Title	Optimal Energy-Efficient Beamforming Designs for Cloud-RANs With Rate-Dependent Fronthaul Power
Authors(s)	Luong, Phuong; Gagnon, François; Despins, Charles; Tran, Le-Nam
Publication date	2019-07
Publication information	IEEE Transactions on Communications, 67 (7): 5099-5113
Publisher	IEEE
Item record/more information	http://hdl.handle.net/10197/11574
Publisher's statement	© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Publisher's version (DOI)	10.1109/TCOMM.2019.2906590

Optimal Energy-Efficient Beamforming Designs for Cloud-RANs with Rate-Dependent Fronthaul Power

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

Abstract

We study the downlink of a limited fronthaul capacity cloud-radio access networks (C-RANs). Three energy efficiency metrics, namely global energy efficiency (GEE), weighted sum energy efficiency (WSEE), and energy efficiency fairness (EEF) are maximized by jointly designing transmit beamforming, remote radio head (RRH) selection, and RRH-user association. Further, we incorporate a rate-dependent fronthaul power model, in which the fronthaul power consumption is proportional to the user sum rate. The formulated problems are difficult to solve. Our first contribution is to customize a branch and reduce and bound (BRB) method based on monotonic optimization to find a globally optimal solutions for the three energy efficiency maximization problems. Then, for a more practical approach, we propose a unified framework based on successive convex approximation (SCA) method that can be applied for all the considered problems. Our novelty lies in the equivalent transformations leading to more tractable problems that are amenable to the SCA. Specifically, appropriate continuous relaxation and convex approximation techniques are employed to arrive at a sequence of second order cone programs (SOCPs) for which dedicated solvers are available. Then, a post-processing algorithm is devised to obtain a high-performance feasible solution from the continuous relaxation. Numerical results demonstrate that the proposed SCA based algorithms converge rapidly and achieve near-optimal performance as well as outperform the known methods. They also highlight the importance of the rate dependent fronthaul power model in designing the energy efficient C-RANs.

Index Terms

Base station selection, beamforming, cloud radio access networks, energy efficiency, limited fronthaul, mixed integer second order cone programming, optimization.

I. INTRODUCTION

The exponential growth of data demand from numerous high-speed applications along with seamless area coverage requirement have been drawing significant attention in the design of the 5G networks [1]. This demand excessively increases the energy consumption due to the additional deployed base stations and transporting networks. According to an estimation in [2], 60% of the total power consumption of a wireless telecommunication network comes from the base station side. This is also the main source of CO_2 emission and tremendous electrical costs for mobile network operators. Consequently, the importance of energy efficient design for mobile wireless networks has been raised in recent years.

Cloud-radio access networks (C-RANs) are promising solution to significantly enhance not only the system spectral efficiency but also the global network EE [3]. Particularly, a C-RAN composes of multiple low-power low-cost remote radio heads (RRHs) which are typically simplified, including only some basic radio frequency (RF) functions to handle the transmission/reception of radio signals to/from users. At the core of a C-RAN, there is a cloud of BBUs which are connected with RRHs through fronthauls and are interconnected to centrally carry out sophisticated baseband signal processing tasks, instead of doing that at RRHs [4]. This novel architecture avails of the powerful computing capability of baseband unit (BBU) pool in the cloud to coordinate signal transmission and reception on a large scale, thereby increasing the throughput [5]. Further, RRHs are placed closer to users to reduce the transmit power while still achieving high data rate. Thus, the global network EE is greatly enhanced.

Despite these clear benefits, C-RANs pose some certain challenges on radio resource allocation problems. Firstly, the RRHs cooperation scale depends on the capacity of the fronthauls that are limited in practice, causing an impact on the potential EE performance of the C-RANs system [6]. Secondly, the power consumption of fronthauls increases proportionally with the number of associated users, but this problem has not been addressed satisfactorily when the user number becomes large [7]. To this end, it is critical to consider a joint design of beamforming, RRH-user association and RRH selection along with a realistic fronthaul power consumption model to attain the optimal EE performance of C-RANs.

The radio resource allocation problem for maximizing EE has been investigated for various wireless communication networks [8]–[13]. For example, an efficient iterative algorithm was proposed in [8] by applying the successive convex approximation (SCA) method to maximize

network EE in multi-user multiple input single output (MISO) systems while the problem of EE maximization in two-tier wireless backhaul hetnets was considered in [9]. The EE maximization problem of joint beamforming and power splitting design for MISO SWIPT systems was studied in [10], where Lagrangian relaxation combined with Dinkelbach's method was proposed. Furthermore, a low-complexity approximation and alternating optimization method was presented in [11] to obtain the maximal EE in multirelay OFDM networks. To solve the weighted sum EE maximization problem for MISO interference channels, an efficient distributed beamforming algorithm based on pricing mechanism was proposed in [12].

To address the weighted EE maximization problem in C-RANs, a generalized weighted minimum mean square error (WMMSE) approach was used in [14] under the Lyapunov framework. By exploiting the ℓ_1/ℓ_2-norm approximation and the block coordinate descent (BCD) method, an iterative algorithm was proposed to maximize the minimum EE in a user-centric green C-RAN [15]. For energy efficient D2D communication in C-RANs, a joint design of channel selection and power allocation was presented in [16], which was then solved by applying the Dinkelbach method. In [17], the authors proposed a virtual base station architecture for improving EE in C-RANs. In [18], an energy consumption minimization algorithm was developed via the design of beamforming and user association in C-RAN by utilizing the reweighted ℓ_1–norm approximation method while [19] employed a smoothed ℓ_p minimization approach for solving the beamforming and RRH selection of power consumption minimization problem. Motivated by the reweighted ℓ_1-norm approximation technique in [18], [20] proposed a joint design of beamforming and energy to minimize the total energy cost in C-RANs where RRHs are equipped with renewable energy resources. The authors in [21] studied a framework for green C-RANs by optimizing computation provisioning in the BBU pool coupled with hybrid clustering. Through a joint design of beamforming and RRH selection in a C-RAN, a tradeoff between sum rate maximization and total power minimization was optimized in [22] using the SCA method, while the global network EE taking into account the virtual computing resources was maximized using the difference of convex method in [23]. For a analytical framework, [24] analyzed the maximal EE gain for a given cache strategy at the base stations under the condition of limited-capacity backhaul.

In this paper, we study resource allocation problems optimizing various EE metrics of limited-capacity fronthaul C-RAN. We consider a more practical model where the power consumed by the fronthaul links depends on the associated user's rate served by the corresponding RRHs [18]. This is different from the existing related works in [15], [25], [26] where the fronthaul

power consumption is simply assumed to be linear, e.g., a rate independent function depending only on the active or sleep modes of the corresponding fronthaul links. In all these related works, the fronthaul power consumption is considered as a constant power dissipation assuming the maximum capacity consumption mode, which is not always the case in reality. Thus, to reflect the fronthaul power consumption more accurately, we need to model it as a function of the sum of users' data rates associated with a fronthaul link, which depends beamforming vectors [18], [27]. The formulated problems are generally combinatorial non-convex, which immediately leads to the following challenges: (i) the fractional and non-convex nature of the objective functions, (ii) the non-convexity of the per-fronthaul capacity constraints and (iii) the combinatorial nature of the introduced binary variables. Another problem is that even if these binary association variables are relaxed to be continuous, the resulting problem is still non-convex. Different from previous works of EE maximization where Dinklebach's method is applied to convert the fractional objective function into the subtraction form requiring two-layer iterative methods, we directly tackle the considered problems by proposing novel transformations to arrive at equivalent but more tractable forms. Based on these, we develop an iterative algorithm using SCA method [28] to arrive at a mixed integer-second order cone program (MI-SOCP) in each iteration, which can be solved optimally by dedicated MI-SOCP solvers to compute a high-quality feasible solution. Our contributions are outlined as follows.

- We study three different EE objectives: (i) GEE which quantifies the EE of the overall network, (ii) WSEE which can balance the EE of each individual RRH, (iii) and EEF which provides the best EE fairness among all RRHs. Unlike the existing models where the fronthaul power is a quadratic or linear function of the respective resources, we consider the rate-dependent power model where the fronthaul power consumption depends on the rate served by the corresponding RRHs. Compared to the previous studies [14]–[21], we are first to consider the GEE, WSEE, and EEF maximization problems which takes into account the rate-dependent power model.

- We first transform the three EE maximization problems into equivalent forms which are more amendable to a customized branch-and-reduce-and-bound (BRB) algorithm to find a globally optimal solution. To develop a low-complexity and unified approach to all the considered problems, we propose transformation and approximation techniques in the light of the SCA method. In particular we employ the continuous relaxation on the binary variables together with appropriate convex approximation techniques to approximate the non-convex problems into a sequence of convex approximated ones. Finally, we propose an efficient post-processing algorithm to compute

a high-quality sub-optimal solution after the SCA-based method converges.

- We carry out extensive numerical experiments to show the effectiveness of our proposed algorithms in terms of convergence rate and achieved EE performance for the three different EE metrics. Finally, the impact of the rate-dependent power consumption on the EE of the C-RANs is demonstrated as it outperforms the linear fronthaul power consumption model which has been commonly adopted in the literature.

A. Organization

The rest of the paper is organized as follows. Section II introduces the system model. Section III formulates our joint transmit beamforming, RRH selection, and RRH-UE association design into three different energy efficiency optimization problems. Section IV provides a global optimization algorithm. In Section V, we introduce our proposed low complexity algorithms based on SCA. Section VI presents the 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 $\mathbf{x}$. $|x|$ represents the modulus of $x \in \mathbb{C}$, while $\|\mathbf{x}\|_2$ is the ℓ_2-norm of the vector $\mathbf{x}$. $\mathbb{E}\{\cdot\}$ denotes the expectation operator. $\mathrm{Re}(.)$ stand for the real part of the argument. $\mathcal{O}$ represents the big O notation.

II. System Model

A. Transmission Model

We consider the downlink of a C-RAN consisting of I RRHs and K single antenna UEs, where ith RRH is equipped with M_i antennas. We denote $\mathcal{I} = \{1, \ldots, I\}$ and $\mathcal{K} = \{1, \ldots, K\}$ as the set of RRHs and UEs, respectively. As shown in Fig. 1, we consider that a cloud of BBUs is connected with all RRHs through the corresponding fronthaul links, e.g., high-speed optical ones, where the ith fronthaul link has a limited maximum capacity $C_i^{\max}$. Each UE is cooperatively served by a specific RRHs cluster, and one RRH can serve more than one user simultaneously. Different clusters for different users may be overlapped. Let us denote by s_k the signal with unit power, i.e., $\mathbb{E}\{s_k s_k^*\} = 1$, intended for the kth UE and by $\mathbf{w}_{i,k} \in \mathbb{C}^{M_i \times 1}$ the transmit beamforming vector from the ith RRH to the kth UE. The vector of channel coefficients

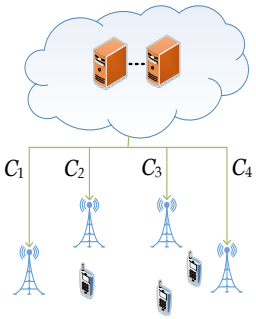

Fig. 1. Limited fronthaul C-RAN.

encompassing small-scale fading and pathloss from the ith RRH to the kth UE is represented by $\mathbf{h}_{i,k} \in \mathbb{C}^{M_i \times 1}$. The received signal at the kth UE is given by

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

where $\mathbf{w}_k \triangleq [\mathbf{w}_{1,k}^T, \mathbf{w}_{2,k}^T, \ldots, \mathbf{w}_{I,k}^T]^T \in \mathbb{C}^{M \times 1}$ represents the set of beamforming vectors intended for the kth UE, $\mathbf{h}_k \triangleq [\mathbf{h}_{1,k}^T, \mathbf{h}_{2,k}^T, \ldots, \mathbf{h}_{I,k}^T]^T \in \mathbb{C}^{M \times 1}$ represents the channel vector from all RRHs to the kth UE, $M = \sum_{i \in \mathcal{I}} M_i$, and $z_k \sim \mathcal{CN}(0, \sigma_0^2)$ is the additive white Gaussian noise (AWGN) with the noise power σ_0^2. 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 $\left\|\mathbf{w}_{i,k}\right\|_2 > 0$. By treating interference as noise, the achievable rate in b/s/Hz at the kth UE is given by

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

where

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

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

B. Fronthaul Capacity Constraint

After the BBU pool performs a relevant radio resource management algorithm to determine the beamforming vectors, data for the kth UE is routed from the BBU pool to the ith RRH via the ith fronthaul link only if $\left\|\mathbf{w}_{i,k}\right\|_2 > 0$. Let us introduce binary variables $a_{i,k} \in \{0, 1\}, \forall i \in \mathcal{I}$ and $k \in \mathcal{K}$ to 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. The fronthaul capacity consumed at the ith RRH is the accumulated rates of the users served by the ith RRH and thus, the per-fronthaul capacity constraints can be given by

$$\sum_{k\in\mathcal{K}} a_{i,k} R_k\left(\mathbf{w}\right) \leq C_i^{\max}, \forall i \in \mathcal{I}. \tag{4}$$

