Interference Management of Analog Function Computation in Multi-Cluster Networks

Haoyu Zhang, Li Chen, Nan Zhao, *Senior Member, IEEE*,
Yunfei Chen, *Senior Member, IEEE*, and F. Richard Yu, *Fellow, IEEE*

***Abstract*—Computation over multiple access channels (CoMAC) has been proposed to solve the problem of spectrum scarcity in wireless networks, which combines communication and computation efficiently using the superposition property of wireless channels. In this paper, we consider a multi-cluster CoMAC network, whose performance is affected by the inter-cluster interference and the non-uniform fading. To minimize the sum mean squared error of signals aggregated at different fusion centers (FCs), we propose a transceiver design for multi-cluster CoMAC. Specifically, we adopt a uniform-forcing transmitter design to formulate the receiver design as a quadratic sum-of-ratios problem with nonconvex quadratic constraints. Then, we propose a branch-and-bound algorithm to find its optimal solution with a given error tolerance. To solve the problem in a decentralized way, we develop a distributed algorithm based on the primal decomposition theory. Each subproblem is solved by using the successive convex approximation method. Further combining Lagrange duality, we derive the optimal solution structure of each subproblem, based on which we can find the solution with lower complexity. Simulation results demonstrate the effectiveness of the proposed distributed transceiver design.**

***Index Terms*—Computation, interference management, multiple access channel, multiple cluster, transceiver design**

I. Introduction

FUTURE wireless networks need to support the ubiquitous deployment of massive nodes [1]. How to quickly collect and process data from these dense nodes with limited spectrum is a challenging issue. The traditional "transmit-then-compute" scheme based on orthogonal multiple access will result in excessive latency and low efficiency in spectrum utilization. Since future wireless networks are more interested in the fusion of massive data rather than the individual data, a seminal scheme, called *computation over multiple access channels* (CoMAC), was proposed to achieve ultrafast data aggregation [2]. It exploits the signal-superposition property of *multiple access channel* (MAC) to compute the desired functions of data at distributed nodes. These desired functions are called nomographic functions, which can be represented as a post-processed sum of pre-processed node readings [3].

The idea of CoMAC originates from a pioneer work on function computation [2]. Instead of fighting the inter-node interference caused by concurrent transmissions [4], it harnesses interference to facilitate computation by designing a structured code. As the number of nodes grows, the coding scheme becomes degenerated. To maintain the superior abilities the coding provides, a user scheduling paradigm was presented in [5] to search good nodes for CoMAC. Coding has been extensively applied in many cases for CoMAC, such as the case of bivariate Gaussian source [6] and correlated Gaussian source [7]. While considering independent Gaussian sources, uncoded analog CoMAC, where transmitted signals are merely scaled from the sensing data, can outperform CoMAC with coding [8]. Many follow-up works on the analog CoMAC have been proposed due to its low complexity and high energy efficiency. One vein of this research focuses on the theoretical properties and the pre-processing and post-processing functions design [9]–[11].

It is worth mentioning that the above research on analog CoMAC does not consider the fading property of practical MAC. CoMAC requires signal alignment, which refers to aligning the received magnitudes of multiuser signals. Different fading at distributed nodes will bring challenges to signal alignment, which can be coped with by the transceiver design. Transceivers for single-function CoMAC were investigated in [12] and [13], where each node is equipped with a single antenna to support uni-modal sensing. In [12], a transmission scheme of channel inversion was adopted to combat the non-uniform fading of different nodes. Then, the computation-optimal transmitting-receiving policy was derived in [13]. Specifically, nodes with the smallest channel coefficients transmitted signals with full power, while others were of a channel-inversion type. These works were all about CoMAC of scalar-valued functions. Future wireless networks equipped with large-scale arrays will make it possible for CoMAC to support vector-valued function computation. Such studies have enabled the *fusion center* (FC) to compute multiple functions simultaneously to realize multi-modal sensing [14]. Besides the non-uniform fading, the implementation of CoMAC faces other practical challenges, such as massive *channel state information* (CSI) requirement for the FC. When the number of nodes is large, traditional individual CSI acquisition schemes will incur large latency [15], [16]. To tackle this challenge, several novel signaling procedures were proposed to reduce

This research was supported by National Key R&D Program of China (Grant No. 2021YFB2900302), and National Natural Science Foundation of China (Grant No. 61601432). (*Corresponding author: Li Chen.*)

Haoyu Zhang and Li Chen are with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China (e-mail: hyzhangY@mail.ustc.edu.cn; chenli87@ustc.edu.cn).

Nan Zhao is with the School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China (e-mail: zhaonan@dlut.edu.cn).

Yunfei Chen is with the School of Engineering, University of Warwick, Coventry CV4 7AL, U.K. (e-mail: yunfei.chen@warwick.ac.uk).

F. Richard Yu is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada (e-mail: richard.yu@carleton.ca).

the latency induced by massive CSI acquisition [17], [18]. Subsequently, the authors in [19] presented an automatic repeat request based CoMAC scheme to avoid massive CSI aggregation. To further enhance the performance of CoMAC, an advanced wireless technology called intelligent reflecting surface was applied to improve poor channel conditions by reconfiguring its propagation environment [20], [21].

Analog CoMAC has recently found a new application in distributed machine learning. Future wireless networks will be required to support ubiquitous artificial intelligence services from the core to the end devices of the networks, which will result in heavy communication overheads [22], [23]. CoMAC can be applied in this area as a communication-efficient technique, e.g., transceivers were proposed to enable each node to concurrently send its gradient to the server via an MAC, based on which fast global model aggregation was realized at the server [24], [25].

From the above discussion, it is noted that the transceiver designs are only developed for the single-cluster CoMAC network so far. In order to realize ubiquitous coverage, practical wireless networks often contain multiple clusters [26], each with a FC to support the fast data aggregation. Thereby, CoMAC for a multi-cluster network is of great interest, where each FC only serves the nodes in its cluster to complete its CoMAC task. Simultaneous tasks in different clusters will make FCs suffer from inter-cluster interference. Compared with conventional multi-cluster networks [27], [28], multi-cluster CoMAC concerns more on the computed functions of simultaneous data streams, which leads to different inter-cluster interference structure. Except for interference, the performance of multi-cluster CoMAC is also affected by signal misalignment errors. How to manage inter-cluster interference while suppressing signal misalignment errors has been studied in very few works [29], [30]. Only considering a two-cluster network with high transmit *signal-to-noise ratio* (SNR), a scheme called signal-and-interference alignment was proposed to simultaneously eliminate inter-cluster interference and realize signals alignment [29]. With given single-antenna receiver, Cao *et al.* in [30] investigated the optimal transmit power policies of each node by utilizing a set of *interference temperature* (IT) constraints. However, for a general multi-cluster CoMAC network, how to design its corresponding transceivers has never been studied before.

Motivated by the above observation, we consider a general multi-cluster CoMAC network and study how to reduce the *mean squared errors* (MSEs) between the target function values and the computed ones at different FCs. Assume that there exists a multi-antenna FC in each cluster to aggregate data streams from nodes it serves. Due to the path loss and the low transmit power of nodes, a node may be connected to more than one FC. The concurrent transmission of all nodes makes each FC exposed to the inter-cluster interference. To tackle this issue, we present a transceiver design to suppress the computation errors induced by signal misalignment, interference and noise. A sum-MSE minimization problem is formulated under per-node power constraints for the joint design of receivers at FCs. A global optimal scheme with a given relative error tolerance ε is proposed to solve it. In addition, to reduce the computational complexity, a suboptimal algorithm is presented to find a high-quality approximate solution. The main contributions of this paper are summarized as follows.

- *Multi-Cluster CoMAC Problem Formulation:* We formulate a *quadratic sum-of-ratios problem* (QSRP) with non-convex quadratic constraints to design new transceivers for multi-cluster CoMAC. The objective function is composed of inter-cluster interference and noise, in which receivers are coupled. Note that this problem is challenging due to the non-convexity of both the objective function and the constraints.
- *ε-Optimal Branch-and-Bound Algorithm:* We design a *branch-and-bound* (BB) scheme to find the ε-optimal solution to the formulated problem. By using the argument cut based relaxation techniques in [31], we formulate a convex relaxation of the QSRP to compute the lower bound of the global optimum. Combined with "branch", "upper bound" and other procedures shown later on, the global optimal solution can be obtained with guaranteed convergence. The proposed BB scheme provides an important benchmark for measuring the performance of suboptimal algorithms for the same problem.
- *Distributed Low-Complexity Algorithm:* Since each FC cannot tolerate the inherent high complexity of the BB algorithm and only has local CSI in the practical implementation, we also propose a distributed low-complexity algorithm based on the primal decomposition theory [32]. More specifically, we introduce slack variables to constrain the transmit power of nodes that interfere with each FC. With given slack variables, the original problem can be decomposed into several parallel subproblems. Each FC can use the *successive convex approximation* (SCA) method to independently solve its own subproblem. Then, the slack variables can be updated based on the master problem. By solving the master problem and subproblems iteratively, the MSE of each cluster can be reduced parallelly until the algorithm converges. When the number of antennas is sufficiently large, we also provide an asymptotically optimal beamformer design in closed-form for the formulated QSRP.

The remainder of the paper is organized as follows. Section II presents the system model and the problem formulation for multi-cluster CoMAC. Section III introduces the proposed centralized global optimal BB algorithm. A distributed algorithm is developed in Section IV. Section V provides the simulation results. Concluding remarks are given in Section VI.

Notations: $\mathbb{R}$ and $\mathbb{C}$ denote the real and complex spaces, respectively. Bold lowercase and bold uppercase letters denote column vectors and matrices, respectively. The operators $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$ correspond to the transpose, conjugate and Hermitian transpose, respectively. $\Re\{x\}$ and $\Im(x)$ refer to the real and imaginary parts of a complex number x, respectively. $\|\mathbf{a}\|$ denotes the Euclidean norm of the vector $\mathbf{a}$. $|\mathcal{C}|$ denotes the number of elements in a set $\mathcal{C}$. Finally, $\mathbf{0}_L$ and $\mathbf{1}_L$ denote the all-zero vector of dimension L and the all-one vector of dimension L, respectively.

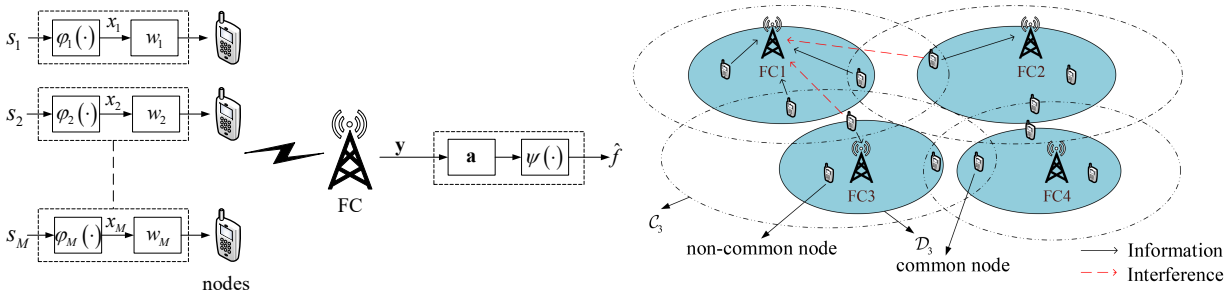


Fig. 1. CoMAC system model. (a) Single-cluster CoMAC. (b) Multi-cluster CoMAC.

II. System Model and Problem Formulation

In this section, we first discuss the transceiver design for single-cluster CoMAC and then extend it to multi-cluster CoMAC networks to combat the inter-cluster interference and non-uniform fading.

A. Single-Cluster CoMAC

As illustrated in Fig. 1a, there are M nodes and a FC in the network. Each node is equipped with a single antenna and the FC is equipped with N_r antennas. The data of the m-th node is $s_m \in \mathbb{R}$. The FC aims at computing the desired function based on the signals transmitted by nodes, which can be expressed as

$$f = \psi\Big[\sum_{m=1}^{M} \varphi_m(s_m)\Big], \tag{1}$$

where $\varphi_m(\cdot) : \mathbb{R} \to \mathbb{R}$ is the pre-processing function of the m-th data and $\psi(\cdot) : \mathbb{R} \to \mathbb{R}$ is the post-processing function of the FC. By choosing specific $\varphi_m(\cdot)$ and $\psi(\cdot)$, some common functions (e.g., arithmetic mean function and polynomial function) can be computed.

The pre-processed signal at the m-th node is $x_m = \varphi_m(s_m)$ with $x_m \in \mathbb{R}$, and the received signal of the FC after concurrent transmissions of all nodes is

$$\mathbf{y} = \sum_{m=1}^{M} \mathbf{h}_m w_m x_m + \mathbf{z}, \tag{2}$$

where $w_m \in \mathbb{C}$ is the transmitter scalar of the m-th data, $\mathbf{h}_m \in \mathbb{C}^{N_r}$ is the wireless channel vector and $\mathbf{z} \in \mathbb{C}^{N_r}$ represents the noise vector with each element distributed as $\mathcal{CN}(0, \sigma_n^2)$. The pre-processed signals are assumed to be normalized to have unit variance, i.e., $\mathbb{E}\{x_m^2\} = 1, m = 1, 2, \ldots, M$, and satisfy $\mathbb{E}\{x_m x_n\} = 0, \forall m \neq n$ and $\mathbb{E}\{x_m z_l\} = 0, l = 1, 2, \ldots, N_r$, where z_l is the l-th element of $\mathbf{z}$. The desired value at the FC is the summation part of the desired function in (1), i.e., $x = \sum_{m=1}^{M} \varphi_m(s_m)$. The estimated value of x, denoted as $\hat{x}$, is given by

$$\hat{x} = \frac{1}{\sqrt{\eta}} \mathbf{a}^H \mathbf{y} = \frac{1}{\sqrt{\eta}} \sum_{m=1}^{M} \mathbf{a}^H \mathbf{h}_m w_m x_m + \frac{\mathbf{a}^H \mathbf{z}}{\sqrt{\eta}}, \tag{3}$$

where $\mathbf{a} \in \mathbb{C}^{N_r}$ is the receiver and η is a normalizing factor. Compared with x, the distortion of $\hat{x}$ is given by

$$\begin{aligned} \mathrm{MSE}(\hat{x}, x) &= \mathbb{E}\big[|\hat{x} - x|^2\big] \\ &= \sum_{m=1}^{M} \left\| \frac{\mathbf{a}^H \mathbf{h}_m w_m}{\sqrt{\eta}} - 1 \right\|^2 + \frac{\sigma_n^2 \|\mathbf{a}\|^2}{\eta}, \end{aligned} \tag{4}$$

which can be used to measure the performance of the single-cluster CoMAC[1].

The non-uniform fading of different nodes may result in unacceptable signal misalignment error in (4). In order to combat it, the novel uniform-forcing transceiver design in [12] is given as follows.

Uniform-forcing transmitter: Fixing the beamforming vector $\mathbf{a}$, each transmitter can be computed as

$$w_m = \sqrt{\eta}\big(\mathbf{a}^H \mathbf{h}_m\big)^{-1}, \forall m. \tag{5}$$

Considering each node's power constraints $|w_m|^2 \leq P_t, \forall m$, η can be computed as

$$\eta = P_t \min_m \left\| \mathbf{a}^H \mathbf{h}_m \right\|^2. \tag{6}$$

Single-cluster receiver: Fixing transmitters $w_m, \forall m$, the beamforming vector $\mathbf{a}$ can be obtained by multiplying the solution to the following problem by a scaling factor.

$$\begin{aligned} \mathrm{P0}: \min_{\mathbf{a}} \ & \|\mathbf{a}\|^2 \\ \text{s.t.} \ & \left\| \mathbf{a}^H \mathbf{h}_m \right\|^2 \geq 1, \forall m. \end{aligned} \tag{7}$$

The above problem has the same form as the single-group multicast beamforming design. It is NP-hard due to its nonconvex quadratic constraints. An approximate solution to it can be obtained by *semidefinite relaxation* (SDR) or SCA method [12], and its global optimal solution can be found by the BB scheme [33].

