving it for publication was Chuan Heng Foh.

CF mMIMO architecture is proposed as a strong candidate to help shape the future of wireless communication systems [6].

Consequently, a substantial body of literatures [1], [7], and [8] has systematically compared the performance of cellular and cell-free (CF) wireless systems.

For example, in [9], the authors studied the total energy efficiency of a CF mMIMO system. In [10], the authors investigated the coherent gains of CF systems and discussed the associated fronthaul overhead. Similarly, [11] and [12] examined the spectral efficiency and energy efficiency, respectively, of CF systems and compared them with those of conventional cellular systems.

A consistent finding across these studies, including more recent works such as [13], [14], and [15], is that CF systems outperform cellular systems due to their macro-diversity and centralized processing capabilities [3], [16]. However, a critical aspect to note is that the deployment of CF systems requires a considerable number of access points (APs) distributed across the region [4], [5]. The envisioned commercial deployment relies on fabricating these numerous APs with mobile-grade components, such as low-cost radio frequency (RF) chains and antennas. For example, print-quality radio stripes are promising candidates for this purpose [17]. The practical performance of CF systems could be significantly affected by the use of these lower-quality components, in contrast to the high-quality RF components commonly deployed in conventional cellular systems. Therefore, when comparing the performance of CF and cellular systems, it is crucial to consider the constraints imposed by the use of these components, a practice that is not commonly observed in the existing literature.

Another important aspect of cell-free architecture is achieving fair and uniform performance among users geographically located in different areas. To achieve fairness, several works in the literature discuss algorithms to optimize spatial resources, such as digital filters and power allocations, to meet this objective [1], [7], [8]. It is worth noting that modern communication systems involve multiple antennas not only at the transmitter side but also at the user-end devices. Having multiple antennas at the user-end device unlocks the potential for multiple spatial streams per user. To optimally achieve performance fairness among users, where each user has multiple receive antennas, it is necessary to optimize the number of data streams, a discrete variable.

A few relevant works that we could identify are [18], [19], [20], [21]. Specifically, [18] considers multi-antenna users; however, it assumes simple conjugate beamforming, does not include an explicit objective function, does not consider multistream transmission, and ignores per-antenna power constraints. Similarly, [19] studies a CF mMIMO system with multi-antenna users employing nonlinear digital filters, but does not address multistream optimization or the impact of antenna radiation power.

In addition, [20] considers a system model similar to ours; however, the study focuses on pilot design and channel estimation. The work most closely related to the proposed work is [21], which investigates a CF mMIMO wireless setup with multi-antenna UEs. However, the data stream optimization in that work is based on maximizing diversity gains rather than optimizing the number of streams per user under a fairness objective, and it primarily focuses on access point clustering, without analyzing fairness degradation.

To the best of our knowledge, achieving fairness with multiple data streams per user has not yet been studied in the literature. We believe that optimizing data streams while achieving fairness is an important aspect that should be studied to unlock the full spatial benefits of CF massive MIMO.

A. CONTRIBUTIONS

The main contributions of this work are summarized as follows:

- Given the per-antenna maximum power constraints of mobile-grade antennas, we formulate a non-convex optimization problem related to fairness performance, whose objective is coupled with a discrete optimization variable due to the possibility of multiple data streams per user. To address this problem, we derive convex formulations for the power allocations, fairness weights per user, and the number of data streams per user.
- Based on the solutions, we propose a novel multistep alternating optimization (mAO) algorithm that solves the problem iteratively, maintains monotonic behavior, and guarantees convergence.
- To reduce the computational complexity, we propose a second algorithm, which is based on closed-form expressions and provides similar performance to mAO algorithm, without relying on general-purpose convex solvers.
- Finally, through numerical analysis, we demonstrate that the fairness of CF wireless systems is substantially influenced when considering per-antenna power constraints.
- We also provide a computational complexity analysis of the proposed algorithms.

Our numerical analysis addresses the following two questions:

- If radio strips [22] are used to implement CF systems, will it impact the SE or energy efficiency (EE) gaps compared to cellular systems, due to lower per-antenna power constraints?
- If it does impact, how many of these print-quality antennas would be needed to maintain a sufficient performance gap?

The remainder of this paper is organized as follows. Section II presents the system model and formulates the problem. Sections III and IV describe the developed solutions and the proposed algorithm. In Section V, we provide a numerical analysis and discussion. Finally, Section VI concludes the paper.

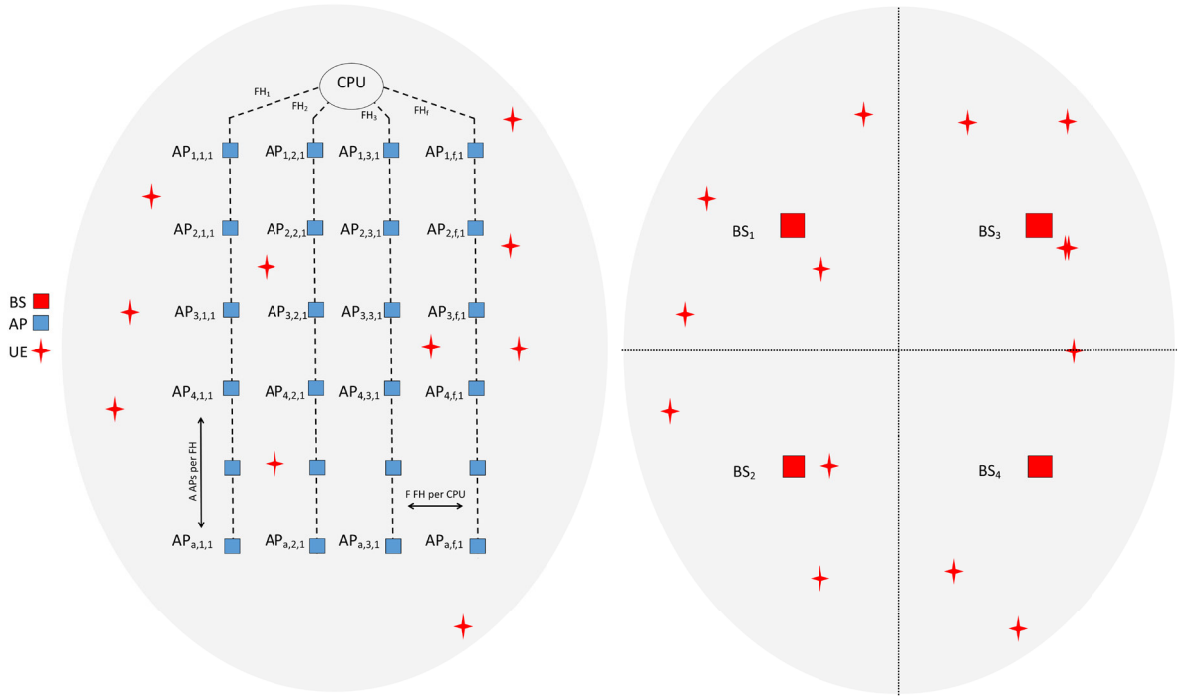

FIGURE 1. System model.

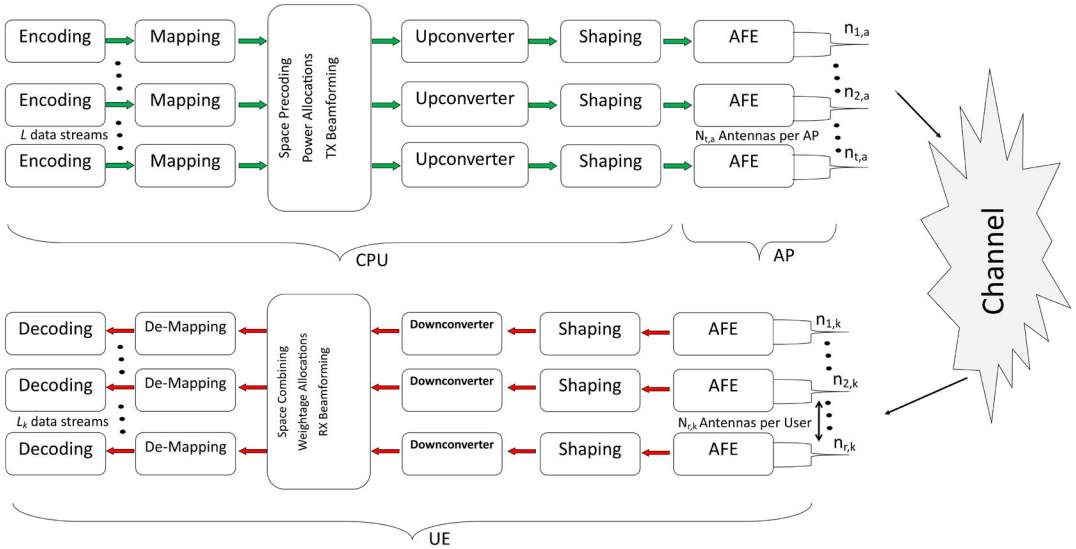

FIGURE 2. Block Diagram of CPU, AP, and UE.

Notations: Scalars are denoted by italic letters, while vectors and matrices are represented by bold-face lower- and upper-case letters, respectively. For a complex-valued vector $\mathbf{v}$ of length K, $\mathbf{v}^H$ denotes the conjugate transpose of $\mathbf{v}$. diag($\mathbf{v}$) denotes a diagonal matrix where each diagonal element is the corresponding element of vector $\mathbf{v}$. max($\mathbf{v}$) represents the maximum value in the vector $\mathbf{v}$. For a complex-valued matrix $\mathbf{M}$, $[\mathbf{M}]_1^2$ denotes the square of the absolute values obtained by column-wise summing the matrix $\mathbf{M}$ and diag($\mathbf{M}$) is a vector containing the diagonal elements of the matrix $\mathbf{M}$. The notation $(\cdot)^*$ refers to the complex conjugate of a complex number. $\hat{e}_{\max}(\mathbf{A}, \mathbf{B})$ denotes the dominant generalized eigenvector of the matrix pair $(\mathbf{A}, \mathbf{B})$. The Hadamard product of two vectors $\mathbf{a}$ and $\mathbf{b}$ is denoted by $\mathbf{a} \odot \mathbf{b}$.

