User-Centric Joint Access-Backhaul Design for Full-Duplex Self-Backhauled Wireless Networks

Erkai Chen, Meixia Tao, and Nan Zhang

Abstract

Full-duplex self-backhauling is promising to provide cost-effective and flexible backhaul connectivity for ultra-dense wireless networks, but also poses a great challenge to resource management between the access and backhaul links. In this paper, we propose a user-centric joint access-backhaul transmission framework for full-duplex self-backhauled wireless networks. In the access link, user-centric clustering is adopted so that each user is cooperatively served by multiple small base stations (SBSs). In the backhaul link, user-centric multicast transmission is proposed so that each user's message is treated as a common message and multicast to its serving SBS cluster. We first formulate an optimization problem to maximize the network weighted sum rate through joint access-backhaul beamforming and SBS clustering when global channel state information (CSI) is available. This problem is efficiently solved via the successive lower-bound maximization approach with a novel approximate objective function and the iterative link removal technique. We then extend the study to the stochastic joint access-backhaul beamforming optimization with partial CSI. Simulation results demonstrate the effectiveness of the proposed algorithms for both full CSI and partial CSI scenarios. They also show that the transmission design with partial CSI can greatly reduce the CSI overhead with little performance degradation.

Index Terms

Full-duplex self-backhauling, joint access-backhaul design, user-centric clustering, successive lower-bound maximization (SLBM), partial CSI.

I. Introduction

Backhaul has emerged as a new challenge of 5G networks due to the ultra-dense deployment of small cells and the explosive growth of data traffic [2], [3]. Traditional fiber-based backhaul can provide high data rates, but the prohibitive cost and the geographical limitations make it impossible to deploy in many practical scenarios. In-band wireless backhauling, also referred to as self-backhauling, is a promising and viable alternative since it utilizes the same spectrum and the same infrastructure with the access link and enables low-cost and plug-and-play installation [3]. Using self-backhauling, a small base station (SBS) can easily receive data from a macro base station (MBS) in the downlink (or a mobile user in the uplink) and then transmit the data to a mobile user (or an MBS) over the same wireless radio spectrum. Combining recent advances in in-band full-duplex (IBFD) technique further facilitates self-backhauled SBSs to transmit and receive at the same time, thereby potentially doubling the spectrum efficiency [4].

Radio resource management across the access and backhaul links is crucial for performance optimization in self-backhauled wireless networks. It is also an important study issue mentioned in the 3GPP technical report on integrated access and backhaul [5]. The challenge lies in the newly introduced cross-link interference by spectrum sharing between the access and backhaul links. Recently, several research efforts have been made to address the resource management issue for full-duplex self-backhauled wireless networks [6]–[8]. The authors in [6] demonstrated that compared with a conventional time-division duplex (TDD)/frequency-division duplex (FDD) self-backhauled network, the downlink rate in the full-duplex self-backhauled heterogeneous network is nearly doubled, but at the expense of reduced coverage due to

higher interference. By employing massive MIMO and transmit beamforming at the MBS, the authors in [7] studied a joint cell association and power allocation problem for energy efficiency maximization. An advanced block digitalization precoding scheme was proposed to eliminate the cross-link interference and multi-user interference. The authors in [8] studied a joint scheduling and interference mitigation problem for network utility maximization. A regularized zero-forcing precoding scheme and the SBS operation mode switching (between full-duplex and half-duplex) were jointly considered to mitigate both the co-link interference and the cross-link interference. Note that in [6]–[8], it is assumed that each user is associated with only one SBS in the access link and no inter-site cooperation is considered.

Multi-cell cooperation (e.g., CoMP) is a promising technique to mitigate the inter-cell interference by allowing the user data to be jointly processed by several interfering cell sites, thus mimicking a large virtual MIMO system [9], [10]. Exploiting the inter-site cooperation, the authors in [11] studied the joint access and backhaul resource management for ultra-dense networks. Both joint transmission (CoMP-JT) and coordinated beamforming (CoMP-CB) are considered in the access link. However, the cross-link interference between the access and backhaul links is ignored. The authors in [12] studied the joint access and backhaul design for the downlink of cloud radio access networks (C-RANs). Therein, a user-centric clustering strategy is assumed in the access link such that each user can be served cooperatively by a cluster of nearby SBSs, and multicast transmission is adopted in the backhaul link by the MBS to deliver the message of each user to the SBSs in its serving cluster simultaneously. The messages of different users are delivered one by one in a time-division manner over the backhaul link. Based on a similar network model in [12], a different transmission scheme is considered in [13] to balance the tradeoff between the cooperation benefits and the coordination overhead, where CoMP-CB is adopted in the access link such that each user is served by only one SBS and accordingly, multi-user beamforming is adopted in the backhaul link. Both works, however, are for out-band wireless backhaul and thus did not have the cross-link interference issue.

In this paper, we propose a joint access-backhaul transmission framework for full-duplex self-backhauled wireless networks by considering the inter-site cooperation. As in [12], an adaptive user-centric clustering strategy is adopted in the access link, where each user is cooperatively served by a cluster of SBSs. To reduce the backhaul traffic load, a user-centric multicast transmission is adopted in the backhaul link, where each user's data is treated as a multicast message and the SBS cluster receiving the same user's message forms a multicast group. Different from [12], we consider IBFD-enabled SBSs which can transmit in the access link while receiving over the backhaul link simultaneously using the same frequency band with potential cross-link interference as well as self-interference. In addition, unlike [12] where all users' messages are fetched in a time-division manner over the backhaul link, we adopt a non-orthogonal multicast transmission scheme over the backhaul link so that the MBS delivers all the multicast messages simultaneously, which is more efficient than the orthogonal time-division manner [14]. Under this user-centric joint design framework, we seek the maximum end-to-end achievable rate of the network under the per-BS power constraints through the joint design of multicast beamforming in the backhaul link as well as SBS clustering and beamforming in the access link.

Enabling the inter-site cooperation in general requires acquiring the channel state information (CSI) and exchanging it among cell sites, which is a challenging issue for ultra-dense networks. With a large number of cell sites and mobile users in the network, acquiring full CSI will bring excessive signaling overhead and counteract the performance gains provided by cooperative transmission [15], [16]. Specifically, obtaining the CSI of all cell sites at the user receivers through downlink pilot training may require a long training period that is comparable to the channel coherence time. Moreover, feeding back all the CSI by the user receivers will occupy plenty of the uplink resources [17]. After collecting the CSI, each cell site will forward it to the central controller or share it among the cooperating cell sites for joint signal processing. The overhead will easily overwhelm the backhaul resources, especially in self-backhauled wireless networks, where the backhaul is a scarce resource. Thus, in this work we consider the user-centric joint access-backhaul design with both full CSI and partial CSI. While the former with full CSI can provide the performance upper bound in the ideal case, the latter with partial CSI offers a more

practical solution. The contributions of this paper are summarized as follows.

We first consider the joint access-backhaul design with full CSI, where the CSI between each user and each cell site is globally available at a central controller. We formulate an optimization problem for the joint design of multicast beamforming in the backhaul link, SBS clustering and beamforming in the access link to maximize the weighted sum rate of all users under per-BS peak power constraints. This problem is a non-convex mixed-integer non-linear programming (MINLP) problem, which is challenging due to the non-smoothness and the non-convexity of the objective function as well as the combinatorial nature of the SBS clustering. To solve the problem, we first consider the joint access-backhaul beamforming design problem under given SBS clustering. Note that this problem can be approximately transformed into a manifold optimization problem and solved via the Riemannian conjugate gradient (RCG) algorithm [18], as shown in our prior conference paper [1]. However, the RCG algorithm may get stuck in unfavorite local points when the MBS peak power is too large. In this paper, we propose to solve the joint access-backhaul beamforming problem via the successive lower-bound maximization (SLBM) approach. A novel concave lower-bound approximation for the achievable rate expression in the objective function is introduced based on signal-to-interference-plus-noise ratio (SINR) convexification. Simulation results show that the proposed SLBM algorithm with the newly introduced SINR-convexification based lower-bound approximation can avoid the high-power issue in the RCG algorithm. It can also achieve better performance than the well-known weighted minimum-mean-square-error (WMMSE) algorithm. We then develop a heuristic algorithm to determine the SBS clustering based on the iterative link removal technique. The effectiveness of the proposed clustering algorithm is also demonstrated via numerical simulations.

We also consider the joint access-backhaul beamforming design with partial CSI, where only part of the CSI in the access link is available. We formulate a stochastic beamforming design problem to maximize the average weighted sum rate of the network under the per-BS power constraints. We develop a stochastic SLBM algorithm to solve it by adopting the introduced SINR-convexification based lower-bound approximation. Moreover, we derive a deterministic lower-bound approximation for the average achievable rate by using Jensen's inequality. The original stochastic optimization problem is then approximately solved via solving the resulting deterministic approximation with low complexity. Simulation results demonstrate the performance of the proposed stochastic SLBM algorithm as well as the effectiveness of the proposed deterministic lower-bound approximation. The results also indicate that with a moderate amount of CSI, the proposed algorithms can achieve good performance that is very close to the full CSI case and significantly reduce the channel estimation overhead.

The rest of the paper is organized as follows. Section II introduces the system model. Section III considers the joint access-backhaul design with full CSI and introduces the proposed joint access-backhaul beamforming and SBS clustering algorithms. The joint access-backhaul design with partial CSI is presented in Section IV. Simulation results are provided in Section V. Finally, we conclude the paper in Section VI.

Notations: Boldface lower-case and upper-case letters denote vectors and matrices, respectively. Calligraphy letters denote sets or problems, depending on the context. $\mathbb{R}$ and $\mathbb{C}$ denote the real and complex domains, respectively. $|\cdot|$ and $\|\cdot\|_2$ denote the absolute value and Euclidean norm, respectively. The operators $(\cdot)^T$ and $(\cdot)^\dagger$ correspond to the transpose and Hermitian transpose, respectively. $\mathcal{CN}(\delta, \sigma^2)$ represents a complex Gaussian distribution with mean δ and variance σ^2. The real part of a complex number x is denoted by $\Re\{x\}$. Finally, $\mathbf{0}_{L \times N}$ denotes the all-zero matrix of dimension $L \times N$.

II. System Model

A. Network Model

Consider the downlink transmission of a wireless network with full-duplex self-backhauling, where N SBSs $\{b_n \mid n \in \mathcal{N} \triangleq \{1, 2, \ldots, N\}\}$ cooperatively serve K users $\{u_k \mid k \in \mathcal{K} \triangleq \{1, 2, \ldots, K\}\}$ and each of them is connected to the core network through an MBS b_0 with in-band wireless backhaul. The MBS is equipped with M antennas. Each SBS is enabled by a full-duplex radio with $L + 1$ antennas:
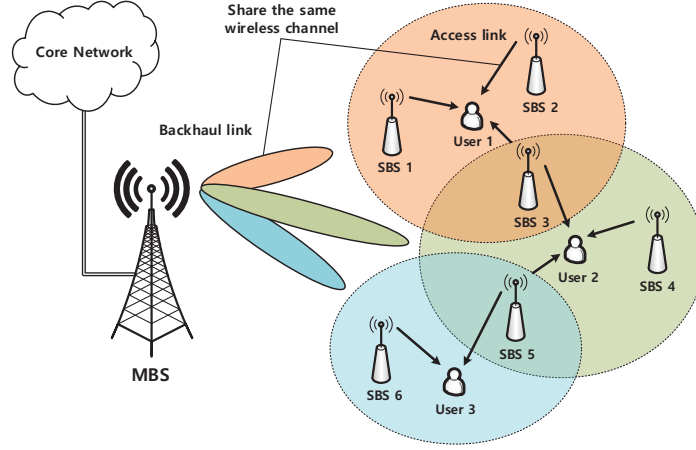
Fig. 1. User-centric joint access-backhaul transmission in an IBFD self-backhauled wireless network.

one for receiving at the wireless backhaul (from the MBS to SBSs) and L for transmitting at the access link (from SBSs to users). Each user is equipped with a single antenna. It is assumed that the users that cannot be served by SBSs will be served by the MBS directly on other orthogonal resource blocks, which is not considered in this paper for simplicity.