B. Multi-Cluster CoMAC Network

In this subsection, we extend the single-cluster CoMAC to a multi-cluster network with M spatially distributed nodes and

[1]The distortion of $\hat{f} = \psi(\hat{x})$ with respect to $f = \psi(x)$ is measured by $\mathrm{MSE}(\hat{f}, f)$. According to Cauchy continuity and the continuity of the function $\psi(\cdot)$, if $\mathrm{MSE}(\hat{x}, x) \to 0$, $\mathrm{MSE}(\hat{f}, f) \to 0$ holds, so that $\mathrm{MSE}(\hat{x}, x)$ can replace $\mathrm{MSE}(\hat{f}, f)$ as the performance metric of single-cluster CoMAC.

K FCs. A quantitative example is shown in Fig. 1b. Each FC is equipped with N_r antennas and each node is equipped with a single antenna. The multi-cluster system model is characterized by the sets $\mathcal{C}_k, \mathcal{D}_k, \mathcal{E}_k, \mathcal{I}_k$ and $\mathcal{I}'_k, \forall k$, i.e.,

1) The set of nodes belonging to the k-th cluster is denoted by $\mathcal{C}_k$ and for each k there exists at least one $k' \neq k$ such that $\mathcal{C}_k \cap \mathcal{C}_{k'} \neq \varnothing$. The nodes belonging to both the k-th and the k'-th clusters (nodes in the overlaps $\mathcal{C}_k \cap \mathcal{C}_{k'}$) are called "common nodes", which will cause interference between clusters, while the interference generated by nodes $i \notin \mathcal{C}_k$ can be negligible at the k-th FC and is treated as noise.
2) The k-th FC takes the nodes in $\mathcal{C}_k$ into consideration, but only serves the nodes in $\mathcal{D}_k \subseteq \mathcal{C}_k$. The common nodes in $\mathcal{D}_k$ that cause interference to the l-th FC can be denoted as $\mathcal{C}_l \cap \mathcal{D}_k$.
3) $\mathcal{E}_k$ denotes the set of the non-common nodes in $\mathcal{D}_k$.
4) $\mathcal{I}_k$ denotes the set of clusters to which the nodes interfering with the k-th FC belong.
5) $\mathcal{I}'_k$ denotes the set of clusters interfered by nodes in $\mathcal{D}_k$.

The i-th node in $\mathcal{D}_k$ can be labeled as the (k,i)-th node with $i \in \{1, 2, ..., |\mathcal{D}_k|\}$ and $k \in \{1, 2, ..., K\}$. The reading of the (k,i)-th node is $s_{k,i}$. The desired function of the k-th FC can be written in the form as

$$f_k = \psi_k\Big[\sum_{i\in\mathcal{D}_k} \varphi_{k,i}\left(s_{k,i}\right)\Big], \tag{8}$$

where $\varphi_{k,i}(\cdot) : \mathbb{R} \to \mathbb{R}$ is the pre-processing function of the (k,i)-th node, and $\psi_k(\cdot) : \mathbb{R} \to \mathbb{R}$ is the post-processing function of the k-th FC.

Let $w_{k,i} \in \mathbb{C}$ be the transmitter scalar of the (k,i)-th node and $\mathbf{h}_{k,i,l} \in \mathbb{C}^{N_r}$ be the wireless channel vector from the (k,i)-th node to the l-th FC. Throughout the rest of the paper, we assume that $\mathbf{h}_{k,i,l} \sim \mathcal{CN}(\mathbf{0}, \kappa_{k,i,l}\mathbf{I}_{N_r})$ and is independent of $\mathbf{h}_{k',i',l}, \forall\big(k,i\big) \neq \big(k^{'},i^{'}\big)$. After concurrent transmissions of the pre-processed signals $x_{k,i} = \varphi_{k,i}\left(s_{k,i}\right), \forall k,i$, the received signal of the k-th FC, denoted as $\mathbf{y}_k$, is given as

$$\mathbf{y}_k = \sum_{i\in\mathcal{D}_k} \mathbf{h}_{k,i,k} w_{k,i} x_{k,i} + \sum_{l\in\mathcal{I}_k} \sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l} \mathbf{h}_{l,j,k} w_{l,j} x_{l,j} + \mathbf{z}_k, \tag{9}$$

where $\mathbf{z}_k \in \mathbb{C}^{N_r}$ is the noise vector with each element distributed as $\mathcal{CN}\left(0, \sigma_n^2\right)$. Each $x_{k,i}$ satisfies $\mathbb{E}\{x_{k,i}^2\} = 1, \forall i \in \mathcal{D}_k, \forall k$, $\mathbb{E}\{x_{k,i}x_{l,j}\} = 0, \forall\left(k,i\right) \neq \left(l,j\right)$ and $\mathbb{E}\left\{x_{k,i}z_{k,l}\right\} = 0$, l=1,2,...,N_r.

C. Problem Formulation

The received signal of the k-th FC is processed by the receiver $\mathbf{a}_k \in \mathbb{C}^{N_r}$. Then, the estimated value before post-processing at the k-th FC can be denoted as

$$\hat{x}_k = \underbrace{\sum_{i\in\mathcal{D}_k} x_{k,i}}_{\text{desired signal}} + \underbrace{\sum_{i\in\mathcal{D}_k}\left(\frac{1}{\sqrt{\eta_k}}\mathbf{a}_k^H\mathbf{h}_{k,i,k}w_{k,i} - 1\right)x_{k,i}}_{\text{signal misalignment error}} + \underbrace{\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\frac{1}{\sqrt{\eta_k}}\mathbf{a}_k^H\mathbf{h}_{l,j,k}w_{l,j}x_{l,j}}_{\text{inter-cluster interference}} + \underbrace{\frac{1}{\sqrt{\eta_k}}\mathbf{a}_k^H\mathbf{z}_k}_{\text{noise}} \tag{10}$$

where $\sqrt{\eta_k}$ is the normalizing factor of the k-th cluster. Similar to the single-cluster CoMAC, the mean squared error of $\hat{x}_k$ can be measured as $\mathbb{E}[\big|\hat{x}_k - \sum_{i\in\mathcal{D}_k} x_{k,i}\big|^2]$. Then, the performance of the multi-cluster CoMAC system is given by (11), as shown at top of the next page.

Following a similar idea of the transceiver design to the single-cluster network, the transceiver for multi-cluster CoMAC can be designed as follows.

Uniform-forcing transmitter: Fixing beamforming vectors $\mathbf{a}_k, \forall k$, each transmitter can be computed as

$$w_{k,i} = \sqrt{\eta_k}\big(\mathbf{a}_k^H\mathbf{h}_{k,i,k}\big)^{-1}, \forall i \in \mathcal{D}_k, \forall k. \tag{12}$$

The transmitters should satisfy $|w_{k,i}|^2 \leq P_t, \forall k,i$, and then η_k can be computed as

$$\eta_k = P_t \min_i \left\|\mathbf{a}_k^H\mathbf{h}_{k,i,k}\right\|^2, \forall i \in \mathcal{D}_k, \forall k. \tag{13}$$

Multi-cluster receiver: Fixing transmitters $w_{k,i}, \forall k,i$, the design of receivers is given in the following proposition.

Proposition 1 *(Receiver Design for Multi-cluster CoMAC):* The receiver of each FC can be obtained by solving the following problem

$$\begin{aligned} \text{P1}: \min_{\{\mathbf{a}_k\}} & \sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\frac{\left\|\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right\|^2}{\left\|\mathbf{a}_l^H\mathbf{h}_{l,j,l}\right\|^2} + \sum_{k=1}^{K}\frac{\sigma_n^2}{P_t}\|\mathbf{a}_k\|^2 \\ \text{s.t. } & \left\|\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\|^2 \geqslant 1, \forall j \in \mathcal{D}_k, \forall k. \end{aligned} \tag{14}$$

Proof: See Appendix A. ∎

Compared with P0 in (7), P1 is more complex due to the non-convexity of the objective function and constraints. In the following sections, we first find a global optimal value of P1 as the network performance benchmark. We then consider a practical scenario where each FC only has the CSI of the nodes within its own cluster, and find an approximate solution with polynomial complexity.

III. BB-Based Optimization Algorithm

In this section, we assume that each FC has global CSI and propose a centralized global optimization algorithm to solve P1 based on the BB scheme, which is often used to obtain the global optimal solutions for nonconvex problems.

A. Overview of the BB Algorithm

The BB algorithm is an enumeration algorithm, which continuously divides the feasible region into smaller subregions. The lower bound of the original problem over a given subregion can be obtained by solving the related convex relaxation problem over the same region. The minimum value among the lower bounds from all obtained subregions is selected as the current optimal lower bound. The current optimal upper bound can be obtained from some other local optimization/heuristic algorithms which can generate the best known feasible solution. A subregion is discarded if the lower bound on it is larger than the current optimal upper bound, and we call it as an inactive subregion. As the number of subregions in the partition increases, the gap between the upper bound and lower bound decreases. The algorithm terminates

$$\begin{aligned}\mathrm{MSE} &= \sum_{k=1}^{K}\mathbb{E}\left[\left|\hat{x}_k-\sum_{i\in\mathcal{D}_k}x_{k,i}\right|^2\right]\\ &= \sum_{k=1}^{K}\sum_{i\in\mathcal{D}_k}\left\|\frac{\mathbf{a}_k^H\mathbf{h}_{k,i,k}w_{k,i}}{\sqrt{\eta_k}}-1\right\|^2+\sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\frac{\left\|\mathbf{a}_k^H\mathbf{h}_{l,j,k}w_{l,j}\right\|^2}{\eta_k}+\sum_{k=1}^{K}\frac{\sigma_n^2\|\mathbf{a}_k\|^2}{\eta_k}.\end{aligned} \tag{11}$$

when the non-decreasing lower bounds and non-increasing upper bounds are close enough to each other. Constructing a tight convex relaxation of the original nonconvex problem is crucial, which determines the quality of the upper and lower bounds. With better lower and upper bounds, more inactive subregions can be deleted, and the algorithm will converge faster.

Recently, the BB algorithm has been applied in the beamforming design of wireless networks, e.g., the single-group multicast beamforming problem [31]. Inspired by the argument cut based relaxation in [31], we propose a novel BB method for globally solving P1.

B. Convex Relaxation Problem

In this subsection, we formulate a convex lower bounding function for the nonconvex objective function in P1 and relax the nonconvex constraints.

***Lemma 1** (Lower Bound of Objective Function):* The lower bounding function of the objective function f_0 in P1, denoted as Lf_0, can be expressed as the following quadratic function.

$$Lf_0=\sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}l_{l,j,l}\left\|\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right\|^2+\sum_{k=1}^{K}\frac{\sigma_n^2}{P_t}\|\mathbf{a}_k\|^2, \tag{15}$$

where $l_{l,j,l}$ and $u_{l,j,l}$ are positive scalars satisfying

$$\frac{1}{\sqrt{l_{l,j,l}}}\geq\left\|\mathbf{a}_l^H\mathbf{h}_{l,j,l}\right\|\geq\frac{1}{\sqrt{u_{l,j,l}}}\geq 1,\forall l,\forall j\in\mathcal{D}_l/\mathcal{E}_l. \tag{16}$$

Proof: It follows from (16) that

$$\frac{\left\|\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right\|^2}{\left\|\mathbf{a}_l^H\mathbf{h}_{l,j,l}\right\|^2}\geq l_{l,j,l}\left\|\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right\|^2>0, \tag{17}$$

and thus $Lf_0\leq f_0$ holds. ■

Before relaxing the nonconvex constraints, we first transform the constraints for non-common nodes into

$$\left\|\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\|\geqslant 1,\forall k,\forall j\in\mathcal{E}_k, \tag{18}$$

If the $(k,|\mathcal{D}_k|)$-th node is a non-common node, we can further convert $\left\|\mathbf{a}_k^H\mathbf{h}_{k,|\mathcal{D}_k|,k}\right\|\geq 1$ to $\mathbf{a}_k^H\mathbf{h}_{k,|\mathcal{D}_k|,k}\geq 1$ to reduce the number of nonconvex constraints. This is because that, for any $\mathbf{a}_k$ satisfying $\left\|\mathbf{a}_k^H\mathbf{h}_{k,|\mathcal{D}_k|,k}\right\|\geq 1$, there always exists a $\theta_k\in\mathbb{R}$ such that $\exp(i\theta_k)\,\mathbf{a}_k^H\mathbf{h}_{k,|\mathcal{D}_k|,k}\geq 1$. Treating $\exp(-i\theta_k)\,\mathbf{a}_k$ as a new optimization variable does not change the structure of the original objective function f_0. Next, we introduce the following lemma to relax the nonconvex constraints (16) and (18).

***Lemma 2** (Convex Relaxation of Constraints):* Define the argument of $\mathbf{a}_k^H\mathbf{h}_{k,j,k}$ as $\varphi_{k,j}$, where $\varphi_{k,j}\in[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]$, $0<\underline{\varphi}_{k,j}\leq\overline{\varphi}_{k,j}<2\pi$. Let $Y_{k,j}^{[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]}$ denote the set of $\mathbf{a}_k$ defined by the inequality (16) and $S_{k,j}^{[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]}$ denote the set of $\mathbf{a}_k$ defined by the inequality (18). Suppose that $\overline{\varphi}_{k,j}-\underline{\varphi}_{k,j}\leq\pi$, then the convex envelope of $S_{k,j}^{[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]}$ is given by (19), and the convex envelope of $Y_{k,j}^{[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]}$ is given by (20), as shown at top of the next page, where $a_{k,j}=(\cos(\underline{\varphi}_{k,j})+\cos(\overline{\varphi}_{k,j}))/2$ and $b_{k,j}=(\sin(\underline{\varphi}_{k,j})+\sin(\overline{\varphi}_{k,j}))/2$.

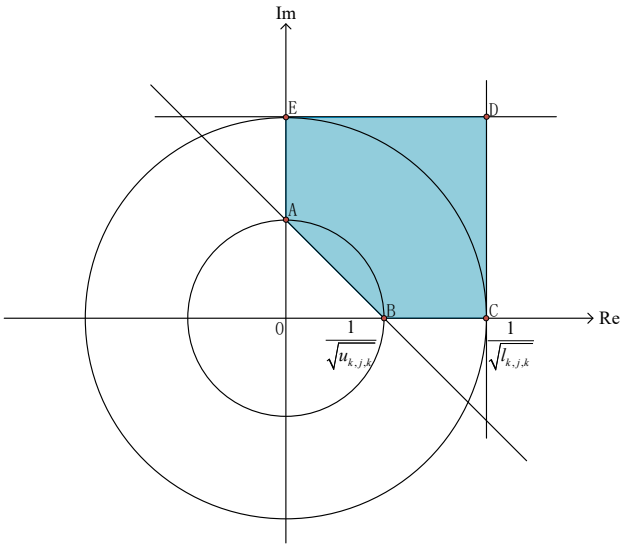


Fig. 2. An illustration of the convex envelope $Conv(Y_{k,j}^{[0,\frac{\pi}{2}]})$.

