

Robust Beamforming Design for Ultra-dense User-Centric C-RAN in the Face of Realistic Pilot Contamination and Limited Feedback

Cunhua Pan, Hong Ren, Maged El Kashlan, Arumugam Nallanathan, *Fellow, IEEE*
and Lajos Hanzo, *Fellow, IEEE*

Abstract

The ultra-dense cloud radio access network (UD-CRAN), in which remote radio heads (RRHs) are densely deployed in the network, is considered. To reduce the channel estimation overhead, we focus on the design of robust transmit beamforming for user-centric frequency division duplex (FDD) UD-CRANs, where only limited channel state information (CSI) is available. Specifically, we conceive a complete procedure for acquiring the CSI that includes two key steps: channel estimation and channel quantization. The phase ambiguity (PA) is also quantized for coherent cooperative transmission. Based on the imperfect CSI, we aim for optimizing the beamforming vectors in order to minimize the total transmit power subject to users' rate requirements and fronthaul capacity constraints. We derive the closed-form expression of the achievable data rate by exploiting the statistical properties of multiple uncertain terms. Then, we propose a low-complexity iterative algorithm for solving this problem based on the successive convex approximation technique. In each iteration, the Lagrange dual decomposition method is employed for obtaining the optimal beamforming vector. Furthermore, a pair of low-complexity user selection algorithms are provided to guarantee the feasibility of the problem. Simulation results confirm the accuracy of our robust algorithm in terms of meeting the rate requirements. Finally, our simulation results verify that using a single bit for quantizing the PA is capable of achieving good performance.

Index Terms

C. Pan, H. Ren, M. El Kashlan and A. Nallanathan are with the Queen Mary University of London, London E1 4NS, U.K. (Email: {c.pan, h.ren, maged.elkashlan}@qmul.ac.uk). H. Ren was with National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China. L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk).

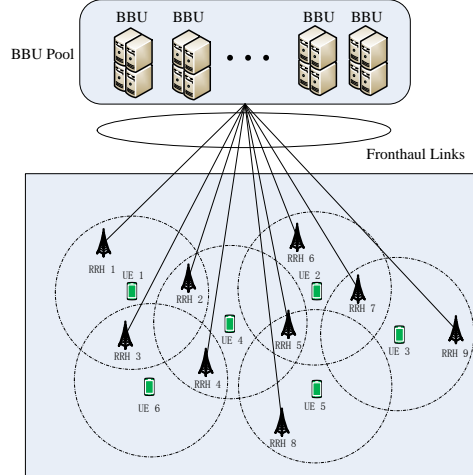


Fig. 1. Illustration of a UD-CRAN with nine RRHs and six UEs, i.e., $I = 9$, $K = 6$. To reduce the complexity, each UE is served by the RRHs within the dashed circle centered around the UE.

Ultra-dense networks (UDN), CRAN, virtual cell, limited feedback, fronthaul capacity constraints.

I. INTRODUCTION

Ultra dense networks (UDNs), where more and more small base stations (BSs) are deployed within a given area, have been widely regarded as one of the most promising techniques of achieving a high system throughput [1]. In UDNs, the average distance between small BSs and users can be dramatically reduced, which can translate into improved link reliability. However, since all small BSs reuse the same frequency, the users are also exposed to severe inter-cell interference, which is a severe performance limiting factor. Hence, the interference should be judiciously managed in order to reap the potential benefits of UDNs. As a result, the cloud radio access network (CRAN) concept has been recently proposed as a promising network architecture [2]. In CRAN, all the signal processing tasks are performed at the BBU pool, and all the conventional small BSs are replaced by low-cost low-power RRHs, which are only responsible for simple transmission/reception functions. The RRHs are connected to the BBU pool through the fronthaul links to support the centralized signal processing. Hence, the interference in the network can be effectively mitigated by employing the coordinated multipoint (CoMP) technique. Furthermore, due to their low-complexity functionalities, the mobile operator can densely deploy the RRHs at a low capital cost. Hence, the CRAN architecture is an ideal platform for supporting

UDNs. This kind of network is generally termed as an UD-CRAN [3], [4]. A simple example of UD-CRAN is illustrated in Fig. 1, where the number of RRHs is larger than that of the UEs.

Most of the existing contributions tend to deal with the various technical issues of conventional CRAN with a limited number of RRHs based on the assumption of the availability of perfect CSI [5]–[12]. In particular, Luong *et al.* [11] considered the transmit power minimization problem for the downlink of C-RANs with limited fronthaul capacity, where a pair of novel iterative algorithms were proposed for solving this problem. In the first one, the classic successive convex approximation framework was adopted for approximating the continuous nonconvex constraints, and the problem was converted into a mixed-integer second order cone program (MI-SOCP). By relaxing the binary variables to continuous valued variables, the second algorithm that is based on the so-called inflation procedure was proposed, which only has to solve a series of SOCP problems. Most recently, the same authors studied in [12] considered the tradeoff between the achievable sum-rate and total power consumption by using the radical multiobjective optimization concept, where the optimization problem was formulated as a mixed-integer nonconvex program. The authors proposed a branch and reduce and bound-based (BRB) algorithm for finding the globally optimal solution for benchmarking purposes, and also provided low-complexity iterative algorithms similar to the ones in [11].

However, the most challenging issue in UD-CRANs is that a large amount of CSI is required for facilitating CoMP transmission. The acquisition of the CSI requires a large amount of training resources that escalate rapidly with the network size. One of the most promising solutions is to consider the availability of only partial CSI. Specifically, each user only has to estimate the CSI of the links from the RRHs in its serving cluster (termed intra-cluster CSI), while only measuring the large-scale channel gains (such as path loss and shadowing) for the CSI of the links from the RRHs beyond its serving cluster (termed inter-cluster CSI). For the example in Fig. 1, UE 1 only needs to estimate the CSI from RRH 1, 2, and 3 to itself, while only the large-scale channel gains are required for the RRHs outside of its cluster. For this kind of scenario, the methods developed in [5]–[10] based on the assumption of perfect CSI cannot be tailored for this case.

Recently, the transmission design relying on partial CSI has attracted extensive research interests [13]–[16]. In particular, a novel compressive CSI acquisition method was proposed in [13] that can adaptively determine the set of instantaneous CSIs that should be estimated. The weighted sum-rate maximization problem was considered in [14], where the Cauchy-Schwarz

inequality was employed for deriving the lower-bound of the accurate data rate. The threshold-based channel matrix sparsification method was proposed in [15] for a UD-CRAN, where the authors demonstrated that only a negligible performance loss will be caused by discarding the channel matrix entries below a certain threshold. Finally, in our recent work [16], we proposed a unified framework to deal with the challenges arising in UD-CRAN, and Jensen's inequality was utilized to obtain a more tight lower bound on the achievable rate than that in [14].

However, in [13]–[16], perfect intra-cluster CSI was assumed to be available at the BBU pool, which is unrealistic for UD-CRANs, especially when the network operates in the frequency division duplex (FDD) mode [17], which is the focus of this paper. Tran *et al.* [18] considered the queue-aware robust beamforming design to minimize the average transmission power in the face of imperfect CSI for the whole C-RAN, while satisfying the outage probability constraint of each user. The classic Lyapunov optimization theory was employed for ensuring the systems stability. The Bernstein-Type Inequality [19] was utilized for transforming the outage probability constraints into a more tractable form that facilitates the application of the Semi-Definite Relaxation (SDR) approach. However, the channel error model is only suitable for the channel estimation error. In FDD UD-CRAN, each user has to estimate the intra-cluster CSI based on the pilot sequences sent from the RRHs within the serving cluster. Then, the user selects a codeword from a pre-designed CSI codebook to quantize the estimated CSI and feeds back its index to the BBU pool through a dedicated feedback channel. This procedure will impose three kinds of channel imperfections: channel estimation error, CSI quantization error and feedback delay. Since UD-CRANs are usually deployed in a limited area such as shopping malls and stadiums where the users move slowly, the effect of channel feedback delay can be ignored [16]. However, the other two error sources are inevitable and remain to be a serious problem in UD-CRANs.

To estimate the intra-cluster CSI, the pilot sequences sent from the RRHs that belong to the same user's serving cluster should be mutually orthogonal so that the user can differentiate the channels associated with different RRHs. For the example in Fig. 1, since RRH 1, 2, and 3 cooperatively serve UE 1, the pilot sequences sent from these RRHs should be mutually orthogonal. A direct method is to assign to all the RRHs mutually orthogonal pilots. However, the number of pilots linearly increases with the number of RRHs, which is excessive in UD-CRANs. To save the pilot resources, one should allow the RRHs serving no common user to reuse the same pilot. The authors [20], [21] provided novel pilot reuse schemes for minimizing

the total number of pilots required based on graph theory. In [22], Nguyen *et al.* proposed an iterative pilot allocation method for multicell massive MIMO networks, where the modified Hungarian method was adopted to solve the pilot allocation problem for each cell by fixing the pilot assignments for all the other cells. However, the beamforming direction was fixed and the computational complexity of the pilot assignment algorithm increases drastically with the number of cells. It is commonly known that the pilot reuse scheme will impose non-negligible pilot contamination, which inevitably leads to sizeable channel estimation error that cannot be eliminated. Hence, the channel estimation error should be taken into account when designing the transmission strategy. A robust beamforming design explicitly considering the channel estimation error was studied in our recent work [23] for time division duplex (TDD) UD-CRANs, where no channel quantization error is imposed as a benefit of the TDD channel's reciprocity.

Since coherent cooperative transmission among RRHs provides higher spectral efficiency than non-coherent transmission, we consider the limited feedback scenario of the former transmission scheme. To reduce the implementation complexity, the authors in [24], [25] advocated the per-RRH limited feedback strategy, where the estimated channels of all the links from all the candidate RRHs to each user are independently quantized rather than quantizing them jointly. However, this feedback strategy will result in phase ambiguity (PA) [24]. To elaborate, the PA is the phase differences between the single-RRH channel direction information (CDI) and the single-RRH quantized CSI codeword, which has no impact on the conventional single-cell channel quantization. However, it was shown in [24] that its adverse effect can be compensated by feeding back the PA information to the transmitter at the cost of a modest feedback overhead.

In this paper, we consider the robust downlink beamforming design of FDD UD-CRAN by taking into account all the channel uncertainties. Specifically, we aim for jointly optimizing the user-RRH associations and beamforming vectors for minimizing the total transmission power subject to users' rate requirements, fronthaul capacity constraints and per-RRH power constraints. This is a mixed integer non-linear programming (MINLP) problem that is generally difficult to solve. For the imperfect CSI considered in this paper, in contrast to the constraints of (6) and (7) in [11], the SINR constraints cannot be transformed into an SOCP format. Due to the same reason, the BRB algorithm in [12] aiming for globally optimal solution cannot be used for the imperfect CSI case. Furthermore, for the low-complexity algorithms developed in [11], [12], one has to solve an MI-SOCP or SOCP problem in each iteration, which incurs

high computational complexity for UD-CRAN. Specifically, the contributions of this paper are summarized as follows:

- 1) We provide a complete and practical procedure for the BBU pool to acquire the CSI required for centralized signal processing, namely for both channel estimation and channel quantization. To the best of our knowledge, this paper is the first attempt to unify these two steps into a joint framework. We derive the closed-form expression of the achievable data rate by exploiting statistical characteristics of the channel estimation error, the per-RRH CDI, the PA quantization errors and partial inter-cluster CSI.
- 2) To address the feasibility issue, we provide a pair of low-complexity user selection algorithms, namely the successive UE deletion method having a complexity order of $O(K)$ and a bisection based search method having a complexity order of $O(\log_2(K))$, where K is the total number of users. Simulation results show that the former algorithm performs better than the latter, and only slightly worse than the exhaustive search based method having an exponentially increasing complexity order of K . The performance loss is roughly 8% in the worst case.
- 3) Based on the feasible set of users given by the user selection algorithms, we propose a low-complexity iterative algorithm for solving the power minimization problem. Specifically, the non-smooth indicator function is approximated as a smooth concave real-valued fractional function, which is iteratively approximated by its first order Taylor expansion. In contrast to [23], this paper additionally considers the impact of CSI quantization errors, hence the semi-definite relaxation approach developed in [23] cannot be guaranteed to generate a rank-one beamforming solution. Instead, we approximate the complex-valued useful signal part in the rate expression by its first-order Taylor expansion with the aid of the \mathcal{T} -transform [26] that transforms complex-valued matrices and vectors into their real-valued equivalents. The transformed optimization problem becomes a convex one, and we derive the optimal beamforming vectors by employing the Lagrange dual decomposition method. Then, the successive convex approximation (SCA) technique is used for iteratively updating the corresponding variables that can guarantee to converge. Note that [11], [12] provided the results of the first-order Taylor expansion for the complex-valued expressions without a strict proof. Furthermore, The special structure of the resultant sub-problem has

TABLE I
THE LIST OF NOTATIONS

I	The number of RRHs	K	The number of UEs
\mathcal{I}	The set of RRHs	$\bar{\mathcal{U}}$	The set of UEs
\mathcal{U}	The set of selected UEs	\mathcal{I}_k	The candidate set of RRHs serving UE k
\mathcal{U}_i	The candidate set of UEs served by RRH i	$\mathbf{h}_{i,k}$	The channel from RRH i to UE k
$\mathbf{w}_{i,k}$	The BF vector from RRH i to UE k	$\alpha_{i,k}$	Large-scale channel gain from RRH i to UE k
$\bar{\mathbf{h}}_{i,k}$	Small-scale fading from RRH i to UE k	σ_k^2	Noise power at UE k
M	The number of antennas at each RRH	τ	The number of time slots for training
\mathcal{Q}	The set of pilot indices	\mathbf{Q}	The orthogonal pilot sequences
n_l	The reuse time of pilot l	n_{\max}	The maximum pilot reuse time
$\hat{\mathbf{h}}_{i,k}$	The MMSE estimation of channel $\mathbf{h}_{i,k}$	$\tilde{\mathbf{h}}_{i,k}$	Channel direction information of $\hat{\mathbf{h}}_{i,k}$
p_t	Pilot power $\mathbf{h}_{i,k}$	$\tilde{\mathbf{h}}_{i,k}$	Channel direction information of $\hat{\mathbf{h}}_{i,k}$
$\mathbf{q}_{i,k}$	The quantized version of $\tilde{\mathbf{h}}_{i,k}$	$\mathcal{C}_{i,k}$	Per-RRH codebook used by UE k
$B_{i,k}^{\text{CDI}}$	The number of bits to quantize CDI	$\phi_{i,k}$	The PA between CDI and its quantized codeword
$\tilde{\phi}_{i,k}$	The PA quantization error	$\hat{\phi}_{i,k}$	The quantized version of the PA $\phi_{i,k}$
$B_{i,k}^{\text{PA}}$	The number of bits to quantize PA	$a_{i,k}$	The quantization error of the CDI $\tilde{\mathbf{h}}_{i,k}$

not been exploited for developing a reduced-complexity algorithm for avoiding the direct solution of the MI-SOCP or SOCP.

The rest of this paper is organized as follows. Section II presents the system model. Section III formulates a two-stage optimization problem. A low-complexity iterative algorithm is provided in Section IV to deal with the transmit power minimization problem when the users are selected to be admitted. Two low-complexity user selection algorithms are presented in Section V. Extensive simulation results are given in Section VI. Finally, our conclusions are drawn in Section VII.

