Energy Efficiency Maximization for Downlink Cloud Radio Access Networks With Data Sharing and Data Compression

Tung Thanh Vu, *Student Member, IEEE*, Duy Trong Ngo, *Member, IEEE*, Minh N. Dao, Salman Durrani, *Senior Member, IEEE*, Duy H. N. Nguyen, *Member, IEEE*, and Richard H. Middleton, *Fellow, IEEE*

***Abstract*— This paper aims to maximize the energy efficiency of a downlink cloud radio access network (C-RAN). Here, data is transferred from a baseband unit in the core network to several remote radio heads via a set of edge routers over capacity-limited fronthaul links. The remote radio heads then send the received signals to their users via radio access links. Both data sharing and compression-based strategies are considered for fronthaul data transfer. New mixed-integer nonlinear problems are formulated, in which the ratio of network throughput and total power consumption is maximized. These challenging problem formulations include practical constraints on routing, predefined minimum data rates, fronthaul capacity, and maximum remote radio head transmit power. By employing the successive convex quadratic programming, iterative algorithms are proposed with guaranteed convergence to the Fritz John solutions of the formulated problems. Significantly, each iteration of the proposed algorithms solves only one simple convex program. Numerical examples with practical parameters confirm that the proposed joint optimization designs markedly improve the C-RAN's energy efficiency compared to benchmark schemes. They also show that the fronthaul data-sharing strategy outperforms its compression-based counterpart in terms of energy efficiency, in both single-hop and multi-hop network scenarios.**

***Index Terms*— C-RAN, energy efficiency, limited-capacity fronthaul, precoding design, user association.**

Manuscript received October 23, 2017; revised January 21, 2018 and April 20, 2018; accepted April 27, 2018. Date of publication May 15, 2018; date of current version August 10, 2018. This work was supported in part by The University of Newcastle through the ECR-HDR Scholarship, in part by the Australian Research Council Discovery Project scheme under Grant DP170100939 and Grant DP160101537, in part by the National Foundation for Science and Technology Development of Vietnam under Grant 101.02-2016.11, and in part by the Startup Fund from San Diego State University. The associate editor coordinating the review of this paper and approving it for publication was W. P. Tay. *(Corresponding author: Tung Thanh Vu.)*

T. T. Vu, D. T. Ngo, and R. H. Middleton are with the School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW 2308, Australia (e-mail: thanhtung.vu@uon.edu.au; duy.ngo@newcastle.edu.au; richard.middleton@newcastle.edu.au).

M. N. Dao is with the Priority Research Centre for Computer-Assisted Research Mathematics and Its Applications, The University of Newcastle, Callaghan, NSW 2308, Australia (e-mail: daonminh@gmail.com).

S. Durrani is with the Research School of Engineering, The Australian National University, Canberra, ACT 2601, Australia (e-mail: salman.durrani@anu.edu.au).

D. H. N. Nguyen is with the Department of Electrical and Computer Engineering, San Diego State University, San Diego, CA 92182 USA (e-mail: duy.nguyen@sdsu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2018.2834370

I. Introduction

THE telecommunications industry is currently facing an exponential growth in mobile data traffic [1]. To meet the traffic demand with limited radio spectrum, it is crucial to improve the spectral efficiency of the communication system [2]. However, focusing only on spectral efficiency may lead to additional energy required to support the increasing number of base station sites [3]. Moreover, serious concerns about global warming and operational cost are rising due to the enormous growth of energy consumed by wireless network deployment [4]. As such, it is necessary to balance the tradeoff between the spectral and energy efficiencies and to redesign the existing cellular networks from the green communication perspective. It is also important to investigate the resource allocation designs that not only achieve high spectral efficiency but also improve the energy efficiency in fifth generation (5G) of mobile communication systems [5]. Specifically, increasing $1,000$-fold data traffic and reducing the total network energy consumption by half are listed among of the top priorities of 5G [6].

Cloud radio access networks (C-RANs) have been proposed and considered as a promising solution to meet 5G's challenging objectives [2], [7]–[10]. In a C-RAN, low-cost low-power remote radio heads (RRHs) replace traditional high-cost high-power base stations, resulting in lower energy consumption in a dense network deployment [11]. A central base band unit (BBU) in the core network is connected to the RRHs via wireline fronthaul links, whilst the RRHs are connected to end users via radio access links. The most important advantage of C-RANs is that large-scale allocation of radio and computing resources across all the RRHs is centrally processed at the same BBU pools. As such, significant spectral and energy efficiency gains over the single-cell processing can be realized [12]. However, due to joint processing C-RANs require significant data sharing on the fronthaul, making these capacity-limited links a main bottleneck in practical C-RANs [13]. It is therefore crucial to effectively utilize the C-RAN fronthaul links [14].

Data transfer over C-RAN fronthaul links is based on data-sharing and compression-based strategies [3], [15]–[17]. Under data-sharing, the BBU multicasts users' data messages to a set of RRHs over the fronthaul links. These RRHs then

pre-code or beamform the messages and cooperatively transmit them to the users via radio access links. With compression, the users' messages are instead pre-coded/beamformed and compressed centrally at the BBU. The BBU then unicasts the compressed messages to the RRHs over the fronthaul links for subsequent wireless transmission to the users. Although the compression-based strategy incurs a lesser amount of fronthaul traffic, their throughput performance is limited by the quantization noise generated during the compression process [15]. In the case of power minimization, [3] shows that the compression-based strategy requires less power consumption than the data-sharing counterpart when the fixed user data rate is high. It therefore remains unclear which of the two strategies is a better option in terms of energy efficiency typically depending on both throughput achievement and power consumption.

Energy efficiency can be measured by the area power consumption metric (watts/unit area) [18] or the economical energy efficiency metric (effective bits/Joule) [19]. A more widely adopted metric is the ratio of the network throughput and the total power consumption (bits/Joule) [3], [20], [21]. To improve energy efficiency, one may opt to (i) increase data rate, (ii) decrease RRH transmit power, (iii) turn off RRHs, (iv) reduce the fronthaul rate for power saving, and (v) implement any combination thereof. Among the indirect approaches, [3] fixes user rates and transforms the energy efficiency maximization problem into a power minimization problem. Although both data-sharing and compression-based strategies are considered, the limited fronthaul capacity constraint is not taken into account. The work of [22] addresses a different power minimization problem for downlink C-RANs under the data-sharing strategy, where users are put into several multicast groups. Applying the random matrix theory, [21] proposes heuristic user association (UA) schemes that maximize the equivalent energy efficiency under data sharing. On the other hand, the study of [23] directly maximizes the energy efficiency under the compression-based method. Nonetheless, the data-sharing scenario and the power consumption at the fronthaul links are not investigated by [23].

It is worth noting that the aforementioned solutions focus on single-hop fronthaul networks only, while a practical BBU is typically connected to RRHs via a number of edge routers over a multi-hop fronthaul network [15], [24]. Very recently, [15] attempts to maximize the network throughput of a downlink multi-hop C-RAN, where network coding is shown necessary to better utilize the finite-capacity fronthaul links in the multi-hop fronthaul case. However, to the best of our knowledge, energy efficiency maximization has not yet been addressed for this practical network scenario.

This paper considers the downlink of a C-RAN with multi-hop and capacity-limited fronthaul. The aim is to directly maximize the network energy efficiency by jointly optimizing user association (UA), RRH activation, data rate allocation and signal precoding. Since these optimization variables are strongly interrelated, it is not straightforward to even formulate these problems in a suitable form, let alone solve them effectively and optimally. We specifically try to answer: (i) Can we do better than existing solutions in terms of energy efficiency? (ii) With our devised solution, is data sharing or compression better? In answering those questions, we make the following research contributions.

- We formulate new problems of energy efficiency maximization under both data-sharing and compression-based strategies. They are subject to routing constraints, limited fronthaul capacities, predefined minimum rates and maximum transmit power at each RRH. Practically, all the power consumption resulting from data transmission, RRH and fronthaul operations is included in the formulations.
- We propose new iterative algorithms to solve the challenging mixed-integer nonlinear problem formulations, where each iteration involves solving only one simple convex program. Specifically, the original problems are first transformed into their epigraph forms. To deal with the binary nature of UA and RRH activation decisions, they are further recast to equivalent problems that include continuous variables only. These problems are finally solved by the successive convex quadratic programming.
- We prove theoretically and verify by numerical examples that the proposed algorithms converge to Fritz John solutions[1] once initialized from a feasible point. Simulation results with practical parameter settings show that our joint optimization approach significantly improves the energy efficiency over existing methods. Importantly, it is confirmed that the data-sharing strategy offers higher energy efficiency than the compression-based strategy in both single-hop and multi-hop cases.

It should be noted that this work substantially extends our initial result in [26], where only the energy-efficient design of the downlink C-RANs under the data-sharing strategy was demonstrated.

The rest of this paper is organized as follows. Section II presents the system model under consideration. Sections III and IV respectively formulate and solve the optimization problem for the data-sharing strategy, whereas Sections V and VI for the compression-based strategy. Section VII verifies the performance of the proposed algorithms through comprehensive numerical examples. It also compares the energy efficiency performance of the data-sharing and compression-based strategies in both single-hop and multi-hop scenarios. Finally, Section VIII concludes the paper.

Notation: The real part of a complex number x is denoted as $\Re\{x\}$. Boldface symbols are used for vectors and capitalized boldface symbols for matrices. $\boldsymbol{X}^H$ is the conjugate transposition of a matrix $\boldsymbol{X}$. $\langle \boldsymbol{X} \rangle$ means the trace of a matrix $\boldsymbol{X}$. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{Q})$ denotes the circularly symmetric complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{Q}$. $\boldsymbol{I}$ and $\boldsymbol{0}$ are the identity and zero matrices with appropriate dimensions, respectively. For ease of reference, the symbols frequently used throughout the paper are listed in Table I.

[1] A point is called a Fritz John solution if it satisfies Fritz John conditions which are necessary conditions for a solution in nonlinear programming to be optimal [25].

TABLE I
FREQUENTLY USED SYMBOLS

Symbols	Definition
$\mathcal{K}_R, K_R$	Set of RRHs and its cardinality
$\mathcal{K}_U, K_U$	Set of users and its cardinality
$\mathcal{M}, M$	Set of routers and its cardinality
$\mathcal{N}, N$	Set of fronthaul links and its cardinality
C_n	Capacity of fronthaul link n
N_r, N_u	Numbers of antenna at each RRH and at each user
M_k	Message intended for user k
R_k	Data transmission rate for message M_k
$\boldsymbol{s}_k$	Encoded symbol of message M_k
d	Number of data streams
$\boldsymbol{F}_{k,i}$	Precoding matrix for symbol $\boldsymbol{s}_k$ at RRH i
$\boldsymbol{x}_i$	Transmit signal at RRH i
$\boldsymbol{Q}_i$	Covariance matrix of quantization noise at RRH i
$\boldsymbol{H}_{k,i}$	Channel matrix from RRH i to user k
$f_{k,i,n}$	Conceptual flow rate of message M_k transferred to RRH i over link n
$r_{k,n}$	Actual flow rate of message M_k over link n
$d_{i,n}$	Data rate of the compressed signal transferred to RRH i over link n
$a_{k,i}$	RRH-user association variable
b_i	RRH activation variable

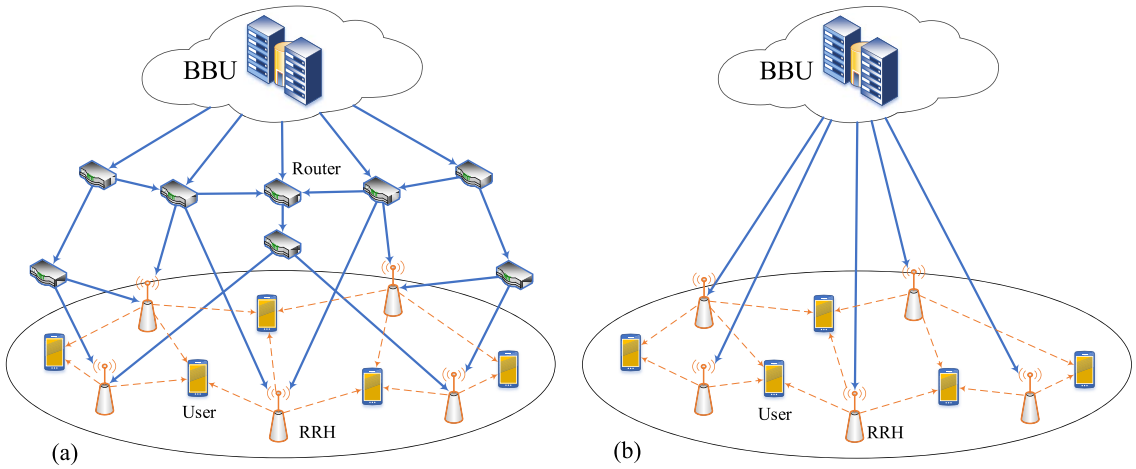


Fig. 1. A general C-RAN architecture with multi-hop fronthaul links [in (a)] and single-hop fronthaul links [in (b)].

II. SYSTEM MODEL

Fig. 1 illustrates the downlink of a general C-RAN model, where the baseband unit (BBU) in the core network connects to a set of RRHs $\mathcal{K}_R \triangleq \{1, \ldots, K_R\}$ via a fronthaul network of M routers and N noiseless fronthaul links [3], [15], [27],[2] where $M > 0$, $N > K_R$ for a multi-hop C-RAN and $M = 0$, $N = K_R$ for a single-hop C-RAN. Denote by $\mathcal{M} \triangleq \{1, \ldots, M\}$ and $\mathcal{N} \triangleq \{1, \ldots, N\}$ the sets of routers and fronthaul links, respectively. Assume that a fronthaul link $n \in \mathcal{N}$ has a limited capacity of $C_n > 0$ (in bits per second).

[2]In practical C-RAN-based LTE scenarios, quantization error always exists on the fronthaul links due to low resolution digital-to-analog conversion (DAC)/analog-to-digital conversion (ADC) at the BBU, RRHs and routers [28]. This paper assumes that the fronthaul link is ideal with high resolution DAC/ADC. We refer the case of noisy fronthaul links with low resolution DAC/ADC to [29] and leave the energy efficiency optimization problem of noisy-fronthaul C-RANs for future work.