Proof: The relaxation of (18) is same as [31, Proposition 1], and we omit its proof here for brevity. To prove that $Conv(Y_{k,j}^{[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]})$ is a relaxation of $Y_{k,j}^{[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]}$, we first give an illustration on how $Y_{k,j}^{[0,\frac{\pi}{2}]}$ and $Conv(Y_{k,j}^{[0,\frac{\pi}{2}]})$ appear in Fig. 2. $Y_{k,j}^{[0,\frac{\pi}{2}]}$ is the filled region between the arc $\overset{\frown}{\mathrm{AB}}$ and the arc $\overline{\mathrm{CE}}$, which is obviously a nonconvex set, and $Conv(Y_{k,j}^{[0,\frac{\pi}{2}]})$ is the region bounded by five lines $\overline{\mathrm{AB}}$, $\overline{\mathrm{BC}}$, $\overline{\mathrm{AE}}$, $\overline{\mathrm{ED}}$ and $\overline{\mathrm{CD}}$. The points A, B, C, E are four extreme points of the set $Conv(Y_{k,j}^{[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]})$, which are $\left(\frac{1}{\sqrt{u_{k,j,k}}}\cos\left(\overline{\varphi}_{k,j}\right),\frac{1}{\sqrt{u_{k,j,k}}}\sin\left(\overline{\varphi}_{k,j}\right)\right)$, $\left(\frac{1}{\sqrt{u_{k,j,k}}}\cos\left(\underline{\varphi}_{k,j}\right),\frac{1}{\sqrt{u_{k,j,k}}}\sin\left(\underline{\varphi}_{k,j}\right)\right)$, $\left(\frac{1}{\sqrt{l_{k,j,k}}}\cos\left(\underline{\varphi}_{k,j}\right),\frac{1}{\sqrt{l_{k,j,k}}}\sin\left(\underline{\varphi}_{k,j}\right)\right)$ $\left(\frac{1}{\sqrt{l_{k,j,k}}}\cos\left(\overline{\varphi}_{k,j}\right),\frac{1}{\sqrt{l_{k,j,k}}}\sin\left(\overline{\varphi}_{k,j}\right)\right)$, respectively. Then the lines $\overline{\mathrm{AB}}$, $\overline{\mathrm{BC}}$ and $\overline{\mathrm{AE}}$ are given by (20c), (20a) and (20b), respectively. And the lines $\overline{\mathrm{CD}}$ and $\overline{\mathrm{ED}}$ are given by (20d) and (20e), respectively, as $\overline{\mathrm{CD}}$ and $\overline{\mathrm{ED}}$ are perpendicular to $\overline{\mathrm{BC}}$ and $\overline{\mathrm{AE}}$, respectively. ■

$$
\begin{aligned}
Conv(S_{k,j}^{[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]}) =&\Big\{\mathbf{a}_k|\sin(\underline{\varphi}_{k,j})\Re\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}-\cos(\underline{\varphi}_{k,j})\Im\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}\le 0,\\
&\sin\left(\overline{\varphi}_{k,j}\right)\Re\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}-\cos\left(\overline{\varphi}_{k,j}\right)\Im\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}\ge 0,\\
&a_{k,j}\Re\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}+b_{k,j}\Im\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}\ge a_{k,j}^2+b_{k,j}^2\Big\}
\end{aligned}\tag{19}
$$

$$
\begin{aligned}
Conv(Y_{k,j}^{[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]}) =&\Big\{\mathbf{a}_k|\sin(\underline{\varphi}_{k,j})\Re\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}-\cos(\underline{\varphi}_{k,j})\Im\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}\le 0, &\text{(20a)}\\
&\sin\left(\overline{\varphi}_{k,j}\right)\Re\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}-\cos\left(\overline{\varphi}_{k,j}\right)\Im\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}\ge 0, &\text{(20b)}\\
&a_{k,j}\Re\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}+b_{k,j}\Im\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}\ge\left(a_{k,j}^2+b_{k,j}^2\right)\frac{1}{\sqrt{u_{k,j,k}}}, &\text{(20c)}\\
&\cos(\underline{\varphi}_{k,j})\Re\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}+\sin(\underline{\varphi}_{k,j})\Im\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}\le\frac{1}{\sqrt{l_{k,j,k}}}, &\text{(20d)}\\
&\cos(\overline{\varphi}_{k,j})\Re\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}+\sin(\overline{\varphi}_{k,j})\Im\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}\le\frac{1}{\sqrt{l_{k,j,k}}}\Big\}. &\text{(20e)}
\end{aligned}
$$

When $\overline{\varphi}_{k,j}-\underline{\varphi}_{k,j}=2\pi$, $\mathbf{a}_k\in Conv(S_{k,j}^{[0,2\pi]})$ is equivalent to $\mathbf{a}_k\in\mathbb{C}$ and $\mathbf{a}_k\in Conv(Y_{k,j}^{[0,2\pi]})$ is equivalent to $\Re\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\}\le 1/\sqrt{l_{k,j,k}}$. It is easy to verify that, when the width of the interval $[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]$ goes to zero, the set $Conv(S_{k,j}^{[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]})$ becomes $S_{k,j}^{[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]}$ and the set $Conv(Y_{k,j}^{[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]})$ becomes $Y_{k,j}^{[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]}$, so the convex envelopes become tight.

Proposition 2 *(Argument Cut based Relaxation Problem):* The convex relaxation of P1 can be formulated as

$$
\begin{aligned}
\text{P2}: \min_{\{\mathbf{a}_k\}}\ &\sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l} l_{l,j,l}\left\|\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right\|^2+\sum_{k=1}^{K}\frac{\sigma_n^2}{P_t}\|\mathbf{a}_k\|^2\\
\text{s.t. }&(19),\forall k,\forall j\in\mathcal{E}_k\\
&(20),\forall k,\forall j\in\mathcal{D}_k/\mathcal{E}_k,
\end{aligned}\tag{21}
$$

which is a *quadratic programming* (QP) problem.

Proof: In Lemma 1, we give a lower bound of the objective function in P1. In Lemma 2, we relax the nonconvex constraints in P1 to convex ones. Therefore, P2 is a convex relaxation problem of P1. ■

C. Proposed BB-Based Algorithm

To solve P2, we need to know the ranges of each $1/\left\|\mathbf{a}_l^H\mathbf{h}_{l,j,l}\right\|^2$ in (16) and each $\varphi_{k,j}$ in Lemma 2. Define $\mathbf{v}=[\mathbf{d},\boldsymbol{\varphi}]\in\mathbb{R}_+^{p+M}$, where the elements in $\mathbf{d}\in\mathbb{R}_+^p$ denote $1/\left\|\mathbf{a}_l^H\mathbf{h}_{l,j,l}\right\|^2,\forall l,\forall j\in\mathcal{D}_l/\mathcal{E}_l$, $\boldsymbol{\varphi}$ consists of the argument variables $\{\varphi_{k,j}\}$ and $p=\sum_{k=1}^{K}|\mathcal{D}_k/\mathcal{E}_k|$ denotes the number of common nodes. Let V, LB^t and UB^t denote the the range of $\mathbf{v}$, the optimal lower bound, and the optimal upper bound of the optimal value of P1 at the t-th iteration, respectively. Define $lb(V)$ and $ub(V)$ as the lower bound and the upper bound over the box V, respectively. Let C^t denote a box list and $\{V,lb(V),\mathbf{c}\}$ denote an item in C^t, where $\mathbf{c}$ is the optimal solution to P2 over the box V. The initial range of $\mathbf{v}$ is defined as $V_{init}=[\underline{\mathbf{v}},\overline{\mathbf{v}}]$. According to Lemma 1 and 2, the lower vertices $\underline{\mathbf{v}}$ and the upper vertices $\overline{\mathbf{v}}$ can be given by

$$
\underline{\mathbf{v}}\to\mathbf{0}_{p+M},\ \overline{\mathbf{v}}=\left[\mathbf{1}_p^T,\ 2\pi\times\mathbf{1}_M^T\right]^T,
$$

where each element in $\underline{\mathbf{v}}$ is a positive number approaching 0.

At the t-th iteration, select a box in C^t such that the lower bound over it is the smallest one in C^t. Split the selected box into two smaller sub-boxes. For each sub-box, find its lower bound and upper bound. The current optimal lower bound is updated as the minimum value among the lower bounds from all obtained sub-boxes. The current optimal upper bound is updated as the minimum value among the upper bounds from all obtained sub-boxes. Discard the sub-boxes whose lower bound is larger than the current optimal upper bound. The detailed steps of the proposed BB algorithm are listed as follows.

a) Branch: The bisection method is a widely used branching rule for rectangular subdivision. Specifically, at the t-th iteration, we select the box with the smallest lower bound in C^t, denoted as V^t, and split it along the longest edge into two smaller ones, i.e., set $l=\arg\max_n\left\{\overline{v}_n^t-\underline{v}_n^t\right\}$ and $j_l^t=\left(\underline{v}_l^t+\overline{v}_l^t\right)/2$, split V^t into the following two smaller boxes

$$
\begin{aligned}
V_l^t&=\left\{\boldsymbol{\theta}\in V^t|\underline{v}_l^t\le\theta_l\le j_l^t\right\}\\
V_r^t&=\left\{\boldsymbol{\theta}\in V^t|j_l^t\le\theta_l\le\overline{v}_l^t\right\},
\end{aligned}\tag{22}
$$

where θ_l is the l-th component of $\boldsymbol{\theta}\in\mathbb{R}^{p+M}$.

b) Lower Bound: For each box $V\in\{V_l^t,V_r^t\}$, compute the lower bound $lb(V)$ by solving P2 over the box V. Then C^{t+1} can be formed by removing V^t from C^t and adding V_l^t and V_r^t if their lower bounds are smaller than or equal to the current best upper bound UB^t, i.e., $C^{t+1}=C^t\backslash\{V^t,lb(V^t),\mathbf{c}^t\}\cup\{V_i^t,lb(V_i^t),\mathbf{c}_i^t|lb(V_i^t)\le UB^t,\ i=l,r\}$. The optimal lower bound at the t-th iteration can be updated according to $LB^{t+1}=\min_{V\in C^{t+1}}lb(V)$. Since we always solve the relaxation problem in a smaller box $V\in\{V_l^t,V_r^t\}$ than V^t, we can ensure that the optimal lower bound does not decrease at each iteration.

c) Upper Bound: We can compute an upper bound by finding a feasible solution to P1 (denoted as $\left\{\overline{\mathbf{a}}_k\right\}$), which can be obtained by appropriately scaling the optimal solution to P2

$$\overline{\mathbf{a}}_k = \frac{\hat{\mathbf{a}}_k}{\min\left\{\underbrace{\left\{\cdots, \left\|\hat{\mathbf{a}}_k^H \mathbf{h}_{k,j,k}\right\|, \cdots\right\}}_{\forall j \in \mathcal{E}_k}, \underbrace{\left\{\cdots, \sqrt{u_{k,j,k}}\left\|\hat{\mathbf{a}}_k^H \mathbf{h}_{k,j,k}\right\|, \cdots\right\}}_{\forall j \in \mathcal{D}_k/\mathcal{E}_k}, 1\right\}}, \ \forall k. \tag{23}$$

(denoted as $\{\hat{\mathbf{a}}_k\}$). Specifically, $\{\hat{\mathbf{a}}_k\}$ may not be feasible to (16) and (18). Therefore, we use (23), as shown at top of the next page, to scale $\{\hat{\mathbf{a}}_k\}$ to satisfy the constraints in (16) and (18). Then, for each $V \in \{V_l^t, V_r^t\}$, the upper bound $ub(V)$ can be given as

$$ub(V) = \sum_{k=1}^{K} \sum_{l \in \mathcal{I}_k} \sum_{j \in \mathcal{C}_k \cap \mathcal{D}_l} \frac{\left\|\overline{\mathbf{a}}_k^H \mathbf{h}_{l,j,k}\right\|^2}{\left\|\overline{\mathbf{a}}_l^H \mathbf{h}_{l,j,l}\right\|^2} + \sum_{k=1}^{K} \frac{\sigma_n^2}{P_t} \|\overline{\mathbf{a}}_k\|^2. \tag{24}$$

If the newly added boxes can provide a smaller upper bound than the current best upper bound UB^t, we can find a better upper bound, i.e., $UB^{t+1} = \min\{UB^t, ub(V_l^t), ub(V_r^t)\}$, so that we can ensure that the optimal upper bound does not increase at each iteration.

d) Termination: For a given relative error tolerance, if

$$\frac{UB^t - LB^t}{LB^t} \leq \varepsilon, \tag{25}$$

the BB algorithm terminates.

The overall BB-based algorithm for solving P1 is summarized in Algorithm 1.

D. Convergence and Computational Complexity

As the length of the longest edge of the box V goes to zero, the BB algorithm converges only when the gap between upper and lower bounds also tends to zero. We provide the following proposition to show that an ε-optimal value can be obtained by the proposed BB algorithm.

Proposition 3 *(ε-Optimal Value):* For any given tolerance ε, an ε-optimal value is always exists when the size of the box $V^t \subseteq V_{init}$ is small enough. More specifically, for any given $\varepsilon > 0$, we can always find a $\delta \in (0, \pi/2)$ satisfying $p^{'}/\cos^2(\delta) + T_m\delta/\cos^2(\delta) - p^{'} + K\tan^2(\delta) \leq \varepsilon$, where $p^{'} = \sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}|\mathcal{C}_k \cap \mathcal{D}_l|$ and T_m is a large positive number such that $\sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l} 2/l_{l,j,l} \leq T_m$. When $\overline{v}_l^t - \underline{v}_l^t \leq 2\delta$, $(ub(V^t) - lb(V^t))/lb(V^t) \leq \varepsilon$ holds, so $(UB^t - LB^t)/LB^t \leq \varepsilon$ holds.

Proof: See Appendix B. ∎

Remark 1 *(Convergence Analysis for the BB Algorithm):* From Proposition 3, when the box V^t shrinks to a point, the gap between the upper and lower bounds over V^t becomes sufficiently small and the proposed BB algorithm converges. More precisely, when $\varepsilon \to 0$, $(UB^t - LB^t)/LB^t$ will converge to 0. Since LB^t is uniformly bounded away from 0, we can further obtain $UB^t - LB^t \to 0$, which indicates that the upper and lower bounds converge to the optimal value of the formulated problem. Therefore, the convergence of the proposed BB algorithm is guaranteed.

Remark 2 *(Complexity Analysis for the BB Algorithm):* The volume of V_{init} is $1^p(2\pi)^M$. Using the conclusion in [31, Theorem 1], we can conclude that, for any given $\varepsilon > 0$, the maximum number of iterations of the BB algorithm is given by

$$T_{\max} = \left\lceil \left(\frac{2\pi}{\delta}\right)^M \left(\frac{1}{\delta}\right)^p \right\rceil + 1, \tag{26}$$

where δ is defined in Proposition 3. In each iteration, the BB algorithm generates two branches, and each branch corresponds to a QP problem P2. In the worst case, the number of branches generated in the whole algorithm is $2 \cdot T_{\max}$. Each QP problem can be solved by a generic *interior-point method* (IPM). The complexity of IPM is on the order of $\mathcal{O}((KN_r)^3 L)$, where L denotes the quantity $\ln(1/\epsilon)$ and ϵ denotes the relative accuracy [34]. The worst-case computational complexity of the proposed BB algorithm is therefore $\mathcal{O}\left(2 \cdot T_{\max}(KN_r)^3 L\right)$.

Algorithm 1 BB-Based Algorithm

Initialization: Set the iteration index $t = 0$, $V^0 = V_{init}$ and the tolerance ε. Solve P2 to find the initial lower bound $LB^0 = lb(V^0)$ and $\mathbf{c}^0$. Obtain the initial upper bound $UB^0 = ub(V^0)$ according to (23) and (24). Initialize $C^0 \leftarrow \{V^0, lb(V^0), \mathbf{c}^0\}$.

Repeat

1: Select a box V^t in C^t such that the lower bound $lb(V^t)$ is the smallest one in C^t. Split V^t along the longest edge into two smaller ones V_l^t and V_r^t according to (22).
2: For each box V_i^t $(i = l, r)$, find its lower bound $lb(V_i^t)$ by solving P2 and its upper bound $ub(V_i^t)$ according to (23) and (24).
3: Update $C^{t+1} = C^t \backslash \{V^t, lb(V^t), \mathbf{c}^t\} \cup \{V_i^t, lb(V_i^t), \mathbf{c}_i^t | lb(V_i^t) \leq UB^t,\ i = l, r\}$.
4: Update $LB^{t+1} = \min_{V \in C^{t+1}} lb(V)$.
5: Update $UB^{t+1} = \min\{UB^t, ub(V_l^t), ub(V_r^t)\}$.
6: Set $t = t + 1$.

