

Computation Capacity Constrained Joint Transmission Design for C-RANs

Vu Nguyen Ha and Long Bao Le

INRS, University of Quebec, Montreal, QC, Canada; e-mails: {hanguyen,long.le}@emt.inrs.ca.

Abstract—This paper considers the joint processing design for the cloud radio access network (C-RAN) with limited cloud computation capacity. This amounts to determine the set of remote radio heads (RRHs) serving each user and the corresponding precoding vectors whose corresponding computation effort (CE) is a non-linear function of the number of antennas pooled from all serving RRHs and the modulation bits. Toward this end, we propose a novel three-step approach to solve the underlying mixed non-linear integer program. First, we transform this problem into a group association problem (GAP) with additional association constraints where each user must be associated with exactly one particular group of RRHs. Second, we study the relaxed power minimization problem (PMP) where the group association integer variables are relaxed and the computational constraint functions are approximated by weighted linear functions of transmission powers. We prove that this relaxed PMP can be solved optimally and the obtained optimal solution satisfies all association constraints of the original GAP problem. Third, we develop an iterative procedure to update the weight parameters of the approximated computational constraint functions to drive the achieved solution to an efficient and feasible solution of the original problem. Finally, we present numerical results to demonstrate the significant gains of our proposed design compared to that due to a fast greedy algorithm.

I. INTRODUCTION

C-RANs have been recently considered as an alternative architecture for next-generation wireless cellular networks. By realizing various processing functions in the cloud, this network architecture allows more efficient utilization of computational and radio resources to provide better network capacity and reduce both the network expenditure and operating expense. To realize these benefits, one has to address many technical challenges, such as, to efficiently utilize the computational resource in the cloud, fronthaul capacity, and design suitable communication schemes in the access network [1].

Recent literature on C-RANs has tackled some technical problems which are described in the following. The work in [2, 3] proposed different design approaches to utilize the fronthaul network capacity efficiently. The energy efficiency benefits of C-RAN is discussed in [4]. Moreover, the joint transmission for RRHs is studied in [5–7] through designing beamforming for all RRHs to minimize the total network power consumption or to maximize the network rate. However, there is very limited literature on the optimization of computational resources of C-RANs which is a major challenge for the large-scale network where the cloud must support a large number of users.

In [8], we studied the MIMO beamforming for C-RANs assuming that the required CE for any user is the same regardless of the number of serving RRHs and required QoS,

which is an over-simplified model. In [9], the computation model considering turbo decoding of users' data is adopted for C-RAN uplink design. In general, this CE model for uplink transmission cannot be applied to the downlink since the cloud must perform different tasks in the encoding process for the downlink, which are very different from the decoding process in the uplink. In fact, the authors of [10] have shown that the required CE to support a particular user's downlink communication is a non-linear function of different parameters including number of antennas, number of modulation bits, coding rate. This computation model has motivated us to fill the gap of the literature where we consider the joint transmission design for C-RANs considering limited cloud computation capacity and more realistic CE model.

Specifically, we consider downlink joint processing for transmission power minimization in C-RANs subject to the practical constraints on users' quality-of-service (QoS) and cloud computational capacity. We propose a novel solution approach to deal with this complex problem. We first transform this problem into an equivalent *group association problem* (GAP) where each user must be associated with exactly one of all possible groups of RRHs. This transformation enables us to express the computational constraint in the linear form. Then, we relax GAP to obtain a new PMP by relaxing the user and RRH group association variables into real variables and approximating computational constraint functions by corresponding weighted linear functions. By applying the duality approach, we prove that the global optimal solution of this PMP can be obtained and this optimal solution satisfies all association constraints of GAP. Finally, we propose to iteratively update the weights of the approximated computational constraint functions and solve the corresponding PMP to obtain a good and feasible solution for the original problem. We demonstrate through numerical studies that our proposed design significantly outperforms the fast greedy algorithm.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider the multiuser downlink transmission in C-RAN system consisting three main parts as in [8]: base-band units (BBU) pool, fronthaul network, and K RRHs which serve M users on the same frequency band. In this system, the joint processing scheme is employed to efficiently exploit available antennas at different RRHs for the interference mitigation. Specifically, it is assumed that each user is served by a specific group of RRHs. Let us define \mathcal{K} and \mathcal{U} as the sets of RRHs and UEs in the network, respectively. Denote \mathcal{K}_u ($\mathcal{K}_u \subseteq \mathcal{K}$) as the set of nearby RRHs which can be chosen to serve user u .

We now define binary variables $t_{u,k}$ where $t_{u,k} = 1$ if RRH k serves user u and $t_{u,k} = 0$ otherwise. Then, we have $t_{u,k} = 0$ for all $k \notin \mathcal{K}_u$. In addition, the set of RRHs that serve user u can also be expressed as $\mathcal{R}_u = \{k | k \in \mathcal{R}, t_{u,k} = 1\}$.

