

Joint Fronthaul Link Selection and Transmit Precoding for Energy Efficiency Maximization of Multiuser MIMO-Aided Distributed Antenna Systems

Hong Ren, *Student Member, IEEE*, Nan Liu, Cunhua Pan, *Member, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

Abstract—We jointly select the fronthaul links and optimize the transmit precoding matrices for maximizing the energy efficiency of a multiuser multiple-input multiple-output aided distributed antenna system. The fronthaul link's power consumption is taken into consideration, which is assumed to be proportional to the number of active fronthaul links quantified by using indicator functions. Both the rate requirements and the power constraints of the remote access units are considered. Under realistic power constraints some of the users cannot be admitted. Hence, we formulate a two-stage optimization problem. In Stage I, a novel user selection method is proposed for determining the maximum number of admitted users. In Stage II, we deal with the energy efficiency optimization problem. First, the indicator function is approximated by a smooth concave logarithmic function. Then, a triple-layer iterative algorithm is proposed for solving the approximated energy efficiency optimization problem, which is proved to converge to the Karush-Kuhn-Tucker conditions of the smoothed energy efficiency optimization problem. To further reduce the complexity, a single-layer iterative algorithm is conceived, which guarantees convergence. Our simulation results show that the proposed user selection algorithm approaches the performance of the exhaustive search method. Finally, the proposed algorithms is capable of achieving an order of magnitude higher energy efficiency than its conventional counterpart operating without considering link selection.

Index Terms—Distributed antenna system, fronthaul link selection, user selection, MIMO, energy efficiency.

I. INTRODUCTION

Distributed antenna systems (DASs) are capable of enhancing the bandwidth efficiency, whilst improving the link reliability [1], [2]. In a DAS, the remote access units (RAUs) are spatially distributed in a cell and each RAU is connected to

the central processing unit (CPU) through high-speed optical fibers. The links between the CPU and the RAUs are usually termed as fronthaul links. Through this distributed deployment, the average distance between the users and the RAUs can be significantly reduced, which leads to reduced transmit power. At the same time, large-scale multiple-input multiple-output (MIMO) systems have received extensive interests, since they can mitigate the hostile channel effects, such as fast fading [3]. However, deploying a large number of densely-packed antennas results in correlated signals, which will significantly limit the attainable system performance [4]. Hence, placing the antennas sufficiently far apart is important for eliminating their correlations.

The energy efficiency (EE) of wireless systems has attracted substantial research attention [5] due to energy shortage and greenhouse effects, and it is one of the key performance metrics in next-generation mobile networks [6]. The EE is usually defined as the ratio of the total power consumption to the total capacity or in a reciprocal for us as the number of successfully transmitted bits per Joule of energy consumption. Recently, the EE of DASs has also been extensively studied. But most of the existing literature has been predominantly focused on power allocation for *single-antenna* aided DASs [7]–[13]. Nevertheless, there are a few papers studying the EE optimization problem of *multiple-antenna* assisted DASs as well, which aimed for designing beamforming for the users [14]–[16]. More specifically, Kim *et al.* [14] presented joint beamforming and power allocation optimization for a single-user multiple-input single-output (MISO) DAS, in which they proved that - as expected - the optimal beam-directions should be matched to the channel state information (CSI) and the globally optimal power allocation was obtained by satisfying the Karush-Kuhn-Tucker (KKT) conditions of the problem. We then extended the solutions in [10] to the scenario, where both the RAUs and the users are equipped with multiple antennas for spatial multiplexing by additionally considering rate constraints and RAU selection [15]. The EE optimization problem of a multicast MISO DAS was studied by Ren *et al.* in [16], where an iterative algorithm was proposed by alternately optimizing the power allocation and beam-directions. However, the authors of [14]–[16] focused their attention on single-user or single-group scenarios, and the benefits of multiple antennas have not been fully exploited, since multi-antenna systems can simultaneously support multiple users for improv-

This paper has been accepted in IEEE Transactions on Communications. This work is partially supported by the National Natural Science Foundation of China under Grants 61571123 and 61521061, the National 863 Program (2014AA01A702), and the Research Fund of National Mobile Communications Research Laboratory, Southeast University (No. 2017A03), China Scholarship Council (CSC), the Scientific Research Foundation of Graduate School of Southeast University under Grant YBJJ1654, and Program Sponsored for Scientific Innovation Research of College Graduate in Jiang-su Province under Grant KYLX16_0223.

H. Ren and N. Liu are with National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China. (Email: {renhong, nanliu}@seu.edu.cn).

C. Pan is with School of Electronic Engineering and Computer Science, Queen Mary University of London, London, E1 4NS, U.K. (Email: c.pan@qmul.ac.uk).

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, U.K. (email: lh@ecs.soton.ac.uk).

ing the bandwidth efficiency. Recently, Yoon *et al.* [17] studied the EE optimization problem for a dynamic time division duplex (TDD) multiuser MISO DAS, where the downlink and uplink beamforming weights are alternatively optimized. Yoon *et al.* [17] directly applied the classic fractional programming method to the EE function, although the denominator of the EE function is non-convex. Hence, the convergence of this algorithm cannot be guaranteed. Furthermore, Yoon *et al.* [17] only considered the case, where each user is equipped with a single antenna, which can be solved by the iterative beamforming algorithm of Venturino *et al.* [18]. Unfortunately, their iterative algorithm cannot be readily extended to the scenario, where the users have multiple antennas.

Against this background, in this paper, we consider the EE maximization problem of a multiuser MIMO DAS, where both the RAUs and the users are equipped with multiple antennas. In contrast to the centralized antenna systems (CAS), in a DAS each user can adaptively select its serving RAUs for additionally reducing the power consumption. More specifically, if all RAUs are transmitting data for each user, the power consumption of the fronthaul links may become excessive. Additionally, the RAUs that are far away from the user may contribute little to the user's received power due to the large path loss, yet they incur significant data delivery power consumption on the fronthaul links. Thus, fronthaul link selection constitutes an important issue to consider in the EE maximization problem of a multiuser MIMO DAS. To the best of our knowledge, the problem of fronthaul link selection has not been considered in the EE maximization of a DAS, although, it has been considered in papers that aim for minimizing the network's power consumption supported by a DAS [19]–[22]. In these papers, the minimum power consumption of the network is achieved when the users' optimal rates are equal to their minimum rate requirements, which simplifies the analysis. Unfortunately, when maximizing the EE, the optimal rates derived for the users are not necessarily equal to their minimum rate requirements, which complicates the analysis and needs further investigation. Moreover, the authors of [19]–[22] assumed that each user is equipped with a single antenna and the techniques developed in these papers cannot be directly applied to the multi-antenna user case. The reason is that in the single-antenna user case, the rate expression can be transformed into a second-order cone programming (SOCP) problem, which is a convex problem. By contrast, the rate expression of the multi-antenna user scenario is more complicated and hence cannot be transformed into an SOCP problem. New techniques have to be developed to deal with the multiple-antenna user case. Thus, we believe that this is the first paper that considers fronthaul link selection in maximizing the EE of a multiuser MIMO DAS.

The other avenue to improve the EE of the DAS is to control the operating status of RAUs by switching off some RAUs having a low traffic load. Some related contributions can be found in [23]–[25]. Jiang *et al.* considered the EE maximization problem of a simple MIMO point-to-point system. Only a single data stream is transmitted with the aid of simple maximal ratio transmission (MRT), while maximal ratio combining (MRC) is applied at the transmitter and

receiver. Both an exhaustive search method and an antenna ordering algorithm were proposed for maximizing the EE. The total network power consumption minimization problem was considered in [24] for a MISO multi-user C-RAN, and EE maximization was carried out in [25] for MISO single-cell multiple-user networks, where both RAU selection and antenna selection were considered. However, the RAU/antenna selection method is different from our fronthaul link selection, because when a RAU is switched off, all the links supported by this RAU will be disconnected. Hence, the beamforming vectors at this RAU for all users will become zero vectors. Hence, the RAU selection or antenna selection problems of [24], [25] can be reformulated as a group sparsity technique known from compression sensing in order to transform the original integer optimization problem into a more readily solvable problem. However, in this paper, when the link spanning from a user to a RAU is switched off, the other users can still be connected to this RAU. Hence, our optimization problem cannot be formulated as a group sparsity structure and thus the methods of [24], [25] cannot be adopted. In this paper, we do not consider the RAU selection issue, since it is impractical to switch off the RAUs and to wake up them in a very short time [22]. The decision of the “on/off” status of RAUs should be made on a longer time scale than that considered in this paper. On the other hand, the data routing or RAU-user association can be promptly changed, which is considered in this paper. However, the algorithm proposed in this paper is sufficiently general to incorporate the sleep mode capability of RAUs, similarly to [20].

A. Other Related Works

Although the EE maximization problem has been extensively studied for multiple-antenna systems other than the DASs, the methods developed cannot be directly applied to the problem studied in this paper. More specifically, the EE optimization problem of MIMO broadcast channels (BCs) was studied in [26], where the duality between the BC and the multiple access channels (MAC) was exploited for transforming the original optimization problem into a more tractable problem. Unfortunately, due to the power constraints imposed on the RAUs in DAS, the duality method cannot be readily extended to the MIMO multiuser DAS scenario. He *et al.* [27] studied the EE optimization problem of the multicell MISO systems, and proposed a two-layer optimization scheme to solve it. Then, they extended [27] to heterogeneous multicell systems in [28] by incorporating the user-rate constraints. However, the fronthaul link selection was not considered in [28], and each user was assumed to be equipped with only a single antenna. Furthermore, even though naturally the power constraints and rate requirements are in conflict with each other, the feasibility issues were not analyzed in [28], rather it was always assumed that the original problem is feasible. Transmit precoder matrices were optimized for maximizing the EE of the family of MIMO interference channels in [29] under rate constraints, where an iterative algorithm was proposed for maximizing the lower bound of the EE. However, link selection did not have to be considered in [29], since each user

is served by a dedicated transmitter. Furthermore, the power constraints at each transmitter were not considered. To check the feasibility of the problem, a heuristic method based on the interference alignment technique was proposed in [29] under the assumption of having infinite transmit power, which is not realistic. References [30] and [31] adopted the interference alignment technique and the zero-forcing method to eliminate the multiuser interference, respectively. Then, the original beam-vector optimization problems were transformed into a more tractable scalar power allocation optimization problem. Since the beam directions in [30] and [31] are heuristically chosen and not jointly optimized with the power allocation, the algorithms developed in [30] and [31] will suffer from a substantial EE loss. Furthermore, neither the feasibility nor the link selection issues were addressed in [30] and [31].

B. Main Contributions

Again, we consider the EE maximization problem of a multiuser MIMO DAS by jointly selecting fronthaul links and optimizing the precoding matrices, subject to the users' data rate requirements and per-RAU power constraints. Due to the power constraints, not all users' rate requirements may be satisfied. Hence, some users should be removed or rescheduled for the next transmission time slots. We solve this optimization problem by decomposing it into a two-stage optimization problem: in Stage I, the user selection is optimized, by finding the largest subset of users that can be admitted into the network. This deals with the feasibility issue of the optimization problem, which was not considered in [7]–[17]; in Stage II, the EE optimization problem is dealt with by jointly optimizing the link selection and the transmit precoding (TPC) matrices for the users selected in Stage I, where each user is equipped with multiple antennas. To the best of our knowledge, there is no contribution in the open literature, dealing with the EE maximization problem, not even for a *single-antenna* aided user scenario in conjunction with fronthaul link selection. The main reason is that the associated optimization problem cannot be transformed into an SOCP problem for solving the power minimization problem. Specifically, for the power minimization problem, the optimal rate of each user is equal to its specific rate requirement, which can be transformed to an SOCP problem. By contrast, for the EE maximization problem, the optimal rate is generally not equal to the rate requirement, which hence cannot be transformed into an SOCP problem. Actually, the EE maximization may be deemed to be reminiscent of the rate maximization problem, which has been proved to be NP-hard [32].

The main contributions of this paper are summarized as follows:

- 1) In Stage I, a beneficial user selection algorithm is proposed for determining the maximum number of users that can be supported. In each iteration of the algorithm, we solve an alternative problem with the aid of several additional auxiliary variables. This alternative problem is always feasible and the auxiliary variables can be regarded as the admission indicators of the users. By replacing the rate expression with its lower bound, an

iterative so-called block descent algorithm is proposed for solving this alternative problem. In each iteration of the algorithm, the optimal auxiliary variables and precoder matrices are obtained in closed form by employing the Lagrangian dual method. The algorithm is formally shown to converge.

- 2) In Stage II, we first approximate the indicator function by a smooth concave logarithmic function, and then conceive a pair of algorithms for solving the approximated EE optimization problem, namely, a triple-layer and a reduced-complexity single-layer iterative algorithm. We will show that these two algorithms have almost the same performance, and they have a significantly higher EE than the existing rate maximization method, the power minimization method and the EE maximization method operating without link selection. Specifically, in some cases the proposed algorithms achieve an order of magnitude higher EE than the existing methods.

The remainder of this paper is organized as follows. Section II introduces the system model, including the user selection problem of Stage I and the EE maximization problem of Stage II. Section III considers the user selection algorithm, while Section IV develops the iterative triple-layer and single-layer algorithms to solve the EE maximization problem. Our simulation results are provided in Section V and our conclusions are drawn in Section VI.

II. SYSTEM MODEL

A. System model

Consider the downlink of a single-cell distributed antenna system (DAS) relying on I RAUs jointly serving K users, where each RAU has M antennas and each user has N antennas. Let us denote the set of RAUs and users as $\mathcal{I} = \{1, \dots, I\}$ and $\mathcal{U} = \{1, \dots, K\}$, respectively. As shown in Fig. 1, we assume that all the RAUs are connected to the central processing unit (CPU) through a high speed fiber-optic fronthaul and all the user data and channel state information (CSI) are available at the CPU.

Let $\mathcal{S} = \{u_1, \dots, u_{|\mathcal{S}|}\} \subseteq \mathcal{U}$ represent the subset of users that are selected for transmission in the DAS. The number of data streams transmitted to each user $k \in \mathcal{S}$, is denoted by d , where d satisfies $d \leq \min\{MI, N\}$. Let $\mathbf{V}_{i,k} \in \mathbb{C}^{M \times d}$ denote the TPC matrix used by the i th RAU to transmit the data vector $\mathbf{s}_k \in \mathbb{C}^{d \times 1}$ to user k , with $\mathbb{E}[\mathbf{s}_k \mathbf{s}_k^H] = \mathbf{I}_d$. Then we introduce a network-wide TPC matrix $\mathbf{V}_k = [\mathbf{V}_{1,k}^H, \mathbf{V}_{2,k}^H, \dots, \mathbf{V}_{I,k}^H]^H \in \mathbb{C}^{M_t \times d}$ for each user $k \in \mathcal{S}$, where $M_t = MI$.

