Beamforming Design and BBU Computation Resource Allocation for Power Minimization in Green C-RAN

Xiaojun Yue[1], Kai Sun[1], Wei Huang[1], Xuemin Liu[1], Haijun Zhang[2]
[1] *College of Electronic Information Engineering, Inner Mongolia University, Hohhot, China*
[2] *University of Science and Technology Beijing*, Beijing, China
Email: sunkai@imu.edu.cn

ICC 2021 - IEEE International Conference on Communications | 978-1-7281-7122-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICC42927.2021.9500980

***Abstract*—This article focuses on the joint optimization of beamforming and baseband unit (BBU) computing resource allocation, with the goal of minimizing the network power consumption for the downlink cloud radio access network (C-RAN). To reduce the computational complexity, we split the joint optimization problem into two subproblems for transmission and computation respectively. The subproblem of minimizing power consumption for transmission is a network-wide beamforming design problem. By using positive semi-definite relaxation (SDR) technology, we transform it into a convex positive semi-definite programming (SDP) problem, which can be solved with effect. For the second subproblem, which intends to minimize the power consumption for computation, a computing resource allocation scheme based on the Simulated Annealing algorithm is proposed, which minimizes the active servers in the BBU pool to save power while meeting each user's computing resource requirement. The simulation results demonstrate that the algorithm proposed in this paper has a better performance compared with the existing algorithms.**

***Index Terms*—C-RAN, power consumption, BBU computation resource allocation, beamforming, Simulated Annealing**

I. Introduction

In the past few decades, the ever-increasing demands for data services and high-data-rate applications have placed higher demands on operators' network operations and management. Cloud radio access network (C-RAN) was proposed for the sake of reducing the energy consumption of communication system, maintaining operating costs and improving the network resource utilization [1]. As a new system architecture, C-RAN is composed of three parts: low-cost remote radio heads (RRHs), high-speed optical fiber transmission links and central baseband unit (BBU) pool with powerful computing capability [2]. In C-RAN, BBU pool aggregates a large amount of computing resources to perform the baseband processing functions of traditional base station, while RRH only performs basic signal processing functions and connects to BBU pool via fronthaul links. Centralized resource management and signal processing technology can be achieved under C-RAN architecture to significantly improve the network performance [3]. Moreover, unnecessary energy consumption can be reduced by dynamically adjusting resource allocation. In addition, RRHs are densely deployed around user equipment terminals (UEs) at low operating costs, which will significantly reduce the transmission power.

Despite the above-mentioned advantages, C-RAN also brings new challenges to wireless communication networks [4]. The dense deployment of RRHs may cause serious inter cell interference (ICI) and make C-RAN consume more power. Hence designing energy efficient beamforming matrix is a significant issue. Furthermore, the power consumed for baseband processing is also quite large, which depends on the allocation of computing resources. Unnecessary energy consumption and low resource utilization may occur due to the unreasonable resource allocation. Therefore, optimizing the beamforming matrix and designing reasonable resource allocation scheme according to UEs' different computing resource demands can significantly lower the network power consumption (NPC), which is the focus of our article.

In recent years, the problem of NPC minimization for different network architectures has been extensively studied in [5]–[14]. In [5] and [6], the beamforming matrix design and RRH selection were jointly optimized to reduce the power consumption of wireless transmission, i.e., the power consumption of RRHs and fronthaul links. Optimal resource allocation of BBUs were investigated in [7] and [8] to minimize the power consumed at the cloud side. Nevertheless, the above-mentioned researches did not consider the power consumption of wireless transmission and baseband processing at the same time. How to optimize NPC based on the overall architecture of C-RAN is still an vital issue, which has drawn a lot of attention [9]–[13]. Sigwele *et al.* [9] studied the pico base stations switching-off strategy at the radio side and the power optimization problem at the cloud side to maximize the energy efficiency in the heterogeneous downlink C-RAN. [10] and [11] both adopt the idea of queuing theory and divided the data transmission into cloud processing queue and wireless transmission queue. In order to meet UE's quality of service (QoS) while minimizing the NPC, Yao *et al.* [10] investigated the joint optimization of BBU-RRH mapping and the downlink user association. Tang *et al.* [11] minimized the system cost by jointly optimizing virtual machine (VM) activation and beamforming design. Wang *et al.* [12] studied the joint allocation of wireless resources and computing resources. An iterative resource allocation algorithm is proposed in [14] to solve the problem of wireless backhaul bandwidth allocation and power allocation in heterogeneous small cell networks, so as to maximize energy efficiency. Although the aforementioned papers attempted to minimize the NPC by solving problems such as beamforming design, RRH selection, and computing resource allocation, they did not take users' actual computing resource requirements into consideration. In order to characterize each user's computing resource demand, Xia *et al.* [13] modeled the NPC problem as a mixed time-scale issue and introduced a large-scale system analysis method. However, they assumed that all the users have identical computing resource requirements, which was somewhat idealized. In addition, when solving the BBU resource allocation problem, many works transformed it into a classic bin packing problem and used greedy algorithms such as best fitting decreasing (BFD) to solve it [10], [12]. Although these algorithms have low time complexity, they may fall into local optimal solutions.

