Resource Allocation for Uplink OFDMA C-RANs with Limited Computation and Fronthaul Capacity

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 fronthaul resource and rate allocation for the OFDMA uplink cloud radio access networks (C-RANs). This amounts to determine users' transmission rates and quantization bit allocation for I/Q baseband signals, which must be transferred from remote radio heads (RRHs) to the cloud over the capacity-limited fronthaul network. Our design aims at maximizing the system sum rate through optimal allocation of fronthaul capacity and cloud computation resources. Toward this end, we propose a novel two-stage approach to solve the underlying non-linear integer problem. In the first stage, we relax the integer variables to attain a relaxed problem, which is solved by employing a pricing-based method. Interestingly, we show that the pricing-based problem is convex with respect to each optimization variable, which can be, therefore, solved efficiently. In addition, we develop a novel mechanism to iteratively update the pricing parameter which is proved to converge. In the second stage, we propose two different rounding strategies, which are applied to the obtained continuous solution of the relaxed problem to achieve a feasible solution for the original problem. Finally, we present numerical results to demonstrate the significant sum-rate gains of our proposed design with respect to a standard greedy algorithm.**

I. Introduction

C-RAN has been recently considered as an architecture evolution for the next-generation wireless network. By realizing various communications and processing functions in cloud, C-RAN enables more efficient utilization of computational and radio resources, which results in better network throughput as well as reduced network deployment and operation costs. To realize these benefits, one has to address many technical challenges, e.g., to efficiently utilize computation resource in the cloud, fronthaul capacity, and to design suitable communication schemes in the access network [1].

Recent literature on C-RANs has tackled some of these technical problems which are described in the following. The work [2] proposed different design approaches to utilize the fronthaul network capacity efficiently. The energy efficiency benefits of C-RANs are studied in [3]. Moreover, the downlink joint transmission design for RRHs is conducted in [4–6] where beamforming for all RRHs are optimized to minimize the total power consumption or maximize the system sum rate. In [7], it has been shown that the computation effort (CE) required to support a user's downlink communication is a non-linear function of different parameters including number of antennas, modulation bits, and coding rate. Based on this CE model, our previous work in [8] has considered the joint processing design for downlink C-RANs with limited computation capacity. On the other hand, the authors in [9] have performed performance analysis for the uplink C-RAN system with limited cloud computation for turbo decoding. This work, however, assumes unlimited fronthaul capacity and does not optimize the resource allocation. To our best knowledge, uplink C-RAN design considering both limited fronthaul capacity and cloud computation resource has not been studied. The current work aims to fill this gap in the existing literature.

Specifically, we consider the joint optimization of users' communication rates and quantization bits allocation for sum rate maximization in the uplink OFDMA C-RANs subject to constraints on cloud computation resource and fronthaul capacity. We propose a novel two-stage solution approach to optimize the integer variables related to the allocation of quantization bits and discrete rate adaptation for different users. In the first stage, we relax the discrete variables and consider a related pricing problem of the relaxed problem. We then develop an iterative algorithm in which we update the pricing parameter and then solve the pricing problem in each iteration. In the second stage, we propose two methods to round the obtained solution of the relaxed problem and achieve feasible solutions for the original problem. We also discuss a fast greedy algorithm. Finally, we demonstrate through numerical studies that our proposed design significantly outperforms the greedy algorithm.

II. System Model

We consider the uplink OFDMA C-RAN with K RRHs and each RRH serves a separate group of users. Without loss of generality, we call RRH k and its corresponding coverage area as cell k. We further assume that both RRHs and users are equipped with single antenna. Each of the cells utilizes the whole spectrum of S sub-channels (SCs) (frequency reuse factor of one) and let $\mathcal{S}$ denote the set of all SCs. Moreover, we assume that SC allocation (SCA) to users in each cell and the corresponding power allocation have been predetermined.

Let the baseband signal transmitted by user k on an assigned SC s be $x_k^{(s)} \in \mathbb{C}$, which has the unit power. Then, the signal received at RRH k on SC s can be written as

$$y_k^{(s)} = \sum_{j \in \mathcal{K}} h_{k,j}^{(s)} \sqrt{p_j^{(s)}} x_j^{(s)} + \eta_k^{(s)}, \tag{1}$$

where $p_j^{(s)}$ denotes the transmission power of the user which is assigned SC s in cell j, $h_{k,j}^{(s)}$ is the channel gain from that user in cell j to RRH k on SC s, and $\eta_k^{(s)} \sim \mathcal{CN}\left(0, \sigma_k^{(s)2}\right)$ denotes the Gaussian thermal noise.