C. Rate-Dependent Power consumption

According to [25], the power consumption at a RRH consists of two parts: (i) the power dispatched at the power amplifiers in each RRH which is a function of transmitted signals; (ii) the static power due to electronic components which are powers required to keep the ith RRH and its corresponding fronthau link in active and sleep mode, denoted by P_i^{ra} and P_i^{ri}, respectively. To represent the operation mode of the ith RRH, we introduce a binary variable $b_i = \{0, 1\}, \forall i \in \mathcal{I}$. In particular, $b_i = 0$ states that the ith RRH is in sleep mode and $b_i = 1$ means otherwise. The power consumption at the ith RRHs can be expressed as below

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

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

In this paper, we incorporate a rate-dependent power consumption, in which the fronthaul power depends on the rate served by the corresponding RRHs. In particular, the ith fronthaul transfers the total transmission rates of users served by the corresponding RRH, thus the power consumed for the ith fronthaul is proportional to the sum achievable rates at the ith RRH [18]. Hence, the rate-dependent fronthaul power consumption at the ith fronthaul is given by

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

where ρ_i is a constant scaling factor and $\mathbf{a}_i = \left[\mathbf{a}_{i,1}, \ldots, \mathbf{a}_{i,K}\right]^T$. In summary, the total power consumption of all RRHs and fronthaul links is given by

$$P^{\text{tot}}\big(\mathbf{w}, \mathbf{a}, \mathbf{b}\big) = \sum_{i\in\mathcal{I}} \left(\underbrace{P_i^{\text{FH}}(\mathbf{w}, \mathbf{a}_i) + P_i^{\text{rrh}}(\mathbf{w}, b_i)}_{P_i(\mathbf{w}, \mathbf{a}_i, \mathbf{b}_i)} \right) \tag{7}$$

where we denote $\mathbf{b} = \left[b_1, \ldots, b_I\right]^T$, and $\mathbf{a} = \left[\mathbf{a}_1^T, \ldots, \mathbf{a}_I^T\right]^T$.

III. Problem Formulation

In this paper, we study various EE metrics via a joint design of beamforming, RRH-UE association and RRH selection. First, global energy efficiency (GEE) is referred as the ratio of network sum rate and the total network power consumption, which is presented as

$$f_{\text{GEE}}(\mathbf{w},\mathbf{a},\mathbf{b}) = \frac{R^{\text{tot}}(\mathbf{w})}{P^{\text{tot}}\big(\mathbf{w},\mathbf{a},\mathbf{b}\big)} \tag{8}$$

where the total rate over all RRHs is given by

$$R^{\text{tot}}(\mathbf{w}) = \sum_{i}\sum_{k} a_{i,k}R_k(\mathbf{w}). \tag{9}$$

GEE is the common EE objective in the literature used to quantify the system EE performance, however GEE is not able to balance the maximization of individual EEs which have different priorities. To handle this issue, the weighted sum of individual energy efficiency (WSEE) over all RRHs is studied and given by

$$f_{\text{WSEE}}(\mathbf{w},\mathbf{a},\mathbf{b}) = \sum_{i\in\mathcal{I}} \lambda_i \frac{\sum_{k\in\mathcal{K}} a_{i,k}R_k(\mathbf{w})}{P_i(\mathbf{w},\mathbf{a}_i,\mathbf{b}_i)} \tag{10}$$

where $\lambda = [\lambda_1^T,\ldots,\lambda_I^T]^T$ is the vector of the weights associated to the corresponding RRHs. It is worth mentioning that WSEE fails to guarantee the fairness between RRHs, leading the introduction of fair EE metric which aims at maximizing the minimum EE across all RRHs and is mathematically stated as

$$f_{\text{FEE}}(\mathbf{w},\mathbf{a},\mathbf{b}) = \min_{i\in\mathcal{I}}\{\frac{\sum_{k\in\mathcal{K}} a_{i,k}R_k(\mathbf{w})}{P_i(\mathbf{w},\mathbf{a}_i,\mathbf{b}_i)}\} \tag{11}$$

We are now ready to formulate a joint design of beamforming, RRH-UE association and RRH selection that maximizes the three different EE merits in the considered C-RAN system as

$$(\mathcal{E}_{\text{X}}):\ \max_{\mathbf{b},\mathbf{a},\mathbf{w},\boldsymbol{\nu}}\ f_{\text{X}}(\mathbf{w},\mathbf{a},\mathbf{b}) \tag{12a}$$

$$\text{s.t.}\quad \Gamma_k(\mathbf{w}) \geq \Gamma_k^{\min}, \forall k\in\mathcal{K} \tag{12b}$$

$$\sum_{k\in\mathcal{K}} \big\|\mathbf{w}_{i,k}\big\|_2^2 \leq b_i P^{\max}, \forall i\in\mathcal{I} \tag{12c}$$

$$\big\|\mathbf{w}_{i,k}\big\|_2^2 \leq a_{i,k}\nu_{i,k}, \forall i\in\mathcal{I}, \forall k\in\mathcal{K} \tag{12d}$$

$$\nu_{i,k} \leq a_{i,k}P^{\max}, \forall i\in\mathcal{I}, \forall k\in\mathcal{K} \tag{12e}$$

$$a_{i,k} \leq b_i, \forall i\in\mathcal{I}, \forall k\in\mathcal{K} \tag{12f}$$

$$\sum_{k\in\mathcal{K}} a_{i,k} R_k(\mathbf{w}) \le C_i^{\max}, \forall i \in \mathcal{I} \tag{12g}$$

$$a_{i,k} \in \{0,1\}, b_i \in \{0,1\}, \forall i \in \mathcal{I}, \forall k \in \mathcal{K} \tag{12h}$$

where X can be GEE, WSEE, and EEF representing for the performance metrics of EE introduced in (8), (10), and (11), respectively. In addition, the introduction of the set of auxiliary variables $\boldsymbol{\nu} = \{\nu_{i,k} \ge 0, \forall i \in \mathcal{I}, k \in \mathcal{K}\}$ and the constraints in (12) deserve further explanation. Intuitively, $\nu_{i,k}$ represents the *soft* power transmitted from the ith RRH to UE k. Constraint (12b) is to ensure the QoS requirement for the kth user, where $\Gamma_k^{\min}$ is the predetermined SINR requirement for the kth user. Moreover, constraint (12c) implies that the total transmit power at each RRH is limited by a given budget power $P^{\max}$. The constraints (12c) and (12d) are to make sure that when the ith RRH is in sleep mode, e.g., $b_i = 0$, no power will be transmitted from it. This can be easily seen as $b_i = 0$, then $a_{i,k} = 0$ for all $k \in \mathcal{K}$ and $\sum_{k\in\mathcal{K}} \left\|\mathbf{w}_{i,k}\right\|_2^2 = 0$. Similarly, in (12d) we also guarantee that the transmit power $\left\|\mathbf{w}_{i,k}\right\|_2^2$ from the ith RRH to the kth user is zero if $a_{i,k} = 0$. The constraint in (12e) means that the soft power from the ith RRH to the kth user should not exceed $P^{\max}$. Finally, the per-fronthaul capacity constraint is explicitly presented in (12g). Note that, the constraint (12d) is called a rotated SOC which can be reformulated as an SOC

$$\left(a_{i,k} + \nu_{i,k}\right)/2 \ge \left\| \left[\left(a_{i,k} - \nu_{i,k}\right)/2, \mathbf{w}_{i,k}^T\right]^T \right\|_2 \tag{13}$$

We note that in this paper, problem $(\mathcal{E}_{\mathrm{X}})$ is basically formulated for the star-like topology in which all RRHs are directly connected to the BBU pool. However, we can extend our work to cope with tree topology C-RANs, where each RRH is directly connected to a single ancestor RRH and only top-level RRHs are directly connected to the BBU pool [29]. Particularly, the ith RRH can be connected to the BBU pool through a set of ancestor RRHs, denoted by $\mathcal{A}_i$, $\forall i \in \mathcal{I}$. According to the tree topology, if the data of the kth user is routed from the BBU pool to the ith RRH, it is also delivered to all RRHs in $\mathcal{A}_i$. Thus, the tree topology can be considered by posing the extra constraints, i.e., $a_{i,k} \le \prod_{\forall j\in\mathcal{A}_i} a_{j,k}$, $\forall i \in \mathcal{I}, \forall k \in \mathcal{K}$ to problem $(\mathcal{E}_{\mathrm{X}})$ to ensure that when the kth user's data is available at the ith RRH, it must be so at all ancestor RRHs of the ith RRH. In this way, the fronthaul data rate consumed at each RRH must be the accumulated rate of users served by itself and its all ancestor RRHs. It will become clear later that our proposed solutions for $(\mathcal{E}_{\mathrm{X}})$ presented in the following sections can also be easily modified to deal with tree topology C-RANs.

IV. Globally Optimal Solution

In this section we present a solution to solve $(\mathcal{E}_{\mathrm{X}})$ in optimally. Before proceeding further, we provide some comments on the complexity of $(\mathcal{E}_{\mathrm{X}})$. First, as we mentioned above that problem $(\mathcal{E}_{\mathrm{X}})$ is generally NP-hard. Moreover, even when binary variables $\mathbf{a}$ and $\mathbf{b}$ are relaxed to be continuous, the obtained problem is still non-convex because of the non-convexity of the objective function (12a) and the constraint (12g). More precisely, in mathematical programming, (12) is categorized as a mixed integer non-convex problem for which such a method in [18], [21], [30], [31] is not applicable to find a globally optimal solution. To the best of our knowledge, there is no off-the-shelf solver for (12). In what follows, we present an equivalent formulation of $(\mathcal{E}_{\mathrm{X}})$, based on which a binary branch and reduce and bound (BRB) algorithm using monotonic optimization (MO) is customized to solve $(\mathcal{E}_{\mathrm{X}})$ optimally.

A. Equivalent Formulation

Let us introduce the slack variables $\boldsymbol{\tau} = \{\tau_k \geq 0\}_{\forall k \in \mathcal{K}}$, $\mathbf{t} = [t_0, t_1, \ldots, t_I]^T$, and compact the variables $\mathbf{a}$, $\boldsymbol{\tau}$ and $\mathbf{t}$ into a vector denoted as $\mathbf{s} = [\mathbf{a}^T, \boldsymbol{\tau}^T, \mathbf{t}^T]^T \in R_+^S$ with dimension $S = N + K + I + 1$ where $N = IK$ is the length of vector $\mathbf{a}$. Thus, we are able to rewrite $(\mathcal{E}_{\mathrm{X}})$ into the following equivalent problem

$$\max_{\mathbf{a},\mathbf{b},\mathbf{w},\boldsymbol{\nu},\boldsymbol{\tau},\mathbf{t}} \quad f_{\mathrm{X}}(\mathbf{s}) \tag{14a}$$

$$\text{s.t.} \quad R_k(\mathbf{w}) \geq \tau_k, \forall k \in \mathcal{K} \tag{14b}$$

$$\tau_k \geq \log_2(1 + \Gamma_k^{\min}), \forall k \in \mathcal{K} \tag{14c}$$

$$\begin{cases} \tilde{P}^{\mathrm{tot}}(\mathbf{w}, \mathbf{a}, \mathbf{b}, \boldsymbol{\tau}) \leq 1/t_0, & \text{if X is GEE} \\ \tilde{P}_i(\mathbf{w}, \mathbf{a}, \mathbf{b}, \boldsymbol{\tau}) \leq 1/t_i, \forall i \in \mathcal{I} & \text{otherwise} \end{cases} \tag{14d}$$

$$\sum_{k \in \mathcal{K}} a_{i,k}\tau_k \leq C_i^{\max}, \forall i \in \mathcal{I} \tag{14e}$$

$$(12c), (12e), (12f), (12h), (13). \tag{14f}$$

where three different EE objectives are expressed in the form of $f_{\mathrm{X}}(\mathbf{s})$ as below

$$f_{\mathrm{X}}(\mathbf{s}) = \begin{cases} t_0 \sum_i \sum_k a_{i,k}\tau_k & \text{if X is GEE} \\ \sum_{\forall i \in \mathcal{I}} \lambda_i t_i \sum_k a_{i,k}\tau_k & \text{if X isWSEE} \\ \min_{\forall i \in \mathcal{I}} \{t_i \sum_k a_{i,k}\tau_k\} & \text{if X is FEE} \end{cases} \tag{15}$$

According to the slack variable introduction of τ and $\mathbf{t}$ and the equivalent transformation, the total power consumption function $P^{\text{tot}}\big(\mathbf{w},\mathbf{a},\mathbf{b}\big)$ and power consumption function at each RRH and corresponding fronthaul $P_i(\mathbf{w},\mathbf{a}_i,\mathbf{b}_i)$ in the denominator of objective function in (12a) are respectively rewritten in the newly additional constraints in (14d) as

$$\tilde{P}^{\text{tot}}\left(\mathbf{w},\mathbf{a},\mathbf{b},\boldsymbol{\tau}\right)=\sum_{i\in\mathcal{I}}\left(\rho_i\sum_{k\in\mathcal{K}}a_{i,k}\tau_k+P_i^{\text{rrh}}(\mathbf{w},b_i)\right) \tag{16}$$

$$\tilde{P}_i\left(\mathbf{w},\mathbf{a},\mathbf{b},\boldsymbol{\tau}\right)=\rho_i\sum_{k\in\mathcal{K}}a_{i,k}\tau_k+P_i^{\text{rrh}}(\mathbf{w},b_i),\forall i\in\mathcal{I} \tag{17}$$

The following lemma is to characterize the property of problem (14).

