This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2799859, IEEE Transactions on Wireless Communications

Joint Virtual Computing and Radio Resource Allocation in Limited Fronthaul Green C-RANs

Phuong Luong, *Student Member, IEEE*, François Gagnon, *Senior Member, IEEE*, Charles Despins, *Senior Member, IEEE*, and Le-Nam Tran, *Senior Member, IEEE*

***Abstract*—We consider the virtualization technique in the downlink transmission of limited fronthaul capacity cloud-radio access networks (C-RANs). A novel virtual computing resource allocation (VCRA) method which can dynamically split the users workload into smaller fragments to be served by virtual machines is presented. Under the proposed scheme, we aim at maximizing the network energy efficiency by a joint design of virtual computing resources, transmit beamforming, remote radio head (RRH) selection, and RRH-user association. Moreover, we construct a more realistic fronthaul power consumption model, which is directly proportional to users' rate transmitted by the corresponding RRHs. The formulated problem is combinatorial and difficult to solve in general. Our first contribution is to customize a branch and reduce and bound (BnRnB) method to attain a globally optimal solution. To compute a high-quality approximate solution, a standard routine is used to deal with the continuous relaxation of the original problem. However, the proposed continuous relaxation is non-convex which implies another challenge. For a practically appealing solution approach, we resort to a local optimization method, namely the difference of convex algorithm (DCA). Our second contribution is on the use of Lipschitz continuity to arrive at a sequence of convex quadratic programs, which can be solved efficiently by modern convex solvers. Finally, a post-processing procedure is proposed to obtain a high-performance feasible solution from the continuous relaxation. Extensive numerical results demonstrate that the proposed algorithms converge rapidly and achieve near-optimal performance as well as outperform other known methods. Moreover, we numerically show that the VCRA scheme significantly improves the system energy efficiency compared to the existing schemes.**

***Index Terms*—Beamforming, cloud radio access networks, computational provisioning, limited fronthaul, optimization, virtualization, network slicing.**

I. Introduction

The development of the next generation wireless network, commonly referred to as 5G, is underway. In general, 5G is expected to meet a set of challenging requirements that can solve many problems in existing wireless systems [1]. Some key technologies for 5G are introduced in [2]. From a network architecture viewpoint, cloud-radio access networks (C-RANs) have received growing attention as a powerful candidate to implement 5G standards. In particular, C-RANs can significantly enhance both system spectral and energy efficiency (EE), and satisfy other quality-of-service (QoS) requirements. In C-RANs, signals transmitted or received by low-power remote radio heads (RRHs) are processed by the centralized baseband unit (BBU) pool, comprising several physical servers (PSs) on a cloud-computing platform [1], [3]. A RRH is typically simplified with only radio frequency (RF) functions to handle the transmission/reception of radio signals to/from the users. On the other hand, sophisticated baseband signal processing tasks are migrated to the BBU pool, i.e., on the cloud. In this way, the virtualization and network slicing technology can be used to deliver dynamic and powerful resource allocation. As a main feature of cloud-computing, virtualization at the BBU pool enables each PS to dynamically split the dedicated computing resources into various virtual machines (VMs) sizes, depending on data traffic. Some noticeable results in this regard were reported in [4].

Although the benefits of C-RAN technology, with the use of distributed RRHs, are relatively convincing from a viewpoint of centralized resource management, it still raises a serious concern over the energy power consumption. In fact, there are many power consumption sources in a C-RAN, ranging from the circuit powers and RF transmission power to the fronthaul data transportation and processing, fronthaul maintenance power, computational power from the cloud, etc. In light of this, designing greener resource allocation which maximizes the EE of C-RANs has become an essential criteria for viable C-RAN solutions. A variety of our studies on the energy efficiency maximization, the power consumption minimization as well as the maximization of the trade-off between total power consumption and throughput in the C-RAN by jointly optimizing the radio resource allocation given a fixed setting of cloud computing capability have been carried out in [5]–[8]. Despite many potential benefits, the understanding on energy-efficiency of virtualized C-RANs with limited-capacity fronthaul is far from comprehensive, this paper should contribute to thoroughly clarify the issue. From the virtualization standpoint, an energy-efficient design must be able to adapt computing resources to elastic traffic [9]–[11]. From the energy-efficiency maximization perspective, it should maximize the system spectral efficiency while consuming the least power. Thus, PSs on the cloud can be activated wisely, otherwise the power consumption for signal processing taking place in the BBU pool dominates other power consumption sources in the network, leading to a poor energy-efficiency performance [12]. More explicitly, some PSs should be switched OFF to save power while others must be active to maintain the system quality-of-service (QoS). Moreover, the number

Phuong Luong, Charles Despins, and François Gagnon are with École de Technologie Supérieure (ÉTS), Montreal, QC, Canada. (email: {thi-thu-phuong.luong.1@ens.etsmtl.ca}, {francois.gagnon,charles.despins}@etsmtl.ca).

Le-Nam Tran is with the School of Electronical and Electronic Engineering, University College Dublin, Ireland. (email: nam.tran@ucd.ie).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2799859, IEEE Transactions on Wireless Communications

of RRHs and its associated fronthaul links can be very large in a dense C-RAN, generating a huge pressure on the total network power consumption. As a result, designing a green C-RAN must also consider the RRH selection together with the RRH-user association problem, concerning the limited-capacity fronthaul [13]. These motivations call for new radio resource management methods to design energy-efficient C-RANs.

Energy-efficient approaches for C-RANs have been presented in the recent literature, especially for joint design of RRH selection and RRH clustering [10], [14]–[20]. Specifically, to minimize network power consumption, the authors in [14] considered a joint design of cell activation and spectrum allocation, using the reweighted ℓ_1-norm approximation technique. Similarly, a smoothed ℓ_p-norm minimization method was involved to deal with the RRH selection and user admission problem in a multi-cast C-RAN [16]. The work of in [18] studied joint optimization of beamforming and user association for green C-RANs. In [21], stochastic optimization was applied to solve the problem of queue-aware joint RRH selection and beamforming to minimize the network power consumption. Queue-aware energy efficiency maximization was also studied for heterogeneous C-RAN in [20]. A stochastic game approach was proposed in [10] to allow virtual base stations to learn the cellular traffic variation and thus enable to switch OFF some RRHs to reduce the overall energy consumption in C-RANs. Unlike these, a recent and pragmatic technology which deploys a cached server at each base station to alleviate the fronthaul congestion and improve the system energy efficiency was considered in [22]. Given a cache strategy, the authors developed a framework to analyze and derive the EE closed-form expression on top of achieving an optimal cache policy which maximizes the network EE [22]. The energy-efficient resource allocation considering limited fronthaul capacity was studied in [23], [24]. Specifically, in [23], Z. Wang *et al.* aimed to maximize the sigmoidal function of user's SINR under imperfect CSI condition and limited backhaul capacity. In [24], D. Ng *et al.* studied an EE resource allocation in multi-cell limited backhaul OFDMA downlink networks, where zero-forcing beamforming and semi-orthogonal user selection policies were employed prior to the EE problem formulation. The trade-off between EE and spectral efficiency was investigated in [25]. There also exists the trade-off between the EE and the system delay as shown in [26].

In recent years, virtualization for wireless communications has proved to be a powerful technology to fully and flexibly exploit the computing resources of C-RANs [4]. In [27], the authors considered a framework of dynamic request allocation of VMs to minimize the VM computing cost. A heuristic algorithm was developed in [28] to design the VM placement and network element switching ON/OFF to conserve more computing power. In [29], an integer linear program was formulated for the problem of dynamic VM provisioning and allocation. For a joint design of radio resource management and network virtualization for C-RANs, the work of [9] proposed a virtual base station (VBS) cluster associated with a subset of RRHs which can adapt to traffic dynamics. Likewise, the authors in [30] leveraged the VBS concept and optimized VBS formation using an mixed-integer linear program. A joint design of VM computation capacity, RRH selection, and beamforming to minimize the total power consumption in C-RANs was proposed in [11], [31]. Similarly, the authors in [32] considered the joint computing capacity and beamforming design for energy minimization problem in a mobile cloud computing network. Throughput maximization in C-RAN was studied in [33], taking into account the constraint of computing resource capacity of VBS pool.

In this paper we study an energy-efficient design of a virtualized C-RAN with limited-capacity fronthaul. Unlike [23], [24], where the virtual computing resource at the central unit or BBU pool is not accounted for, we consider a joint optimization of transmit beamforming, virtual computing resource allocation, RRH selection, and the RRH-user association to maximize the global network energy efficiency. Compared to the recent literature [9]–[11], [31], [32], we first propose a novel virtual computing resource allocation (VCRA) scheme. In particular, to best exploit the VMs in each PS, the proposed VCRA method splits the users' workload into smaller pieces that can be served by different VMs in parallel. The distinguishing features of the proposed VCRA scheme are as follows: (i) data traffic of a user can be processed by heterogeneous VMs; (ii) VM's computing capacity is dynamically allocated according to the traffic condition; (iii) the assignment of VMs to PSs is done in such a way that the number of unused PSs is maximal. Furthermore, to quantify the power consumption more accurately, we introduce a new power consumption model which includes the rate dependent fronthaul power consumption. More specifically, the fronthaul power consumption model is computed based on the total transmission rates served by the corresponding RRH, which is more realistic and different from that in [16], [18], [31], where fronthaul power is a quadratic or linear function of involved variables.

We formulate the problem as a mixed-integer non-convex program, for which is generally difficult to find an optimal solution. Even if possible, the complexity is prohibitively high since a mixed-integer program is commonly known to be NP-hard. Solving the problem is far more challenging for several reasons: (i) the non-convexity of the cost function, (ii) the non-convexity of the limited-capacity fronthaul and the cross-layer delay constraints, and (iii) the combinatorial nature of the selection and assignment procedure. The said non-convexity actually implies that the continuous relaxation of the problem is non-convex. In fact, this attribute makes known mixed-integer optimization solvers of no use to solve the considered problem, which motivates us to develop an optimal algorithm. To this end we first propose multiple novel transformations to reformulate the original problem into a form that amenable to monotonic optimization [34]. For a more practically appealing method, we develop a low complexity algorithm based on difference of convex algorithm (DCA) [35], approximating the original problem by a series of convex quadratic programs, using Lipschitz continuity [36]. Our contributions are the following:

- We propose a novel VCRA strategy at the BBU pool and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2799859, IEEE Transactions on Wireless Communications

consider a joint optimization problem of the beamforming, VCRA, RRH selection and RRH-user association. To formulate the problem of interest we introduce several binary preference variables. The objective is to maximize the overall network energy efficiency under explicitly limited-capacity fronthaul constraints. We customize a branch-and-reduce-and-bound (BnRnB) algorithm to compute a globally optimal solution to the formulated problem, which is a mixed-integer non-convex program.

- To find a high-quality low-complexity solution, we first deal with the continuous relaxation of the problem and then propose a post-processing procedure to recover binary variables. This way seems to be a standard mixed-integer programming but, as mentioned earlier, the relaxed problem is non-convex and still difficult to solve. Our aim is to solve the relaxed problem using a local optimization method called DCA [35], which has been shown to be very effective in may applications. The proposed DCA method can cope with the non-convex limited fronthaul constraints without assigning a fixed rate as done in [23] or employing zero forcing beamforning and user selection as in [24] before solving the optimization problem. To this end, we invoke the concept of Lipschitz continuity to rewrite the relaxed problem as a DC program. This reformation offers two benefits. First, no slack variable is introduced in the DC program, which is different from previous publications in the similar context [5], [37]. Second, the resulting DC program can be easily approximated by a sequence of convex quadratic optimization problems using DCA. Finally, a post-processing algorithm is then carried out to search for a high-performance binary solution.
- Extensive numerical results are presented to show the efficiency of our proposed algorithms in terms of the convergent rate and achievable energy efficiency performance, compared to other existing methods. In particular, the numerical results also demonstrate that the proposed VCRA scheme significantly outperform the known methods.

The rest of the paper is organized as follows. Section II introduces the system model. Section III formulates the joint design of the energy efficiency maximization problem and Section IV presents a global optimization algorithm. In Section V, we propose a low-complexity algorithm to find a high-quality feasible solution. Section VI presents numerical results and insight discussions under different simulation setups. Finally, the conclusion of the paper is given in Section VII.

Notation: We use bold uppercase and lowercase letters to denote matrices and vectors, respectively. $\mathbb{C}$ and $\mathbb{R}$ represent the space of complex and real numbers. $\mathbf{x}^T$ and $\mathbf{x}^H$ stand for the transpose and Hermitian operation of vector $\mathbf{x}$. $|x|$ and x^* represent the modulus and complex conjugate of $x \in \mathbb{C}$, respectively, while $\|\mathbf{x}\|_2$ is the ℓ_2-norm of the vector $\mathbf{x}$. In addition, $\|\mathbf{X}\|_F$ is the Frobenius norm of matrix $\mathbf{X}$. $\mathbb{E}\{\cdot\}$ denotes the expectation operator. $\mathrm{Re}(.)$ stands for the real part of the argument. $\mathcal{O}$ and $blkdiag(\mathbf{X_1}, \mathbf{X_2}, \ldots)$ represent the big O notation and block diagonal matrix whose diagonal elements are matrices $\mathbf{X_1}, \mathbf{X_2}, \ldots$, respectively.

II. System Model

A. Transmission Model

We consider the downlink of a C-RAN consisting of I RRHs and K single-antenna user equipments (UEs). We denote by $\mathcal{I} = \{1, \ldots, I\}$ and $\mathcal{K} = \{1, \ldots, K\}$ the set of RRHs and UEs, respectively. The ith RRH is equipped with M_i antennas,$\forall i \in \mathcal{I}$. As shown in Fig. 1, we assume that all the RRHs are connected to BBU pool via the fronthaul links, e.g., high-speed optical ones, where the ith link has a maximum capacity C_i^{FH}. Each UE is served by a specific group of RRHs but one RRH can serve more than one UEs simultaneously. Let s_k be the signal with unit power, i.e., $\mathbb{E}\{s_k s_k^*\} = 1$, intended for the kth UE and $\mathbf{w}_{i,k} \in \mathbb{C}^{M_i \times 1}$ be the beamforming vector from the ith RRH to the kth UE. The vector of channel coefficients from the ith RRH to the kth UE is represented by $\mathbf{h}_{i,k} \in \mathbb{C}^{M_i \times 1}$. In this work, we assume perfect channel state information (CSI) between the RRHs and the UEs.[1] For notational convenience, we denote the set of beamforming vectors intended for the kth UE as $\mathbf{w}_k \triangleq [\mathbf{w}_{1,k}^T, \mathbf{w}_{2,k}^T, \ldots, \mathbf{w}_{I,k}^T]^T \in \mathbb{C}^{M \times 1}$, and the vector including the channels from all RRHs to the kth UE as $\mathbf{h}_k \triangleq [\mathbf{h}_{1,k}^T, \mathbf{h}_{2,k}^T, \ldots, \mathbf{h}_{I,k}^T]^T \in \mathbb{C}^{M \times 1}$, where $M = \sum_{i \in \mathcal{I}} M_i$. Using these notations, the received signal at the kth UE is given by

$$y_k = \mathbf{h}_k^H \mathbf{w}_k s_k + \sum\nolimits_{j \in \mathcal{K} \setminus k} \mathbf{h}_k^H \mathbf{w}_j s_j + z_k \tag{1}$$

where $z_k \sim \mathcal{CN}(0, \sigma_0^2)$ is the additive white Gaussian noise (AWGN) and σ_0^2 is the noise power. We normalize the noise power factor to 1 for the sake of notational simplicity in the rest of the paper. Note that in (1), we have assumed that the kth UE is connected to all the RRHs, but the ith RRH serves the kth UE only if $\left\|\mathbf{w}_{i,k}\right\|_2^2 > 0$. By treating interference as noise, the achievable rate in b/s/Hz for a given set of channel realizations at the kth UE is given by

$$R_k(\mathbf{w}) = \log_2(1 + \Gamma_k(\mathbf{w})), \tag{2}$$

where

$$\Gamma_k(\mathbf{w}) = \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum_{j \in \mathcal{K} \setminus k} |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma_0^2} \tag{3}$$

where $\mathbf{w} \triangleq [\mathbf{w}_1^T, \mathbf{w}_2^T, \ldots, \mathbf{w}_k^T]^T \in \mathbb{C}^{(KM) \times 1}$ is vector stacking the beamformers for all UEs.