Motivated by the above facts, our goal is to minimize the NPC while meeting users' demands for different computing resources. Hence, this paper studies the joint optimization of beamforming design and computing resource allocation in C-RAN. The main contributions of our work are summarized as follows:

1) The total power consumption of baseband processing and wireless transmission are comprehensively considered when studying the NPC problem for downlink C-RAN. In addition, we suppose that each UE has its specific computing task, and the physical server in the BBU pool creates a VM for each UE according to its computing resource requirements to perform this task.
2) We first split the joint optimization problem into two subproblems for transmission and computation respectively. The problem of minimizing the transmission power is a network-wide beamforming design problem. Since this problem is non-convex, we use positive semi-definite relaxation (SDR) technology to redesign it as a convex positive semi-definite programming (SDP) problem, the purpose of which is to design optimal beamformers for every UE and calculate their computing resource requirements. The second subproblem is the BBU computation resource allocation problem. Different from the algorithms in [10] and [12], this paper proposes a novel method for BBU computation resource allocation based on the Simulated Annealing algorithm, which aims to minimize the active VMs to save power under the premise of meeting users' different computing resource requirements.
3) The simulation demonstrates that the proposed scheme minimizes the NPC of the C-RAN system while satisfying UEs' computing resource requirements and practical constraints, and is superior to the competitive algorithms based on greedy strategy.

The remainder of this paper is organized as follows. System model and problem formulation is introduced in Section II. Next, our proposed solution for the formulated problem is presented in Section III. Simulation results are shown in Section IV and the conclusion is described in Section V.

II. System Model and Problem Formulation

Consider a downlink C-RAN system which includes K single-antenna users with various QoS requirements and L RRHs. Assuming that each RRH is equipped with N transmit antennas and serves users in the same time-frequency resource, as depicted in Fig. 1. Let $\mathcal{K} = \{1, \ldots, K\}$ and $\mathcal{L} = \{1, \ldots, L\}$ denote the set of UEs and RRHs, respectively. Moreover, we use $\mathcal{S} = \{1, \ldots, S\}$ to represent the collection of BBU servers and each server has identical maximum computation capability $C_s^{\max}$. Each RRH is assumed to be connected to the BBU pool through the fronthaul link and the l-th link has a maximum capability C_l^{FH}.

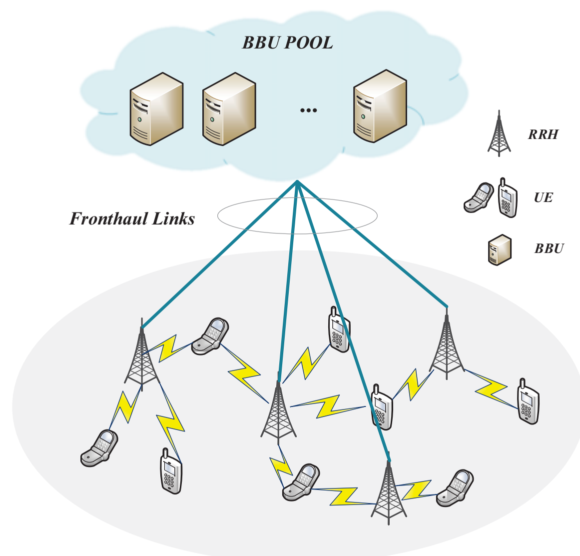

Fig. 1. C-RAN system

A. Transmission Model Description

Let s_k denote the information signal for UE k, which has a unit power, i.e., $\mathbb{E}\left[|s_k|^2\right] = 1$, and $\mathbf{w}_{lk} \in \mathbb{C}^{N\times 1}$ denote the transmit beamformer from RRH l to UE k. $\mathbf{h}_{lk} \in \mathbb{C}^{N\times 1}$ is the channel vector between RRH l and UE k. Assuming that each user can receive signals from multiple RRH and each RRH can also serve numerous users at the same time. Then the received signal at arbitrary UE k can be expressed as

