Joint Pilot Allocation and Robust Transmission Design for Ultra-Dense User-Centric TDD C-RAN With Imperfect CSI

Cunhua Pan, *Member, IEEE*, Hani Mehrpouyan, *Member, IEEE*, Yuanwei Liu, *Member, IEEE*, Maged Elkashlan, *Member, IEEE*, and Arumugam Nallanathan, *Fellow, IEEE*

Abstract—This paper considers the unavailability of complete channel state information (CSI) in ultra-dense cloud radio access networks. The user-centric cluster is adopted to reduce the computational complexity, while the incomplete CSI is considered to reduce the heavy channel training overhead, where only large-scale inter-cluster CSI is available. Channel estimation for intra-cluster CSI is also considered, where we formulate a joint pilot allocation and user equipment (UE) selection problem to maximize the number of admitted UEs with fixed number of pilots. A novel pilot allocation algorithm is proposed by considering the multi-UE pilot interference. Then, we consider robust beam-vector optimization problem subject to UEs' data rate requirements and fronthaul capacity constraints, where the channel estimation error and incomplete inter-cluster CSI are considered. The exact data rate is difficult to obtain in closed form, and instead we conservatively replace it with its lower-bound. The resulting problem is non-convex, combinatorial, and even infeasible. A practical algorithm, based on UE selection, successive convex approximation and semi-definite relaxation approach, is proposed to solve this problem with guaranteed convergence. We strictly prove that the semidefinite relaxation is tight with probability 1. Finally, extensive simulation results are presented to show the fast convergence of our proposed algorithm and demonstrate its superiority over the existing algorithms.

Index Terms—Ultra-dense networks (UDN), C-RAN, imperfect CSI, pilot reuse, virtual cell, fronthaul capacity constraints.

I. INTRODUCTION

THE fifth-generation (5G) wireless system is excepted to offer a 1000x increase in capacity before 2020 to meet the ever-growing capacity demand in mobile wireless systems [1]. Cloud radio access network (C-RAN) is an emerging network architecture that shows significant promises in achieving this goal [2]. A typical C-RAN has three main components:

1) Baseband unit (BBU) pool hosted in a cloud data center; 2) Radio remote heads (RRHs) geographically distributed over the coverage area; 3) High-bandwidth low-latency fronthaul links that connect the RRHs to the BBU pool. The key feature of C-RAN is that the baseband signal processing in traditional base stations (BSs) are migrated to the BBU pool so that the conventional BSs can be replaced by low-functionality RRHs, which transmit/receive radio frequency signals. Due to their simple functionalities, RRHs can be densely deployed to provide ubiquitous service access for a large number of user equipment (UEs) in hot spots such as shopping malls, stadium, etc. Unlike the conventional ultra-dense small cells where co-channel interference (CCI) is a limiting factor [3], under ultra-dense C-RAN architecture, centralized signal processing such as coordinated multi-point (CoMP) technique can be adopted to effectively mitigate the CCI thanks to the powerful computational capability at the BBU pool. Ultra-dense C-RANs are usually defined as those networks that there are more RRHs than the UEs [4]–[6], i.e., $\lambda_r > \lambda_u$, where λ_r and λ_u represent the density of RRHs and UEs, respectively. However, these works have not provided quantitative values for those densities. Ge *et. al.* [7] showed that the density of RRHs in 5G ultra-dense networks is expected to be up to 40-50 RRHs/km^2 to satisfy the seamless coverage requirements.

Recently, transmission design has been extensively studied to deal with various technical issues in conventional C-RAN such as reducing network power consumption in [8], or/and tackling limited fronthaul capacity constraints in [9], or/and reducing the computational complexity [10] where user-centric cluster method is adopted. However, the most troublesome challenge is that dense C-RAN requires a large amount of CSI to facilitate centralized signal processing. The acquisition of these CSI requires large amount of training overhead that increases with the network size. Results in [11] showed that the network performance may even decrease with increasing number of RRHs when taking into account the cost of acquiring CSI. One promising way to reduce the training overhead is to consider the incomplete CSI case, where each UE only measures its CSI from the RRHs in its serving cluster (named intra-cluster CSI) and only tracks the large-scale fading (path loss and shadowing) for the CSI outside its cluster (named inter-cluster CSI). Recently, transmission design considering the incomplete CSI has attracted extensive research interests [12]–[14].

However, intra-cluster CSI was assumed to be perfectly known in [12]–[14], which is impractical for dense C-RAN. The reasons are given as follows. To estimate the intra-cluster CSI in time-division duplex (TDD) C-RAN, the uplink training pilot sequences sent from the UEs that share at least one serving RRH should be mutually orthogonal so that the BBU pool can differentiate the CSI of the shared RRH to the corresponding UEs. One naive method is to assign all the UEs with mutually orthogonal pilot sequences. However, the number of time slots required for training will increase linearly with the number of UEs, which is unaffordable for ultra-dense C-RAN with large number of UEs. Hence, to support more UEs, one should allow some UEs to reuse the same pilots. The pilot reuse scheme will incur the pilot contamination issue, which results in sizeable channel estimation error. Hence, in this paper, besides the incomplete inter-cluster CSI, we also consider channel estimation procedure of intra-cluster CSI for a user-centric ultra-dense TDD C-RAN by designing pilot reuse scheme, and then design beam-vectors based on imperfect intra-cluster CSI. To the best of our knowledge, this paper is the first attempt to unify the channel estimation and transmit beam-vector design for ultra-dense C-RAN into one general framework.

Pilot reuse design has been extensively studied in massive MIMO [15]–[19], where the proposed schemes are mainly based on the idea of assigning the same pilot to the UEs with different angles due to the special feature of massive MIMO. Due to the limited number of antennas at the RRHs, the pilot reuse schemes designed for massive MIMO cannot be extended to the ultra-dense C-RAN network. Recently, [20] proposed a novel pilot allocation scheme for user-centric C-RAN based on the graph coloring algorithm such as the Dsatur algorithm [21]. The aim of this algorithm is to minimize the number of required pilots for a given set of UEs under some practical constraints. However, for the sake of low implementation cost, the current LTE standards [22] suggests that the proportion of pilots allocated to the UEs for training is fixed during one coherence time, i.e., where the training overhead of 1% is considered in LTE specification with a 10 ms training period [22]. Hence, we should consider the joint pilot allocation and UE selection method to maximize the number of admitted UEs under a fixed number of available pilots. In addition, the transmission beam-vector design is not considered in [20]. In this paper, we also consider robust beam-vector design under the pilot reuse scheme by taking into account the joint effects of pilot contamination and incomplete inter-cluster CSI. Unfortunately, due to the imperfect intra-cluster CSI, the beam-vector design based on the well-known weighted minimum mean square error (WMMSE) method [23], that has been widely applied in the C-RAN setup [9], [10], [13], [24]–[26] with perfect intra-cluster CSI, cannot be extended to solve the robust beam-vector optimization problem considered in this paper. Specifically, the contributions of this paper are given as follows:

1) We consider a two-stage optimization problem for dense C-RAN, i.e., channel estimation for intra-cluster CSI in Stage I and robust transmit beam-vector design in Stage II. Due to the constraint that the UEs served by at least one common RRH should be allocated with different pilots, some UEs may not be supported for given small number of available pilots τ, considering large number of UEs in ultra-dense C-RAN. Hence, in Stage I, we formulate a joint UE selection and pilot allocation optimization problem to maximize the number of admitted UEs under several practical constraints. To solve this problem, we first apply the Dsatur algorithm to find the minimum number of required pilots n^* for supporting all UEs in the network. If $n^* > \tau$, one novel UE removal method is proposed by considering the multi-UE pilot interference. If $n^* < \tau$, all UEs can be admitted and a novel algorithm is proposed to fully reallocate all available pilots to further reduce pilot contamination.

2) Based on the results from Stage I, in Stage II transmit beam-vectors are designed to minimize the transmit power based on the imperfect intra-cluster CSI and incomplete inter-cluster CSI. Both UEs' data rate and fronthaul capacity constraints are considered. It is observed that the expectation with respect to the channel estimation error and small-scale inter-cluster CSI makes it difficult to derive closed-form data rate expressions. To deal with this difficulty, we conservatively replace the data rate of each UE with its lower bound obtained by the Jensen's inequality. Furthermore, considering that the problem may be infeasible, we construct an alternative optimization problem that simultaneously maximizes the number of admitted UEs and minimizes the transmit power. Then, one iterative algorithm based on successive convex approximation (SCA) technique [27] and semi-definite relaxation approach [28] is proposed to solve this optimization problem with guaranteed convergence. We prove that semidefinite relaxation is tight with probability 1.

This paper is organized as follows. In Section II, we introduce the signal transmission model along with the channel estimation for intra-cluster CSI. In Section III, we propose the joint pilot allocation and UE selection algorithm. In Section IV, we develop a robust transmit beam-vector optimization algorithm while taking the channel estimation error into account. Extensive simulation results are provided in Section V. Finally, this paper is concluded in Section VI.

Notations: For a set $\mathcal{A}$, $|\mathcal{A}|$ is the cardinality of $\mathcal{A}$, while for a complex number x, $|x|$ represents the magnitude of x. 's.t.' denotes 'subject to'. $\mathbb{E}_{\{x\}}\{y\}$ denotes the expectation of y over x. The complex Gaussian distribution is denoted as $\mathcal{CN}(\cdot, \cdot)$ and $\mathbb{C}$ is used to represent the complex set. The lower-case bold letters means vectors and upper-case bold letters denote matrices.

II. System Model

A. Signal Transmission Model

Consider a downlink dense TDD C-RAN with I RRHs and K UEs as shown in Fig. 1. Each RRH and each UE are equipped with M transmit antennas and a single receive antenna, respectively. Denote the set of RRHs and UEs as
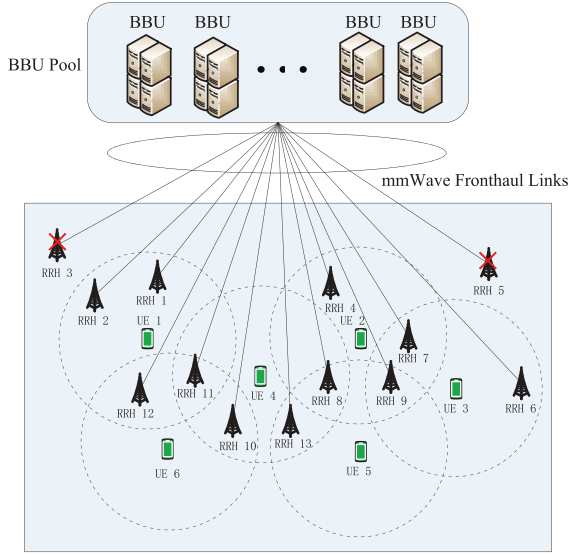
Fig. 1. Illustration of a C-RAN with thirteen RRHs and six UEs, i.e., $I = 13$, $K = 6$. To reduce the operation cost, mmWave communication is used to establish the fronthaul link that connects each RRH to the BBU pool. To reduce the complexity, each UE is served by the RRHs within the dashed circle centered around the UE.

$I = \{1, \cdots, I\}$ and $\bar{\mathcal{U}} = \{1, \cdots, K\}$, respectively. Each RRH is connected to the BBU pool through the wireless fronthaul links employing mmWave communication technologies, which are represented by dark lines in Fig. 1. The BBU pool is assumed to have all UEs' data and distribute each UE's data to a carefully selected set of RRHs through the wireless fronthaul links. We further assume that all the RRHs transmit to the UEs with the carrier frequency below 6 GHz. As a result, the transmission from the BBU pool to the RRHs and the transmission from RRHs to the UEs can take place at the same time without interfering with each other.

The set of UEs that are admitted in this network is denoted by $\mathcal{U} \subseteq \bar{\mathcal{U}}$. To reduce the computational cost for dense C-RAN, the user-centric cluster method is adopted, e.g., each UE is only served by its nearby RRHs since distant RRHs contribute little to the UE's signal quality due to the severe path loss. Denote $I_k \subseteq I$ and $\mathcal{U}_i \subseteq \mathcal{U}$ as the set of RRHs that potentially serve UE k and the set of UEs that are potentially served by RRH i, respectively. Note that for ultra-dense C-RAN with many UEs and RRHs, the clusters for the UEs may overlap with each other, i.e., there may exist two distinct UEs k and k' that $I_k \cap I_{k'} \neq \emptyset$, for $\forall k, k' \in \mathcal{U}$. In practice, the clusters for each UE are formed based on the long term channel state information (CSI) such as large-scale fading [29]. Specifically, each UE measures its average channel gains to the RRHs and sends them to the BBU pool. Then, based on the obtained average channel gains, the BBU pool decides the cluster for each UE based on cluster formation techniques such as the ones in [29]. The dense C-RAN can be deployed in hot areas such as shopping mall and stadium, where the average channel gains change slowly and depends mainly on UEs' locations [3]. Hence, the clusters can be formed on a long time scale and are assumed to be fixed.

Denote $\mathbf{h}_{i,k} \in \mathbb{C}^{M \times 1}$ and $\mathbf{w}_{i,k} \in \mathbb{C}^{M \times 1}$ as the channel vector and the beam-vector from RRH i to UE k, respectively. Then,

the baseband received signal at UE k is given by

$$y_k = \underbrace{\sum\nolimits_{i \in I_k} \mathbf{h}_{i,k}^{\mathrm{H}} \mathbf{w}_{i,k} s_k}_{\text{desired signal}} + \underbrace{\sum\nolimits_{l \neq k, l \in \mathcal{U}} \sum\nolimits_{i \in I_l} \mathbf{h}_{i,k}^{\mathrm{H}} \mathbf{w}_{i,l} s_l}_{\text{interference}} + z_k,$$

(1)

where s_k is the data symbol for UE k, z_k is the additive complex white Gaussian noise that is assumed to follow the distribution $\mathcal{CN}(0, \sigma_k^2)$. It is assumed that $\mathbb{E}\{|s_k|^2\} = 1$ and $\mathbb{E}\{s_{k_1} s_{k_2}\} = 0$ for $k_1 \neq k_2, \forall k_1, k_2 \in \mathcal{U}$. The channel vector $\mathbf{h}_{i,k}$ can be decomposed as $\mathbf{h}_{i,k} = \sqrt{\alpha_{i,k}} \bar{\mathbf{h}}_{i,k}$, where $\alpha_{i,k}$ denotes the large-scale channel gain that includes the path loss and shadowing, and $\bar{\mathbf{h}}_{i,k}$ denotes the small-scale fading following the distribution of $\mathcal{CN}(\mathbf{0}, \mathbf{I})$.