In C-RAN systems, CSI between RRHs and UEs is exchanged between RRHs and the BBU pool via the fronthaul links. The cost of the incurred overhead scales with the amount of CSI fed back to the cloud via fronthaul links. Thus, a certain portion of fronthaul capacity is reserved for this overhead information transportation to and from the BBU pool. This reduces each effective fronthaul capacity budget C_i^{FH} for the data transmission, which may degrade the overall network performance in terms of network throughput and EE. For simplicity, we assume that this overhead fronthaul capacity

[1] In practice, CSI between RRHs and UEs is estimated by exploiting the channel reciprocity between the UL and DL transmissions in the time division duplexing system or by the feedback channels in the frequency division duplexing system.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2799859, IEEE Transactions on Wireless Communications

reservation is done prior to our problem formulation. Each limited fronthaul capacity budget C_i^{FH} is now reserved for the data transportation. The benefit of RRHs coordination is limited by the overhead of pilot-assisted channel estimation [38]. Estimating a subset of channel coefficients, rather than all the channel coefficients from all UEs can simply reduce the overhead of CSI and signaling, but restricts the cooperation within a limited number of RRHs, resulting in the loss of system throughput. We note that data for the kth UE is routed from the BBU pool to the ith RRH via the ith fronthaul link only if $\|\mathbf{w}_{i,k}\|_2^2 > 0$. Let binary variables $a_{i,k} \in \{0,1\}, \forall i \in \mathcal{I}$ and $k \in \mathcal{K}$ represent the association status between the ith RRH and the kth UE, i.e., $a_{i,k} = 1$ implies that the kth UE is served by the ith RRH and $a_{i,k} = 0$, otherwise. Then, the per-fronthaul capacity constraints can be

$$\sum\nolimits_{k\in\mathcal{K}} a_{i,k} R_k(\mathbf{w}) \le C_i^{\text{FH}}, \forall i \in \mathcal{I}. \tag{C1}$$

B. Proposed Virtual Machine Computing Model

We consider a BBU pool consisting a set of $\mathcal{S} = \{1, \ldots, S\}$ physical servers (PSs). The proposed VCRA scheme is described as follows. Assume that each PS is capable of creating multiple virtual machines (VMs) to process the incoming packets in parallel. Unlike the work in [31], we consider a VM assignment in which one VM can only process the packets to one user but one user's packets can be served by several VMs with different computing capacities. To model this assignment scheme, we introduce the binary variables $c_{s,k} \in \{0,1\}$, $\forall k \in \mathcal{K}$ and $\forall s \in \mathcal{S}$, where $c_{s,k} = 1$ states that the packets of the kth user are processed by a VM in the sth PS and $c_{s,k} = 0$, otherwise. In addition, let binary variable $d_s \in \{0,1\}$ and $\mathbf{d} = \{d_s, \forall s \in \mathcal{S}\}$ denote the operation mode of the sth PS, where $d_s = 0$ means the sth PS is turned off and $d_s = 1$ otherwise.

C. Processing Queue Model

The packet arrival of the kth UE is assumed to follow a Poisson process with arrival rate Λ_k. For simplicity, we assume each packet has identical length. As illustrated in Fig. 1, packets of the kth UE first arrive at the dispatcher and are subsequently split into smaller fragments that are then routed to VMs in different PSs for parallel processing. It is worth mentioning that each small fragment from the kth UE's packets assigned to the VM in the sth PS also follows a Poisson process with arrival rate $\lambda_{s,k}$, where we have

$$\begin{cases} \sum_{s\in\mathcal{S}} \lambda_{s,k} = \Lambda_k \\ \lambda_{s,k} \le c_{s,k}\Lambda_k \end{cases}, \forall k \in \mathcal{K}, s \in \mathcal{S} \tag{C2}$$

We assume that the baseband processing of each VM on each UE packets can be described as a $M/M/1$ processing queue, where the service time at the VM of the sth PS follows an exponential distribution with mean $1/\mu_{s,k}$, where $\mu_{s,k}$ represents the computing capacity that the VM of sth PS can process the kth UE's packets. Note that since each PS has a maximum computing capacity $C_s^{\text{PS}}, \forall s \in \mathcal{S}$, we have the following constraints

$$\begin{cases} \sum_{k\in\mathcal{K}} \mu_{s,k} \le d_s C_s^{\text{PS}} \\ \mu_{s,k} \le c_{s,k} C_s^{\text{PS}} \\ c_{s,k} \le d_s \end{cases}, \forall k \in \mathcal{K}, s \in \mathcal{S} \tag{C3}$$

Based on these, the average response time to process each packet for the kth UE at the VM of the sth PS is computed as $\frac{c_{s,k}}{\mu_{s,k}-\lambda_{s,k}}$, where $\lambda_{s,k} < \mu_{s,k}, \forall s \in \mathcal{S}, k \in \mathcal{K}$. Since the packets for the kth UE can be processed by multiple VMs of different computing capacities, the effective response time τ_k to process all packets of the kth UE in the BBU pool should be larger than the worst average response time among its serving VMs, leading to the following constraint

$$\tau_k \ge \frac{c_{s,k}}{\mu_{s,k} - \lambda_{s,k}}; \mu_{s,k} \ge \lambda_{s,k}, \forall k \in \mathcal{K}, s \in \mathcal{S} \tag{C4}$$

D. Transmission Queue Model

After being processed by the VMs, the outcome packets from the processing queue are aggregated at a virtual switching node. Then, they are transported via the corresponding fronthaul links to the RRHs and eventually transmitted to the UEs. For simplicity, we neglect the transportation delay. By Burke's Theorem [39], the arrival process of transmission queue for the kth UE, (i.e., the departure process of processing queue for the kth UE) is still Poisson with rate Λ_k. Therefore, the data transmission to the kth UE from its serving RRHs can be modeled as a $M/M/1$ transmission queue service time $1/R_k(\mathbf{w})$ [14] (cf. Fig. 1). Therefore, the average response time in the wireless transmission queue for the kth UE is simply given by

$$t_k(\mathbf{w}) = 1/(R_k(\mathbf{w}) - \Lambda_k), \forall k \in \mathcal{K} \tag{4}$$

where $R_k(\mathbf{w}) > \Lambda_k$ should be guaranteed for the queue stability. In this paper, we restrict the total response time of the processing and transmission queue by a delay value D_k to ensure a low-latency transmission for each UE, which is expressed as

$$\tau_k + 1/(R_k(\mathbf{w}) - \Lambda_k) \le D_k, \forall k \in \mathcal{K} \tag{C5}$$

It is noteworthy that virtual computing constraints are coupled with the physical constraints via QoS delay constraint in (C5), motivating the cross-layer joint design considered in this paper.

E. Power Consumption Model

1) RRH power consumption: According to [16], [18], [31], the power consumption at each RRH is categorized into two types: data-dependent power, which is related to the transmitted signal, and data-independent power. The data-independent power can be further sub-categorized into two types: power to keep each ith RRH active, denoted as P_i^{ra}, and power to keep each ith RRH idle, denoted as P_i^{ri}. To formulate the design problem, we introduce a binary variable $b_i = \{0,1\}, \forall i \in \mathcal{I}$ to represent the operation mode of each ith RRH, where $b_i = 0$ indicates that the ith RRH is in sleep mode and $b_i = 1$ otherwise. The total power consumption at the ith RRH is written as

$$P_i^{\text{RRH}}(\mathbf{w}, b_i) = \frac{1}{\eta_i} \sum\nolimits_{k\in\mathcal{K}} \left\|\mathbf{w}_{i,k}\right\|_2^2 + b_i P_i^{\text{ra}} + \left(1 - b_i\right) P_i^{\text{ri}} \tag{5}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2799859, IEEE Transactions on Wireless Communications

where $\eta_i \in [0,1]$ is the power amplifier efficiency.

2) Fronthaul power consumption: We adopt the model in [40] where the fronthaul link power consumption directly depends on transmission rates served by the corresponding RRH. Specifically, the power consumption of the ith fronthaul link for forwarding information data and beamformers is written as

$$P_i^{\text{FH}}(\mathbf{w}, \mathbf{a}_i) = \rho_i \sum\nolimits_{k\in\mathcal{K}} a_{i,k} R_k(\mathbf{w}) \tag{6}$$

where $\mathbf{a}_i = [a_{i,1} \ldots, a_{i,K}]^T$ and $\rho_i = P_{i,\max}^{\text{FH}}/C_i^{\text{FH}}$ is the constant scaling factor associated to the ith fronthaul with $P_{i,\max}^{\text{FH}}$ is the power dissipation of ith fronthaul.

3) BBU power consumption: Let us define P_s^{PS} and $\kappa_s \mu_{s,k}^{\alpha_s}, \forall s \in \mathcal{S}, k \in \mathcal{K}$ as the power spent by the sth PS and the associated VMs for processing the kth UE's traffic, respectively, where in the polynomial approximation $\kappa_s \mu_{s,k}^{\alpha_s}$(c.f., [11], [27]), $\kappa_s > 0$, $\alpha_s > 1$ are the positive multiplication and exponent factors.Thus, by denoting $\boldsymbol{\mu} = \{\mu_{s,k}, \forall s \in \mathcal{S}, \forall k \in \mathcal{K}\}$, the overall power consumption in the BBU pool is

$$P^{\text{BBU}}(\mathbf{d}, \boldsymbol{\mu}) = \sum\nolimits_{s\in\mathcal{S}} d_s P_s^{\text{PS}} + \sum\nolimits_{s\in\mathcal{S}} \sum\nolimits_{\forall k\in\mathcal{K}} \kappa_s \mu_{s,k}^{\alpha_s} \tag{7}$$

4) Total power consumption: Finally, the entire network power consumption in the considered system model is formulated as

$$P(\mathbf{w}, \boldsymbol{\mu}, \mathbf{a}, \mathbf{b}, \mathbf{d}) = \sum\nolimits_{i\in\mathcal{I}} \left(P_i^{\text{RRH}}(\mathbf{w}, b_i) + P_i^{\text{FH}}(\mathbf{w}, \mathbf{a}_i)\right) + \phi P^{\text{BBU}}(\mathbf{d}, \boldsymbol{\mu}) \tag{8}$$

where $\phi > 0$ is a parameter to strike a balance between the power consumption of RRHs, fronthaul and BBU pool, $\mathbf{b} = [b_1, \ldots, b_I]^T$ and $\mathbf{a} = \left[\mathbf{a}_1^T, \ldots, \mathbf{a}_I^T\right]^T$. The symbols used through out the paper are defined in Table I.

III. Problem Formulation

We aim at jointly optimizing the virtual computing resource allocation with beamforming, RRH selection and RRH-UE association to maximize the global network energy efficiency. To guarantee the stability of the transmission queue as shown in (4) and the minimum QoS UE rate requirement $R_k^{\min}$ for each UE k, we impose the following constraint

$$R_k(\mathbf{w}) \geq \max\left\{R_k^{\min}, \Lambda_k\right\} \tag{C6}$$

Moreover, the total transmit power at each RRH is limited by a power budget $P^{\max}$, which is expressed as

$$\sum\nolimits_{k\in\mathcal{K}} \left\|\mathbf{w}_{i,k}\right\|_2^2 \leq b_i P^{\max};\ \left\|\mathbf{w}_{i,k}\right\|_2^2 \leq a_{i,k} P^{\max};\ a_{i,k} \leq b_i \tag{C7}$$

where $\mathbf{a}$ and $\mathbf{b}$ are defined in (C1) and (5). The above constraint implies that when the ith RRH is in sleep mode, e.g., $b_i = 0$, no power will be transmitted from it. Similarly, we also guarantee that the transmit power $\left\|\mathbf{w}_{i,k}\right\|_2^2$ from the ith RRH to the kth UE is zero if $a_{i,k} = 0$. Also, whereas $b_i = 0$, then $a_{i,k} = 0$ for all $k \in \mathcal{K}$ and $\sum_{k\in\mathcal{K}} \left\|\mathbf{w}_{i,k}\right\|_2^2 = 0$. Now the considered problem is formulated as

$$(\mathcal{P}_0): \underset{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d},\boldsymbol{\lambda},\boldsymbol{\tau},\mathbf{w},\boldsymbol{\mu}}{\text{maximize}} \quad \frac{\sum_{k\in\mathcal{K}} R_k(\mathbf{w})}{P(\mathbf{w}, \boldsymbol{\mu}, \mathbf{a}, \mathbf{b}, \mathbf{d})} \tag{9a}$$

$$\text{subject to} \quad \text{(C1)}; \text{(C2)}; \text{(C3)}; \text{(C4)}; \text{(C5)}; \text{(C6)}; \text{(C7)} \tag{9b}$$

where $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ are implicitly understood to be binary. To solve (9), we first customize a branch-and-reduce-and-bound (BnRnB) based algorithm which is presented in the next section.

IV. Proposed Global Optimization Method

We present an algorithm to solve (9) optimally. Before proceeding further, we provide some comments on the complexity of (9). First, problem (9) is generally NP-hard due to the presence of binary variables $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{d}$. Moreover, even when these binary variables are relaxed to be continuous, the obtained problem is still non-convex because of the non-convexity of the objective function (9a) and the constraints in (C1) and (C5). In mathematical programming, (9) is categorized as a mixed-integer non-convex program for which such a method in [11], [31], [41] is not applicable to find a globally optimal solution. To the best of our knowledge, there is no off-the-shelf solver for (9). In what follows, we present an equivalent formulation of (9), based on which a BnRnB algorithm using monotonic optimization (MO) is customized to solve it optimally. We note that there are also other global optimization techniques such as inner and outer approximation, cutting-plane methods, etc. These optimal algorithms will yield the same optimal objective value. In this paper, we adopt the branch-and-reduce-and-bound (BnRnB) method to find a globally optimal solution for the considered problem since it lends itself to the considered problem, especially with a novel reformulation presented next.

