

Weighted Max-min Fairness for C-RAN Multicasting under Limited Fronthaul Constraints

Quang-Doanh Vu, Kien-Giang Nguyen, *Student Member, IEEE*, and Markku Juntti, *Senior Member, IEEE*

Abstract—We consider downlink transmission in cloud radio access networks (C-RANs) with multiple cochannel multicasting groups served by a group of remote radio heads (RRHs), which receive information from a base band unit via finite-capacity fronthaul links. Our aim is to jointly design RRH selection and beamforming vectors such that the minimum weighted data rate among users are maximized under the constraints of maximum transmit power and fronthaul capacity at each specific RRH. The problem is intractable due to the numerical difficulties of combination and nonconvex functions. Based on a semidefinite relaxation technique, bisection search, and branch-and-bound method, we develop an upper bound which is also the optimal solution to the original problem if the relaxation is tight. More importantly, we propose a heuristic low-complexity iterative procedure for practical applications based on the state-of-the-art sequential convex approximation. Subsequently, we modify the proposed methods for the uncertain channel state information case. To be specific, the upper bound and its suboptimal solution are altered based on the S -lemma while the low-complexity algorithm is tailored by using two different approximations of intractable robust counterpart. The validity of the proposed methods in the region of limited fronthaul capacity is confirmed by numerical results.

Index Terms—Cloud radio access networks, multicasting, channel uncertainties, transmit beamforming, mixed integer semidefinite program, sequential convex approximation, robust optimization.

I. INTRODUCTION

The evolution of wireless communications and the development of user devices (e.g. smartphones) with high-quality media capabilities lead to the explosion in mobile data demand, in which videos contribute nearly 70 percent to the total mobile traffic [1]. On the other hand, since the number of connected devices is growing rapidly and is expected to exceed more than 28 billions by 2021 [1], more and more often multiple user devices located in a specific area access the same media content. In such a case, point-to-multipoint transmission or *multicasting* would be much more efficient than point-to-point *unicasting* [2]. The benefits of multicasting have been exploited in the Long-Term Evolution (LTE) networks via the mechanism called evolved multimedia broadcast and multicast service (eMBMS) [2]. The service is also likely to

This work was supported in part by the projects “Wireless Connectivity for Internet of Everything - Energy Efficient Transceiver and System Design (WiConIE)” funded by Academy of Finland under Grant 297803, and “Flexible Uplink-Downlink Resource Management for Energy and Spectral Efficiency Enhancing in Future Wireless Networks (FURMESFuN)” funded by Academy of Finland under Grant 31089.

The authors are with Centre for Wireless Communications, University of Oulu, FI-90014, Finland. Email: {doanh.vu, giang.nguyen, markku.juntti}@oulu.fi.

be embedded in the Fifth Generation (5G) mobile networks [3].

Coordinated multi-point (CoMP) transmission is one of the key technologies to improve the capacity of wireless networks, particularly for the cell edge users. There are two main strategies of CoMP namely coordinated beamforming (CB) and joint transmission (JT) [4]. To enable CoMP-CB, the channel state information (CSI) for receivers needs to be shared between transmitters. However, the data transmitted to a user is available at only one transmitter. The gain of CoMP-CB comes from the interference mitigation, i.e. CoMP-CB reduces the interference level experienced by a user (via designing the appropriate precoding). On the other hand, the same data content is simultaneously transmitted to a user from multiple transmitters in the CoMP-JT, i.e. transmitters collaborate in designing precoders operating as a large virtual multiple-input multiple-output (MIMO) system [5]. Therefore, CoMP-JT requires not only CSI but also the transmitted data to be available at multiple transmitters. The main practical challenge of CoMP implementation is that the strict network synchronization accuracy is required, especially for CoMP-JT [6]. Consequently, CoMP typically cannot achieve its full potential in the conventionally deployed mobile networks due to the large backhaul latency [7].

Cloud (or centralized) radio access network (C-RAN) is a novel network architecture which effectively supports the low-latency deployments [8]. The central idea is to divide the functionalities of the conventional base station (BS) into two parts called base band unit (BBU) and remote radio head (RRH); the BBU performs base-band processing while the RRH includes the radio frequency operations. Due to the separation, the BBU can be located at a central location called BBU pool while the RRH is placed close to the antennas, and they are connected with each other by, e.g., fiber link called fronthaul [9]. Such centralized base band processing allows the full coordination between RRHs, since the backhaul latency can be ignored, and thus CoMP-JT is effectively supported by the C-RAN. However, it should be noted that the capacities of the fronthaul links are limited regardless of the deployed physical medium (e.g. optical fiber). Consequently, the fronthaul links may become the bottlenecks, if their capacities are smaller than those of the wireless channels. An effective approach addressing the issue is to properly select the set of serving transmitters for each typical user [10]. This gives rise to the problems of joint beamforming design and RRH selection in the C-RAN.

A. Related Works

Multicasting has been intensively studied over the last decade. The problems with a single multicasting group were considered in [11], [12] while those with multiple multicasting groups were considered in [13]–[16]. There are two main objectives in designing multicasting transmission. The first one is to minimize the transmit power subject to the quality of service (QoS) of each specific user; the second one is to maximize the minimum weighted signal-to-interference-plus-noise ratio (SINR), i.e. weighted fairness, among users under the constraints on maximum transmit power. In general, a multicasting problem is NP-hard [13], and thus to find its optimal solution is a difficult task. Consequently, low-complexity efficient suboptimal solutions are of interest. There are two main approaches to design suboptimal linear beamforming for multicasting namely the semidefinite relaxation (SDR) [13] and the sequential convex approximation (SCA) [12]. The SDR is a sort of outer approximation technique which provides the exact optimal solution to the original problem—when relaxation is tight—or the performance bound based on that a suboptimal solution can be obtained using, e.g., randomization techniques. On the other hand, the SCA is an inner approximation method which achieves a solution to the problem via solving a series of convex subproblems. It is shown in [12] that the SCA achieves better performance with a lower complexity in large scale networks compared to the SDR. Multicasting transmission under imperfect CSI was studied in [14] where the error sets are modeled as ellipsoids.

There are many works investigating joint beamforming design and RRH selection for unicasting downlink CoMP-JT. The works in [17]–[19] focused on minimizing the total consumed power under the constraint of QoS for each specific user while those in [20], [21] aimed at maximizing the (weighted) sum rate. For multicasting, [22] proposed methods minimizing a network cost, which is the combination of the total transmit power and the fronthaul cost, under the constraints of QoS of each specific user. [23] investigated multicasting transmission under imperfect CSI where the consumed power is minimized subject to the QoS for each user and maximum transmit power at each RRH. Generally, a problem of joint beamforming design and RRH selection is cast as a mixed integer nonconvex program which is highly intractable. Thus, the mentioned works mainly focus on low-complexity suboptimal schemes.

B. Contributions

In this paper, we investigate the problem of joint beamforming design and RRH selection for C-RAN multicasting. Different from [22], [23], our goal is to maximize the minimum weighted SINR among users subject to the limitation of fronthaul capacity and the maximum transmit power at each specific RRH. We formulate the problem as a mixed Boolean nonconvex program by applying the Big-M formulation. The distinguishing feature of the proposed formulation is that its structure is amenable to developing a tight upper bound and efficient suboptimal solutions. In particular, our contributions include the following:

- We derive a tight upper bound and achieve a near optimal solution to the considered problem. To do so, we first follow the principles of the SDR to overcome the quadratic nonconvex parts and arrive at a relaxation. We then globally solve the relaxation problem via a procedure built based on the bisection search and the branch-and-bound (BnB) technique. Then, feasible solutions to the original problem are yielded via the randomization/rescaling procedure (if necessary).
- For more practical implementation, we resort to the SCA framework to develop a low-complexity iterative algorithm. More specifically, we introduce a regularization formulation arrived upon the idea of the exact penalty in which all the elements are smooth [24]. Then, an iterative procedure is built where a second-order cone program (SOCP), which is an inner approximate subproblem of the regularization problem, is solved in each iteration.
- The proposed approaches are then leveraged to the scenario wherein the CSI is imperfect. In particular, we follow the regular worst case strategy to overcome the channel uncertainty. The intractable robust counterpart is tackled by applying the relaxation (via introducing SD matrices and applying the \mathcal{S} -lemma) and the safety approximations.

C. Organization and Notations

The rest of the paper is organized as follows. Section II describes the system model and the problem formulation of joint beamforming design and RRH selection. Section III presents the proposed approaches including the upper bound and the suboptimal method for the perfect CSI scenario. Section IV extends the methods for the case of imperfect CSI. Section V provides the computational complexity estimation of the proposed methods. Numerical results and discussions are provided in Section VI. Finally, Section VII concludes the work.

Notation: We follow the standard notations in this paper. Bold lower and upper case letters represent vectors and matrices, respectively; $\|\cdot\|_2$ represents the l_2 norm; $|\cdot|$ represents the absolute value; $\mathbb{C}^{a \times b}$ represents the space of complex matrices of dimensions given in superscript; $\mathcal{CN}(0, c)$ denotes a zero mean circularly symmetric complex Gaussian random variable with variance c ; $\Re(\cdot)$ represents real part of the argument; $\mathbb{E}\{\cdot\}$ denotes the expectation operator. \mathbf{A}^T and \mathbf{A}^H stand for the transpose and the Hermitian transpose of \mathbf{A} , respectively; $\text{Tr}(\mathbf{A})$ and $\text{rank}(\mathbf{A})$ are the trace and rank of \mathbf{A} , respectively. \mathbf{I}_M represents an $M \times M$ identity matrix. The notation $\mathbf{A} \succeq 0$ means that \mathbf{A} is positive semidefinite. Other notations are defined at their first appearance.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider downlink transmission in C-RAN where a set of RRHs denoted by $\mathcal{T} = \{1, 2, \dots, T\}$ jointly transmit data to a set of multicast groups denoted by $\mathcal{G} = \{1, 2, \dots, G\}$. Let $\mathcal{U}_g = \{1, 2, \dots, U_g\}$ be the set of users belonging to group g , $g \in \mathcal{G}$, and each user belongs to only one group, i.e. $\mathcal{U}_g \cap \mathcal{U}_k = \emptyset$ for all $k \neq g$ [13], [22], [23]. We denote by g_i , $i \in \mathcal{U}_g$, user

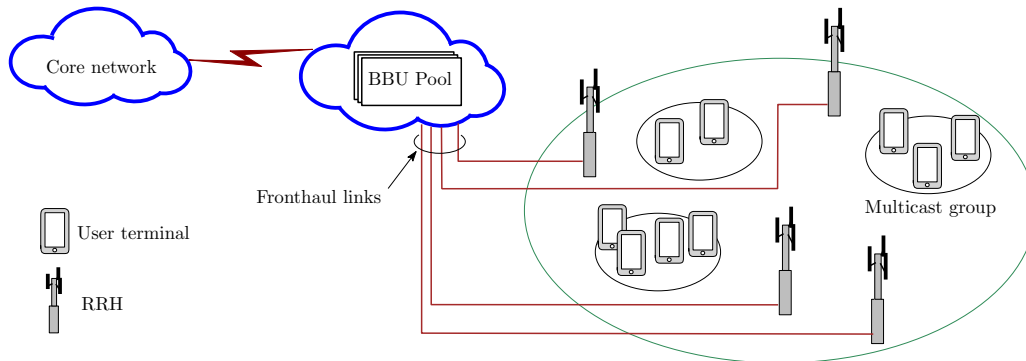


Figure 1. An example of collaborated transmission for multiple cochannel multicast groups in C-RAN where RRHs are connected to a BBU pool via fronthaul links.

i in group g . An example of the considered communication system is displayed in Fig. 1. We assume that each user is equipped with a single-antenna, and RRH t , $t \in \mathcal{T}$, is equipped with L_t transmit antennas. Let us denote by $\mathbf{h}_{g_i t} \in \mathbb{C}^{1 \times L_t}$ the channel (row) vector between user g_i and RRH t , and by $\mathbf{w}_{gt} \in \mathbb{C}^{L_t \times 1}$ the beamforming vector for group g at RRH t . We define $\mathcal{T}_g \subseteq \mathcal{T}$ to be the set of RRHs transmitting data to group g . With these introduced notations and under flat fading channels, the received signal at user g_i is

$$y_{g_i} = \left(\sum_{t \in \mathcal{T}_g} \mathbf{h}_{g_i t} \mathbf{w}_{gt} \right) x_g + \underbrace{\sum_{k \in \mathcal{G} \setminus g} \left(\sum_{m \in \mathcal{T}_k} \mathbf{h}_{g_i m} \mathbf{w}_{km} \right) x_k}_{\text{interference}} + z_{g_i} \quad (1)$$

where $z_{g_i} \sim \mathcal{CN}(0, \sigma_{g_i}^2)$ is the additive white Gaussian noise (AWGN) at user g_i with variance $\sigma_{g_i}^2$, and x_g is the normalized complex data symbol, i.e. $\mathbb{E}\{x_g x_g^*\} = 1$, intended for group g . We assume that the cochannel interference in (1) is treated as noise, then the signal-to-interference-plus-noise ratio (SINR) at user g_i can be expressed based on (1) as

$$\gamma_{g_i} = \frac{|\sum_{t \in \mathcal{T}_g} \mathbf{h}_{g_i t} \mathbf{w}_{gt}|^2}{\sum_{k \in \mathcal{G} \setminus g} |\sum_{m \in \mathcal{T}_k} \mathbf{h}_{g_i m} \mathbf{w}_{km}|^2 + \sigma_{g_i}^2}. \quad (2)$$

