Spectral and Energy Efficiency Maximization for Content-Centric C-RANs with Edge Caching

Tung T. Vu, Duy T. Ngo, Minh N. Dao, Salman Durrani and Richard H. Middleton

***Abstract*—This paper aims to maximize the spectral and energy efficiencies of a content-centric cloud radio access network (C-RAN), where users requesting the same contents are grouped together. Data are transferred from a central baseband unit to multiple remote radio heads (RRHs) equipped with local caches. The RRHs then send the received data to each group's user. Both multicast and unicast schemes are considered for data transmission. We formulate mixed-integer nonlinear problems in which user association, RRH activation, data rate allocation and signal precoding are jointly designed. These challenging problems are subject to minimum data rate requirements, limited fronthaul capacity and maximum RRH transmit power. Employing successive convex quadratic programming, we propose iterative algorithms with guaranteed convergence to Fritz John solutions. Numerical results confirm that the proposed joint designs markedly improve the spectral and energy efficiencies of the considered content-centric C-RAN compared to benchmark schemes. Importantly, they show that unicasting outperforms multicasting in terms of spectral efficiency in both cache and cache-less scenarios. In terms of energy efficiency, multicasting is the best choice for the system without cache whereas unicasting is best for the system with cache. Finally, edge caching is shown to improve both spectral and energy efficiencies.**

***Index Terms*—C-RAN, data rate allocation, edge caching, limited-capacity fronthaul, precoding design, user association**

I. Introduction

Mobile data traffic has been growing exponentially in recent years. A report by Ericsson shows that while the global mobile traffic grew 66% in 2015, it is still predicted to increase more than tenfold by 2022 [1]. However, the rapid increase of the associated operating expenditure and energy consumption becomes problematic, especially with the slow growth of mobile operators' profits and the serious concerns on green-house gas emissions [2]. The fifth-generation (5G) of mobile communication systems is expected to address these issues by offering substantially higher spectral and energy efficiencies than the traditional long-term evolution (LTE) networks. Telecommunications operators specifically target a $1,000$-fold data traffic increase at the same time as reducing the total network energy consumption by half [3].

To meet these ambitious objectives of 5G networks, a promising solution termed as cloud radio access network (C-RAN) has been proposed and developed [4]. In a C-RAN, traditional high-cost high-power base stations (BSs) are replaced by low-cost low-power remote radio heads (RRHs), resulting in less construction space and lower energy consumption in a dense network setting [2], [5]. Users are connected to the RRHs via radio access links, whereas RRHs are connected to a central base band unit (BBU) in the core network via wireline fronthaul links. With C-RANs, large-scale allocation of radio and computing resources across all the RRHs can be centrally processed at the same BBU pools. The effective interference management at the BBUs enables significant performance gains over single-cell processing [6]. However, performing fully joint processing requires tremendous data sharing on the fronthaul links, while the current fronthaul solutions are yet to catch up with. The finite-capacity fronthaul links remain a bottleneck of practical C-RANs, causing severe latency and performance degradation in both spectral and energy efficiencies [7].

Edge caching and user association (UA) have recently been introduced to address the bottleneck problem on the fronthaul links [2], [8]. The main idea of edge caching is to exploit the popularity of the requested contents in modern communications such as video streaming, push media, mobile applications and mobile TV [9]. Since one copy of content may be requested by multiple users in these services, prefetching the most frequently requested content at the local cache of RRHs during off-peak time can significantly reduce the fronthaul data traffic [10], [11]. On the other hand, the idea of user association is to distribute data to an appropriate RRH in order to serve the right user. It has been shown to greatly release the traffic burden on the fronthaul links [12]. Therefore, *a content-centric transmission for C-RANs in which edge caching, UA and the statistics of users' requested contents are taken into account should be carefully designed.*

An example of a content-centric CRAN with edge caching is illustrated in Fig. 1. To further utilize the popularity of the content, the users requesting the same content are grouped together [13]. In this way, the group's requested content only needs to be delivered once to the serving RRHs, helping reduce the waste of resources on the fronthaul links and improve the power efficiency in the access links compared to the user-centric schemes [13]–[15]. Here, data content is delivered to groups in two phases. In the content placement

Manuscript received February 20, 2018; revised April 23, 2018; accepted August 10, 2018. The editor coordinating the review of this paper and approving it for publication was V. Aggarwal. This work is supported in part by an ECR-HDR scholarship from The University of Newcastle, in part by the Australian Research Council Discovery Project grants DP170100939 and DP160101537.

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, richard.middleton}@newcastle.edu.au).

M. N. Dao is with the Priority Research Centre for Computer-Assisted Research Mathematics and its Applications (CARMA), 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).

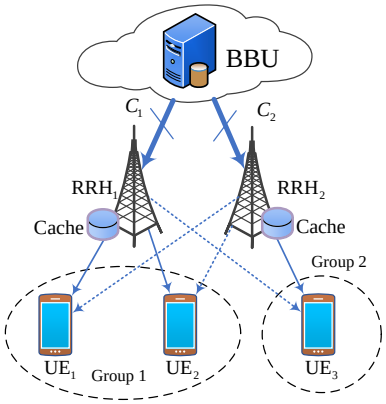


Fig. 1. An example of a content-centric C-RAN with a BBU in the core network and edge caching at RRHs, where users requesting the same contents are grouped together.

(or pre-fetching) phase, the groups/users' potentially requested files are pre-stored in the RRHs' caches via some caching strategies. In the data request-delivery phase, if a group/user's requested file is available at the local caches of its serving RRHs, the file is directly retrieved from the caches instead of from the BBU [16]. If the file is not cached, it is fetched from the BBU to the RRHs before being sent to the groups/users.

In principle, the content placement phase generally takes place in a larger timescale (e.g., on an hourly basis) while the data request-delivery phase happens in a much shorter timescale [13], [17], [18]. This timescale mismatch warrants different design approaches for each phase. For content placement, [19]–[23] devise a number of efficient caching strategies for a given data delivery scheme. These studies exploit the statistics of user content demand to pre-store the most frequently requested files at the appropriate RRHs.

For data delivery, multicast or unicast scheme with group/user association, data rate allocation and beamforming/precoding designs can further be performed to improve system performance [8], [13] for a given caching strategy. Specifically, [8] proposed a joint design of data rate allocation, precoding to maximize the minimum-user rate of a cache-enabled C-RAN under limited fronthaul capacity constraints. However, users are heuristically assigned to RRHs to avoid an exhaustive search, and hence, the full potential of UA may not be realized. Moreover, [8] only focuses on a user-centric scenario where the statistics of users' requested content cannot be exploited. Recently, [13] proposed a joint design of multicast beamforming and BS clustering (a UA problem, essentially) to minimize the fronthaul traffic and transmission power for the content-centric C-RANs with edge caching. However, a full comparison between multicasting and unicasting in terms of spectral and energy efficiencies has not been carried out. On the other hand, groups or users are typically assigned to appropriate RRHs that store the requested files and/or have good channel conditions, and the inactive RRHs with no assigned groups/users are put into sleep mode. By doing this way, the fronthaul traffic is reduced at the same time as less power is expended in the fronthaul and radio access links [24]. Therefore, RRH activation should also be taken into account in the design of this phase.

One may also consider optimizing both content placement and data delivery at the same time. Compared to optimizing each task individually, it is much more challenging to jointly design a caching strategy together with UA, data rate allocation and precoders/beamformers. For example, [17] developed a mixed-timescale precoding and cache control policy for multi-input multi-output wireless systems. However, the same caching strategy must be applied to all BSs, and each user must be assigned to either one BS or all BSs.

In this paper, we optimize the data delivery for a content-centric C-RAN with edge caching and limited fronthaul capacity. To enhance both spectral and energy efficiencies, UA, RRH activation, data rate allocation and signal precoding are jointly optimized for any fixed caching strategy. Our aim is to answer the following two questions: (i) In the data request-delivery phase, which is better: multicasting or unicasting? and (ii) How would the cache affect the spectral and energy efficiencies obtained by these schemes? In answering those questions, we make the following research contributions.

- We formulate the design problems of maximizing the spectral and energy efficiencies. The energy efficiency is defined as a ratio of the spectral efficiency and total power consumption. The designs are constrained on meeting the predefined data rate requirements, the limited fronthaul capacity and the maximum transmit power at each RRH. While particularly relevant, our formulated problems are challenging to solve due to the combinatorial nature as well as the tight coupling among the variables.
- We propose an iterative algorithm to solve the challenging mixed-integer nonlinear problems, where only one simple convex program is involved in each iteration. In particular, we first convert the formulated problems into their epigraph forms. To deal with the binary UA and RRH activation variables, we further recast them into approximated problems with continuous variable only. Finally, these problems are solved by the successive convex quadratic programming.
- We prove analytically and verify by numerical examples that the proposed algorithm converges to at least a Fritz John solution[1] of the approximated problem once initialized from a feasible point. Simulation results with practical parameter settings show that the developed joint optimization substantially improves both spectral and energy efficiencies over benchmarking approaches. Importantly, it is confirmed that the unicast scheme always offers higher spectral efficiency than the multicast scheme. In terms of energy efficiency, the unicast scheme is the best in the system with cache while the multicast scheme for the system without cache. Also, edge caching is shown to contribute to both the spectral and energy efficiency improvements.

It is worth noting that this work substantially extends our

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

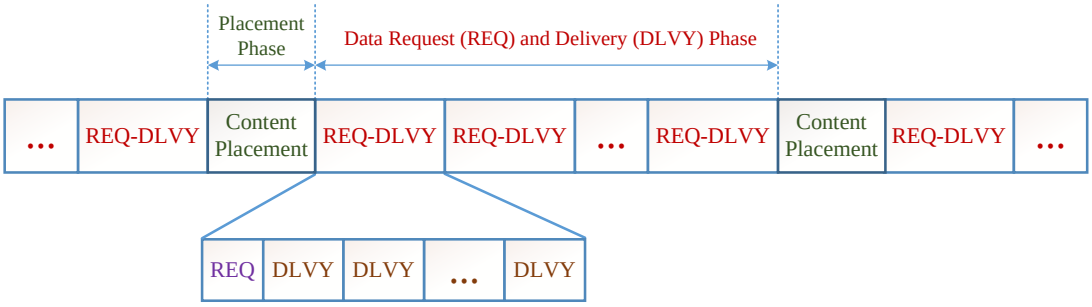


Fig. 2. Cache-assisted communication in a cache-enabled C-RAN.

initial result in [26], where the joint problem of maximizing spectral and energy efficiencies is introduced under a user-centric configuration and only the advantages of the joint design over the data rate allocation and precoder design without any UA optimization were demonstrated.

Organization and Notation: The rest of this paper is organized as follows: Section II presents the considered system model and assumptions. Sections III formulates and solves the spectral efficiency maximization problem, whereas Sections IV deals with the energy efficiency maximization problem. Section V verifies the performance of the developed algorithms through comprehensive numerical examples. Finally, Section VI concludes the paper.

In this paper, the real part of a complex number x is denoted as $\Re\{x\}$. For a scalar x, $\lfloor x \rfloor$ denotes the largest integer that is not larger than x. Boldfaced symbols are used for vectors and capitalized boldfaced 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 $\boldsymbol{X}$. $\boldsymbol{I}$ and $\boldsymbol{0}$ are the identity and zero matrices with appropriate dimensions, respectively. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{Q})$ denotes the circularly symmetric complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{Q}$. $|\mathcal{G}|$ stands for the number of elements in set $\mathcal{G}$.

II. System Model and Assumptions

Consider the general content-centric C-RAN model, where the BBU in the core network connects to a set of RRHs $\mathcal{K}_R \triangleq \{1, \ldots, K_R\}$ via the fronthaul links. Assume that the fronthaul link $i \in \mathcal{K}_R$ has a limited capacity of $C_i > 0$ (bps). 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 is equipped with N_u antennas while each RRH $i \in \mathcal{K}_R$ is equipped with N_r antennas. The library of the BBU contains F files with equal size. These assumption help simplify the problem formulation and solution development, without loss of generality. The same assumptions have been widely used in the literature, e.g., [8], [13], [24].

