Joint Computing and Radio Resource Allocation in Cloud Radio Access Networks

Fatemeh Shirzad and Majid Ghaderi
Department of Computer Science, University of Calgary, Calgary, AB, Canada
Email: {fatemeh.shirzad,mghaderi}@ucalgary.ca

***Abstract*—This paper considers the problem of joint radio and computing resource allocation in Cloud Radio Access Network (C-RAN) architecture. We develop a resource allocation scheme to maximize weighted sum-rate of the system, while minimizing total power consumption. For power consumption, we consider both static and dynamic power consumption in Remote Radio Heads (RRHs), fronthaul links, and Base Band processing Units (BBUs). Our model considers quality of service requirements, fronthaul capacity, maximum transmission power, and computing capacity constraints in a comprehensive formulation. The joint resource allocation problem is non-convex, which is shown to be NP-hard, and thus we apply a number of techniques to convexify the problem. Then, using the Karush-Kuhn-Tucker (KKT) conditions, we show that the problem can be decomposed into two sub-problems that can be efficiently solved using an iterative Quadratically Constrained Quadratic Program (QCQP) and a bin packing algorithm, respectively. The performance of the proposed scheme is evaluated through simulation studies, which shows the proposed scheme outperforms the existing approaches in BBU minimization, total power consumption, and system utility which is defined as the weighted sum-rate minus power consumption.**

***Index Terms*—Cloud Radio Access Networks, Joint Resource Allocation, BBU-RRH Mapping, Joint Transmission**

I. Introduction

Wireless traffic is fast growing which requires increasing the number of base stations and supporting advanced transmission techniques such as Coordinated Multi-Point (CoMP). Increasing the number of base stations however leads to substantial increase in power consumption, capital expenditure (CAPEX) and operational expenditure (OPEX), for mobile network operators [1], [2]. At the same time, it is challenging to implement advanced techniques such as CoMP in traditional wireless Radio Access Network (RAN) architectures because of the high amount of control signaling required in the network [3], [4]. Moreover, an increase in the number of base stations to respond to the peak data traffic for peak hours results in the underutilization of resources in non-peak hours [5].

Cloud Radio Access Network (C-RAN) has been proposed as a promising solution to tackle the above-mentioned problems [3]–[5]. In C-RAN, the Base Band processing Units (BBUs) are separated from the Remote Radio Heads (RRHs) and centralized in a cloud computing based infrastructure called the *BBU pool*. The RRHs are connected to the BBU pool using high bandwidth fronthaul connections (e.g., over fiber optics). This way, computing resources are shared via virtualization in the BBU pool among multiple RRHs, which can reduce energy consumption due to statistical multiplexing of computing resources. Also, RRHs consume less power than traditional base stations. Furthermore, CoMP techniques can be realized in C-RAN due to its centralized architecture, which allows suppressing interference in the network efficiently.

While C-RAN has the potential to bring high energy efficiency and resource utilization, several technical challenges must be addressed to realize this architecture by managing transmission power and resources in radio and BBU side of the architecture [6]. Jointly optimizing transmission power from the RRHs to User Equipment (UE), active RRHs, and BBU-RRH mappings to determine the active BBU servers has a significant effect on power efficiency and meeting Quality of Service (QoS) requirements of users. Several research efforts have been carried out to solve the resource allocation problem in C-RAN [1], [2], [5], [7]–[13]. However, the existing works do not take a fully joint approach in allocating RRH and BBU side resources. There have been a few studies to model the joint problem [12], [13], however, these works end up separating the two sides of the problem at some point, which leads to less efficient solutions. Also, there are several restrictions with studies such as [1], [2], [5], which use simplified rate models, or do not consider Joint Transmission (JT), which is a main technique in 5G C-RAN to improve user throughput. Restrictions of the current work are discussed in more details in Section II.

In this work, we develop a joint model of radio and computing resource allocation in C-RAN to optimize system utilization, which includes both weighted sum-rate maximization and power consumption minimization in RRH and BBU side at the same time. We also design an algorithm to determine beamforming vectors from RRHs to UEs, active RRHs, active BBU servers, and BBU-RRH mappings in a JT-based C-RAN, in which several RRHs are clustered to send data to the same user at the same time-frequency resource block for enhanced data transmission rate.

Our contributions in this paper can be summarized as follows:

- We develop a comprehensive framework to model the problem of joint radio and computing resource allocation in C-RAN considering JT, fronthaul constraints, and QoS requirements. The objective is to maximize weighted sum-rate, while minimizing power consumption of the system.

- After relaxing the integer variables and using several techniques to make the problem convex, we establish the Lagrangian of the joint problem and analyze it using Karush-Kuhn-Tucker (KKT) conditions. Using this analysis, we obtain a modified radio resource allocation problem which includes the BBU side parameters.
- We propose a Quadratically Constrained Quadratic Programming (QCQP) based algorithm to solve the modified radio resource allocation problem.
- We evaluate the performance of the proposed algorithm through simulation. Our simulations show that the proposed algorithm outperforms the existing disjoint (such as [7], [8]) and partially joint (such as [12]) schemes in the number of active BBUs, power consumption, and overall utilization of the system.

The remainder of the paper is organized as follows. In Section II, the related work is discussed. We present the system model and the problem formulation in Section III. Analysis of the problem and the proposed algorithm are described in Section IV. Simulation results are presented in Section V. Section VI concludes the paper.

II. Related Work

Studies on resource allocation in C-RAN can be categorized into three groups, radio resource allocation, BBU resource allocation, and joint radio and BBU resource allocation. Here, we review current work on these three categories. Since our proposed scheme belongs to the joint resource allocation category, the studies in this category are reviewed in more details, and we briefly review recent work in other categories.

Radio resource allocation. The works in this category consider only radio resource allocation. In [14], a Deep Reinforcement Learning-based (DRL) algorithm is designed to minimize the power consumption in RRHs. An RRH selection problem is solved in [15] based on traffic density. The authors in [16] investigated Physical Resource Block (PRB) allocation and admission control problem in C-RAN subject to data rate requirements, fronthaul capacity, and transmission power constraints.

BBU resource allocation. Associating the minimum number of BBUs to the RRHs is known as the BBU-RRH mapping problem, which has been studied in several works. A combination of a bin packing based algorithm and simulated annealing is proposed in [9] to minimize power consumption in the BBU pool. The work in [10] is another study proposing three bin packing based algorithms to solve the BBU-RRH mapping problem. In [7], a borrow-and-lend approach is used to balance the load of the BBU servers by reallocating RRHs from highly utilized BBUs to the BBUs with lower load. The majority of the studies on the BBU-RRH mapping problem do not consider JT-clusters (groups of RRHs to transmit data in joint transmission technique). All RRHs in one JT-cluster need to be assigned to the same BBU so that data from all RRHs in a JT-cluster to any associated user is processed in the same BBU [8], [12]. In [8], a heuristic algorithm is proposed for mapping JT-clusters to BBUs.

