Coordinated Multipoint Transmission Design for Cloud-RANs With Limited Fronthaul Capacity Constraints

Vu Nguyen Ha, *Student Member, IEEE*, Long Bao Le, *Senior Member, IEEE*, and Ngọc-Dũng Đào, *Member, IEEE*

Abstract—In this paper, we consider the coordinated multipoint (CoMP) transmission design for the downlink cloud radio access network (Cloud-RAN). Our design aims to optimize the set of remote radio heads (RRHs) serving each user and the precoding and transmission power to minimize the total transmission power while maintaining the fronthaul capacity and users' quality-of-service (QoS) constraints. The fronthaul capacity constraint involves a nonconvex and discontinuous function that renders the optimal exhaustive search method unaffordable for large networks. To address this challenge, we propose two low-complexity algorithms. The first pricing-based algorithm solves the underlying problem through iteratively tackling a related pricing problem while appropriately updating the pricing parameter. In the second iterative linear-relaxed algorithm, we directly address the fronthaul constraint function by iteratively approximating it with a suitable linear form using a conjugate function and solving the corresponding convex problem. For performance evaluation, we also compare our proposed algorithms with two existing algorithms in the literature. Finally, extensive numerical results are presented, which illustrate the convergence of our proposed algorithms and confirm that our algorithms significantly outperform the state-of-the-art existing algorithms.

Index Terms—Cloud radio access network (Cloud-RAN), limited fronthaul capacity, power minimization, precoding, resource allocation.

I. INTRODUCTION

THE next-generation wireless cellular network is expected to provide significantly enhanced capacity to support the exponential growth of mobile data traffic [1]–[5]. Toward this end, coordinated multipoint (CoMP) transmission/reception techniques provide promising solutions, which have been adopted in the Long-Term Evolution Advanced standard [6]. In fact, CoMP employs different forms of base-station (BS) coordination with dynamic sharing of channel state information (CSI) and/or data information among BSs, as well as efficient transmission, precoding, and resource-allocation algorithms

[7]. However, implementation of CoMP in wireless cellular networks typically requires costly high-speed backhauls connecting different BSs for CSI and information exchanges and distributed computation [8], [9].

Cloud-RAN has been recently proposed as an alternative network architecture that can achieve the performance gains of the CoMP techniques while effectively exploiting the computation power of the cloud computing technology. In a Cloud-RAN, the cloud comprising a pool of baseband processing units (BBUs) performs most baseband processing tasks while transmission functions are realized by simple remote radio heads (RRHs) using the processed baseband signals received from the cloud through a fronthaul transport network [10]–[13]. With a centralized cloud processing center (CPC), complex resource-allocation optimization, such as precoding, power control, user scheduling, interference management algorithms, can be realized, which can translate into significant network performance improvements. In addition, Cloud-RAN with simple RRHs enables us to reduce both capital and operational expenditures of the network [12]. Moreover, this emerging network architecture also allows to deploy dense small cells efficiently [14].

Despite these benefits, there are various technical challenges one must resolve in designing and deploying the Cloud-RAN architecture. In particular, suitable mechanisms that efficiently utilize computing resources in the cloud, fronthaul capacity to realize advanced communications, baseband signal processing, and resource-allocation schemes must be developed. Some recent works have addressed some of these issues, as can be described in the following. In [15], an efficient clustering algorithm was proposed to reduce the number of computations in the centralized pool of BBUs where the number of RRHs is very large. In [16] and [17], compression techniques to minimize the amount of data transmitted over the fronthaul transport network were developed.

Several other papers focus on precoding/decoding design for CoMP considering different design aspects of Cloud-RAN. In particular, in [18], the joint optimization of antenna selection, regularization, and power allocation was studied to maximize the average weighted sum rate. The random matrix theory was utilized to decompose the considered nonconvex problem into subproblems that can be tackled more easily. In [19] and [20], the precoding vectors were optimized for all RRHs to minimize the total network power consumption; the downlink was considered in [19], whereas both downlink and uplink communications were addressed in [20]. Total power

to support radio transmission and operations of fronthaul links and RRHs is accounted for in these research works, where the authors show how to transform the underlying problems into the sparse beamforming problems. Then, these papers combine the solution techniques employed to address the traditional power minimization problem in [21] and [22] and the fronthaul capacity minimization problem in [23] to tackle the transformed problem. This solution technique is also closely related to that employed in [24]. In fact, the design problems considered in [23] and [24] for CoMP and in [19] and [20] for Cloud-RAN with standard convex constraints do not explicitly model the fronthaul capacity constraints. Consequently, they can be solved directly by employing the compressed sensing techniques [26]–[28]. We would like to emphasize that the related work [23] aims at minimizing the number of active links between base stations (RRHs) and users. Therefore, the limited fronthaul capacity is not explicitly imposed as constraints in [23], but it is rather considered in the design objective. Note that the greedy principle proposed in [23] can be employed to solve our problem (although our problem is not the same with the problem in [23]); however, such a greedy approach may not achieve very good performance.

This paper aims to study the CoMP joint transmission design problem for Cloud-RAN that explicitly considers the fronthaul capacity and users' quality-of-service (QoS) constraints. In particular, we make the following contributions.

- We formulate the joint transmission design problem for Cloud-RAN, which optimizes the set of RRHs serving each user together with their precoding and power-allocation solutions to minimize the total transmission power considering users' QoS and fronthaul capacity constraints. In particular, we assume that certain fronthaul capacity, which depends on the user's required QoS, is consumed to transfer the processed baseband signal from the cloud to one particular RRH serving that user and that the fronthaul transport network has limited capacity. The considered problem is indeed nonconvex, and it comprises the discontinuous fronthaul capacity constraint and other traditional power and users' QoS constraints. Therefore, addressing this problem through exhaustive search requires exponential computation complexity, which is not affordable for practically large networks.

- To tackle the considered problem, we then develop two different low-complexity algorithms. For the first algorithm, which is called *pricing-based algorithm*, we consider the related pricing problem and devise a novel mechanism to iteratively update the pricing parameter to obtain a good feasible solution. The second algorithm, which is called *iterative linear-relaxed algorithm*, is developed by regularizing the fronthaul constraint function into an approximated linear form and iteratively solving the approximated optimization problem.

- For performance comparison, we also describe two existing algorithms proposed in [23] and [25] (with appropriate modifications due to different considered settings), which can also solve our considered problem. We then study the extended setting where there are multiple individual fronthaul capacity constraints. We show that it
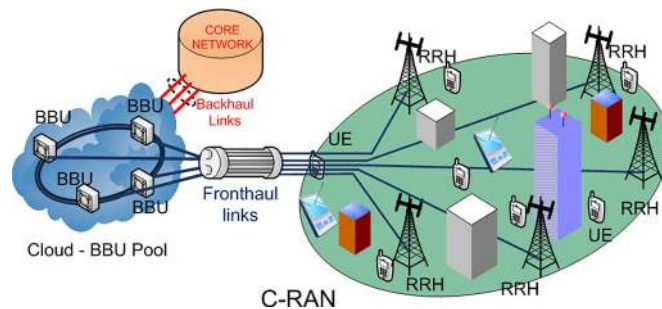

Fig. 1. Cloud-RAN architecture.

is possible to adapt the proposed low-complexity algorithms to solve the corresponding problem in this setting. In addition, we indicate the need to introduce multiple pricing parameters, each of which is required to deal with one corresponding fronthaul capacity constraint in the extended *iterative linear-relaxed algorithm*.

- Numerical results are presented to demonstrate the efficacy of the proposed algorithms and their relative performance compared with the optimal algorithm and existing algorithms. We show the convergence and the impacts of the number of users, the number of antennas, and users' required QoS on the network performance. Specifically, while both two proposed algorithms perform remarkably well in most cases, the pricing-based algorithm achieves slightly better performance than the iterative linear-relaxed algorithm at the cost of longer convergence time. In addition, both proposed algorithms significantly outperform the existing algorithms in [23] and [25] for all considered simulation scenarios.

The remaining of this paper is organized as follows. We describe the system model and problem formulation in Section II. In Section III, we present two low-complexity algorithms and two existing algorithms. Next, we extend our proposed algorithms to the setting with multiple individual fronthaul capacity constraints and the multiple-input–multiple-output (MIMO) systems where multiple data streams can be transmitted to each user in Section IV. Numerical results are presented in Section V followed by the conclusion in Section VI. Some preliminary results of this paper have been published in [29] and [30].

For notation, we use $\mathbf{X}^T$, $\mathbf{X}^H$, $\mathrm{Tr}(\mathbf{X})$, and $\mathrm{rank}(\mathbf{X})$ to denote transpose, Hermitian transpose, trace, and rank of matrix $\mathbf{X}$, respectively. $\mathbf{1}_{x \times y}$ and $\mathbf{0}_{x \times y}$ denote the matrix of ones and the matrix of zeros whose dimension are $x \times y$, respectively. $|\mathcal{S}|$ denotes the cardinality of set $\mathcal{S}$, and $\mathrm{diag}(\mathbf{x})$ is the diagonal matrix constructed from the elements of vector $\mathbf{x}$.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

The general architecture of Cloud-RAN under consideration, which is shown in Fig. 1, consists of three main components: 1) CPC consisting a BBU pool; 2) the optical fronthaul transport network (i.e., fronthaul links); and 3) RRH access units with antennas located at remote sites. Specifically, the CPC comprising a certain number of BBUs is the heart of this architecture where BBUs act as virtual base stations that process

baseband signals for users and optimize the network-resource-allocation tasks. The fronthaul transport network connecting the CPC and distributed RRHs is usually deployed by using optical fibers. In addition, RRHs communicate with users in the downlink via RF signals, which are formed by using the baseband signals and the precoding vectors received from the BBU pool.

In this paper, we consider the joint transmission design for CoMP downlink communications in the Cloud-RAN with K RRHs and M users. Let $\mathcal{K}$ and $\mathcal{U}$ be the sets of RRHs and users in the network, respectively. Suppose that RRH k is equipped with N_k antennas ($k \in \mathcal{K}$) and each user has a single antenna. We assume that each user is served by a specific group of assigned RRHs, and one RRH can serve a number of users. When RRH k is assigned to serve user u, this RRH receives the user's processed baseband signal from the cloud. Then, the RRH converts and transmits the corresponding RF signal using a suitably designed precoding vector. Moreover, denote $\mathbf{v}_u^k \in \mathbb{C}^{N_k \times 1}$ as the precoding vector at RRH k corresponding to the signal transmitted to user u. Then, the transmission power used by RRH k to serve user u can be expressed as

$$p_u^k = \mathbf{v}_u^{kH} \mathbf{v}_u^k. \tag{1}$$

Let (k, u) represent the communication link between RRH k and user u. For simplicity, we assume that each user is virtually served by all RRHs in all expressions and problem formulations. However, a user is actually served by one particular RRH if the corresponding transmission power is strictly larger than zero. Let $x_u \in \mathbb{C}$ denote the signal symbols with unit power for user u, which are transmitted by RRHs in set $\mathcal{K}$ upon receiving the processed baseband signals from the cloud. Then, the baseband signal y_u received at user u can be written as

$$y_u = \underbrace{\sum_{k \in \mathcal{K}} \mathbf{h}_u^{kH} \mathbf{v}_u^k x_u}_{\text{desired signal}} + \underbrace{\sum_{i=1, \neq u}^{M} \sum_{l \in \mathcal{K}} \mathbf{h}_u^{lH} \mathbf{v}_i^l x_i}_{\text{interference}} + \eta_u \tag{2}$$

where $\mathbf{h}_u^k \in \mathbb{C}^{N_k \times 1}$ denotes the channel vector between RRH k and user u, and η_u describes the noise at user u. Then, the signal-to-interference-plus-noise ratio (SINR) achieved by user u can be described as

$$\Gamma_u = \frac{\left| \sum_{k \in \mathcal{K}} \mathbf{h}_u^{kH} \mathbf{v}_u^k \right|^2}{\sum_{i=1, \neq u}^{M} \left| \sum_{l \in \mathcal{K}} \mathbf{h}_u^{lH} \mathbf{v}_i^l \right|^2 + \sigma^2} \tag{3}$$

where σ^2 is the noise power. In this paper, we are interested in design the precoding to satisfy users' QoS. This QoS constraint of user u is expressed as

$$\Gamma_u \geq \bar{\gamma}_u \quad \forall u \in \mathcal{U} \tag{4}$$

where $\bar{\gamma}_u$ denotes the target SINR of user u.

To manage the computation complexity in large networks, we can impose additional constraints where each user u can only be served by a set of near RRHs $\mathcal{R}_u$ (e.g., clustering constraints).

