Robust Beamforming Design in C-RAN with Sigmoidal Utility and Capacity-Limited Backhaul

Zehua Wang, *Member, IEEE*, Derrick Wing Kwan Ng, *Member, IEEE*, Vincent W.S. Wong, *Fellow, IEEE*, and Robert Schober, *Fellow, IEEE*

***Abstract*—In the paper, we study the robust beamforming design in cloud radio access networks where remote radio heads (RRHs) are connected to a cloud server that performs signal processing and resource allocation in a centralized manner. Different from traditional approaches adopting a concave increasing function to model the utility of a user, we model the utility by a sigmoidal function of the signal-to-interference-plus-noise ratio (SINR) to capture the diminishing utility returns for very small and very large SINRs in real-time applications (e.g. video streaming). Our objective is to maximize the aggregate utility of the users while taking into account the imperfection of channel state information (CSI), limited backhaul capacity, and minimum quality of service requirements. Because of the sigmoidal utility function and some of the constraints, the formulated problem is non-convex. To efficiently solve the problem, we introduce a maximum interference constraint, transform the CSI uncertainty constraints into linear matrix inequalities, employ convex relaxation to handle the backhaul capacity constraints, and exploit the sum-of-ratios form of the objective function. This leads to an efficient resource allocation algorithm which outperforms several baseline schemes and closely approaches a performance upper bound for large CSI uncertainty or large number of RRHs.**

Index Terms: C-RAN, beamforming design, imperfect CSI, sigmoidal utility function, capacity-limited backhaul.

I. INTRODUCTION

Deploying base stations (BSs) densely is a viable approach to meet the tremendous data traffic demands in the fifth generation (5G) wireless communication networks [2]. However, this may also increase the capital and operational expenditure of mobile network operators (MNOs) and the user equipments (UEs) may suffer from severe multicell interference caused by simultaneous transmissions in adjacent cells [3]. The recently proposed cloud radio access network (C-RAN) architecture is considered to be a promising architecture to overcome these problems in 5G wireless networks [4]. In C-RAN, the radio signal transceiver module and the baseband signal processing module of conventional BSs are detached. The baseband signal processing module is located at a cloud server and is referred to as *baseband unit* (BBU). The BS, which is only composed of radio signal transceivers in C-RAN, is referred to as *remote radio head* (RRH). Ideally, the backhaul communication between the RRHs and the BBUs is implemented by optical fibers. Multiple BBUs running on a cloud server can form a computationally powerful BBU pool, where the baseband signals are processed in a centralized manner. Thus, not only the cost of deploying a new BS can be significantly reduced, but also coordinated multipoint (CoMP) transmission can be seamlessly applied to mitigate the interference caused by nearby BSs.

Due to these advantages inherent to C-RAN, coordinated beamforming design has been studied for C-RAN in the literature [5]–[9] to improve system performance. The authors in [5] formulated an optimization problem for the beamforming design in CoMP networks with the objective of minimizing the backhaul data traffic. Beamforming design for reducing the energy consumption and increasing the energy efficiency of C-RAN was studied in [6]–[8] for various network scenarios. The work in [6] assumed imperfect channel state information (CSI) for beamforming design for downlink data transmission. The work in [7] investigated beamforming design for both uplink and downlink for minimization of the energy consumption in C-RAN. Maximizing the energy efficiency via cooperative beamforming was proposed in [8]. The authors in [9] formulated a multi-objective optimization problem to jointly reduce the backhaul data traffic and the energy consumption of the RRHs. However, it was assumed in [6]–[9] that an unlimited amount of control signals, user CSI, and precoding data can be exchanged over the backhaul.

In practice, the backhaul capacity is limited. Taking this constraint into account for beamforming design is crucial in C-RAN. In the literature, there are two strategies to limit the amount of backhaul data traffic in C-RAN, namely, the compression strategy and the data sharing strategy [10]. For the compression strategy, the backhaul data traffic is reduced by adopting source coding techniques. Specifically, the resolution of the compressed signals is adjusted according to the backhaul capacity. The compression strategy has been investigated for both uplink and downlink transmission in C-RAN. In particular, the authors of [11] and [12] assumed that the quantization noises at different RRHs are uncorrelated and

Manuscript received September 26, 2016; revised April 3, 2017; accepted May 23, 2017. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). Part of this paper was presented at the *IEEE International Conference on Communications (ICC)*, Kuala Lumpur, Malaysia, May 2016 [1]. The review of this paper was coordinated by Prof. Pierluigi Salvo Rossi.

Zehua Wang and Vincent W.S. Wong are with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, V6T 1Z4, Canada (e-mail: zwang@ece.ubc.ca; vincentw@ece.ubc.ca).

Derrick Wing Kwan Ng is with School of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney, Australia (email: w.k.ng@unsw.edu.au). He is supported under Australian Research Council's Discovery Early Career Researcher Award funding scheme (DE170100137).

Robert Schober is with the Institute for Digital Communications, Friedrich-Alexander University of Erlangen–Nuremberg, Germany (email: robert.schober@fau.de)

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

Digital Object Identifier 10.1109/TWC.2017.XXXXXXX

used independent compression for uplink data transmission. By exploiting the correlation of the quantization noises at different RRHs, distributed compression schemes have been proposed for uplink data transmission in [13]–[18]. On the other hand, the authors of [19] studied downlink data transmission employing independent compression. Independent compression for downlink data transmission with imperfect CSI was studied in [20]. Furthermore, the authors of [21] investigated distributed compression for downlink data transmission in C-RAN. However, different mobile applications running on the UEs may require different resolutions for the received signals, which increases the complexity of the baseband signal processing if the compression strategy is employed. For the data sharing strategy, the amount of backhaul data traffic of an RRH is determined by the data traffic of the UEs that are associated with the RRH. The association problem between UEs and RRHs can be solved either by a clustering approach [22] or a user-centric approach [10]. The former approach allows multiple RRHs to form a cluster for serving multiple UEs. However, geographic boundaries exist between adjacent RRH clusters. The UEs located at the boundaries of RRH clusters may suffer from strong co-channel interference. The latter approach, in contrast, dynamically selects suitable RRHs to serve individual UE by exploiting the benefits of interference management. In fact, the user-centric approach can effectively reduce the co-channel interference by associating each UE to multiple RRHs and employing an appropriate beamforming design. Thus, different from [19]–[21], in this paper, we adopt the data sharing strategy and use the user-centric approach for the association of UEs and RRHs.

Mobile applications running on UEs require different amounts of network resources to achieve the desired quality of service (QoS). For example, the signal-to-interference-plus-noise ratios (SINRs) needed for online video and audio streaming applications to ensure smooth video and audio services are different. With the C-RAN architecture, the MNO can allocate the limited network resources efficiently via cooperative beamforming. Although many existing works [23], [24] target the maximization of the system sum rate characterized by a sum of concave increasing functions of the SINRs, sigmoidal functions with the received SINR as the input parameter constitute better models for the utility achieved by mobile users [25], [26]. However, sigmoidal functions are non-convex and thus determining the optimal beamforming vectors is a challenging task. Furthermore, due to the channel noise, interference, and time varying nature of wireless channels, only imperfect CSI can be obtained and exploited for beamforming design in practice. Note that the baseband signal processing in C-RAN is performed by the BBU pool on a cloud server. Thus, the CSI estimated by the RRHs needs to be first conveyed to the cloud server via the capacity-limited backhaul links. Then, the precoded signals are transmitted from the BBU pool to the RRHs. The resulting round trip delay in the backhaul and the associated signal processing delay further add to the imperfection of the estimated CSI used for resource allocation. If the actual link quality between the RRHs and a UE is worse than the estimated value, then the UE may not be able to decode the signal received from the RRHs. In this case, the utility of the serving UE may be significantly reduced.

To address above issues, in this paper, we focus on the utility based beamforming design in C-RAN where we take into account both the imperfection of the CSI and the capacity-limited backhaul. To the best of our knowledge, beamforming design for aggregate utility maximization of mobile users in C-RAN with imperfect CSI and capacity-limited backhaul links has not yet been studied in the existing literature [5]–[10], [18], [21], [22]. We first formulate the robust beamforming design as an optimization problem. The problem is generally intractable since it has a non-convex objective function, non-convex combinatorial constraints due to the limited backhaul capacity, and infinitely many constraints due to the channel uncertainty. To strike a balance between system performance and the computational complexity of solving the problem, we focus on the design of a computationally efficient resource allocation algorithm. In particular, we first introduce an additional robust maximum interference constraint for each mobile user to simplify the considered problem. Subsequently, we transform the infinitely many constraints in our problem to a finite number of linear matrix inequality (LMI) constraints. We then adopt the convex relaxation technique to handle the non-convex combinatorial constraints, such that the transformed problem can be solved in an iterative manner. In each iteration, we introduce an inner loop that exploits the sum-of-ratios form of the objective function to decompose the problem into two subproblems and tackles them with semidefinite programming (SDP) and the damped Newton's method iteratively. Simulation results show that the beamforming design obtained with our proposed algorithm can increase the aggregate utility in C-RAN compared with existing beamforming designs that either maximize the weighted system sum rate (WSSR) or the weighted sum of SINRs of the users.

The remainder of this paper is organized as follows. In Section II, we introduce the system model and present the problem formulation. In Section III, we transform our problem and propose an iterative algorithm to obtain an efficient suboptimal solution. Simulation results are provided in Section IV. Conclusions are drawn in Section V.

Notations: In this paper, the following notations are adopted: $\mathbf{X}^{\mathsf{T}}$, $\mathbf{X}^{\mathsf{H}}$, $\mathsf{Tr}(\mathbf{X})$, and $\mathsf{Rank}(\mathbf{X})$ represent the transpose, conjugate transpose, trace, and rank of matrix $\mathbf{X}$, respectively; $\mathbb{C}$ is the set of complex numbers, $\mathbb{C}^{m\times n}$ represents the set of $m \times n$ complex matrices, $\mathbb{H}^n$ denotes the set of $n \times n$ Hermitian matrices; $|\cdot|$ is the absolute value. $\|\cdot\|_x$ is the ℓ_x-norm. In particular, $\|\cdot\|_0$ is the ℓ_0-norm of a vector and denotes the number of non-zero entries in the vector; $\mathbb{E}[\cdot]$ denotes statistical expectation, $\Re\{x\}$ denotes the real part of complex number x; $\mathbf{x} \succeq \mathbf{0}$ means that each element in vector $\mathbf{x}$ is non-negative, $\mathbf{X} \succeq \mathbf{0}$ (or $\mathbf{X} \succ \mathbf{0}$) means that matrix $\mathbf{X}$ is positive semidefinite (or positive definite), $\mathbf{x}_{[m:n]}$ returns a vector containing the m^{th} to the n^{th} elements of vector $\mathbf{x}$; $\mathbf{I}_n$ is the $n\times n$ identity matrix, $\mathbf{O}_n$ is the $n\times n$ all-zero matrix, $\mathbf{0}_n$ denotes the $n \times 1$ all-zero vector; $\otimes$ stands for the Kronecker product, and $\mathcal{CN}(0, \sigma^2)$ is the zero-mean complex Gaussian distribution with variance σ^2.

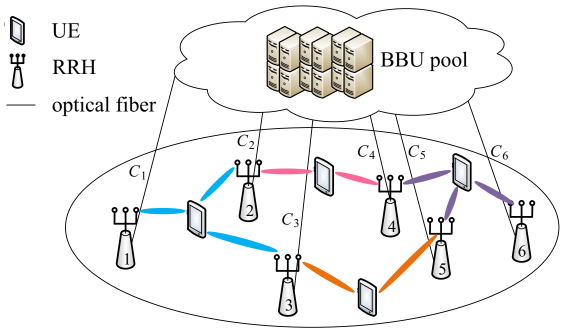


Fig. 1. An example of a C-RAN, where four UEs are served by six RRHs via cooperative beamforming. A BBU pool is hosted by a cloud server. The RRHs communicate with the BBU pool on the cloud server over backhaul links implemented by optical fibers having limited capacities denoted by $C_1, C_2, \ldots, C_6$. The MNO can control the RRHs and allocate network resources to UEs in a centralized manner.

II. System Model and Problem Formulation

We now present our system model and formulate the problem. We consider downlink data transmission in a C-RAN. An example of the considered system is shown in Fig. 1.

A. System Model

Let $\mathcal{M} = \{1, \ldots, M\}$ denote the set of RRHs in the C-RAN. Each RRH is equipped with $N \geq 1$ antennas. We assume that each mobile user has one UE. Therefore, in the sequel, we use the terms "mobile user" and "UE" interchangeably. Let $\mathcal{K} = \{1, \ldots, K\}$ denote the set of UEs. We assume that each UE in set $\mathcal{K}$ is equipped with a single antenna to limit the receiver complexity.. As beamforming and CoMP are employed, a UE can be associated with multiple RRHs simultaneously. The precoded signal transmitted from RRH $m \in \mathcal{M}$ to UE $k \in \mathcal{K}$ is given by $\mathbf{w}_{m,k} s_k$, where $\mathbf{w}_{m,k} \in \mathbb{C}^{N\times 1}$ is the beamforming vector for UE k employed by RRH m and $s_k \in \mathbb{C}$ denotes the data symbol for UE k. Without loss of generality, we assume that $\mathbb{E}\left[|s_k|^2\right] = 1, \forall\, k \in \mathcal{K}$. We note that when $\mathbf{w}_{m,k} \neq \mathbf{0}_N$, UE k is associated with RRH m. Otherwise, $\mathbf{w}_{m,k} = \mathbf{0}_N$ holds. Therefore, the signal received at UE $k \in \mathcal{K}$ can be written as

$$\underbrace{\sum_{m\in\mathcal{M}} \mathbf{h}_{m,k}^{\mathsf{H}} \mathbf{w}_{m,k} s_k}_{\text{desired signal}} + \underbrace{\sum_{u\in\mathcal{K}\setminus\{k\}} \sum_{m\in\mathcal{M}} \mathbf{h}_{m,k}^{\mathsf{H}} \mathbf{w}_{m,u} s_u}_{\text{interfering signals}} + n_k, \tag{1}$$

where $\mathbf{h}_{m,k} \in \mathbb{C}^{N\times 1}$ denotes the instantaneous channel vector from RRH m to UE k and $n_k \sim \mathcal{CN}(0, \sigma_k^2)$ denotes the noise at UE k with power σ_k^2. The received SINR at UE k is given by

$$\gamma_k = \frac{\left|\sum_{m\in\mathcal{M}} \mathbf{h}_{m,k}^{\mathsf{H}} \mathbf{w}_{m,k}\right|^2}{\sum_{u\in\mathcal{K}\setminus\{k\}} \left|\sum_{m\in\mathcal{M}} \mathbf{h}_{m,k}^{\mathsf{H}} \mathbf{w}_{m,u}\right|^2 + \sigma_k^2}. \tag{2}$$

Due to the non-negligible round-trip delay in the backhaul and the imperfection of CSI estimation, the actual CSI, $\mathbf{h}_{m,k}$, from RRH $m \in \mathcal{M}$ to UE $k \in \mathcal{K}$ in (1) may deviate from the estimated CSI used by the BBU pool for resource allocation. Similar to [27], [28], we adopt a deterministic model to capture the CSI uncertainty. Let $\widehat{\mathbf{h}}_{m,k} \in \mathbb{C}^{N\times 1}$ denote the estimated CSI from RRH m to UE k that is used for the beamforming design at the cloud server. For notational simplicity, we introduce $\mathbf{h}_k \triangleq \left[\mathbf{h}_{1,k}^{\mathsf{H}} \ldots \mathbf{h}_{M,k}^{\mathsf{H}}\right]^{\mathsf{H}}$ and $\widehat{\mathbf{h}}_k \triangleq \left[\widehat{\mathbf{h}}_{1,k}^{\mathsf{H}} \ldots \widehat{\mathbf{h}}_{M,k}^{\mathsf{H}}\right]^{\mathsf{H}}$. In the following, we assume $\mathbf{h}_k \neq \mathbf{0}_{MN}$ and $\widehat{\mathbf{h}}_k \neq \mathbf{0}_{MN}, \forall\, k \in \mathcal{K}$. According to the deterministic model in [27], [28], we can capture the CSI uncertainty as follows:

$$\mathbf{h}_k = \widehat{\mathbf{h}}_k + \Delta\mathbf{h}_k, \qquad \forall\, k \in \mathcal{K}, \tag{3}$$

$$\Omega_k \triangleq \left\{\Delta\mathbf{h}_k : \Delta\mathbf{h}_k^{\mathsf{H}} \Delta\mathbf{h}_k \leq \varepsilon_k^2\right\}, \qquad \forall\, k \in \mathcal{K}, \tag{4}$$

where $\Delta\mathbf{h}_k \in \mathbb{C}^{MN\times 1}$ denotes the CSI uncertainty of the channel from the RRHs in set $\mathcal{M}$ to UE k. Constant ε_k is the radius of the uncertainty region Ω_k, which depends on the degree of imperfection of the channel estimation, the coherence time of the wireless channels from the RRHs to user k, and the round-trip delay of the backhaul from the BBU pool to the RRHs.

We assume that each UE executes a single mobile application[1]. In general, for mobile users running real-time applications, the utility increases with the received SINR. However, when the SINR achieved by a user is either very small or very large, the marginal utility benefits for increasing SINR may be negligible. For example, when the SINR achieved by a user using an online video application is lower than a threshold value such that the video cannot be smoothly played for the user even with the lowest possible quality, the utility of the user will not be notably increased until the SINR achieved by the user is greater than that threshold value. On the other hand, when the SINR achieved by the user is high enough such that the video can be smoothly played with the highest possible quality, allocating additional network resources to this user cannot further increase his utility. In some cases, users may always achieve higher utility when their received SINRs are increased, such as when downloading a file or buffering a video. Nevertheless, even in these cases, the improvement in data rate will be limited as the received data rate is a logarithmic function of the SINR [29]. Besides, as buffering a video for a period of time can avoid video freezing and interruptions, the shorter the amount of time that a user has to wait for buffering before the video can be played smoothly, the higher the utility that the user may obtain. However, when the SINR is sufficiently large such that the downloading data rate is high enough, the amount of time required for buffering becomes negligible. Hence, even in this case, the room for further improving the user's utility by further reducing the time of buffering is limited. Therefore, based on the above analysis, the utility of a user may saturate or have little room for improvement at high SINR. Thus, it is of fundamental importance to incorporate the sigmoidal behaviour of the users' utility into the resource allocation algorithm design [25],

[1]The case that a UE is running multiple mobile applications can be modeled by defining multiple virtual UEs at the same location where each of them runs a single application. We note that this approach may not be efficient when the compression strategy is used to overcome the capacity-limited backhaul, as, in this case, the correlation of the channels of the virtual devices should be exploited to improve performance [21].