Joint radio and BBU resource allocation. The works mentioned above consider radio resource allocation and BBU resource allocation separately, which leads to a less efficient solution compared to the joint approaches. Recently, several studies have considered joint allocation of BBU and RRH resources [1], [2], [5], [11]–[13]. The work in [11] is for uplink transmission and cannot be adopted for downlink and JT-based transmission. In [12], after formulating a joint radio and BBU resource allocation problem, the authors decompose the problem into two sub-problems, in which BBU allocation is separated from the radio side of the problem for simplicity. A Weighted Minimum Mean Squared Error (WMMSE) based algorithm is used to solve the beamforming part, and a Best Fit Decreasing (BFD) algorithm is applied to find BBU-RRH mappings. While data processing rates are included in the beamforming side of the problem, the separation of BBU-RRH mapping from the rest of the problem results in a sub-optimal solution. Besides, BBU capacities are modeled based on the number of UEs a BBU can accommodate, which is a simplification that results in a less accurate model compared to the rate-based BBU capacity modeling. Two other works [1], [5] study joint BBU-RRH mapping and user association problem, however, both studies have numerous restrictions such as assumption of single connectivity mode (i.e., non-JT), a simplified formulation with no transmit power allocation, beamforming vector, and data rate calculation. In [13], a joint BBU allocation and beamforming problem is solved by decomposing the problem into two sub-problems. However, the fronthaul constraints are ignored, and similar to [12], the BBU power consumption model is based on the number of RRHs mapped to each BBU instead of actual data rates of users. Besides, the objective function is only power consumption without considering throughput maximization of the system. A swarm intelligence based scheme is proposed in [2] to solve a joint user association and BBU-RRH mapping problem. However, [2] uses a simplified rate formulation and do not consider calculation of beamforming vectors.

We aim to overcome the restrictions and shortcomings of the above mentioned work. To this end, we propose a comprehensive model for joint beamforming, RRH activation, BBU activation, and BBU-RRH mapping. The objective function we study is based on both weighted sum-rate maximization and power minimization. We consider the QoS requirements of users, limited fronthaul capacity, BBU constraints, and maximum transmit power at RRHs. Our main contribution is developing a joint algorithm (rather than only a joint formulation and partially joint algorithms as in the exsiting work [12], [13]) that considers the interplay between the BBU side and the radio/beamforming side of the problem. This approach makes the resulting solution closer to optimal, as we will show in our evaluation results in Section V.

III. System Model

In this section, we describe the models considered for different aspects of the system and formulate the problem.

A. System Description

We consider downlink transmission of a C-RAN with multiple RRHs connected to a BBU pool consisting of multiple servers (called BBU servers). The set of RRHs and BBU servers are denoted as $\mathcal{R}$ and $\mathcal{S}$, respectively. The set of UEs (also called users throughout this paper) is represented by $\mathcal{K}$. We also consider JT as the CoMP technique, thus each user can get connected to multiple RRHs at the same time. We use the term *JT-cluster* to refer to any group of RRHs that jointly transmit data to a group of users. The notation $\mathcal{X}$ is used to denote the set of JT-clusters. Each user is associated to one JT-cluster based on the signals that a user receives from different RRHs. Also, without loss of generality, each RRH belongs to only one JT-cluster. Base band processing for each user is implemented by a Virtual Machine (VM) on a BBU server. The notation VM_k is used to refer to the VM of user k. After processing the traffic of each user k on VM_k (mapped to a BBU server), the processed traffic is transmitted to the RRHs in JT-Cluster of user k via fronthaul links, which connect RRHs to the BBU pool. For each fronthaul link connecting an RRH to the BBU pool, notation C_{fh} denotes its data rate capacity.

B. Service Model

We model QoS requirements of each user k with a minimum and maximum data rate requirement, denoted by r^k_{min} and r^k_{max}, respectively. These requirements are based on the Guaranteed Bit Rate (GBR), Maximum Bit Rate (MBR) and Aggregate-MBR (AMBR) defined in LTE specification [17]. Each user k has a service priority denoted by π_k, which allows the system to implement some notion of fairness [17]. In our model, higher value of π_k means higher priority for user k.

C. RRH Model

The data rate of user k is calculated as

$$r_k = B \cdot \log(1 + \gamma \cdot \text{SINR}_k), \tag{1}$$

in which B is the system bandwidth, $\log(.)$ function is logarithm with base 2, and γ is the coding efficiency [18]. Without loss of generality, γ is assumed to be one. In this formula, SINR_k is the Signal-to-Interference-plus-Noise Ratio (SINR) for user k, defined as

$$\text{SINR}_k = \frac{|\sum\limits_{l \in \mathcal{R}_k} \boldsymbol{h}^l_k \boldsymbol{m}^l_k|^2}{\sum\limits_{\substack{k' \in \mathcal{K} \\ k' \neq k}} |\sum\limits_{l \in \mathcal{R}_{k'}} \boldsymbol{h}^l_k \boldsymbol{m}^l_{k'}|^2 + \sigma^2_k}. \tag{2}$$

Throughout the paper, the vectors are presented by boldface lower case letters. The variable $\boldsymbol{m}^l_k \in \mathbb{C}^{M \times 1}$ is a vector of complex numbers with the length of M (number of transmit antennas in each RRH) denoting the transmit beamformer of RRH l to user k. Each UE is assumed to be equipped with one receive antenna [12], [19]. The term $\boldsymbol{h}^l_k \in \mathbb{C}^{1 \times M}$ denotes the channel vector between RRH l and user k. The noise power of receiver at user k is represented by σ^2_k. We use $\mathcal{R}_k$ to denote the JT-cluster of user k, which only contains those RRHs that have non-zero transmit beamforming vectors to user k. For any RRH l, $\mathcal{K}_l$ is the set of all users with non-zero beamforming vectors from RRH l (i.e., RRH l belongs to the JT-cluster of each user in $\mathcal{K}_l$). The RRHs can be in either active or sleep mode. The set of active RRHs is denoted as $\mathcal{R}_{on}$. We use P^{rrh}_{on} to denote the constant power consumption by each active RRH and its associated fronthaul link. The notation P^{tx}_l is the transmission power consumption in RRH l, which depends on the beamforming vectors of users served by RRH l. There is a limit on the total transmit power consumption in each RRH indicated by P^{tx}_{max}.

D. BBU Model

A server in the BBU pool can be in either active or sleep mode. The active state is when there is at least one VM (corresponding to one user) accommodated on that server. For any BBU server s, the binary variable y_s is 1 if BBU server s is active, and 0 otherwise. We use P^{bbu}_s to denote the power consumption in active server s. This power is composed of a constant portion and a dynamic portion. Constant power consumption of each active server s is denoted by P^{bbu}_c, and the dynamic portion of the power consumption depends on the processing rates of VMs scheduled on the server. Each server in the BBU pool has a limited processing capacity characterized by C_{bbu}. The function $\phi_k(\mu_k)$ is used to refer to the power consumption of VM_k. This function is a convex and increasing function of processing rate (denoted by μ_k) required by VM_k [19]. Similar to [12], we assume a linear model to relate power consumption to processing rates for each VM_k:

$$\phi_k(\mu_k) = \beta_k \cdot \mu_k, \tag{3}$$

where β_k is a positive constant. The processing rate μ_k of VM_k is also assumed to be linear with the data rate of user k (as in [12], [20]),

$$\mu_k = \gamma_1 \cdot r_k + \gamma_2, \tag{4}$$

where γ_1 and γ_2 are constants. Without loss of generality, we assume $\gamma_2 = 0$.

The power consumption of servers in sleep mode is negligible with respect to the power consumption in active mode and is assumed to be zero.