Each user sends requests to its serving RRHs asking for some files from a known file library. The users asking for the same file are grouped together and served by either a unicast or a multicast scheme. Conventionally, the requested files need to be fetched from the BBU to the serving RRHs before being transferred to the requesting users via the wireless access links. However, this two-hop communication lengthens the end-to-end delay as well as putting a traffic burden on the fronthaul links. The idea of edge caching is to bring the requested files closer to the users by pre-storing part of the file library at the RRHs. In this way, if a requested file is located in the serving RRHs of a requesting user, it will be directly sent in the downlink from the RRHs to that user.

Fig. 2 shows the cache-assisted communication scenario in a C-RAN. The two phases are elaborated below.

A. Placement Phase

The placement phase sees a subset of the file library pre-stored at the RRHs' caches before any data request and delivery happen. Our model assumes that each RRH $i \in \mathcal{K}_R$ is equipped with a local cache that can store up to $B_i > 0$ files. Uncoded caching [8], [13], [20], [23] and coded caching [21], [22] strategies can be employed. They are based on the long-term state information of the popularity statistics of the requested files, the cache size, and the fronthaul capacity. The cache contents are assumed to remain unchanged in the RRHs until the next content placement, regardless of variations in user requests and channel conditions. Here, we suppose that the cache placement is given and thus focus on the optimization of the transmission in the next phase. Similar assumptions can be found in the literature, e.g., [8], [13].

B. Data Request and Delivery Phase

This phase consists of multiple data request and delivery (REQ-DLVY) intervals. In each REQ-DLVY interval, each user group g (denoted as $\mathcal{G}_g$) first sends the same request (REQ) for a file f_g from the library $\mathcal{F} \triangleq \{1, \ldots, F\}$ stored in the BBU, followed by a delivery action (DLVY) to transfer all the requested files. Denote by $\mathcal{K}_G \triangleq \{1, \ldots, K_G\}$ the set of all groups, where $1 \leq K_G \leq \min\{K_U, F\}$. It is assumed that each user requests one file at a time, and hence $\mathcal{G}_m \cap \mathcal{G}_g = \varnothing, \forall m \neq g$ and $\sum_{g \in \mathcal{K}_G} |\mathcal{G}_g| \leq K_U$ [13]. For the fixed cache content from the placement phase and the known requested files, the cache state information

$$c_{g,i} \triangleq \begin{cases} 1, & \text{if } f_g \in \mathcal{C}_i, \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

shows whether file f_g is cached at the local cache $\mathcal{C}_i$ of RRH i. The value of $c_{g,i}, \forall g \in \mathcal{K}_G \triangleq \{1, \ldots, K_G\}, i \in \mathcal{K}_R$

is available to the BBU and fixed during one REQ-DLVY interval.

To deliver the requested file to each group, the multicast or unicast schemes are both considered and briefly introduced as follows.

1) Multicast scheme: At the BBU, the message M_g of file f_g requested by group $g, \forall g \in \mathcal{K}_G$ is uniformly distributed in the set $\{1, \ldots, 2^{nR_g}\}$, where n is the block length and R_g (bps/Hz) is the same data rate of message M_g assigned for every user in group g. After being transferred to the RRHs via noiseless fronthaul links at rate R_g, each message M_g is encoded into a symbol $\boldsymbol{s}_g \in \mathbb{C}^{d\times 1}$ for group g, where $d \leq \min\{N_r, N_u\}$ is the number of data streams and $\boldsymbol{s}_g \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$.

Given the limited cache size, not all requested files are available at the RRHs, in which case they must be fetched from the BBU to the RRHs via the fronthaul links. For the cache state information given in (1), each multicast group is served by a selected subset of RRHs and all the missing files are transferred to these serving RRHs. Here, the RRH-group associations are modeled by the following binary variables

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

From (1) and (2), the total fronthaul rate to RRH $i \in \mathcal{K}_R$ in the current DLVY interval is expressed as

$$R_i^{FH,\text{multicast}} \triangleq W \sum_{g\in\mathcal{K}_G} a_{g,i}(1-c_{g,i})R_g \ \text{(bps)}, \tag{3}$$

where W is the total available bandwidth. In (3), $c_{g,i} = 0$ corresponds to the case that the file f_g is fetched from the BBU to the RRHs via the fronthaul link before being sent to the requesting group. On the other hand, $c_{g,i} = 1$ implies that the file f_g is sent directly from the RRH's cache to the requesting group and the fronthaul link is not used.

With the requested files coming from the BBU or already in the caches, an RRH i generates the transmitted baseband signal $\boldsymbol{x}_i$ as

$$\boldsymbol{x}_i = \sum_{g\in\mathcal{K}_G} \boldsymbol{F}_{g,i}\, \boldsymbol{s}_g, \tag{4}$$

where $\boldsymbol{F}_{g,i} \in \mathbb{C}^{N_r\times d}$ is the precoding matrix for $\boldsymbol{s}_g$. Each RRH i is assumed to be subjected to the average transmit power constraint expressed as

$$\mathbb{E}\left\{\|\,\boldsymbol{x}_i\,\|^2\right\} \leq P_i. \tag{5}$$

Denote by $\boldsymbol{H}_{k,i} \in \mathbb{C}^{N_u\times N_r}$ the flat-fading channel matrix from RRH i to user k and by $\boldsymbol{H}_k \triangleq [\boldsymbol{H}_{k,1}, ..., \boldsymbol{H}_{k,K_R}] \in \mathbb{C}^{N_u\times N_R}$ the channel matrix from all RRHs to user k, where $N_R \triangleq K_R N_r$. Assume that the channel states $\boldsymbol{H}_{k,i},\ k \in \mathcal{K}_U,\ i \in \mathcal{K}_R$ remain unchanged during each DLVY interval and are made known to the BBU and RRHs [13]. Upon defining $\bar{\boldsymbol{F}}_g \triangleq \left[\boldsymbol{F}_{g,1}^H, \boldsymbol{F}_{g,2}^H, \ldots, \boldsymbol{F}_{g,K_R}^H\right]^H \in \mathbb{C}^{N_R\times d}$, the received signal $\boldsymbol{y}_k \in \mathbb{C}^{N_u\times 1}$ at a user k in a multicast group g requesting a file f_g can thus be written as

$$\boldsymbol{y}_{g,k} = \boldsymbol{H}_k \bar{\boldsymbol{F}}_g \boldsymbol{s}_g + \sum_{m\in\mathcal{K}_G\setminus\{g\}} \boldsymbol{H}_k \bar{\boldsymbol{F}}_m \boldsymbol{s}_m + \boldsymbol{n}_k, \tag{6}$$

where $\boldsymbol{n}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma}_k)$ is the additive noise term.

The data rate R_g of the file f_g for the group g under the multicast scheme is always achievable in Shannon's sense as

$$R_g \leq \min_{k\in\mathcal{G}_g} r_{g,k}(\bar{\boldsymbol{F}}) \triangleq \log_2 \left|\boldsymbol{I}_{N_{U,k}} + \boldsymbol{\Pi}_{g,k}\boldsymbol{\Pi}_{g,k}^H \boldsymbol{\Xi}_{g,k}^{-1}\right|, \forall g \in \mathcal{K}_G, \tag{7}$$

where $\bar{\boldsymbol{F}} \triangleq \{\bar{\boldsymbol{F}}_g\}_{g\in\mathcal{K}_G}$, $\boldsymbol{\Pi}_{g,k} \triangleq \boldsymbol{H}_k \bar{\boldsymbol{F}}_g$, and

$$\boldsymbol{\Xi}_{g,k} \triangleq \sum_{m\in\mathcal{K}_G\setminus\{g\}} \boldsymbol{H}_k \bar{\boldsymbol{F}}_m \bar{\boldsymbol{F}}_m^H \boldsymbol{H}_k^H + \boldsymbol{\Sigma}_k\,. \tag{8}$$

The spectral efficiency under the multicast scheme is then defined by the following sum rate:

$$\eta_{\text{SE}}^{\text{multicast}} \triangleq \sum_{g\in\mathcal{K}_G} |\,\mathcal{G}_g\,| R_g \ \text{(bps/Hz)}. \tag{9}$$

2) Unicast scheme: Here, the users in a group g requesting a file f_g have different data rates. The message M_g of file f_g is transferred from the BBU to the RRH i at the maximum rate rate of the associated users in the group g [13], i.e., $R_{g,i} = \max_{k\in\mathcal{G}_g}\{\tilde{a}_{k,i}R_k\}$, where

$$\tilde{a}_{k,i} \triangleq \begin{cases} 1, & \text{if RRH } i \text{ serves user } k, \\ 0, & \text{otherwise,} \end{cases} \tag{10}$$

and R_k (bps/Hz) is the data rate of the user k. The total rate of the fronthaul link connecting with RRH $i \in \mathcal{K}_R$ is thus expressed as

$$R_i^{FH,\text{unicast}} \triangleq W \sum_{g\in\mathcal{K}_G} (1-c_{g,i}) \max_{k\in\mathcal{G}_g}\{\tilde{a}_{k,i}R_k\} \ \text{(bps)}. \tag{11}$$

At the RRHs, each message $M_g, \forall g \in \mathcal{K}_G$ is encoded into a symbol $\boldsymbol{s}_k \in \mathbb{C}^{d\times 1}$ for a user $k \in \mathcal{G}_g$, where $\boldsymbol{s}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$. Upon defining $\widetilde{\boldsymbol{F}}_k \triangleq \left[\boldsymbol{F}_{k,1}^H, \boldsymbol{F}_{k,2}^H, \ldots, \boldsymbol{F}_{k,K_R}^H\right]^H \in \mathbb{C}^{N_R\times d}$. The data rate R_k of user k in group g under the unicast scheme is always achievable in Shannon's sense as

$$R_k \leq \widetilde{r}_k(\tilde{\boldsymbol{F}}) \triangleq \log_2 \left|\boldsymbol{I}_{N_{U,k}} + \boldsymbol{\Pi}_k \boldsymbol{\Pi}_k^H \boldsymbol{\Xi}_k^{-1}\right|, \forall k \in \mathcal{K}_U, \tag{12}$$

where $\widetilde{\boldsymbol{F}} \triangleq \{\widetilde{\boldsymbol{F}}_k\}_{k\in\mathcal{K}_U}$, $\boldsymbol{\Pi}_k \triangleq \boldsymbol{H}_k \bar{\boldsymbol{F}}_k$, and

$$\boldsymbol{\Xi}_k \triangleq \sum_{m\in\mathcal{K}_U\setminus\{k\}} \boldsymbol{H}_k \widetilde{\boldsymbol{F}}_m \widetilde{\boldsymbol{F}}_m^H \boldsymbol{H}_k^H + \boldsymbol{\Sigma}_k\,. \tag{13}$$

The spectral efficiency under the unicast scheme is then defined by the following sum rate:

$$\eta_{\text{SE}}^{\text{unicast}} \triangleq \sum_{k\in\mathcal{K}_U} R_k \ \text{(bps/Hz)}. \tag{14}$$

C. RRH-Group/User Associations and RRH Activation

We denote $\boldsymbol{a} \triangleq \{a_{g,i}\}_{g\in\mathcal{K}_G, i\in\mathcal{K}_R}$, $\tilde{\boldsymbol{a}} \triangleq \{\tilde{a}_{k,i}\}_{k\in\mathcal{K}_U, i\in\mathcal{K}_R}$ and $\boldsymbol{b} \triangleq \{b_i\}_{i\in\mathcal{K}_R}$, where

$$b_i \triangleq \begin{cases} 0, & \text{if RRH } i \text{ serves no group or user,} \\ 1, & \text{otherwise.} \end{cases} \tag{15}$$

It should be noted that an RRH i is assigned to serve a group g or a user k if and only if the corresponding precoder of the message symbol $\boldsymbol{s}_g$ or $\boldsymbol{s}_k$ is not a zero matrix, i.e., $\boldsymbol{F}_{g,i} = \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_g \neq \boldsymbol{0}$ or $\boldsymbol{F}_{k,i} = \bar{\boldsymbol{E}}_i^H \tilde{\boldsymbol{F}}_k \neq \boldsymbol{0}$; $\bar{\boldsymbol{E}}_i \in \mathbb{C}^{N_R \times N_r}$ is zero everywhere except an identity matrix of size N_r from row $(i-1)N_r + 1$ to row iN_r. Following this fact, (2), (10) and (15) can be expressed as