Notations: $\mathbb{E}_{\{x\}}\{y\}$ denotes the expectation of y over random variable x . $\mathcal{CN}(\mathbf{x}, \Sigma)$ denotes the complex Gaussian distribution with mean \mathbf{x} and variance Σ . The complex set is denoted as \mathbb{C} . \mathbf{I} and $\mathbf{0}$ are an identity matrix and a zero matrix, respectively. The transpose, conjugate transpose and the pseudo-inverse of matrix \mathbf{A} are denoted as \mathbf{A}^T , \mathbf{A}^H and \mathbf{A}^\dagger , respectively. $\mathbf{B} = \text{blkdiag}\{\mathbf{A}_i, i \in \mathcal{I}\}$ means that matrix \mathbf{B} is formed by performing the block diagonalization over \mathbf{A}_i . $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ represent real and imaginary parts of a variable, respectively. $f'_\theta(x)$ denotes the first-order derivative of $f_\theta(x)$. The other notations are summarized in Table I.

II. SYSTEM MODEL

A. Signal Transmission Model

Consider a downlink FDD UD-CRAN shown in Fig. 1, which has I RRHs and K UEs. Each RRH is equipped with M transmit antennas and each UE has a single receive antenna. The sets of RRHs and UEs are denoted as $\mathcal{I} = \{1, \dots, I\}$ and $\bar{\mathcal{U}} = \{1, \dots, K\}$, respectively. Each RRH is connected to the BBU pool through the wired/wireless fronthaul links. Let $\mathcal{U} \subseteq \bar{\mathcal{U}}$ represent the subset of UEs that can be admitted by the system. To reduce the computational complexity associated with the UD-CRAN, the user-centric cluster technique is considered, where each UE is exclusively served by its nearby RRHs, since the signals arriving from distant RRHs are weak at the UE due to the severe path loss. For the example of Fig. 1, UE 1 is only potentially served by RRH 1, RRH 2 and RRH 3. The set of RRHs that potentially serve UE k is denoted as $\mathcal{I}_k \subseteq \mathcal{I}$, or equivalently the candidate set of RRHs that serve UE k is denoted as \mathcal{I}_k . It should be emphasized that the set of RRHs that finally serve UE k may not be the same as \mathcal{I}_k , which needs to be optimized in the following sections, while the RRHs out of its cluster, i.e., $\mathcal{I} \setminus \mathcal{I}_k$, will not serve UE k . Additionally, let us denote $\mathcal{U}_i \subseteq \mathcal{U}$ as the set of UEs that are potentially served by RRH i . Note that the clusters for the UEs may overlap with each other, which means that each RRH can simultaneously serve multiple UEs. These clusters are assumed to be predetermined based on the large-scale channel gains that vary slowly.

Let us denote by $\mathbf{h}_{i,k} \in \mathbb{C}^{M \times 1}$ and $\mathbf{w}_{i,k} \in \mathbb{C}^{M \times 1}$ the channel vector and beamforming vector of the links spanning from RRH i to UE k , respectively. Then, the signal received at UE k is

$$y_k = \underbrace{\sum_{i \in \mathcal{I}_k} \mathbf{h}_{i,k}^H \mathbf{w}_{i,k} s_k}_{\text{desired signal}} + \underbrace{\sum_{l \neq k, l \in \mathcal{U}} \sum_{i \in \mathcal{I}_l} \mathbf{h}_{i,k}^H \mathbf{w}_{i,l} s_l}_{\text{interference}} + z_k, \quad (1)$$

where s_l denotes the transmission data for UE l and z_k is the zero-mean additive complex white Gaussian noise with variance σ_k^2 . It is assumed that the data destined for each UE is independent of each other and it has a zero mean and unit variance, i.e., we have $\mathbb{E}\{|s_k|^2\} = 1$ 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 $\alpha_{i,k}$ represents the large-scale channel gains of the links spanning from RRH i to UE k that accounts both for the shadowing and path loss, while $\bar{\mathbf{h}}_{i,k}$ is the small-scale channel fading with the distribution of $\mathcal{CN}(\mathbf{0}, \mathbf{I})$.

B. Channel Estimation for Intra-cluster CSI

To design the beam-vectors for the UEs, the overall CSI should be available at the BBU pool for the facilitation of joint transmission. However, it is an unaffordable task to estimate the CSI from all RRHs to all UEs due to the limited availability of training resources. An appealing approach is that each UE only estimates the CSI within its cluster, named intra-cluster CSI. For the CSI beyond this cluster, it is assumed that only large-scale channel gains are available, i.e., $\{\alpha_{i,k}, \forall i \in \mathcal{I} \setminus \mathcal{I}_k, k \in \mathcal{U}\}$. The out-cluster large-scale channel gains are used to control the multiuser interference.

In this paper, we assume that τ time slots are used for CSI training, thus the length of pilot sequences is τ , or equivalently the number of orthogonal pilot sequences is equal to τ . Let us denote the set of pilot indices as $\mathcal{Q} = \{1, 2, \dots, \tau\}$, and the corresponding orthogonal pilot sequences as $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_\tau] \in \mathbb{C}^{\tau \times \tau}$ that satisfies the orthogonal condition $\mathbf{Q}^H \mathbf{Q} = \mathbf{I}$.

For the channel estimation in an FDD UD-CRAN system, the RRHs first send the training sequences to the UEs, then the UEs estimate their channels based on their received signals. Specifically, the training signals received at UE k can be written as

$$\mathbf{y}_k = \sum_{i \in \mathcal{I}_k} \sqrt{p_t} \mathbf{h}_{i,k}^H \mathbf{X}_i^H + \sum_{i \in \mathcal{I} \setminus \mathcal{I}_k} \sqrt{p_t} \mathbf{h}_{i,k}^H \mathbf{X}_i^H + \mathbf{n}_k, \quad (2)$$

where p_t is the pilot transmit power at each transmit antenna, $\mathbf{n}_k \in \mathbb{C}^{1 \times \tau}$ is the additive Gaussian noise vector during the training time slots, whose elements are independently generated and follow the distributions of $\mathcal{CN}(0, \sigma_k^2)$, $\mathbf{X}_i \in \mathbb{C}^{\tau \times M}$ is the pilot training matrix sent from RRH i . The training matrix \mathbf{X}_i can be written as $\mathbf{X}_i = [\mathbf{q}_{\pi_i^1}, \dots, \mathbf{q}_{\pi_i^M}]$, where $\mathbf{q}_{\pi_i^m} \in \mathbb{C}^{\tau \times 1}$ denotes the pilot sequence used for estimating the channels spanning from the m th antenna of RRH i to the UEs.

To conserve the pilot resources, a pilot reuse scheme is considered, which should satisfy the following constraints: 1) The pilot sequences from different RRHs in the same cluster should also be orthogonal, i.e. $\mathbf{X}_m^H \mathbf{X}_n = \mathbf{0}$ for $m, n \in \mathcal{I}_k, m \neq n, \forall k \in \mathcal{U}$; 2) The maximum reuse time for each pilot sequence should be restricted to a small value for reducing the channel estimation error. Let us denote the reuse time for pilot l as n_l . Then this condition can be expressed as $n_l \leq n_{\max}, \forall l \in \mathcal{Q}$. 3) The pilot sequences used by all antennas at the same RRH should be mutually orthogonal, i.e. $\mathbf{X}_i^H \mathbf{X}_i = \mathbf{I}$. The first constraint means that the RRHs serving the same UE should use an orthogonal pilot matrix. A natural pilot allocation approach to satisfy the

above three constraints is the orthogonal pilot allocation scheme, where all antennas and RRHs are allocated orthogonal pilots. Obviously, the number of pilots required is MI , which occupies lots of time slots for UD-CRANs having a large number of RRHs. Hence, if we allow some RRHs to reuse the same set of pilots, the number of pilot sequences required will be reduced. In this paper, we aim for minimizing the number of pilots required, while guaranteeing the above three conditions. This pilot allocation problem has been studied in [20], where the Dsatur algorithm from graph theory was proposed to solve it. The computational complexity of the Dsatur algorithm is given by $\mathcal{O}(I^2)$ [20]. When some RRHs are allocated the same color, these RRHs can reuse the same pilot. Denote c^* as the number of different colors required by the Dsatur algorithm to color all the RRHs. Then the total number of pilots required is given by $\tau = Mc^*$, since the antennas in each RRH use different pilots.

Let us define $\mathcal{K}_{\mathbf{X}} = \{i : \mathbf{X}_i = \mathbf{X}\}$ as the set of RRHs that reuse the same pilots \mathbf{X} obtained by using the Dsatur algorithm. Then, the MMSE estimation of channel $\mathbf{h}_{i,k}$ is given by [27]

$$\hat{\mathbf{h}}_{i,k} = \frac{\alpha_{i,k}}{\sum_{m \in \mathcal{K}_{\mathbf{X}_i}} \alpha_{m,k} + \hat{\sigma}^2} \frac{1}{\sqrt{p_t}} \mathbf{X}_i^H \mathbf{y}_k^H, \quad (3)$$

where $\hat{\sigma}^2 = \sigma_k^2/p_t$. It can be readily derived from (3) that the channel estimate $\hat{\mathbf{h}}_{i,k}$ obeys the distribution of $\mathcal{CN}(\mathbf{0}, \omega_{i,k} \mathbf{I})$ with $\omega_{i,k}$ given by

$$\omega_{i,k} = \frac{\alpha_{i,k}^2}{\sum_{m \in \mathcal{K}_{\mathbf{X}_i}} \alpha_{m,k} + \hat{\sigma}^2}. \quad (4)$$

According to the property of MMSE estimation [27], the channel estimate error $\mathbf{e}_{i,k} = \mathbf{h}_{i,k} - \hat{\mathbf{h}}_{i,k}$ is independent of the channel estimate $\hat{\mathbf{h}}_{i,k}$, which follows the distribution of $\mathcal{CN}(\mathbf{0}, \delta_{i,k} \mathbf{I})$, with $\delta_{i,k}$ given by

$$\delta_{i,k} = \frac{\alpha_{i,k} \left(\sum_{m \in \mathcal{K}_{\mathbf{X}_i} \setminus i} \alpha_{m,k} + \hat{\sigma}^2 \right)}{\sum_{m \in \mathcal{K}_{\mathbf{X}_i}} \alpha_{m,k} + \hat{\sigma}^2}. \quad (5)$$

Note that even when RRH i does not reuse any pilots of any other RRHs, there is still some channel estimation error for channel $\mathbf{h}_{i,k}$ with $\delta_{i,k} = \alpha_{i,k} \hat{\sigma}^2 / (\alpha_{i,k} + \hat{\sigma}^2)$.

C. Limited Feedback Model

In this paper, we consider the limited per-RRH codebook feedback strategy [24], where each UE uses different codebooks to independently quantize its per-RRH CDI, i.e., $\tilde{\mathbf{h}}_{i,k} = \hat{\mathbf{h}}_{i,k} / \|\hat{\mathbf{h}}_{i,k}\|$. Then UE k feeds back the indices of codewords to its corresponding serving RRHs. The BBU

pool will collect all the indices sent from different RRHs and will design beamforming vectors based on these indices. Specifically, the quantized version of the CDI $\tilde{\mathbf{h}}_{i,k}$ is given by

$$\mathbf{q}_{i,k} = \arg \max_{\mathbf{c}_{i,k,n} \in \mathcal{C}_{i,k}} \left| \tilde{\mathbf{h}}_{i,k}^H \mathbf{c}_{i,k,n} \right|, \quad (6)$$

where $\mathcal{C}_{i,k}$ is the per-RRH codebook used by UE k to quantize the CSI spanning from RRH i , which consists of unit-norm codewords $\mathbf{c}_{i,k,n} \in \mathbb{C}^{M \times 1}$, $n = 1, \dots, 2^{B_{i,k}^{\text{CDI}}}$, with $B_{i,k}^{\text{CDI}}$ denoting the number of bits used for quantizing the CDI $\tilde{\mathbf{h}}_{i,k}$.

Coherent joint transmission is assumed in this paper. Then, another important parameter namely the phase ambiguity (PA) is also required at the BBU pool [24], [25]. The PA is defined as the angle between the per-RRH CDI and its quantized codeword, i.e., $e^{j\phi_{i,k}} = \tilde{\mathbf{h}}_{i,k}^H \mathbf{q}_{i,k} / \left\| \tilde{\mathbf{h}}_{i,k}^H \mathbf{q}_{i,k} \right\|$ with $j = \sqrt{-1}$. The PA knowledge is not required for single-point limited feedback MIMO systems, but affects the co-phasing of the coherent joint transmission in UD-CRAN, as detailed in [24], [25]. The PA can be fed back with the aid of a few bits by using scalar quantization. Since the codeword is chosen by maximizing the magnitude of $\tilde{\mathbf{h}}_{i,k}^H \mathbf{c}_{i,k,n}$ and the CDI is isotropically distributed, the PA $\phi_{i,k}$ will be uniformly distributed in $[0, 2\pi]$. Hence, it is optimal to quantize the PA employing a uniform scalar quantizer. Let us denote by $\tilde{\phi}_{i,k}$ and $\hat{\phi}_{i,k}$ the PA quantization error and the quantized version of the PA $\phi_{i,k}$, respectively. Then, the PA $\phi_{i,k}$ can be represented as $\phi_{i,k} = \hat{\phi}_{i,k} + \tilde{\phi}_{i,k}$. If we use $B_{i,k}^{\text{PA}}$ bits to quantize PA $\phi_{i,k}$, the PA quantization error $\tilde{\phi}_{i,k}$ is uniformly distributed within $\left[-\frac{\pi}{2^{B_{i,k}^{\text{PA}}}}, \frac{\pi}{2^{B_{i,k}^{\text{PA}}}} \right]$.

Let us define by $a_{i,k} \triangleq 1 - \left| \tilde{\mathbf{h}}_{i,k}^H \mathbf{q}_{i,k} \right|^2$ the quantization error of the CDI $\tilde{\mathbf{h}}_{i,k}$. For simplicity, random vector quantization (RVQ) is considered for quantizing the per-RRH CDIs in this paper. Then, according to [28], the per-RRH CDI $\tilde{\mathbf{h}}_{i,k}$ can be rewritten as

$$\tilde{\mathbf{h}}_{i,k} = \sqrt{1 - a_{i,k}} e^{j\phi_{i,k}} \mathbf{q}_{i,k} + \sqrt{a_{i,k}} \mathbf{u}_{i,k}, \quad (7)$$

where $\mathbf{u}_{i,k}$ is channel quantization error, which is a unit-norm vector isotropically distributed in the nullspace of $\mathbf{q}_{i,k}$.

In this paper, we assume that there are dedicated error-free feedback channels for feeding back all quantized versions of CDIs and PAs to the BBU pool. Then, the BBU pool determines the beamforming vectors based on the feedback information.

III. PROBLEM FORMULATION

In this section, we first provide the mathematical model for the constraints of the UD-CRAN, which include each UE's data rate requirement, the per-RRH power constraint and limited fronthaul capacity constraint. Then, based on these constraints, we formulate the UE selection problem and the transmit power minimization problem in a two-stage form.