A. Fronthaul Capacity Constraint for Multicasting

We suppose that either perfect or imperfect CSI is available at the BBU pool for central network management [17]. After base-band processing is performed, data is delivered from the BBU pool to the RRHs via fronthaul links. As mentioned above, the fronthaul links are limited in capacity. Herein, multicast data is transferred to the RRHs after the BBU pool determined beamforming vectors $\{\mathbf{w}_{gt}\}$. Let R_g be the data rate corresponding to group g which is delivered over the fronthaul link to RRH t (for all $t \in \mathcal{T}_g$). We note that, in multicasting applications, the data rate transmitted to a specific user would be different from that to other users in the group depending on the corresponding grades of service [13]. Thus, R_g should be the highest data rate of the users in group g , i.e. $R_g \triangleq \max_{i \in \mathcal{U}_g} R_{g_i}$ where R_{g_i} is the data rate transmitted to user g_i . For viable transmission, the total data rate transmitted by a RRH (to users) should be smaller or equal to the capacity of the fronthaul link connecting to the RRH. In particular, let us

denote by C_t the capacity of the fronthaul link connecting to RRH t , and by $\mathcal{G}_t \subseteq \mathcal{G}$ the set of multicast groups receiving data from RRH t . Then, constraints on data rate transmitted over the fronthaul links can be written as [22]

$$\sum_{g \in \mathcal{G}_t} R_g \leq C_t, \forall t \in \mathcal{T}. \quad (3)$$

It can be observed from (3) that, given C_t , R_g would be improved if \mathcal{G}_t is grouped suitably. In other words, properly choosing the sets of serving RRHs for the multicast groups would improve the common information rate.

B. Problem Statement

We are interested in the problem of joint beamforming design and RRH selection for maximizing weighted max-min fairness transmission. In order to formulate the problem, let us denote by Γ_{g_i} the grade of service with respect to user g_i where larger Γ_{g_i} corresponds to higher grade [13]. Let us also denote by γ_0 the worst case scaled SINR, i.e. $\frac{\gamma_{g_i}}{\Gamma_{g_i}} \geq \gamma_0$ for all g_i [13], [14]. Based on the discussion in the previous subsection, in order to achieve some γ_0 , the data rate with respect to group g transmitted from the BBU pool to the serving RRHs should be larger than or equal to $R_g = \log(1 + \gamma_0 \bar{\Gamma}_g)$ where $\bar{\Gamma}_g \triangleq \max_{i \in \mathcal{U}_g} \Gamma_{g_i}$. With this observation, the problem of interest can be formulated as

$$\text{maximize}_{\{\mathbf{w}_{gt}\}, \{\mathcal{G}_t\}, \gamma_0} \gamma_0 \quad (4a)$$

$$\text{subject to } \frac{\gamma_{g_i}}{\Gamma_{g_i}} \geq \gamma_0, \forall g \in \mathcal{G}, i \in \mathcal{U}_g \quad (4b)$$

$$\sum_{g \in \mathcal{G}_t} \log(1 + \gamma_0 \bar{\Gamma}_g) \leq C_t, \forall t \in \mathcal{T} \quad (4c)$$

$$\sum_{g \in \mathcal{G}_t} \|\mathbf{w}_{gt}\|_2^2 \leq \bar{P}_t, \forall t \in \mathcal{T}. \quad (4d)$$

Constraints in (4c) are arrived following (3), and that in (4d) represents transmit power restriction where \bar{P}_t is the maximum transmit power at RRH t .

To solve (4) globally is a difficult task, since (4) combines the numerical challenges of handling discrete variables and nonconvex functions. In particular, (4) is a combinatorial optimization problem for which an optimal solution may

require an exhaustive search to be found. Moreover, even if $\{\mathcal{G}_t\}$ are fixed, the resulting problem is still NP-hard due to the nonconvex constraints in (4c) and the intractable form of $\{\gamma_{g_i}\}$ [13]. These observations promote finding low-complexity efficient suboptimal solutions to (4).

C. Alternative Formulation with Boolean Variables

In order to develop solutions to (4), we equivalently reformulate it into a more tractable form where RRH selection is controlled by Boolean variables. In fact, this approach has been considered in many works related to BS selection, e.g. [18], [21], [25], [26]. For notational convenience, let us define vectors $\mathbf{w}_g \triangleq [\mathbf{w}_{g1}^T, \mathbf{w}_{g2}^T, \dots, \mathbf{w}_{gT}^T]^T \in \mathbb{C}^{(L \times 1)}$ and $\mathbf{h}_{g_i} \triangleq [\mathbf{h}_{g_i1}, \mathbf{h}_{g_i2}, \dots, \mathbf{h}_{g_iT}] \in \mathbb{C}^{(1 \times L)}$ where $L = (\sum_{t=1}^T L_t)$ is the total number of transmit antennas at all RRHs. Then the SINR at user g_i in (2) can be rewritten as

$$\gamma_{g_i} = \frac{\mathbf{w}_g^H \mathbf{H}_{g_i} \mathbf{w}_g}{\sum_{k=1, k \neq g}^G \mathbf{w}_k^H \mathbf{H}_{g_i} \mathbf{w}_k + \sigma_{g_i}^2} \quad (5)$$

where $\mathbf{H}_{g_i} \triangleq \mathbf{h}_{g_i}^H \mathbf{h}_{g_i}$. To this end, let us introduce the Boolean variables $\{d_{gt}\}_{g \in \mathcal{G}, t \in \mathcal{T}}$ such that $d_{gt} = 1$ indicates that RRH t serves group g , and $d_{gt} = 0$ otherwise. Then, problem (4) can be equivalently reformulated as

$$\begin{aligned} & \text{maximize} && \gamma_0 && (6a) \\ & \text{subject to} && \{\mathbf{w}_g\}, \{d_{gt}\}, \gamma_0 \end{aligned}$$

$$\frac{\gamma_{g_i}}{\Gamma_{g_i}} \geq \gamma_0, \forall g \in \mathcal{G}, i \in \mathcal{U}_g \quad (6b)$$

$$\sum_{g=1}^G d_{gt} \log(1 + \gamma_0 \bar{\Gamma}_g) \leq C_t, \forall t \in \mathcal{T} \quad (6c)$$

$$\|\mathbf{B}_t \mathbf{w}_g\|_2^2 \leq d_{gt} \bar{P}_t, \forall t \in \mathcal{T}, g \in \mathcal{G} \quad (6d)$$

$$\sum_{g=1}^G \|\mathbf{B}_t \mathbf{w}_g\|_2^2 \leq \bar{P}_t, \forall t \in \mathcal{T} \quad (6e)$$

$$d_{gt} \in \{0, 1\}, \forall t \in \mathcal{T}, g \in \mathcal{G} \quad (6f)$$

where $\mathbf{B}_t \triangleq [\mathbf{0}_{L_t \times \sum_{b=1}^{t-1} L_b}, \mathbf{I}_{L_t}, \mathbf{0}_{L_t \times \sum_{b=t+1}^T L_b}]$. Additional constraints in (6d) are arrived based on the Big-M formulation [27], which indicate that RRH t , $t \in \mathcal{T}$, does not allocate power for group g , $g \in \mathcal{G}$, if the RRH does not serve the group, i.e. $\mathbf{w}_{gt} = \mathbf{0}$ when $d_{gt} = 0$.

It should be mentioned that a natural idea dealing with a mixed integer problem is to solve the problem where the integer variables are continuously relaxed, then round the value of the relaxed variables to the nearest integers, and subsequently solve the problem with the value of the integer variables fixed for satisfying the constraints and refinement. However, for (6), to relax or fix the Boolean variables does not lead to a more tractable formulation, i.e. the problems at the first and the third steps in the mentioned three-step procedure are still nonconvex. More importantly, it is not guaranteed that the value of $\{d_{gt}\}$ are close to 0 or 1 after solving the continuous relaxation problem. And thus, the final obtained objective value might be unacceptably deteriorated. This motivates us to develop the methods presented in the following sections.

III. PROPOSED JOINT BEAMFORMING DESIGN AND RRH SELECTION

A. Upper Bound and Suboptimal Design via SDR

A common γ_0 approach tackling the intractable format of γ_{g_i} is to lift the problem to the semidefinite domain. Let us introduce semidefinite matrices $\mathbf{W}_g \triangleq \mathbf{w}_g \mathbf{w}_g^H, \forall g$. Utilizing the principle of the SDR [28], we arrive at a relaxation of (6) which is written as

$$\begin{aligned} & \text{maximize} && \gamma_0 && (7a) \\ & \text{subject to} && \{\mathbf{W}_g \succeq 0\}, \{d_{gt}\}, \gamma_0 \end{aligned}$$

$$\frac{\text{Tr}(\mathbf{H}_{g_i} \mathbf{W}_g)}{\sum_{k=1, k \neq g}^G \text{Tr}(\mathbf{H}_{g_i} \mathbf{W}_k) + \sigma_{g_i}^2} \geq \Gamma_{g_i} \gamma_0, \quad \forall g \in \mathcal{G}, i \in \mathcal{U}_g \quad (7b)$$

$$\sum_{g=1}^G d_{gt} \log(1 + \gamma_0 \bar{\Gamma}_g) \leq C_t, \forall t \in \mathcal{T} \quad (7c)$$

$$\text{Tr}(\bar{\mathbf{B}}_t \mathbf{W}_g) \leq d_{gt} \bar{P}_t, \forall t \in \mathcal{T}, g \in \mathcal{G} \quad (7d)$$

$$\sum_{g=1}^G \text{Tr}(\bar{\mathbf{B}}_t \mathbf{W}_g) \leq \bar{P}_t, \forall t \in \mathcal{T} \quad (7e)$$

$$d_{gt} \in \{0, 1\}, \forall t \in \mathcal{T}, g \in \mathcal{G} \quad (7f)$$

where $\bar{\mathbf{B}}_t \triangleq \mathbf{B}_t^T \mathbf{B}_t$. Solving (7) globally is of particular interest. Let $(\{\mathbf{W}_g^*\}, \{d_{gt}^*\}, \gamma_0^*)$ be the optimal solution of (7). Then γ_0^* is the upper bound of the objective in (6). Moreover, if $\text{rank}(\mathbf{W}_g^*) = 1$ for all g , $(\{\mathbf{w}_g^*\}, \{d_{gt}^*\}, \gamma_0^*)$ is an optimal solution of (6) where \mathbf{w}_g^* is extracted from \mathbf{W}_g^* using the eigenvalue decomposition [29]. We note that (7) is still a mixed Boolean nonconvex program. However, its optimal solution can be obtained by the procedure presented below.

1) *Optimal Solution to (7)*: We make some useful observations regarding to (7). First, when γ_0 is fixed, problem (7) reduces to a mixed Boolean SDP (MI-SDP), i.e. the problem becomes SDP as the Boolean variables are fixed or relaxed, which can be solved globally by, e.g., BnB method [30]. Second, let $\mathcal{Q}(\gamma_0)$ be the feasible set of (7) corresponding to some fixed γ_0 , i.e.

$$\mathcal{Q}(\gamma_0) \triangleq \{ \{\mathbf{W}_g\}, \{d_{gt}\} | (7b) - (7f) \}. \quad (8)$$

Given γ_0 and γ_0' such that $\gamma_0 < \gamma_0'$, then we have $\mathcal{Q}(\gamma_0) \neq \emptyset$ if $\mathcal{Q}(\gamma_0') \neq \emptyset$; on the other hand, if $\mathcal{Q}(\gamma_0) = \emptyset$, $\mathcal{Q}(\gamma_0') = \emptyset$. These are due to the fact that $\mathcal{Q}(\gamma_0') \subseteq \mathcal{Q}(\gamma_0)$. Consequently, we can perform a bisection search over γ_0 . We develop a method based on these two remarks which provides the optimal solution to (7).

The main steps of the proposed method are outlined in Algorithm 1. At the initial stage, the starting value of the upper bound, γ_0^U , and lower bound, γ_0^L , of the objective γ_0 are required. For the lower bound, we simply set $\gamma_0^L = 0$ due to $\gamma_0 > 0$. The transmit data rate depends not only on the wireless channels but also on the fronthaul capacities. Thus the upper bound can be simply determined as $\gamma_0^U = \min \left(\left\{ P_{\text{tol}} \frac{\|\mathbf{h}_{g_i}\|_2^2}{\Gamma_{g_i} \sigma_{g_i}^2}; \frac{\exp(\bar{C}) - 1}{\Gamma_{g_i}} \right\}_{g_i} \right)$ where $P_{\text{tol}} = \sum_{t=1}^T \bar{P}_t$ and $\bar{C} = \max_t C_t$. The first term in the braces is arrived by applying the Cauchy-Schwartz inequality on the constraints in (7b)

which represents the upper bound dominated by the capacity of the wireless channels [13]. The second term represents the bound dominated by the capacity of the fronthaul links (i.e. (7c)) which corresponds to the case that the RRH associated to the largest fronthaul capacity serves only one group. Parameter ϵ_{bs} is the error tolerance, i.e. the possible maximum distance between the obtained objective value and the optimal one. The bisection search with respect to γ_0 is performed in lines 2–5. In line 4, a BnB algorithm is carried out to check the feasibility and find $(\{\mathbf{W}_g^*\}, \{d_{gt}^*\})$.