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

$$\tilde{a}_{k,i} = \begin{cases} 0, & \text{if } \langle \bar{\boldsymbol{E}}_i^H \widetilde{\boldsymbol{F}}_k \widetilde{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i \rangle = 0, \ \forall k \in \mathcal{K}_U, i \in \mathcal{K}_R, \\ 1, & \text{otherwise,} \end{cases} \tag{17}$$

$$\begin{aligned} b_i &= \begin{cases} 0, & \text{if } a_{g,i} = 0, \forall g \in \mathcal{K}_G, \\ 1, & \text{otherwise} \end{cases} \\ &= \begin{cases} 0, & \text{if } \sum_{g \in \mathcal{K}_G} \langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_g \bar{\boldsymbol{F}}_g^H \bar{\boldsymbol{E}}_i \rangle = 0, \\ 1, & \text{otherwise} \end{cases} \\ &= \begin{cases} 0, & \text{if } \widetilde{a}_{k,i} = 0, \forall k \in \mathcal{K}_U, \\ 1, & \text{otherwise} \end{cases} \\ &= \begin{cases} 0, & \text{if } \sum_{k \in \mathcal{K}_U} \langle \bar{\boldsymbol{E}}_i^H \widetilde{\boldsymbol{F}}_k \widetilde{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i \rangle = 0, \ \forall i \in \mathcal{K}_R \,. \\ 1, & \text{otherwise,} \end{cases} \end{aligned} \tag{18}$$

(19)

From (16)–(19), the interdependence among $\boldsymbol{a}$, $\tilde{\boldsymbol{a}}$ and $\boldsymbol{b}$ is expressed as

$$a_{g,i} \leq b_i \leq \sum_{g \in \mathcal{K}_G} a_{g,i}, \forall g \in \mathcal{K}_G, i \in \mathcal{K}_R, \tag{20}$$

$$\tilde{a}_{k,i} \leq b_i \leq \sum_{k \in \mathcal{K}_U} \tilde{a}_{k,i}, \forall k \in \mathcal{K}_U, i \in \mathcal{K}_R, \tag{21}$$

ensuring that no group or user is assigned to an inactive RRH.

D. Power Consumption Model

This paper adopts a practical power consumption model that is applicable to different types of BSs [24], [27]. The power consumed by RRH $i \in \mathcal{K}_R$ in the given transmission interval is expressed as

$$P_i^{RRH,X} \triangleq \begin{cases} \beta_i P_i^{Tx,X} + P_{i,a}, & \text{if } 0 < P_i^{tx} \leq P_i, \\ P_{i,s}, & \text{if } P_i^{tx} = 0, \end{cases} \tag{22}$$

where $X = \{\text{multicast}, \text{unicast}\}$ indicates the transmission scheme, the constant $\beta_i > 0$, $i \in \mathcal{K}_R$ reflects the power amplifier efficiency, feeder loss and other loss factors due to power supply and cooling for RRH i [24]; $P_i^{Tx,X}$ is the transmit power required to deliver all requested files from RRH i under scheme X as

$$P_i^{Tx,X} \triangleq \begin{cases} P_i^{Tx,\text{multicast}} = \sum_{g \in \mathcal{K}_G} \langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_g \bar{\boldsymbol{F}}_g^H \bar{\boldsymbol{E}}_i \rangle, \\ P_i^{Tx,\text{unicast}} = \sum_{k \in \mathcal{K}_U} \langle \bar{\boldsymbol{E}}_i^H \widetilde{\boldsymbol{F}}_k \widetilde{\boldsymbol{F}}_k^H \bar{\boldsymbol{E}}_i \rangle, \end{cases} \tag{23}$$

in which $P_{i,a}$ is the power required to support RRH i in the active mode; and $P_{i,s} < P_{i,a}$ is the power consumption in the sleep mode.

On the other hand, the fronthaul link from the BBU to RRH $i \in \mathcal{K}_R$ is modeled as a set of communication channels with a total capacity C_i and total power dissipation $P_{i,\max}^{FH}$. Its power consumption is given by [24]

$$P_i^{FH} \triangleq \frac{R_i^{FH}}{C_i} P_{i,\max}^{FH,X} = \alpha_i R_i^{FH,X}, \tag{24}$$

where $\alpha_i \triangleq P_{i,\max}^{FH}/C_i$ and R_i^{FH} is already defined in (3).

From (22) and (24), the total network power consumption is

$$\begin{aligned} P_{\text{total}}^{\text{Full},X} &\triangleq \sum_{i \in \mathcal{K}_R} (P_i^{RRH,X} + P_i^{FH,X}) \\ &= \sum_{i \in \mathcal{K}_R} \left(\beta_i P_i^{Tx,X} + b_i P_{i,\Delta} + \alpha_i R_i^{FH,X} \right) + P_s, \end{aligned} \tag{25}$$

where $P_{i,\Delta} \triangleq P_{i,a} - P_{i,s}$, $P_s \triangleq \sum_{i \in \mathcal{K}_R} P_{i,s}$. Since P_s is a part of the power consumption that is constant even when the network serves no user, this part has no contribution on the data transmission (DT) process. Therefore, the total power consumption on the downlink is

$$\begin{aligned} P_{\text{total}}^{\text{DT},X} &\triangleq P_{\text{total}}^{\text{Full},X} - P_s \\ &= \sum_{i \in \mathcal{K}_R} \left(\beta_i P_i^{Tx,X} + b_i P_{i,\Delta} + \alpha_i R_i^{FH,X} \right). \end{aligned} \tag{26}$$

III. Spectral Efficiency Maximization

First, we are interested in maximizing the spectral efficiency in (9)/(14) by jointly optimizing the RRH-group/user association, RRH activation, data rate allocation and precoders. Let us define $\boldsymbol{R} \triangleq \{R_g\}_{g \in \mathcal{K}_G}$. For a given cache state information $\{c_{g,i}\}_{g \in \mathcal{K}_G, i \in \mathcal{K}_R}$, the design problem of interest under the multicast scheme is formulated as

$$\max_{\boldsymbol{R}, \bar{\boldsymbol{F}}, \boldsymbol{a}, \boldsymbol{b}} \ \eta_{\text{SE}}^{\text{multicast}} \tag{27a}$$

$$\text{s.t. } (16), (18) \tag{27b}$$

$$\sum_{i \in \mathcal{K}_R} a_{g,i} \geq 1, \forall g \in \mathcal{K}_G \tag{27c}$$

$$\sum_{g \in \mathcal{K}_G} \langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_g \bar{\boldsymbol{F}}_g^H \bar{\boldsymbol{E}}_i \rangle \leq P_i, \forall i \in \mathcal{K}_R \tag{27d}$$

$$R_{\text{QoS}} \leq R_g, \forall g \in \mathcal{K}_G \tag{27e}$$

$$R_g \leq r_{g,k}(\bar{\boldsymbol{F}}), \forall g \in \mathcal{K}_G, k \in \mathcal{G}_g \tag{27f}$$

$$W \sum_{g \in \mathcal{K}_G} a_{g,i}(1 - c_{g,i}) R_g \leq C_i, \forall i \in \mathcal{K}_R \,. \tag{27g}$$

Here, constraint (27c) guarantees that there exists at least one active RRH to serve each multicast group. Constraint (27d) is the per-RRH power constraint (5) via (4). Constraint (27e) imposes a minimum rate $R_{\text{QoS}} \geq 0$. Constraint (27f) is indeed (7). Constraint (27g) expresses the bottleneck at fronthaul link $i \in \mathcal{K}_R$ with the limited capacity $C_i \geq 0$, i.e., $R_i^{FH,\text{multicast}} \leq C_i$, with $R_i^{FH,\text{multicast}}$ found in (3).

Similarly, the design problem under the unicast scheme is formulated as follows

$$\max_{\boldsymbol{R}, \bar{\boldsymbol{F}}, \tilde{\boldsymbol{a}}, \boldsymbol{b}} \ \eta_{\text{SE}}^{\text{unicast}} \tag{28a}$$

$$\text{s.t. } (17), (19) \tag{28b}$$

$$\sum_{i\in\mathcal{K}_R}\tilde{a}_{k,i}\geq 1,\forall k\in\mathcal{K}_U \tag{28c}$$

$$\sum_{k\in\mathcal{K}_U}\langle\bar{\boldsymbol{E}}_i^H\widetilde{\boldsymbol{F}}_k\widetilde{\boldsymbol{F}}_k^H\bar{\boldsymbol{E}}_i\rangle\leq P_i,\forall i\in\mathcal{K}_R\,. \tag{28d}$$

$$R_{\text{QoS}}\leq R_k,\forall k\in\mathcal{K}_U \tag{28e}$$

$$R_k\leq\tilde{r}_k(\tilde{\boldsymbol{F}}),\forall k\in\mathcal{K}_U \tag{28f}$$

$$W\sum_{g\in\mathcal{K}_G}(1-c_{g,i})\max_{k\in\mathcal{G}_g}\{\tilde{a}_{k,i}R_k\}\leq C_i,\forall i\in\mathcal{K}_R \tag{28g}$$

Due to the combinatorial and nonconvex nature of the mixed-integer nonlinear programs (27) and (28), finding their globally optimal solutions is challenging. Instead of aiming for global optimality, this paper proposes the following approaches that are suitable for practical implementation.

We rewrite problem (27) in its epigraph form [28] as follows.

$$\max_{\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b}}\ \sum_{g\in\mathcal{K}_G}|\,\mathcal{G}_g\,|R_g \tag{29a}$$

$$\text{s.t. }(20),(27c),(27e),(27f) \tag{29b}$$

$$\langle\bar{\boldsymbol{E}}_i^H\bar{\boldsymbol{F}}_g\bar{\boldsymbol{F}}_g^H\bar{\boldsymbol{E}}_i\rangle\leq u_{g,i},\forall g\in\mathcal{K}_G,i\in\mathcal{K}_R \tag{29c}$$

$$u_{g,i}\leq a_{g,i}P_i,\forall g\in\mathcal{K}_G,i\in\mathcal{K}_R \tag{29d}$$

$$\sum_{g\in\mathcal{K}_G}u_{g,i}\leq P_i,\forall i\in\mathcal{K}_R \tag{29e}$$

$$0\leq v_{g,i}\leq(1-c_{g,i})C_i,\forall g\in\mathcal{K}_G,i\in\mathcal{K}_R \tag{29f}$$

$$\sum_{g\in\mathcal{K}_G}v_{g,i}\leq C_i,\forall i\in\mathcal{K}_R \tag{29g}$$

$$a_{g,i}(1-c_{g,i})R_g\leq v_{g,i},\forall g\in\mathcal{K}_G,i\in\mathcal{K}_R \tag{29h}$$

$$a_{g,i}\in\{0,1\},b_i\in\{0,1\},\forall g\in\mathcal{K}_G,i\in\mathcal{K}_R \tag{29i}$$

where $\boldsymbol{\delta}\triangleq(\bar{\boldsymbol{F}},\boldsymbol{u},\boldsymbol{v}),\boldsymbol{u}\triangleq\{u_{g,i}\}_{g\in\mathcal{K}_G,i\in\mathcal{K}_U},\boldsymbol{v}\triangleq\{v_{g,i}\}_{g\in\mathcal{K}_G,i\in\mathcal{K}_R}$; (20), (29c)–(29e) and (29i) follow from (16), (18) and (27d); (29f)-(29h) follow from (27g). Solving problem (29) is challenging due to the nonconvex nonlinear constraints (27f), (29h) and (29i). Similarly, problem (28) can be rewritten as

$$\max_{\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b}}\ \sum_{k\in\mathcal{K}_U}R_k \tag{30a}$$

$$\text{s.t. }(21),(28c),(28e),(28f) \tag{30b}$$

$$\langle\bar{\boldsymbol{E}}_i^H\bar{\boldsymbol{F}}_k\bar{\boldsymbol{F}}_k^H\bar{\boldsymbol{E}}_i\rangle\leq u_{k,i},\forall k\in\mathcal{K}_U,i\in\mathcal{K}_R \tag{30c}$$

$$u_{k,i}\leq a_{k,i}P_i,\forall k\in\mathcal{K}_U,i\in\mathcal{K}_R \tag{30d}$$

$$\sum_{k\in\mathcal{K}_U}u_{k,i}\leq P_i,\forall i\in\mathcal{K}_R \tag{30e}$$

$$v_{g,i}\leq(1-c_{g,i})C_i,\forall g\in\mathcal{K}_G,i\in\mathcal{K}_R \tag{30f}$$

$$\sum_{g\in\mathcal{K}_G}v_{g,i}\leq C_i,\forall i\in\mathcal{K}_R \tag{30g}$$

$$a_{k,i}(1-c_{g,i})R_k\leq v_{g,i},\forall k\in\mathcal{G}_g,i\in\mathcal{K}_R \tag{30h}$$

$$a_{k,i}\in\{0,1\},b_i\in\{0,1\},\forall k\in\mathcal{K}_U,i\in\mathcal{K}_R \tag{30i}$$

where (30f)–(30h) follow from (28g).