A. Equivalent Formulation

Let us introduce the slack variables $\boldsymbol{\nu} = \{\nu_i \geq 0, \forall i \in \{0 \cup \mathcal{K}\}\}$ and $\boldsymbol{\zeta} = \{\zeta_k \geq 0, \forall k \in \mathcal{K}\}$ and rewrite (9) as the following problem

$$\underset{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d},\boldsymbol{\lambda},\boldsymbol{\tau},\mathbf{w},\boldsymbol{\mu},\boldsymbol{\nu},\boldsymbol{\zeta}}{\text{maximize}} \quad f(\boldsymbol{\nu}) = \nu_0 \sum\nolimits_{k\in\mathcal{K}} \nu_k \tag{10a}$$

$$\text{subject to} \quad R_k(\mathbf{w}) \geq \nu_k \tag{10b}$$

$$\nu_k \geq \max\left\{R_k^{\min}, \Lambda_k\right\} \tag{10c}$$

$$\hat{P}(\mathbf{w}, \boldsymbol{\mu}, \mathbf{a}, \mathbf{b}, \mathbf{d}, \boldsymbol{\nu}) \leq 1/\nu_0 \tag{10d}$$

$$\sum\nolimits_{k\in\mathcal{K}} a_{i,k}\nu_k \leq C_i^{\text{FH}} \tag{10e}$$

$$\tau_k \geq c_{s,k}^2/(\mu_{s,k} - \lambda_{s,k}) \tag{10f}$$

$$\zeta_k \geq 1/(\nu_k - \Lambda_k) \tag{10g}$$

$$\tau_k + \zeta_k \leq D_k \tag{10h}$$

$$\text{(C2)}, \text{(C3)}, \text{(C7)} \tag{10i}$$

where $\hat{P}(\mathbf{w}, \boldsymbol{\mu}, \mathbf{a}, \mathbf{b}, \mathbf{d}, \boldsymbol{\nu}) = \sum_{i\in\mathcal{I}} \rho_i \sum_{k\in\mathcal{K}} a_{i,k}\nu_k + \sum_{i\in\mathcal{I}} P_i^{\text{RRH}}(\mathbf{w}, b_i) + \phi P^{\text{BBU}}(\mathbf{d}, \boldsymbol{\mu})$. To arrive at (10), several slack variables have been introduced and the idea behind this step is justified as follows. First, we have replaced $c_{s,k}$ in (C4) by $c_{s,k}^2$, resulting in (10f). However, this maneuver still maintains the equivalence between (C4) and (10f) since $c_{s,k} = c_{s,k}^2$ for $c_{s,k} \in \{0,1\}$. The benefit of considering (10f) is that it is a convex constraint and particularly can be recast as a second order cone constraint, while (C4) is a non-convex one. This property will be exploited to develop a global optimization algorithm to (10). Second and more importantly, $R_k(\mathbf{w})$ has been replaced by ν_k or Λ_k at various places in (9). We remark that this move

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2799859, IEEE Transactions on Wireless Communications

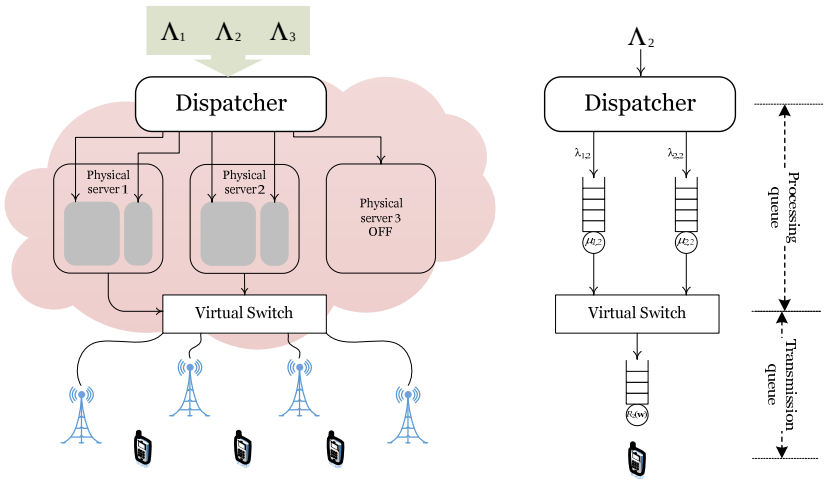


Fig. 1. (a) Limited fronthaul C-RANs with VCRA scheme, (b) Queuing model, i.e., for UE 2

TABLE I
DEFINITION OF MATHEMATICAL SYMBOLS

Symbol	Definition	Symbol	Definition
$a_{i,k}$	Association status of UE k and RRH i	b_i	Operation mode of RRH i
$c_{s,k}$	VM assignment status in PS s to UE k	d_s	Operation mode of PS s
$\lambda_{s,k}$	Arrival rate allocated to VM in PS s for UE k	Λ_k, s_k	Total traffic arrival rate, signal for UE k
$\mu_{s,k}$	VM computing capacity of PS s assigned for UE k	C_s^{PS}	Maximum computing capacity of PS s
$R_k(\mathbf{w})$	Achievable rate of UE k w.r.t variable $\mathbf{w}$	C_i^{FH}	Maximum capacity of fronthaul link i
τ_k	Average delay to process the queue for UE k	D_k	Maximum delay for UE k
ρ_i	Scaling factor for fronthaul link i	η_i	Power amplifier efficiency of RRH i
κ_s	Multiplication factor of the approximation $\kappa_s \mu_{s,k}^{\alpha_s}$	α_s	Exponent factor of the approximation $\kappa_s \mu_{s,k}^{\alpha_s}$
$R_k^{\min}$	Minimum rate requirement for UE k	$P^{\max}$	Maximum transmit power at each RRH
P_i^{ra}, P_i^{ri}	Power to keep RRH i active, idle	$P_{i,\max}^{\text{FH}}$	Power dissipation of fronthaul link i
P_s^{PS}	Power to keep PS s active	ϕ	Parameter to balance two parts of power in (8)
ν_i, ζ_k	Slack variables to support BnRnB algorithm	ξ_k,γ_k	Constants used in the DCA-based method
z_k, σ_0^2	AWGN, noise power	I, M_i	Number of RRHs, antenna of RRH i
$\mathbf{w}_{i,k}$	Beamforming vector from RRH i to UE k	$\Gamma_k(\mathbf{w})$	Received SINR for UE k w.r.t variable $\mathbf{w}$
$\mathbf{h}_{i,k}$	Channel vector from RRH i to UE k	S, K	Number of PSs, UEs
$E_{i,k}$	Incentive measure to decide binary variables	$I_{s,k}$	Incentive measure to decide binary variables

does not follow the standard reformulation technique based on epigraph form and thus the equivalence between (9) and (10) is not guaranteed in general. In this regard one of our main contributions is stated in the following lemma.

Lemma 1. *The formulations in* (9) *and* (10) *are equivalent in the sense that they have the same optimal solution set and objective.*

Proof: The proof is presented in Appendix A. □

B. Optimal Solution based BnRnB Algorithm

The key benefits of the reformulation given in (10) are two-fold: first, it facilitates the customization of the BnRnB algorithm based on the MO framework to solve (10); and secondly in this regard, the search space for an optimal solution is reduced from $(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\zeta})$ in (9) to only $\boldsymbol{\nu}$ in (10), which then results in less computational complexity. Application of MO to solve (10) is possible due to the following two important observations.

- The objective in (10a) monotonically increases with respect to each entry of $\boldsymbol{\nu}$, which is obvious from the expression of $f(\boldsymbol{\nu})$ in (10a).

- For a given $\boldsymbol{\nu}$, the following feasibility problem

$$\text{find } \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\zeta} \quad (11a)$$
$$\text{s.t. } (10b) - (10i). \quad (11b)$$

is a mixed-integer second order cone (MISOC) feasibility problem, which can be solved optimally by dedicated MISOCP solvers such as MOSEK.[2] Note that for a given ν_k, (10b) can be reformulated as a SOC constraint as

$$c' \operatorname{Re}(\mathbf{h}_k^H \mathbf{w}_k) \geq \left\| \mathbf{h}_k^H \mathbf{w}_1, \ldots, \mathbf{h}_k^H \mathbf{w}_K, \sigma_0 \right\|_2 \quad (12)$$

where $c' = \sqrt{\frac{1}{2^{\nu_k}-1} + 1}$.

From the above two facts, we can develop a BnRnB method to solve (10) optimally as done in [34]. The detailed content and description of the BnRnB algorithm is similar to the presentation in [34, Algorithm 1], which is skipped here due to space constraint. Herein, we briefly present some important steps and definitions required to solve the considered

[2]https://www.mosek.com/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2799859, IEEE Transactions on Wireless Communications

problem (10). Specifically, we first define the compact normal set $\mathcal{Q} = \{\boldsymbol{\nu} \in \mathbb{R}_+^{K+1} | (10b) - (10i)\}$, i.e., $\forall \boldsymbol{\nu} \in \mathcal{Q}$ such that problem (11) is feasible. We also define $\mathcal{V} = [\underline{\boldsymbol{\nu}}; \overline{\boldsymbol{\nu}}]$ to be the box that contains all $\boldsymbol{\nu}$ feasible to (10). Note that $Q \subset \mathcal{V}$. The calculation of $\underline{\boldsymbol{\nu}}$ and $\overline{\boldsymbol{\nu}}$ is presented in Appendix B. Problem (10) can now be abstractly expressed as $\max\{f(\boldsymbol{\nu}) | \boldsymbol{\nu} \in \mathcal{Q} \subset \mathcal{V}\}$.

The main idea to solve problem (10) optimally using MO framework is to check if a given $\boldsymbol{\nu}$ belongs to $\mathcal{Q}$ or not, which amounts to solving the MISOC feasibility problem in (11). At the beginning of the proposed algorithm, we check whether $\underline{\boldsymbol{\nu}}$ (i.e., the lower corner of $\mathcal{V}$) is feasible or not. If so, we apply the BnRnB method to find a globally optimal solution to (10). The proposed method recursively branches a box $\mathcal{B}$, which has the largest upper bound compared to others, into two smaller boxes, checks the feasibility of each new box, updates the current upper and lower bounds by the box reduction and bound computation process, and removes the boxes that do not contain an optimal solution. The details of these operations can be found in [34], and thus omitted here for the sake of brevity. As mentioned earlier, these steps are performed over $\boldsymbol{\nu}$, not over all optimization variables. This dimension reduction significantly reduce the overall complexity. Moreover, due to the monotonicity of the objective, the upper and lower bound of a box $\mathcal{B} = [\underline{\boldsymbol{\nu}}, \bar{\boldsymbol{\nu}}]$ can be quickly found as $U(\mathcal{B}) = f(\bar{\boldsymbol{\nu}})$ and $L(\mathcal{B}) = f(\underline{\boldsymbol{\nu}})$, respectively.According to [34], the proposed algorithm is bound improving and terminates after finitely many iterations for a given desired accuracy level ϵ.

To conclude this section, we remark that the proposed optimal Algorithm BnRnB presented in this section requires extremely high computational complexity for two apparent reasons. First, the MISOC feasibility problem in (11) is NP-hard in general and thus the complexity can increase exponentially with the problem size in the worst case. Second, the BnRnB algorithm (even when all binary variables are relaxed to be continuous) generally requires a large number of iterations to terminate. As a result, Algorithm BnRnB is practically useful for networks of relatively small size and is mainly used for benchmarking purpose in this paper. For a more practically appealing solution we propose a low-complexity method based on the framework of DC programming in the next section.

V. Low-complexity Method

Given the inherent non-convexity and combinatorial nature of $(\mathcal{P}_0)$, a pragmatic goal is to find a sufficiently good feasible solution in a reasonable amount of time. To this end we will present a low-complexity algorithm in this section based on the following steps:

- Binary variables are relaxed to be continuous to obtain the continuous relaxation problem of $(\mathcal{P}_0)$, denoted as $(\mathcal{P}_1)$. This step is routine to handle the discreteness of the considered problem.

- As mentioned above, $(\mathcal{P}_1)$ is still non-convex and solving it is difficult. Although finding a globally optimal solution to $(\mathcal{P}_1)$ is possible by slightly modifying Algorithm BnRnB, the run time will be prohibitively high which is not suitable for real-time applications. The main idea of the proposed low-complexity method is to solve $(\mathcal{P}_1)$ using a local optimization method to compute a high-quality estimate of $(\mathcal{P}_0)$. Thus, we resort to the DC programming framework.

- The last step is to devise a post-processing procedure to map the solution produced by solving $(\mathcal{P}_1)$ which is not binary in general into a binary one.

In the next subsections we will present the details of the steps listed above.

A. DC Decomposition

The main target of the proposed low-complexity algorithm is to solve $(\mathcal{P}_1)$ efficiently. We recall that the non-convexity of $(\mathcal{P}_1)$ is due to that of function $R_k(\mathbf{w})$ and also the term $a_{i,k} R_k(\mathbf{w})$. Based on the concept of DC programming, we will express each of the non-convex functions as a difference of two convex ones. To illustrate this point let us first consider the rate function $R_k(\mathbf{w})$. Note that the following decomposition holds

$$R_k(\mathbf{w}) = \underbrace{R_k(\mathbf{w}) + \xi_k \|\mathbf{w}\|_2^2}_{f_k(\mathbf{w})} - \xi_k \|\mathbf{w}\|_2^2 \tag{13}$$

for any ξ_k. Intuitively if ξ_k is sufficiently large, the quadratic term $\xi_k \|\mathbf{w}\|_2^2$ will dominate $R_k(\mathbf{w})$ and thus $f_k(\mathbf{w})$ becomes convex eventually. We remark that this kind of DC decomposition is not entirely new. But the problem is that finding a proper value for ξ_k to make (13) a DC expression is very challenging and problem-specific. In this regard our contribution is the following lemma.

Lemma 2. *For $\xi_k > \bar{\xi}_k$, where $\bar{\xi}_k$ is given in* (42) *in Appendix X, $f_k(\mathbf{w})$ is strongly convex .*

Proof: The proof and the derivation of $\bar{\xi}_k$ in Lemma 2 are involved and all the detailed algebra is presented in Appendix X. The idea is to show that $R_k(\mathbf{w})$ is $\bar{\xi}_k$-smooth, i.e.,

$$\|\nabla R_k(\mathbf{x}) - \nabla R_k(\mathbf{y})\|_2 \leq \bar{\xi}_k \|\mathbf{x} - \mathbf{y}\|_2 \tag{14}$$

where $\nabla f(\mathbf{x})$ is the gradient of $f(\mathbf{x})$ with respect to $\mathbf{x}$. Equivalently, $\nabla R_k(\mathbf{x})$ is Lipschitz continuous with a constant $\bar{\xi}_k$. □

We now turn the attention to a DC decomposition of the term $a_{i,k} R_k(\mathbf{w})$. By the same way, we consider the following DC decomposition

$$\begin{aligned} a_{i,k} R_k(\mathbf{w}) =& \gamma_k \left(\|\mathbf{w}\|_2^2 + a_{i,k}^2 \right) \\ & - \underbrace{\left(\gamma_k \left(\|\mathbf{w}\|_2^2 + a_{i,k}^2 \right) - a_{i,k} R_k(\mathbf{w}) \right)}_{u_k(\mathbf{w}, a_{i,k})} \end{aligned} \tag{15}$$

and the following lemma is in order.

Lemma 3. *For $\gamma_k > \bar{\gamma}_k$ where $\bar{\gamma}_k$ is given in* (58) *in Appendix XI, $u_k(\mathbf{w}, a_{i,k})$ is strongly convex.*

Proof: The proof of Lemma 3 follows the same steps as those for that of Lemma 2 and is provided in Appendix XI. □

Based on the above DC decomposition, we are now in a position to describe the proposed algorithm to solve $(\mathcal{P}_1)$ efficiently. First $(\mathcal{P}_1)$ can be equivalently rewritten as

$$(\mathcal{P}_2): \max_{\substack{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d} \\ \boldsymbol{\lambda},\boldsymbol{\tau},\mathbf{w},\boldsymbol{\mu}}} \frac{\sum_{k \in \mathcal{K}} f_k(\mathbf{w}) - \xi_k \|\mathbf{w}\|_2^2}{\tilde{P}(\mathbf{w}, \boldsymbol{\mu}, \mathbf{a}, \mathbf{b}, \mathbf{d})} \tag{16a}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2799859, IEEE Transactions on Wireless Communications

$$\text{s.t. } f_k(\mathbf{w}) - \xi_k \|\mathbf{w}\|_2^2 \geq \max\left\{R_k^{\min}, \Lambda_k\right\} \quad (16b)$$

$$\sum_{k\in\mathcal{K}}\left(\gamma_k\left(\|\mathbf{w}\|_2^2 + a_{i,k}^2\right) - u_k(\mathbf{w}, a_{i,k})\right) \leq C_i^{\text{FH}} \quad (16c)$$

$$\frac{(\tau_k + \mu_{s,k} - \lambda_{s,k})^2}{4} \geq c_{s,k} + \frac{(\tau_k - \mu_{s,k} + \lambda_{s,k})^2}{4} \quad (16d)$$

$$f_k(\mathbf{w}) - \xi_k \|\mathbf{w}\|_2^2 - \Lambda_k \geq \frac{1}{D_k - \tau_k} \quad (16e)$$

$$a_{i,k}, b_i, c_{s,k}, d_s \in [0,1] \quad (16f)$$

$$\mu_{s,k} \geq \lambda_{s,k} \quad (16g)$$

$$\text{(C2); (C3); (C7)} \quad (16h)$$

where $\tilde{P}(\mathbf{w}, \boldsymbol{\mu}, \mathbf{a}, \mathbf{b}, \mathbf{d}) = \sum_{i\in\mathcal{I}} P_i^{\text{RRH}}(\mathbf{w}, b_i) + \phi P^{\text{BBU}}(\mathbf{d}, \boldsymbol{\mu}) + \sum_{i\in\mathcal{I}} \rho_i \sum_{k\in\mathcal{K}} (\gamma_k\left(\|\mathbf{w}\|_2^2 + a_{i,k}^2\right) - u_k(\mathbf{w}, a_{i,k}))$. Note that we have equivalently rewritten constraints (C4) and (C5) as (16d) and (16e), respectively. The purpose of these reformulations is to express $(\mathcal{P}_2)$ as a DC program that is amenable to application of DCA, which is presented next subsection.