B. Channel Estimation for Intra-Cluster CSI

To design the beam-vectors for all the UEs, the overall CSI over the network should be available at the BBU pool. However, it is an unaffordable task to acquire all CSI for dense C-RAN with large number of RRHs and UEs due to limited training resources. To handle this issue, we assume that each UE k only measures the CSI to the RRHs in its cluster I_k, while for the RRHs out of the cluster we assume that the BBU pool only knows the large-scale channel gains, i.e., $\{\alpha_{i,k}, \forall i \in I \setminus I_k, k \in \mathcal{U}\}$. Note that the large-scale channel gains are used to manage the multiuser interference. If these gains are set to zero, the resulting solution may be too optimistic and may not satisfy UEs' quality of service (QoS) requirements as shown in [14].

In this paper, we assume that τ time slots are used for channel training or equivalently the length of the pilot sequences is τ, which is assumed to be smaller than the number of total UEs K for the dense network, i.e., $\tau < K$. Note that the maximum number of orthogonal pilot sequences is equal to the pilot length τ. Hence, in orthogonal pilot assignment scheme, the maximum number of UEs that can be simultaneously served by C-RAN is upper bounded by τ. To support more UEs, we propose a pilot reuse scheme in this paper.

Denote the available pilot set as $\mathcal{Q} = \{1, 2, \cdots, \tau\}$, and the corresponding orthogonal pilot sequences as $\mathbf{Q} = [\mathbf{q}_1, \cdots, \mathbf{q}_\tau] \in \mathbb{C}^{\tau \times \tau}$, which satisfies the orthogonal condition, i.e., $\mathbf{Q}^H \mathbf{Q} = \mathbf{I}$. We denote an arbitrary pilot reuse scheme as $\mathcal{P}(\mathcal{U}, \mathcal{Q}) = \{(k, \pi_k) : k \in \mathcal{U}, \pi_k \in \mathcal{Q}\}$, where (k, π_k) denotes that UE k is allocated with pilot sequence $\mathbf{q}_{\pi_k}$. In addition, define $\mathcal{K}_\pi = \{k : \pi_k = \pi\}$ as the set of UEs that reuse the πth pilot sequences.

Given the pilot reuse scheme $\mathcal{P}(\mathcal{U}, \mathcal{Q})$, the UEs transmit their pilot sequences to the RRHs to estimate the channels during the uplink training phase. Specifically, the received pilot signal at RRH i is

$$\mathbf{Y}_i = \sum\nolimits_{k \in \mathcal{U}} \sqrt{p_t} \mathbf{h}_{i,k} \mathbf{q}_{\pi_k}^{\mathrm{H}} + \mathbf{N}_i,$$

(2)

where p_t is the pilot transmit power at each UE, and $\mathbf{N}_i \in \mathbb{C}^{M \times \tau}$ is the additive Gaussian noise matrix received during the training phase, whose elements are independently generated and follow the distributions of $\mathcal{CN}(0, \sigma^2)$.

To obtain channel $\mathbf{h}_{i,k}, i \in I_k$, the BBU pool first multiplies $\mathbf{Y}_i$ by $\mathbf{q}_{\pi_k}$, which yields

$$\mathbf{y}_{i,k} = \frac{1}{\sqrt{p_t}}\mathbf{Y}_i\mathbf{q}_{\pi_k}$$

$$= \mathbf{h}_{i,k} + \sum_{l \in \mathcal{K}_{\pi_k}\backslash\{k\}} \mathbf{h}_{i,l} + \mathbf{n}_i, \qquad (3)$$

where $\mathbf{n}_i = \frac{1}{\sqrt{p_t}}\mathbf{N}_i\mathbf{q}_{\pi_k}$. Since $\mathbf{q}_{\pi_k}$ is a unit-norm vector, it is easy to show that $\mathbf{n}_i$ is still Gaussian distribution whose elements are independently and identically distributed as $\mathcal{CN}(0, \frac{\sigma^2}{p_t})$. The MMSE estimate of channel $\mathbf{h}_{i,k}$ is given by

$$\hat{\mathbf{h}}_{i,k} = \frac{\alpha_{i,k}}{\sum_{l \in \mathcal{K}_{\pi_k}} \alpha_{i,l} + \hat{\sigma}^2}\mathbf{y}_{i,k} \qquad (4)$$

with $\hat{\sigma}^2 = \sigma^2/p_t$. According to the property of MMSE estimate [30], channel estimation error $\tilde{\mathbf{h}}_{i,k} = \mathbf{h}_{i,k} - \hat{\mathbf{h}}_{i,k}$ is independently distributed as $\mathcal{CN}(\mathbf{0}, \delta_{i,k}\mathbf{I})$, where $\delta_{i,k}$ is given by

$$\delta_{i,k} = \frac{\alpha_{i,k}\left(\sum_{l \in \mathcal{K}_{\pi_k}\backslash\{k\}} \alpha_{i,l} + \hat{\sigma}^2\right)}{\sum_{l \in \mathcal{K}_{\pi_k}} \alpha_{i,l} + \hat{\sigma}^2}. \qquad (5)$$

In the next two sections, we propose a two-stage optimization method to optimize the transmission in dense C-RAN. Specifically, in Stage I, we provide a joint pilot allocation and UE selection algorithm to maximize the number of admitted UEs with fixed number of available pilots in Section III; in Stage II, based on the results from Stage I, a robust beam-vector optimization algorithm is proposed to minimize the transmit power in Section IV, while considering the pilot contamination.

III. STAGE I: PILOT ALLOCATION AND UE SELECTION DESIGN

In this stage, we aim to design the pilot allocation scheme and UE selection method to maximize the number of admitted UEs under several practical constraints with fixed number of pilot sequences.

A. Problem Formulation

First, in conventional TDD communication systems, all UEs in the same macro-cell should be allocated with orthogonal training resources for the macro-BS to distinguish the channels from the UEs [31]. Similarly, in dense C-RAN, the UEs served by the same RRH should be allocated with orthogonal training sequences. This kind of constraints can be mathematically expressed as follows:

$$C1: \mathbf{q}_{\pi_k}^H\mathbf{q}_{\pi_{k'}} = 0, \quad \text{for } k, k' \in \mathcal{U}_i, \ k \neq k', \ \forall i \in I. \qquad (6)$$

For example, in Fig. 1, RRH 11 jointly serves UE 1, UE 4 and UE 6. Then the following conditions should be satisfied: $\mathbf{q}_{\pi_1}^H\mathbf{q}_{\pi_4} = 0, \mathbf{q}_{\pi_1}^H\mathbf{q}_{\pi_6} = 0$, and $\mathbf{q}_{\pi_4}^H\mathbf{q}_{\pi_6} = 0$.

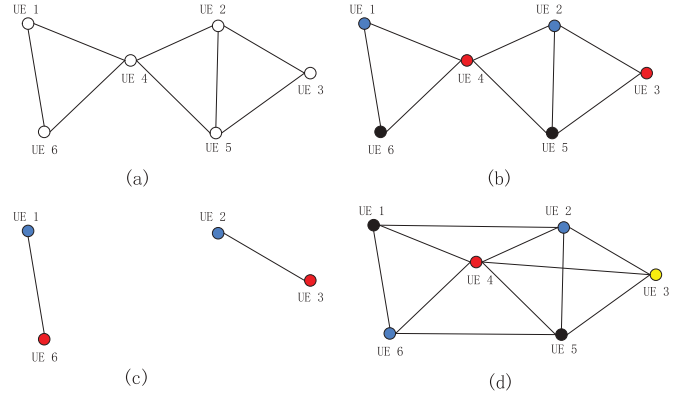Second, to reduce the channel estimation error, the reuse times for each pilot sequence should be restricted under a

Fig. 2. (a) Construction of the undirected graph for the network in Fig. 1, where the vertexes denotes the UEs and any two UEs sharing at least one RRH should be connected with each other; (b) The colored graph after applying the Dsatur algorithm [20] with $n_{\max} = 2$, the minimum number of pilot sequences required is $n^* = 3$; (c) The UE selection and pilot reallocation result after applying Algorithm 1 when $\tau = 2$ and $n_{\max} = 2$, the number of selected UEs is four; (d) The pilot reallocation result after applying Algorithm 2 when $\tau = 4$ and $n_{\max} = 2$. In this colored graph, UE 1 and UE 2 use different pilots to avoid the pilot interference, and the same holds for UE 5 and UE6. In addition, UE 3 and UE 4 are allocated with different pilots, which additionally reduces the pilot interference.

predefined value. Denote the number of UEs that share pilot l as n_l, then the constraints can be expressed as

$$C2: n_l \leq n_{\max}, \quad \forall l \in Q. \qquad (7)$$

where $n_{\max}$ is the maximum reuse time for each pilot. Constraint C2 means that we should reuse the pilot sequences in a fair way to avoid the extreme case where one pilot sequence is reused many times while there are several unused pilot sequences. This extreme case will deteriorate the system performance.

Here, we aim to find the maximum number of UEs that can be served with a fixed number of pilot sequences. Hence, the joint pilot allocation and UE selection problem at Stage I can be formulated as

$$\mathcal{P}_1: \max_{\mathcal{U} \subseteq \bar{\mathcal{U}}, \mathcal{P}(\mathcal{U}, Q)} |\mathcal{U}|$$
$$\text{s.t. } C1, C2. \qquad (8a)$$

Problem $\mathcal{P}_1$ can be considered as a resource-assignment problem, which is difficult to solve in general. The optimal solution can be obtained by the exhaustive search method, where one should check all possible pilot allocation schemes and choose one with the maximum number of admitted UEs while satisfying the constraints C1 and C2. However, the complexity of the exhaustive search method increases exponentially with the number of UEs, which is not practical for dense C-RAN. In the following, we provide a low-complexity pilot allocation scheme, which is suitable for practical applications.

B. Pilot Allocation Scheme

Constraints C1 can be represented by a $K \times K$ binary matrix $\mathbf{B}$, where the kth row and k'th column of matrix $\mathbf{B}$ is given by

$$b_{k,k'} = \begin{cases} 1, & \text{if } I_k \cap I_{k'} \neq \emptyset \text{ and } k \neq k' \\ 0, & \text{otherwise.} \end{cases} \qquad (9)$$

When any two UEs are served by at least one common RRH, the corresponding element in $\mathbf{B}$ should be one and these two UEs should be allocated with different pilots. Otherwise, the element is zero and the UEs can share the same pilot. Based on the matrix $\mathbf{B}$, we can construct an undirected graph to describe the relationship between any two UEs for constraint C1, where any two UEs that are served by at least one common RRH should be connected with each other. For the network in Fig. 1, the corresponding graph is shown in Fig. 2 (a), where any two UEs that are served by at least one common RRH are connected together. The graph coloring algorithm such as the Dsatur algorithm [21], which aims for coloring the vertexes of a graph with the minimum number of different colors under the same constraints in C1 and C2 for a given set of UEs, has been used in [20] for user-centric C-RAN to design the pilot allocation. For the undirected graph in Fig. 2 (a), the final colored graph is shown in Fig. 2 (b) after using the Dsatur algorithm. In this example, the minimum number of colors (pilots) required is three. The computational complexity of this algorithm is low. However, given the set of pilot sequences, how to design the pilot allocation scheme jointly with the UE selection process that maximizes the number of admitted UEs needs further investigation. To resolve this issue, we first adopt the Dsatur algorithm to find the minimum number of colors that are required for all the UEs in $\bar{u}$ while satisfying constraints C1 and C2. If the minimum number of required pilots is larger than the number of available pilots, some UEs should be removed. Otherwise, all UEs can be serve. For the latter case, although all users can be admitted, the pilot allocation results may be that some pilots have not been allocated by the Dsatur algorithm while some pilots are reused by up to $n_{\max}$ users, which wastes the pilot resources. In this case, we can reallocate all of the available pilots to the UEs to reduce the pilot contamination. The details of each case will be discussed in the following.

Denote the minimum number of pilots required by the Dsatur algorithm as n^*. In the following, we discuss two cases: 1) $n^* > \tau$; 2) $n^* < \tau$. When $n^* = \tau$, no additional operation is required.

Case I ($n^* > \tau$): In this case, some UEs should be removed so that the minimum number of required pilots is no larger than τ. Denote $\theta_k \triangleq \sum_{k' \neq k, k' \in \bar{u}} b_{k,k'}$ as the total number of different UEs to which UE k is connected to. In general, the UE with the largest θ_k should be removed since many UEs should use different pilots from that used by this UE, which will increase the number of required pilots. However, there are some cases that different UEs share the same largest θ_k value, and randomly removing one UE may incur inferior performance. Intuitively, the UE that incurs the highest pilot contamination should be removed. To this end, we first define a metric $\eta_{k,k'}$ to measure the level of pilot contamination between any two unconnected UEs when they are allocated with the same pilot,

$$\eta_{k,k'} = \log\left(1 + \frac{\sum_{i \in I_{k'}} \alpha_{i,k}}{\sum_{i \in I_k} \alpha_{i,k}}\right) + \log\left(1 + \frac{\sum_{i \in I_k} \alpha_{i,k'}}{\sum_{i \in I_{k'}} \alpha_{i,k'}}\right). \tag{10}$$

The above definition of $\eta_{k,k'}$ is inspired by the channel estimation error in (5). Obviously, the larger $\eta_{k,k'}$ means more severe pilot contamination between UE k and UE k' when the same pilot is allocated to them. Then, define $\xi_k = \sum_{k' \in \mathcal{K}_{\pi_k} \setminus \{k\}} \eta_{k,k'}$ as the value to measure the level of pilot contamination when keeping UE k. Then, the UE with the largest value of ξ_k should be removed.

Based on this idea, we provide a UE selection and pilot reallocation process in Algorithm 1. By using this algorithm for the case of $\tau = 2$ and $n_{\max} = 2$ for the network in Fig. 1, the UE selection and pilot allocation result is shown in Fig. 2 c), where the number of selected UEs is four.