Lemma 1. *The formulations in the problem* (12) *and* (14) *are equivalent in the sense that they have the same optimal solution set and objective.*

Proof: The equivalence between (12) and (14) is due to the observation that at optimality of (14), the inequalities (14b) and (14d) must hold with equalities. In addition, an optimal solution set to (14) is also optimal solution set to (12). The detail of proof are presented in Appendix A. □

B. Optimal Solution based BRB Algorithm for Problem (14)

In this subsection, we aim at solving the problem (14) optimally using the BRB algorithm based on MO framework. This is possible due to two following important observations.

- The objective in (15) monotonically increases with respect to (w.r.t) each entry of $\mathbf{s}$. Particularly, we can observe that the objective increases if we keep increasing each of $\mathbf{a}$,$\mathbf{t}$ or $\boldsymbol{\tau}$ as long as it is still feasible to (14).
- For given $\boldsymbol{s}$, the resulted problem of (14) becomes a feasibility checking problem, which is recognized as mixed-integer second order cone program (MI-SOCP) feasibility problem and can be solved optimally by dedicated solvers such as MOSEK.

Herein we present the customized steps required for solving the considered problem (14). Specifically, we define the compact normal set $\mathcal{S}=\{\mathbf{s}\in\mathbb{R}_+^L|(14\text{b})-(14\text{f})\}$ as the feasible set of (14) and $\mathcal{U}=[\underline{\mathbf{s}};\overline{\mathbf{s}}]$ as the box that contains all $\mathbf{s}$ feasible to (14). The calculation of box $\mathcal{U}$ is presented detailed in Appendix B. The general idea to solve the problem (14) optimally

in a BRB algorithm using MO framework is to check if a given $\mathbf{s} \in \mathcal{U}$ belongs to $\mathcal{S}$ or not. Mathematically we need to solve the following feasibility problem for a given $\mathbf{s}$

$$\text{find} \quad \mathbf{b}, \mathbf{w}, \boldsymbol{\nu} \tag{18a}$$

$$\text{s.t.} \quad (14b) - (14f). \tag{18b}$$

It is easy to check that when $\boldsymbol{\tau}$ is fixed, (14b) can be reformulated as SOC constraint as

$$c\,\mathrm{Re}(\mathbf{h}_k^H \mathbf{w}_k) \geq \left\| \mathbf{h}_k^H \mathbf{w}_1, \ldots, \mathbf{h}_k^H \mathbf{w}_K, \sigma_0 \right\|_2 \tag{19}$$

where $c = \sqrt{\frac{1}{2^{\tau_k}-1} + 1}$. Similarly, for given $\mathbf{s}$ the constraint (14d) is SOC representable as

$$\begin{cases} \frac{1/t_0 - \hat{P}^{\text{tot}}(\mathbf{b})+1}{2} \geq \left\| \left[\frac{\mathbf{w}_{1,1}^T}{\sqrt{\eta_1}}, \ldots, \frac{\mathbf{w}_{I,K}^T}{\sqrt{\eta_I}}, \frac{1/t_0 - \hat{P}^{\text{tot}}(\mathbf{b})-1}{2} \right]^T \right\|_2 & \text{if X is GEE} \\ \frac{1/t_i - \hat{P}_i(\mathbf{b})+1}{2} \geq \left\| \left[\frac{\mathbf{w}_{1,1}^T}{\sqrt{\eta_1}}, \ldots, \frac{\mathbf{w}_{I,K}^T}{\sqrt{\eta_I}}, \frac{1/t_i - \hat{P}_i(\mathbf{b})-1}{2} \right]^T \right\|_2, \forall i \in \mathcal{I} & \text{otherwise} \end{cases} \tag{20}$$

where

$$\hat{P}^{\text{tot}}(\mathbf{b}) = \sum_{i \in \mathcal{I}} \rho_i \sum_{k \in \mathcal{K}} a_{i,k} \tau_k + \sum_{i \in \mathcal{I}} (b_i P_i^{\text{ra}} + \big(1 - b_i\big) P_i^{\text{ri}}) \tag{21}$$

$$\hat{P}_i(\mathbf{b}) = \rho_i \sum_{k \in \mathcal{K}} a_{i,k} \tau_k + b_i P_i^{\text{ra}} + \big(1 - b_i\big) P_i^{\text{ri}}, \forall i \in \mathcal{I} \tag{22}$$

Thus, the feasibility checking problem (18) is in fact a MI-SOCP problem. Based on the above analysis, problem (14) can now be expressed as $\max\{f_{\text{X}}(\mathbf{s}) | \mathbf{s} \in \mathcal{S} \subset \mathcal{U}\}$. First, we check whether $\underline{\mathbf{s}}$ is feasible or not. If so, we apply the proposed BRB method to find a globally optimal solution to (14). The proposed method recursively branches a box $\mathcal{U}$ into two smaller boxes, checks the feasibility of each new box, update 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. Note that, the variable of interest $\mathbf{s}$ comprises both binary variables $\mathbf{a}$ and continuous variables τ and $\mathbf{t}$, thus the branching and reduction procedures need to be adjusted to guarantee the exact solution of binary variables. The operations of BRB algorithm is presented as follows.

- Box branching: At each iteration, the box which currently contains the largest upper bound is selected to branch. This box is split into two smaller boxes by a standard bi-partition along the longest edge [32]. In particular, supposed that the box $\mathcal{B} = [\mathbf{x}, \mathbf{y}]$ is chosen to branch, then it is divided into two following smaller boxes

$$\mathcal{B}^{(1)} = \begin{cases} [\mathbf{x}, \mathbf{y} - \mathbf{e}_l] & \text{if } l \leq N \\ [\mathbf{x}, \mathbf{y} - \mathbf{e}_l (y_l - x_l)/2] & \text{if } l > N \end{cases} \text{ and } \mathcal{B}^{(2)} = \begin{cases} [\mathbf{x} + \mathbf{e}_l, \mathbf{y}] & \text{if } l \leq N \\ [\mathbf{x} + \mathbf{e}_l (y_l - x_l)/2, \mathbf{y}] & \text{if } l > N \end{cases} \tag{23}$$

where $l = \arg\max_{l\in\{0,\ldots,S\}} (y_l - x_l)$ is the index of the longest edge of $\mathcal{B}$ and $\mathbf{e}_l$ is a $S \times 1$ vector where all entries are zero except that the lth entry is 1. By using the above branching rules, the branched binary variables are guaranteed to lie in binary cutting plane.

- Box reduction: Supposed that we perform the reduction process on the box $\mathcal{B}^{(1)} = [\mathbf{p}, \mathbf{q}]$ to obtain the reduced box redu$(\mathcal{B}^{(1)})$ without any loss of optimality. Note that a similar argument can be applied to $\mathcal{B}^{(2)}$ as well. First, we check whether $\mathcal{B}^{(1)}$ contains at least one feasible solution to problem (14) or not by checking the feasibility of (18) for the given lower bound set $\mathbf{p}$. If (18) is infeasible, it means that $\mathbf{p} \in \mathcal{U} \setminus \mathcal{S}$ and $\mathcal{B}^{(1)}$ does not contain any feasible solution to (14) so that it can be discarded. Otherwise, the reduction process is performed to find redu$(\mathcal{B}^{(1)}) = [\mathbf{p}', \mathbf{q}']$. In particular, we calculate $\mathbf{p}' = \mathbf{q} - \sum_{l\in\{0,\ldots,S\}} \lambda_l(q_l - p_l)\mathbf{e}_l$, where $\lambda_l = \sup\{\lambda|\lambda \in [0,1], \mathbf{q} - \lambda(q_l - p_l)\mathbf{e}_l \in \mathcal{U}\setminus\mathcal{S}\}$, $\forall l \in \{0,\ldots,S\}$ and $\mathbf{q}' = \mathbf{p}' + \sum_{l=\{0,\ldots,S\}} \beta_l(q_l - p'_l)\mathbf{e}_l$, where $\beta_l = \sup\{\beta|\beta \in [0,1], \mathbf{p}' + \beta(q_l - p'_l)\mathbf{e}_l \in \mathcal{S}\}$, $\forall l \in \{0,\ldots,S\}$. The problem of finding λ_l and β_l can be solved easily using a bisection procedure over $\lambda \in [0,1]$ and $\beta \in [0,1]$, respectively. To guarantee p'_l , $q'_l \in \{0,1\}$, $\forall l \leq N$ while reducing, we do

$$p'_l = \begin{cases} 1 & \text{if } \mathbf{q} - \mathbf{e}_l \in \mathcal{U} \setminus \mathcal{S} \\ 0 & \text{otherwise} \end{cases} \text{ and } q'_l = \begin{cases} 1 & \text{if } \mathbf{p}' + \mathbf{e}_l \in \mathcal{S} \\ 0 & \text{otherwise} \end{cases}. \tag{24}$$

It is shown in [32] that if $\mathcal{B}^{(1)}$ contains an optimal solution, then redu$(\mathcal{B}^{(1)})$ also contains this optimal solution.

- Bounding and Pruning: due to the monotone objective function, the upper and lower bound of a box $\mathcal{B} = [\underline{\mathbf{s}}, \bar{\mathbf{s}}]$ is simply computed $U(\mathcal{B}) = f(\bar{\mathbf{s}})$ and $L(\mathcal{B}) = f(\underline{\mathbf{s}})$, respectively. After updating the current best lower bound ζ_n, the pruning is performed to delete the boxes whose upper bounds are smaller than ζ_n. According to [32], the proposed algorithm is bound improving and terminates after finitely many iterations for a given desired accuracy level ϵ. The BRB algorithm is summarized in Algorithm 1 to find a globally optimal solution to (14). We now explain how Algorithm 1 can be modified to deal with a tree topology C-RAN. First, as mentioned above, we need to impose the additional constraint $a_{i,k} \leq \prod_{\forall j\in\mathcal{A}_i} a_{j,k}$, $\forall i \in \mathcal{I}, \forall k \in \mathcal{K}$ to problem $(\mathcal{E}_{\mathrm{X}})$. Since $a_{i,k} \in \{0,1\}$, this constraint is equivalent to $a_{i,k} \leq \sqrt[|\mathcal{A}_i|]{\prod_{\forall j\in\mathcal{A}_i} a_{j,k}}$ where $|\mathcal{A}_i|$ is the cardinality of the set $\mathcal{A}_i$. Note that the constraint $a_{i,k} \leq \sqrt[|\mathcal{A}_i|]{\prod_{\forall j\in\mathcal{A}_i} a_{j,k}}$ is convex and can be represented as an SOC as the right hand side of the inequality is a concave function if $a_{j,k}$'s are relaxed to be continuous. Thus the resulting feasibility problem (18) can also solved by a

Algorithm 1 Proposed BRB algorithm.