Before proceeding further, we remark that to deal with the non-convexity of such a problem as $(\mathcal{P}_1)$, a class of existing methods introduce some slack variables to expose the hidden convexity of non-convex objective and/or constraints, and then apply successive convex approximation to solve the resulting problem [5], [37]. The drawback of such a method is that the eventual number of optimization variables increases (quickly in many cases) with the problem size. In this regard, the DC form in $(\mathcal{P}_2)$ does not introduce any new auxiliary variable, which certainly achieves more favorable scalability property and thus makes it more suitable for C-RANs.

B. DCA-based Method

In this section we apply DCA to solve $(\mathcal{P}_2)$. The main idea of DCA can be briefly explained as follows. Let us consider the following general DC constraint

$$p(\mathbf{x}) - q(\mathbf{x}) \leq 0 \quad (17)$$

where $p(\mathbf{x})$ and $q(\mathbf{x})$ are convex with respect to $\mathbf{x}$. It is obvious that the non-convex part in the above constraint is $-q(\mathbf{x})$ which is concave. Assuming $q(\mathbf{x})$ is differentiable (which is true for all constraints in $(\mathcal{P}_2)$), DCA linearizes $q(\mathbf{x})$ around the current iteration $\mathbf{x}^{(n)}$ to arrive at the following constraint

$$p(\mathbf{x}) - q(\mathbf{x}^{(n)}) - \left\langle \nabla q(\mathbf{x}^{(n)}), \mathbf{x} - \mathbf{x}^{(n)} \right\rangle \leq 0 \quad (18)$$

Note that (18) implies (17) as a concave function (i.e., $-q(\mathbf{x})$ as mentioned above) is upper bounded by its linearization. In other words, DCA arrives at an inner approximation of the feasible set of the considered nonconvex program and updates the point $\mathbf{x}^{(n)}$ until convergence.

Let us deal with the DC constraints in $(\mathcal{P}_2)$ first. According to the philosophy of DCA, we can approximate (16b) as

$$F_k(\mathbf{w}; \mathbf{w}^{(n)}) - \xi_k \|\mathbf{w}\|_2^2 \geq \max\left\{R_k^{\min}, \Lambda_k\right\} \quad (19)$$

where $F_k(\mathbf{w}; \mathbf{w}^{(n)})$ is given in (20) shown at the top of the next page. Note that $F_k(\mathbf{w}; \mathbf{w}^{(n)})$ is simply a linearization of $f_k(\mathbf{w})$ around $\mathbf{w}^{(n)}$ and its derivation is in fact a by-product of Appendix X. Thus, constraint (16e) can be approximated by the following constraint

$$F_k(\mathbf{w}; \mathbf{w}^{(n)}) - \xi_k \|\mathbf{w}\|_2^2 - \Lambda_k \geq \frac{1}{D_k - \tau_k} \quad (21)$$

which is a convex constraint since $1/(D_k - \tau_k)$ is convex and $F_k(\mathbf{w}; \mathbf{w}^{(n)}) - \xi_k \|\mathbf{w}\|_2^2 - \Lambda_k$ is concave with respect to all feasible variables $\mathbf{w}, \tau_k$. In the same way, we can also derive the upper bound convex approximation of the non–convex DC function $\gamma_k\left(\|\mathbf{w}\|_2^2 + a_{i,k}^2\right) - u_k(\mathbf{w}, a_{i,k})$ by deriving the lower bound concave approximation of $u_k(\mathbf{w}, a_{i,k})$ as follow

$$u_k(\mathbf{w}, a_{i,k}) \geq \tilde{F}_k(\mathbf{w}, a_{i,k}; \mathbf{w}^{(n)}, a_{i,k}^{(n)}) \quad (22)$$

where $\tilde{F}_k(\mathbf{w}, a_{i,k}; \mathbf{w}^{(n)}, a_{i,k}^{(n)})$ is given as in (23).Thus, constraint in (16c) can be approximated by its concave upper bound as

$$\sum_{k\in\mathcal{K}}\left(\gamma_k\left(\|\mathbf{w}\|_2^2 + a_{i,k}^2\right) - \tilde{F}_k(\mathbf{w}, a_{i,k}; \mathbf{w}^{(n)}, a_{i,k}^{(n)})\right) \leq C_i^{\text{FH}} \quad (24)$$

By applying the above approximations, we can formulate the approximation of problem $(\mathcal{P}_2)$ at iteration $n+1$ as

$$\max_{\substack{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}\\ \boldsymbol{\lambda},\boldsymbol{\tau},\mathbf{w},\boldsymbol{\mu}}} \frac{\sum_{k\in\mathcal{K}}\left(F_k(\mathbf{w}; \mathbf{w}^{(n)}) - \xi_k \|\mathbf{w}\|_2^2\right)}{\hat{P}\left(\mathbf{w}, \boldsymbol{\mu}, \mathbf{a}, \mathbf{b}, \mathbf{d}; \mathbf{w}^{(n)}, \mathbf{a}^{(n)}\right)} \quad (25a)$$

$$\text{s.t. } F_k(\mathbf{w}; \mathbf{w}^{(n)}) - \xi_k \|\mathbf{w}\|_2^2 \geq \max\left\{R_k^{\min}, \Lambda_k\right\} \quad (25b)$$

$$\sum_{k\in\mathcal{K}} \gamma_k\left(\|\mathbf{w}\|_2^2 + a_{i,k}^2\right) - \sum_{k\in\mathcal{K}} \tilde{F}_k(\mathbf{w}, a_{i,k}; \mathbf{w}^{(n)}, a_{i,k}^{(n)}) \leq C_i^{\text{FH}} \quad (25c)$$

$$c_{s,k} + \frac{(\tau_k - \mu_{s,k} + \lambda_{s,k})^2}{4} \leq \frac{\tau_k^{(n)} + \mu_{s,k}^{(n)} - \lambda_{s,k}^{(n)}}{2} \times (\tau_k + \mu_{s,k} - \lambda_{s,k}) - \frac{(\tau_k^{(n)} + \mu_{s,k}^{(n)} - \lambda_{s,k}^{(n)})^2}{4} \quad (25d)$$

$$F_k(\mathbf{w}; \mathbf{w}^{(n)}) - \xi_k \|\mathbf{w}\|_2^2 - \Lambda_k \geq \frac{1}{D_k - \tau_k} \quad (25e)$$

$$a_{i,k}, b_i, c_{s,k}, d_s \in [0,1] \quad (25f)$$

$$\text{(C2); (C3); (C7); (16g)} \quad (25g)$$

where $\hat{P}\left(\mathbf{w}, \boldsymbol{\mu}, \mathbf{a}, \mathbf{b}, \mathbf{d}; \mathbf{w}^{(n)}, a_{i,k}^{(n)}\right) = \sum_{i\in\mathcal{I}} P_i^{\text{RRH}}(\mathbf{w}, b_i) + \phi P^{\text{BBU}}(\mathbf{d}, \boldsymbol{\mu}) + \sum_{i\in\mathcal{I}} \rho_i \sum_{k\in\mathcal{K}} \gamma_k \times \left(\|\mathbf{w}\|_2^2 + a_{i,k}^2\right) - \sum_{i\in\mathcal{I}} \rho_i \sum_{k\in\mathcal{K}} \tilde{F}_k(\mathbf{w}, a_{i,k}; \mathbf{w}^{(n)}, a_{i,k}^{(n)})$. Note that the fractional objective (25a) can be easily transformed into a linear subtractive form using Dinkelback approach [24]. This subsequently makes (25a) convex and can be solved by the DCA-based algorithm, which is outlined in Algorithm 1.

Algorithm 1 DCA-based Algorithm.

1: Set $n := 0$ and initialize starting points of $\mathbf{w}^{(n)}, \boldsymbol{a}^{(n)}, \boldsymbol{\tau}^{(n)}, \boldsymbol{\mu}^{(n)}, \boldsymbol{\lambda}^{(n)}$;
2: **repeat**
3: Solve the approximated problem (25) at $\mathbf{w}^{(n)}, \boldsymbol{a}^{(n)}, \boldsymbol{\tau}^{(n)}, \boldsymbol{\mu}^{(n)}, \boldsymbol{\lambda}^{(n)}$ to achieve the optimal solution $\mathbf{a}^\star, \mathbf{b}^\star, \mathbf{c}^\star, \mathbf{d}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\tau}^\star, \mathbf{w}^\star, \boldsymbol{\mu}^\star$;
4: Set $n := n + 1$;
5: Update $\mathbf{w}^{(n)} = \mathbf{w}^\star, \mathbf{a}^{(n)} = \mathbf{a}^\star, \boldsymbol{\tau}^{(n)} = \boldsymbol{\tau}^\star, \boldsymbol{\mu}^{(n)} = \boldsymbol{\mu}^\star, \boldsymbol{\lambda}^{(n)} = \boldsymbol{\lambda}^\star$
6: **until** Convergence of the objective (25a);

Convergence analysis: We now prove that Algorithm 1 is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2799859, IEEE Transactions on Wireless Communications

$$F_k(\mathbf{w};\mathbf{w}^{(n)}) = f_k(\mathbf{w}^{(n)}) + \frac{\sum_{j\in\mathcal{K}}\left(2\,\mathrm{Re}\big(\mathbf{w}_j^{(n)H}\mathbf{H}_k\mathbf{w}_j\big) - 2\mathbf{w}_j^{(n)H}\mathbf{H}_k\mathbf{w}_j^{(n)}\right)}{\sum_{j\in\mathcal{K}}|\mathbf{h}_k^H\mathbf{w}_j^{(n)}|^2+\sigma_0^2} - \frac{\sum_{j\in\mathcal{K}\setminus k}\left(2\,\mathrm{Re}\big(\mathbf{w}_j^{(n)H}\mathbf{H}_k\mathbf{w}_j\big) - 2\mathbf{w}_j^{(n)H}\mathbf{H}_k\mathbf{w}_j^{(n)}\right)}{\sum_{j\in\mathcal{K}\setminus k}|\mathbf{h}_k^H\mathbf{w}_j^{(n)}|^2+\sigma_0^2} + 2\xi_k\,\mathrm{Re}\big(\mathbf{w}^{(n)H}\mathbf{w}\big) - 2\xi_k\left\|\mathbf{w}^{(n)}\right\|_2^2 \quad (20)$$

$$\tilde{F}_k(\mathbf{w},a_{i,k};\mathbf{w}^{(n)},a_{i,k}^{(n)}) = \tilde{f}_k(\mathbf{w}^{(n)},a_{i,k}^{(n)}) + \left[\log(\sum_{j\in\mathcal{K}}|\mathbf{h}_k^H\mathbf{w}_j^{(n)}|^2+\sigma_0^2) - \log(\sum_{j\in\mathcal{K}\setminus k}|\mathbf{h}_k^H\mathbf{w}_j^{(n)}|^2+\sigma_0^2)\right](a_{i,k}-a_{i,k}^{(n)})$$
$$+ a_{i,k}^{(n)}\left[\frac{\sum_{j\in\mathcal{K}}\left(2\,\mathrm{Re}\big(\mathbf{w}_j^{(n)H}\mathbf{H}_k\mathbf{w}_j\big) - 2\mathbf{w}_j^{(n)H}\mathbf{H}_k\mathbf{w}_j^{(n)}\right)}{\sum_{j\in\mathcal{K}}|\mathbf{h}_k^H\mathbf{w}_j^{(n)}|^2+\sigma_0^2} - \frac{\sum_{j\in\mathcal{K}\setminus k}\left(2\,\mathrm{Re}\big(\mathbf{w}_j^{(n)H}\mathbf{H}_k\mathbf{w}_j\big) - 2\mathbf{w}_j^{(n)H}\mathbf{H}_k\mathbf{w}_j^{(n)}\right)}{\sum_{j\in\mathcal{K}\setminus k}|\mathbf{h}_k^H\mathbf{w}_j^{(n)}|^2+\sigma_0^2}\right]$$
$$+ 2\gamma_k\,\mathrm{Re}\big(\mathbf{w}^{(n)H}\mathbf{w} + a_{i,k}^{(n)}a_{i,k}\big) - 2\gamma_k\left(\left\|\mathbf{w}^{(n)}\right\|_2^2 + a_{i,k}^{(n)2}\right) \quad (23)$$

guaranteed to converge. This can be established by showing that the sequence of objectives returned by Algorithm 1 is monotonically convergent. Towards this end, let $\theta^{(n)}$ and $\mathbf{\Theta}^{(n)}$ denote the optimal objective value and the achieved optimal solution at the nth iteration of Algorithm 1, respectively. We will show that $\mathbf{\Theta}^{(n)}$ is also feasible to problem (25) at the $(n+1)$th iteration. To see this let us focus on the general DC constraint in (17). Due to the concavity of the term $-q(\mathbf{x})$, the following inequality holds

$$p(\mathbf{x}) - q(\mathbf{x}) \le p(\mathbf{x}) - q(\mathbf{x}^{(n)}) - \left\langle \nabla q(\mathbf{x}^{(n)}), \mathbf{x}-\mathbf{x}^{(n)}\right\rangle \quad (26)$$

for all $\mathbf{x}$. Note that the right hand side of the above inequality stands for the resulting approximate constraint in the problem considered at the $(n+1)$th iteration of Algorithm 1. The inequality in (26) means that if $\mathbf{x}$ satisfies the approximate constraint in (18), then it also satisfies the DC constraint in (17). Thus, Algorithm 1 produces a sequence of iterates $\{\mathbf{x}^{(k)}\}$ that are feasible to the original problem, i.e., $p(\mathbf{x}^{(k)})-q(\mathbf{x}^{(k)}) \le 0$. Substituting $\mathbf{x}$ by $\mathbf{x}^{(n)}$ in (18) we have

$$p(\mathbf{x}^{(n)}) - q(\mathbf{x}^{(n)}) - \left\langle \nabla q(\mathbf{x}^{(n)}), \mathbf{x}^{(n)}-\mathbf{x}^{(n)}\right\rangle = p(\mathbf{x}^{(n)}) - q(\mathbf{x}^{(n)}) \le 0. \quad (27)$$

The above inequality holds because $\mathbf{x}^{(n)}$ is feasible to the original problem. Thus, the solution of the nth iteration is feasible to the problem at iteration $(n+1)$. This leads to $\theta^{(n+1)} \ge \theta^{(n)}$, meaning that Algorithm 1 generates a non-decreasing sequence of objective values. Due to the power budget constraint (C7), the sequence of objectives $\{\theta^{\{n\}}\}$ is upper bounded and thus, is convergent.

C. An Accelerated Version of Algorithm 1: Practical Choices of ξ_k and γ_k

As being shown in the previous section, the sufficient conditions of ξ_k and γ_k to ensure the DC forms of functions (13) and (15) (and thus the convergence of Algorithm 1) are that $\xi_k \ge \bar{\xi}_k$ and $\gamma_k \ge \bar{\gamma}_k$, where $\bar{\xi}_k$ and $\bar{\gamma}_k$ are analytically computed in Appendix X and XI. However, smaller values of ξ_k and γ_k may significantly increase the convergence rate of Algorithm 1 in practice since they can lead to tighter approximation in each iteration of Algorithm 1. This will be easily seen from (13) where $f_k(\mathbf{w})$ is close to $R_k(\mathbf{w})$ for small ξ_k. Based on this observation, we set ξ_k and γ_k to a small value (elaborated in the numerical results section) in each iteration of Algorithm 1. If monotonic increase of the objective is not achieved, we then set ξ_k and γ_k to $\bar{\xi}_k$ and $\bar{\gamma}_k$, respectively. This variant is numerically shown to remarkably improve the convergence of Algorithm 1 and thus referred to as an accelerated version of Algorithm 1.

