IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 9, 2019, accepted January 21, 2019, date of publication February 1, 2019, date of current version February 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2895410

Sparse Beamforming for an Ultradensely Distributed Antenna System With Interlaced Clustering

XINJIANG XIA[1], YU ZHANG[1], JIAMIN LI[1], PENGCHENG ZHU[1], (Member, IEEE), YUANXUE XIN[2], DONGMING WANG[1], (Member, IEEE), AND XIAOHU YOU[1], (Fellow, IEEE)

[1]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
[2]College of IoT Engineering, Hohai University, Changzhou 213022, China

Corresponding authors: Jiamin Li (lijiamin@seu.edu.cn), Pengcheng Zhu (p.zhu@seu.edu.cn), and Dongming Wang (wangdm@seu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61871122, Grant 61801168, and Grant 61571120, in part by the National Key Special Program under Grant 2018ZX03001008-002, and in part by the Six Talent Peaks Project in Jiangsu Province.

ABSTRACT Recently, a novel network architecture for a distributed antenna system with interlaced clustering has been proposed to mitigate the cell-edge problem. Under this network architecture, we propose a more practical implementation for ultradensely deployed remote antenna units (RAUs) with large numbers of users. Furthermore, we focus on the user selection (USC) and sparse beamforming technologies to optimize the weighted sum rate (WSR) with both backhaul and power constraints. First, we divide each cluster pattern (CP) into several adaptive cells, where RAUs are connected to a central processor via finite-capacity backhaul links. Aiming at reducing the computational complexity for large numbers of users and RAUs, we solve the original problem with two steps. In the first stage, we propose an efficient USC algorithm to find the largest user subset that satisfies the quality of service requirement. In the second stage, we adopt a CP-based weighted sum minimum mean square error algorithm to optimize the WSR problem for the selected users in the previous stage. Moreover, two decomposition algorithms, named primal decomposition and dual decomposition, are exploited to further reduce the computational complexity. Furthermore, based on the adaptive cells, we provide a low-complexity alternating optimization method for sparse beamforming. Finally, simulation results show that the proposed algorithms can achieve a significant performance gain on edge-user rates without losing much performance gain. At the same time, the backhaul information exchange is largely reduced, and approximately 90% of the RAUs consumes less than 66.7% of backhaul for each RAU.

INDEX TERMS Distributed antenna system, sparse beamforming, interlaced clustering, CP-based weighted minimum mean square error (CP-WMMSE), user selection (USC), primal decomposition and dual decomposition.

I. INTRODUCTION

With a dramatic increase in mobile devices and the high data-rate transmission demand for video-based applications, the next-generation cellular network is expected to experience a 1000-fold growth in mobile traffic in the coming decade [1]–[4]. As one of the most important technologies in 5G cellular networks, ultradense small cell deployment has the potential to significantly decrease the transmit power consumption. The reason behind this potential improvement is that active base stations (BSs) can transmit to users with a lower power due to their reduced cell size. However, closer neighboring BSs also create stronger intercell interference, leading to a more serious cell-edge problem.

The distributed antenna system (DAS) is a promising network that provides high data-rate coverage via coordinated multipoint (CoMP) processes [5]. In a DAS, remote antenna units (RAUs) are geographically dispersed within a small cell and are connected to a central processor via high-speed backhaul links [6]–[8]. With the cooperation among RAUs,

The associate editor coordinating the review of this manuscript and approving it for publication was Chunlong He.

the DAS can serve single or multiple users in the same time-frequency resource [9], [10]. Therefore, the DAS is also called a distributed multiple-input multiple-output (MIMO) system or cooperative MIMO system. Unlike full cooperation among all RAUs in the network, which requires massive channel state information (CSI) [11]–[13] exchange and high scheduling complexity, CoMP clustering can mitigate the intracluster interference and improve cell-edge performance with relatively low overhead [14]. Optimal cluster selection or design is the key to maximizing its benefits. Generally, CoMP clustering can be categorized into static clustering, dynamic clustering and semidynamic clustering. In static clustering [15]–[18], CoMP clusters are formed statically, i.e., the RAU clusters remain fixed; hence, the performance is limited. Dynamic clustering can change pace with the network [19]–[22]. This scheme can achieve better performance than static clustering but with higher complexity of scheduling and beamforming algorithms.

Interlaced clustering with a number of cluster patterns (CPs) [23]–[25] is regarded as a compromise for the aforementioned two schemes. The interlaced clustering can be considered as a general architecture, and some other network architectures such as centralized tiered networks and fractional frequency reuse can be interpreted as instances of interlaced clustering. In interlaced clustering, several different CPs coexist on orthogonal frequency bands, and each user can be served by either of them. Furthermore, this kind of semidynamic clustering significantly reduces the intercluster interference for distributed transmissions. To achieve higher performance gain, the network architecture requires extensive backhaul connectivity.

Recently, the backhaul requirement has increased significantly with the evolution of cellular networks [26], and various backhaul technologies have been introduced in [27]–[29]. In addition, the problem of minimizing the data transfer in the backhaul with quality of service (QoS) constraints and per-BS power constraints was studied in [30]. Furthermore, a user-centric clustering scheme with finite-capacity backhaul was proposed by optimizing a sparse beamforming vector for each user in the cloud radio access (C-RAN) network [22]. In practice, sparse beamforming vector design is implemented by the l_0-norm optimization problem, in which the nonzero elements in the beamforming vector correspond to the user's service cluster. As a result, the beamforming vector can help to determine the best set of serving RAUs for each user.

The number of optimized variables will increase as the number of CPs increases. Fortunately, the decomposition method can be used to transform the original problem into independent per-CP problems [31], [32]. Decomposition in optimization has appeared in early works on large-scale linear programs (LPs) since the 1960s [33]. Primal decomposition and dual decomposition are two classical methods to separate a complex problem into several subproblems, which can be solved in parallel [31]. Primal decomposition can achieve the upper bound without great difficulty. However, primal decomposition is not suitable for more than two CPs in this paper. Instead, dual decomposition is a good alternative to solve the original problem although its initialization should be further considered.

A DAS with interlaced clustering is a promising network architecture for the ultradense users scenario, while the complexity of sparse beamforming for large numbers of users also increases. User selection (USC) is a critical method to ease the computational burden of sparse beamforming. For example, an iterative user pool shrinking algorithm [22] can significantly reduce the total number of variables to be optimized. Furthermore, this algorithm requires running the optimization problem at least once. Pan *et al.* [34] proposed a novel USC algorithm to remove the users that could not satisfy the power constraints and rate requirements. Furthermore, for ultradense networks serving large numbers of users, it is challenging to design a linear precoding scheme that can ensure as many users as possible be served in a time-frequency (T-F) slot.

In this paper, we consider the ultradense DAS with ultradense users under an interlaced clustering network architecture. We first decouple the problem of the transmission methods design into user selection and beamforming design, and propose a two-stage optimize problem. In the first stage, we propose a low complexity USC algorithm as a solution to the high-dimensional optimization problem for systems with ultradense users. In the second stage, with the users selected by first stage, we propose a CP-based weighted sum MSE minimization (CP-WMMSE) algorithm, to solve the joint sparse beamforming design and power allocation problem for DAS with interlaced clustering from a network utility maximization perspective. Moreover, two decomposition algorithms, named primal decomposition and dual decomposition, are exploited to further reduce the computational complexity. Furthermore, based on the adaptive cells, we provide a low-complexity alternating optimization method for sparse beamforming. To the best of our knowledge, this paper is the first attempt to design sparse beamforming vectors for DAS with interlaced clustering with a multicell, multi-antenna, multiuser cellular network. The main contributions of this paper are summarized as follows:

- We first introduce sparse beamforming for DAS with interlaced clustering to reduce backhaul consumption and increase spatial multiplexing. Thus, by applying sparse beamforming, the users in interlaced clustering can achieve higher coverage compared with the strategy in which all corresponding RAUs serve a single user in one adaptive cell.
- To reduce the complexity of the computations of sparse beamforming with interlaced clustering, we introduce an efficient USC algorithm with alternating optimization as the first stage to select the users for the next stage. Its complexity is much lower than the traditional USC method, especially for dense users and RAUs. Simulation results show that our algorithm achieves similar performance as the traditional algorithm.

- We propose an adaptive cell-based scheme to ease the heavy computational burden, which requires significantly fewer candidate RAUs and users. The proposed scheme, together with alternating optimization, can be generally applied in interlaced clustering or similar networks. In this paper, we apply this strategy in the first stage to further reduce the computations of the USC algorithm and the popular WSR problem. The results show that it can converge quickly without losing obvious gains.
- We solve the popular WSR problem with power allocation in interlaced clustering. To adapt to the architecture of interlaced clustering, the CP-WMMSE algorithm is proposed, where we consider both the sparse beamforming design and power allocation among different CPs. Furthermore, decomposition methods are proposed to address the heavy computational burden problem. Primal decomposition can achieve the upper bound without assistant technology but cannot be easily extended to more than two CPs. Dual decomposition can solve the problem with the proper initiation values. Finally, alternating optimization associated with the primal decomposition algorithm, dual decomposition algorithm and CP-WMMSE algorithm is considered in this paper to further reduce the computational complexity of the WSR problem in interlaced clustering. Simulation results show that the proposed algorithms can achieve similar performance.

The remainder of this paper is organized as follows. Section II presents the channel model, problem formulation and introduces sparse beamforming for DAS with the interlaced clustering architecture. Then, we propose the USC algorithm for interlaced clustering in Section III. Based on the WMMSE algorithm, Section IV describes the proposed five iterative algorithms for sparse beamforming under the CP network architecture. The performance of the proposed algorithms is evaluated by simulations in Section V. Finally, Section VI summarizes the paper.

Notation: Throughout this paper, scalars are represented by lowercase letters (e.g., i); matrices are represented by uppercase bold letters (e.g., $\mathbf{H}$); vectors are represented by lowercase bold letters (e.g., $\mathbf{v}$). The matrix inverse, conjugate transpose and l_p-norm of a vector are denoted as $(.)^{-1}$, $(.)^H$ and $|.|_p$, respectively. We use $\mathbb{C}^{M\times N}$ to denote the set of complex $M\times N$ matrices. The complex Gaussian distribution is represented by $\mathcal{CN}(0,\sigma^2)$. Calligraphy letters are used to denote sets.

II. SYSTEM MODEL

A. DAS WITH INTERLACED CLUSTERING ARCHITECTURE

In the DAS, RAUs are grouped into disjoint cooperation clusters. Each snapshot of a classification is called a CP. The idea behind interlaced clustering is to assign different CPs to orthogonal frequency bands. For simplicity, we choose two CPs. As an illustration, Fig.1 shows two such possible CPs

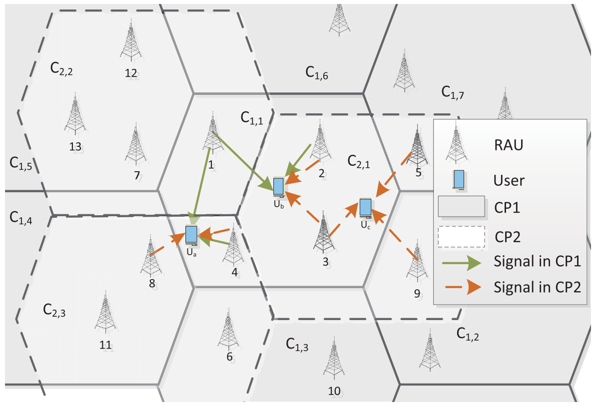


FIGURE 1. An example layout of sparse beamforming for DAS with interlaced clustering.

(CP_1 and CP_2). The total system bandwidth is $B = B_1 + B_2$, where B_1 and B_2 represent the orthogonal frequency bands of CP_1 and CP_2, respectively.

In [23], when considering the linear beamforming design scheme, the analysis is limited to one scheduled user per T-F slot in a cluster to reduce the complexity of the analysis. Therefore, not all users can be served simultaneously on the same T-F resource. For example, in Fig 1, the possible combination scheme is $\{U_a, U_b\}$, $\{U_b, U_c\}$ and $\{U_a, U_c\}$. Hence, the service coverage is 66.7%. However, with sparse beamforming, there are two pairs of users in CP1, the service RAUs for U_a are RAU_1 and RAU_4, the service RAUs for U_b are RAU_1 and RAU_2; there are three pairs of users in CP2, the service RAUs for U_a are RAU_4 and RAU_8, the service RAUs for U_b are RAU_2 and RAU_3, the service RAUs for U_c are RAU_3, RAU_5 and RAU_9. Therefore, the RAUs can serve the three users at the same time, and the service coverage is 100%. The reason is because, in our scheme, we can design the linear sparse beamforming with a backhaul constraint, which does not have complications due to the low complexity of the proposed algorithms, and all the users can be served at one time.

Furthermore, different from [22], we define the cell in each CP as one adaptive cell and apply sparse beamforming as a solution to reduce the backhaul consumption in interlaced clustering, which regards all RAUs and users within the network as the initial sample. We alternate between designing the sparse beamforming for each adaptive cell by fixing the other adaptive cells. Assume that in each adaptive cell, there is a backhaul network connecting the RAUs within the adaptive cell to a central processor. The central processor allocates the CSI or the users' data for all RAUs within the selected adaptive cell. Furthermore, the central processor has access to the global CSI and the user data in its adaptive cell. With such data sharing strategies, user data are restricted to the adaptive cell, and thus, both the complexity and backhaul capacity consumption are effectively reduced.

Fig.1 also shows the affiliation between the adaptive cells and the CPs. For example, $\{RAU_1, RAU_2, RAU_3, RAU_4\} \in C_{1,1}$, $\{RAU_2, RAU_3, RAU_5, RAU_9\} \in C_{2,1}$, where $C_{1,1}$ and

$C_{2,1}$ represent the 1st adaptive cell in CP_1, and the 1st adaptive cell in CP_2, respectively. Note that, in different CPs, the distribution of adaptive cells may not be the same.

B. SYSTEM MODEL

Consider a downlink DAS system that contains $\mathcal{F} = \{f_1, \cdots, f_q, \cdots, f_Q\}$ different CPs operating on disjoint frequency bands $\mathcal{B} = \{B_{f_1}, \cdots, B_{f_q} \cdots, B_{f_Q}\}$, where B_{f_q} denotes the frequency band for CP f_q. For simplicity of notations, we assume that each RAU is equipped with M transmit antennas, and that each user has N receive antennas. Throughout the paper, we use k, m, r to indicate the user index, i, j for the cell index, and d, l, b for the RAU index.