Apply box reduction to $\mathcal{U}$ to obtain redu $(\mathcal{U})$

$n = 1$; $\mathcal{B}_1 =$ redu $(\mathcal{U})$; $\mathcal{D}_1 = \{\mathcal{B}_1\}$;$\zeta_1 = L(\mathcal{B}_1)$;

repeat

  Select the box with the largest upper bound to branch: $\mathcal{B}_n = \arg\max_{\mathcal{B}_i \subset \mathcal{D}_n} U(\mathcal{B}_i)$;

  Branch the box $\mathcal{B}_n$ into two small boxes $\mathcal{B}_n^{(1)}$ and $\mathcal{B}_n^{(2)}$; // *Box Branching* //

  for $j = 1:2$ **do**

    Compute lower bound set of $\mathcal{B}_n^{(j)}$, denoted as $\underline{\mathrm{X}}_n^{(j)} = \left\{\underline{\mathbf{s}}_n^{(j)}\right\}$;

    if $\underline{\mathrm{X}}_n^{(j)}$ is feasible **then**

      Apply box reduction to $\mathcal{B}_n^{(j)}$ to obtain redu $\left(\mathcal{B}_n^{(j)}\right)$; // *Box Reduction* //

      else $\underline{\mathrm{X}}_n^{(j)} = \emptyset$;

    end if

    Compute lower bound $L\left(\text{redu}\left(\mathcal{B}_n^{(j)}\right)\right)$, upper bound $U\left(\text{redu}\left(\mathcal{B}_n^{(j)}\right)\right)$ from the reduced box; // *Bound Computation* //

  end for

  Update the lower bound: $\zeta_{n+1} = \max\left(L\left(\text{redu}\left(\mathcal{B}_n^{(1)}\right)\right), L\left(\text{redu}\left(\mathcal{B}_n^{(2)}\right)\right), \zeta_n\right)$;

  Update the set of boxes: $\mathcal{D}_{n+1} = \left\{\mathcal{D}_n, \mathcal{B}_n^{(1)}, \mathcal{B}_n^{(2)}\right\}$;

  Delete the box that do not contain optimal solution: $\mathcal{D}_{n+1} = \mathcal{D}_n \setminus \{\mathcal{B}_i | \zeta_{n+1} > U(\text{redu}(\mathcal{B}_i)), \forall i = 1, \ldots, \text{cardinal}(\mathcal{D}_n)\}$; // *Pruning* //

  $n = n + 1$;

until $|\max_{\mathcal{B}_i \subset \mathcal{D}_n} U(\text{redu}(\mathcal{B}_i)) - \zeta_n| \leq \epsilon$;

MI-SOCP solver. As a result, Algorithm 1 can be adapted to solve the corresponding problems in a tree topology C-RAN.

To conclude this section we remark that it is also possible to further apply the branch and bound process to the binary variables as done in [13]. In this way, the need of a dedicated MI-SOCP solver to solve (18) is eliminated. In this paper, we still rely on a MI-SOCP solver in each iteration. The intention is that more advanced optimization techniques for MI-SOCP will be developed and integrated in commercial solvers in the future which make them powerful enough to solve MI-SOCP in reasonably time. This will make our proposed method more flexible as it is not associated with a specific method to deal with binary variables. The performance comparison with the method in [13] is left for future work.

V. Proposed SCA-based Low Complexity Algorithms

A. General SCA method

Before elaborating the SCA method to solve the considered EE maximization problems, we take an opportunity to present how SCA method can address a general non-convex problem. Let us consider the general non-convex optimization problem in the following

$$\max_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}) \tag{25a}$$

$$\text{s.t.} \quad g_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, L_1 \tag{25b}$$

$$p_j(\mathbf{x}) \leq q_j(\mathbf{x}),\ j = 0, \ldots, L_2 \tag{25c}$$

where $g_i(\mathbf{x}) : \mathbb{C}^n \to \mathbb{R}$,$i = 1, \ldots, L_1$, $p_j(\mathbf{x}) : \mathbb{C}^n \to \mathbb{R}$,$j = 1, \ldots, L_2$ and $q_j(\mathbf{x}) : \mathbb{C}^n \to \mathbb{R}$,$j = 1, \ldots, L_2$ are convex functions w.r.t $\mathbf{x} \in \mathbb{C}^N$, respectively. The idea to deal with the non-convex constraint (25c) is applying the SCA that approximate the non-convex part of (25c) by its convex one. In particularly, assuming $q_j(\mathbf{x})$ is differentiable (which is mostly true for the constraints in wireless communications), SCA linearizes $q_j(\mathbf{x})$ around the current iteration parameters $\mathbf{x}^{(n)}$ to arrive at the following constraint

$$p(\mathbf{x}) - q(\mathbf{x}^{(n)}) - \left\langle \nabla q(\mathbf{x}^{(n)}), \mathbf{x} - \mathbf{x}^{(n)} \right\rangle \leq 0 \tag{26}$$

Note that (26) implies (25c) as a concave function (i.e., $-q(\mathbf{x})$ as mentioned above) is upper bounded by its linearization. In other words, SCA arrives at an inner approximation of the feasible set of the non-convex program in (25), which is expressed by the following convex subproblem

$$\max_{\mathbf{x}\in\mathcal{X}} \{ f(\mathbf{x}) | \text{ (25b), (26)} \} \tag{27}$$

at the $n+1$th iteration and updates the operating point $\mathbf{x}^{(n)}$ until convergence. The SCA based algorithm to solve the problem (25) is presented in Algorithm 2. In the following subsection, we

Algorithm 2 SCA based Algorithm

1: Set $n := 0$ and initialize starting points of $\mathbf{x}^{(n)}$;
2: **repeat**
3: Solve the approximated convex problem (27) at $\mathbf{x}^{(n)}$ to achieve the optimal solution $\mathbf{x}^\star$;
4: Update $\mathbf{x}^{(n+1)} = \mathbf{x}^\star$;
5: Set $n := n + 1$;
6: **until** Convergence;

present how to apply the SCA based algorithm presented in this subsection V-A to three different formulated EE problems in (12). It is worth mentioning that it is not amendable to directly apply this presented methods to three different EE problems in (12) since it is not ready in the form of (25). Thus, it is necessary to propose the new transformations for each EE problem in (12) into more tractable form that the non-convex constraints are revealed, thus, can be handled by the light of SCA method.

B. SCA based Algorithm for GEE Maximization Problem

1) Equivalent transformations: In this section, we present SCA approach to provide a sub-optimal solution of (12) for GEE maximization problem. The proposed method is developed based on the transformations introduced in [33]. First, we consider the following equivalent formulation

$$\max_{\substack{\mathbf{b},\mathbf{a},\mathbf{w},\boldsymbol{\nu}\\ \mu,\boldsymbol{\gamma},\boldsymbol{\phi},\mathbf{z},\psi,\boldsymbol{\varphi}}} \quad \psi \tag{28a}$$

$$\text{s.t.} \quad \frac{\left(\sum_i \sum_{k\in\mathcal{K}} \varphi_{i,k}^2\right)}{\mu} \geq \psi \tag{28b}$$

$$a_{i,k}\phi_k \geq \varphi_{i,k}^2 \tag{28c}$$

$$\log\left(1+\gamma_k\right) \geq \phi_k \tag{28d}$$

$$\Gamma_k(\mathbf{w}) \geq \gamma_k \tag{28e}$$

$$\gamma_k \geq \Gamma_k^{\min} \tag{28f}$$

$$\tilde{P}^{\text{tot}}\big(\mathbf{w},\mathbf{a},\mathbf{b},\mathbf{z}\big) \leq \mu \tag{28g}$$

$$\sum_{k\in\mathcal{K}} \left(a_{i,k}/z_k\right) \leq C_i^{\max} \tag{28h}$$

$$R_k(\mathbf{w}) \leq 1/z_k \tag{28i}$$

$$\text{(12c), (12e), (12f), (12h), (13).} \tag{28j}$$

where we denote $\tilde{P}^{\text{tot}}\big(\mathbf{w},\mathbf{a},\mathbf{b},\mathbf{z}\big) = \sum_i P_i^{\text{rrh}}(\mathbf{w},b_i) + \sum_{i\in\mathcal{I}} \rho_i \sum_{k\in\mathcal{K}} \left(a_{i,k}/z_k\right)$ and introduce the new slack variables $\psi \geq 0$, $\mu \geq 0$, $\boldsymbol{\varphi} = \{\varphi_{i,k} \geq 0\}_{\forall i\in\mathcal{I},\forall k\in\mathcal{K}}$, $\boldsymbol{\phi} = \{\phi_k \geq 0\}_{\forall k\in\mathcal{K}}$, $\boldsymbol{\gamma} = \{\gamma_k \geq 0\}_{\forall k\in\mathcal{K}}$ and $\mathbf{z} = \{z_k \geq 0\}_{\forall k\in\mathcal{K}}$. Note that the meaning of $\varphi_{i,k}^2$ is considered as the transmission rate of user k transported through the fronthaul link i. It is easy to see that a feasible solution of (28) is also feasible to (12). In addition, with a feasible solution of (12), we

can also find a solution that is feasible to (28). Moreover, the constraints in (28b) and (28g) are active at optimality. Thus, (28) is equivalent to (12).

It is recognized that the newly additional constraint (28c) is SOC representable as $\frac{\phi_k+a_{i,k}}{2} \geq \left\|\varphi_{i,k}, \frac{\phi_k-a_{i,k}}{2}\right\|_2$. Furthermore, a subtle observation $a_{i,k} = a_{i,k}^2$ for $a_{i,k} \in \{0,1\}$ is used to replace $a_{i,k}$ by $a_{i,k}^2$ and rewrite the non-convex constraint (28h) into two equivalent inequalities as

$$a_{i,k}^2 \leq \theta_{i,k} z_k, \tag{29}$$

$$\sum_{k\in\mathcal{K}} \theta_{i,k} \leq C_i^{\max}, \tag{30}$$

with newly introduced variables $\boldsymbol{\theta} = \{\theta_{i,k} \geq 0\}_{\forall i\in\mathcal{I},k\in\mathcal{K}}$. By taking advantage of the transformation in (29), we can rewrite (28g) as

$$\sum_i P_i^{\text{rrh}}(\mathbf{w}, b_i) + \sum_{i\in\mathcal{I}} \rho_i \sum_{k\in\mathcal{K}} \theta_{i,k} \leq \mu \tag{31}$$

Finally, to make (28i) amenable to SCA, we can equivalently rewrite it as

$$1 + \xi_k \leq \exp(1/z_k) \tag{32}$$

$$\frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\xi_k} \leq \sum_{j\neq k}^{K} |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma_0^2, \tag{33}$$

where $\xi = \{\xi_k \geq 0, \forall k \in \mathcal{K}\}$ is the set of newly introduced variables. We note that the constraints (32) and (33) are still non-convex, but both sides of the inequalities in (32) and (33) are convex. Thus, in light of the SCA method, we can immediately linearize the right hand side of (32) and (33) to obtain their approximate convex constraints. By denoting the set of variables as $\Psi_{\text{GEE}} = \{\mathbf{b}, \mathbf{a}, \mathbf{w}, \boldsymbol{\nu}, \boldsymbol{\theta}, \mu, \boldsymbol{\gamma}, \boldsymbol{\phi}, \mathbf{z}, \psi, \boldsymbol{\xi}, \boldsymbol{\varphi}\}$, the final equivalent problem after being transformed is presented as

$$\max_{\Psi_{\text{GEE}}} \ \psi \tag{34a}$$

$$\text{s.t.} \ \ \psi \leq \frac{\left(\sum_i \sum_{k\in\mathcal{K}} \varphi_{i,k}^2\right)}{\mu} \tag{34b}$$

$$a_{i,k}\phi_k \geq \varphi_{i,k}^2 \tag{34c}$$

$$\log(1+\gamma_k) \geq \phi_k \tag{34d}$$

$$\sum_{j\in\mathcal{K}\setminus k} |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma_0^2 \leq \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\gamma_k} \tag{34e}$$

$$\sum_i P_i^{\text{rrh}}(\mathbf{w}, b_i) + \sum_{i\in\mathcal{I}} \rho_i \sum_{k\in\mathcal{K}} \theta_{i,k} \leq \mu \tag{34f}$$

$$1+\xi_k \le \exp(1/z_k) \tag{34g}$$

$$\frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\xi_k} \le \left(\sum_{j\neq k}^{K} |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma_0^2\right) \tag{34h}$$

$$(12c), (12e), (12f), (12h), (13), (28f), (29), (30). \tag{34i}$$

where the constraint (34e) is rewritten from the constraint (28e). We remark that problem (34) is still non-convex but its non-convex constraints in (34b), (34e), (34g) and (34h) are in the form of (25c) and ready to be handled in light of the SCA method as the follows.