Algorithm 1 The proposed method for solving (7) optimally.

```

1: Initialization:  $\gamma_0^L, \gamma_0^U, \epsilon_{\text{bs}} > 0$ 
2: while  $\gamma_0^U - \gamma_0^L \leq \epsilon_{\text{bs}}$  do
3:   set  $\gamma_0 := \frac{\gamma_0^U + \gamma_0^L}{2}$ 
4:   if  $\mathcal{Q}(\gamma_0) \neq \emptyset$  then  $\gamma_0^L := \gamma_0$  and update
      $(\{\mathbf{W}_g^*\}, \{d_{gt}^*\}) \in \mathcal{Q}(\gamma_0)$  else  $\gamma_0^U := \gamma_0$  end if.
5: end while
6: Output:  $(\{\mathbf{W}_g^*\}, \{d_{gt}^*\}, \gamma_0^L)$ 

```

2) *Suboptimal solution to (6):* When \mathbf{W}_g^* is not rank-1 for all g , a feasible point of (6) can be derived from the output $(\{\mathbf{W}_g^*\}, \{d_{gt}^*\}, \gamma_0^L)$ by using randomization/scaling procedures [13], [28]. We note that the optimal value of (6) is smaller than or equal to γ_0^L . Thus the fronthaul capacity constraints in (6c) can be ignored in the scaling step. Specifically, let $\{\mathbf{w}_g^{\text{sdp}}\}_g$ be a set of beamforming vectors obtained from the Gaussian randomization procedure. Then the scaling problem reads [14]

$$\underset{\{p_g\}_g, \alpha}{\text{maximize}} \alpha \quad (9a)$$

$$\text{subject to } \frac{p_g a_{g_i}}{\sum_{k=1, k \neq g}^G p_k b_{g_i, k} + \sigma_{g_i}^2} \geq \Gamma_{g_i} \alpha, \forall g_i \quad (9b)$$

$$\sum_{g=1}^G p_g c_{gt} \leq \bar{P}_t, \forall t \quad (9c)$$

where $a_{g_i} \triangleq (\mathbf{w}_g^{\text{sdp}})^H \mathbf{H}_{g_i} \mathbf{w}_g^{\text{sdp}}$, $b_{g_i, k} \triangleq (\mathbf{w}_k^{\text{sdp}})^H \mathbf{H}_{g_i} \mathbf{w}_k^{\text{sdp}}$, and $c_{gt} \triangleq \|\mathbf{B}_t \mathbf{w}_g^{\text{sdp}}\|_2^2$. Problem (9) is solved by applying bisection search on α over $[0, \gamma_0^L]$ where each supproblem is a linear program.

Remark 1. Although γ_0^L can be achieved by any $(\{\mathbf{W}_g^*\}, \{d_{gt}^*\}) \in \mathcal{Q}(\gamma_0^L)$, choosing an arbitrary feasible point in $\mathcal{Q}(\gamma_0^L)$ might result in inefficient power performance. This is because $\mathcal{Q}(\gamma_0^L)$ might contain multiple points and the chosen one consumes more power than another. Moreover, we numerically observe that the scaling procedure, i.e. (9), is efficient with power-efficient points of $(\{\mathbf{W}_g^*\}, \{d_{gt}^*\})$. Therefore we propose the strategy of choosing $(\{\mathbf{W}_g^*\}, \{d_{gt}^*\})$ for the considered problem such that the total transmit power is minimized, i.e.

$$(\{\mathbf{W}_g^*\}, \{d_{gt}^*\}) \triangleq \underset{(\{\mathbf{W}_g\}, \{d_{gt}\}) \in \mathcal{Q}(\gamma_0^L)}{\text{arg min}} \sum_{g=1}^G \text{Tr}(\mathbf{W}_g) \quad (10)$$

which is also a MI-SDP. The advantages of the strategy are numerically justified in Fig. 3 in Section VI.

B. Low-complexity Solution via Sequential Convex Approximation

The method presented in the previous subsection can provide a tight upper bound and an efficient suboptimal solution to (6). However, this approach is not suitable for large-scale networks due to the exponential increase in the complexity of the BnB search and the high sensitivity of SDP with respect to the problem dimensions [31]. We present herein a heuristic low-complexity algorithm for practical large-scale networks which obtains suboptimal solutions to (6) by solving a series of SOCPs. The idea is to overcome the nonconvex problem by resorting to the SCA framework [32], [33], which has been widely used in wireless communications design, e.g. [12], [34].

1) *Equivalent Transformation:* Problem (6) is currently not in an amenable format for using SCA due to the intractability of nonconvex constraints in (6b) and (6c) as well as the discrete parts in (6f). Thus, as the first step, we transform the problem into an equivalent formulation in which its convexity is more exposed. Let us introduce new optimization variables $\{s_g\}_g$, and equivalently rewrite (6) as

$$\underset{\{\mathbf{w}_g\}, \{d_{gt}\}, \gamma_0, \{s_g\}}{\text{maximize}} \gamma_0 \quad (11a)$$

$$\text{subject to } \frac{\mathbf{w}_g^H \mathbf{H}_{g_i} \mathbf{w}_g}{\Gamma_{g_i} \gamma_0} \geq \sum_{k=1, k \neq g}^G \mathbf{w}_k^H \mathbf{H}_{g_i} \mathbf{w}_k + \sigma_{g_i}^2, \quad \forall g \in \mathcal{G}, i \in \mathcal{U}_g \quad (11b)$$

$$\log(1 + \gamma_0 \bar{\Gamma}_g) \leq s_g, \forall g \in \mathcal{G} \quad (11c)$$

$$\sum_{g=1}^G d_{gt} s_g \leq C_t, \forall t \in \mathcal{T} \quad (11d)$$

$$\sum_{t=1}^T \sum_{g=1}^G (d_{gt}^2 - d_{gt}) \geq 0 \quad (11e)$$

$$0 \leq d_{gt} \leq 1, \forall g \in \mathcal{G}, t \in \mathcal{T} \quad (11f)$$

$$(6d), (6e). \quad (11g)$$

The equivalence between (6) and (11) can be easily justified as follows. The constraints in (6c) are satisfied if and only if there exists a point $(\{\mathbf{w}_g\}, \{d_{gt}\}, \gamma_0, \{s_g\})$ which satisfies the constraints in (11c) and (11d). Similarly, the constraints in (11e) and (11f) are satisfied simultaneously if and only if d_{gt} is 1 or 0 for all g, t [35, Chap. 4].

2) *Regularization Formulation:* We note that the constraints in (11d) can be rewritten as $\sum_{g=1}^G (d_{gt} + s_g)^2 - (d_{gt} - s_g)^2 \leq 4C_t, \forall t \in \mathcal{T}$. Hence, all the nonconvex constraints in (11) have a difference-of-convex (DC) structure which can be convexly approximated using, e.g., the first order Taylor series. However, doing this arrives at the approximated problem which is infeasible due to (11e). Specifically, a convex approximation of (11e) at some $\{d_{gt}^{(n)}\}$ is given as $F(\{d_{gt}\}; \{d_{gt}^{(n)}\}) \triangleq \sum_{t=1}^T \sum_{g=1}^G \left(2d_{gt}^{(n)} d_{gt} - \left(d_{gt}^{(n)} \right)^2 - d_{gt} \right) \geq 0$. It is straightforward to see that the set $\{\{d_{gt}\} | F(\{d_{gt}\}; \{d_{gt}^{(n)}\}) \geq 0, (11f)\}$ is empty for any $\{d_{gt}^{(n)}\} \in (0, 1)^{GT}$. Inspired by the recent results in [36], [37], we tackle this shortcoming

by introducing a nonnegative slack variable ϕ and arriving at a regularization problem of (11) written as

$$\underset{\mathbf{v}}{\text{maximize}} \quad \gamma_0 - k\phi \quad (12a)$$

$$\text{subject to} \quad (6d), (6e), (11b) - (11d), (11f), \phi \geq 0 \quad (12b)$$

$$\sum_{t=1}^T \sum_{g=1}^G (d_{gt}^2 - d_{gt}) + \phi \geq 0 \quad (12c)$$

where $\mathbf{v} = \{\{\mathbf{w}_g\}, \{d_{gt}\}, \gamma_0, \{s_g\}, \phi\}$, and $k > 0$ is the penalty parameter. Clearly, problem (12) reduces to (11) when $\phi = 0$, and thus ϕ represents constraint residual temperature. From now on, we present an iterative procedure based on the SCA technique to solve (12) which reaches a favorable search region (for the solution) during some first iterations and minimizes the constraint residual at convergence.

3) *Proposed Iterative Procedure:* Let $\mathbf{v}^{(n)} = (\{\mathbf{w}_g^{(n)}\}, \{d_{gt}^{(n)}\}, \gamma_0^{(n)}, \{s_g^{(n)}\}, \phi^{(n)})$ be some feasible point of (12). Then a convex approximation of (12) at this point is given as

$$\underset{\mathbf{v}}{\text{maximize}} \quad \gamma_0 - \tau \|\mathbf{v} - \mathbf{v}^{(n)}\|_2^2 - k\phi \quad (13a)$$

$$\text{subject to} \quad \frac{2\Re\{(\mathbf{w}_g^{(n)})^H \mathbf{H}_{g_i} \mathbf{w}_g\}}{\Gamma_{g_i} \gamma_0^{(n)}} - \frac{(\mathbf{w}_g^{(n)})^H \mathbf{H}_{g_i} (\mathbf{w}_g^{(n)})}{\Gamma_{g_i} (\gamma_0^{(n)})^2} \gamma_0$$

$$\geq \sum_{k=1, k \neq g}^G \mathbf{w}_k^H \mathbf{H}_{g_i} \mathbf{w}_k + \sigma_{g_i}^2, \forall g, i \quad (13b)$$

$$\log(1 + \gamma_0^{(n)} \bar{\Gamma}_g) + \frac{\bar{\Gamma}_g (\gamma_0 - \gamma_0^{(n)})}{1 + \gamma_0^{(n)} \bar{\Gamma}_g} \leq s_g, \forall g \quad (13c)$$

$$\sum_{g=1}^G \left((d_{gt} + s_g)^2 + (d_{gt}^{(n)} - s_g^{(n)})^2 - 2(d_{gt}^{(n)} - s_g^{(n)})(d_{gt} - s_g) \right) \leq 4C_t, \forall t \quad (13d)$$

$$\sum_{g,t} \left(2d_{gt}^{(n)} d_{gt} - (d_{gt}^{(n)})^2 - d_{gt} \right) + \phi \geq 0 \quad (13e)$$

$$(6d), (6e), (11f), \phi \geq 0 \quad (13f)$$

where $\tau > 0$. The *proximal term* $-\tau \|\mathbf{v} - \mathbf{v}^{(n)}\|_2^2$ is added to make the objective function strongly concave with respect to \mathbf{v} . It guarantees that the iterates $\{\mathbf{v}^{(n)}\}$ obtained by the proposed iterative algorithm converge to a limit points (cf. [36], [38] for more details). The approximation functions in (13b), (13c), (13d), and (13e) are arrived following the first order Taylor series approach which satisfy the three conditions mentioned in [33].

Our proposed method is to successively solve (13) with feasible point $\mathbf{v}^{(n)}$ and parameter k are updated after each iteration. The main steps of the proposed iterative procedure are outlined in Algorithm 2. At the initial stage, a feasible point $\mathbf{v}^{(0)}$ of (12) is generated and the initial value of k is set. Also, the upper bound and the update step of k denoted by k_{\max} and λ , respectively, are given. In order to provide the relaxation for choosing RRHs, the initial value of the penalty parameter k and the update step λ should be small. Parameter

Algorithm 2 The proposed iterative procedure to solve (12).

- 1: **Initialization:** Given $\lambda > 1, k_{\max}$. Set $n := 0$, generate an initial feasible point $\mathbf{v}^{(0)}$ (of (12)), and set initial value of k .
 - 2: **repeat**
 - 3: Solve (13) to obtain optimal values \mathbf{v}^* .
 - 4: Set $n := n+1, \mathbf{v}^{(n)} := \mathbf{v}^*$. Update $k := \min(\lambda k, k_{\max})$.
 - 5: **until** Convergence
 - 6: **Output:** $(\mathbf{w}_g^{(n)}, \{d_{gt}^{(n)}\}, \gamma_0^{(n)})$
-

k is increased after each iteration until the upper bound k_{\max} which should be large. The purpose is to force ϕ as small as possible when the algorithm converges. The convergence properties of Algorithm 2 are stated in the following claim.