Until $(UB^t - LB^t)/LB^t \leq \varepsilon$

IV. SCA-Based Distributed Algorithm

The BB algorithm in the previous section requires a centralized controller to coordinate all FCs, which may not be practical in some applications. Therefore, we only use it as a network performance benchmark. In this section, we consider a more practical scenario without a centralized controller and each FC only has the CSI of the nodes within its own cluster. We propose a low-complexity distributed algorithm by exploiting SCA and primal decomposition theory. To gain more insights, we also consider a special case of each FC

equipped with massive antennas and develop an asymptotically optimal beamformer design in closed-form for P1.

A. SCA Transformation

Since all the receivers are coupled in the objective function of P1, it is not easy to develop a distributed algorithm. Inspired by the application of IT in beamforming [27], [28], we introduce a set of slack variables $\{\tau_{k,j}\}$ to make P1 decomposable, which satisfy

$$\left\|\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\|^2 \geq \tau_{k,j}, \forall k, \forall j \in \mathcal{D}_k/\mathcal{E}_k. \tag{27}$$

$\tau_{k,j}$ reflects the magnitude of the transmit power of the (k,j)-th node. The larger it is, the smaller the transmit power will be, so the smaller the interference level from the (k,j)-th node to the FCs not serving it. To ease the notation, we denote $\boldsymbol{\tau}$ as a $p \times 1$ vector composed of $\tau_{k,j}$'s, $\forall k, \forall j \in \mathcal{D}_k/\mathcal{E}_k$. Then, an equivalent problem of P1 can be formulated in the following lemma.

Lemma 3 *(Equivalent Problem of* P1*):* P1 can be equivalently transformed into

$$\text{P3}: \min_{\{\mathbf{a}_k\},\boldsymbol{\tau}} \sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\frac{\left\|\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right\|^2}{\tau_{l,j}} + \sum_{k=1}^{K}\frac{\sigma_n^2}{P_t}\|\mathbf{a}_k\|^2 \tag{28a}$$

$$\text{s.t. } \left\|\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\|^2 \geq \tau_{k,j}, \forall k, \forall j \in \mathcal{D}_k/\mathcal{E}_k \tag{28b}$$

$$\left\|\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\|^2 \geq 1, \forall k, \forall j \in \mathcal{E}_k \tag{28c}$$

$$\tau_{k,j} \geq 1, \forall k, \forall j \in \mathcal{D}_k/\mathcal{E}_k. \tag{28d}$$

Proof: See Appendix C. ∎

It can be observed that $\{\mathbf{a}_k\}$ and $\boldsymbol{\tau}$ are coupled in (28a). Thus, we adopt the following lemma to decouple them, i.e.,

Lemma 4 *(Linearization of Hyperbolic Constraint):* The hyperbolic constraint

$$\frac{\left\|\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right\|^2}{\tau_{l,j}} \leq q_{l,j,k} \tag{29}$$

can be converted into the following linear matrix inequality

$$\begin{bmatrix} q_{l,j,k} & \mathbf{a}_k^H\mathbf{h}_{l,j,k} \\ \left(\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right)^H & \tau_{l,j} \end{bmatrix} \succcurlyeq \mathbf{0}. \tag{30}$$

Proof: Applying the Schur complement condition in [35], (30) can be easily obtained from (29). ∎

From Lemma 4, the convexity of the objective function (28a) in P3 can be easily verified. Then, P3 is intractable only due to the nonconvex constraints (28b) and (28c), which is similar to the case of single-cluster CoMAC. It has been shown that SCA outperforms SDR in [12]. Thus, we can apply the similar SCA method in [12] to cope with constraints (28b) and (28c), as given in the following lemma.

Lemma 5 *(Convex Approximation of (28b) and (28c)):* The constraints (28b) and (28c) can be approximated by the iterative relaxed linear constraints, i.e.,

$$2\Re\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\mathbf{h}_{k,j,k}^H\mathbf{c}_k^{(s)}\right\} - \left\|\mathbf{h}_{k,j,k}^H\mathbf{c}_k^{(s)}\right\|^2 \geq \tau_{k,j}, \tag{31}$$

$$2\Re\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\mathbf{h}_{k,j,k}^H\mathbf{c}_k^{(s)}\right\} - \left\|\mathbf{h}_{k,j,k}^H\mathbf{c}_k^{(s)}\right\|^2 \geq 1, \tag{32}$$

where $\mathbf{c}_k^{(s)}$ is the solution at the $(s-1)$-th iterative optimization.

Proof: Considering auxiliary vector $\mathbf{c}_k^{(s)} \in \mathbb{C}^{N_r}$, $\left(\mathbf{a}_k - \mathbf{c}_k^{(s)}\right)^H\mathbf{h}_{k,j,k}\mathbf{h}_{k,j,k}^H\left(\mathbf{a}_k - \mathbf{c}_k^{(s)}\right) \geq 0$ holds. It follows that $\left\|\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\|^2 \geq 2\Re\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\mathbf{h}_{k,j,k}^H\mathbf{c}_k^{(s)}\} - \left\|\mathbf{h}_{k,j,k}^H\mathbf{c}_k^{(s)}\right\|^2$. Thus, we can obtain (31) and (32). ∎

Combining the above lemmas, P3 can be tackled by sequentially solving a series of convex problems in the following proposition.

Proposition 4 *(Convex Approximation of* P3*):* At the s-th iteration optimization, P3 can be approximated as

$$\begin{aligned}\text{P4}\left(\{\mathbf{c}_k^{(s)}\}\right): \min_{\{\mathbf{a}_k\},\mathbf{q}} & \sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l} q_{l,j,k} + \sum_{k=1}^{K}\frac{\sigma_n^2}{P_t}\|\mathbf{a}_k\|^2 \\ \text{s.t. } & (30), \forall k, \forall l \in \mathcal{I}_k, \forall j \in \mathcal{C}_k \cap \mathcal{D}_l \\ & (31), \forall k, \forall j \in \mathcal{D}_k/\mathcal{E}_k \\ & (32), \forall k, \forall j \in \mathcal{E}_k \\ & (28\text{d}),\end{aligned} \tag{33}$$

where the optimization variable $\mathbf{q}$ is a vector composed of $q_{l,j,k}$'s, $\forall k, \forall l \in \mathcal{I}_k, \forall j \in \mathcal{C}_k \cap \mathcal{D}_l$.

Proof: By applying (29) and (30), we give a quadratic representation of (28a). By applying (31) and (32), we convert the nonconvex constraints (28b) and (28c) into convex ones. Therefore, $\text{P4}\left(\{\mathbf{c}_k^{(s)}\}\right)$ is convex. ∎

B. Distributed Beamformer Design

In this subsection, we present a distributed beamforming design based on the existing decomposition methods [32], [36]. Specifically, we alternately optimize $\boldsymbol{\tau}$ and $\{\mathbf{a}_k\}$ so that the objective function value continues to decrease.

When fixing $\boldsymbol{\tau}$, $\text{P4}\left(\{\mathbf{c}_k^{(s)}\}\right)$ can be decomposed into K parallel subproblems, i.e., the k-th subproblem at the s-th iteration can be formulated as

$$\begin{aligned}\mathbf{P}_k^{\text{sub}}\left(\boldsymbol{\tau}, \mathbf{c}_k^{(s)}\right): \min_{\mathbf{a}_k,\mathbf{q}_k} & \sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l} q_{l,j,k} + \frac{\sigma_n^2}{P_t}\|\mathbf{a}_k\|^2 \\ \text{s.t. } & (30), \forall l \in \mathcal{I}_k, \forall j \in \mathcal{C}_k \cap \mathcal{D}_l \\ & (31), \forall j \in \mathcal{D}_k/\mathcal{E}_k \\ & (32), \forall j \in \mathcal{E}_k,\end{aligned} \tag{34}$$

where $\mathbf{q}_k$ is composed of $q_{l,j,k}$'s, $\forall l \in \mathcal{I}_k, \forall j \in \mathcal{C}_k \cap \mathcal{D}_l$.

We can observe that the k-th FC only needs the CSI from the nodes in $\mathcal{C}_k$ to it. Each subproblem is convex and can be solved by generic convex programming solver SeDuMi. Let $\left(\mathbf{a}_k^{(s)}, \mathbf{q}_k^{(s)}\right)$ denote the s-th solution of the SCA procedure. Once getting $\left(\mathbf{a}_k^{(s)}, \mathbf{q}_k^{(s)}\right)$, we then solve $\mathbf{P}_k^{\text{sub}}\left(\boldsymbol{\tau}, \mathbf{c}_k^{(s+1)}\right)$ with $\mathbf{c}_k^{(s+1)} = \mathbf{a}_k^{(s)}$, and repeat the process until a stopping condition is met. The initial $\mathbf{c}_k^{(1)}$ can be obtained by randomly generating $\mathbf{a}_k^{(0)}$ satisfying (28b) and (28c).

When the SCA procedure terminates, each FC can update $\boldsymbol{\tau}$ by solving the following master problem

$$\begin{aligned}\mathbf{P}^{\text{mas}}: \min_{\boldsymbol{\tau}} & \sum_{k=1}^{K} P_k^{\star}(\boldsymbol{\tau}) \\ \text{s.t. } & \boldsymbol{\tau} \geq \mathbf{1}_p,\end{aligned} \tag{35}$$

where $P_k^\star(\boldsymbol{\tau})$ is the objective function value at the k-th FC, and the constraint comes from (28d).

According to the primal decomposition theory, we apply the subgradient projection method to solve the master problem. Define the subgradient vector of $P_k^\star(\boldsymbol{\tau})$ as $\mathbf{g}_k \in \mathbb{R}^{p\times 1}$ and the global subgradient vector as $\mathbf{g} \in \mathbb{R}^{p\times 1}$. Each FC can get the subgradient vector $\mathbf{g}_k$ by solving its own subproblems and the global subgradient vector $\mathbf{g}$ can be obtained using the following proposition.

Proposition 5 *(Global Subgradient):* The vector $\mathbf{g}$ is given by

$$\mathbf{g} = \sum_{k=1}^{K} \mathbf{g}_k. \tag{36}$$

Proof: See Appendix D. ∎

Due to the partial connection characteristics of the system, it can be observed from each subproblem that each FC only needs the slack variables that reflect the transmit power of the common nodes in its cluster, e.g., the k-th FC only needs $\tau_{k,j}, \forall j \in \mathcal{D}_k/\mathcal{E}_k$ and $\tau_{l,j}, \forall l \in \mathcal{I}_k, \forall j \in \mathcal{C}_k \cap \mathcal{D}_l$. Therefore, after receiving the subgradient vectors of its neighbor clusters, i.e., $\mathcal{I}_k \cup \mathcal{I}_k'$, the k-th FC can start computing its global subgradient $\mathbf{g}$ without waiting for the subgradient vectors of all clusters to be received. Then, the vector $\boldsymbol{\tau}$ is updated as

$$\boldsymbol{\tau}(n+1) = \left[\boldsymbol{\tau}(n) - \mu(n)\cdot \frac{\mathbf{g}}{\|\mathbf{g}\|} - \mathbf{1}_p\right]^+ + \mathbf{1}_p, \tag{37}$$

where n denotes the iteration index, μ is the step size which decreases with the iteration process, i.e., $\mu(n) = \mu(1)/\sqrt{n}$ ($\mu(1) > 0$ is the initial step size), and $[\cdot]^+$ denotes the projection onto the nonnegative orthant.

Finally, the decentralized algorithm for solving P1 is summarized in Algorithm 2.

Algorithm 2 SCA-Based Distributed Algorithm

Initialization: Set $\boldsymbol{\tau}(0)$ to a vector satisfying (28d) and the iteration index n to 0

Repeat

1: Each FC uses SCA method to compute its subgradient vector and broadcasts it to other FCs (via the backhaul link between each FC).
2: Upon receiving the subgradient vectors from its neighbor clusters, each FC computes the global subgradient vector $\mathbf{g}$ and updates $\boldsymbol{\tau}$ according to (36) and (37), respectively.
3: Set $n = n + 1$.

Until termination criterion is met

C. Optimal Beamforming Structure

Algorithm 2 can find good feasible solutions to P1. However, it cannot offer a fundamental understanding of the beamforming structure for multi-cluster CoMAC. Besides, the complexity of the SCA method can be prohibitively high if the number of antennas is large. To gain more insight, we analyze the optimal beamforming structure of $\mathbf{P}_k^{\mathrm{sub}}\left(\boldsymbol{\tau}, \mathbf{c}_k^{(s)}\right)$ in the following Lemma 6.

Lemma 6 *(The Optimal Solution Structure of* $\mathrm{P}_k^{\mathrm{sub}}\left(\boldsymbol{\tau}, \mathrm{c}_k^{(s)}\right)$*):* The optimal solution of $\mathbf{P}_k^{\mathrm{sub}}\left(\boldsymbol{\tau}, \mathbf{c}_k^{(s)}\right)$ is a linear combination of the channels between the k-th FC and the nodes in the same cluster, i.e.,

$$\mathbf{a}_k^\star = \sum_{j\in\mathcal{D}_k} v_{k,j,k}^\star \mathbf{h}_{k,j,k} + \sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l} v_{l,j,k}^\star \mathbf{h}_{l,j,k}, \tag{38}$$

where $v_{k,j,k}^\star = \frac{P_t}{\sigma_n^2}\lambda_{k,j}^\star \mathbf{h}_{k,j,k}^H \mathbf{c}_k^{(s)}$ and $v_{l,j,k}^\star = \frac{P_t}{\sigma_n^2}\left(t_{l,j,k}^{(2)\star}\right)^*$ with $\lambda_{k,j}^\star$ and $t_{l,j,k}^{(2)\star}$ being the optimal Lagrange multipliers.

Proof: The Lagrangian of $\mathbf{P}_k^{\mathrm{sub}}\left(\boldsymbol{\tau}, \mathbf{c}_k^{(s)}\right)$ is defined in (68). Ignoring items not related to $\mathbf{a}_k$, we can rewrite it as

$$\begin{aligned}\mathcal{L}_k(\mathbf{a}_k) = \frac{\sigma_n^2}{P_t}\|\mathbf{a}_k\|^2 - \sum_{j\in\mathcal{D}_k} 2\lambda_{k,j}\Re\left\{\mathbf{a}_k^H \mathbf{h}_{k,j,k}\mathbf{h}_{k,j,k}^H \mathbf{c}_k^{(s)}\right\} \\ - \sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\left(t_{l,j,k}^{(2)}\left(\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right)^H + t_{l,j,k}^{(3)}\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right).\end{aligned} \tag{39}$$

By the Karush-Kuhn-Tucker conditions, at the optimality of $\mathbf{P}_k^{\mathrm{sub}}\left(\boldsymbol{\tau}, \mathbf{c}_k^{(s)}\right)$, the gradient of $\mathcal{L}_k(\mathbf{a}_k)$ with respect to $\mathbf{a}_k$ satisfies $\nabla_{\mathbf{a}_k}\mathcal{L}_k(\mathbf{a}_k) = 0$. Since $\nabla_{\mathbf{x}}\Re\left\{\mathbf{c}^H\mathbf{x}\right\} = \mathbf{c}^*/2$ for the complex vector $\mathbf{c}$ and $\nabla_{\mathbf{x}}\left(\mathbf{x}^H\mathbf{C}\mathbf{x}\right) = (\mathbf{C}\mathbf{x})^*$ for the Hermitian matrix $\mathbf{C}$, we obtain

$$\mathbf{a}_k^\star = \sum_{j\in\mathcal{D}_k}\frac{P_t}{\sigma_n^2}\lambda_{k,j}^\star \mathbf{h}_{k,j,k}^H \mathbf{c}_k^{(s)}\mathbf{h}_{k,j,k} + \sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\frac{P_t}{\sigma_n^2}\left(t_{l,j,k}^{(2)\star}\right)^*\mathbf{h}_{l,j,k}. \tag{40}$$

∎

Define $\mathbf{H}_k = \left[\cdots, \mathbf{h}_{k,j,k}, \cdots, \mathbf{h}_{l,j,k}, \cdots\right] \in \mathbb{C}^{N_r\times|\mathcal{C}_k|}$ as the channel matrix for the k-th cluster and $\mathbf{v}_k = \left[\cdots, v_{k,j,k}, \cdots, v_{l,j,k}, \cdots\right]^T$ as the corresponding linear combination coefficients vector. We can rewrite (38) as $\mathbf{a}_k^\star = \mathbf{H}_k\mathbf{v}_k^\star$. As the SCA method iteratively updates $\mathbf{c}_k^{(s)}$, $\mathbf{a}_k^\star$ is updated accordingly without changing its structure. Assume that the rank of $\mathbf{H}_k$ satisfies $r_k \le \min(N_r, |\mathcal{C}_k|)$. Then, $\mathbf{H}_k$ can be decomposed as $\mathbf{U}_k\boldsymbol{\Sigma}_k\mathbf{V}_k$ by the truncated singular-value decomposition, where $\mathbf{U}_k \in \mathbb{C}^{N_r\times r_k}$ is composed of the orthonormal vectors that can span the column space of $\mathbf{H}_k$. Thus, the optimal solution of $\mathbf{P}_k^{\mathrm{sub}}\left(\boldsymbol{\tau}, \mathbf{c}_k^{(s)}\right)$, i.e., $\mathbf{a}_k^\star = \mathbf{H}_k\mathbf{v}_k^\star$, can be converted to

$$\mathbf{a}_k^\star = \mathbf{U}_k\mathbf{b}_k^\star, \tag{41}$$

where $\mathbf{b}_k^\star$ is defined as $\mathbf{b}_k^\star = \boldsymbol{\Sigma}_k\mathbf{V}_k\mathbf{v}_k^\star$.