2) SCA-GEE Algorithm: In this method we preserve the Boolean variables, and only approximate the continuous non-convex parts of (34). In particular, we do so by applying the framework of SCA. Let us first consider the non-convex constraint (34b) and (34e). These functions of $\frac{\left(\sum_i \sum_{k\in\mathcal{K}} \varphi_{i,k}^2\right)}{\mu}$ and $\frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\gamma_k}$ have the same form of function $h(p,q) = \frac{|p|^2}{q}$, $\forall p \in \mathbb{C}$,$q \in \mathbb{R}_+$, i.e., in (34b) it is $h(\boldsymbol{\varphi}, \mu)$ and in (34e), it is $h(\mathbf{w}_k, \gamma_k)$. Thus, at iteration $n+1$ th the proposed algorithm we apply the first order Taylor approximation to $h(p,q)$ around the point of $p^{(n)}$, $q^{(n)}$ by

$$H\big(p, q; p^{(n)}, q^{(n)}\big) = \frac{2\,\mathrm{Re}\big(p^{(n)} p\big)}{q^{(n)}} - \frac{\big|p^{(n)}\big|^2}{q^{(n)2}} q \tag{35}$$

where we have denoted $q^{(n)2} = \big(q^{(n)}\big)^2$ to lighten the notation. Particularly, $H(\boldsymbol{\varphi}, \mu; \boldsymbol{\varphi}^{(n)}, \mu^{(n)})$ is concave upper bound function of $h(\boldsymbol{\varphi}, \mu)$ around the point of $\boldsymbol{\varphi}^{(n)}$ and $\mu^{(n)}$ given by

$$H(\boldsymbol{\varphi}, \mu; \boldsymbol{\varphi}^{(n)}, \mu^{(n)}) = \frac{\sum_i \sum_{k\in\mathcal{K}} 2\varphi_{i,k}^{(n)} \varphi_{i,k}}{\mu^{(n)}} - \frac{\sum_i \sum_{k\in\mathcal{K}} \varphi_{i,k}^{(n)2}}{\mu^{(n)2}} \mu$$

In the same way, we have $H(\mathbf{w}_k, \gamma_k; \mathbf{w}_k^{(n)}, \gamma_k^{(n)})$ is simply a linearization of function $h(\mathbf{w}_k, \gamma_k) = \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\gamma_k}$ around the point of $\mathbf{w}_k^{(n)}$ and $\gamma_k^{(n)}$ given by

$$H(\mathbf{w}_k, \gamma_k; \mathbf{w}_k^{(n)}, \gamma_k^{(n)}) = \frac{2\,\mathrm{Re}\big(\mathbf{w}_k^{(n)H} \mathbf{H}_k \mathbf{w}_k\big)}{\gamma_k^{(n)}} - \frac{\big|\mathbf{h}_k^H \mathbf{w}_k^{(n)}\big|^2}{\gamma_k^{(n)2}} \gamma_k \tag{36}$$

Here, we have denoted $\mathbf{H}_k \triangleq \mathbf{h}_k \mathbf{h}_k^H$, $\mathbf{w}_k^{(n)H} = \big(\mathbf{w}_k^{(n)}\big)^H$ to lighten the notation. The non-convex constraints (34g) and (34h) can also be approximated by its concave upper bound as

$$1 + \xi_k - F\big(z_k; z_k^{(n)}\big) \le 0 \tag{37}$$

$$\frac{\big|\mathbf{h}_k^H \mathbf{w}_k\big|^2}{\xi_k} - G\big(\mathbf{w}; \mathbf{w}^{(n)}\big) \le 0 \tag{38}$$

where $F(z_k; z_k^{(n)})$ given in (39) is lower bound concave approximation of $f(z_k) = \exp(1/z_k)$ around the point of $z_k^{(n)}$ and $G\big(\mathbf{w}; \mathbf{w}^{(n)}\big)$ given in (40) is lower bound concave approximation of $g(\mathbf{w}) = \sum_{j\neq k}^{K} |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma_0^2$ around the point of $\mathbf{w}^{(n)}$.

$$F\big(z_k; z_k^{(n)}\big) = e^{(1/z_k^{(n)})} - \frac{e^{(1/z_k^{(n)})}}{z_k^{(n)2}}\big(z_k - z_k^{(n)}\big) \tag{39}$$

$$G\big(\mathbf{w}; \mathbf{w}^{(n)}\big) = \sum_{j\neq k}^{K} 2\,\mathrm{Re}\big(\mathbf{w}_j^{(n)H}\mathbf{H}_k\mathbf{w}_j\big) - \sum_{j\neq k}^{K} \mathbf{w}_j^{(n)H}\mathbf{H}_k\mathbf{w}_j^{(n)} + \sigma_0^2 \tag{40}$$

By applying these approximations we can obtain a mixed integer convex approximation of problem (34) at iteration $n+1$ of the algorithm as

$$\max_{\Psi_{\text{GEE}}} \ \psi \tag{41a}$$

$$\text{s.t.} \ \ \psi - H(\boldsymbol{\varphi}, \mu; \boldsymbol{\varphi}^{(n)}, \mu^{(n)}) \leq 0 \tag{41b}$$

$$\frac{a_{i,k} + \phi_k}{2} \geq \left\| \varphi_{i,k}, \frac{a_{i,k} - \phi_k}{2} \right\|_2 \tag{41c}$$

$$\log\left(1 + \gamma_k\right) \geq \phi_k \tag{41d}$$

$$\sum_{j\in\mathcal{K}\setminus k} |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma_0^2 - H\big(\mathbf{w}_k, \gamma_k; \mathbf{w}_k^{(n)}, \gamma_k^{(n)}\big) \leq 0 \tag{41e}$$

$$\sum_i P_i^{\text{rrh}}(\mathbf{w}, b_i) + \sum_{i\in\mathcal{I}} \rho_i \sum_{k\in\mathcal{K}} \theta_{i,k} \leq \mu \tag{41f}$$

$$1 + \xi_k - F\big(z_k; z_k^{(n)}\big) \leq 0 \tag{41g}$$

$$\left|\mathbf{h}_k^H \mathbf{w}_k\right|^2 / \xi_k - G\big(\mathbf{w}; \mathbf{w}^{(n)}\big) \leq 0 \tag{41h}$$

$$\text{(12c), (12e), (12f), (12h), (13), (28f), (29), (30)} \tag{41i}$$

where $\mathbf{w}^{(n)}, \mathbf{z}^{(n)}, \boldsymbol{\gamma}^{(n)}, \mu^{(n)}, \boldsymbol{\varphi}^{(n)}$ are the parameters to be updated at the $(n+1)$th iteration.

Remark 2. Note that all the continuous constraints in (41), except (41d), are convex quadratic representable. However, due to the presence of the exponential cone in (41d), (41) is still recognized as a generic convex mixed-integer program for which dedicated solvers are quite limited. To avail of powerful modern MI-SOCP solvers such as MOSEK or GUROBI, our idea is to approximate (41d) by a conic constraint and this should be done in light of SCA. More specifically, a conic lower bound of the left hand side of (41d) is desired.

The key is due to the following inequality. For any $\gamma_k \geq 0$ it holds that

$$\log(1+\gamma_k) \geq U\big(\gamma_k;\gamma_k^{(n)}\big) = \log\big(1+\gamma_k^{(n)}\big) + \frac{1}{1+\gamma_k^{(n)}}\big(\gamma_k-\gamma_k^{(n)}\big) - \frac{1}{2}\big(\gamma_k-\gamma_k^{(n)}\big)^2. \quad (42)$$

In fact $U\big(\gamma_k;\gamma_k^{(n)}\big)$ is a quadratic lower bound of $\log(1+\gamma_k)$ around $\gamma_k^{(n)}$, which is derived from the Lipschitz continuity of the derivative of $\log(1+\gamma_k)$. The proof is given in Appendix C. To obtain an MI-SOCP formulation of (41), we replace (41d) by

$$U(\gamma_k;\gamma_k^{(n)})) \geq \phi_k \quad (43)$$

which is SOC representable. Thus, the problem (41) becomes MI-SOCP program, which can be solved by SCA based algorithm outlined in Algorithm 2.

C. SCA-based Algorithm for WSEE Maximization Problem

In this section we present the method to apply SCA to solve the problem WSEE maximization problem in (12). By introducing the similar slack variables $\boldsymbol{\psi} = \{\psi_i \geq 0\}_{\forall i\in\mathcal{I}}$, $\boldsymbol{\mu} = \{\mu_i \geq 0\}_{\forall i\in\mathcal{I}}$, $\boldsymbol{\varphi} = \{\varphi_{i,k} \geq 0\}_{\forall i\in\mathcal{I},\forall k\in\mathcal{K}}$, $\boldsymbol{\phi} = \{\phi_k \geq 0\}_{\forall k\in\mathcal{K}}$, $\boldsymbol{\gamma} = \{\gamma_k \geq 0\}_{\forall k\in\mathcal{K}}$, and $\mathbf{z} = \{z_k \geq 0\}_{\forall k\in\mathcal{K}}$, as presented in the GEE maximization in the section V-B1, we are able to equivalently transform the maximization problem $(\mathcal{E}_{\text{WSEE}})$ as the following

$$\max_{\substack{\mathbf{b},\mathbf{a},\mathbf{w},\boldsymbol{\nu}\\ \mu,\boldsymbol{\gamma},\boldsymbol{\phi},\mathbf{z},\psi,\boldsymbol{\varphi}}} \quad \sum_i \lambda_i\psi_i \quad (44a)$$

$$\text{s.t.} \quad \psi_i \leq \frac{\sum_k \varphi_{i,k}^2}{\mu_i} \quad (44b)$$

$$\varphi_{i,k}^2 \leq a_{i,k}\phi_k \quad (44c)$$

$$\log\left(1+\gamma_k\right) \geq \phi_k \quad (44d)$$

$$\Gamma_k(\mathbf{w}) \geq \gamma_k \quad (44e)$$

$$\gamma_k \geq \Gamma_k^{\min} \quad (44f)$$

$$P_i^{\text{rrh}}(\mathbf{w},b_i) + \rho_i \sum_{k\in\mathcal{K}} a_{i,k}/z_k \leq \mu_i \quad (44g)$$

$$\sum_{k\in\mathcal{K}} (a_{i,k}/z_k) \leq C_i^{\max} \quad (44h)$$

$$R_k(\mathbf{w}) \leq 1/z_k \quad (44i)$$

$$(12c), (12e), (12f), (12h), (13). \quad (44j)$$

It is noted that μ_i is the variable representing the upper bound for the ith RRH and its fronthaul power consumption. We have an observation that the problem (44) shares the similar form with

the problem (34). For instance, it is also interesting to show that the right hand side of constraint (44b) which is function $h(\varphi_i, \mu_i) = \frac{\sum_k \varphi_{i,k}^2}{\mu_i}$, has the same form of $h(p, q)$ defined in the previous section. Thus, non-convex constraint (44b) can be approximated into a convex one by applying the linearization to the $h(\varphi_i, \mu_i)$ as

$$H(\boldsymbol{\varphi}_i, \mu_i; \boldsymbol{\varphi}_i^{(n)}, \mu_i^{(n)}) = \frac{\sum_{k\in\mathcal{K}} 2\varphi_{i,k}^{(n)}\varphi_{i,k}}{\mu_i^{(n)}} - \frac{\sum_{k\in\mathcal{K}} \varphi_{i,k}^{(n)2}}{\mu_i^{(n)2}}\mu_i \tag{45}$$

where we denote $\boldsymbol{\varphi}_i = \{\varphi_{i,k}\}_{\forall k\in\mathcal{K}}$. Similarly, we apply the same transformations and convex approximations in the previous section to the constraints (44d), (44e), (44h) and (44i). Thus, the final equivalent MI-SOCP problem of (44) at the $n+1$th iteration is presented as

$$\max_{\Psi_{\text{WSEE}}} \quad \sum_i \lambda_i \psi_i \tag{46a}$$

$$\text{s.t.} \quad \psi_i - H(\varphi_i, \mu_i; \varphi_i^{(n)}, \mu_i^{(n)}) \leq 0, \tag{46b}$$

$$P_i^{\text{rrh}}(\mathbf{w}, b_i) + \rho_i \sum_{k\in\mathcal{K}} \theta_{i,k} \leq \mu_i \tag{46c}$$

$$(12\text{c}), (12\text{e}), (12\text{f}), (12\text{h}), (13), (28\text{f}), (29), (30) \tag{46d}$$

$$(41\text{c}), (41\text{e}), (41\text{g}), (41\text{h}), (43). \tag{46e}$$

where $\mathbf{w}^{(n)}, \mathbf{z}^{(n)}, \boldsymbol{\gamma}^{(n)}, \varphi^{(n)}, \mu^{(n)}$ are the parameters that are updated at the $(n+1)$th iteration and $\Psi_{\text{WSEE}} = \{\mathbf{b}, \mathbf{a}, \mathbf{w}, \boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \mathbf{z}, \boldsymbol{\psi}, \boldsymbol{\xi}, \boldsymbol{\varphi}\}$. The proposed iterative approach to solve problem (12) for WSEE maximization problem is given in Algorithm 2, where the MI-SOCP convex problem (46) is applied in the step 3.