Algorithm 1 UE Selection and Pilot Reallocation Algorithm for Case I

1: Initialize the matrix $\mathbf{B}$, the UE set $\mathcal{U} = \bar{u} = \{1, \cdots, K\}$, the initial number of required pilots n^* obtained from the Dsatur algorithm;
2: While $n^* > \tau$
3:     Find $k^* = \arg\max_{k \in \mathcal{U}} \theta_k$. If there are more than one UE with the same values of θ_k,
        remove the UE with the largest ξ_k;
4:     Remove UE k^* from $\mathcal{U}$, i.e., $\mathcal{U} = \mathcal{U}/k^*$, and update matrix $\mathbf{B}$ with current $\mathcal{U}$;
5:     Use the Dsatur algorithm to calculate n^* with $\mathbf{B}$ and $\mathcal{U}$;

Case II: $n^* < \tau$. In this case, all UEs can be admitted in this stage and only part of the available pilots are allocated. However, all the available pilots should be allocated to UEs to reduce the pilot contamination as much as possible. For example, in Fig. 2 b) with three allocated pilots, there may exist measurable pilot interference between UE 1 and UE 2 since they are not so far away from each other. A similar issue holds for UE 3 and UE 4 or UE 5 and UE 6. When there are four available pilots, the pilots can be fully used and reallocated to resolve this issue. For example, in Fig. 2 d), when there are four pilots, the UEs in each pair are able to be allocated with different pilots. Hence, the pilot contamination can be additionally mitigated.

As the definition of $\eta_{k,k'}$ in (10) shows, larger $\eta_{k,k'}$ means more severe pilot contamination between UE k and UE k' when the same pilot is allocated to them. The pair of UEs with larger $\eta_{k,k'}$ should be allocated with different pilots, or equivalently these two UEs should be connected with each other. For the pair of UEs with smaller $\eta_{k,k'}$, they can be allocated the same pilot. Hence, to reconstruct the undirected graph, a threshold η_{th} is introduced to determine whether two UEs can reuse the same pilot. Specifically, when $\eta_{k,k'} > \eta_{\text{th}}$, no reuse is allowed for these two UEs. Otherwise, these two UEs can reuse the same pilot. Based on η_{th}, the binary matrix $\mathbf{B}$ can be reconstructed as follows

$$b_{k,k'} = \begin{cases} 1, & \text{if } I_k \cap I_{k'} \neq \emptyset \text{ and } k \neq k', \\ 1, & \text{if } \eta_{k,k'} > \eta_{\text{th}}, \ I_k \cap I_{k'} = \emptyset \text{ and } k \neq k', \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Obviously, when η_{th} is small, more UEs will be connected with each other and more orthogonal pilots are required. In the extreme case when $\eta_{\text{th}} < \min\{\eta_{k,k'}\}$, all UEs will be connected with each other and the number of required pilots is equal to the number of total UEs K. On the other hand, when η_{th} is very large, less number of UEs will be connected and the number of required pilots will decrease. In the extreme case when $\eta_{\text{th}} \geq \max\{\eta_{k,k'}\}$, the reconstructed binary matrix $\mathbf{B}$ in (11) reduces to the conventional binary matrix $\mathbf{B}$ in (9), where the pairs of UEs without sharing the same RRHs will not be connected with each other. The number of required pilots in this case will be equal to n^*. As it is assumed that $\tau < K$, there must exist at least one η_{th} between $\min\{\eta_{k,k'}\}$ and $\max\{\eta_{k,k'}\}$ that the number of required pilots is equal to τ. As a result, the bisection search method can be adopted to find the η_{th} such that the required number of pilots is equal to τ. The details are given in Algorithm 2. Fig. 2 d) shows the pilot allocation results after using Algorithm 2.

Algorithm 2 Pilot Reallocation Algorithm for Case II

1: Initialize the lower-bound $\eta_{\text{th,LB}} = \min\{\eta_{k,k'}\}$, the upper-bound $\eta_{\text{th,UB}} = \max\{\eta_{k,k'}\}$, the initial number of required pilots n^* obtained from the Dsatur algorithm;
2: While $n^* \neq \tau$
3:     Set $\eta_{\text{th}} = (\eta_{\text{th,LB}} + \eta_{\text{th,UB}})/2$, update the binary matrix $\mathbf{B}$ in (11). Use the Dsatur
    algorithm to calculate n^* with $\mathbf{B}$.
4:     If $n^* > \tau$, set $\eta_{\text{th,LB}} = \eta_{\text{th}}$; If $n^* < \tau$, set $\eta_{\text{th,UB}} = \eta_{\text{th}}$;

C. Complexity Analysis

Now, we provide the complexity analysis for the proposed pilot assignment and UE selection scheme. The computational complexity of the Dsatur algorithm in [20] is on the order of $O(K^2)$.

For Case I, the user selection algorithm in Algorithm 1 requires the Dsatur algorithm to be executed for at most K times to find the final set of admitted UEs, and thus the total complexity is on the order of $O(K^3)$. For Case II, when η_{th} falls in a specific region $[a, c]$, the number of required pilots n^* obtained from the Dsatur algorithm will be equal to τ. Hence, when $\eta_{\text{th,UB}} - \eta_{\text{th,LB}} < c - a$, Algorithm 2 will terminate. As a result, the maximum number of iterations required by Algorithm 2 is upper bounded by $\log_2 \left(\left|\max\{\eta_{k,k'}\} - \min\{\eta_{k,k'}\}\right|/(c - a)\right)$. As a result, the total complexity of Algorithm 2 is $O\left(K^2\log_2\left(\left|\max\{\eta_{k,k'}\} - \min\{\eta_{k,k'}\}\right|/(c - a)\right)\right)$.

For comparison, we also provide the complexity analysis for the exhaustive search method. For a given set of UEs $\mathcal{U}$ with $|\mathcal{U}| = l$, the total number of pilot allocation results is given by τ^l. For each pilot allocation, one should check whether constraints C1 and C2 are satisfied or not. The total number of possible UE sets is given by $\sum_{l=1}^{K} C_K^l$. As the result, the total complexity of the exhaustive search method is $\sum_{l=1}^{K} C_K^l \tau^l$, which is much higher than the proposed algorithm and be unaffordable for dense C-RAN with large number of UEs.

IV. STAGE II: ROBUST BEAMFORMING DESIGN

In this stage, we first formulate the beamforming optimization problem to minimize the transmit power with the UEs selected from Stage I while considering the effect of pilot contamination due to pilot reuse scheme in Section III. Then, a low-complexity beam-vector optimization algorithm is proposed to solve this optimization problem along with the complexity analysis.

A. Problem Formulation

Denote the set of UEs selected from Stage I as $\widetilde{\mathcal{U}}$. The beam-vector for each UE can be merged into a single large-dimension vector $\mathbf{w}_k = [\mathbf{w}_{i,k}^{\text{H}}, \forall i \in I_k]^{\text{H}} \in \mathbb{C}^{|I_k|M \times 1}$. Similarly, we define a set of new channel vectors $\mathbf{g}_{l,k} = [\mathbf{h}_{i,k}^{\text{H}}, \forall i \in I_l]^{\text{H}} \in \mathbb{C}^{|I_l|M \times 1}$, representing the aggregated perfect CSI from the RRHs in I_l to UE k. Also, define $\tilde{\mathbf{g}}_{k,k} = [\tilde{\mathbf{h}}_{i,k}^{\text{H}}, \forall i \in I_k]^{\text{H}} \in \mathbb{C}^{|I_k|M \times 1}$ and $\hat{\mathbf{g}}_{k,k} = [\hat{\mathbf{h}}_{i,k}^{\text{H}}, \forall i \in I_k]^{\text{H}} \in \mathbb{C}^{|I_k|M \times 1}$ as the aggregated CSI error and estimated CSI from the RRHs in I_k to UE k, respectively. Since channel estimation error is expressed as $\tilde{\mathbf{g}}_{k,k} = \mathbf{g}_{k,k} - \hat{\mathbf{g}}_{k,k}$, the received signal model in (1) can be rewritten as

$$y_k = \underbrace{\hat{\mathbf{g}}_{k,k}^{\text{H}}\mathbf{w}_k s_k}_{\text{Desired signal}} + \underbrace{\tilde{\mathbf{g}}_{k,k}^{\text{H}}\mathbf{w}_k s_k}_{\text{Self-interference}} + \underbrace{\sum_{l \neq k, l \in \widetilde{\mathcal{U}}} \mathbf{g}_{l,k}^{\text{H}}\mathbf{w}_l s_l}_{\text{Multiuser Interference}} + z_k. \tag{12}$$

As in most of existing works [32], [33], we consider the achievable data rate where the term corresponding to the channel estimation error in (12) is regarded as Gaussian noise. Specifically, the achievable data rate for UE k is written as

$$r_k = \rho \mathbb{E}\left\{\log_2\left(1 + \frac{\left|\hat{\mathbf{g}}_{k,k}^{\text{H}}\mathbf{w}_k\right|^2}{\left|\tilde{\mathbf{g}}_{k,k}^{\text{H}}\mathbf{w}_k\right|^2 + \sum_{l \neq k, l \in \widetilde{\mathcal{U}}}\left|\mathbf{g}_{l,k}^{\text{H}}\mathbf{w}_l\right|^2 + \sigma_k^2}\right)\right\}, \tag{13}$$

where $\rho = (T - \tau)/T$ with T denoting the coherence time of the channel in terms of time slots, the expectation is taken with respect to the unknown channel estimation errors $\left\{\tilde{\mathbf{h}}_{i,k}, i \in I_k, \forall k \in \widetilde{\mathcal{U}}\right\}$, and the small-scale inter-cluster CSI $\left\{\mathbf{h}_{i,k}, i \in I \backslash I_k\right\}$. Each UE's data rate should be larger than its minimum data rate requirement $R_{k,\min}$:

$$\text{C3} : r_k \geq R_{k,\min}, \quad \forall k \in \widetilde{\mathcal{U}}. \tag{14}$$

Since the capacity of fronthaul links is upper bounded by the available bandwidth, each fronthaul link should be imposed with the fronthaul capacity constraints:

$$\text{C4} : \sum_{k \in \mathcal{U}_i} \varepsilon\left(\|\mathbf{w}_{i,k}\|^2\right) r_k \leq C_{i,\max}, \quad \forall i \in I, \tag{15}$$

where $\varepsilon(\cdot)$ is an indicator function, defined as $\varepsilon(x) = 1$ if $x \neq 0$, otherwise, $\varepsilon(x) = 0$, and $C_{i,\max}$ is the maximum data rate that can be supported by the ith fronthaul link.

Finally, each RRH has its own power constraint, represented as

$$\text{C5} : \sum_{k \in \mathcal{U}_i} \|\mathbf{w}_{i,k}\|^2 \leq P_{i,\max}, \quad i \in I, \tag{16}$$

where $P_{i,\max}$ is the power constraint of RRH i.

In Stage II, we aim to jointly optimize the UE-RRH associations and beam-vectors to minimize the total transmit power of the dense C-RAN network, while guaranteeing the rate constraints in C3, the fronthaul capacity constraints in C4, and the per-RRH power constraints in C5. Specifically, this optimization problem can be formulated as

$$\mathcal{P}_2: \min_{\mathbf{w}} \sum_{k \in \widetilde{\mathcal{U}}} \sum_{i \in I} \left\| \mathbf{w}_{i,k} \right\|^2$$
$$\text{s.t. C3, C4, C5,} \tag{17}$$

where $\mathbf{w}$ denotes the collection of all beam-vectors.

The imperfect intra-cluster CSI and incomplete inter-cluster CSI make the accurate closed-form expression of the data rate difficult to obtain. In the following, we first obtain the lower-bound of the data rate and replace it with this term to make the optimization problem in (17) more tractable. By using Jensen's inequality, the lower bound of the data rate can be derived as

$$r_k \geq \rho \log_2 \left(1 + \frac{\left| \hat{\mathbf{g}}_{k,k}^H \mathbf{w}_k \right|^2}{\mathbb{E}\left\{ \left| \tilde{\mathbf{g}}_{k,k}^H \mathbf{w}_k \right|^2 \right\} + \sum_{l \neq k, l \in \widetilde{\mathcal{U}}} \mathbb{E}\left\{ \left| \mathbf{g}_{l,k}^H \mathbf{w}_l \right|^2 \right\} + \sigma_k^2} \right)$$

$$= \rho \log_2 \left(1 + \frac{\left| \hat{\mathbf{g}}_{k,k}^H \mathbf{w}_k \right|^2}{\mathbf{w}_k^H \mathbf{E}_{k,k} \mathbf{w}_k + \sum_{l \neq k, l \in \widetilde{\mathcal{U}}} \mathbf{w}_l^H \mathbf{A}_{l,k} \mathbf{w}_l + \sigma_k^2} \right) \tag{18}$$

$$\triangleq \tilde{r}_k, \tag{19}$$

where $\mathbf{E}_{k,k} = \text{blkdiag}\left\{ \delta_{i,k} \mathbf{I}_{M \times M}, i \in I_k \right\}$, and $\mathbf{A}_{l,k} = \mathbb{E}\left\{ \mathbf{g}_{l,k}^H \mathbf{g}_{l,k} \right\} \in \mathbb{C}^{M|I_l| \times M|I_l|}$. To obtain the expression of $\mathbf{A}_{l,k}$, we define the indices of I_l as $I_l = \left\{ s_1^l, \cdots, s_{|I_l|}^l \right\}$. Then, we have

$$\mathbf{A}_{l,k} = \begin{bmatrix} \left(\mathbf{A}_{l,k} \right)_{1,1} & \cdots & \left(\mathbf{A}_{l,k} \right)_{1,|I_l|} \\ \vdots & \ddots & \vdots \\ \left(\mathbf{A}_{l,k} \right)_{|I_l|,1} & \cdots & \left(\mathbf{A}_{l,k} \right)_{|I_l|,|I_l|} \end{bmatrix}, \quad l \neq k, \tag{20}$$

where $\left(\mathbf{A}_{l,k} \right)_{i,j} \in \mathbb{C}^{M \times M}, i, j \in 1, \cdots, |I_l|$ is the block matrix of $\mathbf{A}_{l,k}$ at the ith row and jth column, given by

$$\left(\mathbf{A}_{l,k} \right)_{i,j} = \begin{cases} \hat{\mathbf{h}}_{s_i^l,k} \hat{\mathbf{h}}_{s_j^l,k}^H, & \text{if } s_i^l, s_j^l \in I_k, i \neq j, \\ \hat{\mathbf{h}}_{s_i^l,k} \hat{\mathbf{h}}_{s_j^l,k}^H + \delta_{s_i^l,k} \mathbf{I}_{M \times M}, & \text{if } s_i^l, s_j^l \in I_k, i = j, \\ \alpha_{s_i^l,k} \mathbf{I}_{M \times M}, & \text{if } s_i^l, s_j^l \notin I_k, i = j, \\ \mathbf{0}_{M \times M}, & \text{otherwise.} \end{cases} \tag{21}$$

It can be easily verified that $\mathbf{A}_{l,k}$ is a positive definite matrix. In [13], we provided the accurate data rate expression for the special case of non-overlapped cluster. We showed that the approximation error was within 3% in the considered setups.