[26]. In this paper, we adopt the weighted sigmoidal function to model the utility of UE $k \in \mathcal{K}$ experiencing SINR γ_k as follows[2]:

$$g_k(\gamma_k) = \frac{\eta_k}{1+\exp\big(-a_k(\gamma_k - b_k)\big)}, \tag{5}$$

where constant parameters $a_k, b_k > 0$ depend on the application running on UE k, and constant parameter $\eta_k > 0$ is the weight factor of UE k. Parameter a_k controls the steepness of $g_k(\gamma_k)$. The larger a_k is, the faster $g_k(\gamma_k)$ increases with γ_k. It is reasonable to assume that the utility of UE k approaches 0 if $\gamma_k \to 0$. Thus, we need $g_k(0) \approx 0$, which holds if the product $a_k b_k$ is sufficiently large. We assume that parameters a_k, b_k, and η_k are known once the considered application is launched on UE k.

B. Problem Formulation

We aim to maximize the aggregate utility of the mobile users in set $\mathcal{K}$. However, in the considered system, the actual CSI is not perfectly known when the precoded signals are transmitted from the RRHs to the UEs. Thus, the actual SINR of each UE is not known at the transmitters. Furthermore, the transmit signals are precoded by the BBU pool and transmitted to the RRHs via the capacity-limited backhaul. Therefore, each RRH in set $\mathcal{M}$ may only be used to serve a subset of UEs in set $\mathcal{K}$ due to the limited backhaul capacity. Moreover, the beamforming design has to guarantee that the precoded signals transmitted to the RRHs can be successfully forwarded from the RRHs to the UEs over the wireless channel. Otherwise, if the serving UEs cannot decode the received signals, the utility of the mobile users will be reduced. We thus formulate the following optimization problem for the beamforming design:

$$\underset{\mathbf{w},\,\boldsymbol{\varphi}}{\text{maximize}} \sum_{k\in\mathcal{K}} g_k(\varphi_k) \tag{6a}$$

$$\text{subject to } \varphi_k \le \min_{\Delta\mathbf{h}_k\in\Omega_k} \gamma_k, \quad \forall\, k\in\mathcal{K}, \tag{6b}$$

$$\sum_{k\in\mathcal{K}} \big\| \|\mathbf{w}_{m,k}\|_2 \big\|_0 B\log_2(1+\varphi_k) \le C_m, \quad \forall\, m\in\mathcal{M}, \tag{6c}$$

$$\sum_{k\in\mathcal{K}} \|\mathbf{w}_{m,k}\|_2^2 \le p_m, \quad \forall\, m\in\mathcal{M}, \tag{6d}$$

$$\Gamma_{\text{req},k} \le \min_{\Delta\mathbf{h}_k\in\Omega_k} \gamma_k, \quad \forall\, k\in\mathcal{K}, \tag{6e}$$

where vectors $\mathbf{w} \triangleq \big[\mathbf{w}_{1,1}^{\mathsf{H}} \dots \mathbf{w}_{M,1}^{\mathsf{H}} \dots \mathbf{w}_{1,K}^{\mathsf{H}} \dots \mathbf{w}_{M,K}^{\mathsf{H}}\big]^{\mathsf{H}}$ and $\boldsymbol{\varphi} \triangleq (\varphi_1, \dots, \varphi_K)$ are the optimization variables. The auxiliary optimization variable φ_k in constraint (6b) serves as an SINR lower bound for UE $k \in \mathcal{K}$ given the channel uncertainty. The auxiliary optimization variable φ_k is then used to evaluate sigmoidal function $g_k(\varphi_k)$ in the objective function for UE k. In constraint (6c), B denotes the bandwidth of the wireless channel. Moreover, we have $\big\| \|\mathbf{w}_{m,k}\|_2 \big\|_0 = 0$ if and only if $\mathbf{w}_{m,k} = \mathbf{0}_N$, $\forall\, m\in\mathcal{M}, k\in\mathcal{K}$. Thus, the left-hand side (LHS) of constraint (6c) is the aggregate data rate of the UEs associated with RRH m. Hence, constraint (6c) accounts for the limited backhaul capacity, where constant C_m denotes the backhaul capacity of RRH m. Constraint (6d) restricts the total power used by RRH m for beamforming not to exceed the maximum transmit power p_m. Constant $\Gamma_{\text{req},k}$ in (6e) is the minimum required SINR for UE k. Constraint (6e) is introduced to ensure that the minimum signal strength required for signal detection is achieved or/and the QoS of basic wireless communication services is sufficiently high.

Problem (6) is difficult to solve due to the following reasons: objective function (6a) is in a non-convex sum-of-ratios form; constraints (6b) and (6e) involve infinitely many inequality constraints due to the continuity of the CSI uncertainty region Ω_k, $\forall\, k\in\mathcal{K}$; and constraint (6c) is a combinatorial constraint. In general, there is no systematic and efficient approach to solve this kind of non-convex optimization problem. Besides, finding the globally optimal solution for problem (6) via a brute-force approach entails a prohibitively high computational complexity. Therefore, solving problem (6) directly is a challenging task. Hence, we propose an iterative algorithm to obtain an efficient suboptimal solution in the following section.

III. Problem Transformation and Suboptimal Solution

In this section, we will transform problem (6) into a tractable problem. For notational simplicity, we first introduce $\mathbf{w}_k \triangleq \big[\mathbf{w}_{1,k}^{\mathsf{H}} \dots \mathbf{w}_{M,k}^{\mathsf{H}}\big]^{\mathsf{H}}$ to represent the beamforming vector from all RRHs in set $\mathcal{M}$ to UE $k \in \mathcal{K}$. The beamforming vector of RRH $m \in \mathcal{M}$ for UE k can be expressed as $\mathbf{w}_{m,k} = \mathbf{B}_m \mathbf{w}_k$, where $\mathbf{B}_m$ is a constant matrix defined as $\mathbf{B}_m \triangleq \big(\mathbf{0}_{m-1}^{\mathsf{T}}, 1, \mathbf{0}_{M-m}^{\mathsf{T}}\big) \otimes \mathbf{I}_N$. Problem (6) can be equivalently rewritten as follows:

$$\underset{\mathbf{w},\,\boldsymbol{\varphi}}{\text{maximize}} \sum_{k\in\mathcal{K}} g_k(\varphi_k) \tag{7a}$$

$$\text{subject to } \varphi_k \le \min_{\Delta\mathbf{h}_k\in\Omega_k} \frac{\mathsf{Tr}\big(\mathbf{h}_k\mathbf{h}_k^{\mathsf{H}}\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big)}{\sum_{u\in\mathcal{K}\setminus\{k\}} \mathsf{Tr}\big(\mathbf{h}_k\mathbf{h}_k^{\mathsf{H}}\mathbf{w}_u\mathbf{w}_u^{\mathsf{H}}\big) + \sigma_k^2}, \quad \forall\, k\in\mathcal{K}, \tag{7b}$$

$$\sum_{k\in\mathcal{K}} \big\| \mathsf{Tr}\left(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\right) \big\|_0 B\log_2(1+\varphi_k) \le C_m, \quad \forall\, m\in\mathcal{M}, \tag{7c}$$

$$\sum_{k\in\mathcal{K}} \mathsf{Tr}\left(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\right) \le p_m, \quad \forall\, m\in\mathcal{M}, \tag{7d}$$

$$\Gamma_{\text{req},k} \le \min_{\Delta\mathbf{h}_k\in\Omega_k} \frac{\mathsf{Tr}\left(\mathbf{h}_k\mathbf{h}_k^{\mathsf{H}}\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\right)}{\sum_{u\in\mathcal{K}\setminus\{k\}} \mathsf{Tr}\left(\mathbf{h}_k\mathbf{h}_k^{\mathsf{H}}\mathbf{w}_u\mathbf{w}_u^{\mathsf{H}}\right) + \sigma_k^2}, \quad \forall\, k\in\mathcal{K}. \tag{7e}$$

In the following, we transform problem (7) into a tractable form.

A. Interference Decoupling and Constraints Transformation

To handle the non-convexity of constraint (7b), inspired by [30], we introduce the following robust maximum interference constraint for each UE in set $\mathcal{K}$:

$$\max_{\Delta\mathbf{h}_k\in\Omega_k} \sum_{u\in\mathcal{K}\setminus\{k\}} \mathsf{Tr}\left(\mathbf{h}_k\mathbf{h}_k^{\mathsf{H}}\mathbf{w}_u\mathbf{w}_u^{\mathsf{H}}\right) \le I, \quad \forall\, k\in\mathcal{K}, \tag{8}$$

[2]We note that the resource allocation algorithm developed in this paper for the utility function given in (5) is also applicable for other sigmoidal utility functions such as $g_k(\gamma_k) = \frac{\eta_k}{2}\Big(\frac{a_k(\gamma_k - b_k)}{\sqrt{1+(a_k(\gamma_k-b_k))^2}} + 1\Big)$, cf. Section III-C.

where I is a predefined upper bound on the interference experienced by each mobile user despite the channel uncertainty. That is, I is not an optimization variable. Introducing the additional constraint in (8) has three benefits. First, the C-RAN can control the amount of interference experienced by UEs. Second, the interference is decoupled from the objective function. A suitable value of I can be obtained by offline simulations, cf. Section IV. Third, by replacing the power of the interference signals in the SINRs by their upper bound I, we can obtain a lower bound on the aggregate utility of all UEs. Assuming a suitable value for I, we solve the following problem to obtain a suboptimal solution for problem (7):

$$\underset{\mathbf{w},\,\boldsymbol{\varphi}}{\text{maximize}} \sum_{k\in\mathcal{K}} g_k(\varphi_k) \tag{9a}$$

$$\text{subject to } \varphi_k \leq \min_{\Delta\mathbf{h}_k\in\Omega_k} \frac{\mathsf{Tr}\big(\mathbf{h}_k\mathbf{h}_k^{\mathsf{H}}\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big)}{I+\sigma_k^2}, \quad \forall\, k\in\mathcal{K}, \tag{9b}$$

$$\Gamma_{\text{req},k} \leq \min_{\Delta\mathbf{h}_k\in\Omega_k} \frac{\mathsf{Tr}\left(\mathbf{h}_k\mathbf{h}_k^{\mathsf{H}}\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\right)}{I+\sigma_k^2}, \ \forall\, k\in\mathcal{K}, \tag{9c}$$

$$I \geq \max_{\Delta\mathbf{h}_k\in\Omega_k} \sum_{u\in\mathcal{K}\setminus\{k\}} \mathsf{Tr}\left(\mathbf{h}_k\mathbf{h}_k^{\mathsf{H}}\mathbf{w}_u\mathbf{w}_u^{\mathsf{H}}\right), \ \forall\, k\in\mathcal{K}, \tag{9d}$$

constraints (7c) and (7d).

Constraints (9b), (9c), and (9d) involve infinitely many inequality constraints. We handle constraints (9b), (9c), and (9d) by transforming them into LMI constraints by exploiting the following lemma:

Lemma 1 (*S-procedure [31, pp. 655]*): Let $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{H}^{MN}$, $\mathbf{d}_1, \mathbf{d}_2 \in \mathbb{C}^{MN\times 1}$, and $y_1, y_2 \in \mathbb{R}$. Consider the following two functions of vector $\mathbf{x}\in\mathbb{C}^{MN\times 1}$:

$$f_1(\mathbf{x}) = \mathbf{x}^{\mathsf{H}}\mathbf{A}_1\mathbf{x} + 2\Re\{\mathbf{d}_1^{\mathsf{H}}\mathbf{x}\} + y_1, \tag{10a}$$

$$f_2(\mathbf{x}) = \mathbf{x}^{\mathsf{H}}\mathbf{A}_2\mathbf{x} + 2\Re\{\mathbf{d}_2^{\mathsf{H}}\mathbf{x}\} + y_2. \tag{10b}$$

The implication $f_1(\mathbf{x})\leq 0 \Rightarrow f_2(\mathbf{x})\leq 0$ holds if and only if there exists a $\theta\geq 0$ such that

$$\theta\begin{bmatrix}\mathbf{A}_1 & \mathbf{d}_1\\ \mathbf{d}_1^{\mathsf{H}} & y_1\end{bmatrix} - \begin{bmatrix}\mathbf{A}_2 & \mathbf{d}_2\\ \mathbf{d}_2^{\mathsf{H}} & y_2\end{bmatrix} \succeq \mathbf{0}, \tag{11}$$

provided that there exists a point $\widetilde{\mathbf{x}}$ that satisfies $f_1(\widetilde{\mathbf{x}})<0$.

We first apply Lemma 1 to constraint (9b). We obtain the following implication:

$$\begin{aligned}&\Delta\mathbf{h}_k^{\mathsf{H}}\mathbf{I}_{MN}\Delta\mathbf{h}_k + 2\Re\left\{\mathbf{0}_{MN}^{\mathsf{H}}\Delta\mathbf{h}_k\right\} - \varepsilon_k^2 \leq 0\\ \Rightarrow & -\Delta\mathbf{h}_k^{\mathsf{H}}\big(\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big)\Delta\mathbf{h}_k - 2\Re\big\{\big(\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\,\widehat{\mathbf{h}}_k\big)^{\mathsf{H}}\Delta\mathbf{h}_k\big\}\\ &-\widehat{\mathbf{h}}_k^{\mathsf{H}}\big(\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big)\widehat{\mathbf{h}}_k + \varphi_k\big(I+\sigma_k^2\big)\leq 0,\end{aligned} \tag{12}$$

if and only if there exists a $\vartheta_k\geq 0$ such that the following LMI holds:

$$\underbrace{\begin{bmatrix}\vartheta_k\mathbf{I}_{MN} & \mathbf{0}_{MN}\\ \mathbf{0}_{MN}^{\mathsf{H}} & -\varphi_k\big(I+\sigma_k^2\big)-\vartheta_k\varepsilon_k^2\end{bmatrix}}_{\overline{\mathbf{S}}_{k,1}(\varphi_k,\vartheta_k)} + \mathbf{Q}_k^{\mathsf{H}}\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\mathbf{Q}_k \succeq \mathbf{0}, \tag{13}$$

where $\mathbf{Q}_k = \big[\mathbf{I}_{MN}\ \widehat{\mathbf{h}}_k\big], \forall\, k\in\mathcal{K}$. This is because we have $\mathbf{0}_{MN}\in\Omega_k$ such that $f_1(\mathbf{0}_{MN}) = -\varepsilon_k^2 < 0, \forall\, k\in\mathcal{K}$. Similarly, for constraint (9c), we have the following implication

$$\begin{aligned}&\Delta\mathbf{h}_k^{\mathsf{H}}\mathbf{I}_{MN}\Delta\mathbf{h}_k + 2\Re\left\{\mathbf{0}_{MN}^{\mathsf{H}}\Delta\mathbf{h}_k\right\} - \varepsilon_k^2 \leq 0\\ \Rightarrow & -\Delta\mathbf{h}_k^{\mathsf{H}}\big(\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big)\Delta\mathbf{h}_k - 2\Re\big\{\big(\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\,\widehat{\mathbf{h}}_k\big)^{\mathsf{H}}\Delta\mathbf{h}_k\big\}\\ &-\widehat{\mathbf{h}}_k^{\mathsf{H}}\big(\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big)\widehat{\mathbf{h}}_k + \Gamma_k\big(I+\sigma_k^2\big)\leq 0,\end{aligned} \tag{14}$$

if and only if there exists a $\varrho_k\geq 0$ such that the following LMI holds:

$$\underbrace{\begin{bmatrix}\varrho_k\mathbf{I}_{MN} & \mathbf{0}_{MN}\\ \mathbf{0}_{MN}^{\mathsf{H}} & -\Gamma_{\text{req},k}\big(I+\sigma_k^2\big)-\varrho_k\varepsilon_k^2\end{bmatrix}}_{\overline{\mathbf{S}}_{k,2}(\varrho_k)} + \mathbf{Q}_k^{\mathsf{H}}\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\mathbf{Q}_k \succeq \mathbf{0}, \quad \forall\, k\in\mathcal{K}. \tag{15}$$

Finally, applying Lemma 1 to constraint (9d) yields:

$$\begin{aligned}&\Delta\mathbf{h}_k^{\mathsf{H}}\mathbf{I}_{MN}\Delta\mathbf{h}_k + 2\Re\left\{\mathbf{0}_{MN}^{\mathsf{H}}\Delta\mathbf{h}_k\right\} - \varepsilon_k^2 \leq 0\\ \Rightarrow\, &\Delta\mathbf{h}_k^{\mathsf{H}}\big(\textstyle\sum_{u\in\mathcal{K}\setminus\{k\}}\mathbf{w}_u\mathbf{w}_u^{\mathsf{H}}\big)\Delta\mathbf{h}_k\\ &+2\Re\big\{\big(\big(\textstyle\sum_{u\in\mathcal{K}\setminus\{k\}}\mathbf{w}_u\mathbf{w}_u^{\mathsf{H}}\big)\widehat{\mathbf{h}}_k\big)^{\mathsf{H}}\Delta\mathbf{h}_k\big\}\\ &+\widehat{\mathbf{h}}_k^{\mathsf{H}}\big(\textstyle\sum_{u\in\mathcal{K}\setminus\{k\}}\mathbf{w}_u\mathbf{w}_u^{\mathsf{H}}\big)\widehat{\mathbf{h}}_k - I\leq 0,\end{aligned} \tag{16}$$

if and only if there exists a $\xi_k\geq 0$ such that the following LMI holds:

$$\underbrace{\begin{bmatrix}\xi_k\mathbf{I}_{MN} & \mathbf{0}_{MN}\\ \mathbf{0}_{MN}^{\mathsf{H}} & I-\xi_k\varepsilon_k^2\end{bmatrix}}_{\overline{\mathbf{S}}_{k,3}(\xi_k)} - \mathbf{Q}_k^{\mathsf{H}}\big(\textstyle\sum_{u\in\mathcal{K}\setminus\{k\}}\mathbf{w}_u\mathbf{w}_u^{\mathsf{H}}\big)\mathbf{Q}_k \succeq \mathbf{0}, \quad \forall\, k\in\mathcal{K}. \tag{17}$$

Thus, problem (9) can be equivalently rewritten as follows:

$$\underset{\mathbf{w},\,\boldsymbol{\varphi},\,\boldsymbol{\vartheta},\,\boldsymbol{\varrho},\,\boldsymbol{\xi}}{\text{maximize}} \sum_{k\in\mathcal{K}} g_k(\varphi_k) \tag{18a}$$

$$\text{subject to } \overline{\mathbf{S}}_{k,1}(\varphi_k,\vartheta_k) + \mathbf{Q}_k^{\mathsf{H}}\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\mathbf{Q}_k \succeq \mathbf{0}, \forall\, k\in\mathcal{K}, \tag{18b}$$

$$\overline{\mathbf{S}}_{k,2}(\varrho_k) + \mathbf{Q}_k^{\mathsf{H}}\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\mathbf{Q}_k \succeq \mathbf{0}, \quad \forall\, k\in\mathcal{K}, \tag{18c}$$

$$\overline{\mathbf{S}}_{k,3}(\xi_k) - \mathbf{Q}_k^{\mathsf{H}}\big(\textstyle\sum_{u\in\mathcal{K}\setminus\{k\}}\mathbf{w}_u\mathbf{w}_u^{\mathsf{H}}\big)\mathbf{Q}_k \succeq \mathbf{0}, \quad \forall\, k\in\mathcal{K}, \tag{18d}$$

$$\vartheta_k\geq 0, \varrho_k\geq 0, \xi_k\geq 0, \quad \forall\, k\in\mathcal{K}, \tag{18e}$$

constraints (7c) and (7d),

where functions $\overline{\mathbf{S}}_{k,1}(\varphi_k,\vartheta_k)$, $\overline{\mathbf{S}}_{k,2}(\varrho_k)$, and $\overline{\mathbf{S}}_{k,3}(\xi_k)$ are defined in (13), (15), and (17), respectively; $\boldsymbol{\vartheta}\succeq\mathbf{0}$, $\boldsymbol{\varrho}\succeq\mathbf{0}$, and $\boldsymbol{\xi}\succeq\mathbf{0}$ are auxiliary optimization variable vectors whose elements ϑ_k, ϱ_k, and ξ_k, $\forall\, k\in\mathcal{K}$, are introduced in (13), (15), and (17), respectively.