II. SYSTEM MODEL

We consider a generalized CF mMIMO wireless communication system as shown in Fig. 1, which is equipped with

A APs per fronthaul and F fronthauls per CPU. Each AP is equipped with $N_{t,a}$ transmit antennas, where $a = 1, \ldots, A$. Similarly, there are a total of K users in the system, each equipped with $N_{r,k}$ receive antennas and L_k data streams, where $k = 1, \ldots, K$. This results in a total of $N_t = \sum_{a=1}^{A} F \cdot N_{t,a}$, $N_r = \sum_{k=1}^{K} N_{r,k}$, and $L = \sum_{k=1}^{K} L_k$ transmit antennas, receive antennas, and data streams in the system, respectively. Note that the index of each AP is denoted by $AP_{a,f,1}$ in Fig. 1, where a, f, and 1 stand for the index of the a-th AP connected to the f-th fronthaul and the 1st CPU. In a generalized CF MIMO system, processing power is envisioned to be distributed across the network; however, for brevity, we show only one CPU here, corresponding to index 1. Additionally, we depict a block diagram of the CPU, AP, and UE in Fig. 2. Specifically, Fig. 2 illustrates how the L data streams in the system are mapped, precoded, and projected onto the $N_{t,a}$ antennas of an AP, and, upon reception, how UE k retrieves its L_k data streams. Specifically, the CPU+AP and UE consist of several blocks, namely coding, mapping, conversion, shaping, and the analog front end (AFE). To avoid intricate mathematical formulations, the additional indices for fronthaul and CPU are omitted for the sake of our analysis. The downlink channel between AP a and user k is denoted by $\mathbf{H}_{k,a}^{H} \in \mathbb{C}^{N_{r,k} \times N_{t,a}}$, which follows a standard block fading model [23]. The total channel from all APs to user k is then given by $\mathbf{H}_{k}^{H} \in \mathbb{C}^{N_{r,k} \times N_t}$, i.e., $\mathbf{H}_{k}^{H} = \left[\mathbf{H}_{k,1}, \mathbf{H}_{k,2}, \ldots, \mathbf{H}_{k,A}\right]^{H}$.

Let $s_{k,l}$ and $p_{k,l}$ denote the data symbol and power allocation for the l-th data stream of user k, respectively. The received signal at user k can be expressed as shown in (1), at the bottom of the page, where $\mathbf{w}_{k,a,l} \in \mathbb{C}^{N_{t,a} \times 1}$ denotes the transmit digital filter (beamforming vector) at AP a for the l-th data stream of user k, and $\mathbf{z}_k$ denotes the noise vector. The first column of $\mathbf{W}$ in (1) can be conceptualized as the cumulative transmit digital filter, i.e., $\mathbf{w}_{k,l}$, from A APs for the l-th data stream of the k-th user.

After receiving the signal $\mathbf{y}_k$, user k applies the receive combining matrix, i.e., the receive digital filter, to retrieve $\hat{\mathbf{s}}_k$. Hence, the retrieved signal is expressed as

$$\begin{aligned}\hat{\mathbf{s}_k} &= \underbrace{\left[\mathbf{f}_{k,1}, \mathbf{f}_{k,2}, \cdots, \mathbf{f}_{k,L_k}\right]}_{\mathbf{F}_k}^{H} \mathbf{y}_k \\ &= \underbrace{\left[\mathbf{f}_{k,1}, \mathbf{f}_{k,2}, \cdots, \mathbf{f}_{k,L_k}\right]}_{\mathbf{F}_k}^{H} \mathbf{H}_k^{H} \mathbf{WPs} \\ &\quad + \underbrace{\left[\mathbf{f}_{k,1}, \mathbf{f}_{k,2}, \cdots, \mathbf{f}_{k,L_k}\right]}_{\mathbf{F}_k}^{H} \mathbf{z}_k, \end{aligned} \tag{2}$$

where $\mathbf{f}_{k,l}^{H} \in \mathbb{C}^{1 \times N_{r,k}}$ denotes the receive combining vector for the l-th data stream of user k.

Given the signal model from (2), the spectral efficiency (SE) for user k, denoted as c_k, provided by L_k data streams, can be written as (3), shown at the bottom of the next page.

The meaning of each expression in (3) is as follows: The term

$$\left| \mathbf{f}_{k,l}^{H} \sum_{a=1}^{A} \mathbf{H}_{k,a}^{H} \mathbf{w}_{k,a,l} \sqrt{p_{k,a,l}} \right|^2$$

denotes the received signal strength for the l-th data stream of user k, while

$$\sum_{i \neq k}^{K} \sum_{j=1}^{L_i} \left| \mathbf{f}_{k,l}^{H} \sum_{a=1}^{M} \mathbf{H}_{i,a}^{H} \mathbf{w}_{i,a,j} \sqrt{p_{i,a,j}} \right|^2$$

and

$$\sum_{j \neq l}^{L_k} \left| \mathbf{f}_{k,l}^{H} \sum_{a=1}^{A} \mathbf{H}_{k,a}^{H} \mathbf{w}_{k,a,j} \sqrt{p_{k,a,j}} \right|^2$$

denote the inter-user and intra-user interference, respectively.

A. PROBLEM FORMULATION

We aim to ensure performance fairness among users when multiple data streams per user are involved, while considering the maximum power constraints per antenna, per AP, and per system. To achieve this, we need to optimize $\mathbf{w}_{k,l}$, $\mathbf{f}_{k,l}$, $p_{k,l}$, and L_k, for all k and l. Thus, the formulated problem is given in (P1), as shown at the bottom of the next page. Note that the objective function of (P1) contains log-sum expressions and fractions, and is coupled with the discrete variable L_k, making it a non-convex mixed-integer problem.

The meaning of each expression in (P1) is as follows: Constraint 1, i.e.,

$$\sum_{k=1}^{K} \sum_{l=1}^{L_k} \left[\mathbf{w}_{k,1,l}, \mathbf{w}_{k,2,l}, \cdots, \mathbf{w}_{k,A,l} \right] \odot \left[\sqrt{p_{k,1,l}}, \sqrt{p_{k,2,l}}, \cdots, \sqrt{p_{k,A,l}} \right] \leq p_{\max,ant}$$

$$\mathbf{y}_k = \underbrace{\begin{bmatrix} \mathbf{H}_{k,1} \\ \mathbf{H}_{k,2} \\ \vdots \\ \mathbf{H}_{k,a} \\ \vdots \\ \mathbf{H}_{k,A} \end{bmatrix}^{H}}_{\mathbf{H}_k^H} \underbrace{\begin{bmatrix} \mathbf{w}_{1,1,1} & \mathbf{w}_{1,1,2} & \cdots & \mathbf{w}_{k,1,l} & \cdots & \mathbf{w}_{K,1,L_K} \\ \mathbf{w}_{1,2,1} & \mathbf{w}_{1,2,2} & \cdots & \mathbf{w}_{k,2,l} & \cdots & \mathbf{w}_{K,2,L_K} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \mathbf{w}_{1,a,1} & \mathbf{w}_{1,a,2} & \cdots & \mathbf{w}_{k,a,l} & \cdots & \mathbf{w}_{K,a,L_K} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \mathbf{w}_{1,A,1} & \mathbf{w}_{1,A,2} & \cdots & \mathbf{w}_{k,A,l} & \cdots & \mathbf{w}_{K,A,L_K} \end{bmatrix}}_{\mathbf{W}} \underbrace{\begin{bmatrix} \sqrt{p_{1,1}} & & & & & \\ & \sqrt{p_{1,2}} & & & & \\ & & \ddots & & & \\ & & & \sqrt{p_{k,l}} & & \\ & & & & \ddots & \\ & & & & & \sqrt{p_{K,L_K}} \end{bmatrix}}_{\mathbf{P}} \underbrace{\begin{bmatrix} s_{1,1} \\ s_{1,2} \\ \vdots \\ s_{k,l} \\ \vdots \\ s_{K,L_K} \end{bmatrix}}_{\mathbf{s}} + \mathbf{z}_k \tag{1}$$

denotes the power allocated to each antenna at the APs, where $p_{\max,ant}$ is the maximum power radiation constraint per antenna.

For instance, consider the first term on the left-hand side of Constraint 1 in (P1), where $\mathbf{w}_{k,1,l}$ is element-wise multiplied with $\sqrt{p_{k,1,l}}$ and summed over L_k and K. This results in the precoder coefficients for the first antenna at the AP:

$$\sum_{k=1}^{K}\sum_{l=1}^{L_k}\left(\mathbf{w}_{k,1,l}\sqrt{p_{k,1,l}}\right).$$

Subsequently, by taking the squared absolute value, we obtain the power allocation for each antenna at that AP:

$$\left|\sum_{k=1}^{K}\sum_{l=1}^{L_k}\left(\mathbf{w}_{k,1,l}\sqrt{p_{k,1,l}}\right)\right|^2.$$

This formulation is similarly applied to the antennas at other APs.

Now, the second and third constraints correspond to the per-AP and per-system power limits. We note that a per-antenna power constraint implicitly enforces maximum power restrictions on both the AP and the overall system. However, in a generalized system, antennas may have a high capacity for fine power tuning, while the total power budget per AP could be lower than the combined maximum radiation capability of its antennas. Therefore, we retain all three constraints in our problem formulation.

Before deriving a solution for (P1), we first review the relevant literature. Most studies addressing the resource allocation problem in the context of CF systems propose algorithms similar to [24] or modified versions of these approaches. In particular, these fairness-based algorithms are typically limited to single-stream-per-user expressions and do not account for the optimization of the variable L_k to find the optimal trade-off between diversity and multiplexing per user while achieving fairness.

As discussed earlier, when both the AP and UE are equipped with multiple antennas, it is necessary to optimize L_k for all k to ensure optimal fairness among users. In the following two sections, we study (P1), derive convex formulations for mixed-integer variables, and introduce two multistep AO algorithms. Additionally, in Section III, we discuss how an existing method in [24], i.e., the fixed-point iteration algorithm, can be modified to solve (P1).