$$y_k = \sum_{l\in\mathcal{L}} \mathbf{h}_{lk}^H \mathbf{w}_{lk} s_k + \sum_{l\in\mathcal{L}} \sum_{i\in\mathcal{K}\setminus k} \mathbf{h}_{lk}^H \mathbf{w}_{li} s_i + n_k, \tag{1}$$

where the first term represents the desired signal for UE k, the second term is the interference caused by signals sent by RRHs to other UEs, and $n_k \sim \mathcal{CN}\left(0, \sigma_k^2\right)$ denotes the additive white Gaussian noise (AWGN).

The signal-to-interference-plus-noise ratio (SINR) at UE k is given by

$$\gamma_k = \frac{\sum\limits_{l\in\mathcal{L}} \left|\mathbf{h}_{lk}^H \mathbf{w}_{lk}\right|^2}{\sum\limits_{i\in\mathcal{K}\setminus k} \sum\limits_{l\in\mathcal{L}} \left|\mathbf{h}_{lk}^H \mathbf{w}_{li}\right|^2 + \sigma_k^2}. \tag{2}$$

Consequently, the achievable rate at UE k is given by

$$R_k = B\log_2(1+\gamma_k), \tag{3}$$

where B is the bandwidth.

The maximum transmit power budget of RRH l is denoted as $P_l^{\max}$, then the transmit power constraints of RRH l is given by

$$\sum_{k\in\mathcal{K}} \|\mathbf{w}_{lk}\|_2^2 \le P_l^{\max}, \forall l \in \mathcal{L}. \tag{4}$$

Considering that if the BBU pool does not transmit user k's signal data to RRH l, the beamforming vector $\|\mathbf{w}_{lk}\|_2^2 = 0$, then l-th fronthaul link should meet capacity constraint

$$\sum_{k\in\mathcal{K}} \mathbb{1}\left\{\|\mathbf{w}_{lk}\|_2^2\right\} R_k \le C_l^{FH}, \forall l \in \mathcal{L}, \tag{5}$$

where $\mathbb{1}\left\{\|\mathbf{w}_{lk}\|_2^2\right\}$ is an indicative function, which is expressed by

$$\mathbb{1}\left\{\|\mathbf{w}_{lk}\|_2^2\right\} = \begin{cases} 0, & \|\mathbf{w}_{lk}\|_2^2 = 0, \\ 1, & otherwise. \end{cases} \tag{6}$$

B. BBU Computation Model

Like [13] and [15], it is assumed that each UE has one computation-intensive task U_k to be conducted in the BBU pool, which can be modeled as

$$U_k = (D_k, \tau_k, F_k), \forall k \in \mathcal{K}, \tag{7}$$

where D_k indicates the amount of output data of task U_k, τ_k is the total delay constraint and F_k denotes the total number of CPU cycles to complete the task. Tasks are scheduled to be executed on different servers in the BBU pool. Each server creates more than one VM with computation capacity A_{sk} for the purpose of dynamically providing each UE with computing resources required to accomplish task U_k.

We introduce the assignment matrix $\mathbf{X} = [x_{sk}] \in \mathbb{C}^{S\times K}$ to denote the schedule plan of tasks. If UE k is served by server s, $x_{sk} = 1$ and $x_{sk} = 0$ otherwise. Computing resources allocated to different users are limited by the total amount of computing resources on the server. We can express these constraints as

$$\sum_{s\in S} x_{sk} = 1, \forall k \in \mathcal{K}, \tag{8}$$

$$\sum_{k\in\mathcal{K}} x_{sk}A_{sk} \le C_s^{\max}, \forall s \in \mathcal{S}. \tag{9}$$

The constraint (8) indicates that each UE can only be served by one server during the task execution and the latter constraint is the total computing capacity of server s.