It is noteworthy that the mathematical structures of problems (29) and (30) are similar. Therefore, the following algorithm devised for problem (29) of the multicast scheme can be straightforwardly adapted to solve problem (30) for the unicast scheme.

Noting that $x-x^2\geq 0,\forall x\in[0,1]$ and that [29]

$$x\in\{0,1\}\ \Leftrightarrow\ x-x^2=0\ \Leftrightarrow\ (x\in[0,1]\ \&\ x-x^2\leq 0), \tag{31}$$

we rewrite (29i) as

$$\sum_{i\in\mathcal{K}_R}\sum_{g\in\mathcal{K}_G}(a_{g,i}-a_{g,i}^2)+\sum_{i\in\mathcal{K}_R}(b_i-b_i^2)\leq 0 \tag{32}$$

$$0\leq a_{g,i}\leq 1,0\leq b_i\leq 1,\forall g\in\mathcal{K}_G,i\in\mathcal{K}_R\,. \tag{33}$$

Now, problem (29) is equivalent to the following problem with continuous variables $a_{g,i},b_i\in[0,1],\forall g\in\mathcal{K}_G,i\in\mathcal{K}_R$ with (32) and (33):

$$\min_{(\boldsymbol{R},\boldsymbol{p},\boldsymbol{a},\boldsymbol{b})\in\mathcal{H}}\ -\sum_{g\in\mathcal{K}_G}|\,\mathcal{G}_g\,|R_g, \tag{34}$$

where $\mathcal{H}\triangleq\{(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b})|(20),(27c),(27e),(27f),(29c)-(29h),(32),(33)\}$.

Let $\widehat{\mathcal{H}}\triangleq\{(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b})|(20),(27c),(27e),(27f),(29c)-(29h),(33)\}$ be the compact and feasible set of problem (34) without the nonconvex constraint (32). Problem (34) can then be seen as $\min_{(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}}\max_{\mu\geq 0}\mathcal{L}(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b},\mu)$ and its Lagrangian duality is $\sup_{\mu\geq 0}\min_{(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}}\mathcal{L}(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b},\mu)$, where

$$\begin{aligned}\mathcal{L}(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b},\mu)\triangleq&-\sum_{g\in\mathcal{K}_G}|\,\mathcal{G}_g\,|R_g\\&+\mu\Big(\sum_{g\in\mathcal{K}_G}\sum_{i\in\mathcal{K}_R}(a_{g,i}-a_{g,i}^2)+\sum_{i\in\mathcal{K}_R}(b_i-b_i^2)\Big)\end{aligned} \tag{35}$$

is the Lagrangian of (34) with the single constraint (32), and μ is the Lagrangian multiplier corresponding to (32). Next, we consider the problem

$$\min_{(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}}\mathcal{L}(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b},\mu) \tag{36}$$

which is related to (34) by the below proposition.

Proposition 1. The following statements hold:

(i) The value sequence of $S\triangleq\sum_{g\in\mathcal{K}_G}\sum_{i\in\mathcal{K}_R}(a_{g,i}-a_{g,i}^2)+\sum_{i\in\mathcal{K}_R}(b_i-b_i^2)$ at the solutions of (36) corresponding to μ is decreasing to 0 as $\mu\to+\infty$.

(ii) Problem (34) has strong duality, i.e.,

$$\min_{(\boldsymbol{R},\boldsymbol{p},\boldsymbol{a},\boldsymbol{b})\in\mathcal{H}}-\sum_{g\in\mathcal{K}_G}|\,\mathcal{G}_g\,|R_g=\sup_{\mu\geq 0}\min_{(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}}\mathcal{L}(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b},\mu) \tag{37}$$

and is therefore equivalent to (36) at the optimal solution $\mu^*\geq 0$ of the sup-min problem in (37).

Proof: See Appendix A. ∎

Theoretically, it is required to have $S_\mu=0$ in order to obtain an optimal μ^*. According to Proposition 1, S_μ decreases to 0 as $\mu\to+\infty$. Since there is always a numerical tolerance in computation, it is sufficient to accept $S_\mu<\varepsilon$ for some small ε with a large enough value of μ chosen. In our numerical experiment, for $\varepsilon=0.005$, we see that $\mu=200$ is enough to guarantee $S_\mu\leq\varepsilon$. Note that this way of choosing μ has been

widely used in the literature, e.g., [30], [31].

To handle the nonconvex constraint (29g), we note that if $c_{g,i} = 1$, (29g) becomes a convex constraint $0 \leq v_{g,i}$. Then, if $c_{g,i} = 0$, (29g) becomes $a_{g,i}R_g \leq v_{g,i}, \forall g \in \mathcal{K}_G$ and thus can be further rewritten as

$$(R_g + a_{g,i})^2 - (R_g - a_{g,i})^2 - 4v_{g,i} \leq 0, \quad \forall g \in \mathcal{K}_G, i \in \mathcal{K}_R\,. \tag{38}$$

We observe that a function $f(x,y) \triangleq (x-y)^2$ is jointly convex in (x,y). Upon applying the first-order Taylor series expansion at a given point $(x^{(n)}, y^{(n)})$, its convex lower bound is given by $2(x^{(n)}-y^{(n)})(x-y)-(x^{(n)}-y^{(n)})^2 \leq (x-y)^2$. Therefore, (38) can be approximated by

$$(R_g + a_{g,i})^2 - 2(R_g^{(n)} - a_{g,i}^{(n)})(R_g - a_{g,i}) + (R_g^{(n)} - a_{g,i}^{(n)})^2 - 4v_{g,i} \leq 0, \forall g \in \mathcal{K}_G, i \in \mathcal{K}_R, \tag{39}$$

where any point $(\boldsymbol{R}, \boldsymbol{\delta}, \boldsymbol{a}, \boldsymbol{b})$ satisfying (39) will also satisfy (38).

To handle the nonconvex constraint (27f), we approximate (27f) at a given point $(\boldsymbol{R}^{(n)}, \boldsymbol{\delta}^{(n)}, \boldsymbol{a}^{(n)}, \boldsymbol{b}^{(n)})$ by the convex constraint [32]

$$R_g \leq \Gamma_{g,k}^{(n)}(\bar{\boldsymbol{F}}), \forall g \in \mathcal{K}_G, k \in \mathcal{K}_U\,. \tag{40}$$

Here, $\Gamma_{g,k}(\bar{\boldsymbol{F}})$ is the concave lower bound of the nonconcave function $r_{g,k}(\bar{\boldsymbol{F}})$ in (27f) and given as where $\boldsymbol{\Phi}_{g,k} \triangleq \boldsymbol{\Pi}_{g,k}\boldsymbol{\Pi}_{g,k}^H + \boldsymbol{\Xi}_{g,k}$. The derivation of $\Gamma_k^{(n)}(\bar{\boldsymbol{F}})$ in (41) (see the top of the next page) and the proof of its concavity follow from the results of [32] and thus are omitted for brevity.

To handle the nonconvex cost function $\mathcal{L}(\boldsymbol{R}, \boldsymbol{\delta}, \boldsymbol{a}, \boldsymbol{b}, \mu)$ of (36), we observe that

$$a_{g,i}^2 \geq 2a_{g,i}^{(n)}a_{g,i} - (a_{g,i}^{(n)})^2 \;\text{ and }\; b_i^2 \geq 2b_i^{(n)}b_i - (b_i^{(n)})^2. \tag{42}$$

We then obtain a convex upper bound $\tilde{\mathcal{L}}(\boldsymbol{R}, \boldsymbol{\delta}, \boldsymbol{a}, \boldsymbol{b}, \mu)$ of $\mathcal{L}(\boldsymbol{R}, \boldsymbol{\delta}, \boldsymbol{a}, \boldsymbol{b}, \mu)$ at a given point $(\boldsymbol{R}^{(n)}, \boldsymbol{\delta}^{(n)}, \boldsymbol{a}^{(n)}, \boldsymbol{b}^{(n)})$ as follows

$$\begin{aligned}
&\tilde{\mathcal{L}}(\boldsymbol{R}, \boldsymbol{\delta}, \boldsymbol{a}, \boldsymbol{b}, \mu) \\
&\triangleq -\sum_{g\in\mathcal{K}_G} |\,\mathcal{G}_g\,| R_g + \mu\Big(\sum_{g\in\mathcal{K}_G}\sum_{i\in\mathcal{K}_R}\Big((1-2a_{g,i}^{(n)})a_{g,i} + (a_{g,i}^{(n)})^2\Big) + \sum_{i\in\mathcal{K}_R}\Big((1-2b_i^{(n)})b_i + (b_i^{(n)})^2\Big)\Big) \\
&\geq \mathcal{L}(\boldsymbol{R}, \boldsymbol{\delta}, \boldsymbol{a}, \boldsymbol{b}, \mu).
\end{aligned} \tag{43}$$

Now, problem (36) is approximated at a given point $(\boldsymbol{R}^{(n)}, \boldsymbol{\delta}^{(n)}, \boldsymbol{a}^{(n)}, \boldsymbol{b}^{(n)})$ as

$$\min_{(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}^{(n)}} \tilde{\mathcal{L}}(\boldsymbol{R}, \boldsymbol{\delta}, \boldsymbol{a}, \boldsymbol{b}, \mu) \tag{44}$$

where $\widehat{\mathcal{H}}^{(n)} \triangleq \{(\boldsymbol{R}, \boldsymbol{\delta}, \boldsymbol{a}, \boldsymbol{b}) | (20), (27c), (27e), (29c) - (29g), (33), (39), (40)\}$ is the convex feasible set of problem (44).

The steps to find the solution of problem (36) are outlined in Algorithm 1. Starting from a feasible initial point with an empirically chosen λ, we find an optimal solution $(\boldsymbol{R}^*, \boldsymbol{\delta}^*, \boldsymbol{a}^*, \boldsymbol{b}^*)$ of problem (44). This solution is then used as the initial point for the next iteration. The process stops as soon as no improvement of the objective function $\tilde{\mathcal{L}}$ of problem (44) is achieved. The following proposition provides insights into the convergence of Algorithm 1.

Algorithm 1 Spectral efficiency maximization for downlink C-RANs with content-centric multicast and edge caching

1: **Initialization**: Set $n := 1$. Choose a value of μ and choose an initial point $(\boldsymbol{R}^{(0)}, \boldsymbol{\delta}^{(0)}, \boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)})$ by Subroutine 1.
2: **repeat**
3: Update $n := n + 1$
4: Find the optimal solution $(\boldsymbol{R}^*, \boldsymbol{\delta}^*, \boldsymbol{a}^*, \boldsymbol{b}^*)$ by solving convex problem (44)
5: Update $(\boldsymbol{R}^{(n)}, \boldsymbol{\delta}^{(n)}, \boldsymbol{a}^{(n)}, \boldsymbol{b}^{(n)}) := (\boldsymbol{R}^*, \boldsymbol{\delta}^*, \boldsymbol{a}^*, \boldsymbol{b}^*)$
6: **until** convergence

Subroutine 1 Finding an initial point for Algorithm 1

1: **Initialization**: Set $n := 1$ and randomly select a point $(\boldsymbol{R}^{(0)}, \boldsymbol{\delta}^{(0)}, \boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}) \in \widehat{\mathcal{H}}$
2: **repeat**
3: Update $n := n + 1$
4: Find the optimal solution $(\boldsymbol{R}^*, \boldsymbol{\delta}^*, \boldsymbol{a}^*, \boldsymbol{b}^*)$ by solving convex problem (45)
5: Update $(\boldsymbol{R}^{(n)}, \boldsymbol{\delta}^{(n)}, \boldsymbol{a}^{(n)}, \boldsymbol{b}^{(n)}) := (\boldsymbol{R}^*, \boldsymbol{\delta}^*, \boldsymbol{a}^*, \boldsymbol{b}^*)$
6: **until** convergence

Proposition 2. Algorithm 1 converges to a Fritz John solution of problem (36).