E. Problem Formulation

Based on the above models, the joint resource allocation problem is formulated as follows. In the rest of this section, we discuss different components of this formulation.

$$\underset{d^x_s, y_s, \boldsymbol{m}^l_k, \mathcal{R}_{on}}{\textit{maximize}} \quad f(y_s, \boldsymbol{m}^l_k, \mathcal{R}_{on}) \tag{5a}$$

$$\textit{s.t.} \quad r_k \geq r^k_{min}, \quad \forall k \in \mathcal{K} \tag{5b}$$

$$r_k \leq r^k_{max}, \quad \forall k \in \mathcal{K} \tag{5c}$$

$$\sum_{k \in \mathcal{K}_l} r_k \leq C_{fh}, \quad \forall l \in \mathcal{R} \tag{5d}$$

$$\sum_{x \in \mathcal{X}} (d^x_s \cdot \sum_{k \in \mathcal{K}^{jt}_x} \mu_k) \leq C_{bbu} \cdot y_s, \quad \forall s \in \mathcal{S} \tag{5e}$$

$$\sum_{k\in\mathcal{K}_l} \left\|\boldsymbol{m}_k^l\right\|_2^2 \le P_{max}^{tx}, \quad \forall l \in \mathcal{R} \tag{5f}$$

$$\sum_{s\in\mathcal{S}} d_s^x = 1, \quad \forall x \in \mathcal{X} \tag{5g}$$

$$d_s^x, y_s \in \{0,1\}, \quad \forall x \in \mathcal{X}, s \in \mathcal{S}. \tag{5h}$$

Objective Function. We describe the formulation starting with the objective function (5a) defined as

$$\begin{aligned} f(y_s, \boldsymbol{m}_k^l, \mathcal{R}_{on}) &= \alpha \cdot \sum_{k\in\mathcal{K}} w_k \cdot r_k \\ -(\sum_{l\in\mathcal{R}} (\mathbb{1}\{|\mathcal{K}_l| > 0\} &\cdot (P_{on}^{rrh} + P_l^{tx})) + \sum_{s\in\mathcal{S}} y_s \cdot P_s^{bbu}). \end{aligned} \tag{6}$$

This function is composed of two components. The first component is the weighted sum-rate which should be maximized and the second component is the power consumption in the RRHs, fronthaul links, and BBU pool, which should be minimized. The parameter α is a normalizing factor which is used when optimizing sum-rate and power consumption at the same time to make the corresponding values comparable. We use the notation $w_k = \frac{\pi_k}{r_{avg}^k}$ as the weight of each user k, taking both user priorities and long-term average data rate (denoted as r_{avg}^k) into account.

In the second component of the objective function, the notation $\mathbb{1}\{|\mathcal{K}_l| > 0\}$ for any RRH l is the indicator function, where $\mathbb{1}\{|\mathcal{K}_l| > 0\}$ is 1 if $|\mathcal{K}_l| > 0$, and 0 otherwise. An RRH can be in either active or sleep mode. The active state is when there is at least one user with a non-zero beamforming vector from that RRH. Therefore, $\mathbb{1}\{|\mathcal{K}_l| > 0\}$ is used to indicate if an RRH is active or not.

Constraints. Constraints (5b) and (5c) guarantee the minimum and maximum data rate requirements of each user, respectively. The fronthaul link capacities of RRHs are represented by constraints (5d). Constraints (5e) enforce the limits on the processing capacity of the BBU servers. The integer variable d_s^x for any JT-cluster x and BBU server s is defined as

$$d_s^x = \begin{cases} 1 & \text{if JT-cluster } x \text{ is scheduled on server } s, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

As mentioned earlier, in Section II, all RRHs in one JT-cluster must be scheduled on the same BBU [8], [12]. Therefore, VM of all users of those RRHs are scheduled on the same BBU accordingly. We use $\mathcal{K}_x^{jt}$ to denote the set of users in JT-cluster x. Constraints (5f) express the maximum power limit on each RRH. Constraints (5g) ensure that VM of each user is assigned to only one BBU server by assigning each JT-cluster to only one BBU server. Constraints (5h) reflect integer-valued variables d_s^x and y_s.

NP-hardness. The problem is non-convex because of its combinatorial nature due to the integer variables associated with the RRH and server activation. Even ignoring the integer variables, the problem can be reduced to the weighted sum-rate maximization problem which is known to be NP-hard [21] because of non-convex rate expressions in the objective function and constraints. Therefore, an approximation solution is required to solve the problem.

IV. Proposed Solution

In this section, problem (5) is analyzed. We convexify the problem in several steps and use Lagrangian and KKT conditions to develop an algorithm to solve the problem. We call the proposed algorithm Joint Radio and BBU resource allocation in C-RAN (JRBC).

Recasting the power components. The first step in analyzing problem (5) is rewriting the power components of the objective function as follows:

$$\begin{aligned} &\sum_{l\in\mathcal{R}} \mathbb{1}\{|\mathcal{K}_l| > 0\} \cdot (P_{on}^{rrh} + P_l^{tx}) \\ &= \frac{1}{\eta} \sum_{l\in\mathcal{R}_{on}} \sum_{k\in\mathcal{K}_l} \left\|\boldsymbol{m}_k^l\right\|_2^2 + |\mathcal{R}_{on}| P_{on}^{rrh}, \end{aligned} \tag{8}$$

and,

$$\sum_{s\in\mathcal{S}} y_s \cdot P_s^{bbu} = \sum_{k\in\mathcal{K}} \phi_k(\mu_k) + \sum_{s\in\mathcal{S}} y_s \cdot P_c^{bbu} \ . \tag{9}$$

In (8), the expression $\frac{1}{\eta} \sum_{l\in\mathcal{R}_{on}} \sum_{k\in\mathcal{K}_l} \left\|\boldsymbol{m}_k^l\right\|_2^2$ is the variable portion of the power consumption in RRHs which is, in fact, the sum of transmission powers from active RRHs to the users. The notation $\eta \in (0,1)$ is the inefficiency coefficient of the amplifier in each RRH [19].

Overall, using (8) and (9) and plugging (3) and (4) in (9), the objective function can be recast as

$$\begin{aligned} f(y_s, \boldsymbol{m}_k^l, \mathcal{R}_{on}) &= \sum_{k\in\mathcal{K}} (\alpha \cdot w_k - \beta_k \cdot \gamma_1) \cdot r_k \\ - \frac{1}{\eta} \sum_{l\in\mathcal{R}_{on}} \sum_{k\in\mathcal{K}_l} \left\|\boldsymbol{m}_k^l\right\|_2^2 &- |\mathcal{R}_{on}| P_{on}^{rrh} - \sum_{s\in\mathcal{S}} y_s \cdot P_c^{bbu} \ . \end{aligned} \tag{10}$$

Convexifying the components related to the RRH activation. For this part, we use the well-known ℓ_l-norm relaxation [19], [22]. To this end, we first need to define the following vector:

$$\boldsymbol{g} = \Big[\sum_{k\in\mathcal{K}_1} \left\|\boldsymbol{m}_k^1\right\|_2^2, \sum_{k\in\mathcal{K}_2} \left\|\boldsymbol{m}_k^2\right\|_2^2, \ \dots\ , \sum_{k\in\mathcal{K}_{|\mathcal{R}|}} \left\|\boldsymbol{m}_k^{|\mathcal{R}|}\right\|_2^2 \Big]. \tag{11}$$