The RRHs then serve a set of users $\mathcal{K}_U \triangleq \{1, \ldots, K_U\}$ via radio access links, where a user is allowed to connect to multiple RRHs. Each user $k \in \mathcal{K}_U$ is equipped with N_u antennas while each RRH $i \in \mathcal{K}_R$ is equipped with N_r antennas.

At the BBU, a message M_k intended for the user k is uniformly distributed in the set $\{1, \ldots, 2^{uR_k}\}$, where u is the block length and R_k (in bits per second) is the data rate of message M_k [30]. The message M_k is then encoded into symbol $\boldsymbol{s}_k \in \mathbb{C}^{d \times 1}$, where $\boldsymbol{s}_k$ is taken from a Gaussian channel codebook $\mathcal{C}_k^{\text{CH}}$ so that $\boldsymbol{s}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$ and $d \triangleq \min(N_u, N_r)$ is the number of data streams. The network throughput is defined as the following sum rate:

$$R_{\text{sum}} \triangleq \textstyle\sum_{k \in \mathcal{K}_U} R_k. \tag{1}$$

To transfer the information signal from the BBU to the RRHs via the fronthaul network, data-sharing and

compression-based strategies are used [3], [15], [16].[3] Using either strategy, each user's intended message symbol is routed through the fronthaul and delivered to a set of RRHs. Finally, at the radio access links, the RRH-user associations and RRH activation are respectively expressed by the following binary variables

$$a_{k,i} \triangleq \begin{cases} 1, & \text{if RRH } i \text{ serves user } k, \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

$$b_i \triangleq \begin{cases} 0, & \text{if RRH } i \text{ serves no user,} \\ 1, & \text{otherwise.} \end{cases} \tag{3}$$

Denote the set of RRHs that serves user k as $\mathcal{D}_k \triangleq \{i|a_{k,i} = 1, i \in \mathcal{K}_R\}$. Then the BBU sends $\boldsymbol{s}_k$ to $\mathcal{D}_k$ over the multi-hop fronthaul network at the rate R_k.

This paper assumes that the features of the remote radio heads (RRHs) in C-RANs are similar to those of the conventional RRHs. We apply the power consumption model based on existing statistics of [31] for each RRH. This model has been widely used in the literature (see, e.g., [3], [32]). Specifically, the power consumed by an RRH $i \in \mathcal{K}_R$ in a given transmission interval is expressed as

$$P_i^{\text{RRH,T}} \triangleq \begin{cases} \beta_i P_i^{\text{Tx,T}} + P_{i,a}, & \text{if } 0 < P_i^{\text{Tx,T}} \leq P_i, \\ P_{i,s}, & \text{if } P_i^{\text{Tx,T}} = 0, \end{cases} \tag{4}$$

where the constant $\beta_i > 0,\ i \in \mathcal{K}_R$ represents the power amplifier efficiency, feeder loss and other loss factors due to power supply and cooling for RRH i [3]; $P_i^{\text{Tx,T}}$ is the transmit power required to deliver all requested files from RRH i under a strategy T, in which $\text{T} \in \{\text{DS}, \text{C}\}$, and DS and C respectively refer to the data-sharing strategy and the compression-based strategy; $P_{i,a}$ is the power required to support the RRH i in the active mode; $P_{i,s} < P_{i,a}$ is the power consumption in the sleep mode; and P_i is the maximum transmit power of the RRH i.

III. Problem Formulation for Fronthaul Transfer With Data Sharing Strategy

In the data-sharing strategy, users' intended signals are multicast by the BBU to a selected set of RRHs which then cooperatively perform precoding and transmit the precoded signals to the users [3], [15], [27]. In what follows, we will describe each stage of the communication in detail.

A. Data Multicasting Routing With Network Coding at BBU

For information multicast, one may allow each router to replicate and forward its received information to the next routers, or instead use network coding to make each router an encoding and decoding point for the information flows [33]. As shown in [34], the replicate-and-forward approach is not optimal whilst the coding operation at routers can achieve multicast capacity. Following [15], this paper employs a network coding scheme that consists of a flow routing scheme and a code assignment to determine the rate and content of each data flow being delivered across the network. The study of [35] shows that intersession coding only provides marginal throughput gain over independent coding while making the routing problem an NP-hard one. Therefore, this paper assumes that each multicast session is routed and coded independently but not jointly.

In our model, there are K_U multicast sessions for K_U users' messages. Denote by $r_{k,n}$ the routing variable that determines the flow rate on a link n for a multicast session k. If $r_{k,n} = 0$, the multicast session k is not routed on the link n. The network coding theorem in [34] shows that if the rate R_k is achievable at each destination in $\mathcal{D}_k$ independently, it is also achievable for the entire multicast session k. Hence, the multicast flow on the link n to an RRH $i \in \mathcal{D}_k$ can be viewed as including the independent conceptual flows $f_{k,i,n} \leq r_{k,n}, \forall k \in \mathcal{K}_U, i \in \mathcal{K}_R$ [33]. The routing constraints for the multi-hop fronthaul network can be formulated as follows [15], [33]

$$f_{k,i,n} \leq r_{k,n}, \quad \forall k \in \mathcal{K}_U,\ i \in \mathcal{K}_R,\ n \in \mathcal{N} \tag{5}$$

$$r_{k,n} \leq a_{k,i} C_n, \quad \forall k \in \mathcal{K}_U,\ n \in \mathcal{I}_i^{\mathcal{K}_R} \tag{6}$$

$$a_{k,i} R_k \leq \sum_{n \in \mathcal{I}_i^{\mathcal{K}_R}} f_{k,i,n}, \quad \forall k \in \mathcal{K}_U,\ i \in \mathcal{K}_R \tag{7}$$

$$\sum_{n \in \mathcal{O}_m^{\mathcal{M}}} f_{k,i,n} = \sum_{n \in \mathcal{I}_m^{\mathcal{M}}} f_{k,i,n}, \quad \forall k \in \mathcal{K}_U, \ i \in \mathcal{K}_R,\ m \in \mathcal{M} \tag{8}$$

$$\sum_{k \in \mathcal{K}_U} r_{k,n} \leq C_n, \quad \forall n \in \mathcal{N} \tag{9}$$

$$R_k \geq R_{\text{QoS}}, \quad r_{k,n} \geq 0,\ f_{k,i,n} \geq 0, \quad \forall k \in \mathcal{K}_U,\ i \in \mathcal{K}_R,\ n \in \mathcal{M}, \tag{10}$$

where $\mathcal{I}_i^{\mathcal{K}_R}$ denotes the set of incoming links at an RRH i while $\mathcal{I}_m^{\mathcal{M}}$ and $\mathcal{O}_m^{\mathcal{M}}$ are the sets of incoming and outgoing links at a router m, respectively. Constraint (5) shows that the actual flow rate on link n for multicast session k is a MAX operation, i.e., $r_{k,n} = \max_{i \in \mathcal{D}_k} f_{k,i,n}$, instead of a SUM operation, i.e., $r_{k,n} = \sum_{i \in \mathcal{D}_k} f_{k,i,n}$, of the conceptual flows. This is the benefit of network coding in which the amount of information conveyed on a fixed-capacity link is increased by splitting each multicast session into subsessions and sending only an XOR version of the subsessions on this link.[4] Constraint (6) makes sure there is no data transmission to the unassigned or inactive RRHs. Constraint (7) guarantees that each RRH $i \in \mathcal{D}_k$ can receive the information flow at rate R_k when $a_{k,i} = 1$. Constraint (8) follows the law of flow conservation for conceptual flows at a router m. Constraint (9) ensures that the information flow for all K multicast sessions does not exceed each link capacity. Constraint (10) guarantees a quality-of-service (QoS) rate $R_{\text{QoS}} \geq 0$ for each user as well as nonnegative flow rates on all links for all the multicast sessions. For any given flow rates that satisfy the routing constraints (5)–(10), a code assignment scheme can be found to design the content of each flow [15], [37], [38].

[3] In [16], these fundamentally different strategies are respectively referred to as hard-transfer and soft-transfer modes.

[4] This paper defines network coding as the coding at a node in a network where there are arbitrary mappings from inputs to output of the node. This definition is different from that of coding at a node in a packet network where data are divided into packets and the network coding is applied to the contents of the packets [36].

B. Signal Precoding and Transmission by RRHs

After receiving the message symbols from all K multicast sessions via the multi-hop fronthaul network, an RRH i generates a baseband signal $\boldsymbol{x}_i \in \mathbb{C}^{N_r \times 1}$ as

$$\boldsymbol{x}_i = \sum_{k \in \mathcal{K}_U} \boldsymbol{F}_{k,i}\, \boldsymbol{s}_k, \tag{11}$$

where $\boldsymbol{F}_{k,i} \in \mathbb{C}^{N_r \times d}$ is the precoding matrix applied by the RRH i to $\boldsymbol{s}_k$. The RRH i is subjected to the average transmit power constraint P_i as

$$\mathbb{E}\left\{\| \boldsymbol{x}_i \|^2\right\} \le P_i. \tag{12}$$

Denote by $\boldsymbol{H}_{k,i} \in \mathbb{C}^{N_u \times N_r}$ the flat-fading channel matrix from the RRH i to a user k and by $\boldsymbol{H}_k \triangleq [\boldsymbol{H}_{k,1}, \ldots, \boldsymbol{H}_{k,K_R}] \in \mathbb{C}^{N_u \times N_R}$ the channel matrix from all RRHs to the user k, where $N_R \triangleq K_R N_r$. Assume that channel states $\boldsymbol{H}_{k,i}, k \in \mathcal{K}_U, i \in \mathcal{K}_R$ remain unchanged during the transmission interval and are available at the BBU and RRHs [15], [16]. Upon defining $\bar{\boldsymbol{F}}_k \triangleq \left[(\boldsymbol{F}_{k,1})^H, (\boldsymbol{F}_{k,2})^H, \ldots (\boldsymbol{F}_{k,K_R})^H\right]^H \in \mathbb{C}^{N_R \times d}$, the received signal $\boldsymbol{y}_k \in \mathbb{C}^{N_u \times 1}$ at the user k can be written as

$$\begin{aligned}
\boldsymbol{y}_k &= \sum_{i \in \mathcal{K}_R} \boldsymbol{H}_{k,i}\, \boldsymbol{x}_i + \boldsymbol{n}_k \\
&= \sum_{i \in \mathcal{K}_R} \boldsymbol{H}_{k,i}\, \boldsymbol{F}_{k,i}\, \boldsymbol{s}_k + \sum_{i \in \mathcal{K}_R} \sum_{\ell \in \mathcal{K}_U \setminus \{k\}} \boldsymbol{H}_{\ell,i}\, \boldsymbol{F}_{\ell,i}\, \boldsymbol{s}_\ell + \boldsymbol{n}_k \\
&= \boldsymbol{H}_k\, \bar{\boldsymbol{F}}_k\, \boldsymbol{s}_k + \underbrace{\sum_{\ell \in \mathcal{K}_U \setminus \{k\}} \boldsymbol{H}_\ell\, \bar{\boldsymbol{F}}_\ell\, \boldsymbol{s}_\ell}_{\text{interference}} + \boldsymbol{n}_k,
\end{aligned} \tag{13}$$

where $\boldsymbol{n}_k \in \mathbb{C}^{N_u \times 1}$ is the additive noise term with $\boldsymbol{n}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma}_k)$. By treating the interference in (13) as additive Gaussian noise, the rate R_k of the message symbol $\boldsymbol{s}_k$ is always achievable in the Shannon's sense as follows

$$R_k \le g_k(\bar{\boldsymbol{F}}) \triangleq W \log_2 \left| \boldsymbol{I}_{N_u} + \boldsymbol{\Pi}_k\, \boldsymbol{\Pi}_k^H\, \boldsymbol{\Xi}_k^{-1} \right|, \tag{14}$$

where W is the total available bandwidth, $\bar{\boldsymbol{F}} \triangleq \{\bar{\boldsymbol{F}}_k\}_{k \in \mathcal{K}_U}$, $\boldsymbol{\Pi}_k \triangleq \boldsymbol{H}_k\, \bar{\boldsymbol{F}}_k$, and

$$\boldsymbol{\Xi}_k \triangleq \sum_{\ell \in \mathcal{K}_U \setminus \{k\}} \boldsymbol{H}_k\, \bar{\boldsymbol{F}}_\ell \bar{\boldsymbol{F}}_\ell^H\, \boldsymbol{H}_k^H + \boldsymbol{\Sigma}_k\,. \tag{15}$$

C. User Association and RRH Activation at Radio Access Links

Define $\boldsymbol{a} \triangleq \{a_{k,i}\}_{k \in \mathcal{K}_U, i \in \mathcal{K}_R}$ and $\boldsymbol{b} \triangleq \{b_i\}_{i \in \mathcal{K}_R}$. Let $\bar{\boldsymbol{E}}_i \in \mathbb{C}^{N_R \times N_r}$ be zero everywhere except an identity matrix of size N_r from row $(i-1)N_r + 1$ to row iN_r. The interdependence among $\boldsymbol{a}$, $\boldsymbol{b}$ and $\bar{\boldsymbol{F}}$ is modeled as

$$a_{k,i} = \begin{cases} 0, & \text{if } \langle \bar{\boldsymbol{E}}_i^H\, \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^H\, \bar{\boldsymbol{E}}_i \rangle = 0, \\ 1, & \text{otherwise,} \end{cases} \quad \forall k \in \mathcal{K}_U,\ i \in \mathcal{K}_R \tag{16}$$

$$b_i = \begin{cases} 0, & \text{if } a_{k,i} = 0, \quad \forall k \in \mathcal{K}_U \\ 1, & \text{otherwise,} \end{cases} \tag{17}$$

$$= \begin{cases} 0, & \text{if } \sum_{k \in \mathcal{K}_U} \langle \bar{\boldsymbol{E}}_i^H\, \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^H\, \bar{\boldsymbol{E}}_i \rangle = 0, \\ 1, & \text{otherwise,} \end{cases} \quad \forall i \in \mathcal{K}_R, \tag{18}$$

where (16) and (17) are indeed (2) and (3) in the sense that an RRH i is assigned to serve a user k if and only if the corresponding precoder of the message symbol $\boldsymbol{s}_k$ is not a zero matrix, i.e., $\boldsymbol{F}_{k,i} = \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_k \neq \boldsymbol{0}$. From (16) and (17), the relationship between $\boldsymbol{a}$ and $\boldsymbol{b}$ is expressed as

$$a_{k,i} \le b_i \le \sum_{k \in \mathcal{K}_U} a_{k,i}, \quad \forall k \in \mathcal{K}_U,\ i \in \mathcal{K}_R, \tag{19}$$

which guarantees that no user be assigned to an inactive RRH.