We consider a user-centric clustering strategy in the access link. Each user u_k is served by a cluster of SBSs cooperatively, denoted as $\mathcal{N}_k$. An example is shown in Fig. 1, where the serving SBS clusters of the three users are $\mathcal{N}_1 = \{1, 2, 3\}$, $\mathcal{N}_2 = \{3, 4, 5\}$, and $\mathcal{N}_3 = \{5, 6\}$, respectively. Let the binary variable $c_{k,n} = 1$ indicate that SBS b_n belongs to the SBS cluster of user u_k and $c_{k,n} = 0$ otherwise. Thus, we have $\mathcal{N}_k \triangleq \{n \in \mathcal{N} \mid c_{k,n} = 1\}$. The data intended to user u_k should be fetched at all the SBSs in $\mathcal{N}_k$ from the MBS via the backhaul link. Note that the SBS clusters $\{\mathcal{N}_k\}_{k=1}^K$ may overlap with each other, which means each SBS may serve multiple users at the same time. We denote $\mathcal{K}_n \triangleq \{k \in \mathcal{K} \mid c_{k,n} = 1\}$ as the set of users served by SBS b_n.

In the backhaul link, the MBS adopts a user-centric multicast transmission. Specifically, the MBS treats the message intended to each user u_k as a multicast message and transmits it to the multicast group formed by the SBS cluster $\mathcal{N}_k$ associated with this user. Each SBS may receive multiple multicast messages due to the potentially overlapped SBS clusters. All the messages of the users are transmitted simultaneously at the MBS and each SBS decodes the set of messages for its served users using successive interference cancellation (SIC)-based receiver.

B. Signal Model

Let $x_k^{\mathrm{A}} \in \mathbb{C}$ and $x_k^{\mathrm{B}} \in \mathbb{C}$ denote the transmitted signals in the access link and the backhaul link, respectively, that carry the message intended for user u_k. All these signals have normalized power of 1. Let $\mathbf{w}_{k,n} \in \mathbb{C}^{L \times 1}$ denote the beamforming vector at SBS b_n for message x_k^{A} in the access link and $\mathbf{v}_k \in \mathbb{C}^{M \times 1}$ denote the multicast beamforming vector at the MBS for message x_k^{B} in the backhaul link. Note that $\mathbf{w}_{k,n} = \mathbf{0}_{L \times 1}$ if $c_{k,n} = 0$, which implies that SBS b_n does not participate in the transmission of message x_k^{A}. The main notations in the system model are summarized in Table I.

With full-duplex capability, each SBS can deliver messages to its served users in access while receiving messages from the MBS in backhaul simultaneously using the same frequency band.

1) Access Link: The received signal at user u_k is given by

$$y_{u_k} = \underbrace{\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j} x_k^{\mathrm{A}}}_{\text{desired signal}} + \underbrace{\sum_{i=1}^K \mathbf{h}_{u_k}^{(b_0)\dagger} \mathbf{v}_i x_i^{\mathrm{B}}}_{\text{cross-link interference}} + \underbrace{\sum_{i=1, i \neq k}^K \left(\sum_{j \in \mathcal{N}_i} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{i,j} \right) x_i^{\mathrm{A}}}_{\text{co-link interference}} + \underbrace{z_{u_k}}_{\text{noise}}, \tag{1}$$

TABLE I
SUMMARY OF NOTATIONS

Notation	Description
$\mathcal{N}$	set of all SBSs
$\mathcal{K}$	set of all users
$c_{k,n}$	binary variable indicating whether or not SBS b_n belongs to the cluster of user u_k
$\mathcal{K}_n$	set of users served by SBS b_n
$\mathcal{N}_k$	set of SBSs serving user u_k
x_k^{A} (x_k^{B})	message intended for user u_k in the access (backhaul) link
z_{u_k} (z_{b_n})	additive white Gaussian noise at user u_k (SBS b_n)
$\mathbf{w}_{k,n}$	beamforming vector at SBS b_n for message x_k^{A}
$\mathbf{v}_k$	beamforming vector at the MBS b_0 for message x_k^{B}
$\mathbf{h}_{u_k}^{(b_0)}$ $(\mathbf{h}_{u_k}^{(b_j)})$	channel vector between the MBS b_0 (SBS b_j) and user u_k
$\mathbf{h}_{b_n}^{(b_0)}$ $(\mathbf{h}_{b_n}^{(b_j)}, j \neq n)$	channel vector between the MBS b_0 (SBS b_j) and SBS b_n
$\mathbf{h}_{b_n}^{(b_n)}$	SI channel at SBS b_n
β_{SI}	SI cancellation capability of the SBSs
P^{M} (P_n^{S})	peak transmit power of the MBS (SBS b_n)

where $\mathbf{h}_{u_k}^{(b_0)} \in \mathbb{C}^{M \times 1}$ $(\mathbf{h}_{u_k}^{(b_j)} \in \mathbb{C}^{L \times 1})$ is the channel vector between the MBS b_0 (SBS b_j) and user u_k, and $z_{u_k} \sim \mathcal{CN}(0, \sigma_{u_k}^2)$ is the additive white Gaussian noise at user u_k. The channel coefficient $\mathbf{h}_{u_k}^{(b_0)}$ is modeled as $\mathbf{h}_{u_k}^{(b_0)} = \sqrt{\beta_{u_k}^{(b_0)}} \mathbf{g}_{u_k}^{(b_0)}$, where $\beta_{u_k}^{(b_0)} \in \mathbb{R}$ is the large-scale fading coefficient that includes the path loss and shadowing, and $\mathbf{g}_{u_k}^{(b_0)} \sim \mathcal{CN}(\mathbf{0}, \mathbf{1})$ is the small-scale fading coefficient modeled as an independent and identically distributed (i.i.d.) random vector. The channel vector $\mathbf{h}_{u_k}^{(b_j)}$ is modeled in the same manner. In (1), the first term is the desired signal transmitted cooperatively by all the SBSs in the cluster $\mathcal{N}_k$, the second term is the cross-link interference transmitted by the MBS over the backhaul link, and the third term presents the co-link interference transmitted by all the SBSs in the same access link but intended to other users. Note that the cross-link interference term includes the message for user u_k in the backhaul link, i.e., x_k^{B}, but the user does not intend to decode it directly due to the potentially weak channel condition between the MBS and this user.

Based on (1), the achievable data rate of user u_k in the access link can be expressed as

$$R_k^{\mathrm{A}} = \log\left(1 + \frac{|\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j}|^2}{\Phi_k + \sigma_{u_k}^2}\right), \tag{2}$$

where $\Phi_k = \sum_{i=1}^{K} |\mathbf{h}_{u_k}^{(b_0)\dagger} \mathbf{v}_i|^2 + \sum_{i \neq k}^{K} |\sum_{j \in \mathcal{N}_i} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{i,j}|^2$.

2) Backhaul Link: The received signal at SBS b_n is given by

$$y_{b_n} = \underbrace{\sum_{k \in \mathcal{K}_n} \mathbf{h}_{b_n}^{(b_0)\dagger} \mathbf{v}_k x_k^{\mathrm{B}}}_{\text{desired signals}} + \underbrace{\sum_{i=1, i \notin \mathcal{K}_n}^{K} \mathbf{h}_{b_n}^{(b_0)\dagger} \mathbf{v}_i x_i^{\mathrm{B}}}_{\text{co-link interference}} + \underbrace{\sum_{i=1}^{K} \left(\sum_{j \in \mathcal{N}_i} \mathbf{h}_{b_n}^{(b_j)\dagger} \mathbf{w}_{i,j}\right) x_i^{\mathrm{A}}}_{\text{cross-link interference}} + \underbrace{z_{b_n}}_{\text{noise}}, \tag{3}$$

where $\mathbf{h}_{b_n}^{(b_0)} \in \mathbb{C}^{M \times 1}$ $(\mathbf{h}_{b_n}^{(b_j)} \in \mathbb{C}^{L \times 1}, j \neq n)$ is the channel vector between the MBS b_0 (SBS b_j) and SBS b_n and $z_{b_n} \sim \mathcal{CN}(0, \sigma_{b_n}^2)$ is the additive white Gaussian noise at SBS b_n. The channel coefficient $\mathbf{h}_{b_n}^{(b_0)}$ $(\mathbf{h}_{b_n}^{(b_j)})$ is modeled as $\mathbf{h}_{b_n}^{(b_0)} = \sqrt{\beta_{b_n}^{(b_0)}} \mathbf{g}_{b_n}^{(b_0)}$ $(\mathbf{h}_{b_n}^{(b_j)} = \sqrt{\beta_{b_n}^{(b_j)}} \mathbf{g}_{b_n}^{(b_j)})$. In (3), the first term represents the desired signals intended to the set of users $\mathcal{K}_n$ served by SBS b_n, the second term is the co-link interference caused by the signals intended to other users, and the third term is the cross-link interference transmitted by all the SBSs over the access link. Note that the cross-link interference term in (3) includes the self-interference (SI) term $\sum_{i \in \mathcal{K}_n} \mathbf{h}_{b_n}^{(b_n)\dagger} \mathbf{w}_{i,n} x_i^{\mathrm{A}}$, where $\mathbf{h}_{b_n}^{(b_n)}$ is the SI channel at SBS b_n and models the residual SI due to imperfect SI cancellation. The cross-link interference term also contains the desired signals by users in

$\mathcal{K}_n$, i.e., $\{x_i^\text{A} \mid i \in \mathcal{K}_n\}$ but transmitted by other SBSs. In this work, since the location of SBSs are fixed, we assume that the CSI between SBSs changes very slowly, and hence can be perfectly estimated and made available at all SBSs. Moreover, each SBS b_n already has the knowledge of its transmitted signals, i.e., $\{x_i^\text{A} \mid i \in \mathcal{K}_n\}$, the cross-link interference from other SBSs that transmit the same signals can be perfectly canceled at SBS b_n.

Since each SBS has only one receiving antenna for the wireless backhaul, SIC-based multi-user detection is adopted to decode the multiple signals of its served users. In general, the decoding order of these signals at each SBS can be optimized to improve the system performance. In this paper, for simplicity, we assume that the same decoding order is applied for all SBSs, which is given heuristically according to the aggregative channel gain of each user from all SBSs in the access link (i.e., $H_k \triangleq \sum_{n=1}^N \|\mathbf{h}_{u_k}^{(b_n)}\|_2^2$). This implies that the user with better channel condition has a higher priority to be decoded.