D. SCA-based Algorithm for EEF Maximization Problem

Inspired by the SCA method for solving the problems of GEE and WSEE maximization, here we provide the main steps to tackle the EEF maximization problem. First, with the similar set of newly introduced slack variables $\psi \geq 0$, $\boldsymbol{\mu} = \{\mu_i \geq 0, \forall i \in \mathcal{I}\}$, $\boldsymbol{\varphi} = \{\varphi_{i,k} \geq 0, \forall i \in \mathcal{I}, \forall k \in \mathcal{K}\}$,$\boldsymbol{\phi} = \{\phi_i \geq 0, \forall i \in \mathcal{I}\}$, $\boldsymbol{\gamma} = \{\gamma_k \geq 0, \forall k \in \mathcal{K}\}$ and $\mathbf{z} = \{z_k \geq 0, \forall k \in \mathcal{K}\}$, the EEF maximization problem in (12) can be equivalently transformed into the following problem

$$\max_{\substack{\mathbf{b},\mathbf{a},\mathbf{w},\boldsymbol{\nu} \\ \boldsymbol{\mu},\boldsymbol{\gamma},\boldsymbol{\phi},\mathbf{z},\psi,\boldsymbol{\varphi}}} \quad \psi \tag{47a}$$

$$\text{s.t.} \quad \psi \leq \frac{\sum_{k\in\mathcal{K}} \varphi_{i,k}^2}{\mu_i} \tag{47b}$$

$$\varphi_{i,k}^2 \leq a_{i,k}\phi_k \tag{47c}$$

$$P_i^{\text{rrh}}(\mathbf{w},b_i) + \rho_i \sum_{k\in\mathcal{K}} a_{i,k}/z_k \leq \mu_i \tag{47d}$$

$$(12c), (12e), (12f), (12h), (13). \tag{47e}$$

$$(44d), (44e), (44f), (44h), (44i). \tag{47f}$$

Note that, ψ is considered to equivalently replace for the minimum EE among all RRHs. The formulation of problem (47) now is similar to the form of problem (44). Thus, the same transformations and approximations can be applied to (47). As a result, the problem (47) can be approximated to a MI-SOCP problem at the $n+1$th iteration as the following

$$\max_{\Psi_{\text{FEE}}} \left\{ \psi \middle| \psi - H(\boldsymbol{\varphi}_i, \mu_i; \boldsymbol{\varphi}_i^{(n)}, \mu_i^{(n)}) \leq 0, (46c), (46d), (46e) \right\} \tag{48}$$

where $\mathbf{w}^{(n)}, \mathbf{z}^{(n)}, \boldsymbol{\gamma}^{(n)}, \varphi^{(n)}, \mu^{(n)}$ are the parameters that are updated at the $(n+1)$th iteration and $\Psi_{\text{FEE}} = \{\mathbf{b}, \mathbf{a}, \mathbf{w}, \boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \mathbf{z}, \boldsymbol{\psi}, \boldsymbol{\xi}, \boldsymbol{\varphi}\}$ is the set of variables. The proposed iterative approach to solve problem (12) for EEF maximization problem is given in Algorithm 2 where the problem (48) is applied to the step 3.

E. Relaxed based Algorithm for GEE, WSEE, and EEF Maximization Problems

To develop an algorithm with polynomial time, we further consider the continuous relaxation of binary variables, i.e., $0 \leq b_i \leq 1, 0 \leq a_{i,k} \leq 1$ for $\forall i \in \mathcal{I}, \forall k \in \mathcal{K}$. As a result, the continuous relaxation of (41), (46), and (48) denoted as $(\mathcal{E}_{\text{X}}^r)$, becomes an SOCP which can be solved in polynomial time, with X representing for GEE, WSEE, and EEF, respectively. The relaxed based algorithm generally combines two stages: (i) continuous relaxation and (ii) post-processing. In the first stage, we follow Algorithm 2, but simply solve $(\mathcal{E}_{\text{X}}^r)$ in Step 3. The post-processing process is then used to map the obtained b_i's and $a_{i,k}$'s to the binary values, which is required due to the continuous relaxation. In particular, we rely on the solution to the continuous relaxation at convergence as an incentive measure to make a decision on the binary value of $\mathbf{a}$ and $\mathbf{b}$. Let us denote $\tilde{\mathbf{a}}, \tilde{\mathbf{b}}$ and $\tilde{\mathbf{w}}$ as the solution achieved after the first stage. Intuitively, the connection between the ith RRH and the kth UE is more likely if the channel of the link is in better condition and the power consumed to transmit fronthaul data $P_i^{\text{FH}}(\mathbf{w}, \mathbf{a}_i)$ is smaller than the others. Consequently, solving the continuous relaxation would possibly yield higher $\widetilde{b}_i$ for the ith RRH and higher $\widetilde{a}_{i,k}$ for the connection between the ith RRH and the kth UE. Based on the

above intuitive observations, we propose an iterative procedure to determine the set of active RRHs and RRH-UE association based on $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$. The process starts by assuming that all the RRHs are off and there is no association between RRH and UE. In each iteration, $(\mathcal{E}_{\mathrm{X}}^{r})$ is solved given a set of active RRHs and RRH-UE association that is connected. The RRH-UE association with the largest $\tilde{a}_{i,k}$ will be made connected and the resulting RRH will be set active, following the relationship in (12f). The overall algorithm is presented in Algorithm 3. We remark that

Algorithm 3 Relaxed algorithm

1: Set $m := 0$, $\pi^{(m)}$ is significantly small, and initialize the set $\mathcal{R}_{\text{off}}^{(m)} = \{(i,k) \times i \in (\mathcal{I}, \mathcal{K}) \times \mathcal{I}\}$.
2: **repeat**
3: Set $m := m + 1$;
4: Algorithm 2 is used to solve $(\mathcal{E}_{\mathrm{X}}^{r})$ given $a_{i',k'} = 1$ and $b_{i'} = 1, \forall \{(i',k') \times i'\} \notin \mathcal{R}_{\text{off}}^{(m-1)}$;
5: Update $\mathcal{R}_{\text{off}}^{(m)} = \mathcal{R}_{\text{off}}^{(m-1)} \setminus \left\{(i',k') \times i' = \arg\max_{i,k \in \mathcal{R}_{\text{off}}^{(m-1)}} \widetilde{a}_{i,k}\right\}$;
6: Algorithm 2 is used to solve $(\mathcal{E}_{\mathrm{X}}^{r})$ given $a_{i',k'} = 1$,$b_{i'} = 1, \forall \{(i',k') \times i'\} \notin \mathcal{R}_{\text{off}}^{(m)}$ and $a_{i,k} = 0, b_i = 0, \forall \{(i,k) \times i\} \in \mathcal{R}_{\text{off}}^{(m)}$. If it is feasible, set $\pi^{(m)}$as the value of objective function achieved at the convergence. If not, set $\pi^{(m)} = \pi^{(0)}$.
7: **until** $(\mathcal{E}_{\mathrm{X}}^{r})$ starts to be infeasible or it is feasible but $\pi^{(m)} < \pi^{(m-1)}$;
8: Algorithm 2 is used to solve $(\mathcal{E}_{\mathrm{X}}^{r})$ given $a_{i',k'} = 1$,$b_{i'} = 1, \forall \{(i',k') \times i'\} \notin \mathcal{R}_{\text{off}}^{(m-1)}$ and $a_{i,k} = 0, b_i = 0, \forall \{(i,k) \times i\} \in \mathcal{R}_{\text{off}}^{(m-1)}$ to obtain final solution $\Psi_{\mathrm{X}}^{\star}$;

there also exist other approaches to find feasible binary solutions from a continuous relaxation. For example, in the continuous relaxation based branch and bound method, the tree search is performed over the space of the variables $a_{i,k}$ and b_i, $\forall i,k$. These are successfully fixed binary values of some unfixed $a_{i,k}$ or b_i at the corresponding nodes of the tree. The tree search is terminated when all relaxed binary variables indeed become binary. However, this method is only attractive if the number of binary variables is relatively small and it generally requires higher computation complexity compared to our proposed method in Algorithm 3. There is also another way that expresses binary variables by a set of continuous constraints, and then uses continuous optimization techniques. Such a method is also applicable to our considered problems and is left as future work.

F. Convergence and Complexity Analysis

The convergence of SCA-based algorithms has been well studied [28]. In particular, due to the use of convex approximations, the optimal solution obtained at iteration n is feasible to the convex

TABLE I

SIMULATION PARAMETERS

Description	Notation	Value	Description	Notation	Value
Number of antennas per RRH	M_i	2	Active power for RRH	P_i^{ra}	38.5 dBW
Power amplifier efficiency	η_i	0.35	Sleep power for RRH	P_i^{ri}	36.5 dBW
Maximum transmit power	$P^{\max}$	10 dBW	Fronthaul power factor	$\rho_i, \forall i$	1
Maximum fronthaul capacity	$C_i^{\max} = C, \forall i$	20 b/s/Hz	Weight vector	λ	$[\frac{1}{I}, \ldots, \frac{1}{I}]$

problems (41), (46), and (48) at iteration $n+1$ for GEE, WSEE, and EEF maximization problems, respectively. This results in a non-decreasing sequences of objectives. Since the objective is upper bounded due to the power budget, the iterative algorithms are probably convergent. We now discuss the complexity of the proposed algorithms in this section. For Algorithm 2, the overall complexity mainly depends on that of solving the MI-SOCP problem in (41), (46), and (48), respectively, which is indeed a combinatorial optimization problem. In particular, there are IK binary variables $a_{i.k}$'s and I binary variables b_i's, resulting in 2^{IK+I} combinations for all the binary variables. Thus, the worst-case complexity of Algorithm 2 in each iteration can be written as $\mathcal{O}\left(2^{IK+K}(K^4M^3I)\right)$. For Algorithm 3, first we remark that in the worst case, Algorithm 3 must iteratively solve and update the resulting parameters for the SOCP problem $(\mathcal{E}_{\text{X}}^r)$ for $(I-1)K$ times, resulting the overall complexity of $\mathcal{O}\left(2(I-1)K\left(K^4M^3I\right)\right)$. In Section VI, Algorithm 3 is shown to yield a performance very close to that of SCA based algorithms but with polynomial computation time.

VI. NUMERICAL RESULTS

In this section, the extensive numerical results are presented to evaluate the performance of the proposed algorithms. For most numerical experiments, we use the simulation parameters listed in Table I [18], [31]. For the spatial model, we assume a network consisting of I RRHs that are uniformly located around the considered coverage and K UEs are randomly scattered across the considered network coverage. Moreover, we assume shadowing channel and the path-loss component is calculated as $(d_{ik}/d_0)^{-3}$where d_{ik} is the distance between the ith RRH and the kth user and $d_0 = 100$ m is the reference distance. In our simulations, Algorithm 2 is terminated when the increase in the objective between two consecutive iterations is less than 10^{-5}.

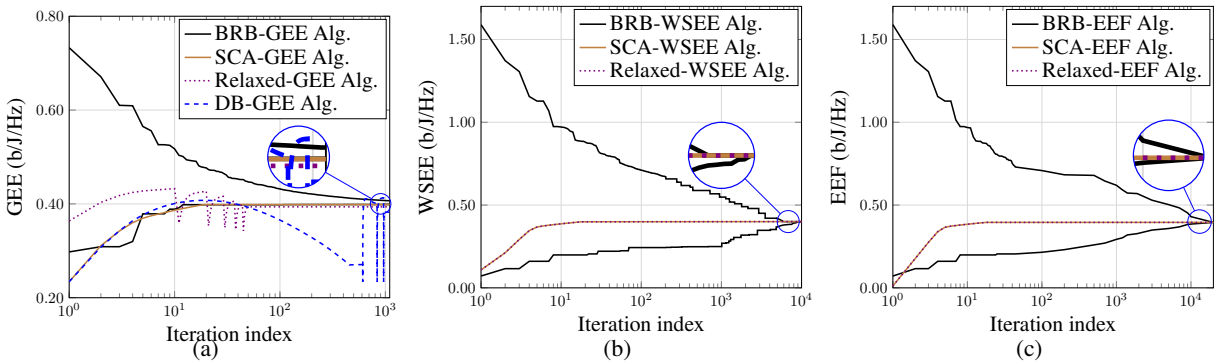

Fig. 2. (a)-(b)-(c): Convergent behavior of different algorithms for GEE, WSEE and EEF maximization problems.