In this paper, we consider multiple-inputs single-output transmissions where RRH k is equipped with N_k antennas ($k \in \mathcal{K}$) and each user has a single antenna. We assume that each user (e.g., user u) has a single data stream which is represented by symbol sequence $x_u \in \mathbb{C}$ of unit power. Let $\mathbf{v}_{u,k} \in \mathbb{C}^{N_k \times 1}$ be the precoding vector at RRH k applied to signal of UE u . For a certain RRH-user allocation, the BB signal y_u received at user u can be written as

$$y_u = \sum_{k \in \mathcal{K}_u} t_{u,k} \mathbf{h}_{u,k}^H \mathbf{v}_{u,k} x_u + \sum_{i \in \mathcal{U}/u} \sum_{l \in \mathcal{K}_i} t_{i,l} \mathbf{h}_{u,l}^H \mathbf{v}_{i,l} x_i + z_u, \quad (1)$$

where $\mathbf{h}_{u,k} \in \mathbb{C}^{N_k \times 1}$ denotes the channel coefficient vector between RRH k and user u , and z_u describes the noise at user u . Also, the SINR achieved by user u can be expressed as

$$\Gamma_u = \frac{\left| \sum_{k \in \mathcal{K}_u} t_{u,k} \mathbf{h}_{u,k}^H \mathbf{v}_{u,k} \right|^2}{\sum_{i \in \mathcal{U}/u} \left| \sum_{l \in \mathcal{K}_i} t_{i,l} \mathbf{h}_{u,l}^H \mathbf{v}_{i,l} \right|^2 + \sigma^2}, \quad (2)$$

where σ^2 is the noise power.

A. Cloud Computation Effort

In this section, we first describe the model that quantifies the CE required for processing the baseband signal for a user. Practically, the CE (in Giga Operations Per Second - GOPS) required by physical layer can be determined based on the Frequency Domain processing (FDP) and Coding processing (CP) [10]. The CE for FDP scales with the number of antennas supporting a user. While the CE for CP depends on the number of bits, the type of coding, and encoding side or decoding side. For example, in uplink system, the base station has to perform decoding and if the turbo code is employed, then the CE for CP can be calculated as in equation (3) of [9]. In the downlink, the CE for CP scales with the number of bits at a coding rate C [10]. Generally, CE required to serve a specific user is a function of number of utilized antennas, the modulation bits, coding rate, number of data streams, and number of allocated resource blocks [4]. The CE required for each user u can be expressed in a general form as follows:

$$X_u = f(N_u, b_u), \quad (3)$$

where $f(\cdot, \cdot)$ is a non-linear increasing function with its variables, N_u denotes the total number of antennas serving user u , and b_u is the modulation bits per symbol in the data stream x_u . Here, the number of antennas serving user u is the total antennas of all RRHs in set \mathcal{R}_u , which can be written as

$$N_u = \sum_{k \in \mathcal{R}_u} N_k = \sum_{k \in \mathcal{K}_u} t_{u,k} N_k. \quad (4)$$

One particular model for $f(\cdot, \cdot)$ proposed in [10] is

$$f_{\text{ex}}(N_u, b_u) = (27N_u + 9N_u^2 + 3Cb_u) / 5, \quad (5)$$

Our design in this paper is, however, applicable for any other CE models (i.e., any other models for the CE function $f(\cdot)$). From (3), the total CE required by the BBU pool to process all users' signals can be expressed as

$$X_{\text{total}} = \sum_{u \in \mathcal{U}} X_u = \sum_{u \in \mathcal{U}} f(N_u, b_u). \quad (6)$$

Studies from the literature [5, 6] have shown that allowing bigger group of RRHs to serve each user results in better network performance in term of total transmission power or transmission rate. However, enabling bigger group of RRHs to serve each user consumes more computation resource in the cloud. This observation motivates us to investigate the joint transmission design for C-RAN considering limited computational capacity of the BBU pool.

B. Computation Constrained Power Minimization Problem

Our design aims to determine the set of RRHs serving each UE and the corresponding precoders for RRHs to minimize the total transmission power under the constraints on the computational capacity in the cloud and users' QoS. Here, the required QoS of UE u is described as

$$(C1) : \Gamma_u \geq \gamma_u, \quad \forall u \in \mathcal{U}. \quad (7)$$

We assume that the cloud has computation capacity of X^{\max} (GOPS). The cloud computational constraint is captured as

$$(C2) : X_{\text{total}} = \sum_{u \in \mathcal{U}} f\left(\sum_{k \in \mathcal{K}_u} t_{u,k} N_k, b_u\right) \leq X^{\max}. \quad (8)$$

Then, the Computation Constrained Power Minimization problem (CCPMP) can be stated as

$$\begin{aligned} \min_{\{\mathbf{v}_{u,k}\}, \{t_{u,k}\}} & \sum_{u \in \mathcal{U}} \sum_{k \in \mathcal{K}} \mathbf{v}_{u,k}^H \mathbf{v}_{u,k} \\ \text{s. t. } & (C1), (C2), \\ & (C3) : t_{u,k} \in \{0, 1\}, \forall (u, k) \in \mathcal{U} \times \mathcal{K}. \end{aligned} \quad (9)$$

The optimization problem (9) is indeed a non-linear mixed integer program (MIP), which is very hard to solve. In the following, we propose a novel low-complexity algorithm to solve this problem.