Upon using the linear TPC matrices at all the RAUs, the signal received by user k , denoted as $\mathbf{y}_k \in \mathbb{C}^{N \times 1}$, is given by

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{V}_k \mathbf{s}_k + \sum_{j \in \mathcal{S}, j \neq k} \mathbf{H}_k \mathbf{V}_j \mathbf{s}_j + \mathbf{n}_k, \quad (1)$$

where $\mathbf{H}_k \in \mathbb{C}^{N \times M_t}$ the CSI matrix representing the links between all the RAUs and user k , \mathbf{n}_k denotes the noise at user k , which is an additive Gaussian noise vector subject to $\mathcal{CN}(\mathbf{0}, \sigma_k^2 \mathbf{I}_N)$. We assume that the different users have independent data and receiver noise processes.

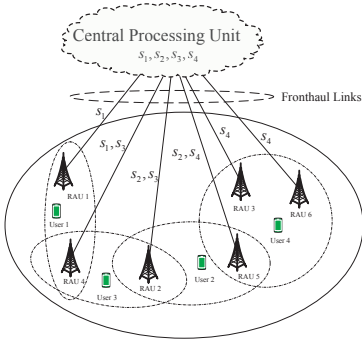


Fig. 1. Example of DAS with six RAUs and four users. Each user only selects its nearby RAUs for transmission, which is named link selection. In this example, user 1 selects link 1 and link 4, user 2 selects link 2 and link 5, user 3 selects link 2 and link 3, user 4 selects link 3, link 5 and link 6.

Assuming single-user detection at each user, the achievable rate of user k (nat/s/Hz) can be written as [33]

$$R_k(\mathbf{V}) = \log |\mathbf{I} + \mathbf{H}_k \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{J}_k^{-1}|, \quad (2)$$

where $\mathbf{J}_k = \sum_{j \neq k, j \in \mathcal{S}} \mathbf{H}_k \mathbf{V}_j \mathbf{V}_j^H \mathbf{H}_k^H + \sigma_k^2 \mathbf{I}$ is the interference-plus-noise covariance matrix of user k , and $\mathbf{V} = [\mathbf{V}_1^H, \mathbf{V}_2^H, \dots, \mathbf{V}_K^H]^H \in \mathbb{C}^{M_t K \times d}$. Each user k , $k = 1, \dots, K$, should satisfy its rate requirement $R_{k,\min}$, which is formulated as:

$$R_k(\mathbf{V}) \geq R_{k,\min}, \forall k \in \mathcal{S}. \quad (3)$$

To solve our EE optimization problem, the total power consumption of the DAS should be considered, which mainly includes the transmit power consumption of the RAUs, the fronthaul link power required for transmitting data to the users and the signal processing power consumption of the CPU. Hence, the total power consumption is modeled as

$$P_{\text{total}}(\mathbf{V}) = P_{\text{Tr}}(\mathbf{V}) + P_C + P_F(\mathbf{V}), \quad (4)$$

where $P_{\text{Tr}}(\mathbf{V})$ denotes the total transmit power of RAUs given by $P_{\text{Tr}}(\mathbf{V}) = \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{S}} \|\mathbf{V}_{i,k}\|_F^2$, where $\|\cdot\|_F$ represents the Frobenius norm, P_C denotes the total fixed power consumption of the DAS system that includes both the circuit power consumption at the RAUs and the signal processing power consumption at the CPU. Finally, $P_F(\mathbf{V})$ denotes the fronthaul power consumption, which is proportional to the number of active links [19] and it is given by

$$P_F(\mathbf{V}) = \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{S}} \varepsilon(\|\mathbf{V}_{i,k}\|_F^2) P_{\text{fh}}, \quad (5)$$

with P_{fh} being the fronthaul power required for forwarding data to each user at each fronthaul link and $\varepsilon(\cdot)$ is the link-selection indicator function defined as

$$\varepsilon(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

B. Problem formulation

In practical systems, each RAU has its own power constraint. Due to this kind of constraints, the DAS may not always be able to support all the users' transmission. As a result, some users should be removed or rescheduled for transmission on

the other orthogonal slots. Thus, we propose to solve a two-stage optimization problem. The first stage finds the largest subset of users that can be supported by the DAS. The second stage optimizes the TPC matrices for maximizing the energy efficiency of the DAS for the users selected during the first stage.

Specifically, the optimization problem in Stage I can be formulated as

$$\max_{\mathbf{V}, \mathcal{S} \subseteq \mathcal{U}} |\mathcal{S}| \quad (7a)$$

$$\text{s.t. } R_k(\mathbf{V}) \geq R_{k,\min}, \forall k \in \mathcal{S}, \quad (7b)$$

$$\sum_{k \in \mathcal{S}} \|\mathbf{V}_{i,k}\|_F^2 \leq P_{i,\max}, \forall i \in \mathcal{I}, \quad (7c)$$

where $P_{i,\max}$ is the power limit of the i th RAU.

In the second stage, we aim for jointly selecting the active links for the users selected during Stage I and optimize the TPC matrices for maximizing the energy efficiency of the DAS system, which is defined as the ratio of the sum rate to the total power consumption, while guaranteeing all RAUs' power constraints and all the selected users' minimum rate requirements. The optimization problem of Stage II is formulated as follows:

$$\max_{\mathbf{V}} \frac{\sum_{k \in \mathcal{S}^*} R_k(\mathbf{V})}{P_{\text{total}}(\mathbf{V})} \quad (8a)$$

$$\text{s.t. } R_k(\mathbf{V}) \geq R_{k,\min}, \forall k \in \mathcal{S}^*, \quad (8b)$$

$$\sum_{k \in \mathcal{S}^*} \|\mathbf{V}_{i,k}\|_F^2 \leq P_{i,\max}, \forall i \in \mathcal{I}, \quad (8c)$$

where \mathcal{S}^* is the optimal solution found during Stage I.

The user selection problem may be viewed as a combinatorial optimization problem, which is NP-hard. Additionally, the non-smooth link-selection indicator functions in the objective function (OF) of the optimization problem in Stage II makes the problem also NP-hard. In the following two sections, we propose a pair of efficient algorithms for solving the above two optimization problems.

III. STAGE I: USER SELECTION ALGORITHM

In this section, we propose a novel user selection algorithm for maximizing the number of admitted users. Inspired by Matskani *et al.* [34], we construct the following optimization problem for a given subset of users \mathcal{S} :

$$\min_{\{\alpha_k\}_{k \in \mathcal{S}}, \mathbf{V}} \sum_{k \in \mathcal{S}} (\alpha_k - 1)^2 \quad (9a)$$

$$\text{s.t. } R_k(\mathbf{V}) \geq \alpha_k^2 R_{k,\min}, \forall k \in \mathcal{S}, \quad (9b)$$

$$\sum_{k \in \mathcal{S}} \|\mathbf{V}_{i,k}\|_F^2 \leq P_{i,\max}, \forall i \in \mathcal{I}, \quad (9c)$$

where $\{\alpha_k\}_{k \in \mathcal{S}}$ represent the introduced auxiliary variables. Note that the problem in (9) is always feasible. It can be seen that the optimal α_k cannot be larger than one. This can be readily proved by the method of contradiction. On the other hand, user k can only be admitted if the optimal α_k is equal to one. By setting as many α_k values to ones as possible, the number of admitted users is maximized. It is worth noting that if all the α_k 's are equal to ones, then all the users in \mathcal{S} can be admitted to the DAS.

Based on the above analysis, we formally state our user selection procedure in Algorithm 1 to solve the problem in (9). The main core of the user selection algorithm is an iterative procedure, where in each iteration the user associated with the least $\alpha_k < 1$ will be removed until all the optimal α_k s become equal to ones. The intuition behind removing the user having the lowest α_k is that this user has the largest discrepancy with respect to its target rate and more resources will be required for satisfying this user's rate requirement. Recall that there is a total of K users. For the worst case scenario, when no user can be supported, Problem (9) will be solved K times. Hence, the number of times we have to solve Problem (9) increases linearly with K . By contrast, for the exhaustive search method that searches all feasible sets of users and selects the largest one, the total number of solving Problem (9) increases exponentially with K , which has a much higher complexity than Algorithm 1. However, we will demonstrate by simulations that the performance gain of the exhaustive search method over the proposed algorithm is marginal.

Algorithm 1 User Selection Algorithm

- 1: Initializing the set of selected users $\mathcal{S} = \{1, \dots, K\}$;
 - 2: With the user candidate set \mathcal{S} , solve Problem (9) by using the block descent algorithm of Subsection III-A to obtain $\{\alpha_k\}_{k \in \mathcal{S}}$ and \mathbf{V} .
 - 3: If $\alpha_k = 1, \forall k \in \mathcal{S}$, generate the output \mathbf{V} and set $\mathcal{S}^* = \mathcal{S}$ for the initialization of Stage II, then terminate; Otherwise, find $k^* = \arg \min_{k \in \mathcal{S}} \alpha_k$, remove user k^* by updating $\mathcal{S} = \mathcal{S} \setminus k^*$ and go to step 2.
-

A. Algorithm to solve Problem (9)

Without loss of generality, it is assumed that \mathcal{S} is given by $\mathcal{S} = \{1, \dots, S\}$, where S denotes the number of selected users. To solve Problem (9), we first employ the weighted minimum mean square error (WMMSE) method [35] to transform the rate expression into a more tractable form. Then, the Lagrange method [36] is invoked for obtaining the closed-form of the TPC matrices.

More specifically, the linear receiver filter $\mathbf{U}_k \in \mathbb{C}^{N \times d}$ is invoked for decoding the received signal vector of user k , which yields

$$\hat{\mathbf{s}}_k = \mathbf{U}_k^H \mathbf{y}_k, \forall k \in \mathcal{S}. \quad (10)$$

By using the assumption that the signal vectors \mathbf{s}_k and noise vectors \mathbf{n}_k are independent, the mean square error (MSE) matrix at the k th user can be calculated as

$$\begin{aligned} \mathbf{E}_k &= \mathbb{E}_{\mathbf{s}, \mathbf{n}} \left[(\hat{\mathbf{s}}_k - \mathbf{s}_k) (\hat{\mathbf{s}}_k - \mathbf{s}_k)^H \right] \\ &= (\mathbf{U}_k^H \mathbf{H}_k \mathbf{V}_k - \mathbf{I}_d) (\mathbf{U}_k^H \mathbf{H}_k \mathbf{V}_k - \mathbf{I}_d)^H \\ &\quad + \sum_{j \neq k, j \in \mathcal{S}} \mathbf{U}_k^H \mathbf{H}_k \mathbf{V}_j \mathbf{V}_j^H \mathbf{H}_k^H \mathbf{U}_k + \sigma_k^2 \mathbf{U}_k^H \mathbf{U}_k. \end{aligned} \quad (11)$$

By introducing a set of auxiliary matrices $\{\mathbf{W}_k \succeq \mathbf{0}\}$, we define the following functions

$$h_k(\mathbf{V}, \mathbf{U}_k, \mathbf{W}_k) = \log |\mathbf{W}_k| - \text{Tr}(\mathbf{W}_k \mathbf{E}_k) + d, \forall k \in \mathcal{S}. \quad (12)$$

Based on the above definitions, we provide the following lemma, whose proof can be found in [35].

Lemma 1: We have the following results:

- 1) Given any two matrices of \mathbf{V} , \mathbf{U}_k and \mathbf{W}_k , $h_k(\mathbf{V}, \mathbf{U}_k, \mathbf{W}_k)$ is a concave function of the remaining matrix.
- 2) When the TPC matrices \mathbf{V} are fixed, the data rate $R_k(\mathbf{V})$ in (2) is lower-bounded by $h_k(\mathbf{V}, \mathbf{U}_k, \mathbf{W}_k)$, for any $\mathbf{U}_k, \mathbf{W}_k$. i.e., we have:

$$R_k(\mathbf{V}) \geq h_k(\mathbf{V}, \mathbf{U}_k, \mathbf{W}_k), \quad \forall \mathbf{U}_k, \mathbf{W}_k \succeq \mathbf{0};$$

- 3) Given \mathbf{V} , consider the following problem

$$\max_{\mathbf{U}_k, \mathbf{W}_k \succeq \mathbf{0}} h_k(\mathbf{V}, \mathbf{U}_k, \mathbf{W}_k); \quad (13)$$

The optimal \mathbf{U}_k and \mathbf{W}_k that solves the maximization problem in (13) are given by

$$\begin{aligned} \mathbf{U}_k^* &= \left(\sum_{j=1}^K \mathbf{H}_k \mathbf{V}_j \mathbf{V}_j^H \mathbf{H}_k^H + \sigma_k^2 \mathbf{I} \right)^{-1} \mathbf{H}_k \mathbf{V}_k, \quad (14) \\ \mathbf{W}_k^* &= \mathbf{E}_k^{*-1}, \forall k \in \mathcal{S}, \end{aligned} \quad (15)$$

where \mathbf{E}_k^* is obtained by inserting \mathbf{U}_k^* into user k 's MSE in (11), yielding

$$\mathbf{E}_k^* = \mathbf{I}_d - \mathbf{V}_k^H \mathbf{H}_k^H \left(\sum_{j=1}^K \mathbf{H}_k \mathbf{V}_j \mathbf{V}_j^H \mathbf{H}_k^H + \sigma_k^2 \mathbf{I} \right)^{-1} \mathbf{H}_k \mathbf{V}_k. \quad (16)$$

Furthermore, we have

$$h_k(\mathbf{V}, \mathbf{U}_k^*, \mathbf{W}_k^*) = R_k(\mathbf{V}). \quad (17)$$

□

By adopting Lemma 1, the non-convex rate constraints in (9b) of the original Problem (9) can be effectively handled. Specifically, by replacing the rate expression in (9b) with its lower-bound $h_k(\mathbf{V}, \mathbf{U}_k, \mathbf{W}_k)$, we have the following optimization problem

$$\min_{\{\alpha_k\}_{k \in \mathcal{S}}, \mathbf{V}, \mathbf{U}, \mathbf{W}} \sum_{k \in \mathcal{S}} (\alpha_k - 1)^2 \quad (18a)$$

$$\text{s.t.} \quad h_k(\mathbf{V}, \mathbf{U}_k, \mathbf{W}_k) \geq \alpha_k^2 R_{k, \min}, \forall k \in \mathcal{S}, \quad (18b)$$

$$\sum_{k \in \mathcal{S}} \|\mathbf{V}_{i, k}\|_F^2 \leq P_{i, \max}, \forall i \in \mathcal{I}, \quad (18c)$$

where \mathbf{U} and \mathbf{W} are the collection of matrices $\mathbf{U}_k, \forall k \in \mathcal{S}$ and $\mathbf{W}_k, \forall k \in \mathcal{S}$, respectively.

