
Qiang Liu, Tao Han
University of North Carolina at Charlotte
Charlotte, 28223, United State
Email: liuqiang12040913@gmail.com, than3@uncc.edu

Gang Wu
University of Electronic Science and Technology of China
Chengdu, 611731, China
Email: wugang99@uestc.edu.cn

Abstract—Cloud radio access network (Cloud-RAN) has been recognized as one of the most promising architectures for next generation wireless network. The features of Cloud-RAN include the central signal processing mechanism and the flexible information interaction among cloud platforms. The General-Purpose Processors (GPP) is a more efficient architecture to multiplex computing resource in the cloud platform. In this paper, we study the impact of computing resource sharing on the total power consumption in downlink Cloud-RAN. Aiming to minimize the total power consumption of Cloud-RAN, we formulate the optimization problem with the consideration of the power consumption in the cloud platform, fronthaul links and base stations. By leveraging the iterative $l_0$ approximation method, we transform the optimization problem into a second-order conic programming (SOCP) problem. Then, we propose a two-loop iterative algorithm to obtain the optimal solution. Simulation results show that our proposed algorithm significantly reduces power consumption of Cloud-RAN as compared to the static algorithm in which the serve cluster assigned to each UE is static, and the GPP-based Cloud-RAN achieves better power consumption performance as compared to the traditional network and Virtual Base Station (VBS)-based network.

Keywords—Cloud-RAN; Computing resource; Energy saving; GPP;

I. INTRODUCTION

The fifth generation mobile communication system (5G) is expected to support higher data rates with massive communication connections, lower latency, ultra-higher reliability, and higher energy efficiency [1] than the fourth generation mobile communication system (4G), e.g., 3GPP long-term-evolution (LTE). To meet the exponential growing mobile data traffic demand, network densification has become a feasible and inevitable trend [2]. But the power consumption of traditional wireless networks will become very high and definitely increase the burden of the network operator. Therefore, the research on energy efficient network architecture is critical for the next generation wireless network.

A. Cloud Radio Access Network

Cloud radio access network (Cloud-RAN) is recognized as the most promising approach to achieve those goals and concepts [3], in which the baseband signal processing is shifted to a central baseband unit pool (BBU pool) such that efficient resource allocation is enabled. The function of fronthaul links is to deliver the data from the BBU pool to low-power remote radio heads (RRHs), mostly carried by huge-capacity and low-latency optical fiber. Figure 1 demonstrates the network architecture of Cloud-RAN. Leveraging this unique network architecture, the energy efficiency (EE) and spectrum efficiency (SE) can be improved simultaneously in
wireless networks, which is a remarkable advantage as compared with the traditional network architecture [4].

The Cloud-RAN has several advantages and gets extensive attentions. It is estimated that Cloud-RAN reduces capital expenditure by 15% and operational expenditure by 50% as compared to a traditionally distributed network [3]. Suryaprakash et al. proposed a model to analyze the deployment cost of a typical Cloud-RAN, and the results show that the Cloud-RAN system is more cost-effective than traditional LTE system [5]. The notable reduction in deployment costs mainly comes from the better adaptation to network traffic and the more efficient computing resource allocation. But it also indicated that the increment of traffic on fronthaul links is inevitable. Peng et al. presented the network architecture and key techniques of Cloud-RAN [6]. The authors presented that Cloud-RAN enables large-scale cooperative signal processing and networking techniques, which provides sufficient network capacity with a large number of communication links for 5G.

There are various challenges in deploying Cloud-RAN. The limited fronthaul capacity will restrict the performance of cooperative communications in Cloud-RAN, especially when the fronthaul traffic volume becomes unsustainable. Dai et al. studied the impact of limited fronthaul capacity on network downlink cooperative beamforming and investigated the tradeoff between the sum transmit power consumption and the sum fronthaul capacity [7]. Ha et al. studied the downlink joint transmission scheme in Cloud-RAN considering the fronthaul capacity constraints [8]. Meanwhile, Bartelt et al. investigated the heterogeneous fronthaul links in Cloud-RAN, aiming to alleviate the fronthaul constraint by exploiting mmWave communications [9]. In our previous work [10], we investigated the impact of heterogeneous fronthaul on network energy efficiency with comprehensive power consumption model. The results indicated that dynamic wireless fronthaul capacity allocation can effectively alleviates the damage of limited fronthaul capacity on network EE.