Let us denote the beamforming vectors from all RRHs in \mathcal{I}_k by $\mathbf{w}_k = [\mathbf{w}_{i,k}^H, \forall i \in \mathcal{I}_k]^H \in \mathbb{C}^{|\mathcal{I}_k|M \times 1}$, and the aggregated channel vectors from RRHs in \mathcal{I}_l to UE k by $\mathbf{g}_{l,k} = [\mathbf{h}_{i,k}^H, \forall i \in \mathcal{I}_l]^H \in \mathbb{C}^{|\mathcal{I}_l|M \times 1}$. In addition, define $\tilde{\mathbf{g}}_{k,k} = [\mathbf{e}_{i,k}^H, \forall i \in \mathcal{I}_k]^H \in \mathbb{C}^{|\mathcal{I}_k|M \times 1}$ and $\hat{\mathbf{g}}_{k,k} = [\hat{\mathbf{h}}_{i,k}^H, \forall i \in \mathcal{I}_k]^H \in \mathbb{C}^{|\mathcal{I}_k|M \times 1}$ as the overall CSI error and estimated CSI of the links spanning from the RRHs in \mathcal{I}_k to UE k , respectively. Then, the channel estimation error can be rewritten as $\tilde{\mathbf{g}}_{k,k} = \mathbf{g}_{k,k} - \hat{\mathbf{g}}_{k,k}$, while the received signal model in (1) can be reformulated as

$$y_k = \underbrace{\hat{\mathbf{g}}_{k,k}^H \mathbf{w}_k s_k}_{\text{Desired signal}} + \underbrace{\tilde{\mathbf{g}}_{k,k}^H \mathbf{w}_k s_k}_{\text{Residual-interference}} + \underbrace{\sum_{l \neq k, l \in \mathcal{U}} \mathbf{g}_{l,k}^H \mathbf{w}_l s_l}_{\text{Multi-user Interference}} + z_k, \forall k \in \mathcal{U}. \quad (8)$$

As in most existing papers [29], [30], we consider the achievable data rate, where the residual-interference term in (8) due to the channel estimation error is treated as uncorrelated Gaussian noise. Additionally, for the sake of reducing the decoding complexity, the multi-user interference term is also regarded as uncorrelated Gaussian noise. By considering the time slots allocated for channel training, the net achievable data rate of UE k can be expressed as [30]

$$r_k = \frac{T - \tau}{T} \log_2 \left(1 + \frac{\mathbb{E} \left\{ |\hat{\mathbf{g}}_{k,k}^H \mathbf{w}_k|^2 \right\}}{\mathbb{E} \left\{ |\tilde{\mathbf{g}}_{k,k}^H \mathbf{w}_k|^2 \right\} + \sum_{l \neq k, l \in \mathcal{U}} \mathbb{E} \left\{ |\mathbf{g}_{l,k}^H \mathbf{w}_l|^2 \right\} + \sigma_k^2} \right), \forall k \in \mathcal{U}, \quad (9)$$

where τ is the total number of time slots required by the Dsatur algorithm, T denotes the total number of time slots in each time frame, and the expectation is taken over multiple random processes, namely, the fast fading of the unknown CSI in $\mathcal{I} \setminus \mathcal{I}_k$, the channel estimation errors $\{\mathbf{e}_{i,k}, i \in \mathcal{I}_k\}$, the CDI quantization errors $\{\mathbf{u}_{i,k}, \forall i \in \mathcal{I}_k\}$ and the PA quantization errors $\{\tilde{\phi}_{i,k}, \forall i \in \mathcal{I}_k\}$. Each UE's data rate should be higher than its minimum rate requirement:

$$\text{C1} : r_k \geq R_{k,\min}, \forall k \in \mathcal{U}, \quad (10)$$

where $R_{k,\min}$ is the rate target of UE k .

The second constraint is the per-RRH power constraint, which can be expressed as

$$\text{C2} : \sum_{k \in \mathcal{U}_i} \|\mathbf{w}_{i,k}\|^2 \leq P_{i,\max}, i \in \mathcal{I}, \quad (11)$$

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

Finally, each fronthaul link has a capacity constraint, since we consider a limited bandwidth. Specifically, this kind of constraint can be expressed as

$$\text{C3} : \sum_{k \in \mathcal{U}_i} \varepsilon(\|\mathbf{w}_{i,k}\|^2) r_k \leq C_{i,\max}, \forall i \in \mathcal{I}, \quad (12)$$

where $C_{i,\max}$ is the capacity limit of the fronthaul link spanning from the BBU pool to RRH i , and $\varepsilon(\cdot)$ is an indicator function, defined as

$$\varepsilon(x) = \begin{cases} 1, & \text{if } x \neq 0, \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

Due to the constraints of the system (C2 and C3), some UEs' rate requirements (C1) may not be satisfied. Hence, some UEs should be removed in order to guarantee the QoS requirements of the remaining UEs. Similar to [10], [16], we formulate a two-stage optimization.

Specifically, in Stage I, we aim for maximizing the number of UEs admitted to the dense network, which is formulated as

$$\begin{aligned} \mathcal{P}_1 : \quad & \max_{\mathbf{w}, \mathcal{U} \subseteq \mathcal{U}} |\mathcal{U}| \\ \text{s.t.} \quad & \text{C1, C2, C3,} \end{aligned} \quad (14)$$

where \mathbf{w} denotes the set of all beamforming vectors and $|\mathcal{U}|$ is the cardinality of the set \mathcal{U} .

In Stage II, our goal is to optimize the beamforming vectors for minimizing the total transmit power with the UEs selected from Stage I. Let us denote by \mathcal{U}^* the specific solution from Stage I, where the corresponding \mathcal{U}_i becomes \mathcal{U}_i^* . Then the optimization problem in Stage II is

$$\begin{aligned} \mathcal{P}_2 : \quad & \min_{\mathbf{w}} \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{U}_i^*} \|\mathbf{w}_{i,k}\|_2^2 \\ \text{s.t.} \quad & \text{C1, C2, C3.} \end{aligned} \quad (15)$$

In constraints C1-C3, \mathcal{U} and \mathcal{U}_i are replaced by \mathcal{U}^* and \mathcal{U}_i^* , respectively.

Problems \mathcal{P}_1 and \mathcal{P}_2 in (14) and (15) are difficult to solve. The reasons are given as follows. Firstly, the exact data rate r_k is difficult to derive, since the expectation is taken over multiple uncertain terms. Secondly, both the objective function and the fronthaul capacity constraint C3 of Problem \mathcal{P}_1 contain the non-smooth and non-differential indicator functions, which is recognized as a mixed-integer non-linear programming (MINLP) problem. The exhaustive search method can be adopted to solve this kind of optimization problem. However, it has an exponential complexity order, which becomes excessive for UD-CRAN with large number of UEs.

In the following section, we first deal with the power minimization Problem \mathcal{P}_2 by assuming that the set of admitted UEs has already been determined by solving Problem \mathcal{P}_1 . Then, we will conceive low-complexity methods to deal with Problem \mathcal{P}_1 in Section V.

IV. LOW-COMPLEXITY ALGORITHM TO DEAL WITH PROBLEM \mathcal{P}_2

In this section, we provide a low-complexity algorithm for solving Problem \mathcal{P}_2 , when the UEs to be admitted have already been selected by using the UE selection algorithms in Section V, and denote the subset of UEs that have been selected as \mathcal{U} . In the following, we first simplify the rate expression.

The multiple random processes in the rate expression make the accurate closed-form expression of the achievable data rate of UE k in (9) difficult to derive. In Appendix A of [31], we derived the achievable data rate as

$$r_k = \frac{T - \tau}{T} \log_2 \left(1 + \frac{\mathbf{w}_k^H \mathbf{A}_{k,k} \mathbf{w}_k}{\mathbf{w}_k^H \mathbf{E}_{k,k} \mathbf{w}_k + \sum_{l \neq k, l \in \mathcal{U}} \mathbf{w}_l^H \mathbf{A}_{l,k} \mathbf{w}_l + \sigma_k^2} \right) \quad (16)$$

$$= \frac{T - \tau}{T} \log_2 (1 + \text{SINR}_k), \quad (17)$$

where we have $\mathbf{E}_{k,k} = \mathbb{E} \{ \tilde{\mathbf{g}}_{k,k} \tilde{\mathbf{g}}_{k,k}^H \} \in \mathbb{C}^{M|\mathcal{I}_k| \times M|\mathcal{I}_k|}$, $\mathbf{A}_{k,k} = \mathbb{E} \{ \hat{\mathbf{g}}_{k,k} \hat{\mathbf{g}}_{k,k}^H \} \in \mathbb{C}^{M|\mathcal{I}_k| \times M|\mathcal{I}_k|}$ and $\mathbf{A}_{l,k} = \mathbb{E} \{ \mathbf{g}_{l,k} \mathbf{g}_{l,k}^H \} \in \mathbb{C}^{M|\mathcal{I}_l| \times M|\mathcal{I}_l|}$. The matrix $\mathbf{E}_{k,k}$ can be readily computed as

$$\mathbf{E}_{k,k} = \text{blkdiag} \{ \delta_{i,k} \mathbf{I}_M, i \in \mathcal{I}_k \}, \quad (18)$$

while $\mathbf{A}_{k,k}$ and $\mathbf{A}_{l,k}$ are given in Appendix A, respectively. Note that the matrices $\mathbf{E}_{k,k}$, $\mathbf{A}_{k,k}$ and $\mathbf{A}_{l,k}$ are semi-definite matrices, since they represent the expectations over semi-definite matrices [32]. The achievable signal to interference plus noise ratio (SINR) of UE k is given by

$$\text{SINR}_k = \frac{\mathbf{w}_k^H \mathbf{A}_{k,k} \mathbf{w}_k}{\mathbf{w}_k^H \mathbf{E}_{k,k} \mathbf{w}_k + \sum_{l \neq k, l \in \mathcal{U}} \mathbf{w}_l^H \mathbf{A}_{l,k} \mathbf{w}_l + \sigma_k^2}. \quad (19)$$

By exploiting the fact that the rate constraints hold with equality at the optimal point [23], Problem \mathcal{P}_2 can be transformed as

$$\mathcal{P}_3 : \min_{\mathbf{w}} \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{U}_i} \|\mathbf{w}_{i,k}\|_2^2 \quad (20a)$$

$$\text{s.t.} \quad \text{C2, C4} : \text{SINR}_k \geq \eta_{k,\min}, \forall i \in \mathcal{I}, \quad (20b)$$

$$\text{C5} : \sum_{k \in \mathcal{U}_i} \varepsilon (\|\mathbf{w}_{i,k}\|^2) R_{k,\min} \leq C_{i,\max}, \forall i \in \mathcal{I}, \quad (20c)$$

where $\eta_{k,\min} = 2^{\frac{T}{T-\tau} R_{k,\min}} - 1$.

A. Smooth Approximation of the Indicator Function

We first deal with the non-smooth nature of the indicator function in C5. Similar to [16], the indicator function is approximated by the smooth function $f_\theta(x) = \frac{x}{x+\theta}$, where θ is a small constant. By replacing the indicator function with $f_\theta(x)$, Problem \mathcal{P}_3 can be approximated as

$$\mathcal{P}_4 : \min_{\mathbf{w}} \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{U}_i} \|\mathbf{w}_{i,k}\|_2^2 \quad (21a)$$

$$\text{s.t.} \quad \text{C2, C4, C6} : \sum_{k \in \mathcal{U}_i} f_\theta(\|\mathbf{w}_{i,k}\|^2) R_{k,\min} \leq C_{i,\max}, \forall i \in \mathcal{I}. \quad (21b)$$

The successive convex approximation (SCA) method [33] is used to deal with the non-convex constraint C7. Specifically, by exploiting the concavity of $f_\theta(x)$, we have

$$f_\theta(\|\mathbf{w}_{i,k}\|^2) \leq f_\theta(\|\mathbf{w}_{i,k}(t)\|^2) + \beta_{i,k}(t)(\|\mathbf{w}_{i,k}\|^2 - \|\mathbf{w}_{i,k}(t)\|^2), \quad (22)$$

where $\mathbf{w}_{i,k}(t)$ is the beamforming vector at the t th iteration, $\beta_{i,k}(t) = f'_\theta(\|\mathbf{w}_{i,k}(t)\|^2)$. By replacing $f_\theta(\|\mathbf{w}_{i,k}\|^2)$ in Problem \mathcal{P}_4 with the right hand side of (22), we arrive at

$$\mathcal{P}_5 : \min_{\mathbf{w}} \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{U}_i} \|\mathbf{w}_{i,k}\|_2^2 \quad (23a)$$

$$\text{s.t.} \quad \text{C2, C4, C7} : \sum_{k \in \mathcal{U}_i} \tau_{i,k}(t) \|\mathbf{w}_{i,k}\|^2 \leq \tilde{C}_i(t), \forall i \in \mathcal{I}, \quad (23b)$$

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

However, Problem \mathcal{P}_5 is still difficult to solve due to Constraint C4, although it has been simplified from Constraint C1. The reasons are given as follows. Due to the channel estimation error, each user suffers from residual interference, as seen from the right hand side of C4, i.e. $\mathbf{w}_k^H \mathbf{E}_{k,k} \mathbf{w}_k$. Although the classic weighted minimum mean square error (WMMSE) method has been successfully applied in UD-CRANs under the idealized simplifying assumptions of having perfect intra-cluster CSI [7], [10], [16], it cannot be adopted in this realistic optimization problem due to the residual-interference. Furthermore, note that the rank of matrix $\mathbf{A}_{k,k}$ is in general higher than one, the Semi-definite (SDP) relaxation method used in [23] cannot be adopted here, since the resultant solution is no longer guaranteed to be of rank one. In the following, we propose a novel method to deal with Constraint C4.

B. Method to Deal with Constraint C4

In this following, we propose a novel method based on the first-order Taylor approximation to deal with Constraint C4 and then propose the Lagrange dual decomposition algorithm for solving this problem.

Constraint C4 is non-convex, because $\mathbf{w}_k^H \mathbf{A}_{k,k} \mathbf{w}_k$ is a convex function of \mathbf{w}_k ¹. Similar to the successive convex approximation method dealing with the concave fractional function, we approximate it by its first-order Taylor expansion and make Constraint C4 convex. Since $\mathbf{w}_k^H \mathbf{A}_{k,k} \mathbf{w}_k$ is convex, we have

$$\mathbf{w}_k^H \mathbf{A}_{k,k} \mathbf{w}_k \geq \mathbf{w}_k^H(t) \mathbf{A}_{k,k} \mathbf{w}_k(t) + 2\text{Re} \left\{ \mathbf{w}_k^H(t) \mathbf{A}_{k,k} (\mathbf{w}_k - \mathbf{w}_k(t)) \right\}, \quad (24)$$

where $\mathbf{w}_k(t)$ is the beamforming vector at the t th iteration. The above derivation is not direct since $\mathbf{w}_k^H \mathbf{A}_{k,k} \mathbf{w}_k$ is a function of complex-valued vector \mathbf{w}_k . The Taylor expansion developed for the functions over real-valued variables cannot be directly extended to the complex case. In Appendix B, we derived the above result relying on the so-called \mathcal{T} -transform [26] that transforms complex-valued matrices and vectors into their real-valued equivalents.