Claim 1. The iterate $\{\mathbf{v}^{(n)}\}_{n=1}^{\infty}$ is guaranteed to converge, i.e. $\|\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}\|_2$ tends to zero when n tends to infinity, and $\mathbf{v}^{(\infty)}$ is a Karush–Kuhn–Tucker (KKT) point of (12).

The proof for the claim follows the same arguments as those in [32, Section 3]. It is not guaranteed that the output of Algorithm 2 is a feasible point of (6), i.e. $\{d_{gt}^{(n)}\}_{g,t}$ might be only nearly binary. If this is the case, we round $\{d_{gt}^{(n)}\}_{g,t}$ to binary and run the algorithm with respect to the continuous variables only until convergence for refinement. It is worth mentioning that in the numerical section with the considered simulation systems and parameters, we obtained $\phi \approx 0$ when convergence. The examples are described in Section VI and shown in Fig. 2.

4) *Discussion on Practical Implementation:* As mentioned above, the initial value of k and the update step λ are set to be small. However, this might lead to the issue that the required number of iterations for convergence is large. A possible approach overcoming this shortcoming is to use two different update steps λ_1 and λ_2 where $\lambda_1 < \lambda_2$; λ_2 is applied after some predefined number of iterations.

At some first iterations when k is small, the constraint in (12c) has an insignificant impact on the solutions. Therefore, to arrive at a formulation which is tight as the Boolean variables are relaxed might improve the desired performance. A common approach is to apply perspective formulation [25], [39], [40]. In particular, let us introduce nonnegative variables $\{\rho_{gt}\}_{g,t}$. Then the constraints in (6d) and (6e) are replaced by

$$\|\mathbf{B}_t \mathbf{w}_g\|_2^2 \leq d_{gt} \rho_{gt}, \forall t \in \mathcal{T}, g \in \mathcal{G} \quad (14)$$

$$\sum_{g=1}^G \rho_{gt} \leq \bar{P}_t, \forall t \in \mathcal{T} \quad (15)$$

respectively. We note that the constraints in (14) can be written in SOC as $\|[\sqrt{2} \mathbf{B}_t \mathbf{w}_g; d_{gt}; \rho_{gt}]\|_2 \leq d_{gt} + \rho_{gt}$. A tighter formulation using the quadratic form of the Boolean variables can also be applied [39], in which (14) is substituted by $\|\mathbf{B}_t \mathbf{w}_g\|_2^2 \leq d_{gt}^2 \rho_{gt}$. This constraint is nonconvex which can be also convexly approximated by applying the first order Taylor series similar to those in (13).

The objective (13a) is strongly convex for all $\tau > 0$. However, τ should be small, otherwise the algorithm converges slowly. An alternative approach for practical implementation

is that the proximal term $\tau \|\mathbf{v} - \mathbf{v}^{(n)}\|_2^2$ only joins when there is no variation in γ_0 (cf. [38]).

IV. JOINT ROBUST BEAMFORMING DESIGNS AND RHH SELECTION UNDER CHANNEL UNCERTAINTIES

In this section, we extend the proposed solutions to systems with imperfect channel state. This is significant because to ensure the error-free channel is still a challenge in practice, and the performance would be reduced due to the sensitivity of the beamforming to the errors [41].

There exist two approaches broadly used for modeling the channel uncertainty [42]. The first approach assumes that the error resides in prespecified sets. Then, robust optimization is applied with the aim at guaranteeing the performance level. The second one is suitable for errors which are unbounded, e.g. estimation errors. This approach assumes that the error parts are random variables whose statistics are known. In this case, stochastic optimization is applied exploiting the statistical information for guaranteeing probability for some performance target (i.e. outage performance). However, the reformulated problems include chance constraints which are usually highly intractable [43]. In this paper, we consider the bounded uncertainty, and optimize the worst case performance.

A. Imperfect CSI Model and Robust Counterpart Formulation

We consider the common additive error model of uncertain channels given as

$$\mathbf{h}_{g_i} = \tilde{\mathbf{h}}_{g_i} + \boldsymbol{\theta}_{g_i}, \forall g \in \mathcal{G}, i \in \mathcal{U}_g \quad (16)$$

where $\tilde{\mathbf{h}}_{g_i} \in \mathbb{C}^{(1 \times L)}$ is the known value vector, and $\boldsymbol{\theta}_{g_i} \in \mathbb{C}^{(1 \times L)}$ is the vector of random errors [14], [23], [44]. We assume that $\{\boldsymbol{\theta}_{g_i}\}$ lie in bounded sets. In particular, we concentrate on the ellipsoidal uncertainty sets given as

$$\mathcal{E}_{g_i} = \{\boldsymbol{\theta}_{g_i} | \boldsymbol{\theta}_{g_i} \mathbf{E}_{g_i} \boldsymbol{\theta}_{g_i}^H \leq \delta_{g_i}\}, \forall g \in \mathcal{G}, i \in \mathcal{U}_g \quad (17)$$

where \mathbf{E}_{g_i} is a nonsingular positive definite matrix. The uncertainty model (17) means that the true channel \mathbf{h}_{g_i} randomly lies in the ellipsoid parameterized by \mathbf{E}_{g_i} and δ_{g_i} and centered at $\tilde{\mathbf{h}}_{g_i}$. The motivation of considering the ellipsoidal uncertainty sets was discussed in [41], [44], [45].

Following the worst case robust optimization strategy, which has been broadly used in the existing literature [23], [41], [46], the robust counterpart of (6) takes the form

$$\begin{aligned} & \text{maximize } \gamma_0 \\ & \{\mathbf{w}_g\}, \{d_{gt}\}, \gamma_0 \end{aligned} \quad (18a)$$

subject to

$$\begin{aligned} & \min_{\boldsymbol{\theta}_{g_i} \in \mathcal{E}_{g_i}} \frac{|(\tilde{\mathbf{h}}_{g_i} + \boldsymbol{\theta}_{g_i}) \mathbf{w}_g|^2}{\sum_{k=1, k \neq g}^G |(\tilde{\mathbf{h}}_{g_i} + \boldsymbol{\theta}_{g_i}) \mathbf{w}_k|^2 + \sigma_{g_i}^2} \geq \Gamma_{g_i} \gamma_0, \\ & \forall g \in \mathcal{G}, i \in \mathcal{U}_g \end{aligned} \quad (18b)$$

$$(6c) - (6f) \quad (18c)$$

where (18b) follows from (16). To solve (18) optimally is a major challenge because (18) inherits the numerical difficulties of (6) in addition to the semi-infinite number of constraints in (18b).

B. Upper bound and suboptimal solution to (18) via SDR

A common approach overcoming the semi-infinite number of intractable constraints in (18b) is to lift the problem to higher dimensions with SD variables [23]. Specifically, let us define $\bar{\mathbf{W}}_{g_i} = \mathbf{W}_g - \Gamma_{g_i} \gamma_0 \sum_{k=1, k \neq g}^G \mathbf{W}_k$ for all g_i , where the positive semidefinite matrices $\{\mathbf{W}_g\}$ have been defined in Section III-A. We equivalently rewrite (18b) as

$$\begin{aligned} \boldsymbol{\theta}_{g_i} \mathbf{E}_{g_i} \boldsymbol{\theta}_{g_i}^H \leq \delta_{g_i} & \Rightarrow \boldsymbol{\theta}_{g_i} \bar{\mathbf{W}}_{g_i} \boldsymbol{\theta}_{g_i}^H + 2\Re(\tilde{\mathbf{h}}_{g_i} \bar{\mathbf{W}}_{g_i} \boldsymbol{\theta}_{g_i}^H) \\ & + \tilde{\mathbf{h}}_{g_i} \bar{\mathbf{W}}_{g_i} \tilde{\mathbf{h}}_{g_i}^H - \Gamma_{g_i} \gamma_0 \sigma_{g_i}^2 \geq 0, \forall g, i \end{aligned} \quad (19)$$

By straightforwardly applying the well-known technique called \mathcal{S} -Lemma [46] on (19) and ignoring the nonconvex rank-1 constraints on $\{\mathbf{W}_g\}$, we arrive at a relaxation of (18) given as

$$\begin{aligned} & \text{maximize } \gamma_0 \\ & \{\mathbf{W}_g \succeq 0\}, \{d_{gt}\}, \\ & \gamma_0, \{\mu_{g_i} \geq 0\} \end{aligned} \quad (20a)$$

subject to

$$\begin{pmatrix} \bar{\mathbf{W}}_{g_i} + \mu_{g_i} \mathbf{E}_{g_i} & & & \\ & \tilde{\mathbf{h}}_{g_i} \bar{\mathbf{W}}_{g_i} \tilde{\mathbf{h}}_{g_i}^H & & \\ & & \bar{\mathbf{W}}_{g_i} \tilde{\mathbf{h}}_{g_i}^H & \\ & & & \bar{\mathbf{W}}_{g_i} \tilde{\mathbf{h}}_{g_i}^H - \sigma_{g_i}^2 \Gamma_{g_i} \gamma_0 - \mu_{g_i} \delta_{g_i} \end{pmatrix} \succeq 0, \quad \forall g \in \mathcal{G}, i \in \mathcal{U}_g \quad (20b)$$

$$(7c) - (7f) \quad (20c)$$

where $\{\mu_{g_i}\}_{g_i}$ are additional slack variables. Since (20b) is the exact representation of (18b), the relaxation (20) is tight if $\text{rank}(\mathbf{W}_g) = 1$ for all g . As can be observed, nonconvex constraints in (20b) reduce to tractable linear matrix inequalities (LMIs) when γ_0 is fixed. Consequently, the procedure similar to that in Algorithm 1 can achieve the optimal solution to (20), i.e. replacing $\mathcal{Q}(\gamma_0)$ in step 4 by $\bar{\mathcal{Q}}(\gamma_0)$ where $\bar{\mathcal{Q}}(\gamma_0) \triangleq \{\{\mathbf{W}_g\}, \{d_{gt}\}, \{\mu_{g_i}\} | (7c) - (7f), (20b)\}$. Again, when the rank-1 constraints are not satisfied for all obtained $\{\mathbf{W}_g\}$, a randomization/scaling procedure is run to yield suboptimal solutions to (18).

C. Low-complexity Robust Designs via SCA

We now apply the SCA framework to design low-complexity robust solution to (18). To be consistent with the worst case strategy, we consider a safety approximation of (18b) given as

$$\begin{aligned} & \min_{\boldsymbol{\theta}_{g_i} \in \mathcal{E}_{g_i}} \frac{|(\tilde{\mathbf{h}}_{g_i} + \boldsymbol{\theta}_{g_i}) \mathbf{w}_g|^2}{\sum_{k=1, k \neq g}^G |(\tilde{\mathbf{h}}_{g_i} + \boldsymbol{\theta}_{g_i}) \mathbf{w}_k|^2 + \sigma_{g_i}^2} \geq \Gamma_{g_i} \gamma_0, \\ & \max_{\boldsymbol{\theta}_{g_i} \in \mathcal{E}_{g_i}} \end{aligned} \quad \forall g \in \mathcal{G}, i \in \mathcal{U}_g. \quad (21)$$