Directly obtaining the weighted vector $\mathbf{b}_k^\star$ is difficult. Thus, we compute $\mathbf{b}_k^\star$ through a simple numerical algorithm. Specifically, substituting (41) into (34), $\mathbf{b}_k^\star$ can be computed by solving the following problem.

$$\begin{aligned}\mathbf{S}_k^{\mathrm{sub}}\left(\boldsymbol{\tau}, \mathbf{d}_k^{(s)}\right): \min_{\mathbf{b}_k, \mathbf{q}_k} \ & \sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l} q_{l,j,k} + \frac{\sigma_n^2}{P_t}\|\mathbf{U}_k\mathbf{b}_k\|^2 \\ \text{s.t. } & g_j^{(s)}(\mathbf{b}_k) \ge 1, \forall j \in \mathcal{E}_k \\ & g_j^{(s)}(\mathbf{b}_k) \ge \tau_{k,j}, \forall j \in \mathcal{D}_k/\mathcal{E}_k \\ & \begin{bmatrix} q_{l,j,k} & \mathbf{b}_k^H\mathbf{U}_k^H\mathbf{h}_{l,j,k} \\ \mathbf{h}_{l,j,k}^H\mathbf{U}_k\mathbf{b}_k & \tau_{l,j}\end{bmatrix} \succcurlyeq \mathbf{0}, \forall l \in \mathcal{I}_k, \forall j \in \mathcal{C}_k\cap\mathcal{D}_l,\end{aligned} \tag{42}$$

where $g_j^{(s)}(\mathbf{b}_k) = 2\Re\left\{\mathbf{b}_k^H\mathbf{U}_k^H\mathbf{h}_{k,j,k}\mathbf{h}_{k,j,k}^H\mathbf{U}_k\mathbf{d}_k^{(s)}\right\} - \left\|\mathbf{h}_{k,j,k}^H\mathbf{U}_k\mathbf{d}_k^{(s)}\right\|^2$ with $\mathbf{d}_k^{(s)}$ being the solution at the $(s-1)$-th iterative optimization.

Compared with directly optimizing each receiver in (34), solving the weighted vectors $\mathbf{b}_k$ in (42) does not depend on the number of antennas but only on the rank of $\mathbf{H}_k$, which leads to lower computational complexity when FCs are equipped with large antenna arrays.

To gain more insights of $\{\mathbf{a}_k\}$, we consider a special case where each FC has a very large number of antennas and use the following lemma.

Lemma 7 *(Convergence of $\boldsymbol{h}_{k,i,k}^H\boldsymbol{h}_{k',i',k}/N_r$):* When $N_r \to \infty$, we have

$$\lim_{N_r\to\infty}\frac{\mathbf{h}_{k,i,k}^H\mathbf{h}_{k',i',k}}{N_r} \overset{\text{a.s.}}{=} \begin{cases} 0, & \text{if } (k,i)\neq(k',i') \\ \kappa_{k,i,k}, & \text{else} \end{cases} \tag{43}$$

where $\overset{\text{a.s.}}{=}$ denotes the almost sure convergence.

Proof: According to [37, Equation (4)], we can derive (43) by applying the law of large numbers and using the fact that $\mathbf{h}_{k,i,k} \sim \mathcal{CN}(\mathbf{0}, \kappa_{k,i,k}\mathbf{I}_{N_r})$ and is independent of $\mathbf{h}_{k',i',k}, \forall(k,i)\neq(k',i')$, i.e.,

$$\begin{aligned}\lim_{N_r\to\infty}\frac{\mathbf{h}_{k,i,k}^H\mathbf{h}_{k',i',k}}{N_r} &= \lim_{N_r\to\infty}\frac{\sum_{j=1}^{N_r}\left|p_j^*q_j\right|}{N_r} \overset{\text{a.s.}}{=} \mathbb{E}\left\{\left|p_j^*q_j\right|\right\} \\ &= \begin{cases} 0, & \text{if } (k,i)\neq(k',i') \\ \kappa_{k,i,k}, & \text{else} \end{cases}\end{aligned}$$

where p_j and q_j denote the j-th element of $\mathbf{h}_{k,i,k}$ and $\mathbf{h}_{k',i',k}$, respectively. ∎

Based on Lemma 7, a closed-form asymptotically optimal receive beamformer can be derived in the following proposition.

Proposition 6 *(The Asymptotically Optimal Beamformer of* P1*):* When N_r is sufficiently large, the asymptotically optimal receiver at each FC is approximated as a linear combination of the channels between it and its served nodes, i.e.,

$$\mathbf{a}_k \approx \sum_{j\in\mathcal{D}_k}\frac{1}{\kappa_{k,j,k}N_r}\mathbf{h}_{k,j,k}, \forall k. \tag{44}$$

Proof: First, inspired by Lemma 6, we assume that the asymptotically optimal receivers can be given by $\mathbf{a}_k = \sum_{i\in\mathcal{D}_k} v_{k,i,k}\mathbf{h}_{k,i,k}/N_r, \forall k$, which will lead to gradually vanishing inter-cluster interference and noise as $N_r \to \infty$:

$$\begin{aligned}&\lim_{N_r\to\infty}\hat{x}_k \\ &\overset{(a)}{=} \lim_{N_r\to\infty}\left(\sum_{i\in\mathcal{D}_k}x_{k,i} + \sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\frac{\mathbf{a}_k^H\mathbf{h}_{l,j,k}}{\mathbf{a}_l^H\mathbf{h}_{l,j,l}}x_{l,j} + \mathbf{a}_k^H\mathbf{z}_k\right) \\ &= \lim_{N_r\to\infty}\left(\sum_{i\in\mathcal{D}_k}x_{k,i} + \sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\frac{\sum_{i\in\mathcal{D}_k}v_{k,i,k}^*\frac{\mathbf{h}_{k,i,k}^H\mathbf{h}_{l,j,k}}{N_r}}{\sum_{i\in\mathcal{D}_l}v_{l,i,l}^*\frac{\mathbf{h}_{l,i,l}^H\mathbf{h}_{l,j,l}}{N_r}}x_{l,j}\right. \\ &\quad \left. + \sum_{i\in\mathcal{D}_k}v_{k,i,k}^*\frac{\mathbf{h}_{k,i,k}^H\mathbf{z}_k}{N_r}\right) \\ &\overset{\text{a.s.}}{=} \sum_{i\in\mathcal{D}_k}x_{k,i},\end{aligned} \tag{45}$$

where the equation (a) comes from (10) and (12). By substituting the assumed receivers $\{\mathbf{a}_k\}$ into P1, we have

$$\lim_{N_r\to\infty}\text{MSE} = \lim_{N_r\to\infty}\frac{\sigma_n^2}{P_t}\sum_{k=1}^{K}\sum_{j\in\mathcal{D}_k}\frac{\left|v_{k,j,k}\right|^2\mathbf{h}_{k,j,k}^H\mathbf{h}_{k,j,k}}{N_r^2} \overset{\text{a.s.}}{=} 0. \tag{46}$$

Next, if the receiver is not a linear combination of channels, it can be written as

$$\mathbf{a}_k = \sum_{j\in\mathcal{D}_k}\frac{v_{k,j,k}}{N_r}\mathbf{h}_{k,j,k} + \sum_{t=1}^{N_r-|\mathcal{D}_k|}\frac{\bar{v}_{k,t}}{N_r}\mathbf{u}_{k,t}, \tag{47}$$

where $\{\mathbf{u}_{k,t}\}_{t=1}^{N_r-|\mathcal{D}_k|}$ is an orthonormal basis for the orthogonal complement of the space spanned by $\{\mathbf{h}_{k,j,k}\}_{j\in\mathcal{D}_k}$. It can be observed that the term $\sum_{t=1}^{N_r-|\mathcal{D}_k|}\bar{v}_{k,t}\mathbf{u}_{k,t}/N_r$ in (47) will bring additional interference and noise as $N_r \to \infty$, causing degraded MSE performance. Therefore, the asymptotically optimal beamformer of P1 has the structure as $\mathbf{a}_k = \sum_{i\in\mathcal{D}_k} v_{k,i,k}\mathbf{h}_{k,i,k}/N_r, \forall k$.

When N_r is sufficiently large, MSE is approximated to

$$\text{MSE} \approx \frac{\sigma_n^2}{P_tN_r}\sum_{k=1}^{K}\sum_{j\in\mathcal{D}_k}\left|v_{k,j,k}\right|^2\kappa_{k,j,k}. \tag{48}$$

And the constraints in P1 can be converted to

$$\left|v_{k,j,k}^*\kappa_{k,j,k}\right| \geq 1, \forall j\in\mathcal{D}_k, \forall k. \tag{49}$$

Apparently, when $v_{k,j,k} = 1/\kappa_{k,j,k}$, (48) reaches minimum, i.e.,

$$\text{MSE} \approx \frac{\sigma_n^2}{P_tN_r}\sum_{k=1}^{K}\sum_{j\in\mathcal{D}_k}\frac{1}{\kappa_{k,j,k}}. \tag{50}$$

∎

It can be seen that the inter-cluster interference vanishes as $N_r \to \infty$. Similar conclusions have also appeared in massive MIMO multicast networks [38]. In this case, each FC only needs the CSI of the nodes it serves, which indicates that the FCs do not need to cooperate with each other, information exchange overhead is thus avoided. Considering a special scenario of $K = 1$ and $\kappa_{k,j,k} = \kappa, \forall(k,j)$, we obtain $\text{MSE} \approx \sigma_n^2|\mathcal{D}|/(\kappa P_tN_r)$, which has the same form with that in [39, Theorem 2].

Remark 3 *(Convergence Analysis for the SCA Algorithm):* P3 is solved by alternately optimizing the receivers $\{\mathbf{a}_k\}$ and the vector $\boldsymbol{\tau}$. With fixed $\boldsymbol{\tau}$, $\{\mathbf{a}_k\}$ can be obtained by SCA. At each iteration optimization of the SCA method, several parallel convex subproblems $\mathbf{P}_k^{\text{sub}}(\boldsymbol{\tau}, \mathbf{c}_k^{(s)})$ or $\mathbf{S}_k^{\text{sub}}(\boldsymbol{\tau}, \mathbf{d}_k^{(s)})$ needs to be solved,. Each FC solves its own subproblem in parallel to ensure that the value of the objective function continues to decrease. With fixed $\{\mathbf{a}_k\}$, $\boldsymbol{\tau}$ is updated by solving the master problem $\mathbf{P}^{\text{mas}}$. It can be seen from Proposition 5 that the subgradient projection method can also guarantee a continuously decreasing objective function value. Considering the fact that the objective function is lower bounded by 0 and its value can be reduced by alternately optimizing the vector $\boldsymbol{\tau}$ and receivers $\{\mathbf{a}_k\}$, the proposed SCA-based distributed algorithm is guaranteed to converge to a finite value. Several

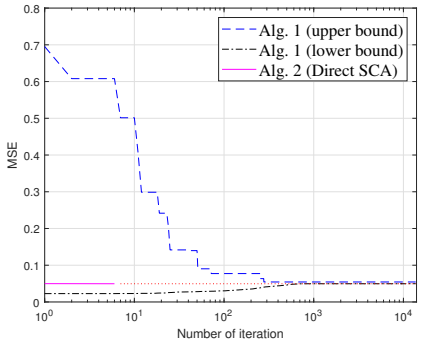
numerical simulations will be provided later in Section V to validate the convergence of the proposed algorithm.

Remark 4 *(Signaling Overhead Analysis):* We use the number of channel use to characterize the signaling overhead. The main information exchange among all FCs in the above distributed algorithms is the real-valued subgradient vectors. During each iteration, for the k-th FC, it should broadcast its subgradient vector $\mathbf{g}_k$, which only has $|\mathcal{D}_k/\mathcal{E}_k|+\sum_{l\in\mathcal{I}_k}|\mathcal{C}_k\cap\mathcal{D}_l|$ nonzero scalars and needs $|\mathcal{D}_k/\mathcal{E}_k|+\sum_{l\in\mathcal{I}_k}|\mathcal{C}_k\cap\mathcal{D}_l|$ channel uses to transmit these scalars. The sum signaling overhead among the FCs in one iteration is thus $(p+p^{'})(K-1)$. Note that the centralized BB-based algorithm requires each FC to share its local CSI with other FCs. In this case, the k-th FC has to send $2|\mathcal{C}_k|N_r$ real values per link, which yields a total overhead of $2N_r(K-1)\left(M+p^{'}\right)$.

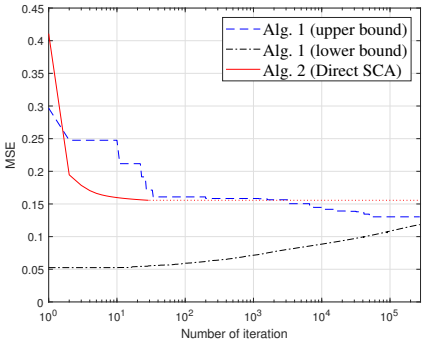
Remark 5 *(Computational Complexity for the SCA Algorithm):* The complexity of IPM for solving each subproblem (42) is on the order of $\sqrt{\beta_k}\cdot\left(C^k_{form}+C^k_{fact}\right)\cdot\ln(1/\upsilon)$, where $\sqrt{\beta_k}\ln(1/\upsilon)$ denotes the number of iterations required in IPM to reach an υ-optimal solution and $C^k_{form}+C^k_{fact}$ denotes per-iteration computation cost [40]. Specifically, $\beta_k=2\cdot\sum_{l\in\mathcal{I}_k}|\mathcal{C}_k\cap\mathcal{D}_l|+|\mathcal{D}_k|$, $C^k_{form}=n_k\cdot\left(8\cdot\sum_{l\in\mathcal{I}_k}|\mathcal{C}_k\cap\mathcal{D}_l|+|\mathcal{D}_k|\right)+n_k^2\cdot\left(4\cdot\sum_{l\in\mathcal{I}_k}|\mathcal{C}_k\cap\mathcal{D}_l|+|\mathcal{D}_k|\right)$ and $C^k_{fact}=n_k^3$, where $n_k=r_k+\sum_{l\in\mathcal{I}_k}|\mathcal{C}_k\cap\mathcal{D}_l|$ for (42). Assume that the number of subproblems solved by each FC is N_a during the iteration. Then the total computational complexity for the distributed algorithms is $\sum_{k=1}^{K}N_a\sqrt{\beta_k}\cdot\left(C^k_{form}+C^k_{fact}\right)\cdot\ln(1/\upsilon)$. Specially, if $r_k=N_r$, we obtain the computational complexity for solving each subproblem (34). Due to the fact that $r_k\le\min(N_r,|\mathcal{C}_k|)$, the complexity for solving the weighted vector $\mathbf{b}_k$ is less than directly optimizing the receiver $\mathbf{a}_k$, especially in the case of large-scale antenna systems. Moreover, since $T_{\max}\gg KN_a\sqrt{\beta_k}\ln(1/\upsilon)$, the computational complexity of the SCA-based distributed algorithm is much less than that of the BB algorithm.