Unlike the traditional wireless radio access network in which base stations are the major power consumers, the power consumption of Cloud-RAN is determined by that of base stations, fronthaul links and cloud platform [11]. Considering the power consumption of base stations and fronthaul link, Y. Shi et al. studied energy minimization problem in a downlink Cloud-RAN and proposed low-complexity algorithms to jointly select RRHs and coordinate beamforming [12]. The results demonstrated that the power consumption of fronthaul links cannot be ignored and the Cloud-RAN can significantly reduce the total network power consumption. However, this work did not take the cloud platform into consideration. J. Tang et al. investigated the problem of minimizing the total power consumption including the power consumption of the cloud platform, fiber link and RRHs in Cloud-RAN [13]. It has been shown that jointly considering the power consumption of the cloud platform, fronthaul links and base stations can significantly reduce total network power consumption [11].

B. Related Work

In Cloud-RAN, cloud platform is the most critical part which undertakes all the signal processing and information sharing [14]. With the densification of base stations, the cloud platform of Cloud-RAN should support more user equipment, provide higher data rate and handle massive computation work. Traditional communication signal processing is undertaken by Digital Signal Processing (DSP) or Field-Programmable Gate Array (FPGA), which is inconvenient to update and maintain. The implementation of General-Purpose Processors (GPP) based communication signal processing has been realized [15]. GPP-based platform can effectively increases computing resource utilization ratio and can be compatible with virtualization technology. Generally, the computation capability of the cloud platform is constrained by its hardware and software, such as the GPP, memory, network interfaces and operation system. The impact of limited computation capacity in cloud platform cannot be ignored.

Tran et al. studied a downlink Cloud-RAN with computing resource sharing and tried to maximize the weighted sum-rate system utility, in which the maximum computing capacity of Virtual Base Station (VBS) pool is given and shared by all VBSs [16]. The results show that the improvement in computing-resource utilization of Cloud-RAN outperforms that of traditional RANs with the distributed computing-resource. Valenti et al. introduced the concept of computational outage and investigated its role in Cloud-RAN [17]. The computational outage caused by the limited computation capacity in cloud platform may lead to the transmission failure or inefficient resource utilization. The results demonstrated that if the limited computation complexity is properly considered, the computational outage can be reduced.

In fact, the cloud platform plays a significant role in reducing total network power consumption. The impact of limited computation capacity is not well studied, especially with the comprehensive network power con-
sumption model. Different from those work mentioned above, in this paper, we focus on total power consumption of Cloud-RAN including cloud platform, fronthaul links and base stations. We first build a comprehensive power consumption model for Cloud-RAN and consider GPP-based cloud platform with the limited computation capacity. To minimize total power consumption, we formulate the energy minimization problem as a non-convex optimization problem. In order to solve the problem, the objective function is transformed to a second-order conic programming (SOCP) problem by the iterative $l_0$ approximation method, and then we propose a two-loop iterative algorithm to obtain the optimal beamforming vector. The simulation results show that our proposed algorithm significantly reduces the power consumption of Cloud-RAN.

The reminder of this paper is organized as follows. In Section II, we describe the system model of downlink Cloud-RAN, including the network model and power consumption model. In Section III, the optimization problem is formulated, and a two-loop iterative algorithm is proposed to solve the problem. The simulation results presented in Section IV verify the effectiveness of our proposed algorithm. Finally, we draw the conclusion in Section V.