B. Convex Relaxation for Backhaul Constraint

Next, we handle the combinatorial constraint (7c) by applying the *convex relaxation* technique. We note that this technique has also been used in [10] for the design of a computationally efficient resource allocation. We first approximate the LHS of (7c) as follows:

$$\begin{aligned}&\sum_{k\in\mathcal{K}}\big\|\mathsf{Tr}\big(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big)\big\|_0 B\log_2(1+\varphi_k)\\ \approx&\sum_{k\in\mathcal{K}}\big\|q_{m,k}\mathsf{Tr}\big(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big)\big\|_1 R_k \qquad (19a)\\ =&\sum_{k\in\mathcal{K}} q_{m,k}R_k\mathsf{Tr}\left(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\right), \quad \forall\, m\in\mathcal{M}, \qquad (19b)\end{aligned}$$

where $q_{m,k} \geq 0$ is a weight factor, which corresponds to the transmission power from RRH m to user k, and $R_k = B\log_2(1+\varphi_k)$ denotes the downlink data rate of user k, $\forall m \in \mathcal{M}, k \in \mathcal{K}$. In (19a), the ℓ_0-norm is approximated by its convex hull given by the ℓ_1-norm. This approximation is commonly used in compressed sensing to handle problems with ℓ_0-norm [5], [32], [33]. In particular, for $\mathsf{Tr}\big(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big) \neq 0$ and $q_{m,k} = \big(\mathsf{Tr}(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}})\big)^{-1}$, we have $\big\|q_{m,k}\mathsf{Tr}\big(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big)\big\|_1 = \big\|\mathsf{Tr}\big(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big)\big\|_0 = 1$. For $\mathsf{Tr}\big(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big) = 0$, we have $\big\|q_{m,k}\mathsf{Tr}\big(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big)\big\|_1 = \big\|\mathsf{Tr}\big(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big)\big\|_0 = 0, \forall q_{m,k} \in [0,\infty)$. Thus, by letting $q_{m,k} = \big(\mathsf{Tr}(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}) + \tau\big)^{-1}$ with a small regulation factor $\tau > 0$, we have $\big\|q_{m,k}\mathsf{Tr}\big(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big)\big\|_1 \approx \big\|\mathsf{Tr}\big(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big)\big\|_0$. A suboptimal solution of problem (18) can thus be obtained by solving a transformed problem in an iterative manner. Specifically, let $\mathbf{w}_k^{(i)} \triangleq \big[\mathbf{w}_{1,k}^{(i)\,\mathsf{H}} \ldots \mathbf{w}_{M,k}^{(i)\,\mathsf{H}}\big]^{\mathsf{H}}$ and $\varphi_k^{(i)}$ denote the beamforming vector and the guaranteed SINR of UE $k \in \mathcal{K}$ in the solution of the transformed problem in the i^{th} iteration $(i = 0, 1, 2, \ldots)$, respectively. The transformed problem in the $(i+1)^{\text{th}}$ iteration is given as follows:

$$\mathrm{P}^{(i+1)}: \underset{\mathbf{w},\boldsymbol{\varphi},\boldsymbol{\vartheta},\boldsymbol{\varrho},\boldsymbol{\xi}}{\text{maximize}} \sum_{k\in\mathcal{K}} g_k(\varphi_k) \tag{20a}$$

$$\text{subject to} \sum_{k\in\mathcal{K}} q_{m,k}^{(i)} R_k^{(i)} \mathsf{Tr}\big(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big) \leq C_m, \quad \forall m \in \mathcal{M}, \tag{20b}$$

$$\text{constraints (7d), (18b)} - \text{(18e)},$$

where $q_{m,k}^{(i)} \triangleq \big(\mathsf{Tr}(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k^{(i)}\mathbf{w}_k^{(i)\,\mathsf{H}}) + \tau\big)^{-1}$, $\forall m \in \mathcal{M}, k \in \mathcal{K}$, and $R_k^{(i)} \triangleq B\log_2\big(1+\varphi_k^{(i)}\big)$, $\forall k \in \mathcal{K}$. Hence, $q_{m,k}^{(i)}$ and $R_k^{(i)}$ are constants in problem $\mathrm{P}^{(i+1)}$. We note that we omit the superscript $(i+1)$ for the other constants and the optimization variables in problem $\mathrm{P}^{(i+1)}$ for notational simplicity.

The rationale behind handling constraint (7c) by solving problem $\mathrm{P}^{(i+1)}$ is as follows. Without loss of generality, we consider problem $\mathrm{P}^{(i+1)}$ after solving problem $\mathrm{P}^{(i)}$ and obtaining the intermediate solution $\boldsymbol{\Xi}^{(i)} \triangleq \big(\mathbf{w}^{(i)}, \boldsymbol{\varphi}^{(i)}, \boldsymbol{\vartheta}^{(i)}, \boldsymbol{\varrho}^{(i)}, \boldsymbol{\xi}^{(i)}\big)$ for problem $\mathrm{P}^{(i)}$. We have $\mathsf{Tr}\big(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k^{(i)}\mathbf{w}_k^{(i)\,\mathsf{H}}\big) = \|\mathbf{w}_{m,k}^{(i)}\|_2^2$, $\forall m \in \mathcal{M}, k \in \mathcal{K}$, so the value of $q_{m,k}^{(i)}$ is inversely proportional to the transmission power from RRH m to UE k. Since $\mathbf{w}_{m,k}^{(i)}$ is obtained by solving problem $\mathrm{P}^{(i)}$, if the transmission power from RRH m to UE k is smaller than the transmission power from RRH m to UE $u \in \mathcal{K}\setminus\{k\}$, i.e., $\|\mathbf{w}_{m,k}^{(i)}\|_2^2 < \|\mathbf{w}_{m,u}^{(i)}\|_2^2$, $\forall u \in \mathcal{K}\setminus\{k\}$, this indicates that the channel quality from RRH m to UE k is worse than the channel quality from RRH m to the other UEs; so that the aggregate utility would decrease if a higher transmission power was assigned to RRH m for serving UE k. In other words, if the quality of the channel from RRH m to UE k is poor compared with the quality of the channels from RRH m to the other UEs, letting RRH m serve UE k with a high transmission power will increase the interference to other UEs and the resulting total loss of the aggregate utility at other UEs will outweigh the utility increment at UE k. Meanwhile, the smaller the value of $\|\mathbf{w}_{m,k}^{(i)}\|_2^2$ obtained by solving problem $\mathrm{P}^{(i)}$ is, the larger the value of weight factor $q_{m,k}^{(i)}$ that is used in problem $\mathrm{P}^{(i+1)}$. Therefore, solving problem $\mathrm{P}^{(i+1)}$ will force $\|\mathbf{w}_{m,k}^{(i+1)}\|_2^2$ to decrease further in the intermediate solution $\boldsymbol{\Xi}^{(i+1)}$. As the iterations continue, a subset of UEs with relatively poor channel conditions compared to other UEs from a given RRH will be eliminated from being served by this RRH. Second, we note that the term $R_k^{(i)}$ in the first constraint of problem $\mathrm{P}^{(i+1)}$ is the downlink data rate obtained by UE $k \in \mathcal{K}$ after problem $\mathrm{P}^{(i)}$ is solved. Moreover, if UE k is not served by RRH $m \in \mathcal{M}$, we have $\mathsf{Tr}\big(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\big) = 0$. Thereby, only the downlink data rate of the UEs that are served by RRH m is taken into account for the backhaul capacity constraint at RRH m. The proposed iterative procedure generates sparsity in the beamforming vectors and guarantees that the obtained solution after iteratively solving problem $\mathrm{P}^{(i+1)}$ fulfills non-convex combinatorial constraint (7c). It is worth noting that $q_{m,k}^{(0)}$, $\forall m \in \mathcal{M}, k \in \mathcal{K}$, and $R_k^{(0)}$, $\forall k \in \mathcal{K}$, are required for problem $\mathrm{P}^{(1)}$ in the first iteration. In Section IV, we will present a method to determine an initial vector $\mathbf{w}^{(0)}$, so as to obtain suitable values for $q_{m,k}^{(0)}$ and $R_k^{(0)}$.

C. Non-convex Objective Function Transformation

We note that the values of $q_{m,k}^{(i)}$ and $R_k^{(i)}$ in (20b) are known and fixed in problem $\mathrm{P}^{(i+1)}$. Thus, the constraints in problem $\mathrm{P}^{(i+1)}$ are either convex or LMI constraints. However, problem $\mathrm{P}^{(i+1)}$ is still difficult to solve because of the non-convexity of its objective function. We now transform problem $\mathrm{P}^{(i+1)}$ to an equivalent problem based on the following theorem:

Theorem 1: If $\boldsymbol{\Xi}^{(i+1)}$ is the optimal solution to $\mathrm{P}^{(i+1)}$, there exist two vectors $\boldsymbol{\beta}^{(i+1)} = \big(\beta_1^{(i+1)}, \ldots, \beta_K^{(i+1)}\big)$ and $\boldsymbol{\nu}^{(i+1)} = \big(\nu_1^{(i+1)}, \ldots, \nu_K^{(i+1)}\big)$ such that $\boldsymbol{\Xi}^{(i+1)}$ is also an optimal solution of problem (21) which is given as follows:

$$\underset{\mathbf{w},\boldsymbol{\varphi},\boldsymbol{\vartheta},\boldsymbol{\varrho},\boldsymbol{\xi}}{\text{maximize}} \sum_{k\in\mathcal{K}} \nu_k^{(i+1)}\Big(\eta_k - \beta_k^{(i+1)}\Big(1+\exp\big(-a_k(\varphi_k - b_k)\big)\Big)\Big) \tag{21}$$

subject to constraints (7d), (18b) – (18e), and (20b).

Meanwhile, vector $\boldsymbol{\varphi}^{(i+1)}$ in solution $\boldsymbol{\Xi}^{(i+1)}$ satisfies the following system of equations:

$$\beta_k^{(i+1)}\Big(1+\exp\big(-a_k(\varphi_k^{(i+1)} - b_k)\big)\Big) - \eta_k = 0, \ \forall k \in \mathcal{K}, \tag{22a}$$

$$\nu_k^{(i+1)}\Big(1+\exp\big(-a_k(\varphi_k^{(i+1)} - b_k)\big)\Big) - 1 = 0, \quad \forall k \in \mathcal{K}. \tag{22b}$$

Proof: Please refer to Appendix A for a proof of Theorem 1. ■

Remark 1: Theorem 1 can be extended to other objective functions in sum-of-ratios form. As long as each ratio has a concave numerator and a convex denominator, we can obtain a similar proof as in Appendix A and arrive at conclusions similar to those in Theorem 1. For example, $g_k(\gamma_k) = \frac{\eta_k}{2}\Big(\frac{a_k(\gamma_k - b_k)}{\sqrt{1+(a_k(\gamma_k - b_k))^2}} + 1\Big)$ can also be used to define the utility of UE $k \in \mathcal{K}$ with received SINR γ_k. Moreover, for this new utility function, our proposed problem transformation still applies and leads to a similar algorithm as the one presented in Section III-E.

Vectors $\boldsymbol{\beta}^{(i+1)}$ and $\boldsymbol{\nu}^{(i+1)}$ used in Theorem 1 to solve problem $\mathrm{P}^{(i+1)}$ are not known *a priori*. However, these two vectors can be determined in an iterative manner. Since problem $\mathrm{P}^{(i+1)}$ is solved in the $(i+1)^{\text{th}}$ iteration of a loop, we refer to this existing loop as the "outer" loop and to the new loop used to determine vectors $\boldsymbol{\beta}^{(i+1)}$ and $\boldsymbol{\nu}^{(i+1)}$ for problem $\mathrm{P}^{(i+1)}$ as the "inner" loop. For notational simplicity, we refer to the j^{th} $(j=1,2,\ldots)$ iteration of the inner loop in the $(i+1)^{\text{th}}$ iteration of the outer loop as iteration $(i+1,j)$ or the $(i+1,j)^{\text{th}}$ iteration.

In iteration $(i+1,j)$, two subproblems need to be solved. Specifically, before the appropriate vectors $\boldsymbol{\beta}^{(i+1)}$ and $\boldsymbol{\nu}^{(i+1)}$ are found, we denote vectors $\boldsymbol{\beta}^{(i+1,j)}$ and $\boldsymbol{\nu}^{(i+1,j)}$ as the intermediate values of $\boldsymbol{\beta}^{(i+1)}$ and $\boldsymbol{\nu}^{(i+1)}$ in iteration $(i+1,j)$, respectively. Then, we refer to the problem obtained after substituting vectors $\boldsymbol{\beta}^{(i+1,j)}$ and $\boldsymbol{\nu}^{(i+1,j)}$ for $\boldsymbol{\beta}^{(i+1)}$ and $\boldsymbol{\nu}^{(i+1)}$ in problem (21) as the *primary subproblem* in iteration $(i+1,j)$. Let $\boldsymbol{\Xi}^{(i+1,j)} \triangleq \big(\mathbf{w}^{(i+1,j)}, \boldsymbol{\varphi}^{(i+1,j)}, \boldsymbol{\vartheta}^{(i+1,j)}, \boldsymbol{\varrho}^{(i+1,j)}, \boldsymbol{\xi}^{(i+1,j)}\big)$ denote the solution of the primary subproblem. To facilitate the presentation, we first define the following $2K$ functions of $\boldsymbol{\beta}^{(i+1,j)}$ and $\boldsymbol{\nu}^{(i+1,j)}$ with vector $\boldsymbol{\varphi}^{(i+1,j)}$ given in solution $\boldsymbol{\Xi}^{(i+1,j)}$:

$$\chi_k^{(i+1,j)}\big(\beta_k^{(i+1,j)}\big) \tag{23a}$$
$$\triangleq \beta_k^{(i+1,j)}\Big(1+\exp\big(-a_k(\varphi_k^{(i+1,j)}-b_k)\big)\Big)-\eta_k, \quad \forall\, k\in\mathcal{K},$$
$$\chi_{K+k}^{(i+1,j)}\big(\nu_k^{(i+1,j)}\big) \tag{23b}$$
$$\triangleq \nu_k^{(i+1,j)}\Big(1+\exp\big(-a_k(\varphi_k^{(i+1,j)}-b_k)\big)\Big)-1, \quad \forall\, k\in\mathcal{K}.$$

We then define a $2K\times 1$ vector $\boldsymbol{\chi}^{(i+1,j)}\big(\boldsymbol{\beta}^{(i+1,j)},\boldsymbol{\nu}^{(i+1,j)}\big) \triangleq \big(\chi_1^{(i+1,j)}(\beta_1^{(i+1,j)}),\ldots,\chi_K^{(i+1,j)}(\beta_K^{(i+1,j)}),\chi_{K+1}^{(i+1,j)}(\nu_1^{(i+1,j)}),\ldots,\chi_{2K}^{(i+1,j)}(\nu_K^{(i+1,j)})\big)$. Now, we can use the damped Newton's method to update the parameter vectors $\boldsymbol{\beta}^{(i+1,j)}$ and $\boldsymbol{\nu}^{(i+1,j)}$ to reduce the ℓ_2-norm of $\boldsymbol{\chi}^{(i+1,j)}\big(\boldsymbol{\beta}^{(i+1,j)},\boldsymbol{\nu}^{(i+1,j)}\big)$. This is referred to as the *secondary subproblem* in iteration $(i+1,j)$. Problem $\mathrm{P}^{(i+1)}$ is solved when $\boldsymbol{\beta}^{(i+1,j)}$, $\boldsymbol{\nu}^{(i+1,j)}$, and $\boldsymbol{\varphi}^{(i+1,j)}$ satisfy $\boldsymbol{\chi}^{(i+1,j)}\big(\boldsymbol{\beta}^{(i+1,j)},\boldsymbol{\nu}^{(i+1,j)}\big)=\mathbf{0}_{2K}$. It should be noted that solving problem $\mathrm{P}^{(i+1)}$ does not lead to the solution of problem (18). We need to continue solve problems $\mathrm{P}^{(i+2)}$, $\mathrm{P}^{(i+3)}$, $\cdots$ in the outer loop. The proposed algorithm to tackle problem (18) is explained in detail in the following subsections.

D. Subproblems for the Inner Iterations

Without loss of generality, we present and solve the primary and secondary subproblems in the $(i+1,j)^{\text{th}}$ iteration.

1) Primary Subproblem: For given parameter vectors $\boldsymbol{\beta}^{(i+1,j)}$ and $\boldsymbol{\nu}^{(i+1,j)}$ in the $(i+1,j)^{\text{th}}$ iteration, the primary subproblem is given as follows:

$$\underset{\mathbf{w},\boldsymbol{\varphi},\boldsymbol{\vartheta},\boldsymbol{\varrho},\boldsymbol{\xi}}{\text{maximize}} \sum_{k\in\mathcal{K}}\nu_k^{(i+1,j)}\Big(\eta_k-\beta_k^{(i+1,j)}\Big(1+\exp\big(-a_k(\varphi_k-b_k)\big)\Big)\Big) \tag{24}$$

subject to constraints (7d), (18b) – (18e), and (20b).