V. Simulation Results And Discussion

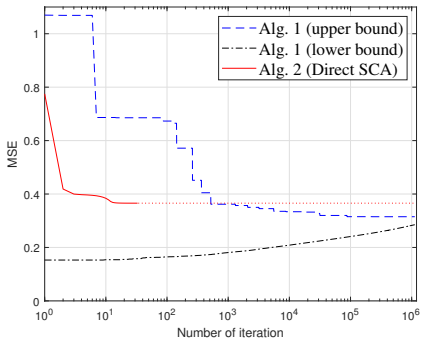
In this section, we provide simulation results to show the convergence and the MSE performance of the proposed algorithms. The channels are Rayleigh, i.e., the elements of each channel vector are modeled as independent and identically distributed complex Gaussian random variables with zero mean and unit variance. For interfering channels, we introduce $\varsigma\in[0,1]$ to represent the strength ratio between the interfering and the desired signal. $\varsigma=0$ denotes that there is no inter-cluster interference, which happens when undesired nodes are farther away, thereby making the interference very weak. And $\varsigma=1$ denotes that the interfering signals have the same strength as the desired signals. Through adjusting the strength ratio ς, we can characterize the distance between FC and its undesired nodes. We also assume that each common node only affects one FC and each cluster has the same system configuration, denoted as $(N_r,|\mathcal{D}|)$, where each FC is equipped with N_r antennas and serves $|\mathcal{D}|$ nodes. For the BB-based algorithm, the relative error tolerance ε is set to 0.1.



(a)



(b)



(c)

Fig. 3. Convergence of the BB algorithm. (a) Two-cluster case. (b) Three-cluster case. (c) Four-cluster case.

For the SCA-based distributed algorithm, it terminates when the successive difference of the MSE value is less than 10^{-4}, and its initial step size $\mu(1)$ is set to 2. We name Algorithm 2 and the optimal-beamforming-structure-based distributed algorithm as Direct SCA and OptSCA, respectively. The plots in Fig. 3a-5a are based on a random channel realization. In Fig. 5b-8, 100 independent channel realizations are simulated.

A. Convergence Analysis of the Proposed Algorithms

In this subsection, we show the convergence behaviors of the proposed BB-based and SCA-based algorithms. The transmit SNR P_t/σ_n^2 is set to 20dB and ς is set to 0.4. In Fig. 3, we consider a multi-cluster network with $(N_r,|\mathcal{D}|)=(2,3)$, where $|\mathcal{D}_i\cap\mathcal{C}_j|=1,\forall i\neq j$. We observe that the upper bound

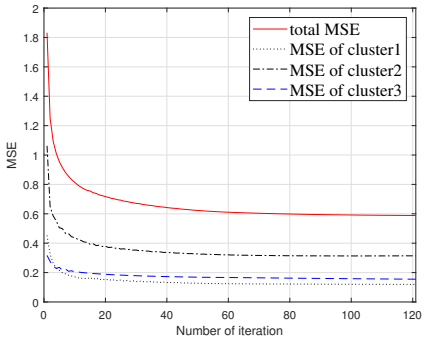


Fig. 4. Convergence of the Direct SCA algorithm in the three-cluster case.

is non-increasing and the lower bound is non-decreasing. During the first few iterations, the gap between the upper bound and lower bound is reduced rapidly due to a large number of infeasible subregions being discarded. The gap becomes smaller as the number of iterations increases until the convergence. As expected, the BB-based algorithm requires many iterations to converge. We observe that the number of iterations increases sharply as the number of clusters increases. This is because that the maximum number of iterations of the BB algorithm in (26) is an exponential function of the number of nodes in the network. In contrast, the SCA algorithm converges much faster and achieves a objective value very close to the BB-based algorithm. Fig. 3 demonstrates the effectiveness of the SCA algorithm. The BB-based algorithm is not practical due to its high complexity, but it can be used as a network performance benchmark.

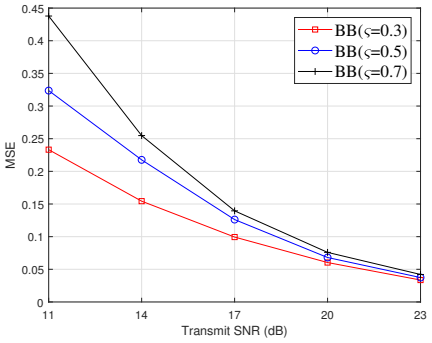
In Fig. 4, we consider a three-cluster network with $(N_r, |\mathcal{D}|) = (4, 8)$, where $|\mathcal{D}_i \cap \mathcal{C}_j| = 3, \forall i \neq j$. We can see that the MSE of each FC is monotonically non-increasing over iteration, which demonstrates that the SCA algorithm can reduce the overall MSE in a pairwise manner.

B. Computation Performance of Multi-Cluster CoMAC

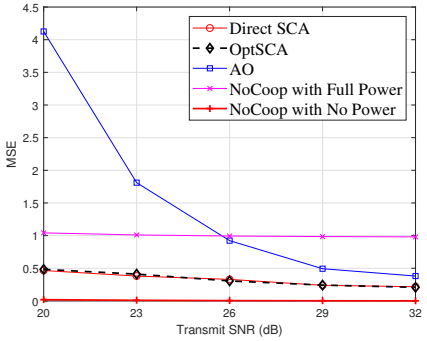
In this subsection, we analyze the performance of the SCA algorithms for different numbers of nodes, numbers of receive antennas and transmit SNRs in the three-cluster networks with $\varsigma = 0.5$. Since the complexity of BB algorithm is high in large-scale networks, we only consider the effect of different transmit SNRs on the MSE performance in a two-cluster network with $\varsigma = 0.5$. We also consider the following for comparison:

a) Transceiver design without cooperative interference management: Since the FCs don't cooperate with each other, each FC is unaware of the actual signal power of its undesired nodes. When these nodes transmit interfering signals at full power, P1 is converted to

$$\begin{aligned} \text{P5}: \min_{\{\mathbf{a}_k\}} & \sum_{k=1}^{K} \sum_{l \in \mathcal{I}_k} \sum_{j \in \mathcal{C}_k \cap \mathcal{D}_l} \left\| \mathbf{a}_k^H \mathbf{h}_{l,j,k} \right\|^2 + \sum_{k=1}^{K} \frac{\sigma_n^2}{P_t} \|\mathbf{a}_k\|^2 \\ \text{s.t.} & \left\| \mathbf{a}_k^H \mathbf{h}_{k,j,k} \right\|^2 \geqslant 1, \forall j \in \mathcal{D}_k, \forall k. \end{aligned} \tag{51}$$



(a)



(b)

Fig. 5. The effects of transmit SNR on the MSE of multi-cluster CoMAC.

When these nodes transmit signals at very low power, the inter-cluster interference is approximately zero, then P1 is converted to

$$\begin{aligned} \text{P6}: \min_{\{\mathbf{a}_k\}} & \sum_{k=1}^{K} \frac{\sigma_n^2}{P_t} \|\mathbf{a}_k\|^2 \\ \text{s.t.} & \left\| \mathbf{a}_k^H \mathbf{h}_{k,j,k} \right\|^2 \geqslant 1, \forall j \in \mathcal{D}_k, \forall k. \end{aligned} \tag{52}$$

P5 and P6 are quadratically constrained quadratic programming problems, which can be solved by the BB algorithm in [31] to get a optimal value with the relative error tolerance set to 10^{-3}. We name these two cases NoCoop with Full Power and NoCoop with No Power, respectively.

b) Transceiver design based on alternating optimization (AO for short): Introduce new optimizing variables $\tilde{\mathbf{a}}_k = \mathbf{a}_k / \sqrt{\eta_k}, \forall k$. The MSE in (11) is convex over each of the transmit scalars $\{w_{k,i}\}$ or receive vectors $\{\tilde{\mathbf{a}}_k\}$, but not jointly convex. Then we can adopt an alternating optimization method in [41] to design transceivers for multi-cluster CoMAC, i.e., with fixed transmit scalars, the optimal receive vectors can be expressed as

$$\tilde{\mathbf{a}}_k = (\mathbf{B}_k)^{-1} \left(\sum_{j \in \mathcal{D}_k} \mathbf{h}_{k,j,k} w_{k,j} \right) \tag{53}$$

where $\mathbf{B}_k = \sigma_n^2 \mathbf{I} + \sum_{l \in \mathcal{I}_k} \sum_{j \in \mathcal{C}_k \cap \mathcal{D}_l} |w_{l,j}|^2 \mathbf{h}_{l,j,k} \mathbf{h}_{l.j.k}^H + \sum_{j \in \mathcal{D}_k} |w_{k,j}|^2 \mathbf{h}_{k,j,k} \mathbf{h}_{k,j,k}^H$. And with fixed receive vectors, the optimal transmit scalars are given by

$$w_{k,j} = \left(\left\| \mathbf{h}_{k,j,k}^H \tilde{\mathbf{a}}_k \right\|^2 + \mu_{k,j} \right)^{-1} \mathbf{h}_{k,j,k}^H \tilde{\mathbf{a}}_k, \forall j \in \mathcal{E}_k, \forall k, \tag{54}$$

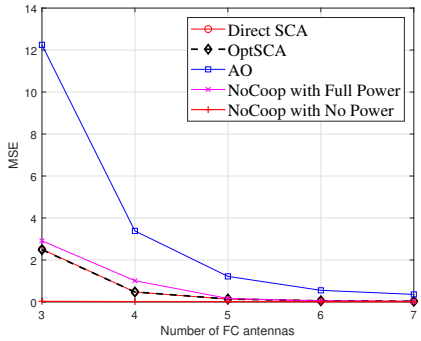


Fig. 6. The effects of antenna numbers on the MSE of multi-cluster CoMAC.

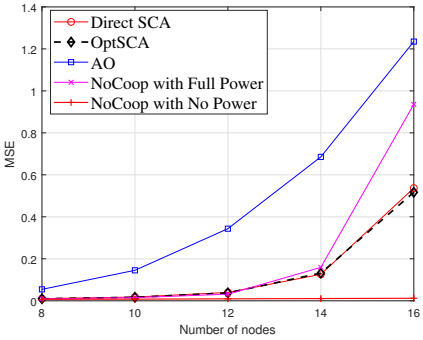


Fig. 7. The effects of node numbers on the MSE of multi-cluster CoMAC.

$$w_{k,j}=\left(\left\|\mathbf{h}_{k,j,k}^{H}\tilde{\mathbf{a}}_k\right\|^2+\mu_{k,j}+\sum_{l\in\mathcal{I}_{k,j}^{'}}\left\|\mathbf{h}_{k,j,l}^{H}\tilde{\mathbf{a}}_l\right\|^2\right)^{-1}\mathbf{h}_{k,j,k}^{H}\tilde{\mathbf{a}}_k,\quad \forall j\in\mathcal{D}_k/\mathcal{E}_k,\forall k, \tag{55}$$

where $\mu_{k,j}\geq 0$ and satisfies $\mu_{k,j}\left(|w_{k,j}|^2-P_t\right)=0$, and $\mathcal{I}_{k,j}^{'}$ denotes the set of clusters interfered by the (k,j)-th node. It can be noticed that the global CSI acquisition of all nodes is essential for the AO method. The iterative algorithm is set to have the same termination conditions as the SCA algorithms.

Fig. 5a illustrates the effect of different transmit SNRs and different ς on the MSE performance in $(2,2)$ networks, where $|\mathcal{D}_i\cap\mathcal{C}_j|=1,\forall i\neq j$. Fig. 5b compares the MSE performance of different transceiver design schemes under different transmit SNRs in $(4,10)$ networks, where $|\mathcal{D}_i\cap\mathcal{C}_j|=2,\forall i\neq j$. It can be observed that MSE is a decreasing function of transmit SNR, and the four curves approach a lower minima when transmit SNR is high, which means that further increasing transmit SNR will not significantly suppress the inter-cluster interference, and the MSE performance will only be marginally improved. Besides, the performance of OptSCA method is near-identical to its direct method (Direct SCA), and is much better than the AO method in low transmit SNR conditions. The performance of transceiver design based on AO is even worse than that of cases of NoCoop with Full Power at low transmit SNRs, which demonstrates the effectiveness of the uniform-forcing transmitter design.

In Fig. 6, we study the MSE performance of different transceiver designs for different numbers of antennas, where transmit $\mathrm{SNR}=20\mathrm{dB}, |\mathcal{D}|=10$ and $|\mathcal{D}_i\cap\mathcal{C}_j|=2,\forall i\neq j$. It is observed that the MSEs monotonically decrease with the receive antenna number. This is because that more receive antennas enable FCs to obtain higher diversity gain. When the receive antenna number is small, the proposed transceiver design can achieve significantly better MSE performance than the AO method, which demonstrates the effectiveness of our proposed scheme.

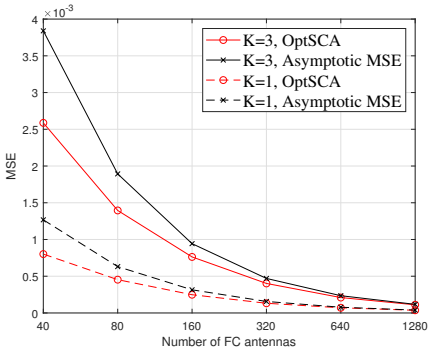


Fig. 8. The computation MSE performance of (50) versus N_r.

The MSE versus different numbers of nodes is shown in Fig. 7, where transmit $\mathrm{SNR}=20\mathrm{dB}$, $|N_r|=9$. The number of non-common nodes served by each FC is set to 6, and that of common nodes increases from 2 to 10. Fig. 7 shows that the MSE is an increasing function of the node number. This is because more nodes make it harder to design one common receiver to equalize the channels of different nodes. We can also observe that the proposed transceiver design is better than that in the case of NoCoop with Full Power, and the MSE gap between these two cases will increase with the node number. This is because that the cooperative interference management among FCs can significantly reduce the MSE for multi-cluster CoMAC networks, especially when there are a large number of common nodes in each cluster, making inter-cell interference the main reason that affects the MSE performance.

In Fig. 8, we analyze the accuracy of the asymptotic MSE in (50) by comparing with the one obtained from OptSCA, where we set transmit $\mathrm{SNR}=20\mathrm{dB}$, $|\mathcal{D}|=10$ and $|\mathcal{D}_i\cap\mathcal{C}_j|=2,\forall i\neq j$. It can be observed from Fig. 8 that the gap between the asymptotic MSE and the one computed by SCA decreases as N_r increases. When the number of antennas is not large enough, the asymptotic MSE performance in (50) is slightly worse than that obtained from the SCA method, which makes sense because (50) holds in the case of massive antennas. When $N_r\geq 320$, the gap becomes negligible, which verifies our conclusion in Proposition 6.

VI. Conclusion

In this work, we have studied the transceiver design of CoMAC in multi-cluster wireless networks for combating inter-cluster interference and non-uniform fading. We have formulated a QSRP to minimize the sum-MSE of signals aggregated at different FCs subject to the peak power constraints of nodes. The formulated nonconvex problem has

been optimally solved by the proposed BB-based scheme. We have also developed a distributed algorithm, in which the master problem and subproblems have been solved iteratively to obtain a high-performance solution with lower complexity. Through asymptotic analysis, we have provided simple closed-form asymptotically optimal receivers for the formulated problem. According to numerical results, our proposed distributed algorithm can significantly reduce the sum-MSE of multi-cluster CoMAC networks in a cooperative way.