This set of RRHs can be determined based on the distance or channel gain from them to each user.[1] In the following, we assume that only RRH–user communication links in predetermined set $\mathcal{L}$ are allowed, where $\mathcal{L} = \{(k, u) | u \in \mathcal{U}, k \in \mathcal{R}_u\}$. When no such constraint exists, set $\mathcal{L}$ comprises all possible MK links between K RRHs and M users. We can express the relation between transmission power and communication link (k, u) as follows:

$$p_u^k = \mathbf{v}_u^{kH} \mathbf{v}_u^k = 0 \text{ if } (k, u) \notin \mathcal{L}. \tag{5}$$

Remark 1: In this paper, we assume that perfect CSI between all RRHs and users is available at the CPC. In practice, the CSI can be estimated by the corresponding users and RRHs and then transferred to the CPC through the fronthaul network.[2] In general, the transfer of CSI to the CPC over the fronthaul network consumes much smaller fraction of fronthaul capacity compared with the transfer of users' data (i.e., I/Q data sequences) from the cloud to the RRHs. Moreover, the transfer of CSI can be performed periodically since the CSI would change slowly over time, whereas the data transfer over the fronthaul network occurs continuously. Therefore, we do not explicitly consider the fronthaul capacity consumption due to CSI transfer in this paper, which is left for our future work.

1) Fronthaul Capacity: Let $\mathbf{p}^k = [p_1^k, \dots, p_M^k]^T$ be the transmission power vector of RRH k whose elements $p_u^k \geq 0$ represent transmission power p_u^k given in (1). We also define the vector $\mathbf{p} = [\mathbf{p}^{1T}, \dots, \mathbf{p}^{KT}]^T$ to describe the transmission power of all RRHs for all users. Recall that $p_u^k = 0$ implies that RRH k does not serve user u. In contrast, if $p_u^k > 0$, the fronthaul link from RRH k to user u is activated for carrying the baseband signal to serve user u at its required target SINR. Therefore, the total capacity of the fronthaul links can be indicated by the value of vector $\mathbf{p}$ and the target SINR of users, which can be written mathematically as

$$G(\mathbf{p}) = \sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{U}} \delta\left(p_u^k\right) R_u^{k, \text{fh}} \tag{6}$$

where $\delta(\cdot)$ denotes the step function, and $R_u^{k, \text{fh}}$ represents the required capacity for transferring the signal of user u over the fronthaul network to the RRH k, The required fronthaul capacity $R_u^{k, \text{fh}}$ depends on the actual required data rate corresponding to the target SINR $\bar{\gamma}_u$, the specific design of Cloud-RAN, and the specific quantization technique employed to process the baseband signals, which are discussed in the following remark.

Remark 2: In the Cloud-RAN downlink system, the CPC processes the baseband signals and optimizes the precoding vectors. As described in [38] and [39], there are two strategies to realize such design, namely "compression-after-precoding" (CAP) and "compression-before-precoding" (CBP). In the CAP, which is employed in [16] and [17], the CPC precodes the

---

[1]For example, each user is potentially associated with r nearest RRHs as rth-order Voronoi tessellation [34] or RRHs with strong channel gain as in [35].

[2]Detailed realization of CSI channel estimation varies depending if the time-division duplex (TDD) or frequency-division duplex (FDD) strategy is employed. In particular, in the TDD system, RRHs perform CSI estimation; then, they transfer the estimated CSI to the CPC.

data streams of users with their corresponding precoding vectors, then compresses the resulting signals, and finally forwards them to the corresponding RRHs over the fronthaul network. For the CBP, the CPC directly compresses the precoding vectors and forwards the compressed precoding vectors and users' data streams to the corresponding RRHs [38]. Then, the RRHs precode the signals and transmit to users.

In this paper, we assume that the CBP strategy is employed, which is motivated by its advantages for the Cloud-RAN system with preliminary clustering [39]. In this strategy, information streams and precoding vectors are compressed and transmitted from the cloud to RRHs separately. The fronthaul capacity consumed by each RRH depends on the group of users it serves, which is given in (6). In particular, the required fronthaul capacity to support the communication link between user u to RRH k can be calculated as

$$R_u^{k,\mathrm{fh}} = R_u^{k,\mathrm{pr}} + R_u^{\mathrm{dt}}(\bar{\gamma}_u) \qquad (7)$$

where $R_u^{\mathrm{dt}}(\bar{\gamma}_u)$ and $R_u^{k,\mathrm{pr}}$ correspond to the fronthaul capacity consumption for transferring the information stream and precoding vector $\mathbf{v}_u^k$ for user u, as given in [38, Eq. 3], respectively. In fact, $R_u^{\mathrm{dt}}(\bar{\gamma}_u)$ is a function of the target SINR (e.g., it can be expressed as $R_u^{\mathrm{dt}}(\bar{\gamma}_u) = \log_2(1 + \bar{\gamma}_u)$ if the capacity can be achieved), and $R_u^{k,\mathrm{pr}}$ can be predetermined based on the desirable quantization quality of precoding vectors. When the precoding vectors are quantized and then employed at RRHs, there are the quantization errors. However, if the corresponding quantization noise is sufficiently small compared with the interference and noise at receivers, the quantization errors can be omitted, which is what to be assumed in this paper.

Note that our model allows to capture the scenario where users demand services with different rate requirements (e.g., voice and video services). We are now ready to describe the considered fronthaul-constrained power minimization (FCPM) problem in the following.

B. Problem Formulation

We are interested in determining the set of RRHs serving each user and the corresponding precoding vectors (i.e., $\{\mathbf{v}_u^k\}$) to minimize the total transmission power considering the constraints on fronthaul capacity limit, transmission power, and users' QoS. In addition to the SINR constraints in (4), we also impose the constraint on total transmission power for each RRH k as

$$\sum_{u\in\mathcal{U}} p_u^k = \|\mathbf{p}^k\|_1 = \sum_{u\in\mathcal{U}} \mathbf{v}_u^{kH}\mathbf{v}_u^k \le P_k \quad \forall k\in\mathcal{K} \qquad (8)$$

where P_k $(k\in\mathcal{K})$ denotes the maximum power of RRH k, and $\|\mathbf{x}\|_1$ represents the $\ell1$-norm of vector $\mathbf{x}$. Furthermore, we assume that the capacity of fronthaul transport network between the cloud and all RRHs is limited, and we denote C as the fronthaul capacity limit. Then, we have to impose the following fronthaul capacity constraint:

$$G(\mathbf{p}) = \sum_{k\in\mathcal{K}}\sum_{u\in\mathcal{U}} \delta\left(p_u^k\right) R_u^{k,\mathrm{fh}} \le C. \qquad (9)$$

This indeed represents the sum fronthaul capacity constraint. We will discuss the extended setting with multiple individual fronthaul capacity constraints in Section IV. Now, it is ready to state the FCPM problem as follows:

$$(\mathcal{P}_{\mathrm{FCPM}}) \quad \min_{\{\mathbf{v}_u^k\},\mathbf{p}} \quad \|\mathbf{p}\|_1$$
$$\text{s. t.} \quad \text{constraints } (1),(4),(5),(8), \text{ and } (9). \quad (10)$$

We will describe how to transform this problem into an appropriate form, based on which we can determine its optimal solution in the following.

1) Optimal Exhaustive Search Algorithm: The main challenge involved in solving problem $\mathcal{P}_{\mathrm{FCPM}}$ comes from the fronthaul capacity constraint (9). Let us now define variables $a_u^k \in \{0,1\}$, where $a_u^k = 1$ if RRH k serves user u and $a_u^k = 0$ otherwise. Denote $\mathfrak{a}$ as the allocation vector $[a_1^1,\ldots,a_M^K]$ and $\mathcal{S}_\mathfrak{a}$ as the set of all possible $\mathfrak{a}$. Then, the consumed fronthaul capacity corresponding to a given vector $\mathfrak{a}$ can be written as

$$G(\mathfrak{a}) = \sum_{k\in\mathcal{K}}\sum_{u\in\mathcal{U}} a_u^k R_u^{k,\mathrm{fh}}. \qquad (11)$$

Then, we have to optimize over both integer variables $\mathfrak{a}$ in $\mathcal{S}_\mathfrak{a}$ and continuous variables $\{\mathbf{v}_u^k\}$ to find the optimal solution of the considered problem so that $G(\mathfrak{a}) \le C$. Note, however that the number of elements in set $\mathcal{S}_\mathfrak{a}$ is finite. Moreover, suppose that the values of $\mathfrak{a}$ that satisfy constraint (9) are given, then we only need to solve following precoding optimization problem:

$$(\mathcal{P}_\mathfrak{a}) \quad \min_{\{\mathbf{v}_u^k\}} \quad \sum_{k\in\mathcal{K}}\sum_{u\in\mathcal{U}} \mathbf{v}_u^{kH}\mathbf{v}_u^k$$
$$\text{s. t.} \quad \text{constraints } (4) \text{ and } (8)$$
$$\mathbf{v}_u^{kH}\mathbf{v}_u^k = 0 \text{ if } a_u^k = 0 \quad \forall(u,k). \qquad (12)$$

This problem is indeed the sum-power minimization problem (SPMP), which can be transformed into a solvable convex semi-definite program (SDP). This transformation is described in Appendix A. Here, we refer to such an SPMP for given $\mathfrak{a}$ as a sparse PMP. It is now clear that the optimal solution for problem $\mathcal{P}_{\mathrm{FCPM}}$ can be determined as follows. First, we enumerate all possible $\mathfrak{a}$ that satisfy the fronthaul capacity constraints (9). Second, for each such feasible $\mathfrak{a}$, we solve the corresponding problem $(\mathcal{P}_\mathfrak{a})$ and obtain the optimal precoding vectors and objective value. Finally, the feasible $\mathfrak{a}$ that achieves the minimum total transmission power together with the corresponding optimal precoding vectors are the optimal solution of problem $\mathcal{P}_{\mathrm{FCPM}}$.

Such exhaustive search method, however, has exponentially high complexity. This motivates us to develop low-complexity algorithms to solve problem $\mathcal{P}_{\mathrm{FCPM}}$, which is the focus of the following section.

III. LOW-COMPLEXITY ALGORITHMS

Here, we describe two low-complexity algorithms with different levels of complexity to solve problem $\mathcal{P}_{\mathrm{FCPM}}$. For the

performance evaluation purposes, we also present two existing algorithms in [23] and [25].

A. Pricing-Based Algorithm

The first low-complexity algorithm is developed by employing the penalty method to deal with the step-function fronthaul capacity constraint. Specifically, we consider the so-called *pricing fronthaul capacity and power minimization* (PFCPM) problem, which is given in the following:

$$(\mathcal{P}_{\text{PFCPM}}) \quad \min_{\{\mathbf{v}_u^k\},\mathbf{p}} \quad \|\mathbf{p}\|_1 + qG(\mathbf{p})$$

$$\text{s. t. constraints } (1), (4), (5), \text{ and } (8). \quad (13)$$

In this problem, the consumed fronthaul capacity $G(\mathbf{p})$ scaled by a pricing parameter q is added to the objective function of problem $\mathcal{P}_{\text{FCPM}}$, and we have removed the fronthaul capacity constraint (9). We can obtain a good feasible solution for problem $\mathcal{P}_{\text{FCPM}}$ by iteratively solving problem $\mathcal{P}_{\text{PFCPM}}$ while adaptively adjusting the pricing parameter q to satisfy the fronthaul capacity constraint (9). We describe this procedure in details in the following.

1) Relationship Between FCPM and PFCPM Problems: We now establish the relationship between problems $\mathcal{P}_{\text{FCPM}}$ and $\mathcal{P}_{\text{PFCPM}}$, based on which we can develop an efficient algorithm to solve problem $\mathcal{P}_{\text{FCPM}}$. Let $G_{\text{PFCPM}}(q)$ be the consumed fronthaul capacity resulted from solving problem $\mathcal{P}_{\text{PFCPM}}$ for a given pricing parameter q. Then, it can be verified that the optimal precoding vectors and the transmission power of problems $\mathcal{P}_{\text{FCPM}}$ and $\mathcal{P}_{\text{PFCPM}}$ are the same if $G_{\text{PFCPM}}(q) = C$. This fact enables us to develop an iterative algorithm to solve problem $\mathcal{P}_{\text{FCPM}}$ through tackling problem $\mathcal{P}_{\text{PFCPM}}$ while adjusting q iteratively in attempting to attain $G_{\text{PFCPM}}(q) = C$. Our proposed algorithm is developed based on the assumption that problem $\mathcal{P}_{\text{PFCPM}}$ can be solved, which will be addressed in the following. In the following, we establish some theoretical results based on which we develop the mechanism to update the pricing parameter.

Proposition 1: Let $\sigma_{\min}$ be the smallest nonzero value of $|G(\mathfrak{a}) - G(\mathfrak{a}')|$, where $\{\mathfrak{a}, \mathfrak{a}'\} \subset \mathcal{S}_\mathfrak{a}$, and $\bar{q} = \sum_{k \in \mathcal{K}} P_k / \sigma_{\min}$. We have the following.

1) $G_{\text{PFCPM}}(q)$ is a decreasing function of q.
2) If we increase $q \geq \bar{q}$, then $G_{\text{PFCPM}}(q)$ cannot be further decreased.
3) When $q \geq \bar{q}$, if $G_{\text{PFCPM}}(q) > C$, problem $\mathcal{P}_{\text{FCPM}}$ is infeasible.