We transform problem (24) into an equivalent rank-constrained SDP problem. We define Hermitian matrix $\mathbf{W}_k \triangleq \mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}$, $\forall\, k\in\mathcal{K}$, so problem (24) is equivalent to the following problem:

$$\underset{\mathbf{W}_{\mathcal{K}},\boldsymbol{\varphi},\boldsymbol{\vartheta},\boldsymbol{\varrho},\boldsymbol{\xi}}{\text{maximize}} \sum_{k\in\mathcal{K}}\nu_k^{(i+1,j)}\Big(\eta_k-\beta_k^{(i+1,j)}\Big(1+\exp\big(-a_k(\varphi_k-b_k)\big)\Big)\Big) \tag{25a}$$
$$\text{subject to} \sum_{k\in\mathcal{K}}\mathsf{Tr}(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{W}_k)\le p_m, \quad \forall\, m\in\mathcal{M}, \tag{25b}$$
$$\overline{\mathbf{S}}_{k,1}(\varphi_k,\vartheta_k)+\mathbf{Q}_k^{\mathsf{H}}\mathbf{W}_k\mathbf{Q}_k\succeq\mathbf{0}, \ \forall\, k\in\mathcal{K}, \tag{25c}$$
$$\overline{\mathbf{S}}_{k,2}(\varrho_k)+\mathbf{Q}_k^{\mathsf{H}}\mathbf{W}_k\mathbf{Q}_k\succeq\mathbf{0}, \quad \forall\, k\in\mathcal{K}, \tag{25d}$$
$$\overline{\mathbf{S}}_{k,3}(\xi_k)-\mathbf{Q}_k^{\mathsf{H}}\big(\textstyle\sum_{u\in\mathcal{K}\setminus\{k\}}\mathbf{W}_u\big)\mathbf{Q}_k\succeq\mathbf{0}, \forall\, k\in\mathcal{K}, \tag{25e}$$
$$\sum_{k\in\mathcal{K}} q_{m,k}^{(i)}R_k^{(i)}\mathsf{Tr}\big(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{W}_k\big)\le C_m, \forall\, m\in\mathcal{M}, \tag{25f}$$
$$\mathbf{W}_k\succeq\mathbf{0}, \forall\, k\in\mathcal{K}, \tag{25g}$$
$$\vartheta_k\ge 0, \varrho_k\ge 0, \xi_k\ge 0, \forall\, k\in\mathcal{K}, \tag{25h}$$
$$\mathsf{Rank}(\mathbf{W}_k)=1, \forall\, k\in\mathcal{K}, \tag{25i}$$

where optimization variable $\mathbf{W}_{\mathcal{K}}$ is defined as $\mathbf{W}_{\mathcal{K}} \triangleq \big\{\mathbf{W}_k \,|\, \mathbf{W}_k\in\mathbb{H}^{MN},\, k\in\mathcal{K}\big\}$. Problem (25) is still non-convex due to constraint (25i). To arrive at a tractable problem, we relax problem (25) by removing constraint (25i) and obtain the following problem in SDP form:

$$\underset{\mathbf{W}_{\mathcal{K}},\boldsymbol{\varphi},\boldsymbol{\vartheta},\boldsymbol{\varrho},\boldsymbol{\xi}}{\text{maximize}} \sum_{k\in\mathcal{K}}\nu_k^{(i+1,j)}\Big(\eta_k-\beta_k^{(i+1,j)}\Big(1+\exp\big(-a_k(\varphi_k-b_k)\big)\Big)\Big) \tag{26}$$

subject to constraints (25b) – (25h).

Problem (26) can be efficiently solved by convex programming solvers (e.g., CVX [34]) to obtain a numerical solution. In Fig. 2, we briefly summarize the transformation and relaxation steps that we have used from problem (6) to problem (26). In Fig. 2, a bidirectional arrow represents a transformation into an equivalent problem. A unidirectional arrow represents a transformation involving approximations.

Remark 2: The proposed problem transformations can also be applied if the objective function of problem (6) is replaced by a mixture of sigmoidal and concave functions. With such an objective function, we can, for example, jointly maximize the users' utility and the received data rate. Specifically, for each concave function added in the objective function of problem (6), e.g. $f(\varphi_k)=B\log_2(1+\varphi_k)$, we can introduce an auxiliary optimization variable x_k to substitute for $f(\varphi_k)$ in the objective function and include a constraint $f(\varphi_k)\ge x_k$ in the problem. Then, using the same steps as shown in Fig. 2, we can obtain an SDP problem similar to problem (26).

We denote the solution of problem (26) as $\big(\mathbf{W}_{\mathcal{K}}^{(i+1,j)}, \boldsymbol{\varphi}^{(i+1,j)}, \boldsymbol{\vartheta}^{(i+1,j)}, \boldsymbol{\varrho}^{(i+1,j)}, \boldsymbol{\xi}^{(i+1,j)}\big)$. If the Hermitian matrices in set $\mathbf{W}_{\mathcal{K}}^{(i+1,j)}$ are all rank-one matrices, then problems (25) and (26) have the same optimal solution and the same objective values. Otherwise, the optimal objective value of problem (26) is an upper bound for the objective value of problem (25), since problem (26) has a

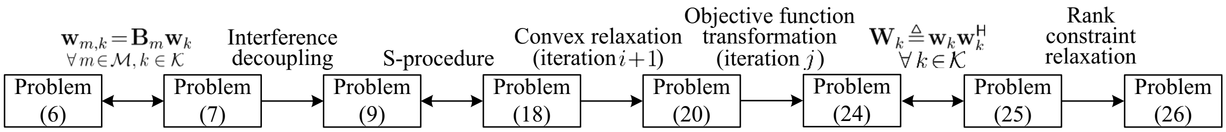


Fig. 2. The transformation and relaxation steps taken from problem (6) to problem (26).

larger feasible set. We now reveal the tightness of the SDP relaxation adopted in problem (26) in the following theorem:

Theorem 2: Assuming problem (26) is feasible, an optimal solution $\big(\overline{\mathbf{W}}_{\mathcal{K}}^{(i+1,j)}, \boldsymbol{\varphi}^{(i+1,j)}, \overline{\boldsymbol{\vartheta}}^{(i+1,j)}, \overline{\boldsymbol{\varrho}}^{(i+1,j)}, \overline{\boldsymbol{\xi}}^{(i+1,j)}\big)$ for problem (26), where $\overline{\mathbf{W}}_{\mathcal{K}}^{(i+1,j)} = \big\{\overline{\mathbf{W}}_k^{(i+1,j)} \,|\, \overline{\mathbf{W}}_k^{(i+1,j)} \in \mathbb{H}^{MN}, k \in \mathcal{K}\big\}$, can always be constructed such that $\mathsf{Rank}\big(\overline{\mathbf{W}}_k^{(i+1,j)}\big) = 1, \forall k \in \mathcal{K}$.

Proof: Please refer to Appendix B for a proof of Theorem 2. ■

That is, after solving problem (26), if the solution to problem (26) does not satisfy the rank-one constraint (25i), as outlined in the proof of Theorem 2, we can solve the SDP problem given in (32) in Appendix B to obtain the optimal beamforming matrices $\overline{\mathbf{W}}_{\mathcal{K}}^{(i+1,j)}$ for problem (25) that satisfy the rank-one constraint[3]. Eventually, the solution to problem (24) is given by $\boldsymbol{\Xi}^{(i+1,j)} = \big(\mathbf{w}^{(i+1,j)}, \boldsymbol{\varphi}^{(i+1,j)}, \overline{\boldsymbol{\vartheta}}^{(i+1,j)}, \overline{\boldsymbol{\varrho}}^{(i+1,j)}, \overline{\boldsymbol{\xi}}^{(i+1,j)}\big)$ where $\mathbf{w}_{[(k-1)MN+1\,:\,kMN]}^{(i+1,j)}$ is the principal eigenvector of matrix $\overline{\mathbf{W}}_k^{(i+1,j)} \in \overline{\mathbf{W}}_{\mathcal{K}}^{(i+1,j)}, \forall k \in \mathcal{K}$.

2) Secondary Subproblem: We now update parameter vectors $\boldsymbol{\beta}^{(i+1,j)}$ and $\boldsymbol{\nu}^{(i+1,j)}$ by using vector $\boldsymbol{\varphi}^{(i+1,j)}$ given by solution $\boldsymbol{\Xi}^{(i+1,j)}$. The updated parameter vectors, denoted by $\boldsymbol{\beta}^{(i+1,j+1)}$ and $\boldsymbol{\nu}^{(i+1,j+1)}$, will be used in the $(i+1,j+1)^{\text{th}}$ iteration. Recall the definition of vector $\boldsymbol{\chi}^{(i+1,j)}\big(\boldsymbol{\beta}^{(i+1,j)}, \boldsymbol{\nu}^{(i+1,j)}\big)$. By Theorem 1, if $\boldsymbol{\chi}^{(i+1,j)}\big(\boldsymbol{\beta}^{(i+1,j)}, \boldsymbol{\nu}^{(i+1,j)}\big) = \mathbf{0}_{2K}$, then $\boldsymbol{\beta}^{(i+1,j)}$ and $\boldsymbol{\nu}^{(i+1,j)}$ are the appropriate parameter vectors $\boldsymbol{\beta}^{(i+1)}$ and $\boldsymbol{\nu}^{(i+1)}$ employed in Theorem 1, respectively. Otherwise, i.e., if $\boldsymbol{\chi}^{(i+1,j)}\big(\boldsymbol{\beta}^{(i+1,j)}, \boldsymbol{\nu}^{(i+1,j)}\big) \neq \mathbf{0}_{2K}$, we update $\boldsymbol{\beta}^{(i+1,j)}$ and $\boldsymbol{\nu}^{(i+1,j)}$ by the damped Newton's method to determine a new pair of parameter vectors for the $(i+1,j+1)^{\text{th}}$ iteration. Specifically, let $\boldsymbol{\chi}'\big(\boldsymbol{\beta}^{(i+1,j)}, \boldsymbol{\nu}^{(i+1,j)}\big)$ denote the Jacobian matrix of $\boldsymbol{\chi}^{(i+1,j)}\big(\boldsymbol{\beta}^{(i+1,j)}, \boldsymbol{\nu}^{(i+1,j)}\big)$. We first introduce the following $2K\times 1$ vector:

$$\boldsymbol{\omega}^{(i+1,j)} \triangleq -\Big(\boldsymbol{\chi}'\big(\boldsymbol{\beta}^{(i+1,j)}, \boldsymbol{\nu}^{(i+1,j)}\big)\Big)^{-1} \boldsymbol{\chi}^{(i+1,j)}\big(\boldsymbol{\beta}^{(i+1,j)}, \boldsymbol{\nu}^{(i+1,j)}\big). \tag{27}$$

The first half and the second half of vector $\boldsymbol{\omega}^{(i+1,j)}$ (i.e., $\boldsymbol{\omega}_{[1:K]}^{(i+1,j)}$ and $\boldsymbol{\omega}_{[K+1:2K]}^{(i+1,j)}$) are the directions for updating parameter vectors $\boldsymbol{\beta}^{(i+1,j)}$ and $\boldsymbol{\nu}^{(i+1,j)}$, respectively. We then determine a proper updating step size $\zeta^{(i+1,j)}$ which is the largest value of t^{ℓ} that satisfies the following inequality:

$$\begin{aligned}&\big\|\boldsymbol{\chi}^{(i+1,j)}\big(\boldsymbol{\beta}^{(i+1,j)}+t^{\ell}\boldsymbol{\omega}_{[1:K]}^{(i+1,j)}, \boldsymbol{\nu}^{(i+1,j)}+t^{\ell}\boldsymbol{\omega}_{[K+1:2K]}^{(i+1,j)}\big)\big\|_2 \\ &\leq \big(1-\epsilon t^{\ell}\big)\big\|\boldsymbol{\chi}^{(i+1,j)}\big(\boldsymbol{\beta}^{(i+1,j)}, \boldsymbol{\nu}^{(i+1,j)}\big)\big\|_2,\end{aligned} \tag{28}$$

where $t, \epsilon \in (0,1)$ are predefined parameters, $\ell \in \{1,2,\ldots\}$. We update the parameter vectors as follows:

$$\boldsymbol{\beta}^{(i+1,j+1)} = \boldsymbol{\beta}^{(i+1,j)} + \zeta^{(i+1,j)}\boldsymbol{\omega}_{[1\,:\,K]}^{(i+1,j)}, \tag{29a}$$

$$\boldsymbol{\nu}^{(i+1,j+1)} = \boldsymbol{\nu}^{(i+1,j)} + \zeta^{(i+1,j)}\boldsymbol{\omega}_{[K+1\,:\,2K]}^{(i+1,j)}. \tag{29b}$$

The steps in (27)–(29) are repeated after problem (24) has been solved in the $(i+1,j+1)^{\text{th}}$ iteration by substituting $\boldsymbol{\beta}^{(i+1,j+1)}$ and $\boldsymbol{\nu}^{(i+1,j+1)}$ for $\boldsymbol{\beta}^{(i+1,j)}$ and $\boldsymbol{\nu}^{(i+1,j)}$, respectively, and so forth. It has been shown that the damped Newton's method converges and vectors $\boldsymbol{\beta}^{(i+1)}$, $\boldsymbol{\nu}^{(i+1)}$, and $\boldsymbol{\varphi}^{(i+1)}$ that satisfy the system of equations in (22) are obtained, cf. [36].

E. Outer Iterations and the Overall Algorithm

1) The Outer Iteration: In the outer iteration, we aim to create solution sparsity for $\mathbf{w}_k$ in problem (7). Based on the analysis that we provided after formulating problem $\mathrm{P}^{(i+1)}$ in Section III-B, we solve problem (7) in an iterative manner. Specifically, by iteratively solving the subproblems introduced in Sections III-D1 and III-D2, we obtain the solution $\boldsymbol{\Xi}^{(i+1)}$ for problem $\mathrm{P}^{(i+1)}$ in the $(i+1)^{\text{th}}$ outer iteration. We note that $\mathbf{w}_k^{(i+1)} = \mathbf{w}_{[(k-1)MN+1\,:\,kMN]}^{(i+1)}$ is the principal eigenvector of the optimal beamforming matrix $\overline{\mathbf{W}}_k^{(i+1,j)} \in \overline{\mathbf{W}}_{\mathcal{K}}^{(i+1,j)}, \forall k \in \mathcal{K}$. We then continue to solve problems $\mathrm{P}^{(i+2)}, \mathrm{P}^{(i+3)}, \cdots$ and obtain solutions $\boldsymbol{\Xi}^{(i+2)}, \boldsymbol{\Xi}^{(i+3)}, \cdots$, respectively. The outer iteration stops when either the solutions converge or the maximum number of iterations has been reached. We define $\Delta\mathbf{w}^{(i+1)} \triangleq \mathbf{w}^{(i+1)} - \mathbf{w}^{(i)}$ and $\Delta\boldsymbol{\varphi}^{(i+1)} \triangleq \boldsymbol{\varphi}^{(i+1)} - \boldsymbol{\varphi}^{(i)}$. The outer iteration stops if $\big\|\big[\Delta\mathbf{w}^{(i+1)\,\mathsf{H}} \;\; \Delta\boldsymbol{\varphi}^{(i+1)\,\mathsf{H}}\big]^{\mathsf{H}}\big\|_2 \leq \epsilon'$, where $\epsilon' > 0$ is a predefined small constant.

2) The Overall Algorithm: The proposed algorithm to solve problem (7) is Algorithm 1. We denote the maximum number of inner and outer iterations as $L_{\max}$ and $l_{\max}$, respectively. Let $\epsilon'' \ll 1$ denote the maximum tolerance for satisfying the system of equations in Theorem 1. The values of $L_{\max}$, $l_{\max}$, ϵ, ϵ', ϵ'', t, and τ as well as the maximum interference I, vectors $\mathbf{w}^{(0)}$, $\boldsymbol{\varphi}^{(0)}$, $\Delta\mathbf{w}^{(0)}$, $\Delta\boldsymbol{\varphi}^{(0)}$, and the iteration index i are initialized in Step 1. In each iteration of the outer loop, we determine $q_{m,k}^{(i)}, \forall m \in \mathcal{M}, k \in \mathcal{K}$ and $R_k^{(i)}, \forall k \in \mathcal{K}$ (Steps 3, 4). We then initialize $\nu_k^{(i+1,1)}$ and $\beta_k^{(i+1,1)}$ (Step 5). We solve problem $\mathrm{P}^{(i+1)}$ in an iterative manner in the inner iteration. In the $(i+1,j)^{\text{th}}$ iteration, we solve the relaxed SDP problem in (26) with parameter vectors $\boldsymbol{\beta}^{(i+1,j)}$ and $\boldsymbol{\nu}^{(i+1,j)}$ and obtain solution $\boldsymbol{\Xi}^{(i+1,j)}$ (Step 7). Thus, vector $\boldsymbol{\varphi}^{(i+1,j)}$ in $\boldsymbol{\Xi}^{(i+1,j)}$ is

[3]Since we have shown in Theorem 2 that a set of optimal matrices $\overline{\mathbf{W}}_{\mathcal{K}}^{(i+1,j)}$ with $\mathsf{Rank}\big(\overline{\mathbf{W}}_k^{(i+1,j)}\big) = 1, \forall k \in \mathcal{K}$, always exists, we can find these matrices also by a low-rank matrix completion based on Riemannian optimization [35]. To do so, for UE $k \in \mathcal{K}$, we first need to determine the appropriate subset of indices, Φ_k, from the complete set of indices $\{1, \ldots, MN\}^2$. We then need to tackle the low-rank matrix completion problem to determine $\overline{\mathbf{W}}_k^{(i+1,j)}$, where $\overline{\mathrm{W}}_k^{(i+1,j)}(r,c) = \mathrm{W}_k^{(i+1,j)}(r,c)$ if $(r,c) \in \Phi_k$, and $\overline{\mathrm{W}}_k^{(i+1,j)}(r,c) = 0$, otherwise. Here, $\mathrm{X}(r,c)$ denotes element (r,c) in matrix $\mathbf{X}$.

Algorithm 1: Algorithm to solve problem (7).