By replacing the data rate r_k in data rate constraint C3 with its lower-bound or achievable data rate $\tilde{r}_k$ given in (19), constraint C3 can be rewritten as

$$\text{C6}: \tilde{r}_k \geq R_{k,\min}, \quad \forall k \in \widetilde{\mathcal{U}}, \tag{22}$$

which can be equivalently rewritten as

$$\text{C7}: \frac{\left| \hat{\mathbf{g}}_{k,k}^H \mathbf{w}_k \right|^2}{\mathbf{w}_k^H \mathbf{E}_{k,k} \mathbf{w}_k + \sum_{l \neq k, l \in \mathcal{U}} \mathbf{w}_l^H \mathbf{A}_{l,k} \mathbf{w}_l + \sigma_k^2} \geq \eta_{k,\min},$$
$$\tag{23}$$

where $\eta_{k,\min} = 2^{\frac{R_{k,\min}}{\rho}} - 1$. In addition, by replacing the data rate r_k in fronthaul constraints C4 with its lower-bound $\tilde{r}_k$, constraints C4 can be rewritten as

$$\text{C8}: \sum_{k \in \mathcal{U}_i} \varepsilon \left(\left\| \mathbf{w}_{i,k} \right\|^2 \right) \tilde{r}_k \leq C_{i,\max}, \quad \forall i \in I. \tag{24}$$

Then, Problem $\mathcal{P}_2$ can be approximately transformed as

$$\mathcal{P}_3: \min_{\mathbf{w}} \sum_{k \in \widetilde{\mathcal{U}}} \sum_{i \in I} \left\| \mathbf{w}_{i,k} \right\|^2 \tag{25a}$$
$$\text{s.t. C5, C7, C8.} \tag{25b}$$

To simplify the fronthaul capacity constraints C8, we first provide the following theorem.

Theorem 1: Denote the optimal solution of Problem $\mathcal{P}_3$ as $\mathbf{w}^\star = \left[\mathbf{w}_j^\star, j \in \widetilde{\mathcal{U}} \right]$. Then the minimum rate constraints are met with equality at the optimal point, i.e., $\tilde{r}_k(\mathbf{w}^\star) = R_{k,\min}$, $\forall k \in \widetilde{\mathcal{U}}$.

Proof: Please see Appendix A.   □

Based on Theorem 1, the expressions of rate lower bound $\tilde{r}_k$'s in constraint C8 can be replaced by their rate targets, which yields the following optimization problem

$$\mathcal{P}_4: \min_{\mathbf{w}} \sum_{k \in \widetilde{\mathcal{U}}} \sum_{i \in I} \left\| \mathbf{w}_{i,k} \right\|^2 \tag{26a}$$
$$\text{s.t. C5, C7,} \tag{26b}$$
$$\text{C9}: \sum_{k \in \mathcal{U}_i} \varepsilon \left(\left\| \mathbf{w}_{i,k} \right\|^2 \right) R_{k,\min} \leq C_{i,\max}, \quad \forall i. \tag{26c}$$

However, Problem $\mathcal{P}_4$ is still difficult to solve due to the following reasons. First, due to the per-RRH power constraints of C5 and the fronthaul capacity constraints of C9, the system may not be able to support all the UEs that are selected in Stage I. Second, constraint C9 contains the non-smooth and non-differential indicator function, which is usually named as an mixed-integer non-linear programming (MINLP) problem that is NP-hard to solve. One can solve the problem via the exhaustive search method. Specifically, for each given set of UE-RRH associations, one should check whether Problem $\mathcal{P}_4$ is feasible or not, if feasible, then the problem can be solved. Hence, the complexity is on the order of $O\left(2^{|\widetilde{\mathcal{U}}|I} \right)$, which is prohibitive for dense C-RAN with large number of UEs and RRHs. Third, due to the presence of the channel estimation error, the conventional weighted minimum mean square error (WMMSE) method [23] that has been used in [9], [10], and [13] for the perfect intra-cluster CSI case cannot be used for the problem considered here. Hence, even given the set of selected UEs and UE-RRH associations, it is still difficult to check the feasibility of Problem $\mathcal{P}_4$.

B. Low-Complexity Algorithm

1) UE Selection Method: Denote the set of admitted UEs as $\mathcal{U}$. Inspired by the UE selection algorithm in [34], we construct the following alternative optimization problem

by introducing a series of auxiliary non-negative variables $\{x_k\}_{k\in\mathcal{U}}$:

$$\mathcal{P}_5: \min_{\mathbf{w},\{x_k\geq 0\}_{k\in\mathcal{U}}} \sum_{i\in I}\sum_{k\in\mathcal{U}_i} \|\mathbf{w}_{i,k}\|^2 + \Gamma\sum_{k\in\mathcal{U}} x_k \tag{27a}$$

$$\text{s.t. C5, C9,} \tag{27b}$$

$$\text{C10}: \eta_{k,\min}\left(\mathbf{w}_k^{\mathrm{H}}\mathbf{E}_{k,k}\mathbf{w}_k + \sum_{l\neq k,l\in\mathcal{U}} \mathbf{w}_l^{\mathrm{H}}\mathbf{A}_{l,k}\mathbf{w}_l + \sigma_k^2\right)$$
$$\leq \left|\hat{\mathbf{g}}_{k,k}^{\mathrm{H}}\mathbf{w}_k\right|^2 + x_k, \quad \forall k, \tag{27c}$$

where Γ is a large constant. Obviously, Problem $\mathcal{P}_5$ is always feasible. If the optimal solution $\{x_k, \forall k \in \mathcal{U}\}$ are equal to zeros, i.e., $\{x_k = 0, \forall k \in \mathcal{U}\}$, all UEs can be admitted. Otherwise, some UEs should be removed. The large constant Γ acts as a penalty factor that forces as many x_ks to zero as possible, so that the number of admitted users is maximized. Hence, the objective function of Problem $\mathcal{P}_5$ is to maximize the number of admitted users while simultaneously minimizing the transmit power.

Intuitively, the UE with the largest x_k should have the highest priority to be removed since it has the largest gap from the rate target. Based on this idea, we provide a low-complexity algorithm to solve Problem $\mathcal{P}_5$ in Algorithm 3. The main idea is to remove the UE with the largest x_k in each iteration.

Algorithm 3 Low-Complexity UE Selection Algorithm

1: Initialize the set of users $\mathcal{U} = \widetilde{\mathcal{U}}$ from Stage I;
2: Given $\mathcal{U}$, solve Problem $\mathcal{P}_5$ to obtain $\{x_k\}_{k\in\mathcal{U}}$ and $\mathbf{w}$;
3: If $x_k = 0, \forall k \in \mathcal{U}$, output $\mathbf{w}$ and $\mathcal{U}^\star = \mathcal{U}$, terminate; Otherwise, find $k^\star = \arg\max_{k\in\mathcal{U}} x_k$, remove user $k^\star$ and update $\mathcal{U} = \mathcal{U}\backslash k^\star$, go to step 2.

---

2) Method to Deal With the Indicator Function: Similar to [13], we approximate the non-smooth indicator function as a fractional function $f_\theta(y) = \frac{y}{y+\theta}$, where θ is a small positive value. By replacing the indicator function in Problem $\mathcal{P}_5$ with $f_\theta(y)$, Problem $\mathcal{P}_5$ can be approximated as

$$\mathcal{P}_6: \min_{\mathbf{w},\{x_k\geq 0,\forall k\}} \sum_{k\in\mathcal{U}}\sum_{i\in I} \|\mathbf{w}_{i,k}\|^2 + \Gamma\sum_{k\in\mathcal{U}} x_k \tag{28a}$$
$$\text{s.t. C5, C10,}$$
$$\text{C11}: \sum_{k\in\mathcal{U}_i} f_\theta\left(\|\mathbf{w}_{i,k}\|^2\right) R_{k,\min} \leq C_{i,\max}, \quad \forall i. \tag{28b}$$

Although constraints C11 are still non-convex since $f_\theta(x)$ is a concave function, it is the difference of convex (d.c.) program, which can be efficiently solved by the successive convex approximation (SCA) method [27]. The main idea is to approximate it as its first-order Taylor expansion. Specifically, by using the concavity of $f_\theta(x)$, we have

$$f_\theta\left(\|\mathbf{w}_{i,k}\|^2\right) \leq f_\theta\left(\|\mathbf{w}_{i,k}(t)\|^2\right)$$
$$+ \beta_{i,k}(t)\left(\|\mathbf{w}_{i,k}\|^2 - \|\mathbf{w}_{i,k}(t)\|^2\right) \tag{29}$$

where $\mathbf{w}_{i,k}(t)$ is the beam-vector obtained at the tth iteration, $\beta_{i,k}(t) = f_\theta'\left(\|\mathbf{w}_{i,k}(t)\|^2\right)$.[1] By replacing $f_\theta\left(\|\mathbf{w}_{i,k}\|^2\right)$ in Problem $\mathcal{P}_6$ with the right hand side of (29), the optimization problem to be solved in the $(t+1)^{\mathrm{th}}$ iteration is given by

$$\mathcal{P}_7: \min_{\mathbf{w},\{x_k\geq 0,\forall k\}} \sum_{k\in\mathcal{U}}\sum_{i\in I} \|\mathbf{w}_{i,k}\|^2 + \Gamma\sum_{k\in\mathcal{U}} x_k$$
$$\text{s.t. C5, C10,}$$
$$\text{C12}: \sum_{k\in\mathcal{U}_i} \tau_{i,k}(t)\|\mathbf{w}_{i,k}\|^2 \leq \tilde{C}_i(t), \quad \forall i, \tag{30a}$$

where $\tau_{i,k}(t) = \beta_{i,k}(t)R_{k,\min}$, $\tilde{C}_i(t) = C_{i,\max} - \sum_{k\in\mathcal{U}_i}\left(f_\theta\left(\|\mathbf{w}_{i,k}(t)\|^2\right) - \beta_{i,k}(t)\|\mathbf{w}_{i,k}(t)\|^2\right)R_{k,\min}$.

Based on the above analysis, we provide the SCA algorithm to solve Problem $\mathcal{P}_6$ in Algorithm 4. According to [13], [27], and [35], the convergence of the SCA algorithm can be guaranteed under two conditions: 1) The initial beam-vectors are feasible for Problem $\mathcal{P}_6$; 2) In each iteration of the SCA algorithm, the globally optimal solution of Problem $\mathcal{P}_7$ can be obtained. The first condition can be easily guaranteed, while the second condition is in general difficult to satisfy due to the non-convex constraints of C10. Fortunately, we are able to obtain the globally optimal solution of Problem $\mathcal{P}_7$ as shown in the following subsection.

Algorithm 4 SCA Algorithm to Solve Problem $\mathcal{P}_6$

1: Initialize the iteration number $t = 1$, error tolerance δ. Initialize any feasible $\mathbf{w}(0)$, calculate $\tau_{i,k}(0)$, $\tilde{C}_i(0)$, calculate the objective value of Problem $\mathcal{P}_6$, denoted as $\text{Obj}(\mathbf{w}(0))$.
2: Solve Problem $\mathcal{P}_7$ to get $\mathbf{w}(t)$ with $\tau_{i,k}(t-1)$, $\tilde{C}_i(t-1)$;
3: Update $\tau_{i,k}(t)$, $\tilde{C}_i(t)$ with $\mathbf{w}(t)$;
4: If $|\text{Obj}(\mathbf{w}(t-1)) - \text{Obj}(\mathbf{w}(t))|/\text{Obj}(\mathbf{w}(t)) < \delta$, terminate. Otherwise, set $t \leftarrow t+1$, go to step 2.

---

3) Method to Deal With the Non-Convex Constraint C10: We apply the semi-definite relaxation approach [28] to solve Problem $\mathcal{P}_7$. Specifically, define $\mathbf{W}_k = \mathbf{w}_k\mathbf{w}_k^{\mathrm{H}}, \forall k \in \mathcal{U}$ with the constraints that $\text{rank}(\mathbf{W}_k) = 1, \forall k \in \mathcal{U}$. Then, Problem $\mathcal{P}_7$ can be equivalently transformed as[2]

$$\mathcal{P}_8: \min_{\{\mathbf{W}_k\succeq\mathbf{0}, x_k\geq 0\}_{k\in\mathcal{U}}} \sum_{k\in\mathcal{U}} (\text{tr}(\mathbf{W}_k) + \Gamma x_k) \tag{31a}$$

$$\text{s.t. C13}: \sum_{k\in\mathcal{U}_i} \text{tr}(\mathbf{W}_k\mathbf{B}_{i,k}) \leq P_{i,\max}, \quad \forall i, \tag{31b}$$

$$\text{C14}: \text{tr}\left(\mathbf{W}_k\hat{\mathbf{g}}_{k,k}\hat{\mathbf{g}}_{k,k}^{\mathrm{H}}\right) + x_k$$
$$\geq \eta_{k,\min}\left(\text{tr}(\mathbf{W}_k\mathbf{E}_{k,k})\right.$$
$$\left. + \sum_{l\neq k,l\in\mathcal{U}} \text{tr}(\mathbf{W}_l\mathbf{A}_{l,k}) + \sigma_k^2\right), \quad \forall k, \tag{31c}$$

$$\text{C15}: \sum_{k\in\mathcal{U}_i} \tau_{i,k}\text{tr}(\mathbf{W}_k\mathbf{B}_{i,k}) \leq \tilde{C}_i, \quad \forall i, \tag{31d}$$

---

[1] $f_\theta'(x)$ denotes the first-order derivative of $f_\theta(x)$ w.r.t. x.
[2] For simplicity, the iteration index t is omitted.

$$\text{C16: } \text{rank}(\mathbf{W}_k) = 1, \quad \forall k, \tag{31e}$$

where $\mathbf{B}_{i,k}$'s are the following block diagonal matrices

$$\mathbf{B}_{i,k} = \text{diag} \left\{ \overbrace{\mathbf{0}_{1 \times M}}^{s_1^k}, \cdots, \overbrace{\mathbf{1}_{1 \times M}}^{s_j^k}, \overbrace{\mathbf{0}_{1 \times M}}^{s_{j+1}^k}, \cdots, \overbrace{\mathbf{0}_{1 \times M}}^{s_{|I_k|}^k} \right\},$$

$$\text{if } s_j^k = i, \quad \forall i, k. \tag{32}$$

However, Problem $\mathcal{P}_8$ is still non-convex due to the rank-one constraints in C16. We further relax these non-convex constraints and obtain the following optimization problem:

$$\mathcal{P}_9: \min_{\{\mathbf{W}_k \succeq \mathbf{0}, \alpha_k\}_{k \in \mathcal{U}}} \sum_{k \in \mathcal{U}} (\text{tr}(\mathbf{W}_k) + \alpha_k^2) \tag{33a}$$

$$\text{s.t. C13, C14, C15.} \tag{33b}$$

Obviously, Problem $\mathcal{P}_9$ is a semi-definite programming (SDP) problem [36], which is convex and can be effectively solved by using the standard tools such as CVX.