Based on the above discussion, the achievable rate of decoding the message for user u_k at SBS b_n in the backhaul link is given by

$$R_{k,n}^\text{B} = \log\left(1 + \frac{|\mathbf{h}_{b_n}^{(b_0)\dagger}\mathbf{v}_k|^2}{\Delta_{k,n} + \sigma_{b_n}^2}\right), \tag{4}$$

where $\Delta_{k,n} = \sum_{i \in \mathcal{I}_{k,n}} |\mathbf{h}_{b_n}^{(b_0)\dagger}\mathbf{v}_i|^2 + \sum_{i \notin \mathcal{K}_n} |\sum_{j \in \mathcal{N}_i} \mathbf{h}_{b_n}^{(b_j)\dagger}\mathbf{w}_{i,j}|^2 + \beta_\text{SI} \sum_{i \in \mathcal{K}_n} \|\mathbf{w}_{i,n}\|_2^2$ is a term composed of co-link interference, cross-link interference, and residual SI, β_SI represents the SI cancellation capability of the SBSs, and $\mathcal{I}_{k,n} = \{i \in \mathcal{K} \mid H_i < H_k \text{ or } i \notin \mathcal{K}_n\}$ is an index set that is jointly determined by the SIC decoding order and the set of users served by SBS b_n.

Since the message for user u_k is multicast to the SBS cluster $\mathcal{N}_k$, to ensure all the SBSs in $\mathcal{N}_k$ can decode the message successfully, the overall transmission rate of the message for user u_k in the backhaul link is limited by the SBS with the worst channel condition, given by

$$R_k^\text{B} = \min_{n \in \mathcal{N}_k} \{R_{k,n}^\text{B}\}. \tag{5}$$

Considering the backhaul link from the MBS to SBSs and the access link from SBSs to users, the end-to-end achievable rate of user u_k is given by

$$R_k = \min\{R_k^\text{A}, R_k^\text{B}\}. \tag{6}$$

III. Joint Access-Backhaul Design with Full CSI

In this section, we consider the joint access-backhaul design when all the CSI is available. We first provide the problem formulation for the joint design of multicast beamforming in the backhaul link, SBS clustering and beamforming in the access link. We then develop an effective algorithm to solve it via the SLBM approach and the iterative link removal technique.

A. Problem Formulation

Our objective is to maximize the end-to-end weighted sum rate of the network through joint design of the backhaul multicast beamforming $\mathbf{v} \triangleq \{\mathbf{v}_k \mid k \in \mathcal{K}\}$, the access beamforming $\mathbf{w} \triangleq \{\mathbf{w}_{k,n} \mid k \in \mathcal{K}, n \in$

$\mathcal{N}\}$, and the SBS clustering $\mathbf{c} \triangleq \{c_{k,n} \mid k \in \mathcal{K}, n \in \mathcal{N}\}$. This problem is mathematically formulated as

$$\mathcal{P} : \max_{\mathbf{w},\mathbf{v},\mathbf{c}} \sum_{k=1}^{K} \omega_k R_k \tag{7a}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \|\mathbf{v}_k\|_2^2 \leq P^{\mathrm{M}}, \tag{7b}$$

$$\sum_{k=1}^{K} \|\mathbf{w}_{k,n}\|_2^2 \leq P_n^{\mathrm{S}}, \ \forall \, n \in \mathcal{N}, \tag{7c}$$

$$\mathbf{v}_k \left(1 - \max_{n \in \mathcal{N}}\{c_{k,n}\}\right) = \mathbf{0}_{M \times 1}, \ \forall \, k \in \mathcal{K}, \tag{7d}$$

$$\mathbf{w}_{k,n}(1 - c_{k,n}) = \mathbf{0}_{L \times 1}, \ \forall \, k \in \mathcal{K}, \ n \in \mathcal{N}, \tag{7e}$$

where P^{M} is the peak power of the MBS, P_n^{S} is the peak power of SBS b_n, and $\{\omega_k\}_{k=1}^{K}$ are the weights accounting for possibly different service grades among all users.

Problem $\mathcal{P}$ is a non-convex MINLP problem [19], which is NP-hard in general. Obtaining its optimal solution is challenging due to the non-smoothness and the non-convexity of the rate expression (6) as well as the combinatorial nature of the SBS clustering variable $\mathbf{c}$. Even when the SBS clustering $\mathbf{c}$ is given, problem $\mathcal{P}$ is still non-convex and computationally difficult.

In the following subsections, we first tackle problem $\mathcal{P}$ with a given SBS cluster $\mathbf{c}$, denoted as $\mathcal{P}(\mathbf{c})$. We propose to solve it via the SLBM approach with a newly introduced lower-bound approximation for the achievable rate. We then develop a heuristic algorithm to determine the SBS clustering $\mathbf{c}$ based on the iterative link removal technique.

B. Successive Lower-Bound Maximization for $\mathcal{P}(\mathbf{c})$

For ease of notation, we rewrite problem $\mathcal{P}$ with a given SBS cluster $\mathbf{c}$, i.e., $\mathcal{P}(\mathbf{c})$ as

$$\mathcal{P}(\mathbf{c}) : \max_{\mathbf{x}} \ f(\mathbf{x}) \triangleq \sum_{k=1}^{K} \omega_k R_k(\mathbf{x}) \tag{8}$$

$$\text{s.t.} \ \ \mathbf{x} \in \mathcal{X}_{\mathbf{c}}.$$

where $\mathbf{x} \triangleq \{\mathbf{w}, \mathbf{v}\}$ and $\mathcal{X}_{\mathbf{c}}$ is a closed convex set of $\mathbf{x}$ constructed by constraints (7b), (7c), (7d), and (7e) with the given $\mathbf{c}$.

Problem $\mathcal{P}(\mathbf{c})$ is a non-convex problem with non-smooth and non-convex objective function but convex feasible region. We propose to solve it via the SLBM approach [20]. The main idea of the SLBM approach is to successively maximize a sequence of approximate objective functions [20]. Specifically, starting from a feasible point $\mathbf{x}^0$, the algorithm generates a sequence of $\{\mathbf{x}^t\}$ according to the update rule

$$\mathbf{x}^t \leftarrow \arg\max_{\mathbf{x} \in \mathcal{X}_{\mathbf{c}}} \ \hat{f}(\mathbf{x}, \mathbf{x}^{t-1}), \tag{9}$$

where $\mathbf{x}^{t-1}$ is the point obtained at the $(t-1)$-th iteration and $\hat{f}(\mathbf{x}, \mathbf{x}^{t-1})$ is an approximation of $f(\mathbf{x})$ at the t-th iteration. Typically the approximate function $\hat{f}(\mathbf{x}, \mathbf{x}^{t-1})$ needs to be carefully chosen such that the subproblem (9) is easy to solve. Moreover, to ensure the convergence of the SLBM algorithm, $\hat{f}(\mathbf{x}, \mathbf{x}^{t-1})$ should be a global lower bound for $f(\mathbf{x})$ and also be tight at $\mathbf{x}^{t-1}$, i.e., $\hat{f}(\mathbf{x}, \mathbf{x}^{t-1}) \leq f(\mathbf{x})$ and $\hat{f}(\mathbf{x}^{t-1}, \mathbf{x}^{t-1}) = f(\mathbf{x}^{t-1})$.

Note that the SLBM approach shares the same idea with many important algorithms such as the successive convex approximation (SCA) [21] and the concave-convex procedure (CCP) [22], by successively optimizing an approximate version of the original problem. However, they are different mainly in two aspects [20]:

- The SCA or CCP approximates both the objective functions and the feasible sets. On the contrary, the SLBM approximates only the objective function.
- The SCA or CCP is applicable only to problems with smooth objectives that are differentiable, while the SLBM is able to handle non-smooth objectives.

In this paper, we construct an approximate function $\hat{f}(\mathbf{x}, \mathbf{x}^{t-1})$ by introducing a novel concave lower-bound approximation for the non-convex achievable rate expression in the objective function based on SINR convexification, which allows subproblem (9) to be easily solved with guaranteed convergence. For comparison purpose, we first present an existing lower-bound approximation constructed by the WMMSE method [23], which has been widely used to deal with varieties of sum rate maximization problems.

1) WMMSE-based Concave Lower-Bound Approximation: We take the non-convex achievable rate expression in the access link (2) with given SBS clustering for example. In the WMMSE method, the relationship between the achievable rate and its mean-square-error (MSE) is established. The key transformation is to rewrite it into the following equivalent form [24]:

$$R_k^{\mathrm{A}}(\mathbf{x}) = \max_{\mathbf{a}_k} \left\{ \log(\rho_k) - \rho_k e_k + 1 \right\}, \tag{10}$$

where $\mathbf{a}_k \triangleq \{\alpha_k, \rho_k\}$, $\alpha_k \in \mathbb{C}$ is the receive beamformer, $\rho_k \in \mathbb{R}$ is a scalar variable associated with the k-th user, and $e_k \in \mathbb{R}$ is the MSE function defined as

$$
\begin{aligned}
e_k &= \mathbb{E}\left[|(\alpha_k)^{\dagger} y_{u_k} - x_k^{\mathrm{A}}|^2\right] \\
&= \left(|\textstyle\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j}|^2 + \Phi_k + \sigma_{u_k}^2\right) |\alpha_k|^2 - 2\Re\left\{(\alpha_k)^{\dagger}(\textstyle\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j})\right\} + 1.
\end{aligned} \tag{11}
$$

Note that in (10), with given SBS clustering, the achievable rate R_k^{A} in the access link is a function of the set of beamformers $\mathbf{x}$, denoted as $R_k^{\mathrm{A}}(\mathbf{x})$.

The unconstrained optimization problem in the right-hand side of (10) can be easily solved by using the first-order optimality condition. Its optimal solution $\mathbf{a}_k^*$ is given by

$$
\begin{cases}
\alpha_k^* = \left(|\textstyle\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j}|^2 + \Phi_k + \sigma_{u_k}^2\right)^{-1} \left(\textstyle\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j}\right), \\
\rho_k^* = 1/e_k^*.
\end{cases} \tag{12}
$$

where α_k^* is the minimum-mean-square-error (MMSE) receive beamformer and e_k^* is the resulting MSE.