1 **Initialize** $L_{\max}$, $l_{\max}$, ϵ, ϵ', ϵ'', t, τ, I, $\mathbf{w}^{(0)}$, $\boldsymbol{\varphi}^{(0)}$, $\Delta\mathbf{w}^{(0)}$, $\Delta\boldsymbol{\varphi}^{(0)}$, $i := 0$.
//Outer Iteration
2 **while** $(i < l_{\max})$ and $\big\|\big[\Delta\mathbf{w}^{(i)\,\mathsf{H}}\ \Delta\boldsymbol{\varphi}^{(i)\,\mathsf{H}}\big]^{\mathsf{H}}\big\|_2 > \epsilon'$ **do**
3 Set $q_{m,k}^{(i)} := \big(\mathsf{Tr}(\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m\mathbf{w}_k^{(i)}\mathbf{w}_k^{(i)\,\mathsf{H}})+\tau\big)^{-1}$, $\forall\, m \in \mathcal{M}, k \in \mathcal{K}$.
4 Set $R_k^{(i)} := B\log_2\big(1+\varphi_k^{(i)}\big)$, $\forall\, k \in \mathcal{K}$.
5 Set $j := 1$, $\nu_k^{(i+1,j)} := \frac{1}{1+\exp\left(-a_k(\varphi_k^{(i)}-b_k)\right)}$, $\beta_k^{(i+1,j)} := \eta_k\nu_k^{(i+1,j)}$, $\forall\, k \in \mathcal{K}$.
//Inner Iteration
6 **while** $(j < L_{\max})$ **do**
7 Solve problem (26) with parameter vectors $\boldsymbol{\beta}^{(i+1,j)}$ and $\boldsymbol{\nu}^{(i+1,j)}$ by SDP to obtain solution $\boldsymbol{\Xi}^{(i+1,j)}$.
8 Determine vector $\boldsymbol{\chi}^{(i+1,j)}\big(\boldsymbol{\beta}^{(i+1,j)},\boldsymbol{\nu}^{(i+1,j)}\big)$ according to (23).
9 **if** $\big\|\boldsymbol{\chi}^{(i+1)}\big(\boldsymbol{\beta}^{(i+1,j)},\boldsymbol{\nu}^{(i+1,j)}\big)\big\|_2 \le \epsilon''$ **then**
10 **break**
11 **else**
12 Set $\boldsymbol{\beta}^{(i+1,j+1)} := \boldsymbol{\beta}^{(i+1,j)}+\zeta^{(i+1,j)}\boldsymbol{\omega}_{[1:K]}^{(i+1,j)}$, $\boldsymbol{\nu}^{(i+1,j+1)} := \boldsymbol{\nu}^{(i+1,j)}+\zeta^{(i+1,j)}\boldsymbol{\omega}_{[K+1:2K]}^{(i+1,j)}$.
13 Set $j := j+1$.
14 Set $\boldsymbol{\beta}^{(i+1)} := \boldsymbol{\beta}^{(i+1,j)}$, $\boldsymbol{\nu}^{(i+1)} := \boldsymbol{\nu}^{(i+1,j)}$, $\boldsymbol{\varphi}^{(i+1)} := \boldsymbol{\varphi}^{(i+1,j)}$.
15 Solve problem (32) for solution $\big(\overline{\mathbf{W}}_{\mathcal{K}}^{(i+1)},\boldsymbol{\varphi}^{(i+1)},\overline{\boldsymbol{\vartheta}}^{(i+1)},\overline{\boldsymbol{\varrho}}^{(i+1)},\overline{\boldsymbol{\xi}}^{(i+1)}\big)$ with $\mathsf{Rank}(\overline{\mathbf{W}}_k^{(i+1)})=1, \forall\, k\in\mathcal{K}$.
16 Find the optimal solution $\boldsymbol{\Xi}^{(i+1)}$ for problem $\mathrm{P}^{(i+1)}$, where $\mathbf{w}_{[(k-1)MN+1\,:\,kMN]}^{(i+1)}$ is given by the principal eigenvector of matrix $\overline{\mathbf{W}}_k^{(i+1)}\in\overline{\mathbf{W}}_{\mathcal{K}}^{(i+1)}$, $\forall\, k \in \mathcal{K}$.
17 Set $\Delta\mathbf{w}^{(i+1)} := \mathbf{w}^{(i+1)}-\mathbf{w}^{(i)}$, $\Delta\boldsymbol{\varphi}^{(i+1)} := \boldsymbol{\varphi}^{(i+1)}-\boldsymbol{\varphi}^{(i)}$, $i := i+1$.
18 Set $\mathbf{w}_k^{\star} := \mathbf{w}_{[(k-1)MN+1\,:\,kMN]}^{(i)}$, $\forall\, k \in \mathcal{K}$, $\mathbf{w}_{m,k}^{\star} := (\mathbf{w}_k^{\star})_{[(m-1)N+1\,:\,mN]}$, $\forall\, m \in \mathcal{M}$, $k \in \mathcal{K}$.
19 Employ beamforming vector $\mathbf{w}_{m,k}^{\star}$ for RRH m to serve UE k, $\forall\, m \in \mathcal{M}, k \in \mathcal{K}$.

found. We then determine vector $\boldsymbol{\chi}^{(i+1,j)}\big(\boldsymbol{\beta}^{(i+1,j)},\boldsymbol{\nu}^{(i+1,j)}\big)$ (Step 8). If the ℓ_2-norm of vector $\boldsymbol{\chi}^{(i+1,j)}\big(\boldsymbol{\beta}^{(i+1,j)},\boldsymbol{\nu}^{(i+1,j)}\big)$ is smaller than threshold ϵ'' (Step 9), the appropriate parameter vectors $\boldsymbol{\beta}^{(i+1)}$ and $\boldsymbol{\nu}^{(i+1)}$ employed in Theorem 1 have been found. We thus break the inner loop (Step 10). Otherwise, we update $\boldsymbol{\beta}^{(i+1,j)}$ and $\boldsymbol{\nu}^{(i+1,j)}$ to obtain parameter vectors $\boldsymbol{\beta}^{(i+1,j+1)}$ and $\boldsymbol{\nu}^{(i+1,j+1)}$ for the next inner iteration (Step 12). When the inner loop stops, we set $\boldsymbol{\beta}^{(i+1,j)}$, $\boldsymbol{\nu}^{(i+1,j)}$, and $\boldsymbol{\varphi}^{(i+1,j)}$ to $\boldsymbol{\beta}^{(i+1)}$, $\boldsymbol{\nu}^{(i+1)}$, and $\boldsymbol{\varphi}^{(i+1)}$, respectively (Step 14). We then construct the optimal solution $\big(\overline{\mathbf{W}}_{\mathcal{K}}^{(i+1,j)},\boldsymbol{\varphi}^{(i+1,j)},\overline{\boldsymbol{\vartheta}}^{(i+1,j)},\overline{\boldsymbol{\varrho}}^{(i+1,j)},\overline{\boldsymbol{\xi}}^{(i+1,j)}\big)$ that satisfies the rank-one constraint (Step 15) and obtain optimal solution $\boldsymbol{\Xi}^{(i+1)}$ for problem $\mathrm{P}^{(i+1)}$, where $\mathbf{w}_{[(k-1)MN+1\,:\,kMN]}^{(i+1)}$ is the principal eigenvector of matrix $\overline{\mathbf{W}}_k^{(i+1,j)}\in\overline{\mathbf{W}}_{\mathcal{K}}^{(i+1,j)}$, $\forall\, k\in\mathcal{K}$ (Step 16). At the end of each iteration of the outer loop, we update vectors $\Delta\mathbf{w}^{(i+1)}$, $\Delta\boldsymbol{\varphi}^{(i+1)}$, and i (Step 17). The outer loop stops when either $i=l_{\max}$ or $\big\|\big[\Delta\mathbf{w}^{(i)\,\mathsf{H}}\ \Delta\boldsymbol{\varphi}^{(i)\,\mathsf{H}}\big]^{\mathsf{H}}\big\|_2 \le \epsilon'$. When the outer loop stops, we have beamforming vector $\mathbf{w}^{(i)}$ outside the outer loop. We recover beamforming vector $\mathbf{w}_{m,k}^{\star}$ and employ it for RRH $m\in\mathcal{M}$ to serve UE $k\in\mathcal{K}$ (Steps 18, 19). We note that the output of Algorithm 1 may depend on the initial vector $\mathbf{w}^{(0)}$. In the next section, we will show the performance of Algorithm 1 by using the initial vector $\mathbf{w}^{(0)}$ that maximizes the WSSR.

For the complexity of Algorithm 1, we find that solving the SDP problem in Step 7 in the inner loop is the computationally most intensive task. The SDP problem with optimization variables $\mathbf{W}_k\in\mathbb{H}^{MN}$, $k\in\mathcal{K}$ can be solved with the interior-point method with a computational complexity of $\mathcal{O}\big(M^6N^6\big)$ [37]. Thus, the complexity of each iteration of the outer loop is $\mathcal{O}\big((L_{max}+1)M^6N^6\big)$. Therefore, the overall complexity of Algorithm 1 is $\mathcal{O}\big((L_{max}l_{max}+l_{max})M^6N^6\big)$. We note that the $(i+1)^{\mathrm{th}}/j^{\mathrm{th}}$ iteration of the outer/inner loop of Algorithm 1 takes the output of the previous iteration as the input. Thus, parallelization of Algorithm 1 does not seem possible.

TABLE I
SIMULATION PARAMETERS

Square wireless service area	$25\times10^4\,\mathrm{m}^2$
Reference distance	15 m
User and RRH distribution	Uniformly distributed in a square area
Path loss exponent	3.8
Fading distribution	Rayleigh fading
Bandwidth B	20 MHz
Number of antennas per RRH N	2
$L_{\max}$ and $l_{\max}$ in Algorithm 1	30 and 30
ϵ, ϵ', ϵ'', τ, and t in Algorithm 1	0.1, 0.01, 0.01, 1×10^{-5}, and 0.97
η_k, $\forall\, k \in \mathcal{K}$	Uniformly distributed in $[1, 10]$
p_m, $\forall\, m \in \mathcal{M}$	400 mW

IV. PERFORMANCE EVALUATION

A. Simulation Parameters and Initial Beamforming Vectors

We assume that several RRHs and a number of mobile users are located in a square wireless service area. Unless otherwise stated, the simulation parameters shown in Table I are adopted. In particular, we assume that each mobile user experiences the same noise power given by $\sigma_k^2 = -174\,\mathrm{dBm}+10\log_{10}(20\,\mathrm{MHz}) = -101\,\mathrm{dBm} = \sigma^2$, $\forall\, k \in \mathcal{K}$. To determine parameters a_k and b_k, we assume that the value of the SINR at which a mobile user achieves $\frac{1}{1+\exp(-5)}\times 100\% = 99.33\%$ of its maximum achievable utility (i.e., η_k) is uniformly distributed in $[10, 100]$ (i.e., $[10\,\mathrm{dB}, 20\,\mathrm{dB}]$). In other words, we assume that the mobile users use different applications and the SINRs required by these applications to achieve 99.33% of their maximum achievable utility is uniformly distributed in $[10, 100]$. This assumption is based on the fact that the quality of wireless services for playing online videos and video streaming is considered to be excellent in the Long Term Evolution (LTE) network if the received SINR is greater than or equal to 20 dB, which is expected to be the case when a user is close to the base station [38]. In this case, typical online videos can be played smoothly and users using video streaming applications can achieve their maximum possible utility. On the other hand, the quality of wireless service is regarded as good if the received SINR is greater than 12 dB but less than 20 dB [38]. Within this range, web pages can be smoothly downloaded by UEs, so users using web browsing applications can achieve their maximum possible utility. Since different

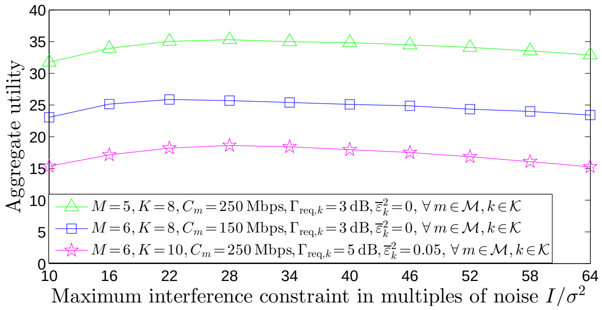


Fig. 3. Aggregate utility for different system parameters versus the normalized maximum interference constraint I/σ^2.

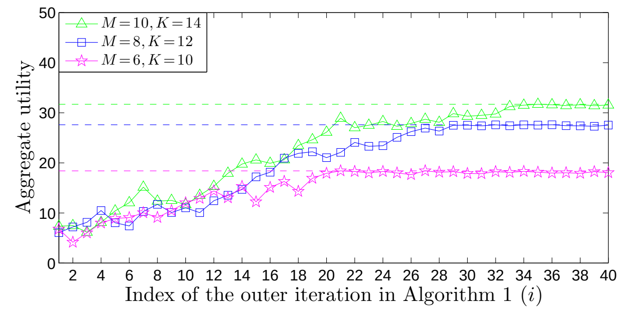


Fig. 4. Convergence of Algorithm 1 for different values of M and K. $C_m = 150\,\mathrm{Mbps}, \Gamma_{\mathrm{req},k} = 3\,\mathrm{dB}, \bar{\varepsilon}_k^2 = 0.05, \forall m \in \mathcal{M}, k \in \mathcal{K}$.

applications (e.g. video streaming, web browsing) are used by mobile users, it is reasonable to assume that $[10\,\mathrm{dB}, 20\,\mathrm{dB}]$ is the range of the SINR at which a given user achieves 99.33% of his maximum possible utility. We thus set a_k via a random variable that follows an inverse uniform distribution on $[0.1, 1]$ and set $b_k = \frac{5}{a_k}$, $\forall\, k \in \mathcal{K}$. Besides, vector $\boldsymbol{\varphi}^{(0)}$ as well as vector $\mathbf{w}^{(0)}$ that comprises K initial beamforming vectors (i.e., $\mathbf{w}_k^{(0)}$, $\forall\, k \in \mathcal{K}$) are determined as follows. We first solve the optimization problem after replacing the objective function in problem (25) by $\sum_{k\in\mathcal{K}} \eta_k B \log_2(1+\varphi_k)$ and retaining constraints (25b) – (25e), (25g), and (25h). That is, we first determine the optimal beamforming matrices that maximize the WSSR without considering the backhaul capacity and the rank-one constraint. We denote the corresponding solution as $\big(\mathbf{W}_{\mathcal{K}}^{(0)}, \boldsymbol{\varphi}^{(0)}, \boldsymbol{\vartheta}^{(0)}, \boldsymbol{\varrho}^{(0)}, \boldsymbol{\xi}^{(0)}\big)$. Thus, the initial vector $\boldsymbol{\varphi}^{(0)}$ is determined. We then apply the approach used in Theorem 2 to determine another set of rank-one matrices, denoted by $\overline{\mathbf{W}}_{\mathcal{K}}^{(0)} = \{\overline{\mathbf{W}}_k^{(0)}, k \in \mathcal{K}\}$, which achieves the same WSSR. The beamforming vector $\mathbf{w}_k^{(0)}$ is initialized by the principal eigenvector of beamforming matrix $\overline{\mathbf{W}}_k^{(0)}$, $\forall\, k \in \mathcal{K}$. By doing so, we not only can check the feasibility of our original problem for the given $(\Gamma_{\mathrm{req},1}, \ldots, \Gamma_{\mathrm{req},K})$ and $(p_1, \ldots, p_M)$, but can also obtain the appropriate $q_{m,k}^{(0)}$, $\forall\, m \in \mathcal{M}, k \in \mathcal{K}$, and $R_k^{(0)}$, $\forall\, k \in \mathcal{K}$, required in the first iteration of the outer loop of our proposed Algorithm 1. To simulate the imperfectness of the CSI estimation, we introduce the normalized maximum channel estimation error $\bar{\varepsilon}_k^2 \triangleq \varepsilon_k^2 / \|\widehat{\mathbf{h}}_k\|_2^2, \forall\, k \in \mathcal{K}$. The CSI uncertainty region of user k is $\Omega_k = \{\Delta\mathbf{h}_k : \Delta\mathbf{h}_k^{\mathsf{H}} \Delta\mathbf{h}_k \leq \bar{\varepsilon}_k^2 \|\widehat{\mathbf{h}}_k\|_2^2\}$. Our results are obtained by averaging the performance of multiple path loss and small scale fading realizations.

B. Simulation Results

In Fig. 3, we show the impact of the maximum interference constraint, I, in (8) on the aggregate utility. We conducted simulations for network scenarios with different system parameters, such as different numbers of users, backhaul capacities, minimum SINRs required by mobile users, and normalized maximum channel estimation errors. For each network scenario, we plotted the aggregate utility versus the maximum interference constraint given in multiples of noise power σ^2. That is, I was changed by tuning the ratio I/σ^2. We find that the aggregate utility first increases and then slightly decreases as I/σ^2 increases. Thus, a suitable I can be obtained by running offline simulations. Even though setting I too high (i.e., $I \geq 34\sigma^2$) or too low (i.e., $I \leq 16\sigma^2$) may degrade the aggregate utility, the aggregate utility is not very sensitive to the choice of I/σ^2 in the considered interval. We use $I = 25\sigma^2$ for the following simulations.

In Fig. 4, we illustrate the convergence of the proposed algorithm for different numbers of RRHs (i.e., M) and users (i.e., K). We set $C_m = 150$Mbps, $\forall\, m \in \mathcal{M}$, and $\Gamma_{\mathrm{req},k} = 3\,\mathrm{dB}$ and $\bar{\varepsilon}_k^2 = 0.05$, $\forall\, k \in \mathcal{K}$. The maximum iterations $L_{\max}$ and $l_{\max}$ for the inner and outer loops in Algorithm 1 are set to 40, respectively. The aggregate utility obtained by Algorithm 1 after each iteration i of the outer loop is shown in Fig. 4. We observe that depending on the system parameters the aggregate utility has converged to the maximum value after about 20 to 33 iterations. In particular, the larger the numbers of RRHs and users in the network, the larger the number of iterations are required for convergence.

In Fig. 5, we present the performance of the proposed resource allocation algorithm as a function of the number of users in the network for six RRHs. For comparison, we also solve problem (6) without backhaul constraint (6c). The resulting problem has a larger feasible set compared to problem (6). The corresponding optimal value constitutes an "upper bound" for the proposed suboptimal solution of problem (6). We also evaluate the aggregate utility of three baseline schemes. In baseline schemes I and II, we maximize the WSSR and the weighed sum of SINRs by solving problem (6) for objective functions $\sum_{k\in\mathcal{K}} \eta_k \log_2(1+\gamma_k)$ and $\sum_{k\in\mathcal{K}} \eta_k \gamma_k$, respectively. Both problems can be solved with similar approaches as that developed in this paper. We then use the beamforming vectors obtained for the two baseline schemes to determine the corresponding aggregate utility for the sigmoidal utility function. In baseline scheme III, we adopt the compression strategy proposed in [21] to solve the WSSR maximization problem with backhaul constraints, where similar to [21], the CSI imperfection is handled by the S-procedure [31]. We then employ the users' received SINRs to determine their aggregate utility. From Fig. 5, it can be observed that the aggregate utility increases with the number of users. When the number of users is increased from 4 to 6, the aggregate utility of Algorithm 1 increases almost linearly. The reason for this behaviour is that if only few users are in the system, the degrees of freedom in the network are sufficient and all users can be properly served. However, when the number of users gets large, the co-channel

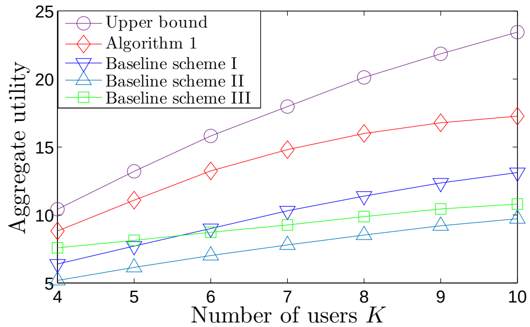


Fig. 5. Aggregate utility versus the number of users. $M = 6$, $C_m = 150\,\text{Mbps}$, $\forall\, m \in \mathcal{M}$, $\bar{\varepsilon}_k^2 = 0.05, \Gamma_{\text{req},k} = 3\,\text{dB}$, $\forall\, k \in \mathcal{K}$.