A. Quantize-and-Forward Processing

We assume that all baseband symbols $y_k^{(s)}$ must be transferred via the fronthaul links to the cloud, which performs all required processing. To accomplish this signal transfer, the "quantize-and-forward" pre-processing strategy is assumed where each RRH first quantizes its received baseband signals and then sends the corresponding quantized codewords to the cloud for being further processed by baseband units (BBU). Assume that RRH k uses the same $b_k^{(s)}$ bits to quantize the real and imaginary parts of received symbol $y_k^{(s)}$. Then, according to the results in [10], the quantization noise power can be approximated as

$$q(b_k^{(s)}) \simeq 2Q_y/2^{b_k^{(s)}}, \tag{2}$$

where $Q_y = \left(\int_{-\infty}^{\infty} f(y)^{1/3} dy\right)^3/12$, and $f(y)$ is the probability density function of both the real and imaginary part of $y_{k,u}^{(s)}$. Assuming a Gaussian distribution of the signal to be quantized, we have [11]

$$q(b_k^{(s)}) \simeq \frac{\sqrt{3}\pi}{2^{2b_k^{(s)}+1}} Y_k^{(s)}, \tag{3}$$

where $Y_k^{(s)}$ is the power of received signal $y_k^{(s)}$, i.e.,

$$Y_k^{(s)} = \sum_{j\in\mathcal{K}} |h_{k,j}^{(s)}|^2 p_j^{(s)} + \sigma_k^{(s)2} = D_k^{(s)} + I_k^{(s)}, \tag{4}$$

where $D_k^{(s)} = |h_{k,k}^{(s)}|^2 p_k^{(s)}$ and $I_k^{(s)} = \sum_{j\in\mathcal{K}/k} |h_{k,j}^{(s)}|^2 p_j^{(s)} + \sigma_k^{(s)2}$. Then, the total number of quantization bits forwarded from RRH k to the cloud can be expressed as

$$b_k = 2W \sum_{s\in\mathcal{S}} b_k^{(s)}, \tag{5}$$

where W is the bandwidth of one SC. Let $\mathbf{b}$ denote the vector that represents the numbers of quantization bits for all users in the network. For a given $\mathbf{b}$, the quantized symbol of $y_k^{(s)}$ can be then written as

$$\tilde{y}_k^{(s)} = y_k^{(s)} + e_k^{(s)}, \tag{6}$$

where $e_k^{(s)}$ represents the quantization error for $y_k^{(s)}$, which has zero mean and variance $q(b_k^{(s)})$. Then, the SINR of the signal corresponding to SC s in cell k can be expressed as

$$\gamma_k^{(s)}(b_k^{(s)}) = \frac{D_k^{(s)}}{I_k^{(s)} + q(b_k^{(s)})} \simeq \frac{D_k^{(s)}}{I_k^{(s)} + \frac{\sqrt{3}\pi Y_k^{(s)}}{2^{2b_k^{(s)}+1}}}. \tag{7}$$

B. Decoding Computation Effort Model

We assume that a data rate $r_k^{(s)}$ (in *"bits per channel use (bit pcu)"*) chosen for transmission over SC s comes from a discrete set with M_R different rates (i.e., different modulation-and-coding schemes) $\mathcal{M}_R = \{R_1, R_2, ..., R_{M_R}\}$. The chosen rate must be less than the link capacity to ensure satisfactory communication reliability, i.e., $r_k^{(s)} \leq \log_2(1+\gamma_k^{(s)}(b_k^{(s)}))$. We assume that the capacity-achieving turbo code is employed, then the computation effort required to successfully decode information bits strongly depends on the number of turbo-iterations. According to [9], this computation effort (in *"bit-iterations"*) for decoding the signal transmitted over SC s in cell k can be expressed as a function of $\gamma_k^{(s)}(b_k^{(s)})$ and $r_k^{(s)}$ as

$$C_k^{(s)} = A r_k^{(s)} \Big[B - 2\log_2\Big(\log_2\Big(1+\gamma_k^{(s)}(b_k^{(s)})\Big) - r_k^{(s)}\Big)\Big], \tag{8}$$

where $A = 1/\log_2(\zeta - 1)$, $B = \log_2\left((\zeta-2)/(\zeta T(\epsilon_{\mathsf{ch}}))\right)$, ζ is a parameter related to the connectivity of the decoder, $T(\epsilon_{\mathsf{ch}}) = -T'/\log_{10}(\epsilon_{\mathsf{ch}})$, T' is another model parameter, and ϵ_{ch} is the target channel outage probability. The set $\{T', \zeta\}$ can be selected by calibrating (8) with an actual turbo-decoder implementation or a message-passing decoder.