Denote

$$G_k^{\mathrm{A\text{-}MMSE}}(\mathbf{x}, \mathbf{a}_k) \triangleq \log(\rho_k) - \rho_k e_k + 1. \tag{13}$$

This function is a global lower bound of $R_k^{\mathrm{A}}(\mathbf{x})$, which is derived by the rate-MSE relationship. Then, the WMMSE-based lower-bound approximation for the achievable rate $R_k^{\mathrm{A}}(\mathbf{x})$ at a feasible point $\mathbf{x}'$ can be constructed as

$$\hat{R}_k^{\mathrm{A\text{-}MMSE}}(\mathbf{x}, \mathbf{x}') \triangleq G_k^{\mathrm{A\text{-}MMSE}}(\mathbf{x}, \mathbf{a}_k(\mathbf{x}')), \tag{14}$$

where $\mathbf{a}_k(\mathbf{x}')$ is the argument that maximizes $G_k^{\mathrm{A\text{-}MMSE}}(\mathbf{x}, \mathbf{a}_k)$ with the given $\mathbf{x}'$, i.e.,

$$\mathbf{a}_k(\mathbf{x}') = \arg\max_{\mathbf{a}_k} G_k^{\mathrm{A\text{-}MMSE}}(\mathbf{x}', \mathbf{a}_k). \tag{15}$$

It is easy to check that

$$R_k^{\mathrm{A}}(\mathbf{x}) = \max_{\mathbf{a}_k} G_k^{\mathrm{A\text{-}MMSE}}(\mathbf{x}, \mathbf{a}_k) \geq G_k^{\mathrm{A\text{-}MMSE}}(\mathbf{x}, \mathbf{a}_k(\mathbf{x}')) = \hat{R}_k^{\mathrm{A\text{-}MMSE}}(\mathbf{x}, \mathbf{x}'), \tag{16}$$

and

$$R_k^{\mathrm{A}}(\mathbf{x}') = \max_{\mathbf{a}_k} G_k^{\mathrm{A\text{-}MMSE}}(\mathbf{x}', \mathbf{a}_k) = G_k^{\mathrm{A\text{-}MMSE}}(\mathbf{x}', \mathbf{a}_k(\mathbf{x}')) = \hat{R}_k^{\mathrm{A\text{-}MMSE}}(\mathbf{x}', \mathbf{x}'). \tag{17}$$

Thus, $\hat{R}_k^{\mathrm{A\text{-}MMSE}}(\mathbf{x}, \mathbf{x}')$ is a lower bound of $R_k^{\mathrm{A}}(\mathbf{x})$ and it is tight at the point $\mathbf{x}'$.

2) SINR-Convexification based Concave Lower-Bound Approximation: The main idea of SINR convexification is to approximate the SINR expressions in the achievable rates with their concave approximations, thus to convexify the weighted sum rate maximization problem. Specifically, for the achievable rate $R_k^{\mathrm{A}}(\mathbf{x})$ in (2), we introduce an auxiliary variable γ_k such that $\gamma_k = \Phi_k + \sigma_{u_k}^2$. Then the SINR expression in $R_k^{\mathrm{A}}(\mathbf{x})$ can be rewritten into a quadratic-over-linear form $\frac{|\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j}|^2}{\gamma_k}$, which is jointly convex in $\mathbf{w}_{k,j}$ and $\gamma_k > 0$ [25, Section 3.1.5]. Taking its first-order Taylor expansion at any feasible point $\{\mathbf{w}'_{k,j}, \gamma'_k\}$, we have

$$\frac{|\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j}|^2}{\gamma_k} \geq \frac{2\Re\{(\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}'_{k,j})^\dagger (\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j})\}}{\gamma'_k} - \frac{|\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}'_{k,j}|^2}{(\gamma'_k)^2} \gamma_k, \quad (18)$$

where the equality holds only when $\mathbf{w}_{k,j} = \mathbf{w}'_{k,j}$ and $\gamma_k = \gamma'_k$.

Denote $u_k^{\mathrm{A}} = \frac{\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}'_{k,j}}{\gamma'_k}$ and replace γ_k with $\Phi_k + \sigma_{u_k}^2$, then the inequality (18) can be rewritten as

$$\frac{|\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j}|^2}{\Phi_k + \sigma_{u_k}^2} \geq 2\Re\{(u_k^{\mathrm{A}})^\dagger (\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j})\} - |u_k^{\mathrm{A}}|^2 (\Phi_k + \sigma_{u_k}^2), \quad (19)$$

which holds for all $u_k^{\mathrm{A}} \in \mathbb{C}$. The equality holds only when $u_k^{\mathrm{A}} = u_k^{\mathrm{A}*}$, where $u_k^{\mathrm{A}*}$ is given by

$$u_k^{\mathrm{A}*} = (\Phi_k + \sigma_{u_k}^2)^{-1} (\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j}). \quad (20)$$

It is easy to check that the right-hand side of (19) is concave in $\mathbf{x}$ and u_k^{A}, respectively. It is also seen that $u_k^{\mathrm{A}*}$ in (20) is a scaled version of the MMSE receive beamformer in (12). By taking u_k^{A} as the receive beamformer, the right-hand side of (19) can be viewed as a lower bound of the SINR expression under the receive beamformer u_k^{A}, which is tight at the MMSE receive beamformer $u_k^{\mathrm{A}*}$.

Substituting (19) into the rate expression $R_k^{\mathrm{A}}(\mathbf{x})$ in (2), we have

$$R_k^{\mathrm{A}}(\mathbf{x}) = \log \left(1 + \frac{|\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j}|^2}{\Phi_k + \sigma_{u_k}^2} \right)$$

$$\geq \log \left(1 + 2\Re\{(u_k^{\mathrm{A}})^\dagger (\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j})\} - |u_k^{\mathrm{A}}|^2 (\Phi_k + \sigma_{u_k}^2) \right), \quad (21)$$

which holds for all $u_k^{\mathrm{A}} \in \mathbb{C}$. Then a global lower bound of $R_k^{\mathrm{A}}(\mathbf{x})$ is given by

$$G_k^{\mathrm{A}}(\mathbf{x}, u_k^{\mathrm{A}}) \triangleq \log \left(1 + 2\Re\{(u_k^{\mathrm{A}})^\dagger (\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j})\} - |u_k^{\mathrm{A}}|^2 (\Phi_k + \sigma_{u_k}^2) \right). \quad (22)$$

Different from the global lower bound $G_k^{\mathrm{A}\text{-MMSE}}(\mathbf{x}, \mathbf{a}_k)$ in (13), which is derived by the rate-MSE relationship, the above global lower bound $G_k^{\mathrm{A}}(\mathbf{x}, u_k^{\mathrm{A}})$ is derived via SINR convexification. Since the log function is concave and nondecreasing, according to the composition rules in [25, Section 3.2.4], $G_k^{\mathrm{A}}(\mathbf{x}, u_k^{\mathrm{A}})$ is concave in $\mathbf{x}$ and u_k^{A}, respectively. There also satisfies that

$$R_k^{\mathrm{A}}(\mathbf{x}) = \max_{u_k^{\mathrm{A}}} G_k^{\mathrm{A}}(\mathbf{x}, u_k^{\mathrm{A}}). \quad (23)$$

By checking its first-order optimality condition, the optimal solution is given by the MMSE receive beamformer in (20).

Then, similar to the WMMSE-based lower-bound approximation in (14), we construct a new concave lower-bound approximation for $R_k^{\mathrm{A}}(\mathbf{x})$ based on the relationship (22) derived via SINR convexification as

$$\hat{R}_k^{\mathrm{A}}(\mathbf{x}, \mathbf{x}') \triangleq G_k^{\mathrm{A}}(\mathbf{x}, u_k^{\mathrm{A}}(\mathbf{x}')), \quad (24)$$

where

$$u_k^A(\mathbf{x}') = \arg\max_{u_k^A} G_k^A(\mathbf{x}', u_k^A). \tag{25}$$

Clearly, there holds

$$R_k^A(\mathbf{x}) = \max_{u_k^A} G_k^A(\mathbf{x}, u_k^A) \geq G_k^A(\mathbf{x}, u_k^A(\mathbf{x}')) = \hat{R}_k^A(\mathbf{x}, \mathbf{x}'), \tag{26}$$

and

$$R_k^A(\mathbf{x}') = \max_{u_k^A} G_k^A(\mathbf{x}', u_k^A) = G_k^A(\mathbf{x}', u_k^A(\mathbf{x}')) = \hat{R}_k^A(\mathbf{x}', \mathbf{x}'). \tag{27}$$

Therefore, $\hat{R}_k^A(\mathbf{x}, \mathbf{x}')$ is a lower bound of $R_k^A(\mathbf{x})$, which is tight at $\mathbf{x}'$.

Similar to the WMMSE-based lower-bound approximation, our newly introduced lower-bound approximation via SINR convexification can also be used to handle varieties of weighted sum rate maximization problems, e.g., [23], [24]. Its superiority over the WMMSE-based approximation will be demonstrated via numerical simulations in Section V.

Applying the similar method, we can obtain a new concave lower-bound approximation for the non-convex rate expression $R_{k,n}^B$ in (4) as

$$\hat{R}_{k,n}^B(\mathbf{x}, \mathbf{x}') \triangleq G_{k,n}^B(\mathbf{x}, u_{k,n}^B(\mathbf{x}')), \tag{28}$$

where the function $G_{k,n}^B(\mathbf{x}, u_{k,n}^B)$ is given by

$$G_{k,n}^B(\mathbf{x}, u_{k,n}^B) \triangleq \log(1 + 2\Re\{(u_{k,n}^B)^\dagger(\mathbf{h}_{b_n}^{(b_0)\dagger}\mathbf{v}_k)\} - |u_{k,n}^B|^2(\Delta_{k,n} + \sigma_{b_n}^2)), \tag{29}$$

and $u_{k,n}^B(\mathbf{x}')$ is

$$u_{k,n}^B(\mathbf{x}') = \arg\max_{u_{k,n}^B} G_{k,n}^B(\mathbf{x}', u_{k,n}^B). \tag{30}$$

By utilizing the proposed SINR-convexification based lower-bound approximations in (24) and (28), we construct a lower bound for the end-to-end achievable rate $R_k(\mathbf{x})$ in (6) as

$$\hat{R}_k(\mathbf{x}, \mathbf{x}') \triangleq \min\left\{\hat{R}_k^A(\mathbf{x}, \mathbf{x}'), \min_{n\in\mathcal{N}_k}\left\{\hat{R}_{k,n}^B(\mathbf{x}, \mathbf{x}')\right\}\right\}. \tag{31}$$

It is easy to verify that $\hat{R}_k(\mathbf{x}, \mathbf{x}') \leq R_k(\mathbf{x})$, where the equality holds when $\mathbf{x} = \mathbf{x}'$. Moreover, $\hat{R}_k(\mathbf{x}, \mathbf{x}')$ is concave in $\mathbf{x}$, since the pointwise minimum of concave functions is also concave [25, Section 3.2.3].

Thus, the subproblem (9) in each iteration of the SLBM algorithm becomes

$$\mathbf{x}^t \leftarrow \arg\max_{\mathbf{x}\in\mathcal{X}_\mathbf{c}} \hat{f}(\mathbf{x}, \mathbf{x}^{t-1}) \triangleq \sum_{k=1}^K \omega_k \hat{R}_k(\mathbf{x}, \mathbf{x}^{t-1}). \tag{32}$$

Since $\hat{f}(\mathbf{x}, \mathbf{x}^{t-1})$ is a lower-bound approximation of the objective function $f(\mathbf{x})$ and is also tight at $\mathbf{x}^{t-1}$, with a feasible initial point, the iterations of the SLBM algorithm converge to a stationary solution of problem $\mathcal{P}(\mathbf{c})$ [20]. Note that problem (32) is a convex problem with log functions in the objective. It can be approximated by a sequence of second-order cone programming (SOCP) problems [26] via the successive approximation method [27]. Each SOCP can then be solved with a worst-case computational complexity of $\mathcal{O}(K^4 N(NL + M)^3)$ via the interior-point methods [26] using a general-purpose solver, e.g., SDPT3 in CVX [27].