An RRH l is active if and only if at least one beamforming vector associated to RRH l is nonzero (i.e., $\sum_{k\in\mathcal{K}_l} \left\|\boldsymbol{m}_k^l\right\|_2^2 > 0$). Then, we can conclude that $|\mathcal{R}_{on}| = \|\boldsymbol{g}\|_0$. Then, the following normalized ℓ_1-norm relaxation is applied on $\|\boldsymbol{g}\|_0$:

$$\xi \cdot \|\boldsymbol{g}\|_1 = \xi \cdot \sum_{l\in\mathcal{R}} \sum_{k\in\mathcal{K}_l} \left\|\boldsymbol{m}_k^l\right\|_2^2, \tag{12}$$

where ξ is the normalizing factor [22]. Since the beamforming vectors associated with the inactive RRHs are zero, the following is concluded:

$$\frac{1}{\eta} \sum_{l\in\mathcal{R}_{on}} \sum_{k\in\mathcal{K}_l} \left\|\boldsymbol{m}_k^l\right\|_2^2 = \frac{1}{\eta} \sum_{l\in\mathcal{R}} \sum_{k\in\mathcal{K}_l} \left\|\boldsymbol{m}_k^l\right\|_2^2. \tag{13}$$

Altogether, the objective function can be transformed to

$$f'(y_s, \boldsymbol{m}_k^l) = \sum_{k \in \mathcal{K}} (\alpha \cdot w_k - \beta_k \cdot \gamma_1) \cdot r_k \\ - (\frac{1}{\eta} + P_{on}^{rrh} \cdot \xi) \sum_{l \in \mathcal{R}} \sum_{k \in \mathcal{K}_l} \left\| \boldsymbol{m}_k^l \right\|_2^2 - \sum_{s \in \mathcal{S}} y_s \cdot P_c^{bbu} \; . \tag{14}$$

Convexifying SINR related expressions. Another component of the problem that makes it non-convex is the SINR expression in r_k. In order to make it convex, we introduce new variables r_k and add $r_k \geq 0$ along with the following constraint set to the problem:

$$r_k \leq B \cdot \log(1 + \mathrm{SINR}_k), \quad \forall k \in \mathcal{K} \, . \tag{15}$$

It can be proved that this constraint is tight (omitted for brevity), that is for any optimum solution, the equality between two sides of (15) holds. Furthermore, the following relation exists among Mean Squared Error (MSE) ϵ_k of user k and transmit and receive beamforming vectors [23]:

$$\epsilon_k = | \sum_{l \in \mathcal{R}_k} u_k \, \boldsymbol{h}_k^l \, \boldsymbol{m}_k^l - 1|^2 \\ + \sum_{\substack{k' \in \mathcal{K} \\ k' \neq k}} | \sum_{l \in \mathcal{R}_{k'}} u_k \, \boldsymbol{h}_k^l \, \boldsymbol{m}_{k'}^l |^2 + (\sigma_k)^2 \, |u_k|^2, \tag{16}$$

where $u_k \in \mathbb{C}$ is the receive beamformer at user k. With given transmit beamformers, the optimum MMSE (Minimum MSE) receive beamformers can be obtained as

$$u_k = \frac{\sum_{l \in R_k} \boldsymbol{h}_k^l \, \boldsymbol{m}_k^l}{\sum_{k' \in \mathcal{K}} | \sum_{l \in R_{k'}} \boldsymbol{h}_k^l \, \boldsymbol{m}_{k'}^l |^2 + (\sigma_k)^2}, \tag{17}$$

Then, the following well-known relation exists between SINR_k and corresponding optimum ϵ_k [24]:

$$\epsilon_k^{-1} = 1 + \mathrm{SINR}_k. \tag{18}$$

Using equations (15) and (18), the following is obtained:

$$r_k \leq -B \cdot \log(\epsilon_k). \tag{19}$$

We use the first order Taylor approximation, in which any function $f(x)$ at any point $\bar{x}$ can be approximated using

$$\tilde{f}(x) \approx f(\bar{x}) + (x - \bar{x}) \cdot f'(\bar{x}). \tag{20}$$

Applying (20) to the function $f(\epsilon_k) = \log(\epsilon_k)$ around any point of approximation $\tilde{\epsilon}_k$, the right-hand side of (19) is linearized, and the following is obtained.

$$\frac{B}{\tilde{\epsilon}_k \cdot \ln 2} \cdot \epsilon_k + B \log \tilde{\epsilon}_k - \frac{B}{\ln 2} \leq -r_k. \tag{21}$$

Altogether and using (4) to transform constraints (5e) to (22f) and with fixed u_k's, the problem is written as

$$\underset{d_s^x, y_s, \boldsymbol{m}_k^l, r_k}{\textit{maximize}} \quad f'(y_s, \boldsymbol{m}_k^l, r_k) \tag{22a}$$

$$\textit{s.t.} \quad \frac{B}{\tilde{\epsilon}_k \cdot \ln 2} \cdot \epsilon_k + B \log \tilde{\epsilon}_k - \frac{B}{\ln 2} \leq -r_k \quad , \forall k \in \mathcal{K} \tag{22b}$$

$$r_k \geq r_{min}^k \quad , \forall k \in \mathcal{K} \tag{22c}$$

$$r_k \leq r_{max}^k \quad , \forall k \in \mathcal{K} \tag{22d}$$

$$\sum_{k \in \mathcal{K}_l} r_k \leq C_{fh} \quad , \forall l \in \mathcal{R} \tag{22e}$$

$$\sum_{x \in \mathcal{X}} (d_s^x \cdot \sum_{k \in \mathcal{K}_x^{jt}} r_k) \leq \frac{C_{bbu}}{\gamma_1} \cdot y_s \quad , \forall s \in \mathcal{S} \tag{22f}$$

$$\sum_{k \in \mathcal{K}_l} \left\| \boldsymbol{m}_k^l \right\|_2^2 \leq P_{max}^{tx} \quad , \forall l \in \mathcal{R} \tag{22g}$$

$$\sum_{s \in \mathcal{S}} d_s^x = 1 \quad , \forall x \in \mathcal{X} \tag{22h}$$

$$d_s^x, y_s \in \{0, 1\} \quad , \forall x \in \mathcal{X}, s \in \mathcal{S} \tag{22i}$$

$$r_k \geq 0 \quad , \forall k \in \mathcal{K}. \tag{22j}$$

Relaxation of integer variables. There are two sets of integer variables d_s^x and y_s, both of which are part of the variables in the BBU side. Note that the other components of power consumption in the BBU side have been incorporated in the sum rate expression in the objective function using (3) and (4) as discussed above. First, we relax these integer variables to continuous variables between 0 and 1. Therefore, the constraint (22i) is replaced with $0 \leq d_s^x \leq 1$ and $0 \leq y_s \leq 1$. However, the constraint $d_s^x \leq 1$ is extra because it can be implied from constraints (22h) and $d_s^x \geq 0$. Also, we can ignore the constraint $y_s \geq 0$ since this constraint can be implied from constraints (22f), (22j), and $d_s^x \geq 0$.

BBU constraints. Now, the only difficulty of the problem is the set of constraints (22f), which is a multi-variable quadratic set of constraints and not convex in general. However, it can be shown (omitted for brevity) that the functions in the left-hand side of constraints (22f) are convex within range $d_s^x \geq 0$ and $r_k^b \geq 0$ (both are among the constraints of the problem).