Define $\mathcal{A} = \{\mathcal{A}_{f_1}, \mathcal{A}_{f_2}, \cdots, \mathcal{A}_{f_Q}\}$ as the collection of all adaptive cells, where $\mathcal{A}_{f_q}$ represents the collection of the cells within CP f_q, and the relationship between different CPs is defined as follows:

$$\begin{cases} \mathcal{A}_{f_1} \cup \mathcal{A}_{f_2} \cup \cdots \cup \mathcal{A}_{f_Q} = \mathcal{A}, \\ \mathcal{A}_{f_1} \cap \mathcal{A}_{f_2} \cap \cdots \cap \mathcal{A}_{f_Q} = \emptyset. \end{cases} \tag{1}$$

where $\emptyset$ denotes a null set.

Let $\mathcal{A}_{i,f_q}$ denote the i-th adaptive cell in $\mathcal{A}_{f_q}$. Then, we define $\mathcal{D}_{i,f_q}$ as the distributed RAUs located in the adaptive cell $\mathcal{A}_{i,f_q}$. Likewise, let $\tilde{\mathcal{K}}_{f_q}, \tilde{\mathcal{K}}_{i,f_q} \subset \tilde{\mathcal{K}}_{f_q}$ be the subset of users assigned to a CP f_q and the adaptive cell $\mathcal{A}_{i,f_q}$, respectively. Let $\mathcal{K}_{i,f_q} \subset \tilde{\mathcal{K}}_{i,f_q}$ be the set of users that should be admitted to the adaptive cell $\mathcal{A}_{i,f_q}$, i.e., $\mathcal{K}_{i,f_q}$ represents the set of users after USC, which we will introduce in Part A of Section III.

Let $\mathbf{G}_{k,l,j,f_q}, \forall k \in \mathcal{K}_{i,f_q}, \forall l \in \mathcal{D}_{i,f_q}, \forall i, j \in \mathcal{A}_{f_q}$ denote the CSI matrix between the lth RAU in adaptive cell j and the user k that in adaptive cell i. Let

$$\mathbf{H}_{k,j,f_q} \triangleq \left[\mathbf{G}_{k,1,j,f_q}, \ldots, \mathbf{G}_{k,\left|\mathcal{D}_{j,f_q}\right|,j,f_q}\right] \in \mathbb{C}^{N \times M\left|\mathcal{D}_{j,f_q}\right|}$$

denote the CSI matrix between all the RAUs in the jth adaptive cell to the user k.

Instead of a network-wide beamforming vector [22], [34], we introduce an adaptive cell-wide beamforming vector

$$\mathbf{v}_{k,i,f_q} \triangleq \left[\mathbf{v}^H_{k,1,i,f_q}, \ldots, \mathbf{v}^H_{k,\left|\mathcal{D}_{i,f_q}\right|,i,f_q}\right]^H \in \mathbb{C}^{M\left|\mathcal{D}_{i,f_q}\right| \times 1}$$

for each user k in adaptive cell i, where $\mathbf{v}_{k,d,i,f_q} \in \mathbb{C}^{M \times 1}, \forall d \in \mathcal{D}_{i,f_q}$ denotes that the dth RAU in adaptive cell i is used to transmit a single stream of data signal $s_{k,i,f_q} \in \mathbb{C}$ to the kth user, where s_{k,i,f_q} is independent and identically distributed according to $\mathcal{CN}(0, 1)$.

The received signal $\mathbf{y}_{k,i,f_q} \in \mathbb{C}^{N \times 1}$ of the user k is

$$\begin{aligned} \mathbf{y}_{k,i,f_q} = & \mathbf{H}_{k,i,f_q}\mathbf{v}_{k,i,f_q}s_{k,i,f_q} \\ & + \underbrace{\sum_{m \neq k, m \in \mathcal{K}_{i,f_q}} \mathbf{H}_{k,i,f_q}\mathbf{v}_{m,i,f_q}s_{m,i,f_q}}_{\text{intracell interference}} \\ & + \underbrace{\sum_{j \neq i, j \in \mathcal{A}_{f_q}} \sum_{r \in \mathcal{K}_{j,f_q}} \mathbf{H}_{k,j,f_q}\mathbf{v}_{r,j,f_q}s_{r,j,f_q}}_{\text{intercell interference}} + \mathbf{n}_{k,i,f_q} \end{aligned} \tag{2}$$

where $\mathbf{n}_{k,i,f_q} \in \mathbb{C}^{N \times 1}$ is the additive white Gaussian noise with its distribution $\mathcal{CN}\left(\mathbf{0}, \sigma^2_{k,i,f_q}\mathbf{I}_N\right)$.

C. PROBLEM FORMULATION

With an increasing number of CPs, the resource allocation should be further considered. Assume that the bandwidth budget of CP f_q is $B_{f_q} = \mu_{f_q}B_{\text{sum}}$, where B_{sum} and μ_{f_q} represent the sum bandwidth budget, the ratio between B_{f_q} and B_{sum}, respectively. Similarly, the power and the backhaul budget of dth RAU within CP f_q are given by

$$P_{d,i,f_q} = \eta_{d,i,f_q}P_{\text{sum},d}, \tag{3a}$$

$$C_{d,i,f_q} = \beta_{d,i,f_q}C_{\text{sum},d} \tag{3b}$$

where $P_{\text{sum},d}$, $C_{\text{sum},d}$, η_{d,i,f_q}, β_{d,i,f_q} are defined as the sum power budget and sum backhaul budget of the dth RAU, the ratio between $P_{\text{sum},d}$ and P_{d,i,f_q}, the ratio between $C_{\text{sum},d}$ and C_{d,i,f_q}, respectively. The constraints of the resource budget are given by

$$\text{C1}: \sum_{f \in \mathcal{F}} \eta_{d,i,f_q} = 1, \ \sum_{f \in \mathcal{F}} \beta_{d,i,f_q} = 1, \ \sum_{f \in \mathcal{F}} \mu_{f_q} = 1, \ \forall d \in \mathcal{D}_{i,f_q}. \tag{4}$$

Due to the orthogonal frequency allocations among the CPs, different CPs could be considered independent of one another. Thus, for the sake of simplicity, we set the variables β_{d,i,f_q}, μ_{f_q} equal to $\frac{1}{N_{cp}}$

$$\beta_{d,i,f_q} = \mu_{f_q} = \frac{1}{N_{cp}} \tag{5}$$

where N_{cp} is expressed as the number of CPs.

The user rate (bit/s/Hz) of the user k in CP f_q is

$$R_{k,i,f_q} = \log\left(1 + \mathbf{v}^H_{k,i,f_q}\mathbf{H}^H_{k,i,f_q}\mathbf{Z}^{-1}_{k,i,f_q}\mathbf{H}_{k,i,f_q}\mathbf{v}_{k,i,f_q}\right), \tag{6}$$

where

$$\mathbf{Z}_{k,i,f_q} = \sum_{(m,j) \neq (k,i)} \mathbf{H}_{k,j,f_q}\mathbf{v}_{m,j,f_q}\mathbf{v}^H_{m,j,f_q}\mathbf{H}^H_{k,j,f_q} + \sigma^2_{k,i,f_q}\mathbf{I} \tag{7}$$

is the interference-plus-noise covariance matrix. In interlaced clustering, each user can locate in the interior of at least one adaptive cell belonging to one of the CPs with a high probability. Thus, the users with a low rate should be served by the other CPs that can provide high signal-to-interference-plus-noise ratio (SINR) for the users. Therefore, the system has the ability to provide high QoS for users, and each user's data rate should be higher than the minimum requirement:

$$\text{C2}: R_{k,i,f_q} \geq R_{\min,k,i,f_q}. \tag{8}$$

With densely deployed RAUs, the power and backhaul consumption on the RAUs may be significant. Therefore, it is a realistic problem to consider the power and the backhaul constraints of RAUs when applying sparse beamforming in interlaced clustering. In addition, the backhaul constraint of the RAUs is the key step for designing sparse beamforming.

The per-RAU power constraint and the per-RAU backhaul constraint are given by

$$\begin{aligned} \text{C3}: &\sum_{k\in\mathcal{K}_{i,f_q}} \left\|\mathbf{v}_{k,d,i,f_q}\right\|_2^2 \le \eta_{d,i,f_q} P_{\text{sum},d}, \forall d \in \mathcal{D}_{i,f_q} \\ \text{C4}: &\sum_{k\in\mathcal{K}_{i,f_q}} \mathbb{1}\left\{\left\|\mathbf{v}_{k,d,i,f_q}\right\|_2^2\right\} R_{k,i,f_q} \le \frac{1}{N_{cp}} C_{\text{sum},d} \forall d \in \mathcal{D}_{i,f_q} \end{aligned} \tag{9}$$

where $\mathbb{1}\left\{\left\|\mathbf{v}_{k,d,i,f_q}\right\|_2^2\right\}$ represents the indicator function with the facility of scheduling choice:

$$\mathbb{1}\left\{\left\|\mathbf{v}_{k,d,i,f_q}\right\|_2^2\right\} = \begin{cases} 0, & \text{if } \left\|\mathbf{v}_{k,d,i,f_q}\right\|_2^2 = 0 \\ 1, & \text{otherwise.} \end{cases} \tag{10}$$

With the minimum rate constraint, the system can allocate the users for each CP in a T−F slot, which will greatly reduce the complexity of the proposed algorithms. Hence, we first formulate a user selection problem to maximize the number of admitted users supported by each CP. The USC can be formulated as

$$\begin{aligned} \max_{\mathbf{v}_{k,d,i,f_q}, \mathcal{K}_{f_q} \subset \tilde{\mathcal{K}}_{f_q}} &\sum_{f_q\in\mathcal{F}} \mathcal{K}_{f_q}, \forall d \in \mathcal{D}_{i,f_q} \\ \text{s.t. } &\text{C2, C3, C4.} \end{aligned} \tag{11}$$

Next, we fix the users selected from the user selection algorithm and apply sparse beamforming in interlaced clustering to solve the weighted sum-rate (WSR) maximization problem

$$\begin{aligned} \max_{\substack{\{\mathbf{v}_{k,d,i,f_q} | k\in\mathcal{K}_{i,f_q} \\ ,d\in\mathcal{D}_{i,f_q}\}}} &\sum_{f_q\in\mathcal{F}} \sum_{i\in\mathcal{A}_{f_q}} \sum_{k\in\mathcal{K}_{i,f_q}} \Psi_{k,i,f_q} R_{k,i,f_q} \\ \text{s.t. } &\text{C3, C4,} \end{aligned} \tag{12}$$

where Ψ_{k,i,f_q} represents the priority of the user k.

The the nonconvex constraint C4 in both problem (11) and (12) motivates us to first convert the nonconvex constraint C4 to a convex constraint. Motivated by compressive sensing technology and its applications in [22], [35], and [36], we can rewrite the constraint C4 of problem (10) as

$$\begin{aligned} \text{C5}: \sum_{k\in\mathcal{K}_{i,f_q}} \Upsilon_{k,d,i,f_q} R_{k,i,f_q}^{(n)} \left\|\mathbf{v}_{k,d,i,f_q}\right\|_2^2 \le \frac{1}{N_{cp}} C_{\text{sum},d}, \\ \forall d \in \mathcal{D}_{i,f_q}, \end{aligned} \tag{13}$$

where $R_{k,i,f_q}^{(n)}$ is the fixed rate obtained from the previous iteration and Υ_{k,d,i,f_q} is a constant weight associated with the dth RAU and the kth user, which is updated iteratively according to

$$\Upsilon_{k,d,i,f_q} = \frac{1}{\left\|\mathbf{v}_{k,d,i,f_q}^{(n)}\right\|_2^2 + \varsigma} \tag{14}$$

with regularization factor $\varsigma \ge 0$ and $\left\|\mathbf{v}_{k,d,i,f_q}^{(n)}\right\|_2^2$ from the previous iteration, and their values are small.

However, although the constraint C4 is converted to a convex constraint, the problems (11) and (12) are both mixed-integer nonlinear programming (MINLP) problems, which are still difficult to solve. Furthermore, the adaptive cell-wide beamforming vector introduced in this paper makes the problems more complex. In the next section, we will try to solve the problems (11) and (12) in turn.

III. THE FIRST STAGE: USER SELECTION ALGORITHM WITH BACKHAUL CONSTRAINTS

A. USC ALGORITHM FOR CP SCHEME

By introducing a series of auxiliary variables $\{\omega_{k,i,f_q}\}$, we approximate the problem (11) as

$$\begin{aligned} \min_{\{\omega_{k,i,f_q}\}_{k\in\tilde{\mathcal{K}}_{f_q}}, \mathbf{v}_{k,i,f_q}} &\sum_{f_q\in\mathcal{F}} \sum_{i\in\mathcal{A}_{f_q}} \sum_{k\in\tilde{\mathcal{K}}_{i,f_q}} (\omega_{k,i,f_q} - 1)^2 \\ \text{s.t. } &\text{C3, C5,} \\ &R_{k,i,f_q} \ge \omega_{k,i,f_q}^2 R_{\min,k,i,f_q}. \end{aligned} \tag{15}$$

To reduce the complexity of the network-wide beamforming design for large numbers of RAUs and users, we introduce an alternating optimization method with a user selection algorithm by using an adaptive cell-wide beamforming vector as follows. First, we initialize the adaptive cell-wide beamforming vector $\mathbf{v}_{k,i,f_q}, \forall i \in \mathcal{A}_{f_q}$, and choose an adaptive cell as the optimized target. At the same time, we consider signals from the other adaptive cells as constant interference by fixing the beamforming vector $\mathbf{v}_{r,j,f_q} (j \ne i)$. Second, we design the beamforming of the fixed adaptive cell and iteratively solve the optimization problems (15) by fixing the adaptive cell in turn until convergence is obtained.

Furthermore, the factor of power allocation is not easy to obtain, which we optimize in the second stage. Thus, in the first stage, for the sake of simplicity, we relax the power constraints by setting $\eta_{d,i,f_q} = 1$ for all CPs. Then, the number of users admitted to the network will be larger than the practice value. Increasing the number of users who were admitted in the first stage will not affect the final result in the second stage, and this conclusion is verified through the simulation results.