By replacing $\mathbf{w}_k^H \mathbf{A}_{k,k} \mathbf{w}_k$ in C4 with the right side of (24), Problem \mathcal{P}_5 is transformed to the following optimization problem

$$\mathcal{P}_6 : \min_{\mathbf{w}} \sum_{k \in \mathcal{U}} \|\mathbf{w}_k\|_2^2 \quad (25a)$$

$$\text{s.t.} \quad \text{C2, C7}, \quad (25b)$$

$$\begin{aligned} \text{C8} : 2\text{Re} \left(\mathbf{w}_k^H(t) \mathbf{A}_{k,k} \mathbf{w}_k \right) - \zeta_k(t) \geq \\ \eta_{k,\min} \left(\mathbf{w}_k^H \mathbf{E}_{k,k} \mathbf{w}_k + \sum_{l \neq k, l \in \mathcal{U}} \mathbf{w}_l^H \mathbf{A}_{l,k} \mathbf{w}_l + \sigma_k^2 \right), \forall k \in \mathcal{U}, \end{aligned} \quad (25c)$$

where $\zeta_k(t) = \mathbf{w}_k^H(t) \mathbf{A}_{k,k} \mathbf{w}_k(t)$. Now, Problem \mathcal{P}_6 is a convex optimization problem. Additionally, in Appendix C, we prove that the Slater's condition [32] of Problem \mathcal{P}_6 is satisfied. Hence, the duality gap between Problem \mathcal{P}_6 and its dual problem is zero. As a result, the original Problem \mathcal{P}_6 can be solved by solving its dual problem instead. In the following, we derive the structure of the optimal beamforming vector by applying the Lagrange dual decomposition method.

Let us represent \mathcal{I}_k as $\mathcal{I}_k = \{s_1^k, \dots, s_{|\mathcal{I}_k|}^k\}$. We first introduce the following block-diagonal matrices

$$\mathbf{B}_{i,k} = \text{diag} \left\{ \overbrace{\mathbf{0}_{1 \times M}}^{s_1^k}, \dots, \overbrace{\mathbf{1}_{1 \times M}}^{s_m^k}, \overbrace{\mathbf{0}_{1 \times M}}^{s_{m+1}^k}, \dots, \overbrace{\mathbf{0}_{1 \times M}}^{s_{|\mathcal{I}_k|}^k} \right\}, \text{ if } s_m^k = i, \forall i \in \mathcal{I}, k \in \mathcal{U}. \quad (26)$$

¹Note that $\mathbf{A}_{k,k}$ is a semi-definite matrix.

Then, Constraints C2 and C7 can be rewritten as

$$\text{C9} : \sum_{k \in \mathcal{U}_i} \mathbf{w}_k^H \mathbf{B}_{i,k} \mathbf{w}_k \leq P_{i,\max}, \forall i \in \mathcal{I} \quad (27)$$

$$\text{C10} : \sum_{k \in \mathcal{U}_i} \tau_{i,k}(t) \mathbf{w}_k^H \mathbf{B}_{i,k} \mathbf{w}_k \leq \tilde{C}_i(t), \forall i \in \mathcal{I}. \quad (28)$$

After some further manipulations, the Lagrangian function of Problem \mathcal{P}_6 can be written as

$$\begin{aligned} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}) &= \sum_{k \in \mathcal{U}} \mathbf{w}_k^H \mathbf{J}_k(t) \mathbf{w}_k - \sum_{k \in \mathcal{U}} v_k [\mathbf{w}_k^H(t) \mathbf{A}_{k,k} \mathbf{w}_k + \mathbf{w}_k^H \mathbf{A}_{k,k} \mathbf{w}_k(t)] - \sum_{i \in \mathcal{I}} \lambda_i P_{i,\max} - \sum_{i \in \mathcal{I}} \mu_i \tilde{C}_i(t) \\ &\quad + \sum_{k \in \mathcal{U}} v_k [\eta_{k,\min} \sigma_k^2 + \zeta_k(t)], \end{aligned}$$

where $\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}$ are the collections of non-negative Lagrangian multipliers associated with Constraint C9, C10 and C8, respectively, the matrix $\mathbf{J}_k(t)$ above is given by

$$\mathbf{J}_k(t) = \mathbf{I} + \sum_{i \in \mathcal{I}_k} (\lambda_i + \mu_i \tau_{i,k}(t)) \mathbf{B}_{i,k} + v_k \eta_{k,\min} \mathbf{E}_{k,k} + \sum_{l \neq k, l \in \mathcal{U}} \eta_{l,\min} \mathbf{A}_{k,l}. \quad (29)$$

Then, the dual function is given by

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}) \quad (30)$$

$$= \min_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}) \quad (31)$$

$$\begin{aligned} &= \min_{\mathbf{w}} \sum_{k \in \mathcal{U}} \mathbf{w}_k^H \mathbf{J}_k(t) \mathbf{w}_k - \sum_{k \in \mathcal{U}} v_k [\mathbf{w}_k^H(t) \mathbf{A}_{k,k} \mathbf{w}_k + \mathbf{w}_k^H \mathbf{A}_{k,k} \mathbf{w}_k(t)] \\ &\quad - \sum_{i \in \mathcal{I}} \lambda_i P_{i,\max} - \sum_{i \in \mathcal{I}} \mu_i \tilde{C}_i(t) + \sum_{k \in \mathcal{U}} v_k [\eta_{k,\min} \sigma_k^2 + \zeta_k(t)]. \end{aligned} \quad (32)$$

Note that $\mathbf{J}_k(t)$ is a positive definite matrix. Hence, Problem (32) is a strictly convex problem and its unique solution can be obtained from its first-order optimality condition as:

$$\mathbf{w}_k = v_k \mathbf{J}_k^{-1}(t) \mathbf{A}_{k,k} \mathbf{w}_k(t). \quad (33)$$

By substituting the optimal solution of \mathbf{w}_k in (33) into (32), the dual function becomes

$$\begin{aligned} g(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}) &= - \sum_{k \in \mathcal{U}} v_k^2 \mathbf{w}_k^H(t) \mathbf{A}_{k,k} \mathbf{J}_k^{-1}(t) \mathbf{A}_{k,k} \mathbf{w}_k(t) - \sum_{i \in \mathcal{I}} \lambda_i P_{i,\max} - \sum_{i \in \mathcal{I}} \mu_i \tilde{C}_i(t) \\ &\quad + \sum_{k \in \mathcal{U}} v_k [\eta_{k,\min} \sigma_k^2 + \zeta_k(t)]. \end{aligned} \quad (34)$$

Then, the dual of Problem \mathcal{P}_6 is given by

$$\max_{\{\lambda_i \geq 0, \mu_i \geq 0, \nu_k \geq 0, \forall k, i\}} g(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}). \quad (35)$$

The classic gradient descent methods such as the subgradient or ellipsoid methods [32] can be employed to solve the dual problem (35) to update the Lagrangian multipliers.

C. Low-complexity Algorithm

Combining Subsection-IV-A and Subsection IV-B, we conceive an iterative algorithm to solve Problem \mathcal{P}_3 based on the first order Taylor approximation (FOTA) method in Algorithm 1. It is readily seen that the optimal solution obtained at the t th iteration is also feasible for Problem \mathcal{P}_3 at the $(t+1)$ th iteration, since the indicator function is smaller than one and it is approximated as the right hand side of (22). This implies that Algorithm 1 generates a non-increasing sequence of objective function values and finally converges to the Karush-Kuhn-Tucker solution of Problem \mathcal{P}_4 , as proved in [34]. Note that the optimal beamforming solution obtained by Algorithm 1 is guaranteed to be rank one.

In Algorithm 1, it is necessary to find the initial feasible set of beamforming vectors $\mathbf{w}(0)$. In Section V, we provide the UE selection algorithm to find the maximum number of admitted UEs. The corresponding obtained beamforming vectors can be set as the initial point of Algorithm 1. The reason is that the constraints of Problem \mathcal{P}_3 and Problem \mathcal{P}_7 are the same.

Algorithm 1 FOTA-based Algorithm to Solve Problem \mathcal{P}_3

- 1: Initialize iteration number $t = 1$, error tolerance δ , small constant θ , feasible $\mathbf{w}(0)$, calculate $\tau_{i,k}(0)$, $\tilde{C}_i(0)$ and $\zeta_k(0)$, calculate the objective value of Problem \mathcal{P}_6 , denoted as $\text{Obj}(0)$.
 - 2: Solve Problem \mathcal{P}_6 by using the Lagrange dual decomposition method to obtain $\{\mathbf{w}_k(t), \forall k\}$ with $\tau_{i,k}(t-1)$, $\tilde{C}_i(t-1)$ and $\zeta_k(t-1)$;
 - 3: With $\{\mathbf{w}_k(t), \forall k\}$, update $\tau_{i,k}(t)$, $\tilde{C}_i(t)$ and $\zeta_k(t)$;
 - 4: If $|\text{Obj}(t-1) - \text{Obj}(t)|/\text{Obj}(t) < \delta$, terminate. Otherwise, set $t \leftarrow t + 1$, go to step 2.
-

D. Complexity Analysis

In this subsection, we analyze the computational complexity of Algorithm 1. For notational simplicity, we assume that candidate set size for each UE is equal to L , $|\mathcal{I}_k| = L, \forall k \in \mathcal{U}$. Note that in general L is much smaller than the total number of RRHs I .

For Algorithm 1, the main complexity lies in solving Problem \mathcal{P}_6 by using the Lagrange dual decomposition method. In each iteration of the Lagrange dual decomposition method, the complexity is dominated by calculating \mathbf{w}_k in (33). Note that the complexity of calculating \mathbf{w}_k mainly lies in the calculation of $\mathbf{J}_k^{-1}(t)$. According to [32], for a complex matrix $\mathbf{A} \in \mathbb{C}^{N \times N}$, the complexity of calculating \mathbf{A}^{-1} is on the order of $O(N^3)$. Hence, the complexity of calculating

\mathbf{w}_k is on the order of $O(M^3 L^3)$. Since there are a total of K UEs, the total complexity of the Lagrange dual decomposition method in each iteration is on the order of $O(KM^3 L^3)$. Since there are a total of $(2I + K)$ dual variables, the total number of iterations required by the ellipsoid methods is upper-bounded by $O[(2I + K)^2]$ [35]. Hence, the total complexity of the Lagrange dual decomposition method is given by $O[(2I + K)^2 KM^3 L^3]$. Let us denote t_{avg} as the average number of iterations required for Algorithm 1 to converge, then the total complexity of Algorithm 1 imposed by solving Problem \mathcal{P}_1 is expressed as $T_{\mathcal{P}_1} = O[t_{\text{avg}}(2I + K)^2 KM^3 L^3]$. Simulation results show that Algorithm 1 converges fast, typically 10 iterations are sufficient for the algorithm to converge.

V. LOW-COMPLEXITY UE SELECTION ALGORITHMS

In this section we solve the UE selection Problem \mathcal{P}_1 . By substituting $r_k = R_{k,\min}, \forall k$ into the fronthaul capacity constraint C3, we obtain an alternative optimization problem to Problem \mathcal{P}_1 , which is expressed as follows:

$$\begin{aligned} \mathcal{P}_7 : \quad & \max_{\mathbf{w}, \mathcal{U} \subseteq \bar{\mathcal{U}}} |\mathcal{U}| \\ \text{s.t.} \quad & \text{C2, C4, C5,} \end{aligned} \quad (36)$$

where C4 and C5 are given in Problem \mathcal{P}_3 . Although Problem \mathcal{P}_7 is not the same as the original UE selection Problem \mathcal{P}_1 , both problems yield the same optimal set of selected UEs, the proof of which can be found in Appendix D. It should be noted that the optimal beamforming vectors obtained from solving Problem \mathcal{P}_7 may not be feasible for Problem \mathcal{P}_1 . However, the aim of solving Problem \mathcal{P}_7 is twofold. Firstly, one can find the optimal set of selected UEs. Second, one can provide the initial feasible point for solving Problem \mathcal{P}_3 in Stage II since both problems have the same set of constraints.

Inspired by the UE selection method of [36], we construct an alternative to Problem \mathcal{P}_7 by introducing a set of auxiliary variables $\{\varphi_k\}_{k \in \bar{\mathcal{U}}}$:

$$\mathcal{P}_8 : \quad \min_{\{\varphi_k \geq 0\}_{k \in \bar{\mathcal{U}}}, \mathbf{w}} \sum_{k \in \bar{\mathcal{U}}} \varphi_k \quad (37a)$$

$$\text{s.t.} \quad \text{C2, C5,} \quad (37b)$$

$$\begin{aligned} \text{C11 : } & \mathbf{w}_k^H \mathbf{A}_{k,k} \mathbf{w}_k + \varphi_k \geq \\ & \eta_{k,\min} \left(\mathbf{w}_k^H \mathbf{E}_{k,k} \mathbf{w}_k + \sum_{l \neq k, l \in \bar{\mathcal{U}}} \mathbf{w}_l^H \mathbf{A}_{l,k} \mathbf{w}_l + \sigma_k^2 \right), \forall k \in \bar{\mathcal{U}}. \end{aligned} \quad (37c)$$

Let us denote the solution of $\{\varphi_k\}_{k \in \bar{\mathcal{U}}}$ by $\{\varphi_k^*\}_{k \in \bar{\mathcal{U}}}$. It is readily seen that Problem \mathcal{P}_8 is always feasible. If the optimal solutions of $\{\varphi_k^*\}_{k \in \bar{\mathcal{U}}}$ are all equal to zero, then all UEs can be admitted to the network. Otherwise, some UEs should be removed from the system and we reschedule them for the next opportunity. Intuitively, the UE having a largest value of φ_k^* has a higher probability to be removed since it has the largest discrepancy from its rate target. One needs to find the initial beamforming vectors for Problem \mathcal{P}_8 , which is detailed as follows. Please note that we can ignore Constraint C11 when solving Problem \mathcal{P}_8 , because for any given beamforming solutions, we can always find feasible φ_k values that satisfy Constraint C11. As a result, we only have to find feasible beamforming vectors that simultaneously satisfy Constraints C2 and C5. Note that $R_{k,\min}$ in Constraint C5 in (20c) is a constant, hence we only have to search the set of users that can be served by each RRH. This can be readily achieved by randomly selecting the subset of users that satisfy Constraint C5. When the set of users served by each RRH is given, the power at each RRH can be equally shared amongst these users. Then, Problem \mathcal{P}_8 can be solved similarly to Problem \mathcal{P}_3 of the above section, hence the details of which are omitted.

There are two low-complexity UE deletion methods. One is the successive UE deletion method that is provided in [10], [36]. The main idea is to remove the UE having the largest φ_k^* each time, until all the remaining optimal values of φ_k^* become equal to zero. The complexity of this algorithm increases linearly with the number of UEs, hence it is on the order of $O(K)$. This algorithm is suitable for medium-sized networks. The other technique is the bisection based search method proposed in [16]. The main idea is to sort $\{\varphi_k^*\}_{k \in \bar{\mathcal{U}}}$ in descending order $\varphi_{\pi_1}^* \geq \dots \geq \varphi_{\pi_K}^*$. Then, one should find a minimum L_0 for ensuring that all the UEs in $\mathcal{U} = \{\pi_{L_0+1}, \dots, \pi_K\}$ can be supported with $L_0 = 1, \dots, K-1$. The bisection search method is used to iteratively find the optimal L_0 by updating its upper-bound and lower-bound. The complexity of the bisection based method is on the order of $\lceil \log_2(1+K) \rceil$, which is suitable for very dense networks supporting a large number of UEs. The details of these two algorithms are not shown here for simplicity.

It should be emphasized that when using the iterative algorithm in the above section to solve Problem \mathcal{P}_8 , the iterative procedure will terminate once the intermediate solutions of $\{\varphi_k\}_{k \in \mathcal{U}}$ are all equal to zero. Hence, the data rates of some UEs with the obtained beamforming solution are strictly larger than their minimum rate requirements.