Lagrangian analysis. Now that the problem is convex, we take its Lagrangian to obtain insights about the optimal solution using KKT conditions. The Lagrangian of the problem is:

$$L(\boldsymbol{d}, \boldsymbol{y}, \boldsymbol{m}, \boldsymbol{r}, \boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}, \boldsymbol{\lambda_3}, \boldsymbol{\lambda_4}, \boldsymbol{\lambda_5}, \boldsymbol{\lambda_6}, \boldsymbol{\lambda_7}, \boldsymbol{\lambda_8}, \boldsymbol{\lambda_9}, \boldsymbol{\nu}) = \\ - \sum_{k \in \mathcal{K}} (\alpha \cdot w_k - \beta_k \cdot \gamma_1) \cdot r_k + (\frac{1}{\eta} + P_{on}^{rrh} \cdot \xi) \sum_{l \in \mathcal{R}} \sum_{k \in \mathcal{K}_l} \left\| \boldsymbol{m}_k^l \right\|_2^2 \\ + \sum_{s \in \mathcal{S}} y_s \cdot P_c^{bbu} + \sum_{k \in \mathcal{K}} \lambda_1^k (\frac{B}{\tilde{\epsilon}_k \cdot \ln 2} \cdot \epsilon_k + B \log \tilde{\epsilon}_k - \frac{B}{\ln 2} + r_k) \\ + \sum_{k \in \mathcal{K}} \lambda_2^k \, (r_{min}^k - r_k) + \sum_{k \in \mathcal{K}} \lambda_3^k \, (r_k - r_{max}^k) \\ + \sum_{l \in \mathcal{R}} \lambda_4^l \, (\sum_{k \in \mathcal{K}_l} r_k - C_{fh}) \\ + \sum_{s \in \mathcal{S}} \lambda_5^s \, (\sum_{x \in \mathcal{X}} (d_s^x \cdot \sum_{k \in \mathcal{K}_x^{jt}} r_k) - \frac{C_{bbu}}{\gamma_1} \cdot y_s) \\ + \sum_{l \in \mathcal{R}} \lambda_6^l \, (\sum_{k \in \mathcal{K}_l} \left\| \boldsymbol{m}_k^l \right\|_2^2 - P_{max}^{tx}) \\ + \sum_{x \in \mathcal{X}} \nu^x \, (\sum_{s \in \mathcal{S}} d_s^x - 1) + \sum_{s \in \mathcal{S}} \sum_{x \in \mathcal{X}} \lambda_7^{sx} \, (-d_s^x) \\ + \sum_{s \in \mathcal{S}} \lambda_8^s \, (y_s - 1) + \sum_{k \in \mathcal{K}} \lambda_9^k \, (-r_k), \tag{23}$$

where $\boldsymbol{\lambda_1}, ..., \boldsymbol{\lambda_9}$ are Lagrange multipliers of inequality constraints and ν is the Lagrange multiplier of the equality constraints. Then, the partial derivative of Lagrangian function L with respect to variable y_s gives the following KKT condition:

$$\frac{\partial L}{\partial y_s} = P_c^{bbu} - \frac{\lambda_5^s \cdot C_{bbu}}{\gamma_1} + \lambda_8^s = 0. \tag{24}$$

Besides, using complementary slackness we obtain

$$\lambda_5^s \; (\sum_{x\in\mathcal{X}} (d_s^x \cdot \sum_{k\in\mathcal{K}_x^{jt}} r_k) - \frac{C_{bbu}}{\gamma_1} \cdot y_s) = 0. \tag{25}$$

From this, either $\lambda_5^s = 0$ or

$$\sum_{x\in\mathcal{X}} (d_s^x \cdot \sum_{k\in\mathcal{K}_x^{jt}} r_k) - \frac{C_{bbu}}{\gamma_1} \cdot y_s = 0. \tag{26}$$

We prove that the first case (i.e., $\lambda_5^s = 0$) does not hold, and (26) is always true. We use contradiction by assuming $\lambda_5^s = 0$. However, plugging this into (24) results in $\lambda_8^s = -P_c^{bbu}$, and since $P_c^{bbu} > 0$ it is concluded that $\lambda_8^s < 0$, which contradicts the KKT condition that the Lagrange multipliers are non-negative values. Therefore, $\lambda_5^s = 0$ does not hold, and the corresponding constraint is tight, that is

$$\sum_{x\in\mathcal{X}} (d_s^x \cdot \sum_{k\in\mathcal{K}_x^{jt}} r_k) = \frac{C_{bbu}}{\gamma_1} \cdot y_s \; . \tag{27}$$

Now that (27) holds for all $s \in \mathcal{S}$, by taking a summation on both sides of it and multiplying by P_c^{bbu}, we obtain

$$P_c^{bbu} \sum_{s\in\mathcal{S}} \sum_{x\in\mathcal{X}} (d_s^x \cdot \sum_{k\in\mathcal{K}_x^{jt}} r_k) = \frac{C_{bbu}}{\gamma_1} \sum_{s\in\mathcal{S}} P_c^{bbu} \cdot y_s \; . \tag{28}$$

Then, the summations on the left-hand side of the equation are rearranged as

$$P_c^{bbu} \sum_{x\in\mathcal{X}} ((\sum_{s\in\mathcal{S}} d_s^x) \cdot (\sum_{k\in\mathcal{K}_x^{jt}} r_k)) = \frac{C_{bbu}}{\gamma_1} \sum_{s\in\mathcal{S}} P_c^{bbu} \cdot y_s \; . \tag{29}$$

From constraint (22h), we know that $\sum_{s\in\mathcal{S}} d_s^x = 1$, hence the following is obtained:

$$\frac{P_c^{bbu} \cdot \gamma_1}{C_{bbu}} \sum_{x\in\mathcal{X}} \sum_{k\in\mathcal{K}_x^{jt}} r_k = \sum_{s\in\mathcal{S}} P_c^{bbu} \cdot y_s \; . \tag{30}$$

The two first summations on the left-hand side can be replaced by one summation since each user belongs to only one JT-cluster, and the result is as follows:

$$\frac{P_c^{bbu} \cdot \gamma_1}{C_{bbu}} \sum_{k\in\mathcal{K}} r_k = \sum_{s\in\mathcal{S}} P_c^{bbu} \cdot y_s \; . \tag{31}$$

Using (31), the objective function is transformed to

$$\begin{aligned} f''(\boldsymbol{m}_k^l, r_k) = &\sum_{k\in\mathcal{K}} (\alpha \cdot w_k - \beta_k \cdot \gamma_1 - \frac{P_c^{bbu} \cdot \gamma_1}{C_{bbu}}) \cdot r_k \\ &- (\frac{1}{\eta} + P_{on}^{rrh} \cdot \xi) \sum_{l\in\mathcal{R}} \sum_{k\in\mathcal{K}_l} \left\| \boldsymbol{m}_k^l \right\|_2^2 . \end{aligned} \tag{32}$$