By combining the first conclusion of Lemma 1 and the fact that $\alpha_k^2 R_{k, \min}, \forall k$ is a convex function of $\alpha_k, \forall k$, the constraints (18b) are convex when fixing \mathbf{U} and \mathbf{W} . Hence, given \mathbf{U} and \mathbf{W} , Problem (18) is a convex optimization problem constructed for the TPC \mathbf{V} , since the OF in (18) is a convex function and the feasible set is a convex set. Then, given the TPC \mathbf{V} , the auxiliary matrices \mathbf{U} and \mathbf{W} can be updated by using (14) and (15). The above observations motivate the application of the block coordinate descent method for solving Problem (18).

Given \mathbf{U} and \mathbf{W} , Problem (18) becomes

$$\min_{\{\alpha_k\}_{k \in \mathcal{S}}, \mathbf{V}} \sum_{k \in \mathcal{S}} (\alpha_k - 1)^2 \quad (19a)$$

$$\text{s.t.} \quad \sum_{j \in \mathcal{S}} \text{Tr}(\mathbf{V}_j^H \mathbf{H}_k^H \mathbf{U}_k \mathbf{W}_k \mathbf{U}_k^H \mathbf{H}_k \mathbf{V}_j) - \text{Tr}(\mathbf{W}_k \mathbf{U}_k^H \mathbf{H}_k \mathbf{V}_k) - \text{Tr}(\mathbf{V}_k^H \mathbf{H}_k^H \mathbf{U}_k \mathbf{W}_k) + \alpha_k^2 R_{k,\min} \leq t_k, \forall k \in \mathcal{S}, \quad (19b)$$

$$\sum_{k \in \mathcal{S}} \text{Tr}(\mathbf{V}_{i,k}^H \mathbf{V}_{i,k}) \leq P_{i,\max}, \forall i \in \mathcal{I}, \quad (19c)$$

where the MSE expression in (11) is substituted into the first set of constraints in Problem (18), $t_k = \log |\mathbf{W}_k| + d - \text{Tr}(\mathbf{W}_k) - \sigma_k^2 \text{Tr}(\mathbf{U}_k^H \mathbf{U}_k \mathbf{W}_k)$.

As argued above, Problem (19) is a convex optimization problem. Additionally, since the power constraints of all RAUs are positive, i.e., we have $P_{i,\max} > 0, \forall i \in \mathcal{I}$, Slater's condition [36] is satisfied for Problem (19). Hence, the gap between Problem (19) and its dual problem is zero [36]. In the following, we derive the closed-form solution of Problem (19) by using the Lagrange dual method. Specifically, by appropriately rearranging the terms, the Lagrangian function of Problem (19) is given by (20) at the top of the next page. where $\boldsymbol{\lambda} = \{\lambda_1, \dots, \lambda_S\}$ and $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_I\}$ are the Lagrangian multipliers associated with the first and second constraint sets of problem (19), respectively, and \mathbf{G} is given by

$$\mathbf{G} = \sum_{j \in \mathcal{S}} \lambda_j \mathbf{H}_j^H \mathbf{U}_j \mathbf{W}_j \mathbf{U}_j^H \mathbf{H}_j + \sum_{i \in \mathcal{I}} \mu_i \mathbf{B}_i,$$

with \mathbf{B}_i formulated as

$$\mathbf{B}_i = \text{diag} \left(\underbrace{0, \dots, 0}_{(i-1)M}, \underbrace{1, \dots, 1}_M, \underbrace{0, \dots, 0}_{(I-i)M} \right). \quad (21)$$

Then the dual function of Problem (19) is given by

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\mathbf{V}, \boldsymbol{\alpha}} \mathcal{L}(\mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\mu}), \quad (22)$$

where $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \dots, \alpha_S\}$, and $\mathcal{L}(\mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is given in (20). Naturally, each constraint in (19b) must be met with equality at the optimum point. Then, according to the complementary slackness condition, the optimal $\boldsymbol{\lambda}$ must be positive. As a result, the globally optimal solution to Problem (22) can be calculated as:

$$\alpha_k = \frac{1}{1 + \lambda_k R_{k,\min}}, \mathbf{V}_k = \lambda_k \mathbf{G}^{-1} \mathbf{H}_k^H \mathbf{U}_k \mathbf{W}_k, \forall k \in \mathcal{S}. \quad (23)$$

Substituting it into Problem (22) yields

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \sum_{k \in \mathcal{S}} \frac{\lambda_k R_{k,\min}}{1 + \lambda_k R_{k,\min}} - \sum_{k \in \mathcal{S}} \lambda_k C_k - \sum_{i \in \mathcal{I}} \mu_i P_{i,\max} - \sum_{k \in \mathcal{S}} \lambda_k^2 \text{Tr}(\mathbf{W}_k^H \mathbf{U}_k^H \mathbf{H}_k \mathbf{G}^{-1} \mathbf{H}_k^H \mathbf{U}_k \mathbf{W}_k). \quad (24)$$

The optimal $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ can be obtained by solving the dual problem of (19), which is given by

$$\max_{\{\lambda_k > 0, \forall k \in \mathcal{S}\}, \{\mu_i \geq 0, \forall i \in \mathcal{I}\}} g(\boldsymbol{\lambda}, \boldsymbol{\mu}). \quad (25)$$

Fortunately, the OF of Problem (25) is differentiable, hence classical descent algorithms such as the gradient descent method [36], [37] can be applied for solving it of a low computational complexity. The details are however omitted for simplicity.

Based on the above analysis, the block descent algorithm conceived for solving Problem (18) is given in Algorithm 2. Its convergence is guaranteed in the following theorem.

Theorem 1: The block descent algorithm is guaranteed to converge.

Proof: The main philosophy of the proof is to prove that in each iteration of the algorithm, the OF value is non-increasing. For more details, please see Appendix A. \square

Algorithm 2 Block Descent Algorithm

- 1: Initialize iterative number $n = 1$, the maximum number of iterations n_{\max} , any feasible precoding matrices $\mathbf{V}^{(0)}$. With fixed $\mathbf{V}^{(0)}$, compute $\mathbf{U}^{(0)}$ and $\mathbf{W}^{(0)}$ by using (14) and (14);
 - 2: With $\mathbf{U}^{(n-1)}$ and $\mathbf{W}^{(n-1)}$, update $\{\alpha_k^{(n)}\}_{k \in \mathcal{S}}$ and $\mathbf{V}^{(n)}$ by using the Lagrange dual method;
 - 3: Update $\mathbf{U}^{(n)}$ and $\mathbf{W}^{(n)}$ as in (14) and (15) with $\mathbf{V}^{(n)}$;
 - 4: If $n < n_{\max}$, set $n = n + 1$ and go to step 2; Otherwise, output $\mathbf{V}^{(n)}$ and $\{\alpha_k^{(n)}\}_{k \in \mathcal{S}}$, terminate;
-

B. Complexity analysis

In this section, we provide our complexity analysis of the block descent technique of Algorithm 1. In each iteration of Algorithm 2, the complexity is dominated by the computation of precoding matrices $\mathbf{V}^{(n)}$ by using the Lagrange dual method. In each iteration of the Lagrange dual method, the complexity is dominated by the computation of $\mathbf{V}_k^{(n)}$ in (23). Let us denote the number of remaining users as S , when executing Algorithm 2. For any two matrices $\mathbf{X} \in \mathbb{C}^{m \times n}$, $\mathbf{Y} \in \mathbb{C}^{n \times p}$, the calculation of \mathbf{XY} involves a complexity on the order of $O(mnp)$ [36]. Generally, the number of data streams is much smaller than the product of the number of RAUs and the number of transmit antennas at each RAU, i.e., we have $d \ll MI$. Hence, the computational complexity of matrix \mathbf{G} is on the order of $O(SM^2I^2d)$. According to [36], the inversion of a $(l \times l)$ -element matrix has a complexity on the order of $O(l^3)$. Hence, the complexity of computing $\mathbf{V}_k^{(n)}$ in (23) is on the order of $O(M^3I^3)$. Since $\mathbf{V}_k^{(n)}$ should be calculated for S users, the overall complexity of each iteration of the Lagrange dual method is on the order of $O(SM^3I^3)$. Given a total of $(S + I)$ dual variables, the number of iterations required for ellipsoid methods to converge is on the order of $O((S + I)^2)$ [38]. Hence, the total complexity of Algorithm 2 is on the order of $T_{\text{Alg2}} = O(t_{\text{Alg2}}(S + I)^2SM^3I^3)$, where t_{Alg2} denotes the average number of iterations required by Algorithm 2. It will be shown in the Simulation Section that Algorithm 2 always converges within 10 iterations for any simulation setups.

To provide a clear view of the main mathematical manipulations, we also post the associated flow chart in Fig. 2, which

$$\begin{aligned} \mathcal{L}(\mathbf{V}, \alpha, \lambda, \mu) = & \sum_{k \in \mathcal{S}} [\text{Tr}(\mathbf{V}_k^H \mathbf{G} \mathbf{V}_k) - \text{Tr}(\lambda_k \mathbf{W}_k \mathbf{U}_k^H \mathbf{H}_k \mathbf{V}_k) - \text{Tr}(\lambda_k \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{U}_k \mathbf{W}_k)] \\ & + \sum_{k \in \mathcal{S}} (\alpha_k - 1)^2 + \sum_{k \in \mathcal{S}} \lambda_k \alpha_k^2 R_{k,\min} - \sum_{k \in \mathcal{S}} \lambda_k t_k - \sum_{i \in \mathcal{I}} \mu_i P_{i,\max}, \end{aligned} \quad (20)$$

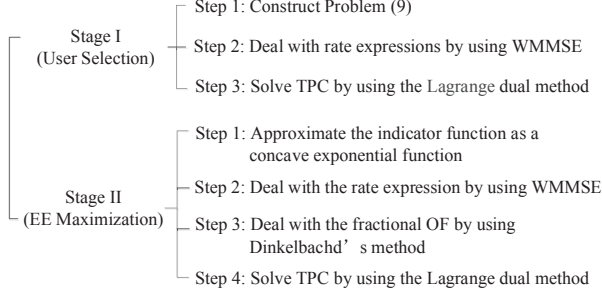


Fig. 2. The main steps and methodologies involved in Stage I and Stage II.

shows the three main techniques of Stage I and the four steps of Stage II.

IV. STAGE II: ALGORITHMS TO SOLVE PROBLEM (8)

In this section, we first propose a triple-layer iterative algorithm to solve Problem (8) for the set of users \mathcal{S}^* selected in Stage I. Then, to further reduce the complexity, we propose a single-layer iterative algorithm, which guarantees convergence. For notational simplicity, the superscript in \mathcal{S}^* is omitted and the selected users are given by $\mathcal{S} = \{1, \dots, S\}$.

A. Triple-layer Iterative Algorithm

In this subsection, we propose a triple-layer iterative algorithm to solve Problem (8), where the first layer deals with the nonsmooth nature of the link selection indicator function, the second layer deals with the nonconvex rate expression, while the third layer deals with the fractional form of the OF.

The non-smooth indicator function in the OF of Problem (8) makes the optimization problem a challenging one to solve. A potential solution is to invoke the exhaustive search method, where one checks all possible combination of selected links and opts for the specific set of links yielding the maximum EE. However, the total number of links is IK , thus the computational complexity is on the order of $O(2^{IK})$, which is unaffordable for practical applications. In the following, we propose a low-complexity algorithm to handle our non-smooth indicator function by approximating it as a concave continuous smooth function, denoted by $f(x)$. In particular, we consider the following four commonly used smooth concave functions [39], [40]:

$$f_\delta(x) = \begin{cases} \frac{x}{x+\delta}, & \text{fractional-function} \\ \frac{2}{\pi} \arctan\left(\frac{x}{\delta}\right), & \text{arctan-function} \\ 1 - \exp\left(-\frac{x}{\delta}\right), & \text{exp-function} \\ \frac{\log(x/\delta+1)}{\log(1/\delta+1)}, & \text{log-function.} \end{cases}$$

The evolution of the different functions is shown in Fig. 3. It is seen from this figure that the first three functions have

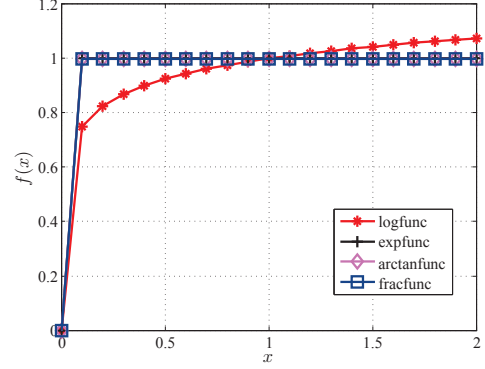


Fig. 3. Approximating the indicator function by smooth functions, where we set $\delta = 10^{-4}$.

better approximation accuracy than the log-function. However, the log-function is much more smooth than the other three functions.

If we choose a sufficiently small δ and replace the indicator functions in the OF of Problem (8) with function $f_\delta(x)$, the original EE optimization Problem (8) can be approximated as

$$\max_{\mathbf{V}} \quad \widetilde{EE}(\mathbf{V}) = \frac{\sum_{k \in \mathcal{S}} R_k(\mathbf{V})}{\tilde{P}_{\text{total}}(\mathbf{V})} \quad (26a)$$

$$\text{s.t.} \quad R_k(\mathbf{V}) \geq R_{k,\min}, \forall k \in \mathcal{S}, \quad (26b)$$

$$\sum_{k \in \mathcal{S}} \|\mathbf{V}_{i,k}\|_F^2 \leq P_{i,\max}, \forall i \in \mathcal{I}, \quad (26c)$$

where $\tilde{P}_{\text{total}}(\mathbf{V})$ is given by

$$\tilde{P}_{\text{total}}(\mathbf{V}) = \sum_{k \in \mathcal{S}} \text{Tr}(\mathbf{V}_k \mathbf{V}_k^H) + P_{\text{th}} \sum_{k \in \mathcal{S}} \sum_{i \in \mathcal{I}} f_\delta(\|\mathbf{V}_{i,k}\|_F^2) + P_C. \quad (27)$$

However, the denominator of (26a) is still concave. In order to support the application of fractional programming, we should replace these concave functions with their first-order approximations. More specifically, given an arbitrary point $\{\bar{\mathbf{V}}_{i,k}, \forall i, k\}$, due to the concavity of the function $f_\delta(x)$, we have

$$f_\delta(\|\mathbf{V}_{i,k}\|_F^2) \leq f_\delta(\|\bar{\mathbf{V}}_{i,k}\|_F^2) + f_\delta'(\|\bar{\mathbf{V}}_{i,k}\|_F^2) (\|\mathbf{V}_{i,k}\|_F^2 - \|\bar{\mathbf{V}}_{i,k}\|_F^2), \quad (28)$$

where $f_\delta'(\|\bar{\mathbf{V}}_{i,k}\|_F^2)$ denotes the first derivative of $f_\delta(x)$ at $x = \|\bar{\mathbf{V}}_{i,k}\|_F^2$.