III. ALGORITHM DESIGN

A. Group Association Reformulation

It can be observed that the required CE for user u can be calculated if the set of serving RRHs \mathcal{R}_u is given. Let \mathcal{G}_u denote the set of all possible groups of RRHs serving user u each of which is a sub-set of \mathcal{K}_u and $A_u = |\mathcal{G}_u|$. Then, any specific RRH group in \mathcal{G}_u , denoted as \mathcal{R}_u^a ($1 \leq a \leq A_u$), can be chosen to serve user u . Let us define $\mathcal{R}_u^a = \{k_1, \dots, k_{e_u^a}\}$ where $e_u^a = |\mathcal{R}_u^a|$ and $\{k_1, \dots, k_{e_u^a}\} \subset \mathcal{K}_u$, represent the indices of serving RRHs for user u in \mathcal{R}_u^a . If user u is associated with \mathcal{R}_u^a , CE for user u can be given as

$$X_u^a = f\left(\sum_{k \in \mathcal{R}_u^a} N_k, b_u\right). \quad (10)$$

Let \mathbf{w}_u^a denote a vector, which is formed by concatenating precoding vectors of all RRHs in group \mathcal{R}_u^a for user u . Then, we can write $\mathbf{w}_u^a = [\mathbf{v}_{u,k_1}^T, \dots, \mathbf{v}_{u,k_{e_a}}^T]^T$. We also denote $\mathbf{h}_i^a = [\mathbf{h}_{i,k_1}^T, \dots, \mathbf{h}_{i,k_{e_a}}^T]^T$ as the concatenated channel vector for RRHs in \mathcal{R}_u^a and user i . Then, the SINR of user u if it is served by RRH group \mathcal{R}_u^a can be rewritten as

$$\Gamma_u^a = \frac{|\mathbf{h}_u^{aH} \mathbf{w}_u^a|^2}{\sum_{i \in \mathcal{U}/u} \sum_{m=1}^{A_i} |\mathbf{h}_u^{mH} \mathbf{w}_i^m|^2 + \sigma^2}. \quad (11)$$

Then, we can define GAP as

$$\begin{aligned} \min_{\{\mathbf{w}_u^a\}, \{r_u^a\}} \quad & \sum_{u \in \mathcal{U}} \sum_{a=1}^{A_u} \mathbf{w}_u^{aH} \mathbf{w}_u^a \quad (12) \\ \text{s. t. (C4):} \quad & \sum_{a=1}^{A_u} r_u^a \Gamma_u^a \geq \gamma_u, \forall u \in \mathcal{U}, \\ \text{(C5):} \quad & \sum_{u \in \mathcal{U}} \sum_{a=1}^{A_u} r_u^a X_u^a \leq X^{\max}, \\ \text{(C6):} \quad & \sum_{a=1}^{A_u} r_u^a = 1, \forall u \in \mathcal{U}, \\ \text{(C7):} \quad & r_u^a \in \{0, 1\}, \forall u \in \mathcal{U}, 1 \leq a \leq A_u, \end{aligned}$$

where binary variable $\{r_u^a\}$, ($u \in \mathcal{U}$, $1 \leq a \leq A_u$), represents the association between user u and RRH group $\mathcal{R}_u^a \in \mathcal{G}_u$, and the constraint (C6, C7) ensure that each user is associated with only one group of RRHs. Let $\{\mathbf{w}_u^{a*}\}, \{r_u^{a*}\}$ denote the optimal precoding vectors and group association variables obtained by solving GAP (12). Then, the optimal precoding vectors solution for CCPMP (9) can be determined by decomposing \mathbf{w}_u^{a*} into corresponding precoding vectors for RRHs in \mathcal{R}_u^a with $r_u^{a*} = 1$. Unfortunately, GAP is still a non-convex MIP. It needs to be noticed that the case of that $r_u^a = 1$ but $\mathbf{w}_u^{aH} \mathbf{w}_u^a = 0$ is equivalent to that of $r_u^a = 0$. Based on this, we propose an algorithm for solving GAP in the next section.

B. Group Association based Algorithm

We now present a novel method based on which we can achieve the solution for GAP. In this method, two different relaxation approaches are applied to address the two constraints (C4) and (C5), respectively. In particular, a three-step solution approach is employed as follows. In the first step, we tackle constraints (C4) by transforming GAP into a general form where we set all variables $\{r_u^a\}$ equal to one. Interestingly, if solution of the new problem satisfies the some special condition for each user then there is only one group of RRHs having transmission power greater than zero for each user, and this solution is also a feasible solution of GAP. In the second step, we approximate the constraint (C5) in the new problem to obtain an approximated weight-based PMP from which we can determine a feasible solution for GAP. In the last step, we propose a method to update the weights to drive the obtained solution of PMP to an efficient solution of the CCPMP. These design steps are presented in details in the following.