Proof: See Appendix B. ∎

The computational complexity of solving (44) at each iteration of Algorithm 1 is polynomial in the number of variables and constraints. In particular, (44) can be transformed into an equivalent optimization problem that involves $N_v \triangleq (K_G + 3K_GK_R + 2K_GN_Rd + 1)$ real-valued scalar decision variables, $N_l \triangleq (4K_GK_R + 2K_G + 4K_R + 1)$ linear constraints and $N_q \triangleq (2K_GK_R + K_GK_U)$ quadratic constraints. Therefore, (44) requires a complexity of $\mathcal{O}(\sqrt{N_l + N_q}[N_v + N_l + N_q]N_v^2)$ [29], [33].

To find the initial point $(\boldsymbol{R}^{(0)}, \boldsymbol{\delta}^{(0)}, \boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}) \in \widehat{\mathcal{H}}$ of Algorithm 1, we solve problem (34) without constraint (32), which can be approximated by the following *convex* problem

$$\min_{(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}^{(n)}} \; -\sum_{g\in\mathcal{K}_G} |\,\mathcal{G}_g\,| R_g. \tag{45}$$

The steps of solving problem (34) without constraint (32) are detailed in Subroutine 1. After taking a random point $(\boldsymbol{R}^{(0)}, \boldsymbol{\delta}^{(0)}, \boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}) \in \widehat{\mathcal{H}}$, the initial point obtained by Subroutine 1 is located close to a solution of problem (34). Due to the equivalence between (34) and (36), the initial point obtained by Subroutine 1 will improve the solution obtained by solving (44), which is an approximation of (36).

IV. Energy Efficiency Maximization

In Section III, the problem of maximizing the spectral efficiency is addressed. However, only improving the spectral efficiency may lead to high power consumption, especially in dense networks [34]. Therefore, the aim of this section

$$\Gamma_{g,k}^{(n)}(\bar{\boldsymbol{F}}) \triangleq r_{g,k}(\bar{\boldsymbol{F}}^{(n)}) + \frac{2W}{\ln 2}\Re\left\{\left\langle \left(\left(\boldsymbol{\Phi}_{g,k}^{(n)} - \boldsymbol{\Pi}_{g,k}^{(n)}(\boldsymbol{\Pi}_{g,k}^{(n)})^H\right)^{-1}\boldsymbol{\Pi}_{g,k}^{(n)}\right)^H (\boldsymbol{\Pi}_{g,k}(\bar{\boldsymbol{F}}_g) - \boldsymbol{\Pi}_{g,k}^{(n)})\right\rangle\right\}$$
$$- \frac{W}{\ln 2}\left\langle \left(\left(\boldsymbol{\Phi}_{g,k}^{(n)} - \boldsymbol{\Pi}_{g,k}^{(n)}(\boldsymbol{\Pi}_{g,k}^{(n)})^H\right)^{-1} - (\boldsymbol{\Phi}_{g,k}^{(n)})^{-1}\right)^H (\boldsymbol{\Phi}_{g,k}(\bar{\boldsymbol{F}}) - \boldsymbol{\Phi}_{g,k}^{(n)})\right\rangle \leq r_{g,k}(\bar{\boldsymbol{F}}) \quad (41)$$

is to improve the spectral efficiency at the same time as reducing the power consumption. To this end, we maximize the energy efficiency of the considered content-centric C-RAN, which is defined as the ratio of the spectral efficiency and the total power consumption. With the cache state information $\{c_{g,i}\}_{g\in\mathcal{K}_G, i\in\mathcal{K}_R}$ known, the energy-efficient design problems under the multicast scheme is formulated as

$$\max_{\boldsymbol{R},\bar{\boldsymbol{F}},\boldsymbol{a},\boldsymbol{b}} \ \frac{\eta_{\text{SE}}^{\text{multicast}}}{P_{\text{total}}^{\text{DT,multicast}}} \quad (46a)$$
$$\text{s.t. } (16), (18), (27c) - (27g). \quad (46b)$$

Similarly, the energy-efficient design problem under the unicast scheme is formulated as

$$\max_{\boldsymbol{R},\bar{\boldsymbol{F}},\boldsymbol{a},\boldsymbol{b}} \ \frac{\eta_{\text{SE}}^{\text{unicast}}}{P_{\text{total}}^{\text{DT,unicast}}} \quad (47a)$$
$$\text{s.t. } (17), (19), (28c) - (28g). \quad (47b)$$

Note that problems (46) and (47) are similar in the mathematical structure. Therefore, the algorithm devised in the following can be adapted straightforwardly to find the solution for (47).

Since problem (46) is combinatorial, nonconvex and fractional, it is challenging to find its global optimality. Here, we propose an algorithm that is suitable for practical implementation. To this end, we rewrite problem (46) in an epigraph form as [28]

$$\max_{\phi,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b}} \ \phi \quad (48a)$$
$$\text{s.t. } (20), (27c), (27e), (27f), (29c) - (29i), \quad (48b)$$
$$\phi\rho \leq \sum_{g\in\mathcal{K}_U} |\,\mathcal{G}_g\,| R_g \quad (48c)$$
$$\rho \geq \sum_{i\in\mathcal{K}_R}\Big(\beta_i \sum_{g\in\mathcal{K}_G} \langle \bar{\boldsymbol{E}}_i^H \bar{\boldsymbol{F}}_g \bar{\boldsymbol{F}}_g^H \bar{\boldsymbol{E}}_i\rangle + b_i P_{i,\Delta}\Big) + \sum_{i\in\mathcal{K}_R}\alpha_i \sum_{g\in\mathcal{K}_G} v_{g,i} + P_s \quad (48d)$$

where $\boldsymbol{w} = (\boldsymbol{R}, \bar{\boldsymbol{F}}, \boldsymbol{u}, \boldsymbol{v}, \rho)$. Problem (48) is still challenging to solve due to the nonconvex constraints (27f), (29h), (29i) and (48c).

Following the procedure discussed in Section III, the binary constraint (29i) is replaced by (32) and (33), problem (48) is then rewritten as

$$\min_{(\phi,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})\in\mathcal{T}} \ -\phi, \quad (49)$$

where $\mathcal{T} \triangleq \{(\phi,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})|(20), (27c), (27e), (27f), (29c) - (29g), (32), (33), (48c), (48d)\}$. Similar to Proposition 1, with an empirically chosen value of μ, problem (49) is equivalent to the following problem with continuous variables $\boldsymbol{a}, \boldsymbol{b}$:

$$\min_{(\phi,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{T}}} \ \bar{\mathcal{L}}(\phi,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b},\mu) = -\phi + \mu\left(\sum_{i\in\mathcal{K}_R}\sum_{g\in\mathcal{K}_G}(a_{g,i} - a_{g,i}^2) + \sum_{i\in\mathcal{K}_R}(b_i - b_i^2)\right) \quad (50)$$

where $\widehat{\mathcal{T}} \triangleq \{(\phi,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})|(20), (27c), (27e), (27f), (29c) - (29g), (33), (48c), (48d)\}$.

With $a_{g,i}, b_i \in [0,1], \forall g \in \mathcal{K}_G, i \in \mathcal{K}_R$, (27f) and (29h) are approximated by the convex constraints (40) and (39), respectively. Here, the remaining nonconvex constraint (48c) is rewritten as

$$(\phi+\rho)^2 - (\phi-\rho)^2 - 4\sum_{g\in\mathcal{K}_G}|\,\mathcal{G}_g\,|R_g \leq 0. \quad (51)$$

Similar to (38), we further approximate (51) by the following convex constraint:

$$(\phi+\rho)^2 - 2(\phi^{(n)} - \rho^{(n)})(\phi-\rho) + (\phi^{(n)} - \rho^{(n)})^2 - 4\sum_{g\in\mathcal{K}_G}|\,\mathcal{G}_g\,|R_g \leq 0. \quad (52)$$

Hence, at a given point $(\phi^{(n)}, \boldsymbol{w}^{(n)}, \boldsymbol{a}^{(n)}, \boldsymbol{b}^{(n)})$, problem (49) can be approximated by the following *convex* problem

$$\min_{(\phi,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{T}}^{(n)}} \ \widehat{\mathcal{L}}(\phi,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b},\mu), \quad (53)$$

where $\widehat{\mathcal{T}}^{(n)} \triangleq \{(\phi,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})|(20), (27c), (27e), (29c) - (29g), (33), (39), (40), (48d), (52)\}$ is the compact, convex feasible set of problem (53) and

$$\widehat{\mathcal{L}}(\phi,\boldsymbol{w},\boldsymbol{a},\boldsymbol{b},\mu) \triangleq -\phi + \mu\Big(\sum_{g\in\mathcal{K}_G}\sum_{i\in\mathcal{K}_R}\Big((1-2a_{g,i}^{(n)})a_{g,i} + (a_{g,i}^{(n)})^2\Big) + \sum_{i\in\mathcal{K}_R}\Big((1-2b_i^{(n)})b_i + (b_i^{(n)})^2\Big)\Big). \quad (54)$$

We propose Algorithm 2 to solve problem (50) for maximizing the energy efficiency of the downlink of the content-centric C-RANs with edge caching. Starting from a feasible initial point with an empirically chosen μ, problem (53) is solved to obtain the optimal solution. This solution is then used as the initial point for the next iteration. The process terminates as soon as no improvement in the objective function $\widetilde{\mathcal{L}}$ of (53) is achieved. By a similar argument to the proof of Proposition 2, Algorithm 2 converges to a Fritz John solution of problem (50). Note that (53) involves $M_v \triangleq (3K_GK_R + 2K_GN_Rd + K_G + 2)$ real-valued scalar decision variables, $M_l \triangleq (4K_GK_R + 2K_G + 4K_R + 1)$ linear constraints and $M_q \triangleq (2K_GK_R + K_GK_U + 2)$ quadratic constraints.

Algorithm 2 Energy efficiency maximization of the downlink C-RANs with content-centric multicast and edge caching

1: **Initialization**: Set $n := 1$. Choose a value of λ and choose an initial point $(\phi^{(0)}, \boldsymbol{w}^{(0)}, \boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)})$ by Subroutine 2.
2: **repeat**
3: Update $n := n + 1$
4: Find the optimal solution $(\phi^*, \boldsymbol{w}^*, \boldsymbol{a}^*, \boldsymbol{b}^*)$ by solving convex problem (53)
5: Update $(\phi^{(n)}, \boldsymbol{w}^{(n)}, \boldsymbol{a}^{(n)}\, \boldsymbol{b}^{(n)}) := (\phi^*, \boldsymbol{w}^*, \boldsymbol{a}^*, \boldsymbol{b}^*)$
6: **until** convergence

Subroutine 2 Find an initial point for Algorithm 2

1: **Initialization**: Set $n := 1$ and choose randomly a point $(\phi^{(0)}, \tilde{\boldsymbol{w}}^{(0)}, \boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}) \in \widehat{\mathcal{T}}$
2: **repeat**
3: Update $n := n + 1$
4: Find the optimal solution $(\phi^*, \boldsymbol{w}^*, \boldsymbol{a}^*, \boldsymbol{b}^*)$ by solving convex problem (55)
5: Update $(\phi^{(n)}, \boldsymbol{w}^{(n)}, \boldsymbol{a}^{(n)}, \boldsymbol{b}^{(n)}) := (\phi^*, \boldsymbol{w}^*, \boldsymbol{a}^*, \boldsymbol{b}^*)$
6: **until** convergence

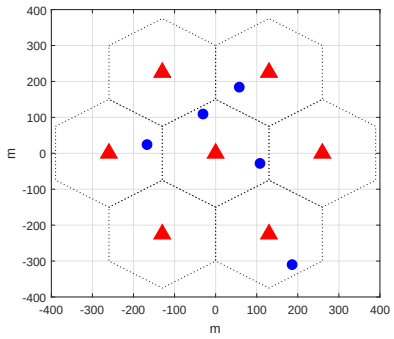


Fig. 3. Network simulation scenario with $K_R = 7$ fixed eRHHs and $K_U = 5$ randomly positioned users.

The computational complexity required to solve (53) is thus $\mathcal{O}(\sqrt{M_l + M_q}[M_v + M_l + M_q]M_v^2)$ [29], [33].