Based on the optimal beamformers, UE k's computation resource demand is determined by

$$A_k = \frac{F_k}{\tau_k - \frac{D_k}{R_k}}, \forall k \in \mathcal{K}. \tag{10}$$

It is obviously that for such a user who has a better QoS requirements needs more computation resource.

C. Power Consumption Model

As in [6], the power consumption of RRH l can be expressed as

$$P_l = \frac{1}{\eta}\sum_{k\in\mathcal{K}} \|\mathbf{w}_{lk}\|_2^2 + NP_l^N + P_c, \forall l \in \mathcal{L}, \tag{11}$$

where η accounts for the power amplifier efficiency of RRH l and P_l^N represents the power consumption for each antenna. As in [12], the fronthaul link power consumption of each RRH is modeled as a constant P_c for simplicity, which can be neglected.

The power consumption of s-th server can be formulated as

$$P_s = E_s + \sum_{k\in\mathcal{K}} x_{sk}\chi_{sk}A_{sk}, \forall s \in \mathcal{S}. \tag{12}$$

Here E_s represents the power consumption for an active server, and χ_{sk} is the weight factor. $\sum_{k\in\mathcal{K}} A_{sk} = 0$ states that server s does not provide computation resource to any user, so $P_s = 0$.

Then the overall system power consumption is formulated as

$$P_{total} = \sum_{l\in\mathcal{L}} P_l + \sum_{s\in\mathcal{S}} P_s. \tag{13}$$

D. Problem Formulation

Considering the aforementioned constraints especially computing resource requirements of different users, the power consumption optimization problem can be constructed as follows:

$$\begin{aligned} (\mathbf{P0}) \quad & \min_{\mathbf{x},\mathbf{w}} P_{total} \\ \text{s.t.} \; & \begin{cases} R_k \ge R_k^{\min}, \forall k \in \mathcal{K}, \\ x_{sk} \in \{0,1\}, \forall k \in \mathcal{K}, \forall s \in \mathcal{S}, \\ (4), (5), (8), (9). \end{cases} \end{aligned} \tag{14}$$

The first constraint expresses the minimum achievable rate for each UE k, which is related to the computing resources provided by the BBU server. Problem (P0) is a non-convex optimization problem as well as NP-hard. A simple but effective method to solve this problem is to split it into several subproblems which can be optimized independently. In our study, the optimization problem (P0) is decomposed into two parts: network-wide beamforming design and BBU computation resource allocation. In the following, details of the solutions to these two problems are given.

III. Problem Solutions

A. Network-Wide Beamforming Optimization

In order to meet each UE's QoS requirement while minimizing power consumption of wireless transmission part, this section proposes the optimal beamforming design without considering computation resource allocation. The network-wide beamforming optimization problem (P1) can be cast as:

$$\begin{aligned} (\mathbf{P1}) \quad & \min_{\mathbf{w}} \sum_{l\in\mathcal{L}} \frac{1}{\eta} \sum_{k\in\mathcal{K}} \|\mathbf{w}_{lk}\|_2^2 + P_c \\ \text{s.t.} \; & \begin{cases} R_k \ge R_k^{\min}, \forall k \in \mathcal{K}, \\ \sum_{k\in\mathcal{K}} \mathbb{1}\left\{\|\mathbf{w}_{lk}\|_2^2\right\} R_k \le C_l^{FH}, \forall l \in \mathcal{L}, \\ \sum_{k\in\mathcal{K}} \|\mathbf{w}_{lk}\|_2^2 \le P_l^{\max}, \forall l \in \mathcal{L}. \end{cases} \end{aligned} \tag{15}$$

Obviously, problem (P1) is non-convex due to the binary indicator in the second constraint and the interference term in the data rate expression. Hence, for the sake of converting (P1) into a form that is easy to deal with, we adopt the reweighted ℓ_1-norm method to approximate the non-convex ℓ_0-norm. The second constraint is then approximated by

$$\sum_{k\in\mathcal{K}} t_{lk} \left\|\mathbf{w}_{lk}\right\|_2^2 \tilde{R}_k \leq C_l^{FH}, \forall l \in \mathcal{L}, \tag{16}$$

with t_{lk} a constant weight, which is updated iteratively as follows

$$t_{lk} = \frac{1}{\left\|\mathbf{w}_{lk}\right\|_2^2 + \tau}, \tag{17}$$

and $\tau > 0$ is a small positive constant. $\tilde{R}_k$ denotes the achievable transmission rate of UE k at the algorithm's previous iteration.