With the above description, we rewrite the optimization problems (15) as a series of subproblems:

$$\min_{\substack{\{\omega_{k,i,f_q}\}_{k\in\tilde{\mathcal{K}}_{i,f_q}}, \\ \mathbf{v}_{k,i,f_q}}} \sum_{k\in\tilde{\mathcal{K}}_{i,f_q}} (\omega_{k,i,f_q} - 1)^2 \tag{16a}$$

$$\text{s.t. C5, } \sum_{k\in\tilde{\mathcal{K}}_{i,f_q}} \left\|\mathbf{v}_{k,d,i,f_q}\right\|_2^2 \le P_{\text{sum},d}, \forall d \in \mathcal{D}_{i,f_q}, \tag{16b}$$

$$R_{k,i,f_q} \ge \omega_{k,i,f_q}^2 R_{\min,k,i,f_q}. \tag{16c}$$

It is obvious that (16a) and (16b) are already in convex form, so we only need to deal with the constraint (16c). To proceed, we introduce the following lemma:

Lemma 1: Given expression

$$g(\boldsymbol{v}) = \boldsymbol{b}^H(\boldsymbol{v}) \boldsymbol{A}^{-1}(\boldsymbol{v}) \boldsymbol{b}(\boldsymbol{v}) \tag{17}$$

where $\boldsymbol{b}^H(\boldsymbol{v})$ and $\boldsymbol{A}(\boldsymbol{v})$ are arbitrary given functions: $\mathbb{C}^{r_1} \to \mathbb{C}^{r_2}$ and $\mathbb{C}^{r_1} \to \mathbb{S}_{++}^{r_2\times r_2}$, respectively, and $\boldsymbol{v} \in \boldsymbol{\mathcal{V}}$ where $\mathcal{V} \subseteq \mathbb{C}^{r_1}$, $r_1, r_2 \in \mathbb{N}$.

We can obtain the following inequivalent

$$\boldsymbol{b}^H(\boldsymbol{v})\boldsymbol{A}^{-1}(\boldsymbol{v})\boldsymbol{b}(\boldsymbol{v}) \geq G(\boldsymbol{v}) \tag{18}$$

where $G(\boldsymbol{v}) = 2\Re\left\{\boldsymbol{\chi}^H\boldsymbol{b}(\boldsymbol{v})\right\} - \boldsymbol{\chi}^H\boldsymbol{A}(\boldsymbol{v})\boldsymbol{\chi}$ and $\boldsymbol{\chi} \in \mathbb{C}^{r_2}$ refers to an auxiliary variable. The proof of Lemma *1* is relegated to Appendix B.

According to Lemma *1*, we can rewrite constraint (16c) as

$$1 + G\left(\boldsymbol{v}_{k,d,i,f_q}\right) \geq 2^{\omega_{k,i,f_q}^2 R_{\min,k,i,f_q}}, \tag{19}$$

where

$$G\left(\boldsymbol{v}_{k,d,i,f_q}\right) = 2\Re\left\{\boldsymbol{\chi}_{k,i,f_q}^H\boldsymbol{H}_{k,i,f_q}\boldsymbol{v}_{k,i,f_q}\right\} - \boldsymbol{\chi}_{k,i,f_q}^H\boldsymbol{Z}_{k,i,f_q}\boldsymbol{\chi}_{k,i,f_q}.$$

It is obvious that, according to Lemma *1*, by transforming the problem (17) into a multidimensional quadratic transform, the SINR term in (16c) is converted to a concave function of $\mathbf{v}_{k,d,i,f_q}$. Then, (19) can be further equivalently transformed into a more tractable form as

$$\sum_{m\neq k}\boldsymbol{\chi}_{k,i,f_q}^H\boldsymbol{H}_{k,i,f_q}\boldsymbol{v}_{m,i,f_q}\boldsymbol{v}_{m,i,f_q}^H\boldsymbol{H}_{k,i,f_q}^H\boldsymbol{\chi}_{k,i,f_q} \leq \phi_{k,i,f_q} \tag{20}$$

where

$$\begin{aligned}\phi_{k,i,f_q} = {} & 1 + 2\Re\{\boldsymbol{\chi}_{k,i,f_q}^H\boldsymbol{H}_{k,i,f_q}\boldsymbol{v}_{k,i,f_q}\} \\ & - 2^{\omega_{k,i,f_q}^2 R_{\min,k}} - \sigma_{k,i,f_q}^2\boldsymbol{\chi}_{k,i,f_q}^H\boldsymbol{I}\boldsymbol{y}_{k,i,f_q} \\ & - \sum_{j\neq i}\boldsymbol{\chi}_{k,j,f_q}^H\boldsymbol{H}_{k,j,f_q}\boldsymbol{v}_{r,j,f_q}\boldsymbol{v}_{r,j,f_q}^H\boldsymbol{H}_{k,j,f_q}^H\boldsymbol{\chi}_{k,j,f_q}.\end{aligned} \tag{21}$$

Then, we iteratively optimize problem (16) over $\mathbf{v}_{k,d,i,f_q}$ and $\boldsymbol{\chi}_{k,i,f_q}$. First, we set optimal $\tilde{\boldsymbol{\chi}}_{k,i,f_q}$ to its optimal value for given $\mathbf{v}_{k,d,i,f_q}$ as (see Lemma *1*):

$$\tilde{\boldsymbol{\chi}}_{k,i,f_q} = \boldsymbol{Z}_{k,i,f_q}^{-1}\mathbf{H}_{k,i,f_q}\mathbf{v}_{k,i,f_q} \tag{22}$$

Second, for a given $\boldsymbol{\chi}_{k,i,f_q}$, the problem (16) can be rewritten as a second-order cone programming (SOCP) problem:

$$\begin{aligned}\min_{\substack{\{\omega_{k,f_q}\}_{k\in\tilde{\mathcal{K}}_{i,f_q}},\\ \mathbf{v}_{k,i,f_q}}} \quad & \sum_{k\in\tilde{\mathcal{K}}_{i,f_q}}(\omega_{k,i,f_q}-1)^2 \\ \text{s.t.} \quad & \left\|\boldsymbol{\theta}_{k,i,f_q}\right\|_2 \leq \sqrt{\phi_{k,i,f_q}}, \\ & \left\|\boldsymbol{\varphi}_{d,i,f_q}\right\|_2 \leq \sqrt{P_{\text{sum},d}}, \\ & \left\|\boldsymbol{\rho}_{d,i,f_q}\right\|_2 \leq \sqrt{\frac{1}{N_{cp}}C_{\text{sum},d}},\end{aligned} \tag{23}$$

where $\boldsymbol{\theta}_{k,i,f_q}$, $\boldsymbol{\varphi}_{d,i,f_q}$, $\boldsymbol{\rho}_{d,i,f_q}$ are given by

$$\boldsymbol{\theta}_{k,i,f_q} = \left[\boldsymbol{\chi}_{k,i,f_q}^H\boldsymbol{H}_{k,i,f_q}\boldsymbol{v}_{1,i,f_q},\ldots,\boldsymbol{\chi}_{k,i,f_q}^H\boldsymbol{H}_{k,i,f_q}\boldsymbol{v}_{\left|\mathcal{K}_{i,f_q}\right|,i,f_q}\right]^H, \tag{24}$$

$$\boldsymbol{\varphi}_{d,i,f_q} = \left[\mathbf{v}_{1,d,i,f_q}^H,\ldots,\mathbf{v}_{\mathcal{K}_{\left|\mathcal{K}_{d,i,f_q}\right|,d,i,f_q},d,i,f_q}^H\right]^H \tag{25}$$

$$\begin{aligned}\boldsymbol{\rho}_{d,i,f_q} = {} & \left[\sqrt{\Upsilon_{1,d,i,f_q}R_{1,i,f_q}^{(n)}}\mathbf{v}_{1,d,i,f_q}^H,\ldots,\right. \\ & \left.\sqrt{\Upsilon_{\left|\mathcal{K}_{d,i,f_q}\right|,d,i,f_q}R_{\left|\mathcal{K}_{d,i,f_q}\right|,i,f_q}^{(n)}}\mathbf{v}_{\left|\mathcal{K}_{d,i,f_q}\right|,d,i,f_q}^H\right]^H,\end{aligned} \tag{26}$$

where $\mathcal{K}_{d,i,f_q}$ is the candidate set of users served by the dth RAU. Thus, $\left|\mathcal{K}_{d,i,f_q}\right|$ and $\left|\mathcal{K}_{i,f_q}\right|$ represent the number of users served by the dth RAU and the total number of users admitted in adaptive cell i of CP f_q, respectively, and $\mathcal{K}_{i,f_q} = \{\mathcal{K}_{1,i,f_q},\cdots,\mathcal{K}_{\left|\mathcal{K}_{d,i,f_q}\right|,i,f_q}\}$.

To obtain $\mathcal{K}_{i,f_q}$, we initialize $\mathcal{K}_{i,f_q} = \tilde{\mathcal{K}}_{i,f_q}$. Then, the problem (23) has a global optimal solution by traditional techniques such as the interior point method [37], [38].

After solving problem (23), we obtain $\mathcal{K}_{i,f_q}$ according to

$$\mathcal{K}_{i,f_q} = \begin{cases}\tilde{\mathcal{K}}_{i,f_q}, & \text{if } \omega_{k,i,f_q} = 1, \forall k \in \mathcal{K}_{i,f_q} \\ \tilde{\mathcal{K}}_{i,f_q}\backslash k_{k,i,f_q}^*, & \text{otherwise.}\end{cases} \tag{27}$$

where $k_{k,i,f_q}^* = \arg\min_{k\in\tilde{\mathcal{K}}_{k,i,f_q}}\omega_{k,i,f_q}$.

Based on the above description and analysis, the iterative algorithm for problem (15) is summarized in Algorithm 1.

Algorithm 1 User Selection for Interlaced Clustering

Input:
$\mathbf{v}_{k,i,f_q}^{(0)}$, $R_{k,i,f_q}^{(0)}$, $\mathcal{K}_{i,f_q}^{(0)} = \tilde{\mathcal{K}}_{i,f_q}$, $\forall k \in \mathcal{K}_{i,f_q}^{(0)}$, $\forall d \in \mathcal{D}_{i,f_q}$, $\forall i$.
fix f_q, $\mathbf{v}_{r,j,f_q}^{(0)}$, $\mathcal{K}_{j,f_q}^{(0)}$, $\forall j$
Set $n \to 0$
Set adaptive cell i as the target optimal cell.

1: **repeat**
2:   **repeat**
3:     Given $\mathcal{K}_{i,f_q}^{(n)}$, fix $\mathbf{v}_{k,i,f_q}^{(n)}$, $\forall k \in \mathcal{K}_{i,f_q}$, compute $\boldsymbol{\chi}_{k,i,f_q}^{(n)}$ according to (22), $\forall k \in \mathcal{K}_{i,f_q}^{(n)}$.
4:     Find the optimal transmit beamformer $\mathbf{v}_{k,i,f_q}^{(n+1)}$ and auxiliary variable $\omega_{k,i,f_q}^{(n+1)}$ under fixed $\boldsymbol{\chi}_{k,i,f_q}^{(n)}$ by solving the SOCP problem (23) with CVX [39].
5:     Update $R_{k,i,f_q}^{(n+1)}$ and $\Upsilon_{k,d,i,f_q}^{(n+1)}$ according to (14) and (6) respectively, $\forall d \in \mathcal{D}_{i,f_q}$, $\forall k \in \mathcal{K}_{i,f_q}^{(n)}$.
6:     Update $\mathcal{K}_{i,f_q}^{(n+1)}$ and $\boldsymbol{\chi}_{k,i,f_q}^{(n+1)}$ according to (27) and (22) respectively.
7:     Set $n \to n+1$
8:   **until** $\omega_{k,i,f_q}^{(n+1)} = 1, \forall k \in \mathcal{K}_{i,f_q}^{(n+1)}$.
9:   Set adaptive cell $j \to i, j \neq i, j \in \mathcal{A}_{f_q}$ as the target optimal cell. Go to step 3.
10: **until** Go through all adaptive cells and CPs.
11: **return** $\mathbf{v}_{k,i,f_q}$ and ω_{k,i,f_q} and go to Stage II.

B. COMPLEXITY ANALYSIS OF USC ALLOCATION

In this section, we analyze the complexity of Algorithm 1, and the main complexity of each iteration lies in step 2, which mainly solves the SOCP problem. Due to the alternating optimization methods, we first limit the analysis within an adaptive cell. In each adaptive cell, the problem has $2M|\mathcal{D}_{i,f_q}||\tilde{\mathcal{K}}_{i,f_q}| + |\tilde{\mathcal{K}}_{i,f_q}|$ variables, $|\tilde{\mathcal{K}}_{i,f_q}|$ SOC constraints with $2|\tilde{\mathcal{K}}_{i,f_q}|+1$ dimensions, and $2|\mathcal{D}_{i,f_q}|$ SOC constraints with $2M|\tilde{\mathcal{K}}_{i,f_q}|$ dimensions. Then, we obtain the complexity $O(V_{com})$ [40], where

$$V_{com} = \left(2M|\mathcal{D}_{i,f_q}||\tilde{\mathcal{K}}_{i,f_q}| + |\tilde{\mathcal{K}}_{i,f_q}|\right)^2 \times \left(2|\tilde{\mathcal{K}}_{i,f_q}|^2+|\tilde{\mathcal{K}}_{i,f_q}| + 4M|\tilde{\mathcal{K}}_{i,f_q}||\mathcal{D}_{i,f_q}|\right)$$

and the total number of iterations is $O\left(\sqrt{|\tilde{\mathcal{K}}_{i,f_q}| + |\mathcal{D}_{i,f_q}|}\right)$. Therefore, the total complexity to solve the SOCP problem (23) is given by $O\left(\sqrt{|\tilde{\mathcal{K}}_{i,f_q}| + |\mathcal{D}_{i,f_q}|}V_{com}\right)$. Next, we can see that Algorithm 1 should run $|\tilde{\mathcal{K}}_{i,f_q}|$ times at most, and the maximal total complexity of the USC algorithm within each adaptive cell is $O\left(|\tilde{\mathcal{K}}_{i,f_q}|\sqrt{|\tilde{\mathcal{K}}_{i,f_q}| + |\mathcal{D}_{i,f_q}|}V_{com}\right)$. Finally, as there exists $|\mathcal{A}|$ adaptive cells and $|\mathcal{F}|$ CPs, running K times at most, we obtain the overall complexity of the first stage as $O\left(\sum_{f_q\in\mathcal{F}}\sum_{i\in\mathcal{A}_{f_q}}\left(|\tilde{\mathcal{K}}_{i,f_q}|\sqrt{|\tilde{\mathcal{K}}_{i,f_q}| + |\mathcal{D}_{i,f_q}|}V_{com}\right)\right)$.