Appendix A
Proof of Proposition 1

Substituting (12) and (13) into (11), MSE can be computed as

$$\mathrm{MSE}=\sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\frac{\left(\min\limits_{i\in\mathcal{D}_l}\left\|\mathbf{a}_l^H\mathbf{h}_{l,i,l}\right\|^2\right)\left\|\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right\|^2}{\left(\min\limits_{i\in\mathcal{D}_k}\left\|\mathbf{a}_k^H\mathbf{h}_{k,i,k}\right\|^2\right)\left\|\mathbf{a}_l^H\mathbf{h}_{l,j,l}\right\|^2}+\sum_{k=1}^{K}\frac{\sigma_n^2\|\mathbf{a}_k\|^2}{P_t\min\limits_{i\in\mathcal{D}_k}\left\|\mathbf{a}_k^H\mathbf{h}_{k,i,k}\right\|^2}. \tag{56}$$

By introducing auxiliary variables $\gamma_k=\min\limits_{i\in\mathcal{D}_k}\left\|\mathbf{a}_k^H\mathbf{h}_{k,i,k}\right\|^2$ for each k, the minimum problem of MSE in (56) can be transformed into

$$\begin{aligned}\min_{\{\mathbf{a}_k\},\{\gamma_k\}}\ &\sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\frac{\gamma_l\left\|\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right\|^2}{\gamma_k\left\|\mathbf{a}_l^H\mathbf{h}_{l,j,l}\right\|^2}+\sum_{k=1}^{K}\frac{\sigma_n^2\|\mathbf{a}_k\|^2}{P_t\gamma_k}\\ \text{s.t. }&\left\|\mathbf{a}_k^H\mathbf{h}_{k,j,k}\right\|^2\geqslant\gamma_k,\forall j\in\mathcal{D}_k,\forall k.\end{aligned} \tag{57}$$

Then introducing new optimizing variables $\tilde{\mathbf{a}}_k=\mathbf{a}_k/\sqrt{\gamma_k}$ for each k, the above problem can be converted to

$$\begin{aligned}\min_{\{\tilde{\mathbf{a}}_k\}}\ &\sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\frac{\left\|\tilde{\mathbf{a}}_k^H\mathbf{h}_{l,j,k}\right\|^2}{\left\|\tilde{\mathbf{a}}_l^H\mathbf{h}_{l,j,l}\right\|^2}+\sum_{k=1}^{K}\frac{\sigma_n^2}{P_t}\|\tilde{\mathbf{a}}_k\|^2\\ \text{s.t. }&\left\|\tilde{\mathbf{a}}_k^H\mathbf{h}_{k,j,k}\right\|^2\geqslant 1,\forall j\in\mathcal{D}_k,\forall k,\end{aligned} \tag{58}$$

which completes the proof.

Appendix B
Proof of Proposition 3

For each $j\in\mathcal{E}_k$, since $\hat{\mathbf{a}}_k\in Conv(S_{k,j}^{[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]})$, we can easily conclude that $\min\left\|\hat{\mathbf{a}}_k^H\mathbf{h}_{k,j,k}\right\|=\cos\left((\overline{\varphi}_{k,j}-\underline{\varphi}_{k,j})/2\right)$, which implies that

$$\left\|\hat{\mathbf{a}}_k^H\mathbf{h}_{k,j,k}\right\|^2\geq\cos^2\left(\frac{\overline{\varphi}_{k,j}-\underline{\varphi}_{k,j}}{2}\right)\geq\cos^2(\delta). \tag{59}$$

And for each $j\in\mathcal{D}_k/\mathcal{E}_k$, since $\hat{\mathbf{a}}_k\in Conv\left(Y_{k,j}^{[\underline{\varphi}_{k,j},\overline{\varphi}_{k,j}]}\right)$, we can also conclude that $\min\left\|\hat{\mathbf{a}}_k^H\mathbf{h}_{k,j,k}\right\|=\frac{1}{\sqrt{u_{k,j,k}}}\cdot\cos\left((\overline{\varphi}_{k,j}-\underline{\varphi}_{k,j})/2\right)$, which implies that

$$\left\|\hat{\mathbf{a}}_k^H\mathbf{h}_{k,j,k}\right\|^2\geq\frac{1}{u_{k,j,k}}\cos^2\left(\frac{\overline{\varphi}_{k,j}-\underline{\varphi}_{k,j}}{2}\right)\geq\frac{1}{u_{k,j,k}}\cos^2(\delta). \tag{60}$$

Combined with (23), we obtain

$$\|\overline{\mathbf{a}}_k\|^2\leq\frac{\|\hat{\mathbf{a}}_k\|^2}{\cos^2(\delta)},\forall k. \tag{61}$$

Since $\{\hat{\mathbf{a}}_k\}$ is the optimal solution to P2, the upper bound $ub(V^t)$ is given in (24) and the lower bound is given as

$$lb\left(V^t\right)=\sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}l_{l,j,l}\left\|\hat{\mathbf{a}}_k^H\mathbf{h}_{l,j,k}\right\|^2+\sum_{k=1}^{K}\frac{\sigma_n^2}{P_t}\|\hat{\mathbf{a}}_k\|^2. \tag{62}$$

Finally, the gap between the upper bound $ub(V^t)$ and the lower bound $lb(V^t)$ is given by (63), as shown at top of the next page. The fourth inequality in (63) uses the fact that $l_{l,j,l}>0$ holds for all common nodes and we can always find a positive number satisfying $\sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}2/l_{l,j,l}\leq T_m$.

Since the function $f(\delta)=p^{'}/\cos^2(\delta)+T_m\delta/\cos^2(\delta)-p^{'}+K\tan^2(\delta)$ is monotonically increasing for all $\delta\in(0,\pi/2)$ and $f(0)=0$, we can always find a δ satisfying $f(\delta)\leq\varepsilon$ for any given ε by bisection search.

Appendix C
Proof of Lemma 3

Define

$$f_0\left(\{\mathbf{a}_k\}\right)=\sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\frac{\left\|\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right\|^2}{\left\|\mathbf{a}_l^H\mathbf{h}_{l,j,l}\right\|^2}+\sum_{k=1}^{K}\frac{\sigma_n^2}{P_t}\|\mathbf{a}_k\|^2 \tag{64}$$

and

$$f_1\left(\{\mathbf{a}_k\},\boldsymbol{\tau}\right)=\sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap D_l}\frac{\left\|\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right\|^2}{\tau_{l,j}}+\sum_{k=1}^{K}\frac{\sigma_n^2}{P_t}\|\mathbf{a}_k\|^2. \tag{65}$$

Let $(\{\mathbf{a}_k^\star\},\boldsymbol{\tau}^\star)$ be a global optimal solution to P3. From (27) we can get

$$\frac{\left\|\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right\|^2}{\tau_{l,j}}\geq\frac{\left\|\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right\|^2}{\left\|\mathbf{a}_l^H\mathbf{h}_{l,j,l}\right\|^2},\forall k,\forall l\in\mathcal{I}_k,\forall j\in\mathcal{C}_k\cap\mathcal{D}_l, \tag{66}$$

so $f_0(\{\mathbf{a}_k^\star\})\leq f_1(\{\mathbf{a}_k^\star\},\boldsymbol{\tau}^\star)$ holds. Let $\hat{\tau}_{l,j}=\left\|(\mathbf{a}_l^\star)^H\mathbf{h}_{l,j,l}\right\|^2,\forall l,\forall j\in\mathcal{D}_l/\mathcal{E}_l$, then $(\{\mathbf{a}_k^\star\},\hat{\boldsymbol{\tau}})$ is a feasible solution to P3. Since $(\{\mathbf{a}_k^\star\},\boldsymbol{\tau}^\star)$ is globally optimal, this implies that $f_0(\{\mathbf{a}_k^\star\})=f_1(\{\mathbf{a}_k^\star\},\hat{\boldsymbol{\tau}})\geq f_1(\{\mathbf{a}_k^\star\},\boldsymbol{\tau}^\star)$ holds. Therefore, $f_0(\{\mathbf{a}_k^\star\})=f_1(\{\mathbf{a}_k^\star\},\boldsymbol{\tau}^\star)$ holds. Combined with (27) and (66), it follows that

$$\left\|(\mathbf{a}_l^\star)^H\mathbf{h}_{l,j,l}\right\|^2=\tau_{l,j}^\star,\forall l,\forall j\in\mathcal{D}_l/\mathcal{E}_l. \tag{67}$$

Let $\tau_{l,j}=\left\|\mathbf{a}_l^H\mathbf{h}_{l,j,l}\right\|^2,\forall l,\forall j\in\mathcal{D}_l/\mathcal{E}_l$, then $(\{\mathbf{a}_k\},\boldsymbol{\tau})$ is a feasible solution to P3. Combined with (67), $f_0(\{\mathbf{a}_k^\star\})\leq f_0(\{\mathbf{a}_k\})$ holds, which implies $\{\mathbf{a}_k^\star\}$ is a global optimal solution to P1.

Conversely, let $\{\mathbf{a}_k^\star\}$ be a global optimal solution to P1, and let $\tau_{l,j}^\star=\left\|(\mathbf{a}_l^\star)^H\mathbf{h}_{l,j,l}\right\|^2,\forall l,\forall j\in\mathcal{D}_l/\mathcal{E}_l$, then $(\{\mathbf{a}_k^\star\},\boldsymbol{\tau}^\star)$ is a feasible solution to P3. For some feasible solution $(\{\mathbf{a}_k\},\boldsymbol{\tau})$ to P3, due to (66), we have $f_0(\{\mathbf{a}_k\})\leq f_1(\{\mathbf{a}_k\},\boldsymbol{\tau})$. Since $\{\mathbf{a}_k^\star\}$ is a global optimal solution to P1, it follows that $f_0(\{\mathbf{a}_k^\star\})\leq f_0(\{\mathbf{a}_k\})$. Combined with the fact $f_0(\{\mathbf{a}_k\})\leq$

$$
\begin{aligned}
\frac{ub\left(V^t\right)-lb\left(V^t\right)}{lb\left(V^t\right)} &\leq \sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\frac{\left\|\bar{\mathbf{a}}_k^H\mathbf{h}_{l,j,k}\right\|^2-l_{l,j,l}\left\|\hat{\mathbf{a}}_k^H\mathbf{h}_{l,j,k}\right\|^2}{\left\|\bar{\mathbf{a}}_l^H\mathbf{h}_{l,j,l}\right\|^2-l_{l,j,l}\left\|\hat{\mathbf{a}}_k^H\mathbf{h}_{l,j,k}\right\|^2}+\sum_{k=1}^{K}\frac{\left\|\bar{\mathbf{a}}_k\right\|^2-\left\|\hat{\mathbf{a}}_k\right\|^2}{\left\|\hat{\mathbf{a}}_k\right\|^2}\\
&\leq \sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\frac{u_{l,j,l}\frac{\left\|\hat{\mathbf{a}}_k^H\mathbf{h}_{l,j,k}\right\|^2}{\cos^2(\delta)}-l_{l,j,l}\left\|\hat{\mathbf{a}}_k^H\mathbf{h}_{l,j,k}\right\|^2}{l_{l,j,l}\left\|\hat{\mathbf{a}}_k^H\mathbf{h}_{l,j,k}\right\|^2}+\sum_{k=1}^{K}\frac{\frac{\left\|\hat{\mathbf{a}}_k\right\|^2}{\cos^2(\delta)}-\left\|\hat{\mathbf{a}}_k\right\|^2}{\left\|\hat{\mathbf{a}}_k\right\|^2}\\
&\leq \sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\left(\frac{1}{\cos^2(\delta)}+\frac{2\delta}{l_{l,j,l}\cos^2(\delta)}-1\right)+\sum_{k=1}^{K}\tan^2(\delta)\\
&\leq \frac{p^{'}}{\cos^2(\delta)}+\frac{T_m\delta}{\cos^2(\delta)}-p^{'}+K\tan^2(\delta)\\
&= \varepsilon
\end{aligned}
\tag{63}
$$

$f_1\left(\{\mathbf{a}_k\},\boldsymbol{\tau}\right)$, we can get $f_0\left(\{\mathbf{a}_k^\star\}\right)\leq f_1\left(\{\mathbf{a}_k\},\boldsymbol{\tau}\right)$. Since $\tau_{l,j}^\star=\left\|(\mathbf{a}_l^\star)^H\mathbf{h}_{l,j,l}\right\|^2,\forall l,\forall j\in\mathcal{D}_l/\mathcal{E}_l$, it follows that $f_1\left(\{\mathbf{a}_k^\star\},\boldsymbol{\tau}^\star\right)\leq f_1\left(\{\mathbf{a}_k\},\boldsymbol{\tau}\right)$, which implies that $\left(\{\mathbf{a}_k^\star\},\boldsymbol{\tau}^\star\right)$ is a global optimal solution to P3 and the proof is completed.

Appendix D
Proof of Proposition 5

Define $f_2^j\left(\mathbf{a}_k\right)=2\Re\left\{\mathbf{a}_k^H\mathbf{h}_{k,j,k}\mathbf{h}_{k,j,k}^H\mathbf{c}_k^{(s)}\right\}-\left\|\mathbf{h}_{k,j,k}^H\mathbf{c}_k^{(s)}\right\|^2$. The Lagrangian of $\mathbf{P}_k^{\mathrm{sub}}\left(\boldsymbol{\tau},\mathbf{c}_k^{(s)}\right)$ is given by

$$
\begin{aligned}
&\mathcal{L}_k\left(\mathbf{a}_k,\mathbf{q}_k,\boldsymbol{\lambda}_k,\{\mathbf{T}_{l,j,k}\}\right)=\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}q_{l,j,k}+\frac{\sigma_n^2}{P_t}\|\mathbf{a}_k\|^2\\
&+\sum_{j\in\mathcal{E}_k}\lambda_{k,j}\left[1-f_2^j\left(\mathbf{a}_k\right)\right]+\sum_{j\in\mathcal{D}_k/\mathcal{E}_k}\lambda_{k,j}\left[\tau_{k,j}-f_2^j\left(\mathbf{a}_k\right)\right]\\
&-\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\mathrm{tr}\left(\mathbf{T}_{l,j,k}\begin{bmatrix}q_{l,j,k} & \mathbf{a}_k^H\mathbf{h}_{l,j,k}\\ \left(\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right)^H & \tau_{l,j}\end{bmatrix}\right),
\end{aligned}
\tag{68}
$$

where $\boldsymbol{\lambda}_k$ and $\{\mathbf{T}_{l,j,k}\}$ denote the lagrange multipliers, $\boldsymbol{\lambda}_k$ is composed of $\lambda_{k,j},\forall j\in\mathcal{D}_k$, and

$$
\mathbf{T}_{l,j,k}=\begin{bmatrix}t_{l,j,k}^{(1)} & t_{l,j,k}^{(2)}\\ t_{l,j,k}^{(3)} & t_{l,j,k}^{(4)}\end{bmatrix}\succcurlyeq\mathbf{0}.
\tag{69}
$$