1) *Transformation of GAP*: In the first step, we describe how to transform GAP into the PMP with discontinuous constraint. To deal with (C4), we set all $\{r_u^a\}$ equal to 1's while for constraint (C5), we replace $\{r_u^a\}$ by the step function $\delta(\mathbf{w}_u^{aH} \mathbf{w}_u^a)$ ($\delta(x) = 1$ if $x > 0$ and $\delta(x) = 0$ if $x = 0$), which will be further approximated in next step. In particular, GAP is relaxed to the following problem

$$\begin{aligned} \min_{\{\mathbf{w}_u^a\}, \{r_u^a\}} \quad & \sum_{u \in \mathcal{U}} \sum_{a=1}^{A_u} \mathbf{w}_u^{aH} \mathbf{w}_u^a \quad (13) \\ \text{s. t. (C8):} \quad & \sum_{a=1}^{A_u} \Gamma_u^a \geq \gamma_u, \forall u \in \mathcal{U}, \\ \text{(C9):} \quad & \sum_{u \in \mathcal{U}} \sum_{a=1}^{A_u} \delta(\mathbf{w}_u^{aH} \mathbf{w}_u^a) X_u^a \leq X^{\max}. \end{aligned}$$

The relationship between this problem and GAP is characterized in the following Proposition 1.

Proposition 1. *Let $\{\mathbf{w}_u^{a*}\}$ denote the optimal precoding vectors obtained by solving problem (13). For this solution, if there is only one a_u , $1 \leq a_u \leq A_u$ for each user u such that $\mathbf{w}_u^{a_u^*H} \mathbf{w}_u^{a_u^*} > 0$ and $\mathbf{w}_u^{a^*H} \mathbf{w}_u^{a^*} = 0$, $\forall a \neq a_u$ then $\{\mathbf{w}_u^{a*}\}$ and $\{r_u^{a*}\}$ where $r_u^{a*} = \delta(\mathbf{w}_u^{a^*H} \mathbf{w}_u^{a^*})$, $\forall (u, a)$ are the optimal precoding and RRH allocation solutions for GAP.*

Proof. We can see that (C8) is expanding version of (C4), and (C9) is equivalent to (C5). Thus, the optimal objective value of (13) is upper bounded by that of GAP. On the other hand, if $\{\mathbf{w}_u^{a*}\}$ satisfies the condition in this proposition, $\{\mathbf{w}_u^{a*}\}$ is also a feasible solution for GAP; hence, it must be the optimal solution for GAP. \square

2) *Approximation for problem (13)*: Problem (13) is difficult to solve because of the non-smooth step functions in constraint (C9). In this step, we apply the approximation method in [6] to deal with this constraint. Specifically, constraint (C9) can be relaxed into the following form

$$\text{(C10):} \quad \sum_{u \in \mathcal{U}} \sum_{a=1}^{A_u} \beta_u^a \mathbf{w}_u^{aH} \mathbf{w}_u^a X_u^a \leq X^{\max}, \quad (14)$$

where β_u^a is a weight parameter, which can be updated to achieve good approximation. Let us introduce vector β to represent all $\{\beta_u^a\}$. Then, we study the following problem

$$\begin{aligned} \min_{\{\mathbf{w}_u^a\}, \{r_u^a\}} \quad & \sum_{u \in \mathcal{U}} \sum_{a=1}^{A_u} \mathbf{w}_u^{aH} \mathbf{w}_u^a \quad (15) \\ \text{s. t.} \quad & \text{(C8) and (C10)}. \end{aligned}$$

In the following, we show how to obtain the optimal solution of problem (15).

3) *Optimal Solution of (15)*: We now apply the Lagrangian method to solve problem (15). In addition, we will show that the strong duality holds for problem (15). Interestingly, the optimal solution of the problem (15) satisfies the condition in Proposition 1; hence, it is a feasible solution of GAP.

a. Dual Problem of (15): The Lagrangian of problem (15) is expressed in (16) on the top of the page 4 where $\alpha_u^a = \beta_u^a X_u^a$,