IV. THE SECOND STAGE: ITERATIVE ALGORITHMS FOR SPARSE BEAMFORMING

After the USC algorithm, we consider solving problem (12) for the selected users. Under the CP network architecture, we develop four low-complexity algorithms, named the CP-WMMSE algorithms, including the CP-WMMSE algorithm for sparse beamforming, the primal decomposition-based power allocation algorithm for sparse beamforming, the dual decomposition-based power allocation algorithm for sparse beamforming and a fast iterative algorithm for sparse beamforming.

In the CP-WMMSE algorithm for sparse beamforming, since the WMMSE algorithm [41] can be jointly applied between different CPs, we combine each CP jointly to solve the WSR problem. However, applying the algorithm in this way will consume a significant amount of computing resources. To further reduce the operation complexity, we introduce the decomposition algorithm-based power allocation for sparse beamforming. The reason is that further research revealed that the WMMSE algorithm can be applied to solve the WSR problem in different CPs in parallel. To use primal decomposition, we introduce a series of variables that represent the power allocation. By using these auxiliary variables, we can divide the original problem into two subproblems, and then we can update the resource allocation using a subgradient master algorithm. To further extend the decomposition to more CPs, we propose a dual decomposition-based power allocation algorithm for sparse beamforming.

Since the users and RAUs may be densely deployed in the practical environment, an additional fast iterative algorithm is required for reducing the computational complexity of the systems. By alternately optimizing the target adaptive cell and considering the other adaptive cells as constant interference, the algorithm reduces the complexity further without significant performance loss. Note that the fast iterative algorithm can be combined with both the primal decomposition- or dual decomposition-based power allocation algorithm for sparse beamforming and the CP-based scheme independently.

A. CP-WMMSE ALGORITHM FOR SPARSE BEAMFORMING

In traditional communication systems, the WSR problem has been solved efficiently, but in interlaced clustering, the problem may be difficult to solve due to resource allocation. Next, we introduce the CP-WMMSE algorithm for sparse beamforming and show it more clearly; we first rewrite C3 as

$$\text{C7: } \sum_{f_q\in\mathcal{F}}\sum_{i\in\mathcal{A}_{f_q}}\sum_{k\in\mathcal{K}_{i,f_q}} \|\boldsymbol{v}_{k,d,i,f_q}\|_2^2 \le P_{\text{sum},d}, \quad \forall d \in \mathcal{D}_{i,f_q}. \tag{28}$$

Now, we introduce the following proposition, which extends the WMMSE algorithm to interlaced clustering.

Proposition 1: Assume that interlaced clustering is assigned to different CPs with orthogonal frequency bands. For DAS with interlaced clustering, the original WMMSE algorithm can be solved jointly, i.e., the original WSR problem (12) can be rewritten as

$$\begin{aligned} &\max_{\boldsymbol{v}_{k,i,f_q}} \sum_{f_q\in\mathcal{F}}\sum_{i\in\mathcal{A}_{f_q}}\sum_{k\in\mathcal{K}_{i,f_q}} \Psi_{k,i,f_q}R_{k,i,f_q} \\ &\text{s.t. } \text{C5, C7.} \end{aligned} \tag{29}$$

The corresponding CP-WMMSE expression is

$$\begin{aligned} &\min_{\{\varepsilon_{k,i,f_q},\boldsymbol{u}_{k,i,f_q},\boldsymbol{v}_{k,i,f_q}\}} \sum_{f_q\in\mathcal{F}}\sum_{i\in\mathcal{A}_{f_q}}\sum_{k\in\mathcal{K}_{i,f_q}} \Psi_{k,i,f_q}\left(\varepsilon_{k,i,f_q}e_{k,i,f_q} - \log\varepsilon_{k,i,f_q}\right) \\ &\text{s.t. C5, C7.} \end{aligned} \tag{30}$$

where e_{k,i,f_q}, ε_{k,i,f_q}, and $\mathbf{u}_{k,i,f_q}$ are the MSE, the MSE weight, and the optimal receiver for user k, respectively, which are defined as

$$e_{k,i,f_q} = \mathbf{u}_{k,i,f_q}^H\mathbf{T}_{k,i,f_q}\mathbf{u}_{k,i,f_q} - 2\Re\{\mathbf{u}_{k,i,f_q}^H\mathbf{H}_{k,i,f_q}\mathbf{v}_{k,i,f_q}\} + 1, \tag{31a}$$

$$\varepsilon_{k,i,f_q} = e_{k,i,f_q}^{-1}, \tag{31b}$$

$$\mathbf{u}_{k,i,f_q} = \mathbf{T}_{k,i,f_q}^{-1}\mathbf{H}_{k,i,f_q}\mathbf{v}_{k,i,f_q}, \tag{31c}$$

where $\mathbf{T}_{k,i,f_q}$ is

$$\mathbf{T}_{k,i,f_q} = \sum_{\substack{r\in\mathcal{K}_{j,f_q},\\ j\in\mathcal{A}_{f_q}}} \mathbf{H}_{k,j,f_q}\mathbf{v}_{r,j,f_q}\mathbf{v}_{r,j,f_q}^H\mathbf{H}_{k,j,f_q}^H + \sigma_{k,i,f_q}^2\mathbf{I}, \forall j \in \mathcal{A}_{f_q}.$$

The proof of Proposition *1* is shown in Appendix A. In the following, we try to solve problem (30) using the block coordinate descent (BCD) method with a series of variables of three sets over ε_{k,i,f_q}, $\mathbf{u}_{k,i,f_q}$ and $\mathbf{v}_{k,i,f_q}$. Given $\mathbf{v}_{k,i,f_q}$, synchronously update ε_{k,i,f_q} and $\mathbf{u}_{k,i,f_q}$ by

$$\begin{cases} e_{k,i,f_1} = \boldsymbol{u}^H_{k,i,f_1}\boldsymbol{T}_{k,i,f_q}\boldsymbol{u}_{k,i,f_1} - 2\Re\left\{\boldsymbol{u}_{k,i,f_1}\boldsymbol{H}_{k,i,f_1}\boldsymbol{v}_{k,i,f_1}\right\}+1, \\ \dots \\ e_{k,i,f_q} = \boldsymbol{u}^H_{k,i,f_q}\boldsymbol{T}_{k,i,f_q}\boldsymbol{u}_{k,i,f_q} - 2\Re\left\{\boldsymbol{u}_{k,i,f_q}\boldsymbol{H}_{k,i,f_q}\boldsymbol{v}_{k,i,f_q}\right\}+1, \end{cases}$$
$$\begin{cases} \varepsilon_{k,i,f_1} = e^{-1}_{k,i,f_1}, \\ \dots \\ \varepsilon_{k,i,f_q} = e^{-1}_{k,i,f_q}, \end{cases} \begin{cases} \boldsymbol{u}_{k,i,f_1} = \boldsymbol{T}^{-1}_{k,i,f_q}\boldsymbol{H}_{k,i,f_1}\boldsymbol{v}_{k,i,f_1}, \\ \dots \\ \boldsymbol{u}_{k,i,f_q} = \boldsymbol{T}^{-1}_{k,i,f_q}\boldsymbol{H}_{k,i,f_q}\boldsymbol{v}_{k,i,f_q}, \end{cases} \tag{32}$$

By applying the BCD method with fixed $\mathbf{u}_{k,i,f_q}$ and ε_{k,i,f_q}, the problem (12) can be expressed as a CP-based quadratically constrained quadratic programming (CP-QCQP) problem

$$\min_{\{\boldsymbol{v}_{k,i,f_q}|k\in\mathcal{K}_{i,f_q},d\in\mathcal{D}_{i,f_q}\}} \sum_{f_q\in\mathcal{F}}\sum_{i\in\mathcal{A}_{f_q}} J_{k,i,f_q}$$
$$\text{s.t. C5, C7}, \tag{33}$$

where J_{k,i,f_q} is defined as

$$\begin{aligned} J_{k,i,f_q} &= -2\sum_{k\in\mathcal{K}_{i,f_q}} \Psi_{k,i,f_q}\varepsilon_{k,i,f_q}\Re\left\{\mathbf{u}^H_{k,i,f_q}\mathbf{H}_{k,i,f_q}\mathbf{v}_{k,i,f_q}\right\} \\ &\quad + \sum_{k\in\mathcal{K}_{i,f_q}} \mathbf{v}^H_{k,i,f_q}\Bigg(\sum_{j\in\mathcal{A}_{f_q}}\sum_{m\in\mathcal{K}_{j,f_q}} \Psi_{m,j,f_q}\varepsilon_{m,j,f_q}\mathbf{H}^H_{m,j,f_q} \\ &\quad \times \mathbf{u}_{m,j,f_q}\mathbf{u}^H_{m,j,f_q}\mathbf{H}_{m,j,f_q}\Bigg)\mathbf{v}_{k,i,f_q}. \end{aligned} \tag{34}$$

The problem in (29) involves two loops: an inner loop to design sparse beamforming $\mathbf{v}_{k,i,f_q}$ by solving the multi-CP optimization problem (33) with the BCD method, and an outer loop to synchronously update the optimal receiver ε_{k,i,f_q} and the optimal MSE weight $\mathbf{u}_{k,i,f_q}$. The two loops are summarized in Algorithm 2.

Algorithm 2 CP-WMMSE Algorithm for Sparse Beamforming.

Input:
$\Upsilon^{(0)}_{k,d,i,f_q}$, $\mathbf{v}^{(0)}_{k,i,f_q}$, $R^{(0)}_{k,i,f_q}$, $\forall k \in \mathcal{K}_{i,f_q}$, $\forall d \in \mathcal{D}_{i,f_q}$, $\forall i$.
Set $n \to 0$.

1: **repeat**
2: Fix $\mathbf{v}^{(n)}_{k,i,f_q}$, $\forall k \in \mathcal{K}_{i,f_q}$, compute the MMSE receiver $\mathbf{u}^{(n+1)}_{k,i,f_q}$ and the corresponding MSE $e^{(n+1)}_{k,i,f_q}$ according to (32), $\forall k \in \mathcal{K}_{i,f_q}$.
3: Find the optimal transmit beamformer $\mathbf{v}^{(n+1)}_{k,i,f_q}$ under fixed $\mathbf{u}^{(n+1)}_{k,i,f_q}$, $e^{(n+1)}_{k,i,f_q}$ by solving QCQP problem (33).
4: Update $R^{(n+1)}_{k,i,f_q}$ and $\Upsilon^{(n+1)}_{k,d,i,f_q}$ according to (14) and (6) respectively, $\forall d \in \mathcal{D}_{i,f_q}$, $\forall k \in \mathcal{K}_{i,f_q}$.
5: Set $n \to n+1$.
6: **until** Convergence.
7: **return** $\mathbf{v}_{k,i,f_q}$, $\forall k$.

Although, the original problem (12) can be solved by applying Algorithm 2, the high computational complexity has to be considered in practical applications, especially when the number of CPs is high. To further reduce the complexity, we propose the primal decomposition-based power allocation algorithm for sparse beamforming in the next subsection.

B. PRIMAL DECOMPOSITION-BASED POWER ALLOCATION ALGORITHM FOR SPARSE BEAMFORMING

In this section, we go one step further to design an algorithm with lower complexity. First, we assume that there are two CPs in our system. Obviously, problem (30) is convex and decomposable. Hence, the duality gap between constraint C7 of problem (30) and its optimal dual problem is zero. Then, we can decompose the original problem into two subproblems and solve it by computing the subgradient.

As assumed before, we rewrite the problem (30) as two subproblems; the first subproblem becomes

$$\begin{aligned} \min_{\substack{\{\varepsilon_{k,i,f_1},\boldsymbol{u}_{k,i,f_1},\\ \boldsymbol{v}_{k,i,f_1}\}}} &\sum_{i\in\mathcal{A}_{f_1}}\sum_{k\in\mathcal{K}_{i,f_1}} \Psi_{k,i,f_1}\left(\varepsilon_{k,i,f_1}e_{k,i,f_1} - \log\varepsilon_{k,i,f_1}\right) \\ \text{s.t. C8}: &\sum_{k\in\mathcal{K}_{i,f_1}} \Upsilon_{k,d,i,f_1}\tilde{R}_{k,i,f_1}\left\|\mathbf{v}_{k,d,i,f_1}\right\|_2^2 \le \frac{1}{2}C_{\text{sum},d}, \\ \text{C9}: &\sum_{k\in\mathcal{K}_{i,f_1}} \left\|\mathbf{v}_{k,d,i,f_1}\right\|_2^2 \le \eta_{d,i,f_1}P_{\text{sum},d}, \forall d\in\mathcal{D}_{i,f_1}, \end{aligned} \tag{35}$$

and the second subproblem becomes

$$\begin{aligned} \min_{\substack{\{\varepsilon_{r,j,f_2},\boldsymbol{u}_{r,j,f_2},\\ \boldsymbol{v}_{r,j,f_2}\}}} &\sum_{j\in\mathcal{A}_{f_2}}\sum_{r\in\mathcal{K}_{j,f_2}} \Psi_{r,j,f_2}\left(\varepsilon_{r,j,f_2}e_{r,j,f_2} - \log\varepsilon_{r,j,f_2}\right) \\ \text{s.t. C10}: &\sum_{r\in\mathcal{K}_{j,f_2}} \Upsilon_{r,d,j,f_2}\tilde{R}_{r,j,f_2}\left\|\mathbf{v}_{r,d,j,f_2}\right\|_2^2 \le \frac{1}{2}C_{\text{sum},d} \\ \text{C11}: &\sum_{r\in\mathcal{K}_{j,f_2}} \left\|\mathbf{v}_{r,d,j,f_2}\right\|_2^2 - P_{\text{sum},d} \le -\eta_{d,i,f_1}P_{\text{sum},d}, \\ &\forall d\in\mathcal{D}_{j,f_2} \end{aligned} \tag{36}$$

where η_{d,i,f_1} represents the ratio of power allocated to the first CP, and C8, C9, C10, C11 are the backhaul constraint and power constraint of CP1 and CP2. Note that, the constraint C8, C9, C10, C11 must be for the same RAU d.