II. SYSTEM MODEL

A. Channel Model

We consider a downlink Cloud-RAN with $L$ RRHs and $K$ user equipment (UE), denote as $L$ and $K$ respectively, where each RRH has $M$ transmit antennas and each UE has one receive antenna. Each RRH connects to cloud platform with the ideal fiber fronthaul link. We assume that all UE information and channel information needed for resource allocation can be obtained. Each UE is cooperatively served by its serving cluster through joint downlink beamforming.

For the convenient representation, we introduce a network beamforming vector $w_k = [w^1_k, w^2_k, \ldots, w^L_k]$, where the dimension of $w^l_k$ is $M \times 1$. It is worthwhile to note that $w^j_k$ is not 0 means RRH $l$ is part of UE $k$’s serving cluster. Here, we assume that each RRH can serve every UE through downlink cooperative beamforming. Therefore, with linear transmit beamforming scheme at the RRHs, the received signal at UE $k$, denoted as $y_k$, can be denoted as

$$y_k = H_k w_k s_k + \sum_{j \neq k, j \in K} H_k w_j s_j + n_k$$ (1)

where $H_k$ denotes the channel state information matrix from all the $L M$ transmit antennas to UE $k$, $n_k$ is the received noise at UE $k$. In here, we assume that each UE had only single data stream and that UE’s signal is independent and identically distributed as $CN(0, 1)$. Therefore, the achievable data rate for UE $k$ can be written as

$$R_k = \log \left( 1 + \frac{w^H_k H^H_k H_k w_k}{\sum_{j \neq k, j \in K} H_k w_j w^H_j H^H_k + \sigma^2} \right)$$ (2)

B. Network Model

In this paper, we consider the GPP-based Cloud-RAN in which all signal processing are undertaken by GPPs in the cloud platform. Unlike the static hardware and software in traditional network architecture, the advantage of the GPP-based Cloud-RAN is the dynamic system configuration and flexible computing resource sharing. The diagram of the GPP-based cloud platform is depicted in Figure 2. In the cloud platform, utilizing the virtual machine (VM) techniques, control units (CU) can dynamically allocate GPPs for VMs to finish computing works. Dynamic GPPs allocation can effectively reduce the power consumption of cloud platform, especially when the network traffic volume is low.

Considering the GPP-based Cloud-RAN mentioned above, each VM performs the baseband processing for a set of UEs or base stations. VMs can effectively share the GPPs and achieve a high multiplexing gain. We assume that each VM needs certain computation resouce
to complete a UE’s signal processing. Here, we formulate the computation complexity model as

$$C_k = \phi (R_k)$$ (3)

where $R_k$ is the data rate of UE $k$, and $\phi (\cdot)$ is a increasing function which are widely adopted by [11], [18], [19].

In general, the computation capability of each VM can be modeled as a multi-dimensional function representing the capacities of the GPPs, memory, software algorithm and network interfaces. In order to evaluate the impact of limited computation capability, we formulate the constraints of computation capacity of VM $l$ as

$$C_l = \sum_k \| w_k^l \|_2^2 \phi (R_k) \leq C_{l}^{\text{max}}, \forall l \in \mathcal{L}$$ (4)

where $C_{l}^{\text{max}}$ denotes the maximum computation capacity.

C. Power Consumption Model

Traditional network energy minimization mainly focuses on the transmit power consumption in base stations. The power consumption in Cloud-RAN, however, consists of the power consumption in cloud platform, fronthaul links and base stations. In order to reduce power consumption, it is necessary to model the comprehensive network power consumption. Inspired by [11], we model the power consumption of the GPP-based Cloud-RAN as follows.