III. PROPOSED SOLUTION TO PROBLEM (P1)

We first decouple the problem (P1) with respect to each optimization variable and derive its solution. We begin by assuming an initial setup where the number of data streams for user k, denoted by L_k, and the corresponding power allocations $p_{k,l}$ for all users are set as $\min(N_t, N_{r,k})$ and $\frac{P_{\max}}{L_k}$, respectively. In this case, the optimal linear MMSE filters for the transmit and receive digital filters, $\mathbf{w}_{k,l}$ and $\mathbf{f}_{k,l}$, are given as follows:

$$\mathbf{w}_{k,l} = \frac{\hat{e}_{\max}\left(\mathbf{S}_{k,l}^{\mathbf{f}_{k,l}}, \mathbf{T}_{k,l}^{\mathbf{f}_{k,l}}\right)}{\left\|\hat{e}_{\max}\left(\mathbf{S}_{k,l}^{\mathbf{f}_{k,l}}, \mathbf{T}_{k,l}^{\mathbf{f}_{k,l}}\right)\right\|}, \tag{4}$$

$$\mathbf{f}_{k,l} = \frac{\hat{e}_{\max}\left(\mathbf{S}_{k,l}^{\mathbf{w}_{k,l}}, \mathbf{T}_{k,l}^{\mathbf{w}_{k,l}}\right)}{\left\|\hat{e}_{\max}\left(\mathbf{S}_{k,l}^{\mathbf{w}_{k,l}}, \mathbf{T}_{k,l}^{\mathbf{w}_{k,l}}\right)\right\|}, \tag{5}$$

$$c_k = \sum_{l=1}^{L_k}\log_2\left(1+\frac{\left|\mathbf{f}_{k,l}^H\sum_{a=1}^{A}\mathbf{H}_{k,a}^H\mathbf{w}_{k,a,l}\sqrt{p_{k,a,l}}\right|^2}{\sum_{j\neq l}^{L_k}\left|\mathbf{f}_{k,l}^H\sum_{a=1}^{A}\mathbf{H}_{k,a}^H\mathbf{w}_{k,a,j}\sqrt{p_{k,a,j}}\right|^2+\sum_{i\neq k}^{K}\sum_{j=1}^{L_i}\left|\mathbf{f}_{k,l}^H\sum_{a=1}^{M}\mathbf{H}_{i,a}^H\mathbf{w}_{i,a,j}\sqrt{p_{i,a,j}}\right|^2+\sigma^2}\right) \tag{3}$$

$$\begin{aligned}
\max_{\mathbf{w}_{k,a,l},\mathbf{f}_{k,l},L_k,p_{k,a,l}}\ \min_{k\in\{1,\ldots,K\}}\ & \sum_{l=1}^{L_k}\log_2\left(1+\frac{\left|\mathbf{f}_{k,l}^H\sum_{a=1}^{A}\mathbf{H}_{k,a}^H\mathbf{w}_{k,a,l}\sqrt{p_{k,a,l}}\right|^2}{\sum_{j\neq l}^{L_k}\left|\mathbf{f}_{k,l}^H\sum_{a=1}^{A}\mathbf{H}_{k,a}^H\mathbf{w}_{k,a,j}\sqrt{p_{k,a,j}}\right|^2+\sum_{i\neq k}^{K}\sum_{j=1}^{L_i}\left|\mathbf{f}_{k,l}^H\sum_{a=1}^{A}\mathbf{H}_{i,a}^H\mathbf{w}_{i,a,j}\sqrt{p_{i,a,j}}\right|^2+\sigma^2}\right)\\
\text{s.t.}\ & \left(\sum_{k=1}^{K}\sum_{l=1}^{L_k}\left[\mathbf{w}_{k,1,l},\ \mathbf{w}_{k,2,l},\ \ldots,\ \mathbf{w}_{k,A,l}\right]\odot\left[\sqrt{p_{k,1,l}},\ \sqrt{p_{k,2,l}},\ \ldots,\ \sqrt{p_{k,A,l}}\right]\right)^{|\cdot|^2}\leq p_{\max,ant},\\
& \sum_{k=1}^{K}\sum_{l=1}^{L_k}p_{k,a,l}\leq p_{\max,ap},\quad \forall a,\\
& \sum_{a=1}^{A}\sum_{k=1}^{K}\sum_{l=1}^{L_k}p_{k,a,l}\leq P_{\max}.
\end{aligned} \tag{P1}$$

where the matrix $\mathbf{S}_{k,l}^{\mathbf{f}_{k,l}} = \mathbf{H}_k \mathbf{f}_{k,l} \mathbf{f}_{k,l}^H \mathbf{H}_k^H$ represents the correlation matrix associated with the received signal strength, and $\mathbf{T}_{k,l}^{\mathbf{f}_{k,l}}$ is defined as

$$\mathbf{T}_{k,l}^{\mathbf{f}_{k,l}} = \sum_{i \neq k}^{K} \sum_{j=1}^{L_i} \rho_{i,j} \mathbf{H}_i \mathbf{f}_{i,j} \mathbf{f}_{i,j}^H \mathbf{H}_i^H + \sum_{j \neq l}^{L_k} p_{k,j} \mathbf{H}_k \mathbf{f}_{k,j} \mathbf{f}_{k,j}^H \mathbf{H}_k^H + \sigma^2 \mathbf{I}, \tag{6}$$

which represents the interference-plus-noise matrix, assuming that $\mathbf{f}_{k,l}$ remains constant. Similarly, the correlation matrix for the received signal strength when $\mathbf{w}_{k,l}$ is fixed is given by $\mathbf{S}_{k,l}^{\mathbf{w}_{k,l}} = \mathbf{H}_k^H \mathbf{w}_{k,l} \mathbf{w}_{k,l}^H \mathbf{H}_k$. The corresponding matrix for interference plus noise in this case is represented by

$$\mathbf{T}_{k,l}^{\mathbf{w}_{k,l}} = \sum_{i \neq k}^{K} \sum_{j=1}^{L_i} \rho_{i,j} \mathbf{H}_k^H \mathbf{w}_{i,j} \mathbf{w}_{i,j}^H \mathbf{H}_k + \sum_{j \neq l}^{L_k} p_{k,j} \mathbf{H}_k^H \mathbf{w}_{k,j} \mathbf{w}_{k,j}^H \mathbf{H}_k + \sigma^2 \mathbf{I}, \tag{7}$$

which accounts for the interference and noise when $\mathbf{w}_{k,l}$ is fixed. The optimization process for $\mathbf{w}_{k,l}$ and $\mathbf{f}_{k,l}$ is described in detail in Algorithm 1.

Given $\mathbf{w}_{k,l}$ and $\mathbf{f}_{k,l}$ for all k and l from (4) and (5), respectively, the problem (P1) for L_k for all k and the corresponding $p_{k,l}$ for all k, l can be formulated as

$$\begin{aligned}
&\max_{p_{k,l}, L_k} \min_{k \in \{1,\ldots,K\}} \sum_{l=1}^{\min(N_t, N_{r,k})} \log_2 \left(1 + \frac{p_{k,l} a_{k,l}}{b_{i,m} + c_{k,j} + \sigma^2}\right) \\
&\text{s.t.} \quad a_{k,l} = \mathbf{f}_{k,l}^H \mathbf{H}_k^H \mathbf{w}_{k,l} \mathbf{w}_{k,l}^H \mathbf{H}_k \mathbf{f}_{k,l}, \forall k, l, \\
&\qquad b_{k,l} = \sum_{i \neq k}^{K} \sum_{m=1}^{L_i} p_{i,m} \mathbf{f}_{k,l}^H \mathbf{H}_k^H \mathbf{w}_{i,m} \mathbf{w}_{i,m}^H \mathbf{H}_k \mathbf{f}_{k,l}, \forall k, l, \\
&\qquad c_{k,l} = \sum_{j \neq l}^{L_k} p_{k,j} \mathbf{f}_{k,l}^H \mathbf{H}_k^H \mathbf{w}_{k,j} \mathbf{w}_{k,j}^H \mathbf{H}_k \mathbf{f}_{k,l}, \forall k, l, \\
&\qquad \left(\sum_{k=1}^{K} \sum_{l=1}^{\min(N_t, N_{r,k})} \mathbf{w}_{k,l} \odot \mathbf{p}_{k,l} \right)^{|\cdot|^2} \leq p_{\max,ant}, \\
&\qquad \sum_{k=1}^{K} \sum_{l=1}^{\min(N_t, N_{r,k})} p_{k,a,l} \leq p_{\max,ap}, \quad \forall a, \\
&\qquad \sum_{a=1}^{A} \sum_{k=1}^{K} \sum_{l=1}^{\min(N_t, N_{r,k})} p_{k,a,l} \leq P_{\max}, \qquad \text{(P1.1)}
\end{aligned}$$

where $a_{k,l}$, $b_{k,l}$, and $c_{k,l}$ denote the signal strength, interuser interference, and intrauser interference of the l-th data stream of user k, respectively. The fourth constraint in (P1.1) is the same as the first constraint of (P1), with $\mathbf{w}_{k,l} = [\mathbf{w}_{k,1,l}, \mathbf{w}_{k,2,l}, \cdots, \mathbf{w}_{k,A,l}]$ and $\mathbf{p}_{k,l} = [p_{k,1,l}, p_{k,2,l}, \cdots, p_{k,A,l}]$ for brevity.

To derive a convex formulation of (P1.1), we need to address two main challenges:

1) The log-sum function, which is coupled with the discrete variable L_k.
2) The fractional expressions in the objective function of (P1.1).

To tackle these issues, we introduce auxiliary variables, denoted as $\varsigma_k, \forall k$. These variables represent the user performances. Specifically, by incorporating these auxiliary variables, (P1.1) can be rewritten as follows:

$$\begin{aligned}
&\max_{\mathbf{p}, L_k, \varsigma_k} \varsigma_{k,\min} \\
&\text{s.t.} \quad \sum_{l=1}^{\min(N_t, N_{r,k})} \log_2 \left(1 + \frac{p_{k,l} a_{k,l}}{b_{i,m} + c_{k,j} + \sigma^2}\right) \geq \varsigma_{k,\min}, \\
&\qquad \left(\sum_{k=1}^{K} \sum_{l=1}^{\min(N_t, N_{r,k})} \mathbf{w}_{k,l} \odot \mathbf{p}_{k,l} \right)^{|\cdot|^2} \leq p_{\max,ant}, \\
&\qquad \sum_{k=1}^{K} \sum_{l=1}^{\min(N_t, N_{r,k})} p_{k,a,l} \leq p_{\max,ap}, \quad \forall a, \\
&\qquad \sum_{a=1}^{A} \sum_{k=1}^{K} \sum_{l=1}^{\min(N_t, N_{r,k})} p_{k,a,l} \leq P_{\max}. \qquad \text{(P1.2)}
\end{aligned}$$

where $\varsigma_{k,\min}$ stands for $\min_{k \in \{1,\ldots,K\}} \varsigma_k$, denoting the auxiliary variable of the user with the minimum performance.