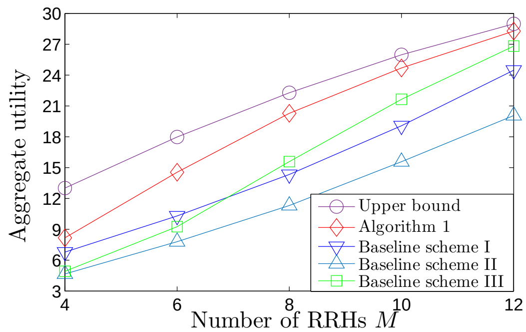


Fig. 6. Aggregate utility versus the number of RRHs. $K = 7$, $\bar{\varepsilon}_k^2 = 0.05, \Gamma_{\text{req},k} = 3\,\text{dB}$, $\forall\, k \in \mathcal{K}$, $C_m = 150\,\text{Mbps}$, $\forall\, m \in \mathcal{M}$.

interference as well as the limited backhaul capacity hinder the RRHs in steering the information beams accurately. Thus, the aggregate utility grows sublinearly. The performance gap between Algorithm 1 and the upper bound increases with the number of users, since the upper bound neglects the limited capacity of the backhaul links. Compared with the baseline schemes, we observe that Algorithm 1 achieves larger gains in the aggregate utility as the number of users grows. This is because the baseline schemes may cause mismatches in resource allocation due to the nonlinear relationship between the WSSR and the considered aggregate utility given by the weighted sum of sigmoidal functions. In particular, for the baseline schemes, an exceedingly large amount of system resources are allocated to a small set of users causing saturation in the sigmoidal functions, which limits the achievable aggregate utility. Moreover, a comparison of baseline schemes I and II shows that maximizing the WSSR results in a higher aggregate utility than maximizing the weighted sum of SINRs since the concavity of the logarithmic utility function in WSSR maximization alleviates the resource allocation mismatch to a certain extent. We also observe that for large numbers of users, the data sharing strategy in baseline scheme II outperforms the compression strategy in baseline scheme III. This is because when there are more users are in the C-RAN, a coarser compression is needed introducing more quantization noise to accommodate all users with the capacity-limited backhaul. The received SINRs of the users in baseline scheme III are thus limited by the quantization noise.

In Fig. 6, we show the aggregate utility as a function of the numbers of RRHs for 7 mobile users. We find that the aggregate utility increases with the number of RRHs for both the proposed algorithm and the baseline schemes. Although the upper bound also increases with the number of RRHs, the gap between the upper bound and Algorithm 1 shrinks as the number of RRHs increases. This is because the number of backhaul links increases with the number of RRHs and the C-RAN can support higher data rates for the mobile users. Thus, the negative impact of the limited capacity of each backhaul link on the aggregate utility is alleviated when the number of RRHs is large. Furthermore, for $M = 4$, the difference between the aggregate utility achieved by Algorithm 1 and the baseline schemes is relatively small. This is because when $M = 4$, the data traffic in the network is limited by the capacity-limited backhaul in all cases. This is especially true for baseline scheme III, as its performance is limited by the large quantization noise introduced by the compression strategy to accommodate all users in the network. Besides, the limited number of antennas also restricts the available degrees of freedom for accurately steering the information beams towards the mobile users while satisfying their SINR requirements. However, as the number of RRHs increases, the gap between Algorithm 1 and the baseline schemes widens. This is because Algorithm 1 can make better use of the additional antennas and backhaul links than the baseline schemes. The performance of baseline scheme III improves fast as the number of RRHs increases. This is due to the increased degrees of freedom. Specifically, the quantization noises introduced at different RRHs for a UE are different. An RRH which has a good CSI estimate for a UE, introduces little quantization noise and employs high transmission power for the UE to increase its SINR. On the other hand, an RRH which is not in a good position to serve the UE, introduces large quantization noise for the UE to satisfy the capacity-limited backhaul constraint and decreases its transmission power for the UE. Thus, baseline scheme III achieves good performance as the number of RRHs increases.

In Fig. 7, we investigate the impact of CSI uncertainty on the aggregate utility. We not only compare Algorithm 1 with the baseline schemes, but also with a non-robust beamforming design for aggregate sigmoidal utility maximization. Specifically, the non-robust beamforming design treats the estimated CSI as perfect information for resource allocation. Then, we optimize $\mathbf{w}$ and $\boldsymbol{\varphi}$ for the maximization problem in (6). In other words, robustness against CSI errors is not provided by this scheme. If the resulting resource allocation does not satisfy the constraints for user k in (6) due to the channel estimation errors, the system is in outage and the utility of user k is set to zero for that channel realization to account for the penalty of violating the constraints. We observe that the aggregate utility severely decreases when the normalized maximum channel estimation error increases. This is because when the channel estimation error increases, it is more difficult for the RRHs to accurately steer the information beams towards the desired users. Besides, the RRHs become less capable of mitigating the multiuser interference. We further observe from Fig. 7 that compared to the baseline scheme, our proposed resource allocation algorithm can significantly increase the aggregate utility in C-RAN. Moreover, the proposed scheme

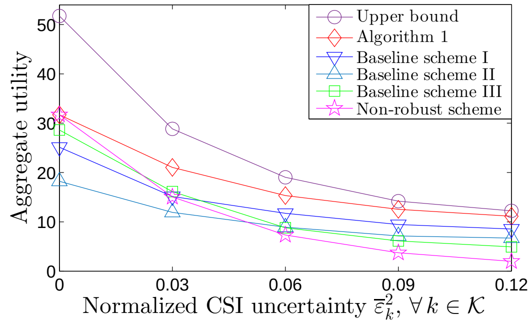


Fig. 7. Aggregate utility versus normalized maximum channel estimation error. $M=6, K=10, \Gamma_{\text{req},k}=3\,\text{dB}, C_m=150\,\text{Mbps}, \forall\, k\in\mathcal{K}, m\in\mathcal{M}$.

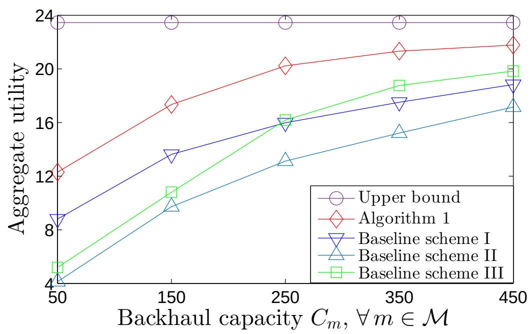


Fig. 8. Aggregate utility versus the backhaul capacity of RRHs in C-RAN. $M=6$, $K=10$, $\Gamma_{\text{req},k}=3\,\text{dB}$, $\overline{\varepsilon}_k^2=0.05$, $\forall\, k\in\mathcal{K}$.

achieves a significantly higher aggregate utility compared to the non-robust beamforming design, especially for large maximum channel estimation errors. In fact, the non-robust resource allocation scheme assumes that the available CSI is perfect and causes saturation in the utility function of some users and under-utilization of other users. Besides, the upper bound on the aggregate utility decreases dramatically when the channel uncertainty increases. Although Algorithm 1 performs better than the baseline schemes, the achieved gain in terms of the aggregate utility decreases as the channel uncertainty increases. This is because the aggregate utilities shown in Fig. 7 are determined based on the SINR lower bounds taking into account the channel uncertainty. When the normalized CSI uncertainty increases, the achievable aggregate utility and also the possible aggregate utility gain obtained by Algorithm 1 are reduced. For baseline scheme III using the compression strategy, the aggregate utility severely decreases as the maximum channel estimation error increases. This is because the robust beamforming design has to accommodate the worst case after the quantization noise is transmitted over uncertain wireless channels. The lower bound of the user received SINR in the compression strategy is thus significantly reduced for the compression strategy.

In Fig. 8, we show the aggregate utility as a function of the backhaul capacity. The performance upper bound is thus shown as a horizontal line. We observe that the aggregate utility increases with the backhaul capacity of the RRHs. This is because a higher backhaul capacity can support higher data rates and the SINR received by each user also increases due to the beamforming enabled by Algorithm 1. In other words, as the backhaul capacity increases, the spatial degrees of freedom offered by multiple RRHs can be more efficiently utilized to increase the aggregate utility of the C-RAN. Compared with the baseline schemes, Algorithm 1 can significantly improve the aggregate utility. This is because of the resource allocation mismatch due to the nonlinear relationship between the WSSR/weighted sum of SINRs and the aggregate utility. Moreover, the aggregate utility of baseline scheme III, where the compression strategy is adopted, improves significantly when the backhaul capacity increases. This is because when the backhaul capacity is increased, less compression is needed and thus less quantization noise is introduced.

V. Conclusions

In this paper, we studied utility-based cooperative beamforming design in C-RAN. We used a weighted sum of sigmoidal functions to model the aggregate utility and formulated the beamforming design as a non-convex optimization problem for the maximization of the aggregate utility. Our problem formulation took into account both the imperfect CSI and the capacity-limited backhaul. Due to the complexity of the problem, we introduced maximum interference constraints to simplify the optimization problem. Subsequently, an efficient iterative algorithm was proposed to obtain a good suboptimal solution. In each iteration, we tackled a non-convex optimization problem with infinitely many constraints. By exploiting the sum-of-ratios form of the objective function, we transformed the non-convex optimization problem into an equivalent rank-constrained SDP problem which could be solved optimally. Simulation results unveiled that the aggregate utility can be significantly improved by our proposed resource allocation algorithm compared to baseline schemes where either the WSSR or the weighted sum of SINRs is maximized. Furthermore, for large CSI uncertainty and a large number RRHs compared to the number of users, the proposed suboptimal resource allocation algorithm approached the performance upper bound determined in the backhaul capacity unconstrained case.

Appendix A: Proof of Theorem 1

We present a constructive proof. We first introduce the following optimization problem:

$$\underset{\boldsymbol{\beta},\mathbf{w},\,\boldsymbol{\varphi},\,\boldsymbol{\vartheta},\,\boldsymbol{\varrho},\,\boldsymbol{\xi}}{\text{maximize}} \; \sum_{k\in\mathcal{K}} \beta_k \tag{30a}$$

$$\text{subject to } \eta_k \geq \beta_k\big(1+\exp\left(-a_k\varphi_k + a_k b_k\right)\big), \quad \forall\, k\in\mathcal{K}, \tag{30b}$$

$$\text{constraints (7d), (18b)} - \text{(18e), and (20b)},$$

where $\boldsymbol{\beta} \triangleq (\beta_1, \ldots, \beta_K)$, β_k is the auxiliary optimization variable for the utility of user $k \in \mathcal{K}$, and constraint (30b) is obtained from the definition of $g_k(\varphi_k)$, $\forall\, k\in\mathcal{K}$. Problem (30) is equivalent to problem $\text{P}^{(i+1)}$ in the sense that if $\boldsymbol{\Xi}^{(i+1)}$ is the solution of problem $\text{P}^{(i+1)}$, the solution of problem (30) is $\big(\boldsymbol{\beta}^{(i+1)}, \boldsymbol{\Xi}^{(i+1)}\big)$, where $\beta_k^{(i+1)} = g_k(\varphi_k^{(i+1)})$. The Lagrangian of problem (30) is $\mathcal{L}_{\text{theo1}}^{(i+1)}\big(\boldsymbol{\beta},\, \boldsymbol{\Xi},\, \boldsymbol{\nu},\, \boldsymbol{\Psi}\big) \triangleq \sum_{k\in\mathcal{K}} \beta_k + \sum_{i\in\mathcal{K}} \nu_i \left(\eta_k - \beta_k\big(1+\exp\left(-a_k\varphi_k + a_k b_k\right)\big)\right) +$

$\boldsymbol{\Theta}$, where $\boldsymbol{\nu} \triangleq (\nu_1, \ldots, \nu_K)$ ($\boldsymbol{\nu} \succeq \mathbf{0}$) comprises the Lagrangian multipliers for constraint (30b), $\boldsymbol{\Xi}$ collects all optimization variables of problem (30) except vector $\boldsymbol{\beta}$, $\boldsymbol{\Psi}$ contains the Lagrangian multipliers of all constraints in (30) except constraint (30b), and $\boldsymbol{\Theta}$ denotes the sum of all terms which are not related to vectors $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$. Given solution $\big(\boldsymbol{\beta}^{(i+1)}, \boldsymbol{\Xi}^{(i+1)}\big)$ of problem (30), the following Karush-Kuhn-Tucker (KKT) conditions are obtained for $\boldsymbol{\nu}^{(i+1)}$ and $\boldsymbol{\beta}^{(i+1)}$:

$$\eta_k - \beta_k^{(i+1)}\big(1 + \exp\left(-a_k\varphi_k^{(i+1)} + a_k b_k\right)\big) = 0, \forall\, k \in \mathcal{K}, \tag{31a}$$

$$1 - \nu_k^{(i+1)}\big(1 + \exp\left(-a_k\varphi_k^{(i+1)} + a_k b_k\right)\big) = 0, \quad \forall\, k \in \mathcal{K}, \tag{31b}$$

where $\boldsymbol{\nu}_k^{(i+1)} \triangleq \big(\nu_k^{(i+1)}, \ldots, \nu_k^{(i+1)}\big)$ is obtained from the solution of the dual problem of (30). On the other hand, given $\boldsymbol{\beta}^{(i+1)}$ and $\boldsymbol{\nu}^{(i+1)}$, the Lagrangian of problem (21) is $\widehat{\mathcal{L}}_{\text{theo1}}^{(i+1)}\big(\widehat{\boldsymbol{\Xi}}, \widehat{\boldsymbol{\Psi}}\big) \triangleq \sum_{i\in\mathcal{K}} \nu_i^{(i+1)}\Big(\eta_k - \beta_k^{(i+1)}\big(1 + \exp\left(-a_k\varphi_k + a_k b_k\right)\big)\Big) + \widehat{\boldsymbol{\Theta}}$, where $\widehat{\boldsymbol{\Xi}}$ collects the optimization variables in (21), $\widehat{\boldsymbol{\Psi}}$ contains the Lagrangian multipliers for the constraints in (21), and $\widehat{\boldsymbol{\Theta}}$ denotes the sum of all terms related to $\widehat{\boldsymbol{\Xi}}$ and $\widehat{\boldsymbol{\Psi}}$. It is easy to see that $\widehat{\boldsymbol{\Xi}} = \boldsymbol{\Xi}$, $\widehat{\boldsymbol{\Psi}} = \boldsymbol{\Psi}$, and $\widehat{\boldsymbol{\Theta}} = \boldsymbol{\Theta}$. Thus, the KKT conditions for $\boldsymbol{\Xi}^{(i+1)}$ and $\boldsymbol{\Psi}^{(i+1)}$ for the solutions of the primary and dual problems of problem (30), respectively, are exactly the KKT conditions for problem (21). Since $\big(\boldsymbol{\beta}^{(i+1)}, \boldsymbol{\Xi}^{(i+1)}\big)$ is the solution to problem (30) which is a convex optimization problem for given $\boldsymbol{\beta}^{(i+1)}$ and $\boldsymbol{\nu}^{(i+1)}$, the KKT conditions of problem (21) are sufficient for the optimality of $\widehat{\boldsymbol{\Xi}}^{(i+1)}$ in problem (21). Thus, if $\boldsymbol{\Xi}^{(i+1)}$ is the solution to $\mathrm{P}^{(i+1)}$, there exist two vectors $\boldsymbol{\beta}^{(i+1)} = \big(\beta_1^{(i+1)}, \ldots, \beta_K^{(i+1)}\big)$ and $\boldsymbol{\nu}^{(i+1)} = \big(\nu_1^{(i+1)}, \ldots, \nu_K^{(i+1)}\big)$ such that $\boldsymbol{\Xi}^{(i+1)}$ is also an optimal solution of problem (21). Moreover, vector $\boldsymbol{\varphi}^{(i+1)}$ in $\boldsymbol{\Xi}^{(i+1)}$ satisfies the system of equations given in (31), which is the same as the system of equations in (22). This completes the proof. ∎