VI. SIMULATION RESULTS

In this section, we provide simulation results to evaluate the performance of the proposed robust algorithms. Two types of UD-CRAN networks are considered: a small UD-CRAN deployed in a square area of $[400 \text{ m} \times 400 \text{ m}]$ and a larger one of $[700 \text{ m} \times 700 \text{ m}]$. Both the UEs and RRHs are uniformly distributed in these areas. For the small one, the numbers of RRHs and UEs are set to $I = 14$ and $K = 8$ with the densities of 87.5 RRHs/km^2 and 50 UEs/km^2 , respectively. For the large one, the numbers are set to $I = 42$ and $K = 24$ with the densities of 85.7 RRHs/km^2 and 49 UEs/km^2 , respectively. These two scenarios comply with the ultra-dense networks in the fifth-generation (5G) wireless system [37], where the density of BSs will be up to $40\text{-}50 \text{ BSs/km}^2$. The channels are generated according to the LTE specifications [38], which are composed of three elements: 1) the large-scale path loss given by $PL = 148.1 + 37.6 \log_{10} d$ (dB), where d is the distance between a RRH and a UE in km; 2) the log-normal shadowing fading having a zero mean and 8 dB standard deviation; 3) small-scale Rayleigh fading with zero mean and unit variance. For ease of exposition, all UEs are assumed to have the same rate constraints of $R_{\min} = R_{k,\min}, \forall k$, and all RRHs have the same power constraints of $P_{\max} = P_{i,\max}, \forall i$. Furthermore, the fronthaul capacity constraints are assumed to be the same for all RRHs, i.e., $C_{\max} = C_{i,\max}, \forall i$, and we consider the normalized fronthaul capacity constraints (with respect to each UE's rate target), i.e., $\tilde{C}_{\max} = C_{\max}/R_{\min}$. Note that \tilde{C}_{\max} can be interpreted as the maximum number of UEs that can be supported by each fronthaul link. For simplicity, each UE is assumed to choose its nearest L RRHs as its serving candidate set, i.e., $|\mathcal{I}_k| = L, \forall k$. The maximum pilot reuse times for small and large UD-CRANs are $n_{\max} = 2$ and $n_{\max} = 3$, respectively. The total number of time slots in each time frame is $T = 200$, the numbers of CDI and PA quantization bits for each RRH are set as $B^{\text{CDI}} = 4$ and $B^{\text{PA}} = 2$, respectively. Unless otherwise stated, the simulation parameters are given in Tabel II and the following results are obtained by averaging over 100 channel generations.

A. Smaller UD-CRAN

In this section, we evaluate the performance of our algorithm in the small C-RAN network, where the simulation results in Fig. 2-Fig. 9 are based on this scenario.

We first study the impact of the initial points on the convergence behaviour of the FOTA-based Algorithm to solve Problem \mathcal{P}_8 . Fig. 2 shows the objective value of Problem \mathcal{P}_8 versus the number

TABLE II
MAIN SIMULATION PARAMETERS

Parameters	Value	Parameters	Value
Number of antennas M	2	System bandwidth B	20 MHz
Noise power density	-174 dBm/Hz [38]	Error tolerance δ	10^{-5}
Small constant θ	10^{-5}	Pilot power p_t	200 mW
Maximum transmit power P_{\max}	100 mW	Rate target R_{\min}	3 bit/s/Hz
Candidate size L	3	Normalized fronthaul limits \tilde{C}_{\max}	3

of iterations for one randomly generated set of channel realizations for two cases of $R_{\min} = 1$ bit/s/Hz and $R_{\min} = 2$ bit/s/Hz. Since Problem \mathcal{P}_8 is a non-convex problem, different initial points may lead to different solutions. To investigate this effect, we consider two initialization schemes: 1) Rand-initial: In this scheme, both the power allocation and beamforming direction on each beam is randomly generated; 2) CM-initial: For this scheme, the total power on each RRH is equally split among its served UEs and the beamforming direction is set to be the same as its channel direction. It can be observed from Fig. 2 that the algorithm with different initial points will have different convergence speeds, but converge to the same objective value. It is difficult to justify which initialization scheme has faster convergence speed as seen in Fig. 2. For both schemes, five iterations are sufficient for the algorithm to converge. When $R_{\min} = 1$ bit/s/Hz, the algorithm converges to zero, which means all the UEs can be admitted. However, when $R_{\min} = 2$ bit/s/Hz, the algorithm converges to a positive value, which implies that some UEs should be deleted.

Next, we study the convergence behaviour of the proposed two UE selection algorithms. Specifically, Fig. 3 illustrates the number of UEs to be checked versus the number of times to solve Problem \mathcal{P}_8 for a randomly generated network, where the successive UE deletion method and the bisection search method are labeled as ‘Suc’ and ‘Bis’, respectively. It can be found from Fig. 3 that the number of UEs to be checked for the ‘Suc’ algorithm always decreases with the number of times Problem \mathcal{P}_8 is solved, while that of the ‘Bis’ algorithm fluctuates during the procedure. These observations are consistent with the features of these two algorithms. Interestingly, for both rate targets, the numbers of times by the ‘Bis’ algorithm are fixed to five. However, the number of times Problem \mathcal{P}_8 is solved by the ‘Suc’ algorithm depends on

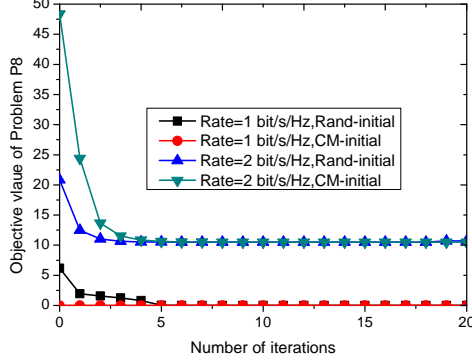


Fig. 2. Convergence behaviour of the FOTA-based Algorithm to solve Problem \mathcal{P}_8 under different initial points.

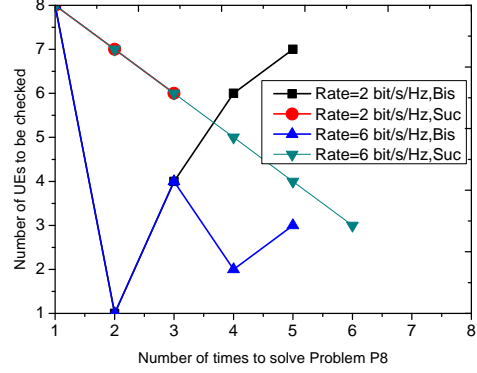


Fig. 3. Convergence behaviour of the proposed two UE selection algorithms.

the rate targets. For the example in Fig. 3, the ‘Suc’ algorithm only needs three times when $R_{\min} = 2$ bit/s/Hz, while six times when $R_{\min} = 6$ bit/s/Hz.

In Fig. 4, we plot the number of UEs admitted by the various algorithms versus the total number of UEs to be checked. To guarantee fairness, the length of the square area is set to guarantee the same UE density, which yields $[283\text{ m}, 346\text{ m}, 400\text{ m}, 447\text{ m}, 490\text{ m}]$ for each corresponding total number of UEs to be checked in Fig. 5. Then, the number of RRHs is set to $[7, 10, 14, 17, 21]$ for each case to ensure having almost the same RRH density. The exhaustive UE search algorithm (labeled as ‘Exhaustive search’) is used as a performance benchmark, which checks all subsets of UEs and chooses the one having the largest number of admitted UEs. For each given user-RRH association set, we adopt Equ. (24) to tackle each optimization problem. It should be emphasized that the exhaustive search method cannot be guaranteed to achieve the globally optimal solution due to the non-convexity of Constraint C11. The exhaustive search method is employed to show the effectiveness of approximating the indicator function in Constraint C5 as the smooth fractional function. The same holds for the exhaustive search method adopted to solve Problem \mathcal{P}_3 in Fig. 7. Note that the computational complexity of the exhaustive search is on the order of $O(2^K)$. As expected, the number of UEs admitted by all the algorithms is reduced upon increasing the total number of UEs to be checked. The exhaustive search method performs better than the other two algorithms, which comes at the expense of a high computational complexity. However, its performance gain over the successive UE deletion method is marginal, especially when the total number of UEs is small. On the other hand, the

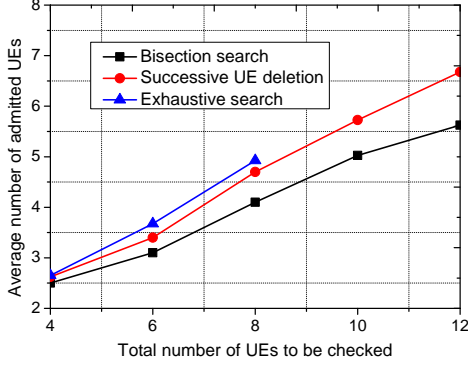


Fig. 4. Number of UEs admitted by various algorithms versus the total number of UEs to be checked.

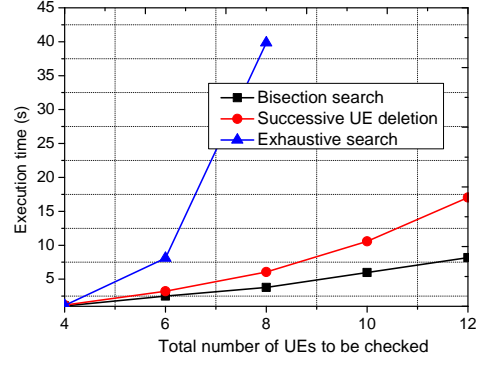


Fig. 5. Execution time for various UE selection algorithms.

successive UE deletion method significantly outperforms the bisection search method, and the performance gain increases with the total number of UEs.

Fig. 5 compares the execution time for various UE selection algorithms versus the total number of UEs to be checked by using an i7-7700 CPU operating at 3.6GHz for the same setup as in Fig. 4. As expected, this figure shows that the execution time of all algorithms increases with the system size, and the exhaustive search algorithm imposes a significantly higher execution time than both the proposed UE selection algorithms. The operation time required by the successive UE deletion algorithm is higher than that of the bisection search algorithm, and the gap increases with the system size. This means that the successive UE deletion is a good option for moderate-sized UD-CRANs, while the bisection based search method is more suitable for larger UD-CRAN, as a benefit of its lowest complexity.

Fig. 6 shows the convergence behaviour of the FOTA-based Algorithm under different rate targets, where the bisection search algorithm is employed for selecting the admitted UEs. The average numbers of admitted UEs for different rate targets are shown in this figure. It is seen from this figure that our proposed algorithm converges rapidly and generally three iterations are sufficient for the algorithm to converge under all considered rate targets, which is appealing for practical applications. Since the number of UEs admitted for the larger R_{\min} is smaller, the larger R_{\min} may not yield higher transmit power.

In Fig. 7, we compare the performance of the FOTA-based Algorithm with that of the exhaustive search method. For the latter algorithm, if $|\mathcal{U}_i| \geq \tilde{C}_{\max}$, the algorithm checks all

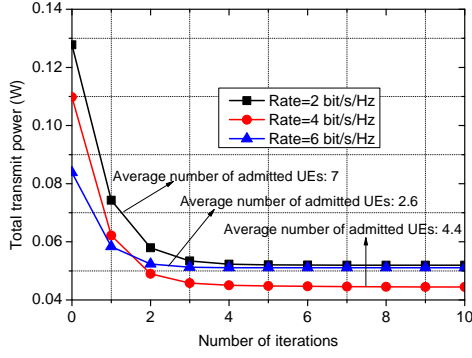


Fig. 6. Convergence behaviour of the FOTA-based Algorithm to solve Problem \mathcal{P}_3 under different rate targets.

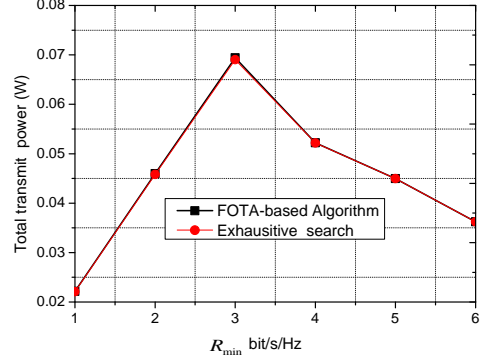


Fig. 7. Total transmit power versus the rate targets for the proposed algorithm and the exhaustive search method.

possible subsets of \mathcal{U}_i with size \tilde{C}_{\max} , and chooses the one with the minimum transmit power. It is observed from Fig. 7 that our proposed algorithm achieves almost the same performance as that of the exhaustive search method, which confirms the effectiveness of our proposed algorithm. The corresponding execution time for these two algorithms is shown in Fig. 8. We can observe from Fig. 8 that the execution time of the exhaustive search method requires much more time than the proposed FOTA-based Algorithm for the small R_{\min} , and almost the same for large R_{\min} . The reason can be explained as follows. For the case of small R_{\min} , more UEs can be admitted in the network, so that more RRHs will satisfy the condition $|\mathcal{U}_i| \geq \tilde{C}_{\max}$. Then the number of checking times is large, which leads to high computational complexity. However, for the case of large R_{\min} , only a small number of UEs can be admitted as seen in Fig. 4. Then, almost all the RRHs satisfy the fronthaul capacity constraint, and it is not necessary for the exhaustive search method to enumerate the UE-RRH associations, leading to almost the same complexity of our algorithm. Note that the time required by the proposed algorithm is within one second and the algorithm converges within five iterations as seen in Fig. 6, then the execution time for each iteration of the FOTA-based Algorithm is within 0.2 second.

Now, we study the robustness of the proposed algorithm against the following four algorithms:

- 1) Only Robust to Channel Quantization (labeled as ‘Robust CH Quan.’): This method only takes into account the effect of channel quantization, when designing the beamforming vectors, regardless of the channel estimation errors.
- 2) Only Robust to Channel Estimation Error (labeled as ‘Robust CH Esti.’): As the termi-

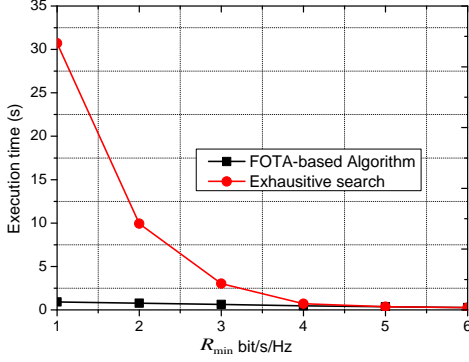


Fig. 8. Execution time for the proposed FOTA-based Algorithm and the exhaustive search method.

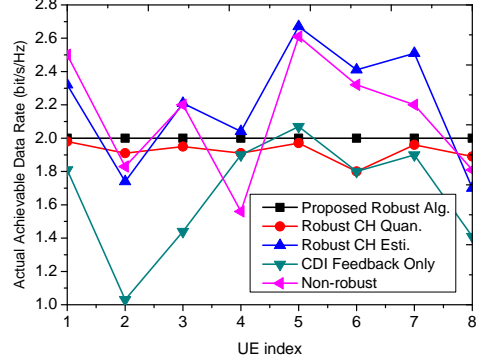


Fig. 9. The achievable data rate for various algorithms, where the rate target is set as $R_{\min} = 2$ bit/s/Hz.

TABLE III
POWER CONSUMPTION AND NUMBERS OF ADMITTED UES FOR VARIOUS METHODS

	Proposed Robust	Robust CH Quan.	Robust CH Esti.	CDI FB Only	Non-robust
Power consumption (mW)	167	130	158	149	114

nology suggests, this method only considers the effects of channel estimation errors, and naively treats the feedback CDI and PA as perfect. Then, the SDP method proposed in [23] can be adopted to solve the resultant optimization problem.