As can be seen from the above, the objective function does not depend on y_s's anymore. Furthermore, to obtain this result, we have already involved two constraints (22f) and (22h). Consequently, we can remove variables y_s and d_s^x and all related constraints from the problem and obtain the following optimization problem, which is a QCQP and can be solved efficiently with the time complexity of $\mathcal{O}(|\mathcal{K}| \cdot |\mathcal{R}_{max}| \cdot M)^{3.5}$ [25], where $|\mathcal{R}_{max}|$ is the maximum number of RRHs in any JT-cluster.

$$\underset{\boldsymbol{m}_k^l, r_k}{\textit{maximize}} \quad f''(\boldsymbol{m}_k^l, r_k) \tag{33a}$$

$$\textit{subject to} \quad \frac{B}{\tilde{\epsilon}_k \cdot \ln 2} \cdot \epsilon_k + B \log \tilde{\epsilon}_k - \frac{B}{\ln 2} \leq -r_k \quad , \forall k \in \mathcal{K} \tag{33b}$$

$$r_k \geq r_{min}^k \quad , \forall k \in \mathcal{K} \tag{33c}$$

$$r_k \leq r_{max}^k \quad , \forall k \in \mathcal{K} \tag{33d}$$

$$\sum_{k\in\mathcal{K}_l} r_k \leq C_{fh} \quad , \forall l \in \mathcal{R} \tag{33e}$$

$$\sum_{k\in\mathcal{K}_l} \left\| \boldsymbol{m}_k^l \right\|_2^2 \leq P_{max}^{tx} \quad , \forall l \in \mathcal{R} \tag{33f}$$

$$r_k \geq 0 \quad , \forall k \in \mathcal{K}. \tag{33g}$$

This is an interesting result as it indicates that we can solve the problem without involving variables y_s and d_s^x making the problem state smaller. Indeed, the main difference between our work and the existing work on joint radio resource and BBU resource allocation is that those studies separate the BBU related variables from the radio side for simplicity without considering their dependency. Using the analysis described above, we proved that the variables can be separated in such a way that the BBU side parameters are included in the radio side problem. This means that the problem in the BBU side indirectly affects the optimal solution values of beamforming and rate variables. Furthermore, it does not affect the running time of the algorithm since we only need to modify the multipliers of the rate variable r_k in the objective function to $\alpha \cdot w_k - \beta_k \cdot \gamma_1 - \frac{P_c^{bbu} \cdot \gamma_1}{C_{bbu}}$.

RRH activation and refining the JT-clusters. By solving problem (33), transmit beamforming vectors $\boldsymbol{m}_k^l$'s are obtained. Using $\boldsymbol{m}_k^l$'s, active RRHs can be determined. To this end, any RRH l with $\sum_{k\in\mathcal{K}_l} \left\| \boldsymbol{m}_k^l \right\|_2^2 \leq \tau$ is switched off (if this does not violate QoS requirements of the users), where τ is a predefined value. The next step after solving problem (33) is refining the JT-cluster set $\mathcal{X}$. We use the set of active RRHs to refine any $x \in \mathcal{X}$ by removing any inactive RRH l from x. The solution can be further improved by resolving problem (33) using fixed active RRHs and refined JT-clusters.

The final step in determining the active BBU servers and BBU-RRH mappings. After solving problem (33), we use the resulted r_k's to calculate μ_k's, and from there, the total required processing rate ω_x of each JT-cluster x is calculated using

$$\omega_x = \gamma_1 \cdot \sum_{k\in\mathcal{K}_x^{jt}} r_k \; . \tag{34}$$

Then, the following bin-packing problem formulation can be used to model the BBU-RRH mapping problem:

$$\underset{d_s^x, y_s}{\textit{minimize}} \quad \sum_{s \in \mathcal{S}} y_s \cdot P_c^{bbu} \tag{35a}$$

$$\textit{subject to} \quad \sum_{x \in \mathcal{X}} d_s^x \cdot \omega_x \leq C_{bbu} \cdot y_s \quad , \forall s \in \mathcal{S} \tag{35b}$$

$$\sum_{s \in \mathcal{S}} d_s^x = 1 \quad , \forall x \in \mathcal{X} \tag{35c}$$

$$d_s^x, y_s \in \{0, 1\} \quad , \forall x \in \mathcal{X}, s \in \mathcal{S}. \tag{35d}$$

Bin-packing is widely studied and has efficient algorithms with provable performance. We can use an existing algorithm to solve this problem efficiently. Note that for the bin-packing problem to be feasible, we need to ensure that $\omega_x \leq C_{bbu}$. Therefore, we incorporate the following constraint into problem (33) when solving that problem:

$$\gamma_1 \cdot \sum_{k \in \mathcal{K}_x^{jt}} r_k \leq C_{bbu}, \quad \forall x \in \mathcal{X}. \tag{36}$$

JRBC algorithm. Now we have all the components to present JRBC (our proposed algorithm). The algorithm is depicted in Algorithm 1.

Algorithm 1 JRBC: Joint Radio and BBU allocation in C-RAN

1: Initialize the beamforming variable $\boldsymbol{m}_k^l$ for each user k and RRH l by dividing available total power consumption P_{max}^{tx} in RRH l equally among the users in $\mathcal{K}_l$
2: Set $i = 1$, and set the convergence parameter.
3: **repeat**
4: Using transmit beamforming vectors $\boldsymbol{m}_k^l$'s and (17), compute the receive beamformer u_k for each user k.
5: Set $j = 1$.
6: **repeat**
7: Compute $\tilde{\epsilon}_k$ using (16), u_k's and $\boldsymbol{m}_k^l$'s.
8: Use QCQP optimization to solve problem (33) while incorporating constraints (36) and using u_k's and $\tilde{\epsilon}_k$'s obtained from lines 4 and 7, respectively.
9: $j = j + 1$.
10: **until** $j > I_{max}$
11: $i = i + 1$.
12: **until** Convergence is reached or $i > J_{max}$
13: Put any RRH l with $\sum_{k \in \mathcal{K}_l} \left\|\boldsymbol{m}_k^l\right\|_2^2 \leq \tau$ into sleep mode and refine JT-clusters.
14: Calculate ω_x's using (34) and solve the bin-packing problem (35).

An iterative approach is taken between two sets of variables. One set of variables is the set of receive beamformers u_k's, and the other set includes $\boldsymbol{m}_k^l$ and r_k. First, we initialize $\boldsymbol{m}_k^l$'s (line 1), and then there is an outer loop (lines 3-12). In each iteration of the outer loop, the receive beamformers u_k's are updated (line 4) using the last updated $\boldsymbol{m}_k^l$'s, which are either from the initialization step or the previous iteration. Then with the fixed updated u_k's, an inner loop (6-10) is executed. In each iteration of the inner loop, the approximation points $\tilde{\epsilon}_k$'s are calculated (line 7), and problem (33) is solved using QCQP optimization. The purpose of the inner iteration is to make the function value $\log(\epsilon_k)$ obtained using Taylor approximation closer to its actual value, and the purpose of the outer loop is the iterative update of u_k's, $\boldsymbol{m}_k^l$'s, and r_k's. After completion of the outer loop, the active RRHs are determined, and the JT-clusters are refined. At the last stage, a bin-packing problem is solved to find the active BBUs and BBU-RRH mappings.