Then we substitute the right hand side of (28) for the function $f_\delta(x)$ in the denominator of the OF in (26), and consequently Problem (8) can be successively approximated

by solving a series of the following optimization problems

$$\max_{\mathbf{V}} \quad \overline{EE}(\mathbf{V}|\bar{\mathbf{V}}) = \frac{\sum_{k \in \mathcal{S}} R_k(\mathbf{V})}{\bar{P}_{\text{total}}(\mathbf{V}|\bar{\mathbf{V}})} \quad (29a)$$

$$\text{s.t.} \quad R_k(\mathbf{V}) \geq R_{k,\min}, \forall k \in \mathcal{S}, \quad (29b)$$

$$\sum_{k \in \mathcal{S}} \|\mathbf{V}_{i,k}\|_F^2 \leq P_{i,\max}, \forall i \in \mathcal{I}, \quad (29c)$$

where $\bar{P}_{\text{total}}(\mathbf{V}|\bar{\mathbf{V}})$ is given by

$$\bar{P}_{\text{total}}(\mathbf{V}|\bar{\mathbf{V}}) = \sum_{k \in \mathcal{S}} \text{Tr}(\mathbf{V}_k \mathbf{V}_k^H) + \sum_{k \in \mathcal{S}} \sum_{i \in \mathcal{I}} \bar{\omega}_{i,k} \|\mathbf{V}_{i,k}\|_F^2 + \bar{\gamma}$$

with

$$\begin{aligned} \bar{\omega}_{i,k} &= P_{\text{fh}} f_{\delta}' \left(\|\bar{\mathbf{V}}_{i,k}\|_F^2 \right), \\ \bar{\gamma} &= P_{\text{fh}} \sum_{k \in \mathcal{S}} \sum_{i \in \mathcal{I}} \left(f_{\delta} \left(\|\bar{\mathbf{V}}_{i,k}\|_F^2 \right) - f_{\delta}' \left(\|\bar{\mathbf{V}}_{i,k}\|_F^2 \right) \|\bar{\mathbf{V}}_{i,k}\|_F^2 \right) + P_C. \end{aligned} \quad (30)$$

Note that $\bar{P}_{\text{total}}(\bar{\mathbf{V}}|\bar{\mathbf{V}}) = \bar{P}_{\text{total}}(\bar{\mathbf{V}})$.

However, Problem (29) still remains nonconvex due to the nonconvex rate expression. To handle this challenge, the relationships between the WMMSE and the rate expression given in Lemma 1 is applied again. Based on Lemma 1, we introduce the following optimization problem

$$\max_{\mathbf{V}, \mathbf{U}, \mathbf{W}} \quad \frac{\sum_{k \in \mathcal{S}} h_k(\mathbf{V}, \mathbf{U}_k, \mathbf{W}_k)}{\bar{P}_{\text{total}}(\mathbf{V}|\bar{\mathbf{V}})} \quad (31a)$$

$$\text{s.t.} \quad h_k(\mathbf{V}, \mathbf{U}_k, \mathbf{W}_k) \geq R_{k,\min}, \forall k \in \mathcal{S}, \quad (31b)$$

$$\sum_{k \in \mathcal{S}} \|\mathbf{V}_{i,k}\|_F^2 \leq P_{i,\max}, \forall i \in \mathcal{I}, \quad (31c)$$

where $h_k(\mathbf{V}, \mathbf{U}_k, \mathbf{W}_k)$ is defined in (12).

The following theorem establishes the equivalence between Problem (29) and Problem (31).

Theorem 2: Problem (29) is equivalent to Problem (31) in the following sense:

1) Denote the solution of Problem (31) as $(\mathbf{U}^*, \mathbf{V}^*, \mathbf{W}^*)$ and let $\boldsymbol{\lambda}^* = \{\lambda_{u_1}^*, \dots, \lambda_{u_S}^*\}$, $\boldsymbol{\mu}^* = \{\mu_1^*, \dots, \mu_J^*\}$ be the corresponding optimal non-negative Lagrange multipliers associated with the user-rate and per-RAU power constraints, respectively. Then the optimal $(\mathbf{U}^*, \mathbf{W}^*)$ must be given in (14) and (15). In addition, \mathbf{V}^* satisfies the Karush-Kuhn-Tucker (KKT) of Problem (29) with the Lagrange multipliers given by $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$.

2) Let \mathbf{V}^* be the solution of Problem (29) and $\boldsymbol{\lambda}^* = \{\lambda_{u_1}^*, \dots, \lambda_{u_S}^*\}$, $\boldsymbol{\mu}^* = \{\mu_1^*, \dots, \mu_J^*\}$ be the corresponding non-negative optimal Lagrange multipliers associated with the rate and power constraints, respectively. Define $(\mathbf{U}^*, \mathbf{W}^*)$ as in (14) and (15), then $(\mathbf{U}^*, \mathbf{V}^*, \mathbf{W}^*)$ satisfies the KKT conditions of Problem (31) with the Lagrange multipliers given by $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$.

Proof: Please see Appendix B. \square

The fact that Problems (29) and (31) are equivalent, as stated in Theorem 2, means that the solution of Problem (29) can be obtained by solving Problem (31). If the solution of Problem (31) is unique, the solution is also the globally optimal solution of Problem (29). In the following, we focus our attention on the more tractable Problem (31).

Similarly to Stage I, the block coordinate descent method of [37] is applied to solve Problem (31). Given the TPC matrices

\mathbf{V} , the optimal solution of \mathbf{U} and \mathbf{W} are given by (14) and (15). Given \mathbf{U} and \mathbf{W} , the numerator and denominator in the OF of Problem (31) are concave and convex with respect to (w.r.t.) \mathbf{V} respectively, and the available power region is convex w.r.t. \mathbf{V} . Hence, it is a standard fractional programming problem, for which the globally optimal solution can be obtained through the following lemma. The corresponding proof is given in [41].

Lemma 2: Let us define the function $G(\eta)$ as ¹

$$G(\eta) \triangleq \max_{\mathbf{V} \in \mathcal{W}} \sum_{k \in \mathcal{S}} h_k(\mathbf{V}) - \eta \bar{P}_{\text{total}}(\mathbf{V}|\bar{\mathbf{V}}), \quad (32)$$

where we have:

$$\begin{aligned} \mathcal{W} &= \left\{ \mathbf{V} \mid h_k(\mathbf{V}) \geq R_{k,\min}, \forall k \in \mathcal{S}, \sum_{k \in \mathcal{S}} \|\mathbf{V}_{i,k}\|_F^2 \leq P_{i,\max}, \forall i \in \mathcal{I} \right\}. \end{aligned} \quad (33)$$

For a fixed η , the solution of Problem (32) is denoted as $\mathbf{V}^*(\eta)$. Then, solving Problem (31) with the aid of the given \mathbf{U} and \mathbf{W} is equivalent to finding the root of the equation

$$G(\eta^*) \triangleq \sum_{k \in \mathcal{S}} h_k(\mathbf{V}^*(\eta^*)) - \eta^* \bar{P}_{\text{total}}(\mathbf{V}^*(\eta^*)|\bar{\mathbf{V}}) = 0 \quad (34)$$

and $\mathbf{V}^*(\eta^*)$ is the optimal solution of Problem (31), when \mathbf{U} and \mathbf{W} are given. \square

Lemma 2 gives us insights into solving Problem (31) for a given \mathbf{U} and \mathbf{W} . We should firstly solve Problem (32) with a given η , and then adopt Dinkelbach's method [41] to update η as follows:

$$\eta^{(m)} = \frac{\sum_{k \in \mathcal{S}} h_k(\mathbf{V}^{(m)})}{\bar{P}_{\text{total}}(\mathbf{V}^{(m)}|\bar{\mathbf{V}})}, \quad (35)$$

where m is the iteration index. Problem (32) associated with a given η constitutes a convex optimization problem, and the globally optimal solution can again be found by the Lagrange dual method [36] as in Stage I, the details of which are omitted owing to space-economy.

Based on the above analysis, the triple-layer iterative algorithm is summarized in Table I, where the initial $\bar{\mathbf{V}}$ is set as a feasible $\{\mathbf{V}_{i,k}^{(0)}, \forall i, k\}$ output of Stage I and then it is updated by the previous iteration's solution. Next, we show that each layer in the triple-layer algorithm is guaranteed to converge. For the third layer, since the numerator and the denominator of the OF of Problem (31)² are concave and convex functions, respectively, Dinkelbach's method is guaranteed to converge to the globally optimal solution [41]. According to Theorem 2, the second layer is guaranteed to converge to the KKT conditions of Problem (29). In the following theorem, we prove the convergence of the first layer's iterative process, hence equivalently the triple-layer iterative algorithm is guaranteed to converge.

Theorem 3: The triple-layer iterative algorithm is guaranteed to converge to the KKT conditions of Problem (26).

Proof: Please see Appendix C. \square

¹Since we focus our attention on the solution of \mathbf{V} with given \mathbf{U} and \mathbf{W} , we omit the dependency of $h_k(\mathbf{U}, \mathbf{V}, \mathbf{W})$ on \mathbf{U} and \mathbf{W} , and $h_k(\mathbf{V})$ is short for $h_k(\mathbf{U}, \mathbf{V}, \mathbf{W})$.

²Both \mathbf{U} and \mathbf{W} are given.

TABLE I
A TRIPLE-LAYER ITERATIVE ALGORITHM TO SOLVE PROBLEM (8)

First Layer Initializations:
Initialize a small δ , error tolerance ε_1 , feasible $\{\mathbf{V}_{i,k}^{(0)}, \forall i, k\}$ output from Stage I, calculate $\bar{\omega}_{i,k}^{(0)}, \bar{\gamma}^{(0)}$ as in (30) and the OF of Problem (26), i.e., $\overline{EE}(\mathbf{V}^{(0)})$.

First Layer Iterations: $n_1 = 1, 2, \dots$
1) *Second Layer Initializations:*
Set $\mathbf{V}^{(n_1-1,0)} = \mathbf{V}^{(n_1-1)}$, error tolerance ε_2 , calculate the objective value of Problem (29), i.e., $\overline{EE}(\mathbf{V}^{(n_1-1,0)}|\mathbf{V}^{(n_1-1)})$, calculate $\mathbf{U}^{(n_1-1,0)}$ and $\mathbf{W}^{(n_1-1,0)}$ by using (14) and (15).
Second Layer Iterations: $n_2 = 1, 2, \dots$
1-1) *Third layer Initializations:*
Set $\mathbf{V}^{(n_1-1,n_2-1,0)} = \mathbf{V}^{(n_1-1,n_2-1)}$, error tolerance ε_3 , calculate $\eta^{(n_1-1,n_2-1,0)}$ by using (35).
Third Layer Iterations: $n_3 = 1, 2, \dots$
1-1-1) Compute the optimal solution $\mathbf{V}^*(\eta^{(n_1-1,n_2-1,n_3)})$ to Problem (32) with $\eta = \eta^{(n_1-1,n_2-1,n_3-1)}$;
1-1-2) Update $\eta^{(n_1-1,n_2-1,n_3)}$ according to (35) with $\mathbf{V}^*(\eta^{(n_1-1,n_2-1,n_3)})$.
1-1-3) If $\frac{|\eta^{(n_1-1,n_2-1,n_3)} - \eta^{(n_1-1,n_2-1,n_3-1)}|}{\eta^{(n_1-1,n_2-1,n_3-1)}} \leq \varepsilon_3$, let $\mathbf{V}^{(n_1-1,n_2)} = \mathbf{V}^*(\eta^{(n_1-1,n_2-1,n_3)})$ and stop, otherwise, go to step 1-1-1);
1-2) Update $\mathbf{U}^{(n_1-1,n_2)}$ and $\mathbf{W}^{(n_1-1,n_2)}$ by using (14) and (15) with $\mathbf{V}^{(n_1-1,n_2)}$.
1-3) If $\frac{|\overline{EE}(\mathbf{V}^{(n_1-1,n_2)}|\mathbf{V}^{(n_1-1)}) - \overline{EE}(\mathbf{V}^{(n_1-1,n_2-1)}|\mathbf{V}^{(n_1-1)})|}{\overline{EE}(\mathbf{V}^{(n_1-1,n_2-1)}|\mathbf{V}^{(n_1-1)})} \leq \varepsilon_2$, let $\mathbf{V}^{(n_1)} = \mathbf{V}^{(n_1-1,n_2)}$ and stop, otherwise, go to step 1-1).
2) Update $\{\bar{\omega}_{i,k}^{(n_1)}, \forall i, k\}, \bar{\gamma}^{(n_1)}$ as in (30) with $\{\mathbf{V}_{i,k}^{(n_1)}, \forall i, k\}$.
3) If $|\overline{EE}(\mathbf{V}^{(n_1)}) - \overline{EE}(\mathbf{V}^{(n_1-1)})|/\overline{EE}(\mathbf{V}^{(n_1-1)}) \leq \varepsilon_1$, stop, otherwise, go to step 1).

B. Single-layer Iterative Algorithm

Although the triple-layer iterative algorithm is guaranteed to converge to the KKT conditions of Problem (26), the computational complexity may become excessive due to the three layers of iterations. To reduce its complexity, we propose a single-layer iterative algorithm by combining the three layers of optimization into a single-layer iterative procedure. To guarantee its convergence, the parameter η is updated according to the following equation instead of (35)³:

$$\eta^{(n)} = \frac{\sum_{k \in \mathcal{S}} R_k(\mathbf{V}^{(n)})}{\tilde{P}_{\text{total}}(\mathbf{V}^{(n)})}. \quad (36)$$

where n is the iteration index. The detailed procedure is formulated in Algorithm 3. Note that the value of $\bar{\gamma}$ is not required in this algorithm, since η is updated in (36), rather than in (35).