In general, the optimal solution obtained from SDP Problem may not satisfy the rank-one constraints. Fortunately, for Problem $\mathcal{P}_9$, we can prove that the relaxation of the non-convex constraints is tight. Denote the optimal solution of $\mathcal{P}_9$ as $\mathbf{W}_k^\star, \alpha_k^\star, \forall k$, we have the following theorem.

Theorem 2: The optimal solution obtained from the SDP Problem $\mathcal{P}_9$ is guaranteed to satisfy the rank-one constraints, i.e., $\text{rank}(\mathbf{W}_k^\star) = 1, \forall k$.

Proof: Please see Appendix B.   □

Since the rank of $\mathbf{W}_k^\star, \forall k$ are equal to one, one can use the simple singular value decomposition operation to obtain the optimal beam-vector of Problem $\mathcal{P}_7$. Hence, the SCA algorithm will be guaranteed to converge.

C. Complexity Analysis

For simplicity, we assume the candidate size for each UE is equal to L, i.e., $|I_k| = L, \forall k \in \mathcal{U}$. The method to solve Problem $\mathcal{P}_4$ consists of two layers. The inner layer is to solve the SDP problem $\mathcal{P}_9$ by using the interior-point method. According to [37], the number of iterations required to reduce the duality gap to a constant fraction is upper-bounded by $O\left(\sqrt{(|\mathcal{U}| + 2I)ML}\right)$. The total number of real variables in Problem $\mathcal{P}_4$ is given by $N_{\text{tot}} = |\mathcal{U}| \left(\frac{ML(ML+1)}{2} + 1\right)$. Then, the complexity for each iteration of the interior-point method is given by $O\left(N_{\text{tot}}^2 |\mathcal{U}| M^2 L^2\right)$ [37]. Hence, the complexity of the interior-point method to solve $\mathcal{P}_9$ is given by $O\left(\sqrt{(|\mathcal{U}| + 2L)ML} N_{\text{tot}}^2 |\mathcal{U}| M^2 L^2\right)$. According to the UE selection algorithm in Algorithm 3, each UE is removed for each time, and thus Problem $\mathcal{P}_9$ needs to be solved at most $|\mathcal{U}|$ times in the worst case. Hence, the total complexity to solve Problem $\mathcal{P}_4$ is on the order of $O\left(\sqrt{(|\mathcal{U}| + 2L)ML} N_{\text{tot}}^2 |\mathcal{U}|^2 M^2 L^2\right)$, which can be solved within polynomial time.

V. Simulation Results

In this section, simulation results are provided to evaluate the performance of the proposed algorithms. Two types of dense C-RAN networks are considered: small dense C-RAN

deployed in a square area of 400 m × 400 m and large one with 700 m × 700 m. The numbers of UEs in small and large dense C-RANs are given by 8 and 24 with the densities of 50 UEs/km^2 and 49 UEs/km^2, respectively. The numbers of RRHs in both dense C-RANs are set as 14 and 40 with the densities of 87.5 RRHs/km^2 and 81.6 RRHs/km^2, respectively. The considered scenarios comply with the 5G ultra-dense network [7], where the density of 5G BS is highly anticipated to come up to 40-50 BS/km^2. Both the UEs and RRHs are assumed to be independently and uniformly placed in this area. The channel gains are composed of three parts: 1) the channel path loss is modeled as $PL = 148.1 + 37.6\log_{10}d$ (dB) [38], where d is the distance measured in km; 2) the log-normal shadowing fading with zero mean and 8 dB standard derivation; 3) Rayleigh fading with zero mean and unit variance. It is assumed that each UE chooses its nearest L RRHs as its serving cluster, i.e., $|I_k| = L, \forall k$. Unless otherwise specified, the other system parameters are set as follows: the number of transmit antennas at each RRH $M = 2$, system bandwidth $B = 20$ MHz, error tolerance $\delta = 10^{-5}$, noise power spectral density is -174 dBm/Hz, each RRH's maximum power $P_{i,\max} = 100$ mW, $\forall i$, large constant $\Gamma = 10^5$, the pilot power at each UE is $p_t = 200$ mW, the parameter θ in the fractional function is $\theta = 10^{-5}$, the minimum rate requirement for each UE and the capacity constraint for each fronthaul link are uniformly generated within the regions of [3, 5] bit/s/Hz and [5, 10] bit/s/Hz, respectively, the cluster size for each UE $L = 3$, the proportion of pilots for training in one coherence time is 1% [22], the pilot maximum reuse times for small dense C-RAN and large one are set as $n_{\max} = 2$ and $n_{\max} = 4$, respectively. The MMSE channel estimation method is used in our simulations. The following results are obtained by averaging over 200 independent trials, where in each trial, both the UEs and RRHs are randomly placed.

We compare our proposed algorithms (with legend "Proposed") with the following algorithms:

1) Orthogonal pilot allocation (with legend "Ortho"): The number of admitted UEs in Stage I is equal to the number of available pilots τ, and τ UEs are randomly selected from K UEs.

2) No reallocation operations for Case II in Stage I (with legend "NoCaseII"): This approach is similar to the proposed pilot allocation method in Stage I, except that when Case II happens, no additional operation is applied to re-allocate the pilots.

3) Conventional pilot allocation method (with legend "Con"): This approach is similar to the above approach, except that when Case I happens, UEs are randomly removed until the minimum number of pilots is equal to τ.

4) Perfect CSI estimation (with legend "Perfect"): In this approach, we assume that the CSI within each UE's serving cluster can be perfectly known. Note that it is not necessary to consider the pilot allocation stage, and the set of UEs selected from Stage I of the proposed algorithm is used as the initial set of UEs in Stage II of this approach.
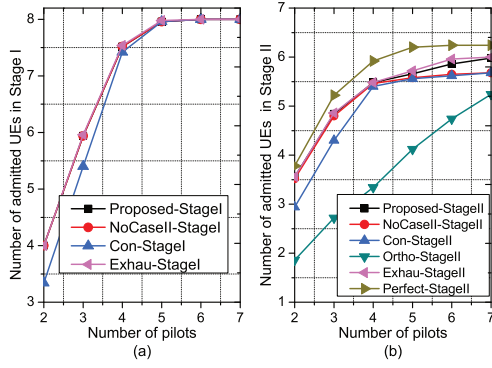
Fig. 3. (a) Number of admitted UEs in Stage I versus the number of available pilots τ for small networks. (b) Number of admitted UEs in Stage II versus the number of available pilots τ for small networks.
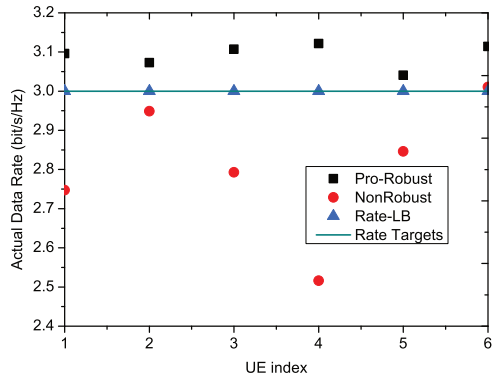

Fig. 4. The actual data rate for various algorithms under small dense C-RAN networks when six UEs are finally admitted in Stage II.

5) Exhaustive search method (with legend "Exhau"): In this approach, when Case I in Stage I happens, exhaustive search method is adopted to find the maximum number of UEs that can be admitted. If Case II happens, exhaustive search method is used to find all pilot allocation results that the number of allocated pilots is equal to τ. In Stage II, exhaustive search method is used to find the maximum number of admitted UEs. Note that the exhaustive search method has an exponential complexity, which is only feasible for small networks.

A. Small C-RAN Networks

We first consider the small C-RAN networks for the facilitation of applying the exhaustive search method. Fig. 3 shows the number of admitted UEs versus the number of available pilots τ for Stage I and Stage II, while Fig. 4 verifies the robustness of our proposed algorithm.

From Fig. 3-(a), we can observe that the numbers of admitted UEs for all algorithms increase with the number of available pilots. When $\tau \leq 4$, our proposed algorithm outperforms the "Con" algorithm. The reason is that the "Con" algorithm randomly removes the UE without considering the pilot interference among the UEs. With the increase of τ, the performance gain shrinks and finally both algorithms have the same performance where all UEs can be admitted by using the Dsatur algorithm. It is observed from Fig. 3-(a) that the

proposed algorithm and the "NoCaseII" have almost the same performance over the entire range of τ. This is not surprising, as when Case I in Stage I occurs, both the proposed algorithm and the "NoCaseII" algorithm employ Algorithm 1 to remove the UEs. On the other hand, when Case II happens, both algorithms can support all the UEs. The superiority of the proposed algorithm over the "NoCaseII" algorithm will be observed in Fig. 3-(b) discussed later. Note that the exhaustive search method performs almost the same as the proposed algorithm. However, the former one requires enormous computational complexity, which is not feasible for large-scale networks. The performance of the "Ortho" algorithm is not shown in this figure. It has the worst performance since no UEs are allowed to reuse the pilots, and the number of admitted UEs is equal to the number of available pilots.

It is also seen from Fig. 3-(b) that the numbers of admitted UEs in Stage II for all the algorithms increase with the number of available pilots. The performance of the proposed algorithm is comparable to that of the exhaustive search method over the entire range of τ, but the former needs huge computational complexity. The proposed algorithm outperforms the "NoCaseII" algorithm when $\tau \geq 4$, and the performance gain increases with τ. This can be explained as follows. When τ increases, Case II is more likely to happen, our proposed algorithm reallocates the unassigned pilots to the UEs to additionally reduce the pilot interference, which leads to more accurate CSI. Thus, more UEs can be admitted. On the other hand, no reallocation operations are considered in "NoCaseII", which will incur severe pilot interference and less UEs can be admitted. This highlights the necessity of reallocating the unassigned pilots to UEs. It is also observed from Fig. 3-(b) that as τ increases, our proposed algorithm approaches the performance of the "Perfect" algorithm, which verifies the effectiveness of the proposed algorithm.

Now, we study the robustness of our proposed algorithm in Fig. 4 for one randomly generated simulation, where the actual data rates achieved by various algorithms are plotted. For the clarity of illustration, the data rate targets and fronthaul capacity constraints are set to be the same for all UEs and fronthaul links, which are given by $R_{\min} = 3$ bit/s/Hz and $C_{\min} = 9$ bit/s/Hz, respectively. For comparison, the performance of non-robust algorithm is also shown (labeled as 'Nonrobust'), where channel estimation error is not taken into account and the estimated channel is naively regarded as the perfect channel. The data rate lower-bound obtained in (19) is also shown in this figure which is labeled as 'Rate-LB'. The bold line in Fig. 4 represents the rate targets of all UEs. It is observed that the rate lower bound is equal to their data rate target, which verifies the correctness of Theorem 1. As expected, the actual data rates achieved by all UEs are above the bold line, which means that our proposed algorithm can provide guaranteed data rates for all UEs. Note that the gap between the rate lower bound and actual data rates achieved by our proposed algorithm are within 0.1 bit/s/Hz, which is acceptable in practice. In contrast, the actual data rates achieved by the non-robust algorithm for almost all UEs are below the bold line, and the maximum gap is up to 0.5 bit/s/Hz for UE 4.
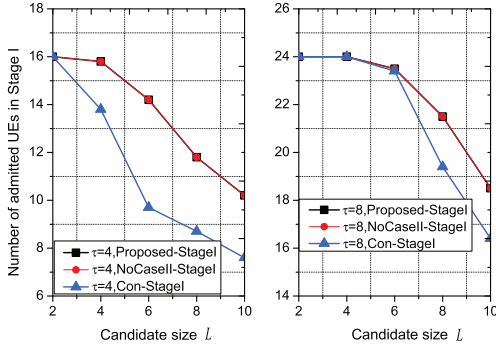
Fig. 5.   Number of admitted UEs in Stage I versus candidate size L. The left subplot corresponds to the case of $\tau = 4$ while the right one is $\tau = 8$.

Fig. 7.   Number of admitted UEs in Stage I versus candidate size L. The left subplot corresponds to the case of $\tau = 4$ while the right one is $\tau = 8$.

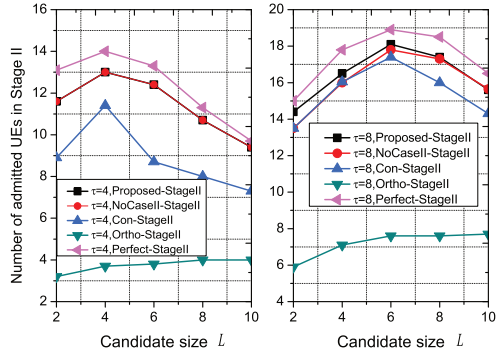
Fig. 6.   Number of admitted UEs in Stage I versus candidate size L. The left subplot corresponds to the case of $\tau = 4$ while the right one is $\tau = 8$.