To find the initial solution $(\phi^{(0)}, \boldsymbol{w}^{(0)}, \boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}) \in \widehat{\mathcal{T}}$ of Algorithm 2, we solve problem (49) without constraint (32), which is approximated by

$$\min_{(\phi, \boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}) \in \widehat{\mathcal{V}}^{(n)}} \quad -\phi. \tag{55}$$

The steps for solving problem (49) without constraint (32) is detailed in Subroutine 2. From a random point $(\phi^{(0)}, \boldsymbol{w}^{(0)}, \boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}) \in \widehat{\mathcal{T}}$, the initial point obtained by Subroutine 2 is located close to a solution of problem (49). Because of the equivalence between (49) and (50), this initial point will improve the solution obtained by solving (53), which is an inner approximation of (50).

V. Numerical Examples

A. Simulation Setup

We simulate a hexagonal multicell C-RAN in Fig. 3 where the locations of the $K_R = 7$ RRHs are fixed. The $K_U = 5$ users are uniformly and independently placed in the RRHs' coverage area, excluding the circular area with radius of 50 m around each RRH [13]. The LTE parameters used in our numerical examples are listed in Table I. We further assume that each RRH is equipped with $N_r = 8$ antennas and each user with $N_u = 1$ antennas. At each RRH, the active mode and the sleep mode consume 84W and 56W of power, respectively. Also the slope of transmit power is taken as $\beta_i = \beta = 2.8$ and $\alpha_i = \alpha = 5$ for all $i \in \mathcal{K}_R$ [24]. We set $d = 2$, $P_i = P$, $C_i = C = 100$ Mbps for all $i \in \mathcal{K}_R$, and $\boldsymbol{\Sigma}_k = \sigma^2 \boldsymbol{I}$ for all $k \in \mathcal{K}_U$. Also, we set $R_{\text{QoS}} = 0.1$ Mbps.

TABLE I
LTE Parameters Used in Numerical Examples [35]

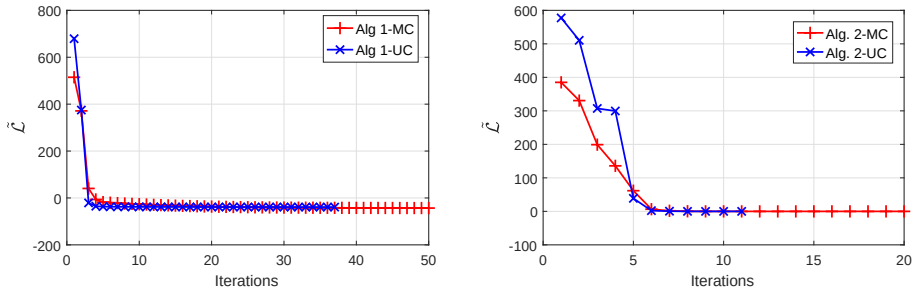
Parameters	Values
Distance between adjacent RRHs	0.3 km
Total bandwidth	10 MHz
Standard deviation of log-normal shadowing	10 dB
Path loss at distance d (km)	$140.7 + 36.7 \log_{10}(d)$ dB
Noise variance $\sigma_k^2 = \sigma^2$	-174 dBm/Hz
Maximum RRH transmit power	24 dBm

Each RRH's cache has the same size of $B_i = B, \forall i \in \mathcal{K}_R$ which can store up to $B = 5$ files randomly chosen from a library of $F = 100$ files. For demonstration purposes, we consider a heuristic caching strategy, namely Popularity-based Caching, at the content placement phase [13]. Each RRH stores the most popular files until its cache is full. Each user then independently requests one file from the library, where the content popularity distributions is described as follows. Among the F files, one file has a request probability of 0.5, and $F - 1$ files share the remaining request probability according to the Zipf distribution as $\Pr(f) = \frac{r_f^{-\gamma}}{\sum_{f \in \mathcal{F}} r_f^{-\gamma}}$, where r_f is the popularity rank of file f, γ is the skewness parameter and $\sum_{f \in \mathcal{F} \setminus \{f | r_f = 1\}} \Pr(f) = 0.5$.

Each numerical result is obtained by averaging over 200 simulation trials. In each trial, we independently generate a new set of user locations, channel realizations and user requests [13]. For the given set $\mathcal{F}_{\text{req}}$ of requested files available at the BBU in each delivery interval, the cache state information $\{c_{g,i}\}$ in (1) are recorded by checking $\mathcal{F}_{\text{req}}$ against the local caches of RRHs. We choose $\mu = 200$ and our simulations confirm that this value of μ produces results with sufficient accuracy.

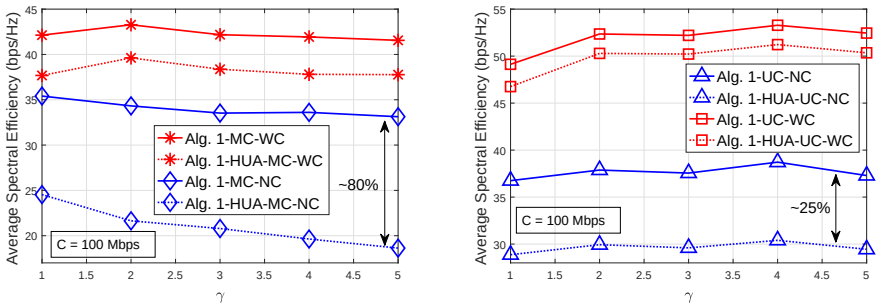
B. Performance of Proposed Joint Optimization under a Fixed Caching Strategy and a Fixed File Popularity Distribution

Using a fixed caching strategy for a specific file popularity distribution, we evaluate the performance of Algorithm 1 and Algorithm 2 (respectively referred to as Alg. 1 and Alg. 2 in the figures) in content delivery. The following 'transmission scheme - with/no cache' scenarios are implemented as examples: *MC-NC* (multicast - no caching) and *UC-WC* (unicast - with caching). For comparison purposes, we also propose two benchmark schemes (referred to as Alg. 1-HUA and Alg. 2-HUA) where heuristic UA strategies are adopted. Here, each

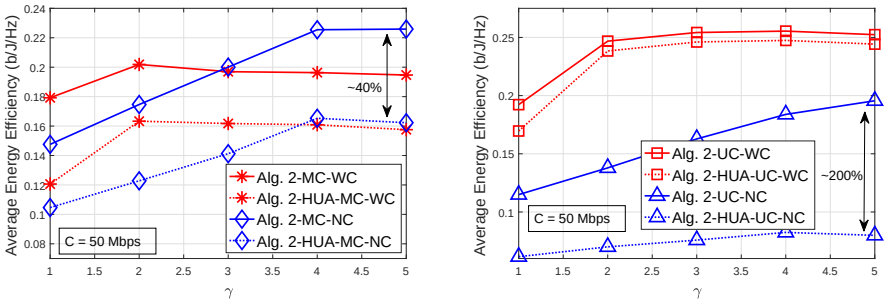


(a) (b)

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



(a) Average spectral efficiency - Multicast (b) Average spectral efficiency - Unicast



(c) Average energy efficiency - Multicast (d) Average energy efficiency - Unicast

Fig. 5. Performance of Algorithm 1 under 'Muticast/Unicast - No/With Cache' scenarios in different skewness parameter γ. Two benchmark algorithms Alg. 1-HUA and Alg. 2-HUA are also presented for comparison.

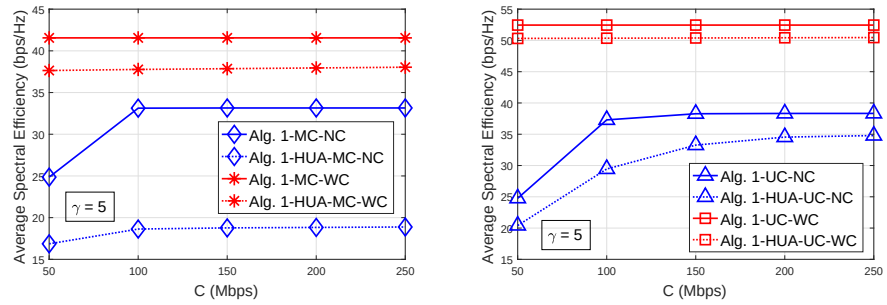
group or user is assigned heuristically to the RRHs that store the requested files and also to the N_c RRHs that have the largest channel gain to the worst user. The parameter N_c is determined empirically.

Figs. 4(a) and 4(b) confirm the theoretical convergence for Algorithm 1 and Algorithm 2, respectively. In the example used to plot Fig. 4, both algorithms converge in fewer than 50 iterations. It is worth noting that each iteration of these proposed algorithms corresponds to solving at most a simple convex program (44) and (53). Therefore, low computational cost can be expected.

Figs. 5(a), 5(b), 6(a) and 6(b) compare the spectral efficiency performance of Alg. 1 against Alg. 1-HUA while Figs. 5(c), 5(d), 6(c) and 6(d) plot the energy efficiency performance of Alg. 2 against Alg. 2-HUA. It can be observed that Alg. 1 and Alg. 2 perform best in all scenarios of multicast/unicast - with/no caching. Particularly, the sum rate obtained by Alg. 1 is up to 180% that by the Alg. 1-HUA scheme while the energy efficiency achieved by Alg. 2 is up to 300% that by the Alg. 2-HUA. This result can be explained by noting the extra dimension of UA optimization of the proposed algorithms, allowing better solutions for data rate allocation and precoders of problems (27) and (46) compared to the heuristic HUA approach.

C. Multicast-Unicast Comparison and Evaluating the Effects of Caching Strategies under Proposed Joint Design

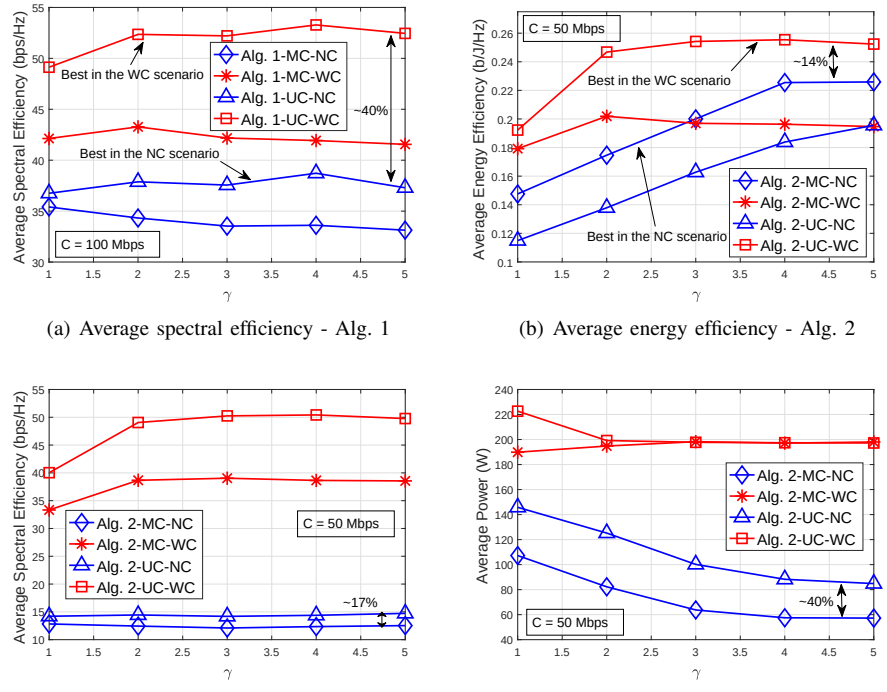
As seen from Figs. 7(a) and 8(a), the unicast scheme outperforms the multicast in terms of spectral efficiency in both cache and cache-less scenarios. This result is reasonable because the performance of the multicast scheme is limited by the condition of the worst user in each multicast group. Moreover, our proposed joint design with UA and cache makes the bottleneck on the fronthaul links less problematic. Although



(a) Average spectral efficiency - Multicast (b) Average spectral efficiency - Unicast

(c) Average energy efficiency - Multicast (d) Average energy efficiency - Unicast

Fig. 6. Performance of Algorithm 1 under 'Muticast/Unicast - No/With Cache' scenarios with different fronthaul capacities. Two baseline algorithms Alg. 1-HUA and Alg. 2-HUA are also presented for comparison.