Theorem 4: The single-layer algorithm is guaranteed to converge.

Proof: Please see Appendix D. \square

C. Complexity Analysis

By using a similar complexity analysis to that of the block descent technique of Algorithm 1, the total complexity of the

Algorithm 3 Single-layer algorithm to solve Problem (26)

- 1: Initialize iteration index $n = 1$, a small δ , error tolerance ε , initialize the feasible $\mathbf{V}^{(0)}$ output from Stage I, calculate $\bar{\omega}_{i,k}^{(0)}$ as in (30), $\mathbf{U}^{(0)}$ and $\mathbf{W}^{(0)}$ by using (14) and (15), $\eta^{(0)}$ according to (36);
- 2: Given $\eta^{(n-1)}, \bar{\omega}_{i,k}^{(n-1)}, \mathbf{U}^{(n-1)}$ and $\mathbf{W}^{(n-1)}$, solve Problem (32) by using the Lagrange dual method to obtain the optimal solution of $\mathbf{V}^{(n)}$;
- 3: Given $\mathbf{V}^{(n)}$, update $\mathbf{U}^{(n)}$ and $\mathbf{W}^{(n)}$ by using (14) and (15);
- 4: Given $\mathbf{V}^{(n)}$, update $\bar{\omega}_{i,k}^{(n)}$ as in (30);
- 5: Given $\mathbf{V}^{(n)}$, update $\eta^{(n)}$ according to (36);
- 6: If $|\eta^{(n)} - \eta^{(n-1)}|/\eta^{(n-1)} \leq \varepsilon$, terminate. Otherwise, set $n \leftarrow n + 1$, go to step 2.

triple-layer iterative algorithm and of the single-layer one are respectively given by

$$T_{\text{Tri}} = O[t_{\text{layer1}} t_{\text{layer2}} t_{\text{layer3}} (S + I)^2 S M^3 I^3] \quad (37)$$

and

$$T_{\text{Single}} = O[t_{\text{sinlayer}} (S + I)^2 S M^3 I^3], \quad (38)$$

where $t_{\text{layer1}}, t_{\text{layer2}}$ and t_{layer3} are the average number of iterations required for the first, second and third layers of the triple-layer iterative algorithm, respectively, and t_{sinlayer} denotes the average number of iterations in the single-layer algorithm. It is seen from the simulations that the total number of solving Problem (32), i.e. $t_{\text{layer1}} t_{\text{layer2}} t_{\text{layer3}}$, is roughly 100, while that for the single-layer algorithm, i.e. t_{layer3} is about 20. Despite its five-fold complexity, the performance gain of the triple-layer iterative algorithm over the single-layer algorithm is not so significant. Hence, in large networks, the single-layer iterative algorithm may be deemed a more suitable approach. To additionally reduce the complexity, some clustering methods can be adopted, such as the user-centric method, which will be considered in our future work.

The main mathematical techniques involved in Stage II are illustrated in Fig.2.

V. SIMULATION RESULTS

In this section, simulation results are presented for evaluating the performance of the proposed algorithms. All the K users are assumed to be independently and uniformly distributed in a circular cell centered at $(0,0)$ with a radius of R . The locations of the RAUs are the same as in [42] and [43]: if the number of RAUs I is less than six, the j -th RAU's polar coordinate is $(r \cos[2\pi(j-1)/I], r \sin[2\pi(j-1)/I])$ for $j = 1, \dots, I$, $r = 2R \sin(\pi/I)/(3\pi/I)$. Otherwise, the first RAU is located at the cell center $(0,0)$, and the other $(I-1)$ RAUs are located at $(r \cos[2\pi(j-2)/(I-1)], r \sin[2\pi(j-2)/(I-1)])$ with $j = 2, \dots, I$. The channel model consists of 1) the long term evolution (LTE) standard path loss model: $PL_{i,k} = 148.1 + 37.6 \log_{10} d_{i,k}$ (dB), where $d_{i,k}$ (in km) is the distance between the i th RAU and the k th user; 2) the log-normal shadowing with zero mean and 8 dB standard deviation; 3) the

³The update in (36) will be used in the proof of Theorem 4.

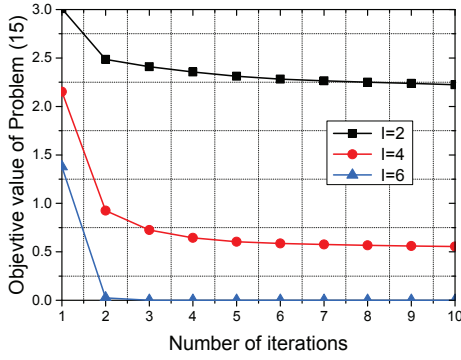


Fig. 4. Convergence behaviour of the block descent algorithm when $K = 6$ for different number of RAUs.

small-scale Rayleigh fading with zero mean and unit variance. For simplicity, we assume that each user has the same rate requirement and each RAU has the same power limit, i.e., we have $R_{\min} = R_{k,\min}, \forall k$ and $P_{\max} = P_{i,\max}, \forall i$. The number of data streams d is set to $d = \min\{MI, N\}$. The small parameter δ in the triple-layer algorithm and the single-layer algorithm is set to $\delta = 10^{-5}$. Unless otherwise stated, the other main system parameters are given as follows: system bandwidth 20 MHz, cell radius $R = 1000$ m, number of receiver antennas $N = 2^4$, number of transmit antennas $M = 2$, noise power spectral density is -174 dBm/Hz, fronthaul power consumption $P_{\text{fh}} = 0.5$ W, signal processing power of CPU $P_C = 5$ W, error tolerance $\varepsilon = 10^{-4}$. The following results are obtained by averaging over 50 independent channel generations.

A. Properties of the User Selection Algorithm

Fig. 4 shows the convergence behaviour of the block descent procedure in Algorithm 2 for different numbers of RAUs. The number of users is $K = 6$ and the rate target is $R_{\min} = 3$ nats/s/Hz. It can be observed from this figure that the OF value of Problem (18) is monotonically decreasing during the iterative procedure and converges rapidly for the all the configurations considered. It is interesting to note that for the case of $I = 6$, only one iteration is enough for the algorithm to converge, while for the cases of $I = 2, 4$, nearly ten iterations are required. The reason is that when $I = 6$, more degrees of freedom are available and the OF will decrease to the optimal zero value after a single iteration (i.e. all six users can be supported in this case). However, for the cases of $I = 2, 4$, the system is not able to support all the users and the algorithm will need several iterations to converge to a fixed non-zero value. In the following simulations, the maximum number of iterations for the block descent algorithm is set to ten.

To study the efficiency of our proposed user selection algorithm (denoted as ‘USC alg.’), we compare it to two existing user selection methods in Fig. 5: the exhaustive search method (denoted as ‘Exhaus alg.’) and the random

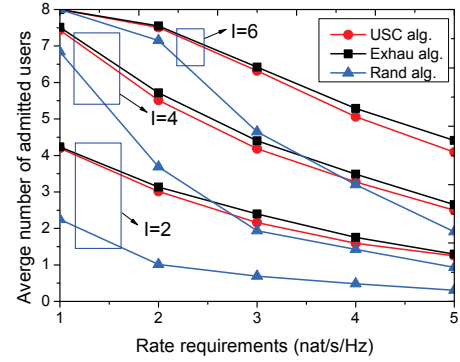


Fig. 5. Average number of admitted users versus the data requirements for our proposed user selection algorithm, for the exhaustive search method and for random user deletion algorithm with $K = 8$.

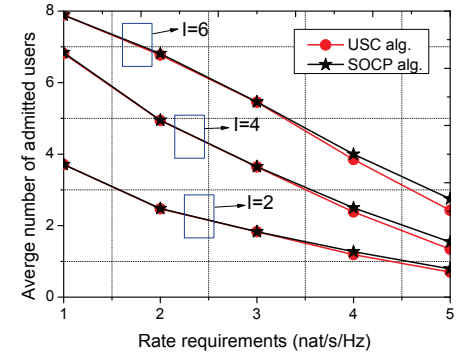


Fig. 6. The average number of admitted users versus the rate requirements for our proposed user selection method and the SOCP-based algorithm of [34] with $N = 1$ and $K = 8$.

user deletion algorithm (denoted as ‘Rand alg.’). Note that the complexity of the exhaustive search method is on the order of 2^K , which is unaffordable for a dense network supporting a large number of users. However, it can serve as the performance benchmark for our proposed algorithm. For the random user deletion algorithm, one should solve Problem (9) in each iteration. If $\alpha_k = 1, \forall k \in \mathcal{S}$, terminate. Otherwise, a user is randomly removed from \mathcal{S} without considering the optimal values of $\alpha_k, \forall k \in \mathcal{S}$. We can find from Fig. 5 that the proposed USC algorithm achieves almost the same performance as the exhaustive search method, especially for low rate requirements. In the high rate regime, the exhaustive search method performs slightly better than the proposed user selection algorithm, which confirms the efficiency of the proposed algorithm. As expected, the number of admitted users decreases upon increasing of the rate requirements. It is also observed that the random user deletion algorithm has a much worse performance than our proposed algorithm, which confirms the efficiency of removing the user having the lowest value of α_k .

In Fig. 6, we compare the performance of our proposed user selection method to that of the SOCP-based algorithm of [34], when each user is equipped with a single antenna. It is seen from Fig. 6 that both algorithms achieve almost the same performance for most of the rate requirement range, and only a slight performance gain is attained for the SOCP algorithm

⁴In fact, the long term evolution -Advanced (LTE-A) standards can support UEs with up to four receive antennas [44], and some field trials have been demonstrated to verify the feasibility of using four antennas at the UEs [44].

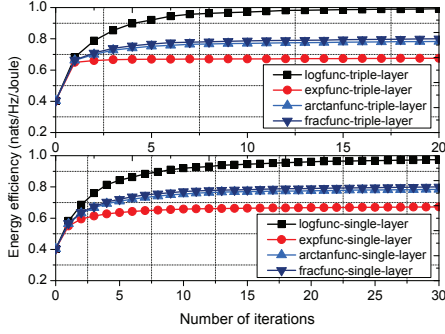


Fig. 7. Convergence behaviour of the various indicator approximation functions for both the triple-layer iteration algorithm (upper subplot) and the single-layer iteration algorithm (lower subplot).

over our proposed algorithm in the high rate requirement region. However, the SOCP is restricted to the single-antenna aided user scenario, while our algorithm is also applicable to multiple-antenna aided users.

B. Properties of the EE Maximization Algorithms

In this subsection, we study the properties of the proposed EE maximization algorithms: the triple-layer algorithm and the single-layer algorithm.

In Fig. 7, we first investigate the convergence behaviours of the two algorithms when using different indicator approximation functions, where $\delta = 10^{-5}$. For the triple-layer algorithm, we only simulate the convergence behaviour of the first layer, where the second and third layer are converged for each iteration of the first layer. It is observed from Fig. 7 that the algorithms converge promptly for all the indicator approximation functions considered. It is interesting to observe that although the algorithm using the exponential function, arctangent function and the fractional function exhibits a slightly faster convergence than the logarithmic function, the EE achieved by the algorithm using these three functions is much lower than that using the logarithmic function. This indicates that the more accurate approximation function may not guarantee a better performance, and the more smooth function may be a good option. Hence, in the following investigations, the logarithmic function is adopted. Note that the triple-layer algorithm and the single-layer algorithm have a similar performance, when using the same function. This phenomenon will be studied in detail in the following.

Fig. 8 and Fig. 9 compare the EE performance and the corresponding number of iterations required for the triple-layer algorithm and the single-layer algorithm under 30 randomly generated channel realizations, respectively. The number of users is set as $K = 5$. Since the triple-layer algorithm is only guaranteed to converge to the KKT-optimal solution, there is no guarantee that it always outperforms the single-layer algorithm owing to the nonconvex nature of Problem (26). However, it can be observed from Fig. 8 that the triple-layer algorithm performs slightly better than the single-layer algorithm in most cases. Fig. 9 shows that the marginal performance gain of the triple-layer algorithm over the single-layer algorithm comes at the cost of a higher computational

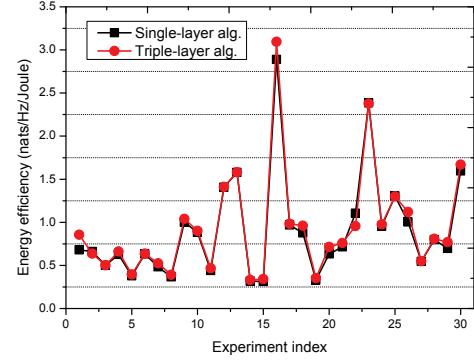


Fig. 8. Energy efficiency achieved by the triple-layer algorithm and the single-layer algorithm.

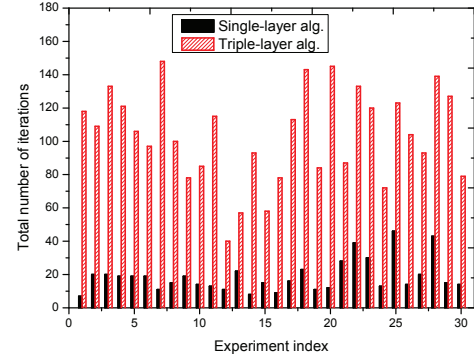


Fig. 9. Total number of iterations required by the triple-layer algorithm and the single-layer algorithm.

complexity, i.e., the single-layer algorithm requires about 20 iterations, while the triple-layer algorithm requires about a total of 100 inner iterations.

In this following, we compare the performance of the proposed triple-layer algorithm and the single-layer algorithm with the following algorithms, where the number of users is set as $K = 6$. For all algorithms, we adopt the same output from Stage I as the input for the initializations for all Stage II algorithms.