Appendix B: Proof of Theorem 2

For the proof, we follow a similar approach as [39]. If the optimal solution $\big(\mathbf{W}_{\mathcal{K}}^{(i+1,j)}, \boldsymbol{\varphi}^{(i+1,j)}, \boldsymbol{\vartheta}^{(i+1,j)}, \boldsymbol{\varrho}^{(i+1,j)}, \boldsymbol{\xi}^{(i+1,j)}\big)$ of problem (26) is obtained and $\mathsf{Rank}\big(\mathbf{W}_k^{(i+1,j)}\big) > 1, \exists\, k \in \mathcal{K}$, we can construct another optimal solution that comprises rank-one matrices as follows. For a given $\boldsymbol{\varphi}^{(i+1,j)}$ obtained from the solution of problem (26), we solve the following problem:

$$\underset{\mathbf{W}_{\mathcal{K}},\, \boldsymbol{\vartheta},\, \boldsymbol{\varrho},\, \boldsymbol{\xi}}{\text{minimize}} \ \textstyle\sum_{k\in\mathcal{K}} \mathsf{Tr}\big(\mathbf{W}_k\big) \tag{32a}$$

$$\text{subject to } \overline{\mathbf{S}}_{k,1}(\varphi_k^{(i+1,j)}, \vartheta_k) + \mathbf{Q}_k^{\mathsf{H}}\mathbf{W}_k\mathbf{Q}_k \succeq \mathbf{0}, \quad \forall\, k \in \mathcal{K}, \tag{32b}$$

$$\text{constraints (25b), (25d) – (25h).}$$

Problem (32) is in SDP form. Let $\big(\overline{\mathbf{W}}_{\mathcal{K}}^{(i+1,j)}, \overline{\boldsymbol{\vartheta}}^{(i+1,j)}, \overline{\boldsymbol{\varrho}}^{(i+1,j)}, \overline{\boldsymbol{\xi}}^{(i+1,j)}\big)$, where $\overline{\mathbf{W}}_{\mathcal{K}}^{(i+1,j)} \triangleq \big\{\overline{\mathbf{W}}_k^{(i+1,j)} \mid k \in \mathcal{K}\big\}$, denote the optimal solution of problem (32). It is easy to show that $\big(\overline{\mathbf{W}}_{\mathcal{K}}^{(i+1,j)}, \boldsymbol{\varphi}^{(i+1,j)}, \overline{\boldsymbol{\vartheta}}^{(i+1,j)}, \overline{\boldsymbol{\varrho}}^{(i+1,j)}, \overline{\boldsymbol{\xi}}^{(i+1,j)}\big)$ satisfies the constraints in problem (26) and yields the same objective value as solution $\big(\mathbf{W}_{\mathcal{K}}^{(i+1,j)}, \boldsymbol{\varphi}^{(i+1,j)}, \boldsymbol{\vartheta}^{(i+1,j)}, \boldsymbol{\varrho}^{(i+1,j)}, \boldsymbol{\xi}^{(i+1,j)}\big)$ for problem (26). We now show $\mathsf{Rank}\big(\overline{\mathbf{W}}_k^{(i+1,j)}\big) = 1, \forall\, k \in \mathcal{K}$. To this end, the Lagrangian of problem (32) is given as follows:

$$\begin{aligned}
&\mathcal{L}_{\text{theo2}}^{(i+1,j)}\big(\mathbf{W}_{\mathcal{K}}, \boldsymbol{\vartheta}, \boldsymbol{\varrho}, \boldsymbol{\xi}, \boldsymbol{\Lambda}\big) \\
&= \sum_{k\in\mathcal{K}} \mathsf{Tr}\Big(\mathbf{W}_k\big(\mathbf{I}_{MN} + \sum_{m\in\mathcal{M}}(\lambda_{m,1} + R_k^{(i)}\lambda_{m,2}q_{m,k}^{(i)})\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m \\
&\qquad - \mathbf{Q}_k(\mathbf{L}_{k,1} + \mathbf{L}_{k,2})\mathbf{Q}_k^{\mathsf{H}} + \sum_{u\in\mathcal{K}\setminus\{k\}}\mathbf{Q}_u\mathbf{L}_{u,3}\mathbf{Q}_u^{\mathsf{H}} - \mathbf{V}_k\big)\Big) \\
&\quad - \sum_{k\in\mathcal{K}} \mathsf{Tr}\big(\overline{\mathbf{S}}_{k,1}(\varphi_k^{(i+1,j)}, \vartheta_k)\mathbf{L}_{k,1}\big) \\
&\quad - \sum_{k\in\mathcal{K}} \mathsf{Tr}\big(\overline{\mathbf{S}}_{k,2}(\varrho_k)\mathbf{L}_{k,2}\big) - \sum_{k\in\mathcal{K}} \mathsf{Tr}\big(\overline{\mathbf{S}}_{k,3}(\xi_k)\mathbf{L}_{k,3}\big) \\
&\quad - \sum_{m\in\mathcal{M}} \big(\lambda_{m,1}p_m + \lambda_{m,2}C_m\big) - \sum_{k\in\mathcal{K}} \rho_{k,1}\vartheta_k \\
&\quad - \sum_{k\in\mathcal{K}} \rho_{k,2}\varrho_k - \sum_{k\in\mathcal{K}} \rho_{k,3}\xi_k,
\end{aligned} \tag{33}$$

where $\boldsymbol{\Lambda} \triangleq \big(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \boldsymbol{\rho}_3, \mathbf{L}_{\mathcal{K},1}, \mathbf{L}_{\mathcal{K},2}, \mathbf{L}_{\mathcal{K},3}, \mathbf{V}_{\mathcal{K}}\big)$ contains the dual variables. Specifically, $\boldsymbol{\lambda}_n \triangleq \big(\lambda_{1,n}, \ldots, \lambda_{M,n}\big) \succeq \mathbf{0}, \forall\, n \in \{1,2\}$, are the vectors of the dual variables for constraints (25b) and (25f) in problem (32), respectively; $\boldsymbol{\rho}_n \triangleq \big(\rho_{1,n}, \ldots, \rho_{K,n}\big) \succeq \mathbf{0}, \forall\, n \in \{1,2,3\}$, are the vectors of the dual variables for the constraints in (25h) in problem (32); $\mathbf{L}_{\mathcal{K},n} \triangleq \big\{\mathbf{L}_{k,n} \mid \mathbf{L}_{k,n} \in \mathbb{H}^{MN}, k \in \mathcal{K}\big\}, \forall\, n \in \{1,2,3\}$, with $\mathbf{L}_{k,n} \succeq \mathbf{0}, \forall\, k \in \mathcal{K}, n \in \{1,2,3\}$, are the sets of dual variable matrices for constraints (32b), (25d), and (25e) in problem (32), respectively; $\mathbf{V}_{\mathcal{K}} \triangleq \big\{\mathbf{V}_k \mid \mathbf{V}_k \in \mathbb{H}^{MN}, k \in \mathcal{K}\big\}$ with $\mathbf{V}_k \succeq \mathbf{0}, \forall\, k \in \mathcal{K}$, is the set of dual variable matrices for constraint (25g) in problem (32). The dual problem of problem (32) is given as follows:

$$\underset{\boldsymbol{\Lambda}}{\text{minimize}} \ \sup_{\mathbf{W}_{\mathcal{K}},\, \boldsymbol{\vartheta},\, \boldsymbol{\varrho},\, \boldsymbol{\xi}} \ \mathcal{L}_{\text{theo2}}^{(i+1,j)}\big(\mathbf{W}_{\mathcal{K}}, \boldsymbol{\vartheta}, \boldsymbol{\varrho}, \boldsymbol{\xi}, \boldsymbol{\Lambda}\big). \tag{34}$$

We focus on the following KKT conditions that are relevant for our proof:

$$\nabla_{\mathbf{W}_k}\mathcal{L}_{\text{theo2}}^{(i+1,j)}\big(\mathbf{W}_{\mathcal{K}}, \boldsymbol{\vartheta}, \boldsymbol{\varrho}, \boldsymbol{\xi}, \boldsymbol{\Lambda}\big)\Big|_{\overline{\boldsymbol{\Upsilon}}^{(i+1,j)},\, \boldsymbol{\Lambda}^{(i+1,j)}} = \mathbf{O}_{MN}, \quad \forall\, k \in \mathcal{K}, \tag{35a}$$

$$\mathbf{V}_k^{(i+1,j)}\overline{\mathbf{W}}_k^{(i+1,j)} = \mathbf{O}_{MN}, \quad \forall\, k \in \mathcal{K}, \tag{35b}$$

$$\big(\overline{\mathbf{S}}_{k,1}(\varphi_k^{(i+1,j)}, \overline{\vartheta}_k^{(i+1,j)}) + \mathbf{Q}_k^{\mathsf{H}}\overline{\mathbf{W}}_k^{(i+1,j)}\mathbf{Q}_k\big)\mathbf{L}_{k,1}^{(i+1,j)} = \mathbf{O}_{MN}, \quad \forall\, k \in \mathcal{K}, \tag{35c}$$

$$\big(\overline{\mathbf{S}}_{k,2}(\overline{\varrho}_k^{(i+1,j)}) + \mathbf{Q}_k^{\mathsf{H}}\overline{\mathbf{W}}_k^{(i+1,j)}\mathbf{Q}_k\big)\mathbf{L}_{k,2}^{(i+1,j)} = \mathbf{O}_{MN}, \quad \forall\, k \in \mathcal{K}, \tag{35d}$$

$$\mathbf{L}_{k,3}^{(i+1,j)} \succeq \mathbf{0}, \quad \forall\, k \in \mathcal{K}, \tag{35e}$$

$$\overline{\mathbf{W}}_k^{(i+1,j)} \succeq \mathbf{0}, \quad \forall\, k \in \mathcal{K}, \tag{35f}$$

$$\overline{\boldsymbol{\vartheta}}^{(i+1,j)} \succeq \mathbf{0}, \ \overline{\boldsymbol{\varrho}}^{(i+1,j)} \succeq \mathbf{0}, \ \overline{\boldsymbol{\xi}}^{(i+1,j)} \succeq \mathbf{0}, \tag{35g}$$

$$\boldsymbol{\lambda}_1^{(i+1,j)} \succeq \mathbf{0}, \ \boldsymbol{\lambda}_2^{(i+1,j)} \succeq \mathbf{0}, \tag{35h}$$

where $\overline{\boldsymbol{\Upsilon}}^{(i+1,j)} \triangleq \big(\overline{\mathbf{W}}_{\mathcal{K}}^{(i+1,j)}, \overline{\boldsymbol{\vartheta}}^{(i+1,j)}, \overline{\boldsymbol{\varrho}}^{(i+1,j)}, \overline{\boldsymbol{\xi}}^{(i+1,j)}\big)$ and $\boldsymbol{\Lambda}^{(i+1,j)} \triangleq \big(\boldsymbol{\lambda}_1^{(i+1,j)}, \boldsymbol{\lambda}_2^{(i+1,j)}, \boldsymbol{\rho}_1^{(i+1,j)},$

$\boldsymbol{\rho}_2^{(i+1,j)}, \boldsymbol{\rho}_3^{(i+1,j)}, \mathbf{L}_{\mathcal{K},1}^{(i+1,j)}, \mathbf{L}_{\mathcal{K},2}^{(i+1,j)}, \mathbf{L}_{\mathcal{K},3}^{(i+1,j)}, \mathbf{V}_{\mathcal{K}}^{(i+1,j)})$ represent the optimal solutions of the primal and dual problems in (34), respectively; $\nabla_{\mathbf{W}_k} \mathcal{L}_{\text{theo2}}^{(i+1,j)}(\mathbf{W}_{\mathcal{K}}, \boldsymbol{\vartheta}, \boldsymbol{\varrho}, \boldsymbol{\xi}, \boldsymbol{\Lambda})\big|_{\overline{\boldsymbol{\Upsilon}}^{(i+1,j)}, \boldsymbol{\Lambda}^{(i+1,j)}}$ denotes the gradient of the Lagrangian function in (33) with respect to $\mathbf{W}_k$ at $\overline{\boldsymbol{\Upsilon}}^{(i+1,j)}$ and $\boldsymbol{\Lambda}^{(i+1,j)}$. By jointly considering (35a) and (35b), we have the following equality:

$$\mathbf{X}_k^{(i+1,j)} \overline{\mathbf{W}}_k^{(i+1,j)} = \mathbf{Q}_k(\mathbf{L}_{k,1}^{(i+1,j)} + \mathbf{L}_{k,2}^{(i+1,j)})\mathbf{Q}_k^{\mathsf{H}} \overline{\mathbf{W}}_k^{(i+1,j)}, \quad \forall k \in \mathcal{K}, \quad (36)$$

where $\mathbf{X}_k^{(i+1,j)} \triangleq \mathbf{I}_{MN} + \sum_{m\in\mathcal{M}}(\lambda_{m,1}^{(i+1,j)} + R_k^{(i)}\lambda_{m,2}^{(i+1,j)} q_{m,k}^{(i)})\mathbf{B}_m^{\mathsf{H}}\mathbf{B}_m + \sum_{u\in\mathcal{K}\setminus\{k\}}\mathbf{Q}_u\mathbf{L}_{u,3}^{(i+1,j)}\mathbf{Q}_u^{\mathsf{H}}$. Moreover, we have $q_{m,k}^{(i)} > 0$ by definition and $R_k^{(i)} > 0$ due to the lower-bounded SINR $\Gamma_{\text{req},k}$, $\forall k \in \mathcal{K}$. Further considering (35e) and (35h), we have $\mathbf{X}_k^{(i+1,j)} \succ \mathbf{0}$, i.e., $\mathsf{Rank}(\mathbf{X}_k^{(i+1,j)}) = MN$. Thus, we have

$$\begin{aligned}
&\mathsf{Rank}(\overline{\mathbf{W}}_k^{(i+1,j)}) \\
&= \mathsf{Rank}(\mathbf{X}_k^{(i+1,j)} \overline{\mathbf{W}}_k^{(i+1,j)}) \\
&= \mathsf{Rank}(\mathbf{Q}_k(\mathbf{L}_{k,1}^{(i+1,j)} + \mathbf{L}_{k,2}^{(i+1,j)})\mathbf{Q}_k^{\mathsf{H}} \overline{\mathbf{W}}_k^{(i+1,j)}) \\
&\leq \min\Big\{\mathsf{Rank}(\mathbf{Q}_k(\mathbf{L}_{k,1}^{(i+1,j)} + \mathbf{L}_{k,2}^{(i+1,j)})\mathbf{Q}_k^{\mathsf{H}}), \\
&\qquad \mathsf{Rank}(\overline{\mathbf{W}}_k^{(i+1,j)})\Big\}, \quad \forall k \in \mathcal{K}.
\end{aligned} \quad (37)$$

To evaluate $\mathsf{Rank}(\mathbf{Q}_k(\mathbf{L}_{k,1}^{(i+1,j)} + \mathbf{L}_{k,2}^{(i+1,j)})\mathbf{Q}_k^{\mathsf{H}})$, $\forall k \in \mathcal{K}$, we post-multiply $\mathbf{Q}_k^{\mathsf{H}}$ to (35c) $\forall k \in \mathcal{K}$, so

$$\begin{aligned}
&\overline{\mathbf{S}}_{k,1}(\varphi_k^{(i+1,j)}, \overline{\vartheta}_k^{(i+1,j)})\mathbf{L}_{k,1}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}} \\
&\quad + \mathbf{Q}_k^{\mathsf{H}}\overline{\mathbf{W}}_k^{(i+1,j)}\mathbf{Q}_k\mathbf{L}_{k,1}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}} = \mathbf{O}_{MN}, \quad \forall k \in \mathcal{K}.
\end{aligned} \quad (38)$$

We then pre-multiply the LHS of (38) by $[\mathbf{I}_{MN}\ \mathbf{0}_{MN}]$. By noting that $\mathbf{Q}_k = [\mathbf{I}_{MN}\ \widehat{\mathbf{h}}_k]$, we have

$$\begin{aligned}
&[\mathbf{I}_{MN}\ \mathbf{0}_{MN}]\overline{\mathbf{S}}_{k,1}(\varphi_k^{(i+1,j)}, \overline{\vartheta}_k^{(i+1,j)})\mathbf{L}_{k,1}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}} \\
&\quad + [\mathbf{I}_{MN}\ \mathbf{0}_{MN}]\mathbf{Q}_k^{\mathsf{H}}\overline{\mathbf{W}}_k^{(i+1,j)}\mathbf{Q}_k\mathbf{L}_{k,1}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}} = \mathbf{O}_{MN} \\
&\overset{\text{(a)}}{\iff} \overline{\vartheta}_k^{(i+1,j)}[\mathbf{I}_{MN}\ \mathbf{0}_{MN}]\mathbf{L}_{k,1}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}} \\
&\quad + \mathbf{I}_{MN}\overline{\mathbf{W}}_k^{(i+1,j)}\mathbf{Q}_k\mathbf{L}_{k,1}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}} = \mathbf{O}_{MN} \\
&\overset{\text{(b)}}{\iff} \overline{\vartheta}_k^{(i+1,j)}\mathbf{Q}_k\mathbf{L}_{k,1}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}} + \overline{\mathbf{W}}_k^{(i+1,j)}\mathbf{Q}_k\mathbf{L}_{k,1}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}} \\
&\quad = \overline{\vartheta}_k^{(i+1,j)}[\mathbf{O}_{MN}\ \widehat{\mathbf{h}}_k] \\
&\iff (\overline{\vartheta}_k^{(i+1,j)}\mathbf{I}_{MN} + \overline{\mathbf{W}}_k^{(i+1,j)})\mathbf{Q}_k\mathbf{L}_{k,1}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}} \\
&\quad = \overline{\vartheta}_k^{(i+1,j)}[\mathbf{O}_{MN}\ \widehat{\mathbf{h}}_k], \qquad \forall k \in \mathcal{K}.
\end{aligned} \quad (39)$$

In step (a), we substituted $\varphi_k^{(i+1,j)}$ and $\overline{\vartheta}_k^{(i+1,j)}$ for φ_k and ϑ_k in $\overline{\mathbf{S}}_{k,1}(\varphi_k, \vartheta_k)$, respectively. Note that $[\mathbf{I}_{MN}\ \mathbf{0}_{MN}]\mathbf{Q}_k^{\mathsf{H}} = \mathbf{I}_{MN}$. Step (b) follows by adding $\overline{\vartheta}_k^{(i+1,j)}[\mathbf{O}_{MN}\ \widehat{\mathbf{h}}_k]$ on both sides of the equation. Following similar steps in (38) and (39), we post-multiply and pre-multiply the LHS of (35d) by $\mathbf{Q}_k^{\mathsf{H}}$ and $[\mathbf{I}_{MN}\ \mathbf{0}_{MN}]$, respectively, and obtain another equality:

$$\begin{aligned}
&(\overline{\mathbf{S}}_{k,2}(\overline{\varrho}_k^{(i+1,j)}) + \mathbf{Q}_k^{\mathsf{H}}\overline{\mathbf{W}}_k^{(i+1,j)}\mathbf{Q}_k)\mathbf{L}_{k,2}^{(i+1,j)} = \mathbf{O}_{MN} \\
&\iff (\overline{\varrho}_k^{(i+1,j)}\mathbf{I}_{MN} + \overline{\mathbf{W}}_k^{(i+1,j)})\mathbf{Q}_k\mathbf{L}_{k,2}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}} \\
&\quad = \overline{\varrho}_k^{(i+1,j)}[\mathbf{O}_{MN}\ \widehat{\mathbf{h}}_k], \qquad \forall k \in \mathcal{K}.
\end{aligned} \quad (40)$$

Without loss of generality, for UE k, we can distinguish three cases for $\overline{\vartheta}_k^{(i+1,j)}$ and $\overline{\varrho}_k^{(i+1,j)}$:

Case I: $\overline{\vartheta}_k^{(i+1,j)} \neq 0$ and $\overline{\varrho}_k^{(i+1,j)} \neq 0$. According to (35f) and (35g), the inverse of matrices $\overline{\vartheta}_k^{(i+1,j)}\mathbf{I}_{MN} + \overline{\mathbf{W}}_k^{(i+1,j)}$ and $\overline{\varrho}_k^{(i+1,j)}\mathbf{I}_{MN} + \overline{\mathbf{W}}_k^{(i+1,j)}$ exist. Further considering (39) and (40), we have

$$\begin{aligned}
&\mathsf{Rank}(\mathbf{Q}_k(\mathbf{L}_{k,1}^{(i+1,j)} + \mathbf{L}_{k,2}^{(i+1,j)})\mathbf{Q}_k^{\mathsf{H}}) \\
&= \mathsf{Rank}(\mathbf{Q}_k\mathbf{L}_{k,1}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}} + \mathbf{Q}_k\mathbf{L}_{k,2}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}}) \\
&= \mathsf{Rank}\Big((\overline{\vartheta}_k^{(i+1,j)}(\overline{\vartheta}_k^{(i+1,j)}\mathbf{I}_{MN} + \overline{\mathbf{W}}_k^{(i+1,j)})^{-1} \\
&\quad + \overline{\varrho}_k^{(i+1,j)}(\overline{\varrho}_k^{(i+1,j)}\mathbf{I}_{MN} + \overline{\mathbf{W}}_k^{(i+1,j)})^{-1})[\mathbf{O}_{MN}\ \widehat{\mathbf{h}}_k]\Big) \\
&\leq \mathsf{Rank}([\mathbf{O}_{MN}\ \widehat{\mathbf{h}}_k]) = 1.
\end{aligned} \quad (41)$$

By substituting (41) into (37), we have $\mathsf{Rank}(\overline{\mathbf{W}}_k^{(i+1,j)}) \leq 1$. On the other hand, we have $\overline{\mathbf{W}}_k^{(i+1,j)} \neq \mathbf{O}_{MN}$ due to $\Gamma_{\text{req},k} > 0$. Thus, for Case I, we have $\mathsf{Rank}(\overline{\mathbf{W}}_k^{(i+1,j)}) = 1$.