By using the triangle and the Cauchy-Schwarz inequalities (c.f. [47] for details) for the numerator in the left side and introducing nonnegative slack variables $\{\nu_{g_i}\}_{g_i}$, we can equivalently rewrite (21) for each g_i as follows

$$\max_{\boldsymbol{\theta}_{g_i} \in \mathcal{E}_{g_i}} \sum_{k=1, k \neq g}^G |(\tilde{\mathbf{h}}_{g_i} + \boldsymbol{\theta}_{g_i}) \mathbf{w}_k|^2 \leq \nu_{g_i}^2 \quad (22)$$

$$\frac{\left(\left[|\tilde{\mathbf{h}}_{g_i} \mathbf{w}_g| - \sqrt{\delta_{g_i}} \|\mathbf{E}_{g_i}^{-1/2} \mathbf{w}_g\| \right]^+ \right)^2}{\Gamma_{g_i} \gamma_0} \geq \nu_{g_i}^2 + \sigma_{g_i}^2 \quad (23)$$

where $[x]^+$ denotes $\max(x, 0)$. To deal with the semi-infinite set of constraints in (22), we first rewrite it as

$$\boldsymbol{\theta}_{g_i} \mathbf{E}_{g_i} \boldsymbol{\theta}_{g_i}^H \leq \delta_{g_i} \Rightarrow \|(\tilde{\mathbf{h}}_{g_i} + \boldsymbol{\theta}_{g_i}) \hat{\mathbf{W}}_g\|^2 \leq \nu_{g_i}^2 \quad (24)$$

where $\hat{\mathbf{W}}_g = [\mathbf{w}_1, \dots, \mathbf{w}_{g-1}, \mathbf{w}_{g+1}, \dots, \mathbf{w}_G]$. Then, by applying the Schur's complement lemma and the \mathcal{S} -lemma on (24) (similar to the arguments in [46, Chap. 6.3]), this constraint is equivalent to

$$\begin{aligned} \exists \kappa_{g_i} \geq 0 : \\ \begin{pmatrix} \nu_{g_i} \mathbf{I} & \sqrt{\delta_{g_i}} (\mathbf{E}_{g_i}^{-1/2} \hat{\mathbf{W}}_g)^H & \hat{\mathbf{W}}_g^H \tilde{\mathbf{h}}_{g_i} \\ \sqrt{\delta_{g_i}} \mathbf{E}_{g_i}^{-1/2} \hat{\mathbf{W}}_g & \kappa_{g_i} \mathbf{I} & \mathbf{0} \\ \tilde{\mathbf{h}}_{g_i} \hat{\mathbf{W}}_g & \mathbf{0} & \nu_{g_i} - \kappa_{g_i} \end{pmatrix} \succeq 0 \end{aligned} \quad (25)$$

which is a LMI with respect to $(\nu_{g_i}, \kappa_{g_i}, \{\mathbf{w}_k\}_{k \neq g})$. We now focus on the nonconvex constraints in (23) which are equivalent to the following set of constraints

$$\frac{\alpha_{g_i}^2}{\Gamma_{g_i} \gamma_0} \geq \nu_{g_i}^2 + \sigma_{g_i}^2, \forall g, i \quad (26)$$

$$|\tilde{\mathbf{h}}_{g_i} \mathbf{w}_g| \geq \tilde{\alpha}_{g_i}, \forall g, i \quad (27)$$

$$\tilde{\alpha}_{g_i} - \alpha_{g_i} \geq \sqrt{\delta_{g_i}} \|\mathbf{E}_{g_i}^{-1/2} \mathbf{w}_g\|, \forall g, i \quad (28)$$

where $\{\alpha_{g_i} \geq 0\}_{g_i}$, and $\{\tilde{\alpha}_{g_i} \geq 0\}_{g_i}$ are newly introduced slack variables. We can see that the nonconvex parts lie in (26) and (27) which can be convexly approximated as

$$2 \frac{\alpha_{g_i}^{(n)}}{\Gamma_{g_i} \gamma_0^{(n)}} \alpha_{g_i} - \frac{(\alpha_{g_i}^{(n)})^2}{\Gamma_{g_i} (\gamma_0^{(n)})^2} \gamma_0 \geq \nu_{g_i}^2 + \sigma_{g_i}^2, \forall g_i \quad (29)$$

$$\begin{aligned} (\mathbf{w}_g^{(n)})^H \tilde{\mathbf{H}}_{g_i} \mathbf{w}_g^{(n)} \\ + 2\Re\{(\mathbf{w}_g^{(n)})^H \tilde{\mathbf{H}}_{g_i} (\mathbf{w}_g - \mathbf{w}_g^{(n)})\} \geq \tilde{\alpha}_{g_i}^2, \forall g, i, \end{aligned} \quad (30)$$

respectively, where $\tilde{\mathbf{H}}_{g_i} \triangleq \tilde{\mathbf{h}}_{g_i}^H \tilde{\mathbf{h}}_{g_i}$. At this point, we are ready to arrive at the problem solved in each iteration of the SCA procedure which is written as

$$\underset{\hat{\mathbf{v}}}{\text{maximize}} \quad \gamma_0 - \tau \|\hat{\mathbf{v}} - \hat{\mathbf{v}}^{(n)}\|^2 - k\phi \quad (31a)$$

$$\text{subject to} \quad (13c) - (13f), (25), (29), (30), (28) \quad (31b)$$

where $\hat{\mathbf{v}} \triangleq \{\mathbf{v}, \{\alpha_{g_i}\}, \{\tilde{\alpha}_{g_i}\}, \{\nu_{g_i}\}, \{\kappa_{g_i}\}\}$. We remark that to find initial points for starting the iterative procedure is not straightforward, since it requires that $|\tilde{\mathbf{h}}_{g_i} \mathbf{w}_g^{(0)}| \geq \sqrt{\delta_{g_i}} \|\mathbf{E}_{g_i}^{-1/2} \mathbf{w}_g^{(0)}\|$ for all g_i . However, we can overcome this issue by using variable ϕ , i.e. (28) is replaced by $\tilde{\alpha}_{g_i} - \alpha_{g_i} + \phi \geq \sqrt{\delta_{g_i}} \|\mathbf{E}_{g_i}^{-1/2} \mathbf{w}_g\|$. In fact, this technique of finding feasible point was used in [48].

Problem (31) is a SDP, and thus the computational complexity might become high when the problems are large-scale. In order to reduce computational effort, let us consider another approximation of (18b) given as

$$\frac{\min_{\boldsymbol{\theta}_{g_i} \in \mathcal{E}_{g_i}} |(\tilde{\mathbf{h}}_{g_i} + \boldsymbol{\theta}_{g_i}) \mathbf{w}_g|^2}{\sum_{k=1, k \neq g}^G \max_{\boldsymbol{\theta}_{g_i} \in \mathcal{E}_{g_i}} |(\tilde{\mathbf{h}}_{g_i} + \boldsymbol{\theta}_{g_i}) \mathbf{w}_k|^2 + \sigma_{g_i}^2} \geq \Gamma_{g_i} \gamma_0,$$

$$\forall g \in \mathcal{G}, i \in \mathcal{U}_g \quad (32)$$

We remark that (21) is tighter than (32). However, (32) leads to the algorithm where a SOCP is solved in each iteration. In particular, by again using the triangle and the Cauchy-Schwarz inequalities, (32) is equivalent to

$$\frac{\left(|[\tilde{\mathbf{h}}_{g_i} \mathbf{w}_g| - \sqrt{\delta_{g_i}} \|\mathbf{E}_{g_i}^{-1/2} \mathbf{w}_g\|]^+ \right)^2}{\sum_{k=1, k \neq g}^G (|\tilde{\mathbf{h}}_{g_i} \mathbf{w}_k| + \sqrt{\delta_{g_i}} \|\mathbf{E}_{g_i}^{-1/2} \mathbf{w}_k\|)^2 + \sigma_{g_i}^2} \geq \Gamma_{g_i} \gamma_0, \quad \forall g_i. \quad (33)$$

Let us introduce some slack variables $\{\alpha_{g_i} \geq 0\}_{g_i}$, $\{\tilde{\alpha}_{g_i}\}_{g_i}$, $\{\nu_{g_i k}\}_{g_i, k \neq g}$, and $\{\tilde{\nu}_{g_i k}\}_{g_i, k \neq g}$. Then (33) is equivalent to the following sets of constraints

$$\frac{\alpha_{g_i}^2}{\Gamma_{g_i} \gamma_0} \geq \sum_{k=1, k \neq g}^G \nu_{g_i k}^2 + \sigma_{g_i}^2, \forall g, i \quad (34)$$

$$|\tilde{\mathbf{h}}_{g_i} \mathbf{w}_g| \geq \tilde{\alpha}_{g_i}, \forall g, i \quad (35)$$

$$\tilde{\alpha}_{g_i} - \alpha_{g_i} \geq \sqrt{\delta_{g_i}} \|\mathbf{E}_{g_i}^{-1/2} \mathbf{w}_g\|, \forall g, i \quad (36)$$

$$|\tilde{\mathbf{h}}_{g_i} \mathbf{w}_k| \leq \tilde{\nu}_{g_i k}, \sqrt{\delta_{g_i}} \|\mathbf{E}_{g_i}^{-1/2} \mathbf{w}_k\| \leq \nu_{g_i k} - \tilde{\nu}_{g_i k}, \forall g, i, k \neq g \quad (37)$$

where the nonconvex parts in (34) and (35) can be convexly approximated as those in (29) and (30), respectively. To summarize, the SOCP solved in iteration $n+1$ of the SCA procedure is given as

$$\underset{\tilde{\mathbf{v}}}{\text{maximize}} \quad \gamma_0 - \tau \|\tilde{\mathbf{v}} - \tilde{\mathbf{v}}^{(n)}\|^2 - k\phi \quad (38a)$$

subject to

$$2 \frac{\alpha_{g_i}^{(n)}}{\Gamma_{g_i} \gamma_0^{(n)}} \alpha_{g_i} - \frac{(\alpha_{g_i}^{(n)})^2}{\Gamma_{g_i} (\gamma_0^{(n)})^2} \gamma_0 \geq \sum_{k=1, k \neq g}^G \nu_{g_i k}^2 + \sigma_{g_i}^2, \forall g, i \quad (38b)$$

$$(13c) - (13f), (30), (36), (37) \quad (38c)$$

where $\tilde{\mathbf{v}} \triangleq \{\mathbf{v}, \{\alpha_{g_i}\}, \{\tilde{\alpha}_{g_i}\}, \{\nu_{g_i k}\}, \{\tilde{\nu}_{g_i k}\}\}$. The theoretical complexity of the presented methods will be discussed in the following section.

V. COMPLEXITY ESTIMATE

A. Solutions for Perfect CSI

We first make some observations on the analytical complexity of the upper bound and the suboptimal solution developed via the SDR. The number of iterations related to the bisection search is $N_{\text{bi}} = \lceil \log_2(\gamma_0^U / \epsilon_{\text{bs}}) \rceil$. The complexity is mainly caused by solving MI-SDPs. Particularly, the number of the Boolean variables is GT leading to 2^{GT} combinations. The worst case complexity solving the problem with the Boolean variables are fixed using a general interior point solver is $\mathcal{O}(G^{3.5} L^{6.5} + UG^{1.5} L^{2.5} + G^{2.5} L^{2.5} T)$ [13]. This number is dominated by the total number of transmit antennas at RRHs and the number of multicast groups. When the randomization/scaling procedure is used to extract a suboptimal solution, the maximum number of solved linear programs (LPs) is $N_{\text{rand}} \lceil \log_2(\gamma_0^L / \epsilon_{\text{bs}}) \rceil$ where N_{rand} is the size of

the randomization. The complexity of solving each LP is $\mathcal{O}(G^2(U+T+G)^{1.5})$ [49], which does not depend on L .

We now consider the SCA design. Since there exists no result of the bound of the required number of iterations, we focus on the complexity of solving a SOCP in each iteration. Specifically, (13) contains $G(2L+T+1)+2$ real variables and $U+G(T+1)+2T$ SOC constraints of different dimensions. Thus, the upper bound complexity is $\mathcal{O}((U+GT+T)^{0.5}G^2(L+T)^2G(U+L))$. As the numbers of multicasting groups and total transmit antennas are large, the upper bound reduces to $\mathcal{O}(G^{3.5}L^3)$ showing that the SCA design is less sensitive to L compared to that of the solution derived via SDR.

B. Solutions for Imperfect CSI

In case of the upper bound and SDR solution, the complexity of solving the problem as the Boolean variables and variable γ_0 are fixed is $\mathcal{O}(L^{0.5}(G+U)^{0.5}(GL^2+U)^2(GL^2+UL^2+(G+1)T))$. The number has the same order of magnitude with respect to G and L compared to that for the perfect CSI. When randomization/scaling procedure is used, the complexity of solving a scaling problem is $\mathcal{O}(L^{0.5}U^{0.5}(G+U)^2(UL^2+T))$ which depends on L (different from that in case of perfect CSI).

For the SCA designs, the complexity of solving the problem (31) is $\mathcal{O}((GL+LU+UG)^{0.5}G^2L^2(TG^2L_t^2+UL^2+UGL))$. Focusing on L , the number reduces to $\mathcal{O}(L^{4.5})$ which is lower (in order of magnitude) compared to that of the design via the SDR. This is because there is no introduction of SD matrices $\{\mathbf{W}_g\}$ in (31). The upper bound complexity of solving (38) is $\mathcal{O}(G^{0.5}(U+T)^{0.5}G^2(L+T+U)^2(G(U+L)+UL))$.¹

VI. NUMERICAL RESULTS

In this section, we report simulation results to evaluate the performances of the proposed methods. We consider the simulation model based on those in [18], [22], [23]. In particular, for each multicast group, the positions of the users are uniformly and independently generated in a circle of radius 15 meters. Similarly, the positions of the RRHs and that of the centers of multicast groups are uniformly and independently generated in a circle of radius 300 meters. The numbers of the RRHs and multicast groups are changed in the experiments. The system parameters are set as follows. The system bandwidth is 10 MHz, and the noise power density is -174 dBm/Hz. The 3GPP path loss model in dB is $145.4 + 37.5 \log_{10}(l)$ where l is the distance in kilometers, and the log normal shadowing standard deviation is 4 [50]. The channel vector from RRH t and user g_i is generated as $\mathbf{h}_{g_i t} = \beta_{g_i t} \bar{\mathbf{h}}_{g_i t}$ where $\beta_{g_i t}$ represents path loss and shadowing, and $\bar{\mathbf{h}}_{g_i t}$ follows $\mathcal{CN}(0, \mathbf{I})$. The maximum transmit power and the number of antennas at the RRHs are set to $\bar{P}_t = 32$ dBm and $L_t = 2$, for all t . The number of users in a multicast group is $U_g = 4$ for all g . Without loss of generality, we set $\Gamma_{g_i} = 1$ for all

g_i . This network setting is considered unless otherwise stated. Other parameters will be specified in each experiment.