- Rate maximization method (denoted as ‘Ratemaxi’): This method aims for maximizing the sum-rate of the users selected during Stage I under the rate constraints and the per-RAU power constraints;
- Power minimization method (denoted as ‘Powermin’): This method aims for minimizing the total transmit power of the users selected during Stage I under the two sets of constraints;
- EE maximization without link selection (denoted as ‘EEnolinksel’): In this method, the link selection is not considered. The triple-layer algorithm can be adapted for solving this problem by removing the first layer.
- Exhaustive search (denoted as ‘Exhaustive’): This method searches through all possible sets of active links and selects the feasible set of links that has the maximum EE. This method can find the globally optimal solution, but incurs an exponentially increasing complexity that is not affordable for practical use. Due to its excessive

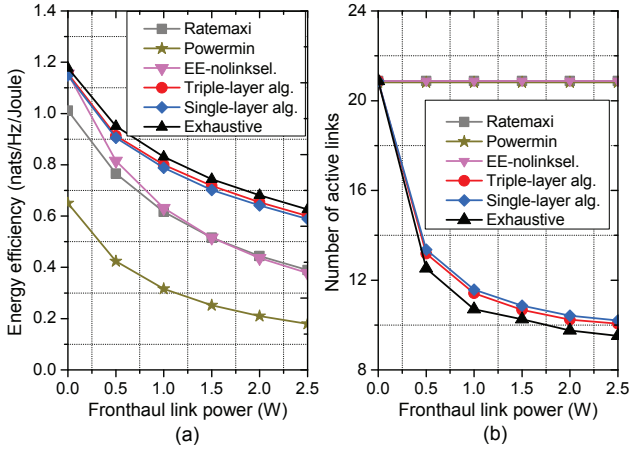


Fig. 10. Average EE and the number of active links versus the fronthaul power consumption P_{fh} . The maximum power constraint is set as $P_{\text{max}} = 2\text{W}$, the number of users is $K = 6$, the number of RAUs is $I = 4$, and the rate requirement is set as $R_{\text{min}} = 1 \text{ nats/s/Hz}$.

complexity, we only report its performance for small-scale systems.

1) *Impact of fronthaul power consumption:* Figs. 10 (a) and (b) illustrate the EE performance and the corresponding number of active links versus the fronthaul power consumption P_{fh} , respectively. As expected, the EE achieved by all the algorithms is reduced upon increasing P_{fh} . It is observed again that the performance of the triple-layer algorithm is close to that of the single-layer algorithm. Interestingly, the proposed triple-layer algorithms' performance is comparable to the optimal EE achieved by the exhaustive search method. More specifically, the triple-layer algorithm achieves at least 94% of the optimal EE in this example. By jointly selecting the active links and optimizing the TPCs, the proposed algorithms significantly outperform 'EE-nolinksel'. The performance gain is increased upon increasing P_{fh} , which highlights the importance of link selection. It is interesting to observe that with the increase of P_{fh} , the EE performance of 'Ratemaxi' is close to that of 'EE-nolinksel'. This is due to the fact that with the increase of P_{fh} , a higher data rate is required for improving the EE as proved in [45], and 'EE-nolinksel' reduces to 'Ratemaxi'. From Fig. 10 (b), we see that the number of active links supported by our proposed algorithms decreases with the increase of P_{fh} in order to save power, while the number of active links for the other algorithms (except the exhaustive search method) remains almost the same.

2) *Impact of the number of RAUs:* Figs. 11 (a) and (b) show the EE and the corresponding number of active links versus the number of RAUs I , respectively. Again, the performance of the proposed algorithms is nearly the same as that of the exhaustive search method, which verifies the efficiency of the proposed algorithms. When the number of RAUs increases from 2 to 6, the EE achieved by the proposed algorithms remains almost the same. The reason for this trend is that with a small increase of I , more users can be admitted and the sum rate will increase. However, to support more users, more transmit power and fronthaul power is required. On the other hand, when the number of RAUs continue to increase

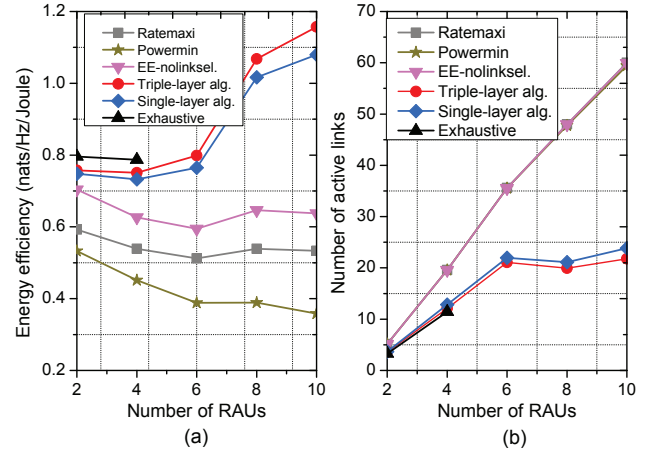


Fig. 11. Average EE and the number of active links versus the number of RAUs I . The maximum power constraint is set as $P_{\text{max}} = 5\text{W}$, the number of users is $K = 6$, the fronthaul power consumption is $P_{\text{fh}} = 0.5\text{W}$, and the rate requirement is set as $R_{\text{min}} = 2 \text{ nats/s/Hz}$.

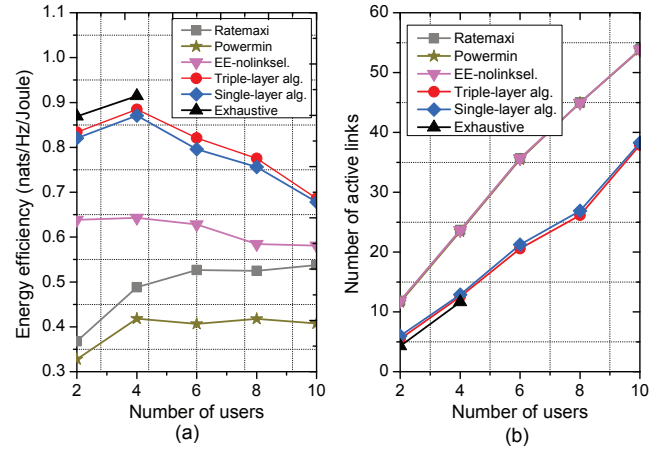


Fig. 12. Average EE and the number of active links versus the number of users. The maximum power constraint is set as $P_{\text{max}} = 5\text{W}$, the number of RAUs is $I = 6$, the fronthaul power consumption is $P_{\text{fh}} = 0.5\text{W}$, and the rate requirement is set as $R_{\text{min}} = 2 \text{ nats/s/Hz}$.

from 6 to 10, there is a significant EE increase for our proposed algorithms. This is mainly due to the fact that when there is a large number of RAUs, all the six users can be supported and the RAU selection diversity gain can be obtained with more RAUs. It is observed that our proposed algorithms significantly outperform the other algorithms (except for the exhaustive search method), and that the performance gain is more prominent, when the number of RAUs is large. The reason is that these algorithms do not take into consideration the link selection and all the links are active, which consumes a large amount of power.

3) *Impact of the number of users:* Finally, let us quantify the impact of the number of users on the performance of the various algorithms. Figs. 12 (a) and (b) show the EE performance and the corresponding number of active links versus the number of users, respectively. It is shown in Fig. 12 (a) that when the number of users increases from 2 to 4, the EE achieved by our proposed algorithms is slightly increased. The

reason for this is that a small number of users can be admitted without requiring much transmit power and the increase of the sum rate can compensate for the increase of the total power consumption. However, when the number of users increases from 4 to 10, our proposed algorithms suffer from a significant EE performance reduction. This is mainly due to the fact that our two-stage procedure tries to maximize the number of admitted users, which comes at the expense of much increased power consumption. In this case, nearly all transmit power will be consumed and all links will be active in an effort to support more users. As a result, the performance gap between our proposed algorithms and the other algorithms will decrease upon increasing the number of users.

VI. CONCLUSIONS

This paper studied the EE maximization problem of a multiuser MIMO DAS through jointly selecting the fronthaul links and optimizing the TPC matrices. This problem was solved in two stages. In Stage I, a novel user selection algorithm was proposed for maximizing the number of admitted users. In Stage II, our triple-layer iterative algorithm and single-layer iterative algorithm were proposed for optimizing the TPC matrices to maximize the EE for the users selected during Stage II. Simulation results show that the proposed user selection algorithm achieves almost the same performance as the exhaustive search method at a much reduced complexity. Furthermore, the single-layer iterative algorithm was observed to converge much faster than the triple-layer iterative algorithm, while its performance is comparable to the latter. This means that the single-layer iterative algorithm is more suitable for practical implementations. By appropriately selecting the fronthaul links, the proposed algorithms significantly outperform the existing methods in terms of their EE performance.

APPENDIX A PROOF OF THEOREM 1

Denote the OF of Problem (18) in the n th iteration as

$$\text{Obj}^{(n)} = \sum_{k \in \mathcal{S}} \left(\alpha_k^{(n)} - 1 \right)^2.$$

In Step 2 of the n th iteration, the optimal $\{\alpha_k^{(n)}\}_{k=1}^K$ and $\mathbf{V}^{(n)}$ can be obtained by solving Problem (19) with given $\mathbf{U}^{(n-1)}$ and $\mathbf{W}^{(n-1)}$. Hence, we

$$h_k \left(\mathbf{V}^{(n)}, \mathbf{U}_k^{(n-1)}, \mathbf{W}_k^{(n-1)} \right) \geq \left(\alpha_k^{(n)} \right)^2 R_{k,\min}, \forall k \in \mathcal{S}.$$

In Step 3 of the n th iteration, $\mathbf{U}^{(n)}$ and $\mathbf{W}^{(n)}$ are updated in (14) and (15) with $\mathbf{V}^{(n)}$. Based on Lemma 1, the following inequality holds

$$\begin{aligned} R \left(\mathbf{V}^{(n)} \right) &= h_k \left(\mathbf{V}^{(n)}, \mathbf{U}_k^{(n)}, \mathbf{W}_k^{(n)} \right) \\ &\geq h_k \left(\mathbf{V}^{(n)}, \mathbf{U}_k^{(n-1)}, \mathbf{W}_k^{(n-1)} \right). \end{aligned} \quad (\text{A.1})$$

Hence, we have

$$h_k \left(\mathbf{V}^{(n)}, \mathbf{U}_k^{(n)}, \mathbf{W}_k^{(n)} \right) \geq \alpha_k^{(n)2} R_{k,\min}. \quad (\text{A.2})$$

In Step 2 of the $(n+1)$ th iteration, $\alpha^{(n+1)}$ and $\mathbf{V}^{(n+1)}$ are obtained with given $\mathbf{U}^{(n)}$ and $\mathbf{W}^{(n)}$ by solving Problem (19). Then we have $\text{Obj}^{(n+1)} \leq \text{Obj}^{(n)}$ due to the fact that $\alpha_k^{(n)}$ and $\mathbf{V}^{(n)}$ is a feasible solution of Problem (19) with given $\mathbf{U}^{(n)}$ and $\mathbf{W}^{(n)}$ according to (A.2). Hence, the OF of Problem (18) is monotonically increasing. In addition, the OF value is lower-bounded by zero. Hence, the block descent algorithm must converge.

APPENDIX B PROOF OF THEOREM 2

We first prove the first part. Introducing non-negative Lagrange multipliers $\boldsymbol{\lambda} = \{\lambda_{u_1}, \dots, \lambda_{u_S}\}$ associated with the rate constraints and $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_I\}$ associated with the power constraints, the Lagrange function of Problem (31) is

$$\begin{aligned} \mathcal{L}(\mathbf{U}, \mathbf{V}, \mathbf{W}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \frac{\sum_{k \in \mathcal{S}} h_k(\mathbf{U}, \mathbf{V}, \mathbf{W})}{\bar{P}_{\text{total}}(\mathbf{V}|\bar{\mathbf{V}})} \\ &\quad + \sum_{k \in \mathcal{S}} \lambda_k (h_k(\mathbf{U}, \mathbf{V}, \mathbf{W}) - R_{k,\min}) \\ &\quad + \sum_{i \in \mathcal{I}} \mu_i \left(P_{i,\max} - \sum_{k \in \mathcal{S}} \|\mathbf{V}_{i,k}\|_F^2 \right). \end{aligned} \quad (\text{B.1})$$

Since $(\mathbf{U}^*, \mathbf{V}^*, \mathbf{W}^*)$ is the solution of problem (31), there exist $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ values, so that $(\mathbf{U}^*, \mathbf{V}^*, \mathbf{W}^*)$ satisfies the KKT conditions [46, Prop. 4.7.2] of Problem (31), shown at the top of the next page, (B.2)-(B.8).

Since we have $\bar{P}_{\text{total}}(\mathbf{V}^*|\bar{\mathbf{V}}) > 0$ and $\{\lambda_k^* \geq 0, \forall k\}$, we can solve $\mathbf{U}_k^*, \forall k$ and $\mathbf{W}_k^*, \forall k$ from (B.2) and (B.3) as in (14) and (15). Then, we have the following chain of inequalities

$$\nabla_{\mathbf{V}_k^*} h_k(\mathbf{U}^*, \mathbf{V}^*, \mathbf{W}^*) = -\text{Tr}(\mathbf{W}_k^* \nabla_{\mathbf{V}_k^*} \mathbf{E}_k^*) \quad (\text{B.9})$$

$$= -\text{Tr}(\mathbf{E}_k^{*-1} \nabla_{\mathbf{V}_k^*} \mathbf{E}_k^*) \quad (\text{B.10})$$

$$= \nabla_{\mathbf{V}_k^*} \log |\mathbf{E}_k^{*-1}| \quad (\text{B.11})$$

$$= \nabla_{\mathbf{V}_k^*} R_k(\mathbf{V}^*). \quad (\text{B.12})$$

where (B.9) follows from the chain rule, and the final equality follows from the Woodbury matrix identity [47] to applied (16). With the aid of similar methods, we can readily prove that $h_k(\mathbf{U}^*, \mathbf{V}^*, \mathbf{W}^*) = R_k(\mathbf{V}^*)$. Hence, equation (B.4) is equivalent to the following equation

$$\begin{aligned} &\frac{\bar{P}_{\text{total}}(\mathbf{V}^*|\bar{\mathbf{V}}) \sum_{k \in \mathcal{S}} \nabla_{\mathbf{V}_k^*} R_k(\mathbf{V}^*) - \nabla_{\mathbf{V}^*} \bar{P}_{\text{total}}(\mathbf{V}^*|\bar{\mathbf{V}}) \sum_{k \in \mathcal{S}} R_k(\mathbf{V}^*)}{\bar{P}_{\text{total}}(\mathbf{V}^*|\bar{\mathbf{V}})^2} + \\ &\sum_{k \in \mathcal{S}} \lambda_k^* \nabla_{\mathbf{V}_k^*} R_k(\mathbf{V}^*) - \sum_{i \in \mathcal{I}} \mu_i^* \nabla_{\mathbf{V}_k^*} \left(\sum_{k \in \mathcal{S}} \|\mathbf{V}_{i,k}\|_F^2 \right) = \mathbf{0}, \forall k. \end{aligned} \quad (\text{B.13})$$