We now define $\varsigma_{k,l}$ as auxiliary variables for each ς_k, for all k, l, such that

$$\sum_{l=1}^{L_k} \varsigma_{k,l} = \varsigma_k.$$

This step will help remove the summation expression from the first constraint of (P1.2) and rewrite it as a separate convex constraint. More specifically, by introducing the additional variables $\varsigma_{k,l}$, $\forall k, l$, (P1.2) can be reformulated as follows: This approach will allow us to eliminate the summation expression from the first constraint of (P1.2), and transform it into a separate convex constraint. Specifically, by introducing the auxiliary variables $\varsigma_{k,l}$, for all k, l, (P1.2) can be restated as:

$$\begin{aligned}
&\max_{\mathbf{p}, L_k, \varsigma_k, \varsigma_{k,l}} \sum_{l=1}^{\min(N_t, N_{r,k})} \varsigma_{k,l,\min} \\
&\text{s.t.} \quad \sum_{l=1}^{\min(N_t, N_{r,k})} \log_2 \left(1 + \frac{p_{k,l} a_{k,l}}{b_{i,m} + c_{k,j} + \sigma^2}\right) \\
&\qquad \geq \sum_{l=1}^{\min(N_t, N_{r,k})} \varsigma_{k,l}, \\
&\qquad \left(\sum_{k=1}^{K} \sum_{l=1}^{\min(N_t, N_{r,k})} \mathbf{w}_{k,l} \odot \mathbf{p}_{k,l} \right)^{|\cdot|^2} \leq p_{\max,ant},
\end{aligned}$$

$$\sum_{k=1}^{K}\sum_{l=1}^{\min(N_t,N_{r,k})} p_{k,a,l} \le p_{\max,\mathrm{ap}}, \quad \forall a,$$
$$\sum_{a=1}^{A}\sum_{k=1}^{K}\sum_{l=1}^{\min(N_t,N_{r,k})} p_{k,a,l} \le P_{\max}. \tag{P1.3}$$

It is important to note that there is now a direct mapping between the left and right-hand sides of the first constraint in (P1.3). As a result, the summation term can be eliminated from the first constraint and instead introduced as an additional constraint, as follows:

$$\max_{\mathbf{p},L_k,\varsigma_k,\varsigma_{k,l}} \sum_{l=1}^{L_k} \varsigma_{k,l,\min}$$
$$\text{s.t.} \quad \log_2\left(1+\frac{p_{k,l}a_{k,l}}{b_{i,m}+c_{k,j}+\sigma^2}\right) \ge \varsigma_{k,l}, \quad \forall l,$$
$$\left(\sum_{k=1}^{K}\sum_{l=1}^{\min(N_t,N_{r,k})} \mathbf{w}_{k,l}\odot\mathbf{p}_{k,l}\right)^{|\cdot|^2} \le p_{\max,\mathrm{ant}},$$
$$\sum_{k=1}^{K}\sum_{l=1}^{\min(N_t,N_{r,k})} p_{k,a,l} \le p_{\max,\mathrm{ap}}, \quad \forall a,$$
$$\sum_{a=1}^{A}\sum_{k=1}^{K}\sum_{l=1}^{\min(N_t,N_{r,k})} p_{k,a,l} \le P_{\max},$$
$$\sum_{l=1}^{\min(N_t,N_{r,k})} \varsigma_{k,l} = \varsigma_k. \tag{P1.4}$$

Now, it is straightforward to show that the convex formulation of the first constraint of (P1.4), i.e.,

$$\log_2\left(1+\frac{p_{k,l}a_{k,l}}{b_{i,m}+c_{k,j}+\sigma^2}\right) \ge \varsigma_{k,l},$$

can be derived as follows:

$$p_{k,l}a_{k,l} \ge \left(2^{\varsigma_{k,l}}-1\right)\left(b_{i,m}+c_{k,j}+\sigma^2\right).$$

Hence, the final convex formulation is given by:

$$\max_{\mathbf{p},L_k,\varsigma_k,\varsigma_{k,l}} \sum_{l=1}^{\min(N_t,N_{r,k})} \varsigma_{k,l,\min}$$
$$\text{s.t.} \quad p_{k,l}a_{k,l} \ge \left(2^{\varsigma_{k,l}}-1\right)\left(b_{i,m}+c_{k,j}+\sigma^2\right), \quad \forall l,$$
$$\left(\sum_{k=1}^{K}\sum_{l=1}^{\min(N_t,N_{r,k})} \mathbf{w}_{k,l}\odot\mathbf{p}_{k,l}\right)^{|\cdot|^2} \le p_{\max,\mathrm{ant}},$$
$$\sum_{k=1}^{K}\sum_{l=1}^{\min(N_t,N_{r,k})} p_{k,a,l} \le p_{\max,\mathrm{ap}}, \quad \forall a,$$
$$\sum_{a=1}^{A}\sum_{k=1}^{K}\sum_{l=1}^{\min(N_t,N_{r,k})} p_{k,a,l} \le P_{\max},$$
$$\sum_{l=1}^{\min(N_t,N_{r,k})} \varsigma_{k,l} = \varsigma_k. \tag{P1.5}$$

Note that the convex formulation in (P1.5) is derived from (P1.1) without any loss in optimality. It can be efficiently solved using general-purpose convex solvers, such as CVX [25]. Solving (P1.5) with CVX yields the optimal power allocations and auxiliary weights, from which the optimal L_k can be determined as the number of data streams assigned non-zero power, i.e., $p_{k,l} > 0$ for the corresponding k. Further details on the optimization of (P1), based on (4), (5), and (P1.5), are provided in Algorithm 1.

Algorithm 1 Multi-Step AO Algorithm for Solving (P1)

1: **Initialize:** Set iteration index $\tau = 0$, and initialize:
- $L_{k,\tau} = \min(N_t, N_{r,k})$, for all k
- $\mathbf{W}_\tau = \frac{1}{\sqrt{N_t}}\left[\mathbf{1}_{1,L_1},\ldots,\mathbf{1}_{K,L_K}\right]^T$
- $\mathbf{p}_\tau = \frac{P_{\max}}{\sum_{k=1}^{K} L_{k,\tau}}\left[\mathbf{1}_{1,L_1},\ldots,\mathbf{1}_{K,L_K}\right]$
- $\varsigma_{k,\tau}$, $\forall k$, random but feasible
- $\varsigma_{k,l,\tau}$, $\forall k, l$, such that $\sum_{l=1}^{L_{k,\tau}} \varsigma_{k,l,\tau} = \varsigma_{k,\tau}$
- Fairness tolerance ε

2: **repeat**
3: **Step 1:** Given $\mathbf{F}_{k,\tau}$, $L_{k,\tau}$, $\varsigma_{k,\tau}$, $\varsigma_{k,l,\tau}$ for all k, l, and $\mathbf{p}_\tau$, compute $\mathbf{W}_{\tau+1}$ using (4).
4: **Step 2:** Given $\mathbf{p}_\tau$, $L_{k,\tau}$, $\varsigma_{k,\tau}$, $\varsigma_{k,l,\tau}$ for all k, l, and $\mathbf{W}_{\tau+1}$, compute $\mathbf{F}_{k,\tau+1}$ for all k using (5).
5: **Step 3:** Given $\mathbf{W}_{\tau+1}$, $\mathbf{F}_{k,\tau+1}$ for all k, $L_{k,\tau}$, $\varsigma_{k,l,\tau+1}$ for all k, l, and $\mathbf{p}_\tau$, compute $\varsigma_{k,\tau+1}$ for all k using (P1.5).
6: **Step 4:** Given $\mathbf{W}_{\tau+1}$, $\mathbf{F}_{k,\tau+1}$ for all k, $L_{k,\tau}$, $\varsigma_{k,\tau+1}$ for all k, and $\mathbf{p}_\tau$, compute $\varsigma_{k,l,\tau+1}$ for all k, l using (P1.5).
7: **Step 5:** Given $\mathbf{W}_{\tau+1}$, $\mathbf{F}_{k,\tau+1}$ for all k, $\varsigma_{k,\tau+1}$ and $\varsigma_{k,l,\tau+1}$ for all k, l, compute $\mathbf{p}_{\tau+1}$ and $L_{k,\tau+1}$ using (P1.5).
8: **Step 6:** Update the iteration index: $\tau \leftarrow \tau + 1$
9: **until** The stopping criterion is satisfied, i.e.,

$$\max_{k\in\{1,\ldots,K\}} \delta_{k,\tau} - \min_{k\in\{1,\ldots,K\}} \delta_{k,\tau} \le \varepsilon,$$

or the maximum number of iterations is reached.

IV. LOW-COMPLEXITY PROPOSED SOLUTION TO PROBLEM (P1)

To solve (P1), Algorithm 1 relies on a general-purpose convex solver [25], making it computationally expensive and thus impractical for real-time applications. To reduce the computational complexity, we propose a modified fixed-point iteration-based alternating optimization algorithm (MFPI-AO), which ensures fairness among users while also accounting for per-user multi-stream allocations. Specifically, we show that, through the introduction and appropriate selection of auxiliary variables, problem (P1.1) can be transformed into a more recognizable formulation, to which the existing fixed-point iteration (FPI) algorithm [14], [26] can be adapted. We refer to this overall approach, which

incorporates auxiliary variables, as MFPI-AO. Let us now reconsider (P1.1).

Similar to (P1.3), we introduce auxiliary variables $\zeta_{k,l}, \forall k, l$ for each data stream in the system. Optimizing $\zeta_{k,l}, \forall k, l$ allows us to transform problem (P1.1) into a per-data-stream performance issue. Formulating an objective function for each data stream simplifies the resolution, treating each data stream as a virtual user in the system.