- 3) Only Feeding back the CDI Information (labeled as ‘CDI FB Only’): In this method, each UE only feeds back the CDI index to the BBU pool, without considering the PA information. The \mathbf{A} matrix derived in Appendix A in [31] can be recalculated without considering the PA quantization information and the statistics of the quantization error.
- 4) Nonrobust Beamforming Design (labeled as ‘Non-robust’): Neither channel quantization errors nor channel estimation errors are considered by this algorithm and the feedback CDI and PA are regarded as perfect.

Table III reports the total power consumption required by the various methods for one random channel generation, where all eight UEs are admitted. It can be seen that our proposed algorithm has the highest power consumption, since it requires more power to compensate for both the channel estimation errors and channel quantization errors. Note that the non-robust algorithm requires the least since these errors are not considered. However, it is important to observe each UE’s actual achievable data rate achieved by these algorithms. Fig. 9 shows each UE’s actual

achievable data rate by all the methods. It is seen that all UEs' data requirements are satisfied by our proposed robust algorithm, which confirms the effectiveness of our proposed algorithm. For the 'Robust CH Quan.' method, all UEs' rate requirements are not fulfilled since the channel estimation errors are not considered. Hence, the channel estimation error cannot be ignored when designing the beamforming vector due to the non-negligible pilot contamination. For 'Robust CH Esti.' method, the statistics information of channel quantization error is not considered and some UEs' actual achievable data rates are lower than the rate target, such as those of UE 2 and UE 8. It is also observed that some UEs have much higher rates, indicating that the power and spatial resources are not properly allocated by the 'Robust CH Esti.' method. For the 'CDI FB Only' method, the actual achievable data rates of all UEs are lower than the rate targets, and UE 2's data rate is even lower than 1 bit/s/Hz. This confirms the importance of feeding back the PA information for coherent transmission. Finally, some UEs' actual achievable data rates are below the rate target by the 'Non-robust' method as it naively treats the feedback CSI as the perfect. However, it is observed in Fig. 5 that, even with non-perfect PA feedback information, the performance of the 'Non-robust' method is even better than that of the 'CDI FB Only' method in terms of the number of UEs that satisfy the rate target. In summary, only our proposed algorithm is capable of maintaining the guaranteed rates for each UE, since it jointly considers the effects of channel estimation errors and channel quantization errors, which are (partially) ignored by the other algorithms.

B. Larger UD-CRANs

The following simulation results are based on the larger UD-CRAN. We investigate the effects of different system parameters on the performance of the proposed algorithm.

Since UD-CRANs will be deployed in hot spots, where the number of UEs is high and the communication resources are limited, maximizing the number of admitted users for each time frame should be a high priority. Additionally, according to the results of Fig. 6, the power consumption may not provide sufficient insights, since its value mainly depends on the number of UEs selected from Stage I. Hence, in the following, we only consider the performance in terms of the number of UEs that can be supported. For comparison, the performance of the algorithm having perfect intra-cluster CSI [16] is also simulated as a performance benchmark. The bisection based search method and the successive UE deletion method of the robust algorithm

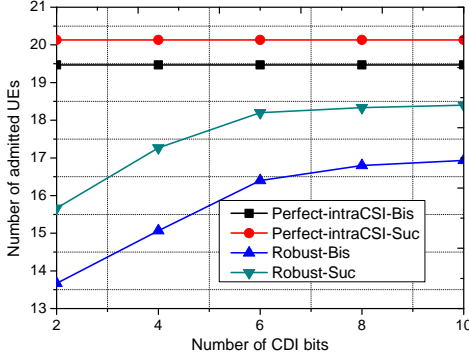


Fig. 10. Number of UEs admitted by various algorithms versus CDI quantization bits B^{CDI} for a large UD-CRAN.

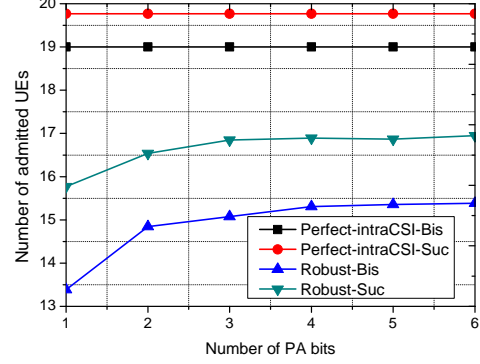


Fig. 11. Number of UEs admitted by various algorithms versus PA quantization bits B^{PA} for a large UD-CRAN.

are denoted as ‘Robust-Bis’ and ‘Robust-Suc’, respectively, while ‘Perfect-intraCSI-Bis’ and ‘Perfect-intraCSI-Suc’ represent the two methods for the case of perfect intra-cluster CSI.

1) *Impact of the number of CDI quantization bits:* Fig. 10 illustrates the impact of CDI quantization bits B^{CDI} on the system performance. Note that when B^{CDI} increases from 2 to 6, three more UEs can be admitted by the proposed robust algorithms and will not increase for $B^{\text{CDI}} \geq 6$. This is due to the fact that each RRH is equipped with two antennas and a small number of CDI quantization bits are sufficient to achieve good performance. A fixed performance gap is observed between the robust algorithm and those for perfect intra-cluster CSI when $B^{\text{CDI}} \geq 6$ due to the additional channel estimation error.

2) *Impact of the number of PA quantization bits:* Now, we study the impact of the important system parameter B^{PA} in Fig. 11. It is seen from this figure that there is a slight increase of the number of UEs admitted by the robust algorithms when B^{PA} increases from 1 to 3 and becomes saturated, when $B^{\text{PA}} \geq 3$. This is a very inspiring result, implying that only a small number of bits is necessary for the PA quantization, which mitigates the feedback overhead, while guaranteeing good performance. In particular, even one bit used for PA quantization can achieve 90% of the performance attained with perfect PA information.

VII. CONCLUSIONS

This paper provided a complete framework for dealing with the unavailability of full CSI in user-centric UD-CRANs, where only partial inter-cluster CSI and a quantized version of the intra-cluster CSI are available at the BBU pool. We derived the achievable data rate expression by

exploiting the statistical characteristics of various channel uncertainties. Based on this, we developed a low-complexity robust beamforming algorithm for minimizing the total transmit power, while guaranteeing each user's rate requirement and fronthaul capacity constraints. In addition, to ensure the feasibility of the problem, a pair of low-complexity user selection algorithms are provided as well. Simulation results show that our proposed robust algorithm significantly outperforms the existing state-of-art algorithms in terms of providing the required guaranteed quality-of-service (QoS) for the users. Furthermore, extensive simulations results are provided to study the impact of different system parameters on the performance. One new important observation was made: One bit for quantizing the each RRH's PA is enough to obtain a large proportion of the performance obtained with perfect PA information.

APPENDIX A

DERIVATIONS OF $\mathbf{A}_{k,k}$ AND $\mathbf{A}_{l,k}$

A. Derivation of $\mathbf{A}_{k,k}$

We first derive the expression of $\mathbf{A}_{k,k}$, which is equal to $\mathbf{A}_{k,k} = \mathbb{E} \{ \hat{\mathbf{g}}_{k,k} \hat{\mathbf{g}}_{k,k}^H \}$. Denote indices of \mathcal{I}_k as $\mathcal{I}_k = \{s_1^k, \dots, s_{|\mathcal{I}_k|}^k\}$. To calculate $\mathbf{A}_{k,k}$, we first have to calculate $\mathbb{E} \{ \hat{\mathbf{h}}_{s_i^k,k} \hat{\mathbf{h}}_{s_i^k,k}^H \}$ and $\mathbb{E} \{ \hat{\mathbf{h}}_{s_i^k,k} \hat{\mathbf{h}}_{s_l^k,k}^H \}$, where $i \neq l$.

The derivation of $\mathbb{E} \{ \hat{\mathbf{h}}_{s_i^k,k} \hat{\mathbf{h}}_{s_i^k,k}^H \}$ is based on the following lemma.

Lemma 1: The random vectors and variables have the following properties:

$$\mathbb{E} \{ \mathbf{q}_{s_i^k,k} \mathbf{u}_{s_i^k,k}^H \} = \mathbf{0}, \quad (\text{A.1})$$

$$\mathbb{E} \{ a_{s_i^k,k} \} = 2^{B_{s_i^k,k}^{\text{CDI}}} \beta \left(2^{B_{s_i^k,k}^{\text{CDI}}}, \frac{M}{M-1} \right) \triangleq \rho_{s_i^k,k}, \quad (\text{A.2})$$

$$\mathbb{E} \{ \mathbf{u}_{i,k} \mathbf{u}_{i,k}^H \} = \frac{1}{M-1} \left(\mathbf{I}_M - \mathbf{q}_{s_i^k,k} \mathbf{q}_{s_i^k,k}^H \right) \triangleq \mathbf{O}_{s_i^k,k}, \quad (\text{A.3})$$

where $\beta(a, b)$ is the beta function.

Proof: (A.1) follows since $\mathbf{q}_{s_i^k,k}$ is orthogonal to $\mathbf{u}_{s_i^k,k}$, while (A.2) and (A.3) are from [39] and [40], respectively. \square

Based on Lemma 1, we embark on deriving $\mathbb{E} \{ \hat{\mathbf{h}}_{s_i^k,k} \hat{\mathbf{h}}_{s_i^k,k}^H \}$. According to Section-II-B and Section-II-C, the quantized channel $\hat{\mathbf{h}}_{s_i^k,k}$ can be rewritten as

$$\hat{\mathbf{h}}_{s_i^k,k} = \left\| \hat{\mathbf{h}}_{s_i^k,k} \right\| \left(\sqrt{1 - a_{s_i^k,k}} e^{j\phi_{s_i^k,k}} \mathbf{q}_{s_i^k,k} + \sqrt{a_{s_i^k,k}} \mathbf{u}_{s_i^k,k} \right). \quad (\text{A.4})$$

Then $\mathbb{E} \left\{ \hat{\mathbf{h}}_{s_i^k, k} \hat{\mathbf{h}}_{s_i^k, k}^H \right\}$ can be derived as

$$\begin{aligned} & \mathbb{E} \left\{ \hat{\mathbf{h}}_{s_i^k, k} \hat{\mathbf{h}}_{s_i^k, k}^H \right\} \\ &= \mathbb{E} \left\{ \left\| \hat{\mathbf{h}}_{s_i^k, k} \right\|^2 \right\} \left(\mathbb{E} \left\{ \left(1 - a_{s_i^k, k} \right) \right\} \mathbf{q}_{s_i^k, k} \mathbf{q}_{s_i^k, k}^H + \mathbb{E} \left\{ a_{s_i^k, k} \right\} \mathbb{E} \left\{ \mathbf{u}_{s_i^k, k} \mathbf{u}_{s_i^k, k}^H \right\} \right) \end{aligned} \quad (\text{A.5})$$

$$= \omega_{s_i^k, k} M \left(\left(1 - \rho_{s_i^k, k} \right) \mathbf{q}_{s_i^k, k} \mathbf{q}_{s_i^k, k}^H + \rho_{s_i^k, k} \mathbf{O}_{s_i^k, k} \right), \quad (\text{A.6})$$

where (A.1) is used in (A.5), (A.2) and (A.3) are used in (A.6). Note that the PA quantization does not affect the value of $\mathbb{E} \left\{ \hat{\mathbf{h}}_{s_i^k, k} \hat{\mathbf{h}}_{s_i^k, k}^H \right\}$.

To calculate $\mathbb{E} \left\{ \hat{\mathbf{h}}_{s_i^k, k} \hat{\mathbf{h}}_{s_l^k, k}^H \right\}$ with $i \neq l$, the following lemma should be employed.

Lemma 2: The random vectors and variables in (A.4) have the following properties:

$$\mathbb{E} \left\{ e^{j\phi_{s_i^k, k}} \right\} = e^{j\hat{\phi}_{s_i^k, k}} \frac{2^{B_{s_i^k, k}^{\text{PA}}}}{\pi} \sin \left(\frac{\pi}{2^{B_{s_i^k, k}^{\text{PA}}}} \right) \triangleq e^{j\hat{\phi}_{s_i^k, k}} \xi_{s_i^k, k} \quad (\text{A.7})$$

$$\mathbb{E} \left\{ \left\| \hat{\mathbf{h}}_{s_i^k, k} \right\| \right\} = \frac{\sqrt{\omega_{s_i^k, k}} \Gamma(M + \frac{1}{2})}{\Gamma(M)} \triangleq \varsigma_{s_i^k, k} \quad (\text{A.8})$$

$$\mathbb{E} \left\{ \sqrt{1 - a_{s_i^k, k}} \right\} = \sum_{m=1}^{2^{B_{s_i^k, k}^{\text{CDI}}}} C_{2^{B_{s_i^k, k}^{\text{CDI}}}}^m (-1)^{m+1} m(M-1) \beta \left(m(M-1), \frac{3}{2} \right) \triangleq \Omega_{s_i^k, k}, \quad (\text{A.9})$$

where $C_n^r = \frac{n!}{r!(n-r)!}$.

Proof: We first prove equality (A.7). Specifically, we have

$$\begin{aligned} \mathbb{E} \left\{ e^{j\phi_{s_i^k, k}} \right\} &= e^{j\hat{\phi}_{s_i^k, k}} \mathbb{E} \left\{ e^{j\tilde{\phi}_{s_i^k, k}} \right\} = e^{j\hat{\phi}_{s_i^k, k}} \int_{-\frac{\pi}{2^{B_{s_i^k, k}^{\text{PA}}}}}^{\frac{\pi}{2^{B_{s_i^k, k}^{\text{PA}}}}} e^{jx} \frac{2^{B_{s_i^k, k}^{\text{PA}}}}{2\pi} dx \\ &= e^{j\hat{\phi}_{s_i^k, k}} \int_{-\frac{\pi}{2^{B_{s_i^k, k}^{\text{PA}}}}}^{\frac{\pi}{2^{B_{s_i^k, k}^{\text{PA}}}}} (\cos x + j \sin x) \frac{2^{B_{s_i^k, k}^{\text{PA}}}}{2\pi} dx \\ &= e^{j\hat{\phi}_{s_i^k, k}} \frac{2^{B_{s_i^k, k}^{\text{PA}}}}{\pi} \sin \left(\frac{\pi}{2^{B_{s_i^k, k}^{\text{PA}}}} \right). \end{aligned}$$

For (A.8), let us define $\hat{\mathbf{h}}_{s_i^k, k} \triangleq \hat{\mathbf{h}}_{s_i^k, k} / \sqrt{\omega_{s_i^k, k}}$. Then, $\left\| \hat{\mathbf{h}}_{s_i^k, k} \right\|$ obeys a scaled (by a factor of $1/\sqrt{2}$) chi distribution with $2M$ degrees of freedom. Hence, $\mathbb{E} \left\{ \left\| \hat{\mathbf{h}}_{s_i^k, k} \right\| \right\} = \frac{\Gamma(M + \frac{1}{2})}{\Gamma(M)}$ [29] and (A.8) is proved.