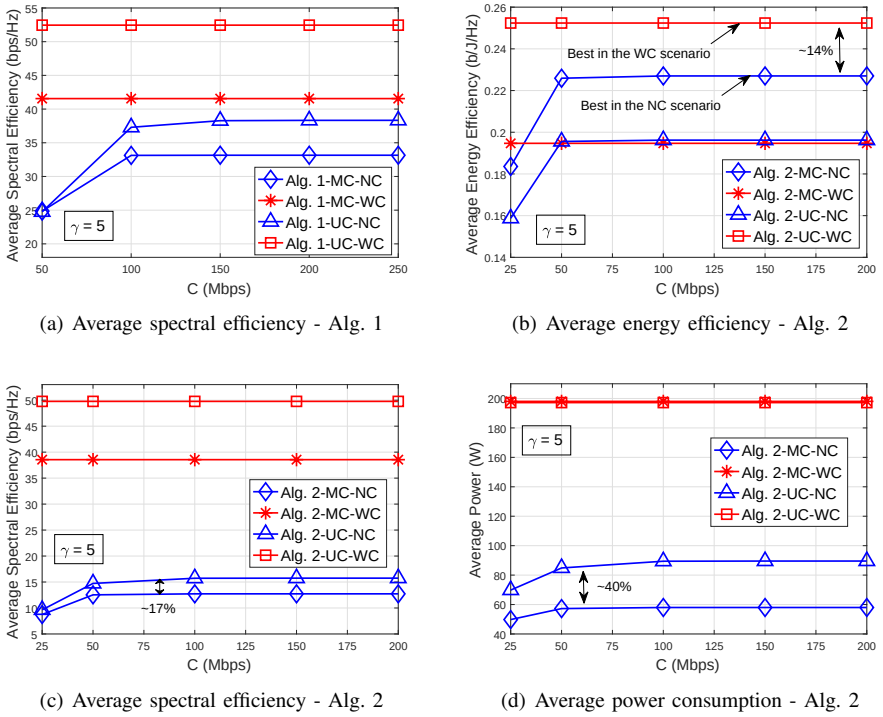
(a) Average spectral efficiency - Alg. 1 (b) Average energy efficiency - Alg. 2

(c) Average spectral efficiency - Alg. 2 (d) Average power consumption - Alg. 2

Fig. 7. Multicast versus unicast in terms of spectral and energy efficiencies - no/with cache in different skewness parameter γ

the unicast scheme requires more fronthaul traffic than the multicast counterpart, this disadvantage seems not to have a significant impact on the spectral efficiency performance of the unicast scheme.

Figs. 7(b) and 8(b) illustrate the energy efficiency comparison between the multicast and unicast schemes. In the case of no cache, multicasting is better than unicasting, especially when γ is large. This is because the multicast scheme requires less power consumption and obtains not much lower spectral efficiency than the unicast scheme. Specifically, the power



(a) Average spectral efficiency - Alg. 1

(b) Average energy efficiency - Alg. 2

(c) Average spectral efficiency - Alg. 2

(d) Average power consumption - Alg. 2

Fig. 8. Multicast versus unicast in terms of spectral and energy efficiencies - no/with cache under different fronthaul capacities

consumed by multicating is 40% lower than unicasting as seen from Figs. 7(d) and 8(d), while the spectral efficiency obtained by multicasting is 17% lower than unicasting as seen from Figs. 7(c) and 8(c). The power consumption of the fronthaul links is a dominating part of the total consumption. Since a larger value of γ means fewer groups, the fronthaul rate is more likely to be reduced due to the worst user's condition in each group. More power on the fronthaul links is thus saved. However, in the system with cache, the unicast scheme outperforms the multicast scheme. In this case, the main bottleneck on the fronthaul links is effectively handled by the UA and caching. Figs. 7(c), 8(c), 7(d) and 8(d) show that while the power consumption of these schemes is almost the same, the unicast scheme achieves higher spectral efficiency than the multicast scheme because it does not suffer from the worst user's condition in each group as in multicasting. Figs. 7(b) and 8(b) also implies that in terms of energy efficiency, the unicast scheme is the best option for the system with cache while the multicast scheme for the system without cache.

Finally, Figs. 7 and 8 shows that the caching contributes to higher spectral and energy efficiencies of the considered C-RAN. Particularly, as seen from Figs. 7(a) and 8(a), a 20% spectral efficiency gain is obtained with cache in comparison with that of a cache-less system. Also, from Figs. 7(b) and 8(b), the energy efficiency achieved by the system with cache is 14% higher than that by the cache-less system.

VI. Conclusion

In this paper, we have jointly designed user association, remote radio head activation, data delivery rate allocation and precoding for the downlink of a content-centric C-RAN with edge caching. Mixed-integer optimization problems have been formulated with the objective of maximizing the spectral and energy efficiencies. The requirements on data delivery rates, limited fronthaul capacity and maximum RRH transmit powers have been included in the design. Upon applying a range of optimization techniques, we have successfully solved these challenging problems and proposed iterative algorithms that are guaranteed to converge to Fritz John solutions. Numerical results with practical parameter settings have shown that our joint designs markedly improve both spectral and energy efficiency performances of the considered C-RAN under the example caching strategy. Moreover, they show that the unicast scheme always offers higher spectral efficiency than its multicast counterpart. For the energy efficiency, unicasting is also the best option for the system with cache while the multicast scheme is best for the system without cache. Edge caching has been shown to improve both the spectral and energy efficiencies.

Appendix A
Proof of Proposition 1

Let $\theta(\mu)$ and $(\boldsymbol{R}_\mu, \boldsymbol{\delta}_\mu, \boldsymbol{a}_\mu, \boldsymbol{b}_\mu)$ be the optimal value and the optimal solution of problem (36) for a given μ, respectively. For simplicity, we use only η for $\eta_{\text{SE}}^{\text{multicast}} = \sum_{g\in\mathcal{K}_G} |\mathcal{G}_g| R_g$.

Due to a duality gap between the optimal value of problem (34) and the optimal value of its dual problem, it holds that

$$\begin{aligned}\sup_{\mu\geq 0}\theta(\mu) &= \sup_{\mu\geq 0}\min_{(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}} \mathcal{L}(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b},\mu)\\ &\leq \theta^* \triangleq \min_{(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b})\in\widehat{\mathcal{H}}}\max_{\mu\geq 0}\mathcal{L}(\boldsymbol{R},\boldsymbol{\delta},\boldsymbol{a},\boldsymbol{b},\mu), \qquad (56)\end{aligned}$$

where θ^* is the optimal value of problem (34). We observe that θ^* is finite since $\widehat{\mathcal{H}}$ is compact. Combining with (56), we have

$$\theta(\mu) \leq \theta^* < +\infty, \quad \forall \mu \geq 0. \tag{57}$$

(i) Let $S_\mu \triangleq \sum_{i\in\mathcal{K}_R}\sum_{g\in\mathcal{K}_G}((a_{g,i})_\mu - (a_{g,i})_\mu^2) + \sum_{i\in\mathcal{K}_R}((b_i)_\mu - (b_i)_\mu^2)$ be the value of S at $\boldsymbol{a}_\mu$ and $\boldsymbol{b}_\mu$. Then, $S_\mu \geq 0, \forall \mu$. Let $0 \leq \mu_1 < \mu_2$. Since $\theta(\mu_1)$ and $\theta(\mu_2)$ are respectively the optimal value of (36) for μ_1 and μ_2,

$$\theta(\mu_1) = -\eta_{\mu_1} + \mu_1 S_{\mu_1} \leq -\eta_{\mu_2} + \mu_1 S_{\mu_2}, \tag{58}$$
$$\theta(\mu_2) = -\eta_{\mu_2} + \mu_2 S_{\mu_2} \leq -\eta_{\mu_1} + \mu_2 S_{\mu_1}. \tag{59}$$

On one hand, adding these two inequalities yields $\mu_1 S_{\mu_1} + \mu_2 S_{\mu_2} \leq \mu_1 S_{\mu_2} + \mu_2 S_{\mu_1}$, which implies that $S_{\mu_2} \leq S_{\mu_1}$. We deduce that S_μ is decreasing and bounded below by 0 when μ is increasing. Hence,

$$S_\mu \to S_* \geq 0 \text{ as } \mu \to +\infty. \tag{60}$$

On the other hand, multiplying (58) and (59) by μ_2 and μ_1, respectively, followed by adding these results together gives

$$\mu_2(-\eta_{\mu_1}) + \mu_1(-\eta_{\mu_2}) \leq \mu_2(-\eta_{\mu_2}) + \mu_1(-\eta_{\mu_1}), \tag{61}$$

or equivalently, $-\eta_{\mu_2} \geq -\eta_{\mu_1}$. Therefore, $-\eta_\mu$ is increasing and hence bounded below as $\mu \to +\infty$.

Now, if $S_* > 0$, then $\theta(\mu) = -\eta_\mu + \mu S_\mu \to +\infty$ as $\mu \to +\infty$, which contradicts to (57). We must therefore have $S_* = 0$ and the proof of Proposition 1 (i) is completed.

(ii) Since the sequence $\{(\boldsymbol{R}_\mu, \boldsymbol{\delta}_\mu, \boldsymbol{a}_\mu, \boldsymbol{b}_\mu)\}_{\mu\geq 0} \subset \widehat{\mathcal{H}}$ is bounded, it has convergent subsequences. Let $(\boldsymbol{R}_*, \boldsymbol{\delta}_*, \boldsymbol{a}_*, \boldsymbol{b}_*)$ be any limit point of $\{(\boldsymbol{R}_\mu, \boldsymbol{\delta}_\mu, \boldsymbol{a}_\mu, \boldsymbol{b}_\mu)\}_\mu$ as $\mu \to +\infty$. We assume without loss of generality that $(\boldsymbol{R}_\mu, \boldsymbol{\delta}_\mu, \boldsymbol{a}_\mu, \boldsymbol{b}_\mu) \to (\boldsymbol{R}_*, \boldsymbol{\delta}_*, \boldsymbol{a}_*, \boldsymbol{b}_*)$. Then $(a_{g,i})_\mu \to (a_{g,i})_*, (b_i)_\mu \to (b_i)_*, \forall g \in \mathcal{K}_G, i \in \mathcal{K}_R$ and $S_\mu \to S_* \triangleq \sum_{i\in\mathcal{K}_R}\sum_{g\in\mathcal{K}_G}((a_{g,i})_* - (a_{g,i})_*^2) + \sum_{i\in\mathcal{K}_R}((b_i)_* - (b_i)_*^2)$. Using (i), we get $S_* = 0$. Hence, $(\boldsymbol{a}_*, \boldsymbol{b}_*)$ satisfies (32). Also, because $(\boldsymbol{R}_\mu, \boldsymbol{\delta}_\mu, \boldsymbol{a}_\mu, \boldsymbol{b}_\mu) \in \widehat{\mathcal{H}}, \forall \mu \geq 0$, it holds that $(\boldsymbol{R}_*, \boldsymbol{\delta}_*, \boldsymbol{a}_*, \boldsymbol{b}_*) \in \widehat{\mathcal{H}}$, which implies that $(\boldsymbol{R}_*, \boldsymbol{\delta}_*, \boldsymbol{a}_*, \boldsymbol{b}_*) \in \mathcal{H}$. Therefore, $(\boldsymbol{R}_*, \boldsymbol{\delta}_*, \boldsymbol{a}_*, \boldsymbol{b}_*)$ is a feasible point of problem (34). By the definition of $\theta(\mu)$, it is true that

$$\sup_{\mu\geq 0} \theta(\mu) \geq \theta(\mu) = -\eta_\mu + \mu S_\mu \geq -\eta_\mu, \forall \mu \geq 0. \tag{62}$$

Letting $\mu \to +\infty$ yields

$$\sup_{\mu\geq 0} \theta(\mu) \geq -\eta_* \geq \theta^*. \tag{63}$$

Combining with (56), we obtain that $\sup_{\mu\geq 0} \theta(\mu) = -\eta_* = \theta^*$, which proves (37) and also implies that $(\boldsymbol{R}_*, \boldsymbol{\delta}_*, \boldsymbol{a}_*, \boldsymbol{b}_*)$ is an optimal solution of (34). The proof is complete.