However, it is noteworthy that the goal is fair performance among users with multiple data streams, not among individual data streams. An appropriate selection of variables $\zeta_{k,l}, \forall k, l$ can ensure that appropriate weight is assigned to each data stream, ensuring that, while achieving the objective of fair performance among users, the weaker eigen directions of a user do not negatively impact the performance of other users' stronger eigen directions. In fairness problems, aiming for equitable performance among all users may lead to the maximization of worst-performance users, causing degradation in the performance of the best-performance users. This issue becomes more significant when dealing with multiple data streams per user. For example, when aiming for fairness between data streams, the weaker eigen-directions or low signal-to-interference-and-noise-ratio (SINR) data streams of some users may negatively impact the performance of other users. The introduced weights, i.e., $\zeta_{k,l}, \forall k, l$, ensure that this scenario is avoided.

More details on the selection of these variables, i.e., $\zeta_{k,l}, \forall k, l$, using a closed-form formula are given in Algorithm 2. More specifically, we propose to determine the values of $\zeta_{k,l}, \forall l$ for user k as being inversely proportional to the normalized eigen strengths of $\mathbf{H}_k^H$, or inversely proportional to the normalized SINRs, as outlined in Algorithm 1. Once $p_{k,l}$ and $\zeta_{k,l}, \forall k, l$ are calculated, L_k for all k can be retrieved using the similar method as we discussed in Algorithm 1.

Nevertheless, given an appropriate number of data streams per user k, i.e., L_k and corresponding weights, i.e., $\zeta_{k,l}, \forall k, l$, allow to convert the objective function of (P1.1) from $\max\limits_{\mathbf{p}, L_k} \min\limits_{k\in\{1,\cdots,K\}} \sum\limits_{l=1}^{L_k} \log_2(1 + \frac{p_{k,l}a_{k,l}}{b_{i,m}+c_{k,j}+\sigma^2})$ to $\max\limits_{\mathbf{P}} \min\limits_{k\in\{1,\cdots,K\}\{1,\cdots,L_k\}} \log_2(1 + \frac{p_{k,l}a_{k,l}\zeta_{k,l}}{b_{i,m}+c_{k,j}+\sigma^2})$, which is essentially a function of $\frac{p_{k,l}a_{k,l'}}{b_{i,m}+c_{k,j}+\sigma^2}$ with terms (log2) and (1 +), common and constant across all data streams. By neglecting these common and constant factors, the final optimization problem for $\mathbf{p}$ is reduced to:

$$\max_{p_{k,l},\forall k,l} \min_{\substack{k\in\{1,\ldots,K\}\\ l\in\{1,\ldots,L_k\}}} \left(\frac{p_{k,l}a'_{k,l}}{b_{i,m}+c_{k,j}+\sigma^2}\right)$$
$$\text{s.t.} \quad a_{k,l} = \mathbf{f}_{k,l}^H \mathbf{H}_k^H \mathbf{w}_{k,l}\mathbf{w}_{k,l}^H \mathbf{H}_k \mathbf{f}_{k,l},$$
$$a'_{k,l} = a_{k,l}\zeta_{k,l},$$
$$b_{i,m} = \sum_{i\neq k}^{K}\sum_{m=1}^{L_i} p_{i,m}\mathbf{f}_{k,l}^H \mathbf{H}_k^H \mathbf{w}_{i,m}\mathbf{w}_{i,m}^H \mathbf{H}_k \mathbf{f}_{k,l},$$
$$c_{k,j} = \sum_{j\neq l}^{L_k} p_{k,j}\mathbf{f}_{k,l}^H \mathbf{H}_k^H \mathbf{w}_{k,j}\mathbf{w}_{k,j}^H \mathbf{H}_k \mathbf{f}_{k,l},$$
$$\left(\sum_{k=1}^{K}\sum_{l=1}^{L_k} \mathbf{w}_{k,l} \odot \mathbf{p}_{k,l}\right)^{|\cdot|^2} \leq p_{\max,ant},$$
$$\sum_{k=1}^{K}\sum_{l=1}^{L_k} p_{k,a,l} \leq p_{\max,ap}, \quad \forall a,$$
$$\sum_{a=1}^{A}\sum_{k=1}^{K}\sum_{l=1}^{L_k} p_{k,a,l} \leq P_{\max}, \tag{P1.6}$$

which can be efficiently solved based on the concept of the fixed point iteration algorithm, i.e., the inner loop of Algorithm 1. Further details on the optimization of $\mathbf{W}$, $\mathbf{F}_k$, L_k, and $p_{k,l}$, $\forall k, l$ are provided in Algorithm 1.

A. mFPI-AO ALGORITHM

The meaning of each step in Algorithm 2 is as follows: Firstly, similar to Algorithm 1, we initialize L_k, $\mathbf{W}$, power allocations, and the fairness accuracy metric. Then, we initialize two loops in Steps 2 and 3, where the nested loop essentially implements the FPI algorithm. More specifically, in Steps 4 and 5, we calculate the current SINR-proportional factor for each user's data stream and then divide the corresponding per-data-stream power by these values—meaning that data streams with higher SINRs are allocated lower power, and those with lower SINRs are allocated higher power. This process continues until the factor $\delta_{k,l,\tau}$ becomes equal for all k and all L_k. The selection of $a'_{k,l,\tau}$ ensures that fairness is maintained among users, rather than among individual data streams.

Steps 6 to 12 of Algorithm 2 check whether the power allocations exceed $P_{\max,\text{ant}}$, $P_{\max}$, or $P_{\max,\text{ap}}$. If any of these limits are violated, we normalize the power allocations to the corresponding maximum values. For example, in Step 10, the expression $\max\left[\sum_{k=1}^{K}\sum_{l=1}^{L_k} p_{k,1:a,l}\right]$ extracts the index of AP a with the highest allocated power budget and checks if it exceeds $P_{\max,a}$; if so, the power is scaled accordingly in Step 11.

In the outer loop (Step 14), we obtain power allocations from the inner loop, which are then used to optimize $\mathbf{W}$, $\mathbf{F}_k$, and L_k, similar to Algorithm 1.

Finally, in Steps 17–18, we update the weights as discussed in the context of (P1.6).

V. NUMERICAL ANALYSIS AND DISCUSSION

A. SIMULATION SETUP

To cover a communication area of 500 × 500m, the cellular network consists of four square cells, as illustrated in [3], with 50 co-located antennas per base station (BS). The CF network is deployed in the same area. Unless specified otherwise, the simulation parameters are as follows:

TABLE 1. Average CPU running time for Algorithms 1 and 2.

	Algorithm 1: Cell-free			Algorithm 2: Cell-free		
	$N_t = 64$	$N_t = 100$	$N_t = 200$	$N_t = 64$	$N_t = 100$	$N_t = 200$
$K = 5$	1.10 s	3.61 s	18.1 s	0.11 s	0.31 s	1.71 s
$K = 10$	3.39 s	7.09 s	25.75 s	0.28 s	0.62 s	3.44 s
$K = 20$	7.12 s	17.50 s	31.13 s	0.63 s	1.34 s	3.72 s

Algorithm 2 mFPI-AO Algorithm for Solving (P1)

1: Set iteration $\tau = 0$ and initialize $L_{k,\tau} = \min\left(N_t, N_{r,k}\right)$, $\mathbf{W}_\tau = \frac{1}{\sqrt{N_t}}\left[\mathbf{1}, \mathbf{1}, \cdots, \mathbf{1}_{1,L_1}, \cdots, \mathbf{1}_{K,L_K}\right]^T$, $\mathbf{p}_\tau = \frac{P_{\max}}{\sum_{k=1}^{K} \min(N_t, N_{r,k})}\left[1, 1, \cdots, 1_{1,L_1}, \cdots, 1_{K,L_K}\right]$, and fairness accuracy ε.
2: **repeat**
3: **repeat**
4: Compute $\delta_{k,l,\tau}$ as $\frac{p_{k,l,\tau} a'_{k,l,\tau}}{b_{i,m,\tau} + c_{k,j,\tau} + \sigma^2}$, $\forall k$, $\forall l$
5: Compute $p_{k,l,\tau}$ as $\frac{p_{k,l,\tau}}{\delta_{k,l,\tau}}$, $\forall k$, $\forall l$
6: Compute $\mathbf{p}_\tau$ as $\frac{p_{\max,ant}}{\max\left[\left(\sum_{k=1}^{K}\sum_{l=1}^{L_k} \mathbf{w}_{k,l} \odot \mathbf{p}_{k,l}\right)^{|\cdot|^2}\right]}\mathbf{p}_\tau$
7: **if** $sum\left(\mathbf{p}_\tau\right) > P_{\max}$ **then**
8: Update $\mathbf{p}_\tau$ as $\frac{P_{\max}}{sum(\mathbf{p}_\tau)}\mathbf{p}_\tau$
9: **end if**
10: **if** $\max\left[\sum_{k=1}^{K}\sum_{l=1}^{L_k} p_{k,1:a,l}\right] \geq p_{\max,ap}$ **then**
11: Update $\mathbf{p}_\tau$ as $\frac{P_{\max}}{\max\left[\sum_{k=1}^{K}\sum_{l=1}^{L_k} p_{k,1:a,l}\right]}\mathbf{p}_\tau$
12: **end if**
13: **until** $\max_{\substack{k\in\{1,\cdots,K\}\\ l\in\{1,\cdots,L_k\}}} \left(\delta_{k,l,\tau}\right) - \min_{\substack{k\in\{1,\cdots,K\}\\ l\in\{1,\cdots,L_k\}}} \left(\delta_{k,l,\tau}\right) > \varepsilon$ or the maximum number of iterations are completed.
14: $\mathbf{p}_{\tau+1} \leftarrow \mathbf{p}_\tau$
15: Given $\mathbf{p}_{\tau+1}$, $L_{k,\tau}$, and $\mathbf{W}_\tau$, calculate $\mathbf{F}_{k,\tau+1}$, $\forall k$ using (5).
16: Given $\mathbf{F}_{k,\tau+1}$, $\forall k$, $L_{k,\tau}$, $\forall k$, and $\mathbf{p}_{\tau+1}$, calculate $\mathbf{W}_{\tau+1}$ using (4).
17: Given $\mathbf{F}_{k,\tau+1}$, $\forall k$, $\mathbf{W}_{\tau+1}$, and $\mathbf{p}_{\tau+1}$, choose the appropriate $L_{k,\tau+1}$, $\forall k$ by neglecting the relatively worst-performing eigen directions, and select corresponding $\zeta_{k,l}$ as the inverse of achieved normalized SINRs, i.e., $[\zeta_{k,1,\tau+1}, \zeta_{k,2,\tau+1}, \cdots, \zeta_{k,L_k,\tau+1}] = \frac{\max\left([\delta_{k,1,\tau}, \delta_{k,2,\tau}, \cdots, \zeta_{k,L_k,\tau}]\right)}{[\delta_{k,1,\tau}, \delta_{k,2,\tau}, \cdots, \zeta_{k,L_k,\tau}]}$, $\forall k$.
18: $\delta_{k,l,\tau+1} \leftarrow \frac{p_{k,l,\tau+1} a'_{k,l,\tau+1}}{b_{i,m,\tau+1} + c_{k,j,\tau+1} + \sigma^2}$, $k = 1, \cdots, K$, $l = 1, \cdots, L_k$
19: Update $\tau = \tau + 1$.
20: **until** $\max_{\substack{k\in\{1,\cdots,K\}\\ l\in\{1,\cdots,L_k\}}} \left(\delta_{k,l,\tau}\right) - \min_{\substack{k\in\{1,\cdots,K\}\\ l\in\{1,\cdots,L_k\}}} \left(\delta_{k,l,\tau}\right) > \varepsilon$ or the maximum number of iterations are completed.