The initial points for starting the proposed iterative procedures are generated as follows. First, we set $d_{gt}^{(0)} = 0.5$ for all g, t . Then the beamforming vectors $\{\mathbf{w}_g^{(0)}\}_g$ are created randomly which are scaled (if necessary) such that the power constraints are satisfied. Based on those, $\gamma_0^{(0)}$, $\{s_g^{(0)}\}$, and $\phi^{(0)}$ are determined such that (11b), (11c), (11d), and (12c) are satisfied. For penalty parameter k , the initial value is set as 10^{-3} , and the maximum value k_{\max} is set as 10^4 ; the updated constant λ is set as 3. The parameter τ (of the proximal term) is set as 10^{-7} .

In this section, beside the derived upper bound serving as a benchmark, we also compare the schemes of joint RRH selection and beamforming design to the conventional scheme of joint cooperation, i.e. $d_{gt} = 1$ for all g, t . The beamforming vectors of the scheme are derived based on the SDR-based method presented in [14]. We refer this scheme as ‘FullCo’ and ‘FullCoRo’ in perfect and imperfect CSI systems, respectively. To this end, for convenience, we denote by ‘JBR-SDR’ and ‘JBR-SCA’ the proposed suboptimal solutions obtained from Algorithm 1 (with the randomization/scaling procedure is applied if necessary) and Algorithm 2, respectively. Similarly, in the case of the imperfect CSI, we denote by ‘JBRRo-SDR’ and ‘JBRRo-SCA’ the solutions derived based on SDR and SCA, respectively. To solve SOCPs and SDPs, we use the modeling package YALMIP [51] with the inner solver MOSEK [52]; to solve MI-SDPs, inner solver BNB (of YALMIP) is used.

A. Performances under Perfect Knowledge of Channel States

In the first experiment, we study the convergence behavior of Algorithm 2 (JBR-SCA) over two randomly generated channels. Specifically, Fig. 2(a) depicts the convergence with respect to the value of γ_0 , and Fig. 2(b) shows the value of ϕ over iterations. As can be seen, the value of γ_0 converges to the points close to the corresponding upper bounds. The value of ϕ converges to zero which implies that the obtained $\{d_{gt}\}_{gt}$ are binary, i.e. the outputs of Algorithm 2 are the feasible solutions to (6). We also observe that the value of γ_0 is larger than the derived upper bound during some iterations. This is due to the relaxation of the discrete constraints. Therefore, we can see in Fig. 2(b) that the value of ϕ (representing the constraint residual temperature) is also large during these iterations. Another observation is that the algorithm converges within 25 iterations with the both channels.

In Fig. 3 we investigate the average performances of the considered schemes as the functions of the fronthaul capacity with network configuration (G, T) set as $(3, 4)$ and $(4, 5)$. The error tolerance of the bisection search ϵ_{bs} is 10^{-2} . The size of the randomization procedure N_{rand} is 100. For JBR-SDR, we provide the performances of two different strategies of choosing feasible points in $\mathcal{Q}(\gamma)$ (i.e. step 4 of Algorithm 1); the first strategy follows (10); the second one is choosing random feasible points.² The iterative procedure of Algorithm 2 stops when the condition $|\varphi^{(n+1)} - \varphi^{(n)}| \leq 10^{-4}$, where

¹We provide the computational time of different schemes in Subsection VI-C, Table I, for numerically comparing the complexity among the schemes.

²The first feasible points obtained by the branch and bound procedure are chosen.

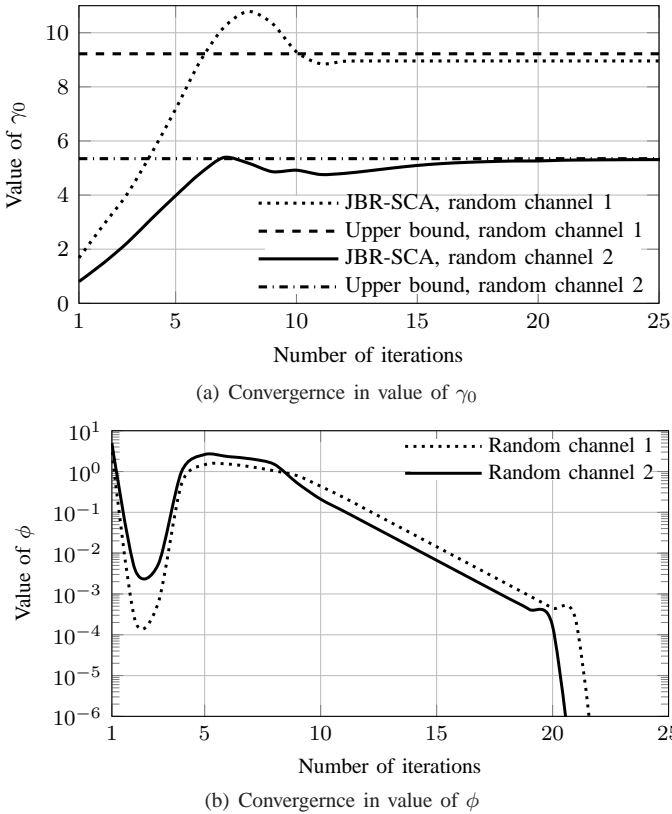


Figure 2. Convergence performances of Algorithm 2 (i.e. JBR-SCA) over two random channel realizations (with perfect CSI). The network configuration (G, T) of random channel 1 is (3, 4), and that of random channel 2 is (4, 5). The fronthaul capacity is set as $C_t = 80$ Mbits/s, $\forall t$.

$\varphi^{(n)} = \gamma_0^{(n)} - \tau \|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|^2 - k\phi^{(n)}$, is satisfied. Figures 3(a) and 3(c) show the performances in terms of average achieved minimum data rate. The first observation from these figures is that the achieved minimum data rate of all considered schemes increases when C_t increases indicating that the network performance does not only depend on the capacity of wireless channels but also on that of the fronthaul links. An expected result is that remarkable gains are achieved by jointly designing beamforming vectors and selecting RRH, especially when the fronthaul capacity is small. On the other hand, when C_t is large enough, the performances of the proposed schemes close to that of FullCo. This is because the network performance now is mainly dominated by the capacity of the wireless channels which is maximized when the RRHs fully cooperate. As such, further increasing C_t does not change the performances. We can also see that the performances of suboptimal schemes JBR-SDR and JBR-SCA are quite close to the upper bound. Another observation is that the performance of JBR-SDR with (10) almost agree with the upper bound when C_t is small and there is a small gap between them when C_t is large. The results can be explained as follows. When C_t is small, the consumed power corresponding to the obtained SD matrices $\{\mathbf{W}_g^*\}$ is small. As the beamforming vectors are extracted from $\{\mathbf{W}_g^*\}$, it is likely that there exist points in the feasible set of the scaling problem (9) which satisfy the obtained upper bound γ_0^L . On the other hand, when C_t is large,

the consumed power corresponding to $\{\mathbf{W}_g^*\}$ is large, and thus the probability of existing a feasible point of (9) satisfying γ_0^L is small. This also explains that JBR-SDR with (10) outperforms JBR-SDR with random chosen (feasible) points. Figures 3(b) and 3(d) depict the average consumed power of the considered schemes. We observe that if beamforming design ignores the limitation of fronthaul capacity, i.e. the scheme labeled ‘ [14] w/o fronthaul constraints’ in the figures, the consumed power is always large while the achievable desired performance is inferior. This result again confirms the necessity of taking into account the capacity of fronthaul links in design. The scheme FullCo consumes the smallest power compared to the others due to the full connection and the low achieved data rate. The figures clearly show the advantage of the strategy in (10) in terms of power efficiency.

In Fig. 4, we show the average achieved minimum data rate performance regarding to different number of RRHs. In particular, we consider a network scenario where RRHs and the centers of multicast groups are randomly placed inside a circle of radius 500 meters. The number of multicast groups is fixed at $G = 10$ (i.e. 40 users), and the number of RRHs is taken as $T = \{12, 16, 20\}$. For the such scenario, we focus on the low-complexity scheme JBR-SCA, and also provide the performance of FullCo for comparison. It is observed that, with large values of C_t , the two schemes provide better performance when more RRHs are added. This is due to the cooperation gain, i.e. the wireless channel capacity increases with respect to T . In addition, it is interesting to observe that the gap between the performance of JBR-SCA and FullCo is larger when T increases. On the other hand, when C_t is small, to add more RRHs does not improve the performance with scheme FullCo. Meanwhile, JBR-SCA can exploit the selection gain, which results in better performance when T increases.

B. Performances under Channel Uncertainties

We now investigate the performance of the robust designs under imperfect CSI conditions. For simplicity, we select the error sets as follows. We first define the uncertainty parameter $\delta' > 0$ and let $\alpha \triangleq \max_{\forall g_i, t} \beta_{g_i t}$; we then set $\delta_{g_i} = \delta'/\alpha$ and $\mathbf{E}_{g_i} = \mathbf{I}$ for all g_i . Other parameters are set as those in the experiments in the previous subsection.

Similar to the case of perfect CSI, we investigate the convergence behavior of the proposed iterative procedures over two random channel realizations which are shown in Fig. 5. We set the network configuration (G, T) as (3, 4) and the uncertainty parameter δ' as 0.3. Again we can observe that ϕ goes to zero in all cases. The procedures converge within 10 and 25 iterations with the random channel 1 and the random channel 2, respectively. We can also observe that the convergence points of γ_0 of scheme (31) are better than those of scheme (38). This is because the approximation in (21) is tighter than that in (32). The reason that γ_0 is larger than the upper bound during some iterations is the same as that for the results in Fig. 2.

Fig. 6 depicts the average performances in terms of the worst case minimum data rate of the considered robust

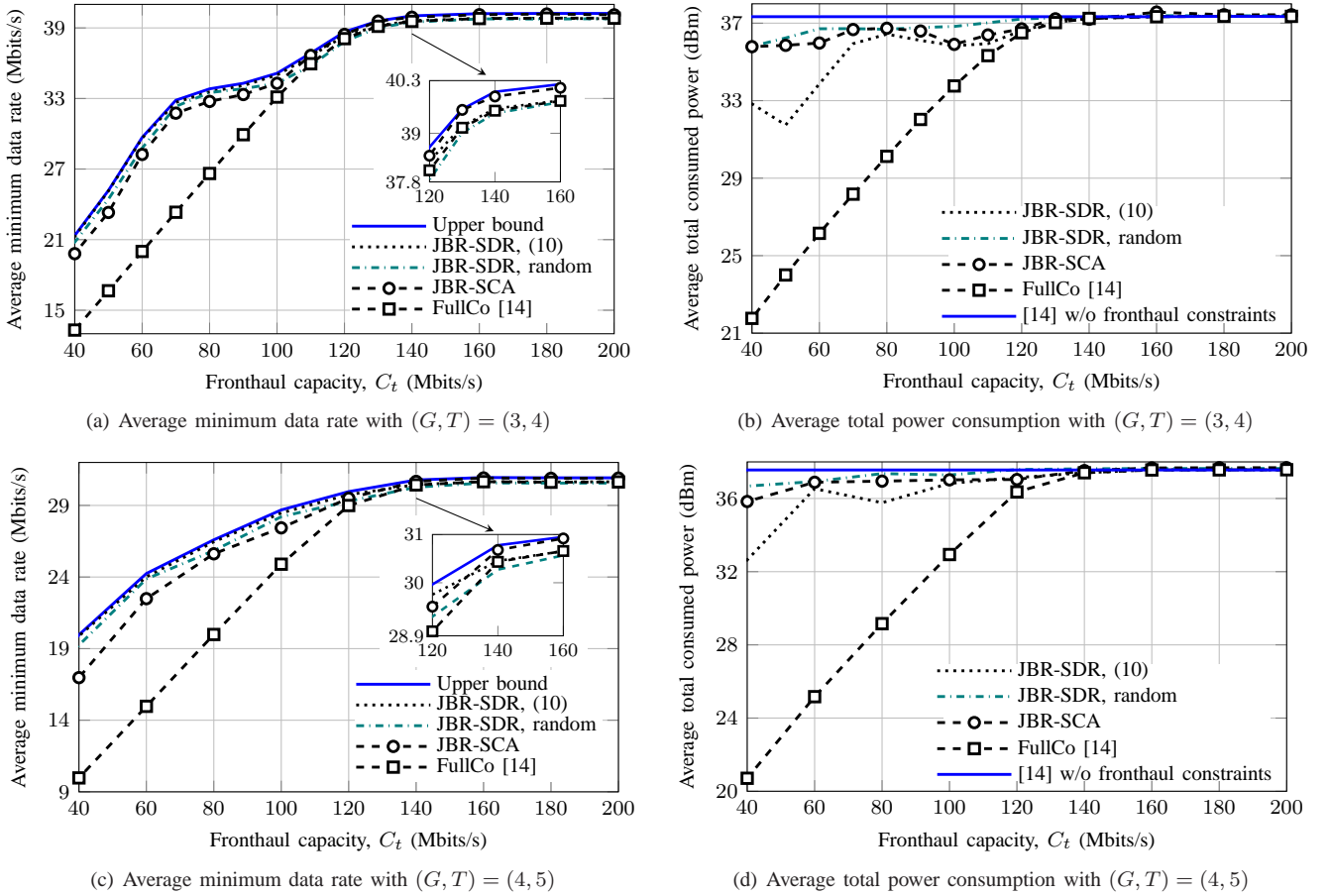


Figure 3. Average performances of the considered schemes versus different fronthaul capacities with two different network configurations of (G, T) those are $(3, 4)$ and $(4, 5)$.