Fig. 8.   Number of admitted UEs in Stage II versus the pilot power. The left subplot corresponds to the case of $\tau = 4$ while the right one is $\tau = 8$.

B. Large-Scale C-RAN Networks

In this subsection, we study the effects of different system parameters on the performance of the algorithms in large-scale dense C-RAN networks.

1) Impacts of Candidate Size: Figs. 5 and 6 illustrate the number of admitted UEs versus the candidate size L in Stage I and Stage II, respectively. It is seen from Fig. 5 that the numbers of admitted UEs achieved by all the algorithms in Stage I decrease with the candidate size. The reason is that with the increase of candidate size, more UEs will be connected with each other when constructing the undirected graph. In this case, more UEs will be removed in this stage to satisfy conditions C1 and C2 in Problem $\mathcal{P}_1$. Fig. 5 also shows that our proposed algorithm achieves superior performance over the "Con" algorithm, highlighting the importance of carefully considering pilot interference when removing UEs.

It is interesting to observe from Fig. 6 that the numbers of admitted UEs in Stage II obtained by all the algorithms (except the "Ortho" algorithm) initially increase with the candidate size, and then decrease. The reason for the former part is due to the increased spatial degrees of freedom with the increased candidate size. However, when the candidate size continues to increase, many UEs have been prohibited to be admitted due to the pilot allocation in Stage I as seen in Fig. 5. Hence, in Stage II, even all the selected UEs from Stage I can be admitted, the number is small. This trend is different from most of the existing papers [10], [13], [14], [39], [40], where the system performance always increases with the candidate size. Hence, the cluster size should be properly optimized and
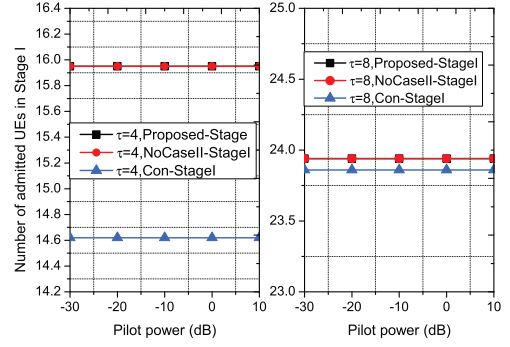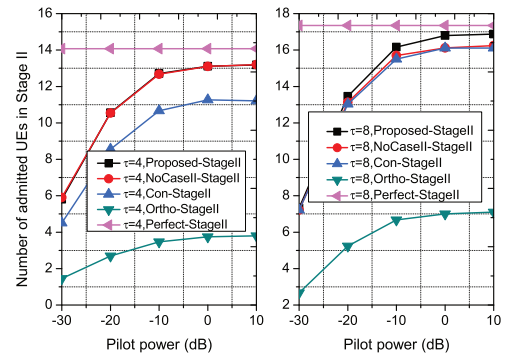
larger cluster size may deteriorate the system performance if channel estimation process is considered. From Fig. 6, we also observe that our proposed algorithm significantly outperforms the "Con" algorithm for both $\tau = 4$ and $\tau = 8$. On the other hand, when $\tau = 4$, both the proposed algorithm and the "NoCaseII" algorithm have similar performance; when $\tau = 8$, our proposed algorithm performs much better than the "NoCaseII" algorithm in the low candidate size regime, i.e., $X < 6$, and this performance gap decreases with X. The reason is that when the candidate size is small, less number of pilots is required. Hence, it is more likely to fall in Case II in Stage I, which leads to superior performance of the proposed algorithm over the "NoCaseII" algorithm due to the additional pilot reallocation step of our proposed algorithm. As expected, the "Ortho" algorithm has the worst performance since no pilot reuse is allowed.

2) Impacts of Pilot Power: Figs. 7 and 8 show the number of admitted UEs versus the pilot power for Stage I and Stage II, respectively. It is again seen from Fig. 7 that our proposed algorithm significantly outperforms the "Con" algorithm in Stage I when $\tau = 4$, and all algorithms can admit all UEs when $\tau = 8$ is Stage I.

Fig. 8 shows that the numbers of admitted UEs obtained by all the algorithms (except the "Perfect" algorithm) increase with the pilot power for both cases of $\tau = 4$ and $\tau = 8$ due to the more accurate CSI. However, there is a fixed gap between the proposed algorithm and the "Perfect" algorithm in the high pilot power regime. The reason can be explained as follows. According to the channel estimation error in (5), when the pilot power p_t increases, $\hat{\delta}$ will approach zero, i.e., $\hat{\delta} \rightarrow 0$.
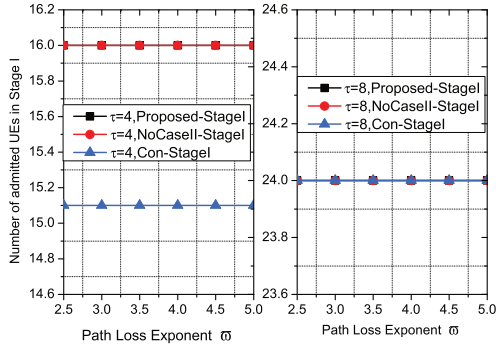
Fig. 9.   Number of admitted UEs in Stage I versus the path loss exponent. The left subplot corresponds to the case of $\tau = 4$ while the right one is $\tau = 8$.
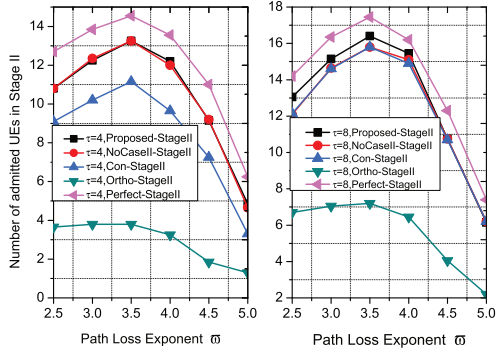

Fig. 10.   Number of admitted UEs in Stage II versus the path loss exponent. The left subplot corresponds to the case of $\tau = 4$ while the right one is $\tau = 8$.

Hence, the channel estimation error will become a fixed value, which is only related to the large-scale parameters and the pilot reuse scheme. In this case, the system will enter a pilot interference limited regime. Fig. 8 again demonstrates that the performance superiority of the proposed algorithm over the existing algorithms.

3) Impacts of Path Loss Exponent: The impact of path loss exponent is studied here. Specifically, the channel path loss is modeled as $PL = 148.1 + 10\varpi \log_{10} d$ (dB), where ϖ denotes the path loss exponent. Note that the other simulation results in this section are obtained by setting $\varpi = 3.76$. Figs. 9 and 10 show the numbers of admitted UEs versus the path loss exponent ϖ for Stage I and Stage II, respectively. It is observed from Fig. 9 that nearly one more UE can be admitted by the proposed algorithm over the "Con" algorithm in Stage I.

It is interesting to find from Fig. 10 that the number of admitted UEs in Stage II for all algorithms (except the "Ortho" algorithm) first increases with ϖ and then decreases with it. The reason can be explained as follows. When ϖ is small, the interference power from other RRHs is significant. As a result, the number of UEs that can satisfy their rate targets is lower. With the increase of ϖ, the interference power is reduced, and the number of UEs admitted is increasing. However, by additionally increasing ϖ, i.e., $\varpi > 3.5$, the performance achieved by all algorithms become worse since the signal power is becoming weaker. Hence, there is an optimal ϖ that all algorithms have the best performance. The performance gain over the existing algorithms is observed in Fig. 10.
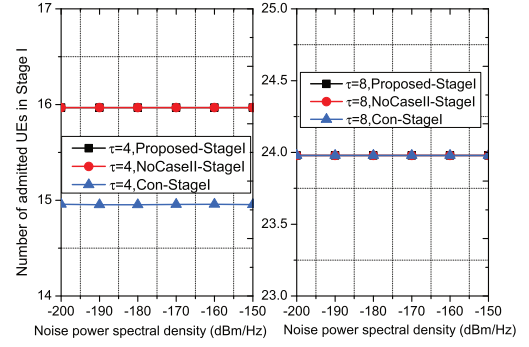

Fig. 11.   Number of admitted UEs in Stage I versus the noise power density. The left subplot corresponds to the case of $\tau = 4$ while the right one is $\tau = 8$.
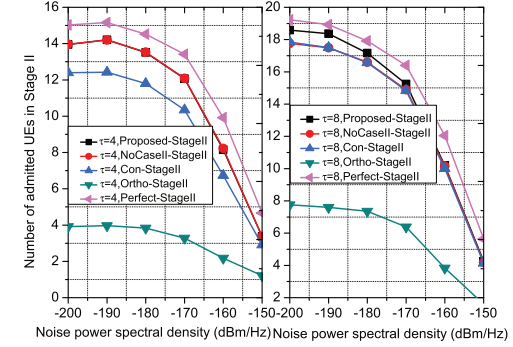

Fig. 12.   Number of admitted UEs in Stage II versus the noise power density. The left subplot corresponds to the case of $\tau = 4$ while the right one is $\tau = 8$.
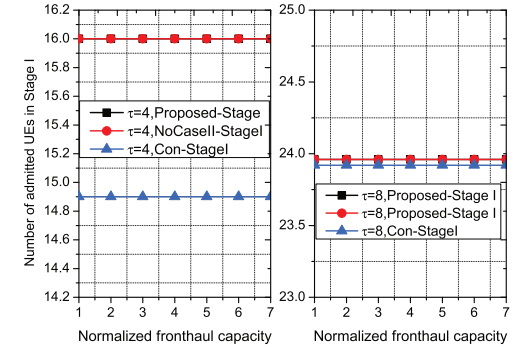

Fig. 13.   Number of admitted UEs in Stage I versus the maximum fronthaul capacity $\tilde{C}_{\max}$. The left subplot corresponds to the case of $\tau = 4$ while the right one is $\tau = 8$.

4) Impacts of Noise Power Spectral Density: Figs. 11 and 12 show the number of admitted UEs versus the noise power density for Stage I and Stage II, respectively. Fig. 11 shows the performance superiority of our proposed algorithm over the 'Con' algorithm and the number of admitted UEs achieved by all algorithms keep fixed in Stage I.

As expected, from Fig. 12, it is observed that the numbers of admitted UEs achieved by all algorithms decrease rapidly with the noise power density in Stage II, and only about five UEs can be admitted by our proposed algorithm when the noise power density is -150 dBm/Hz and $\tau = 8$. Fig. 12 also shows that our proposed algorithm can achieve better performance than the "Con" algorithm for both cases of τ, and than the "NoCaseII" algorithm when $\tau = 8$.

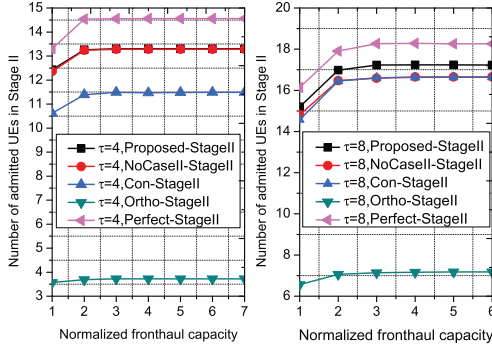*5) Impacts of Maximum Fronthaul Capacity:* Next, we study the impact of fronthaul capacity on the performance

Fig. 14. Number of admitted UEs in Stage II versus the maximum fronthaul capacity $\tilde{C}_{\max}$. The left subplot corresponds to the case of $\tau = 4$ while the right one is $\tau = 8$.
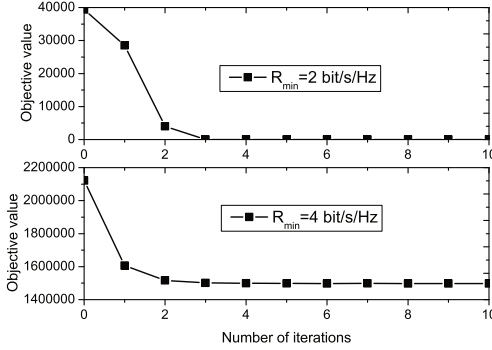

Fig. 15. Convergence behaviour of the SCA algorithm. The upper subplot shows the feasible case with low rate requirements $R_{\min} = 2$ bit/s/Hz, while the lower subplot shows the infeasible case with high rate requirements $R_{\min} = 4$ bit/s/Hz.

of various algorithms. For the sake of illustration, we assume that all UEs have the same rate targets, which are set as $R_{\min} = 4$bit/s/Hz. In addition, all fronthaul links are assumed to have the same fronthaul capacity constraints, i.e., $C_{\max} = C_{i,\max}, \forall i$. For ease of exposition, the normalized fronthaul capacity is considered, i.e., $\tilde{C}_{\max} = C_{\max}/R_{\min}$, which denotes the number of UEs that each fronthaul link can support. Figs. 13 and 14 illustrate the number of admitted UEs versus the maximum normalized fronthaul capacity $\tilde{C}_{\max}$ for Stage I and Stage II, respectively. As expected, from Fig. 13, we can find that all algorithms achieve fixed number of admitted UEs in Stage I, since the pilot allocation process does not depend on $\tilde{C}_{\max}$.

One insightful observation can be found in Fig. 14: the numbers of admitted UEs by all algorithms initially increase with $\tilde{C}_{\max}$ due to the fact that more UEs can be supported by each fronthaul link. However, when additionally increasing $\tilde{C}_{\max}$, these numbers remain fixed and do not vary much in the large $\tilde{C}_{\max}$ domain for both $\tau = 4$ and $\tau = 8$. It is shown that $\tilde{C}_{\max} = 3$ is sufficient for all algorithms to achieve a large portion of the performance of the case $\tilde{C}_{\max} = \infty$. This observation is instructive for the practical implementation of the dense C-RAN, where the mmWave communication technology with limited bandwidth can be employed as the wireless fronthaul link in dense C-RAN network.