To further describe the problem, we redefine the above two subproblems (35) and (36) as

$$\begin{aligned} \min_{\substack{\{\varepsilon_{k,i,f_1},\boldsymbol{u}_{k,i,f_1},\\ \boldsymbol{v}_{k,i,f_1}\}}} &\sum_{i\in\mathcal{A}_{f_1}}\sum_{k\in\mathcal{K}_{i,f_1}} f_1\left(\boldsymbol{v}_{k,i,f_1},\varepsilon_{k,i,f_1},\boldsymbol{u}_{k,i,f_1}\right) \\ \text{s.t. C12}: &\begin{cases}\boldsymbol{v}_{k,d,i,f_1}\in\mathcal{V}_1, \\ \lambda_{d,f_1}: h_{f_1}\left(\boldsymbol{v}_{k,d,i,f_1}\right) \le g_d,\end{cases} \end{aligned} \tag{37}$$

$$\begin{aligned} \min_{\substack{\{\varepsilon_{r,j,f_2},\boldsymbol{u}_{r,j,f_2},\\ \boldsymbol{v}_{r,j,f_2}\}}} &\sum_{j\in\mathcal{A}_{f_2}}\sum_{r\in\mathcal{K}_{j,f_2}} f_2\left(\boldsymbol{v}_{r,j,f_2},\varepsilon_{r,j,f_2},\boldsymbol{u}_{r,j,f_2}\right) \\ \text{s.t. C13}: &\begin{cases}\boldsymbol{v}_{r,d,j,f_2}\in\mathcal{V}_2, \\ \lambda_{d,f_2}: h_{f_2}\left(\boldsymbol{v}_{r,d,j,f_2}\right) \le -g_d,\end{cases} \end{aligned} \tag{38}$$

respectively, where $\mathcal{V}_1$ and $\mathcal{V}_2$ are the two feasible sets of the subproblems, i.e., for CP1, the feasible set is defined in C8, $g_d = \eta_{d,i,f_1} P_{\text{sum},d}$ represents the amount of power allocated to the first CP for the dth RAU, $f_1(\boldsymbol{v}_{k,i,f_1}, \varepsilon_{k,i,f_1}, \boldsymbol{u}_{k,i,f_1})$ and $f_2(\boldsymbol{v}_{r,j,f_2}, \varepsilon_{r,j,f_2}, \boldsymbol{u}_{r,j,f_2})$ are the object functions of the subproblem (35) and the subproblem (36) respectively, and λ_{d,f_1} and λ_{d,f_2} are the optimal dual variables associated with the constraints $h_{f_1}(\boldsymbol{v}_{k,d,i,f_1}) \leq g_d$ and $h_{f_2}(\boldsymbol{v}_{r,d,j,f_2}) \leq -g_d$. In addition, we define

$$\boldsymbol{\lambda}_1 = \left[\lambda_{1,f_1}, \lambda_{2,f_1}, \ldots, \lambda_{D,f_1}\right]^H,$$
$$\boldsymbol{\lambda}_2 = \left[\lambda_{1,f_2}, \lambda_{2,f_2}, \ldots, \lambda_{D,f_2}\right]^H,$$
$$\boldsymbol{g} = \left[\lambda_{1,f_1}, \lambda_{2,f_1}, \ldots, \lambda_{D,f_1}\right]^H.$$

Each subproblem can be solved by using the BCD method: given $\mathbf{v}_{k,i,f_q}$ and $\mathbf{u}_{k,i,f_q}$, update ε_{k,i,f_q} by using (31b); update $\mathbf{u}_{k,i,f_q}$ with $\mathbf{v}_{k,i,f_q}$ and ε_{k,i,f_q} by using (31c); and under fixed $\mathbf{u}_{k,i,f_q}$ and e_{k,i,f_q}, we obtain the QCQP problems

$$\min_{\{\boldsymbol{v}_{k,i,f_1} | k \in \mathcal{K}_{i,f_1}, d \in \mathcal{D}_{i,f_1}\}} \sum_{i \in \mathcal{A}_{f_1}} J_{k,i,f_1} \quad \text{s.t. C12,} \tag{39}$$

and

$$\min_{\{\boldsymbol{v}_{r,j,f_2} | k \in \mathcal{K}_{j,f_2}, d \in \mathcal{D}_{j,f_2}\}} \sum_{j \in \mathcal{A}_{f_2}} J_{r,j,f_2} \quad \text{s.t. C13,} \tag{40}$$

where J_{k,i,f_1} and J_{r,j,f_2} are defined in (34), where $q = 1$ and $q = 2$, respectively.

Then, we can solve the original problem (30) by applying primal decomposition with a subgradient master algorithm, which is given in Algorithm 3.

Algorithm 3 Primal Decomposition-Based Power Allocation Algorithm for Sparse Beamforming.

Input:
$\Upsilon^{(0)}_{k,d,i,f_q}, \mathbf{v}^{(0)}_{k,i,f_q}, R^{(0)}_{k,i,f_q}, g, \forall k \in \mathcal{K}_{i,f_q}, \forall d \in \mathcal{D}_{i,f_q}$, $q = 1, 2, \forall i$
Set $n \to 0$.

1: **repeat**
2: Run 2 in Algorithm 2, where $q = 1, 2$.
3: Find the optimal transmit beamformer $\mathbf{v}^{(n+1)}_{k,i,f_q}$ and λ_{d,f_q} under fixed $\mathbf{u}^{(n+1)}_{k,i,f_q}$, $e^{(n+1)}_{k,i,f_q}$ by solving QCQP problem (39) and (40) respectively, where $q = 1, 2$.
4: Update the power allocation: $\boldsymbol{g} := \boldsymbol{g} - \boldsymbol{\alpha}_n (\boldsymbol{\lambda}_2 - \boldsymbol{\lambda}_1)$, where $\boldsymbol{\alpha}_n$ are series-appropriate step sizes for RAUs.
5: Run 4 in Algorithm 2, where $q = 1, 2$.
6: Set $n \to n + 1$.
7: **until** Convergence.
8: **return** $\mathbf{v}_{k,i,f_q}, \forall k$.

C. DUAL DECOMPOSITION-BASED POWER ALLOCATION ALGORITHM FOR SPARSE BEAMFORMING

The primal decomposition is suited for two CPs; when the system consists of more than two CPs, the primal decomposition in the previous section will not work. This limitation motivated us to further introduce a dual decomposition-based power allocation algorithm for sparse beamforming, which can be used for more CPs. As the subject C7 is convex, we can add it to the objective by using dual variables $\boldsymbol{\lambda}$.

With a few simple manipulations, the Lagrangian function of problem (30) is given by

$$\begin{aligned} L\left(\boldsymbol{v}_{k,i,f_q}, \boldsymbol{\lambda}\right) &= \sum_{f_q \in \mathcal{F}} \sum_{i \in \mathcal{A}_{f_q}} \sum_{k \in \mathcal{K}_{i,f_q}} f_{f_q}\left(\boldsymbol{v}_{k,i,f_q}, \varepsilon_{k,i,f_q}, \boldsymbol{u}_{k,i,f_q}\right) \\ &\quad + \boldsymbol{\lambda}^T \sum_{f_q \in \mathcal{F}} \sum_{i \in \mathcal{A}_{f_q}} \sum_{k \in \mathcal{K}_{i,f_q}} h_{f_q}\left(\boldsymbol{v}_{k,d,i,f_q}\right) \\ &= \sum_{f_q \in \mathcal{F}} \sum_{i \in \mathcal{A}_{f_q}} \sum_{k \in \mathcal{K}_{i,f_q}} \Big(f_{f_q}\left(\boldsymbol{v}_{k,i,f_q}, \varepsilon_{k,i,f_q}, \boldsymbol{u}_{k,i,f_q}\right) \\ &\quad + \boldsymbol{\lambda}^T h_{f_q}\left(\boldsymbol{v}_{k,d,i,f_q}\right)\Big), \end{aligned} \tag{41}$$

where $\boldsymbol{\lambda}$ are the given dual variables, the dual function is defined as

$$g(\boldsymbol{\lambda}) = \sum_{f_q} g_{f_q}(\boldsymbol{\lambda}),$$

and the current resource function is defined as

$$h_{f_q}\left(\boldsymbol{v}_{k,i,f_q}\right) = \sum_{k \in \mathcal{K}_{i,f_1}} \left\|\mathbf{v}_{k,d,i,f_q}\right\|_2^2.$$

Note that $f_q(\boldsymbol{v}_{k,i,f_q}, \varepsilon_{k,i,f_q}, \boldsymbol{u}_{k,i,f_q})$ is defined the same as in (37) and (38), but can handle more than 2 CPs.

To find $g_{f_q}(\boldsymbol{\lambda})$, we solve the subproblems

$$\begin{aligned} \min_{\boldsymbol{v}_{k,i,f_q}} \; & \sum_{i \in \mathcal{A}_{f_q}} \sum_{k \in \mathcal{K}_{i,f_q}} \Big(f_q\left(\boldsymbol{v}_{k,i,f_q}, \varepsilon_{k,i,f_q}, \boldsymbol{u}_{k,i,f_q}\right) \\ & \qquad + \boldsymbol{\lambda}^T h_{f_q}\left(\boldsymbol{v}_{k,d,i,f_q}\right)\Big) \\ \text{s.t.} \; & \boldsymbol{v}_{k,i,f_q} \in \mathcal{V}_q. \end{aligned} \tag{42}$$

The subgradient of $-g_{f_q}$ at $\boldsymbol{\lambda}$ is $h_{f_q}(\tilde{\boldsymbol{v}}_{k,i,f_q})$, where $\tilde{\boldsymbol{v}}_{k,i,f_q}$ are any solution of subproblem (42). By solving the subproblems (42), we obtain solutions $\tilde{\boldsymbol{v}}_{k,d,i,f_q}$ and find the subgradient of g_{f_q}, the master problem objective. Hence, the subgradient of $-g_{f_q}$ is $\sum_{f_q} h_{f_q}(\tilde{\boldsymbol{v}}_{k,i,f_q})$. Finally, we obtain a dual decomposition-based power allocation algorithm for sparse beamforming by using a projected subgradient method to update $\boldsymbol{\lambda}$ and the BCD method to find the optimal $\tilde{\boldsymbol{v}}_{k,i,f_q}$.

In the dual decomposition-based power allocation algorithm for sparse beamforming, we first initialize $\boldsymbol{\lambda}$. Then, we solve each subproblem by using the BCD method, which is similar to Section IV Part C. Under fixed $\mathbf{u}_{k,i,f_q}$ and e_{k,i,f_q}, we obtain the following problem:

$$\min_{\{\boldsymbol{v}_{k,i,f_q}|k\in\mathcal{K}_{i,f_q},d\in\mathcal{D}_{i,f_q}\}} \sum_{i\in\mathcal{A}_{f_q}} \tilde{J}_{k,i,f_q} \quad \text{s.t. } \boldsymbol{v}_{k,i,f_q} \in \mathcal{V}_q, \tag{43}$$

where

$$\tilde{J}_{k,i,f_q} = J_{k,i,f_q} + \boldsymbol{\lambda}^T\left(h_{f_q}\left(\boldsymbol{v}_{k,d,i,f_q}\right)\right), \tag{44}$$

and J_{k,i,f_q} and $\mathcal{V}_q$ are defined in (34) and (37), respectively.

Note that, in the dual decomposition algorithm, we transform the power constraint into an objective function. Therefore, the iterations in the algorithm need not be feasible; however, in practice, we can construct a feasible set of variables by solving additional subproblems [31], [42].

Based on the above analysis, the dual decomposition-based power allocation algorithm for sparse beamforming is given in Algorithm 4.

Algorithm 4 Dual Decomposition-Based Power Allocation Algorithm for Sparse Beamforming.

Input:
$\Upsilon^{(0)}_{k,d,i,f_q}$, $\mathbf{v}^{(0)}_{k,i,f_q}$, $\boldsymbol{\lambda}^{(0)}$, $R^{(0)}_{k,i,f_q}$, $\forall k \in \mathcal{K}_{i,f_q}$, $\forall d \in \mathcal{D}_{i,f_q}$, $q = 1, \ldots, Q$, $\forall i$
Set $n \rightarrow 0$.

1: **repeat**
2: Run 2 in Algorithm 2.
3: Find the optimal transmit beamformer $\mathbf{v}^{(n+1)}_{k,i,f_q}$ under fixed $\mathbf{u}^{(n+1)}_{k,i,f_q}$, $e^{(n+1)}_{k,i,f_q}$ by solving subproblem (42).
4: Update dual variables (the price vector): $\boldsymbol{\lambda}^{(n+1)} := \left(\boldsymbol{\lambda}^{(n)} + \boldsymbol{\alpha}_n \sum_{\boldsymbol{v}^{(n+1)}_{k,d,i,f_q}} h_{f_q}\left(\boldsymbol{v}^{(n+1)}_{k,d,i,f_q}\right)\right)_+$.
5: Run 4 in Algorithm 2.
6: Set $n \rightarrow n+1$.
7: **until** Convergence.
8: **return** $\mathbf{v}_{k,i,f_q}$, $\forall k$.

D. A FAST ITERATIVE ALGORITHM FOR SPARSE BEAMFORMING

The above analysis shows that the primal decomposition and the dual decomposition can solve the original problem with lower complexity; however, with increasing densities, the computational cost must be further reduced. In this section, a fast iterative algorithm is applied to realize this requirement.

Similar to Section III, we can also apply an alternative optimize method in each subproblem of the above two decomposition algorithms. For primal decomposition, we limit the data sharing strategy in the adaptive cell, and the final problem of the two subproblems can be rewritten as

$$\min_{\{\boldsymbol{v}_{k,d,i,f_1}|k\in\mathcal{K}_{i,f_1},d\in\mathcal{D}_{i,f_q}\}} J_{k,i,f_1} \quad \text{s.t. C12}, \tag{45}$$

and

$$\min_{\{\boldsymbol{v}_{k,j,f_2}|r\in\mathcal{K}_{j,f_2},d\in\mathcal{D}_{j,f_q}\}} J_{r,j,f_2} \quad \text{s.t. C13}, \tag{46}$$

where J_{k,i,f_1} and J_{r,j,f_2} are defined in (34), where $q = 1$ and $q = 2$, respectively.

For dual decomposition, we have the final optimization subproblem

$$\min_{\{\boldsymbol{v}_{k,d,i,f_q}|k\in\mathcal{K}_{i,f_q},d\in\mathcal{D}_{i,f_q}\}} \tilde{J}_{k,i,f_q} \quad \text{s.t. } \boldsymbol{v}_{k,d,i,f_q} \in \mathcal{V}_q, \tag{47}$$

where $\tilde{J}_{k,i,f_q}$ is defined in (44).