Proof: The proof is given in Appendix B. ∎

These results form the foundation based on which we can develop an iterative algorithm presented in Algorithm 1. In fact, we have described how the pricing parameter q is updated over iterations, which can be summarized as follows. At the first iteration, $q^{(0)}$ is set equal to $\bar{q}$ to verify the feasibility of problem $\mathcal{P}_{\text{FCPM}}$. If $G_{\text{PFCPM}}(\bar{q}) > C$, we can conclude that problem $\mathcal{P}_{\text{FCPM}}$ is infeasible. Otherwise, we apply the bisection search method to update q until $G_{\text{PFCPM}}(\bar{q}) = C$.

Algorithm 1 Pricing-based Algorithm for FCPM Problem

1: Solve PFCPM problem using Alg. 2 with $q^{(0)} = \bar{q}$.
2: **if** $G_{\text{PFCPM}}(\bar{q}) > C$ **then**
3:    Stop, the FCPM problem is infeasible.
4: **else if** $G_{\text{PFCPM}}(\bar{q}) = C$ **then**
5:    Stop, the solution is achieved.
6: **else if** $G_{\text{PFCPM}}(\bar{q}) < C$ **then**
7:    Set $l = 0$, $q_U^{(l)} = \bar{q}$ and $q_L^{(l)} = 0$.
8:    **repeat**
9:       Set $l = l + 1$ and $q^{(l)} = (q_U^{(l-1)} + q_L^{(l-1)})/2$.
10:      Solve PFCPM problem using Alg. 2 with $q^{(l)}$.
11:      **if** $G_{\text{PFCPM}}(q^{(l)}) > C$ **then**
12:         Set $q_U^{(l)} = q_U^{(l-1)}$ and $q_L^{(l)} = q^{(l)}$.
13:      **else if** $G_{\text{PFCPM}}(q^{(l)}) < C$ **then**
14:         Set $q_U^{(l)} = q^{(l)}$ and $q_L^{(l)} = q_L^{(l-1)}$.
15:      **end if**
16:   **until** $G_{\text{PFCPM}}(q^{(l)}) = C$ or $q_U^{(l)} - q_L^{(l)}$ is too small.
17: **end if**

---

2) PFCPM Problem Solution: We now develop an efficient algorithm to solve problem $\mathcal{P}_{\text{PFCPM}}$ based on the concave approximation of the step function. Note that the step function makes the objective function of problem $\mathcal{P}_{\text{PFCPM}}$ nonsmooth, which is difficult to solve. To overcome this challenge, the step function $\delta(x)$ for $x \geq 0$ can be approximated by a suitable concave function. Denote $f_{\text{apx}}^{(k,u)}(p_u^k)$ as the concave penalty function that approximates the step function $\delta(p_u^k)$ corresponding to link (k, u). Then, problem $\mathcal{P}_{\text{PFCPM}}$ can be approximated by the following problem:

$$\min_{\{\mathbf{v}_u^k\},\mathbf{p}} \sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{U}} p_u^k + q \sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{U}} f_{\text{apx}}^{(k,u)}(p_u^k) R_u^{k,\text{fh}}$$

$$\text{s.t. constraints } (1), (4), (5), \text{ and } (8). \quad (14)$$

The objective function of this approximated problem is concave and the feasible region determined by all the constraints is convex. This problem is of the following general form:

$$\min_{\mathbf{x}} g(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{F} \quad (15)$$

where $g(\mathbf{x})$ is concave with respect to $\mathbf{x}$, and $\mathcal{F}$ is the corresponding feasible region. This problem can be solved by the standard gradient method as follows. We start with an initial solution $\mathbf{x}^{(0)} \in \mathcal{F}$ and then iteratively update its solution as

$$\mathbf{x}^{(n+1)} = \arg\min_{\mathbf{x}} g\left(\mathbf{x}^{(n)}\right)$$

$$+ \nabla g\left(\mathbf{x}^{(n)}\right)\left(\mathbf{x} - \mathbf{x}^{(n)}\right) \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{F}. \quad (16)$$

In particular, $\mathbf{x}^{(n+1)}$ needs to be determined from the following problem:

$$\min_{\mathbf{x}} \nabla g\left(\mathbf{x}^{(n)}\right)\mathbf{x} \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{F}. \quad (17)$$

Return to our problem, and let $\mathbf{x}$ represent the precoding vectors and power, and $g(\mathbf{x})$ denote the objective function of problem (14). Then, we have $\nabla g(\mathbf{x})\mathbf{x} = \sum_{(k,u) \in \mathcal{L}}(1 + q\nabla f_{\text{apx}}(p_u^k)R_u^{k,\text{fh}})p_u^k$. By applying the gradient method, we can

solve problem (14) by iteratively solving the following problem until convergence:

$$\min_{\{\mathbf{v}_u^k\}} \quad \sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{U}} \alpha_u^{k(n)} \mathbf{v}_u^{kH} \mathbf{v}_u^k$$

$$\text{s. t.} \quad \text{constraints } (4), (5), \text{ and } (8) \tag{18}$$

where

$$\alpha_u^{k(n)} = 1 + q \nabla f_{\text{apx}}^{(k,u)} \left(p_u^k \right) R_u^{k,\text{fh}}. \tag{19}$$

Problem (18) is a weighted SPMP, which can be transformed into the convex SDP as presented in Appendix A.

3) Algorithm Design: The algorithm to solve (14) is given in Algorithm 2. We have the following proposition that states the convergence property of Algorithm 2.

Algorithm 2 SDP-Based Algorithm for PFCPM Problem

1: Initialization: Set $n = 0$, and $\alpha_u^{k(0)} = 1$ for all RRH–user links (k, u).

2: Iteration n:

   a. Solve problem (18) with $\{\alpha_u^{k(n-1)}\}$ to obtain $(\mathbf{p}^{(n)}, \{\mathbf{v}_u^k\}^{(n)})$.

   b. Update $\{\alpha_u^{k(n)}\}$ as in (19).

3: Set $n := n + 1$, and go back to Step 2 until convergence.

Proposition 2: Algorithm 2 converges to a local optimal solution of problem (14).

   Proof: The proof is given in Appendix C. ∎

Algorithm 1, which is proposed to solve problem $\mathcal{P}_{\text{FCPM}}$ is based on the solution of problem $\mathcal{P}_{\text{PFCPM}}$, which can be obtained by using Algorithm 2.

B. Iterative Linear-Relaxed Algorithm

Here, we propose an iterative linear-relaxed algorithm to directly deal with the step-function fronthaul capacity constraint. Toward this end, we first propose to approximate the step-function by an approximation function. By doing so, the non-continuous constraint with step functions becomes a continuous but still nonconvex one. To convexify the obtained problem, the nonconvex approximated constraint function is further relaxed to a linear form by using the concave duality method [32]. The linear-relaxed constraint function can be made sufficiently close to the original nonconvex function by iteratively updating its parameters. First, problem $\mathcal{P}_{\text{FCPM}}$ can be approximated by the following problem:

$$\min_{\{\mathbf{v}_u^k\}, \mathbf{p}} \quad \|\mathbf{p}\|_{\mathbf{1}} \tag{20}$$

$$\text{s. t.} \quad \text{constraints } (1), (4), (5), \text{ and } (8)$$

$$\sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{U}} f_{\text{apx}}^{(k,u)} \left(p_u^k \right) R_u^{k,\text{fh}} \leq C. \tag{21}$$

As mentioned earlier, the function $f_{\text{apx}}^{(k,u)}(p_u^k)$ is concave with respect to p_u^k; hence, the constraint (21) is indeed nonconvex. Hence, we approximate it by the corresponding linear form based on the duality properties of conjugate of convex functions [32] as follows. First, it can be verified that the function $f_{\text{apx}}^{(k,u)}(p_u^k)$ can be rewritten by using its concave duality as

$$f_{\text{apx}}^{(k,u)} \left(p_u^k \right) = \inf_{z_u^k} \left[z_u^k p_u^k - f_{\text{apx}}^{(k,u)*} \left(z_u^k \right) \right] \tag{22}$$

where $f_{\text{apx}}^{(k,u)*}(z)$ is the conjugate function of $f_{\text{apx}}^{(k,u)}(w)$, which can be expressed as

$$f_{\text{apx}}^{(k,u)*}(z) = \inf_{w} \left[zw - f_{\text{apx}}^{(k,u)}(w) \right]. \tag{23}$$

According to [32], function $f_{\text{apx}}^{(k,u)*}(z)$ can be determined by the optimal value of w obtained from the right-hand side of (23) for a given z. After substituting the results of (23) into (22), it can be verified that the optimization problem in the right-hand side of (22) achieves its minimum at

$$\hat{z}_u^k = \left. \nabla f_{\text{apx}}^{(k,u)}(w) \right|_{w=p_u^k}. \tag{24}$$

With the representation of $f_{\text{apx}}^{(k,u)}(p_u^k)$ as in (22), the constraints (9) can be rewritten in a linear form for a given $\{\hat{z}_u^k\}$ as

$$\sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{U}} \hat{z}_u^k R_u^{k,\text{fh}} p_u^k \leq C + \sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{U}} R_u^{k,\text{fh}} f_{\text{apx}}^{(k,u)*} \left(\hat{z}_u^k \right). \tag{25}$$

In summary, for a given value of $\{\hat{z}_u^k\}$, the problem in (20) and (21) can be reformulated to

$$\min_{\{\mathbf{v}_u^k\}} \quad \sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{U}} \mathbf{v}_u^{kH} \mathbf{v}_u^k \tag{26}$$

$$\text{s. t.} \quad \text{constraints } (4), (5), (8),$$

$$\sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{U}} \hat{z}_u^k R_u^{k,\text{fh}} \mathbf{v}_u^{kH} \mathbf{v}_u^k \leq C + \sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{U}} R_u^{k,\text{fh}} f_{\text{apx}}^{(k,u)*} \left(\hat{z}_u^k \right). \tag{27}$$

Now, the problem in (26) and (27) is the well-known SPMP, which can be solved by transforming it into the SDP, which is described in Appendix A where the additional constraint (27) can be rewritten as (55), shown in Appendix A below. In summary, we can fulfill our design objectives by updating $\{\hat{z}_u^k\}$ iteratively, based on which we repeatedly solve the precoding problem in (26) and (27). This complete procedure is described in Algorithm 3 whose properties are stated in the following proposition.

Algorithm 3 Iterative Linear-Relaxed Algorithm

1: Start with a feasible solution.

2: Set $l = 0$.

3: **repeat**

4:    Calculate $\{\hat{z}_u^{k,(l)}\}$ as in (24) for all (k, u).

5:    Solve problem (26), (27) with $\{\hat{z}_u^{k,(l)}\}$.

6:    Update $l = l + 1$.

7: **until** Convergence.

TABLE I
STEP FUNCTION APPROXIMATION

$f_{\mathrm{apx}}^{(k,u)}(p_u^k)$ - Exponential function	Conjugate function	Condition in (29)	Condition in (30)
$\lambda_u^k \left(1 - e^{-\Psi_u^k p_u^k}\right); \ (\Psi_u^k \gg 1)$	$\dfrac{z_u^k}{\Psi_u^k}\left[1 - \log(\dfrac{z_u^k}{\lambda_u^k \Psi_u^k})\right] - \lambda_u^k$	$\lambda_u^k = \left(1 - e^{-\Psi_u^k P_k}\right)^{-1}$	$1 + e^{-\Psi_u^k P_k} \geq 2e^{-\Psi_u^k \epsilon \beta_u^k}$

Proposition 3: Algorithm 3 has the following properties.

1) If the FCPM problem is feasible, Algorithm 3 converges.
2) The feasible solution achieved by Algorithm 3 at convergence satisfies all constraints of the problem in (20) and (21).

Proof: The proof is given in Appendix D. ∎

Remark 3: Some careful study of $\hat{z}_u^k$ given in (24) yields $f_{\mathrm{apx}}^{(k,u)}(p_u^k) = \hat{z}_u^k p_u^k - f_{\mathrm{apx}}^{(k,u)*}(\hat{z}_u^k)$, which can be considered the first-order approximation (first-order Taylor expansion) of $f_{\mathrm{apx}}^{(k,u)}(w)$ at $w = p_u^k$. In this paper, we prefer using the conjugate function concept in approximating the concave function. This is because the property of the conjugate function, i.e., $f_{\mathrm{apx}}^{(k,u)}(w) \leq zw - f_{\mathrm{apx}}^{(k,u)*}(z)$ for all z, enables us to prove the convergence of Algorithm 3, which is stated in Proposition 3.

C. Adjusting Precoding and Power Solution

After running Algorithms 1 and 3, the obtained solution may have several small but nonzero power elements in the power vector $\mathbf{p}^*$. Since the approximation function is smooth, there may be significant difference between $C(\mathbf{p}^*)$ and $\sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{M}} R_u^{k,\mathrm{fh}} f_{\mathrm{apx}}^{(k,u)}(p_u^{k*})$. To address this problem, we force the power value and precoding vector to zero for any link (k, u) as follows:

$$p_u^{k*} = 0 \ \text{if} \ f_{\mathrm{apx}}^{(k,u)}\left(p_u^{k*}\right) < \frac{1}{2}. \tag{28}$$

We will show how to design the approximation function so that this adjustment only results in tolerable performance degradation in the following.