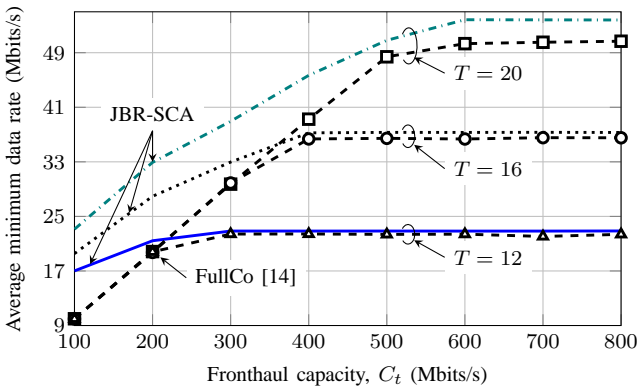


Figure 4. Average achieved minimum data rate of JBR-SCA and FullCo versus different fronthaul capacities for three number of RRHs: $T = 12$, $T = 16$, and $T = 20$. The number of multicast groups is $G = 10$.

schemes as functions of uncertainty parameter δ' . It is observed that the performances of all the proposed schemes reduce when δ' increases due to the effect of the channel uncertainty. Although the safety approximations are used in (21) and (32), JBRo-SCA schemes (with (31) and (38)) are able to achieve more than 90% of the upper bound. We also see that the performance of FullCoRo does not change over the considered range of δ' . This is because the performance

of this scheme is mainly dominated by the fronthaul capacity with respect to the simulation setup, i.e. although the wireless channel capacity reduces because of the uncertainty, it is still higher than the limitation caused by the fronthaul capacity. Clearly, there are remarkable gaps between FullCoRo and the proposed schemes.

Fig. 7 shows the average worst case minimum performance of the considered robust schemes against the fronthaul capacity C_t . We set the network configuration as $(G, T) = (3, 4)$ and the uncertainty parameter as $\delta' = 0.3$. We can see that the performance of all considered schemes increase with respect to C_t . This result is similar to that in the case of perfect CSI. As expected, the joint beamforming design and RRH selection schemes outperform FullCoRo when C_t small. When C_t is large, the performances of JBRo-SDR, JBRo-SCA with (31), and FullCoRo are close to each other and slightly higher than that of JBRo-SCA with (38). This is because the impact of the safety approximation used in (32). However, as will be seen in Subsection VI-C, the numerical effort for computing JBRo-SCA with (38) is remarkably lower compared to that for the others.

In Fig. 8 we study the behavior of the proposed schemes JBRo-SCA with (31) and (38) when the size of the uncertainty sets considered in the designs are smaller than that of the actual uncertainty sets. For this purpose, let us introduce the

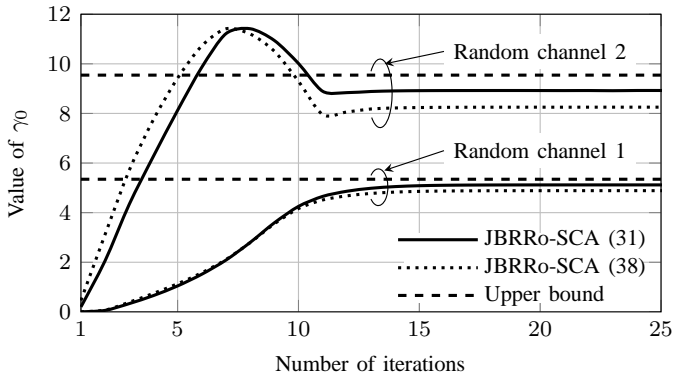
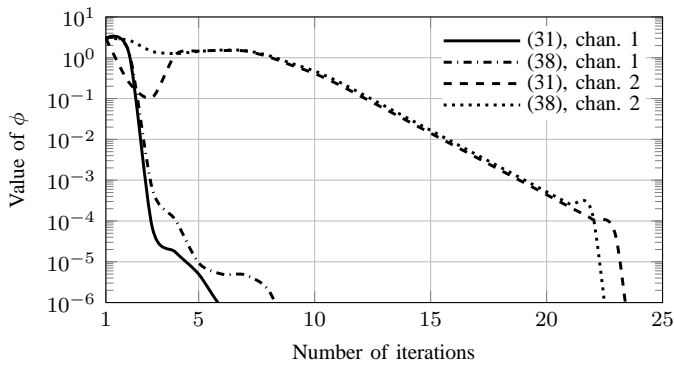
(a) Convergence in value of γ_0 (b) Convergence in value of ϕ

Figure 5. Convergence performances of the proposed iterative procedures, i.e. JBRRo-SCA with (31) and (38), over two different randomly generated channel realizations (under imperfect CSI). The network configuration is $(G, T) = (3, 4)$, and the uncertainty parameter δ' is 0.3. The capacity of the fronthaul links is $C_t = 80$ Mbits/s for all t .

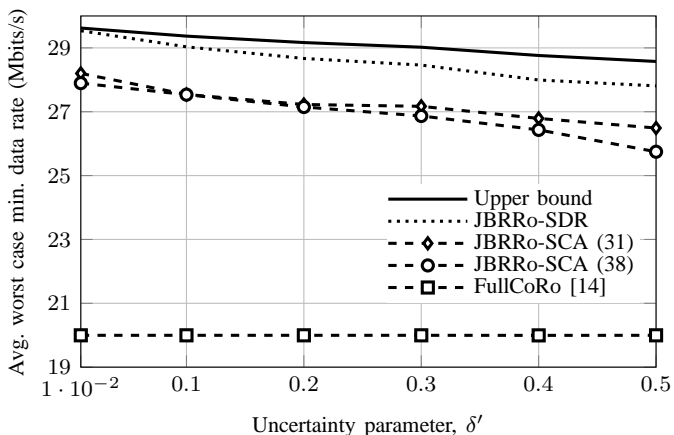


Figure 6. Average worst case minimum data rate of the considered robust schemes versus uncertainty parameter δ' . The network configuration (G, T) is $(3, 4)$ and the capacity of the fronthaul links is $C_t = 60$ Mbits/s for all t .

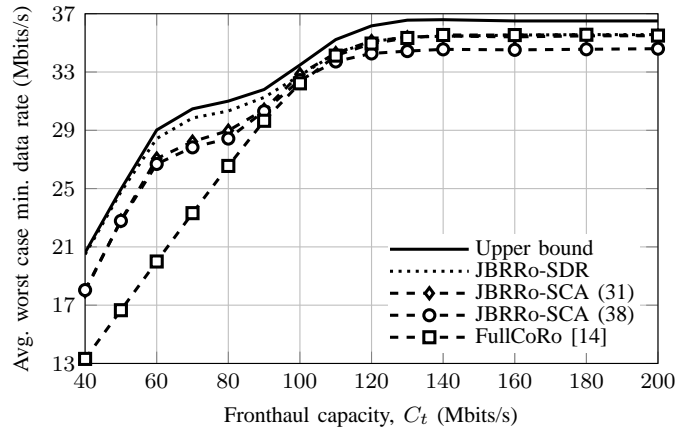
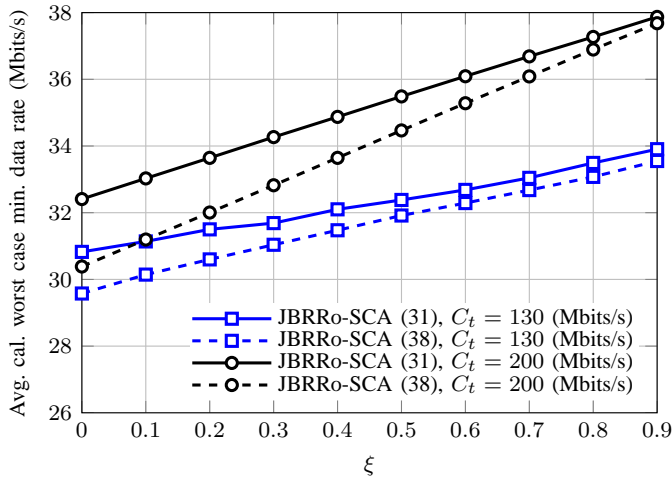


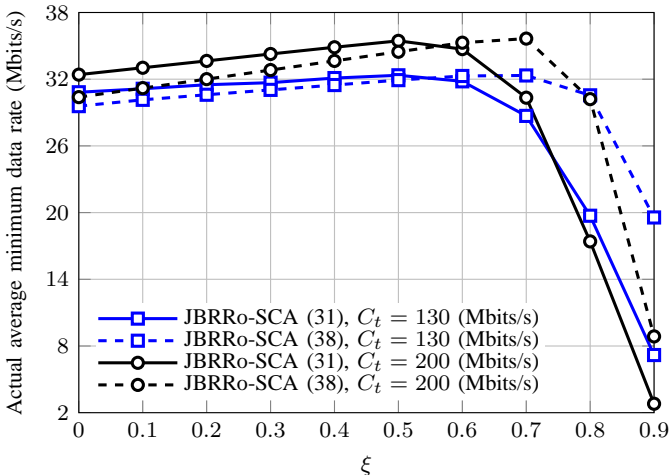
Figure 7. Average worst case minimum data rate of the considered robust schemes versus the fronthaul capacity C_t . The network configuration (G, T) is $(3, 4)$, and the uncertainty parameter δ' is 0.3.

parameter $\xi \in [0, 1)$ which represents the degree of reduction in size of the uncertainty sets in the designs (compared to that of the actual uncertainty sets), i.e. the error vectors in the designs are assumed to lie in hyperspheres of radius $(1 - \xi)\sqrt{\delta_{g_i}}$ for all g_i . We run the algorithms over 1000 channel realizations where the uncertainty parameter is taken as $\delta' = 0.5$. Fig. 8(a) shows the worst case minimum data rate obtained by the algorithms. Fig. 8(b) depicts the actual average minimum data rate. In this figure, if any constraint in (6b) is violated, there will be an outage and the (actual) minimum data rate is set as zero. The probability of outage is showed in Fig. 8(c). As expected, the worst case minimum data rate obtained by the algorithms increases when ξ increases due to the reduction in size of the uncertainty sets (considered in the designs). The important observation is that the actual average minimum data rate first increases then decreases with the increase of ξ . This is because, when ξ is small, the probability of outage is small. Thus, the average minimum data rate performance is improved due to the increase of the worst case minimum data rate. However, after a certain value of ξ , the probability of outage becomes larger leading to performance loss in terms of average minimum data rate. Clearly, the results indicate that there exists a tradeoff between the worst case minimum data rate and the average minimum data rate performances.

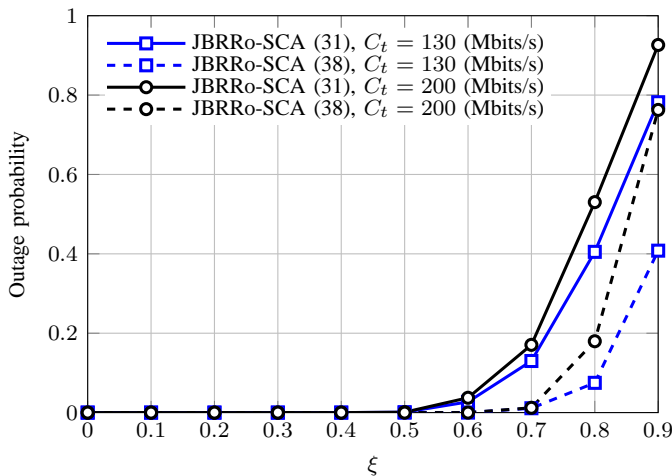
We use JBRRo-SCA with (38) to evaluate the average worst case minimum data rate performance in a large scale scenario. The results are illustrated in Fig. 9. In particular, we adopt the same network settings as those in Fig. 4. We take $G = 10$, $T = \{12, 14, 16\}$, and $\delta' = \{0.1, 0.5\}$. On one hand, we can observe similar results as in Figs. 6 and 7 that the performance improves regarding to the increase of C_t , and the larger δ' results in the lower performance in all cases of C_t . On the other hand, adding more RRHs provides improvement in performance. This is due to the cooperation gain and selection gain exploited by the scheme. We can also see that the gap between the curves of $\delta' = 0.1$ and $\delta' = 0.5$ in the regime of small C_t is smaller than that in the regime of large C_t . Again, this is because, in the regime of small C_t , a reduction



(a) Average calculated worst case minimum data rate



(b) Actual average achieved minimum data rate



(c) Outage probability

Figure 8. Performances of the proposed robust schemes, i.e. JBRRo-SCA with (31) and (38), as functions of ξ . The network configuration is $(G, T) = (4, 6)$. The uncertainty parameter is $\delta' = 0.5$. The fronthaul capacity is taken as $C_t = \{130, 200\}$ (Mbits/s).