D. Power Consumption

The transmit power at an RRH i is computed as

$$\begin{aligned}
P_i^{\text{Tx,DS}} &= b_i \sum_{k \in \mathcal{K}_U} a_{k,i} \langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i \rangle = b_i \sum_{k \in \mathcal{K}_U} \langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i \rangle \\
&= \sum_{k \in \mathcal{K}_U} \langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i \rangle.
\end{aligned} \tag{20}$$

On the other hand, a fronthaul link $n \in \mathcal{N}$ can be modeled as a set of communication channels with a total capacity C_n and total power dissipation $P_{n,\max}^{\text{FH}}$. Its power consumption is given by [3]

$$P_n^{\text{FH,DS}} \triangleq \frac{\sum_{k \in \mathcal{K}_U} r_{k,n}}{C_n} P_{n,\max}^{\text{FH}} = \alpha_n \sum_{k \in \mathcal{K}_U} r_{k,n}, \tag{21}$$

where $\alpha_n \triangleq P_{n,\max}^{FH}/C_n$ and $r_{k,n}$ has previously been defined as the actual flow rate on the link n for a multicast session k. From (4)–(21), the total network power consumption with the data-sharing strategy employed on fronthaul links is

$$\begin{aligned}
P_{\text{total}}^{\text{DS}}(\bar{\boldsymbol{F}}, \boldsymbol{r}, \boldsymbol{b}) &\triangleq \sum_{i \in \mathcal{K}_R} P_i^{\text{RRH,DS}} + \sum_{n \in \mathcal{N}} P_n^{\text{FH,DS}} \\
&= \sum_{i \in \mathcal{K}_R} \left(\beta_i P_i^{\text{Tx,DS}} + b_i P_{i,\Delta}\right) \\
&\quad + \sum_{n \in \mathcal{N}} \alpha_n \sum_{k \in \mathcal{K}_U} r_{k,n} + P_s, \\
&= \sum_{i \in \mathcal{K}_R} \left(\beta_i \sum_{k \in \mathcal{K}_U} \langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i \rangle + b_i P_{i,\Delta}\right) \\
&\quad + \sum_{n \in \mathcal{N}} \alpha_n \sum_{k \in \mathcal{K}_U} r_{k,n} + P_s
\end{aligned} \tag{22}$$

where $\bar{\boldsymbol{F}} \triangleq \{\bar{\boldsymbol{F}}_{k,i}\}_{k \in \mathcal{K}_U, i \in \mathcal{K}_R}$, $\boldsymbol{r} \triangleq \{r_{k,n}\}_{k \in \mathcal{K}_U, n \in \mathcal{N}}$, $P_{i,\Delta} \triangleq P_{i,a} - P_{i,s}$, and $P_s \triangleq \sum_{i \in \mathcal{K}_R} P_{i,s}$.

E. Problem Formulation

Here, we aim to maximize the network energy efficiency defined as the ratio of the achievable sum rate and the total power consumption. The optimization problem for the multi-hop C-RAN with the data-sharing strategy is formulated as follows.

$$\max_{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{R}, \bar{\boldsymbol{F}}, \boldsymbol{f}, \boldsymbol{r}} \quad \mathcal{P}_1^{\text{DS}} \triangleq \frac{R_{\text{sum}}}{P_{\text{total}}^{\text{DS}}} \tag{23a}$$

$$\text{s.t.} \quad (5) - (10), (14), (16), (17) \tag{23b}$$

$$\sum_{k \in \mathcal{K}_U} \langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i \rangle \le P_i, \quad \forall i \in \mathcal{K}_R \tag{23c}$$

$$\sum_{i \in \mathcal{K}_R} a_{k,i} \ge 1, \quad \forall k \in \mathcal{K}_U, \tag{23d}$$

where $\boldsymbol{R} \triangleq \{R_k\}_{k\in\mathcal{K}_U}$; $\boldsymbol{f} \triangleq \{f_{k,i,n}\}_{k\in\mathcal{K}_U, i\in\mathcal{K}_R, n\in\mathcal{N}}$; R_{sum} and $P_{\text{total}}^{\text{DS}}$ are defined in (1) and (22), respectively; constraint (23c) is the per-RRH power constraint (12) via (11) and (20); constraint (23d) guarantees that each user is served by at least one active RRH.

For the case of single-hop fronthaul, the data-sharing strategy does not require the network coding in Section III-A. The formulation (23) above therefore does not cover this case. Instead, (23) has to be modified by replacing the constraints (5)–(10) with the following constraints:

$$0 \le r_{k,i} \le a_{k,i} C_i, \quad \forall k \in \mathcal{K}_U,\ i \in \mathcal{K}_R \tag{24}$$

$$a_{k,i} R_k \le r_{k,i}, \quad \forall k \in \mathcal{K}_U,\ i \in \mathcal{K}_R \tag{25}$$

$$\sum_{k\in\mathcal{K}_U} r_{k,i} \le C_i, \quad \forall i \in \mathcal{K}_R \tag{26}$$

$$R_k \ge R_{\text{QoS}}, \quad r_{k,i} \ge 0,\ \forall k \in \mathcal{K}_U,\ i \in \mathcal{K}_R\,. \tag{27}$$

It is noteworthy that the mathematical structure of constraints (24)–(27) is similar to that of (5)–(10). Therefore, the algorithm devised in the next section for (23) can be straightforwardly adapted to solve this corresponding single-hop problem too.

IV. Proposed Algorithm for Fronthaul Data Sharing Strategy

Since problem (23) is combinatorial, nonconvex and fractional, it is challenging to find its global optimality. Global methods would incur prohibitive computational complexity even for problem instances of small to medium sizes. This paper instead proposes a new method that is suitable for practical implementation. To this end, we first rewrite problem (23) in an epigraph form as follows [39]

$$\max_{t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b}} \ t \tag{28a}$$

$$\text{s.t. } (5)-(10),(14),(19),(23\text{d}) \tag{28b}$$

$$tz \le \sum_{k\in\mathcal{K}_U} R_k \tag{28c}$$

$$z \ge \sum_{i\in\mathcal{K}_R} \Big(\beta_i \sum_{k\in\mathcal{K}_U} \langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i\rangle + b_i P_{i,\Delta}\Big) + \sum_{n\in\mathcal{N}} \alpha_n \sum_{k\in\mathcal{K}_U} r_{k,n} + P_s \tag{28d}$$

$$\langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i\rangle \le u_{k,i}, \quad \forall k \in \mathcal{K}_U,\ i \in \mathcal{K}_R \tag{28e}$$

$$u_{k,i} \le a_{k,i} P_i, \quad \forall k \in \mathcal{K}_U,\ i \in \mathcal{K}_R \tag{28f}$$

$$a_{k,i} \in \{0,1\}, b_i \in \{0,1\}, \quad \forall k \in \mathcal{K}_U,\ i \in \mathcal{K}_R \tag{28g}$$

$$\sum_{k\in\mathcal{K}_U} u_{k,i} \le P_i, \quad \forall i \in \mathcal{K}_R, \tag{28h}$$

where $\boldsymbol{w} \triangleq (\boldsymbol{R}, \bar{\boldsymbol{F}}, z, \boldsymbol{u}, \boldsymbol{f}, \boldsymbol{r})$; $\boldsymbol{u} \triangleq \{u_{k,i}\}_{k\in\mathcal{K}_U, i\in\mathcal{K}_R}$; (19) and (28e)–(28g) follow from (16) and (17), which guarantees no power is consumed if $a_{k,i} = 0$; (28h) is indeed (23c) via (28e). Still, problem (28) is challenging due to the nonconvex constraints (7), (14), (28c), (28d) and (28g).

To deal with the binary nature of constraint (28g), we note that for a real number x,

$$x \in \{0,1\} \Leftrightarrow x - x^2 = 0 \Leftrightarrow (x \in [0,1]) \ \&\ (x - x^2 \le 0). \tag{29}$$

Therefore, (28g) can be rewritten as

$$\sum_{i\in\mathcal{K}_R} \sum_{k\in\mathcal{K}_U} (a_{k,i} - a_{k,i}^2) + \sum_{i\in\mathcal{K}_R} (b_i - b_i^2) \le 0 \tag{30}$$

$$0 \le a_{k,i} \le 1, \quad 0 \le b_i \le 1,\ \forall k \in \mathcal{K}_U,\ i \in \mathcal{K}_R\,. \tag{31}$$

With (30) and (31), (28d) also becomes a convex constraint. Problem (28) is now transformed to the following problem with continuous variables $a_{k,i}, b_i \in [0,1], \forall k \in \mathcal{K}_U$, $i \in \mathcal{K}_R$:

$$\min_{(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})\in\mathcal{H}} \ -t \tag{32}$$

where $\mathcal{H} \triangleq \{(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b}) | (5) - (10),(14),(19),(23\text{d}),(28\text{c}) - (28\text{f}),(28\text{h}),(30),(31)\}$.

Without including the nonconvex constraint (30), let $\widehat{\mathcal{H}} \triangleq \{(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b}) | (5) - (10),(14),(19),(23\text{d}),(28\text{c}) - (28\text{f}),(28\text{h}),(31)\}$ be the compact, feasible set of problem (32). The Lagrangian of (32) is given as

$$\mathcal{L}(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b},\lambda) \triangleq -t + \lambda\Big(\sum_{i\in\mathcal{K}_R}\sum_{k\in\mathcal{K}_U} (a_{k,i} - a_{k,i}^2) + \sum_{i\in\mathcal{K}_R} (b_i - b_i^2)\Big), \tag{33}$$

where $\lambda \ge 0$ is the Lagrangian multiplier to handle the nonconvex constraint (30). Problem (32) can then be expressed as $\min_{(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}} \max_{\lambda\ge 0} \mathcal{L}(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b},\lambda)$ and its dual problem as $\sup_{\lambda\ge 0} \min_{(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}} \mathcal{L}(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b},\lambda)$. The property of (32) is stated in the following result.

Proposition 1: Strong Lagrangian duality holds for problem (32), i.e.,

$$\min_{(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}} \max_{\lambda\ge 0} \mathcal{L}(t,\boldsymbol{w}\,\boldsymbol{a},\boldsymbol{b},\lambda) = \sup_{\lambda\ge 0} \min_{(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}} \mathcal{L}(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b},\lambda). \tag{34}$$

Problem (32) is thus equivalent to the following problem

$$\min_{(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}} \mathcal{L}(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b},\lambda) = -t + \lambda\Big(\sum_{i\in\mathcal{K}_R}\sum_{k\in\mathcal{K}_U} (a_{k,i} - a_{k,i}^2) + \sum_{i\in\mathcal{K}_R} (b_i - b_i^2)\Big) \tag{35}$$

at the optimal $\lambda^* \ge 0$ of the sup-min problem in (34).

Proof: Thanks to a special property of the Lagrangian function (33) of problem (32), i.e., $S = \big(\sum_{i\in\mathcal{K}_R}\sum_{k\in\mathcal{K}_U}(a_{k,i} - a_{k,i}^2) + \sum_{i\in\mathcal{K}_R}(b_i - b_i^2)\big)$ is always nonnegative, it can be proved that S must be zero at an optimal solution, which leads to the same optimum obtained from problem (32) and its dual problem. The detailed proof can be found in Appendix A. ∎

Proposition 1 implies that the optimal solution of problem (32) can be found by solving problem (35) for an appropriately chosen value of λ.

To deal with constraints (7) and (28c), we rewrite them respectively as

$$(R_k + a_{k,i})^2 - (R_k - a_{k,i})^2 - 4\sum_{n\in\mathcal{I}_i^{\mathcal{K}_R}} f_{k,i,n} \le 0, \quad \forall k \in \mathcal{K}_U,\ i \in \mathcal{K}_R \quad (36)$$

$$(t+z)^2 - (t-z)^2 - 4\sum_{k\in\mathcal{K}_U} R_k \le 0. \quad (37)$$

Note that a function $f_1(x,y) \triangleq (x-y)^2$ is jointly convex in (x,y). Upon applying the first-order Taylor series expansion at a given point $(x^{(\kappa)}, y^{(\kappa)})$, its convex lower bound is given as $2(x^{(\kappa)}-y^{(\kappa)})(x-y)-(x^{(\kappa)}-y^{(\kappa)})^2 \le (x-y)^2$. Therefore, constraints (36) and (37) can be approximated at a given point $(t^{(\kappa)}, \boldsymbol{w}^{(\kappa)}, \boldsymbol{a}^{(\kappa)}, \boldsymbol{b}^{(\kappa)})$ by the convex constraints

$$(R_k + a_{k,i})^2 - 2(R_k^{(\kappa)} - a_{k,i}^{(\kappa)})(R_k - a_{k,i}) + (R_k^{(\kappa)} - a_{k,i}^{(\kappa)})^2 - 4\sum_{n\in\mathcal{I}_i^{\mathcal{K}_R}} f_{k,i,n} \le 0, \quad \forall k \in \mathcal{K}_U,\ i \in \mathcal{K}_R \quad (38)$$

$$(t+z)^2 - 2(t^{(\kappa)} - z^{(\kappa)})(t-z) + (t^{(\kappa)} - z^{(\kappa)})^2 - 4\sum_{k\in\mathcal{K}_U} R_k \le 0 \quad (39)$$

in the sense that every point $(t, \boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b})$ that satisfies constraints (38) and (39) would also satisfy constraints (36) and (37).