We should note that one of the issues in the joint optimization problems in wireless networks is the two-timescale problem. This issue arises when different variables of the problem should be determined in different timescales [6]. In our problem, the beamforming variables are determined in fast timescales (milliseconds), while RRH activation and BBU-RRH mappings are performed in slow timescales (minutes to hours) [26]. This issue can be solved using the standard techniques such as sample average approximation [27], in which the unknown channel is approximated using the known channel distribution for use in the slow time scale problem (i.e., BBU-RRH mappings and RRH activation). Then, at the start of each fast timescale period, only the beamforming part of Algorithm 1 is executed with fixed activated RRHs, JT-clusters, and BBU-RRH mappings.

V. Performance Evaluation

We use simulations to study the performance of JRBC and compare it with existing schemes. In the following subsections, the simulation setting and results are described.

A. Simulation Setup

Simulation scenario. We mainly adopt the simulation parameters from [19], [28]. A C-RAN architecture composed of 21 RRHs is considered. The RRHs are uniformly placed over an area of 2000 m by 2500 m resulting in an approximate distance of 500 m between neighboring RRHs. Each RRH is equipped with 2 transmit antennas. The locations of the users are modeled using the spatial Poisson Point Process (PPP) with the mean number of users ranging from 50 to 150 for the entire area. The RRH and user distributions are illustrated in Fig. 1. Each user is equipped with 1 receive antenna. Three categories of users are considered, with minimum data rate requirements of 2 Mb/s, 4 Mb/s, and 6 Mb/s with the priority ratio 1:2:3. We assume the maximum transmit power $P_{max}^{tx} = 40$ dBm for each RRH and $\eta = 0.36$ [13]. System bandwidth is assumed to be 10 MHz. The channel coefficients are calculated following the path-loss model L[dB] $= 128.1 + 37.6 \ \log_{10}(d$ [km]$)$, in which d is the distance between the transmitter and the receiver. Log-normal shadowing variance is set to 10 dB. In addition, it is assumed that the noise spectral density is -83.98 dBm. Static power consumption of each active BBU (i.e., P_c^{bbu}) is set to 40 W [5], and factor γ_1 is assumed to be 1. We also assume that $C_{fh} = 50$ Mb/s and $C_{bbu} = 50$ Mb/s [17]. The normalizing factors α and $\beta_k, \forall k \in \mathcal{K}$ are set to 2.4×10^{-6} and 8×10^{-7} respectively. Each point on the plots in the next sub-sections is averaged over 20 simulation runs.

Implemented Schemes. The performance of JRBC is compared to two baseline schemes:

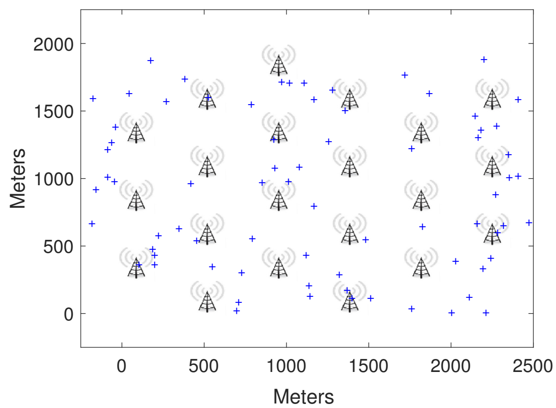

Fig. 1: RRH and user distribution: users are specified by "+".

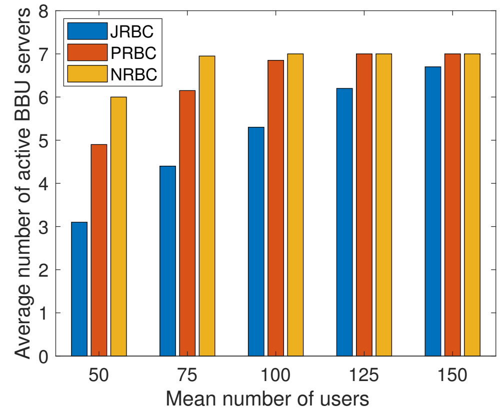

Fig. 2: Number of active BBUs with different number of users.

- *Non-Joint Radio and BBU resource allocation in C-RAN (NRBC)*: In this scheme, allocation of radio resources and allocation of BBU computing resources are completely independent. The works such as [7] and [8] which only study resource allocation in BBU side of C-RAN are based on such an approach.
- *Partially Joint Radio and BBU resource allocation in C-RAN (PRBC)*: This category includes the joint resource allocation studies in the literature, which end up separating BBU-RRH mapping from the rest of the problem. In this type of work, a part of the computing resource allocation in the BBU pool is considered jointly with the radio resource allocation side of the problem. The work in [12] is a representative of studies in this category.

Since RRH activation is not considered in the above mentioned schemes, we set P_{on}^{rrh} to 0 in our simulations to provide a fair comparison to those schemes. Also, the Best-Fit-Decreasing (BFD) algorithm [12] is used as the bin-packing algorithm for all schemes.

B. Results

We study the number of active BBUs under different network loads. Fig. 2 illustrates the results of this experiment. As shown in this figure, JRBC outperforms two other schemes because it jointly considers data rates and BBU parameters, resulting in choosing appropriate data rates which affect the number of active BBUs.

The power consumption of the three schemes is shown in Fig. 3. The power consumption is composed of transmit power consumption, the processing power consumption, and the power required for BBU activation. As it is seen in Fig. 3, JRBC has the lowest power consumption compared to other schemes. This is mainly because of the fewer number of active BBUs and also because of better transmit power management in the RRH side, which also affects the processing power management in the BBU side in our joint approach.

The last performance metric that we study in our simulations is system utility which is the weighted sum-rate minus power consumption. Fig. 4 plots the system utility versus different

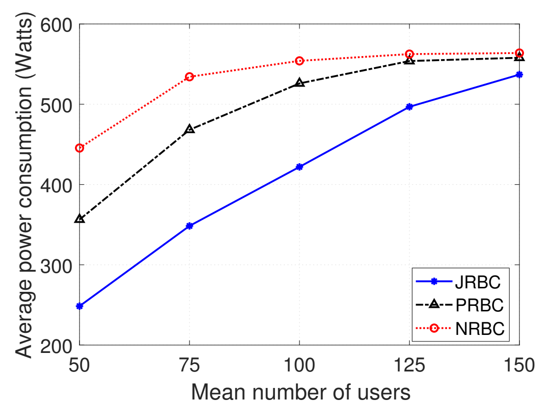

Fig. 3: Power consumption with different number of users.

network loads. As depicted in this figure, JRBC achieves more than 20% of system utility improvement over PRBC and NRBC in low and medium network loads. All three schemes show close performance in high loads. The reason is that the algorithms can utilize user diversity when the number of users is high. However, for the two other approaches, this comes with higher energy consumption as discussed above.

Overall, the simulation studies demonstrate the superiority of JRBC over the schemes that are either non-joint or partially joint approaches. Also, the simulations showed that less than 13 iterations of QCQP are required to achieve convergence in JRBC for all samples in all scenarios. This is shown in Fig. 5.