A. Convergence and achieved EE performance

Fig. 2(a), 2(b), and 2(c) show the convergence of lower and upper bound of BRB algorithms and proposed low-complexity algorithms for GEE, WSEE and EEF maximization problems, respectively. In general for all cases of GEE, WSEE and EEF maximization problems, it can be seen that the lower and upper bound of BRB algorithm require more than 10^4 iterations to converge at the optimal solution while the SCA based algorithms need few iterations to converge to the objective value that is very close to the optimal value returned by the optimal BRB algorithm. In Fig. 2(a), the convergence of each SOCP $\mathcal{E}^r_{\text{GEE}}$ during the relaxed processes is plotted for the relaxed-GEE algorithm, which illustrates the uphill and downhill effect in the figure. For example, at some first iterations the GEE objective achieved in the relaxed-GEE algorithm is higher than in BRB and SCA based algorithms because of the relaxation continuous of binary variables and as expected, it eventually converges to the near-optimal objective value after all relaxed binary variables are mapped into feasible binary values. In addition, to demonstrate the performance gains of our proposed GEE algorithms, we compare our GEE algorithms with the Dinkelbach method, named DB-GEE algorithm. Especially, in DB-GEE algorithm, Dinkelbach approach is used to transform the fractional GEE objective function in (8) into the subtraction form associated with a fixed parameter, then the SCA method is applied to solve this subtraction formulation problem and the parameter in the objective can be updated until convergence. It is shown that DB-GEE algorithm requires more iterations to stabilize. In addition, in Fig. 2(b) and 2(c), the performance of relaxed-WSEE and relaxed-EEF algorithms are exactly same as that of SCA-WSEE and SCA-EEF algorithms, respectively. This can be explained as solving the relaxed problem $\mathcal{E}^r_{\text{WSEE}}$ and $\mathcal{E}^r_{\text{FEE}}$ at the first stage of corresponding relaxed algorithms result the

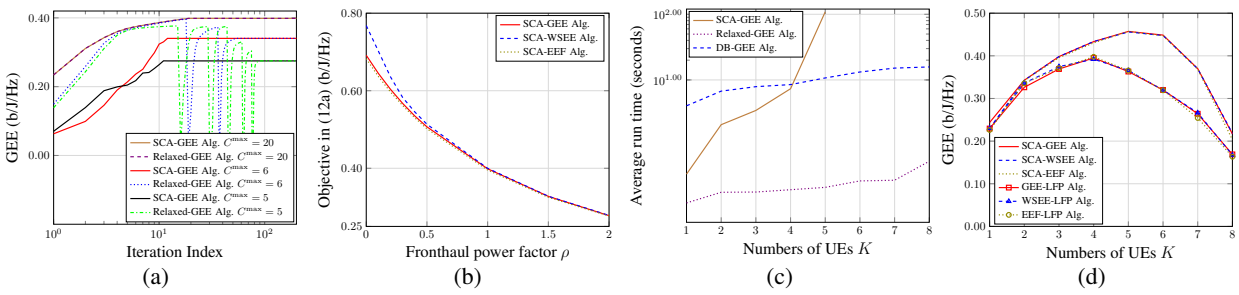

Fig. 3. (a): Convergence rate of our proposed algorithms for GEE against different maximal fronthaul capacity $C^{\max}$, (b): the performance of our proposed algorithms by varying the fronthaul power factor $\rho = \rho_i, \forall i \in \mathcal{I}$, (c)-(d): Average run time and GEE performance of different algorithms by varying the number of UEs K, respectively.

binary values of all relaxed variables. Thus, Fig. 2(a), 2(b) and 2(c) prove the effectiveness of our proposed algorithms in both terms of convergence and achieved EE performance.

In Fig. 3(a) we show the impact of the fronthaul capacity constraint on the convergence rate of our proposed GEE maximization algorithms. As can be seen, when $C^{\max}$ becomes smaller, the SCA-GEE algorithm tends to take a fewer number of iterations to converge, while the relaxed-GEE algorithm needs more iterations during post-processing process to produce a feasible solution. It can be explained that for small $C^{\max}$, the number of users served by a RRH becomes smaller. Thus, the relaxed-GEE algorithm has to carry out the post-processing process several times to refine the RRH-UE association to find a feasible solution. Fig. 3(b) shows the performance of three SCA-based algorithms by varying the fronthaul power factor ρ. It is clear that GEE, WSEE and EEF monotonically decrease when ρ increases. This means that fronthaul power has significant impact on the EE performance in C-RANs.

Fig. 3(c) and 3(d) demonstrate the average run time and the GEE performance of different proposed algorithms versus the number of UEs K. In Fig. 3(c), it is observed that the average run time increases with K which is not surprising according to the complexity analysis in Section V-F. Noticeably, by solving relaxed problem $\mathcal{E}^r_{\text{GEE}}$ which is SOCP, the computation time of relaxed-GEE algorithm is much lower than that of the SCA-GEE and DB-GEE algorithms. On the other hand, the SCA-GEE algorithm requires fewer iterations to converge but its overall run time is very high when K becomes large since the problem in each iteration is a MI-SOCP. In Fig. 3(d), GEE performance first increases and then decreases when K increases for all algorithms of comparison. The fact is that when the number of UE number increases, the total achievable rate first increases due to the multiuser diversity gain, which leads to an increase

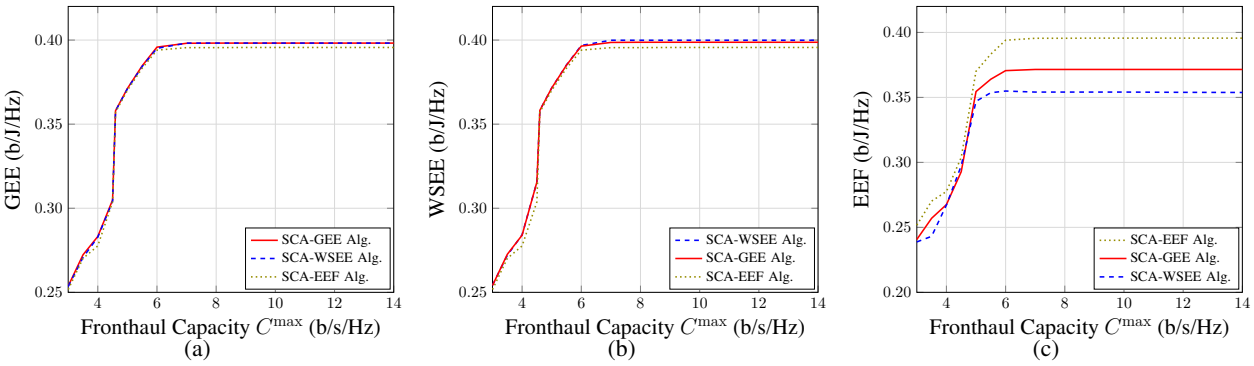

Fig. 4. (a): GEE objective value in (8); (b) WSEE objective value in (10); (c) EEF objective value in (11): calculated from the solutions obtained by applying the SCA-GEE, SCA-WSEE and SCA-EEF algorithms versus $C^{\max}$.

in the GEE. However, when K becomes sufficiently large, more power consumed at all RRHs and the corresponding fronthaul to serve and coordinate the induced interference subsequently decreases the achieved GEE.

Fig. 4(a), 4(b), and 4(c) evaluate the performance of SCA-based algorithms with respect to achieved GEE, WSEE and EEF metrics in (8), (10) and (11), respectively, versus the maximum fronthaul capacity $C_i^{\max} = C^{\max}$, for $\forall i \in \mathcal{I}$. Our first observation is that the achieved GEE, WSEE and EEF performance increases in the low regime of $C^{\max}$ and becomes saturated in the large regime of $C^{\max}$. The reason is that the multi-user interference always exists even as more cooperation can be attained among all RRHs. For interference limited situations, there is an upper bound on the achievable rate for all users so that increasing more fronthaul capacity basically provides no benefit to the system performance. It is obvious that the resource allocation solution obtained from GEE maximization algorithm results the best GEE performance than the WSEE and EEF maximization algorithms in Fig. 4(a). Similarly, WSEE maximization algorithm outperforms the GEE and EEF maximization algorithms in term of the achieved WSEE performance in Fig. 4(b). Meanwhile, EEF maximization algorithm achieves the best minimum EE value compared to the GEE and WSEE maximization algorithms in Fig. 4(c), but incur the small loss of GEE and WSEE performance shown in Fig. 4(a) and 4(b). These observations imply that the EEF metric yields the better minimum EE than GEE and WSEE criteria but still achieves the good performance in terms of GEE and WSEE.

B. Advantages of proposed rate-dependent power model

In this numerical result, we demonstrate the advantages of the GEE, WSEE and EEF performance achieved by our proposed rate dependent power consumption model compared to the linear

power consumption model used mostly in the recent literature. Instead of using our proposed rate dependent fronthaul power model in (6), the linear fronthaul power consumption model (LFP) is expressed as $P_i^{\text{FH}}(\mathbf{a}_i) = \sum_{k\in\mathcal{K}} a_{i,k} P_{i,k}^{\text{fh}}$ where $P_{i,k}^{\text{fh}}$ is the fixed power consumption, i.e., $P_{i,k}^{\text{fh}} = 2$ Watts [21], [34], used for data transmission between the kth user and the ith fronthaul. Then our proposed SCA-based algorithm is employed to optimize the GEE, WSEE and EEF metrics with the LFP, called GEE-LFP, WSEE-LFP and EEF-LFP algorithms, respectively.

Fig. 5(a) and 5(b) plot the GEE and WSEE objective values, respectively, achieved from the solutions of the GEE, WSEE, EEF, GEE-LFP, WSEE-LFP and EEF-LFP maximization algorithms. As expected, achieved GEE and WSEE performance increase and reach a plateau along with the growth of $P^{\max}$. The fact is that the sum achieved rate of users first increases due to the increase of transmit power consumption, which leads to an increase in the achieved GEE and WSEE performance. However, when $P^{\max}$ becomes sufficient large, the gain of sum achieved data rate can not compensate for the quickly increase in the power consumption. Thus, GEE and WSEE algorithms will not use the excess transmit power to further increase the rate to maintain the high values of GEE and WSEE. Another important observation is that our proposed rate dependent power consumption model outperforms the LFP model. This is obvious as the linear power model consumes the fixed power consumption for each active transmission between user and fronthaul, which is not precise in practice.

Fig. 5(c) shows the EEF objective values in (11) calculated from the solutions obtained by applying different algorithms versus $P^{\max}$. Clearly, EEF maximization algorithm that focuses on maximizing the minimum EE, achieves the best EEF objective values than GEE and WSEE maximization algorithms. Moreover, it can be seen that for GEE and WSEE maximization algorithms, EEF value increases and then slightly decreases and eventually get saturated when $P^{\max}$ increases. This can be explained similarly to the explanation of Fig. 5(a) and 5(b) in the previous paragraph. The achieved GEE and WSEE first increase with the increase of $P^{\max}$ because of the increase of users' data rates. However, at a certain point of $P^{\max}$, the increase of sum achieved data rate gain lead to larger transmit beamforming power forced to increase at the RRH which has the worst channel conditions to its served UEs. This results in the decrease of EE at this RRH. Furthermore, at the sufficient large $P^{\max}$ increasing more transmit power provides no benefit to the system throughput, thus GEE and WSEE algorithms will not continue to increase the transmit beamforming at the RRHs to maintain the still good achieved EE performance. Last but not least, the performance in term of EEF achieved from our proposed

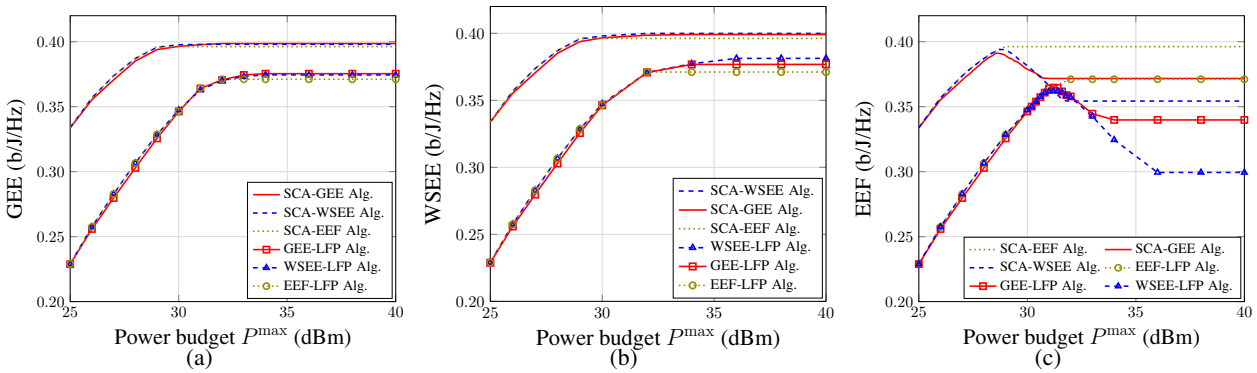

Fig. 5. (a)-(b)-(c): The comparison of GEE, WSEE and EEF performance between our proposed rate-dependent power model and the linear fronthaul power consumption model versus $P^{\max}$.

power model is larger than that of linear power model. This not only proves the benefits of our proposed rate dependent power consumption model, but also emphasizes the important in correctly characterizing the power consumption in C-RAN.