$$\begin{aligned}
\mathcal{L}(\{\mathbf{w}_u^a\}, \boldsymbol{\lambda}, \mu) &= \sum_{u \in \mathcal{U}} \sum_{a=1}^{A_u} \mathbf{w}_u^{aH} \mathbf{w}_u^a - \sum_{u \in \mathcal{U}} \lambda_u \left(\frac{1}{\gamma_u} \sum_{a=1}^{A_u} |\mathbf{h}_u^{aH} \mathbf{w}_u^a|^2 - \sum_{i \in \mathcal{U}/u} \sum_{m=1}^{A_i} |\mathbf{h}_u^{mH} \mathbf{w}_i^m|^2 - \sigma^2 \right) \\
&\quad + \mu \left(\sum_{u \in \mathcal{U}} \sum_{a=1}^{A_u} \beta_u^a \mathbf{w}_u^{aH} \mathbf{w}_u^a X_u^a - X^{\max} \right) \\
&= \sum_{u \in \mathcal{U}} \lambda_u \sigma^2 - \mu X^{\max} + \sum_{u \in \mathcal{U}} \sum_{a=1}^{A_u} \mathbf{w}_u^{aH} \left[\left(1 + \mu \alpha_u^a \right) \mathbf{I}_u^a - \frac{\lambda_u}{\gamma_u} \mathbf{h}_u^a \mathbf{h}_u^{aH} + \sum_{i \in \mathcal{U}/u} \lambda_i \mathbf{h}_i^a \mathbf{h}_i^{aH} \right] \mathbf{w}_u^a.
\end{aligned} \tag{16}$$

$\forall (u, a)$ and $\boldsymbol{\lambda}$ denotes a vector whose elements represent Lagrange multipliers $\{\lambda_u\}$ of constraints (C8). Similarly, μ denotes the Lagrange multiplier of the constraint (C10). The dual function is then

$$g(\boldsymbol{\lambda}, \mu) = \min_{\{\mathbf{w}_u^a\}} \mathcal{L}(\{\mathbf{w}_u^a\}, \boldsymbol{\lambda}, \mu). \tag{17}$$

It can be verified from (16) that if any matrix $\Omega_u^a(\boldsymbol{\lambda}, \mu) = (1 + \mu \alpha_u^a) \mathbf{I}_u^a - \frac{\lambda_u}{\gamma_u} \mathbf{h}_u^a \mathbf{h}_u^{aH} + \sum_{i \in \mathcal{U}/u} \lambda_i \mathbf{h}_i^a \mathbf{h}_i^{aH}$ is not positive semi-definite, there will exist \mathbf{w}_u^a that makes $g(\boldsymbol{\lambda}, \mu)$ unbounded below. For notations, we use $\mathbf{A} \succ \mathbf{0}$, $\mathbf{A} \succeq \mathbf{0}$, and $\mathbf{A} \not\succeq \mathbf{0}$ to indicate that \mathbf{A} is positive definite, positive semi-definite, and positive semi-definite but not positive definite, respectively. Then, the dual problem can be written as

$$\begin{aligned}
&\max_{\boldsymbol{\lambda}, \mu \geq 0} \sum_{u \in \mathcal{U}} \lambda_u \sigma^2 - \mu X^{\max} \\
&\text{s. t. (C11): } \Omega_u^a(\boldsymbol{\lambda}, \mu) \succeq \mathbf{0}, \forall (u, a).
\end{aligned} \tag{18}$$

Note that if the dual problem (18) is unbounded above, i.e., $\lambda_u \rightarrow \infty$, then the problem (15) is infeasible due to the weak duality. Thus, we assume that the optimal objective value of the dual problem (18) is finite. Denote $(\boldsymbol{\lambda}^*, \mu^*)$ as the optimal solution of the dual problem. Note that problem (18) is convex in respect to $\boldsymbol{\lambda}, \mu$ where the objective function and all constraints are linear. Hence, the optimal solution of $\boldsymbol{\lambda}, \mu$ can be achieved by the following searching method [12].

First, we start with any feasible point, i.e., $\Omega_u^a(\boldsymbol{\lambda}, \mu) \succeq \mathbf{0}, \forall (u, a)$. Then, we iteratively perform the following steps:

- We decrease μ until there appears (u, a) such that $\Omega_u^a(\boldsymbol{\lambda}, \mu) = \mathbf{0}$.
- We increase $\boldsymbol{\lambda}$ so that (C11) is not violated.

This searching method not only helps us find the optimal solution of $(\boldsymbol{\lambda}^*, \mu^*)$ but it also finds the optimal precoding solution $\{\mathbf{w}_u^{a*}\} = \operatorname{argmin}_{\{\mathbf{w}_u^a\}} \mathcal{L}(\{\mathbf{w}_u^a\}, \boldsymbol{\lambda}^*, \mu^*)$, which satisfies the following properties.

b. Strong Duality and Properties of Optimal Solution: The following proposition implies that with the optimal solution of the dual problem, $(\boldsymbol{\lambda}^*, \mu^*)$, we can find $\{\mathbf{w}_u^{a*}\} = \operatorname{argmin}_{\{\mathbf{w}_u^a\}} \mathcal{L}(\{\mathbf{w}_u^a\}, \boldsymbol{\lambda}^*, \mu^*)$, which is the optimal solution of (15) due to the strong duality, and $\{\mathbf{w}_u^{a*}\}$ is also a feasible solution of GAP.