1) RRH Power Consumption: We consider the RRH sleep strategy and adopt a practical linear power consumption model:

$$P_{l}^{\text{RRH}} = \begin{cases} P_{l}^{\text{RRH}}_{\text{static}} + \frac{1}{\eta} \sum_k \| w_k^l \|_2^2 \sum_k \| w_k^l \|_2^2 & \neq 0 \\ P_{l}^{\text{RRH}}_{\text{sleep}} & \sum_k \| w_k^l \|_2^2 = 0 \end{cases}$$ (5)

where $P_{l}^{\text{RRH}}_{\text{static}}$ is the static power consumption of each RRH, and $\eta$ is a constant related to the hardware of RRH, such as the effectiveness of the power amplifier. Furthermore, each RRH can be switched into the sleep mode when the RRH only consumes a small amount of power $P_{l}^{\text{RRH}}_{\text{sleep}}$.

2) Fronthaul links Power Consumption: In Cloud-RAN, RRHs are connected to the cloud platform by fronthaul links. Normally, optical fiber is the most promising scheme which provides sufficient channel capacity and acceptable latency. Here, we consider the sleep strategy for fronthaul links [20] and formulate the fronthaul link power consumption model as

$$P_{l}^{FH} = \begin{cases} P_{l}^{FH}_{\text{static}} \sum_k \| w_k^l \|_2^2 & \neq 0 \\ P_{l}^{FH}_{\text{sleep}} \sum_k \| w_k^l \|_2^2 = 0 \end{cases}$$ (6)

where $P_{l}^{FH}_{\text{static}}$ is the static power consumption when fronthaul is in the active mode, and $P_{l}^{FH}_{\text{sleep}}$ is the sleep power consumption when fronthaul is in the sleep mode. We can see that the fronthaul link power consumption model can be unified into the RRH power consumption model.

3) Cloud Platform Power Consumption: Similar to the power consumption model in [19], [21], we model the power consumption of the cloud platform as

$$P_{l}^{C} = \kappa \cdot C_l$$ (7)

where $\kappa$ is the weighted value related to the implementation of the cloud platform.

As we can see from above discussion, in order to save energy, we can reduce the transmit power of RRHs, putting RRHs and fronthaul links into the sleep mode, and reduce the consumption of computing resources in the cloud platform.

III. OPTIMIZATION PROBLEM AND ENERGY SAVING ALGORITHM

A. Optimization Problem Formulation

We aim to find the optimal network beamforming vector $W$ to minimize the total power consumption while satisfying 1) the computation capacity constraints in VMs, the minimum data rate requirements of UEs, and the maximum transmit power constraints of RRHs. Therefore, the optimization problem can be formulated as

$$\begin{array}{ll}
\min_{\{w\}} & \sum_k \kappa \cdot \phi (R_k) + \frac{1}{\eta} \sum_k \| w_k^l \|_2^2 + (P_{FH}^{\text{static}} + P_{\text{RRH}}^{\text{static}}) \sum_l \| w_k^l \|_2^2 \\
\text{s.t.} & P_{k}^{\text{min}} \leq R_k, \forall k \in \mathcal{K} \\
& \sum_k \| w_k^l \|_2^2 \leq P_{l}^{\text{max}}, \forall l \in \mathcal{L} \\
& \sum_k \| w_k^l \|_2^2 \phi (R_k) \leq C_{l}^{\text{max}}, \forall l \in \mathcal{L}
\end{array}$$ (8)

where $C_{l}^{\text{max}}$ denotes the maximum computation capacity constraints of VMs. As we can see, the sleep strategy of
fronthaul links and RRHs are unified. The optimization problem is non-convex, and it is difficult to solve directly with classical convex optimization methods.

However, we notice that the minimum data rate constraints \( C_1 \) is satisfied with the equality at the optimal point. Therefore, the optimization problem (8) can be rewritten as

\[
\min_{\{w\}} \quad \sum_k \kappa \cdot \phi \left( R_{\text{min}}^k \right) + \frac{1}{\eta} \sum_k \|w_k\|_2^2 \\
\text{s.t.} \quad C_1, C_2, C_3
\]

B. Optimization Problem Transformation

The optimization problem (9) is still non-convex. In this section, we try to transform the optimization problem into a solvable one.