To deal with constraint (14), it is observed that the nonconvex function $g_k(\bar{\boldsymbol{F}})$ in (14) has a concave lower bound $\Gamma_k^{(\kappa)}(\bar{\boldsymbol{F}})$ at a specific point $\bar{\boldsymbol{F}}^{(\kappa)}$ as (40), shown at the bottom of this page, where $\boldsymbol{\Phi}_k \triangleq \boldsymbol{\Pi}_k \boldsymbol{\Pi}_k^H + \boldsymbol{\Xi}_k$. The derivation of $\Gamma_k^{(\kappa)}(\bar{\boldsymbol{F}})$ in (40) and the proof of its concavity follows from the results of [40] and thus are omitted for brevity. Constraint (14) can then be approximated at a given point $\bar{\boldsymbol{F}}^{(\kappa)}$ by the following convex constraint

$$R_k \le \Gamma_k^{(\kappa)}(\bar{\boldsymbol{F}}), \quad \forall k \in \mathcal{K}_U\,. \quad (41)$$

To deal with the nonconvex cost function $\mathcal{L}(t, \boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}, \lambda)$ of (35), we note that $a_{k,i}^2 \ge 2a_{k,i}^{(\kappa)} a_{k,i} - (a_{k,i}^{(\kappa)})^2$ and $b_i^2 \ge 2b_i^{(\kappa)} b_i - (b_i^{(\kappa)})^2$. Therefore, $\mathcal{L}(t, \boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}, \lambda)$ has a convex upper bound $\tilde{\mathcal{L}}(t, \boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}, \lambda)$ at a given point $(t^{(\kappa)}, \boldsymbol{w}^{(\kappa)}, \boldsymbol{a}^{(\kappa)}, \boldsymbol{b}^{(\kappa)})$ as

$$\begin{aligned}\tilde{\mathcal{L}}(t, \boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}, \lambda) &\triangleq -t + \lambda\Big(\sum_{k\in\mathcal{K}_U}\sum_{i\in\mathcal{K}_R}\Big((1-2a_{k,i}^{(\kappa)})a_{k,i} + (a_{k,i}^{(\kappa)})^2\Big) + \sum_{i\in\mathcal{K}_R}\Big((1-2b_i^{(\kappa)})b_i + (b_i^{(\kappa)})^2\Big)\Big) \\ &\ge \mathcal{L}(t, \boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}, \lambda). \end{aligned} \quad (42)$$

Algorithm 1 Energy Efficiency Maximization for Downlink C-RANs Under Data-Sharing Strategy

1: **Initialization**: Set $\kappa := 1$. Choose a value of λ and choose an initial point $(t^{(0)}, \boldsymbol{w}^{(0)}, \boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)})$ by Subroutine 1.
2: **repeat**
3: Update $\kappa := \kappa + 1$
4: Find the optimal solution $(t^*, \boldsymbol{w}^*, \boldsymbol{a}^*, \boldsymbol{b}^*)$ by solving convex problem (43)
5: Update $(t^{(\kappa)}, \boldsymbol{w}^{(\kappa)}, \boldsymbol{a}^{(\kappa)}, \boldsymbol{b}^{(\kappa)}) := (t^*, \boldsymbol{w}^*, \boldsymbol{a}^*, \boldsymbol{b}^*)$
6: **until** convergence

As such, for a given point $(t^{(\kappa)}, \boldsymbol{w}^{(\kappa)}, \boldsymbol{a}^{(\kappa)}, \boldsymbol{b}^{(\kappa)})$, problem (35) can be further approximated by the following *convex* problem

$$\min_{(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}^{(\kappa)}} \widetilde{\mathcal{L}}(t, \boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}, \lambda) \quad (43)$$

where $\widehat{\mathcal{H}}^{(\kappa)} \triangleq \{(t, \boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}) \mid (5), (6), (8)-(10), (19), (23d), (28d)-(28f), (28h), (31), (38), (39), (41)\}$ is a convex feasible set.

Now, we are ready to outline the steps to find the solution of problem (35) in Algorithm 1. For an empirically chosen λ and starting from a feasible initial point, we solve problem (43) to obtain the optimal solution $(t^*, \boldsymbol{w}^*, \boldsymbol{a}^*, \boldsymbol{b}^*)$. This solution is then used as an initial point for the next iteration. The loop terminates when there is no improvement in the objective function $\tilde{\mathcal{L}}$ of problem (43). The convergence property of Algorithm 1 is presented in the following proposition.

Proposition 2: Algorithm 1 converges to a solution[5] that satisfies the Fritz John conditions of problem (35).

Proof: See Appendix B. ∎

The initial point $(t^{(0)}, \boldsymbol{w}^{(0)}, \boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}) \in \widehat{\mathcal{H}}$ of Algorithm 1 can be found by Subroutine 1. Here, Subroutine 1 solves problem (32) without constraint (30), which can be approximated by the *convex* problem

$$\min_{(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}^{(\kappa)}} -t. \quad (44)$$

Starting from a random point $(t^{(0)}, \boldsymbol{w}^{(0)}, \boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}) \in \widehat{\mathcal{H}}$, the initial point obtained by Subroutine 1 is located close to a solution of problem (32). Since (32) and (35) are equivalent, the initial point obtained by Subroutine 1 will improve the solution obtained by solving (43) which is an approximation of (35).

[5]In our numerical experiments, since the values of $\boldsymbol{a}$ and $\boldsymbol{b}$ obtained by Algorithm 1 do often satisfy (30), the corresponding solution is feasible to (32). If it is not the case, we will then rerun Algorithm 1. In all our simulations, it is observed that at most two times of doing this are sufficient to have $\boldsymbol{a}$ and $\boldsymbol{b}$ satisfying (30).

$$\Gamma_k^{(\kappa)}(\bar{\boldsymbol{F}}) \triangleq g_k(\bar{\boldsymbol{F}}^{(\kappa)}) + \frac{2W}{\ln 2}\Re\left\{\left\langle\left(\left(\boldsymbol{\Phi}_k^{(\kappa)} - \boldsymbol{\Pi}_k^{(\kappa)}(\boldsymbol{\Pi}_k^{(\kappa)})^H\right)^{-1}\boldsymbol{\Pi}_k^{(\kappa)}\right)^H(\boldsymbol{\Pi}_k(\bar{\boldsymbol{F}}_k) - \boldsymbol{\Pi}_k^{(\kappa)})\right\rangle\right\} - \frac{W}{\ln 2}\left\langle\left(\left(\boldsymbol{\Phi}_k^{(\kappa)} - \boldsymbol{\Pi}_k^{(\kappa)}(\boldsymbol{\Pi}_k^{(\kappa)})^H\right)^{-1} - (\boldsymbol{\Phi}_k^{(\kappa)})^{-1}\right)^H(\boldsymbol{\Phi}_k(\bar{\boldsymbol{F}}) - \boldsymbol{\Phi}_k^{(n)})\right\rangle \le g_k(\bar{\boldsymbol{F}}) \quad (40)$$

Subroutine 1 Finding an Initial Point for Algorithm 1

1: **Initialization**: Set $\kappa := 1$ and randomly select a point $(t^{(0)}, \boldsymbol{w}^{(0)}, \boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}) \in \widehat{\mathcal{H}}$
2: **repeat**
3: Update $\kappa := \kappa + 1$
4: Find the optimal solution $(t^*, \boldsymbol{w}^*, \boldsymbol{a}^*, \boldsymbol{b}^*)$ by solving convex problem (44)
5: Update $(t^{(\kappa)}, \boldsymbol{w}^{(\kappa)}, \boldsymbol{a}^{(\kappa)}, \boldsymbol{b}^{(\kappa)}) := (t^*, \boldsymbol{w}^*, \boldsymbol{a}^*, \boldsymbol{b}^*)$
6: **until** convergence

V. Problem Formulation for Fronthaul Transfer With Compression-Based Strategy

Unlike the data-sharing strategy described in Section III, the compression-based strategy aims to transfer a lesser amount of data over the fronthaul network by compressing the signal data. To do this, after precoding the messages the BBU employs lossy source coding techniques to compress the precoded signals. The compressed signals are then unicast to each RRH via the fronthaul network. The RRHs simply perform data decompression and transmit the decompressed signals to the users over radio access links [3], [15], [16], [30]. In what follows, we will describe each stage of the communication in detail.

A. Signal Beamforming, Data Compression and Unicasting at BBU

Similar to (11), the BBU generates the beamformed signal for an RRH i as

$$\bar{\boldsymbol{x}}_i = \sum_{k \in \mathcal{K}_U} \boldsymbol{F}_{k,i}\, \boldsymbol{s}_k\,. \tag{45}$$

Next, it compresses the beamformed signal $\bar{\boldsymbol{x}}_i$ of block length u into a binary string at a rate R_i^C by selecting a codeword from a quantization codebook $\mathcal{C}_i$ of size $2^{uR_i^C}$. It then sends this binary string to the RRH i via the fronthaul network. We assume that the compression process is implemented independently for each RRH [15]. The case of multivariate compression for beamformed signals at all RRHs [30] can be dealt with in a similar manner. The compression process is information-theoretically modeled by a Gaussian *test channel* [41] in which the decompressed signal at the RRH i can be expressed as

$$\boldsymbol{x}_i = \bar{\boldsymbol{x}}_i + \boldsymbol{q}_i, \tag{46}$$

where the quantization noise $\boldsymbol{q}_i \in \mathbb{C}^{N_r \times 1}$ is independent of $\bar{\boldsymbol{x}}_i$ and randomly distributed according to $\boldsymbol{q}_i \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{Q}_i)$. For the given quantization error covariance matrix $\boldsymbol{Q}_i \in \mathbb{C}^{N_r \times N_r}$, a compression strategy is considered successful if the BBU can find from the quantization codebook $\mathcal{C}_i$ a codeword $\boldsymbol{x}_i$ that is jointly typical[6] with the codeword $\bar{\boldsymbol{x}}_i$. For a sufficiently large block length u, this event happens with an arbitrarily large probability if the following condition is satisfied [41]:

$$\begin{aligned}\theta_i(\bar{\boldsymbol{F}}, \boldsymbol{Q}_i) &\triangleq WI(\boldsymbol{x}_i; \bar{\boldsymbol{x}}_i)\\ &= W \log_2 \left| \sum_{k \in \mathcal{K}_U} \boldsymbol{F}_{k,i} \boldsymbol{F}_{k,i}^H + \boldsymbol{Q}_i \right| - W \log_2 |\boldsymbol{Q}_i|\\ &= W \log_2 \left| \sum_{k \in \mathcal{K}_U} \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i + \boldsymbol{Q}_i \right|\\ &\quad - W \log_2 |\boldsymbol{Q}_i| \le R_i^C, \quad \forall i \in \mathcal{K}_R, \end{aligned} \tag{47}$$

where $\boldsymbol{F}_{k,i} = \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_k$. It should be noted that in order to decompress the compressed signals, the RRHs need not know the precoding matrices and the channel codebooks $\mathcal{C}_k^{\mathrm{CH}}, \forall k \in \mathcal{K}_U$, of the users' intended messages. They only need to know the quantization codebooks $\mathcal{C}_i, \forall i \in \mathcal{K}_R$, of the compressed signals.

Unlike the data-sharing strategy, the compression-based strategy does not require signal multicasting. The BBU simply unicasts a compressed version $\bar{\boldsymbol{x}}_i$ of the beamformed signal to its destination, i.e., RRH i. In total, there are K_R unicast sessions delivered to K_R RRHs via the fronthaul. Denote by $d_{i,n}$ the flow rate of a unicast session $i \in \mathcal{K}_R$ on a link $n \in \mathcal{N}$. To guarantee the unicast session $i \in \mathcal{K}_R$ delivers $\bar{\boldsymbol{x}}_i$ to the RRH i at a rate R_i^C, the following condition must be met:

$$R_i^C \le \sum_{n \in \mathcal{I}_i^{\mathcal{K}_R}} d_{i,n}. \tag{48}$$

By combining (47) and (48), the routing constraints for K_R unicast sessions are formulated as [15]

$$\theta_i(\bar{\boldsymbol{F}}, \boldsymbol{Q}_i) \le \sum_{n \in \mathcal{I}_i^{\mathcal{K}_R}} d_{i,n}, \forall i \in \mathcal{K}_R \tag{49}$$

$$\sum_{n \in \mathcal{O}_m^{\mathcal{M}}} d_{i,n} = \sum_{n \in \mathcal{I}_m^{\mathcal{M}}} d_{i,n}, \forall i \in \mathcal{K}_R, m \in \mathcal{M} \tag{50}$$

$$\sum_{i \in \mathcal{K}_R} d_{i,n} \le C_n, \quad \forall n \in \mathcal{N} \tag{51}$$

$$R_k \ge R_{\mathrm{QoS}}, \quad d_{i,n} \ge 0, \ \forall k \in \mathcal{K}_U, \ i \in \mathcal{K}_R, \ n \in \mathcal{M} \tag{52}$$

$$d_{i,n} \le b_i C_n, \quad \forall i \in \mathcal{K}_R, \ n \in \mathcal{M}\,. \tag{53}$$

Constraint (50) implements the law of flow conservation at each router. Constraint (51) makes sure that the total information flow on link $n \in \mathcal{N}$ for all K_R unicast sessions does not exceed the link capacity. Constraint (52) guarantees nonnegative flow rates for all unicast sessions on all fronthaul links. Constraint (53) means no data transmission to inactive RRHs.