D. Design of Approximation Function

Here, we discuss the design of an approximation function that is employed to approximate the step function in the fronthaul capacity constraint. Let us denote $\mathbf{p}^*$ as the transmission vector achieved by any of our proposed algorithms at convergence. It can be verified that we have $\sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{M}} R_u^{k,\mathrm{fh}} f_{\mathrm{apx}}^{(k,u)}(p_u^{k*}) = C$ and the SINR of any user u satisfies $\Gamma_u^* = \bar{\gamma}_u$.

First, we want the value of the approximation function to be close to one for large power values so that it well approximates the step function. Specifically, it is required that

$$f_{\mathrm{apx}}^{(k,u)}(P_k) = 1. \tag{29}$$

In addition, we would like the adjusting procedure described earlier to result in the deviation of at most ϵ for the achieved SINR of any user. Let $\Gamma_u^*|_{p_u^{k*}=0}$ denote the SINR achieved by user u after we apply the adjustment procedure given in (28) for the obtained power vector $\mathbf{p}^*$. Then, the results in the following proposition provide the guideline to achieve this design goal.

Proposition 4: If the approximation concave function satisfies the following condition:

$$f_{\mathrm{apx}}^{(k,u)}\left(\epsilon \beta_u^k\right) \geq \frac{1}{2} \tag{30}$$

where $\beta_u^k = \bar{\gamma}_u \sigma^2 / |h_u^k|^2$, then we have

$$\frac{\bar{\gamma}_u - \Gamma_u^*|_{p_u^{k*}=0}}{\bar{\gamma}_u} < \epsilon. \tag{31}$$

Proof: The proof is given in Appendix E. ∎

We provide an example of an exponential approximation function and the required conditions on its parameters in Table I. Note that other choice of the approximation functions is possible as long as it satisfies the conditions stated in Proposition 4.

E. Complexity Analysis

The complexity of the exhaustive search and our proposed algorithms are now investigated based on the number of required computations. As can be observed, all these algorithms require us to solve SPMPs several times (i.e., $\mathcal{P}_\mathfrak{a}$ for exhaustive search algorithm, (18) for the pricing-based algorithm, and (26) for the linear-relaxed algorithm) by transferring them into the SDP problems as described in Appendix A. Hence, we first study the complexity of the SDP program corresponding to each algorithm. We assume that all RRHs are equipped with the same number of antennas N, and we consider the worst case where each UE can be served by all RRHs in the network. Then, the numbers of variables of all SDP problems are the same, which is equal to MNK. On other hand, the numbers of constraints of ($\mathcal{P}_\mathfrak{a}$) and (18) are $M + K$, whereas the corresponding number of (26) is $M + K + 1$. As given in [33], the computational complexity involved in solving the SDP is $\mathcal{O}(\max(m,n)^4 n^{1/2} \log(\zeta_{\mathrm{SDP}}^{-1}))$ where n is the number of variables, m is the number of constraints, and ζ_{SDP} represents the solution accuracy. In practice, the number of users is greater than that of RRHs, and they are larger than 1, i.e., $M > K > 1$; hence, we have $MNK > M + K + 1$. Thus, we can express the complexity of solving problem ($\mathcal{P}_\mathfrak{a}$), (18), and (26) as

$$X_{\mathrm{SDP}} = \mathcal{O}\left(\Pi^{4.5} \log\left(\zeta_{\mathrm{SDP}}^{-1}\right)\right) \tag{32}$$

where $\Pi = MNK$. The complexity of the exhaustive search algorithm and our proposed algorithms can be calculated based on the number of iterations and X_{SDP} as follows. First, let us define $R_{\min}^{\mathrm{fh}} = \min_{u \in \mathcal{U}} R_u^{k,\mathrm{fh}}$ and $\bar{C} = C/R_{\min}^{\mathrm{fh}}$. Then, the number of possible $\mathfrak{a}$ (i.e., the number of elements of set $\mathcal{S}_\mathfrak{a}$) is upper bounded by $\binom{MK}{\bar{C}}$. Hence, the complexity of the exhaustive search method can be expressed as

$$X_{\mathrm{exh}} = \binom{MK}{\bar{C}} \times X_{\mathrm{SDP}} = \mathcal{O}\left(\binom{MK}{\bar{C}} \Pi^{4.5} \log\left(\zeta_{\mathrm{SDP}}^{-1}\right)\right). \tag{33}$$

We now quantify the complexity of solving the FCPM problem using Algorithm 1, which is adopted in Algorithm 2. Note that, in each iteration of Algorithm 2, problem (18) is solved with certain values of $\{\alpha_u^k\}$ and q. For a given value of q, the number of iterations for solving PFCPM problem is $\mathcal{O}(\zeta_2^{-2})$ because the gradient method is applied [42], where ζ_2 represents the solution accuracy of Algorithm 1. In addition, using the bisection searching method, the value of q is typically determined after $\mathcal{O}(\log(\bar{q}\zeta_1^{-1}))$ searching steps where ζ_1 denotes the solution accuracy of Algorithm 1. Hence, the complexity of Algorithm 1 can be calculated as

$$
\begin{aligned}
X_{\mathrm{PBA}} &= \mathcal{O}\left(\zeta_2^{-2}\right) \times \mathcal{O}\left(\log\left(\frac{\bar{q}}{\zeta_1}\right)\right) \times X_{\mathrm{SDP}} \\
&= \mathcal{O}\left(\Pi^{4.5}\zeta_2^{-2}\log\left(\zeta_{\mathrm{SDP}}^{-1}\right)\log\left(\bar{q}\zeta_1^{-1}\right)\right). \quad (34)
\end{aligned}
$$

Finally, if we let I_{LRA} be the number of iterations required in Algorithm 3 to solve the FCPM problem, the complexity of the Algorithm 3 can be given as

$$
X_{\mathrm{LRA}} = I_{\mathrm{LRA}} \times X_{\mathrm{SDP}} = \mathcal{O}\left(I_{\mathrm{LRA}}\Pi^{4.5}\log\left(\zeta_{\mathrm{SDP}}^{-1}\right)\right). \quad (35)
$$

As illustrated in the numerical results, Algorithm 3 converges very fast after around ten iterations for the small system and 30 iterations for the large system. Hence, Algorithm 3 is less complex than Algorithm 1. Moreover, this complexity study also shows that the complexities of our proposed algorithms are much lower than that of the exhaustive searching method.

F. Existing Algorithms

We present two existing algorithms in [23] and [25] to evaluate the relative performance of our proposed algorithms in Section V. These works also attempt to determine the set of RRH–user links for each user, although their considered problems are different from ours. Therefore, these existing algorithms are modified so that they can be applied to solve our considered problem. This is described in the following.

1) Zhao, Quek, and Lei Algorithm [23]: We first describe the greedy algorithm proposed by Zhao, Quek, and Lei (ZQL), which is called *iterative link removal algorithm* in [23]. In this paper, the authors aim to minimize the number of active BS–user links (i.e., fronthaul capacity minimization) subject to the constraints on BSs' maximum transmission power and users' target SINRs. This greedy iterative algorithm works as follows. In each iteration, the power-minimization precoding problem with a specific set of active BS–user links is solved where the authors allow all possible BS–user links initially. If this problem is feasible (i.e., all power and QoS constraints can be supported), then the active BS–user links are sorted in the increasing order of their required transmission power. Then, the set of removed BS–user links is updated before the whole procedure is repeated in the next iteration.

If the power-minimization precoding problem is not feasible in a particular iteration, then the number of removed BS–user links is reduced compared to that in the previous iteration. Otherwise, we increase the number of removed BS–user links. These updates for the number of removed BS–user links follow the bisection method. This iterative procedure is repeated until

the largest number of removed BS–user links can be determined while all constraints can still be maintained.

This greedy link removal principle can be applied to solve our problem as follows. We initially solve the precoding problem for power minimization where all possible RRH–user links are allowed. Then, the required power values of all RRH–user links are sorted in increasing order based on which we remove the minimum number of links with the lowest power to maintain the fronthaul capacity constraint. Finally, we solve the precoding problem again to determine the precoding and power-allocation solution.

2) Dai and Yu Algorithm [25]: The second algorithm is proposed by Dai and Yu (DY) [25], where the authors consider the sum-rate maximization problem subject to the constraints on RRHs' transmission power and fronthaul capacity similar to (38). The main idea of this algorithm is to approximate the step function in the fronthaul constraint by a simple function and solve the corresponding optimization problem iteratively. Specifically, the fronthaul capacity constraint is approximated as follows [25]:

$$
\sum_{k\in\mathcal{K}}\sum_{u\in\mathcal{U}}\beta_u^k R_u^{k,\mathrm{fh}}p_u^k \le C \quad (36)
$$

where β_u^k is a parameter associated with the link between RRH k and user u, which is updated iteratively as

$$
\beta_u^k = \frac{1}{p_u^k + \tau} \quad (37)
$$

where τ is a small number.

This approximation can be employed to solve our problem by iteratively solving the problem in (26) and (27), where the fronthaul capacity constraint is approximated by (36). The limitation of this approximation is that convergence cannot be established. In addition, we will show later in Section V that our proposed algorithms achieve better performance than this algorithm in all investigated simulation settings.

IV. FURTHER EXTENSION

A. Individual Fronthaul Capacity Constraints

Here, we discuss the extension of the considered CoMP transmission design where there are multiple individual fronthaul capacity constraints. In particular, each of these constraints captures the limited capacity of one particular fronthaul link between the CPC and the corresponding RRH. In this scenario, we have to replace the sum fronthaul capacity constraint (9) with the following set of constraints:

$$
G_k(\mathbf{p}^k) = \sum_{u\in\mathcal{M}} R_u^{k,\mathrm{fh}}\delta\left(p_u^{k*}\right) \le C_k \quad \forall k\in\mathcal{K} \quad (38)
$$

where C_k denotes the capacity limit corresponding to the fronthaul link of RRH k. Then, we have the following individual fronthaul constraint power minimization (IFCPM) problem:

$$
\begin{aligned}
(\mathcal{P}_{\mathrm{IFCPM}}) \quad &\min_{\{\mathbf{v}_u^k\},\mathbf{p}} \quad \|\mathbf{p}\|_1 \\
&\text{s. t.} \quad \text{constraints } (1),(4),(5),(8), \text{ and } (38). \quad (39)
\end{aligned}
$$

Obviously, problem $\mathcal{P}_{\text{IFCPM}}$ is more challenging to address than problem $\mathcal{P}_{\text{FCPM}}$. However, we show in the following that the same design principles adopted earlier can be employed to develop low-complexity algorithms in this scenario as well.

1) Pricing-Based Algorithm: To address multiple fronthaul capacity constraints in this case, we have to consider the pricing problem with multiple corresponding pricing parameters q_k whose objective is given as follows:

$$\min_{\{\mathbf{v}_u^k\},\mathbf{p}} \|\mathbf{p}\|_1 + \sum_{k\in\mathcal{K}} q_k G_k(\mathbf{p}^k). \tag{40}$$

For given pricing parameters q_k, $k \in \mathcal{K}$, the pricing problem can also be transformed into the convex SDP form, based on which we can determine the precoding vectors. The key challenge is how to iteratively adjust the pricing parameters q_k to satisfy all fronthaul capacity constraints. This procedure has been presented in our previous conference work [29], which is omitted here due to the space constraint.

2) Iterative Linear-Relaxed Algorithm: We can apply the same linearization technique to the sum fronthaul capacity constraint in Section III-B to deal with each fronthaul capacity constraint in (38) for problem $\mathcal{P}_{\text{IFCPM}}$. In particular, these constraints can be relaxed into the following form:

$$\sum_{u\in\mathcal{U}} \hat{z}_u^k R_u^{k,\text{fh}} \mathbf{v}_u^{kH} \mathbf{v}_u^k \le C_k + \sum_{u\in\mathcal{U}} R_u^{k,\text{fh}} f_{\text{apx}}^{(k,u)*}(\hat{z}_u^k) \quad \forall k \in \mathcal{K}. \tag{41}$$

Then, these constraints can be transformed into the convex SDP form as in (56), shown in Appendix A below.