Total APs $A = 50$, transmit antennas per AP $N_{t,a} = 4$ for all a, the maximum power budget in both systems is 400 watts, which translates to 100 watts per BS in the cellular case. The number of users $K = 20$, $N_{r,k} = 4$ for all k, $\sigma^2 = -94$ dBm, and the solution accuracy (i.e., ε in Algorithm 1) is set to 0.01. Further details on simulation setups and channel models can be found in [3]. The maximum transmit power constraint per antenna, denoted as $P_{ant,max}$, is varied for both cellular and CF systems to observe performance variations.

The transmit power constraints for mobile-grade quality antennas depend on various factors and applications. For example, antennas in consumer mobiles and household APs have different quality measures and, consequently, different costs. In general, power constraints for mobile-grade quality antennas can span different ranges, ranging from 100 mW to 2000 mW. Similarly, for cellular wireless networks, the BS antennas' transmit power limits start from 10 W and extend to higher values. To assess the impact on performance and the performance gap between cellular and CF systems, we vary these per-antenna transmit power constraints.

All simulation results presented in this study are averaged over 50 channel realizations with 50 setups and computed using MATLAB R2022a on an AMD Ryzen R7-5800H CPU @3.20GHz with 16 GB of RAM. In this context, each setup represents a random geographic distribution of K users in the system.

B. CONVERGENCE BEHAVIOR AND COMPUTATIONAL COMPLEXITY

In Fig. 3, we set $P_{ant,max} = 2$ W and illustrate the convergence behavior of Algorithm 1 and Algorithm 2 for various values of A and K. It can be observed that Algorithm 2 provides similar performance and convergence behavior as Algorithm 1, but with significantly lower computational complexity, as will be discussed later. Additionally, we note that increasing A and K results in a higher number of iterations required for convergence. Specifically, for $[A = 16, K = 5]$, $[A = 25, K = 10]$, and $[A = 50, K = 20]$, Algorithm 1 requires 10, 14, and 200 iterations for convergence, respectively.

The convergence curves in Fig. 3 exhibit a distinctive pattern due to the two-loop structure of Algorithm 2. Notably, the curve initially approaches zero, then ascends, maintaining this pattern throughout the iterations. Each segment of the curve represents the convergence behavior of the inner loop of Algorithm 2 (i.e., the fixed-point iteration algorithm), while the overall convergence reflects the effect of the outer loop. Interestingly, the inner loop in Algorithm 2 typically converges within approximately 2 or 3 iterations across all values of A and K.

Table 1 presents the computational time required for the convergence of Algorithms 1 and 2 for various values of N_t and K. It can be observed that Algorithm 2, based on the mFPI-AO approach, is significantly less complex than Algorithm 1. For example, for $[N_t = 64, K = 10]$, Algorithm 2 requires only 0.28 seconds, whereas Algorithm 1 takes 3.39 seconds—demonstrating nearly 12 times lower computational time. We also provide a detailed analytical computational complexity analysis in Appendix A. Due to the comparable performance but much lower complexity, we utilize only Algorithm 2 in the following numerical results.

We also applied Algorithm 2 in the cellular architecture shown in Fig. 1, and present the corresponding computational time for convergence in both CF and cellular scenarios in Table 2. Interestingly, the computational cost of Algorithm 2 for CF is higher than that for the cellular architecture. This cost gap is more pronounced for smaller values of N_t and K, but diminishes as A and K increase. For example, when $[N_t = 16, K = 5]$, the cellular case requires 63.6% less computational time compared to the CF scenario, whereas the gap narrows to 5.6% when $[N_t = 200, K = 20]$.

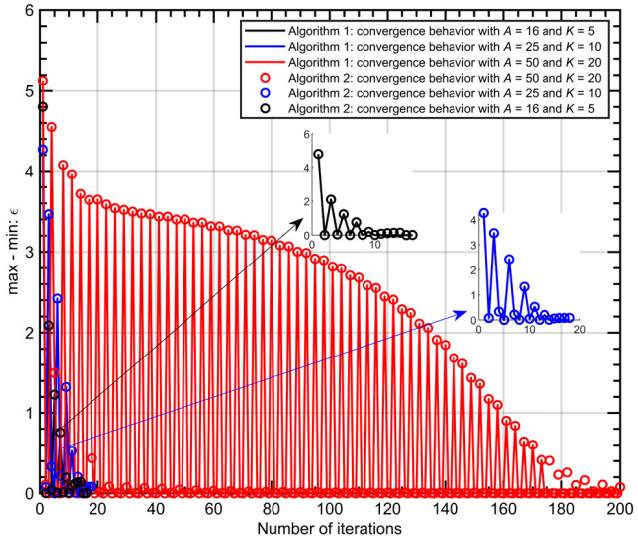

FIGURE 3. Convergence behavior of Algorithm 1 and 2 for various values of A and K.

C. CF VERSUS CELLULAR WITHOUT PER-ANTENNA CONSTRAINTS

Before discussing the impact of $p_{ant,max}$ on the performance of CF compared to cellular, we first examine the performance gains without any per-antenna constraints. In Fig. 4, we present the cumulative distribution function (CDF) of the SE of randomly located users in the system. It can be observed that the performance gap between cellular and CF, represented by the green and red solid curves, is substantial.

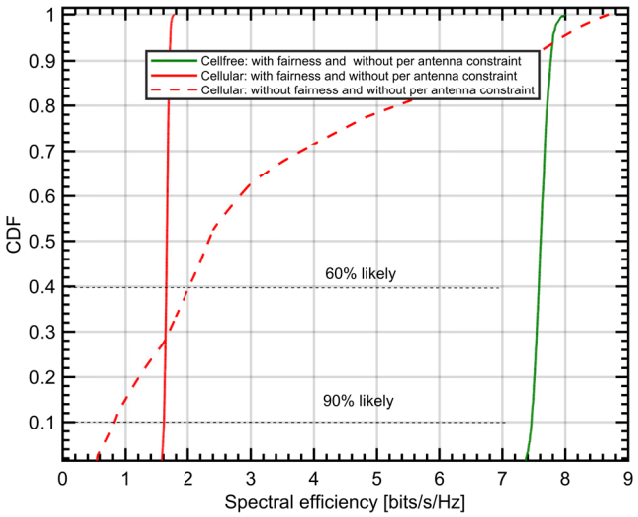

FIGURE 4. CF versus Cellular: no per-antenna constraints.

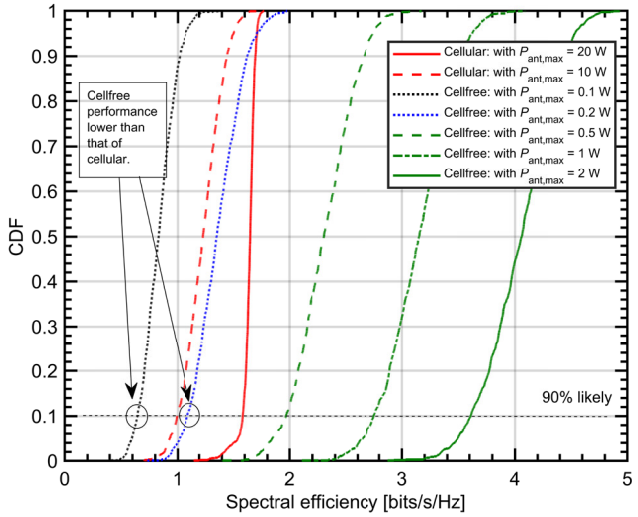

FIGURE 5. CF versus Cellular: with per-antenna constraints.

A significant body of literature suggests this type of ideal performance gap [1]. Interestingly, in Fig. 4, the performance of cellular is much worse than expected [3]. To explain the reason for this worse performance in the cellular case, we also plot the cellular performance by neglecting fairness among the users, i.e., the red dotted line. It can be observed that when fairness among users is neglected, the cellular network performance spans a large range of SE values. For instance, the red dotted curve in Fig. 4 shows that 40% of users achieve lower performance than 2 bits/s/Hz, while roughly 10% of users achieve around 7 bits/s/Hz of SE in extreme cases. Hence, to achieve fair performance among users, i.e., to improve the worst-performance users, the best-performance users are compromised, resulting in the solid red curve, where almost 100% of all users achieve close to 2 bits/s/Hz of performance.

TABLE 2. Average CPU running time for Algorithm 2 in Cell-free and Cellular systems.