Motivated by [16], we use the notation $\mathbf{W}_{lk} = \mathbf{w}_{lk}\mathbf{w}_{lk}^H, \forall l, k$ to achieve a convex reformulation of (P1). According to the equivalence relationship $\mathbf{W}_{lk} = \mathbf{w}_{lk}\mathbf{w}_{lk}^H \Leftrightarrow \mathbf{W}_{lk} \succeq \mathbf{0}$ and $\mathbf{W}_{lk}$ must satisfy $\operatorname{rank}\left(\mathbf{W}_{lk}\right) \leq 1$, we can rewrite (P1) compactly as

$$\begin{aligned} &\min_{\mathbf{W}_{lk}\succeq\mathbf{0}} \sum_{l\in\mathcal{L}} \frac{1}{\eta} \sum_{k\in\mathcal{K}} \operatorname{Tr}\left(\mathbf{W}_{lk}\right) + P_c \\ &\text{s.t.} \begin{cases} \sum\limits_{k\in\mathcal{K}} \operatorname{Tr}\left(\mathbf{W}_{lk}\right) \leq P_l^{\max}, \forall l \in \mathcal{L}, \\ \dfrac{\sum_{l\in\mathcal{L}} \operatorname{Tr}\left(\mathbf{H}_{lk}^H \mathbf{W}_{lk}\right)}{\sum_{i\in\mathcal{K}\backslash k} \sum_{l\in\mathcal{L}} \operatorname{Tr}\left(\mathbf{H}_{lk}^H \mathbf{W}_{li}\right) + \sigma_k^2} \geq \tilde{\gamma}_k, \forall k \in \mathcal{K}, \\ \sum\limits_{k\in\mathcal{K}} t_{lk} \operatorname{Tr}\left(\mathbf{W}_{lk}\right) \tilde{R}_k \leq C_l^{FH}, \forall l \in \mathcal{L}, \\ \operatorname{rank}\left(\mathbf{W}_{lk}\right) \leq 1. \end{cases} \end{aligned} \tag{18}$$

Here we transform the QoS targets into SINR targets of $\tilde{\gamma}_k = 2^{R_k^{\min}-1}, \forall k$. Similarly to [16], we can remove the non-convex constraint $\operatorname{rank}\left(\mathbf{W}_{lk}\right) \leq 1$ by using SDR technology. Then this problem is expressed in a convex semi-definite optimization form which can be derived a solution easily by using CVX, as presented in Algorithm 1.

Algorithm 1 Network-Wide Beamforming Strategy

Initialization: $n = 0$, $n_{\max}$, φ, $\tilde{R}_k^{(0)} = 1$, $t_{lk}^{(0)} = 1$
Repeat:
  For fixed t_{lk}, update $\mathbf{w}_{lk}$ by solving the SDP problem using standard convex optimization tools
  Compute $\tilde{R}_k$ according to (3)
  Update t_{lk} according to (17)
  Let $n = n + 1$
Until $n = n_{\max}$ or convergence

B. BBU Computation Resource Allocation Problem

The main goal of this subproblem is to allocate tasks to different servers for execution to save more power by minimizing the number of active VMs and provide computing resources according to the computing resource requirement of each UE.

According to (14), we can cast the computation resource allocation problem (P2) as

$$\begin{aligned} &(\mathbf{P2}) \quad \min_{\mathbf{x}} \sum_{s\in\mathcal{S}} P_s \\ &\text{s.t.} \begin{cases} x_{sk} \in \{0,1\}, \forall k \in \mathcal{K}, \forall s \in \mathcal{S}, \\ \sum\limits_{s\in S} x_{sk} = 1, \forall k \in \mathcal{K}, \\ \sum\limits_{k\in\mathcal{K}} x_{sk} A_{sk} \leq C_s^{\max}, \forall s \in \mathcal{S}. \end{cases} \end{aligned} \tag{19}$$

Previous studies usually describe the above problem as a bin packing problem, which has been proved to be NP-hard. Approximate heuristic bin packing algorithms such as the next fitting (NF), the first fitting (FF), and the best fitting decreasing (BFD), has been proposed to minimize the number of active servers to lower power. In this paper, we propose a novel computing resource allocation algorithm based on the Simulated Annealing algorithm for its ability to jump out of the local optimal solution. Similarly, we reduce the number of active servers as much as possible to further reduce the power consumption on the premise of meeting the UEs' computing resource requirements. Algorithm 2 presents the details of the SA based algorithm.