The details of the SLBM algorithm using the proposed SINR-convexification based lower-bound approximation for solving problem $\mathcal{P}(\mathbf{c})$ are summarized in Alg. 1, denoted as SINRC-SLBM.

Algorithm 1 The SINRC-SLBM algorithm for solving problem $\mathcal{P}(\mathbf{c})$

Initialization: Find a feasible point $\mathbf{x}^0$ and set $t \leftarrow 1$.
Repeat
   1) Update $\mathbf{u}^{t-1} \triangleq \{u_k^{\mathrm{A}}(\mathbf{x}^{t-1}), u_{k,n}^{\mathrm{B}}(\mathbf{x}^{t-1}), \forall k, n\}$ according to the MMSE receive beamformer (20).
   2) Update $\mathbf{x}^t$ by solving problem (32).
   3) Set $t \leftarrow t + 1$.
Until the convergence criterion is met.

C. Heuristic Algorithm for $\mathcal{P}$

Intuitively, larger SBS cluster size can achieve higher access rate, but result in a lower backhaul rate due to multicast transmission, and vice versa. Thus, by controlling the SBS cluster size of each user, we can make a balance between the access rate and the backhaul rate, such that the end-to-end rate of the two-hop transmission is maximized.

With the above observation, we propose a heuristic algorithm based on the iterative link removal technique. Specifically, starting with full cooperation for each user, i.e., $c_{k,n}^0 = 1$, for all k and n, we shrink the SBS cluster size by deactivating several (denoted as J_Δ) weakest SBS-user links at each iteration and then solve the joint access and backhaul beamforming design problem with the updated SBS cluster. More specifically, at the t-th iteration, we solve problem $\mathcal{P}(\mathbf{c}^t)$ and calculate the transmit power of all the active SBS-user links with $c_{k,n}^t = 1$ as $P_{k,n} = \|\mathbf{w}_{k,n}\|_2^2$. We then sort them in the ascending order and update the SBS cluster $\mathbf{c}^{t+1}$ by deactivating J_Δ SBS-user links with the minimum power. The iterative procedure terminates when all SBS-links are inactive. Comparing the SBS clusters obtained at each iteration, we then choose the one that achieves the maximum objective value to be the final SBS cluster. The details of the algorithm are summarized in Alg. 2.

Algorithm 2 The heuristic algorithm for solving problem $\mathcal{P}$

Initialization: Initialize $\mathbf{c}^0$ with $c_{k,n}^0 = 1$. Set the iteration index $t \leftarrow 0$ and J_Δ.
While $\max\{\mathbf{c}^t\} = 1$
   1) Solve $\mathcal{P}(\mathbf{c}^t)$ in (8) using Alg. 1. Denote the objective value as $R(\mathbf{c}^t)$.
   2) Calculate $P_{k,n} = \|\mathbf{w}_{k,n}\|_2^2$ and sort them in the ascending order.
   3) Update the SBS cluster $\mathbf{c}^{t+1}$ by deactivating J_Δ SBS-user links with the minimum power.
   4) Set $t \leftarrow t + 1$.
End
Obtain the final SBS cluster $\mathbf{c}^{i^*}$, where $i^* = \arg\max_{0 \le i \le t} R(\mathbf{c}^i)$.

Complexity: At the t-th iteration, we need to solve one problem instance $\mathcal{P}(\mathbf{c}^t)$ to determine which J_Δ SBS-user links should be removed. Since the maximum number of iterations is KN/J_Δ, the overall complexity of the algorithm is $\mathcal{O}(K^5 N^2 (NL + M)^3 / J_\Delta)$.

IV. Joint Access-Backhaul Design with Partial CSI

The joint access-backhaul design presented in the previous section requires full CSI. However, in an ultra-dense network with a large number of cell sites and users, it is challenging to obtain all the CSI due to the excessive signaling overhead but limited training resources. To handle this challenge and reduce the channel estimation overhead, we consider the joint access-backhaul beamforming design only with partial CSI in this section. We first present the detailed assumption on the CSI availability, then provide the problem formulation for the stochastic beamforming design. We develop a stochastic SLBM algorithm by using the proposed SINR-convexification based lower-bound approximation and a low-complexity algorithm based on the deterministic lower-bound approximation to solve this problem, respectively.

A. Assumption on CSI Availability and Problem Formulation

Generally, the location of the cell sites is fixed and high above the ground. The CSI between the cell sites changes very slowly and can be tracked easily. The CSI estimation overhead mainly lies in the channel between the users and the cell sites due to the mobility of the users. Thus, one promising approach to reduce the CSI overhead is to acquire part of the instantaneous CSI between users and cell sites. In this work, we only need to acquire the instantaneous information of those links that have main contribution to the performance gain of network cooperation. Specifically, the instantaneous CSI from each user to its few nearby SBSs with strong large-scale channel gain and the instantaneous CSI from each user to the MBS are needed. The remaining channel links, i.e., the instantaneous channel coefficients of the links between each user and the SBSs that have weak large-scale channel gain can be ignored, since the acquisition of these links will only contribute little to the network performance. By using this way, we can reduce the CSI overhead greatly without losing too much of the performance. Throughout this section, the SBS clustering is assumed to be predetermined based on the large-scale fading, which varies slowly enough. Let $\mathcal{N}_k$ denote the predetermined SBS cluster of user u_k.

Recall that we have modeled the channel coefficients $\{\mathbf{h}_{u_k}^{(b_j)}\}$ using the large-scale fading coefficients $\{\beta_{u_k}^{(b_j)}\}$ and the small-scale fading coefficients $\{\mathbf{g}_{u_k}^{(b_j)}\}$ as $\{\mathbf{h}_{u_k}^{(b_j)} = \sqrt{\beta_{u_k}^{(b_j)}} \mathbf{g}_{u_k}^{(b_j)}\}$. Then, the following two kinds of CSI are assumed to be available:

1) Partial instantaneous CSI: The instantaneous CSI between all the cell sites (including the MBS and all SBSs), i.e., $\{\mathbf{h}_{b_n}^{(b_0)}, \mathbf{h}_{b_n}^{(b_j)} \mid j, n \in \mathcal{N}, \ j \neq n\}$, the instantaneous CSI between the MBS and all users, i.e., $\{\mathbf{h}_{u_k}^{(b_0)} \mid k \in \mathcal{K}\}$, and the instantaneous CSI between each user and its serving SBSs, i.e., $\{\mathbf{h}_{u_k}^{(b_j)} \mid j \in \mathcal{N}_k, k \in \mathcal{K}\}$.

2) Statistical CSI: The large-scale fading coefficients of the channel links between each user and the SBSs that do not serve it, i.e., $\{\beta_{u_k}^{(b_j)} \mid j \notin \mathcal{N}_k, k \in \mathcal{K}\}$, whose instantaneous CSI is unavailable.

Let $\Omega = \{\mathbf{h}_{u_k}^{(b_j)} \mid j \notin \mathcal{N}_k, k \in \mathcal{K}\}$ denote the set of unknown instantaneous CSI. Since the unknown CSI in Ω only involves the channel in the access link, we consider the following average achievable rate for the k-th user in the access link:

$$\bar{R}_k^{\mathrm{A}}(\mathbf{x}) = \mathbb{E}_\Omega \left[R_k^{\mathrm{A}}(\mathbf{x}; \Omega) \right], \tag{33}$$

where $R_k^{\mathrm{A}}(\mathbf{x}; \Omega)$ denotes the achievable rate for one realization of Ω and the expectation is performed over Ω. While for the backhaul link, the achievable rate $R_{k,n}^{\mathrm{B}}(\mathbf{x})$ is deterministic and given by (4). Therefore, the end-to-end average achievable rate of user u_k is

$$\bar{R}_k(\mathbf{x}) = \min \left\{ \bar{R}_k^{\mathrm{A}}(\mathbf{x}), \ \min_{n \in \mathcal{N}_k} \left\{ R_{k,n}^{\mathrm{B}}(\mathbf{x}) \right\} \right\}. \tag{34}$$

We aim to maximize the end-to-end average weighted sum rate of all users via joint access-backhaul beamforming design with partial CSI. This problem is formulated as

$$\mathcal{P}_{\mathrm{S}}(\mathbf{c}) : \max_{\mathbf{x} \in \mathcal{X}_{\mathbf{c}}} \ f(\mathbf{x}) \triangleq \sum_{k=1}^{K} \omega_k \bar{R}_k(\mathbf{x}). \tag{35}$$

Note that $\mathcal{P}_{\mathrm{S}}(\mathbf{c})$ is a stochastic programming problem with non-smooth and non-convex objective function. We propose to solve it via the stochastic SLBM approach with the proposed SINR-convexification based lower-bound approximation.

B. Stochastic Successive Lower-Bound Maximization for $\mathcal{P}_S(\mathbf{c})$

A classical method for solving the above stochastic optimization problem $\mathcal{P}_{\mathrm{S}}(\mathbf{c})$ is the sample average approximation (SAA) method [28]. By adopting the SAA method, the stochastic objective function is approximated by an ensemble average and the resulting deterministic optimization problem is then solved

by an appropriate numerical algorithm. However, its computation complexity is high, especially for non-convex stochastic optimization problems. In this paper, we adopt the stochastic SLBM approach by maximizing an approximate ensemble average at each iteration. To ensure convergence and to facilitate computation, the approximate ensemble average should be a locally tight strongly concave lower bound of the expected objective function [29]. To do so, we generalize the SINR-convexification based lower-bound approximation in (22) to a strongly concave version as

$$\tilde{G}_k^{\prime A}(\mathbf{x}, \mathbf{b}_k^A; \Omega) \triangleq \log(1 + 2\Re\{(u_k^A)^\dagger(\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger}\mathbf{w}_{k,j})\} - |u_k^A|^2(\Phi_k + \sigma_{u_k}^2))$$
$$-\frac{\gamma}{2}\sum_{n=1}^{N}\|\tilde{\mathbf{w}}_{k,n} - \mathbf{w}_{k,n}\|_2^2, \tag{36}$$

where $\mathbf{b}_k^A \triangleq \{u_k^A, \tilde{\mathbf{w}}_{k,n}\}$ and $\tilde{\mathbf{w}}_{k,n} \in \mathbb{C}^{L\times 1}$ is an auxiliary variable corresponding to $\mathbf{w}_{k,n}$. Comparing with (22), the quadratic term in (36) is to make $\tilde{G}_k^A$ a strongly concave function in $\mathbf{x}$ with any fixed parameter $\gamma > 0$.