	Algorithm 2: Cellular			Algorithm 2: Cell-free		
	$N_t = 64$	$N_t = 100$	$N_t = 200$	$N_t = 64$	$N_t = 100$	$N_t = 200$
$K = 5$	0.04 s	0.25 s	1.51 s	0.11 s	0.31 s	1.71 s
$K = 10$	0.11 s	0.43 s	3.12 s	0.28 s	0.62 s	3.44 s
$K = 20$	0.28 s	0.67 s	3.51 s	0.63 s	1.34 s	3.72 s

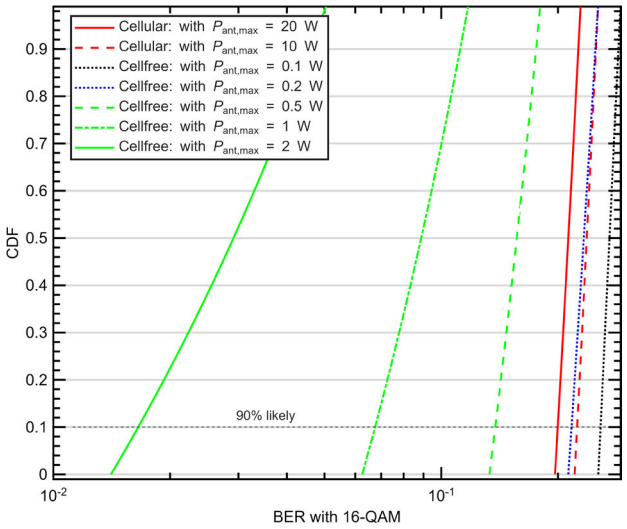

FIGURE 6. CF versus Cellular with per-antenna constraints (erroneous bits with 16-QAM)..

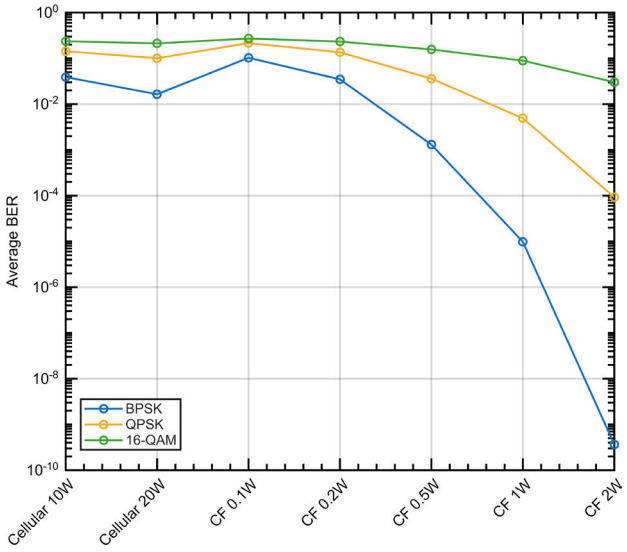

FIGURE 7. CF versus Cellular: with per-antenna constraints (average BER)..

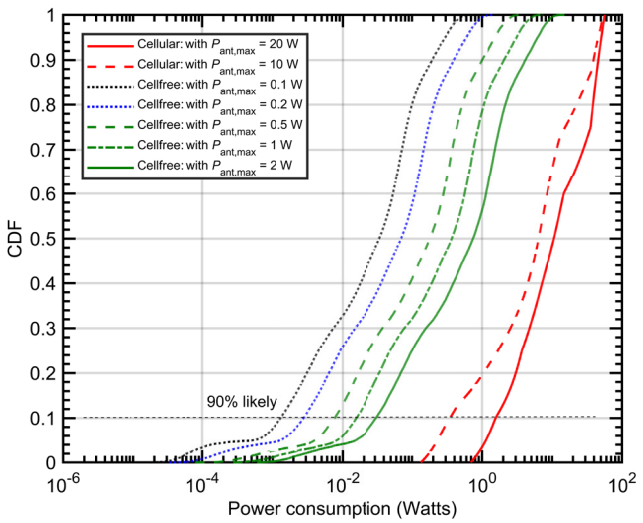

FIGURE 8. Cell-free versus Cellular: actual power radiated.

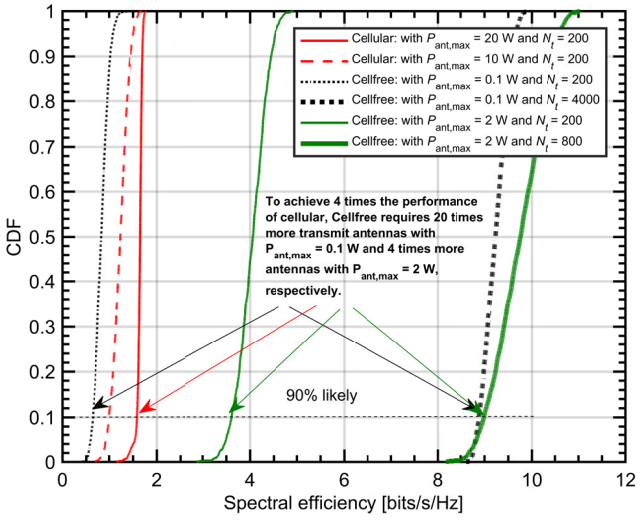

FIGURE 9. How many antennas are required to beat cellular?.

D. CF VERSUS CELLULAR WITH PER-ANTENNA CONSTRAINTS

Now we discuss the performance of cellular and CF with per-antenna power constraints. In Fig. 5, we plot the CDF of SEs of randomly located users in the geographic area for cellular and CF cases with different values of $P_{ant,max}$. Specifically, for the cellular case, we consider 10 and 20 watts of $P_{ant,max}$, shown by red dotted and solid curves in Fig. 5, respectively. Similarly, for the CF case, we consider 100 mW, 200 mW, 500 mW, 1 W, and 2 W of $P_{ant,max}$, plotted in Fig. 5 with green [500 mW, 1 W, and 2 W], blue [200 mW], and black [100 mW] colors. Remember that the total number of transmit antennas in both cases, i.e., cellular and CF, is the same, i.e., 200. In Fig. 5, it can be observed that the performance of CF is worse than cellular when transmitting antennas with $P_{ant,max}$ of 100 mW and 200 mW are used.

Similarly, the performance of CF with $P_{ant,max}$ 500 mW, 1 W, and 2 W is better than cellular but not as much as portrayed in literature in ideal cases [3]. For instance, at 90% likely SE points, CF with $P_{ant,max}$ 500 mW, 1 W, and 2 W gets 0.4, 1.1, and 2 bits/sec/Hz more SE than cellular, which is much lower than the ideal case, i.e., a 5.5 bits/s/Hz gap, as shown in Fig. 4 and portrayed in literature [3]. Corresponding to Fig. 5, we also plot the bit error rate (BER) curves considering 16-QAM modulation. More specifically, in Fig. colorblackCorresponding to Fig. 5, we also plot the bit error rate (BER) curves considering 16-QAM modulation. More specifically, in Fig. 6, we observe similar performance gaps to those discussed in Fig. 5.6, we observe similar performance gaps to those discussed in Fig. 5.

Additionally, in Fig. 7, we demonstrate the average BER for all scenarios and for different modulation schemes, i.e., BPSK, QPSK, and 16-QAM. Specifically, Fig. 7 shows the BER performance corresponding to the Shannon capacity data in Fig. 5. It can be noted from Fig. 7 that, for 16-QAM modulation, the performance is worse in all cases for both CF and cellular systems. This indicates that transmission requires strong channel coding to reach the Shannon capacity values depicted in Fig. 5. Similarly, for lower-order constellations, i.e., BPSK and QPSK, Fig. 7 shows relatively good BER values, e.g., 10^{-5} and 10^{-9} for CF 1W and CF 2W, respectively, with BPSK. Overall, the relative performance gaps between CF and cellular systems follow the same trends as discussed earlier in Fig. 5.

E. IMPACT ON EE AND SE

Notably, the performance difference between CF and cellular with per-antenna constraints is not the full story of this analysis. Interestingly, the actual power radiated per user is also different for both cellular and CF with different $P_{ant,max}$. The reason is to meet the objective function of fair performance among users, both cellular and CF systems limit their power consumption instead of using all $P_{\max}$. Hence, in Fig. 8, we plot the CDF of actual power radiated corresponding to the cases plotted in Fig. 5. It can be observed that CF radiates much less power than cellular with all $P_{ant,max}$ values. This lower radiated power is understandable in cases of 100 mW and 200 mW, where CF can never radiate all power budget with these $P_{ant,max}$ values. However, $P_{ant,max}$ of 2 W allows the use of a $P_{\max}$ budget of 400 W, but CF still limits its consumption to much less power, resulting in a much less performance (SE) gap as compared to that of ideal cases. Nevertheless, Figs. 5 and 8 depict that, regardless of the values of $P_{ant,max}$, CF is always more energy-efficient than cellular.

F. HOW MANY PRINT-QUALITY ANTENNAS DOES CF NEED TO ACHIEVE A SUFFICIENT SE GAP COMPARED TO CELLULAR?

We note that transmit antennas with $P_{ant,max}$ of 2 W are typically used in household APs in Wifi cases. We would like to emphasize that, for the commercial deployment of CF systems, these antennas are unlikely to be considered as low-cost candidates. In the commercial deployment of CF systems, it is envisioned that printable-quality transmit antennas will be fabricated, which are, of course, of much lower quality than the APs used in households for WiFi. Hence, it is more reasonable to assume that $P_{ant,max}$ of 100 mW or 200 mW might be suitable candidates to realize low-cost printable-quality APs. The question then arises: how many of these antennas do we need to surpass cellular systems and maintain a sufficient performance gap? To answer this question, we increase A from 50 to 200 and 1000 for CF cases, resulting in a total of 800 and 4000 transmit antennas (with 100 mW and 2 W of $P_{ant,max}$, i.e., both extreme cases) in CF scenarios. We compare their performance with cellular systems with $A = 50$, i.e., 200 total transmit antennas; all other parameters remain the same as in Fig. 5.

The results are plotted in Fig. 9. It can be observed that to achieve sufficiently better performance than that of the cellular system, the number of antennas required for CF is much more than for cellular systems. Specifically, to achieve a four times performance gap between CF and cellar, comparable to that of ideal cases in Fig. 4, CF requires 4000 low-quality transmit antennas compared to the 200 high-quality antennas of cellular, which is 20 times more than that of cellular systems. Similarly, with $P_{ant,max}$ of 2 W, CF requires 800 transmit antennas to achieve a similar performance gap, which is again 4 times more than cellular systems.