Algorithm 2 SA-based Computation Resource Allocation Strategy

Input:
  Maximum capacity set of servers $\{C_1^{\max}, \ldots, C_S^{\max}\}$
  Computing resource requirements of UEs $\{A_1, \ldots, A_K\}$
Output:
  Minimum power consumption $\sum_{s\in\mathcal{S}} P_s$
  Computation resource allocation $\{A_{s1}, \ldots, A_{sK}\}, \forall s \in \mathcal{S}, \forall k \in \mathcal{K}$
  Assignment of tasks to servers $\{x_{s1}, \ldots, x_{sK}\}, \forall s \in \mathcal{S}, \forall k \in \mathcal{K}$

$Temp = 500$
Generate initial random task assignment $\mathbf{X}$
Calculate power consumption of $\mathbf{X}$, $P_{curr} = \sum_{s=1}^{S} P_s(\mathbf{X})$
while Temp>0.001 **do**
  Generate new assignment $\mathbf{X}'$ by randomly selecting 3 tasks to allocate again, satisfying (8) and (9)
  Power consumption of $\mathbf{X}'$, $P_{new} = \sum_{s=1}^{S} P_s(\mathbf{X}')$
  Calculate power difference $\delta = P_{new} - P_{curr}$
  if $\delta \leq 0$ **then**
    Set task assignment $\mathbf{X}'$ as new state
  else
    Generate a random number $\eta \in \{0, 1\}$
    $P(\delta) = \eta$
    if $P(\delta) < e^{-\Delta/Temp}$ **then**
      Set task assignment $\mathbf{X}'$ as new state
    end if
  end if
  $Temp = \xi * Temp$ where $\xi \in \{0.9, 1\}$
end while

IV. Numerical results

The extensive simulations are performed in this section to evaluate the proposed algorithm. There are L RRHs and K users randomly distributed in the network. For convenience, the simulation parameters of all RRHs are set to be identical. Furthermore, we assume that the output data of each task is $D_k = 10000\text{bits}$ and CPU cycles $F_k = 1000$. Particularly, we assume that different task has different execution delay constraint, i.e., different UE has different computing resource requirement. Table I shows the other simulation parameters.

TABLE I
Simulation Parameters

Parameters	Values
System Bandwidth B	10 MHz
Power Amplifier Efficiency η	1/4
Noise Power Density	-127 dBm/Hz
Log-normal Shadowing	7 dB
Max RRH Transmit power $P_l^{\max}$	1 W
Fronthaul capacity C_l^{FH}	50 Mbps
BBU Computation Capacity $C_s^{\max}$	300
BBU static cost E_s	2 W
Distance-dependent Path Loss	$148.1 + 37.6 \log_{10} d_{lk}$

To better analyze the system performance, the proposed algorithm is compared with the following algorithms in [12]:

- First-Fitting-Based Algorithm: The FF-based resource allocation strategy is simple and has a low time complexity. In the FF algorithm, the first feasible server is chosen to provide computation resource for each UE's task rather than the best feasible one from global perspective.
- Best-Fitting-Decreasing-Based Algorithm: In the BFD algorithm, it tries to assign the task of each UE to the most suitable server with least capacity loss among those active servers.

In Fig. 2, we compare the average number of active servers of different algorithms under different UEs in the system. We can see that the performance of SA-based resource allocation scheme is superior to other schemes and the performance gap among different schemes becomes more obvious as the number of UEs grows. The reason why our proposed SA-based algorithm is better than others is that it allocates the task to the best fit server from global perspective to provide computing resources. In addition, we can see that the BFD algorithm and the FF algorithm have the same performance when the number of UEs is small. As the number of UEs increases, the BFD algorithm uses fewer servers to execute tasks than the FF algorithm.