D. Post-Processing Procedure

A post-processing step is proposed to map the relaxed variables from solving $(\mathcal{P}_2)$ to the binary values, which is required due to the continuous relaxation. The process starts by assuming that all the RRHs and PSs are OFF and there is no association between RRHs, VMs and UEs. In each iteration, $(\mathcal{P}_2)$ is optimally solved given a set of active RRHs, RRH-UE association, active PSs and VM assignment that is connected. Let us denote optimal solution of $(\mathcal{P}_2)$ in the mth iteration of post-processing algorithm by $\{\mathbf{a}^\star,\mathbf{b}^\star,\mathbf{c}^\star,\mathbf{d}^\star,\boldsymbol{\lambda}^\star,\boldsymbol{\tau}^\star,\mathbf{w}^\star,\boldsymbol{\mu}^\star\}$. The RRH-UE association and VM assignment are then gradually updated by fixing untreated relaxed variables to be 1. Apparently, which unfixed relaxed variables are prior to be picked up is a critical decision. Intuitively, the connection between the ith RRH and the kth UE is more likely if the channel link condition is good and the power consumed to transmit fronthaul data is smaller than the others. Similarly, the kth UE is preferred to be processed by VM in the sth PS if the power expended for switching on the sth PS is the smaller than the others and the total signal processing power consumed in the sth PS is larger. Based on the above intuitive observations, we define a virtual energy efficiency for assigning the ith RRH to serve the kth UE (i.e., incentive measure to set $a_{i,k}=1$) as $E_{i,k}$ and a normalized importance index for assigning the VM in the sth PS to process for the kth UE's traffic (i.e., incentive measure to set $c_{s,k}=1$) as $I_{s,k}$, which are given by

$$E_{i,k} = \frac{R_k(\mathbf{w}^\star)}{\frac{1}{\eta_i}\left\|w_{i,k}^\star\right\|^2 + \rho_i R_k(\mathbf{w}^\star)};\; I_{s,k} = \frac{\sum_{k\in\mathcal{K}}\kappa_s\left(\mu_{s,k}^\star\right)^{\alpha_s}}{P_s^{\mathrm{PS}} + \kappa_s\left(\mu_{s,k}^\star\right)^{\alpha_s}} \quad (28)$$

Hence, we propose to fix the untreated relaxed variables to be 1 in the mth iteration of post-processing algorithm, i.e., $a_{i',k'} =$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2799859, IEEE Transactions on Wireless Communications

1, $c_{s',k'}=1$ whose indices $(i',k'),(s',k')$ are selected by

$$(i',k')=\underset{(i,k)\in\mathcal{R}_{\text{off}}^{(m-1)}}{\arg\max}\ E_{i,k}\quad\vee\quad(s',k')=\underset{(s,k)\in\mathcal{R}_{\text{off}}^{(m-1)}}{\arg\max}\ \mathbf{I}_{s,k} \tag{29}$$

where $\mathcal{R}_{\text{off}}^{(m-1)}=\{(i,k),(s,k)|\forall(i,k)\in(\mathcal{I},\mathcal{K}),\forall(s,k)\in(\mathcal{S},\mathcal{K}),a_{i,k}^{(m-1)}=0,c_{s,k}^{(m-1)}=0\}$ denotes the set of unfixed RRH-UE association and VM assignment in the $(m-1)$th iteration of the post-processing algorithm. This selection rule means that the unfixed relaxed binary variable that contributes mostly to the entire energy efficiency is set to be 1. According to constraints in (C3) and (C7), the variables b_i and d_s, $\forall i,s$ are fixed with respect to its associated variables $\mathbf{a}_i=\{a_{i,k},\forall k\in\mathcal{K}\}$ and $\mathbf{c}_s=\{c_{s,k},\forall k\in\mathcal{K}\}$, respectively. In particular, we need to set $b_i=1$ or $d_s=1$ if we fix any $a_{i,k}=1$ or $c_{s,k}=1$. Moreover, if variables $\mathbf{a}_i$ and $\mathbf{c}_s$ are fixed to $\mathbf{0}$, then we need to set $b_i=0$ or $d_s=0$. The RRH-UE association and VM assignment with the largest incentive measures (29) will be made connected and the resulting RRH and PS will be set active, following the relationship in $c_{s,k}\le d_s$ in (C3) and $a_{i,k}\le b_i$ in (C7). The overall algorithm is presented in Algorithm 2.

Algorithm 2 Post-processing algorithm.

1: Set $m:=0$, $\pi^{(m)}$ is significantly small, and initialize $a_{i,k}^{(m)}=0,c_{s,k}^{(m)}=0,b_i^{(m)}=0,d_s^{(m)}=0$, $\forall i\in\mathcal{I},\forall k\in\mathcal{K},\forall s\in\mathcal{S}$ and the set $\mathcal{R}_{\text{off}}^{(m)}=\left\{(i,k),(s,k)|\forall(i,k)\in(\mathcal{I},\mathcal{K}),\forall(s,k)\in(\mathcal{S},\mathcal{K}),a_{i,k}^{(m)}=0,\ c_{s,k}^{(m)}=0\right\}$.
2: **repeat**
3: Set $m:=m+1$;
4: Solve $(\mathcal{P}_2)$ given $a_{i',k'}=b_{i'}=1$, $c_{s',k'}=d_{s'}=1,\forall(i',k'),(s',k')\notin\mathcal{R}_{\text{off}}^{(m-1)}$ until convergence ;
5: Update $\mathcal{R}_{\text{off}}^{(m)}=\mathcal{R}_{\text{off}}^{(m-1)}\setminus\{(i',k'),(s',k')|(29)\}$;
6: Solve $(\mathcal{P}_2)$ until convergence given all binary values: $a_{i',k'}=b_{i'}=1,c_{s',k'}=d_{s'}=1,\forall(i',k'),(s',k')\notin\mathcal{R}_{\text{off}}^{(m)}$ and $a_{i,k}=0,c_{s,k}=0,\forall(i,k),(s,k)\in\mathcal{R}_{\text{off}}^{(m)}$. If $(\mathcal{P}_2)$ is feasible, set $\pi^{(m)}$ as the value of objective function achieved at the convergence. If not, set $\pi^{(m)}=\pi^{(0)}$.
7: **until** $(\mathcal{P}_2)$ starts to be infeasible or it is feasible and $\pi^{(m)}<\pi^{(m-1)}$;

Convergence analysis: Algorithm 2 is provably convergent due to two facts. First, the DCA-based algorithm to solve $(\mathcal{P}_2)$ is guaranteed to converge as proved in the previous section. Second, the post-processing procedure is executed $\max\{(I-1)K,(S-1)K\}$ times in the worst case. Hence, the post-processing algorithm mainly consists in solving finite times the problem $(\mathcal{P}_2)$ and it converges in finite iterations with a polynomial time computational complexity.

E. Complexity Analysis

We now discuss the worst-case per-iteration computational complexity of Algorithms 1 and 2. For Algorithm 1, (25) can be easily rewritten as a second order cone program (SOCP), whose total number of variables is $KM+3SK+IK+K+S+I$ and total number of constraints is $4SK+2IK+3K+2I$. Thus, the worst-case per-iteration computational complexity of Algorithm 1 and accelerated version of Algorithm 1 (ignoring the small orders) can be written as $\mathcal{O}(K^4(M^3+S^3+I^3)(S+I))$. Next, we analyze the worst-case per-iteration complexity of Algorithm 2. First we remark that in the worst case, Algorithm 2 must iteratively solve and update the resulting parameters for the problem $(\mathcal{P}_2)$ for $\max\{(I-1)K,(S-1)K\}$ times. In each step, the worst-case per-iteration complexity of solving $(\mathcal{P}_2)$ is $\mathcal{O}(K^4(M^3+S^3+I^3)(S+I))$. Therefore, the overall worst-case per-iteration computational complexity of Algorithm 2 $\mathcal{O}(\max\{(I-1)K,(S-1)K\}K^4(M^3+S^3+I^3)(S+I))$.

TABLE II
SIMULATION PARAMETERS

Notation	Value	Notation	Value
P_s	17 dBW	$P_{i,k}^{\text{FH}}$	3 dBW
η_i	0.35	$C_s^{\text{PS}}=C^{\text{PS}},\forall s$	2.5×10^3 cycle/s
P_i^{ra}	12.5 dBW	$\rho_i,\ \forall i$	1
M_i	2	$C_i^{\text{FH}}=C^{\text{FH}},\forall i$	15 b/s/Hz
$P^{\max}$	10 dBW	$D_k=D,\forall k$	0.5 s
P_i^{ri}	2.5 dBW	α_s	3

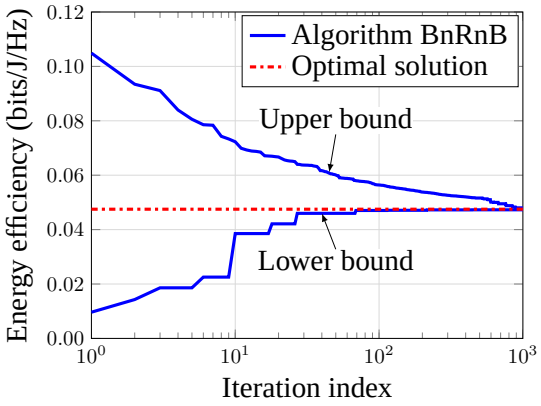
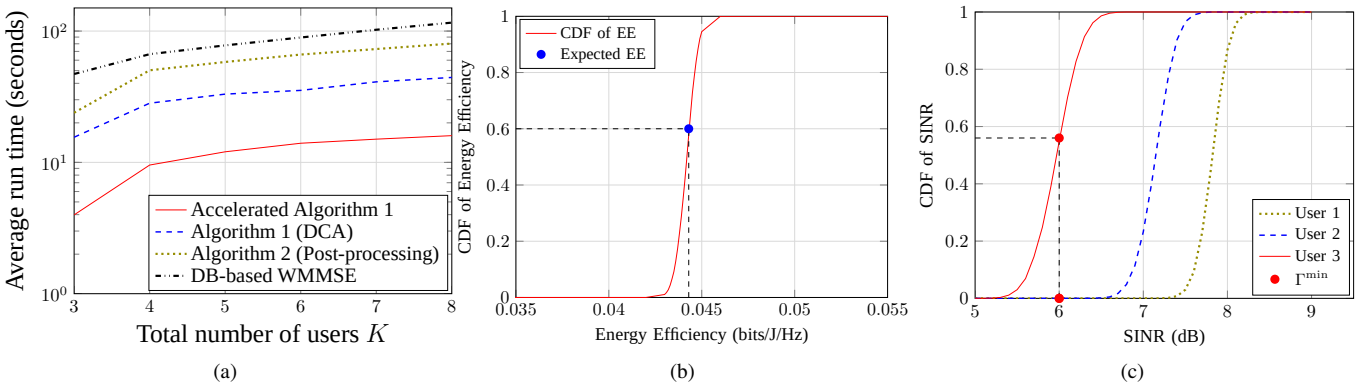


Fig. 2. Convergence of the optimal BnRnB algorithm.

VI. Numerical Results

We carry out extensive numerical experiments to evaluate the performance of the proposed algorithms. Unless mentioned otherwise, we employ the parameters in Table II in our simulations, which are taken from [40]–[42]. We set the number of PSs S to be equal to the number of UEs K to ensure that all UEs's packets are always served in order to satisfy the worst case of schemes B and C in Section VI-B. We consider a network consisting of I RRHs which are uniformly located and K UEs are randomly scattered across the considered network coverage. The path-loss is modelled as $(d_{ik}/d_0)^{-3}$ where d_{ik} is the distance between the ith RRH and the kth UE and $d_0=100$ m is the reference distance. For the VM power consumption, we set $\kappa_s=10^{-26}$ and $\mu_{s,k}$ is in cycle/s [31]. In addition, we consider the conversion calculation $\mu_{s,k}$ b/s $=(8/1900)\times\mu_{s,k}$ cycle/s to compute the processing response time for UEs [31]. Throughout our simulations, the accelerated variant of Algorithm 1 is used where ξ_k and γ_k are both first set to 0.1 in each iteration. Algorithm 1 is terminated when the increase in the objective between two consecutive iterations is less than 10^{-5}.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2799859, IEEE Transactions on Wireless Communications



(a) (b) (c)

Fig. 4. (a) Average run time of low complexity algorithms versus K; (b) CDF of the EE; (c) CDF of each UE's SINR.

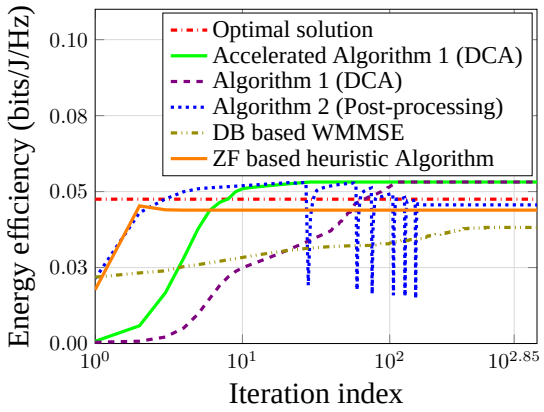


Fig. 3. Convergence of different low complexity algorithms.

A. Convergence Speed and Performance Gains by Proposed Algorithms

In Fig. 2, we show the convergence of the lower and upper bounds returned by Algorithm BnRnB for $I = 4$ and $K = 3$. As can be seen, Algorithm BnRnB requires about 10^3 iterations to compute an optimal solution. Fig. 3 compares the convergence behavior and gains between our proposed low-complexity algorithm (i.e., Algorithm 1), the DB-based WMMSE algorithm in [20], and the zero-forcing (ZF) based heuristic algorithm. Algorithm 1 needs a much smaller number of iterations to converge, compared to the DB-based WMMSE algorithm. As expected, the accelerated version of Algorithm 1 achieves an improved convergence rate due to the reason explained in V-C. Moreover, the convergence of Algorithm 1 combined with the post-processing (i.e., Algorithm 2) is also presented Fig. 3. It is clear that the objective value achieved by Algorithm 1 combined with Algorithm 2 at convergence is very close to the optimal value returned by Algorithm BnRnB. On the other hand, the DB-based WMMSE and ZF based heuristic algorithms converges to a smaller objective compared to that achieved by our proposed algorithms. This demonstrates the superiority of our proposed low-complexity algorithms.

Fig. 4(a) plots the average run time required for the proposed algorithms to obtain the final solution versus K. We observe that the average run time increases with K which is expected from the complexity analysis presented in Section V-E. Noticeably, Accelerated Algorithm 1 achieves lowest run time to return a solution which is consistent with the results shown in Fig. 3. We also observe that Algorithm 2 is much faster than the DB-based WMMSE algorithm. This again illustrates the effectiveness of our proposed algorithms compared to other existing ones.

To evaluate the time sensitivity of the computed solution from our proposed algorithm, in Figs. 4(b) and 4(c), we plot the cumulative distribution function (CDF) of the EE and each UE's SINR. The empirical CDFs are obtained as follows. First (16) is solved for a given set of channel realizations using Algorithm 2. The resulting EE and minimum SINRs are denoted as the expected EE and $\Gamma^{\min}$. Then new channel realizations are generated by adding errors to the given channel realizations. The channel errors are drawn from a zero-mean Gaussian distribution with a variance of 0.01. The obtained solution is used to compute the EE and SINRs for these new channel realizations. As shown from Fig. 4(b), 60% of the cases, the achieved EE is equal and larger than the expected EE. Fig. 4(c) also shows that UE 1 and UE 2 achieve the minimum SINR requirement $\Gamma^{\min} = 6$ dB for 100% of simulated cases. That number for UE 3 is reduced to 50%. These results are quite positive in terms of the time sensitivity of the solution.