6) Convergence Behavior: Finally, we study the convergence behaviour of the SCA algorithm in Algorithm 4. Fig. 15

illustrates the objective value of Problem $\mathcal{P}_6$ versus the number of iterations for two cases of $R_{\min} = 2$ bit/s/Hz and $R_{\min} = 4$ bit/s/Hz for one randomly generated C-RAN network. It is seen from this figure that the objective values for these two cases decrease rapidly and converge within four iterations, which demonstrate the effectiveness and low-complexity associated with our algorithm. When $R_{\min} = 2$ bit/s/Hz, the objective value converges to a very small value, which indicates all UEs can be supported. However, when $R_{\min} = 4$ bit/s/Hz, the objective value finally converges to a very large value, which implies the system is infeasible and some UEs should be removed by using our proposed UE selection algorithm in Algorithm 3.

VI. Conclusions

In this paper, we provided a new framework to handle the new challenges of user-centric ultra-dense C-RAN by considering a two-stage problem: the channel estimation for intra-cluster CSI in Stage I and robust transmit beam-vectors design in Stage II. Specifically, in Stage I, we jointly optimized the UE selection and pilot allocation to maximize the number of admitted UEs with limited number of available pilots. One smart UE selection algorithm was proposed when the number of available pilots is not sufficient to support all the UEs. On the other hand, when the opposite case happens, one novel pilot reallocation was introduced to fully use the available pilots. Under the results of Stage I, robust transmit beam-vector was designed to minimize the transmit power while guaranteeing each UE's rate requirement and fronthaul capacity constraints. One novel algorithm was proposed to solve this problem with convergence guarantee. Simulation results verify the effective of the proposed algorithm in terms of the number of admitted UEs compared with the existing naive pilot allocation method. Some interesting observations have been found in the simulations. For example, increasing the cluster/candidate size may not lead to the increased performance when taking the channel estimation into account. Hence, the cluster size should be carefully decided when designing the transmission scheme. In addition, the maximum fronthaul capacity is not necessary to be extremely high and it is shown that $\tilde{C}_{\max} = 3$ is sufficient to achieve good performance, which is appealing for the application of mmWave communication link as fronthaul links.

Appendix A
Proof of Theorem 1

We prove this theorem by using the contradiction method. Specifically, suppose one of the following two cases occurs:

1) All UEs' data rate lower bounds (LBs) are strictly larger than their rate targets, i.e., $\tilde{r}_k(\mathbf{w}^\star) > R_{k,\min}, \forall k \in \tilde{\mathcal{U}}$;
2) Part of UEs' data rate LBs are larger than their rate targets (denote this subset of UEs as $\tilde{\mathcal{U}}_1$), while those of the rest are equal to their targets (denote as this subset as $\tilde{\mathcal{U}}_2$).

We first prove that the first case cannot happen. To this end, one can construct another feasible solution that yields lower objective value than that of $\mathbf{w}^\star$. Specifically, we scale each

UE's beam-vector by a constant $\sqrt{\kappa}$ and denote the set of new beam-vectors as $\mathbf{w}^{\#} = \left[\sqrt{\kappa}\mathbf{w}_k^{\star}, k \in \widetilde{\mathcal{U}}\right]$, where κ is chosen as

$$\kappa = \max_{k \in \widetilde{\mathcal{U}}} \frac{\sigma_k^2}{\frac{\left|\hat{\mathbf{g}}_{k,k}^{\mathrm{H}}\mathbf{w}_k^{\star}\right|^2}{\eta_{k,\min}} - \mathbf{w}_k^{\star\mathrm{H}}\mathbf{E}_{k,k}\mathbf{w}_k^{\star} - \sum_{l \neq k, l \in \widetilde{\mathcal{U}}} \mathbf{w}_l^{\star\mathrm{H}}\mathbf{A}_{l,k}\mathbf{w}_l^{\star}} \tag{A.1}$$

with $\eta_{k,\min} = 2^{\frac{R_{k,\min}T}{T-\tau}} - 1$. It can be readily verified that the constant κ is strictly smaller than one, i.e., $\kappa < 1$, and all UEs' data rate LBs $\tilde{r}_k(\mathbf{w}^{\#})$ satisfy the following relationships:

$$\tilde{r}_k(\mathbf{w}^{\star}) > \tilde{r}_k(\mathbf{w}^{\#}) \geq R_{k,\min}, \forall k \in \widetilde{\mathcal{U}}. \tag{A.2}$$

Since $\mathbf{w}^{\star}$ is the optimal solution of Problem $\mathcal{P}_3$, then it must be a feasible solution that satisfies constraints C5, C7 and C8. Hence, the new set of beam-vectors can be verified to be a feasible solution of Problem $\mathcal{P}_3$, but yields a lower objective value. Hence, this contradicts the assumption that $\mathbf{w}^{\star}$ is the optimal solution and this case cannot happen.

For the second case, we cannot prove it by using the above method directly. The reason is that if κ is chosen as in (A.1), its value will be equal to one since some UEs' rate LBs are equal to their targets. To deal with this issue, we divide the group of UEs in $\widetilde{\mathcal{U}}$ into two groups, i.e., $\widetilde{\mathcal{U}}_1 = \left\{k | \tilde{r}_k(\mathbf{w}^{\star}) > R_{k,\min}, \forall k \in \widetilde{\mathcal{U}}\right\}$ and $\widetilde{\mathcal{U}}_2 = \left\{k | \tilde{r}_k(\mathbf{w}^{\star}) = R_{k,\min}, \forall k \in \widetilde{\mathcal{U}}\right\}$. Similar to the first case, we introduce a new constant variable defined as follows:

$$\kappa = \max_{k \in \widetilde{\mathcal{U}}_1} \frac{\sigma_k^2}{\frac{\left|\hat{\mathbf{g}}_{k,k}^{\mathrm{H}}\mathbf{w}_k^{\star}\right|^2}{\eta_{k,\min}} - \mathbf{w}_k^{\star\mathrm{H}}\mathbf{E}_{k,k}\mathbf{w}_k^{\star} - \sum_{l \neq k, l \in \widetilde{\mathcal{U}}} \mathbf{w}_l^{\star\mathrm{H}}\mathbf{A}_{l,k}\mathbf{w}_l^{\star}}. \tag{A.3}$$

Then, the constant κ is strictly smaller than one. We scale the beam-vectors of UEs in $\widetilde{\mathcal{U}}_1$ by a constant $\sqrt{\kappa}$, i.e., $\mathbf{w}_k^{\#} = \sqrt{\kappa}\mathbf{w}_k^{\star}, \forall k \in \widetilde{\mathcal{U}}_1$, while keeping the beam-vectors in $\widetilde{\mathcal{U}}_2$ unchanged. Denote the set of new beam-vectors as $\mathbf{w}^{\#} = \left[\mathbf{w}_k^{\#}, \forall k \in \widetilde{\mathcal{U}}_1, \mathbf{w}_k^{\star}, k \in \widetilde{\mathcal{U}}_2\right]$. It can be verified the following relations hold:

$$\tilde{r}_k(\mathbf{w}^{\star}) > \tilde{r}_k(\mathbf{w}^{\#}) \geq R_{k,\min}, \quad \forall k \in \widetilde{\mathcal{U}}_1, \tag{A.4}$$

and

$$\tilde{r}_k(\mathbf{w}^{\#}) > \tilde{r}_k(\mathbf{w}^{\star}) = R_{k,\min}, \quad \forall k \in \widetilde{\mathcal{U}}_2. \tag{A.5}$$

Since the data rate LBs in $\widetilde{\mathcal{U}}_2$ increase, the fronthaul capacity constraints in C8 may not hold. To resolve this problem, we update the new UE sets $\widetilde{\mathcal{U}}_1$ and $\widetilde{\mathcal{U}}_2$, and repeat the above procedure. Since the new constructed beam-vectors decrease the total power consumption and the objective value is lower bounded. Hence, this procedure will converge to a fixed value. Note that if there exists some UEs whose rate LBs are strictly larger than their targets, we can always construct new beam-vectors with reduced total power. Hence, when the procedure converges, all UEs data rates are equal to their targets, and the final beam-vectors satisfy all the constraints C5, C7 and C8. However, the objective value with the final beam-vectors have a lower value than that with $\mathbf{w}^{\star}$, which contradicts the assumption that $\mathbf{w}^{\star}$ is the optimal solution. Hence, this case cannot happen either, and the proof is complete.

APPENDIX B
PROOF OF THEOREM 2

Denote the collection of $\{\mathbf{W}_k, \forall k\}$ as $\mathbf{W}$. Then, the Lagrangian function of Problem $\mathcal{P}_9$ is given in (B.1), shown at the top of the next page, where $\boldsymbol{\lambda} = [\lambda_i, \forall i \in I]$, $\boldsymbol{\mu} = [\mu_k, \forall k \in \mathcal{U}]$, and $\boldsymbol{\nu} = [\nu_i, \forall i \in I]$ are the non-negative Lagrangian multipliers associated with constraints C13, C14 and C15, respectively. $\boldsymbol{\omega} = [\omega_k, \forall k \in \mathcal{U}]$ are the dual variables associated with the nonnegative constraints of $\mathbf{x} = \{x_k, \forall k \in \mathcal{U}\}$, and $\{\mathbf{Z}_k \succeq \mathbf{0}, \forall k \in \mathcal{U}\}$ denote the dual variable matrices associated with the semi-definite constraints on $\{\mathbf{W}_k, \forall k \in \mathcal{U}\}$.

As Problem $\mathcal{P}_9$ is a convex optimization problem, its globally optimal solution must satisfy its first-order optimality condition as:

$$\frac{\partial \mathcal{L}(\mathbf{W}, \lambda, \mu, \nu)}{\partial \mathbf{W}_k}$$
$$= \mathbf{I}_{M|I_k| \times M|I_k|} + \sum_{i \in I} \left(\lambda_i + \nu_i \tau_{i,k}\right) \mathbf{B}_{i,k}$$
$$- \mu_k \hat{\mathbf{g}}_{k,k} \hat{\mathbf{g}}_{k,k}^{\mathrm{H}} + \sum_{k \in \mathcal{U}} \mu_k \eta_{k,\min} \mathbf{E}_{k,k} + \sum_{l \neq k, l \in \mathcal{U}} \mu_l \eta_{l,\min} \mathbf{A}_{k,l}$$
$$- \mathbf{Z}_k = \mathbf{0}, \forall k \in \mathcal{U}. \tag{B.4}$$

Then, $\mathbf{Z}_k$ can be represented as $\mathbf{Z}_k = \mathbf{D}_k - \mu_k \hat{\mathbf{g}}_{k,k} \hat{\mathbf{g}}_{k,k}^{\mathrm{H}}$, where $\mathbf{D}_k$ is given by

$$\mathbf{D}_k = \mathbf{I}_{M|I_k| \times M|I_k|} + \sum_{i \in I} \left(\lambda_i + \nu_i \tau_{i,k}\right) \mathbf{B}_{i,k}$$
$$+ \sum_{k \in \mathcal{U}} \mu_k \eta_{k,\min} \mathbf{E}_{k,k} + \sum_{l \neq k, l \in \mathcal{U}} \mu_l \eta_{l,\min} \mathbf{A}_{k,l}. \tag{B.5}$$

According to the Karush-Kuhn-Tucker (KKT) condition, we have (B.2) and (B.3), shown at the top of the next page.

Before proving the theorem, we first give the following lemma.

Lemma 1: The optimal Lagrangian dual multipliers $\{\mu_k, \forall k\}$ are positive, i.e., $\mu_k > 0, \forall k$.

Proof: This can be proved by using contradiction. Denote the optimal solution of Problem $\mathcal{P}_9$ as $\left\{\mathbf{W}_k^{\star}, x_k^{\star}, \forall k\right\}$ and the corresponding Lagrangian multipliers as $\left[\lambda_i^{\star}, \nu_i^{\star}, \mu_k^{\star}, \forall i \in I, \forall k \in \mathcal{U}\right]$. Assume there exists one $\mu_l^{\star}$ that is zero. Then, according to KKT conditions in (B.2), we have $\mathrm{tr}\left(\mathbf{W}_l^{\star}\hat{\mathbf{g}}_{l,l}\hat{\mathbf{g}}_{l,l}^{\mathrm{H}}\right) + x_l^{\star} > \eta_{l,\min}\left(\mathrm{tr}\left(\mathbf{W}_l^{\star}\mathbf{E}_{l,l}\right) + \sum_{m \neq l, m \in \mathcal{U}} \mathrm{tr}\left(\mathbf{W}_m^{\star}\mathbf{A}_{m,l}\right) + \sigma_l^2\right)$. We consider the following two cases:

1) If $x_l^{\star} > 0$, we can find a new $\bar{x}_l = x_l^{\star} - \Delta$, where $\Delta > 0$ is a small enough positive value such that $\mathrm{tr}\left(\mathbf{W}_l^{\star}\hat{\mathbf{g}}_{l,l}\hat{\mathbf{g}}_{l,l}^{\mathrm{H}}\right) + \bar{x}_l \geq \eta_{l,\min}\left(\mathrm{tr}\left(\mathbf{W}_l^{\star}\mathbf{E}_{l,l}\right) + \sum_{m \neq l, m \in \mathcal{U}} \mathrm{tr}\left(\mathbf{W}_m^{\star}\mathbf{A}_{m,l}\right) + \sigma_l^2\right)$ holds. Then, we find another feasible solution $\left\{\mathbf{W}_k^{\star}, \forall k, x_k^{\star}, \forall k \neq l, \bar{x}_l\right\}$ that yields a lower objective value than that of $\left\{\mathbf{W}_k^{\star}, x_k^{\star}, \forall k\right\}$, which contradicts that $\left\{\mathbf{W}_k^{\star}, x_k^{\star}, \forall k\right\}$ is the optimal solution.