Let $\mathbf{r}$ denote the vector representing all users' rates selected for transmission over all SCs. Then, the total computation effort required by the cloud to successfully decode the signals for all users in the network can be calculated as

$$C_{\mathsf{total}}(\mathbf{r}, \mathbf{b}) = \sum_{k\in\mathcal{K}} \sum_{s\in\mathcal{S}} C_k^{(s)}. \tag{9}$$

C. Problem Formulation

We now present the problem formulation for our study, which is refereed to as *Fronthaul-and-Computation-Constrained Sum Rate Maximization* (FCCRM) problem. Our design aims to determine the transmission rates and the numbers of quantization bits for all users over all SCs that maximize the system sum-rate subject to the constraints on the fronthaul capacity and cloud computation resource. In particular, we are interested in solving the following problem.

$$\max_{\mathbf{r},\mathbf{b}} \quad \sum_{k\in\mathcal{K}} \sum_{s\in\mathcal{S}} r_k^{(s)} \tag{10a}$$

$$\text{s. t.} \quad \sum_{k\in\mathcal{K}} \sum_{s\in\mathcal{S}} C_k^{(s)} \leq \bar{C}_{\mathsf{cloud}}, \tag{10b}$$

$$r_k^{(s)} \leq \log_2\left(1+\gamma_k^{(s)}(b_k^{(s)})\right), \forall k\in\mathcal{K}, \forall s\in\mathcal{S}, \tag{10c}$$

$$\sum_{s\in\mathcal{S}} b_k^{(s)} \leq B_k/(2W), \forall k\in\mathcal{K}, \tag{10d}$$

$$b_k^{(s)} \text{ is integer}, \forall k\in\mathcal{K}, \forall s\in\mathcal{S}, \tag{10e}$$

$$r_k^{(s)} \in \mathcal{M}_R, \forall k\in\mathcal{K}, \forall s\in\mathcal{S}, \tag{10f}$$

Constraint (10b) means that the total computation efforts required by all users should not exceed the cloud computation limit, $\bar{C}_{\mathsf{cloud}}$. The second constraint (10c) ensures that the capacity of end-to-end link from user to cloud over the SC b is greater than rate assigned for the corresponding user over that SC. In constraint (10d), we assume that the fronthaul transport link connecting the cloud and RRH k has the capacity of B_k. It can be verified that this is non-linear integer problem which is very difficult to solve.

III. Algorithm Development

In this section, we describe our low-complexity two-stage algorithm to solve the FCCRM problem. In the first stage, we relax the integer variables $\mathbf{b}$ and discrete variables $\mathbf{r}$ into the continuous ones. Specifically, the constraint (10f) will be relaxed to

$$R_{\mathsf{min}} \leq r_k^{(s)} \leq R_{\mathsf{max}}, \forall k\in\mathcal{K}, \forall s\in\mathcal{S}, \tag{11}$$

where R_{min} and R_{max} are the lowest and highest rates in set $\mathcal{M}_R$, respectively. With this relaxation, the FCCRM problem becomes the relaxed-FCCRM problem which maximizes the objective function (10a) subject to constraints (10b)–(10d) and the new constraint (11). In the second stage, we round the obtained solution of relaxed-FCCRM problem to achieve a feasible solution of the original problem. This low-complexity algorithm is presented in more details in the following.

A. Stage 1 – Solving Relaxed-FCCRM Problem

We employ the penalty method to deal with the complex cloud computation constraint (10c) and solve the relaxed-FCCRM problem. Specifically, we consider a related *Pricing Computation Effort and Rate Maximization* (PCERM) problem, which is presented in the following

$$\max_{\mathbf{r},\mathbf{b}} \quad \Phi(\lambda,\mathbf{r},\mathbf{b}) = \sum_{k\in\mathcal{K}}\sum_{s\in\mathcal{S}} r_k^{(s)} - \lambda\sum_{k\in\mathcal{K}}\sum_{s\in\mathcal{S}} C_k^{(s)} \tag{12}$$
$$\text{s. t.} \quad \text{constraints (10c), (10d), and (11).}$$

In this problem, the weighted required computation effort is added to the objective function of FCCRM problem where the weight is called a pricing parameter λ in the sequel. Note that we have removed the cloud computation constraint (10c) in this pricing-based problem.

We can obtain a good feasible solution for the relaxed-FCCRM problem by iteratively solving the PCERM problem and adaptively adjusting the pricing parameter λ. We describe this procedure in details in the following.

1) Relationship between Relaxed-FCCRM and PCERM Problems: We now establish the relationship between the relaxed-FCCRM and PCERM problems based on which we can develop an efficient algorithm to solve the FCCRM problem. Denote $C_{\mathsf{PCERM}}(\lambda)$ as the consumed computation effort resulted from solving the PCERM problem for a given pricing parameter λ. Then, it can be verified that the optimal solutions of relaxed-FCCRM and PCERM problems are the same if $C_{\mathsf{PCERM}}(\lambda) = \bar{C}_{\mathsf{cloud}}$. This fact enables us to develop an iterative algorithm to solve the relaxed-FCCRM through tackling PCERM problem while adjusting λ iteratively in attempting to attain $C_{\mathsf{PCERM}}(\lambda) = \bar{C}_{\mathsf{cloud}}$. Our proposed algorithm is developed based on the assumption that the PCERM problem can be solved, which will be addressed in the next subsection. We now state some important results based on which we can develop a mechanism to update the pricing parameter in the following proposition.