VI. Conclusion

In this paper, we studied joint radio and BBU resource allocation under C-RAN architecture. We formulated the problem as a weighted sum rate minus power optimization problem considering several RRH and BBU side constraints and user QoS requirements. After convexifying the problem, Lagrangian relaxation and KKT conditions were used to analyze the dependency between the user data rates in RRH side and the BBU side parameters and constraints. Based on this analysis, we designed an iterative algorithm to determine RRH

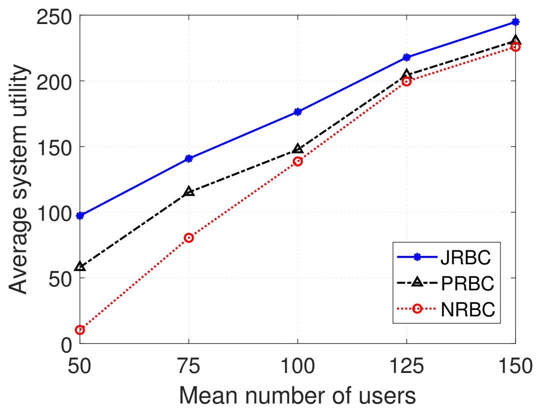

Fig. 4: Overall system utility under different number of users.

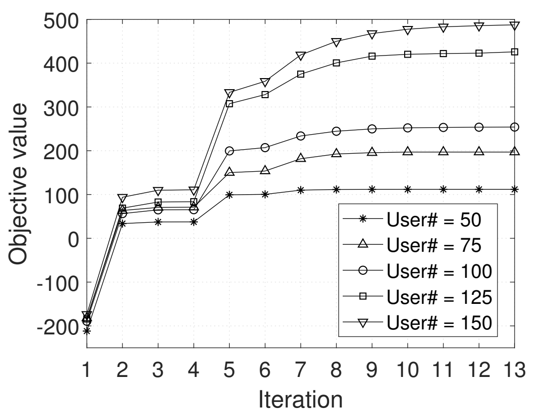

Fig. 5: Convergence behaviour of JRBC under different number of users

configurations and BBU-RRH mappings. Simulation results confirmed that the proposed joint scheme is more resource-efficient than existing approaches. This work can be extended by investigating more detailed models of signal processing on BBUs.

REFERENCES

[1] J. Yao and N. Ansari, "QoS-aware joint BBU-RRH mapping and user association in Cloud-RANs," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 4, pp. 881–889, Dec. 2018.

[2] A. A. A. Ari, A. Gueroui, C. Titouna, O. Thiare, and Z. Aliouat, "Resource allocation scheme for 5G C-RAN: A swarm intelligence based approach," *Computer Networks*, vol. 165, p. 106957, Dec. 2019.

[3] M. Peng, Y. Sun, X. Li, Z. Mao, and C. Wang, "Recent advances in cloud radio access networks: System architectures, key techniques, and open issues," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 2282–2308, 3rd Quart. 2016.

[4] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks: A technology overview," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 405–426, 1st Quart. 2015.

[5] H. Taleb, K. Khawam, S. Lahoud, M. E. Helou, and S. Martin, "A fully distributed approach for joint user association and RRH clustering in cloud radio access networks," *Computer Networks*, vol. 182, p. 107445, Dec. 2020.

[6] R. T. Rodoshi, T. Kim, and W. Choi, "Resource management in cloud radio access network: Conventional and new approaches," *Sensors*, vol. 20, no. 9, p. 2708, 2020.

[7] Y. Chen, W. Chiang, and M. Shih, "A dynamic BBU-RRH mapping scheme using borrow-and-lend approach in cloud radio access networks," *IEEE Syst. J.*, vol. 12, no. 2, pp. 1632–1643, June 2018.

[8] F. Marzouk, T. Akhtar, I. Politis, J. P. Barraca, and A. Radwan, "Power minimizing BBU-RRH group based mapping in C-RAN with constrained devices," in *Proc. IEEE ICC*, Dublin, Ireland, 2020, pp. 1–7.

[9] M. Qian, W. Hardjawana, J. Shi, and B. Vucetic, "Baseband processing units virtualization for cloud radio access networks," *IEEE Commun. Lett.*, vol. 4, no. 2, pp. 189–192, Apr. 2015.

[10] T. Sigwele, A. S. Alam, P. Pillai, and Y. F. Hu, "Evaluating energy-efficient cloud radio access networks for 5G," in *Proc. IEEE DSDIS*, Sydney, NSW, Australia, 2015, pp. 362–367.

[11] L. Ferdouse, A. Anpalagan, and S. Erkucuk, "Joint communication and computing resource allocation in 5G cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 9122–9135, Sep. 2019.

[12] K. Wang, W. Zhou, and S. Mao, "On joint BBU/RRH resource allocation in heterogeneous Cloud-RANs," *IEEE Internet Things J.*, vol. 4, no. 3, pp. 749–759, June 2017.

[13] Q. Liu, T. Han, and N. Ansari, "Energy-efficient on-demand cloud radio access networks virtualization," in *Proc. IEEE GLOBECOM*, Abu Dhabi, UAE, 2018, pp. 1–7.

[14] Z. Xu, Y. Wang, J. Tang, J. Wang, and M. C. Gursoy, "A deep reinforcement learning based framework for power-efficient resource allocation in cloud RANs," in *Proc. IEEE ICC*, Paris, France, 2017, pp. 1–6.

[15] W. Zhao and S. Wang, "Traffic density-based RRH selection for power saving in C-RAN," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3157–3167, Dec. 2016.

[16] M. Y. Lyazidi, N. Aitsaadi, and R. Langar, "Resource allocation and admission control in OFDMA-Based Cloud-RAN," in *Proc. IEEE GLOBECOM*, Washington, DC, USA, 2016, pp. 1–6.

[17] Y. L. Lee, J. Loo, T. C. Chuah, and L. Wang, "Dynamic network slicing for multitenant heterogeneous cloud radio access networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2146–2161, Apr. 2018.

[18] T. X. Tran and D. Pompili, "Dynamic radio cooperation for downlink Cloud-RANs with computing resource sharing," in *Proc. IEEE MASS*, Dallas, TX, USA, 2015, pp. 118–126.

[19] J. Tang, W. Tay, and T. Quek, "Cross-layer resource allocation with elastic service scaling in cloud radio access network," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5068–5081, Sep. 2015.

[20] T. Zhao, J. Wu, S. Zhou, and Z. Niu, "Energy-delay tradeoffs of virtual base stations with a computational-resource-aware energy consumption model," in *Proc. IEEE ICCS*, Macau, China, 2014.

[21] Z. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 57–73, Feb. 2008.

[22] Y. Zeng, E. Gunawan, Y. L. Guan, and J. Liu, "Joint base station selection and linear precoding for cellular networks with multi-cell processing," in *Proc. TENCON 2010*, Fukuoka, Japan, 2010, pp. 1976–1981.

[23] S. Shi, M. Schubert, and H. Boche, "Rate optimization for multiuser MIMO systems with linear processing," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 4020–4030, Aug. 2008.

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

[25] Y. Ye, *Interior Point Algorithms: Theory and Analysis*. Hoboken, NJ, USA: Wiley, 2011, vol. 44.

[26] J. Tang, R. Wen, T. Q. S. Quek, and M. Peng, "Fully exploiting cloud computing to achieve a green and flexible C-RAN," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 40–46, Nov. 2017.

[27] J. Tang, B. Shim, and T. Q. S. Quek, "Service multiplexing and revenue maximization in sliced C-RAN incorporated with URLLC and multicast eMBB," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 881–895, Apr. 2019.

[28] "3GPP TR 36.814, Evolved universal terrestrial radio access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release 9)," 3GPP, Sophia Antipolis, France, Tech. Rep., 2017.