The dual function is given by

$$
\begin{aligned}
&d_k\left(\boldsymbol{\lambda}_k,\{\mathbf{T}_{l,j,k}\}\right)=\min_{\mathbf{a}_k,\mathbf{q}_k}\mathcal{L}_k\left(\mathbf{a}_k,\mathbf{q}_k,\boldsymbol{\lambda}_k,\{\mathbf{T}_{l,j,k}\}\right)\\
&=\left(\sum_{j\in\mathcal{D}_k/\mathcal{E}_k}\lambda_{k,j}\mathbf{e}_{k,j}^T-\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}t_{l,j,k}^{(4)}\mathbf{e}_{l,j}^T\right)\boldsymbol{\tau}\\
&\quad+h_k\left(\boldsymbol{\lambda}_k,\{\mathbf{T}_{l,j,k}\}\right),
\end{aligned}
\tag{70}
$$

where $\tau_{k,j}=\mathbf{e}_{k,j}^T\boldsymbol{\tau},\forall j\in\mathcal{D}_k/\mathcal{E}_k$, $\tau_{l,j}=\mathbf{e}_{l,j}^T\boldsymbol{\tau},\forall l\in\mathcal{I}_k,\forall j\in\mathcal{C}_k\cap\mathcal{D}_l$, and

$$
\begin{aligned}
&h_k\left(\boldsymbol{\lambda}_k,\{\mathbf{T}_{l,j,k}\}\right)=\min_{\mathbf{a}_k,\mathbf{q}_k}\Bigg(\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}q_{l,j,k}+\frac{\sigma_n^2}{P_t}\|\mathbf{a}_k\|^2+\\
&\sum_{j\in\mathcal{E}_k}\lambda_{k,j}\left[1-f_2^j\left(\mathbf{a}_k\right)\right]-\sum_{j\in\mathcal{D}_k/\mathcal{E}_k}\lambda_{k,j}f_2^j\left(\mathbf{a}_k\right)-\sum_{l\in\mathcal{I}_k}\\
&\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}\left(t_{l,j,k}^{(1)}q_{l,j,k}+t_{l,j,k}^{(2)}\left(\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right)^H+t_{l,j,k}^{(3)}\mathbf{a}_k^H\mathbf{h}_{l,j,k}\right)\Bigg).
\end{aligned}
\tag{71}
$$

Denote $\boldsymbol{\lambda}_k^\star$ and $\left\{\mathbf{T}_{l,jk}^\star\right\}$ as the optimal lagrange multipliers for the dual problem. Since each subproblem is convex, we have

$$
\begin{aligned}
P_k^\star\left(\boldsymbol{\tau}\right)&=\max_{\boldsymbol{\lambda}_k,\{\mathbf{T}_{l,j,k}\}}d_k\left(\boldsymbol{\lambda}_k,\{\mathbf{T}_{l,j,k}\}\right)\\
&=d_k\left(\boldsymbol{\lambda}_k^\star,\left\{\mathbf{T}_{l,j,k}^\star\right\}\right)\\
&=\left(\sum_{j\in\mathcal{D}_k/\mathcal{E}_k}\lambda_{k,j}^\star\mathbf{e}_{k,j}^T-\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}t_{l,j,k}^{(4)\star}\mathbf{e}_{l,j}^T\right)\boldsymbol{\tau}\\
&\quad+h_k\left(\boldsymbol{\lambda}_k^\star,\left\{\mathbf{T}_{l,j,k}^\star\right\}\right).
\end{aligned}
\tag{72}
$$

Denoting $\mathbf{g}_k$ as

$$
\mathbf{g}_k=\sum_{j\in\mathcal{D}_k/\mathcal{E}_k}\lambda_{k,j}^\star\mathbf{e}_{k,j}-\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l}t_{l,j,k}^{(4)\star}\mathbf{e}_{l,j},
\tag{73}
$$

we have

$$
\begin{aligned}
P_k^\star\left(\boldsymbol{\tau}\right)&=\mathbf{g}_k^T\boldsymbol{\tau}+h_k\left(\boldsymbol{\lambda}_k^\star,\left\{\mathbf{T}_{l,j,k}^\star\right\}\right)\\
&=\mathbf{g}_k^T\left(\boldsymbol{\tau}-\tilde{\boldsymbol{\tau}}\right)+\mathbf{g}_k^T\tilde{\boldsymbol{\tau}}+h_k\left(\boldsymbol{\lambda}_k^\star,\left\{\mathbf{T}_{l,j,k}^\star\right\}\right)\\
&\leq\mathbf{g}_k^T\left(\boldsymbol{\tau}-\tilde{\boldsymbol{\tau}}\right)+P_k^\star\left(\tilde{\boldsymbol{\tau}}\right),
\end{aligned}
\tag{74}
$$

which can be written as

$$
P_k^\star\left(\tilde{\boldsymbol{\tau}}\right)\geq P_k^\star\left(\boldsymbol{\tau}\right)+\mathbf{g}_k^T\left(\tilde{\boldsymbol{\tau}}-\boldsymbol{\tau}\right).
\tag{75}
$$

It follows that $\mathbf{g}_k$ is the subgradient of $P_k^\star\left(\boldsymbol{\tau}\right)$ and can be obtained from the k-th subproblem.

In the same way, we can compute the global subgradient $\mathbf{g}$ of $\mathbf{P}^{\max}$ as

$$\begin{aligned}\mathbf{g} &= \sum_{k=1}^{K} \sum_{j\in\mathcal{D}_k/\mathcal{E}_k} \lambda_{k,j}^{\star}\mathbf{e}_{k,j} - \sum_{k=1}^{K}\sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l} t_{l,j,k}^{(4)\star}\mathbf{e}_{l,j} \\ &= \sum_{k=1}^{K}\left(\sum_{j\in\mathcal{D}_k/\mathcal{E}_k} \lambda_{k,j}^{\star}\mathbf{e}_{k,j} - \sum_{l\in\mathcal{I}_k}\sum_{j\in\mathcal{C}_k\cap\mathcal{D}_l} t_{l,j,k}^{(4)\star}\mathbf{e}_{l,j}\right) \qquad (76) \\ &= \sum_{k=1}^{K}\mathbf{g}_k,\end{aligned}$$

and the proof is completed.

REFERENCES

[1] M. Agiwal, A. Roy, and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart. 2016.

[2] B. Nazer and M. Gastpar, "Computation Over Multiple-Access Channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.

[3] M. Goldenbaum, H. Boche, and S. Stańczak, "Nomographic Functions: Efficient Computation in Clustered Gaussian Sensor Networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2093–2105, Apr. 2015.

[4] J. Kampeas, A. Cohen, and O. Gurewitz, "The Ergodic Capacity of the Multiple Access Channel Under Distributed Scheduling - Order Optimality of Linear Receivers," *IEEE Trans. Inf. Theory*, vol. 64, no. 8, pp. 5898–5919, Aug. 2018.

[5] O. Shmuel, A. Cohen, and O. Gurewitz, "Compute-and-Forward in Large Relaying Systems: Limitations and Asymptotically Optimal Scheduling," *IEEE Trans. Inf. Theory*, vol. 67, no. 9, pp. 6243–6265, Sep. 2021.

[6] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate Region of the Quadratic Gaussian Two-Encoder Source-Coding Problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938–1961, May 2008.

[7] R. Soundararajan and S. Vishwanath, "Communicating Linear Functions of Correlated Gaussian Sources Over a MAC," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1853–1860, Mar. 2012.

[8] M. Gastpar, "Uncoded Transmission Is Exactly Optimal for a Simple Gaussian "Sensor" Network," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5247–5251, Nov. 2008.

[9] M. Goldenbaum and S. Stanczak, "Robust Analog Function Computation via Wireless Multiple-Access Channels," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3863–3877, Sep. 2013.

[10] M. Goldenbaum, H. Boche, and S. Stańczak, "Harnessing Interference for Analog Function Computation in Wireless Sensor Networks," *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 4893–4906, Oct. 2013.

[11] O. Abari, H. Rahul, and D. Katabi, "Over-the-air Function Computation in Sensor Networks," *arXiv preprint arXiv:1612.02307*, 2016.

[12] L. Chen, X. Qin, and G. Wei, "A Uniform-Forcing Transceiver Design for Over-the-Air Function Computation," *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 942–945, Dec. 2018.

[13] W. Liu, X. Zang, Y. Li, and B. Vucetic, "Over-the-Air Computation Systems: Optimization, Analysis and Scaling Laws," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5488–5502, Aug. 2020.

[14] L. Chen, N. Zhao, Y. Chen, F. R. Yu, and G. Wei, "Over-the-Air Computation for IoT Networks: Computing Multiple Functions With Antenna Arrays," *IEEE Internet Thing J.*, vol. 5, no. 6, pp. 5296–5306, Dec. 2018.

[15] J. Kampeas, A. Cohen, and O. Gurewitz, "On the Outage Probability of Distributed MAC With ZF Detection," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2398–2412, Apr. 2021.

[16] K. Jagannathan, S. Borst, P. Whiting, and E. Modiano, "Efficient Scheduling of Multi-User Multi-Antenna Systems," in *Proc. 4th Int. Symp. Modeling Optimization Mobile Ad Hoc Wireless Netw.*,, 2006, pp. 1–8.

[17] F. Ang, L. Chen, N. Zhao, Y. Chen, and F. R. Yu, "Robust Design for Massive CSI Acquisition in Analog Function Computation Networks," *IEEE Trans. Veh. Technol*, vol. 68, no. 3, pp. 2361–2373, Mar. 2019.

[18] L. Chen, N. Zhao, Y. Chen, X. Qin, and F. R. Yu, "Computation Over MAC: Achievable Function Rate Maximization in Wireless Networks," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5446–5459, Sep. 2020.

[19] L. Chen, N. Zhao, Y. Chen, F. R. Yu, and G. Wei, "Toward Optimal Rate-Delay Tradeoff for Computation Over Multiple Access Channel," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4335–4346, Jul. 2021.

[20] Z. Wang, Y. Shi, Y. Zhou, H. Zhou, and N. Zhang, "Wireless-Powered Over-the-Air Computation in Intelligent Reflecting Surface-Aided IoT Networks," *IEEE Internet Thing J.*, vol. 8, no. 3, pp. 1585–1598, Feb. 2021.

[21] K. Yang, Y. Shi, Y. Zhou, Z. Yang, L. Fu, and W. Chen, "Federated Machine Learning for Intelligent IoT via Reconfigurable Intelligent Surface," *IEEE Netw.*, vol. 34, no. 5, pp. 16–22, Sep. 2020.

[22] Y. Shi, K. Yang, T. Jiang, J. Zhang, and K. B. Letaief, "Communication-Efficient Edge AI: Algorithms and Systems," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2167–2191, 4th Quart. 2020.

[23] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The Roadmap to 6G: AI Empowered Wireless Networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.

[24] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated Learning via Over-the-Air Computation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2022–2035, Mar. 2020.

[25] H. Guo, A. Liu, and V. K. N. Lau, "Analog Gradient Aggregation for Federated Learning Over Wireless Networks: Customized Design and Convergence Analysis," *IEEE Internet Thing J.*, vol. 8, no. 1, pp. 197–210, Jan. 2021.

[26] 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.

[27] Z. Xiang, M. Tao, and X. Wang, "Coordinated Multicast Beamforming in Multicell Networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 12–21, Jan. 2013.

[28] R. Zhang and S. Cui, "Cooperative Interference Management With MISO Beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5450–5458, Oct. 2010.

[29] Q. Lan, H. S. Kang, and K. Huang, "Simultaneous Signal-and-Interference Alignment for Two-Cell Over-the-Air Computation," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1342–1345, Sep. 2020.

[30] X. Cao, G. Zhu, J. Xu, and K. Huang, "Cooperative Interference Management for Over-the-Air Computation Networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2634–2651, Apr. 2021.

[31] C. Lu and Y. Liu, "An Efficient Global Algorithm for Single-Group Multicast Beamforming," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3761–3774, Jul. 2017.

[32] D. P. Palomar and Mung Chiang, "A Tutorial on Decomposition Methods for Network Utility Maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[33] W. Fang, Y. Zou, H. Zhu, Y. Shi, and Y. Zhou, "Optimal Receive Beamforming for Over-the-Air Computation," *arXiv preprint arXiv:2105.05024*, 2021.

[34] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

[35] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.

[36] Y. Shi, J. Zhang, B. O'Donoghue, and K. B. Letaief, "Large-Scale Convex Optimization for Dense Wireless Cooperative Networks," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4729–4743, Sep. 2015.

[37] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

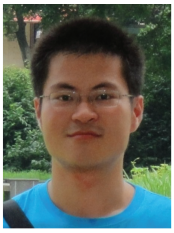
[38] Z. Xiang, M. Tao, and X. Wang, "Massive MIMO Multicasting in Noncooperative Cellular Networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1180–1193, Jun. 2014.

[39] X. Zhai, X. Chen, J. Xu, and D. W. Kwan Ng, "Hybrid Beamforming for Massive MIMO Over-the-Air Computation," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2737–2751, Apr. 2021.

[40] K.-Y. Wang, A. M.-C. So, T.-H. Chang, W.-K. Ma, and C.-Y. Chi, "Outage Constrained Robust Transmit Optimization for Multiuser MISO Downlinks: Tractable Approximations by Conic Optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5690–5705, Nov. 2014.

[41] H. Shen, B. Li, M. Tao, and X. Wang, "MSE-Based Transceiver Designs for the MIMO Interference Channel," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3480–3489, Nov. 2010.

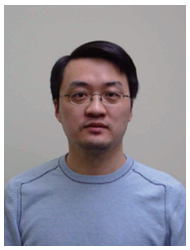
Haoyu Zhang received the B.E. degree in information and communications engineering from Lanzhou University, Lanzhou, China, in 2019. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering and Information Science, University of Science and Technology of China. His research interests include wireless communication and computation, and integrated sensing and communication.



Li Chen received the B.E. degree in electrical and information engineering from Harbin Institute of Technology, Harbin, China, in 2009 and the Ph.D. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2014. He is currently an Associate Professor with the Department of Electronic Engineering and Information Science, University of Science and Technology of China. His research interests include integrated computation and communication, integrated sensing and communication.

Nan Zhao (S'08-M'11-SM'16) is currently a Professor at Dalian University of Technology, China. He received the Ph.D. degree in information and communication engineering in 2011, from Harbin Institute of Technology, Harbin, China.

Dr. Zhao is serving on the editorial boards of IEEE Wireless Communications, IEEE Wireless Communications Letters and IEEE Transactions on Green Communications and Networking. He won the best paper awards in IEEE VTC 2017 Spring, ICNC 2018, WCSP 2018 and WCSP 2019. He also received the IEEE Communications Society Asia Pacific Board Outstanding Young Researcher Award in 2018.



Yunfei Chen (S'02-M'06-SM'10) received his B.E. and M.E. degrees in electronics engineering from Shanghai Jiaotong University, Shanghai, P.R.China, in 1998 and 2001, respectively. He received his Ph.D. degree from the University of Alberta in 2006. He is currently working as an Associate Professor at the University of Warwick, U.K. His research interests include wireless communications, cognitive radios, wireless relaying and energy harvesting.

F. Richard Yu (S'00-M'04-SM'08-F'18) received the PhD degree in electrical engineering from the University of British Columbia (UBC) in 2003. From 2002 to 2006, he was with Ericsson (in Lund, Sweden) and a start-up in California, USA. He joined Carleton University in 2007, where he is currently a Professor. He received the IEEE Outstanding Service Award in 2016, IEEE Outstanding Leadership Award in 2013, Carleton Research Achievement Award in 2012, the Ontario Early Researcher Award (formerly Premiers Research Excellence Award) in 2011, the Excellent Contribution Award at IEEE/IFIP TrustCom 2010, the Leadership Opportunity Fund Award from Canada Foundation of Innovation in 2009 and the Best Paper Awards at IEEE ICNC 2018, VTC 2017 Spring, ICC 2014, Globecom 2012, IEEE/IFIP TrustCom 2009 and Int'l Conference on Networking 2005. His research interests include wireless cyber-physical systems, connected/autonomous vehicles, security, distributed ledger technology, and deep learning.

He serves on the editorial boards of several journals, including Co-Editor-in-Chief for Ad Hoc & Sensor Wireless Networks, Lead Series Editor for IEEE Transactions on Vehicular Technology, IEEE Transactions on Green Communications and Networking, and IEEE Communications Surveys & Tutorials. He has served as the Technical Program Committee (TPC) Co-Chair of numerous conferences. Dr. Yu is a registered Professional Engineer in the province of Ontario, Canada, a Fellow of the Institution of Engineering and Technology (IET), and a Fellow of the IEEE. He is a Distinguished Lecturer, the Vice President (Membership), and an elected member of the Board of Governors (BoG) of the IEEE Vehicular Technology Society.