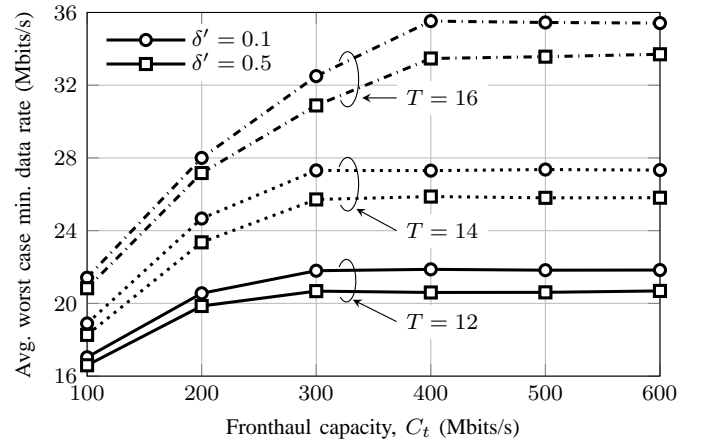


Figure 9. Average worst case minimum data rate of JBRRo-SCA with (38) versus the fronthaul capacity C_t in a large scale scenario where $G = 10$, and $T = 12, 14$, and 16 . We take $\delta' = 0.1$ and 0.5 .

in wireless channel capacity (due to the uncertainty) does not result in a performance loss if the wireless channel capacity is still higher than the limitation caused by the fronthaul capacity.

C. Numerical Efficiency

We now numerically investigate the computational complexity of the proposed solutions. Since the complexity of JBR-SDR and JBRRo-SDR are quite high due to the BnB methods, we mainly focus on the SCA-based designs and compare them to the existing solutions (i.e. FullCo and FullCoRo). Table I presents the average run time (in second) for solving the corresponding convex problems of the considered schemes with different network configurations. Notably, the shown numbers account the total run time of iterations for the algorithms to converge. In order to clarify the effectiveness of the proposed solutions, we also provide the desired performances, i.e. the average (worst case) minimum data rate. We can observe from the table, in case of the perfect CSI, that the average run time of JBR-SCA is smaller than 2 seconds, and that of FullCo is smaller than 3 seconds. Thus, by JBR-SCA, we can yield remarkable gain in terms of the desired performance without putting more computational effort (compared to the existing solution). In case of the imperfect CSI, due to arriving at SOCP in each iteration, the complexity of JBRRo-SCA with (38) is significantly lower compared to the others. On the other hand, the complexity of FullCoRo is much higher than that of FullCo. This is because of the additional U LMIs in (20b) and, more importantly, the randomization/scaling procedure, i.e. the scaling power problems corresponding to FullCoRo are SDPs. This is also the reason that JBRRo-SCA with (31) requires lower computational effort than FullCoRo. Another observation is that the average running time of JBRRo-SCA with (38) is larger than that of JBR-SCA. This is reasonable since the numbers of variables and constraints in (38) are larger than those in (13).

VII. CONCLUSION

We have investigated multicasting transmission in C-RAN where multiple RRHs controlled by a BBU collaborate in

Table I

AVERAGE RUN TIME (IN SECOND) AND AVERAGE (WORST CASE) MINIMUM DATA RATE (IN MBITS/S) OF THE PROPOSED LOW-COMPLEXITY SCHEMES COMPARED TO THE EXISTING SOLUTION (I.E. [14]) WITH DIFFERENT NETWORK CONFIGURATIONS OF (G, T) AND C_t . FOR THE PERFORMANCES UNDER IMPERFECT CSI, THE UNCERTAINTY PARAMETER δ' IS SET AS 0.3.

(G, T)		(3,5)			(4,4)			(4,6)		
C_t (Mbits/s)		60	80	160	60	80	160	60	80	160
JBR-SCA	avg. min. rate	28.4	36.8	45.9	20.6	22.4	26.4	25.5	27.6	37.4
	avg. run time	1.10	1.09	0.51	1.09	0.78	0.56	1.94	1.71	0.51
FullCo [14]	avg. min. rate	20.0	26.7	45.5	15.0	20.0	26.2	15.0	20.0	37.2
	avg. run time	0.65	1.22	2.99	1.15	1.41	1.92	1.18	1.01	2.28
JBRro-SCA (38)	avg. worst case min. rate	28.1	31.1	39.5	17.8	20.1	22.4	21.6	23.6	32.0
	avg. run time	2.04	1.80	1.23	1.83	1.17	1.65	4.08	2.40	1.85
JBRro-SCA (31)	avg. worst case min rate	28.4	31.4	40.6	18.1	20.2	23.4	21.8	24.6	33.5
	avg. run time	33.2	27.7	14.5	23.2	11.1	16.4	58.5	38.4	20.3
FullCoRo [14]	avg. worst case min rate	20.0	26.7	40.6	15.0	19.7	23.5	15.0	20.0	33.6
	avg. run time	65.7	109.2	230.7	83.2	115.7	131.3	100.2	117	226.3

transferring data to multiple cochannel multicast groups. Taking into account the limitation of the fronthaul capacity, we explored the joint RRH selection and beamforming vectors with the aim on weighted max-min fairness among users. In particular, we have formulated the problem as mixed Boolean nonconvex programs. A tight upper bound and an efficient solution have been derived based on the combination of the SDR, bisection search, and branch-and-bound techniques. In addition, we have also developed a low-complexity iterative procedure based on the SCA framework in which a SOCP is solved in each iteration. Afterwards, we have leveraged our proposed methods to the context where the CSI is uncertain. The issue of erroneous CSI has been tackled by the regular robust worst case strategies those are relaxation (via the \mathcal{S} -lemma) and safety approximations. The numerical results have clearly demonstrated that the proposed solutions outperform the existing ones in the regions of limited fronthaul capacity.

REFERENCES

- [1] Ericsson, "Ericsson mobility report," Nov. 2015.
- [2] D. Lecomte and F. Gabin, "Evolved multimedia broadcast/multicast service (eMBMS) in LTE-advanced: overview and rel-11 enhancements," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 68–74, Nov. 2012.
- [3] G. Araniti, M. Condoluci, P. Scopelliti, A. Molinaro, and A. Iera, "Multicasting over emerging 5G networks: Challenges and perspectives," *IEEE Network*, vol. PP, no. 99, pp. 2–11, Feb. 2017.
- [4] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzaresse, S. Nagata, and K. Sayana, "Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 148–155, Feb. 2012.
- [5] D. Gesbert, S. Hanly, H. Huang, S. Shamaï Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.
- [6] 3GPP, "Coordinated multi-point operation for LTE physical layer aspects (Release 11)," 3rd Generation Partnership Project, TR 36.819.
- [7] NGMN Alliance, "RAN evolution project: CoMP evaluation and enhancement," March 2015.
- [8] Ericsson, "Cloud RAN—the benefits of virtualization, centralization and coordination," Sep. 2015.
- [9] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for Mobile Networks—A Technology Overview," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 405–426, Firstquarter 2015.
- [10] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: insights and challenges," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 152–160, April 2015.
- [11] N. Sidiropoulos, T. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, June 2006.
- [12] L.-N. Tran, M. Hanif, and M. Juntti, "A conic quadratic programming approach to physical layer multicasting for large-scale antenna arrays," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 114–117, Jan. 2014.
- [13] E. Karipidis, N. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1268–1279, March 2008.
- [14] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Weighted fair multicast multigroup beamforming under per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5132–5142, Oct. 2014.
- [15] Z. Xiang, M. Tao, and X. Wang, "Coordinated multicast beamforming in multicell networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 12–21, January 2013.
- [16] G. Scutari, F. Facchinei, L. Lampariello, S. Sardellitti, and P. Song, "Parallel and distributed methods for constrained nonconvex optimization-part II: Applications in communications and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 1945–1960, April 2017.
- [17] 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.
- [18] J. Tang, W. P. Tay, and T. Q. S. Quek, "Cross-layer resource allocation with elastic service scaling in cloud radio access network," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5068–5081, Sept 2015.
- [19] 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, April 2016.
- [20] S. H. Park, O. Simeone, O. Sahin, and S. Shamaï, "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.
- [21] 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," vol. 16, no. 2, pp. 872–884, Feb. 2017.
- [22] 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, Sept. 2016.
- [23] Y. Shi, J. Zhang, and K. B. Letaief, "Robust group sparse beamforming for multicast green Cloud-RAN with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4647–4659, Sept. 2015.
- [24] T. V. Do, H. A. L. Thi, and N. T. Nguyen, *Advanced Computational Methods for Knowledge Engineering*. Springer, 2014.
- [25] Y. Cheng, M. Pesavento, and A. Philipp, "Joint network optimization and downlink beamforming for CoMP transmissions using mixed integer conic programming," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3972–3987, Aug 2013.
- [26] Z. Fang, X. Wang, and X. Yuan, "Joint base station activation and downlink beamforming design for heterogeneous networks," in *Proc. IEEE Global Commu. Conf. (GLOBECOM)*, Dec 2015, pp. 1–6.
- [27] J. Lee and S. Leyffer, *Mixed Integer Nonlinear Programming*. New York, USA: Springer, 2011.
- [28] Z. q. Luo, W. k. Ma, A. M. c. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.
- [29] S. Boyd and L. Vandenberghe, *Convex optimization*, 1st ed., S. Boyd, Ed. Cambridge, 2004.

- [30] S. Boyd, "Branch and bound method," <http://www.stanford.edu/class/ee364b/lectures.html>, 2011.
- [31] M. Lobo, L. Vandenberghe, S. Boyd, and H. Le Bret, "Applications of second-order cone programming," *Linear Algebra Appl., Special Issue on Linear Algebra in Control, Signals and Image Processing*, pp. 193–228, Nov. 1998.
- [32] A. Beck, A. Ben-Tal, and L. Tretushvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *Journal of Global Optimization*, vol. 47, no. 1, pp. 29–51, 2010.
- [33] 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, August 1978.
- [34] Q.-D. Vu, L.-N. Tran, R. Farrell, and E.-K. Hong, "An efficiency maximization design for SWIPT," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2189–2193, Dec 2015.
- [35] H. Tuy, *Convex Analysis and Global Optimization*. Springer US, 1998.
- [36] T. P. Dinh and H. A. L. Thi, "Recent advances in DC programming and DCA," *Transactions on Computational Intelligence XIII*, vol. 8342, pp. 1–37, April 2014.
- [37] T. Lipp and S. Boyd, "Variations and extension of convex-concave procedure," *Optimization and Engineering*, no. 3, pp. 1–25, 2015.
- [38] T. Dinh Quoc and M. Diehl, "Sequential Convex Programming Methods for Solving Nonlinear Optimization Problems with DC constraints," *ArXiv e-prints*, Jul. 2011.
- [39] O. Tervo, L. N. Tran, and M. Juntti, "Optimal energy-efficient transmit beamforming for multi-user MISO downlink," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5574–5588, Oct. 2015.
- [40] O. Günlük and J. Linderoth, "Perspective reformulations of mixed integer nonlinear programs with indicator variables," *Math. Programming*, vol. 124, no. 1, pp. 183–205, July 2010.
- [41] M. B. Shoukry and T. N. Davidson, "Convex conic formulations of robust downlink precoder designs with quality of service constraints," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 714–724, Dec 2007.
- [42] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," *SIAM Rev.*, vol. 53, no. 3, pp. 464–501, Aug. 2011.
- [43] M. F. Hanif, L. N. Tran, A. Tölli, and M. Juntti, "Computationally efficient robust beamforming for SINR balancing in multicell downlink with applications to large antenna array systems," vol. 62, no. 6, pp. 1908–1920, June 2014.
- [44] E. Björnson, G. Zheng, M. Bengtsson, and B. Ottersten, "Robust monotonic optimization framework for multicell MISO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2508–2523, May 2012.
- [45] M. F. Hanif, L.-N. Tran, A. Tölli, M. Juntti, and S. Glisic, "Efficient solutions for weighted sum rate maximization in multicellular networks with channel uncertainties," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5659–5674, Nov. 2013.
- [46] A. Ben-Tal, L. E. Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton University Press, 2009.
- [47] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, Feb 2003.
- [48] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar, and N. Sidiropoulos, "Feasible point pursuit and successive approximation of non-convex QCQPs," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 804–808, July 2015.
- [49] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*. Soc. Ind. Appl. Math. (SIAM), 2001.
- [50] 3GPP, "E-UTRA: Further advancements for E-UTRA physical layer aspects (Release 9)," 3rd Generation Partnership Project, TR 36.814.
- [51] J. Löfberg, "YALMIP : A toolbox for modeling and optimization in MATLAB," in *Proc. the CACSD Conference*, Taipei, Taiwan., 2004.
- [52] I. MOSEK ApS, 2014, [Online]. Available: www.mosek.com.