B. Transmission of Decompressed Signals by RRHs

Now, the received signal $\boldsymbol{y}_k \in \mathbb{C}^{N_r \times 1}$ at a user k can be written as

$$\begin{aligned}\boldsymbol{y}_k &= \sum_{i \in \mathcal{K}_R} \boldsymbol{H}_{k,i} \boldsymbol{x}_i + \boldsymbol{n}_k,\\ &= \sum_{i \in \mathcal{K}_R} \boldsymbol{H}_{k,i} \boldsymbol{F}_{k,i} \boldsymbol{s}_k + \sum_{i \in \mathcal{K}_R} \sum_{\ell \in \mathcal{K}_U \setminus \{k\}} \boldsymbol{H}_{\ell,i} \boldsymbol{F}_{\ell,i} \boldsymbol{s}_\ell\\ &\quad + \sum_{i \in \mathcal{K}_R} \boldsymbol{H}_{k,i} \boldsymbol{q}_i + \boldsymbol{n}_k \end{aligned} \tag{54}$$

[6]Two codewords (or sequences) X_1 and X_2 of block length u are jointly typical with respect to a distribution $p(x_1, x_2)$ if their joint empirical distribution does not deviate much from $p(x_1, x_2)$ as $u \to \infty$ [41].

$$= \boldsymbol{H}_k \bar{\boldsymbol{F}}_k \boldsymbol{s}_k + \underbrace{\sum_{\ell \in \mathcal{K}_U \setminus \{k\}} \boldsymbol{H}_\ell \bar{\boldsymbol{F}}_\ell \boldsymbol{s}_\ell + \sum_{i \in \mathcal{K}_R} \boldsymbol{H}_{k,i} \boldsymbol{q}_i}_{\text{interference}} + \boldsymbol{n}_k \,. \tag{55}$$

By treating the interference term in (54) as additive Gaussian noise, the data rate R_k of the message symbol $\boldsymbol{s}_k$ is always achievable in the Shannon's sense as follows

$$R_k \le h_k(\bar{\boldsymbol{F}}, \boldsymbol{Q}) \triangleq W \log_2 \left| \boldsymbol{I}_{N_u} + \boldsymbol{\Pi}_k \boldsymbol{\Pi}_k^H \boldsymbol{\Omega}_k^{-1} \right|, \tag{56}$$

where $\boldsymbol{Q} \triangleq \{\boldsymbol{Q}_i\}_{i \in \mathcal{K}_R}$; $\boldsymbol{\Omega}_k \triangleq \sum_{\ell \in \mathcal{K}_U \setminus \{k\}} \boldsymbol{H}_k \bar{\boldsymbol{F}}_\ell \bar{\boldsymbol{F}}_\ell^H \boldsymbol{H}_k^H + \sum_{i \in \mathcal{K}_R} \boldsymbol{H}_{k,i} \boldsymbol{Q}_i \boldsymbol{H}_{k,i}^H + \boldsymbol{\Sigma}_k = \boldsymbol{\Xi}_k + \boldsymbol{\Upsilon}_k$, and $\boldsymbol{\Upsilon}_k \triangleq \sum_{i \in \mathcal{K}_R} \boldsymbol{H}_{k,i} \boldsymbol{Q}_i \boldsymbol{H}_{k,i}^H$. The relation in (56) implies that the achievable rate R_k depends on the quantization error covariance matrices $\boldsymbol{Q}$.

C. Power Consumption

Here, the transmit power at an RRH i is expressed as

$$\begin{aligned} P_i^{\text{Tx,C}} &= b_i \Big(\sum_{k \in \mathcal{K}_U} a_{k,i} \langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i \rangle + \langle \boldsymbol{Q}_i \rangle \Big) \\ &= \sum_{k \in \mathcal{K}_U} \langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i \rangle + b_i \langle \boldsymbol{Q}_i \rangle. \end{aligned} \tag{57}$$

Similar to (21), the power consumption on a fronthaul link n in this case is modeled as

$$P_n^{\text{FH,C}} \triangleq \alpha_n \sum_{i \in \mathcal{K}_R} d_{i,n}. \tag{58}$$

From (4), (57) and (58), the total power consumption of the network under the compression-based strategy is

$$\begin{aligned} &P_{\text{total}}^{\text{C}}(\bar{\boldsymbol{F}}, \boldsymbol{Q}, \boldsymbol{d}, \boldsymbol{b}) \\ &\triangleq \sum_{i \in \mathcal{K}_R} P_i^{\text{RRH,C}} + \sum_{n \in \mathcal{N}} P_n^{\text{FH,C}} \\ &= \sum_{i \in \mathcal{K}_R} \left[\beta_i P_i^{\text{Tx,C}} + b_i P_{i,\Delta} \right] + \sum_{n \in \mathcal{N}} \alpha_n \sum_{i \in \mathcal{K}_R} d_{i,n} + P_s \\ &= \sum_{i \in \mathcal{K}_R} \left[\beta_i \left(\sum_{k \in \mathcal{K}_U} \langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i \rangle, + b_i \langle \boldsymbol{Q}_i \rangle \right) + b_i P_{i,\Delta} \right] \\ &\quad + \sum_{n \in \mathcal{N}} \alpha_n \sum_{i \in \mathcal{K}_R} d_{i,n} + P_s, \end{aligned} \tag{59}$$

where $\boldsymbol{d} \triangleq \{d_{i,n}\}_{i \in \mathcal{K}_R, n \in \mathcal{N}}$, $P_{i,\Delta} \triangleq P_{i,a} - P_{i,s}$, and $P_s \triangleq \sum_{i \in \mathcal{K}_R} P_{i,s}$.

D. Problem Formulation

Similar to problem (23), the energy efficiency maximization problem for the multi-hop C-RAN under the compression-based strategy is formulated as

$$\max_{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{R}, \bar{\boldsymbol{F}}, \boldsymbol{Q}, \boldsymbol{d}} \quad \mathcal{P}_1^{\text{C}} \triangleq \frac{R_{\text{sum}}}{P_{\text{total}}^{\text{C}}} \tag{60a}$$

$$\text{s.t. } (19), (23\text{d}), (49) - (53), (56) \tag{60b}$$

$$\sum_{k \in \mathcal{K}_U} \langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i \rangle + b_i \langle \boldsymbol{Q}_i \rangle \le P_i, \quad \forall i \in \mathcal{K}_R, \tag{60c}$$

where R_{sum} and $P_{\text{total}}^{\text{DS}}$ are respectively defined in (1) and (59), and (60c) is the per-RRH power constraint (12) via (11) and (57).

The problem formulation for a single-hop C-RAN is obtained from (60) by replacing (49)–(53) with the following constraints

$$\theta_i(\bar{\boldsymbol{F}}, \boldsymbol{Q}_i) \le d_i, \quad \forall i \in \mathcal{K}_R \tag{61}$$

$$R_k \ge R_{\text{QoS}}, \quad d_i \ge 0, \ \forall k \in \mathcal{K}_U, \ i \in \mathcal{K}_R, \tag{62}$$

$$d_i \le b_i C_i, \quad \forall i \in \mathcal{K}_R, \ n \in \mathcal{M}, \tag{63}$$

where d_i is the data rate of a unicast session i from the BBU to an RRH i. Since the structure of (61)-(63) is similar to that of (49)–(53), the algorithm devised in the next section for (60) can be straightforwardly adapted to solve this corresponding single-hop problem as well.

VI. Proposed Algorithm for Fronthaul Compression-Based Strategy

Problem (60) can be rewritten in the following epigraph form

$$\max_{t, \boldsymbol{p}, \boldsymbol{a}, \boldsymbol{b}} \ t \tag{64a}$$

$$\text{s.t. } (19), (23\text{d}), (28\text{c}), (28\text{e}) - (28\text{g}), (49) - (53), (56) \tag{64b}$$

$$z \ge \sum_{i \in \mathcal{K}_R} \Big(\beta_i \Big(\sum_{k \in \mathcal{K}_U} \langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i \rangle + \eta_i \Big) + b_i P_{i,\Delta} \Big) + \sum_{n \in \mathcal{N}} \alpha_n \sum_{i \in \mathcal{K}_R} d_{i,n} + P_s \tag{64c}$$

$$b_i \langle \boldsymbol{Q}_i \rangle \le \eta_i, \quad \forall i \in \mathcal{K}_R \tag{64d}$$

$$\sum_{k \in \mathcal{K}_U} u_{k,i} \le \rho_i, \quad \forall k \in \mathcal{K}_U, \ i \in \mathcal{K}_R \tag{64e}$$

$$\rho_i + \eta_i \le P_i, \quad \forall i \in \mathcal{K}_R \tag{64f}$$

$$\eta_i \le b_i P_i, \quad \forall i \in \mathcal{K}_R, \tag{64g}$$

where $\boldsymbol{p} \triangleq (\boldsymbol{R}, \bar{\boldsymbol{F}}, \boldsymbol{Q}, z, \boldsymbol{u}, \boldsymbol{\eta}, \boldsymbol{\rho}, \boldsymbol{d})$; $\boldsymbol{\eta} \triangleq \{\eta_i\}_{i \in \mathcal{K}_R}$; $\boldsymbol{\rho} \triangleq \{\rho_i\}_{i \in \mathcal{K}_R}$; (64f) is indeed (60c) via (64d) and (64e). Problem (64) is still challenging due to the nonconvex constraints (28c), (28g), (49), (56), (64c) and (64d).

Similar to the solution framework proposed in Section IV, with (28g) replaced by (30) and (31), (28d) becomes a convex constraint. Problem (64) is now transformed to the following problem with continuous variables $a_{k,i}, b_i \in [0,1], \forall k \in \mathcal{K}_U, i \in \mathcal{K}_R$:

$$\min_{(t, \boldsymbol{p}, \boldsymbol{a}, \boldsymbol{b}) \in \mathcal{V}} \ -t \tag{65}$$

where $\mathcal{V} \triangleq \{(t, \boldsymbol{p}, \boldsymbol{a}, \boldsymbol{b}) \mid (19), (23\text{d}), (28\text{c}), (28\text{e}) - (28\text{g}), (30), (31), (49) - (53), (56), (64\text{c}) - (64\text{g})\}$. For an appropriately chosen value of λ, the solution of problem (65) can now be found by solving the following equivalent problem

$$\min_{(t, \boldsymbol{p}, \boldsymbol{d}, \boldsymbol{a}, \boldsymbol{b}) \in \widehat{\mathcal{V}}} \mathcal{L}(t, \boldsymbol{p}\, \boldsymbol{a}, \boldsymbol{b}, \lambda) = -t + \lambda \Big(\sum_{i \in \mathcal{K}_R} \sum_{k \in \mathcal{K}_U} (a_{k,i} - a_{k,i}^2) + \sum_{i \in \mathcal{K}_R} (b_i - b_i^2) \Big) \tag{66}$$

where $\widehat{\mathcal{V}} \triangleq \{(t,\boldsymbol{p},\boldsymbol{a},\boldsymbol{b}) \mid (19),(23d),(28c),(28e)-(28g),(31),(49)-(53),(56),(64c)-(64g)\}$. The proof of this fact is similar to the proof of Proposition 1, and hence, omitted.

For continuous values of $\boldsymbol{a},\boldsymbol{b}$ satisfying (30) and (31), (28c) is approximated by the convex constraint (39), and (64c) also becomes a convex constraint. For (49), its nonconvex part $\varphi_i(\bar{\boldsymbol{F}},\boldsymbol{Q}_i) \triangleq \log_2\left|\sum_{k\in\mathcal{K}_U}\bar{\boldsymbol{E}}_i^H\bar{\boldsymbol{F}}_k\bar{\boldsymbol{F}}_k^H\bar{\boldsymbol{E}}_i+\boldsymbol{Q}_i\right|$ has a convex upper bound $\tilde{\varphi}_i(\bar{\boldsymbol{F}},\boldsymbol{Q}_i)$ at a specific point $(\bar{\boldsymbol{F}}^{(\kappa)},\boldsymbol{Q}_i^{(\kappa)})$ as [15]

$$\begin{aligned}\varphi_i(\bar{\boldsymbol{F}},\boldsymbol{Q}_i) &\le \tilde{\varphi}_i(\bar{\boldsymbol{F}},\boldsymbol{Q}_i)\\ &\triangleq \varphi(\bar{\boldsymbol{F}}^{(\kappa)},\boldsymbol{Q}_i^{(\kappa)})\\ &\quad+\frac{1}{\ln 2}\Big\langle\sum_{k\in\mathcal{K}_U}\bar{\boldsymbol{F}}_k^H\bar{\boldsymbol{E}}_i(\boldsymbol{V}_i^{(\kappa)})^{-1}\bar{\boldsymbol{E}}_i^H\bar{\boldsymbol{F}}_k\Big\rangle\\ &\quad+\frac{1}{\ln 2}\left\langle(\boldsymbol{V}_i^{(\kappa)})^{-1}\boldsymbol{Q}_i-\boldsymbol{I}_{N_r}\right\rangle\end{aligned}\tag{67}$$

where $\boldsymbol{V}_i \triangleq \sum_{k\in\mathcal{K}_U}\bar{\boldsymbol{E}}_i^H\bar{\boldsymbol{F}}_k\bar{\boldsymbol{F}}_k^H\bar{\boldsymbol{E}}_i+\boldsymbol{Q}_i = \sum_{k\in\mathcal{K}_U}\boldsymbol{F}_{k,i}\boldsymbol{F}_{k,i}^H+\boldsymbol{Q}_i$. Moreover, $\log_2|\boldsymbol{Q}_i|$ is a concave function in $\boldsymbol{Q}_i\succeq 0$. At a given point $(\bar{\boldsymbol{F}}^{(\kappa)},\boldsymbol{Q}_i^{(\kappa)})$, (49) can thus be approximated by the following convex constraint

$$W\tilde{\varphi}_i(\bar{\boldsymbol{F}},\boldsymbol{Q}_i)-W\log_2|\boldsymbol{Q}_i|-\sum_{n\in\mathcal{I}_i^{\mathcal{K}_R}}d_{i,n}\le 0,\forall i\in\mathcal{K}_R\,.\tag{68}$$

Constraint (56) can be approximated at a given point $\bar{\boldsymbol{F}}^{(\kappa)}$ by the following convex constraint

$$R_k\le\bar{h}_k^{(\kappa)}(\bar{\boldsymbol{F}}),\quad\forall k\in\mathcal{K}_U,\tag{69}$$

where $\bar{h}_k^{(\kappa)}$ is the concave lower bound of the nonconvex function $h_k(\bar{\boldsymbol{F}},\boldsymbol{Q})$ of (56) as (70), shown at the bottom of this page, with $\boldsymbol{\Theta}_k \triangleq \boldsymbol{\Pi}_k\boldsymbol{\Pi}_k^H+\boldsymbol{\Omega}_k$. Constraint (64d) is approximated by

$$\begin{aligned}(b_i+\langle\boldsymbol{Q}_i\rangle)^2-2(b_i^{(\kappa)}-\langle\boldsymbol{Q}_i\rangle^{(\kappa)})(b_i-\langle\boldsymbol{Q}_i\rangle)+(b_i^{(\kappa)}\\ -\langle\boldsymbol{Q}_i\rangle^{(\kappa)})^2-4\eta_i\le 0.\end{aligned}\tag{71}$$

The cost function $\mathcal{L}(t,\boldsymbol{p},\boldsymbol{a},\boldsymbol{b},\lambda)$ of problem (66) can be approximated by its convex upper bound $\widetilde{\mathcal{L}}(t,\boldsymbol{p},\boldsymbol{a},\boldsymbol{b},\lambda)$ in (42).