B. MIMO Systems With Multi-stream Communications

We now discuss how our proposed algorithms for multiple-input–single-output (MISO) systems can be extended for the MIMO systems where multiple data streams can be transmitted to each user. Let us consider a multi-stream communication setting where user u is equipped with T_u antennas and D_u data streams are transmitted to each user u, where $D_u \le \min(T_u, \min_{k\in\mathcal{K}} N_k)$. Denote $\mathbf{x}_u = [x_{u,1}x_{u,2},\ldots,x_{u,D_u}]^T \in \mathbb{C}^{D_u\times 1}$ as the vector describing D_u data streams for user u, where $\mathbb{E}\{\mathbf{x}_u\mathbf{x}_u^H\} = \mathbf{I}_{D_u\times D_u}$. Moreover, let $\mathbf{V}_u^k = [\mathbf{v}_{u,1}^k\mathbf{v}_{u,2}^k,\ldots, \mathbf{v}_{u,D_u}^k] \in \mathbb{C}^{N_k\times D_u}$ denote the precoding matrix at RRH k corresponding to the signal transmitted to user u as in [40], $\mathbf{U}_u = [\mathbf{u}_{u,1}\mathbf{u}_{u,2},\ldots, \mathbf{u}_{u,D_u}] \in \mathbb{C}^{T_u\times D_u}$ be the decoding matrix of user u, and $\mathbf{H}_u^k \in \mathbb{C}^{T_u\times N_k}$ be the channel matrix between RRH k and user u. Then, the baseband signal $\mathbf{y}_u$ received at user u can be written as

$$\mathbf{y}_u = \mathbf{U}_u^H \left(\sum_{i\in\mathcal{U}}\sum_{l\in\mathcal{K}} \mathbf{H}_u^l \mathbf{V}_i^l \mathbf{x}_i + \Xi_u\right) \tag{42}$$

where $\Xi_u \in \mathbb{C}^{T_u\times 1}$ denotes the additive noise vector, and $\mathbb{E}\{\Xi_u\Xi_u^H\} = \sigma^2 \mathbf{I}_{T_u\times T_u}$. Then, the received baseband signal for

stream m of user u can be expressed as

$$y_{u,m} = \sum_{k\in\mathcal{K}}\sum_{t=1}^{D_u} \mathbf{u}_{u,m}^H \mathbf{H}_u^k \mathbf{v}_{u,t}^k x_{u,t}$$
$$+ \sum_{i=1,\neq u}^{M}\sum_{l\in\mathcal{K}}\sum_{t=1}^{M_i} \mathbf{u}_{u,m}^H \mathbf{H}_u^l \mathbf{v}_{i,t}^l x_{i,t} + \mathbf{u}_{u,m}^H \Xi_u. \tag{43}$$

Moreover, the SINR of stream m of user u can be given as

$$\Gamma_{u,m} = \frac{\left|\sum_{k\in\mathcal{K}} \mathbf{u}_{u,m}^H \mathbf{H}_u^k \mathbf{v}_{u,m}^k\right|^2}{\sum_{(i,t)\neq(u,m)}\left|\sum_{l\in\mathcal{K}} \mathbf{u}_{u,m}^H \mathbf{H}_u^l \mathbf{v}_{i,t}^l\right|^2 + \sigma^2|\mathbf{u}_{u,m}|^2}. \tag{44}$$

Assume that each data stream m of every user u has a QoS requirement, which is represented by target SINR $\gamma_{u,m}$. Our design problem for the MIMO system with multi-stream communications ($\mathcal{P}_{\text{MIMO}}$) can be written as

$$\min_{\{\mathbf{V}_u^k\},\{\mathbf{U}_u\}} \sum_{\forall(k,u)} \text{Tr}\left(\mathbf{V}_u^{kH}\mathbf{V}_u^k\right) \tag{45}$$

$$\text{s. t.} \sum_{u\in\mathcal{U}} \text{Tr}\left(\mathbf{V}_u^{kH}\mathbf{V}_u^k\right) \le P_k \quad \forall k \tag{46}$$

$$\Gamma_{u,m} \ge \bar{\gamma}_{u,m} \quad \forall(u,m) \tag{47}$$

$$\text{Tr}\left(\mathbf{V}_u^{kH}\mathbf{V}_u^k\right) = 0 \text{ if } (k,u) \notin \mathcal{L} \tag{48}$$

$$\sum_{\forall(k,u,m)} \delta\left(\mathbf{v}_{u,m}^{kH}\mathbf{v}_{u,m}^k\right) R_{u,m}^{k,\text{fh}} \le C \tag{49}$$

where these constraints capture the power limitation of RRHs, the QoS requirement of each data stream, the clustering constraint, and the limited fronthaul capacity, respectively. Here, $R_{u,m}^{k,\text{fh}}$ represents the required fronthaul capacity of RRH k for transmitting the precoding vector and data stream m of user u. As described in [41], for a given fixed precoding matrix, the optimal decoding matrix based on the minimum-mean-square-error criterion can be expressed as follows:

$$\mathbf{U}_u^\star = \left(\sum_{i\in\mathcal{U}}\sum_{l\in\mathcal{K}} \mathbf{H}_u^l \mathbf{V}_i^l \mathbf{V}_i^{lH} \mathbf{H}_u^{lH} + \sigma^2 \mathbf{I}_{T_u\times T_u}\right)^{-1} \mathbf{H}_u^k \mathbf{V}_u^k. \tag{50}$$

Then, if $\mathbf{U}_u$ is given, we can treat each user u as D_u virtual users $\{u_1, u_2, \ldots, u_{D_u}\}$. Specifically, each virtual user u_m ($m \in \{1, 2, \ldots, D_u\}$) receives a single data stream represented by data symbol $x_{u,m}$, the virtual channel vector between RRH k and user u_m can be expressed as $\mathbf{h}_{u,m}^k = \mathbf{H}_u^{kH}\mathbf{u}_{u,m}$, and the power noise $\sigma_{u,m}^2 = \sigma^2|\mathbf{u}_{u,m}|^2$. Therefore, the multistream design problem can be reformulated by the corresponding virtual single-stream one. Consequently, our proposed algorithms for the single-stream MISO setting can be applied to determine the precoding matrices for this multi-stream MIMO system. Based on these results, we can easily develop an iterative algorithm, which is similar to the one in [41], to solve the multi-stream problem where we alternatively determine the precoding matrices and decoding matrices in each iteration.
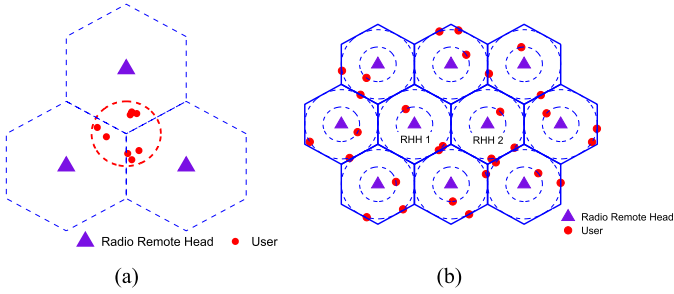
Fig. 2. Network simulation setting. (a) Small network. (b) Large network.

Algorithm 4 MIMO Fronthaul Constraint Algorithm

1: Start with any value of $\mathbf{U}_u^{(0)}$ for all $u \in \mathcal{U}$.
2: Set $l = 0$.
3: **repeat**
4:     Use our proposed algorithm (Algorithm 1 or 3) to find the precoding matrices $\mathbf{V}_u^{k,(l)}$ by determining each precoding vector $\mathbf{v}_{u,m}^{k,(l)}$.
5:     Update the decoding matrices $\mathbf{U}_u^{(l+1)}$ as in (50) for all $u \in \mathcal{U}$.
6:     Update $l = l + 1$.
7: **until** Convergence.

V. NUMERICAL RESULTS

We present numerical results to illustrate the performance achieved by our proposed algorithms. To obtain all numerical results for our proposed algorithms, we employ the exponential approximation function given in Table I. To demonstrate the effectiveness of our design, we consider a small network simulation setting as shown in Fig. 2(a), which allows us to obtain the global optimal solution based on exhaustive search. We then further study the network performance for a larger network setting shown in Fig. 2(b). For ease of exposition, we will present all results using normalized fronthaul capacity, which is defined as $\bar{C} = C/R_{\min}^{\text{fh}}$, where $R_{\min}^{\text{fh}} = \min_{u \in \mathcal{U}} R_u^{k,\text{fh}}$. Furthermore, the initial solution for our iterative linear-relaxed algorithm and DY algorithm [25] is obtained by running Algorithm 2 with the pricing parameter $\bar{q}$.

Unless stated otherwise, the following parameter and simulation setup are adopted for both network settings. The channel gains are generated by considering both Rayleigh fading and path loss, which is modeled as $L_u^k = 36.8 \log_{10}(d_u^k) + 43.8 + 20 \log_{10}(f_c/5)$, where d_u^k is the distance from user u to RRH k; $f_c = 2.5$ GHz. The noise power is set equal to $\sigma^2 = 10^{-13}$ W. We set other parameters as follows: $\epsilon = 10^{-6}$, $\tau = 10^{-8}$, RRH power $P_k = 3$ W for all $k \in \mathcal{K}$, and the number of antennas of each RRH is equal to 4, except for the results in Fig. 6. We denote our proposed algorithms and two existing algorithms as "Pricing-Based Alg.," "Iterative Linear-Relaxed Alg.," "ZQL Alg." [23], and "DY Alg." [25], respectively, in relevant figures.

A. Small Network Simulation Setting

We consider three RRHs in this setting where the distance between their centers is equal to 500 m. Users are randomly placed inside a circle at the center of three RRHs, whose radius
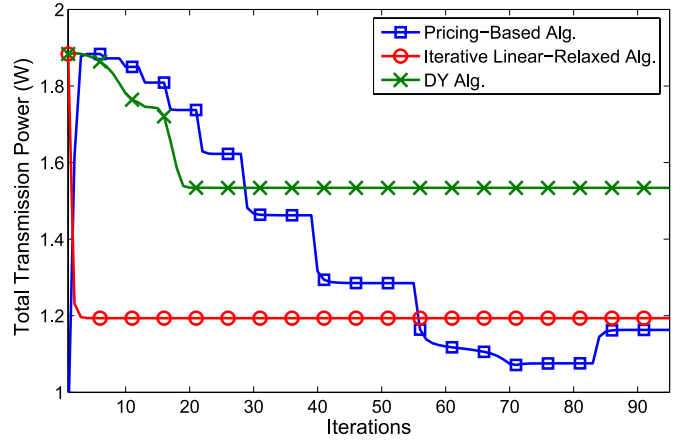

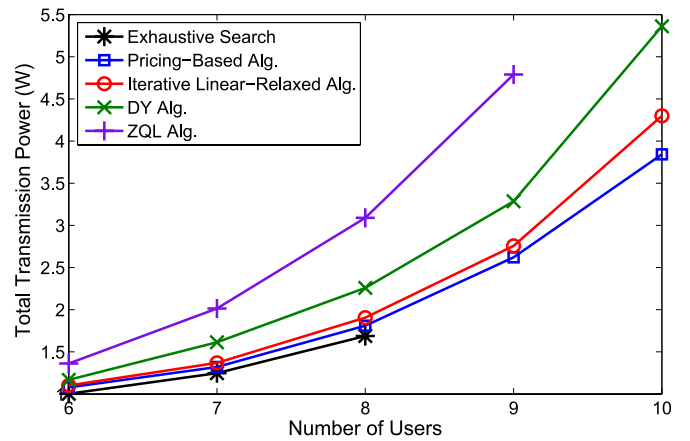Fig. 3. Variations of total power under three algorithms.

Fig. 4. Total power versus number of users in a small network.

is 125 m. First, we examine the convergence of our proposed algorithms and DY algorithm in Fig. 3. Three different curves represent the variations of total transmission power due to the pricing-based algorithm, the iterative linear-relaxed algorithm, and the DY algorithm, respectively. To obtain these simulation results, we set the number of users equal to 7 $(M = 7)$, and user target SINR $\bar{\gamma}_u = 0$ dB for all $u \in \mathcal{M}$. It can be observed that all algorithms converge, although the iterative linear-relaxed algorithm is the fastest and the pricing-based algorithm is the slowest but achieves the lowest transmission power. This figure also illustrates that the DY algorithm converges after 20 iterations; however, the transmission power achieved by this algorithm is much higher than those achieved by ours.