G. DISCUSSION

Here, we discuss the results of the proposed study in the context of distributed processing (DP) for CF systems. Specifically, DP is an alternative approach for driving CF systems, where resource allocation and optimization are performed on distributed nodes with little or no feedback from the CPU. While DP helps reduce the computational burden on the CPU, it comes with performance downsides. The performance gap between centralized and DP processing has been well studied in the literature; however, we aim to discuss the implications of the results for DP:

1) MULTI-STREAM TRANSMISSION WITH DP

Firstly, it is important to note that DP generally suffers from performance degradation due to the absence of joint processing and minimal cooperation between nodes [3]. With a high density of users and unfavorable channels, DP performance is already suboptimal. Adding multiple data streams per user exacerbates these performance challenges. For example, DP struggles with canceling inter-user interference [24], and the introduction of multiple data streams per user leads to both inter-user and intra-user interference. Although having more antennas on the user equipment side would be beneficial, A pertinent question is whether DP can support multiple streams per user. For instance, assuming a reasonable number of receive antennas at the UE, e.g., four antennas, theoretically, up to four data streams could

be supported. However, when optimizing for fairness or capacity (P1), DP may suggest transmitting a single stream, as multiple streams increase interference more than they offer benefits. Nonetheless, further analysis is needed to address these questions and develop novel methods to optimize the number of data streams per user in DP.

2) FAIRNESS DEGRADATION WITH DP

As we discussed earlier, fairness degradation occurs in centralized processing compared to cellular systems. Introducing DP into CF will further degrade fairness. To achieve a sufficient fairness gap compared to cellular systems, more antennas will be required than those discussed in the previous section.

Nevertheless, answering these questions requires careful analysis and the development of novel algorithms, making it a promising direction for future work.

3) POWER MODELING

In addition to the points discussed above, power modeling of CF mMIMO systems is another important direction for future research. Specifically, CF system requires extra hardware, such as fronthaul links and high-performance CPUs, which contribute to total power consumption. Considering these components would provide a more accurate assessment of CF energy efficiency compared to cellular networks. This topic warrants a dedicated study and could be explored in a full-length article.

4) HARDWARE AND CHANNEL IMPAIRMENTS

In addition to per-antenna power constraints, considering hardware impairments arising from mobile-grade components, such as IQ imbalance, nonlinearities, and local oscillator-related imperfections from mixers, nonlinear distortion and saturation from power amplifiers, as well as fronthaul capacity limits, jitter, and ''imperfect channel state information in (P1), can further impact the performance gap between CF and cellular systems. Note that our current analysis considers centralized processing for the CF system, while distributed processing-aided (P1) remains for future work.

VI. CONCLUSION

This paper emphasizes the importance of accounting for the impact of mobile-grade components when comparing cellular and CF wireless systems. Specifically, we focus on the effect of mobile-grade antennas, which are subject to lower power constraints, on the fairness of CF wireless systems, and we compare this performance with that of cellular systems. Our study considers a generalized CF mMIMO wireless network capable of supporting multiple data streams per user. Under per-antenna power constraints, we formulate a complex optimization problem aimed at ensuring fair performance among users. To solve this, we propose a novel AO algorithm that jointly optimizes transmit/receive digital filters, the number of data streams per user, and power allocation.

To reduce the computational complexity associated with this solution, we further propose a low-complexity mFPI-AO algorithm that exploits closed-form expressions while achieving performance comparable to that of Algorithm 1.

Our numerical analysis demonstrates that, for the same total number of transmit antennas and power budget, the fairness gap between cellular and CF systems is not as large as previously suggested in the literature. In particular, under practical $P_{ant,max}$ values, the overall SE of CF systems, despite their energy efficiency, is lower than that of cellular systems. Simulation results further reveal that, in order to match the performance gap often claimed in the literature, CF systems require a significantly higher number of mobile-grade (print-quality) transmit antennas compared to the high-quality antennas typically used in cellular systems. Specifically, with $P_{ant,max}$ values of 100 mW and 2 W, CF systems must deploy approximately 20 and 4 times more transmit antennas, respectively, than cellular systems to achieve a comparable performance gap.

REFERENCES

[1] S. Elhoushy, M. Ibrahim, and W. Hamouda, "Cell-free massive MIMO: A survey," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 492–523, 1st Quart., 2022.

[2] X. Shi, X. Shao, B. Zheng, and R. Zhang, "6DMA-aided cell-free massive MIMO communication," *IEEE Wireless Commun. Lett.*, vol. 14, no. 5, pp. 1361–1365, May 2025.

[3] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 77–90, Jan. 2020.

[4] J. Zheng, J. Zhang, H. Du, D. Niyato, B. Ai, M. Debbah, and K. B. Letaief, "Mobile cell-free massive MIMO: Challenges, solutions, and future directions," *IEEE Wireless Commun.*, vol. 31, no. 3, pp. 140–147, Jun. 2024.

[5] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, Jul. 2020.

[6] M. Munawar, M. Guenach, and I. Moerman, "Performance and architectural tradeoffs in scalable cell-free massive MIMO," *IEEE Access*, vol. 12, pp. 150189–150203, 2024.

[7] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "User-centric cell-free massive MIMO networks: A survey of opportunities, challenges and solutions," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 611–652, 1st Quart., 2022.

[8] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-free massive MIMO: A new next-generation paradigm," *IEEE Access*, vol. 7, pp. 99878–99888, 2019.

[9] H. Q. Ngo, L.-N. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 25–39, Mar. 2018.

[10] G. Interdonato, E. Björnson, H. Quoc Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, pp. 1–13, Dec. 2019.

[11] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[12] H. Yang and T. L. Marzetta, "Energy efficiency of massive MIMO: Cell-free vs. cellular," in *Proc. IEEE 87th Veh. Technol. Conf. (VTC Spring)*, Jun. 2018, pp. 1–5.

[13] H. Q. Ngo, G. Interdonato, E. G. Larsson, G. Caire, and J. G. Andrews, "Ultradense cell-free massive MIMO for 6G: Technical overview and open questions," *Proc. IEEE*, vol. 112, no. 7, pp. 805–831, Jul. 2024.

[14] G. Femenias and F. Riera-Palou, "From cells to freedom: 6G's evolutionary shift with cell-free massive MIMO," *IEEE Trans. Mobile Comput.*, vol. 24, no. 2, pp. 812–829, Feb. 2025.

[15] C. Hao, T. T. Vu, H. Q. Ngo, M. N. Dao, X. Dang, C. Wang, and M. Matthaiou, "Joint user association and power control for cell-free massive MIMO," *IEEE Internet Things J.*, vol. 11, no. 9, pp. 15823–15841, May 2024.
[16] J. Zeng, T. Wu, Y. Song, Y. Zhong, T. Lv, and S. Zhou, "Achieving energy-efficient massive URLLC over cell-free massive MIMO," *IEEE Internet Things J.*, vol. 11, no. 2, pp. 2198–2210, Jan. 2024.
[17] F. Conceição, M. Gomes, V. Silva, and R. Dinis, "Survey on resource allocation for future 6G network architectures: Cell-free and radio stripe technologies," *Electronics*, vol. 13, no. 13, p. 2489, Jun. 2024.
[18] T. C. Mai, H. Quoc Ngo, and T. Q. Duong, "Cell-free massive MIMO systems with multi-antenna users," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2018, pp. 828–832.
[19] T. C. Mai, H. Q. Ngo, and T. Q. Duong, "Downlink spectral efficiency of cell-free massive MIMO systems with multi-antenna users," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4803–4815, Aug. 2020.
[20] G. Femenias and F. Riera-Palou, "Cell-free massive MIMO with multi-antenna users: Do we really need downlink pilots?" *IEEE Trans. Veh. Technol.*, vol. 74, no. 4, pp. 5954–5969, Apr. 2025.
[21] U. K. Ganesan, T. T. Vu, and E. G. Larsson, "Cell-free massive MIMO with multi-antenna users and phase misalignments: A novel partially coherent transmission framework," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 1639–1655, 2024.
[22] Z. H. Shaik, E. Björnson, and E. G. Larsson, "Cell-free massive MIMO with radio stripes and sequential uplink processing," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–6.
[23] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signal Process.*, vol. 11, nos. 3–4, pp. 154–655, 2017.
[24] T. Demir, E. Bjrnson, and L. Sanguinetti, *Foundations of User-Centric Cell-Free Massive MIMO*. Hanover, MA, USA: Now Publishers, 2021.
[25] M. Grant and S. Boyd. (2021). *CVX: MATLAB Software for Disciplined Convex Programming*. [Online]. Available: http://cvxr.com/cvx
[26] U. T. Demir, E. BjErnson, and L. Sanguinetti, "Foundations of user-centric cell-free massive MIMO," *Found. Trends Signal Process.*, vol. 14, nos. 3–4, pp. 162–472, Jan. 2021.

MUTEEN MUNAWAR (Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering, in 2019 and 2021, respectively, and the second master's degree in electrical and information engineering from Seoul National University of Science and Technology, South Korea. He is currently pursuing the Ph.D. degree with UGent, Belgium. He was with the Communication and Signal Processing Laboratory, South Korea, from February 2023 to September 2023. He is with IMEC, Leuven, as a Researcher. His research interests include wireless communications, applied machine learning, and digital multipliers. He received the Best Project Award at the IEEE Riphah Innovation Summit (IRIS), in 2019.

MAMOUN GUENACH (Senior Member, IEEE) received the degree from EMI, Morocco, in 1995, and the Ph.D. degree from UCL, Belgium, in 2002. He was a Postdoctoral Researcher at Ghent University, from 2002 to 2006, where he has been a part-time Visiting Research Professor, since 2015. He was a member of the Nokia Bell Labs Technical Staff, from 2006 to 2019. Since 2019, he has been with IMEC, Leuven, as a Research Scientist. His research interests include coding, modulation, synchronization, and MIMO equalization for high-speed wired and wireless communication technologies.

INGRID MOERMAN (Senior Member, IEEE) received the degree in electrical engineering and the Ph.D. degree from Ghent University, in 1987 and 1992, respectively. She became a part-time Professor at Ghent University, in 2000. She is currently a Staff Member with IDLab, a core research group of IMEC, with research activities embedded at Ghent University and the University of Antwerp. She coordinates the research activities on intelligent wireless networking (iWiNe) at Ghent University, where she leads a team of more than 30 researchers. She is also the Program Manager of the deterministic networking track, part of the Connectivity Program, IMEC. In this role, she coordinates research on end-to-end wired/wireless networking solutions for business and mission-critical applications that must meet strict quality of service requirements in terms of throughput, bounded latency, and reliability in private wireless environments. She has extensive experience in running and coordinating national and EU research-funded projects.