As such, at a given point $(t^{(\kappa)},\boldsymbol{p}^{(\kappa)},\boldsymbol{a}^{(\kappa)},\boldsymbol{b}^{(\kappa)})$, problem (66) can thus be approximated by the *convex* problem

$$\min_{(t,\boldsymbol{p},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{V}}^{(\kappa)}}\ \widetilde{\mathcal{L}}(t,\boldsymbol{p},\boldsymbol{a},\boldsymbol{b},\lambda),\tag{72}$$

Algorithm 2 Energy Efficiency Maximization of the Downlink C-RANs Under Compression-Based Strategy

1: **Initialization**: Set $\kappa:=1$. Choose a value of λ and choose an initial point $(t^{(0)},\boldsymbol{p}^{(0)},\boldsymbol{a}^{(0)},\boldsymbol{b}^{(0)})$ by Subroutine 2
2: **repeat**
3: Update $\kappa:=\kappa+1$
4: Find the optimal solution $(t^*,\boldsymbol{p}^*,\boldsymbol{a}^*,\boldsymbol{b}^*)$ by solving convex problem (72)
5: Update $(t^{(\kappa)},\boldsymbol{p}^{(\kappa)},\boldsymbol{a}^{(\kappa)}\,\boldsymbol{b}^{(\kappa)}):=(t^*,\boldsymbol{p}^*,\boldsymbol{a}^*,\boldsymbol{b}^*)$
6: **until** convergence

Subroutine 2 Find an Initial Point for Algorithm 2

1: **Initialization**: Set $\kappa:=1$ and choose randomly a point $(t^{(0)},\tilde{\boldsymbol{p}}^{(0)},\boldsymbol{a}^{(0)},\boldsymbol{b}^{(0)})\in\widehat{\mathcal{V}}$
2: **repeat**
3: Update $\kappa:=\kappa+1$
4: Find the optimal solution $(t^*,\boldsymbol{p}^*,\boldsymbol{a}^*,\boldsymbol{b}^*)$ by solving convex problem (73)
5: Update $(t^{(\kappa)},\boldsymbol{p}^{(\kappa)},\boldsymbol{a}^{(\kappa)}\,\boldsymbol{b}^{(\kappa)}):=(t^*,\boldsymbol{p}^*,\boldsymbol{a}^*,\boldsymbol{b}^*)$
6: **until** convergence

where $\widehat{\mathcal{V}}^{(\kappa)} \triangleq \{(t,\boldsymbol{p},\boldsymbol{a},\boldsymbol{b})|(19),(23d),(39),(31),(50)-(52),(64c),(64f),(68),(69),(71)\}$ is the compact, convex feasible set of problem (72).

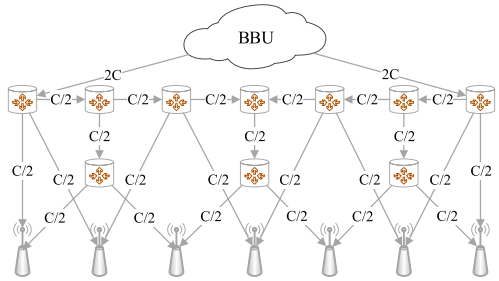
Algorithm 2 outlines the steps to solve problem (66) for maximizing the energy efficiency of the downlink C-RANs under the compression-based strategy. For an empirically chosen λ and starting from a feasible initial point, we solve problem (72) to obtain the optimal solution. This solution is then used as an initial point for the next iteration. The loop terminates when there is no improvement in the objective function $\widetilde{\mathcal{L}}$ of (72). Algorithm 2 converges to a Fritz John solution of problem (66). The proof of this fact is similar to the proof of Proposition 2, and hence omitted.

The initial solution $(t^{(0)},\boldsymbol{p}^{(0)},\boldsymbol{a}^{(0)},\boldsymbol{b}^{(0)})\in\widehat{\mathcal{V}}$ of Alogrithm 2 is generated by Subroutine 2. Here, Subroutine 2 aims to solve problem (65) without constraint (30), which is then approximated by the problem

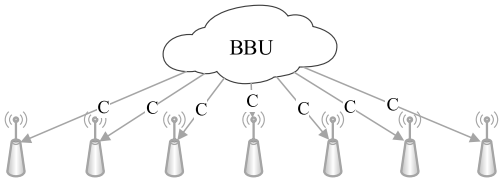
$$\min_{(t,\boldsymbol{p},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{V}}^{(\kappa)}}\ -t.\tag{73}$$

After taking a random point $(t^{(0)},\boldsymbol{p}^{(0)},\boldsymbol{a}^{(0)},\boldsymbol{b}^{(0)})\in\widehat{\mathcal{V}}$, Subroutine 2 obtains an initial point located close to a solution of problem (65). Since (65) and (66) are equivalent, this initial

$$\begin{aligned}\bar{h}_k^{(\kappa)}(\bar{\boldsymbol{F}},\boldsymbol{Q}) \triangleq h_k(\bar{\boldsymbol{F}}^{(\kappa)},\boldsymbol{Q}^{(\kappa)})&+\frac{2W}{\ln 2}\Re\left\{\left\langle\left(\left(\boldsymbol{\Theta}_k^{(\kappa)}-\boldsymbol{\Pi}_k^{(\kappa)}(\boldsymbol{\Pi}_k^{(\kappa)})^H\right)^{-1}\boldsymbol{\Pi}_k^{(\kappa)}\right)^H(\boldsymbol{\Pi}_k(\bar{\boldsymbol{F}}_k)-\boldsymbol{\Pi}_k^{(\kappa)})\right\rangle\right\}\\ &-\frac{W}{\ln 2}\left\langle\left(\left(\boldsymbol{\Theta}_k^{(\kappa)}-\boldsymbol{\Pi}_k^{(\kappa)}(\boldsymbol{\Pi}_k^{(\kappa)})^H\right)^{-1}-(\boldsymbol{\Theta}_k^{(\kappa)})^{-1}\right)^H(\boldsymbol{\Phi}_k(\bar{\boldsymbol{F}})-\boldsymbol{\Phi}_k^{(\kappa)})\right\rangle\\ &-\frac{W}{\ln 2}\left\langle\left(\left(\boldsymbol{\Theta}_k^{(\kappa)}-\boldsymbol{\Pi}_k^{(\kappa)}(\boldsymbol{\Pi}_k^{(\kappa)})^H\right)^{-1}-(\boldsymbol{\Theta}_k^{(\kappa)})^{-1}\right)^H(\boldsymbol{\Upsilon}_k(\boldsymbol{Q})-\boldsymbol{\Upsilon}_k^{(\kappa)})\right\rangle\le h_k(\bar{\boldsymbol{F}},\boldsymbol{Q})\end{aligned}\tag{70}$$



(a) A multi-hop fronthaul network with $M = 10$ routers and $L = 25$ fronthaul links



(b) A single-hop fronthaul network

Fig. 2. Multi-hop and single-hop fronthaul network simulation scenarios with $K_R = 7$ RRHs. In both cases, the total capacity of the incoming information flow to each RRH is C.

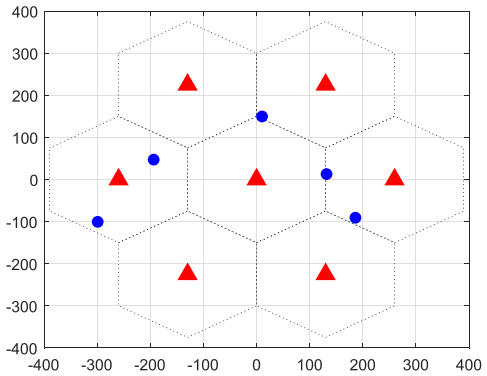


Fig. 3. A radio access network with $K_R = 7$ fixed RHHs and $K_U = 5$ randomly positioned users.

point will improve the solution obtained by solving (72), which is an inner approximation of (66).

VII. Numerical Examples

We consider a C-RAN with a multi-hop fronthaul in Fig. 2(a) and a single-hop fronthaul Fig. 2(b). The radio access part of the considered C-RAN is illustrated in Fig. 3. The locations of the $K_R = 7$ RRHs are fixed, while the $K_U = 5$ users are uniformly and independently placed within the RRHs' coverage area, excluding the circular area of 50 m around each RRH [27]. The LTE parameters used in our numerical examples are listed in Table II. Each RRH is assumed to be equipped with $N_r = 2$ antennas and each user with $N_u = 1$ antennas. The active mode and the sleep mode at each RRH consume 84W and 56W of power, respectively. The slope of transmit power is set as $\beta_i = \beta = 2.8$ and $\alpha_i = \alpha = 5$ for all $i \in \mathcal{K}_R$ [3]. We take $d = 1$, $P_i = P$, and $\mathbf{\Sigma}_k = \sigma^2 \boldsymbol{I}$ for all $k \in \mathcal{K}_U$.

TABLE II
LTE Parameters Used in Numerical Examples [42]

Parameters	Values
Distance between adjacent RRHs	0.3 km
Total bandwidth	10 MHz
Standard deviation of log-normal shadowing	10 dB
Path loss at distance d (km)	$140.7 + 36.7 \log_{10}(d)$ dB
Noise variance	-174 dBm/Hz
Maximum RRH transmit power	24 dBm

Here, we set $R_{\text{QoS}} = 0.1$Mbps for the feasibility of the problem (28) and (64). In our numerical experiments, increasing the required QoS value to around $\frac{C}{2K_U}$ does not affect simulation results. This is different from the case of power minimization where the QoS rate constraints are always met with equality at optimality. On the other hand, setting a higher QoS value may cause an infeasible problem or a smaller feasible set to which a random point in the initialization process has a smaller chance to belong, and it would take Subroutines 1 and 2 much more time to find feasible initial points. All the presented results have been averaged over 100 simulation trials. Our simulations confirm that this number of trials produces results with sufficient accuracy.

Intuitively, λ can be considered as a weight for the binary constraint (30). For a large value of λ, while converging quickly the proposed algorithms would likely give a worse solution compared to that by using a smaller value of λ. Conversely, for a small value of λ, we would expect a better solution but at a slow convergence speed. In this paper, we empirically choose λ for the best performance while keeping the convergence speed relatively fast, i.e., $\lambda = 100$.

To the best of our knowledge, there is no existing work that addresses the energy efficiency maximization for the system model considered in this paper. As such, we will demonstrate the advantages of the proposed algorithms by comparing them with some benchmark heuristic schemes derived from our proposed optimization approach. Specifically, to verify the effectiveness of the proposed algorithms under both the data-sharing and compression-based strategies in the multi-hop C-RANs (referred to as Alg. 1-DS-MH, Alg. 2-C-MH in the figures) and in the single-hop C-RANs (referred to as Alg. 1-DS-SH, Alg. 2-C-SH in the figures), we consider the following benchmark schemes:

- HUA-DS-SH: This scheme applies a heuristic UA scheme [15] for the data-sharing strategy in the single-hop C-RAN. Here, each user is heuristically assigned to N_c RRHs that have the largest channel gains, where N_c is empirically chosen for the best performance.
- HUA-C-SH: The HUA scheme of [15] is applied to the compression-based strategy in the single-hop C-RAN.
- HUA-DS-MH: The HUA scheme of [15] is used for the data-sharing strategy in the multi-hop C-RAN.
- HUA-C-MH: The HUA scheme [15] is used for the compression-based strategy in the multi-hop C-RAN.

For existing HUA solutions [15], the energy efficiency achieved by these HUA schemes above can be found by only jointly optimizing data rate allocation and signal precoding with the similar manners discussed in Section IV and VI.

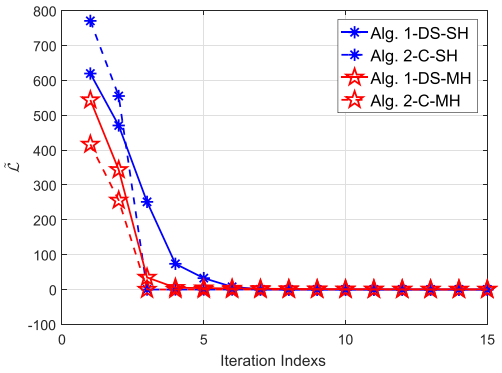
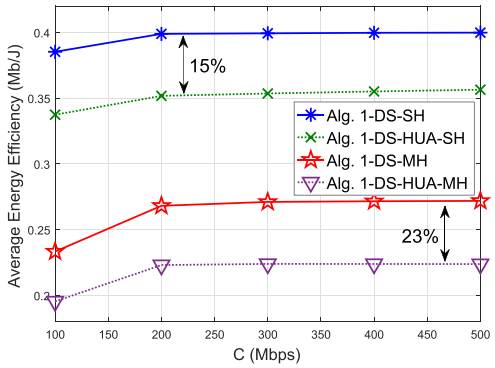
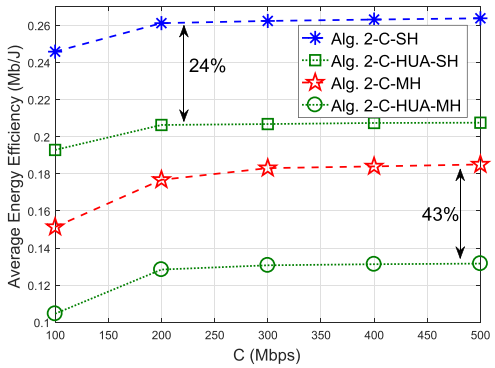


Fig. 4. Convergence process of the proposed Algorithms 1 and 2.

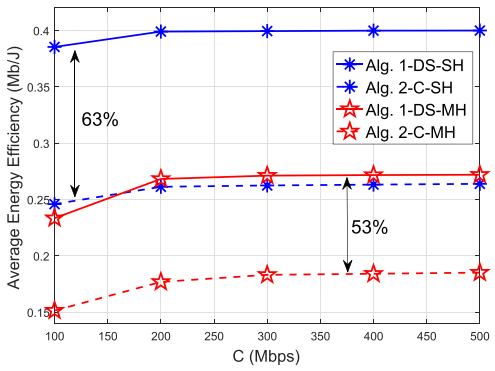


(a) Data-sharing strategy

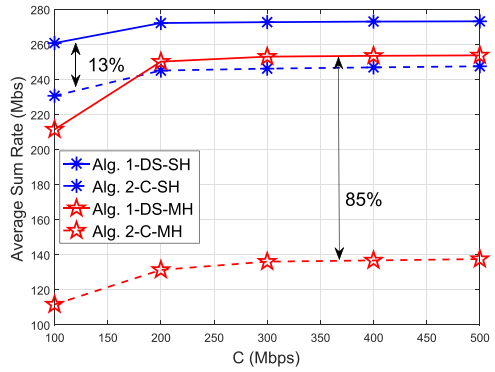


(b) Compression-based strategy

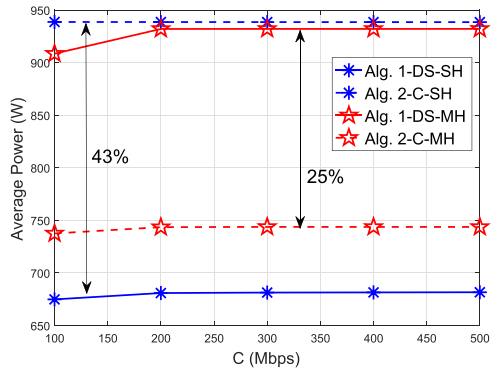
Fig. 5. Performance of the proposed algorithms in comparison with the benchmark schemes.