In Fig. 5, we compare the energy efficiency performance of our proposed algorithms with the DB-based WMMSE in [20], SCA-based algorithms in [33], and the ZF based heuristic algorithm, with respect to the maximum fronthaul capacity C_i^{FH}. Here, we choose $C_i^{\text{FH}} = C^{\text{FH}}, \forall i \in \mathcal{I}$. We observe that when C^{FH} increases, energy efficiency of all methods in comparison increases accordingly. This is because larger fronthaul capacities allows more data to be transported, which requires a smaller number of activated RRHs to serve the demanding UEs and subsequently leads to reduced total power consumption. It is worth mentioning that larger fronthaul capacity promotes more coordination among RRHs so that inter-RRH interference is more effectively managed, and thus improves the overall achievable sum rate. These two factors collectively increase the energy efficiency. However, all the energy efficiency curves saturate at the high fronthaul capacity regime. This can be explained as the multi-user interference always exists despite more cooperation among the RRHs.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2799859, IEEE Transactions on Wireless Communications

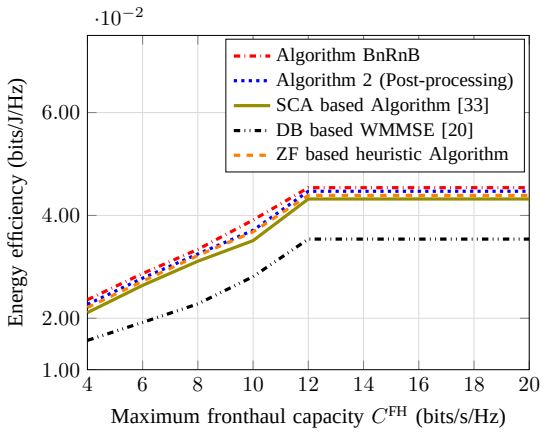


Fig. 5. EE performance of different algorithms.

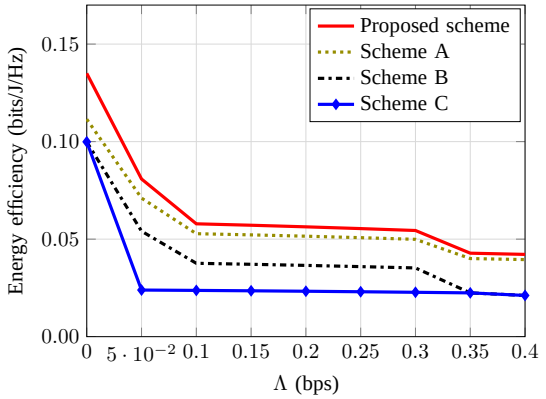


Fig. 6. EE comparison of different algorithms versus Λ

For this situation, there is an upper bound on the achievable sum rate determined by the wireless interface of the network. Thus increasing more fronthaul capacity provides no benefit to the system performance. It is also shown that our proposed algorithms outperforms the DB-based WMMSE, SCA-based, and ZF based heuristic algorithms in terms of achieved energy efficiency, which again justifies the effectiveness of our proposed methods.

B. Advantages of Proposed Computing Model

Next we evaluate the performance of our proposed VCRA scheme and the rate-dependent fronthaul power consumption (RDFP) model. The following schemes are compared:

- Proposed Scheme: the proposed VCRA scheme and RDFP model are considered, where our proposed method in Algorithm 1 and Algorithm 2 is employed.
- Scheme A: the proposed VCRA scheme without the RDFP model is considered, where the SCA-based algorithm in [33] is employed. Instead of using the RDFP model, a linear fronthaul power consumption model in [18], [19] is applied by setting $P_i^{\text{FH}} = \sum_{k\in\mathcal{K}} a_{i,k} P_{i,k}^{\text{FH}}$ with $P_{i,k}^{\text{FH}}$ being a fixed power consumption, i.e., $P_{i,k}^{\text{FH}} = 2$ Watts [19], for each data transmission between the ith fronthaul link and the kth UE.
- Scheme B: This scheme considers that the UE's workload is not split and thus the entire workload of one UE is served by only one VM (c.f., [11], [31]). Additionally, the linear fronthaul power consumption model in Scheme A is applied to this scheme. Here, the DB-based WMMSE combined with reweighted ℓ_1-norm technique in [20], [43] is employed and greedy algorithm in [31] is used to determine the active RRHs and RRH-UE associations.
- Scheme C: no PS switching ON/OFF is considered (which is similar to [11], [32]) and DB-based WMMSE combined with reweighted ℓ_1-norm technique used in Scheme B is applied.

Fig. 6 plots the energy efficiency performance of the above listed schemes as a function of the workload arrival rates when $I = 6, K = 4$. Here, we set the UE's workload arrival rates Λ_k equally to Λ, $\forall k \in \mathcal{K}$. As can be seen from Fig. 6, the energy efficiency attained by all schemes decreases when Λ increases, which can be explained as follows. As the traffic arrival rate grows, more computing resources and active PSs are needed to process the data. This results in the increase of power consumption in the BBU pool which then reduces the overall energy efficiency of the C-RAN system. In addition, we observe that our Proposed Scheme outperforms Scheme A, which verifying the benefit of considering the RDFH model in the formulated C-RAN optimization problem. Moreover, there is a remarkably large performance gap between our Proposed Scheme and Schemes B and C. This is because when Λ becomes larger, splitting the UE's workload into smaller fractions allows for more flexibility in assigning multiple VMs for different tasks and thus consolidate the existence of VMs to active PSs. As a result, more PSs can be switched OFF and more system power consumption can be saved, thereby enhancing the overall system energy efficiency. This again validates the advantages of our Proposed Scheme over the others in comparison.

In Fig. 7, the energy efficiency performance of different schemes is studied with respect to the total number of UEs K where $I = 6$. In this figure, we set the UE's workload arrival rate $\Lambda_k = \Lambda = 0.3$ bps and the delay $D_k = D = 0.5$ seconds, $\forall k \in \mathcal{K}$. It is obvious that when K increases, the energy efficiency first increases and then slightly decreases. The fact is that when the UE number increases, the total achievable rate first increases due to the multiuser diversity gain, which leads to an increase in the energy efficiency. However, when K becomes sufficiently large, a large number of RRHs and the PSs need to be activated to coordinate the induced interference. This in turn produces a huge amount of power consumption which subsequently decreases the achieved energy efficiency. Again, the energy efficiency achieved by our Proposed Scheme is much higher than that by Schemes A, B, and C.

Fig. 8 demonstrates the impact of the joint optimization of virtual computing resource in the cloud and radio resource allocation in the RAN by comparing it with the decoupled optimization problem. Note that we can integrate the coupled and decoupled problem on all the above schemes. It is worth mentioning that the decoupled problem separately optimize each component. For the decoupled problem, to enable the separation between the the virtual computing resource and the radio resource allocation, the cross-layer delay constraint in (C5) are divided into two separated constraints, namely, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2799859, IEEE Transactions on Wireless Communications

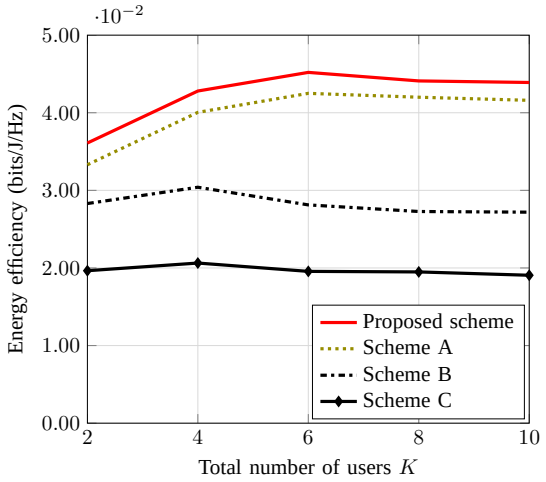


Fig. 7. EE comparison of different algorithms versus K.

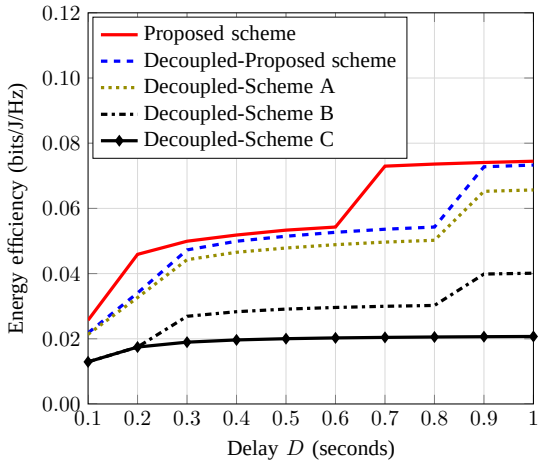


Fig. 8. EE comparison of different schemes versus D.

processing delay constraint $\tau_k \leq D_k/2$ and the transmission delay constraint $\frac{1}{R_k(\mathbf{w})-\Lambda_k} \leq D_k/2$, $\forall k \in \mathcal{K}$. Then, similar transformations and approximation techniques presented in Section V can be straightforwardly applied to solve this decoupled problem. The numerical results in Fig. 8 are obtained by setting the UEs' delay requirement to be $D_k = D$ $\forall k \in \mathcal{K}$. From Fig. 8, we can see that the energy efficiency for both the coupled and decoupled schemes increases when D grows. This is because that large delay requirement offers less computing resources which leads to more idle PSs and more power savings. Moreover, it is obvious from Fig. 8 that the coupled design outperforms the decoupled design for each scheme. Furthermore, the joint design applied to our Proposed Scheme achieves better performance than the decoupled one for Schemes A, B, and C, which again verifies the advantages of our Proposed Scheme compared to other known methods in [11], [20], [33], [43].

VII. Conclusion

We have considered the joint design of VCRA and radio resource allocation in a limited fronthaul C-RAN under a RDFP model for maximizing the network EE. To solve the formulated problem, we have customized a BnRnB algorithm to search for a globally optimal solution. We have also developed a novel low-complexity and more appealing algorithm which is provably convergent. The proposed method is inspired by the DCA method combined with the Lipschitz continuity concept, approximating the non-convex problem into a sequence of convex quadratic ones, which can be efficiently solved by dedicated convex conic solvers. After solving the continuous relaxation, a post-processing routine is executed to find a high-performance solution which is feasible to the original problem. Numerical results showed that our proposed algorithms converge rapidly and achieve a near-optimal performance as well as outperform the other existing methods. Additionally, we have also numerically demonstrated that the proposed VCRA scheme not only efficiently allocates virtual computing in the BBU pool to process the user's workload in parallel but also significantly reduces the total power consumption.

VIII. Proof of the equivalence of (9) and (10)

We prove that the constraints in (10b) and (10d) of problem (10) are hold with equalities at optimality by contradiction. Let $\Theta^\star = \{\mathbf{b}^\star, \mathbf{a}^\star, \mathbf{c}^\star, \mathbf{d}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\tau}^\star, \mathbf{w}^\star, \boldsymbol{\mu}^\star, \boldsymbol{\nu}^\star, \boldsymbol{\zeta}^\star\}$ denote an optimal solution of (10). By contradiction suppose that (10d) is inactive, i.e., $\hat{P}(\mathbf{d}^\star, \boldsymbol{\mu}^\star, \mathbf{w}^\star, \mathbf{a}^\star, \mathbf{b}^\star, \boldsymbol{\nu}^\star) < 1/\nu_0^\star$. Then there exists ν_0' such that $\nu_0' > \nu_0^\star$ and $\hat{P}(\mathbf{d}^\star, \boldsymbol{\mu}^\star, \mathbf{w}^\star, \mathbf{a}^\star, \mathbf{b}^\star, \boldsymbol{\nu}^\star) \leq 1/\nu_0'$. ν_0' is feasible to (10) but yields a strictly larger objective, which contradicts with the fact that $\Theta^\star$ is an optimal solution. Similarly, assume that $R_k(\mathbf{w}^\star) > \nu_k^\star$ for some k. We then create a new set of beamformers as $\mathbf{w}' = [\mathbf{w}_1'^T, \mathbf{w}_2'^T, \ldots, \mathbf{w}_k'^T]^T$ where $\mathbf{w}_i' = \mathbf{w}_i^\star$ if $i \neq k$ and $\mathbf{w}_i' = \xi \mathbf{w}_k^\star$ otherwise, for some $0 < \xi < 1$. Intuitively, the beamforming vector of UE k is scaled down by a factor of ξ and the beamforming vectors of other UEs remain the same. It is easy to see that there exists $\xi \in (0,1)$ such that $R_k(\mathbf{w}') > \nu_k^\star$ for all k. Note that $\|\mathbf{w}'\|_2 < \|\mathbf{w}\|_2$ and thus $\hat{P}(\mathbf{d}^\star, \boldsymbol{\mu}^\star, \mathbf{w}', \mathbf{a}^\star, \mathbf{b}^\star, \boldsymbol{\nu}^\star) < \hat{P}(\mathbf{d}^\star, \boldsymbol{\mu}^\star, \mathbf{w}^\star, \mathbf{a}^\star, \mathbf{b}^\star, \boldsymbol{\nu}^\star) \leq 1/\nu_0^\star$. Consequently, we can find ν_0' such that $\nu_0' > \nu_0^\star$ and $\hat{P}(\mathbf{y}^\star, \boldsymbol{\mu}^\star, \mathbf{w}', \mathbf{a}^\star, \mathbf{b}^\star, \boldsymbol{\nu}^\star) \leq 1/\nu_0'$, meaning that a strictly larger objective can be obtained. Again, this contradicts with the fact that $\Theta^\star$ is an optimal solution, and thus proves that the constraints in (10b) and (10d) of problem (10) hold with equalities at optimality. As a result, for given optimal solution $\Theta^\star$ to (10) we can simply obtain optimal solution $(\mathbf{b}^\star, \mathbf{a}^\star, \mathbf{c}^\star, \mathbf{d}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\tau}^\star, \mathbf{w}^\star, \boldsymbol{\mu}^\star)$ to (9) and the same objective value as that of (10). Similarly, for given optimal solution $(\mathbf{b}^\star, \mathbf{a}^\star, \mathbf{c}^\star, \mathbf{d}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\tau}^\star, \mathbf{w}^\star, \boldsymbol{\mu}^\star)$ to (9), we can easily find a feasible solution $\Theta^\star$ to (10) by setting $\nu_k^\star = R_k(\mathbf{w}^\star)$, $1/\nu_0^\star = \hat{P}(\mathbf{d}^\star, \boldsymbol{\mu}^\star, \mathbf{w}^\star, \mathbf{a}^\star, \mathbf{b}^\star, \boldsymbol{\nu}^\star)$ and $D_k - \tau_k^\star \geq \zeta_k^\star \geq 1/(\nu_k^\star - \Lambda_k)$, $\forall k \in \mathcal{K}$ that achieves the same objective value as that of (9).

IX. Calculation of box $\mathcal{V}$

The values of $\underline{\boldsymbol{\nu}} = \{\underline{\nu}_0, \underline{\nu}_1, \ldots, \underline{\nu}_K\}$ and $\overline{\boldsymbol{\nu}} = \{\overline{\nu}_0, \overline{\nu}_1, \ldots, \overline{\nu}_K\}$ can be computed as follows. From (10c), it holds that $\nu_k \geq \log(1+\Gamma_k^{\min}) = \underline{\nu_k}$, $\forall k \in \mathcal{K}$. Moreover, we have

$$\nu_k \overset{(a)}{\leq} \log\Big(1 + \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sigma_0^2}\Big) \overset{(b)}{\leq} \log\Big(1 + I \times P^{\max} \frac{\|\mathbf{h}_k\|_2^2}{\sigma_0^2}\Big) = \overline{\nu}_k, \forall k \in \mathcal{K}. \quad (30)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2799859, IEEE Transactions on Wireless Communications

where (a) is due to omitting the inter-user interference, (b) is by applying the result of Cauchy-Schwarz inequality to have $|\mathbf{h}_k^H \mathbf{w}_k|^2 \leq \|\mathbf{h}_k\|_2^2 \|\mathbf{w}_k\|_2^2 \leq IP^{\max} \|\mathbf{h}_k\|_2^2$. Similarly, an upper bound and lower bound of ν_0 can be given by $\nu_0 \geq \frac{1}{\overline{P}} = \underline{\nu}_0; \nu_0 \leq \frac{1}{\sum_{i\in\mathcal{I}} P_i^{\mathrm{ri}}} = \overline{\nu}_0$, where $\overline{P} = \sum_{i\in\mathcal{I}} \rho_i \sum_{k\in\mathcal{K}} \overline{\nu}_k + \sum_{i\in\mathcal{I}} (\frac{1}{\eta_i} P^{\max} + P_i^{\mathrm{ra}}) + \sum_{s\in\mathcal{S}} (P_s^{\mathrm{PS}} + \kappa_s C_s^{\alpha_s}))$.