Then, fast iterative algorithms for sparse beamforming with primal decomposition and dual decomposition are given in Algorithm 5 and Algorithm 6, respectively.

Algorithm 5 A Fast Iterative Algorithm for Sparse Beamforming with Primal Decomposition.

Input:
$\Upsilon^{(0)}_{k,d,i,f_q}$, $\mathbf{v}^{(0)}_{k,i,f_q}$, $R^{(0)}_{k,i,f_q}$, $\forall k \in \mathcal{K}_{i,f_q}$, $\forall d \in \mathcal{D}_{i,f_q}$, $\forall i$
Set $n \rightarrow 0$
Set adaptive cell i as the target optimal cell.

1: **repeat**
2: **repeat**
3: Under the fixed target adaptive cell, run 2, 3, 4 and 6 of Algorithm 3, where in 3 and 4, we find the optimal transmit beamformer $\mathbf{v}^{(n+1)}_{k,i,f_q}$ and λ_{d,f_1} by solving problems (45) and (46) instead of (39) and (40).
4: Set adaptive cell $j \rightarrow i, j \neq i, j \in \mathcal{A}_{f_q}$ as the target optimal cell. Go to step 3.
5: **until** Go through all adaptive cells.
6: Run 5 of Algorithm 3.
7: Set $n \rightarrow n+1$.
8: **until** Convergence.
9: **return** $\mathbf{v}_{k,i,f_q}$, $\forall k$.

E. COMPLEXITY ANALYSIS OF THE ALGORITHMS PROPOSED IN THE SECOND STAGE

In this subsection, we analyze the complexity of the algorithms to solve the original problem (12). First, we assume that a practical network with $|\tilde{\mathcal{K}}_{i,f_q}| > |\mathcal{D}_{i,f_q}| > M > N$. In the CP-WMMSE algorithm, the computational complexity of step 2 is $O\Big(\sum_{f_q\in\mathcal{F}}\sum_{i\in\mathcal{A}_{f_q}} |\tilde{\mathcal{K}}_{i,f_q}|^2|\mathcal{D}_{i,f_q}|MN\Big)$, and updating the weights ε_{k,i,f_q} is only $O\Big(\sum_{f_q\in\mathcal{F}}\sum_{i\in\mathcal{A}_{f_q}} |\tilde{\mathcal{K}}_{i,f_q}|\Big)$. In step 2, we solve a QCQP problem, which can also be equivalently reformulated as an SOCP problem. Then, the total number of variables as $\sum_{f_q\in\mathcal{F}}\sum_{i\in\mathcal{A}_{f_q}} |\tilde{\mathcal{K}}_{i,f_q}||\mathcal{D}_{i,f_q}|M$ and the computational complexity is approximately $O\Big(\sum_{f_q\in\mathcal{F}}\sum_{i\in\mathcal{A}_{f_q}} |\tilde{\mathcal{K}}_{i,f_q}||$

Algorithm 6 A Fast Iterative Algorithm for Sparse Beamforming with Dual Decomposition.

Input:
$\Upsilon_{k,d,i,f_q}^{(0)}, \mathbf{v}_{k,i,f_q}^{(0)}, \lambda, R_{k,i,f_q}^{(0)}, \forall k \in \mathcal{K}_{i,f_q}, \forall d \in \mathcal{D}_{i,f_q}, q = 1, \ldots, Q, \forall i.$
Set $n \rightarrow 0$.
Set adaptive cell i as the target optimal cell.

1: **repeat**
2: **repeat**
3: Run 2, 3 and 5 of Algorithm 4, where in 3, we find the optimal transmit beamformer $\mathbf{v}_{k,i,f_q}^{(n+1)}$ according to (47) instead of (42).
4: Set adaptive cell $j \rightarrow i, j \neq i, j \in \mathcal{A}_{f_q}$ as target optimal cell. Go to step 3.
5: **until** Go through all adaptive cells.
6: Run 4 of Algorithm 4.
7: Set $n \rightarrow n + 1$.
8: **until** Convergence.
9: **return** $\mathbf{v}_{k,i,f_q}, \forall k$.

$\mathcal{D}_{i,f_q}|M\Big)^{3.5}$ [43] when solving the problem with the interior-point method. In step 4, the computational complexity of the rate updating procedure is $O\Big(\sum_{f_q \in \mathcal{F}} \sum_{i \in \mathcal{A}_{f_q}} |\tilde{\mathcal{K}}_{i,f_q}|^2 |\mathcal{D}_{i,f_q}| MN\Big)$, which is the same as computing the MSE. Therefore, the computational complexity of the CP-WMMSE algorithm mainly derives from step 2. Suppose Algorithm 2 requires η total number of iterations to converge; the overall complexity is then $O\Big(\Big(\sum_{f_q \in \mathcal{F}} \sum_{i \in \mathcal{A}_{f_q}} |\tilde{\mathcal{K}}_{i,f_q}||\mathcal{D}_{i,f_q}|M\Big)^{3.5} \eta\Big)$. As the computational complexity of the proposed algorithms mainly lies in solving the QCQP problem, we focus on analyzing the QCQP computation in the following.

In Algorithm 3 and Algorithm 4, the number of variables is $\sum_{i \in \mathcal{A}_{f_q}} |\tilde{\mathcal{K}}_{i,f_q}||\mathcal{D}_{i,f_q}|M$; then, the total computational complexity is $O\Big(\sum_{i \in \mathcal{A}_{f_q}} |\tilde{\mathcal{K}}_{i,f_q}||\mathcal{D}_{i,f_q}|M\Big)^{3.5}$, and the overall computational complexity is

$$O\Big(\sum_{f_q \in \mathcal{F}} \Big(\sum_{i \in \mathcal{A}_{f_q}} |\tilde{\mathcal{K}}_{i,f_q}||\mathcal{D}_{i,f_q}|M\Big)^{3.5} \eta\Big).$$

Similar to Algorithm 3 and Algorithm 4, the computational complexity of Algorithm 5 and Algorithm 6 is given by $O\Big(\sum_{f_q \in \mathcal{F}} \Big(\sum_{i \in \mathcal{A}_{f_q}} |\tilde{\mathcal{K}}_{i,f_q}||\mathcal{D}_{i,f_q}|M\Big)^{3.5} \eta\Big)$. From the above analysis, we can conclude that with decomposition, the complexity is reduced, and that with alternating optimization, the complexity is further reduced.

F. PERFORMANCE COMPARISON OF THE ALGORITHMS PROPOSED IN THE SECOND STAGE

Overall, we summarize the advantages and disadvantages of the proposed algorithms as follows:

1) The order of complexity of the proposed algorithms from high to low is Algorithm 2, Algorithm 3 and Algorithm 5. Furthermore, Algorithm 4 and Algorithm 6 have the same complexity as Algorithm 3 and Algorithm 5, respectively.

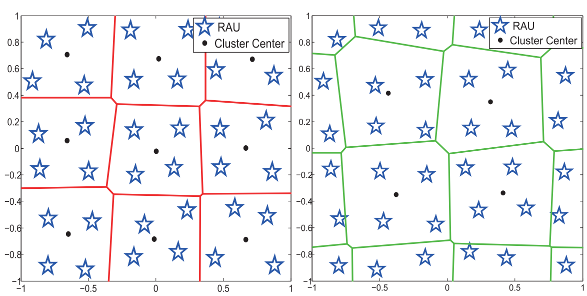
2) Since Algorithm 2 jointly optimizes the beamforming and the power allocation for RAUs between CPs throughout the network, the performance of Algorithm 2 should be higher than that of the other proposed algorithms. Similarly, by jointly optimizing the beamforming and the power allocation for RAUs within each CP simultaneously, the performance of the two decomposition algorithms should be higher than the fast iterative algorithm with corresponding decomposition algorithm, respectively. Thus, the performance order from high to low is Algorithm 2, Algorithm 3, Algorithm 5; similarly, the complexity of the other two algorithms together with Algorithm 2 is ranked as Algorithm 2, Algorithm 4, Algorithm 6.

V. NUMERICAL RESULTS

In this section, we perform a numerical evaluation of our proposed algorithms. Numerical evaluations are conducted to show the effectiveness of the proposed algorithms, with the simulation parameters listed in Table 1.

TABLE 1. Simulation parameters.

Cellular Layout	9-cell
Layout dimensions(km)	2×2
Tot bandwidth	10 MHz
Max.Tx power for RAU	3 W
Backhaul constraint for RAU	600 Mbps
No of CPs/users/antennas(RAU,user)	2/135/(2,2)
Path loss from the BS to the user	140.7+36.7log10(d)
Lognormal shadowing	8 dB
Rayleigh small-scale fading	0 dB
Background noise	-169 dBm/Hz
Reweighting function parameter	$\varsigma = 10^{-10}$



(a) (b)

FIGURE 2. Cluster patterns and simulation layouts. (a) Planar, CP1. (b) Planar, CP2.

Fig 2 shows a large simulation layout containing several adaptive cells. For the sake of brevity, the simulation results are restricted to the important case of interlaced clustering with equal partitioning of bandwidth among the CPs as illustrated in Fig. 2a and Fig.2b.

A. USER SELECTION PERFORMANCE IN THE CLUSTER PATTERN SCHEME

1) Convergence behavior of USC: Fig. 3 shows the convergence behavior of USC with different rate requirements for selected users. For different CP schemes, different channel realizations may lead to different solutions; we run the two CPs differently, and then compute the average of the two CPs. It can be seen from Fig. 3 that the number of users monotonically increases during the iterative procedure for two CP schemes. The algorithm converges very fast, and in general, six iterations can achieve a stable value for different rate requirements and different CP schemes. In addition, it is obvious that the lower the rate requirement is set, the more users that will be selected.

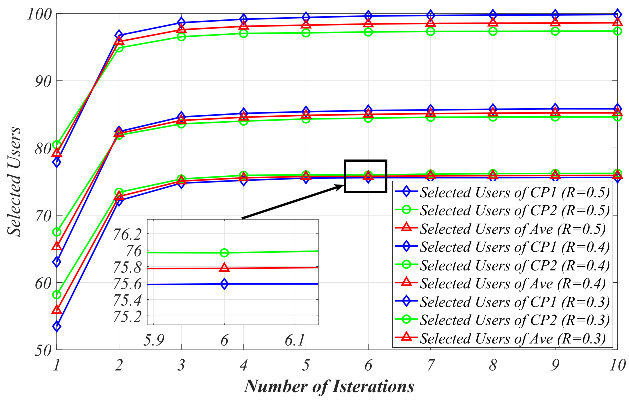


FIGURE 3. Convergence behavior of USC.

2) Sum-rate performance of the USC algorithm: Fig. 4 compares the sum-rate performance of the USC algorithm with our proposed algorithms. Under a fixed rate requirement, we compute the sum-rate of the five proposed algorithms with the same channel realizations. As expected, the sum-rate decreases with the rate requirements for all algorithms. The different algorithms have almost the same rate of decline. Moreover, when the rate requirement increases to 1 bit/s/Hz the sum-rate decreases only 2.03%, which can be negligible. In addition, the gap between the algorithm with the lowest sum-rate, i.e., a fast iterative algorithm for sparse beamforming with dual decomposition and the highest sum-rate, i.e., the CP-WMMSE algorithm is approximately 2.18%, which can also be negligible. Therefore, the proposed algorithms can be considered to have approach performance gains in the sum-rate.

3) The user selection performance and complexity of the USC algorithm along with the proposed algorithms: The user selection performance and complexity of the USC algorithm are shown in Fig. 5a and Fig. 5b respectively. In Fig. 5a, three different schemes are considered, i.e., CP1, CP2, and the number of selected users throughout the network. We can see that the selected users decrease with the rate requirement, and the two CPs have almost the same performance on the number of selected users. When the rate requirement increases to 1 bit/s/Hz, the users selected of each CP decreases to

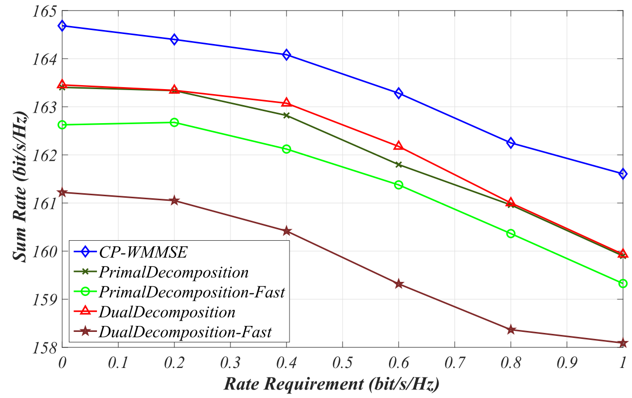


FIGURE 4. Sum-rate change versus rate requirement.

approximately 49.39. However, the number of all the selected users in the system is above 72.54, which is one of the advantages of the CP schemes. Accordingly, the complexity of the USC algorithm along with the proposed algorithms can be seen in Fig. 5b. It can be seen from the figures that the CP-WMMSE and the decomposition algorithms with the same complexity performance decrease with the USC algorithm. When the rate requirement increases to 0.6, i.e., the selected users are below 67.82, the complexity of the two algorithms changes slowly. At the same time, the complexity of the fast decomposition algorithms changes slowly and on a high position as the number of users decreases throughout. Compared to the other algorithms, the fast decomposition algorithms run with lower complexity but can still decrease nearly 27.13% with the USC algorithm when the rate requirement increase to 0.6. When the rate requirement increases to 1 bit/s/Hz, the complexity can be reduced by approximately 30%.

B. PERFORMANCE OF THE PROPOSED SPARSE BEAMFORMING ALGORITHMS

After considering the tradeoff between the complexity and the sum-rate performance loss, and based on the above analysis, we choose 0.6 bit/s/Hz as the users' rate requirement for the USC algorithm and analyze the performance of the proposed algorithms. In this subsection, we compare the performance of the proposed algorithms with the dynamic cluster (DCA) algorithm [22]. The complexity of the DCA is $O\Big(\sum_{f_q\in\mathcal{F}}\Big(\sum_{i\in\mathcal{A}_{f_q}}\big|\tilde{\mathcal{K}}_{i,f_q}\big|\,|\mathcal{D}|\,M\Big)^{3.5}\eta\Big)$, which is higher than the CP-WMMSE. Thus, the additional techniques are necessary to deal with the high complexity. We choose the iterative user pool shrinking technique when running the DCA algorithm, i.e., the user rate below 0.01 bps/Hz will be negligible in each iteration.