(a) Average energy efficiency



(b) Average sum rate



(c) Average total power

Fig. 6. Performance comparison between the data-sharing and compression-based strategy in the considered single-hop and multi-hop C-RANs.

Fig. 4 plots the convergence process of the proposed algorithms under both single-hop and multi-hop fronthaul cases. In the considered example, they converge quickly in fewer than 15 iterations. It should be emphasized that each iteration of the algorithms corresponds to solving only one simple convex program (43) or (72), which results in a low computational cost.

Fig. 5 answers the first question posed at the beginning of the paper "*(i) Can we do better than existing solutions in terms of energy efficiency?*" It shows that the proposed algorithms outperform the benchmark schemes in terms of the average energy efficiency in all cases. From Fig. 5(a), the improvements by Algorithm 1 in the data-sharing case are 15% and 23% in the single-hop and the multi-hop fronthaul scenarios, respectively. The gains offered by Algorithm 2 in the compression-based case are even more pronounced, 24% in the single-hop scenario and 43% in the multi-hop scenario as shown in Fig. 5(b). Such enhancement is brought about by the extra dimension of UA optimization in the proposed joint optimization algorithms compared to the benchmark schemes.

Under the proposed designs for energy efficiency maximization, Fig. 6 answers the second question "*(ii) With our devised solution, is data sharing or compression better?*" Fig. 6(a) confirms that the data-sharing strategy provides up to 63% and 53% energy-efficiency gains over the compression-based strategy in the single-hop and multi-hop fronthaul

cases, respectively. This fact can be explained by considering throughput in Fig. 6(b) and power consumption in Fig. 6(c). Specifically, it is clear from Fig. 6(b) that data sharing always outperforms compression in throughput achievement. This is because the compression may provide worse data rate than the data sharing due to the quantization noise (see (14) and (54)). The only time compression outperforms is with power consumption in the multi-hop fronthaul case—an improvement of 25% as illustrated in Fig. 6(c). However, such power saving amount is not enough to compensate the loss of 85% in throughput over the data sharing strategy as shown in Fig. 6(b). Finally, it can be concluded that from an energy-efficient perspective, data multicasting with/without network coding for multi-hop/single-hop fronthaul is always preferable over unicasting the compressed signal data over the fronthaul links.

In [3], the compression-based strategy is shown to consume less power than the data-sharing strategy with single-hop fronthaul when the data rate on fronthaul links is high. Here, power minimization is the optimization objective. However, Fig. 6(c) shows that compression actually requires 43% more power than data sharing in our considered scenario. This is because the current optimization solution tries to improve the throughput at the same time as saving power for the maximum energy efficiency.

VIII. Conclusion

This paper has jointly designed UA, RRH activation, data rate and signal precoding to maximize the energy efficiency of a downlink C-RAN, under both data-sharing and compression-based fronthaul strategies. The formulated mixed-integer nonlinear optimization problems take into account routing constraints at the fronthaul links, minimum data rate requirements, limited fronthaul capacity and maximum RRH transmit power. Using optimization techniques, new iterative algorithms with guaranteed convergence to Fritz John solutions have been proposed to solve the difficult problem formulations. Numerical results confirm the significant performance advantages of the developed solutions over baseline schemes. In both single-hop and multi-hop network examples, the data-sharing fronthaul strategy has been shown to offer much higher energy efficiency than does its compression-based counterpart.

Appendix A
Proof of Proposition 1

Denote by $\mathcal{F}(\lambda)$ and $(t_\lambda, \boldsymbol{w}_\lambda, \boldsymbol{a}_\lambda, \boldsymbol{b}_\lambda)$ the optimal value and the optimal solution of problem (35) for a given λ, respectively. In general, there always exists a duality gap between the optimal value of problem (32) and the optimal value of its dual problem, i.e.,

$$\begin{aligned}\sup_{\lambda\geq 0}\mathcal{F}(\lambda) &= \sup_{\lambda\geq 0}\min_{(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}}\mathcal{L}(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b},\lambda)\\ &\leq \mathcal{F}^* \triangleq \min_{(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}}\max_{\lambda\geq 0}\mathcal{L}(t,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b},\lambda). \quad (74)\end{aligned}$$

Here, $\mathcal{F}^*$ is the optimal value of problem (32). Since $\widehat{\mathcal{H}}$ is compact, $\mathcal{F}^*$ is finite, and by (74),

$$\mathcal{F}(\lambda) \leq \mathcal{F}^* < +\infty, \quad \forall\lambda \geq 0. \quad (75)$$

As the sequence $\{(t_\lambda, \boldsymbol{w}_\lambda, \boldsymbol{a}_\lambda, \boldsymbol{b}_\lambda)\}_{\lambda\geq 0} \subset \widehat{\mathcal{H}}$ is bounded, there exists a convergent subsequence of $\{(t_\lambda, \boldsymbol{w}_\lambda, \boldsymbol{a}_\lambda, \boldsymbol{b}_\lambda)\}$. Without loss of generality, it can be assumed that $(t_\lambda, \boldsymbol{w}_\lambda, \boldsymbol{a}_\lambda, \boldsymbol{b}_\lambda) \to (t_*, \boldsymbol{w}_*, \boldsymbol{a}_*, \boldsymbol{b}_*)$ as $\lambda \to +\infty$. Then, $(a_{k,i})_\lambda \to (a_{k,i})_*, (b_i)_\lambda \to (b_i)_*, \forall k \in \mathcal{K}_U, i \in \mathcal{K}_R$ and $t_\lambda \to t_*$. Since $S_\lambda \triangleq \sum_{i\in\mathcal{K}_R}\sum_{k\in\mathcal{K}_U}((a_{k,i})_\lambda - (a_{k,i})_\lambda^2) + \sum_{i\in\mathcal{K}_R}((b_i)_\lambda - (b_i)_\lambda^2) \geq 0, \forall\lambda \geq 0$, one must have $S_* \triangleq \sum_{i\in\mathcal{K}_R}\sum_{k\in\mathcal{K}_U}((a_{k,i})_* - (a_{k,i})_*^2) + \sum_{i\in\mathcal{K}_R}((b_i)_* - (b_i)_*^2) \geq 0$.

We will show that $S_* = 0$ as follows. Suppose to the contrary that $S_* > 0$. Noting that $t_\lambda \to t_*$ and $S_\lambda \to S_*$ as $\lambda \to +\infty$, we have

$$\mathcal{F}(\lambda) = \mathcal{L}(t_\lambda, \boldsymbol{w}_\lambda, \boldsymbol{a}_\lambda, \boldsymbol{b}_\lambda, \lambda) = -t_\lambda + \lambda S_\lambda \to +\infty, \quad (76)$$

which contradicts (75). Hence, $S_* = 0$, which also means that $(\boldsymbol{a}_*, \boldsymbol{b}_*)$ satisfies (30). Since $(t_\lambda, \boldsymbol{w}_\lambda, \boldsymbol{a}_\lambda, \boldsymbol{b}_\lambda) \in \widehat{\mathcal{H}}, \forall\lambda \geq 0$, it is true that $(t_*, \boldsymbol{w}_*, \boldsymbol{a}_*, \boldsymbol{b}_*) \in \widehat{\mathcal{H}}$. Therefore, $(t_*, \boldsymbol{w}_*, \boldsymbol{a}_*, \boldsymbol{b}_*)$ is a feasible point of problem (32).

We note that

$$\sup_{\lambda\geq 0}\mathcal{F}(\lambda) \geq \mathcal{F}(\lambda) = -t_\lambda + \lambda S_\lambda \geq -t_\lambda, \quad \forall\lambda \geq 0. \quad (77)$$

Taking the limit as $\lambda \to +\infty$, it follows that

$$\sup_{\lambda\geq 0}\mathcal{F}(\lambda) \geq -t_* \geq \mathcal{F}^*, \quad (78)$$

which together with (74) implies (34). If $\sup_{\lambda\geq 0}\mathcal{F}(\lambda)$ attains at $\lambda^* \geq 0$, problems (32) and (35) are equivalent in the sense that they share the same optimal value.

Appendix B
Proof of Proposition 2

Note that $(t^{(\kappa+1)}, a_{k,i}^{(\kappa+1)}, b_i^{(\kappa+1)})$ is the optimal solution of problem (43) at the given point $(t^{(\kappa)}, a_{k,i}^{(\kappa)}, b_i^{(\kappa)})$. Following the approximation step (42) at iteration $(\kappa+1)$, we have

$$\begin{aligned}&\widetilde{\mathcal{L}}(t^{(\kappa+1)}, \boldsymbol{a}^{(\kappa+1)}, \boldsymbol{b}^{(\kappa+1)}, \lambda)\\ &= -t^{(\kappa+1)} + \lambda\Big(\sum_{k\in\mathcal{K}_U}\sum_{i\in\mathcal{K}_R}\Big((1-2a_{k,i}^{(\kappa)})a_{k,i}^{(\kappa+1)} + (a_{k,i}^{(\kappa)})^2\Big)\\ &\qquad + \sum_{i\in\mathcal{K}_R}\Big((1-2b_i^{(\kappa)})b_i^{(\kappa+1)} + (b_i^{(\kappa)})^2\Big)\Big)\\ &\leq -t^{(\kappa)} + \lambda\Big(\sum_{k\in\mathcal{K}_U}\sum_{i\in\mathcal{K}_R}\Big((1-2a_{k,i}^{(\kappa)})a_{k,i}^{(\kappa)} + (a_{k,i}^{(\kappa)})^2\Big)\\ &\qquad + \sum_{i\in\mathcal{K}_R}\Big((1-2b_i^{(\kappa)})b_i^{(\kappa)} + (b_i^{(\kappa)})^2\Big)\Big)\\ &= \widetilde{\mathcal{L}}(t^{(\kappa)}, \boldsymbol{a}^{(\kappa)}, \boldsymbol{b}^{(\kappa)}, \lambda), \quad (79)\end{aligned}$$

which means that once initialized from a feasible point $\bar{\boldsymbol{F}}^{(0)}$ given by Subroutine 1, the Algorithm 1 generates a monotone sequence $\{\widetilde{\mathcal{L}}(t^{(\kappa)}, \boldsymbol{a}^{(\kappa)}, \boldsymbol{b}^{(\kappa)}, \lambda)\}$ of improved feasible solutions for (43). As $\{\widetilde{\mathcal{L}}(t^{(\kappa)}, \boldsymbol{a}^{(\kappa)}, \boldsymbol{b}^{(\kappa)}, \lambda)\}$ is bounded from below by constraint (10), it is a convergent sequence. As such, Algorithm 1 converges in the sense that $\lim_{\kappa\to+\infty}\Big(\widetilde{\mathcal{L}}(t^{(\kappa+1)}, \boldsymbol{a}^{(\kappa+1)}, \boldsymbol{b}^{(\kappa+1)}, \lambda) - \widetilde{\mathcal{L}}(t^{(\kappa)}, \boldsymbol{a}^{(\kappa)}, \boldsymbol{b}^{(\kappa)}, \lambda)\Big) = 0$.

On the other hand, following the procedure in [43], we rewrite (35) as the following problem for ease of presentation

$$\min_{\boldsymbol{x}} \ h_0(\boldsymbol{x}) \tag{80a}$$
$$\text{s.t. } h_j(\boldsymbol{x}) \le 0, \quad \forall j \in \mathcal{J} \tag{80b}$$

where $\boldsymbol{x}$ represents the variables of (35); $h_0(\boldsymbol{x})$ is the cost function of problem (35); $h_j(\boldsymbol{x}) \le 0$, $\forall j \in \mathcal{J}_1 = \{1, \ldots, U\}$ are the convex constraints in the feasible set $\widehat{\mathcal{H}}^{(\kappa)}$, $h_j(\boldsymbol{x}) \le 0, \forall j \in \mathcal{J}_2 = \{U+1, \ldots, J\}$ are the nonconvex constraints in $\widehat{\mathcal{H}}^{(\kappa)}$ and $\mathcal{J} = \mathcal{J}_1 \cup \mathcal{J}_2$.