X. Proof of Lemma 2

In this section, we show that $\bar{\xi}_k$ in Lemma 2 is a Lipschitz constant of $\nabla R_k(\mathbf{w})$, which is then used to prove Lemma 2. Since there is a mapping rule from a complex-valued vector into a real domain and also due to the space limitation, we will treat $\mathbf{w}$ as a real-valued vector in the following. For ease of mathematical presentation, let us rewrite $R_k(\mathbf{w})$ as

$$R_k(\mathbf{w}) = \log\left(\sigma_0^2 + \mathbf{w}^H \mathbf{H}_k \mathbf{w}\right) - \log\left(\sigma_0^2 + \mathbf{w}^H \mathbf{G}_k \mathbf{w}\right) \quad (31)$$

where $\tilde{\mathbf{H}}_k = \mathbf{h}_k \mathbf{h}_k^H$, $\mathbf{H}_k = \mathrm{blkdiag}(\underbrace{\tilde{\mathbf{H}}_k, \ldots, \tilde{\mathbf{H}}_k}_{K \text{ elements}})$, and $\mathbf{G}_k = \mathrm{blkdiag}(\tilde{\mathbf{H}}_k, \ldots, \underbrace{\mathbf{0}}_{k\text{th position}}, \ldots, \tilde{\mathbf{H}}_k)$. Then the gradient of $R_k(\mathbf{w})$ is given by

$$\nabla R_k(\mathbf{w}) = \frac{2\mathbf{w}^H \mathbf{H}_k}{\mathbf{w}^H \mathbf{H}_k \mathbf{w} + \sigma_0^2} - \frac{2\mathbf{w}^H \mathbf{G}_k}{\mathbf{w}^H \mathbf{G}_k \mathbf{w} + \sigma_0^2} \quad (32)$$

$$= h_1(\mathbf{w}) h_2(\mathbf{w}) - g_1(\mathbf{w}) g_2(\mathbf{w}) \quad (33)$$

where $h_1(\mathbf{w}) = 2\mathbf{w}^H \mathbf{H}_k$, $h_2(\mathbf{w}) = \frac{1}{\mathbf{w}^H \mathbf{H}_k \mathbf{w} + \sigma_0^2}$, $g_1(\mathbf{w}) = 2\mathbf{w}^H \mathbf{G}_k$, and $g_2(\mathbf{w}) = \frac{1}{\mathbf{w}^H \mathbf{G}_k \mathbf{w} + \sigma_0^2}$. Next we will find a Lipschitz constant of each term in (33). From the definition of $h_1(\mathbf{w})$, the following inequality holds

$$\|h_1(\mathbf{w}) - h_1(\bar{\mathbf{w}})\|_2 = \left\|\left(\mathbf{w}^H - \bar{\mathbf{w}}^H\right) \mathbf{H}_k\right\|_2 \quad (34)$$

$$\leq \|\mathbf{H}_k\|_F \|(\mathbf{w} - \bar{\mathbf{w}})\|_2 \quad (35)$$

In words, a Lipschitz constant of $h_1(\mathbf{w})$ is $\|\mathbf{H}_k\|_F$. Next we have

$$\begin{aligned}
&\|h_2(\mathbf{w}) - h_2(\bar{\mathbf{w}})\|_2 = \left\| \frac{\bar{\mathbf{w}}^H \mathbf{H}_k \bar{\mathbf{w}} - \mathbf{w}^H \mathbf{H}_k \mathbf{w}}{(\mathbf{w}^H \mathbf{H}_k \mathbf{w} + \sigma_0^2)(\bar{\mathbf{w}}^H \mathbf{H}_k \bar{\mathbf{w}} + \sigma_0^2)} \right\|_2 \\
&\leq \left\| \bar{\mathbf{w}}^H \mathbf{H}_k \bar{\mathbf{w}} - \mathbf{w}^H \mathbf{H}_k \mathbf{w} \right\|_2 \\
&\leq \left\| \bar{\mathbf{w}}^H \mathbf{H}_k \bar{\mathbf{w}} - \bar{\mathbf{w}}^H \mathbf{H}_k \mathbf{w} + \bar{\mathbf{w}}^H \mathbf{H}_k \mathbf{w} - \mathbf{w}^H \mathbf{H}_k \mathbf{w} \right\|_2 \\
&\leq \left\| \bar{\mathbf{w}}^H \mathbf{H}_k (\bar{\mathbf{w}} - \mathbf{w}) \right\|_2 + \left\| \left(\bar{\mathbf{w}}^H - \mathbf{w}^H\right) \mathbf{H}_k \mathbf{w} \right\|_2 \\
&\leq ||\bar{\mathbf{w}}||_2 \|\mathbf{H}_k\|_F \|\mathbf{w} - \bar{\mathbf{w}}\|_2 + ||\mathbf{w}||_2 \|\mathbf{H}_k\|_F \|\mathbf{w} - \bar{\mathbf{w}}\|_2 \\
&\leq 2P \|\mathbf{H}_k\|_F \|\mathbf{w} - \bar{\mathbf{w}}\|_2 \quad (36)
\end{aligned}$$

where $P = \sqrt{IP^{\max}}$. Note that the last inequality occurs due to the sum power constraint. Using (35) and (36) we can find a Lipschitz constant of the product $h_1(\mathbf{w}) h_2(\mathbf{w})$ as

$$\begin{aligned}
&\|h_1(\mathbf{w}) h_2(\mathbf{w}) - h_1(\bar{\mathbf{w}}) h_2(\bar{\mathbf{w}})\|_2 \\
&\leq \|h_2(\mathbf{w})\|_2 \|h_1(\mathbf{w}) - h_1(\bar{\mathbf{w}})\|_2 + \\
&\qquad \|h_1(\bar{\mathbf{w}})\|_2 \|h_2(\mathbf{w}) - h_2(\bar{\mathbf{w}})\|_2 \\
&\leq \left(\|\mathbf{H}_k\|_F \|h_2(\mathbf{w})\|_2 + 2P \|\mathbf{H}_k\|_F \|h_1(\bar{\mathbf{w}})\|_2\right) \|(\mathbf{w} - \bar{\mathbf{w}})\|_2 \\
&\leq \left(\|\mathbf{H}_k\|_F + \left(2P \|\mathbf{H}_k\|_F\right)^2\right) \|(\mathbf{w} - \bar{\mathbf{w}})\|_2 \quad (37)
\end{aligned}$$

We now study the Lipschitz continuity of the term $g_1(\mathbf{w}) g_2(\mathbf{w})$ in (33). Following the same algebraic manipulations the following inequalities are obtained

$$\|g_1(\mathbf{w}) - g_1(\bar{\mathbf{w}})\|_2 = \left\|\left(\mathbf{w}^H - \bar{\mathbf{w}}^H\right) \mathbf{G}_k\right\|_2$$

$$\leq \|\mathbf{G}_k\|_F \|(\mathbf{w} - \bar{\mathbf{w}})\|_2 \quad (38)$$

$$\|g_2(\mathbf{w}) - g_2(\bar{\mathbf{w}})\|_2 \leq 2P \|\mathbf{G}_k\|_F \|\mathbf{w} - \bar{\mathbf{w}}\|_2 \quad (39)$$

Thus a Lipschitz constant of $g_1(\mathbf{w}) g_2(\mathbf{w})$ is simply given by

$$\|g_1(\mathbf{w}) g_2(\mathbf{w}) - g_1(\bar{\mathbf{w}}) g_2(\bar{\mathbf{w}})\| \leq \left(\|\mathbf{G}_k\|_F + 4P^2 \|\mathbf{G}_k\|_F^2\right) \|(\mathbf{w} - \bar{\mathbf{w}})\|_2 \quad (40)$$

Combining (37) and (40) results in

$$\begin{aligned}
&\|\nabla R_k(\mathbf{w}) - \nabla R_k(\bar{\mathbf{w}})\|_2 \leq \|h_1(\mathbf{w}) h_2(\mathbf{w}) - h_1(\bar{\mathbf{w}}) h_2(\bar{\mathbf{w}})\|_2 \\
&\quad + \|g_1(\mathbf{w}) g_2(\mathbf{w}) - g_1(\bar{\mathbf{w}}) g_2(\bar{\mathbf{w}})\|_2 \leq \bar{\xi}_k \|(\mathbf{w} - \bar{\mathbf{w}})\|_2 \quad (41)
\end{aligned}$$

where

$$\bar{\xi}_k = \|\mathbf{H}_k\|_F + \left(2P \|\mathbf{H}_k\|_F\right)^2 + \|\mathbf{G}_k\|_F + \left(2P \|\mathbf{G}_k\|_F\right)^2. \quad (42)$$

In other words $R_k(\mathbf{w})$ is a $\bar{\xi}_k$-smooth function.

We now show that $f_k(\mathbf{w}) = R_k(\mathbf{w}) + \xi_k \|\mathbf{w}\|_2^2$ is strongly convex. Since $R_k(\mathbf{w})$ is $\bar{\xi}_k$-smooth, the following inequality holds

$$\left| R_k(\mathbf{w}) - R_k(\bar{\mathbf{w}}) - \nabla R_k(\bar{\mathbf{w}})^T (\mathbf{w} - \bar{\mathbf{w}}) \right| \leq \frac{\bar{\xi}_k}{2} \|\mathbf{w} - \bar{\mathbf{w}}\|^2 \quad (43)$$

which implies

$$R_k(\mathbf{w}) \geq -\frac{\bar{\xi}_k}{2} \|\mathbf{w} - \bar{\mathbf{w}}\|^2 + R_k(\bar{\mathbf{w}}) + \nabla R_k(\bar{\mathbf{w}})^T (\mathbf{w} - \bar{\mathbf{w}}) \quad (44)$$

Due to the strong convexity of $\xi_k \|\mathbf{w}\|_2^2$ we have

$$\xi_k \|\mathbf{w}\|_2^2 \geq \xi_k \|\bar{\mathbf{w}}\|_2^2 + 2\xi_k \bar{\mathbf{w}}^T (\mathbf{w} - \bar{\mathbf{w}}) + \frac{\xi_k}{2} \|\mathbf{w} - \bar{\mathbf{w}}\|^2 \quad (45)$$

Combining (44) and (45) we obtain

$$\begin{aligned}
&R_k(\mathbf{w}) + \xi_k \|\mathbf{w}\|_2^2 \geq \frac{\xi_k - \bar{\xi}_k}{2} \|\mathbf{w} - \bar{\mathbf{w}}\|^2 + R_k(\bar{\mathbf{w}}) + \\
&\qquad \xi_k \|\bar{\mathbf{w}}\|_2^2 + \mathrm{Re}\left\{ \left(\nabla R_k(\bar{\mathbf{w}}) + 2\xi_k \bar{\mathbf{w}}\right)^T (\mathbf{w} - \bar{\mathbf{w}}) \right\} \quad (46)
\end{aligned}$$

which is equivalent to

$$f_k(\mathbf{w}) \geq \frac{\xi_k - \bar{\xi}_k}{2} \|\mathbf{w} - \bar{\mathbf{w}}\|^2 + f_k(\bar{\mathbf{w}}) + \nabla f_k(\bar{\mathbf{w}})^T (\mathbf{w} - \bar{\mathbf{w}}) \quad (47)$$

(47) implies that $f_k(\mathbf{w})$ is $(\xi_k - \bar{\xi}_k)$-strongly convex, $\forall \xi_k > \bar{\xi}_k$ which completes the proof.

XI. Proof of Lemma 3

Similar to the proof of Lemma 2, $u_k(\mathbf{w}, a_{i,k})$ is strongly convex if $\nabla\left(a_{i,k} R_k(\mathbf{w})\right)$ has a Lipschitz constant of γ_k. First we have $\nabla\left(a_{i,k} R_k(\mathbf{w})\right) = [R_k(\mathbf{w}); a_{i,k} \nabla R_k(\mathbf{w})]$ which leads to

$$\begin{aligned}
&\left\|\nabla\left(a_{i,k} R_k(\mathbf{w})\right) - \nabla\left(\bar{a}_{i,k} R_k(\bar{\mathbf{w}})\right)\right\|_2^2 = \left(R_k(\mathbf{w}) - R_k(\bar{\mathbf{w}})\right)^2 \\
&\qquad + \left\|a_{i,k} \nabla R_k(\mathbf{w}) - \bar{a}_{i,k} \nabla R_k(\bar{\mathbf{w}})\right\|_2^2 \quad (48)
\end{aligned}$$

From (33) we can write

$$\begin{aligned}
&\left\|a_{i,k} \nabla R_k(\mathbf{w}) - \bar{a}_{i,k} \nabla R_k(\bar{\mathbf{w}})\right\|_2^2 \\
&= \|a_{i,k} \nabla R_k(\mathbf{w}) - a_{i,k} \nabla R_k(\bar{\mathbf{w}}) + \\
&\qquad a_{i,k} \nabla R_k(\bar{\mathbf{w}}) - \bar{a}_{i,k} \nabla R_k(\bar{\mathbf{w}})\|_2^2 \quad (49) \\
&\leq 2\left(\left\|a_{i,k} \nabla R_k(\mathbf{w}) - a_{i,k} \nabla R_k(\bar{\mathbf{w}})\right\|_2^2 + \right. \\
&\qquad \left. \left\|a_{i,k} \nabla R_k(\bar{\mathbf{w}}) - \bar{a}_{i,k} \nabla R_k(\bar{\mathbf{w}})\right\|_2^2\right) \quad (50) \\
&\leq 2\left(|a_{i,k}|^2 \left\|\nabla R_k(\mathbf{w}) - \nabla R_k(\bar{\mathbf{w}})\right\|_2^2 + \right. \\
&\qquad \left. |a_{i,k} - \bar{a}_{i,k}|^2 \left\|\nabla R_k(\bar{\mathbf{w}})\right\|_2^2\right) \quad (51) \\
&\leq 2\left(\left\|\nabla R_k(\mathbf{w}) - \nabla R_k(\bar{\mathbf{w}})\right\|_2^2 + \right. \\
&\qquad \left. |a_{i,k} - \bar{a}_{i,k}|^2 \left\|\nabla R_k(\bar{\mathbf{w}})\right\|_2^2\right) \quad (52)
\end{aligned}$$

Also from (33) the following inequality holds

$$\begin{aligned}
&\left\|\nabla R_k(\mathbf{w})\right\|_2^2 \\
&= \left\| \frac{2\mathbf{w}^H \mathbf{H}_k}{\mathbf{w}^H \mathbf{H}_k \mathbf{w} + \sigma_0^2} - \frac{2\mathbf{w}^H \mathbf{G}_k}{\mathbf{w}^H \mathbf{G}_k \mathbf{w} + \sigma_0^2} \right\|_2^2 \quad (53a)
\end{aligned}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2018.2799859, IEEE Transactions on Wireless Communications

$$\leq 2\Big(\Big\|\frac{2\mathbf{w}^H\mathbf{H}_k}{\mathbf{w}^H\mathbf{H}_k\mathbf{w}+\sigma_0^2}\Big\|_2^2+\Big\|\frac{2\mathbf{w}^H\mathbf{G}_k}{\mathbf{w}^H\mathbf{G}_k\mathbf{w}+\sigma_0^2}\Big\|_2^2\Big) \quad (53b)$$

$$\leq 4\left(||\mathbf{w}^H\mathbf{H}_k||_2^2+||\mathbf{w}^H\mathbf{G}_k||_2^2\right) \quad (53c)$$

$$\leq 4\left(||\mathbf{H}_k||_F^2+||\mathbf{G}_k||_F^2\right)||\mathbf{w}||_2^2 \quad (53d)$$