1) Convergence behavior of the proposed algorithms based on power allocation: Fig. 6 shows the convergence behavior of the proposed algorithms along with the DCA algorithm. As we can see, in the CP-based scheme, the proposed algorithms converge in approximately $30-40$ iterations under the same settings, which is a faster convergence than the DCA

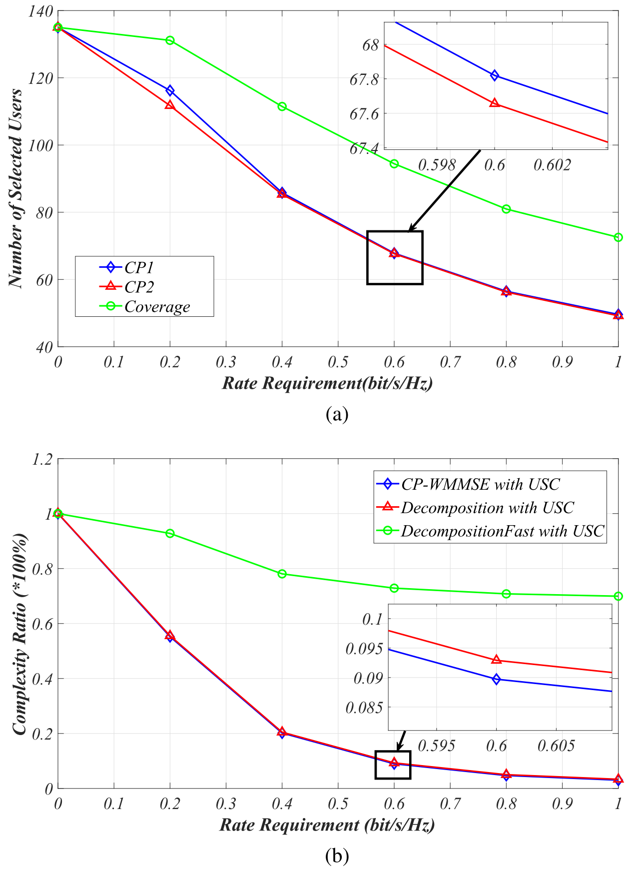


FIGURE 5. The complexity and the selected user selection performance under the rate requirement increase. (a) The number of the selected users versus the rate requirement. (b) The complexity of the proposed algorithms versus the rate requirement.

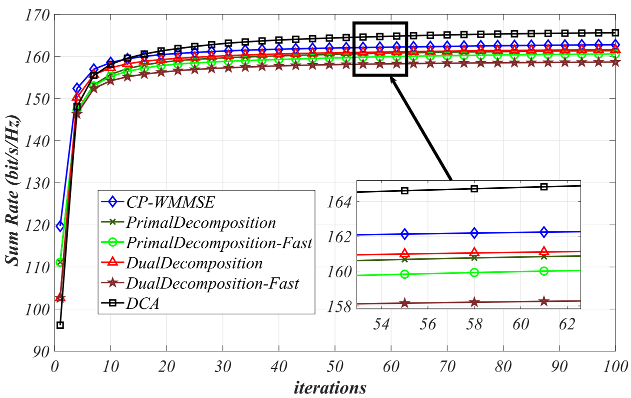


FIGURE 6. Convergence behavior of the proposed algorithms in the second stage.

algorithm, although the sum-rate of the proposed algorithms is slightly lower than that of the DCA algorithm.

2) Cumulative distribution function of the long-term average user rates: In Fig 7, we compare the cumulative distribution function of the long-term average user rates between the proposed algorithms with the CP-based scheme and the DCA algorithm. As expected, the performance of the DCA

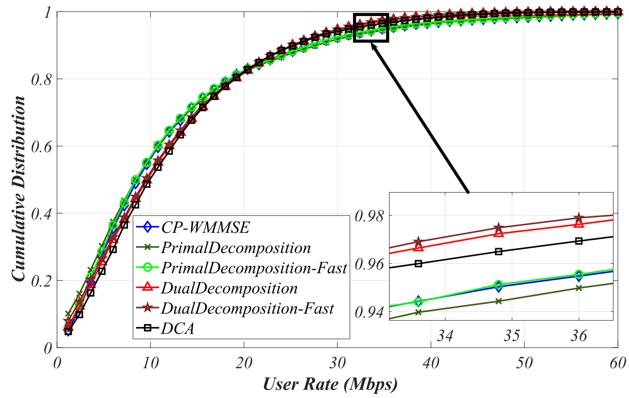


FIGURE 7. Cumulative distribution function of the average user data rate compared with that of the DCA algorithm.

algorithm is slightly better than that of the proposed algorithms in the user rate. The slightly lower performance gains are caused by more users being served in the CP-based scheme with the adaptive cell-wide beamforming design in our scheme rather than the network-wide design in the DCA algorithm.

3) Performance of the user rate distribution: The long-term average user rates distributions are given in Fig. 8. We observe from the results that the cell-edge performance of the CP-WMMSE algorithm that the primal decomposition algorithm and the primal decomposition-fast algorithm are similar to each other, while the two dual decomposition algorithms are similar. The reason is that the above three algorithms have the same power constraints, while the dual decomposition algorithms have no power constraints. However, the proposed algorithms have a similar distribution of cell-edge user convergence. In addition, we conclude that the proposed algorithms enhance the edge-user rates compared to the DCA algorithm (see Fig. 8g and Fig. 8h'). The reason is that the DCA algorithm is based on a user-centric cluster strategy throughout the network, i.e., the strategy does not affect the cell-edge user when the CSI is not better than other users. At the same time, the proposed algorithms jointly consider the cell-edge users' performance and user-centric cluster strategy. In addition, some users located in the cluster center are also well served compared with the DCA algorithm and benefit from the sparse beamforming design with interlaced clustering.

4) Long-term users' convergence performance: Since the USC algorithm and the iterative user pool shrinking technique are applied in this paper, the long-term users' convergence performance needs to be carefully analyzed. Fig. 9 shows the long-term users' convergence performance of the proposed algorithms and the DCA algorithm. The network with the CP-based algorithms can serve 91.12% of the users after the 6 T-F slot, which is higher than applying the DCA algorithm, which only serves 80.71% of the users. Note that the proposed algorithms have different convergence performance; this is because, after designing sparse precoding, there are

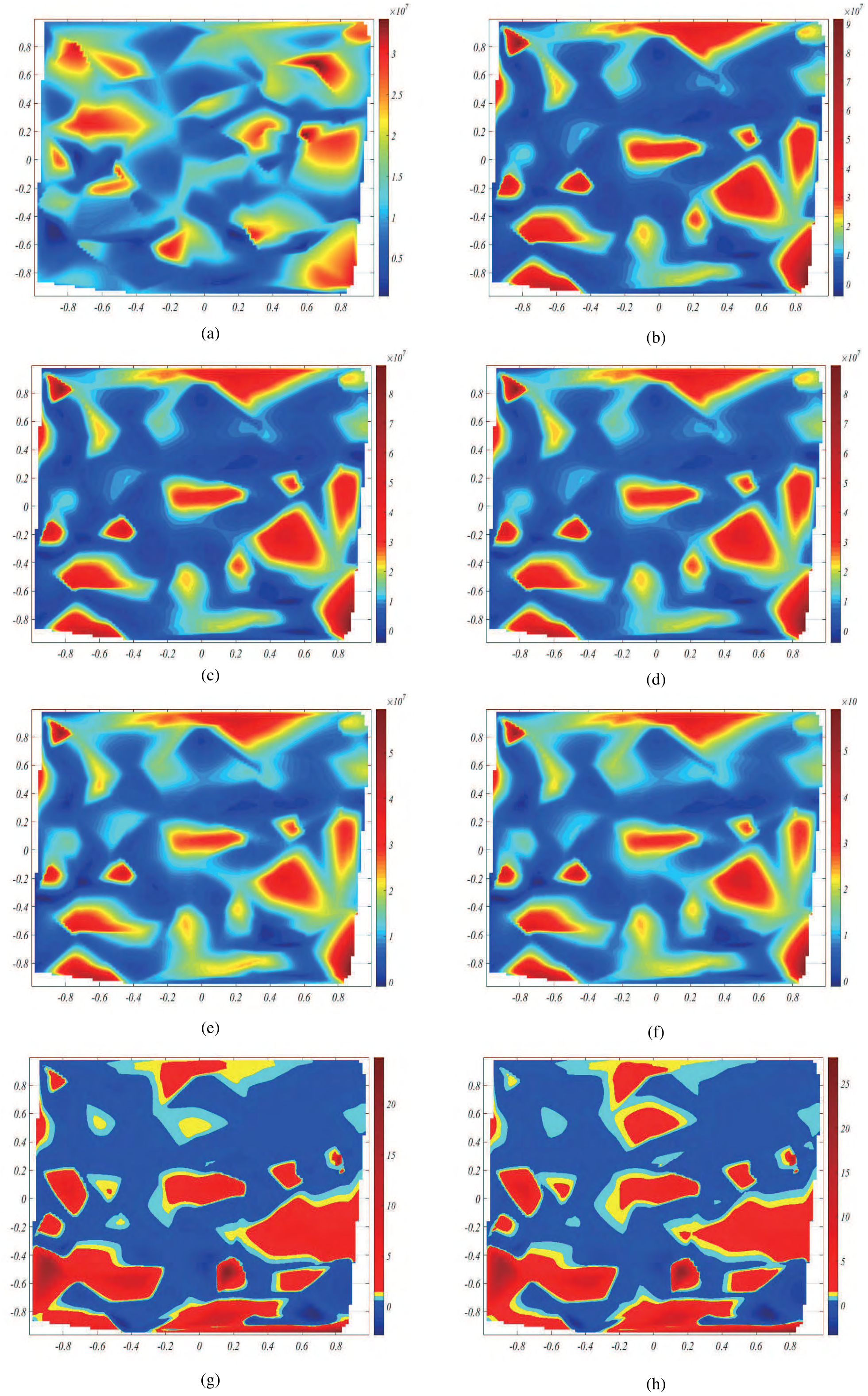


FIGURE 8. User rates distribution performance of the proposed algorithms. (a) Rates with DCA algorithm. (b) Rates with the CP-WMMSE algorithm. (c) Rates with the primal decomposition algorithm. (d) Rates with the primal decomposition-fast algorithm. (e) Rates with the dual decomposition algorithm. (f) Rates with the dual decomposition-fast algorithm. (g) Rate improvement factor $\frac{R_{\text{CP-WMMSE}}-R_{\text{DCA}}}{R_{\text{DCA}}}$. (h) Rate improvement factor $\frac{R_{\text{Dual}}-R_{\text{DCA}}}{R_{\text{DCA}}}$.

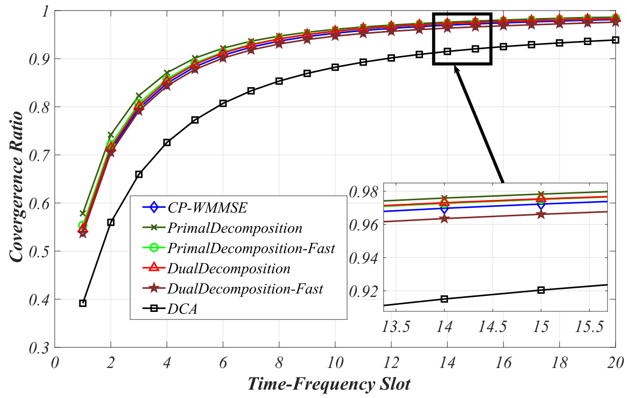


FIGURE 9. Long-term users convergence performance.

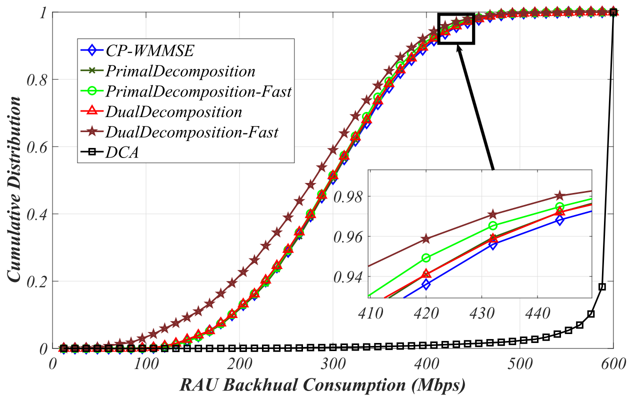


FIGURE 10. Cumulative distribution function of the long-term average RAU backhaul consumption.

few user rates below 0.01 bps/Hz, which can be ignored in this paper.

5) Cumulative distribution function of the long-term average RAU backhaul consumption: Fig. 10 shows the cumulative distribution function of the long-term average RAU backhaul consumption of the proposed algorithms and the DCA algorithm. We can see that approximately 90% of the RAUs consume less than 400 Mbps backhaul capacity, while the DCA algorithm consumes almost all of the 600 Mbps backhaul capacity. The reason is that the CP-based sparse beamforming design does not require user rate sharing throughout the network; instead, the sharing is limited in adaptive cells.

VI. CONCLUSION

In this paper, we studied the sparse beamforming for an ultradense DAS with interlaced clustering under power constraints and imperfect backhaul. Then, a power allocation and sparse beamforming design were jointly considered by applying WMMSE with different CPs. To solve the problem more efficiently, we formulated a two-stage optimization problem. The first stage selects the largest subset of users that can be admitted to the network by a fast USC algorithm with an alternating optimization method. Then, the second stage solves the WSR problem with the CP-WMMSE algorithm, the primal decomposition-based power allocation algorithm and the dual decomposition-based power allocation algorithm. Finally, we applied the alternating optimization method to cooperate with the three algorithms, which decreased the complexity of the three algorithms. Simulation results showed that the proposed algorithms converge quickly, which is attractive for practical implementation. With the USC algorithms, the sum-rate performance of the proposed algorithms only slightly decreased, but the complexity was significantly reduced, e.g., when the rate requirement increased to 1 bit/s/Hz, the sum-rate decreased only 2.03% while the complexity reduced at least 30%. Moreover, our proposed algorithms were shown to enhance the performance gains of the cell-edge users and reduce the backhaul consumption without losing significant performance gains.