Additionally, (B.5) and (B.7) are equivalent to the following equations:

$$\lambda_k^* (R_k(\mathbf{V}^*) - R_{k,\min}) = 0, \forall k, \quad (\text{B.14})$$

$$R_k(\mathbf{V}^*) \geq R_{k,\min}, \forall k. \quad (\text{B.15})$$

As a result, the system equations (B.13), (B.6), (B.8), (B.14) and (B.15) represent the KKT conditions of Problem (29). Therefore, the first conclusion holds. By using a similar procedure, it may be readily shown that the second conclusion also holds.

$$\nabla_{\mathbf{U}_k^*} \mathcal{L} = \frac{\sum_{k \in \mathcal{S}} \nabla_{\mathbf{U}_k^*} h_k(\mathbf{U}^*, \mathbf{V}^*, \mathbf{W}^*)}{\bar{P}_{\text{total}}(\mathbf{V}^* | \bar{\mathbf{V}})} + \sum_{k \in \mathcal{S}} \lambda_k^* \nabla_{\mathbf{U}_k^*} h_k(\mathbf{U}^*, \mathbf{V}^*, \mathbf{W}^*) = \mathbf{0}, \forall k, \quad (\text{B.2})$$

$$\nabla_{\mathbf{W}_k^*} \mathcal{L} = \frac{\sum_{k \in \mathcal{S}} \nabla_{\mathbf{W}_k^*} h_k(\mathbf{U}^*, \mathbf{V}^*, \mathbf{W}^*)}{\bar{P}_{\text{total}}(\mathbf{V}^* | \bar{\mathbf{V}})} + \sum_{k \in \mathcal{S}} \lambda_k^* \nabla_{\mathbf{W}_k^*} h_k(\mathbf{U}^*, \mathbf{V}^*, \mathbf{W}^*) = \mathbf{0}, \forall k, \quad (\text{B.3})$$

$$\begin{aligned} \nabla_{\mathbf{V}_k^*} \mathcal{L} = & \frac{\bar{P}_{\text{total}}(\mathbf{V}^* | \bar{\mathbf{V}}) \sum_{k \in \mathcal{S}} \nabla_{\mathbf{V}_k^*} h_k(\mathbf{U}^*, \mathbf{V}^*, \mathbf{W}^*) - \nabla_{\mathbf{V}_k^*} \bar{P}_{\text{total}}(\mathbf{V}^* | \bar{\mathbf{V}}) \sum_{k \in \mathcal{S}} h_k(\mathbf{U}^*, \mathbf{V}^*, \mathbf{W}^*)}{\bar{P}_{\text{total}}(\mathbf{V}^* | \bar{\mathbf{V}})^2} \\ & + \sum_{k \in \mathcal{S}} \lambda_k^* \nabla_{\mathbf{V}_k^*} h_k(\mathbf{U}^*, \mathbf{V}^*, \mathbf{W}^*) - \sum_{i \in \mathcal{I}} \mu_i^* \nabla_{\mathbf{V}_k^*} \left(\sum_{k \in \mathcal{S}} \|\mathbf{V}_{i,k}^*\|_F^2 \right) = \mathbf{0}, \forall k, \end{aligned} \quad (\text{B.4})$$

$$\lambda_k^* (h_k(\mathbf{U}^*, \mathbf{V}^*, \mathbf{W}^*) - R_{k,\min}) = 0, \forall k, \quad (\text{B.5})$$

$$\mu_i^* \left(P_{i,\max} - \sum_{k \in \mathcal{S}} \|\mathbf{V}_{i,k}^*\|_F^2 \right) = 0, \forall i, \quad (\text{B.6})$$

$$h_k(\mathbf{U}^*, \mathbf{V}^*, \mathbf{W}^*) \geq R_{k,\min}, \forall k, \quad (\text{B.7})$$

$$\sum_{k \in \mathcal{S}} \|\mathbf{V}_{i,k}^*\|_F^2 \leq P_{i,\max}, \forall i. \quad (\text{B.8})$$

APPENDIX C PROOF OF THEOREM 3

We first show that given the initial point gleaned from Stage I, the triple-layer algorithm will converge to a unique point ⁵.

As mentioned before, Dinkelbach's method [41] is guaranteed to converge to the globally optimal solution [41]. Hence, given the initial point from the second layer, the third layer is guaranteed to converge to a unique point. Additionally, $\mathbf{U}^{(n_1-1, n_2)}$ and $\mathbf{W}^{(n_1-1, n_2)}$ are uniquely determined by (14) and (15), given $\mathbf{V}^{(n_1-1, n_2)}$. Hence, the second layer will also converge to a unique point, when the initial points output by the first layer are given. In Step 2) of the first layer, $\{\bar{\omega}_{i,k}^{(n_1)}, \forall i, k\}$, $\bar{\gamma}^{(n_1)}$ are uniquely determined according to (30) with $\{\mathbf{V}_{i,k}^{(n_1)}, \forall i, k\}$. As a result, if the first layer can be shown to converge, then the three-layer algorithm is guaranteed to converge to a unique point with the initial point gleaned from Stage I.

The convergence of the first layer can be guaranteed by the set of inequalities, shown at the top of next page. where (C.1) follows due to (28), and (C.2) is guaranteed by the monotonically increasing property of the second layer. It is plausible that the OF value of Problem (26) is upper-bounded. Thus, the triple-layer algorithm is guaranteed to converge to a unique solution.

Let us denote the unique solution as \mathbf{V}^* . In the following, we prove that this solution satisfies the KKT condition of Problem (26). According to Theorem 2, the converged solution \mathbf{V}^* satisfies the KKT condition of Problem (29). The Lagrangian

of Problem (29) can be written as

$$\begin{aligned} \mathcal{L}(\mathbf{V}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = & \frac{\sum_{k \in \mathcal{S}} R_k(\mathbf{V})}{\bar{P}_{\text{total}}(\mathbf{V} | \mathbf{V}^*)} + \sum_{k \in \mathcal{S}} \lambda_k (R_k(\mathbf{V}) - R_{k,\min}) + \\ & \sum_{i \in \mathcal{I}} \mu_i \left(P_{i,\max} - \sum_{k \in \mathcal{S}} \|\mathbf{V}_{i,k}\|_F^2 \right). \end{aligned} \quad (\text{C.10})$$

The converged solution \mathbf{V}^* satisfies the KKT conditions of (C.4)-(C.8) at the top of the next page:

where $\{\lambda_k^*, \mu_i^*, \forall i, k\}$ are the corresponding optimal Lagrange dual multipliers.

It can be readily verified that the following two equations hold:

$$\bar{P}_{\text{total}}(\mathbf{V}^* | \mathbf{V}^*) = \tilde{P}_{\text{total}}(\mathbf{V}^*) \quad (\text{C.11})$$

$$\begin{aligned} \nabla_{\mathbf{V}_{i,k}^*} \bar{P}_{\text{total}}(\mathbf{V}^* | \mathbf{V}^*) &= \mathbf{V}_{i,k}^* + P_{\text{th}} f_{\delta}' \left(\|\mathbf{V}_{i,k}^*\|_F^2 \right) \mathbf{V}_{i,k}^* \\ &= \nabla_{\mathbf{V}_{i,k}^*} \tilde{P}_{\text{total}}(\mathbf{V}^*). \end{aligned} \quad (\text{C.12})$$

By substituting (C.11) and (C.12) into (C.4), we have (C.9) at the top of next page.

Combining (C.9) with (C.5), (C.5), (C.6), (C.7) and (C.8), we can find that the converged solution \mathbf{V}^* again with $\{\lambda^*, \mu^*\}$ satisfies the KKT conditions of Problem (26).

APPENDIX D PROOF OF THEOREM 4

To prove the convergence, we first define some functions as follows:

$$\begin{aligned} f_1(\mathbf{V}, \mathbf{U}, \mathbf{W}, \bar{\omega}, \eta) &= \sum_{k \in \mathcal{S}} h_k(\mathbf{V}, \mathbf{U}_k, \mathbf{W}_k) \\ &\quad - \eta \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{S}} \left[(1 + \bar{\omega}_{i,k}) \|\mathbf{V}_{i,k}\|_F^2 \right] \end{aligned} \quad (\text{D.1})$$

$$f_2(\mathbf{V}, \bar{\omega}, \eta) = \sum_{k \in \mathcal{S}} R_k(\mathbf{V}) - \eta \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{S}} \left[(1 + \bar{\omega}_{i,k}) \|\mathbf{V}_{i,k}\|_F^2 \right] \quad (\text{D.2})$$

⁵It is emphasized here that since Problem (26) is a nonconvex optimization problem, the converged solution depends on the initialization point from Stage I.

$$\begin{aligned} & \widetilde{EE}(\mathbf{V}^{(n_1)}) \\ & \geq \frac{\sum_{k \in \mathcal{S}} R_k(\mathbf{V}^{(n_1)})}{\sum_{k \in \mathcal{S}} \text{Tr}(\mathbf{V}_k^{(n_1)} \mathbf{V}_k^{(n_1)H}) + \sum_{k \in \mathcal{S}} \sum_{i \in \mathcal{I}} \bar{\omega}_{i,k}^{(n_1-1)} \left\| \mathbf{V}_{i,k}^{(n_1)} \right\|^2 + \bar{\gamma}^{(n_1-1)}} \end{aligned} \quad (\text{C.1})$$

$$\begin{aligned} & \geq \frac{\sum_{k \in \mathcal{S}} R_k(\mathbf{V}^{(n_1-1)})}{\sum_{k \in \mathcal{S}} \text{Tr}(\mathbf{V}_k^{(n_1-1)} \mathbf{V}_k^{(n_1-1)H}) + \sum_{k \in \mathcal{S}} \sum_{i \in \mathcal{I}} \bar{\omega}_{i,k}^{(n_1-1)} \left\| \mathbf{V}_{i,k}^{(n_1-1)} \right\|^2 + \bar{\gamma}^{(n_1-1)}} \end{aligned} \quad (\text{C.2})$$

$$= \widetilde{EE}(\mathbf{V}^{(n_1-1)}), \quad (\text{C.3})$$

$$\begin{aligned} \nabla_{\mathbf{V}_{i,k}^*} \mathcal{L}(\mathbf{V}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \frac{\bar{P}_{\text{total}}(\mathbf{V}^* | \mathbf{V}^*) \sum_{k \in \mathcal{S}} \nabla_{\mathbf{V}_{i,k}^*} R_k(\mathbf{V}^*) - \sum_{k \in \mathcal{S}} R_k(\mathbf{V}^*) \nabla_{\mathbf{V}_{i,k}^*} \bar{P}_{\text{total}}(\mathbf{V}^* | \mathbf{V}^*)}{\bar{P}_{\text{total}}(\mathbf{V}^* | \mathbf{V}^*)^2} \\ &+ \sum_{k \in \mathcal{S}} \lambda_k^* \nabla_{\mathbf{V}_{i,k}^*} R_k(\mathbf{V}^*) - \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{S}} \mu_i^* \mathbf{V}_{i,k}^* = \mathbf{0}, \forall i, k \end{aligned} \quad (\text{C.4})$$

$$\lambda_k^* (R_k(\mathbf{V}^*) - R_{k,\min}) = 0, \forall k \quad (\text{C.5})$$

$$\mu_i^* \left(P_{i,\max} - \sum_{k \in \mathcal{S}} \left\| \mathbf{V}_{i,k}^* \right\|_F^2 \right) = 0, \forall i \quad (\text{C.6})$$

$$R_k(\mathbf{V}^*) \geq R_{k,\min}, \forall k \quad (\text{C.7})$$

$$\sum_{k \in \mathcal{S}} \left\| \mathbf{V}_{i,k}^* \right\|_F^2 \leq P_{i,\max}, \forall i, \quad (\text{C.8})$$

$$\begin{aligned} \nabla_{\mathbf{V}_{i,k}^*} L(\mathbf{V}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \frac{\tilde{P}_{\text{total}}(\mathbf{V}^*) \sum_{k \in \mathcal{S}} \nabla_{\mathbf{V}_{i,k}^*} R_k(\mathbf{V}^*) - \sum_{k \in \mathcal{S}} R_k(\mathbf{V}^*) \nabla_{\mathbf{V}_{i,k}^*} \tilde{P}_{\text{total}}(\mathbf{V}^*)}{\tilde{P}_{\text{total}}(\mathbf{V}^*)^2} \\ &+ \sum_{k \in \mathcal{S}} \lambda_k^* \nabla_{\mathbf{V}_{i,k}^*} R_k(\mathbf{V}^*) - \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{S}} \mu_i^* \mathbf{V}_{i,k}^* = \mathbf{0}, \forall i, k. \end{aligned} \quad (\text{C.9})$$

$$\begin{aligned} f_3(\mathbf{V}, \eta) &= \sum_{k \in \mathcal{S}} R_k(\mathbf{V}) \\ &- \eta \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{S}} \left[\left\| \mathbf{V}_{i,k} \right\|_F^2 + P_{\text{fh}} f_\delta \left(\left\| \mathbf{V}_{i,k} \right\|_F^2 \right) \right], \end{aligned} \quad (\text{D.3})$$

where $\bar{\omega}$ denotes the collection of $\bar{\omega}_{i,k}, \forall i, k$ with $\bar{\omega}_{i,k}$ given in (30).

Consider Step 2 of the first iteration (i.e., $n = 1$) of the single-layer iterative algorithm. Since $\mathbf{V}^{(1)}$ is the optimal solution of Problem (32) with given $\bar{\omega}^{(0)}$, $\mathbf{U}^{(0)}$ and $\mathbf{W}^{(0)}$, we have

$$\begin{aligned} & f_1(\mathbf{V}^{(1)}, \mathbf{U}^{(0)}, \mathbf{W}^{(0)}, \bar{\omega}^{(0)}, \eta^{(0)}) \\ & \geq f_1(\mathbf{V}^{(0)}, \mathbf{U}^{(0)}, \mathbf{W}^{(0)}, \bar{\omega}^{(0)}, \eta^{(0)}) \\ & = f_2(\mathbf{V}^{(0)}, \bar{\omega}^{(0)}, \eta^{(0)}), \end{aligned} \quad (\text{D.4})$$

where (D.4) follows by using Lemma 1.