We now study the Lipschitz continuity of $R_k(\mathbf{w})$. To this end we will show that the function $\log(1+x)$ for $x \geq 0$ is Lipschitz continuous with a Lipschitz constant of 1. That is, for $u \geq 0$ and $v \geq 0$, $|\log(1+u)-\log(1+v)| \leq |u-v|$. Obviously, we only need to prove this for the case $u > v \geq 0$. Since $\log(1+x)$ is continuous and differentiable in the interval $[v,u]$, by the mean-value theorem [44], there exists $v < x_0 < u$ such that $\frac{1}{1+x_0} = \frac{\log(1+u)-\log(1+v)}{u-v}$, and thus

$$\log(1+u)-\log(1+v) = (u-v)/(1+x_0) \leq u-v \quad (54)$$

Now we have

$$\begin{aligned}&\left(R_k(\mathbf{w})-R_k(\bar{\mathbf{w}})\right)^2\\ &=\big(\log\left(1+\mathbf{w}^H\mathbf{H}_k\mathbf{w}\right)-\log\left(1+\bar{\mathbf{w}}^H\mathbf{H}_k\bar{\mathbf{w}}\right)+\\ &\quad\log\left(1+\bar{\mathbf{w}}^H\mathbf{G}_k\bar{\mathbf{w}}\right)-\log\left(1+\mathbf{w}^H\mathbf{G}_k\mathbf{w}\right)\big)^2 \quad (55a)\\ &\leq 2\left(\log\left(1+\mathbf{w}^H\mathbf{H}_k\mathbf{w}\right)-\log\left(1+\bar{\mathbf{w}}^H\mathbf{H}_k\bar{\mathbf{w}}\right)\right)^2+\\ &\quad 2\left(\log\left(1+\bar{\mathbf{w}}^H\mathbf{G}_k\bar{\mathbf{w}}\right)-\log\left(1+\mathbf{w}^H\mathbf{G}_k\mathbf{w}\right)\right)^2 \quad (55b)\end{aligned}$$

Applying (54) results in

$$\begin{aligned}&\left(R_k(\mathbf{w})-R_k(\bar{\mathbf{w}})\right)^2\\ &\leq 2|\mathbf{w}^H\mathbf{H}\mathbf{w}-\mathbf{w}^H\mathbf{H}\bar{\mathbf{w}}+\mathbf{w}^H\mathbf{H}\bar{\mathbf{w}}-\bar{\mathbf{w}}^H\mathbf{H}\bar{\mathbf{w}}|^2+\\ &2|\mathbf{w}^H\mathbf{G}_k\mathbf{w}-\mathbf{w}^H\mathbf{G}_k\bar{\mathbf{w}}+\mathbf{w}^H\mathbf{G}_k\bar{\mathbf{w}}-\bar{\mathbf{w}}^H\mathbf{G}_k\bar{\mathbf{w}}|^2 \quad (56a)\\ &\leq 4|\mathbf{w}^H\mathbf{H}(\mathbf{w}-\bar{\mathbf{w}})|^2+4|(\mathbf{w}^H-\bar{\mathbf{w}})\mathbf{H}\bar{\mathbf{w}}|^2+\\ &\qquad 4|\mathbf{w}^H\mathbf{G}_k(\mathbf{w}-\bar{\mathbf{w}})|^2+4|(\mathbf{w}^H-\bar{\mathbf{w}})\mathbf{G}_k\bar{\mathbf{w}}|^2 \quad (56b)\\ &\leq 8P^2(||\mathbf{H}_k||_F^2+||\mathbf{G}_k||_F^2)||\mathbf{w}-\bar{\mathbf{w}}||_2^2 \quad (56c)\end{aligned}$$

Combining (41),(52), (53d) and (56c) we obtain

$$\begin{aligned}&\left\|\nabla(a_{i,k}R_k(\mathbf{w}))-\nabla(\bar{a}_{i,k}R_k(\bar{\mathbf{w}}))\right\|_2 \leq\\ &\qquad\bar{\gamma}_k\sqrt{|a_{i,k}-\bar{a}_{i,k}|^2+||\mathbf{w}-\bar{\mathbf{w}}||_2^2} \quad (57)\end{aligned}$$

where

$$\bar{\gamma}_k = \sqrt{2\bar{\xi}_k^2+8\left(||\mathbf{H}_k||_F^2+||\mathbf{G}_k||_F^2\right)P^2} \quad (58)$$

This completes the proof.

REFERENCES

[1] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[2] V. W. S. Wong, R. Schober, D. W. K. Ng, and L.-C. Wang, *Key Technologies for 5G Wireless Systems*. Cambridge, UK: Cambridge University Press, 2017.

[3] P. Rost *et al.*, "Cloud technologies for flexible 5G radio access networks," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 68–76, May 2014.

[4] O. Simeone, A. Maeder, M. Peng, O. Sahin, and W. Yu, "Cloud radio access network: Virtualizing wireless access for dense heterogeneous systems," *IEEE J. Commun. Net.*, vol. 18, no. 2, pp. 135–149, Apr. 2016.

[5] P. Luong, C. Despins, F. Gagnon, and L.-N. Tran, "Designing green C-RAN with limited fronthaul via mixed-integer second order cone programming," in *Proc. IEEE Int. Conf. Communications (ICC'17)*, Paris, France, May 2017, pp. 1–6.

[6] P. Luong, L.-N. Tran, C. Despins, and F. Gagnon, "Joint beamforming and remote radio head selection in limited fronthaul C-RAN," in *Proc. IEEE VTC Fall-2016*, Montreal, Canada, Sep. 2016, pp. 1–6.

[7] P. Luong, C. Despins, F. Gagnon, and L.-N. Tran, "A fast converging algorithm for limited fronthaul C-RANs design: Power and throughput trade-off," in *Proc. IEEE Int. Conf. Communications (ICC'17)*, Paris, France, May 2017, pp. 1–6.

[8] P. Luong, F. Gagnon, C. Despins, and L.-N. Tran, "Optimal joint remote radio head selection and beamforming design for limited fronthaul C-RAN," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp. 5605–5620, Nov. 2017.

[9] 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.

[10] N. Saxena, A. Roy, and H. Kim, "Traffic-aware cloud RAN: A key for green 5G networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 1010–1021, Apr. 2016.

[11] J. Tang, W. P. Tay, and T. Q. S. 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.

[12] C.-C. Lin, P. Liu, and J.-J. Wu, "Energy-aware virtual machine dynamic provision and scheduling for cloud computing," in *Proc. IEEE Inter. Conf. Cloud Computing*, 2011, pp. 736–737.

[13] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: Insights and challenges," *IEEE Wireless Commun. Mag.*, vol. 22, no. 2, pp. 152–160, Apr. 2015.

[14] B. Zhuang, D. Guo, and M. L. Honig, "Energy-efficient cell activation, user association, and spectrum allocation in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 823–831, Apr. 2016.

[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] Y. Shi *et al.*, "Smoothed ℓ_1-minimization for green Cloud-RAN with user admission control," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 1022–36, Apr. 2016.

[17] B. Niu, Y. Zhou, H. Shah-Mansouri, and V. W. S. Wong, "A dynamic resource sharing mechanism for cloud radio access networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8325–8338, Dec. 2016.

[18] S. Luo, R. Zhang, and T. J. Lim, "Downlink and uplink energy minization through user association and beamforming in C-RAN," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 494–508, Jan. 2015.

[19] 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.

[20] M. Peng *et al.*, "Energy-efficient resource allocation optimization for multimedia heterogeneous cloud radio access networks," *IEEE Trans. Multimedia*, vol. 18, no. 5, pp. 879–892, May 2016.

[21] J. Li, J. Wu, M. Peng, W. Wang, and V. K. N. Lau, "Queue-aware joint remote radio head activation and beamforming for green cloud radio access networks," in *IEEE Global Commun. Conf. (Globecom)*, San Diego, Dec. 2015, pp. 1–6.

[22] D. Liu and C. Yang, "Energy efficiency of downlink networks with caching at base stations," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 907–922, Apr. 2016.

[23] Z. Wang, D. W. K. Ng, V. W. S. Wong, and R. Schober, "Robust beamforming design in C-RAN with sigmoidal utility and Capacity-Limited backhaul," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5583–5598, Sep. 2017.

[24] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-Efficient resource allocation in Multi-Cell OFDMA systems with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3618–3631, Oct. 2012.

[25] Y. Wu *et al.*, "Green transmission technologies for balancing the energy efficiency and spectrum efficiency trade-off," *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 112–120, Nov. 2014.

[26] Y. Li, M. Sheng, C.-X. Wang, X. Wang, Y. Shi, and J. Li, "Throughput-delay tradeoff in interference-free wireless networks with guaranteed energy efficiency," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1608–1621, Mar. 2015.

[27] J. Tang, W. P. Tay, and Y. Wen, "Dynamic request redirection and elastic service scaling in cloud-centric media networks," *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1434–1445, Aug. 2014.

[28] J. Xu and J. A. B. Fortes, "Multi-objective virtual machine placementin virtualized data center environments," in *Proc. IEEE/ACM GREENCOM-CPSCOM*, 2010, pp. 179–188.

[29] M. M. Nejad, L. Mashayekhy, and D. Grosu, "Truthful greedy mechanisms for dynamic virtual machine provisioning and allocation in clouds," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 2, pp. 594–603, Feb. 2015.

[30] X. Wang, S. Thota, M. Tornatore, H. S. Chung, H. H. Lee, S. Park, and B. Mukherjee, "Energy-efficient virtual base station formation in optical-access-enabled Cloud-RAN," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1130–1139, May 2016.

[31] K. Guo, M. Sheng, J. Tang, T. Q. S. Quek, and Z. Qiu, "Exploiting hybrid clustering and computation provisioning for green C-RAN," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 4063–4076, Dec. 2016.

[32] K. Wang, K. Yang, and C. S. Magurawalage, "Joint energy minimization and resource allocation in C-RAN with mobile cloud," *IEEE Trans. Cloud Computing*, vol. pp, no. 99, pp. 1–11, Jan. 2016.

[33] T. X. Tran and D. Pompili, "Dynamic radio cooperation for User-Centric Cloud-RAN with computing resource sharing," *IEEE Trans. Wireless Commun.*, To appear.

[34] O. Tervo, L.-N. Tran, and M. Juntti, "Optimal energy-efficient transmit beamforming for multi-user MISO downlink," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5574–5587, Oct. 2015.

[35] T. D. Pham and H. A. L. Thi, *Recent Advances in DC Programming and DCA*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.

[36] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in optimization*, vol. 1, no. 3, Nov. 2014.

[37] D. Nguyen, L.-N. Tran, P. Pirinen, and M. Latva-aho, "On the spectral efficiency of full-duplex small cell wireless systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 4896–4910, Sep. 2014.

[38] C. Fan, Y. J. Zhang, and X. Yuan, "Dynamic nested clustering for parallel PHY-layer processing in cloud-RANs," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1881–1894, Mar. 2016.

[39] P. J. Burke, "The output of a queuing system," *Oper. Res.*, vol. 4, no. 6, pp. 699–704, Dec. 1956.

[40] B. Dai and W. Yu, "Energy efficiency of downlink transmission strategies for cloud radio access networks," *IEEE J. Sel. Areas Commun.*, vol. PP, no. 99, pp. 1–14, Mar. 2016.

[41] Y. Cheng, M. Pesavento, and A. Phillip, "Joint network optimization and downlink beamforming for CoMP transmission using mixed integer conic programming," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3972–3987, Aug. 2013.

[42] A. Kansal, F. Zhao, N. Kothari, and A. A. Bhattacharya, "Virtual machine power metering and provisioning," in *Proc. ACM Symp. on Cloud Computing (SoCC'10)*, 2010, pp. 39–50.

[43] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access*, vol. 31, no. 2, pp. 1326–1339, Oct. 2014.

[44] R. M. Mcleod, "Mean value theorems for vector valued functions," *Edinburgh Math. Soc.*, vol. 14, pp. 197–209, 1965.

Phuong Luong (S'15) received the B.Eng. degree in telecommunications and electrical engineering from the Hanoi University of Science and Technology, Vietnam, in 2009, and the M.E. degree in the Department of Computer Science from Kyung Hee University, Yongin, South Korea, in 2012. She is currently working towards the Ph.D. degree with the École de Technologie Supérieure (ÉTS), Montréal, QC, Canada. Her current research interests include radio resource management in cloud radio access networks (C-RAN), cloud computing, network virtualization, green communications.

François Gagnon holds a Bachelor of Engineering degree and a Doctorate in Electrical Engineering from the École Polytechnique de Montréal, and has been a professor in the Department of Electrical Engineering at the École de technologie Supérieure (ÉTS) since 1991. He served as director of this department from 1999 to 2001. He has held industrial research chairs since 2001. In addition to holding the Richard J. Marceau Industrial Research Chair for Wireless Internet in developing countries, François Gagnon also holds the NSERC-Ultra Electronics Chair in Wireless Emergency and Tactical Communication, the most prestigious industrial chair program in Canada. He also founded the Communications and Microelectronic Integration Laboratory (LACIME) and was its first director. He has been very involved in the creation of the new generation of high-capacity line-of-sight military radios offered by the Canadian Marconi Corporation, which is now Ultra Electronics Tactical Communication Systems. Ultra-Electronics TCS and ÉTS have obtained the NSERC Synergy prize for this collaboration. Professor Gagnon serves on the boards of funding agencies and companies, he specializes in wireless communications, modulation, coding, microelectronics, signal processing, equalization, software defined radio, mobile communication and fading channels. He is actively involved in the SmartLand project of UTPL, Ecuador, the STARACOM strategic research network and the Réseau Québec Maritime.

Charles Despins received the bachelor's degree in electrical engineering from McGill University, Montreal, QC, Canada, in 1984, and the master's and Ph.D. degrees from Carleton University, Ottawa, ON, Canada, in 1987 and 1991, respectively. He was with CAE Electronics as a member of the Technical Staff from 1984 to 1985, the Department of Electrical and Computer Engineering, École Polytechnique de Montréal, Canada, as a Lecturer and a Research Engineer from 1991 to 1992, and a Faculty Member with the Institut National de la Recherche Scientifique, Université du Québec, Montréal, from 1992 to 1996. From 1996 to 1998, he was with Microcell Telecommunications Inc., a Canadian GSM operator, and was responsible for industry standard and operator working groups, as well as for technology trials and technical support for joint venture deployments in China and India. From 1998 to 2003, he was Vice President and Chief Technology Officer of Bell Nordiq Group Inc., a wireless and wireline network operator in northern and rural areas of Canada. From 2003 to 2016, he was the President and CEO of Prompt Inc., a university-industry research consortium in the field of information and communications technologies. He is currently a professor of electrical engineering at the École de Technologie Supérieure (Université du Québec), with research interests in wireless communications. He is also a Guest Lecturer in the M.B.A. program with McGill University, Montreal. He received the IEEE Vehicular Technology Society Best Paper of the Year prize in 1993, and the Outstanding Engineer Award in 2006 from the IEEE Canada. He is a member of the Order of Engineers of Québec and is also a Fellow (2005) of the Engineering Institute of Canada.

Le-Nam Tran (M'10–SM'17) received the B.S. degree in electrical engineering from Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, in 2003 and the M.S. and Ph.D. degrees in radio engineering from Kyung Hee University, Seoul, Korea, in 2006 and 2009, respectively. He is currently a Lecturer/Assistant Professor at the School of Electrical and Electronic Engineering, University College Dublin, Ireland. Prior to this, he was a Lecturer at the Department of Electronic Engineering, Maynooth University, Co. Kildare, Ireland. From 2010 to 2014, he had held postdoc positions at the Signal Processing Laboratory, ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm, Sweden (2010-2011), and at Centre for Wireless Communications and the Department of Communications Engineering, University of Oulu, Finland (2011-2014). His research interests are mainly on applications of optimization techniques on wireless communications design. Some recent particular topics include energy-efficient communications, cloud radio access networks, massive MIMO, and full-duplex transmission. He has authored or co-authored in some 70 papers published in international journals and conference proceedings. Dr. Tran is an Associate Editor of EURASIP Journal on Wireless Communications and Networking. He was Symposium Co-Chair of Cognitive Computing and Networking Symposium of International Conference on Computing, Networking and Communication (ICNC 2016).