Proposition 1. *We have:*

1) $C_{\mathsf{PCERM}}(\lambda)$ *is a decreasing function of* λ.
2) *There exists* $\bar{\lambda}$ *so that if we increase* $\lambda \geq \bar{\lambda}$ *then* $C_{\mathsf{PCERM}}(\lambda)$ *cannot be further decreased.*

Proof. The proof is given in Appendix A. □

These results form the foundation based on which we can develop an iterative algorithm presented in Algorithm 1. In this algorithm, we have described how λ is updated over iterations, which can be summarized as follows. Initially, we find $\bar{\lambda}$ and solve the PCERM problem with $\lambda = \bar{\lambda}$ to verify the feasibility of the relaxed-FCCRM problem. If $C_{\mathsf{PCERM}}(\bar{\lambda}) > \bar{C}_{\mathsf{cloud}}$, we can conclude that the relaxed-FCCRM problem is infeasible. Otherwise, we apply the bisection search method to update λ until $C_{\mathsf{PCERM}}(\lambda) = \bar{C}_{\mathsf{cloud}}$.

In the following, we develop an efficient algorithm to solve the PFCPM problem by decoupling this problem into K sub-

Algorithm 1 Pricing-based Algorithm for FCCRM Problem

```
Initialization: Find λ̄ by increasing λ.
if C_PCERM(λ̄) > C̄_cloud then
    Stop, the FCCRM problem is infeasible.
else if C_PCERM(λ̄) = C̄_cloud then
    Stop, the solution is achieved.
else if C_PCERM(λ̄) < C̄_cloud then
    Set l = 0, λ^(0) = λ̄, λ_U^(l) = λ̄ and λ_L^(l) = 0.
    repeat
        Set l = l + 1 and λ^(l) = (λ_U^(l-1) + λ_L^(l-1))/2.
        Solve PFCPM problem with λ^(l).
        if C_PCERM(λ^(l)) > C̄_cloud then
            Set λ_U^(l) = λ_U^(l-1) and λ_L^(l) = λ^(l).
        else if C_PCERM(λ^(l)) < C̄_cloud then
            Set λ_U^(l) = λ^(l) and λ_L^(l) = λ_L^(l-1).
        end if
    until C_PCERM(λ^(l)) = C̄_cloud or λ_U^(l) − λ_L^(l) is too small.
end if
```

problems corresponding to K cells and iteratively solving this problem with respect to each variable sets $\left\{r_k^{(s)}\right\}$ and $\left\{b_k^{(s)}\right\}$.

2) Decoupling PCERM Problem: The objective function of the PCERM problem can be rewritten as

$$\Phi(\lambda,\mathbf{r},\mathbf{b}) = \sum_{k\in\mathcal{K}} \Phi_k(\lambda,\mathbf{r}_k,\mathbf{b}_k), \tag{13}$$

where $\Phi_k(\lambda,\mathbf{r}_k,\mathbf{b}_k)$ is expressed in (14) on the top of the next page, and $\mathbf{r}_k, \mathbf{b}_k$ represent the vectors of all rates and quantization bits corresponding to cell k. In addition, all constraints in the PCERM problem can be divided into K independent groups corresponding to K cells. Therefore, we can decouple the PFCPM problem into K independent problems, $(\mathcal{P}_k)$'s, which are given as

$$(\mathcal{P}_k) \quad \max_{\mathbf{r}_k,\mathbf{b}_k} \quad \Phi_k(\lambda,\mathbf{r}_k,\mathbf{b}_k) \tag{15a}$$
$$\text{s. t.} \quad r_k^{(s)} \leq \log_2\left(1+\gamma_k^{(s)}(b_k^{(s)})\right), \forall s\in\mathcal{S}, \tag{15b}$$
$$\sum_{s\in\mathcal{S}} b_k^{(s)} \leq B_k/(2W), \tag{15c}$$
$$R_{\mathsf{min}} \leq r_k^{(s)} \leq R_{\mathsf{max}}, \forall s\in\mathcal{S}, \tag{15d}$$

Even though each problem $(\mathcal{P}_k)$ does not contain any constraints with integer variables, it is not jointly convex over all optimization variables. Moreover, it has a complex objective function, which is separately convex over each variable $\mathbf{r}_k$ and $\mathbf{b}_k$. Therefore, it is possible to solve problem $(\mathcal{P}_k)$ by alternately optimizing over one variable while keeping the other fixed.

3) Solving $(\mathcal{P}_k)$ for given $\mathbf{b}_k$: For a given $\mathbf{b}_k$, problem $(\mathcal{P}_k)$ becomes

$$\max_{\mathbf{r}_k} \sum_{s\in\mathcal{S}}\left[E_k^{(s)} r_k^{(s)} + 2\lambda A r_k^{(s)} \log_2\left(1-\frac{r_k^{(s)}}{t(b_k^{(s)})}\right)\right] \tag{16a}$$
$$\text{s. t.} \quad R_{\mathsf{min}} \leq r_k^{(s)} \leq \min\left(t(b_k^{(s)}), R_{\mathsf{max}}\right), \forall s\in\mathcal{S}, \tag{16b}$$