Firstly, the constraint \( C_1 \) is a non-convex constraint, we are here to transform it into a convex one. Since the phase of \( w_l \) will not change the objective function and constraints of (9), the minimum data rate constraints are equivalent to the following second order conic (SOC) constraints [12, 22]:

\[
\|r_k\|_2 \leq \sqrt{1 + \frac{1}{(2R_{\text{min}}^k/B - 1)}} \Re \{ R_{kk} \}, \forall k
\]

where \( r_k = [R_{k1}, R_{k2}, ..., R_{kK}, \sigma_k]^T \), \( R_{ki} = H_{ki} w_k \).

Secondly, since the \( l_0 \)-norm is non-convex, it is hard to solve the optimization problem. Hence, we using reweighted convex \( l_1 \)-norm to iteratively approximate the non-convex \( l_0 \)-norm [23, 24]:

\[
\sum_k \|w'_k\|_2^2 = \alpha_k \|w'_k\|_2^2
\]

\[
\sum_k \|w_k\|_2^2 = \beta_l \sum_k \|w_k\|_2^2
\]

with weighted value \( \alpha, \beta \) iteratively updated according to

\[
\alpha_k = \frac{1}{\|w'_k\|_2^2 + \varepsilon_1}, \forall k
\]

\[
\beta_l = \frac{1}{\sum_k \|w'_k\|_2^2 + \varepsilon_2}, \forall l
\]

where \( \varepsilon_1, \varepsilon_2 \) is constant value used to control precision.

C. Proposed Energy Saving Algorithm

Therefore, we propose a two-loop algorithm to solve the optimization problem [8]. The outer loop iteratively updates weighted value \( \alpha, \beta \) and the inner loop solves the approximated problem. The optimization problem after \( l_1 \)-norm approximation is rewritten as follow:

\[
\min_{\{w\}} \quad \sum_k \kappa \cdot \phi \left( R_{\text{min}}^k \right) + \frac{1}{\eta} \sum_k \|w_k\|_2^2 \\
\text{s.t.} \quad C_1, C_2, C_3
\]

\[
\sum_k \|w'_k\|_2^2 = \alpha_k \|w'_k\|_2^2
\]

\[
\sum_k \|w_k\|_2^2 = \beta_l \sum_k \|w_k\|_2^2
\]

We can see that the approximated problem (15) is a second-order cone programing (SOCP) problem, which can be effectively solved by classical convex tools.

Now, we summarize our proposed algorithm that solves the energy saving problem (8) in Algorithm 1.

Algorithm 1 Energy Saving Algorithm

1: **Initialization**: Set a feasible \( w_k^1 \) and initialize weighted value \( \alpha, \beta \) according to (13) and (14);

2: **repeat**

3: **for** the fixed \( \alpha, \beta \), solve the SOCP optimization problem (15) and obtain the optimal \( w_k^l \) using classical convex tools;

4: Update weighted value \( \alpha, \beta \) according to (13) and (14);

5: **until** convergence

IV. SIMULATION RESULTS AND ANALYSIS

In this section, the energy saving performance of proposed algorithm is evaluated with system simulation. We consider the GPP-based Cloud-RAN in a hexagonal deployment with the inter-site distance (ISD) of 500 meters, within which RRHS and UEs are randomly and uniformly distributed. The carrier center frequency of RRHs is 3.5 GHz. The total system bandwidth is 1 MHz. Each RRH equips 4 transmit antennas and each UE has one receive antenna. Maximum transmit power of RRHs is 6.3 Watt, the circuit power is 56 Watt and sleep power is 39 Watt. It is assumed that the path loss model is [140.7 + 36.7 \log_{10}(d)] for the links between RRHs and UEs. \( \eta, \lambda, \) and \( \kappa \) is 2.6, 1, 0.66, respectively. The noise power spectral density \( N_0 \) is -104 dBm.