Proposition 2. Consider the set of constraints corresponding to $\forall a \in \{1, \dots, A_u\}$ with optimal $(\boldsymbol{\lambda}^*, \mu^*)$ for (18). Then,

- There exists a_u such that $\Omega_u^{a_u}(\boldsymbol{\lambda}^*, \mu^*) \neq \mathbf{0}$.
- It almost surely happens that all other matrix inequalities are strict, i.e., $\Omega_u^a(\boldsymbol{\lambda}^*, \mu^*) \succ \mathbf{0}, \forall a \neq a_u$.
- We can obtain the optimal solution $(\{\mathbf{w}_u^{a*}\}, \boldsymbol{\lambda}^*, \mu^*)$.

Proof. The proof sketch is described briefly as follows:

- We apply the searching method above to update $\boldsymbol{\lambda}$. Assume there exists $u \in \mathcal{U}$ such that $\Omega_u^a(\boldsymbol{\lambda}^*, \mu^*) \succ \mathbf{0}$ for all a . Then, we can keep all λ_i^* ($i \neq u$) and μ^* unchanged, and increase λ_u to a value λ'_u so that there is one a_u with $\Omega_u^{a_u}(\boldsymbol{\lambda}^* |_{\lambda_u = \lambda'_u}, \mu^*) \neq \mathbf{0}$. Hence, we can find another $\boldsymbol{\lambda}$ satisfying constraint (C11) and $\sum_{u \in \mathcal{U}} \lambda_u > \sum_{u \in \mathcal{U}} \lambda_u^*$, which is a contradiction. Hence, the first statement of Proposition 2 must hold.
- The way to update λ_u suggests that the probability to have two or more a_u at a specific value of λ_u satisfying the condition in *i.* is almost zero for most practical systems, except for the cases where the channels from two groups are exactly symmetric.
- When *i.* and *ii.* hold, for every a that the strict inequality in statement *ii.* holds, we can set \mathbf{w}_u^{a*} to all-0 vector. For each user u , we can find only one a_u so that $\Omega_u^{a_u}(\boldsymbol{\lambda}^*, \mu^*) \neq \mathbf{0}$. Hence, there exists $\hat{\mathbf{w}}_u^{a_u}, \|\hat{\mathbf{w}}_u^{a_u}\| = 1$, such that $\hat{\mathbf{w}}_u^{a_uH} \Omega_u^{a_u} \hat{\mathbf{w}}_u^{a_u} = 0$. Then, $\{\mathbf{w}_u^{a*}\}$ can be found via the matrix inversion method (see for example [13]). \square

The proposition states that $\{\mathbf{w}_u^{a*}\}$ can be determined when $(\boldsymbol{\lambda}^*, \mu^*)$ are finite. According to *Theorem 7* in [11] about the strong duality of the non-convex quadratic programming, it is implied that problems (15) and (18) satisfy the strong duality. Therefore, $\{\mathbf{w}_u^{a*}\}$ must be the optimal solution of problem (15). Moreover, the optimal solution of (15) satisfied the condition stated in Proposition 1; hence, it is a feasible solution of GAP (12) which would be close to the optimal solution of GAP if the relaxed constraint (C10) is sufficiently close to the original constraint (C5). Based on these interesting results, we devise an algorithm to solve GAP in the following.

4) Group Association Based Algorithm to Solve (12): We propose an efficient iterative algorithm to solve GAP by iteratively solving problem (15) and updating weight parameters β , which is summarized in Algorithm 1. Specifically, for a given β , the optimal solution of (15) satisfies that each user u is associated with only one group in \mathcal{G}_u . In each iteration, we update the weights β_u^a corresponding to RRH groups that

Algorithm 1 GROUP ASSOCIATION BASED ALGORITHM

- 1: Set $\Phi = \{(u, a) | u \in \mathcal{U}, 1 \leq a \leq A_u\}$, $\beta_u^a = 0, \forall (u, a)$.
 - 2: **repeat**
 - 3: Solve problem (15) with current $\{\beta_u^a\}$.
 - 4: **for** all (u, a) **do**
 - 5: **if** $\mathbf{w}_u^{aH} \mathbf{w}_u^a > 0$ **then**
 - 6: Set $\Phi = \Phi / (u, a)$ and $\beta_u^a = \frac{1}{\mathbf{w}_u^{aH} \mathbf{w}_u^a + \tau}$.
 - 7: **else if** $(u, a) \notin \Phi$ **then** Set $\beta_u^a = 1/\tau$ **end if**
 - 8: **end if**
 - 9: **end for**
 - 10: **until** Convergence
-

have been chosen in the current or previous iterations. These weights are updated by the well-known re-weighted method [6] $\beta_u^a = 1/(\mathbf{w}_u^{aH} \mathbf{w}_u^a + \tau)$ ($\tau \ll 1$), while other weights are remained to be zero. In fact, the weights of *not-chosen-yet* RRH groups are maintained at zero so that the “best” of these remaining RRH groups can be selected if current user-RRH group associations cannot maintain the cloud computational constraint. Then, user-RRH group associations can be determined through solving the power minimization problem (15) with the updated weights. The convergence of this proposed algorithm is stated in the following proposition.