$$\Phi_k(\lambda, \mathbf{r}_k, \mathbf{b}_k) = \sum_{s\in\mathcal{S}} r_k^{(s)} - \lambda \sum_{s\in\mathcal{S}} C_k^{(s)} = \sum_{s\in\mathcal{S}} \left[(1-\lambda AB) r_k^{(s)} + 2\lambda A r_k^{(s)} \log_2 \left(\log_2\left(1+\gamma_k^{(s)}(b_k^{(s)})\right) - r_k^{(s)}\right)\right]. \quad (14)$$

where $E_k^{(s)} = \left(1 - \lambda AB + 2\lambda A \log_2 t(b_k^{(s)})\right)$ and $t(b_k^{(s)}) = \log_2\left(1+\gamma_k^{(s)}(b_k^{(s)})\right)$. In fact, the objective function of this problem is concave with respect to $\mathbf{r}_k$. In addition, the constraints are all linear. Hence, this problem is convex and its optimal solution can be determined easily by using the KKT condition as:

$$r_k^{(s)\star} = \max\Big[R_{\mathsf{min}}, \min\Big(t(b_k^{(s)}), R_{\mathsf{max}}, r|_{\frac{\partial w(r)}{\partial r}=-E_k^{(s)}}\Big)\Big]. \quad (17)$$

where $w(r) = 2\lambda A r \log_2\left(1 - r/t(b_k^{(s)})\right)$.

4) Solving $(\mathcal{P}_k)$ for given $\mathbf{r}_k$: For given $\mathbf{r}_k$, problem $(\mathcal{P}_k)$ becomes

$$\max_{\mathbf{b}_k} \quad \sum_{s\in\mathcal{S}} z\left(b_k^{(s)}\right) \quad (18a)$$

$$\text{s. t.} \quad b_k^{(s)} \geq t^{-1}\left(r_k^{(s)}\right), \forall s \in \mathcal{S}, \quad (18b)$$

$$\sum_{s\in\mathcal{S}} b_k^{(s)} \leq B_k/(2W), \quad (18c)$$

where $z(b_k^{(s)}) = G_k^{(s)} \log_2\left(\log_2\left(1+\gamma_k^{(s)}(b_k^{(s)})\right) - r_k^{(s)}\right)$, $G_k^{(s)} = 2\lambda A r_k^{(s)}$, and $t^{-1}(r)$ is the inverse function of $t(b)$. The objective function of this problem is concave with respect to $\mathbf{b}_k$. Hence, this problem is convex which can be solved by the Lagrangian method. The optimal solution of this problem is characterized in the following proposition.

Proposition 2. *The optimal solution of problem (18a)-(18c) can be expressed as*

$$b_k^{(s)\star} = \max\left(t^{-1}\left(r_k^{(s)}\right), b|_{\frac{\partial z(b)}{\partial b}=\mu}\right), \quad (19)$$

where μ is a constant so that $\sum_{s\in\mathcal{S}} b_k^{(s)\star} = B_k/(2W)$.

Proof. The proof is given in Appendix B. □

5) Proposed Algorithm to Solve PCERM Problem: We now describe an efficient algorithm to solve PCERM problem, which is summarized in Algorithm 2. In this algorithm, we iteratively and alternatively update one of the two variables $\mathbf{b}_k$ and $\mathbf{r}_k$ while keeping the other fixed. Because the optimal solution for each variable can be obtained, the objective function increases over iterations which ensures the convergence for this algorithm.

B. Stage 2 – Rounding

After running Algorithms 1 and 2, we obtain a feasible solution $\left\{r_k^{(s)\bigstar}\right\}$ and $\left\{b_k^{(s)\bigstar}\right\}$ of the relaxed problem, which are real numbers. To satisfy the discrete constraints (10e) and (10f), the continuous variables must be rounded to discrete values. This rounding must be designed carefully because the discrete results may not satisfy the original cloud computation and fronthaul constraints. To address this issue, we propose two rounding methods as follows.

Algorithm 2 ALGORITHM FOR PCERM PROBLEM

Initialization: Set $r_k^{(s)} = R_{\mathsf{min}}$ for all $k \in \mathcal{K}$ and $s \in \mathcal{S}$.
for $k \in \mathcal{K}$ **do**
  repeat
    Fix $\mathbf{r}_k$ and update $\mathbf{b}_k$ as in (19).
    Fix $\mathbf{b}_k$ and update $\mathbf{r}_k$ as in (17).
  until Convergence.
end for

Algorithm 3 FAST GREEDY ALGORITHM

Calculate $\{b_k^{(s)\prime}\}$ as in (21).
Set $b_k^{(s)} = \lfloor b_k^{(s)\prime} \rfloor$, for all (k, s).
Set $r_k^{(s)} = \max_{r\in\mathcal{M}_R} r$ s. t. $r \leq t(b_k^{(s)})$, for all (k, s).
while $C_{\mathsf{total}}(\mathbf{r}, \mathbf{b}) > \bar{C}_{\mathsf{cloud}}$ **do**
  Find $(k^*, s^*) = \arg\max C_k^{(s)}$.
  Reduce $r_{k^*}^{(s^*)}$ to the nearest value in $\mathcal{M}_R$.
end while