Case II: $\overline{\vartheta}_k^{(i+1,j)} = 0$, $\overline{\varrho}_k^{(i+1,j)} \neq 0$ or $\overline{\vartheta}_k^{(i+1,j)} \neq 0$, $\overline{\varrho}_k^{(i+1,j)} = 0$. For $\overline{\vartheta}_k^{(i+1,j)} \neq 0$, $\overline{\varrho}_k^{(i+1,j)} = 0$, we have

$$\overline{\mathbf{W}}_k^{(i+1,j)}\mathbf{Q}_k\mathbf{L}_{k,2}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}} = \mathbf{O}_{MN}. \quad (42)$$

Further, considering (35a) and (35b), we have

$$\overline{\mathbf{W}}_k^{(i+1,j)}(\mathbf{X}_k^{(i+1,j)} - \mathbf{Q}_k\mathbf{L}_{k,1}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}}) = \mathbf{O}_{MN}. \quad (43)$$

Since $\overline{\vartheta}_k^{(i+1,j)} \neq 0$, $(\overline{\vartheta}_k^{(i+1,j)}\mathbf{I}_{MN} + \overline{\mathbf{W}}_k^{(i+1,j)})^{-1}$ exists due to (35f) and (35g). From (39), we have

$$\begin{aligned}
&\mathsf{Rank}(\mathbf{Q}_k\mathbf{L}_{k,1}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}}) \\
&= \mathsf{Rank}(\overline{\vartheta}_k^{(i+1,j)}(\overline{\vartheta}_k^{(i+1,j)}\mathbf{I}_{MN} + \overline{\mathbf{W}}_k^{(i+1,j)})^{-1}[\mathbf{O}_{MN}\ \widehat{\mathbf{h}}_k]) \\
&\leq \mathsf{Rank}([\mathbf{O}_{MN}\ \widehat{\mathbf{h}}_k]) = 1.
\end{aligned} \quad (44)$$

We now introduce the following lemma that facilitates our proof:

Lemma A.1: For matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ of the same size, $\mathsf{Rank}(\mathbf{C}_1 - \mathbf{C}_2) \geq \mathsf{Rank}(\mathbf{C}_1) - \mathsf{Rank}(\mathbf{C}_2)$.

Proof: As $\mathsf{Rank}(\mathbf{C}_1) + \mathsf{Rank}(\mathbf{C}_2) \geq \mathsf{Rank}(\mathbf{C}_1 + \mathbf{C}_2)$, it follows that $\mathsf{Rank}(\mathbf{C}_1 - \mathbf{C}_2) + \mathsf{Rank}(\mathbf{C}_2) \geq \mathsf{Rank}(\mathbf{C}_1)$. Thus, we have $\mathsf{Rank}(\mathbf{C}_1 - \mathbf{C}_2) \geq \mathsf{Rank}(\mathbf{C}_1) - \mathsf{Rank}(\mathbf{C}_2)$, which completes the proof. ■

Applying Lemma A.1 to the second term of the LHS of (43), we have

$$\begin{aligned}
&\mathsf{Rank}(\mathbf{X}_k^{(i+1,j)} - \mathbf{Q}_k(\mathbf{L}_{k,1}^{(i+1,j)})\mathbf{Q}_k^{\mathsf{H}}) \\
&\geq \mathsf{Rank}(\mathbf{X}_k^{(i+1,j)}) - \mathsf{Rank}(\mathbf{Q}_k(\mathbf{L}_{k,1}^{(i+1,j)})\mathbf{Q}_k^{\mathsf{H}}) \geq MN - 1.
\end{aligned} \quad (45)$$

Thus, $\overline{\mathbf{W}}_k^{(i+1,j)}$ lies in the null space of matrix $\mathbf{X}_k^{(i+1,j)} - \mathbf{Q}_k\mathbf{L}_{k,1}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}}$ which has at least rank $MN - 1$. Thus, $\mathsf{Rank}(\overline{\mathbf{W}}_k^{(i+1,j)}) \leq 1$ for UE k. We also have $\overline{\mathbf{W}}_k^{(i+1,j)} \neq \mathbf{O}_{MN}$

for $\Gamma_{\text{req},k} > 0$. Thus, $\mathsf{Rank}\big(\overline{\mathbf{W}}_k^{(i+1,j)}\big) = 1$ for case $\overline{\vartheta}_k^{(i+1,j)} \neq 0$ and $\overline{\varrho}_k^{(i+1,j)} = 0$. A similar approach can be applied for $\overline{\vartheta}_k^{(i+1,j)} = 0$ and $\overline{\varrho}_k^{(i+1,j)} \neq 0$ in Case II, and we conclude that $\mathsf{Rank}\big(\overline{\mathbf{W}}_k^{(i+1,j)}\big) = 1$ for Case II.

Case III: $\overline{\vartheta}_k^{(i+1,j)} = 0$ and $\overline{\varrho}_k^{(i+1,j)} = 0$. We show by contradiction that $\overline{\vartheta}_k^{(i+1,j)} = 0$ and $\overline{\varrho}_k^{(i+1,j)} = 0$ cannot occur. Assume UE $k \in \mathcal{K}$ such that $\overline{\vartheta}_k^{(i+1,j)} = 0$ and $\overline{\varrho}_k^{(i+1,j)} = 0$. We substitute $\overline{\vartheta}_k^{(i+1,j)} = 0$ and $\overline{\varrho}_k^{(i+1,j)} = 0$ into (39) and (40) for UE k, respectively. We have

$$\begin{aligned}\overline{\mathbf{W}}_k^{(i+1,j)}\mathbf{Q}_k\mathbf{L}_{k,1}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}} = \mathbf{O}_{MN},\\ \overline{\mathbf{W}}_k^{(i+1,j)}\mathbf{Q}_k\mathbf{L}_{k,2}^{(i+1,j)}\mathbf{Q}_k^{\mathsf{H}} = \mathbf{O}_{MN}.\end{aligned} \tag{46}$$

Thus, $\overline{\mathbf{W}}_k^{(i+1,j)}\mathbf{X}_k^{(i+1,j)} = \mathbf{O}_{MN}$, cf. (36). Since $\mathbf{X}_k^{(i+1,j)} \succ \mathbf{0}$, we have $\overline{\mathbf{W}}_k^{(i+1,j)} = \mathbf{O}_{MN}$. However, $\overline{\mathbf{W}}_k^{(i+1,j)} = \mathbf{O}_{MN}$ cannot be in the optimal solution of problem (32) due to the minimum SINR requirement of UE k. This is a contradiction. Thus, $\overline{\vartheta}_k^{(i+1,j)} = 0$ and $\overline{\varrho}_k^{(i+1,j)} = 0$ cannot occur.

Thus, each UE $k \in \mathcal{K}$ belongs to either Case I or Case II. For both cases, $\mathsf{Rank}\big(\overline{\mathbf{W}}_k^{(i+1,j)}\big) = 1$ has been proven, which completes the proof. ∎

REFERENCES

[1] Z. Wang, D. W. K. Ng, V. W.S. Wong, and R. Schober, "Transmit beamforming for QoE improvement in C-RAN with mobile virtual network operators," in *Proc. of IEEE Int'l Commun. Conf. (ICC)*, Kuala Lumpur, Malaysia, May 2016.

[2] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2015-2020," Feb. 2016. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[3] D. Gesbert, S. Hanly, H. Huang, S. Shamai, 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.

[4] China Mobile, "C-RAN: The road towards green radio access networks, version 2.5," *White Paper*, Jun. 2012.

[5] J. Zhao, T. Quek, and Z. Lei, "Coordinated multipoint transmission with limited backhaul data transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2762–2775, Jun. 2013.

[6] Y. Shi, J. Zhang, and K. Letaief, "Robust group sparse beamforming for multicast green cloud-RAN with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4647–4659, Sep. 2015.

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

[8] C. Yoon and D. Cho, "Energy efficient beamforming and power allocation in dynamic TDD based C-RAN system," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1806–1809, Oct. 2015.

[9] F. Zhuang and V. Lau, "Backhaul limited asymmetric cooperation for MIMO cellular networks via semidefinite relaxation," *IEEE Trans. on Signal Process.*, vol. 62, no. 3, pp. 684–693, Feb. 2014.

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

[11] J. Segel and M. Weldon, "Lightradio portfolio — Technical overview," *Technology White Paper 1*, Murray Hill, NJ, USA: Alcatel-Lucent, 2011.

[12] M. Grieger, S. Boob, and G. Fettweis, "Large scale field trial results on frequency domain compression for uplink joint detection," in *Proc. of IEEE Globecom Workshop on Multicell Cooperation*, Anaheim, CA, Dec. 2012.

[13] A. Sanderovich, O. Somekh, H. V. Poor, and S. Shamai, "Uplink macro diversity of limited backhaul cellular network," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3457–3478, Aug. 2009.

[14] A. Sanderovich, S. Shamai, and Y. Steinberg, "Distributed MIMO receiver — Achievable rates and upper bounds," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4419–4438, Oct. 2009.

[15] A. del Coso and S. Simoens, "Distributed compression for MIMO coordinated networks with a backhaul constraint," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4698–4709, Sep. 2009.

[16] L. Zhou and W. Yu, "Uplink multicell processing with limited backhaul via per-base-station successive interference cancellation," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 1981–1993, Oct. 2013.

[17] S. H. Park, O. Simeone, O. Sahin, and S. Shamai, "Robust and efficient distributed compression for cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 692–703, Feb. 2013.

[18] Y. Zhou and W. Yu, "Optimized beamforming and backhaul compression for uplink MIMO cloud radio access networks," in *Proc. of IEEE Wireless Optical Network Convergence in Support of Cloud Architectures*, Austin, TX, Dec. 2014.

[19] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai, "Downlink multicell processing with limited-backhaul capacity," *EURASIP J. Adv. Signal Process*, vol. 2009, pp. 3:1–3:10, Feb. 2009.

[20] P. Marsch and G. Fettweis, "On downlink network MIMO under a constrained backhaul and imperfect channel knowledge," in *Proc. of IEEE Global Commun. Conf. (GLOBECOM)*, Honolulu, HI, Nov. 2009.

[21] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "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.

[22] H. Huang, M. Trivellato, A. Hottinen, M. Shafi, P. Smith, and R. Valenzuela, "Increasing downlink cellular throughput with limited network MIMO coordination," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2983–2989, Jun. 2009.

[23] C. W. Tan, S. Friedland, and S. Low, "Spectrum management in multiuser cognitive wireless networks: Optimality and algorithm," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 421–430, Feb. 2011.

[24] I.-W. Lai, L. Zheng, C.-H. Lee, and C. W. Tan, "Beamforming duality and algorithms for weighted sum rate maximization in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 5, pp. 832–847, May 2015.

[25] M. Xiao, N. B. Shroff, and E. K. P. Chong, "A utility-based power-control scheme in wireless cellular systems," *IEEE/ACM Trans. Networking*, vol. 11, no. 2, pp. 210–221, Apr. 2003.

[26] J.-W. Lee, R. R. Mazumdar, and N. B. Shroff, "Downlink power allocation for multi-class wireless systems," *IEEE/ACM Trans. Networking*, vol. 13, no. 4, pp. 854–867, Aug. 2005.

[27] J. Wang and D. P. Palomar, "Worst-case robust MIMO transmission with imperfect channel knowledge," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3086–3100, Aug. 2009.

[28] C. Shen, T.-H. Chang, K.-Y. Wang, Z. Qiu, and C.-Y. Chi, "Distributed robust multicell coordinated beamforming with imperfect CSI: An ADMM approach," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2988–3003, Jun. 2012.

[29] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[30] G. Zheng, K. K. Wong, and B. Ottersten, "Robust cognitive beamforming with bounded channel uncertainties," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4871–4881, Dec. 2009.

[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[32] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[33] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[34] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," Jun. 2015.

[35] B. Vandereycken, "Low-rank matrix completion by Riemannian optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1214–1236, Jun. 2013.

[36] Y.-C. Jong, "An efficient global optimization algorithm for nonlinear sum of ratios problem," May 2012. [Online]. Available: http://www.optimization-online.org/DB_FILE/2012/08/3586.pdf

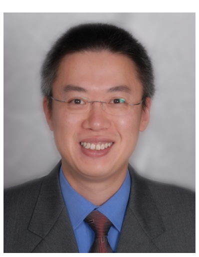
[37] L. Vandenberghe, V. R. Balakrishnan, R. Wallin, A. Hansson, and T. Roh, "Interior-point algorithms for semidefinite programming problems derived from the KYP lemma," *Positive Polynomials in Control*, vol. 312, pp. 195–238, Feb. 2005.

[38] CableFree. LTE metrics including RSRP, RSRQ and SINR. [Online]. Available: http://www.cablefree.net/wirelesstechnology/4glte/lte-rsrq-sinr/

[39] E. Boshkovska, A. Koelpin, D. W. K. Ng, N. Zlatanov, and R. Schober, "Robust beamforming for swipt systems with non-linear energy harvesting model," in *Proc. of IEEE Int'l Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Edinburgh, United Kingdom, Jul. 2016.

Zehua Wang (S'11, M'17) received the B.Eng. degree in Software Engineering from Wuhan University, Wuhan, China, in 2009, the M.Eng. degree in Electrical and Computer Engineering from Memorial University of Newfoundland, St John's, NL, Canada, in 2011, and the Ph.D. degree from the University of British Columbia (UBC), Vancouver, BC, Canada, in 2016. He is currently a postdoctoral research fellow at UBC, Vancouver, BC, Canada and the applied research engineering scientist at Left Inc., BC, Canada. His research interests include convex optimization, device-to-device communications, social networks, machine-type communications, and routing and forwarding in mobile ad hoc networks. He received the Chinese Government Award for Outstanding Self-Financed Students Abroad in 2015. He was the recipient of the Four Year Doctoral Fellowship (FYF) at UBC from 2012 to 2016. He was also awarded the Graduate Support Initiative (GSI) Award at UBC in 2014 and 2015. Dr. Wang served as the technical program committee (TPC) Co-chair of 2017 IEEE International Workshop on Smart Multimedia (SmartMM). He also served as the technical program committee (TPC) member for various conferences including IEEE International Conference on Communications (ICC), IEEE Global Communications Conference (Globecom), and IEEE Vehicular Technology Conference (VTC).

Derrick Wing Kwan Ng (S'06, M'12) received his Ph.D. degree from the University of British Columbia (UBC) in 2012. He was a senior postdoctoral fellow at the Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany. He is now working as a Lecturer and an ARC DECRA Research Fellow at the University of New South Wales, Sydney, Australia. His research interests include convex and non-convex optimization, physical layer security, wireless information and power transfer, and green (energy-efficient) wireless communications.



Vincent W.S. Wong (S'94, M'00, SM'07, F'16) received the B.Sc. degree from the University of Manitoba, Winnipeg, MB, Canada, in 1994, the M.A.Sc. degree from the University of Waterloo, Waterloo, ON, Canada, in 1996, and the Ph.D. degree from the University of British Columbia (UBC), Vancouver, BC, Canada, in 2000. From 2000 to 2001, he worked as a systems engineer at PMC-Sierra Inc. (now Microsemi). He joined the Department of Electrical and Computer Engineering at UBC in 2002 and is currently a Professor. His research areas include protocol design, optimization, and resource management of communication networks, with applications to wireless networks, smart grid, mobile cloud computing, and Internet of Things. Dr. Wong is an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS. He has served as a Guest Editor of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and IEEE WIRELESS COMMUNICATIONS. He has also served on the editorial boards of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and *Journal of Communications and Networks*. He was a Technical Program Co-chair of IEEE SmartGridComm'14, as well as a Symposium Co-chair of IEEE SmartGridComm ('13, '17) and IEEE Globecom'13. He is the Chair of the IEEE Communications Society Emerging Technical Sub-Committee on Smart Grid Communications and the IEEE Vancouver Joint Communications Chapter. He received the 2014 UBC Killam Faculty Research Fellowship.

Robert Schober (S'98, M'01, SM'08, F'10) was born in Neuendettelsau, Germany, in 1971. He received the Diplom (Univ.) and the Ph.D. degrees in electrical engineering from Friedrich-Alexander Universität Erlangen-Nürnberg (FAU), Erlangen, Germany, in 1997 and 2000, respectively. From 2001 to 2002, he was a Post-Doctoral Fellow with the University of Toronto, Canada, sponsored by the German Academic Exchange Service (DAAD). From 2002 to 2011, he was a Professor and Canada Research Chair with The University of British Columbia (UBC), Vancouver, Canada. Since 2012, he has been an Alexander von Humboldt Professor and the Chair for Digital Communication with the FAU. His research interests include the broad areas of communication theory, wireless communications, and statistical signal processing.

Dr. Schober is a fellow of the Canadian Academy of Engineering and a fellow of the Engineering Institute of Canada. From 2012 to 2015, he served as an Editor-in-Chief of the IEEE TRANSACTIONS ON COMMUNICATIONS and since 2014, he is the Chair of the Steering Committee of the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL, AND MULTISCALE COMMUNICATIONS. Furthermore, he is a member-at-large of the Board of Governors of the IEEE Communications Society. He has received several awards for his work, including the 2002 Heinz MaierLeibnitz Award of the German Science Foundation (DFG), the 2004 Innovations Award of the Vodafone Foundation for Research in Mobile Communications, the 2006 UBC Killam Research Prize, the 2007 Wilhelm Friedrich Bessel Research Award of the Alexander von Humboldt Foundation, the 2008 Charles McDowell Award for Excellence in Research from UBC, a 2011 Alexander von Humboldt Professorship, and a 2012 NSERC E.W.R. Steacie Fellowship. In addition, he has received several best paper awards for his research.