In Fig. 4, we show the total transmission power of all RRHs achieved by the exhaustive searching method, our proposed algorithms, and two reference algorithms, versus the number of users where the user target SINR is $\bar{\gamma}_u = 0$ dB. As can be seen, our proposed algorithms result in lower total transmission power compared to DY and ZQL algorithms. Moreover, the pricing-based algorithm is slightly better than the iterative linear-relaxed algorithm, and both proposed algorithms require marginally higher total transmission power than that due to the optimal exhaustive search algorithm. In addition, the DY algorithm outperforms the ZQL algorithm, but these existing algorithms demand considerably higher power compared with our algorithms. Moreover, this figure shows that the total
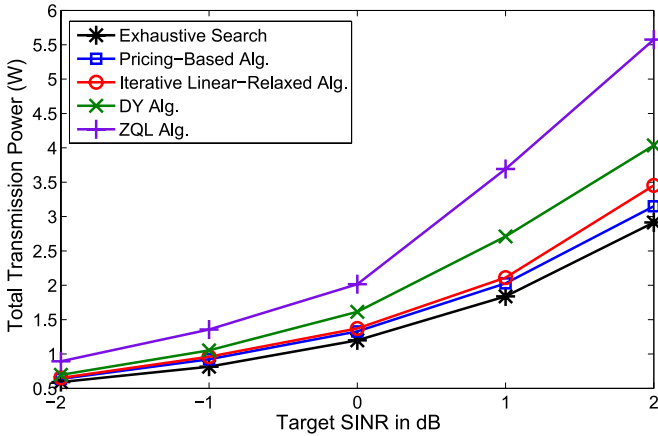
Fig. 5. Total power versus user target SINR in a small network.
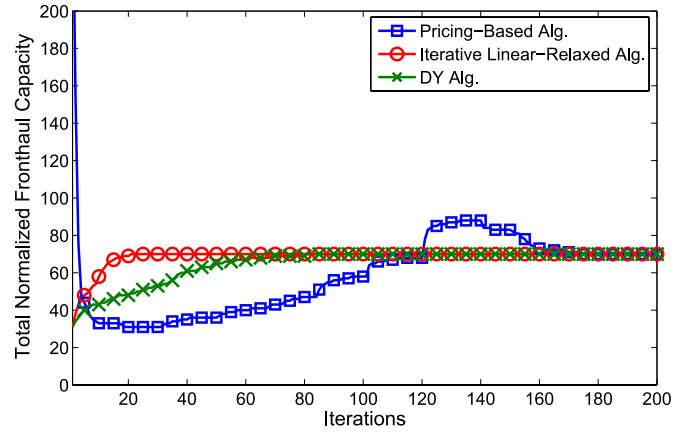

Fig. 7. Variations of utilized fronthaul capacity due to Algorithms 1 and 3 and DY Algorithm over iterations.
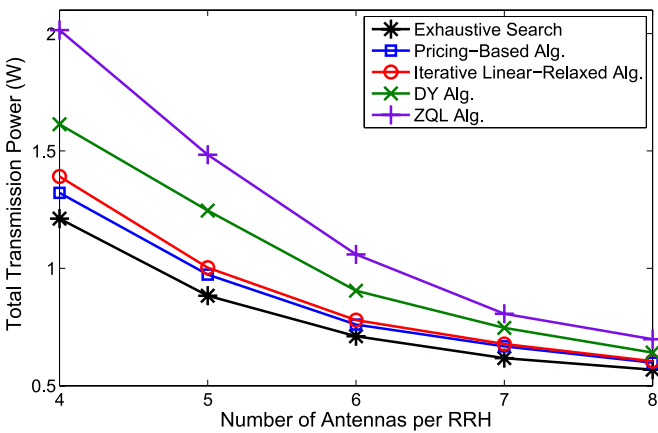

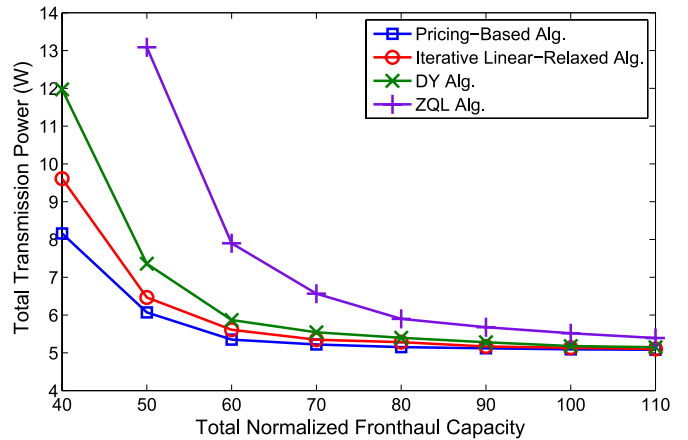Fig. 6. Total power versus number of antennas per RRH in a small network.

Fig. 8. Total power versus total fronthaul capacity.

transmission power increases as the number of users becomes larger under all algorithms, as expected.

Figs. 5 and 6 shows the variations of total transmission power versus user target SINR and the number of antennas per RRH for $M = 7$, respectively. The target SINR of all users for the scheme in Fig. 6 is set as $\bar{\gamma}_u = 0$ dB. As expected, larger power is needed as users require a higher target SINR, and a larger number of equipped antennas result in reduction of total transmission power due to the increasing network degrees of freedom. In addition, our proposed algorithms can achieve the total transmission power that is close to that obtained by the exhaustive search method. Moreover, as can be observed, these figures also confirm the superiority of the proposed algorithms compared to the reference algorithms. Once again, the pricing-based algorithm slightly outperforms the iterative linear-relaxed algorithm.

B. Large Network Simulation Setting

We now study the performance of the proposed algorithms for the large network simulation setting where there are ten hexagonal cells ($K = 10$) and the distance between the centers of two nearest RRHs is 500 m. In this setting, three users are randomly placed inside each cell so that the distance between each of them and their nearest RRH is either 250 or 125 m. Overall, we generate 30 users ($M = 30$) in this network, and

we set $\bar{\gamma}_u = 0$ dB for all users, except for the simulation in Fig. 9. Fig. 7 shows the convergence of two proposed algorithms and DY algorithm where it confirms that our proposed algorithms also perform as expected in this large network setting. Moreover, the pricing-based algorithm requires more convergence time compared with the iterative linear-relaxed algorithm and DY algorithm, although the former achieves smaller transmission power than the latter. Once again, the iterative linear-relaxed algorithm converges faster than the DY algorithm.

Figs. 8 and 9 show the total transmission power versus the normalized fronthaul capacity and users' target SINR, respectively. Our proposed algorithms again outperform the other existing algorithms. The iterative linear-relaxed algorithm results in the total transmission power, which is slightly higher than that due to the pricing-based algorithm. Moreover, as shown in Fig. 8, the performance gap between these algorithms are more significant if the available fronthaul capacity is smaller. Furthermore, the three algorithms, except the ZQL algorithm, achieve similar performance for sufficiently large fronthaul capacity or small target SINR values.

In Fig. 10, we present the variations of total transmission power versus the cluster size. Here, the cluster size is equal to the number of nearest RRHs, which are allowed to serve each user. Specifically, the cluster size of m means that each user can
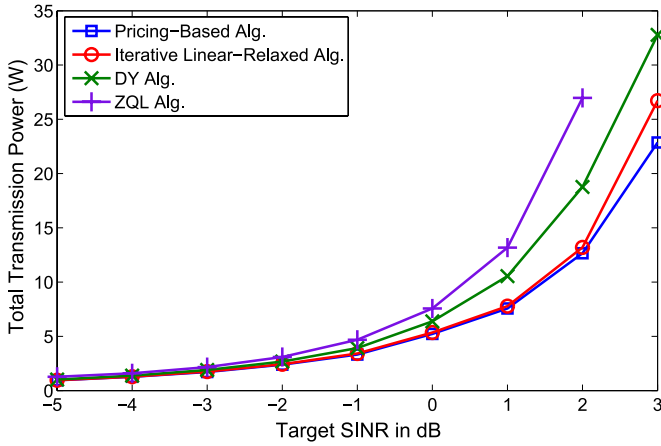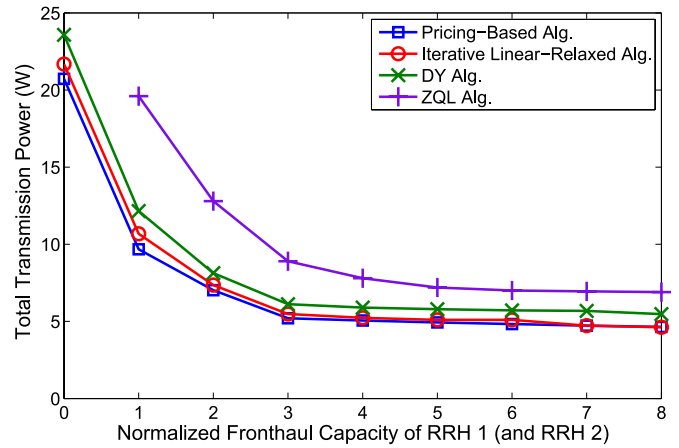
Fig. 9. Total power versus user target SINR.

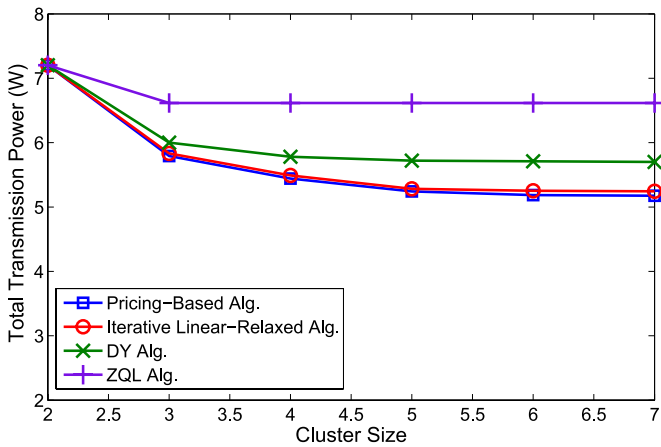Fig. 11. Total power versus fronthaul capacity per RRH 1/2.

Fig. 10. Total power versus cluster size.

Fig. 12. Total power versus the number of data streams per user.

be served by, at most, m nearest RRHs. In this simulation, we set $\bar{C} = 70$ and the target SINR $\bar{\gamma} = 0$ dB. As can be seen, the larger cluster size results in the lower total transmission power. However, total transmission power achieved by all algorithms become flat as the cluster size is sufficiently large. These results imply that activating weak links between users and RRHs does not significantly improve the network performance. This figure again confirms that our proposed algorithms significantly outperform other existing algorithms, and the pricing-based algorithm is slightly better than iterative linear-relaxed algorithm.

C. Individual Fronthaul Capacity Constraints

Fig. 11 shows variations of total transmission power achieved by our proposed algorithms and two reference algorithms as there are multiple individual fronthaul capacity constraints. In this simulation, we keep the normalized fronthaul capacity $\bar{C}_k = 8$ for all $k > 2$ while varying the normalized capacity of fronthaul links corresponding RRHs 1 and 2, which is denoted $\bar{C}_{1-2}$ (i.e., $\bar{C}_1 = \bar{C}_2 = \bar{C}_{1-2}$). Again, the individual normalized fronthaul capacity is defined as $\bar{C}_k = C_k / R_{\min}^{\text{fh}}$. In fact, if we set $\bar{C}_{1-2} = 0$, then RRHs 1 and 2 are turned off, which cannot serve any user in the network. As can be observed, the total transmission power decreases with the increase in $\bar{C}_{1-2}$, which shows a similar trend as in Fig. 8. Interestingly, all users in cells 1 and 2 can still be supported at their target SINR by running pricing-based, iterative linear-relaxed, and DY algorithms,
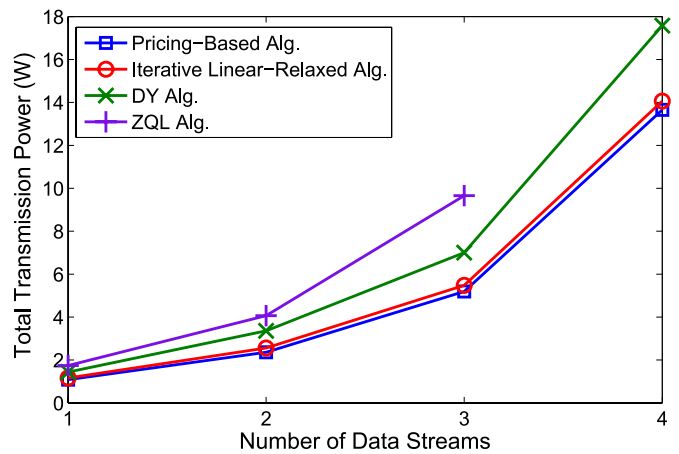
even when RRHs 1 and 2 are turned off, which is not the case with the ZQL algorithm. Overall, our proposed algorithms still outperform the two existing algorithms.

D. MIMO Systems With Multi-stream Communications

For this simulation, we consider the large network, but we allocate only two users in each cell to ensure the feasibility of the network. Each RRH is equipped with eight antennas and each user has four antennas. The total normalized fronthaul capacity is set at 100, and the target SINR for each stream is 0 dB. Fig. 12 shows the total transmission power of all RRHs versus the number of data streams for each users. As shown, the proposed algorithms still work well for the MIMO multi-stream and achieve better solution than that of the algorithm in the literature. Interestingly, the higher number of streams per user requires the higher transmission power of all RRHs.

VI. CONCLUSION

We have proposed efficient and low-complexity algorithms to solve the downlink joint transmission problem in Cloud-RAN that aims to minimize the total transmission power subject to constraints on transmission power, fronthaul capacity, and users' QoS. We have considered both scenarios with sum and individual fronthaul capacity constraints. Numerical results have illustrated the efficacy of our proposed algorithms and the impacts

of different parameters on the network performance. In particular, the pricing-based and iterative linear-relaxed algorithms achieve the performance very close to that due to the optimal exhaustive search. In addition, our proposed algorithms significantly outperform two existing algorithms proposed in [23] and [25] in all investigated scenarios. There are still several interesting directions for future study. First, joint design and optimization of CoMP techniques and fronthaul signal compression is an interesting problem. Furthermore, consideration of CoMP design, CSI estimation, and feedback deserves further effort.