- *Iteratively Rounding (IR) Method:* We must run Algorithm 1 several times and perform the following task after each run. We choose the value of $r_k^{(s)\bigstar}$ (or $b_k^{(s)\bigstar}$) as close to one value in $\mathcal{M}_R$ (or an integer) as possible and set it to that closest value in all following iterations. This is repeated until convergence.
- *One-time Rounding and Adjusting (RA) Method:* This method has two phases, namely rounding and adjusting. First, we round all $\left\{r_k^{(s)\bigstar}\right\}$ and $\left\{b_k^{(s)\bigstar}\right\}$ to their nearest values in $\mathcal{M}_R$ and integers, respectively. Then, if any constraints are violated, we round down the variables one-by-one where each "rounding-down" action is performed for the one that affects the violated constraints the most.

C. Fast Greedy Algorithm

For comparison purposes, we describe another fast greedy algorithm in Algorithm 3 which has two phases. In phase one, since the users' rates are upper bounded by the link capacity as in (10b), we attempt to optimize the quantization bits allocation in order to maximize the total capacity of all users. To do so, we first relax these variables to continuous ones and solve the following problem

$$\max_{\mathbf{b}} \quad \sum_{k\in\mathcal{K}} \sum_{s\in\mathcal{S}} \log_2\left(1+\gamma_k^{(s)}(b_k^{(s)})\right) \quad (20)$$

$$\text{s. t.} \quad \text{constraint (10d).}$$

Similar to problem (18a)-(18c), this problem is convex because of its concave objective function. Using the Lagrangian method, we can obtain the optimal solution as follows.

$$b_k^{(s)\prime} = \max\left(0, b|_{\frac{\partial t(b)}{\partial b}=\nu}\right), \quad (21)$$

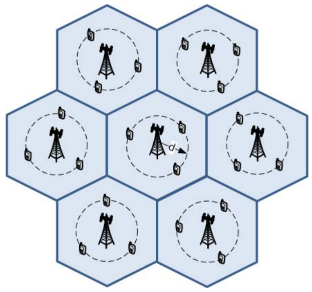

Fig. 1. Simulation model.

where ν is a constant that leads to $\sum_{s\in\mathcal{S}} b_k^{(s)\prime} = B_k/(2W)$. Then, we set $\mathbf{b}$ as the floor value of $\mathbf{b}'$. In phase two, we determine the users' rates so that they satisfy constraints (10b) and (10b). We start by setting $r_k^{(s)}$ to the highest value in the set $\mathcal{M}_R$ which is smaller than $\log_2\left(1+\gamma_k^{(s)}(b_k^{(s)})\right)$. Then, we iteratively reduce the rate that requires the highest computation effort if the cloud computation constraint is violated. All users' rates are achieved when that constraint is satisfied.

IV. Numerical Results

We consider a simple 7-cell network where the distance between two nearest RRHs is $400\,m$ as in Fig. 2. In each cell, we randomly place three users so that the distance from the cell center to every user is d (m). We set $S = 12$ and $W = 15\ kHz$. The channel gains are generated by considering both Rayleigh fading and path loss which is modelled as $L_{j,u}^k = 36.8\log_{10}(d_{j,u}^k) + 43.8 + 20\log_{10}(\frac{f_c}{5})$ where $d_{j,u}^k$ is the distance from user u in cell j to RRH k and $f_c = 2.5\,GHz$. The noise power is set equal to $\sigma^2 = 10^{-13}\ W$ and $p_k^{(s)} = 0.1\ W$. To generate $\mathcal{M}_R$, we consider 27 distinct modulation and coding modes with turbo coding as in the LTE standard [12]. For the computation model, we set $T' = 0.2$, $\zeta = 6$, and $\epsilon_{\text{ch}} = 10\%$.

First, we examine the convergence of our proposed algorithms by showing the variations of system sum rate over iterations by using the combination of Algorithm 1 and Algorithm 2 in Fig. 2. To obtain these simulation results, we set the fronthaul capacity for each RRH as $B_k = 1080\ Kbps$, the cloud computation limit as $\bar{C}_{\text{cloud}} = 700\ bit-iterations$. It can be observed that the algorithms converge for both inner loop and outer loops.