Fig. 3 shows the NPC versus different number of UEs. With the increasing number of UEs, on the one hand, more transmission power is needed to meet the QoS requirements of all UEs. On the other hand, the power consumption of the

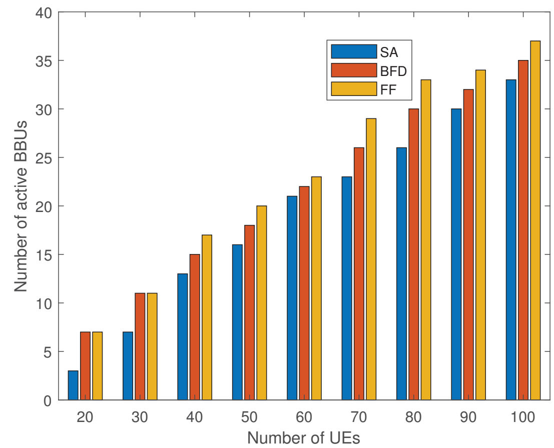

Fig. 2. BBU number in working mode versus the number of UEs

cloud side, i.e., the BBU pool, will increase with the increase of computing tasks, so the NPC increases gradually as the number of users increases. At the same time, the SA based resource allocation scheme reduces the power consumption by 3.3% and 4.5% respectively compared with the BFD algorithm and the FF algorithm when the number of UE is set to 80.

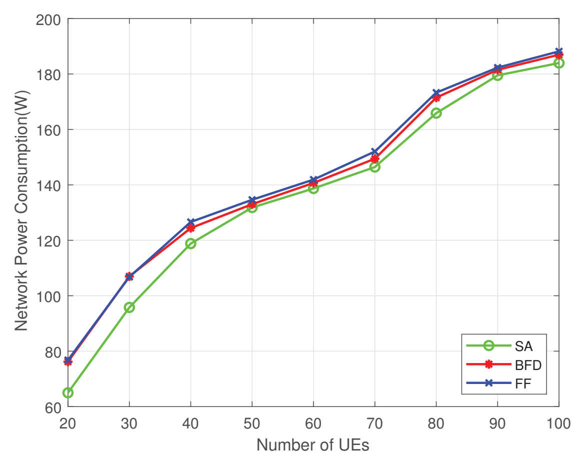

Fig. 3. Network power consumption versus the number of UEs

Moreover, different numbers of transmit antennas will affect dynamic emitted power and static hardware power consumption, thereby affect the NPC. As shown in Fig. 4, adding more antennas will substantially decrease the transmitting power consumption under different cases: all the users have the same computing resource requirements, denoted as "RRH-TRANS-SM"; different users have different computing resource requirements, denoted as "RRH-TRANS-DF". This means that more antennas will bring better energy-focusing and less propagation loss. However, with the number of transmit antennas increases when $N \leq 9$, the power consumption of the wireless transmission part (represented as "RRH-TOTAL") first decreases and then increases when $N > 9$. This is because, with the increase of the number of transmit antennas, the power consumption of static hardware will increase consistently according to the RRH power consumption model in (11), which exceeds the reduction of dynamic transmission power.

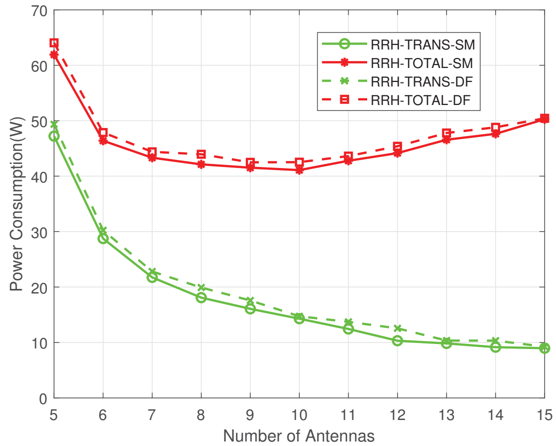

Fig. 4. Wireless transmission power consumption versus transmit antennas

Furthermore, Fig. 5 shows the influence of transmit antennas upon the NPC. It can be seen that with the number of transmit antennas increase, NPC achieved by all schemes decrease sharply to an optimal state and then increase slowly. This is because more antennas will increase the static hardware power consumption. Observation again shows that the SA-based algorithm has better performance than the classic algorithm.