By checking the first-order optimality condition, it can be verified that

$$R_k^A(\mathbf{x}; \Omega) = \max_{\mathbf{b}_k^A} \tilde{G}_k^{\prime A}(\mathbf{x}, \mathbf{b}_k^A; \Omega), \tag{37}$$

of which the optimal solution is given by

$$\begin{cases} u_k^{A*} = (\Phi_k + \sigma_{u_k}^2)^{-1}(\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger}\mathbf{w}_{k,j}), \\ \tilde{\mathbf{w}}_{k,n}^* = \mathbf{w}_{k,n}. \end{cases} \tag{38}$$

Then, we can construct a strongly concave lower-bound approximation for the rate expression $R_k^A(\mathbf{x}; \Omega)$ with a given feasible point $\mathbf{x}'$:

$$\hat{R}_k^A(\mathbf{x}, \mathbf{x}'; \Omega) \triangleq \tilde{G}_k^A(\mathbf{x}, \mathbf{b}_k^A(\mathbf{x}'; \Omega); \Omega), \tag{39}$$

where

$$\mathbf{b}_k^A(\mathbf{x}'; \Omega) = \arg\max_{\mathbf{b}_k^A} \tilde{G}_k^A(\mathbf{x}', \mathbf{b}_k^A; \Omega). \tag{40}$$

By utilizing the stochastic SLBM approach, at the t-th iteration, $\bar{R}_k^A(\mathbf{x})$ is approximated by

$$\bar{R}_k^A(\mathbf{x}) = \mathbb{E}_\Omega\left[R_k^A(\mathbf{x}; \Omega)\right] \simeq \frac{1}{t}\sum_{i=1}^{t} \hat{R}_k^A(\mathbf{x}, \mathbf{x}^{i-1}; \Omega^i), \tag{41}$$

where Ω^i is the i-th channel realization and $\mathbf{x}^{i-1}$ is the beamformers obtained from the $(i-1)$-th iteration. While for the achievable rate $R_{k,n}^B(\mathbf{x})$, it is replaced by its concave lower bound $\hat{R}_{k,n}^B(\mathbf{x}, \mathbf{x}^{t-1})$ derived in (28).

Therefore, at the t-th iteration of the stochastic SLBM algorithm, we solve the following subproblem

$$\mathbf{x}^t \leftarrow \arg\max_{\mathbf{x} \in \mathcal{X}_\mathbf{c}} \hat{f}(\mathbf{x}, \mathbf{x}^{t-1}) \triangleq \sum_{k=1}^{K} \omega_k \hat{R}_k(\mathbf{x}, \mathbf{x}^{t-1}), \tag{42}$$

where

$$\hat{R}_k(\mathbf{x}, \mathbf{x}^{t-1}) \triangleq \min\left\{\frac{1}{t}\sum_{i=1}^{t} \hat{R}_k^A(\mathbf{x}, \mathbf{x}^{i-1}; \Omega^i), \min_{n \in \mathcal{N}_k}\left\{\hat{R}_{k,n}^B(\mathbf{x}, \mathbf{x}^{t-1})\right\}\right\}. \tag{43}$$

Since $\hat{f}(\mathbf{x}, \mathbf{x}^{t-1})$ is a strongly concave lower-bound approximation of the original objective function $f(\mathbf{x})$ and is a locally tight at $\mathbf{x}^{t-1}$, the iterations generated by the stochastic SLBM algorithm converge to the set of stationary points of $\mathcal{P}_S(\mathbf{c})$ almost surely [29]. The subproblem (42) in the each iteration of the stochastic SLBM algorithm is convex, which can be approximated by a sequence of SOCPs and solved using the interior-point methods with a worst-case computational complexity of $\mathcal{O}(K^4N(NL + M)^3)$.

Finally, we summarize the stochastic SLBM algorithm using the proposed SINR-convexification based lower-bound approximation for solving problem $\mathcal{P}_S(\mathbf{c})$ in Alg. 3, denoted as SINRC-SSLBM.

Algorithm 3 The SINRC-SSLBM algorithm for solving problem $\mathcal{P}_S(\mathbf{c})$

Initialization: Randomly generate one feasible point $\mathbf{x}^0$ and set the iteration index $t \leftarrow 1$.

Repeat

  1) Obtain a new channel realization Ω^t.
  2) Update $\mathbf{b}_k^A(\mathbf{x}^{t-1}; \Omega^t)$ according to (38) for all k.
  3) Update $\mathbf{x}^t$ by solving problem (42).
  4) Set $t \leftarrow t + 1$.

Until the stopping criterion is met.

C. Low-Complexity Algorithm via Deterministic Approximation

In general, solving problem $\mathcal{P}_S(\mathbf{c})$ via the stochastic optimization algorithms often needs a large number of iterations due to the presence of random parameters, which suffers from high computational complexity. In this subsection, we derive a deterministic lower-bound approximation for the average achievable rate $\bar{R}_k^A(\mathbf{x})$ in (33) and solve the resulting deterministic optimization problem via the SLBM algorithm with low complexity.

Specifically, since $\log(1 + \frac{c}{x})$ is a convex function for any $x > 0$ with a given positive constant c, by using Jensen's inequality [25, Section 3.1.8] as in [30], a lower bound of the rate $\bar{R}_k^A(\mathbf{x})$ can be derived as

$$\bar{R}_k^A(\mathbf{x}) \geq \log\left(1 + \frac{|\sum_{j \in \mathcal{N}_k} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{k,j}|^2}{\bar{\Phi}_k + \sigma_{u_k}^2}\right), \tag{44}$$

where

$$\bar{\Phi}_k = \mathbb{E}_\Omega[\Phi_k] = \mathbb{E}_\Omega\left[\sum_{i=1}^K |\mathbf{h}_{u_k}^{(b_0)\dagger} \mathbf{v}_i|^2 + \sum_{i \neq k}^K |\sum_{j \in \mathcal{N}_i} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{i,j}|^2\right]$$

$$= \sum_{i=1}^K |\mathbf{h}_{u_k}^{(b_0)\dagger} \mathbf{v}_i|^2 + \sum_{i \neq k}^K \mathbf{w}_i^\dagger \mathbf{A}_{u_k} \mathbf{w}_i. \tag{45}$$

Here, $\mathbf{w}_i \triangleq [\mathbf{w}_{i,1}^\dagger, \ldots, \mathbf{w}_{i,N}^\dagger]^\dagger$ is the aggregative access beamforming vector of the i-th user. Similarly, define $\mathbf{h}_{u_k} \triangleq [\mathbf{h}_{u_k}^{(b_1)\dagger}, \ldots, \mathbf{h}_{u_k}^{(b_N)\dagger}]^\dagger$ as the aggregative channel of the k-th user in the access link. Note that since $\mathbf{w}_{i,j} = \mathbf{0}_{L \times 1}$ if $c_{i,j} = 1$, for all $i \in \mathcal{K}$ and $j \in \mathcal{N}$, there holds $\sum_{j \in \mathcal{N}_i} \mathbf{h}_{u_k}^{(b_j)\dagger} \mathbf{w}_{i,j} = \mathbf{h}_{u_k}^\dagger \mathbf{w}_i$. Then $\mathbf{A}_{u_k} = \mathbb{E}_\Omega[\mathbf{h}_{u_k} \mathbf{h}_{u_k}^\dagger] \in \mathbb{C}^{NL \times NL}$ can be expressed as

$$\mathbf{A}_{u_k} = \begin{bmatrix} (\mathbf{A}_{u_k})_{1,1} & \cdots & (\mathbf{A}_{u_k})_{1,N} \\ (\mathbf{A}_{u_k})_{2,1} & \cdots & (\mathbf{A}_{u_k})_{2,N} \\ \vdots & \ddots & \vdots \\ (\mathbf{A}_{u_k})_{N,1} & \cdots & (\mathbf{A}_{u_k})_{N,N} \end{bmatrix}, \tag{46}$$

where $(\mathbf{A}_{u_k})_{i,j} \in \mathbb{C}^{L \times L}$ is the block matrix of $\mathbf{A}_{u_k}$ at the i-th row and j-th column, given by

$$(\mathbf{A}_{u_k})_{i,j} = \begin{cases} \mathbf{h}_{u_k}^{(b_i)} \mathbf{h}_{u_k}^{(b_j)\dagger}, & \text{if } c_{k,i} = c_{k,j} = 1, \\ \beta_{u_k}^{(b_i)} \mathbf{I}_{L \times L}, & \text{if } i = j, \ c_{k,i} = 0, \\ \mathbf{0}_{L \times L}, & \text{otherwise.} \end{cases} \tag{47}$$

It can be easily verified that $\mathbf{A}_{u_k}$ is a positive definite matrix. Thus, we obtain a deterministic lower-bound approximation of $\bar{R}_k^{\text{A}}(\mathbf{x})$ as

$$
\tilde{R}_k^{\text{A}}(\mathbf{x}) = \log\left(1 + \frac{|\sum_{j\in\mathcal{N}_k}\mathbf{h}_{u_k}^{(b_j)\dagger}\mathbf{w}_{k,j}|^2}{\bar{\bar{\Phi}}_k + \sigma_{u_k}^2}\right)
$$

$$
= \log\left(1 + \frac{|\sum_{j\in\mathcal{N}_k}\mathbf{h}_{u_k}^{(b_j)\dagger}\mathbf{w}_{k,j}|^2}{\sum_{i=1}^K|\mathbf{h}_{u_k}^{(b_0)\dagger}\mathbf{v}_i|^2 + \sum_{i\neq k}^K \mathbf{w}_i^\dagger \mathbf{A}_{u_k}\mathbf{w}_i + \sigma_{u_k}^2}\right). \tag{48}
$$

Instead of solving the stochastic optimization problem $\mathcal{P}_{\text{S}}(\mathbf{c})$ in (35), we now solve the following deterministic optimization problem

$$
\mathcal{P}_{\text{D}}(\mathbf{c}) : \max_{\mathbf{x}\in\mathcal{X}_{\mathbf{c}}} \sum_{k=1}^K \omega_k \tilde{R}_k(\mathbf{x}), \tag{49}
$$

where

$$
\tilde{R}_k(\mathbf{x}) = \min\left\{\tilde{R}_k^{\text{A}}(\mathbf{x}), \min_{n\in\mathcal{N}_k}\left\{R_{k,n}^{\text{B}}(\mathbf{x})\right\}\right\}. \tag{50}
$$

Note that similar to problem $\mathcal{P}(\mathbf{c})$ in (8), problem $\mathcal{P}_{\text{D}}(\mathbf{c})$ can be efficiently solved via the SLBM algorithm. A concave lower-bound approximation of the rate $\tilde{R}_k^{\text{A}}(\mathbf{x})$ can be obtained as in (24) just by replacing Φ_k with $\bar{\bar{\Phi}}_k$. The overall procedure is similar to Alg. 1, which is ignored for simplicity.

V. SIMULATION RESULTS

In this section, we provide numerical simulations to demonstrate the effectiveness of the proposed algorithms. We consider a wireless network covering a square region of 1 km × 1 km, as shown in Fig. 2. The region is divided into a 3-by-3 grid, where one MBS, equipped with $M = 32$ antennas, is located at the center of the region, and $N = 8$ SBSs, each equipped with $L = 2$ antennas for the access link, is located at the center of the rest 8 small squares. The mobile users served by the SBSs are randomly and uniformly distributed within this region, excluding an inner circle of 250 m (50 m) around the MBS (each SBS). In each scheduling interval, $K = 3$ users are scheduled. All the users are scheduled in a round-robin manner. The parameter settings are mainly from [31] and summarized in Table II. In particular, the non-line-of-sight (NLOS) path loss models are adopted for the access channel and the line-of-sight (LOS) path loss models are adopted for the backhaul channel. The iteration of the SINRC-SLBM algorithm in Alg. 1 stops when the relative increase of the objective value is less than 10^{-3} or when a maximum of 30 iterations is reached. For the SINRC-SSLBM algorithm in Alg. 3, the iteration stops when a maximum of 300 iterations is reached. For simplicity, the weights $\{\omega_k\}_{k=1}^K$ in the objective function are set to 1 for all users, and all SBSs have the same peak power, i.e., $P_n^{\text{S}} = P^{\text{S}}$, for all $n \in \mathcal{N}$. The SI cancellation capability β_{SI} is set to 100 dB according to [4]. The convex subproblems of the proposed algorithms, e.g., (32) and (42) are solved using the CVX package via interior-point solver SDPT3. The experiments are carried out on a Windows x64 machine with 3.3 GHz CPU and 24 GB of RAM. All the plots are obtained by averaging over 100 channel realizations, if not specified otherwise.