Proposition 3. *Algorithm 1 converges to a feasible solution satisfying all constraints in GAP.*

Proof. The proof is given briefly as follows. First, when an RRH group, which had been chosen for user association, is dropped, its weight is set to a very large value (τ can be very small). This RRH group will not be chosen in all future iterations since it will lead to violation of constraint (C10), otherwise. Hence, if one RRH group are dropped for a particular user, a new group in Φ will be selected in that iteration. This implies that the number of potential RRH groups in Φ decreases over iterations. Because the number of RRH groups for all users is finite, there must exist an iteration from which the RRH group-user association for every user remains unchanged. It can be verified that the employed weight update guarantees that the achieved solution satisfies all constraints in GAP. \square

C. Fast Greedy Algorithm

For comparison purposes, we describe another fast greedy algorithm in Algorithm 2. It can be seen that for a given association solution between RRH groups and users, our CCPMP becomes the power minimization problem with QoS constraints which can be solved optimally [8]. Moreover, we should associate each user with RRHs as possible to improve the objective value. Here, the link between RRH k and user u is called “activated” if $t_{u,k}$ is set to 1, and vice versa. Assume all links are inactivated at the beginning and the algorithm comprises three steps. First, for each user, we activate one link with the RRH having the strongest channel gain. This is to ensure that each user is served by at least one RRH. Second, we sort all remaining RRH-user links based on their channel

Algorithm 2 FAST GREEDY ALGORITHM

- 1: Set $t_{u,k} = 0$ for all $u \in \mathcal{U}$ and $k \in \mathcal{K}$.
 - 2: Let $\Theta = \{(u, k) | u \in \mathcal{U}, k \in \mathcal{K}_u\}$
 - 3: **for** $u = 1$ to M **do**
 - 4: Set $t_{u,k^*} = 1$ if $k^* = \operatorname{argmax}_{k \in \mathcal{K}_u} \mathbf{h}_{u,k}^H \mathbf{h}_{u,k}$.
 - 5: Set $\Theta = \Theta / (u, k^*)$.
 - 6: **end for**
 - 7: Sort Θ based on channel gains from strongest to lowest.
 - 8: Let (n) indicates index n in sorted Θ . Find $\bar{m} = \max m$ s.t. $f(N_u, s_u) \leq X^{\max}$ if $t_{u(n),k(n)} = 1, \forall n \in \{1, \dots, \bar{m}\}$.
 - 9: Set $t_{u(m),k(m)} = 1$ for all $m \in \{1, \dots, \bar{m}\}$.
 - 10: Solve power minimization problem with defined $\{t_{u,k}\}$.
-

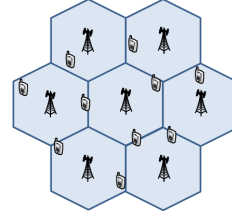


Fig. 1. Simulation model

TABLE I
TARGET SINRS FOR DIFFERENT s -QAM WITH $\bar{P}_e = 10^{-3}$.

s -QAM	2	4	8	16	32
bits/symbol	1	2	3	4	5
$\gamma(s)$	4.77	9.55	27.65	45.11	113.90

gains from strongest to lowest. Third, we activate as many links as possible starting from the strongest ones as long as the computational constraint is still satisfied. Finally, we solve the power minimization problem for the obtained RRH-user association solution to determine all precoding vectors.

IV. NUMERICAL RESULTS

We consider the 7-cell network where the distance between two nearest RRHs are 500 m as in Fig. 1. Users are randomly located inside the cells. The channel gains are generated by considering both Rayleigh fading and path loss which is modeled as $L_u^k = 36.8 \log_{10}(d_u^k) + 43.8 + 20 \log_{10}(\frac{f_c}{5})$ where d_u^k is the distance from user u to RRH k and $f_c = 2.5$ GHz. The noise power is set equal to $\sigma^2 = 10^{-13}$ W and $\tau = 10^{-6}$. The CE required for each user is simulated based on the function given in (5). We also assume that the BBU pool comprises B Intel Xeon E5590 processors where the computational capacity of each processor is 53.328 GOPS [14]. Hence, the total cloud computational capacity is $53.328 \times B$ GOPS. In addition, the same modulation scheme is chosen for all users, which is from the set $\{2-, 4-, 8-, 16-, 32-\}$ QAM and corresponding target SINRs are given in Table I.

In Figs. 2–4, we present the variations of total transmission power (in dB) obtained by the two algorithms versus the constellation size (s_u), the number of processors, the number of users, and the number of antennas per RRH, respectively where “GAB Alg.” represents Algorithm 1 and “(k -RRHs)”

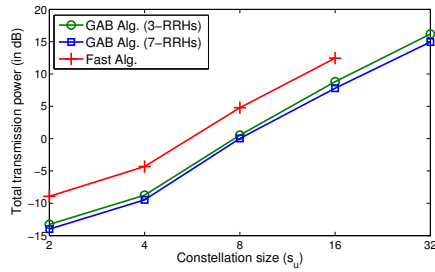


Fig. 2. Total transmission power versus constellation size where $M = 10$, $B = 40$, $N = 4$.