APPENDIX A
WEIGHTED SUM-POWER MINIMIZATION SOLUTION

Denote $\mathcal{K}_u$ as the set of RRHs serving user u, i.e., $\mathcal{K}_u = \{k | (k, u) \in \mathcal{L}\}$. Let $\mathbf{v}_u$ denote the precoding solution over all RRHs in $\mathcal{K}_u$, which is defined as $\mathbf{v}_u = [\mathbf{v}_u^{u_1 T}, \mathbf{v}_u^{u_2 T}, \ldots, \mathbf{v}_u^{u_{a_u} T}]^T$, where $\{u_1, \ldots, u_{a_u}\} = \mathcal{K}_u$, and $a_u = |\mathcal{K}_u|$. Therefore, we have $\mathbf{v}_u \in \mathbb{C}^{N_u \times 1}$, where $N_u = \sum_{k \in \mathcal{K}_u} N_k$. Here, we are interested in determining $\mathbf{v}_u$ because we have $\mathbf{v}_u^i = \mathbf{0}$ for all $i \notin \mathcal{K}_u$ (i.e., the precoding vector of any RRH that does not serve the underlying user is equal to zero).

Let us define $\mathbf{W}_u = \mathbf{v}_u \mathbf{v}_u^H$, we have $\mathbf{W}_u \in \mathbb{C}^{N_u \times N_u}$. It is positive semi-definite ($\mathbf{W}_u \succeq \mathbf{0}$) and has rank one because it is generated from vector $\mathbf{v}_u$. We also define the channel vector $\mathbf{h}_{u,i} = [\mathbf{h}_u^{i_1 T}, \mathbf{h}_u^{i_2 T}, \ldots, \mathbf{h}_u^{i_{a_i} T}]^T$ and $\mathbf{H}_{u,i} = \mathbf{h}_{u,i} \mathbf{h}_{u,i}^H$ for all $u, i \in \mathcal{U}$. Then, the SINR constraint for user u in (4) and the power constraint for RRHs can be rewritten as

$$\mathrm{Tr}(\mathbf{H}_{u,u} \mathbf{W}_u) - \bar{\gamma}_u \sum_{i \in \mathcal{U}/u} \mathrm{Tr}(\mathbf{H}_{u,i} \mathbf{W}_i) \geq \bar{\gamma}_u \sigma^2 \quad \forall u \in \mathcal{U} \quad (51)$$

$$\sum_{u \in \mathcal{U}} \mathrm{Tr}\left(\mathbf{E}_u^k \mathbf{W}_u\right) \leq P_k \quad \forall k \in \mathcal{K} \quad (52)$$

where $\mathbf{E}_u^k = \mathrm{diag}(\mathbf{0}_{N_{u_1} \times 1}, \ldots, \mathbf{1}_{N_{u_i} \times 1}, \ldots, \mathbf{0}_{N_{u_{a_u}} \times 1})$ if $u_i = k$. Therefore, the weighted sum-power minimization can be formulated as the following SDP problem:

$$\min_{\{\mathbf{W}_u\}_{u=1}^M} \quad \sum_{u=1}^M \mathrm{Tr}(\mathbf{F}_u \mathbf{W}_u) \quad (53)$$

$$\text{s.t.} \quad \text{constraints } (51), (52)$$

$$\mathbf{W}_u \succeq \mathbf{0}, \mathrm{rank}(\mathbf{W}_u) = 1 \quad \forall u \quad (54)$$

where $\mathbf{F}_u = \mathrm{diag}(\alpha_u^{u_1}(n) \mathbf{1}_{N_{u_1} \times 1}, \ldots, \alpha_u^{u_{a_u}}(n) \mathbf{1}_{N_{u_{a_u}} \times 1})$. This transformation reveals a special structure of the precoding design problem. Specifically, if we remove the rank-one constraints in (54) from (53), then the new problem is convex. In fact, this relaxed problem is the convex SDP, which can be therefore solved easily by using standard tools such as the CVX solver [31]. As given in [21, Th. 3.1] and [36, Lemma 2], if (53) is feasible, then it has at least one solution, where $\mathrm{rank}(\mathbf{W}_u) = 1$ for all $u \in \mathcal{U}$.

Because the relaxed SDP problem is convex, $\mathbf{v}_u$ can be calculated as the eigenvector of $\mathbf{W}_u$ if the optimum solution is unique. Such a unique optimum solution satisfies the rank-one constraint. As discussed in [21] and [36], the relaxed SDP may have more than one optimum solution, which means that the CVX solver cannot ensure to return the rank-one solution in

general. However, this situation almost surely never happens in practice, except for cases where the channels from two groups are exactly symmetric. Nevertheless, if the algorithm does produce one of these solutions where $\mathbf{W}_u$ does not have rank one, we can still obtain a rank-one optimum solution from that solution by using the method described in [36, Lemma 5], by applying the solution method for solving the rank-constrained problem presented in [37, Alg. 2], or by utilizing the best rank-one approximation based on the largest eigenvalue and the corresponding eigenvector, as discussed in [33, Sec. II].

In addition, this SDP-based transformation can be applied to solve the problem (26), (27), and (39), where the additional linear-relaxed fronthaul constraints (27) and (41) can also be rewritten, respectively, into the SDP form as follows:

$$\sum_{u \in \mathcal{U}} R_u^{k,\mathrm{fh}} \mathrm{Tr}(\mathbf{Z}_u \mathbf{W}_u) \leq C + \sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{U}} R_u^{k,\mathrm{fh}} f_{\mathrm{apx}}^{(k,u)*}\left(\hat{z}_u^k\right) \quad (55)$$

$$\sum_{u \in \mathcal{U}} R_u^{k,\mathrm{fh}} \mathrm{Tr}\left(\mathbf{Z}_u^k \mathbf{W}_u\right) \leq C_k + \sum_{u \in \mathcal{U}} R_u^{k,\mathrm{fh}} f_{\mathrm{apx}}^{(k,u)*}\left(\hat{z}_u^k\right) \quad (56)$$

where $\mathbf{Z}_u = \mathrm{diag}(z_u^{u_1} \mathbf{1}_{N_{u_1} \times 1}, \ldots, z_u^{u_{a_u}} \mathbf{1}_{N_{u_{a_u}} \times 1})$, and $\mathbf{Z}_u^k = \mathrm{diag}(\mathbf{0}_{N_{u_1} \times 1}, \ldots, z_u^{u_i} \mathbf{1}_{N_{u_i} \times 1}, \ldots, \mathbf{0}_{N_{u_{a_u}} \times 1})$ if $u_i = k$.

APPENDIX B
PROOF OF PROPOSITION 1

A. Proof of Statement 1

Let us consider two values of the pricing parameter q and q', where $q' > q$. Let $(\mathbf{p}, \{\mathbf{v}_u^k\})$ and $(\mathbf{p}', \{\mathbf{v}_u^{k'}\})$ be the solutions of the PFCPM problem with q and q', respectively. Then, we have

$$\sum_{\forall (u,j)} p_u^j + q G_{\mathrm{PFCPM}}(q) \leq \sum_{\forall (u,j)} p_u^{j'} + q G_{\mathrm{PFCPM}}(q') \quad (57)$$

$$\sum_{\forall (u,j)} p_u^{j'} + q' G_{\mathrm{PFCPM}}(q') \leq \sum_{\forall (u,j)} p_u^j + q' G_{\mathrm{PFCPM}}(q). \quad (58)$$

After combining and simplifying these two inequalities, we have $(q' - q) G_{\mathrm{PFCPM}}(q') \leq (q' - q) G_{\mathrm{PFCPM}}(q)$. Therefore, we have finished the proof for the first statement.

B. Proof of Statement 2

We prove that, for any $(\mathbf{p}, \{\mathbf{v}_u^k\})$ satisfying constraints (1), (4), and (8), we always have $G(\mathbf{p}) \geq G_{\mathrm{PFCPM}}(q)$ if $q \geq \bar{q}$. Let $P^{\mathrm{opt}}(q)$ be the total power of the PFCPM problem corresponding to q. Because $(\mathbf{p}, \{\mathbf{v}_u^k\})$ satisfy constraints (1), (4), and (8), we have

$$P^{\mathrm{opt}}(q) + q G_{\mathrm{PFCPM}}(q) \leq \sum_{(k,u) \in \mathcal{L}} p_u^k + q G(\mathbf{p}) \quad (59)$$

$$\to G_{\mathrm{PFCPM}}(q) - G(\mathbf{p}) \leq \frac{\sum_{(k,u) \in \mathcal{L}} p_u^k - P^{\mathrm{opt}}(q)}{q} < \frac{\sum_{k \in \mathcal{K}} P_k}{\bar{q}}. \quad (60)$$

Thus, we have $G_{\mathrm{PFCPM}}(q) - G(\mathbf{p}) < \sigma_{\min}$ if $q \geq \bar{q}$. Based on the definition of $\sigma_{\min}$, we must have that $G(\mathbf{p})$ cannot be smaller than $G_{\mathrm{PFCPM}}(q)$.

C. Proof of Statement 3

From the proof of statement 2, we can see that there exists no $(\mathbf{p}, \{\mathbf{v}_u^k\})$ satisfying constraints (1), (8), and (4) so that $G(\mathbf{p}) \leq C < G_{\text{PFCPM}}(q)$. Hence, the last statement of Proposition 1 is proved.

APPENDIX C
PROOF OF PROPOSITION 2

To prove the convergence of Algorithm 2, we will prove that the value of the objective function of problem (14) (which is represented by $g(\mathbf{x})$) decreases after each step of Algorithm 2. For any vector $\mathbf{x}$ representing all precoding vectors and power, the following inequality holds because of the concavity of function $g(\mathbf{x})$

$$g(\mathbf{x}) \leq g\left(\mathbf{x}^{(n)}\right) + \nabla g\left(\mathbf{x}^{(n)}\right)\left(\mathbf{x} - \mathbf{x}^{(n)}\right). \quad (61)$$

In addition, the solution achieved in the $(n + 1)$th iteration is

$$\mathbf{x}^{(n+1)} = \arg\min \nabla g(\mathbf{x}^{(n)})\mathbf{x} \text{ s.t. constraints (4), (5), (8).}$$

Hence, we have $\nabla g(\mathbf{x}^{(n)}\mathbf{x}^{(n+1)} \leq \nabla g(\mathbf{x}^{(n)})\mathbf{x}^{(n)}$. Substituting $\mathbf{x}^{(n+1)}$ into the inequality (61) yields

$$g\left(\mathbf{x}^{(n+1)}\right) \leq g\left(\mathbf{x}^{(n)}\right) + \nabla g\left(\mathbf{x}^{(n)}\right)\left(\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\right) \leq g\left(\mathbf{x}^{(n)}\right). \quad (62)$$

Thus, the objective function of problem (14) corresponding to the solution in each iteration of Algorithm 2 is monotonically decreasing. Therefore, Algorithm 2 must converge to a local optimum point.

APPENDIX D
PROOF OF PROPOSITION 3

A. Proof of Statement 1

Denote Ω_l as the optimum objective value of problem (26) and (27) in iteration l, which is obtained after performing steps 8–12. We will prove that Ω_l decreases over each iteration; hence, our proposed algorithm converges. Denote $\mathcal{F}_l$ as the feasible set of (26) and (27) in iteration l, which corresponds to the value of $\{\hat{z}_u^{k,(l)}\}$. Let us define $(\mathbf{p}^{(l)}, \{\mathbf{v}_u^{k,(l)}\})$ as the optimal solution of the problem in (26) and (27) corresponding to $\{\hat{z}_u^{k,(l)}\}$ as we run steps 8–12 in Algorithm 3. By setting $\{\hat{z}_u^{k,(l+1)}\}$ as in (24), we always have $(\mathbf{p}^{(l)}, \{\mathbf{v}_u^{k,(l)}\}) \in \mathcal{F}_{l+1}$. Hence, we must have

$$\Omega_l \geq \Omega_{l+1} \quad \forall l > 0 \quad (63)$$

which completes the proof.

B. Proof of Statement 2

Denote $\mathcal{F}$ as the feasible set of problems (20) and (21). Because of (22), if $\sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{U}} \hat{z}_u^k R_u^{k,\text{fh}} p_u^k \leq C + \sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{U}} R_u^{k,\text{fh}} f_{\text{apx}}^{(k,u)*}(\hat{z}_u^k)$, we have $\sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{U}} R_u^{k,\text{fh}} f_{\text{apx}}^{(k,u)}(p_u^k) \leq C$ for any value of $\{\hat{z}_u^{k,(l+1)}\}$. Therefore, we have $\mathcal{F}_l \subseteq \mathcal{F}$ for any

iteration l, which means the optimal solution of problem (26) and (27) in any iteration l satisfies all constraints of the problem in (20) and (21). Hence, Algorithm 3 returns the solution that satisfies all constraints of the problem in (20) and (21).