In Step 3 of the single-layer iterative algorithm, $\mathbf{U}^{(1)}$ and $\mathbf{W}^{(1)}$ are updated by using (14) and (15) with fixed $\mathbf{V}^{(1)}$. Then, by Lemma 1, we have

$$\begin{aligned} \sum_{k \in \mathcal{S}} R_k(\mathbf{V}^{(1)}) &= \sum_{k \in \mathcal{S}} h_k(\mathbf{V}^{(1)}, \mathbf{U}_k^{(1)}, \mathbf{W}_k^{(1)}) \\ &\geq \sum_{k \in \mathcal{S}} h_k(\mathbf{V}^{(1)}, \mathbf{U}_k^{(0)}, \mathbf{W}_k^{(0)}). \end{aligned} \quad (\text{D.5})$$

Then, by combining (D.4) and (D.5), we have

$$\begin{aligned} & f_2(\mathbf{V}^{(1)}, \bar{\omega}^{(0)}, \eta^{(0)}) \\ &= \sum_{k \in \mathcal{S}} R_k(\mathbf{V}^{(1)}) - \eta^{(0)} \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{S}} \left((1 + \bar{\omega}^{(0)}) \left\| \mathbf{V}_{i,k}^{(1)} \right\|_F^2 \right) \end{aligned} \quad (\text{D.6})$$

$$\begin{aligned} & \geq \sum_{k \in \mathcal{S}} h_k(\mathbf{V}^{(1)}, \mathbf{U}_k^{(0)}, \mathbf{W}_k^{(0)}) \\ & \quad - \eta^{(0)} \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{S}} \left((1 + \bar{\omega}^{(0)}) \left\| \mathbf{V}_{i,k}^{(1)} \right\|_F^2 \right) \end{aligned} \quad (\text{D.7})$$

$$= f_1(\mathbf{V}^{(1)}, \mathbf{U}^{(0)}, \mathbf{W}^{(0)}, \bar{\omega}^{(0)}, \eta^{(0)}) \quad (\text{D.8})$$

$$\geq f_2(\mathbf{V}^{(0)}, \bar{\omega}^{(0)}, \eta^{(0)}). \quad (\text{D.9})$$

Based on above equalities, we have the following chain

inequalities

$$\begin{aligned} & f_3(\mathbf{V}^{(1)}, \eta^{(0)}) \\ & \geq \sum_{k \in \mathcal{S}} R_k(\mathbf{V}^{(1)}) \\ & \quad - \eta^{(0)} \left(\sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{S}} (1 + \bar{\omega}_{i,k}^{(0)}) \|\mathbf{V}_{i,k}^{(1)}\|_F^2 + C_{i,k}^{(0)} \right) \quad (\text{D.10}) \end{aligned}$$

$$\begin{aligned} & \geq \sum_{k \in \mathcal{S}} R_k(\mathbf{V}^{(0)}) - \\ & \quad \eta^{(0)} \left(\sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{S}} (1 + \bar{\omega}_{i,k}^{(0)}) \|\mathbf{V}_{i,k}^{(0)}\|_F^2 + C_{i,k}^{(0)} \right) \quad (\text{D.11}) \end{aligned}$$

$$= f_3(\mathbf{V}^{(0)}, \eta^{(0)}), \quad (\text{D.12})$$

where (D.10) follows by using (28), (D.11) follows by using (D.9), and $C_{i,k}^{(0)}$ is given by

$$C_{i,k}^{(0)} = P_{\text{th}} \left[f_\delta \left(\|\mathbf{V}_{i,k}^{(0)}\|_F^2 \right) - f_\delta' \left(\|\mathbf{V}_{i,k}^{(0)}\|_F^2 \right) \|\mathbf{V}_{i,k}^{(0)}\|_F^2 \right]. \quad (\text{D.13})$$

Hence, we have

$$f_3(\mathbf{V}^{(1)}, \eta^{(0)}) \geq f_3(\mathbf{V}^{(0)}, \eta^{(0)}). \quad (\text{D.14})$$

Subtracting for both sides of (D.14) with $\eta^{(0)} P_C$ yields

$$\begin{aligned} & \sum_{k \in \mathcal{S}} R_k(\mathbf{V}^{(1)}) - \eta^{(0)} \tilde{P}_{\text{total}}(\mathbf{V}^{(1)}) \\ & \geq \sum_{k \in \mathcal{S}} R_k(\mathbf{V}^{(0)}) - \eta^{(0)} \tilde{P}_{\text{total}}(\mathbf{V}^{(0)}) = 0, \quad (\text{D.15}) \end{aligned}$$

where the last equality holds, since $\eta^{(0)}$ is updated according to (36). Then we have

$$\eta^{(1)} = \frac{\sum_{k \in \mathcal{S}} R_k(\mathbf{V}^{(1)})}{\tilde{P}_{\text{total}}(\mathbf{V}^{(1)})} \geq \eta^{(0)}. \quad (\text{D.16})$$

By using a similar method, we can show that

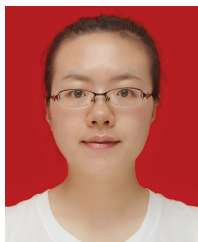
$$\eta^{(0)} \leq \eta^{(1)} \leq \eta^{(2)} \leq \dots \quad (\text{D.17})$$

Due to the limited power, the sum rate is also limited. Hence, the EE value is upper-bounded. As a result, the single-layer iterative algorithm is guaranteed to converge.

REFERENCES

- [1] X. H. You, D. M. Wang, B. Sheng, X. Q. Gao, X. S. Zhao, and M. Chen, "Cooperative distributed antenna systems for mobile communications [coordinated and distributed MIMO]," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 35–43, 2010.
- [2] R. Heath, S. Peters, Y. Wang, and J. Zhang, "A current perspective on distributed antenna systems for the downlink of cellular systems," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 161–167, 2013.
- [3] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.
- [4] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.
- [5] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 524–540, Fourth 2011.
- [6] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.
- [7] C. He, G. Y. Li, F. C. Zheng, and X. You, "Energy-efficient resource allocation in OFDM systems with distributed antennas," *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1223–1231, 2014.
- [8] W. Feng, Y. Chen, N. Ge, and J. Lu, "Optimal energy-efficient power allocation for distributed antenna systems with imperfect CSI," *IEEE Trans. Veh. Technol.*, vol. PP, no. 99, pp. 1–1, 2015.
- [9] C. He, G. Y. Li, F. C. Zheng, and X. You, "Power allocation criteria for distributed antenna systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5083–5090, Nov 2015.
- [10] H. Kim, S. R. Lee, C. Song, K. J. Lee, and I. Lee, "Optimal power allocation scheme for energy efficiency maximization in distributed antenna systems," *IEEE Trans. Commun.*, vol. 63, no. 2, pp. 431–440, 2015.
- [11] J. Wu, J. Liu, W. Li, and X. You, "Low-complexity power allocation for energy efficiency maximization in DAS," *IEEE Commun. Lett.*, vol. 19, no. 6, pp. 925–928, 2015.
- [12] Y. Dong, H. Zhang, M. J. Hossain, J. Cheng, and V. C. M. Leung, "Energy efficient resource allocation for OFDMA full duplex distributed antenna systems with energy recycling," in *Proc. IEEE GLOBECOM*, 2015, pp. 1–6.
- [13] X. Dang, Y. Wang, and X. Yu, "Suboptimal energy efficient power allocation scheme for distributed antenna system with antenna selection," *Electronics Letters*, vol. 52, no. 4, pp. 264–266, 2016.
- [14] H. Kim, E. Park, H. Park, and I. Lee, "Beamforming and power allocation designs for energy efficiency maximization in MISO distributed antenna systems," *IEEE Communications Letters*, vol. 17, no. 11, pp. 2100–2103, 2013.
- [15] H. Ren, N. Liu, C. Pan, and C. He, "Energy efficiency optimization for MIMO distributed antenna systems," *IEEE Trans. Veh. Technol.*, vol. PP, no. 99, pp. 1–1, 2016.
- [16] H. Ren, N. Liu, and C. Pan, "Energy efficient transmission for multicast services in MISO distributed antenna systems," *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 756–759, 2016.
- [17] C. Yoon and D. H. Cho, "Energy efficient beamforming and power allocation in dynamic TDD based C-RAN system," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1806–1809, 2015.
- [18] L. Venturino, N. Prasad, and X. Wang, "Coordinated linear beamforming in downlink multi-cell wireless networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 4, pp. 1451–1461, 2010.
- [19] Y. Cheng, M. Pesavento, and A. Philipp, "Joint network optimization and downlink beamforming for CoMP transmissions using mixed integer conic programming," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3972–3987, 2013.
- [20] B. Dai and W. Yu, "Energy efficiency of downlink transmission strategies for cloud radio access networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 1037–1050, 2016.
- [21] D. W. K. Ng, Y. Wu, and R. Schober, "Power efficient resource allocation for full-duplex radio distributed antenna networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2896–2911, 2016.
- [22] P. Cao, W. Liu, J. S. Thompson, C. Yang, and E. A. Jorswieck, "Semidynamic green resource management in downlink heterogeneous networks by group sparse power control," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1250–1266, 2016.
- [23] C. Jiang and L. J. Cimini, "Antenna selection for energy-efficient mimo transmission," *IEEE Wireless Communications Letters*, vol. 1, no. 6, pp. 577–580, December 2012.
- [24] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green Cloud-RAN," *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2809–2823, 2014.
- [25] S. He, Y. Huang, J. Wang, L. Yang, and W. Hong, "Joint antenna selection and energy-efficient beamforming design," *IEEE Signal Processing Letters*, vol. 23, no. 9, pp. 1165–1169, Sept 2016.
- [26] J. Xu and L. Qiu, "Energy efficiency optimization for MIMO broadcast channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 690–701, 2013.
- [27] S. He, Y. Huang, S. Jin, and L. Yang, "Coordinated beamforming for energy efficient transmission in multicell multiuser systems," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 4961–4971, 2013.
- [28] S. He, Y. Huang, H. Wang, S. Jin, and L. Yang, "Leakage-aware energy-efficient beamforming for heterogeneous multicell multiuser systems," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1268–1281, 2014.
- [29] Y. Li, Y. Tian, and C. Yang, "Energy-efficient coordinated beamforming under minimal data rate constraint of each user," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2387–2397, 2015.

- [30] J. Tang, D. K. C. So, E. Alsusa, K. A. Hamdi, and A. Shojafard, "Energy efficiency optimization with interference alignment in multi-cell MIMO interfering broadcast channels," *IEEE Trans. Commun.*, vol. 63, no. 7, pp. 2486–2499, 2015.
- [31] Q. D. Vu, L. N. Tran, R. Farrell, and E. K. Hong, "Energy-efficient zero-forcing precoding design for small-cell networks," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 790–804, 2016.
- [32] Y. F. Liu, Y. H. Dai, and Z. Q. Luo, "Coordinated beamforming for miso interference channel: Complexity analysis and efficient algorithms," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1142–1157, March 2011.
- [33] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.
- [34] E. Matakani, N. Sidiropoulos, Z.-Q. Luo, and L. Tassiulas, "Convex approximation techniques for joint multiuser downlink beamforming and admission control," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2682–2693, July 2008.
- [35] 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 Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, 2011.
- [36] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [37] D. P. Bertsekas, *Nonlinear programming*. Belmont: Athena scientific, 1999.
- [38] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001.
- [39] F. Rinaldi, F. Schoen, and M. Sciafrone, "Concave programming for minimizing the zero-norm over polyhedral sets," *Computational Optimization and Applications*, vol. 46, no. 3, pp. 467–486, 2010.
- [40] M. Tao, E. Chen, H. Zhou, and W. Yu, "Content-centric sparse multi-cast beamforming for cache-enabled cloud ran," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6118–6131, Sept 2016.
- [41] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, 1967.
- [42] X. Wang, P. Zhu, and M. Chen, "Antenna location design for generalized distributed antenna systems," *IEEE Communications Letters*, vol. 13, no. 5, pp. 315–317, May 2009.
- [43] H. Ren, N. Liu, C. Pan, and C. He, "Energy efficiency optimization for MIMO distributed antenna systems," in *2015 IEEE Globecom Workshops (GC Wkshps)*, 2015, pp. 1–7.
- [44] L. Liu, R. Chen, S. Geirhofer, K. Sayana, Z. Shi, and Y. Zhou, "Downlink mimo in lte-advanced: Su-mimo vs. mu-mimo," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 140–147, 2012.
- [45] C. Pan, W. Xu, J. Wang, H. Ren, W. Zhang, N. Huang, and M. Chen, "Totally distributed energy-efficient transmission in MIMO interference channels," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6325–6338, 2015.
- [46] D. P. Bertsekas, A. Nedi, A. E. Ozdaglar *et al.*, "Convex analysis and optimization," 2003.
- [47] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.



Hong Ren received the M.S. degree in information science and engineering from Southeast University, Nanjing, China, in 2014. She is currently working toward the Ph.D. degree at the National Mobile Communications Research Laboratory (NCRL), Southeast University, China. Since October 2016, she has been working as a visiting student in the School of Electronics and Computer Science, University of Southampton, UK. Her research interests lie in the areas of communication and signal processing, including green communication systems, cooperative transmission and cross layer transmission optimization.



Nan Liu received the B.Eng. degree in electrical engineering from Beijing University of Posts and Telecommunications, Beijing, P. R. China in 2001, and the Ph.D. degree in electrical and computer engineering from University of Maryland, College Park, MD in 2007.

From 2007–2008, she was a postdoctoral scholar in the Wireless Systems Lab, Department of Electrical Engineering, Stanford University. In 2009, she became a professor in the National Mobile Communications Research Laboratory, School of Information Science and Engineering in Southeast University, Nanjing, China. Her research interests are in network information theory for wireless networks, SON algorithms for next generation cellular networks and energy-efficient communications.



Cunhua Pan received his B.S. and Ph.D. degrees in school of Information Science and Engineering, Southeast University, Nanjing, China, in 2010 and 2015, respectively. From 2015 to 2016, he worked as a research associate in University of Kent, UK. He is currently a postdoc with Queen Mary University of London, UK.

His research interests include C-RAN, mm-Wave communications, NOMA, D2D, large-scale MIMO and cloud computing. He is a TPC member of IEEE

ICC and Globecom during 2015–2017.



Lajos Hanzo (<http://www-mobile.ecs.soton.ac.uk>) FREng, FIEEE, FIET, Fellow of EURASIP, DSc received his degree in electronics in 1976 and his doctorate in 1983. In 2009 he was awarded an honorary doctorate by the Technical University of Budapest and in 2015 by the University of Edinburgh. In 2016 he was admitted to the Hungarian Academy of Science. During his 40-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the School of

Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised 111 PhD students, co-authored 18 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1676 research contributions at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 60-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council's Advanced Fellow Grant and the Royal Society's Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. During 2008 - 2012 he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing. For further information on research in progress and associated publications please refer to <http://www-mobile.ecs.soton.ac.uk> Lajos has 30 000+ citations and an H-index of 68.