In Figs. 3–4, we present the variations of system sum rate (in bit pcu) obtained by the two Algorithm 1 and 2 without rounding (relaxed-FCCRM), with IR-rounding and RA-rounding methods (IR-FCCRM and RA-FCCRM), and the Fast Greedy Algorithm (Greedy Alg.), versus the values of the cloud computation limit and fronthaul capacity, respectively. As can been observed, our proposed algorithm outperforms the greedy algorithm in all studied scenarios. In addition, the iterative rounding (IR) method results in a better solution than that achieved by rounding and adjusting (RA) method. Interestingly, our proposed rounding methods can achieve the performance very close to that due to the relaxed one. Moreover, we can see that higher cloud computation limit and larger fronthaul capacity result in the greater system sum rate

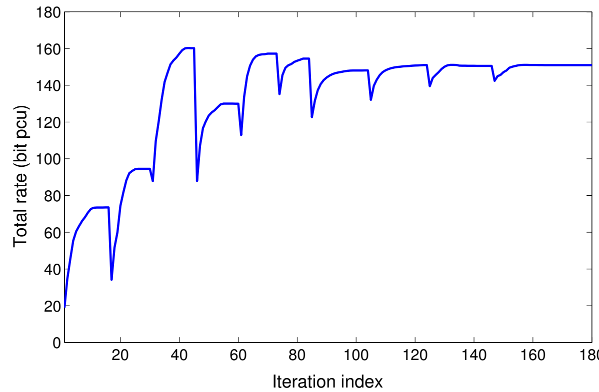

Fig. 2. Total rate versus the iteration index where $B_k = 1080\ Kbps$ and $\bar{C}_{\text{cloud}} = 700\ bit-iterations$.

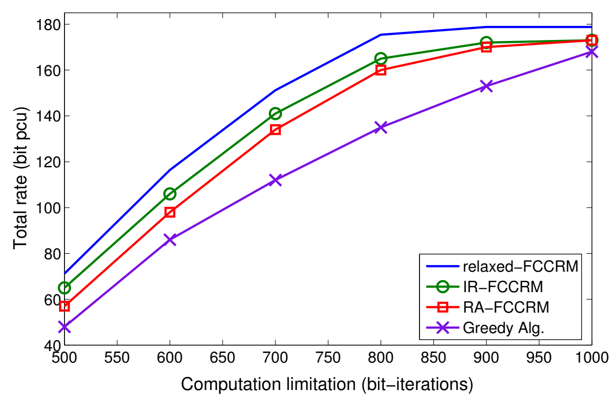

Fig. 3. Total rate versus $\bar{C}_{\text{cloud}}$ where $B_k = 1080\ Kbps$.

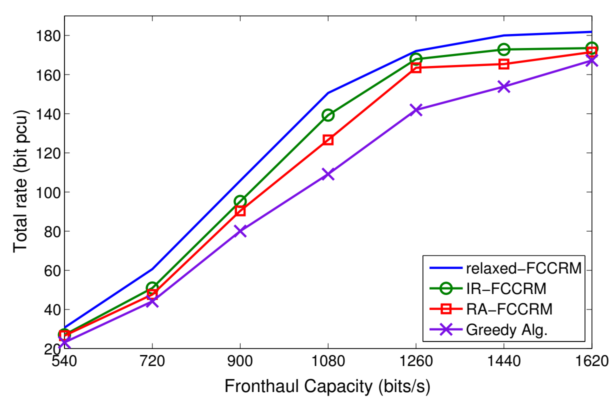

Fig. 4. Total rate versus B_k where $\bar{C}_{\text{cloud}} = 700\ bit-iterations$.

as expected. In addition, the sum rate becomes saturated as the cloud computation limit or fronthaul capacity is sufficiently large.

Fig. 5 shows the sum rate (due to relaxed-FCCRM and IR-FCCRM) versus the user-RRH distance d. In this simulation, we consider two different parameter settings: $B_k = 1080\ Kbps$ and $\bar{C}_{\text{cloud}} = 700\ bit-iterations$; and $B_k =$

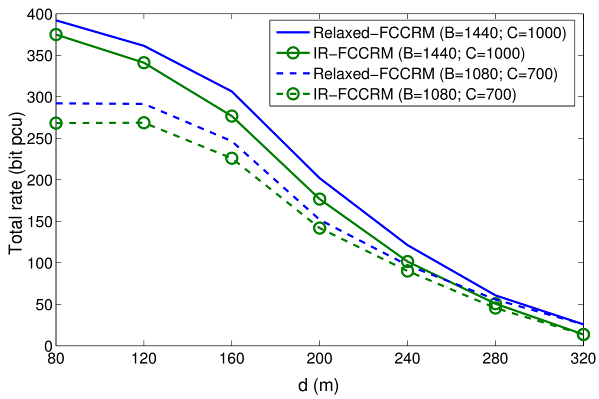

Fig. 5. Total rate vs distance from RRHs to their users.

$1440\ Kbps$ and $\bar{C}_{\mathsf{cloud}} = 1000\ bit-iterations$. When d is small, the received signal is stronger and the multi-cell interference is weaker, which leads to higher user capacity. This explains why the achieved sum rate decreases as d increases. Interestingly, for small values of d, the network cannot achieve higher sum rate as d decreases if the cloud computation limit and/or fronthaul capacity is too low. This means that there exists certain performance bottleneck due to limited cloud computation and/or fronthaul capacity resources.

V. Conclusion

We have proposed a novel algorithm for joint rate and quantization bits allocation design in uplink C-RANs which aims to maximize the system sum rate subject to constraints on the fronthaul capacity and cloud computation limit. Numerical results have illustrated the efficacy of our proposed algorithms and the impacts of different parameters on the sum-rate performance.