By (38), (39), (41) and (42), problem (35) is approximated by convex problem (43) which is rewritten in a simplified form as

$$\min_{\boldsymbol{x}} \ \tilde{h}_0(\boldsymbol{x}, \boldsymbol{x}^{(\kappa)}) \tag{81a}$$
$$\text{s.t. } h_j(\boldsymbol{x}) \le 0, \quad \forall j \in \mathcal{J}_1 \tag{81b}$$
$$\tilde{h}_j(\boldsymbol{x}, \boldsymbol{x}^{(\kappa)}) \le 0, \quad \forall j \in \mathcal{J}_2, \tag{81c}$$

where the cost function $h_0(\boldsymbol{x})$ and each nonconvex constraint $h_j(\boldsymbol{x}) \le 0, \forall j \in \mathcal{J}_2$ respectively are approximated by a convex function $\tilde{h}_0(\boldsymbol{x}, \boldsymbol{x}^{(\kappa)})$ and a convex constraint $\tilde{h}_j(\boldsymbol{x}, \boldsymbol{x}^{(\kappa)})$ for given point $\boldsymbol{x}^{(\kappa)}$. Importantly, it can be shown that $\tilde{h}_j(\boldsymbol{x}, \boldsymbol{x}^{(\kappa)}) \le 0, \forall j \in \mathcal{J}_2 \cup \{0\}$ obtained from (38), (39), (41), (42) satisfy the following conditions:

$$h_j(\boldsymbol{x}) \le \tilde{h}_j(\boldsymbol{x}, \boldsymbol{x}^{(\kappa)}) \tag{82}$$
$$h_j(\boldsymbol{x}^{(\kappa)}) = \tilde{h}_j(\boldsymbol{x}^{(\kappa)}, \boldsymbol{x}^{(\kappa)}) \tag{83}$$
$$\nabla h_j(\boldsymbol{x}^{(\kappa)}) = \nabla \tilde{h}_j(\boldsymbol{x}^{(\kappa)}, \boldsymbol{x}^{(\kappa)}). \tag{84}$$

Assume that $\boldsymbol{x}^{(\kappa)}$ is the solution obtained from Algorithm 1. Since $\boldsymbol{x}^{(\kappa)}$ is the optimal solution of problem (81), $\boldsymbol{x}^{(\kappa)}$ is a Fritz John point that satisfies the following conditions [25], [44, Lemma 2.1]

$$\lambda_0 \nabla \tilde{h}_0(\boldsymbol{x}^{(\kappa)}, \boldsymbol{x}^{(\kappa)}) + \sum_{j \in \mathcal{J}_1} \lambda_j \nabla h_j(\boldsymbol{x}^{(\kappa)}) + \sum_{j \in \mathcal{J}_2} \lambda_j \nabla \tilde{h}_j(\boldsymbol{x}^{(\kappa)}, \boldsymbol{x}^{(\kappa)}) = 0 \tag{85}$$
$$\lambda_j h_j(\boldsymbol{x}^{(\kappa)}) = 0, \quad \forall j \in \mathcal{J}_1 \tag{86}$$
$$\lambda_j \tilde{h}_j(\boldsymbol{x}^{(\kappa)}, \boldsymbol{x}^{(\kappa)}) = 0, \quad \forall j \in \mathcal{J}_2, \tag{87}$$

where λ_j is the dual variable associated with constraint j. Upon applying (83) and (84) to (85)–(87), we have

$$\lambda_0 \nabla h_0(\boldsymbol{x}^{(\kappa)}) + \sum_{j \in \mathcal{J}} \lambda_j \nabla h_j(\boldsymbol{x}^{(\kappa)}) = 0 \tag{88}$$
$$\lambda_j h_j(\boldsymbol{x}^{(\kappa)}) = 0, \quad \forall j \in \mathcal{J}. \tag{89}$$

This fact implies that $\boldsymbol{x}^{(\kappa)}$ is a Fritz John solution of problem (80) which is indeed (35).

REFERENCES

[1] Ericsson. (2017). *Ericsson Mobility Report*. [Online]. Available: https://www.ericsson.com/assets/local/mobility-report/documents/2017/ericsson-mobility-report-june-2017.pdf

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

[3] B. Dai and W. Yu, "Energy efficiency of downlink transmission strategies for cloud radio access networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 1037–1050, Apr. 2016.

[4] X. Cao, L. Liu, Y. Cheng, and X. S. Shen, "Towards energy-efficient wireless networking in the big data era: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 303–332, 1st Quart., 2017.

[5] K. Singh, A. Gupta, T. Ratnarajah, and M. L. Ku, "A general approach toward green resource allocation in relay-assisted multiuser communication networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 848–862, Feb. 2017.

[6] Ericsson. (2015). *White Paper: 5G Energy Performance*. [Online]. Available: https://www.ericsson.com/assets/local/publications/white-papers/wp-5g-energy-performance.pdf

[7] Y. Zhong, T. Q. S. Quek, and W. Zhang, "Complementary networking for C-RAN: Spectrum efficiency, delay and system cost," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4639–4653, Jul. 2017.

[8] B. Hu, C. Hua, J. Zhang, C. Chen, and X. Guan, "Joint fronthaul multicast beamforming and user-centric clustering in downlink C-RANs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5395–5409, Aug. 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] L. Liu and R. Zhang, "Optimized uplink transmission in multi-antenna C-RAN with spatial compression and forward," *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 5083–5095, Oct. 2015.

[11] V. Suryaprakash, P. Rost, and G. Fettweis, "Are heterogeneous cloud-based radio access networks cost effective?" *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2239–2251, Oct. 2015.

[12] T. Q. S. Quek, M. Peng, O. Simeone, and W. Yu, *Cloud Radio Access Networks: Principles, Technologies, and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[13] D. Liu *et al.*, "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, 2nd Quart., 2016.

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

[15] L. Liu and W. Yu, "Cross-layer design for downlink multihop cloud radio access networks with network coding," *IEEE Trans. Signal Process.*, vol. 65, no. 7, pp. 1728–1740, Apr. 2017.

[16] S.-H. Park, O. Simeone, and S. Shamai (Shitz), "Joint optimization of cloud and edge processing for fog radio access networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7621–7632, Nov. 2016.

[17] J. Kang, O. Simeone, J. Kang, and S. Shamai (Shitz), "Fronthaul compression and precoding design for C-RANs over ergodic fading channels," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5022–5032, Jul. 2016.

[18] F. Richter, A. J. Fehske, and G. P. Fettweis, "Energy efficiency aspects of base station deployment strategies for cellular networks," in *Proc. IEEE Veh. Technol. Conf. Fall (VTC-Fall)*, Sep. 2009, pp. 1–5.

[19] Z. Yan, M. Peng, and C. Wang, "Economical energy efficiency: An advanced performance metric for 5G systems," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 32–37, Feb. 2017.

[20] P. Gandotra, R. K. Jha, and S. Jain, "Green communication in next generation cellular networks: A survey," *IEEE Access*, vol. 5, pp. 11727–11758, 2017.

[21] J. Zuo, J. Zhang, C. Yuen, W. Jiang, and W. Luo, "Energy efficient user association for cloud radio access networks," *IEEE Access*, vol. 4, pp. 2429–2438, 2016.

[22] R. Sun, M. Hong, and Z.-Q. Luo, "Joint downlink base station association and power control for max-min fairness: Computation and complexity," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1040–1054, Jun. 2015.

[23] K.-G. Nguyen, Q.-D. Vu, M. Juntti, and L.-N. Tran, "Energy efficient preceding C-RAN downlink with compression at fronthaul," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[24] S. H. Park, O. Simeone, O. Sahin, and S. Shamai (Shitz), "Multihop backhaul compression for the uplink of cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3185–3199, May 2016.

[25] O. L. Mangasarian, *Nonlinear Programming*. Philadelphia, PA, USA: SIAM, 1994.

[26] T. T. Vu, D. T. Ngo, M. N. Dao, S. Durrani, D. H. N. Nguyen, and R. H. Middleton, "Energy-efficient design for downlink cloud radio access networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

[27] M. Tao, E. Chen, H. Zhou, and W. Yu, "Content-centric sparse multicast beamforming for cache-enabled cloud RAN," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6118–6131, Sep. 2016.
[28] A. D. L. Oliva, J. A. Hernandez, D. Larrabeiti, and A. Azcorra, "An overview of the CPRI specification and its application to C-RAN-based LTE scenarios," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 152–159, Feb. 2016.
[29] J. Park, S. Park, A. Yazdan, and R. W. Heath, Jr., "Optimization of mixed-ADC multi-antenna systems for cloud-RAN deployments," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3962–3975, Sep. 2017.
[30] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai (Shitz), "Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5646–5658, Nov. 2013.
[31] G. Auer *et al.*, "How much energy is needed to run a wireless network?" *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.
[32] N. Yu, Y. Miao, L. Mu, H. Du, H. Huang, and X. Jia, "Minimizing energy cost by dynamic switching ON/OFF base stations in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7457–7469, Nov. 2016.
[33] J. Yuan, Z. Li, W. Yu, and B. Li, "A cross-layer optimization framework for multihop multicast in wireless mesh networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 11, pp. 2092–2103, Nov. 2006.
[34] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.
[35] Z. Li, B. Li, D. Jiang, and L. C. Lau, "On achieving optimal throughput with network coding," in *Proc. IEEE INFOCOM*, vol. 3. Mar. 2005, pp. 2184–2194.
[36] T. Ho and D. Lun, *Network Coding: An Introduction*. New York, NY, USA: Cambridge Univ. Press, 2008.
[37] S. Jaggi *et al.*, "Polynomial time algorithms for multicast network code construction," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1973–1982, Jun. 2005.
[38] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, Oct. 2003.
[39] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.
[40] H. H. M. Tam, H. D. Tuan, and D. T. Ngo, "Successive convex quadratic programming for quality-of-service management in full-duplex MU-MIMO multicell networks," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2340–2353, Jun. 2016.
[41] A. El Gamal and Y.-H. Kim, *Network Information Theory*. New York, NY, USA: Cambridge Univ. Press, 2011.
[42] *3GPP Technical Specification Group Radio Access Network, Evolved Universal Terrestrial Radio Access (E-UTRA): Further Advancements for E-UTRA Physical Layer Aspects*, document 3GPP TS 36.814 V9.0.0, 2010.
[43] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, pp. 681–683, Jul. 1978.
[44] M. N. Dao, "Bundle method for nonconvex nonsmooth constrained optimization," *J. Convex Anal.*, vol. 22, no. 4, pp. 1061–1090, Dec. 2015.

Tung Thanh Vu (S'17) received the B.Sc. degree (Hons.) in telecommunications and networking from the Ho Chi Minh City University of Science in 2012 and the M.Sc. degree in telecommunications engineering from the Ho Chi Minh City University of Technology in 2016. He is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Computing, The University of Newcastle, Australia. His current research interests include optimization designs for cloud radio access networks and multi-access edge computing.

Duy Trong Ngo (S'08–M'15) received the B.Eng. degree (Hons.) in telecommunication engineering from The University of New South Wales, Australia, in 2007, the M.Sc. degree in electrical engineering (communication) from the University of Alberta, Canada, in 2009, and the Ph.D. degree in electrical engineering from McGill University, Canada, in 2013.

In 2013, he joined the School of Electrical Engineering and Computing, The University of Newcastle, Australia, as a Lecturer, where he became a Senior Lecturer in 2017. He currently leads the research effort in design and optimization for 5G and beyond wireless communications networks. His research interests include cloud radio access networks, multi-access edge computing, simultaneous wireless information and power transfer, and vehicle-to-everything communications for intelligent transportation systems.

Dr. Ngo received the NICTA Telecommunications Excellence Award in 2006, the University Medal of the University of New South Wales in 2007, the two prestigious Post-Doctoral Fellowships of the Natural Sciences and Engineering Research Council of Canada and the Fonds de recherche du Québec—Nature et technologies in 2013, the 2015 Vice-Chancellor's Award for Research and Innovation Excellence, the 2015 Pro Vice-Chancellor's Award for Research Excellence, and the 2017 Pro Vice-Chancellor's Award for Teaching Excellence in the Faculty of Engineering and Built Environment from The University of Newcastle.

Minh N. Dao received the B.Sc. (Hons.) and M.Sc. degrees in mathematics from the Hanoi National University of Education, Vietnam, in 2004 and 2006, respectively, and the Ph.D. degree in applied mathematics from the University of Toulouse, France, in 2014.

He was as a Lecturer with the Hanoi National University of Education, Vietnam, from 2004 to 2010, a Lecturer and Research Assistant with the National Institute of Applied Sciences, Toulouse, France, from 2013 to 2014, and a Post-Doctoral Fellow with The University of British Columbia, Canada, from 2014 to 2016. He is currently a Research Associate with the Priority Research Centre for Computer-Assisted Research Mathematics and Its Applications, The University of Newcastle, Australia. His research interests include nonlinear optimization, nonsmooth analysis, iterative methods, control theory, operations research, and signal processing. In 2017, he received the Annual Best Paper Award from the *Journal of Global Optimization*.

Salman Durrani (S'00–M'05–SM'10) received the B.Sc. degree (Hons.) in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2000, and the Ph.D. degree in electrical engineering from The University of Queensland, Brisbane, Australia, in 2004. Since 2005, he has been with The Australian National University, Canberra, Australia, where he is currently an Associate Professor with the Research School of Engineering, College of Engineering and Computer Science. His research interests include wireless communications and signal processing, including machine-to-machine and device-to-device communication, wireless-energy-harvesting systems, stochastic geometry modeling of finite area networks, and synchronization in communication systems. He is a member of Engineers Australia and a Senior Fellow of The Higher Education Academy, U.K. From 2015 to 2016, he was the Chair of the ACT Chapter of the IEEE Signal Processing Society and the IEEE Communications Society. He was a recipient of the 2016 IEEE ComSoc Asia Pacific Outstanding Paper Award. He currently serves as an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS.

Duy H. N. Nguyen (S'07–M'14) received the B.Eng. degree (Hons.) from the Swinburne University of Technology, Hawthorn, VIC, Australia, in 2005, the M.Sc. degree from the University of Saskatchewan, Saskatoon, SK, Canada, in 2009, and the Ph.D. degree from McGill University, Montréal, QC, Canada, in 2013, all in electrical engineering. From 2013 to 2015, he held a joint appointment as a Research Associate with McGill University and a Post-Doctoral Research Fellow with the Institut National de la Recherche Scientifique, Université du Québec, Montréal. He was a Research Assistant with the University of Houston in 2015 and a Post-Doctoral Research Fellow with The University of Texas at Austin in 2016. Since 2016, he has been an Assistant Professor with the Department of Electrical and Computer Engineering, San Diego State University, San Diego, CA, USA. His current research interests include resource allocation in wireless networks, signal processing for communications, convex optimization, and game theory. He was a recipient of the Australian Development Scholarship, the FRQNT Doctoral Fellowship and Post-Doctoral Fellowship, and the NSERC Post-Doctoral Fellowship.

Richard H. Middleton (SM'86–F'99) received the Ph.D. degree from The University of Newcastle, Australia, in 1987. From 2007 to 2011, he was a Research Professor with the Hamilton Institute, The National University of Ireland, Maynooth, Ireland. He is currently a Professor and the Head of the School of Electrical Engineering and Computing with The University of Newcastle. His research interests include a broad range of control systems theory and applications, including communications systems, control of distributed systems, and systems biology. He is a fellow of the IFAC. He has served as the Program Chair (CDC 2006) and the Co-General Chair (CDC 2017) for CSS Vice President Membership Activities, and the Vice President for Conference Activities. In 2011, he was the President of the IEEE Control Systems Society.