Finally, we embark on proving (A.9). Define $\nu \triangleq 1 - a_{s_i^k, k}$. According to Lemma 1 in [39], the probability density function of ν with M antennas using a 2^B RVQ codebook is given by

$$f_\nu(\nu) = \sum_{i=1}^{2^B} C_{2^B}^i (-1)^{i+1} i (M-1) (1-\nu)^{i(M-1)-1}. \quad (\text{A.10})$$

Define $x \triangleq \sqrt{\nu} \in [0, 1]$, then the Jacobian of the transformation is given by $J = \frac{dx}{d\nu} = \frac{1}{2} \nu^{-\frac{1}{2}}$.

Hence, the PDF of x is given by

$$f_x(x) = 2 \sum_{i=1}^{2^B} C_{2^B}^i (-1)^{i+1} i (M-1) (1-x^2)^{i(M-1)-1} x. \quad (\text{A.11})$$

Then the expectation of x is given by

$$\begin{aligned} \mathbb{E}\{x\} &= \int_0^1 x f_x(x) dx \\ &= 2 \sum_{m=1}^{2^B} C_{2^B}^m (-1)^{m+1} m (M-1) \int_0^1 (1-x^2)^{m(M-1)-1} x^2 dx \\ &= \sum_{m=1}^{2^B} C_{2^B}^m (-1)^{m+1} m (M-1) \beta \left(m(M-1), \frac{3}{2} \right) \end{aligned}$$

where the last equality is due to the fact that the beta function can be represented as [28], [41]

$$\beta \left(c, \frac{a}{b} \right) = b \int_0^1 x^{a-1} (1-x^b)^{c-1} dx \quad (\text{A.12})$$

with $a = 3$, $b = 2$ and $c = m(M-1)$. Hence, (A.9) follows with $B = B_{s_i^k, k}^{\text{CDI}}$. \square

Based on Lemma 2, $\mathbb{E} \left\{ \hat{\mathbf{h}}_{s_i^k, k} \hat{\mathbf{h}}_{s_l^k, k}^H \right\}$, $i \neq l$, can be calculated as

$$\mathbb{E} \left\{ \hat{\mathbf{h}}_{s_i^k, k} \hat{\mathbf{h}}_{s_l^k, k}^H \right\} = \mathbb{E} \left\{ \hat{\mathbf{h}}_{s_i^k, k} \right\} \mathbb{E} \left\{ \hat{\mathbf{h}}_{s_l^k, k}^H \right\} \quad (\text{A.13})$$

$$= \varsigma_{s_i^k, k} \varsigma_{s_l^k, k} \Omega_{s_i^k, k} \Omega_{s_l^k, k} \xi_{s_i^k, k} \xi_{s_l^k, k}^j e^{j \left(\hat{\phi}_{s_i^k, k} - \hat{\phi}_{s_l^k, k} \right)} \mathbf{q}_{s_i^k, k} \mathbf{q}_{s_l^k, k}^H \quad (\text{A.14})$$

where (A.13) follows since $\mathbf{h}_{s_i^k, k}$ is independent of $\mathbf{h}_{s_l^k, k}$, (A.14) follows since $\mathbb{E} \left\{ \mathbf{u}_{s_i^k, k} \right\} = \mathbf{0}_M$, and Lemma 2 is used. \square

Based on the above results, $\mathbf{A}_{k, k}$ is given by

$$\mathbf{A}_{k, k} = \begin{bmatrix} (\mathbf{A}_{k, k})_{1,1} & \cdots & (\mathbf{A}_{k, k})_{1, |\mathcal{I}_k|} \\ \vdots & \ddots & \vdots \\ (\mathbf{A}_{k, k})_{|\mathcal{I}_k|, 1} & \cdots & (\mathbf{A}_{k, k})_{|\mathcal{I}_k|, |\mathcal{I}_k|} \end{bmatrix}, \quad (\text{A.15})$$

where $(\mathbf{A}_{k, k})_{i, l} \in \mathbb{C}^{M \times M}$, $i, l \in 1, \dots, |\mathcal{I}_k|$ is the block matrix of $\mathbf{A}_{k, k}$ at the i th row and l th column. If $i = l$, then $(\mathbf{A}_{k, k})_{i, l} = \mathbb{E} \left\{ \hat{\mathbf{h}}_{s_i^k, k} \hat{\mathbf{h}}_{s_i^k, k}^H \right\}$. Otherwise, $(\mathbf{A}_{k, k})_{i, l} = \mathbb{E} \left\{ \hat{\mathbf{h}}_{s_i^k, k} \hat{\mathbf{h}}_{s_l^k, k}^H \right\}$.

B. Derivation of $\mathbf{A}_{l,k}$

Let us denote the indices of \mathcal{I}_l by $\mathcal{I}_l = \{s_1^l, \dots, s_{|\mathcal{I}_l|}^l\}$. Then $\mathbf{A}_{l,k}$ can be represented as

$$\mathbf{A}_{l,k} = \begin{bmatrix} (\mathbf{A}_{l,k})_{1,1} & \cdots & (\mathbf{A}_{l,k})_{1,|\mathcal{I}_l|} \\ \vdots & \ddots & \vdots \\ (\mathbf{A}_{l,k})_{|\mathcal{I}_l|,1} & \cdots & (\mathbf{A}_{l,k})_{|\mathcal{I}_l|,|\mathcal{I}_l|} \end{bmatrix}, \quad (\text{A.16})$$

where $(\mathbf{A}_{l,k})_{i,j} = \mathbb{E} \left\{ \mathbf{h}_{s_i^l,k} \mathbf{h}_{s_j^l,k}^H \right\}$. To derive $\mathbb{E} \left\{ \mathbf{h}_{s_i^l,k} \mathbf{h}_{s_j^l,k}^H \right\}$, we should discuss four cases: 1) $s_i^l, s_j^l \in \mathcal{I}_k, i \neq j$; 2) $s_i^l, s_j^l \in \mathcal{I}_k, i = j$; 3) $s_i^l, s_j^l \notin \mathcal{I}_k, i = j$; 4) $s_i^l \notin \mathcal{I}_k$ or $s_j^l \notin \mathcal{I}_k, i \neq j$.

For Case 1), both RRH s_i^l and RRH s_j^l belong to UE k 's cluster \mathcal{I}_k , but they are not the same RRH. Then, $\mathbf{h}_{s_i^l,k}$ can be rewritten as

$$\mathbf{h}_{s_i^l,k} = \left\| \hat{\mathbf{h}}_{s_i^l,k} \right\| \left(\sqrt{1 - a_{s_i^l,k}^2} e^{j\phi_{s_i^l,k}} \mathbf{q}_{s_i^l,k} + \sqrt{a_{s_i^l,k}^2} \mathbf{u}_{s_i^l,k} \right) + \mathbf{e}_{s_i^l,k}. \quad (\text{A.17})$$

By exploiting the fact that $\mathbb{E} \left\{ \mathbf{u}_{s_i^l,k} \right\} = \mathbf{0}_M$, $\mathbb{E} \left\{ \mathbf{e}_{s_i^l,k} \right\} = \mathbf{0}_M$ and Lemma 2, $\mathbb{E} \left\{ \mathbf{h}_{s_i^l,k} \mathbf{h}_{s_j^l,k}^H \right\}$ can be derived similarly to $\mathbb{E} \left\{ \hat{\mathbf{h}}_{i,k} \hat{\mathbf{h}}_{l,k}^H \right\}$ in (A.14), which is given by

$$\mathbb{E} \left\{ \mathbf{h}_{s_i^l,k} \mathbf{h}_{s_j^l,k}^H \right\} = \varsigma_{s_i^l,k} \varsigma_{s_j^l,k} \Omega_{s_i^l,k} \Omega_{s_j^l,k} \xi_{s_i^l,k} \xi_{s_j^l,k} e^{j(\hat{\phi}_{s_i^l,k} - \hat{\phi}_{s_j^l,k})} \mathbf{q}_{s_i^l,k} \mathbf{q}_{s_j^l,k}^H. \quad (\text{A.18})$$

For Case 2), RRH s_i^l and RRH s_j^l represent the same RRH belonging to \mathcal{I}_k . By using (A.17) and the facts that $\mathbb{E} \left\{ \mathbf{q}_{s_i^l,k} \mathbf{e}_{s_i^l,k}^H \right\} = \mathbf{0}$, $\mathbb{E} \left\{ \mathbf{u}_{s_i^l,k} \mathbf{e}_{s_i^l,k}^H \right\} = \mathbf{0}$ and Lemma 1, $\mathbb{E} \left\{ \mathbf{h}_{s_i^l,k} \mathbf{h}_{s_j^l,k}^H \right\}$ can be derived similarly to $\mathbb{E} \left\{ \hat{\mathbf{h}}_{i,k} \hat{\mathbf{h}}_{l,k}^H \right\}$ in (A.6), which is given by

$$\mathbb{E} \left\{ \mathbf{h}_{s_i^l,k} \mathbf{h}_{s_j^l,k}^H \right\} = \omega_{s_i^l,k} M \left[\left(1 - \rho_{s_i^l,k} \right) \mathbf{q}_{s_i^l,k} \mathbf{q}_{s_i^l,k}^H + \rho_{s_i^l,k} \mathbf{O}_{s_i^l,k} \right] + \delta_{s_i^l,k} \mathbf{I}_M, \quad (\text{A.19})$$

where $\delta_{s_i^l,k}$ is the channel estimation error given in (5).

For Case 3), both RRHs represent the same RRH that is not in UE k 's cluster \mathcal{I}_k . Then, $\mathbb{E} \left\{ \mathbf{h}_{s_i^l,k} \mathbf{h}_{s_j^l,k}^H \right\}$ is given by $\mathbb{E} \left\{ \mathbf{h}_{s_i^l,k} \mathbf{h}_{s_j^l,k}^H \right\} = \alpha_{s_i^l,k} \mathbf{I}_M$, since we have assumed that only large-scale fading gains are available for the out-cluster RRHs in the BBU pool.

For the latter case, it can be readily shown that $\mathbb{E} \left\{ \mathbf{h}_{s_i^l,k} \mathbf{h}_{s_j^l,k}^H \right\} = \mathbf{0}_M$, since at least one RRH does not belong to \mathcal{I}_k and they are not the same RRH.

APPENDIX B

DERIVATION OF INEQUALITY (24)

Before deriving (24), we first introduce the concept of \mathcal{T} -transform. For a complex vector $\mathbf{x} \in \mathbb{C}^n$ and a complex matrix $\mathbf{X} \in \mathbb{C}^{n \times m}$, a one-to-one mapping function of $\mathbb{C}^n \rightarrow \mathbb{R}^{2n}$ and $\mathbb{C}^{n \times m} \rightarrow \mathbb{R}^{2n \times 2m}$, is defined as

$$\mathcal{T}(\mathbf{x}) = \begin{bmatrix} \text{Re}(\mathbf{x}) \\ \text{Im}(\mathbf{x}) \end{bmatrix}, \mathcal{T}(\mathbf{X}) = \begin{bmatrix} \text{Re}(\mathbf{x}) & -\text{Im}(\mathbf{x}) \\ \text{Im}(\mathbf{x}) & \text{Re}(\mathbf{x}) \end{bmatrix}. \quad (\text{B.1})$$

The \mathcal{T} -transform establishes the relationship between complex-valued vectors or matrices and their counterparts, which facilitates the derivation of the first-order Taylor expansion for a complex-valued function.

Let us define $\hat{\mathbf{x}} \triangleq \mathcal{T}(\mathbf{x})$ and $\hat{\mathbf{X}} \triangleq \mathcal{T}(\mathbf{X})$ as the \mathcal{T} -transform results of the complex-valued vector \mathbf{x} and matrix \mathbf{X} , respectively. Then we have the following two properties for the \mathcal{T} -transform [26]:

$$\mathbf{y} = \mathbf{A}\mathbf{x} \Leftrightarrow \hat{\mathbf{y}} = \hat{\mathbf{A}}\hat{\mathbf{x}}, \quad (\text{B.2})$$

$$\text{Re}(\mathbf{x}^H \mathbf{y}) = \hat{\mathbf{x}}^T \hat{\mathbf{y}}. \quad (\text{B.3})$$

Based on the above results, $\mathbf{w}_k^H \mathbf{A}_{k,k} \mathbf{w}_k$ can be equivalently written for the real-valued vector $\hat{\mathbf{w}}_k$ and for the matrix $\hat{\mathbf{A}}_{k,k}$ as

$$\mathbf{w}_k^H \mathbf{A}_{k,k} \mathbf{w}_k \stackrel{(a)}{=} \text{Re}(\mathbf{w}_k^H \mathbf{A}_{k,k} \mathbf{w}_k) \stackrel{(b)}{=} \hat{\mathbf{w}}_k^T \mathcal{T}(\mathbf{A}_{k,k} \mathbf{w}_k) \stackrel{(c)}{=} \hat{\mathbf{w}}_k^T \hat{\mathbf{A}}_{k,k} \hat{\mathbf{w}}_k \triangleq g(\hat{\mathbf{w}}_k), \quad (\text{B.4})$$

where (a) follows since $\mathbf{w}_k^H \mathbf{A}_{k,k} \mathbf{w}_k$ is a real value, (b) follows by using (B.3) and (c) follows by using (B.2). Since $g(\hat{\mathbf{w}}_k)$ is a convex function of $\hat{\mathbf{w}}_k$, we have

$$\mathbf{w}_k^H \mathbf{A}_{k,k} \mathbf{w}_k = g(\hat{\mathbf{w}}_k) \quad (\text{B.5})$$

$$\geq g[\hat{\mathbf{w}}_k(t)] + \nabla g[\hat{\mathbf{w}}_k(t)]^H [\hat{\mathbf{w}}_k - \hat{\mathbf{w}}_k(t)] \quad (\text{B.6})$$

$$= g[\hat{\mathbf{w}}_k(t)] + 2\hat{\mathbf{w}}_k^H \hat{\mathbf{A}}_{k,k} [\hat{\mathbf{w}}_k - \hat{\mathbf{w}}_k(t)] \quad (\text{B.7})$$

$$= \mathbf{w}_k^H(t) \mathbf{A}_{k,k} \mathbf{w}_k(t) + 2\text{Re}[\mathbf{w}_k^H(t) \mathbf{A}_{k,k} (\mathbf{w}_k - \mathbf{w}_k(t))], \quad (\text{B.8})$$

where (B.7) follows since $\nabla g(\hat{\mathbf{w}}_k(t)) = 2\hat{\mathbf{A}}_{k,k} \hat{\mathbf{w}}_k(t)$, (B.2) and (B.3) are used in (B.8) similarly to (B.4). Hence, the proof is complete.

APPENDIX C

PROOF OF SLATER'S CONDITION OF PROBLEM \mathcal{P}_6

Without loss of generality, we consider Problem \mathcal{P}_6 in the first iteration of Algorithm 1, i.e., $t = 1$. As explained in Subsection IV-C, the beamforming obtained from solving Problem \mathcal{P}_7 in Section V (denoted as \mathbf{w}^*) is set as the initial beamforming in Algorithm 1, i.e., $\mathbf{w}(0) = \mathbf{w}^*$. Hence, \mathbf{w}^* is a feasible solution to Problem \mathcal{P}_6 .