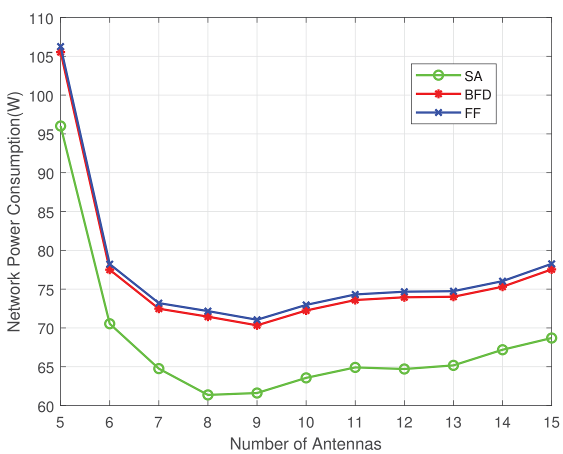

Fig. 5. Network power consumption versus transmit antennas

V. Conclusion

In our paper, a novel scheme of joint computation resource allocation and beamforming design is proposed, which aims to minimize the NPC of C-RAN, while ensuring UEs' different computation resource requirements. The optimization problem is decomposed into two subproblems: network-wide beamforming design and computation resource allocation. The non-convex beamforming optimization problem is transformed into SDP form and then solved by CVX. For the second subproblem, we propose a SA based resource allocation algorithm to minimize the power consumption of BBU pool by reducing the number of active BBU servers. Simulation results validate the advantages of the proposed algorithm.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (Nos. 61861034, 61461035), and in part by the Natural Science Foundation of Inner Mongolia Autonomous Region of China (Nos. 2019MS06009, 2018MS06004).

References

[1] 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, thirdquarter 2016.
[2] J. Wu, ”Green wireless communications: from concept to reality,” *IEEE Wireless Commun.*, vol. 19, no. 4, pp. 4-5, 2012
[3] D. Pompili, A. Hajisami and T. X. Tran, “Elastic resource utilization framework for high capacity and energy efficiency in cloud RAN,” *IEEE Commun. Mag.,* vol. 54, no. 1, pp. 26-32, Jan. 2016.
[4] M. F. Hossain, A. U. Mahin, T. Debnath, et al., “Recent research in cloud radio access network (C-RAN) for 5G cellular systems - A survey,” *J. Netw. Comput. Appl.*, vol. 139, pp. 31-48, Aug. 2019.
[5] Y. Shi, J. Zhang and K. B. Letaief, “Group sparse beamforming for green Cloud-RAN,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809-2823, May 2014.
[6] C. Pan, H. Zhu, N. J. Gomes and J. Wang, “Joint precoding and RRH selection for user-centric green MIMO C-RAN,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2891-2906, May 2017.
[7] C. Lin, W. Chien, J. Chen, C. Lai and H. Chao, ”Energy efficient fog RAN (F-RAN) with flexible BBU resource assignment for latency aware mobile edge computing (MEC) services,” in *Proc. IEEE VTC*, Sep. 2019, pp. 1-6.
[8] E. Aqeeli, A. Moubayed and A. Shami, “Power-aware optimized RRH to BBU allocation in C-RAN,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1311-1322, Feb. 2018.
[9] T. Sigwele, Y. F. Hu, and M.Susanto, “Energy-efficient 5G cloud RAN with virtual BBU server consolidation and base station sleeping,” *J. Netw. Comput.*, vol. 177, pp. 107302, 2020.
[10] 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.
[11] J. Tang, W. P. Tay, T. Q. S. Quek and B. Liang, “System cost minimization in cloud RAN with limited fronthaul capacity,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3371-3384, May 2017.
[12] K. Wang, W. Zhou, and S. Mao, ”On joint BBU/RRH resource allocation in heterogeneous Cloud-RANs,” *IEEE Internet Things J.*, pp. 749-759, Jun. 2017.
[13] W. Xia, J. Zhang, T. Q. S. Quek, S. Jin and H. Zhu, “Power minimization-based joint task scheduling and resource allocation in downlink C-RAN,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7268-7280, Nov. 2018.
[14] H. Zhang, H. Liu, J. Cheng and V. C. M. Leung, “Downlink energy efficiency of power allocation and wireless backhaul bandwidth allocation in heterogeneous small cell networks,” *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1705-1716, Apr. 2018.
[15] L. Shi, Z. Zhang and T. Robertazzi, “Energy-aware scheduling of embarrassingly parallel jobs and resource allocation in cloud,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 28, no. 6, pp. 1607-1620, Jun. 2017.
[16] E. Björnson, M. Kountouris and M. Debbah, “Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination,” in *Proc. ICT*, pp. 1-5, May 2013.