A. Convergence Behavior of the Proposed Algorithms

In this subsection, we first demonstrate the convergence behavior of the SINRC-SLBM algorithm in Alg. 1 for solving problem $\mathcal{P}(\mathbf{c})$. We consider a static SBS clustering scheme, where each user u_k is cooperatively served by the cluster of SBSs $\mathcal{N}_k$ that have the largest large-scale fading coefficients to the user with a fixed cluster size $|\mathcal{N}_k| = C \in \{1, 4, 8\}$. The WMMSE algorithm in [23], [24] can be generalized to serve as a benchmark, where $\mathcal{P}(\mathbf{c})$ is solved via the SLBM approach using the WMMSE-based lower-bound approximation (14), denoted as WMMSE-SLBM. We generate a problem instance and
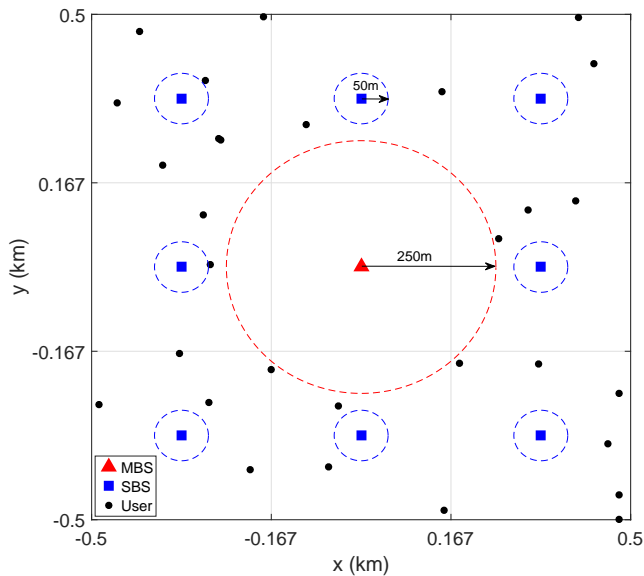
Fig. 2. Network topology of a wireless network with one MBS, 8 SBSs, 30 mobile users.

TABLE II
PARAMETER SETTINGS

Parameter	Value
Bandwidth	10 MHz
Macro (small) BS antenna gain	15 dBi (5 dBi)
Path loss from MBS to user	$128.1 + 37.6 \log_{10}(d)$ dB
Path loss from SBS to user	$140.7 + 36.7 \log_{10}(d)$ dB
Path loss from MBS to SBS	$103.4 + 24.2 \log_{10}(d)$ dB
Path loss from SBS to SBS	$103.8 + 20.9 \log_{10}(d)$ dB
Macro (small) cell shadowing std. dev.	8 dB (10 dB)
Small-scale fading	$\mathcal{CN}(\mathbf{0}, \mathbf{1})$
Noise power spectral density $\sigma_{u_k}^2$ $(\sigma_{b_n}^2)$	-174 dBm/Hz

TABLE III
COMPARISON OF THE AVERAGE PER-ITERATION SIMULATION TIME (SECONDS)

	$C = 1$	$C = 4$	$C = 8$
SINRC-SLBM	2.08	3.51	4.12
WMMSE-SLBM	1.75	3.16	3.91

solve it using both the SINRC-SLBM and WMMSE-SLBM algorithms. The maximum transmit power of the MBS and the SBSs are set as $P^{\mathrm{M}} = 40$ dBm and $P^{\mathrm{S}} = 30$ dBm, respectively. Fig. 3 shows the sum rate performance achieved by both algorithms with different cluster sizes. The average per-iteration simulation running time is shown in Table III. From Fig. 3, it is seen that the WMMSE-SLBM algorithm requires about 100 iterations to achieve a good convergence accuracy. In contrast, the proposed SINRC-SLBM algorithm can converge much faster, all within 10 iterations. It is also seen that the proposed SINRC-SLBM algorithm can converge to the solutions with higher sum rates than those of the WMMSE-SLBM algorithm. Although the per-iteration complexity of the proposed SINRC-SLBM algorithm is a little higher than the WMMSE-SLBM algorithm, as shown in Table III, the number of iterations required to converge is 10 times smaller. Therefore, the proposed SINRC-SLBM algorithm is more efficient than the WMMSE-SLBM algorithm in terms of both complexity and rate performance, indicating that the proposed SINR-convexification based lower-bound approximation is superior to the WMMSE-based lower-bound approximation.
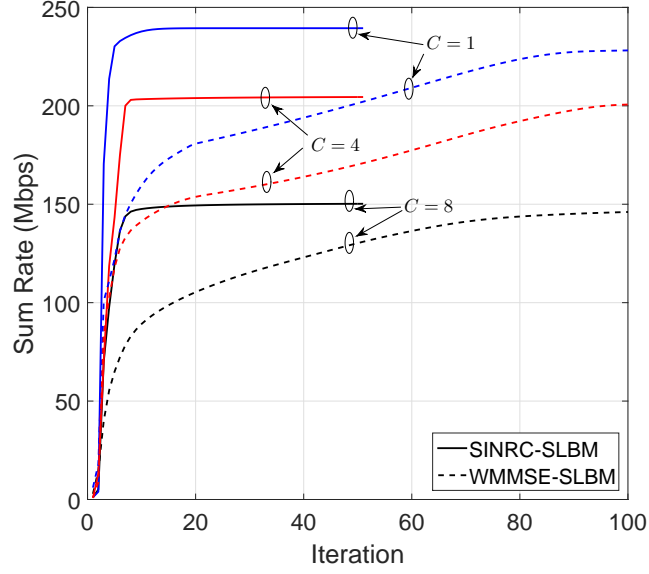
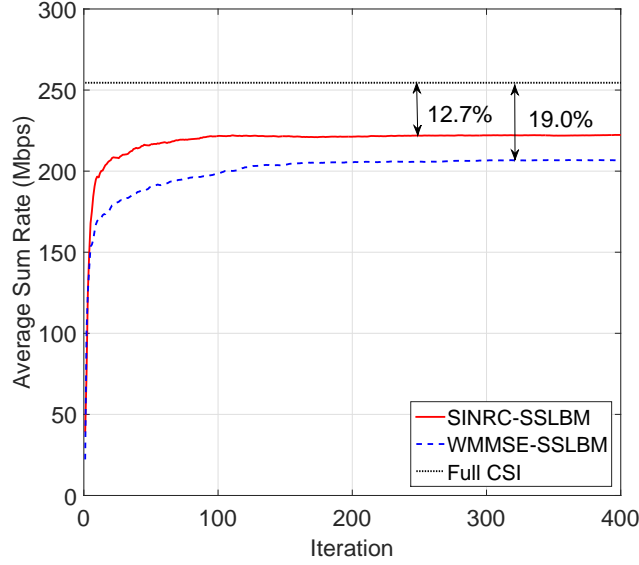Fig. 3. Convergence behavior of the proposed SINRC-SLBM algorithm.

Fig. 4. Convergence behavior of the proposed SINRC-SSLBM algorithm.

We also demonstrate the convergence behavior of the SINRC-SSLBM algorithm in Alg. 3 for solving problem $\mathcal{P}_S(\mathbf{c})$. The stochastic SLBM algorithm proposed in [29] using the WMMSE-based lower-bound approximation can serve as a benchmark, denoted as WMMSE-SSLBM. Fig. 4 shows the average sum rate achieved by both algorithms with cluster size $C = 3$. The average sum rate in each iteration is approximated by the sample mean of 500 independent channel realizations. As can be seen from Fig. 4, both stochastic algorithms can converge within 200 iterations. However, the proposed SINRC-SSLBM algorithm converges to a solution with better performance than the WMMSE-SSLBM algorithm.

B. Effectiveness of the Proposed Algorithms with Full CSI

We first demonstrate the rate performance of the proposed SINRC-SLBM algorithm under the static clustering scheme in Fig. 5. Besides the WMMSE-SLBM, the RCG algorithm proposed in our prior conference paper [1] is also considered for comparison. From Fig. 5, it is clearly seen that the proposed
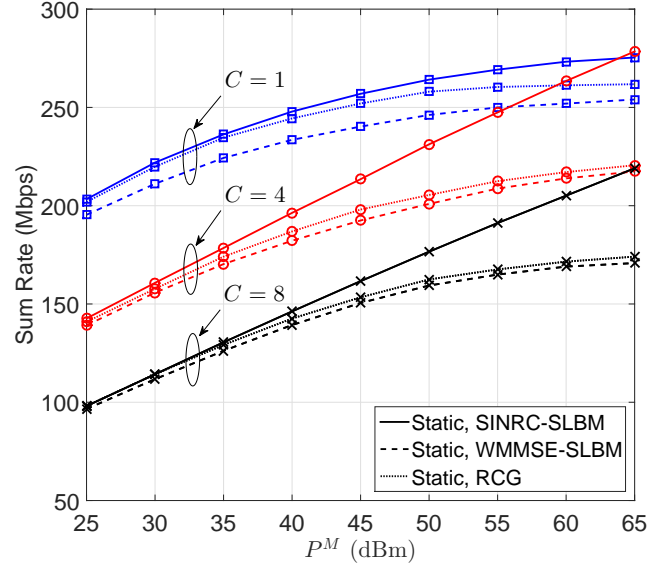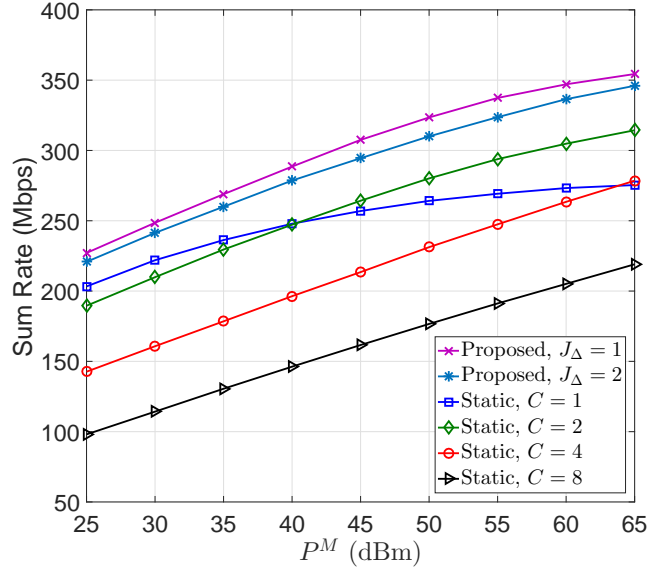
Fig. 5. Sum rate versus P^{M} with $P^{\text{S}} = 30$ dBm.