The default value for the number of RRHs and UEs is 5 and 10, respectively. The minimum data rate requirement of UEs is 2 Mbps and VM’s maximum
computation capacity is 100 GFLOPS (Giga Floating-point Operations Per Second) with medium constraints. In this paper, we adopt the computation complexity model \( \phi(R_k) = \lambda(R_k)^3 \), which are widely adopted by [19] [26].

In here, we explain the following networks used in our simulation:

- **Traditional network**: Base stations transmit signals with their maximum transmit power and cannot switch base stations into the sleep mode, the power consumption always stays high.
- **VBS-based Cloud-RAN**: The cloud platform uses Virtual Base Stations (VBSs) to deal with digital signal processing. Each RRH corresponding to a VBS in the cloud platform, but each VBS are independent of each other and the computing resource cannot be shared flexibly.
- **GPP-based Cloud-RAN**: The cloud platform uses GPPs to carry out digital signal processing. By using the virtual machine technology, computing resources can be shared flexibly. As described in Section II, UE can be served by RRHs dynamically. Especially, we also consider static algorithm in which the serving cluster of each UE is static.

Figure 3 shows the impact of the number of UE on the total power consumption under different network architectures. When the number of UE increases, the total power consumption increases gradually because of the network traffic volume growth. Meanwhile, the gap between the total power consumption of the proposed algorithm and that of the static algorithm becomes larger due to the power saving of the dynamic cluster mechanism. When the network traffic volume is low, the proposed algorithm can reduce more power consumption than the other algorithms. This is because the proposed algorithm enables efficient computing resource sharing in the cloud platform. However, the performance difference among different algorithms and networks becomes less significant with a large number of UE, e.g., 20 UEs. Because higher traffic volume means that Cloud-RAN should switch on more RRHs and computing resources in order to support UEs’ traffic demand. Furthermore, the total power consumption of the traditional network stays high because it cannot effectively adapt to network traffic dynamics.

Figure 4 compares the total power consumption of different algorithms and network architectures under different numbers of RRH with the same number of users. It can be observed that the total power consumption is almost linear versus the number of RRH. This is because even RRH is switched into sleep mode, it still consumes sleep power \( P_{RRH}^{static} \), so the more deployment of RRH means more power consumption mostly consists of sleep power of RRHs. With the increment of the number of RRH, the gap of total power consumption between proposed algorithm and other algorithms becomes larger. The reason behind is that more RRHs can be switched into sleep mode and more computing resource can be shared effectively. The total power consumption of the proposed algorithm is less than the
other algorithms, which also can be interpreted as the benefit of the computing resource sharing and dynamic clustering mechanism. Therefore, the GPP-based Cloud-RAN can effectively reduce total power consumption in response to the network traffic dynamics.

Figure 5 shows the impact of the maximum computation capacity of VMs on the power consumption of the cloud platform. Here, the computation capacity of a VM refers to as the capability of the VM in handling the signal processing and network traffic. We can see that the power consumption of cloud platform decreases when the maximum computation capacity increases. This because a larger computation capacity allows higher flexibility in optimizing the computing resource allocation. When the computation capacity is sufficient, the power consumption of the cloud platform will not further decrease for all the algorithm. This indicates that the minimum power consumption is reached. Based on this observation, we can see that GPP-based Cloud-RAN remarkably reduces the energy consumption by dynamically sharing the computing resource.

V. CONCLUSION

In this paper, we investigated the impact of the computing resource sharing on the total power consumption in downlink Cloud-RAN. Aiming to minimize the total power consumption of Cloud-RAN, we formulated the optimization problem with a comprehensive power consumption model considering the power consumption of the cloud platform, fronthaul links and base stations. By leveraging the iterative $l_0$ approximation method, we transformed the optimization problem into a SOCP problem, and then proposed a two-loop iterative algorithm to obtain the optimal solution. Simulation results show that the proposed algorithm has significantly reduced the total power reduction of Cloud-RAN as compared to the static algorithm. In addition, we have showed that the GPP-based Cloud-RAN can achieve higher power consumption gain that the traditional network and the VBS-based network.

REFERENCES