APPENDIX E
PROOF OF PROPOSITION 4

Let $\mathbf{p}^*$ and $\{\mathbf{v}_u^{k*}\}$ be the power vector and precoding vectors achieved after running our proposed algorithms. Because these results are obtained by solving a weighted sum-power minimization with an SINR constraint for each user, we have

$$\bar{\gamma}_u = \frac{\left| \sum_{j \in \mathcal{K}} \mathbf{h}_u^{jH} \mathbf{v}_u^{j*} \right|^2}{\sum_{i=1,\neq u}^{M} \left| \sum_{l \in \mathcal{K}} \mathbf{h}_u^{lH} \mathbf{v}_i^{l*} \right|^2 + \sigma^2}. \quad (64)$$

Applying the triangle inequality yields

$$\frac{\left| \sum_{j \in \mathcal{K}} \mathbf{h}_u^{jH} \mathbf{v}_u^{j*} \right|^2}{\sum_{i=1,\neq u}^{M} \left| \sum_{l \in \mathcal{K}} \mathbf{h}_u^{lH} \mathbf{v}_i^{l*} \right|^2 + \sigma^2} \leq \frac{\left| \sum_{j \in \mathcal{K}/k} \mathbf{h}_u^{jH} \mathbf{v}_u^{j*} \right|^2 + \left| \mathbf{h}_u^{kH} \mathbf{v}_u^{k*} \right|^2}{\sum_{i=1,\neq u}^{M} \left| \sum_{l \in \mathcal{K}} \mathbf{h}_u^{lH} \mathbf{v}_i^{l*} \right|^2 + \sigma^2}. \quad (65)$$

Now, if we suppose that the power of user u and RRH k is forced to be zero (i.e., $p_u^{k*} = 0$), then its updated SINR becomes

$$\Gamma_u^*|_{p_u^{k*}=0} = \frac{\left| \sum_{j \in \mathcal{K}/k} \mathbf{h}_u^{jH} \mathbf{v}_u^{j*} \right|^2}{\sum_{i=1,\neq u}^{M} \left| \sum_{l \in \mathcal{K}} \mathbf{h}_u^{lH} \mathbf{v}_i^{l*} \right|^2 + \sigma^2}. \quad (66)$$

Consider the contribution of the underlying power p_u^{k*}, we have

$$\frac{\left| \mathbf{h}_u^{kH} \mathbf{v}_u^{k*} \right|^2}{\sum_{i=1,\neq u}^{M} \left| \sum_{l \in \mathcal{K}} \mathbf{h}_u^{lH} \mathbf{v}_i^{l*} \right|^2 + \sigma^2} \leq \frac{\left| \mathbf{h}_u^{k} \right|^2 \left| \mathbf{v}_u^{k*} \right|^2}{\sigma^2} = \frac{\left| \mathbf{h}_u^{k} \right|^2 p_u^{k*}}{\sigma^2}. \quad (67)$$

Combining all these results in (64)–(67) yields

$$\frac{\bar{\gamma}_u - \Gamma_u^*|_{p_u^{k*}=0}}{\bar{\gamma}_u} \leq \frac{\left| \mathbf{h}_u^{k} \right|^2 p_u^{k*}}{\bar{\gamma}_u \sigma^2}. \quad (68)$$

Because we have assumed that p_u^{k*} is forced to be zero as in (28) and $f_{\text{apx}}^{(k,u)}(x)$ is an increasing function, we have $p_u^{k*} < f_{\text{apx}}^{(k,u)-1}(1/2)$, where $f_{\text{apx}}^{(k,u)-1}(x)$ is the inverse function of $f_{\text{apx}}^{(k,u)}(x)$. Therefore, if $f_{\text{apx}}^{(k,u)}(x)$ satisfies (30), which implies $f_{\text{apx}}^{(k,u)-1}(1/2) \leq \epsilon\beta_u^k = (\epsilon\bar{\gamma}_u\sigma^2/|\mathbf{h}_u^k|^2)$, we have

$$\frac{\bar{\gamma}_u - \Gamma_u^*|_{p_u^{k*}=0}}{\bar{\gamma}_u} < \frac{\left| \mathbf{h}_u^{k} \right|^2}{\bar{\gamma}_u \sigma^2} \frac{\epsilon\bar{\gamma}_u\sigma^2}{\left| \mathbf{h}_u^{k} \right|^2} = \epsilon. \quad (69)$$

This completes the proof of Proposition 4.

References

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018, White Paper, Cisco, San Jose, CA, USA, Feb. 2014.

[2] "Mobility report," Ericsson, Stockholm, Sweden, Tech. Rep., Nov. 2013.

[3] E. Hossain, L. B. Le, and D. Niyato, *Radio Resource Management in Multi-Tier Cellular Wireless Networks*. New York, NY, USA: Wiley, 2013.

[4] N. Saquib, E. Hossain, L. B. Le, and D. I. Kim, "Interference management in OFDMA femtocell networks: Issues and approaches," *IEEE Wireless Commun.*, vol. 19, no. 3, pp. 86–95, Jun. 2012.

[5] L. B. Le *et al.*, "Enabling 5G mobile wireless technologies," *EURASIP J. Wireless Commun. Netw.*, vol. 2015, no. 1, p. 218, Sep. 2015.

[6] 4G Mobile Broadband Evolution 3GPP Release 10 and Beyond, 4G Americas, Bellevue, WA, USA, Jan. 2012.

[7] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno, "Coordinated multipoint transmission/reception techniques for LTE-advanced [coordinated and distributed MIMO]," *IEEE Wireless Commun.*, vol. 17, no. 3, pp. 26–34, Jun. 2010.

[8] R. Irmer *et al.*, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[9] P. Marsch and G. Fettweis, "Uplink CoMP under a constrained backhaul and imperfect channel knowledge," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1730–1742, Jun. 2011.

[10] C-RAN: The Road Towards Green RAN, White Paper, China Mobile, Hong Kong, 2011.

[11] Suggestion on Potential Solution to C-Ran, NGMN Alliance, Frankfurt, Germany, Jan. 2013.

[12] A. Checko, H. Christiansen, and M. S. Berger, "Evaluation of energy and cost savings in mobile cloud-RAN," in *Proc. OPNETWORK Conf.*, 2013, pp. 1–7.

[13] J. Wu, "Green wireless communications: From concept to reality [industry perspectives]," *IEEE Wireless Commun.* vol. 19, no. 4, pp. 4–5, Aug. 2012.

[14] C. Liu, K. Sundaresan, M. Jiang, S. Rangarajan, and G.-K. Chang, "The case for re-configurable backhaul in cloud-RAN based small cell networks," in *Proc. IEEE INFOCOM*, 2013, pp. 1124–1132.

[15] C. Fan, Y. J. Zhang, and X. Yuan, "Dynamic nested clustering for parallel PHY-layer processing in cloud-RANs," 2014. [Online]. Available: http://arxiv.org/abs/1408.0876

[16] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai (Shitz), "Robust and efficient distributed compression for cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 692–703, Feb. 2013.

[17] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai (Shitz), "Robust layered transmission and compression for distributed uplink reception in cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 204–215, Jan. 2014.

[18] A. Liu and V. K. N. Lau, "Joint power and antenna selection optimization in large cloud radio access networks," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1319–1328, Mar. 2014.

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

[20] S. Luo, R. Zhang, and T. J. Lim, "Downlink and uplink energy minimization through user association and beamforming in cloud RAN," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 494–508, Jan. 2015.

[21] M. Bengtsson and B. Ottersten, "Optimal downlink beamforming using semidefinite optimization," in *Proc. Annual Allerton Conf. Commun., Control Comput.*, Monticello, NY, USA, Sep. 1999, pp. 987–996.

[22] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1426–1438, Aug. 2006.

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

[24] Y. Cheng, M. Pesavento, and A. Philipp, "Joint network optimization and downlink beamforming for CoMP transmissions using mixed integer conic programming," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3972–3987, Aug. 2013.

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

[26] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[27] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[28] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Found. Trends Mach. Learn.*, vol. 4, no. 1, pp. 1–106, Jan. 2012.

[29] V. N. Ha, L. B. Le, and N.-D. Đào, "Energy-efficient coordinated transmission for Cloud-RANs algorithm design and tradeoff," in *Proc. IEEE CISS*, Princeton, NJ, USA, Mar. 2014, pp. 1–6.

[30] V. N. Ha and L. B. Le, "Joint coordinated beamforming and admission control for fronthaul constrained cloud-RANs," in *Proc. IEEE GLOBECOM*, Dec. 2014, pp. 4054–4059.

[31] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming," Jun. 2015. [Online]. Available: http://cvxr.com/cvx/

[32] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1970.

[33] Z. Q. Luo, W. K. Ma, A. M. C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problem," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[34] N. Lee, D. M. Jimenez, A. Lozano, and R. W. Heath, "Spectral efficiency of dynamic coordinated beamforming: A stochastic geometry approach," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 230–241, Jan. 2015.

[35] J. Gong, S. Zhou, Z. Niu, L. Geng, and M. Zheng, "Joint scheduling and dynamic clustering in downlink cellular networks," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2011, pp. 1–5.

[36] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," in *Handbook of Antennas in Wireless Communications*, L. C. Godara, Ed. Boca Raton, FL, USA, CRC, Aug. 2001, ch. 18, pp. 18-1–18.33.

[37] Y. Huang and D. P. Palomar, "Rank-constrained separable semidefinite programming with applications to optimal beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp 664–678, Feb. 2010.

[38] S. Park, C.-B. Chae, and S. Bahk, "Before/after precoded massive MIMO in cloud radio access networks," *Proc. IEEE Int. Conf. Commun.*, Jun. 2013, pp. 169–173.

[39] J. Kang, O. Simeone, J. Kang, and S. Shamai, "Fronthaul compression and precoding design for C-RANs over ergodic fading channel," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5022–5032, Jul. 2016.

[40] L. Choi and R. D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 773–786, Jul. 2003.

[41] S. W. Peters and R. W. Heath, Jr., "Cooperative algorithms for MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 206–218, Jan. 2011.

[42] C. Cartis, N. I. M. Gould, and P. H. L. Toint, "On the complexity of steepest descent, Newton's and regularized Newton's methods for nonconvex unconstrained optimization problems," *SIAM J. Optim.*, vol. 20, no. 6, pp. 2833–2852, Oct. 2010.

Vu Nguyen Ha (S'11) received the B.Eng. degree from Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, and the Addendum degree from de École Nationale Supérieure des Télécommunications de Bretagne, Plouzané, France, in 2007. He is currently working toward the Ph.D. degree with the Institut National de la Recherche ScientifiqueÉnergie, Matériaux et élécommunications (INRS-EMT), Université du Québec, Montréal, QC, Canada.

From 2008 to 2011, he was a Research Assistant with the School of Electrical Engineering, University of Ulsan, Ulsan, Korea. His research interests include radio resource management and emerging enabling technologies for fifth-generation wireless systems, with special emphasis on heterogeneous small-cell networks, cloud radio access networks, and multiple-input–multiple-output communications.

Long Bao Le (S'04–M'07–SM'12) received the B.Eng. degree in electrical engineering from Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, in 1999; the M.Eng. degree in telecommunications from Asian Institute of Technology, Pathumthani, Thailand, in 2002; and the Ph.D. degree in electrical engineering from the University of Manitoba, Winnipeg, MB, Canada, in 2007.

From 2007 to 2008, he was a Postdoctoral Researcher with the University of Waterloo, Waterloo, ON, Canada, and, from 2008 to 2010, with Massachusetts Institute of Technology, Cambridge, MA, USA. Since 2010, he has been with the Institut National de la Recherche Scientifique, Université du Québec, Montréal, QC, Canada, where he is currently an Associate Professor. He is a coauthor of the book *Radio Resource Management in Multi-Tier Cellular Wireless Networks* (Wiley, 2013). His current research interests include smart grids, cognitive radio, radio resource management, network control and optimization, and emerging enabling technologies for fifth-generation wireless systems.

Dr. Le serves as a member of the editorial board of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, and IEEE WIRELESS COMMUNICATIONS LETTERS. He has served as the Technical Program Committee Chair/Co-Chair for several IEEE conferences, including the IEEE Wireless Communications and Networking Conference, the IEEE Vehicular Technology Conference, and the IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications.

Ngọc-Dũng Đào (S'02–M'07) received the Ph.D. degree from the University of Alberta, Edmonton, AB, Canada, in 2006.

From 1995 to 2000, he has been with Siemens Communications Networks, Vietnam, and, from 2007 to 2010, with Toshiba Research Europe, U.K. He is currently a Senior Research Engineer with Huawei Technologies Canada Company, Ltd, Ottawa, ON, Canada. His recent research interests include software-defined networks and network function virtualization for radio access networks, information-centric networking, heterogeneous and dense networks, and reliable video communications for fifth-generation (5G) mobile networks.

Dr. Đào serves as an Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, as an Associate Technical Editor for IEEE COMMUNICATIONS MAGAZINE, and as a Guest Editor for the 5G special issue of the *EURASIP Journal on Wireless Communications and Networking*.