Fig. 6. Sum rate versus P^{M} with $P^{\text{S}} = 30$ dBm.

SINRC-SLBM algorithm achieves better performance than the WMMSE-SLBM algorithm as well as the RCG algorithm for all $C \in \{1, 4, 8\}$. Especially when the peak power of the MBS P^{M} goes large, the performance gap will increase. This is probably because both the WMMSE-SLBM algorithm and the RCG algorithm are more likely to get stuck in unfavorite local points when P^{M} is too large, while the proposed SINRC-SLBM algorithm can avoid this high-power issue.

We also demonstrate the performance of the heuristic SBS clustering scheme in Fig. 6. For comparison purpose, we consider the static SBS clustering scheme as a benchmark. From Fig. 6, it can be seen that the proposed heuristic SBS clustering scheme is superior to the static scheme for all $C \in \{1, 2, 4, 8\}$. We also see that among all the considered static SBS clustering schemes, the best cluster size C varies at different values of P^{M}. This is due to the fact that higher transmit power in the backhaul link allows more SBSs to cooperate in the access link to increase the end-to-end rate of the two-hop transmission.

Fig. 7 illustrates the achievable sum rate performance obtained by the proposed algorithm when both
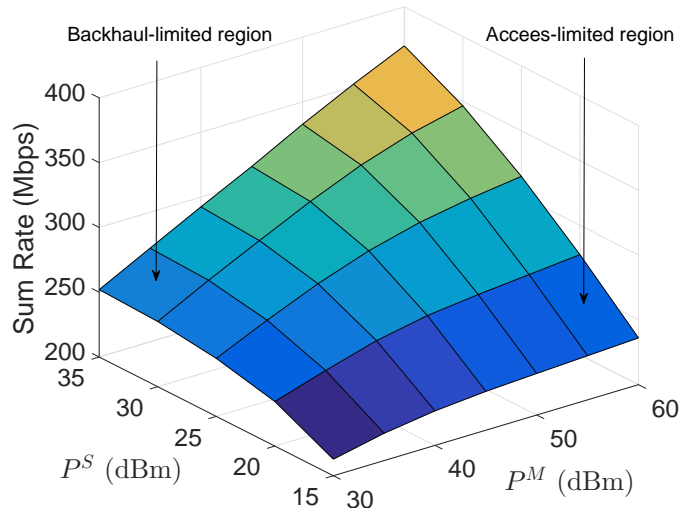
Fig. 7. Sum rate versus P^{M} and P^{S}.

TABLE IV
SUM RATE COMPARISON WITH PARTIAL CSI.

Cluster size	$C = 2$				$C = 3$				$C = 4$			
P^{M} (dBm)	30	40	50	60	30	40	50	60	30	40	50	60
SINRC-SSLBM (%)	81.2	74.4	70.0	67.5	94.3	87.3	84.1	81.2	98.5	98.1	97.9	97.0
DLB-SLBM (%)	78.4	72.4	68.0	65.5	92.4	87.2	83.7	80.6	98.1	97.7	97.6	96.8
WMMSE-SSLBM (%)	78.3	69.2	63.0	60.7	86.0	81.0	74.1	68.3	89.1	87.0	82.8	79.4
SAA-SLBM (%)	82.4	75.8	71.9	69.3	96.1	89.5	85.7	83.2	98.5	98.2	98.1	97.8
Full CSI (Mbps)	238.1	282.5	314.3	326.3	208.3	254.5	295.2	323.0	180.3	214.6	249.8	283.5

P^{M} and P^{S} vary. It is observed that when P^{S} is small (≤ 25 dBm), the sum rate increases slowly when P^{M} increases. This means that the system performance is limited by the access link. Similarly, when P^{M} is small (≤ 40 dBm), the sum rate increases slowly with P^{S}. This indicates that the system is backhaul-limited. The above observation is expected, since for the two-hop transmission, the end-to-end available rate of a user depends on the minimum rate between the access link and the backhaul link.

C. Effectiveness of the Proposed Algorithms with Partial CSI

Finally, we demonstrate the performance of the two proposed algorithms for partial CSI, namely, the SINRC-SSLBM in Alg. 3 and the low-complexity algorithm via deterministic lower-bound approximation in Section IV-C, denoted as DLB-SLBM. We compare them with the following two benchmark schemes:

- **WMMSE-SSLBM**: The stochastic WMMSE algorithm proposed in [29], which essentially solves problem $\mathcal{P}_{\mathrm{S}}(\mathbf{c})$ via the stochastic SLBM algorithm using the WMMSE-based lower-bound approximation.
- **SAA-SLBM**: The SAA method, in which the stochastic objective function is approximated by an ensemble average of 200 channel realizations and the resulting deterministic problem is solved by the deterministic SLBM algorithm.

The performance is measured by the percentage of the sum rate achieved by the considered algorithms with partial CSI over the sum rate achieved by the SINRC-SLBM algorithm with full CSI. The results are shown in Table IV, where each result is obtained by averaging over 500 independent channel realizations. The static SBS clustering scheme is adopted with a fixed cluster size C. The peak power of the SBSs is set as $P^{\mathrm{S}} = 30$ dBm. From Table. IV, it can be seen that the proposed SINRC-SSLBM algorithm

achieves almost the same performance as SAA-SLBM with lower computational complexity. Moreover, it is much better than WMMSE-SSLBM. It is also seen that the proposed DLB-SLBM algorithm can achieve a good performance that is very close to SAA-SLBM with much lower computational complexity. This clearly demonstrates the effectiveness of the proposed deterministic lower-bound approximation for the average achievable rate. From Table. IV, we also see that when $C = 4$, with a moderate amount of CSI, the proposed algorithms can maintain good performance that is very close to the full CSI case. Even when $C = 3$, the proposed algorithms can achieve performance within $80\% \sim 94\%$ of that with full CSI for varieties of P^{M}. Therefore, the channel estimation overhead can be significantly reduced with little performance degradation by considering the transmission design with partial CSI.

VI. CONCLUSION

This paper presents a user-centric joint access-backhaul transmission framework for full-duplex self-backhauled wireless networks by exploiting the inter-site cooperation. We formulate a weighted sum rate maximization problem for joint design of multicast beamforming in the backhaul link, SBS clustering and beamforming in the access link. This problem is solved via the SLBM approach with the newly introduced SINR-convexification based lower-bound approximation and the heuristic iterative link removal technique. We also study the stochastic joint access-backhaul beamforming problem with partial CSI. We develop a stochastic SLBM algorithm and a low-complexity deterministic algorithm to tackle this problem, respectively. Simulation results demonstrate the effectiveness of the proposed algorithms for both full CSI and partial CSI scenarios. They also show that with a moderate amount of CSI, the proposed algorithms can achieve good performance close to that with full CSI. By considering the joint transmission design with partial CSI, the proposed algorithms can significantly reduce the CSI overhead without losing too much of the performance.

REFERENCES

[1] E. Chen, M. Tao, and N. Zhang, "User-centric joint access-backhaul design for full-duplex self-backhauled cooperative networks," in *to appear Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018.

[2] Y. Yuan and X. Zhao, "5G: Vision, scenarios and enabling technologies," *ZTE Communications*, vol. 14, no. S1, pp. 55–60, Dec. 2015.

[3] X. Ge, H. Cheng, M. Guizani, and T. Han, "5G wireless backhaul networks: Challenges and research advances," *IEEE Network*, vol. 28, no. 6, pp. 6–11, Nov. 2014.

[4] R. A. Pitaval, O. Tirkkonen, R. Wichman, K. Pajukoski, E. Lahetkangas, and E. Tiirola, "Full-duplex self-backhauling for small-cell 5G networks," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 83–89, Oct. 2015.

[5] 3GPP TR 38.874, "NR; Study on integrated access and backhaul," Tech. Rep., 2018.

[6] A. Sharma, R. K. Ganti, and J. K. Milleth, "Joint backhaul-access analysis of full duplex self-backhauling heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1727–1740, Mar. 2017.

[7] L. Chen, F. R. Yu, H. Ji, B. Rong, X. Li, and V. C. M. Leung, "Green full-duplex self-backhaul and energy harvesting small cell networks with massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3709–3724, Dec. 2016.

[8] T. K. Vu, M. Bennis, S. Samarakoon, M. Debbah, and M. Latva-aho, "Joint in-band backhauling and interference mitigation in 5G heterogeneous networks," in *Proc. 22nd Eur. Wireless Conf.*, May 2016, pp. 1–6.

[9] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[10] W. Li, Y. Zhang, and L.-K. Huang, "Interference-cancellation scheme for multilayer cellular systems," *ZTE Communications*, vol. 13, no. 1, pp. 43–49, Mar. 2015.

[11] H. Zhuang, J. Chen, and D. O. Wu, "Joint access and backhaul resource management for ultra-dense networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[12] B. Hu, C. Hua, J. Zhang, C. Chen, and X. Guan, "Joint fronthaul multicast beamforming and user-centric clustering in downlink C-RANs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5395–5409, Aug. 2017.

[13] B. Hu, C. Hua, C. Chen, and X. Guan, "Joint beamformer design for wireless fronthaul and access links in C-RANs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2869–2881, May 2018.

[14] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C. L. I, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[15] G. Wang, Q. Liu, R. He, F. Gao, and C. Tellambura, "Acquisition of channel state information in heterogeneous cloud radio access networks: Challenges and research directions," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 100–107, Jun. 2015.

[16] C. Pan, M. Elkashlan, J. Wang, J. Yuan, and L. Hanzo, "User-centric C-RAN architecture for ultra-dense 5G networks: Challenges and methodologies," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 14–20, Jun. 2018.

[17] Y. Shi, J. Zhang, and K. B. Letaief, "CSI overhead reduction with stochastic beamforming for cloud radio access networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 5154–5159.

[18] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton, NJ: Princeton University Press, 2008.

[19] S. Burer and A. N. Letchford, "Non-convex mixed-integer nonlinear programming: A survey," *Surv. Oper. Res. Manage. Sci.*, vol. 17, no. 2, pp. 97–106, Jul. 2012.

[20] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.

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

[22] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Comput.*, vol. 15, no. 4, pp. 915–936, 2003.

[23] Q. Shi, M. Razaviyayn, Z. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[24] B. Dai and W. Yu, "Backhaul-aware multicell beamforming for downlink cloud radio access network," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 2689–2694.

[25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univercity Press, 2004.

[26] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, no. 1, pp. 193–228, Nov. 1998.

[27] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[28] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory, Second Edition*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2014.

[29] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, "A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver design in wireless communication networks," *Math. Program.*, vol. 157, no. 2, pp. 515–545, Jun. 2016.

[30] C. Pan, H. Zhu, N. J. Gomes, and J. Wang, "Joint user selection and energy minimization for ultra-dense multi-channel C-RAN with incomplete CSI," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 8, pp. 1809–1824, Aug. 2017.

[31] 3GPP TR 36.814 V9.0.0, "Further advancements for E-UTRA physical layer aspects," Tech. Rep., 2010.