APPENDIX B
PROOF OF PROPOSITION 2

Because $(\boldsymbol{R}^{(n+1)}, \boldsymbol{a}^{(n+1)}, \boldsymbol{b}^{(n+1)})$ is the optimal solution of problem (44) at a given point $(\boldsymbol{R}^{(n)}, \boldsymbol{a}^{(n)}, \boldsymbol{b}^{(n)})$, the approximation step (43) at iteration $(n+1)$ gives

$$\begin{aligned}
&\widetilde{\mathcal{L}}(\boldsymbol{R}^{(n+1)}, \boldsymbol{a}^{(n+1)}, \boldsymbol{b}^{(n+1)}, \mu) \\
&= -\sum_{g\in\mathcal{K}_G} |\,\mathcal{G}_g\,| R_g^{(n+1)} \\
&\quad + \mu\Big(\sum_{k\in\mathcal{K}_U}\sum_{i\in\mathcal{K}_R}\Big((1-2a_{k,i}^{(n)})a_{k,i}^{(n+1)} + (a_{k,i}^{(n)})^2\Big) \\
&\quad + \sum_{i\in\mathcal{K}_R}\Big((1-2b_i^{(n)})b_i^{(n+1)} + (b_i^{(n)})^2\Big)\Big) \\
&\leq -\sum_{g\in\mathcal{K}_G} |\,\mathcal{G}_g\,| R_g^{(n)} + \mu\Big(\sum_{k\in\mathcal{K}_U}\sum_{i\in\mathcal{K}_R}\Big((1-2a_{k,i}^{(n)})a_{k,i}^{(n)} \\
&\quad + (a_{k,i}^{(n)})^2\Big) + \sum_{i\in\mathcal{K}_R}\Big((1-2b_i^{(n)})b_i^{(n)} + (b_i^{(n)})^2\Big)\Big) \\
&= \widetilde{\mathcal{L}}(\boldsymbol{R}^{(n)}, \boldsymbol{a}^{(n)}, \boldsymbol{b}^{(n)}, \mu).
\end{aligned} \tag{64}$$

Therefore, once initialized from a feasible point $\bar{\boldsymbol{F}}^{(0)}$ given by Subroutine 1, Algorithm 1 generates a monotone sequence $\{\widetilde{\mathcal{L}}(\boldsymbol{R}^{(n)}, \boldsymbol{a}^{(n)}, \boldsymbol{b}^{(n)}, \mu)\}$ of improved feasible solutions for (44). On the other hand, since the sequence $\{\widetilde{\mathcal{L}}(\boldsymbol{R}^{(n)}, \boldsymbol{a}^{(n)}, \boldsymbol{b}^{(n)}, \mu)\}$ is bounded from below by constraint (27e), it converges. As a result, the convergence of Algorithm 1 is guaranteed in the sense that $\Big(\widetilde{\mathcal{L}}(\boldsymbol{R}^{(n+1)}, \boldsymbol{a}^{(n+1)}, \boldsymbol{b}^{(n+1)}, \mu) - \widetilde{\mathcal{L}}(\boldsymbol{R}^{(n)}, \boldsymbol{a}^{(n)}, \boldsymbol{b}^{(n)}, \mu)\Big) \to 0$ as $n \to +\infty$.

Now, for simplicity and by following the procedure in [36], we rewrite (36) in the following form

$$\min_{\boldsymbol{z}} \quad \varphi_0(\boldsymbol{z}) \tag{65a}$$
$$\text{s.t.} \quad \psi_j(\boldsymbol{z}) \leq 0, \forall j \in \{1, ..., r\} \tag{65b}$$
$$\varphi_j(\boldsymbol{z}) \leq 0, \forall j \in \{1, ..., s\} \tag{65c}$$

where $\boldsymbol{z}$ represents the variables of (36); φ_0 is the objective function; ψ_j are the convex functions and φ_j are the nonconvex functions.

Following (39), (40) and (43), problem (44) which is a convex approximation of problem (36) can be recast into a simplified form as

$$\min_{\boldsymbol{z}} \quad \tilde{\varphi}_0(\boldsymbol{z}, \boldsymbol{z}^{(n)}) \tag{66a}$$
$$\text{s.t.} \quad \psi_j(\boldsymbol{z}) \leq 0, \forall j \in \{1, ..., r\} \tag{66b}$$
$$\tilde{\varphi}_j(\boldsymbol{z}, \boldsymbol{z}^{(n)}) \leq 0, \forall j \in \{1, ..., s\}, \tag{66c}$$

where the objective function φ_0 and each nonconvex function φ_j are respectively approximated by convex functions $\tilde{\varphi}_0$ and $\tilde{\varphi}_j$ for a given point $\boldsymbol{z}^{(n)}$. We also note from (39), (40) and (43) that, for every $j \in \{0, 1, ..., s\}$,

$$\varphi_j(\boldsymbol{z}) \leq \tilde{\varphi}_j(\boldsymbol{z}, \boldsymbol{z}^{(n)}), \tag{67a}$$
$$\varphi_j(\boldsymbol{z}^{(n)}) = \tilde{\varphi}_j(\boldsymbol{z}^{(n)}, \boldsymbol{z}^{(n)}), \tag{67b}$$
$$\nabla\varphi_j(\boldsymbol{z}^{(n)}) = \nabla\tilde{\varphi}_j(\boldsymbol{z}^{(n)}, \boldsymbol{z}^{(n)}). \tag{67c}$$

Assume that Algorithm 1 converges to solution $\boldsymbol{z}^{(n)}$. Then, $\boldsymbol{z}^{(n)}$ is the optimal solution of problem (66), and hence, a Fritz John point that satisfies the following conditions [25], [37, Lemma 2.1].

$$\lambda_0\nabla\tilde{\varphi}_0(\boldsymbol{z}^{(n)}, \boldsymbol{z}^{(n)}) + \sum_{j=1}^{r}\nu_j\nabla\psi_j(\boldsymbol{z}^{(n)})$$

$$+\sum_{j=1}^{s}\lambda_j\nabla\tilde{\varphi}_j(\boldsymbol{z}^{(n)},\boldsymbol{z}^{(n)})=0, \tag{68a}$$

$$\nu_j\psi_j(\boldsymbol{z}^{(n)})=0,\forall j\in\{1,...,r\}, \tag{68b}$$

$$\lambda_j\tilde{\varphi}_j(\boldsymbol{z}^{(n)},\boldsymbol{z}^{(n)})=0,\forall j\in\{1,...,s\} \tag{68c}$$

where ν_j and λ_j is the dual variable associated with convex and nonconvex constraints j, respectively. Substituting (67b) and (67c) to (68), it holds that

$$\lambda_0\nabla\varphi_0(\boldsymbol{z}^{(n)})+\sum_{j=1}^{r}\nu_j\nabla\psi_j(\boldsymbol{z}^{(n)}) +\sum_{j=1}^{s}\lambda_j\nabla\varphi_j(\boldsymbol{z}^{(n)})=0, \tag{69a}$$

$$\nu_j\psi_j(\boldsymbol{z}^{(n)})=0,\forall j\in\{1,...,r\}, \tag{69b}$$

$$\lambda_j\varphi_j(\boldsymbol{z}^{(n)})=0,\forall j\in\{1,...,s\}, \tag{69c}$$

which means that $\boldsymbol{z}^{(n)}$ is a Fritz John solution of problem (65) (actually problem (36)). The proof is completed.

REFERENCES

[1] Ericsson, "Ericsson mobility report," https://www.ericsson.com/assets/local/mobility-report/documents/2017/ericsson-mobility-report-june-2017.pdf, 2017.

[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, 2016.

[3] Ericsson, "White paper: 5G energy performance," https://www.ericsson.com/assets/local/publications/white-papers/wp-5g-energy-performance.pdf, 2015.

[4] 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.

[5] 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.

[6] T. Quek, M. Peng, O. Simeone, and W. Yu, *Cloud Radio Access Networks: Principles, Technologies, and Applications*. Cambridge University Press, 2017.

[7] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K. K. Wong, R. Schober, and L. Hanzo, "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, 2016.

[8] S. H. Park, O. Simeone, and S. S. 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.

[9] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update 2016-2021," *White Paper*, 2017.

[10] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. M. Leung, "Cache in the air: exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, Feb. 2014.

[11] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug. 2014.

[12] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K. K. Wong, R. Schober, and L. Hanzo, "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, Second–quarter 2016.

[13] 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.

[14] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green Cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.

[15] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access*, vol. 2, pp. 1326–1339, 2014.

[16] M. Peng and K. Zhang, "Recent advances in fog radio access networks: Performance analysis and radio resource allocation," *IEEE Access*, vol. 4, pp. 5003–5009, 2016.

[17] A. Liu and V. K. N. Lau, "Mixed-timescale precoding and cache control in cached MIMO interference network," *IEEE Trans. Signal Process.*, vol. 61, no. 24, pp. 6320–6332, Dec. 2013.

[18] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, Dec. 2013.

[19] K. Poularakis, G. Iosifidis, V. Sourlas, and L. Tassiulas, "Exploiting caching and multicast for 5G wireless networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2995–3007, Apr. 2016.

[20] B. Azari, O. Simeone, U. Spagnolini, and A. M. Tulino, "Hypergraph-based analysis of clustered co-operative beamforming with application to edge caching," *IEEE Wireless Commun. Lett.*, vol. 5, no. 1, pp. 84–87, Feb. 2016.

[21] Y. Ugur, Z. H. Awan, and A. Sezgin, "Cloud radio access networks with coded caching," in *Proc. IEEE. Int. ITG Workshop Smart Antennas (WSA)*, Mar. 2016, pp. 1–5.

[22] V. Bioglio, F. Gabry, and I. Land, "Optimizing MDS codes for caching at the edge," in *Proc. IEEE Conf. Global Telecommun. (GLOBECOM))*, Dec. 2015, pp. 1–6.

[23] M. A. Maddah-Ali and U. Niesen, "Cache-aided interference channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 809–813.

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

[25] O. Mangasarian, *Nonlinear Programming*. Society for Industrial and Applied Mathematics, 1994.

[26] T. T. Vu, D. T. Ngo, L. Ong, S. Durrani, and R. H. Middleton, "Joint optimization of user association, data delivery rate and precoding for cache-enabled F-RANs," *in Proc. IEEE Conf. Global Telecommun. (GLOBECOM)*, pp. 1–6, Dec. 2017.

[27] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004.

[29] H. H. M. Tam, H. D. Tuan, D. T. Ngo, T. Q. Duong, and H. V. Poor, "Joint load balancing and interference management for small-cell heterogeneous networks with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 2, pp. 872–884, Feb. 2017.

[30] E. Che, H. D. Tuan, and H. H. Nguyen, "Joint optimization of cooperative beamforming and relay assignment in multi-user wireless relay

networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5481–5495, Oct. 2014.

[31] U. Rashid, H. D. Tuan, H. H. Kha, and H. H. Nguyen, "Joint optimization of source precoding and relay beamforming in wireless MIMO relay networks," *IEEE Trans. Commun.*, vol. 62, no. 2, pp. 488–499, Feb. 2014.

[32] 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.

[33] Y. Nesterov and A. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics, 1994.

[34] 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.

[35] 3GPP TS 36.814 V9.0.0, "3GPP technical specification group radio access network, evolved universal terrestrial radio access (E-UTRA): Further advancements for E-UTRA physical layer aspects (release 9)," 2010.

[36] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operations Research*, vol. 26, no. 4, pp. 681–683, 1978.

[37] 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.

Minh Ngoc Dao received the B.Sc. (Hons.) and M.Sc. degrees in mathematics from 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 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. (1st class honours) degree in Electrical Engineering from the University of Engineering & Technology, Lahore, Pakistan in 2000. He received the PhD degree in Electrical Engineering from the University of Queensland, Brisbane, Australia in Dec. 2004. He has been with the Australian National University, Canberra, Australia, since 2005, where he is currently Associate Professor in the Research School of Engineering, College of Engineering & Computer Science.

His research interests are in wireless communications and signal processing, including machine-to-machine and device-to-device communication, wireless energy harvesting systems, stochastic geometry modelling of finite area networks and synchronization in communication systems. He has co-authored more than 130 publications to date in refereed international journals and conferences. He was a recipient of the 2016 IEEE ComSoc Asia Pacific Outstanding Paper Award. He was the Chair of the ACT Chapter of the IEEE Signal Processing and Communications Societies from 2015 to 2016. He currently serves as an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS. He was awarded the 2018 ANU VC Award for Excellence in Supervision and the 2012 ANU VC Award for Excellence in Education. He is a Member of Engineers Australia, a Senior Fellow of IEEE, USA and a Senior Fellow of The Higher Education Academy, UK.

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.