VII. Conclusion

We investigated energy efficient resource allocation in the downlink of a limited fronthaul capacity C-RAN with the constraints where the rate-dependent power model was proposed. Three different EE metrics, namely GEE, WSEE of all RRHs, and EEF across all RRHs were considered. In this work, we customize the BRB algorithm to find the globally optimal solutions for all the visited problems. For the low-complexity solution approach, we presented some novel techniques to transform the combinatorial and non-convex EE problems into more tractable forms. Further, the unified framework based on SCA method and relaxation method was developed to approximate the EE problems into a sequence of SOCP problems. Then, a low-complexity algorithm based on SCA and relaxation framework was conducted to iteratively compute the resource allocation solution at convergence. Numerical results show that our proposed SCA- and relaxed-based algorithm significantly outperform all the existing methods in terms of convergent speed and can achieve near-to-optimal compared to the BRB algorithm. It was shown that WSEE provided more freedoms on the individual EE of each RRH for resource allocation design compared to popular GEE metric. Also, EEF yielded the best balanced EE performance on system design compared to others. Additionally, we have numerically illustrated the significance of our proposed rate-dependent power model in achieve higher EE compared to the existing power model in the literature.

APPENDIX A
PROOF OF THE EQUIVALENCE OF (12) AND (14)

We prove that the constraints in (14b) and (14d) of problem (14) are hold with equalities at optimality by contradiction. Let $\Theta^\star = \{\mathbf{b}^\star, \mathbf{a}^\star, \mathbf{w}^\star, \boldsymbol{\nu}^\star, \boldsymbol{\tau}^\star, \mathbf{t}^\star\}$ denote an optimal solution of (14). By contradiction suppose that (14d) is inactive, i.e., $\tilde{P}^{\text{tot}}(\mathbf{w}^\star, \mathbf{a}^\star, \mathbf{b}^\star, \boldsymbol{\tau}^\star) < 1/t_0^\star$ for GEE maximization problem and $\tilde{P}_i(\mathbf{w}^\star, \mathbf{a}^\star, \mathbf{b}^\star, \boldsymbol{\tau}^\star) < 1/t_i^\star$ for WSEE and EEF maximization problems. Then, for GEE, there exists t_0' such that $t_0' > t_0^\star$ and $\tilde{P}^{\text{tot}}(\mathbf{w}^\star, \mathbf{a}^\star, \mathbf{b}^\star, \boldsymbol{\tau}^\star) \leq 1/t_0'$. That is, t_0' is feasible to (14) but yields a strictly larger objective, which contradicts with the fact that $\Theta^\star$ is an optimal solution. Similarly, assume that $R_k\big(\mathbf{w}^\star\big) > \tau_k^\star$ 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 $\mathbf{w}_i' = \mathbf{w}_i^\star$ for $i \neq k$ and $\mathbf{w}_i' = \xi \mathbf{w}_k^\star$ for $i = k$ for some $0 < \xi < 1$. Intuitively, the beamforming vector of user k is scaled down by a factor of ξ and the beamforming vectors of other users are remain the same. It is easy to see that there exists $\xi \in (0, 1)$ such that $R_k\big(\mathbf{w}'\big) > \tau_k^\star$ for all k. Note that $\big\|\mathbf{w}'\big\|_2 < \big\|\mathbf{w}\big\|_2$ and thus $\tilde{P}^{\text{tot}}(\mathbf{w}', \mathbf{a}^\star, \mathbf{b}^\star, \boldsymbol{\tau}^\star) < \tilde{P}^{\text{tot}}(\mathbf{w}^\star, \mathbf{a}^\star, \mathbf{b}^\star, \boldsymbol{\tau}^\star) \leq 1/t_0{}^\star$. Consequently, we can find t_0' such that $t_0' > t_0^\star$ and $\tilde{P}^{\text{tot}}(\mathbf{w}', \mathbf{a}^\star, \mathbf{b}^\star, \boldsymbol{\tau}^\star) \leq 1/t_0'$, 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 (14b) and (14d) of problem (14) are hold with equalities at optimality. As a result, for given optimal solution $\Theta^\star$ to (14) we can simply obtain optimal solution $(\mathbf{b}^\star, \mathbf{a}^\star, \mathbf{w}^\star, \boldsymbol{\nu}^\star)$ to (12) and the same objective value as that of (14) and vice versa for given optimal solution $(\mathbf{b}^\star, \mathbf{a}^\star, \mathbf{w}^\star, \boldsymbol{\nu}^\star)$ to (12), we can easily find the corresponding optimal solution $\Theta^\star$ to (14) by calculating $\tau_k^\star = R_k(\mathbf{w}^\star)$, $\forall k \in \mathcal{K}$ and $1/t_0^\star = \tilde{P}^{\text{tot}}(\mathbf{w}^\star, \mathbf{a}^\star, \mathbf{b}^\star, \boldsymbol{\tau}^\star)$, and also achieve the same objective value as that of (12). Thus, Lemma 1 is proved. Similar proof steps can be applied for WSEE and EEF maximization problems.

APPENDIX B
CALCULATION OF BOX $\mathcal{U}$

The values of $\underline{\mathbf{s}} = [\underline{\mathbf{a}}^T, \underline{\boldsymbol{\tau}}^T, \underline{\mathbf{t}}^T]^T$ and $\overline{\mathbf{s}} = [\overline{\mathbf{a}}^T, \overline{\boldsymbol{\tau}}^T, \overline{\mathbf{t}}^T]^T$ can be computed as follows. It is easily seen that $\underline{\mathbf{a}} = \mathbf{0}_N$ and $\overline{\mathbf{a}} = \mathbf{1}_N$ where $N = IK$ since $a_{i,k} \in \{\underline{a}_{i,k}, \overline{a}_{i,k}\} = \{0, 1\}$, where $\mathbf{0}_N$ and $\mathbf{1}_N$ denote vectors of all zeros and ones with length of vector is N. Additionally, from (14b), it holds that $\tau_k \geq \log(1 + \Gamma_k^{\min}) = \underline{\tau}_k$, $\forall k \in \mathcal{K}$. Moreover, we have

$$\tau_k \overset{(a)}{\leq} \log(1 + \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sigma_0^2}) \overset{(b)}{\leq} \log\left(1 + \frac{\big\|\mathbf{h}_k\big\|_2^2 \big\|\mathbf{w}_k\big\|_2^2}{\sigma_0^2}\right) \overset{(c)}{\leq} \log\left(1 + I \times P^{\max} \frac{\big\|\mathbf{h}_k\big\|_2^2}{\sigma_0^2}\right) = \overline{\tau}_k, \forall k \in \mathcal{K}. \tag{49}$$

where (a) is due to omitting the inter-user interference, (b) is the result of applying Cauchy–Schwarz inequality, and (c) is obvious from the power constraint for each $\mathbf{w}_{i,k}$. Similarly, an upper bound and lower bound of t_i can be given by$t_i \geq \frac{1}{\overline{p}_i} = \underline{t}_i$ and $t_i \leq \frac{1}{\underline{p}_i} = \overline{t}_i$ where

$$\begin{cases} \overline{p}_0 = \sum_{i\in\mathcal{I}} \rho_i \sum_{k\in\mathcal{K}} \overline{\tau}_k + \sum_{i\in\mathcal{I}} (\frac{1}{\eta_i} P^{\max} + P_i^{\text{ra}}) & \text{if X is GEE} \\ \overline{p}_i = \rho_i \sum_{k\in\mathcal{K}} \overline{\tau}_k + \frac{1}{\eta_i} P^{\max} + P_i^{\text{ra}} & \text{otherwise} \end{cases} \tag{50}$$

$\underline{p_0} = \sum_{i\in\mathcal{I}} P_i^{\text{ri}}$ if X is GEE, and $\underline{p_i} = P_i^{\text{ri}}, \forall i \in \mathcal{I}$, otherwise.

APPENDIX C
PROOF OF (42)

We first show that the gradient of the function $g(x) = -\log(1+x)$ for $x \geq 0$ is Lipschitz continuous with parameter $L = 1$. This can be easily proved since

$$\|\nabla g(x_1) - \nabla g(x_2)\|_2 = \left| -\frac{1}{1+x_1} + \frac{1}{1+x_2} \right| = \left| \frac{x_1 - x_2}{(1+x_1)(1+x_2)} \right| \overset{(a)}{\leq} |x_1 - x_2| \tag{51}$$

where (a) is due to $(1+x_1)(1+x_2) > 1$ for $x_1, x_2 > 0$. Due to the Lipschitz continuity of $\nabla g(x)$, it holds that [35]

$$g(\gamma_k) \leq g(\gamma_k) + \nabla g(\gamma_k^{(n)}) \left(\gamma_k - \gamma_k^{(n)}\right) + \frac{1}{2\lambda} \left(\gamma_k - \gamma_k^{(n)}\right)^2 \tag{52}$$

for $\lambda \in (0, 1]$, and thus completes the proof by noting that (52) is actually (42) when $\lambda = 1$.

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] G. Auer *et al.*, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[3] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud radio access network (C-RAN): A primer," *IEEE Network*, vol. 29, no. 1, pp. 35–41, Jan. 2015.

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

[5] 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.

[6] 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.

[7] S. B. *et al.*, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 697–709, Apr. 2016.

[8] 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.

[9] T. M. Nguyen, A. Yadav, W. Ajib, and C. Assi, "Energy efficiency with adaptive decoding power and wireless backhaul small cell selection," in *IEEE Global Commun. Conf. (Globecom)*, Dec. 2016, pp. 1–6.

[10] Q. Shi, C. Peng, W. Xu, M. Hong, and Y. Cai, "Energy efficiency optimization for MISO SWIPT systems with Zero-Forcing beamforming," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 842–854, Feb. 2016.

[11] K. Xiong, P. Fan, Y. Lu, and K. B. Letaief, "Energy efficiency with proportional rate fairness in multirelay OFDM networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1431–1447, May 2016.

[12] C. Pan, W. Xu, J. Wang, H. Ren, W. Zhang, N. Huang, and M. Chen, "Pricing-based distributed energy-efficient beamforming for MISO interference channels," *IEEE Trans. Veh. Technol.*, vol. 34, no. 4, pp. 710–722, Apr. 2016.

[13] K.-G. Nguyen, Q.-D. Vu, M. Juntti, and L.-N. Tran, "Energy efficiency maximization for C-RANs: Discrete monotonic optimization, penalty, and ℓ_0-Approximation methods," *IEEE Trans. Signal Process.*, vol. 66, no. 17, pp. 4435–4449, Sep. 2018.

[14] M. P. *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.

[15] J.-Y. Lin, C.-H. Lee, and H.-W. Tsao, "On the optimization of user-centric energy-efficient C-RAN," in *IEEE Intern. Conf. Commun. (ICC)*, Malaysia, May 2016, pp. 1962–1967.

[16] Z. Zhou, M. Dong, K. Ota, G. Wang, and L. T. Yang, "Energy-efficient resource allocation for D2D communications underlaying Cloud-RAN-based LTE-A networks," *IEEE Internet of Things Jour.*, vol. 3, no. 3, pp. 428–438, Jun. 2016.

[17] 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.

[18] 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.

[19] 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.

[20] W. N. S. F. W. Ariffin, X. Zhang, and M. R. Nakhai, "Sparse beamforming for real-time resource management and energy trading in green C-RAN," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 2022–2031, Jul. 2017.

[21] 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.

[22] 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.

[23] ——, "Joint virtual computing and radio resource allocation in limited fronthaul green C-RANs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2602–2617, Apr. 2018.

[24] 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.

[25] 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.

[26] 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.

[27] Q. Zhang, C. Yang, H. Haas, and J. S. Thompson, "Energy efficient downlink cooperative transmission with bs and antenna switching off," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 5183–5195, Sep. 2014.

[28] B. Marks and G. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, 1977.

[29] J. Kalevay, M. Bandez, A. Tolliy, M. Junttiy, and V. V. Veeravalli, "Sum rate maximizing joint processing with limited backhaul and tree topology constraints," in *Proc. IEEE Int. Workshop on Sig. Proces. Advances in Wireless Commun. (SPAWC'16)*, Edinburgh, UK, June 2016, pp. 1–5.

[30] 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.

[31] 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.

[32] H. Tuy and F. A. khayya-and P. Thach, "Monotonic optimization: Branch and cut methods," *Essays and Surveys in Global Optimization*, pp. 39–78, C. Audet, P. Hansen, and G. Savard, Eds. , Srpinger US 2005.

[33] 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.

[34] J. Tang, W. P. Tay, and T. Q. S. Quek, "System cost minimization in cloud ran with limited fronthaul capacity," *IEEE Trans. Wireless Commun.*, vol. PP, no. 99, pp. 1–15, Dec. 2017.

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