APPENDIX A
PROOF OF LEMMA 1

We first rewrite the right-hand side of (18) by completing the square as

$$\boldsymbol{b}^H(\boldsymbol{v})\boldsymbol{A}^{-1}(\boldsymbol{v})\boldsymbol{b}(\boldsymbol{v}) - \left(\boldsymbol{\chi}-\boldsymbol{A}^{-1}(\boldsymbol{v})\boldsymbol{b}(\boldsymbol{v})\right)^H \boldsymbol{A}(\boldsymbol{v})\left(\boldsymbol{\chi}-\boldsymbol{A}^{-1}(\boldsymbol{v})\boldsymbol{b}(\boldsymbol{v})\right) \quad (48)$$

Then, it is easy to find that the maximized value of (48) is $\boldsymbol{b}^H(\boldsymbol{v})\boldsymbol{A}^{-1}(\boldsymbol{v})\boldsymbol{b}(\boldsymbol{v})$ when $\tilde{\boldsymbol{\chi}} = \boldsymbol{A}^{-1}(\boldsymbol{v})\boldsymbol{b}(\boldsymbol{v})$ [38]. In addition, the inequation

$$\left(\boldsymbol{\chi}-\boldsymbol{A}^{-1}(\boldsymbol{v})\boldsymbol{b}(\boldsymbol{v})\right)^H \boldsymbol{A}(\boldsymbol{v})\left(\boldsymbol{\chi}-\boldsymbol{A}^{-1}(\boldsymbol{v})\boldsymbol{b}(\boldsymbol{v})\right) \geq 0$$

is always established for every $\boldsymbol{\chi}$, $\boldsymbol{A}(\boldsymbol{v})$ and $\boldsymbol{b}(\boldsymbol{v})$. Therefore, inequivalence to (18) is therefore established, and the two sides can be equal if and only if $\tilde{\boldsymbol{\chi}} = \boldsymbol{A}^{-1}(\boldsymbol{v})\boldsymbol{b}(\boldsymbol{v})$. The proof is therefore complete.

APPENDIX B
PROOF OF PROPOSITION 1

Since interlaced clustering assigns different CPs to orthogonal frequency bands, the sum rate for the users in different interlaced clusters are independent except for constraint C7, i.e., the beamforming for users in different CPs will not affect each other. Therefore, there are no variables coupled together in the objective function. Moreover, C7 has a convex form, so we only need to deal with the objective function for each CP independently. Thus, first, we alternatively fixed the other CPs and the other variables of the target CP; the objective function in (30) is convex with respect to ε_{k,i,f_1}, $\ldots$, ε_{k,i,f_q}. Then, by alternatively checking the first-order optimality condition for ε_{k,i,f_q}, $\forall f_q \in \mathcal{F}$, we obtain

$$\begin{cases} \varepsilon_{k,i,f_1}^{opt} = e_{k,i,f_1}^{-1}, \\ \quad \cdots \\ \varepsilon_{k,i,f_q}^{opt} = e_{k,i,f_q}^{-1}, \end{cases} \quad (49)$$

Substituting the optimal ε_{k,i,f_1}, $\ldots$, ε_{k,i,f_q} and $\mathbf{u}_{k,i,f_1}$, $\ldots$, $\mathbf{u}_{k,i,f_q}$, we have the following equivalent optimization problem:

$$\max_{\{\varepsilon_{k,i,f_q},\boldsymbol{u}_{k,i,f_q},\boldsymbol{v}_{k,i,f_q}\}} \sum_{f_q\in\mathcal{F}} \sum_{i\in\mathcal{A}_{f_q}} \sum_{k\in\mathcal{K}_{i,f_q}} \Psi_{k,i,f_q} \log e_{k,i,f_q}^{-1}$$
$$\text{s.t. C5, C7} \tag{50}$$

Bringing formulas (31) of each CP into (50) completes the proof.

REFERENCES

[1] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu, "5G network capacity: Key elements and technologies," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 71–78, Mar. 2014.

[2] H. Wei, D. Wang, H. Zhu, J. Wang, S. Sun, and X. You, "Mutual coupling calibration for multiuser massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 606–619, Jan. 2016.

[3] C. Li, K. Song, D. Wang, F.-C. Zheng, and L. Yang, "Optimal remote radio head selection for cloud radio access networks," *Sci. China Inf. Sci.*, vol. 59, no. 10, p. 102315, 2016.

[4] Y. Huo, X. Dong, and W. Xu, "5G cellular user equipment: From theory to practical hardware design," *IEEE Access*, vol. 5, pp. 13992–14010, 2017.

[5] D. Wang, Y. Zhang, H. Wei, X. You, X. Gao, and J. Wang, "An overview of transmission theory and techniques of large-scale antenna systems for 5G wireless communications," *Sci. China Inf. Sci.*, vol. 59, no. 8, p. 081301, 2016.

[6] Y. Xin, R. Zhang, D. Wang, J. Li, L. Yang, and X. You, "Antenna clustering for bidirectional dynamic network with large-scale distributed antenna systems," *IEEE Access*, vol. 5, pp. 4037–4047, 2017.

[7] D. Wang, J. Wang, X. You, Y. Wang, M. Chen, and X. Hou, "Spectral efficiency of distributed MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 2112–2127, Oct. 2013.

[8] D. Wang, H. Gu, H. Wei, X. Duan, C. Li, and X. You, "Design of pilot assignment for large-scale distributed antenna systems," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. 99, no. 9, pp. 1674–1682, 2016.

[9] J. Wang and L. Dai, "Downlink rate analysis for virtual-cell based large-scale distributed antenna systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1998–2011, Mar. 2016.

[10] J. Wang and L. Dai, "Asymptotic rate analysis of Downlink multi-user systems with co-located and distributed antennas," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3046–3058, Jun. 2015.

[11] H. Wei, D. Wang, J. Wang, and X. You, "Impact of RF mismatches on the performance of massive MIMO systems with ZF precoding," *Sci. China Inf. Sci.*, vol. 59, no. 2, pp. 1–14, 2016.

[12] H, Wei, D. Wang, J. Wang, and X. You, "TDD reciprocity calibration for multi-user massive MIMO systems with iterative coordinate descent," *Sci. China Inf. Sci.*, vol. 59, no. 10, p. 102306, 2016.

[13] P. Wan, M. McGuire, and X. Dong, "Near-optimal channel estimation for OFDM in fast-fading channels," *IEEE Trans. Veh. Technol.*, vol. 60, no. 8, pp. 3780–3791, Oct. 2011.

[14] S. Bassoy, H. Farooq, M. A. Imran, and A. Imran, "Coordinated multipoint clustering schemes: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 743–764, 2nd Quart., 2017.

[15] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.

[16] S. S. Ali and N. Saxena, "A novel static clustering approach for CoMP," in *Proc. 7th Int. Conf. Comput. Converg. Technol. (ICCCT)*, Dec. 2012, pp. 757–762.

[17] H. Shimodaira, G. K. Tran, K. Sakaguchi, K. Araki, S. Nanba, and S. Konishi, "Diamond cellular network–optimal combination of small power basestations and CoMP cellular networks," *IEICE Trans. Commun.*, vol. 99, no. 4, pp. 917–927, 2016.

[18] P. Marsch and G. Fettweis, "Static clustering for cooperative multipoint (CoMP) in mobile communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2011, pp. 1–6.

[19] J. Liu and D. Wang, "An improved dynamic clustering algorithm for multiuser distributed antenna system," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Nov. 2009, pp. 1–5.

[20] V. Garcia, Y. Zhou, and J. Shi, "Coordinated multipoint transmission in dense cellular networks with user-centric adaptive clustering," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4297–4308, Aug. 2014.

[21] F. Guidolin, L. Badia, and M. Zorzi, "A distributed clustering algorithm for coordinated multipoint in LTE networks," *IEEE Wireless Commun. Lett.*, vol. 3, no. 5, pp. 517–520, Oct. 2014.

[22] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access*, vol. 2, pp. 1326–1339, 2014.

[23] V. V. Ratnam, A. F. Molisch, and G. Caire, "Capacity analysis of interlaced clustering in a distributed transmission system with/without CSIT," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2629–2641, Apr. 2016.

[24] D. Liu, S. Han, C. Yang, and Q. Zhang, "Semi-dynamic user-specific clustering for downlink cloud radio access network," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2063–2077, Apr. 2016.

[25] S. A. Ramprashad, G. Caire, and H. C. Papadopoulos, "A joint scheduling and cell clustering scheme for MU-MIMO downlink with limited coordination," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2010, pp. 1–6.

[26] M. Jaber, M. A. Imran, R. Tafazolli, and A. Tukmanov, "5G backhaul challenges and emerging research directions: A survey," *IEEE Access*, vol. 4, pp. 1743–1766, 2016.

[27] S. Chia, M. Gasparroni, and P. Brick, "The next challenge for cellular networks: Backhaul," *IEEE Microw. Mag.*, vol. 10, no. 5, pp. 54–66, Aug. 2009.

[28] O. Tipmongkolsilp, S. Zaghloul, and A. Jukan, "The evolution of cellular backhaul technologies: Current issues and future trends," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 1, pp. 97–113, May 2011.

[29] H. Raza, "A brief survey of radio access network backhaul evolution: Part II," *IEEE Commun. Mag.*, vol. 51, no. 5, pp. 170–177, May 2013.

[30] J. Zhao, T. Q. S. Quek, and Z. Lei, "Coordinated multipoint transmission with limited backhaul data transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2762–2775, Jun. 2013.

[31] S. Boyd, L. Xiao, A. Mutapcic, and J. Mattingley, "Notes on decomposition methods," Stanford Univ., Stanford, CA, USA, Tech. Note EE364B, 2007, pp. 1–36.

[32] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *Proc. IEEE*, vol. 95, no. 1, pp. 255–312, Jan. 2007.

[33] G. B. Dantzig and P. Wolfe, "Decomposition principle for linear programs," *Oper. Res.*, vol. 8, no. 1, pp. 101–111, 1960.

[34] C. Pan, H. Zhu, N. J. Gomes, and J. Wang, "Joint precoding and RRH selection for user-centric green MIMO C-RAN," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2891–2906, May 2017.

[35] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted $\ell 1$ minimization," *J. Fourier Anal. Appl.*, vol. 14, nos. 5–6, pp. 877–905, 2008.

[36] B. K. Sriperumbudur, D. A. Torres, and G. R. G. Lanckriet, "A majorization-minimization approach to the sparse generalized eigenvalue problem," *Mach. Learn.*, vol. 85, nos. 1–2, pp. 3–39, Oct. 2011.

[37] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[38] K. Shen and W. Yu, "Fractional programming for communication systems—Part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, May 2018.

[39] M. Grant, S. Boyd, and Y. Ye, "CVX: MATLAB software for disciplined convex programming," Tech. Rep., 2008.

[40] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, nos. 1–3, pp. 193–228, Nov. 1998.

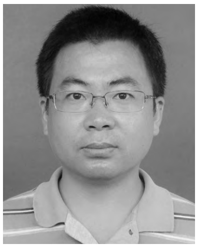
[41] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[42] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[43] Y. Ye, "Interior point algorithms: Theory and analysis," *SIAM Rev.*, vol. 41, no. 2, pp. 409–411, 1999.

XINJIANG XIA received the B.S. and M.S. degrees in communication and information systems from Hohai University, China, in 2009 and 2012, respectively. He is currently pursuing the Ph.D. degree with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. From 2012 to 2016, he was a Software Development Engineer with China Postal Express & Logistics, Nanjing. His research interests include massive multiple-input multiple-output, signal processing, and distributed antenna systems.

YU ZHANG received the B.S. degree in electrical engineering from Soochow University, in 2010, and the M.S. degree in electrical engineering from the Nanjing University of Aeronautics and Astronautics, in 2014. He is currently pursuing the Ph.D. degree with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. From 2014 to 2015, he was a Baseband Software Development Engineer with ZTE Corporation, Shanghai. In 2017, he joined the University of Victoria, Canada, as a Visiting Ph.D. Student. His current research interests include the areas of signal processing and wireless communications with emphasis on channel estimation and hybrid precoding in millimeter-wave systems.

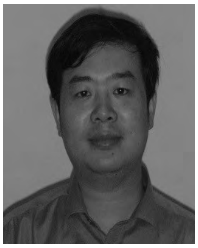


JIAMIN LI received the B.S. and M.S. degrees in communication and information systems from Hohai University, Nanjing, China, in 2006 and 2009, respectively, and the Ph.D. degree in information and communication engineering from Southeast University, Nanjing, in 2014. He has been a Lecturer with the National Mobile Communications Research Laboratory, Southeast University, since 2014. His research interests include massive multiple-input multiple-output, distributed antenna systems, and cooperative communications.

PENGCHENG ZHU (M'10) received the B.S. and M.S. degrees in electrical engineering from Shandong University, Jinan, China, in 2001 and 2004, respectively, and the Ph.D. degree in communication and information science from Southeast University, Nanjing, China, in 2009. He has been a Lecturer with the National Mobile Communications Research Laboratory, Southeast University, since 2009. His research interests include the areas of communication and signal processing, including limited feedback techniques, and distributed antenna systems.

YUANXUE XIN received the B.S. and M.S. degrees in communication and information systems from Hohai University, China, in 2009 and 2012, respectively, and the Ph.D. degree from Southeast University, China, in 2017. She joined the College of IoT Engineering, Hohai University. Her current research interests include massive multiple-input multiple-output systems and energy-efficient system design.

DONGMING WANG (M'06) received the B.S. degree from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 1999, the M.S. degree from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2002, and the Ph.D. degree from Southeast University, Nanjing, in 2006. In 2006, he joined the National Mobile Communications Research Laboratory, Southeast University, where he has been an Associate Professor, since 2010. His research interests include turbo detection, channel estimation, distributed antenna systems, and large-scale multiple-input multiple-output systems.



XIAOHU YOU (F'11) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Nanjing Institute of Technology, Nanjing, China, in 1982, 1985, and 1989, respectively. From 1987 to 1989, he was a Lecturer with the Nanjing Institute of Technology. Since 1990, he has been with Southeast University, first as an Associate Professor and later as a Professor. He is the Chief of the Technical Group of the China 3G/B3G Mobile Communication Research and Development Project. His research interests include mobile communications, adaptive signal processing, and artificial neural networks with applications to communications and biomedical engineering. He received the Excellent Paper Prize from the China Institute of Communications, in 1987, and the Elite Outstanding Young Teacher Award from Southeast University, in 1990, 1991, and 1993. He was also a recipient of the 1989 Young Teacher Award of the Fok Ying Tung Education Foundation, State Education Commission of China.