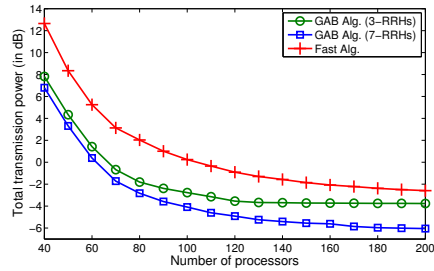


Fig. 3. Total transmission power versus the number of processors where $M = 10$, $s_u = 16$, $N = 4$.

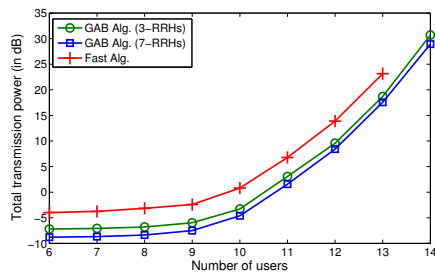


Fig. 4. Total transmission power versus the number of users where $B = 100$, $s_u = 16$, $N = 4$.

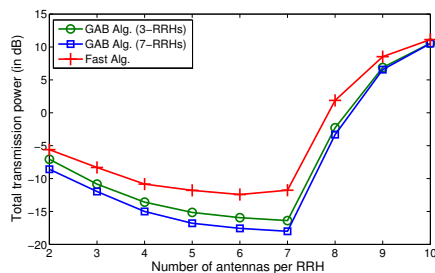


Fig. 5. Total transmission power versus the number of antennas where $B = 100$, $s_u = 2$, $M = 10$.

captures the case in which each user is allowed to associate with at most k nearest RRHs. As evident from Fig. 2, larger transmission power is required as the higher constellation size is employed. In addition, we can achieve better performance when the cloud computational capacity of BBU pool is larger

(i.e., the larger number of processors) as illustrated in Fig. 3. Also, larger number of users results in higher total transmission power as confirmed by Fig. 4.

Interestingly, Fig. 5 demonstrates that when larger number of antennas is equipped at each RRH, we can achieve lower transmission power. However, when number of antennas becomes too high, the required transmission power will increase. This is because of that limited cloud computational capacity would force the number of RRHs serving each user to be smaller when the number of antennas of each RRH becomes sufficiently high. It is also evident that our proposed GAB algorithm outperforms the Fast Greedy algorithm in all studied simulation scenarios. Moreover, allowing more RRHs to serve each user results in reducing the total transmission power.

V. CONCLUSION

We have presented a novel algorithm for downlink joint processing design in C-RAN which aims to minimize the total transmission power subject to constraints on the cloud computational capacity and users' QoS. Numerical results have illustrated the efficacy of our proposed algorithms and the impacts of different parameters on the network performance.

REFERENCES

- [1] A. Checko, H. L. Christiansen, Y. Yany, et al., "Cloud RAN for mobile networks – A technology overview," *IEEE Commun. Surv. Tutor.*, vol. 17, no. 1, First Quarter 2015.
- [2] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: Insights and challenges," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 152–160, Apr. 2015.
- [3] V. N. Ha, L. B. Le, and N.-D. Dao, "Coordinated multipoint transmission design for Cloud-RANs with limited fronthaul capacity constraints," *IEEE Trans. Veh. Technol.*, to appear.
- [4] D. Sabella, A. D. Domenico, E. Katranaras, et al., "Energy efficiency benefits of RAN-as-a-service concept for a cloud-based 5G mobile network infrastructure," *IEEE Access*, vol. 2, pp. 1586–1597, Dec. 2014.
- [5] 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, June 2010.
- [6] 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, June 2013.
- [7] V. N. Ha, L. B. Le, and N.-D. Dao, "Energy-efficient coordinated transmission for Cloud-RANs: Algorithm design and tradeoff," in *Proc. IEEE CISS'2014*, Mar. 2014.
- [8] V. N. Ha, L. B. Le, and N.-D. Dao, "Cooperative transmission in Cloud RAN considering fronthaul capacity and cloud processing constraints," in *Proc. IEEE WCNC'2014*, Apr. 2014.
- [9] P. Rost, S. Talarico, and M. Valenti, "The complexity-rate tradeoff of centralized radio access networks," *IEEE Trans. Wireless Commun.*, accepted for publication.
- [10] T. Werthmann, H. Grob-Lipski, and P. Proebster, "Multiplexing gains achieved in pools of baseband computation units in 4G cellular networks," in *Proc. IEEE 24th Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, pp. 3328–3333, Sept. 2013.
- [11] H. Tuy and H. Tuan, "Generalized s-lemma and strong duality in nonconvex quadratic programming," *J. Global Opt.*, vol. 56, pp. 1045–1072, 2013.
- [12] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [13] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.
- [14] Intel Xeon Processor 5500 Series. [Online]. Available: http://download.intel.com/support/processors/xeon/sb/xeon_5500.pdf, accessed Sept. 2015.