Appendix A
Proof of Proposition 1

A. Proof of Statement 1

Let us consider two values of the pricing parameter λ and λ' where $\lambda' > \lambda$. Let $(\mathbf{b}, \mathbf{r})$ and $(\mathbf{b}', \mathbf{r}')$ be the solutions of PCERM problem with λ and λ', respectively. Then, we have

$$\sum_{k\in\mathcal{K}}\sum_{s\in\mathcal{S}} r_k^{(s)} - \lambda C_{\mathsf{PCERM}}(\lambda) \geq \sum_{k\in\mathcal{K}}\sum_{s\in\mathcal{S}} r_k^{(s)\prime} - \lambda C_{\mathsf{PCERM}}(\lambda'),$$
$$\sum_{k\in\mathcal{K}}\sum_{s\in\mathcal{S}} r_k^{(s)\prime} - \lambda' C_{\mathsf{PCERM}}(\lambda') \geq \sum_{k\in\mathcal{K}}\sum_{s\in\mathcal{S}} r_k^{(s)} - \lambda' C_{\mathsf{PCERM}}(\lambda),$$

After combining and simplifying these two inequalities, we have $(\lambda' - \lambda)C_{\mathsf{PCERM}}(\lambda) \geq (\lambda' - \lambda)C_{\mathsf{PCERM}}(\lambda')$. Therefore, we have finished the proof for the first statement.

B. Proof of Statement 2

In order to prove the second statement, we prove that $C_{\mathsf{PCERM}}(\lambda)$ is lower bounded. According to the first statement, there must exist a $\bar{\lambda}$ as given in this statement. After some careful examination of the expression $C_k^{(s)}$ in (8), we have

$$C_k^{(s)} \geq AR_{\mathsf{min}}\left[B - 2\log_2\left(\log_2\left(1+\gamma_k^{(s)}(B_k/2)\right) - R_{\mathsf{min}}\right)\right].$$

Therefore, we have finished the proof for the second statement.

Appendix B
Proof of Proposition 2

The Lagrangian of problem (18a)-(18c) can be expressed as

$$\mathcal{L}(\mathbf{b}, \mu) = \sum_{s\in\mathcal{S}} z\left(b_k^{(s)}\right) - \mu\left(\sum_{s\in\mathcal{S}} b_k^{(s)} - B_k/(2W)\right), \quad (22)$$

where μ is the Lagrangian multiplier associated with the fronthaul constraint of problem (18a)-(18c). Then, the dual function of problem (18a)-(18c) can be written as

$$g(\mu) = \max_{\mathbf{b}_k} \mathcal{L}(\mathbf{b}, \mu) \text{ s. t. } b_k^{(s)} \geq J_k^{(s)}, \forall s \in \mathcal{S}. \quad (23)$$

This problem can be decoupled into S parallel sub-problems each of which corresponds to one SC. In addition, all these sub-problems have the same structure. Since its objective function is concave, each sub-problem can be solved by using KKT condition $\partial\mathcal{L}(\mathbf{b}, \mu)/\partial b_k^{(s)} = 0$, which is equivalent to

$$\partial z(b_k^{(s)})/\partial b_k^{(s)} = \mu. \quad (24)$$

Due to the constraint (18b), the optimal solution of $b_k^{(s)}$ must satisfy (19). In addition, the objective function is an increasing function with respect to $\mathbf{b}_k$; hence, the fronthaul constraint (18c) must be met with equality. Therefore, μ can be determined so that $\sum_{s\in\mathcal{S}} b_k^{(s)} = B_k/(2W)$. Therefore, we have finished the proof for Proposition 2.

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] 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.

[4] 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.

[5] V. N. Ha and L. B. Le, “Joint coordinated beamforming and admission control for fronthaul constrained Cloud-RANs,” in *Proc. IEEE GLOBE-COM’14,* Dec. 2014

[6] 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.

[7] 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.

[8] V. N. Ha and L. B. Le, “Computation capacity constrained joint transmission design for C-RANs,” in *Proc. IEEE WCNC’16,* Apr. 2016.

[9] P. Rost, S. Talarico, and M. Valenti, “The complexity-rate tradeoff of centralized radio access networks,” *IEEE Trans. Wireless Commun.,* 2015.

[10] J. A. Bucklew and N. C. Gallagher Jr., “Two-dimensional quantization of bivariate circularly symmetric densities,” *IEEE Trans. Inf. Theory,* vol. 25, no. 6, pp. 667–671, Nov. 1979.

[11] P. Baracca, S. Tomasin, and N. Benvenuto, “Backhaul rate allocation in uplink SC-FDMA systems with multicell processing,” *IEEE Trans. Wireless Commun.,* vol. 13, no. 3, pp. 1264–1273, Mar. 2014.

[12] 3GPP, “LTE evolved universal terrestrial radio access (e-utra): physical channels and modulation (3GPP TS 36.211 version 12.7.0 release 12),” 3GPP, Tech. Rep., Oct. 2015.