The idea of the proof is to construct a new set of beam-vectors from \mathbf{w}^* such that Constraints C2, C7 and C8 in Problem \mathcal{P}_6 hold with strict inequalities [32]. As stated at the end of Section V, the iterative algorithm to solve Problem \mathcal{P}_7 will terminate once the intermediate solutions of $\{\varphi_k\}_{k \in \mathcal{U}}$ are all equal to zero. Hence, the data rates achieved by some UEs with the obtained solution \mathbf{w}^* will be strictly larger than its minimum rate requirements. We assume that UE k is one of those UEs, which satisfies

$$2\text{Re}(\mathbf{w}_k^H(0)\mathbf{A}_{k,k}\mathbf{w}_k^*) - \zeta_k(0) > \eta_{k,\min} \left(\mathbf{w}_k^{*H}\mathbf{E}_{k,k}\mathbf{w}_k^* + \sum_{l \neq k, l \in \mathcal{U}} \mathbf{w}_l^{*H}\mathbf{A}_{l,k}\mathbf{w}_l^* + \sigma_k^2 \right). \quad (\text{C.1})$$

We then scale UE k 's beam-vector by a constant $0 < \sqrt{\chi_k} < 1$ and denote the new beam-vector as $\mathbf{w}_k^\# = \sqrt{\chi_k}\mathbf{w}_k^*$. One should find such a χ_k that satisfies the following inequality:

$$2\text{Re}(\mathbf{w}_k^H(0)\mathbf{A}_{k,k}\mathbf{w}_k^\#) - \zeta_k(0) > \eta_{k,\min} \left(\mathbf{w}_k^{\#H}\mathbf{E}_{k,k}\mathbf{w}_k^\# + \sum_{l \neq k, l \in \mathcal{U}} \mathbf{w}_l^{*H}\mathbf{A}_{l,k}\mathbf{w}_l^* + \sigma_k^2 \right). \quad (\text{C.2})$$

By substituting the expressions of $\zeta_k(0)$ and $\mathbf{w}_k^\#$ into (C.2), we have

$$\mathbf{w}_k^{*H}\mathbf{A}_{k,k}\mathbf{w}_k^* > \eta_{k,\min} \left[\frac{\chi_k}{2\sqrt{\chi_k} - 1} \mathbf{w}_k^{*H}\mathbf{E}_{k,k}\mathbf{w}_k^* + \frac{1}{2\sqrt{\chi_k} - 1} \left(\sum_{l \neq k, l \in \mathcal{U}} \mathbf{w}_l^{*H}\mathbf{A}_{l,k}\mathbf{w}_l^* + \sigma_k^2 \right) \right]. \quad (\text{C.3})$$

Hence, when $\frac{1}{4} < \chi_k < 1$, $0 < \frac{\chi_k}{2\sqrt{\chi_k} - 1} < 1$ and $0 < \frac{1}{2\sqrt{\chi_k} - 1} < 1$ hold. Then, one can always find a χ_k that is very close to one such that (C.3) is satisfied.

By keeping the beam-vectors of all other UEs fixed, we immediately have

$$2\text{Re}(\mathbf{w}_l^H(0)\mathbf{A}_{l,l}\mathbf{w}_l^*) - \zeta_l(0) > \eta_{l,\min} \left(\mathbf{w}_l^{*H}\mathbf{E}_{l,l}\mathbf{w}_l^* + \sum_{j \neq l, j \in \mathcal{U}} \mathbf{w}_j^{*H}\mathbf{A}_{j,l}\mathbf{w}_j^* + \chi_k \mathbf{w}_k^{*H}\mathbf{A}_{k,l}\mathbf{w}_k^* + \sigma_l^2 \right), \forall l \neq k, l \in \mathcal{U}. \quad (\text{C.4})$$

Hence, Constraint C8 in Problem \mathcal{P}_6 with the new set of beam-vectors $\{\mathbf{w}_k^\#, \mathbf{w}_l^*, \forall l \neq k, \}$ hold with strict inequality for all UEs.

The remaining task is to prove that Constraint C2 and C6 hold with strict inequality. Unfortunately, with the new beam-vectors $\{\mathbf{w}_k^\#, \mathbf{w}_l^*, \forall l \neq k, \}$, we only guarantee the following strict inequalities corresponding to the RRHs in \mathcal{I}_k :

$$\sum_{l \neq k, l \in \mathcal{U}_i} \|\mathbf{w}_{i,l}^*\|^2 + \chi_k \|\mathbf{w}_{i,k}^*\|^2 < P_{i,\max}, i \in \mathcal{I}_k, \quad (\text{C.5})$$

$$\sum_{l \neq k, l \in \mathcal{U}_i} \tau_{i,l}(0) \|\mathbf{w}_{i,l}^*\|^2 + \tau_{i,k}(0) \chi_k \|\mathbf{w}_{i,k}^*\|^2 < \tilde{C}_i(0), i \in \mathcal{I}_k. \quad (\text{C.6})$$

To deal with this issue, we randomly select one RRH from $\mathcal{I} \setminus \mathcal{I}_k$, say RRH i . Then, randomly select one UE served by RRH i , say UE l . We perform the same scaling operation as UE k for UE l , i.e., $\mathbf{w}_l^\# = \sqrt{\chi_l} \mathbf{w}_l^*$. One can find a χ_l ($\frac{1}{4} < \chi_l < 1$) such that

$$2\text{Re} \left(\mathbf{w}_l^H(0) \mathbf{A}_{l,l} \mathbf{w}_l^\# \right) - \zeta_l(0) > \eta_{l,\min} \left(\mathbf{w}_l^{\#H} \mathbf{E}_{l,l} \mathbf{w}_l^\# + \sum_{j \neq l, k, j \in \mathcal{U}} \mathbf{w}_j^{*H} \mathbf{A}_{j,l} \mathbf{w}_j^* + \chi_k \mathbf{w}_k^{*H} \mathbf{A}_{k,l} \mathbf{w}_k^* + \sigma_l^2 \right). \quad (\text{C.7})$$

Obviously, with the new set of beam-vectors $\{\mathbf{w}_k^\#, \mathbf{w}_l^\#, \mathbf{w}_j^*, \forall j \neq k, j \neq l\}$, Constraint C8 corresponding to the other UEs hold with strict inequality. Then, Constraint C2 and C6 corresponding to the RRHs in \mathcal{I}_l hold with strict inequality. Repeat this step until Constraint C2 and C6 of all the RRHs in \mathcal{I} hold with strict inequality. Then, the final constructed set of beam-vectors remain in the interior of the feasible region of Problem \mathcal{P}_6 . Hence, according to Page 226 in [32], the Slater's condition of Problem \mathcal{P}_6 is satisfied. For Problem \mathcal{P}_6 in the subsequent iterations of Algorithm 1, the similar proof applies.

APPENDIX D

THE EQUIVALENCE BETWEEN PROBLEM \mathcal{P}_1 AND PROBLEM \mathcal{P}_7

Denote the optimal solution of Problem \mathcal{P}_1 and Problem \mathcal{P}_7 as $\{\mathcal{U}^*, \mathbf{w}^*\}$ and $\{\mathcal{U}^\#, \mathbf{w}^\#\}$, respectively. We first prove that the optimal solution of Problem \mathcal{P}_1 is feasible for Problem \mathcal{P}_7 . It is obvious that $\{\mathcal{U}^*, \mathbf{w}^*\}$ is feasible for Constraints C2 and C4 of Problem \mathcal{P}_7 since Constraint C4 is the equivalent transformation of Constraint C1 in Problem \mathcal{P}_1 . Now we show that $\{\mathcal{U}^*, \mathbf{w}^*\}$ is also feasible for Problem \mathcal{P}_7 . Specifically, we have the following chain inequalities:

$$\sum_{k \in \mathcal{U}_i^*} \varepsilon \left(\|\mathbf{w}_{i,k}^*\|^2 \right) R_{k,\min} \leq \sum_{k \in \mathcal{U}_i^*} \varepsilon \left(\|\mathbf{w}_{i,k}^*\|^2 \right) r_k^* \leq C_{i,\max}, \forall i \in \mathcal{I}, \quad (\text{D.1})$$

where r_k^* is obtained by substituting \mathbf{w}^* into (9). Then, $\{\mathcal{U}^*, \mathbf{w}^*\}$ satisfies Constraint C5 of Problem \mathcal{P}_7 . Hence, $\{\mathcal{U}^*, \mathbf{w}^*\}$ is feasible for Problem \mathcal{P}_7 .

For Problem \mathcal{P}_7 , if $r_k^\# = R_{k,\min}, \forall k$, where $r_k^\#$ is obtained by substituting $\mathbf{w}^\#$ into (9), then we have

$$\sum_{k \in \mathcal{U}_i^\#} \varepsilon \left(\left\| \mathbf{w}_{i,k}^\# \right\|^2 \right) r_k^\# = \sum_{k \in \mathcal{U}_i^\#} \varepsilon \left(\left\| \mathbf{w}_{i,k}^\# \right\|^2 \right) R_{k,\min} \leq C_{i,\max}, \forall i \in \mathcal{I} \quad (\text{D.2})$$

which satisfies Constraint C3 of Problem \mathcal{P}_1 . It is readily verified that $\{\mathcal{U}^\#, \mathbf{w}^\#\}$ satisfies Constraints C1 and C2 of Problem \mathcal{P}_1 . Hence, $\{\mathcal{U}^\#, \mathbf{w}^\#\}$ is also feasible for Problem \mathcal{P}_1 . On the other hand, if there exists at least one UE whose data rate is strictly larger than its rate requirement, i.e., $r_k^\# > R_{k,\min}$. Then, we can adopt the iterative scaling algorithm given in Appendix A of [23] to construct another set of beamforming vectors $\mathbf{w}^{\#\#}$ such that $r_k^{\#\#} = R_{k,\min}, \forall k$, where $r_k^{\#\#}$ is obtained by substituting $\mathbf{w}^{\#\#}$ into (9). As a result, $\{\mathcal{U}^\#, \mathbf{w}^{\#\#}\}$ is feasible for Problem \mathcal{P}_1 .

Based on the above discussions, we arrive at the conclusion that Problem \mathcal{P}_1 and Problem \mathcal{P}_7 can achieve the same optimal set of selected UEs.

REFERENCES

- [1] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [2] M. Peng, Y. Sun, X. Li, Z. Mao, and C. Wang, "Recent advances in cloud radio access networks: System architectures, key techniques, and open issues," *IEEE Commun. Surveys Tut.*, vol. 18, no. 3, pp. 2282–2308, thirdquarter 2016.
- [3] Y. Shi, J. Zhang, K. B. Letaief, B. Bai, and W. Chen, "Large-scale convex optimization for ultra-dense cloud-RAN," *IEEE Wireless Commun. Mag.*, vol. 22, no. 3, pp. 84–91, Jun. 2015.
- [4] R. G. Stephen and R. Zhang, "Joint millimeter-wave fronthaul and OFDMA resource allocation in ultra-dense CRAN," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1411–1423, Mar. 2017.
- [5] Y. Shi, J. Zhang, and K. Letaief, "Group sparse beamforming for green Cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.
- [6] B. Dai and W. Yu, "Energy efficiency of downlink transmission strategies for cloud radio access networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 1037–1050, Apr. 2016.
- [7] —, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access*, vol. 2, pp. 1326–1339, Oct. 2014.
- [8] V. N. Ha, L. B. Le, and N. D. Dao, "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.
- [9] A. Abdelnasser and E. Hossain, "Resource allocation for an OFDMA Cloud-RAN of small cells underlying a macrocell," *IEEE Trans. Mobile Comput.*, vol. 15, no. 11, pp. 2837–2850, Nov. 2016.
- [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] P. Luong, L. N. Tran, C. Despins, and F. Gagnon, "Joint beamforming and remote radio head selection in limited fronthaul C-RAN," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Sept 2016, pp. 1–6.

- [12] P. Luong, F. Gagnon, C. Despins, and L. N. Tran, "Optimal joint remote radio head selection and beamforming design for limited fronthaul C-RAN," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp. 5605–5620, Nov. 2017.
- [13] Y. Shi, J. Zhang, and K. B. Letaief, "CSI overhead reduction with stochastic beamforming for cloud radio access networks," in *2014 IEEE International Conference on Communications (ICC)*, 2014, pp. 5154–5159.
- [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] C. Fan, Y. J. Zhang, and X. Yuan, "Dynamic nested clustering for parallel PHY-layer processing in Cloud-RANs," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1881–1894, Mar. 2016.
- [16] 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.
- [17] Z. Bai, "Evolved universal terrestrial radio access (E-UTRA); physical layer procedures," *3GPP, Sophia Antipolis, Technical Specification 36.213 v. 11.4. 0*, 2013.
- [18] T. X. Tran, A. Hajisami, and D. Pompili, "QuaRo: A queue-aware robust coordinated transmission strategy for downlink C-RANs," in *2016 13th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, June 2016, pp. 1–9.
- [19] V. K. N. Lau, F. Zhang, and Y. Cui, "Low complexity delay-constrained beamforming for multi-user mimo systems with imperfect csit," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 4090–4099, Aug. 2013.
- [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] J. Zhang, X. Yuan, and Y. J. Zhang, "Locally orthogonal training design for Cloud-RANs based on graph coloring," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6426–6437, Oct. 2017.
- [22] T. M. Nguyen and L. B. Le, "Joint pilot assignment and resource allocation in multicell massive MIMO network: Throughput and energy efficiency maximization," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2015, pp. 393–398.
- [23] C. Pan, H. Mehrpouyan, Y. Liu, M. El Kashlan, and N. Arumugam, "Joint pilot allocation and robust transmission design for ultra-dense user-centric TDD C-RAN with imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2038–2053, Mar. 2018.
- [24] D. Su, X. Hou, and C. Yang, "Quantization based on per-cell codebook in cooperative multi-cell systems," in *2011 IEEE Wireless Communications and Networking Conference*, March 2011, pp. 1753–1758.
- [25] F. Yuan and C. Yang, "Bit allocation between per-cell codebook and phase ambiguity quantization for limited feedback coordinated multi-point transmission systems," *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2546–2559, Sep. 2012.
- [26] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.
- [27] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*. Prentice Hall Upper Saddle River, NJ, 2000, vol. 1.
- [28] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.
- [29] 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.
- [30] T. V. Chien, E. Bjornson, and E. G. Larsson, "Joint power allocation and user association optimization for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6384–6399, Sep. 2016.

- [31] C. Pan, H. Ren, M. El Kashlan, A. Nallanathan, and L. Hanzo, "Robust beamforming design for ultra-dense user-centric C-RAN in the face of realistic pilot contamination and limited feedback," *to appear in arXiv*, 2018.
- [32] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [33] 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*. Springer, 2010, pp. 93–102.
- [34] C. Pan, W. Xu, W. Zhang, J. Wang, H. Ren, and M. Chen, "Weighted sum energy efficiency maximization in ad hoc networks," *IEEE Wireless Commun. Lett.*, vol. 4, no. 3, pp. 233–236, Jun. 2015.
- [35] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001.
- [36] E. Manskani, 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.
- [37] X. Ge, S. Tu, G. Mao, C. X. Wang, and T. Han, "5G ultra-dense cellular networks," *IEEE Wireless Commun.*, vol. 23, no. 1, pp. 72–79, Feb. 2016.
- [38] E. U. T. R. Access, "Further advancements for E-UTRA physical layer aspects," *3GPP TR 36.814, Tech. Rep.*, 2010.
- [39] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 458–462, Feb. 2007.
- [40] C. Zhang, W. Xu, and M. Chen, "Robust MMSE beamforming for multiuser MISO systems with limited feedback," *IEEE Signal Process. Lett.*, vol. 16, no. 7, pp. 588–591, Jul. 2009.
- [41] A. K. Gupta and S. Nadarajah, *Handbook of beta distribution and its applications*. CRC press, 2004.