2) If $x_l^{\star} = 0$, we can find a new precoding matrix $\bar{\mathbf{W}}_l = \rho \mathbf{W}_l^{\star}$ with $0 < \rho < 1$ such that $\mathrm{tr}\left(\bar{\mathbf{W}}_l \hat{\mathbf{g}}_{l,l} \hat{\mathbf{g}}_{l,l}^{\mathrm{H}}\right) \geq \eta_{l,\min}\left(\mathrm{tr}\left(\bar{\mathbf{W}}_l \mathbf{E}_{l,l}\right) + \sum_{m \neq l, m \in \mathcal{U}} \mathrm{tr}\left(\mathbf{W}_m^{\star}\mathbf{A}_{m,l}\right) + \sigma_l^2\right)$ holds

$$\mathcal{L}\left(\mathbf{W}, \mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\omega}\right) = \sum_{k \in \mathcal{U}} \text{tr}\left(\mathbf{W}_k\right) + \sum_{k \in \mathcal{U}} x_k + \sum_{i \in I} \lambda_i \left(\sum_{k \in \mathcal{U}_i} \text{tr}\left(\mathbf{W}_k \mathbf{B}_{i,k}\right) - P_{i,\max}\right)$$

$$+ \sum_{k \in \mathcal{U}} \mu_k \left(\eta_{k,\min} \left(\text{tr}\left(\mathbf{W}_k \mathbf{E}_{k,k}\right) + \sum_{l \neq k, l \in \mathcal{U}} \text{tr}\left(\mathbf{W}_l \mathbf{A}_{l,k}\right) + \sigma_k^2\right) - \text{tr}\left(\mathbf{W}_k \hat{\mathbf{g}}_{k,k} \hat{\mathbf{g}}_{k,k}^{\text{H}}\right) - x_k\right),$$

$$+ \sum_{i \in I} \nu_i \left(\sum_{k \in \mathcal{U}_i} \tau_{i,k} \text{tr}\left(\mathbf{W}_k \mathbf{B}_{i,k}\right) - \tilde{C}_i\right) - \sum_{k \in \mathcal{U}} \text{tr}\left(\mathbf{W}_k \mathbf{Z}_k\right) - \sum_{k \in \mathcal{U}} \omega_k x_k, \tag{B.1}$$

$$\mu_k \left(\eta_{k,\min} \left(\text{tr}\left(\mathbf{W}_k \mathbf{E}_{k,k}\right) + \sum_{l \neq k, l \in \mathcal{U}} \text{tr}\left(\mathbf{W}_l \mathbf{A}_{l,k}\right) + \sigma_k^2\right) - \text{tr}\left(\mathbf{W}_k \hat{\mathbf{g}}_{k,k} \hat{\mathbf{g}}_{k,k}^{\text{H}}\right) - x_k\right) = 0, \quad \forall k, \tag{B.2}$$

$$\mathbf{W}_k \mathbf{Z}_k = \mathbf{0}, \quad \forall k. \tag{B.3}$$

since $\mathbf{A}_{m,l}, \forall m \neq l$ are positive definite matrices. Then, we find a new feasible solution $\left\{\mathbf{W}_k^\star, \forall k \neq l, \tilde{\mathbf{W}}_l, x_k^\star, \forall k\right\}$ that yields a lower objective value than that of $\left\{\mathbf{W}_k^\star, x_k^\star, \forall k\right\}$, which contradicts that $\left\{\mathbf{W}_k^\star, x_k^\star, \forall k\right\}$ is the optimal solution. $\square$

The first term of $\mathbf{D}_k$ in (B.5) is an identity matrix. In addition, the Lagrange multipliers $\{\lambda_i, \nu_i, \mu_k, \forall i, k\}$ and $\{\tau_{i,k}, \forall i, k\}$ are nonnegative values, and $\{\mathbf{B}_{i,k}, \mathbf{E}_{k,k}, \mathbf{A}_{k,l}\}$ are positive definite matrices. As a result, $\mathbf{D}_k$ is a positive definite matrix with rank$(\mathbf{D}_k) = M |I_k|$. Since rank$(\mathbf{Z}_k) \geq$ rank$(\mathbf{D}_k) -$ rank$(\mu_k \hat{\mathbf{g}}_{k,k} \hat{\mathbf{g}}_{k,k}^{\text{H}})$, we obtain rank$(\mathbf{Z}_k) \geq M |I_k| - 1$, where we use the fact that rank$(\mu_k \hat{\mathbf{g}}_{k,k} \hat{\mathbf{g}}_{k,k}^{\text{H}}) = 1$ since μ_k is positive according to Lemma 1. Furthermore, according to (B.3), we have rank$(\mathbf{W}_k) \leq M |I_k| -$ rank$(\mathbf{Z}_k)$. Then, we have rank$(\mathbf{W}_k) \leq 1$. Obviously, the optimal precoder $\mathbf{W}_k$ is not a zero matrix with rank$(\mathbf{W}_k) = 0$. Hence, the optimal precoder $\mathbf{W}_k$ has rank one, i.e., rank$(\mathbf{W}_k) = 1$, which completes the proof.

REFERENCES

[1] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud radio access network (C-RAN): A primer," *IEEE Netw.*, vol. 29, no. 1, pp. 35–41, Jan. 2015.

[3] S. Chen, F. Qin, B. Hu, X. Li, and Z. Chen, "User-centric ultra-dense networks for 5G: Challenges, methodologies, and directions," *IEEE Wireless Commun. Mag.*, vol. 23, no. 2, pp. 78–85, Apr. 2016.

[4] D. López-Pérez, M. Ding, H. Claussen, and A. H. Jafari, "Towards 1 Gbps/UE in cellular systems: Understanding ultra-dense small cell deployments," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2078–2101, 4th Quart., 2015.

[5] S. Stefanatos and A. Alexiou, "Access point density and bandwidth partitioning in ultra dense wireless networks," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3376–3384, Sep. 2014.

[6] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2522–2545, 4th Quart., 2016.

[7] X. Ge, S. Tu, G. Mao, C. X. Wang, and T. Han, "5G ultra-dense cellular networks," *IEEE Wireless Commun. Mag.*, vol. 23, no. 1, pp. 72–79, Feb. 2016.

[8] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.

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

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

[11] G. Caire, S. A. Ramprashad, and H. C. Papadopoulos, "Rethinking network MIMO: Cost of CSIT, performance analysis, and architecture comparisons," in *Proc. Inf. Theory Appl. Workshop (ITA)*, 2010, pp. 1–10.

[12] Y. Shi, J. Zhang, and K. B. Letaief, "CSI overhead reduction with stochastic beamforming for cloud radio access networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 5154–5159.

[13] C. Pan, H. Zhu, N. J. Gomes, and J. Wang, "Joint user selection and energy minimization for ultra-dense multi-channel C-RAN with incomplete CSI," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 8, pp. 1809–1824, Aug. 2017.

[14] T. R. Lakshmana, A. Tolli, R. Devassy, and T. Svensson, "Precoder design with incomplete feedback for joint transmission," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1923–1936, Mar. 2016.

[15] L. You, X. Gao, X. G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, Jun. 2015.

[16] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[17] H. Yin, L. Cottatellucci, D. Gesbert, R. R. Müller, and G. He, "Robust pilot decontamination based on joint angle and power domain discrimination," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2990–3003, Jun. 2016.

[18] X. Zhu, L. Dai, and Z. Wang, "Graph coloring based pilot allocation to mitigate pilot contamination for multi-cell massive MIMO systems," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1842–1845, Oct. 2015.

[19] P. Zhao, Z. Wang, C. Qian, L. Dai, and S. Chen, "Location-aware pilot assignment for massive MIMO systems in heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6815–6821, Aug. 2016.

[20] Z. Chen, X. Hou, and C. Yang, "Training resource allocation for user-centric base station cooperation networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2729–2735, Apr. 2016.

[21] D. Brélaz, "New methods to color the vertices of a graph," *Commun. ACM*, vol. 22, no. 4, pp. 251–256, 1979.

[22] *Summary of Reflector Discussions on E-UTRA UL RS*, document G. R1-074068, 3GPP TSG RAN WG1 Meeting 50 bis, Oct. 2007.

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

[24] M. Hong, R. Sun, H. Baligh, and Z. Q. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 226–240, Feb. 2013.

[25] W.-C. Liao, M. Hong, Y.-F. Liu, and Z.-Q. Luo, "Base station activation and linear transceiver design for optimal resource management in heterogeneous networks," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3939–3952, Aug. 2014.

[26] X. Huang, G. Xue, R. Yu, and S. Leng, "Joint scheduling and beamforming coordination in cloud radio access networks with QoS guarantees," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5449–5460, Jul. 2016.

[27] Q. T. Dinh and M. Diehl, "Local convergence of sequential convex programming for nonconvex optimization," in *Recent Advances in Optimization and its Applications in Engineering*. Berlin, Germany: Springer, 2010, pp. 93–102.

[28] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 62–75, May 2010.

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

[30] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, vol. 1. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

[31] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[32] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[33] W. Zhang *et al.*, "Large-scale antenna systems with UL/DL hardware mismatch: Achievable rates analysis and calibration," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1216–1229, Apr. 2015.

[34] E. Matskani, N. D. 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, Jul. 2008.

[35] C. Pan *et al.*, "Pricing-based distributed energy-efficient beamforming for MISO interference channels," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 710–722, Apr. 2016.

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

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

[38] *Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects*, document 3GPP TR 36.814, 2010.

[39] V. N. Ha, L. B. Le, and N.-D. Đào, "Coordinated multipoint transmission design for Cloud-RANs with limited fronthaul capacity constraints," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 7432–7447, Sep. 2016.

[40] J. Kim, H.-W. Lee, and S. Chong, "Virtual cell beamforming in cooperative networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1126–1138, Jun. 2014.

Cunhua Pan (M'16) received the B.S. and Ph.D. degrees from the School of Information Science and Engineering, Southeast University, Nanjing, China, in 2010 and 2015, respectively. From 2015 to 2016, he was a Research Associate with the University of Kent, U.K. He holds a post-doctoral position with the Queen Mary University of London, U.K.

His research interests include ultra-dense C-RAN, UAV, Internet of Things (IoT), NOMA, and mobile edge computing. He will serve as the Student Travel Grant Chair for ICC 2019, and a TPC member for many conferences, such as ICC and GLOBECOM.
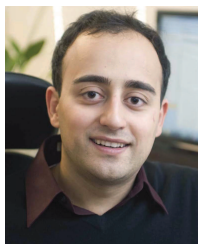
Hani Mehrpouyan (S'05–M'10) received the B.Sc. degree (Hons.) from Simon Fraser University, Burnaby, BC, Canada, in 2004, and the Ph.D. degree in electrical engineering from Queen's University, Kingston, ON, Canada, in 2010.

From 2010 to 2012, he held a postdoctoral position at the Chalmers University of Technology, Gothenburg, Sweden, where he led the MIMO aspects of the microwave backhauling for next-generation wireless networks project. In 2012, he joined the University of Luxembourg, Luxembourg, as a Research Associate, where he was responsible for new interference cancellation and synchronization schemes for satellite communication links. From 2012 to 2015, he was an Assistant Professor with California State University, Bakersfield, CA, USA. Since 2015, he has been an Assistant Professor with Boise State University, Boise, ID, USA. His research interests include applied signal processing and physical layer of wireless communication systems, including millimeter-wave systems, reconfigurable antennas, energy harvesting systems, synchronization, and channel estimation.

He has been involved with industry leaders, such as Ericsson AB and Blackberry. He has close to 60 publications in prestigious IEEE journals and conferences. He has received many scholarships and awards, including the IEEE Globecom Early Bird Student Award, NSERC-IRDF, NSERC PGS-D, NSERC CGS-M Alexander Graham Bell, and B.C. Wireless Innovation. He served as a TPC Member for the IEEE Globecom, ICC, and VTC. He is an Associate Editor of the the IEEE COMMUNICATION LETTERS.
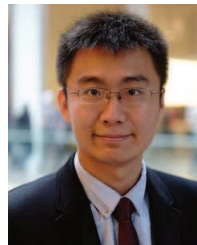
Yuanwei Liu (S'13–M'16) received the B.S. and M.S. degrees from the Beijing University of Posts and Telecommunications, in 2011 and 2014, respectively, and the Ph.D. degree in electrical engineering from the Queen Mary University of London, U.K., in 2016. He has been a Lecturer (Assistant Professor) with the School of Electronic Engineering and Computer Science, Queen Mary University of London, since 2017. From 2016 to 2017, he was with the Department of Informatics, King's College London, where he was a Post-Doctoral Research Fellow.

His research interests include 5G wireless networks, Internet of Things, stochastic geometry, and matching theory. He received the Exemplary Reviewer Certificate of the IEEE WIRELESS COMMUNICATION LETTERS in 2015 and the IEEE TRANSACTIONS ON COMMUNICATIONS in 2017. He has served as a TPC member for many IEEE conferences, such as GLOBECOM and ICC. He currently serves as an Editor of the IEEE COMMUNICATIONS LETTERS and the IEEE ACCESS.

Maged Elkashlan (M'06) received the Ph.D. degree in electrical engineering from The University of British Columbia, Canada, in 2006. From 2007 to 2011, he was with the Commonwealth Scientific and Industrial Research Organization, Australia. During this time, he held visiting appointments with the University of New South Wales and the University of Technology Sydney. In 2011, he was with the School of Electronic Engineering and Computer Science, Queen Mary University of London, U.K. His research interests are in the broad areas of communication theory and statistical signal processing.

He received the Best Paper Awards at the IEEE International Conference on Communications, in 2016 and 2014, the International Conference on Communications and Networking in China, in 2014, and the IEEE Vehicular Technology Conference, in 2013. He currently serves as an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

Arumugam Nallanathan (S'97–M'00–SM'05–F'17) was an Assistant Professor with the Department of Electrical and Computer Engineering, National University of Singapore, from 2000 to 2007. He was with the Department of Informatics, King's College London from 2007 to 2017, where he was a Professor of wireless communications from 2013 to 2017, and a Visiting Professor in 2017. Since 2017, he has been a Professor of wireless communications and the Head of the Communication Systems Research Group with the School of Electronic Engineering and Computer Science, Queen Mary University of London. He has published nearly 400 technical papers in scientific journals and international conferences. His research interests include 5G wireless networks, Internet of Things (IoT), and molecular communications. He is a co-recipient of the Best Paper Awards presented at the IEEE International Conference on Communications in 2016 and the IEEE Global Communications Conference in 2017. He is an IEEE Distinguished Lecturer. He has been selected as a Web of Science Highly Cited Researcher in 2016.

He received the IEEE Communications Society SPCE Outstanding Service Award in 2012 and the IEEE Communications Society RCC Outstanding Service Award in 2014. He served as the Chair for the Signal Processing and Communication Electronics Technical Committee of the IEEE Communications Society and Technical Program Chair and a member of the technical program committees in numerous IEEE conferences. He was an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2006 to 2011, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2006 to 2017, the IEEE WIRELESS COMMUNICATIONS LETTERS, and the IEEE SIGNAL PROCESSING LETTERS. He is an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS.