Open access • Journal Article • DOI:10.1109/TSP.2017.2742984

# Multigroup Multicast Beamformer Design for MISO-OFDM With Antenna Selection



Ganesh Venkatraman, Antti Tolli, Markku Juntti, Le-Nam Tran

**Institutions:** University of Oulu, University College Dublin

**Published on:** 15 Nov 2017 - IEEE Transactions on Signal Processing (IEEE)

**Topics:** Relaxation (approximation), Semidefinite programming, Quadratic programming, Computational complexity theory and Multicast

| | |
|---|---|
| ***Title*** | Multigroup Multicast Beamformer Design for MISO-OFDM With Antenna Selection |
| ***Authors(s)*** | Venkatraman, Ganesh; Tolli, Antti; Juntti, Markku; Tran, Le-Nam |
| ***Publication date*** | 2017-08-21 |
| ***Publication information*** | IEEE Transactions on Signal Processing, 65 (22): 5832-5847 |
| ***Publisher*** | IEEE |
| ***Item record/more information*** | http://hdl.handle.net/10197/10330 |
| ***Publisher's statement*** | © 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. |
| ***Publisher's version (DOI)*** | 10.1109/tsp.2017.2742984 |

# Multi-Group Multicast Beamformer Design for MISO-OFDM with Antenna Selection

Ganesh Venkatraman *Student Member, IEEE*, Antti Tölli *Senior Member, IEEE*, Markku Juntti *Senior Member, IEEE*, and Le-Nam Tran *Senior Member, IEEE*

***Abstract*—We study the problem of designing transmit beamformers for a multi-group multicasting by considering a multiple-input single-output (MISO) orthogonal frequency division multiplexing (OFDM) framework. The design objective involves either minimizing the total transmit power for certain guaranteed quality-of-service (QoS) or maximizing the minimum achievable rate among the users for a given transmit power budget. The problem of interest can be formulated as a nonconvex quadratically constrained quadratic programming (QCQP) for which the prevailing semidefinite relaxation (SDR) technique is inefficient for at least two reasons. At first, the relaxed problem cannot be reformulated as a semidefinite programming. Secondly, even if the relaxed problem is solved, the so-called randomization procedure should be used to generate a feasible solution to the original QCQP, which is difficult to derive for the considered problem. To overcome these shortcomings, we adopt successive convex approximation (SCA) framework to find multicast beamformers directly. The proposed method not only avoids the need of randomization search but also incurs less computational complexity compared to an SDR approach. In addition, we also extend multicasting beamformer design problem with an additional constraint on the number of active elements, which is particularly relevant when the number of antennas is larger than that of radio frequency (RF) chains. Numerical results are used to demonstrate the superior performance of our proposed methods over the existing solutions.**

## I. Introduction

Physical layer multicasting is gaining significant attention in the upcoming standards due to services like audio and video streaming, which simultaneously deliver the same content to multiple users. Current wireless standards such as the 3rd Generation Partnership Project (3GPP) Long Term Evolution (LTE) provide dedicated subframes to deliver multicast contents in addition to regular unicast transmissions due to immense data requirements from on-demand multicast services [1]–[3]. In order to provide both unicast and multicast services over cellular networks, evolved Multimedia Broadcast Multicast Service (eMBMS) is specified in the LTE standard. The challenge is to identify a proper share of wireless resources for providing the mix of unicast and multicast services. Depending on the type of service required for each multimedia application and the number of users requesting for it, the network will determine the appropriate quality-of-service (QoS).

Physical layer multicasting for both single and multiple groups has been studied extensively from a signal processing perspective [4]–[14]. The main challenges that have been addressed in the context of multicasting problems are the grouping of users and the design of transmit beamformers for each multicast group. In both problems, the knowledge of channel state information (CSI) is assumed at the base stations (BSs). While determining the multicast groups, transmit beamformers are also designed to utilize the available spatial and frequency dimensions at each transmission instant. The beamformers are designed with the objective of either minimizing the total transmit power or maximizing the minimum achievable rate among the multiplexed multicast groups, where the minimum rate is defined by the weakest link, i.e., the user with the minimum rate. However, the design of transmit beamformers in turn depends on the selection of users for various multicast groups, which has drawn significant attention in the literature.

### A. Multicast Scheduling

In each time slot, the BS involved in multicast transmission transmits to the user group at a rate determined by the weakest link. Even though associating users requesting the same content in a same multicast group is beneficial in terms of the resource utilization, it may also deteriorate the overall performance if the users have heterogeneous channel conditions, since the transmission rate is guided by the weakest link. Thus, user association and resource scheduling for multicast transmission is not a trivial extension of unicast schedulers.

Multicast scheduling based on a proportional fair metric has been considered extensively to provide fairness among multiple multicast groups [15]–[17]. In [15], two variants of proportional fairness have been proposed based on the achievable rate, namely, inter-group proportional fairness, based on the sum of all user rates, and multicast proportional fairness, aims at maximizing the sum log of user rates. Similarly, [16] designed a scheduler to provide proportional fair utility for both unicast and multicast users over multiple BSs.

### B. Related Work

Upon determining users for various multicast groups, a scheduler must utilize both spatial and frequency resources

This work has been supported by the Finnish Funding Agency for Innovation (Tekes), Nokia Networks, and the Academy of Finland. This work has been co-funded by the Irish Government and the European Union under Irelands EU Structural and Investment Funds Program 2014-2020 through the SFI Research Centre Program under Grant 13/RC/2077.

G. Venkatraman, A. Tölli and M. Juntti are with Centre for Wireless Communications (CWC) - Radio Technologies, P.O. Box 4500, University of Oulu, Finland, FI-90014, (e-mail: firstname.lastname@oulu.fi)

L.-N. Tran was with the Department of Electronic Engineering, Maynooth University, Maynooth, Co. Kildare, Ireland. He is now with the School of Electrical and Electronic Engineering, University College Dublin, Ireland, (e-mail:nam.tran@ucd.ie)

provided by multiple-input single-output (MISO) orthogonal frequency division multiplexing (OFDM) framework efficiently to satisfy certain design criterion. Designing transmit beamformers for single multicast group with perfect CSI at the BS was introduced in [4]. Due to the nonconvex nature of the problem formulation, the beamformers were designed using semidefinite relaxation (SDR) and the resulting problem is solved by semidefinite programming (SDP) in [4]. Briefly, instead of finding a beamformer vector, say, $\mathbf{m}$, the SDR technique defines a hermitian matrix $\mathbf{M} = \mathbf{m}\mathbf{m}^{\mathrm{H}}$ and poses the original problem as an SDP with $\mathbf{M}$ as a variable. If the solution has rank greater than one, then a randomization procedure proposed in [18] is used to extract a rank-one solution. An extension to multiple multicast groups was studied in [6]. Alternatively, [19] employed successive convex approximation (SCA) technique to solve the multicast problem for a single group. Unlike the SDP based designs, the SCA technique solves for beamformers directly, thereby avoiding the need for any randomization procedure. However, for a multi-group multicasting, the problem becomes difficult as it transforms into a nonconvex quadratically constrained quadratic programming (QCQP) wherein finding an initial feasible point itself is difficult. To overcome this, [20] proposed a feasible point pursuit SCA (FPP-SCA) algorithm by adding slack variables to nonconvex constraints and a penalty to ensure that they were all forced to zero, thereby ensuring feasibility of all operating points (and solutions) throughout the SCA procedure.

A closely related problem of maximizing the minimum signal-to-interference-plus-noise ratio (SINR) of all users was also studied extensively in [5], [6], [8], [9], [11], [12], [21]. In [5], [6], [9], [11] papers, max-min fairness based beamformers were designed using the SDR approach. An iterative beamformer design for a single multicast group was proposed in [12] based on weighted SINR gradients. Alternatively, [8], [22] proposed beamformer designs based on the SCA technique for multi-group multicasting. An extension to multiple cells was considered in [21] based on fractional programming. Usually, beamformer design for multi-group multicasting with certain QoS requirement is often not possible when the channel vectors of users in different groups are collinear. This was addressed in [23] as a joint beamformer and admission control design with the objective of maximizing the admitted users.

In addition, various other extensions have been considered in the literature. A distributed multi-cell beamformer design for multicasting was proposed in [24] for both min-power and max-min fairness objectives. An extension to single antenna multi-group multicasting for relay networks was analyzed in [25] with a min-power objective. In [26], a robust beamformer design with imperfect CSI was addressed with a min-power objective for cloud radio access network (Cloud-RAN). In [27] and [28], a weighted fair multicast beamforming was proposed with per antenna power constraints using both the SDR and the SCA techniques, respectively. A multi-group multicasting with antenna selection was introduced in [9] based on bisection search. However, a beamformer design with antenna selection based on biconvex formulation proposed in [10] was shown to outperform [9] in terms of total transmit power. The capacity limits of various multicasting schemes was discussed in [29] by scaling the number of users and transmit antennas for a fairness objective. An extension to antenna subset selection was analyzed in [30] based on the average capacity scaling.

Finally, an extension to multiple-input multiple-output (MIMO) scenario was considered in [13] wherein a non-iterative algorithm for designing multicast precoders was proposed to maximize the minimum user rate for a single multicast group. The multiplexing of users over each sub-channel is based on their channel similarities and the precoders were evaluated by a weighted sum of the right singular vectors of the multiplexed users. In [14], a two stage resource allocation was proposed for multi-group multicasting by performing sub-carrier assignment followed by a power allocation step over all sub-carriers to maximize the overall multicasting throughput.

### C. Main Contribution

In this paper, we consider the problem of physical layer resource allocation for multi-group multicasting in a MISO-OFDM in an isolated cell. In this context, we address the problem of designing transmit beamformers so as to provide certain guaranteed QoS in the form of minimum rate. Due to the presence of multiple sub-channels, the SDR method proposed in [6], [9]–[11] cannot be used directly as the SINR requirement for each sub-channel is not fixed. Inspired by the superior performance of the SCA based solutions in [19], [20], [22], [28], we adopt the SCA method to solve our problem. Furthermore, we extend this technique to solve multi-group multicasting with antenna selection as studied in [9], [10], but under a MISO-OFDM model. Unlike the approaches such as the SDR with $\ell_1/\ell_\infty$ norm in [9] and the exact penalty method in [10], we solve the antenna selection problem by assigning a binary variable for each element to denote its selection status, similar to [10]. However, we adopt the SCA based design as in [19], [20], [22] instead of the SDR based technique proposed in [10]. Finally, we study the problem of maximizing the minimum achievable rate by all users in multicast groups for a given transmit power. Unlike [5], [6], [11], [12], [27], [28], we consider the fairness problem in a multi-group multiple sub-channel framework. The performance of proposed schemes are demonstrated using extensive numerical simulations, including a uniform linear array (ULA) model for illustrative reasons.

The rest of the paper is organized as follows. Section II presents both system model and problem formulation, which is followed by Section III, where the beamformer design for multicast groups is proposed by employing the SCA technique. In Section IV, the problem of selecting a subset of antennas is presented for a power minimization objective. Finally, the problem of maximizing the minimum achievable rate among multicast groups for a given transmit power is analyzed in Section V. The numerical examples are presented in Section VI together with the complexity figures. Finally, conclusions are drawn in Section VII. The following notations are used in this paper. Bold lower and upper case letters denote vectors and matrices, respectively. $(.)^{\mathrm{T}}$, $(.)^{\mathrm{H}}$, $\mathrm{tr}(.)$, $\|.\|_q$ represent the transpose, Hermitian, the trace operator, and the $\ell_q$ norm, respectively. The $i\mathrm{th}$ entry of a vector $\mathbf{x}$ is denoted by $x_i$.

## II. System Model and Problem Formulation

### A. System Model

We consider a single-cell multi-user MISO system with $N_T$ transmit antennas transmitting $N_G$ independent multicast data streams to $K$ single-antenna receivers over $N$ OFDM sub-channels (or coherence bands).[1] Each user belongs to one of the $N_G$ multicast groups, where the users in each group receive a common data stream. Let $\mathcal{G} = \{1, 2, \ldots, N_G\}$ denote the set of all multicast groups present in the system and $\mathcal{N} = \{1, 2, \ldots, N\}$ be the set of all OFDM sub-channels. The set of all users associated with multicast group $g$ is denoted by $\mathcal{G}_g$ and we denote the respective group of user $k$ by a positive integer $g_k$. The received symbol $y_{k,n}$ on the $n$th sub-channel for user $k$ belonging to multicast group $g_k$ is given by

$$y_{k,n} = \mathbf{h}_{k,n}\mathbf{m}_{g_k,n}d_{g_k,n} + \sum_{g' \in \mathcal{G}\setminus\{g_k\}} \mathbf{h}_{k,n}\mathbf{m}_{g',n}d_{g',n} + e_{k,n} \quad (1)$$

where $\mathbf{h}_{k,n} \in \mathbb{C}^{1\times N_T}$ is the channel seen by user $k$ on the $n$th sub-channel, and $\mathbf{m}_{g,n} \in \mathbb{C}^{N_T\times 1}$ is the beamformer for multicast group $g$ on sub-channel $n$. The data symbol $d_{g,n}$, transmitted for all users in $\mathcal{G}_g$, is normalized as $\mathbb{E}[|d_{g,n}|^2] = 1$ and $e_{k,n}$ is the additive complex white Gaussian noise drawn from $\mathcal{CN}(0, N_0)$. The SINR seen by user $k$ on the $n$th sub-channel, which is represented as $\Gamma_{k,n}(\{\mathbf{m}\})$ be, is given by

$$\Gamma_{k,n}(\{\mathbf{m}\}) = \frac{|\mathbf{h}_{k,n}\mathbf{m}_{g_k,n}|^2}{N_0 + \sum_{g'\in\mathcal{G}\setminus\{g_k\}} |\mathbf{h}_{k,n}\mathbf{m}_{g',n}|^2} \quad (2)$$

where $\{\mathbf{m}\} \triangleq \{\mathbf{m}_{g,n}\}$, $\forall g \in \mathcal{G}, \forall n \in \mathcal{N}$ denotes the collection of all transmit beamformers. We remark that $\Gamma_{k,n}(\{\mathbf{m}\})$ is a function of all transmit beamformers as shown in (2), but for simplicity, we express it as $\Gamma_{k,n}$ in the following discussions.

### B. Problem Formulation

We address three closely related problems on designing beamformers for a multicast transmission. At first, we study the problem of minimizing the total transmit power required to guarantee certain QoS for all users in each multicast group. Formally, the beamformer design problem is given by

$$P_1 \triangleq \begin{cases} \underset{\{\mathbf{m}\}}{\text{minimize}} & \sum_{g\in\mathcal{G}} \sum_{n=1}^{N} \|\mathbf{m}_{g,n}\|^2 & (3a) \\ \text{subject to} & \sum_{n=1}^{N} \log(1+\Gamma_{k,n}) \geq \bar{r}_{g_k}, \forall k & (3b) \end{cases}$$

where $\bar{r}_{g_k}$ is the minimum multicast service rate for all users belonging to group $g_k \in \mathcal{G}$. In unicast transmission, both joint encoding across all sub-channels and link adaptation by varying the coding scheme based on the user CSI are optimal from the information theoretic perspective [31]. Whereas for multicast transmission, link adaptation is not optimal, since the code rate is limited by $\log(1 + \min_{k\in\mathcal{G}_g}\{\Gamma_{k,n}\}), \forall k \in \mathcal{G}_g$ for each $n \in \mathcal{N}$ and $g \in \mathcal{G}$ . Thus, only joint coding across all the sub-channels with code rate $\bar{r}_g$ is optimal for each group $g \in \mathcal{G}$. Despite coding jointly across all the sub-channels, beamformers are designed specifically for each sub-channel and multicast group based on the CSI of respective users. Thus, by varying the SINR on each sub-channel independently, the overall achievability of joint coding is ensured $\forall g \in \mathcal{G}$, $\forall k$.

As an extension, we consider a design requirement wherein the number of available radio frequency (RF) chains is smaller than the number of transmit elements. Such a constraint can be achieved by forcing certain entries of transmit beamformers to zero as the power on each antenna is dictated by the respective beamformer entry. To do so, let us define a vector $\mathbf{w}$ as

$$\mathbf{w} = [w_1, \ldots, w_{N_T}]^{\mathrm{T}}, \quad w_t = \textstyle\sum_{g\in\mathcal{G}} \sum_{n=1}^{N} |m_{g,n,t}|^2 \quad (4)$$

where $m_{g,n,t}$ is the complex entry corresponding to antenna index $t \in \{1, 2, \ldots, N_T\}$ of the beamformer vector used to serve multicast group $g$ on the $n$th sub-channel and $w_t$ is the total transmit power from antenna element $t$. With the above notations, the second problem formulation is given as

$$P_2 \triangleq \begin{cases} \underset{\{\mathbf{m}\}}{\text{minimize}} & \sum_{t=1}^{N_T} w_t & (5a) \\ \text{subject to} & \|\mathbf{w}\|_0 \leq N_{RF} & (5b) \\ & \sum_{n=1}^{N} \log(1+\Gamma_{k,n}) \geq \bar{r}_{g_k}, \ \forall k & (5c) \end{cases}$$

where $N_{RF} < N_T$ is the total number of available RF chains.

Finally, we study above two problems with the objective of providing fairness, i.e., maximizing the minimum multicast group rate, for a given power budget. It can be modeled as

$$P_3 \triangleq \begin{cases} \underset{\{\mathbf{m}\}}{\text{maximize}} & \min\limits_{g\in\mathcal{G}, k\in\mathcal{G}_{g_k}} \Big\{ \sum\limits_{n=1}^{N} \log(1+\Gamma_{k,n}) \Big\} & (6a) \\ \text{subject to} & \|\mathbf{w}\|_0 \leq N_{RF} & (6b) \\ & \sum_{t=1}^{N_T} w_t \leq P_{\text{tot}} & (6c) \end{cases}$$

where $P_{\text{tot}}$ is the available transmit power budget. We discuss above problems and their solutions in subsequent sections.

## III. Proposed Solution for $P_1$

All the problems outlined in Section II-B are nonconvex due to SINR expression (2). In order to handle the nonconvexity, various approaches have been proposed in literature based on the SDR technique [4]–[6]. However, we resort to the SCA method as in [20] wherein nonconvex constraints are relaxed by a sequence of convex ones, which is then solved iteratively until convergence. Before proceeding further with the proposed SCA based design, we discuss some drawbacks in extending the SDR technique to multi-carrier scenario.

### A. Limitations of Semidefinite Relaxation

The SDR technique is a powerful signal processing tool that has been employed widely in wireless communications. For example, it has been used to demodulate higher order constellations and physical layer beamformers for single and multiple groups [4], [6], [32]. Unfortunately, the SDR method is not applicable directly to a multi-carrier multicasting problem. To

[1] Sub-channel refers to a group of frequency resources for which the channel is assumed to be relatively constant. Thus, beamformers are designed for a group of sub-carriers over which multiple data symbols are transmitted.

understand this, let us introduce a positive semidefinite matrix $\mathbf{M}_{g,n} = \mathbf{m}_{g,n}\mathbf{m}_{g,n}^{\mathrm{H}}$ as an optimization variable along with a constraint $\mathrm{rank}(\mathbf{M}_{g,n}) = 1$ so as to extract $\mathbf{m}_{g,n}$ from $\mathbf{M}_{g,n}$. Now, by using $\mathbf{M}_{g,n}$, we can express the SINR $\Gamma_{k,n}$ in (2) as

$$\Gamma_{k,n} = \frac{\mathrm{tr}\left(\mathbf{H}_{k,n}\mathbf{M}_{g,n}\right)}{N_0 + \sum_{g' \in \mathcal{G}\backslash\{g\}} \mathrm{tr}\left(\mathbf{H}_{k,n}\mathbf{M}_{g',n}\right)},\ \forall k \in \mathcal{G}_g \quad (7)$$

where $\mathbf{H}_{k,n} = \mathbf{h}_{k,n}^{\mathrm{H}}\mathbf{h}_{k,n}$ is the channel matrix related to $\mathbf{h}_{k,n}$. Using (7), an equivalent formulation for $P_1$ is written as

$$\begin{aligned}
\underset{\{\mathbf{M}\}}{\text{minimize}} \quad & \sum_{g\in\mathcal{G}}\sum_{n=1}^{N} \mathrm{tr}\left(\mathbf{M}_{g,n}\right) && \text{(8a)}\\
\text{subject to} \quad & \mathrm{rank}\left(\mathbf{M}_{g,n}\right) = 1,\ \forall g \in \mathcal{G}, \forall n \in \mathcal{N} && \text{(8b)}\\
& \sum_{n=1}^{N} \log\left(1 + \Gamma_{k,n}\right) \geq \bar{r}_g,\ \forall k \in \mathcal{G}_g && \text{(8c)}\\
& \mathbf{M}_{g,n} \succeq 0,\ \ \forall g \in \mathcal{G}, \forall n \in \mathcal{N} && \text{(8d)}
\end{aligned}$$

where $\{\mathbf{M}\} \triangleq \{\mathbf{M}_{g,n}\}$, $\forall g \in \mathcal{G}, \forall n \in \mathcal{N}$ is the collection of all transmit beamformer matrices. The minimum guaranteed QoS requirement of all users is ensured by the constraint (8c).

The problem (8) is still nonconvex with $\mathbf{M}_{g,n}$ even when the rank-one constraint (8b) for all beamformers are omitted. It follows due to the nonconvex nature of the QoS constraint (8c). However, for a single sub-channel scenario, i.e., when $N = 1$, (8) can be modeled as an SDP problem by discarding the rank-one constraint (8b) as discussed in [6], [9]. Even though the rank relaxation of (8) can be solved for a single sub-channel case, we may still require to extract a rank-one solution if (8) yields a result with the rank greater than one. This step is carried out by a randomization procedure in [18]. The best known randomization algorithm for this case was proposed in [6], [18] and it requires solving a series of linear programs. Therefore, as the number of sub-channels and multicast groups increases, the complexity of (8) scales-up quickly. Furthermore, designing beamformers for multi-group multi-carrier multicasting as in (8) by the SDR method is not a trivial problem, since the QoS constraint in (8c) is defined over all sub-channels, and therefore cannot be solved for each sub-channel independently by fixing the SINR arbitrarily by satisfying (8c). Nonetheless, an iterative solution based on the SDR and the SCA technique is proposed in [33].

### *B. Solution based on Successive Convex Approximation*

Due to the issues involved with the SDR technique explained above, we propose an alternative approach to solve $P_1$ by employing the SCA technique, thereby ensuring a rank-one solution upon finding the multicast beamformers. In order to do so, we relax the minimum guaranteed rate constraint in (3b) with SINR term $\Gamma_{k,n}$ by the following inequalities as

$$\begin{aligned}
& \sum_{n=1}^{N} \log(1 + \gamma_{k,n}) \geq \bar{r}_g,\ \forall k \in \mathcal{G}_g && \text{(9a)}\\
& \frac{|\mathbf{h}_{k,n}\mathbf{m}_{g,n}|^2}{N_0 + \sum_{r\in\mathcal{G}\backslash\{g\}} |\mathbf{h}_{k,n}\mathbf{m}_{r,n}|^2} \geq \gamma_{k,n},\ \forall k \in \mathcal{G}_g && \text{(9b)}
\end{aligned}$$

where the newly introduced optimization variable $\gamma_{k,n}$ is an under-estimator for the actual SINR $\Gamma_{k,n}$ as $\Gamma_{k,n} \geq \gamma_{k,n}$. By adding one on both sides of (9b), we rewrite it as

$$\frac{N_0 + \sum_{r\in\mathcal{G}} |\mathbf{h}_{k,n}\mathbf{m}_{r,n}|^2}{\gamma_{k,n} + 1} \geq \sum_{r\in\mathcal{G}\backslash\{g\}} |\mathbf{h}_{k,n}\mathbf{m}_{r,n}|^2 + N_0 \quad (10)$$

where user $k$ belongs to group $g \in \mathcal{G}$. Even after replacing the QoS constraints in (3b) by two inequalities (9a) and (10), the problem is still nonconvex due to the nonconvexity of the constraint in (10). Therefore, we adopt the SCA technique in [34] wherein the nonconvex set (10) is relaxed by a convex subset around a fixed operating point, which is used in $P_1$ instead of (10). Upon finding a solution, a new feasible set is updated by using the current solution as an operating point for the next iteration and solved for an optimal solution.

In order to find a convex approximation for the nonconvex constraint (10), we introduce two new stacked channel vectors $\bar{\mathbf{h}}_{k,n}$ and $\tilde{\mathbf{h}}_{k,n}$, which are defined as

$$\begin{aligned}
\bar{\mathbf{h}}_{k,n} &\triangleq [\underbrace{\mathbf{h}_{k,n}, \mathbf{h}_{k,n}, \ldots, \mathbf{h}_{k,n}}_{N_G \text{ terms}}] \in \mathbb{C}^{1\times N_T\cdot N_G} && \text{(11a)}\\
\tilde{\mathbf{h}}_{k,n} &\triangleq [\underbrace{\mathbf{h}_{k,n}, \mathbf{h}_{k,n}, \ldots, \mathbf{h}_{k,n}}_{(N_G-1) \text{ terms}}] \in \mathbb{C}^{1\times N_T(N_G-1)} && \text{(11b)}
\end{aligned}$$

where $\bar{\mathbf{h}}_{k,n}$ and $\tilde{\mathbf{h}}_{k,n}$ are the vectors formed by repeating the channel seen by user $k$ on the $n$th sub-channel by $N_G$ and $N_G - 1$ times, respectively. In addition, we also introduce two new stacked vectors $\mathbf{m}_n$ and $\tilde{\mathbf{m}}_{g,n}$ such that $k \in \mathcal{G}_g$ as

$$\begin{aligned}
\mathbf{m}_n &\triangleq [\mathbf{m}_{1,n}^{\mathrm{T}}, \mathbf{m}_{2,n}^{\mathrm{T}}, \ldots, \mathbf{m}_{N_G,n}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{N_T\cdot N_G\times 1} && \text{(12a)}\\
\tilde{\mathbf{m}}_{g,n} &\triangleq [\mathbf{m}_{1,n}^{\mathrm{T}}, \mathbf{m}_{2,n}^{\mathrm{T}}, \ldots, \mathbf{m}_{g-1,n}^{\mathrm{T}}, \mathbf{m}_{g+1,n}^{\mathrm{T}},\\
&\qquad \ldots, \mathbf{m}_{N_G,n}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{N_T(N_G-1)\times 1} && \text{(12b)}
\end{aligned}$$

where $\mathbf{m}_n$ is formed by stacking all multicast beamformers for sub-channel $n$ and the vector $\tilde{\mathbf{m}}_{g,n}$ is obtained by stacking transmit beamformers corresponding to multicast groups in $\mathcal{G}\backslash\{g\}$ for sub-channel $n$, i.e., by excluding $\mathbf{m}_{g,n}$ from $\mathbf{m}_n$. The newly defined vectors in (12) are just a rearrangement of optimization variables $\{\mathbf{m}\}$.

Now, by using (12), the constraint in (10) becomes

$$\frac{N_0 + |\bar{\mathbf{h}}_{k,n}\mathbf{m}_n|^2}{1 + \gamma_{k,n}} \geq N_0 + |\tilde{\mathbf{h}}_{k,n}\tilde{\mathbf{m}}_{g_k,n}|^2 \quad (13)$$

where the l.h.s of (13) is of *quadratic-over-linear*, i.e., a convex function, and thus can be bounded from below by a linear first order Taylor approximation $\mathcal{L}(\mathbf{m}_n, \gamma_{k,n}; \mathbf{m}_n^{(i)}, \gamma_{k,n}^{(i)})$ as

$$\begin{aligned}
&\mathcal{L}(\mathbf{m}_n, \gamma_{k,n}; \mathbf{m}_n^{(i)}, \gamma_{k,n}^{(i)}) \triangleq\\
&\frac{N_0 + |\bar{\mathbf{h}}_{k,n}\mathbf{m}_n^{(i)}|^2}{1 + \gamma_{k,n}^{(i)}} + 2\,\Re\left\{\frac{\mathbf{m}_n^{(i)\,\mathrm{H}}\bar{\mathbf{h}}_{k,n}^{\mathrm{H}}\bar{\mathbf{h}}_{k,n}}{1 + \gamma_{k,n}^{(i)}}\left(\mathbf{m}_n - \mathbf{m}_n^{(i)}\right)\right\}\\
&\quad - \frac{N_0 + |\bar{\mathbf{h}}_{k,n}\mathbf{m}_n^{(i)}|^2}{\left(1 + \gamma_{k,n}^{(i)}\right)^2}\left(\gamma_{k,n} - \gamma_{k,n}^{(i)}\right) \leq \text{l.h.s (13)} \quad (14)
\end{aligned}$$

where $\mathbf{m}_n^{(i)}$ and $\gamma_{k,n}^{(i)}$ are fixed operating points upon which the approximation is carried out. Moreover, $\mathbf{m}_n^{(i)}$ and $\gamma_{k,n}^{(i)}$ are the solutions obtained for $\mathbf{m}_n$ and $\gamma_{k,n}$ from the $(i-1)$th SCA iteration, respectively.

Finally, by replacing the l.h.s of (13) with (14), the convex

subproblem for the $i$th SCA iteration is given $\forall k \in \mathcal{G}_g$ by

$$\underset{\{\mathbf{m}\},\{\boldsymbol{\gamma}\}}{\text{minimize}} \quad \sum_{g\in\mathcal{G}}\sum_{n=1}^{N} \|\mathbf{m}_{g,n}\|^2 \tag{15a}$$

$$\text{subject to} \quad \sum_{n=1}^{N} \log(1+\gamma_{k,n}) \geq \bar{r}_g,\ \gamma_{k,n} \geq 0,\ \forall k, \forall g \tag{15b}$$

$$\mathcal{L}(\mathbf{m}_n, \gamma_{k,n}; \mathbf{m}_n^{(i)}, \gamma_{k,n}^{(i)}) \geq N_0 + |\tilde{\mathbf{h}}_{k,n}\tilde{\mathbf{m}}_{g_k,n}|^2,\ \forall k, \forall n, \forall g. \tag{15c}$$

We remark that (15) can be formulated as a second-order cone programming (SOCP) by rewriting the QoS constraint (15b) as

$$\prod_{n=1}^{N}(1+\gamma_{k,n}) \geq \exp(\bar{r}_{g_k}),\ \forall k \in \mathcal{G}_g, \forall g \in \mathcal{G} \tag{16}$$

which can be expressed by a system of second-order cone (SOC) constraints as in [35], and thus (15) reduces to a SOCP problem. The fixed operating points $\{\mathbf{m}_{g,n}^{(i)}\}$ and $\{\boldsymbol{\gamma}_{k,n}^{(i)}\}$ are updated with the solution obtained from (15) in the $(i-1)$th SCA iteration. Then, by updating the operating point recursively, (15) is solved until convergence to obtain transmit beamformers, which achieves the required rate for all users in the system. A discussion on the convergence analysis of (15) is given in Appendix A and its complexity in Appendix B. The superior performance of the SCA method over the SDP technique was demonstrated for a multicast problem with $N = 1$ in [20]. Thus, if the SDP based formulation presented in (8) yields a rank-one vector for $N = 1$, the solution is the same as that of (15) as there is no need for randomization search to find a rank-one vector. However, if (8) yields a high rank matrix, the randomization search is required to extract a rank-one beamformer. In the later cases, the SCA schemes find a better solution as they find beamforming vectors directly.

### C. Choice of SCA Initialization Points

Finding an initial operating point $\{\mathbf{m}_g^{(0)}\}$ to start the SCA procedure is not a trivial problem. For the single multicast group case, initialization can be obtained easily by merely scaling the beamformers until all the QoS constraints are satisfied as shown in [19]. However, it is not applicable to $P_1$, since the multi-group beamformers are present in both the numerator and the denominator of (2). Therefore, to find an initial operating point for (15), we consider a relaxed problem

$$\underset{\{\mathbf{m}\},\{\boldsymbol{\gamma}\},\tilde{R}}{\text{minimize}} \quad \sum_{g\in\mathcal{G}}\sum_{n=1}^{N} \|\mathbf{m}_{g,n}\|^2 + \delta\tilde{R} \tag{17a}$$

$$\text{subject to} \quad \exp(\bar{r}_{g_k}) - \prod_{n=1}^{N}(1+\gamma_{k,n}) \leq \tilde{R},\ \forall k \tag{17b}$$

$$\mathcal{L}(\mathbf{m}_n, \gamma_{k,n}; \mathbf{m}_n^{(i)}, \gamma_{k,n}^{(i)}) \geq N_0 + |\tilde{\mathbf{h}}_{k,n}\tilde{\mathbf{m}}_{g_k,n}|^2,\ \gamma_{k,n} \geq 0,\ \forall k,\ \forall n. \tag{17c}$$

where $\mathbf{m}_n$ and $\tilde{\mathbf{m}}_{g,n}$ are defined in (12). The constant $\delta$ determines a trade-off between the two objectives and $\tilde{R}$ is a slack variable, used to relax the strict rate constraint. Problem (17) is feasible for any randomly initialized beamformers, since (15b) is relaxed by a slack variable $\tilde{R}$ in (17) [20], [22]. Algorithm 1 outlines the iterative procedure to solve (15).

**Algorithm 1** Proposed iterative algorithm for solving $P_1$

**Initialization:** Set $i = 1$, generate $\{\mathbf{m}^{(0)}\}$ randomly and compute $\{\boldsymbol{\gamma}^{(0)}\}$ by evaluating (10) with equality. Set $\delta$ to be large enough to ensure feasibility
**repeat**
  solve (17) for optimal $\{\mathbf{m}_*^{(i)}\}$ and $\{\boldsymbol{\gamma}_*^{(i)}\}$
  update $\{\mathbf{m}^{(i+1)}\} = \{\mathbf{m}_*^{(i)}\}$ and $\{\boldsymbol{\gamma}^{(i+1)}\} = \{\boldsymbol{\gamma}_*^{(i)}\}$
  $i = i + 1$
**until** $\tilde{R} < 0$
Let $\{\mathbf{m}_*\}$ and $\{\boldsymbol{\gamma}_*\}$ be a solution of (17) when $\tilde{R} < 0$
Set $j = 1$, $\{\mathbf{m}^{(0)}\} = \{\mathbf{m}_*\}$ and $\{\boldsymbol{\gamma}^{(0)}\} = \{\boldsymbol{\gamma}_*\}$
**repeat**
  solve (15) for optimal $\{\mathbf{m}_*^{(j)}\}$ and $\{\boldsymbol{\gamma}_*^{(j)}\}$
  update $\{\mathbf{m}^{(j+1)}\} = \{\mathbf{m}_*^{(j)}\}$ and $\{\boldsymbol{\gamma}^{(j+1)}\} = \{\boldsymbol{\gamma}_*^{(j)}\}$
  $j = j + 1$
**until** convergence or for fixed number of iterations, $I_{\max}$

If $\delta$ is chosen to be infinite, then (17) reduces to a feasibility problem. However, if $\delta$ is finite but sufficiently large, then (17) solves for multicast beamformers with the objective of minimizing total transmit power and the slackness term $\tilde{R}$ in each iteration. Once $\tilde{R} < 0$, the above procedure can be terminated and the solution can be used as an initial feasible point to solve (15). For instance, if the feasible set of (15) is nonempty, based on our observation from extensive numerical simulations, choosing $\delta \approx 10^3$ will yield a feasible point in $2-3$ iterations itself. Finally, we note that (15) reduces to

- [19] when $N = 1$ and $N_G = 1$, since $\gamma_{k,n}$ becomes a constant, thereby eliminating (15b) and (17b) from (15) and (17), respectively, or
- [20] when $N = 1$ and $N_G \geq 1$. It follows from the fact that when $N = 1$, $\gamma_{k,n}$ becomes a constant, thereby removing (17b) from (17). Thus, the FPP-SCA method in [20] is similar to (17) if it is performed until convergence.

In order to perform efficient multicast transmission over multiple sub-channels, a group specific coding scheme must be designed jointly across all the sub-channels with code rate $\bar{r}_g$ defined by the QoS requirement [31]. On the contrary, if the coding scheme is designed for each sub-channel separately, then the QoS constraint in (16) must be replaced $\forall g \in \mathcal{G}$ by

$$\sum_{n=1}^{N} \log(1 + \min_{k\in\mathcal{G}_g}\{\gamma_{k,n}\}) \geq \bar{r}_g \tag{18}$$

which is equivalent to (16) when the number of sub-channels is $N = 1$ and inferior when $N > 1$.

## IV. Proposed Solution for $P_2$

In this section, we extend the solution of $P_1$ to the antenna selection problem introduced in $P_2$. Since the power on each antenna is determined by the beamformers, the number of active elements is restricted to, say, $N_{\mathrm{RF}}$, by constraint (5b) in $P_2$. However, due to the presence of $\ell_0$ norm in (5b), $P_2$ cannot be solved directly. Hence, to find $N_{\mathrm{RF}}$ active

elements from $N_T$ transmit antennas, we could resort to the re-weighted $\ell_1$ norm minimization approach as in [9]. However, due to the presence of multiple sub-channels, even without the constraint (5b), $P_2$ can only be solved by employing the SCA technique to find beamformers for all sub-channels. Therefore, if we adopt the re-weighted $\ell_1$ norm minimization technique proposed in [9], then we need a bisection search to find a suitable regularization constant in order to find $N_{\mathrm{RF}}$ active elements, thereby leading to two nested loops, i.e., an inner loop to perform the SCA based iterative procedure and an outer loop for the bisection search.

Thus, to avoid the above mentioned problem of two nested loops while solving $P_2$, (5b) can be equivalently rewritten by introducing a binary optimization variable $\mathbf{a}$ as

$$w_t \leq a_t \, v_t, \quad \forall t \in \{0, \ldots, N_T\} \tag{19}$$

where $a_t \in \{0,1\}$ and $v_t \in \mathbb{R}$ are newly introduced variables together with $\mathbf{w}$ as in (4). The binary variable $a_t$ denotes the status of antenna elements, i.e., whether it is enabled or not, and $v_t$ represents the total power on each antenna $t$ due to multicast transmission. Upon replacing $\ell_0$ norm in (5b) by a binary vector $\mathbf{a}$ together with (19), we have

$$\underset{\{\mathbf{m}\},\{\mathbf{a}\},\{\mathbf{v}\}}{\text{minimize}} \quad \sum_{t=1}^{N_T} v_t \tag{20a}$$

$$\text{subject to} \quad \sum_{n=1}^{N} \log(1+\Gamma_{k,n}) \geq \bar{r}_{g_k}, \ \forall k \tag{20b}$$

$$(19) \text{ and } \sum_{t=1}^{N_T} a_t = N_{\mathrm{RF}}, \ a_t \in \{0,1\}, \ \forall t \tag{20c}$$

where the SINR $\Gamma_{k,n}$ is given by expression (2).

### A. Mixed Integer Solution

Following the approximations proposed in Section III-B, we relax $\Gamma_{k,n}$ in (2) by an under-estimator $\gamma_{k,n}$ as in (9). Due to the nonconvex nature of the newly introduced constraint (9b), (20) cannot be solved directly. Therefore, we approximate the continuous nonconvex constraint (13) by (15c) to arrive at a mixed-integer quadratic programming (MIQP) subproblem as

$$\underset{\{\mathbf{m}\},\{\mathbf{a}\},\{\mathbf{v}\},\{\boldsymbol{\gamma}\},}{\text{minimize}} \quad \sum_{t=1}^{N_T} v_t \tag{21a}$$

$$\text{subject to} \quad \sum_{t=1}^{N_T} a_t = N_{\mathrm{RF}}, \ a_t \in \{0,1\}, \ \forall t \tag{21b}$$

$$(15\text{b}), (15\text{c}) \text{ and } (19). \tag{21c}$$

The MIQP subproblem in (21) can be solved optimally by the available MIQP solvers such as [36], [37], however, it involves huge computational complexity. This procedure of iteratively solving (21) is referred to as iterative MIQP solution. Even though (21) can be solved optimally in each iteration, due to the iterative nature of the problem, the complex entries of the SCA operating point $\{\mathbf{m}_{g,n}^{(i)}\}$ will be forced either to zero or to a non-zero complex values depending on the respective values of $\mathbf{w}$ by (19). The equivalent formulation of $P_2$ in (20) consists of both binary and continuous variables. However, in the iterative MIQP approach, we only approximate the continuous nonconvex constraints by a sequence of convex subsets using the principles of the SCA, while keeping the binary variables untouched. In this manner, the domain of the continuous variables is approximated by a feasible subset in each iteration. Due to this approximation, it cannot be guaranteed that the feasible subset can cover the entire continuous region of the original feasible set when the SCA procedure converges. Thus, the iterative MIQP cannot guarantee an optimal solution in general even upon the convergence of the SCA procedure.

### B. Regularized Relaxed Binary Formulation

As an alternative to the combinatorial problem (20), we relax the binary variable $a_t \in \{0,1\}$ by a linear constraint as $a_t \in [0,1]$, $\forall t$ in order to find a low complexity formulation. Due to the linear relaxation, binary outcome for the solution of $a_t, \forall t$ cannot be guaranteed. Hence, to promote a binary solution for the relaxed linear variable $\mathbf{a}$, we regularize the objective function with a penalty term as in [38], [39], which depends only on the relaxed variable $\mathbf{a}$. Thus, an approximate solution with a regularization term for (20) is given by

$$\widehat{P}_2 \triangleq \begin{cases} \underset{\substack{\{\mathbf{m}\},\{\mathbf{a}\},\\ \{\mathbf{v}\},\{\boldsymbol{\gamma}\}}}{\text{minimize}} & \displaystyle\sum_{t=1}^{N_T} v_t + \psi \, f(\mathbf{a}) & (22\text{a}) \\ \text{subject to} & \displaystyle\sum_{t=1}^{N_T} a_t = N_{\mathrm{RF}}, a_t \in [0,1], \forall t & (22\text{b}) \\ & (15\text{b}), (15\text{c}) \text{ and } (19) & (22\text{c}) \end{cases}$$

where $f(\mathbf{a})$ is a penalty function to encourage sparse solution for $\mathbf{a}$. Due to linear relaxation in (22b), (19) can be expressed as a rotated SOC constraint. The parameter $\psi$ determines a trade-off between the desired objective and the degree of sparsity in the solution of $\mathbf{a}$. Since the quality of $\mathbf{a}$ depends on the choice of regularization parameter $\psi$, binary outcome for the solution of $\mathbf{a}$ cannot be guaranteed, therefore, (22) can only be shown as an approximate problem for (20) and not an equivalent formulation, which can be expressed as $P_2 \subset \widehat{P}_2$. Upon solving $\widehat{P}_2$ with a sparsity promoting penalty function, if $\mathbf{a} \in \{0,1\}^{N_T}$, then the solution obtained is also a feasible solution of $P_2$. However, if $\mathbf{a} \notin \{0,1\}^{N_T}$, then a feasible point for $P_2$ can be found by setting $N_{\mathrm{RF}}$ maximum entries from sorted vector $\mathbf{a}$ to unity while forcing the remaining to zero. Thus, the solution found by solving (22) can be used to obtain a feasible point for (20) and original problem $P_2$.

*1) Log-Sum as Penalty Function:* We begin with the well known sparsity inducing penalty function proposed in [40] to promote binary solution for vector $\mathbf{a}$, as given by

$$f(\mathbf{a}) = \sum_{t=1}^{N_T} \log(a_t + \epsilon) \tag{23}$$

where $\epsilon$ is a small positive constant used to limit the dynamic range of the $\log$ function. With this penalty term, $\widehat{P}_2$ becomes

$$\underset{\{\mathbf{m}\},\{\mathbf{a}\},\{\mathbf{v}\},\{\boldsymbol{\gamma}\}}{\text{minimize}} \quad \sum_{t=1}^{N_T} v_t + \psi \sum_{t=1}^{N_T} \log(a_t + \epsilon) \tag{24a}$$

$$\text{subject to} \quad (15b), (15c), (19) \text{ and } (22b). \tag{24b}$$

The use of $\log(a_t+\epsilon)$ in (24a) is justified by the fact that $\log(a_t+\epsilon)$ has slope $1/\epsilon$ at origin that becomes infinitely large as $\epsilon \rightarrow 0$, which is greater than the $\ell_1$ norm that has unit slope at the origin. Therefore, similar to the $\ell_0$ norm that has infinite slope at the origin, the $\log(a_t+\epsilon)$ term allows for a large penalty for relatively small values of $a_t$, compared to the $\ell_1$ norm [40]. It is similar to the re-weighted $\ell_1$ minimization in [9] but with better convergence properties.

However, due to the presence of $-\log(a_t+\epsilon)$ in (24a), the objective is a difference of convex (DC) function. Therefore, by adopting SCA framework, we linearize the $\sum_{t=1}^{N_T}\log(a_t+\epsilon)$ function by its first order Taylor approximation as

$$\hat{f}(\mathbf{a};\mathbf{a}^{(i)}) = f(\mathbf{a}^{(i)}) + \sum_{t=1}^{N_T} \frac{a_t - a_t^{(i)}}{a_t^{(i)}+\epsilon} \geq f(\mathbf{a}),\ \forall \mathbf{a} \in \{0,1\}^{N_T} \tag{25}$$

around some fixed operating point $\mathbf{a}^{(i)}$ such that $f(\mathbf{a}^{(i)}) = \hat{f}(\mathbf{a}^{(i)};\mathbf{a}^{(i)})$. The expression in (25) is similar to the one used in the iterative re-weighted $\ell_1$ algorithm to obtain a sparse solution as discussed in [9] for the SDR formulation of the antenna selection problem. In summary, the convex subproblem in each iteration of the proposed iterative algorithm to solve $\widehat{P}_2$ using the log-sum penalty function is given by

$$\underset{\{\mathbf{m}\},\{\mathbf{a}\},\{\mathbf{v}\},\{\boldsymbol{\gamma}\}}{\text{minimize}} \quad \sum_{t=1}^{N_T} v_t + \psi\, \hat{f}(\mathbf{a};\mathbf{a}^{(i)}) \tag{26a}$$

$$\text{subject to} \quad (15b), (15c), (19) \text{ and } (22b). \tag{26b}$$

Upon finding a binary solution for vector $\mathbf{a}$, the approximate penalty term $\hat{f}(\mathbf{a};\mathbf{a}^{(i)})$ in (26a) becomes a constant, thereby leading to the original objective as in (20a). Finally, by using the discussions presented in Appendix A, we can ensure that (26a) converges to (24a) upon the SCA convergence, since $\hat{f}(\mathbf{a},\mathbf{a}^{(i)}) \rightarrow f(\mathbf{a})$ as $i \rightarrow \infty$. Thus, upon solving subproblem (26) in each of the SCA iterations with the updated feasible convex subset, we obtain a stationary point of (24) upon the convergence of the SCA procedure.

*2) Entropy as Penalty Function:* Another choice for the penalty function is an entropy measure of $\mathbf{a}$, defined as [41]

$$f(\mathbf{a}) = -\sum_{t=1}^{N_T} a_t \log(a_t). \tag{27}$$

The slope of $\mathbf{a}$ in (27) also goes to infinity when $a_t \rightarrow 0$, which is the same as that of the $\ell_0$ norm. Specifically, the entropy of an event with probability $x$, defined as $-x\log(x)$, is minimized (i.e., to be zero) when $x = \{0,1\}$. Due to the fact that binary variable $\mathbf{a}$ is relaxed to take values in $[0,1]$ similar to probability mass function, $\widehat{P}_2$ becomes

$$\underset{\{\mathbf{m}\},\{\mathbf{a}\},\{\mathbf{v}\},\{\boldsymbol{\gamma}\}}{\text{minimize}} \quad \sum_{t=1}^{N_T} v_t - \psi \sum_{t=1}^{N_T} a_t \log(a_t) \tag{28a}$$

$$\text{subject to} \quad (15b), (15c), (19) \text{ and } (22b). \tag{28b}$$

where the constant $\psi$ is used to induce sparsity in $\mathbf{a}$.

Due to the concave nature of $f(\mathbf{a})$, the objective function is a DC function. Therefore, it is solved by employing the SCA technique for $f(\mathbf{a})$ around the $\mathbf{a}^{(i)}$ as

$$\hat{f}(\mathbf{a};\mathbf{a}^{(i)}) = f(\mathbf{a}^{(i)}) - \sum_{t=1}^{N_T} \left[1+\log\left(a_t^{(i)}\right)\right]\left(a_t - a_t^{(i)}\right) \tag{29}$$

where $\hat{f}(\mathbf{a};\mathbf{a}^{(i)})$ is the first order Taylor approximation of $f(\mathbf{a})$ at fixed $\mathbf{a}^{(i)}$. By using the regularized objective, we can formulate the convex subproblem for SCA iteration $i$ as

$$\underset{\{\mathbf{m}\},\{\mathbf{a}\},\{\mathbf{v}\},\{\boldsymbol{\gamma}\},}{\text{minimize}} \quad \sum_{t=1}^{N_T} v_t + \psi\, \hat{f}(\mathbf{a};\mathbf{a}^{(i)}) \tag{30a}$$

$$\text{subject to} \quad (15b), (15c), (19) \text{ and } (22b). \tag{30b}$$

Upon obtaining a binary solution for vector $\mathbf{a}$, the approximate penalty term $\hat{f}(\mathbf{a};\mathbf{a}^{(i)})$ in (29) becomes zero, thereby leading to the original objective as in (20a). Finally, the iterative convex subproblems (30) can be shown to find a stationary point of the nonconvex problem $\widehat{P}_2$ using the discussions in Appendix A. The iterative procedure for solving $P_2$ by using either (26) or (30) is outlined in Algorithm 2. In order to find an initial operating point, we run Algorithm 1 to find $\{\mathbf{m}\}$ and $\{\boldsymbol{\gamma}\}$ such that all the QoS constraints are satisfied. Then, the antenna selection vector is initialized with $\mathbf{a}^{(0)} = \frac{N_{\mathrm{RF}}}{N_T}\mathbf{1}^{\mathrm{T}}$, since Algorithm 1 utilizes all the available number of antennas. Upon initializing $\{\mathbf{m}^{(0)}\},\{\boldsymbol{\gamma}^{(0)}\}$ and $\mathbf{a}^{(0)}$ with the solution from Algorithm 1, feasibility of the operating point can be guaranteed to begin Algorithm 2.

Finally, the choice of regularization parameter $\psi$ plays a vital role in determining the quality of solution obtained for the relaxed vector $\mathbf{a}$ as it provides a trade-off between the binary outcome to a solution with less transmit power. Due to the iterative nature of the proposed algorithms, choosing $\psi$ very large may lead to a sub-optimal solution, since the feasible subsets during first few SCA iterations are too restrictive as they depend on the initial operating point. On the contrary, choosing the parameter $\psi$ very small may lead to a non-integer solution, which is not a desired one. In such cases, we can choose the antenna elements corresponding to the indices of maximum $N_{\mathrm{RF}}$ entries from the sorted $\mathbf{a}$ vector. Once the transmit elements are chosen, steps 6-12 of Algorithm 1 is performed by retaining only those channel coefficients corresponding to the chosen $N_{\mathrm{RF}}$ maximum entries. If the solution obtained by solving (22) is binary for some fixed value of $\psi$, then it is also a feasible point for (20).

Alternatively, we can adopt the method in [38], [42], which was also used in the multicasting context by [10], [11], where $\psi$ is increased in each SCA step to emphasize the penalty term over the original objective for some $\psi^{(0)} \geq 0$. It is obtained by modifying the objective functions in (26a) and (30a) as

$$\underset{\{\mathbf{m}\},\{\mathbf{a}\},\{\mathbf{v}\},\{\boldsymbol{\gamma}\}}{\text{minimize}} \quad \sum_{t=1}^{N_T} v_t + \psi^{(i)}\, \hat{f}(\mathbf{a};\mathbf{a}^{(i)}) \tag{31}$$

where $\psi^{(i+1)} = \sigma\psi^{(i)} \rightarrow \Psi_{\max}$ as $i \rightarrow \infty$ for some very large $\Psi_{\max}$ together with $\sigma > 1$. By this way, the desired objective is emphasized during initial SCA iterations and the sparsity promoting term will be introduced gradually as the SCA iteration increases. Heuristically, we can choose $\psi$ such that it satisfies $\psi \geq \frac{P_{\mathrm{all}}}{N_T}$, where $P_{\mathrm{all}}$ is the total power required

to satisfy the QoS requirements while using all antennas. The factor $N_T$ is used to normalize the value of penalty function $f(\mathbf{a};\mathbf{a}^{(i)})$ when $\mathbf{a}$ is not binary. Therefore, to emphasize the original objective of power minimization during the first few SCA iterations with $N_{\mathrm{RF}}$ active antennas, $\psi$ includes the scaling of $N_T$ so that $f(\mathbf{a};\mathbf{a}^{(i)})$ does not dominate when initialized with $\mathbf{a}^{(0)} = \frac{N_{\mathrm{RF}}}{N_T}\mathbf{1}^{\mathrm{T}}$. The convergence analysis of Algorithm 2 is similar to the one presented in the Appendix for Algorithm 1. The solution obtained by solving either (26) or (30) iteratively, converges to a stationary point of $\widehat{P}_2$. Thus, if the solution obtained for $\mathbf{a}$ has a binary outcome, it is also a feasible point for the original problem $P_2$, otherwise an approximate solution can be found by using step 6 in Algorithm 2.

**Algorithm 2** Proposed iterative algorithm for solving $P_2$

1: Perform Algorithm 1 until convergence.
2: Assign the solution $\{\mathbf{m}_*\}$ and $\{\boldsymbol{\gamma}_*\}$ to $\{\mathbf{m}^{(0)}\}$ and $\{\boldsymbol{\gamma}^{(0)}\}$, respectively.
3: Initialize $\mathbf{a}^{(0)} = \frac{N_{\mathrm{RF}}}{N_T}\,\mathbf{1}^{\mathrm{T}}$.
4: **repeat**
5: Solve (26) for the log-sum penalty or (30) for the entropy penalty function. Denote the solution $\{\mathbf{m}\}$,$\{\boldsymbol{\gamma}\}$ and $\mathbf{a}$ as $\{\mathbf{m}_*^{(i)}\}$, $\{\boldsymbol{\gamma}_*^{(i)}\}$ and $\mathbf{a}_*^{(i)}$, respectively.
6: Set $\{\mathbf{m}^{(i+1)}\}$,$\{\boldsymbol{\gamma}^{(i+1)}\}$ and $\mathbf{a}^{(i+1)}$ to $\{\mathbf{m}_*^{(i)}\}$, $\{\boldsymbol{\gamma}_*^{(i)}\}$, and $\mathbf{a}_*^{(i)}$, respectively and update $i = i+1$.
7: **until** convergence or for maximum number of iterations
8: Select $N_{\mathrm{RF}}$ maximum entries from the sorted $\mathbf{a}$ and perform steps 6-12 of Algorithm 1 by retaining $N_{\mathrm{RF}}$ channel entries $\mathbf{h}_{k,n}, \forall k, \forall n$ for which $a_t = 1$, $\forall t$

## V. Proposed Solution for $P_3$

In this section, we study problem $P_3$ wherein the objective is to maximize the minimum rate seen by the users of various multicast groups for a given transmit power budget. In addition, $P_3$ also restricts the number of active antennas used for transmission. Again, due to the presence of the $\ell_0$ norm in (6b), it is not possible to solve $P_3$ directly. Therefore, by following the technique as in $P_2$, we relax the continuous nonconvex constraint (13) by (15c), which leads to the problem

$$\begin{aligned}
\underset{r_g,\{\mathbf{m}\},\{\mathbf{a}\},\{\mathbf{v}\},\{\boldsymbol{\gamma}\}}{\text{maximize}} \quad & \min_{g\in\mathcal{G}} \{r_g\} && \text{(32a)}\\
\text{subject to} \quad & \sum_{n=1}^{N} \log(1+\gamma_{k,n}) \geq r_{g_k}, \forall k && \text{(32b)}\\
& \sum_{t=1}^{N_T} a_t = N_{\mathrm{RF}},\ a_t \in \{0,1\},\ \forall t && \text{(32c)}\\
& \sum_{t=1}^{N_T} v_t \leq P_{\mathrm{tot}} && \text{(32d)}\\
& \text{(15c) and (19).} && \text{(32e)}
\end{aligned}$$

where the total transmit power is restricted to $P_{\mathrm{tot}}$ by (32d). Even though $r_{g_k}$ is treated as a variable in (32b), it can still be formulated as a system of SOC constraints by following [35]. Due to the binary constraint on $\mathbf{a}$, the subproblem (32) for the $i$th SCA iteration is a MIQP, which can be solved efficiently by the existing solvers such as MOSEK and Gurobi [36], [37].

In contrast to the MIQP subproblem, which has huge computational complexity, we consider an approximate version of $P_3$ by relaxing the binary constraint on vector $\mathbf{a}$ by a linear one as discussed in Section IV-B. Again, to encourage sparsity in the solution of $\mathbf{a}$, we regularize the objective as

$$\widehat{P}_3 \triangleq \begin{cases} \underset{\substack{r_g,\{\mathbf{m}\},\{\mathbf{a}\},\\ \{\mathbf{v}\},\{\boldsymbol{\gamma}\}}}{\text{maximize}} & \min\limits_{g\in\mathcal{G}} \{r_g\} - \psi\, f(\mathbf{a}) && \text{(33a)}\\ \text{subject to} & \sum\limits_{t=1}^{N_T} a_t = N_{\mathrm{RF}},\ \mathbf{a} \in [0,1]^{N_T} && \text{(33b)}\\ & \text{(15c), (19), (32b) and (32d)} && \text{(33c)} \end{cases}$$

where $f(\mathbf{a})$ is a penalty function which is defined as either (23) or (27) and the vector $\mathbf{w}$ is given in (4). The newly introduced variable $r_g$ is the minimum guaranteed rate of multicast group $g \in \mathcal{G}$. The problem now is to allocate the total power $P_{\mathrm{tot}}$ among the sub-channels to maximize the minimum achievable rate of users in the multicast groups.

To solve (33), we again use the SCA technique as discussed in the previous subsection. In particular, the approximate convex subproblem in the $i$th SCA step is given by

$$\begin{aligned}
\underset{r_g,\{\mathbf{m}\},\{\mathbf{a}\},\{\mathbf{v}\},\{\boldsymbol{\gamma}\}}{\text{maximize}} \quad & \underset{g\in\mathcal{G}}{\text{minimum}} \{r_g\} - \psi\, \hat{f}(\mathbf{a};\mathbf{a}^{(i)}) && \text{(34a)}\\
\text{subject to} \quad & \text{(15c), (19), (32b), (32d) and (33b)} && \text{(34b)}
\end{aligned}$$

where $\hat{f}(\mathbf{a};\mathbf{a}^{(i)})$ is the first order Taylor approximation of $f(\mathbf{a})$ around a fixed operating point $\mathbf{a}^{(i)}$, given in (25) or (29) depending upon the choice of $f(\mathbf{a})$, which is either (23) or (27). Upon determining the functions $f(\mathbf{a})$ and $\hat{f}(\mathbf{a};\mathbf{a}^{(i)})$, the resulting subproblem (34) is solved by following the discussions presented in Section IV. Unlike $\widehat{P}_1$ and $\widehat{P}_2$, an initial operating point can be chosen randomly as there is no guaranteed QoS requirement in the constraint set of (34). Thus, finding an initial feasible point is a simple procedure for (34).

Due to the iterative nature of problem (33), finding minimum rate $r_g$ through bisection search is not an efficient method as it introduces nested iterations, i.e., outer bisection search and the inner feasibility check SCA loop for a fixed minimum guaranteed rate $r_g, \forall g \in \mathcal{G}$. Therefore, to avoid nested iterations, the minimum rate $r_g$ is considered as an optimization variable and it is solved along with the other variables iteratively. The algorithmic description for designing beamformers with the objective of providing fairness among multicast group rates is outlined in Algorithm 3 and the convergence behavior of (33) is discussed in Appendix A.

## VI. Numerical Simulations

In this section, we compare the performance of the proposed algorithms to those of the known solutions in [6], [9], [10], [20]. We divide this section into three parts wherein the first subsection discuss the design of multicast beamformers for multiple groups with single sub-channel, i.e., with $N = 1$. Then, we discuss the performance of the proposed methods for a multi-group multiple sub-channels setting. Finally, we discuss the performance of max-min beamformer design for

**Algorithm 3** Proposed iterative algorithm for solving $P_3$

1: Initialize $\{\mathbf{m}^{(0)}\}$ and $\{\boldsymbol{\gamma}^{(0)}\}$ randomly to perform Algorithm 1 until convergence by replacing (15a) and (15b) with (32a) and (32b), respectively.
2: Assign the obtained solution $\{\mathbf{m}_*\}$ and $\{\boldsymbol{\gamma}_*\}$ to $\{\mathbf{m}^{(0)}\}$ and $\{\boldsymbol{\gamma}^{(0)}\}$, respectively.
3: Set $\mathbf{a}^{(0)} = \frac{N_{\mathrm{RF}}}{N_T}\mathbf{1}^{\mathrm{T}}$ to solve (34).
4: **repeat**
5: Solve (34) with $\hat{f}(\mathbf{a};\mathbf{a}^{(i)})$ as (25) for log-sum penalty or (29) for entropy penalty function. Let the solution $\{\mathbf{m}\}$,$\{\boldsymbol{\gamma}\}$ and $\mathbf{a}$ obtained from the $i$th SCA step be denoted as $\{\mathbf{m}_*^{(i)}\}$, $\{\boldsymbol{\gamma}_*^{(i)}\}$ and $\mathbf{a}_*^{(i)}$, respectively.
6: Set $\{\mathbf{m}^{(i+1)}\}$,$\{\boldsymbol{\gamma}^{(i+1)}\}$ and $\mathbf{a}^{(i+1)}$ to $\{\mathbf{m}_*^{(i)}\}$, $\{\boldsymbol{\gamma}_*^{(i)}\}$, and $\mathbf{a}_*^{(i)}$, respectively and update $i = i+1$.
7: **until** convergence or for maximum number of iterations
8: Select $N_{\mathrm{RF}}$ maximum entries from the sorted $\mathbf{a}$ and perform steps 6-12 of Algorithm 1 by retaining $N_{\mathrm{RF}}$ channel entries $\mathbf{h}_{k,n}, \forall k, \forall n$ for which $a_t = 1$, $\forall t$

multi-group multicasting with multiple sub-channels. For simplicity, the path loss seen by all users is fixed to 0 dB in all simulations and the channels are drawn from a zero-mean complex Gaussian distribution unless stated otherwise. The noise variance is normalized to $N_0 = 1$ and the transmit power is measured in dBm units. All the convex problems are modeled with YALMIP [43] and solved by using MOSEK solver [36]. For the SDP problems, if the solution obtained has rank greater than one, then the randomization technique with 500 samples as in [6] is used to generate rank-one beamforming vectors from the high rank solution and choosing the one that requires minimal transmission power.

Apart from the proposed methods in Sections IV and V, we have included two more schemes for better comparison. The first one is the highly complex exhaustive search wherein all possible combinations of $\binom{N_T}{N_{\mathrm{RF}}}$ are searched to find a best pattern that either minimizes the transmit power or maximizes the minimum achievable rate. Secondly, we have included a low complexity antenna selection by choosing $N_{\mathrm{RF}}$ entries of $\mathbf{a}$ that have high per-antenna power compared to the remaining $N_T - N_{\mathrm{RF}}$ entries, labeled as sorted antenna powers in the figures. It is equivalent of setting $\psi = 0$ in $\widehat{P}_2$ and $\widehat{P}_3$. Unlike $\widehat{P}_2$ with $\psi > 0$, selecting $N_{\mathrm{RF}}$ maximum entries with $\psi = 0$ may not yield a solvable problem for $\widehat{P}_2$ on all cases as there is no guarantee in satisfying the QoS requirements imposed by $\widehat{P}_2$ with the chosen subset. However, $\widehat{P}_3$ is always solvable by setting $\psi = 0$ as there is no guaranteed QoS for the users. To solve $\widehat{P}_2$ and $\widehat{P}_3$, the parameter $\psi$ is fixed as $\psi = 2$ irrespective of the system model and $\epsilon$ is set to $1e^{-5}$ for numerical stability.

*1) Multi-Group Single Sub-Channel Scenario:* We consider the design of multicast beamformers with multiple groups for a single sub-channel, due to which the minimum rate constraint is reduced to a SINR requirement. In order to compare the transmit power required by the various schemes for a fixed QoS demand, we consider an ULA with $N_T = 12$ antennas serving $N_G = 3$ groups, consisting of $|\mathcal{G}_g| = 40$ users each. The three multicast groups are centered around $0°$, $45°$ and $-45°$, respectively. The users within each group are separated

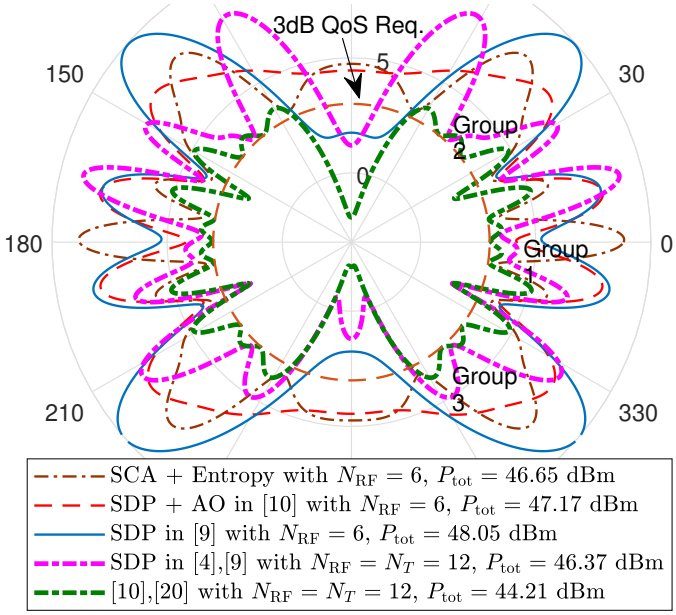



Fig. 1. Radiation plot for a ULA with $\{N_T, N, K, N_G\} = \{12, 1, 120, 3\}$, $N_{\mathrm{RF}} = 6$ and multicast groups centered at $0°$, $45°$, and $-45°$.

by $1°$. The SINR requirement is set to 3 dB for all users. The channel is generated as $[1\, e^{j\theta_k} \ldots e^{j(N_T-1)\theta_k}]$, where $\theta_k$ is the angular frequency related to the location of user $k$ with respect to the array boresight and the gain is normalized to unity.

Fig. 1 shows the radiation pattern of the proposed and existing SDP schemes in [9] and [10] with $N_{\mathrm{RF}} = 6$ antennas for a ULA model. The radial axis in Fig. 1 corresponds to the uniform received power in logarithmic scale. Due to the multi-group interference, the actual transmit power is larger than the desired QoS of 3 dB. Fig. 1 shows that all schemes ensure the SINR of 3 dB to all users. The total power required to achieve the minimum SINR requirement of 3 dB is less for SCA based solutions while comparing with the SDP approach proposed in [9]. Fig. 1 includes only the entropy minimization scheme as a reference model for the SCA based design with $N_{\mathrm{RF}} = 6$. As mentioned in Fig. 1, the total transmit power required by the proposed SCA based scheme is 0.5 dB less than the SDR based approach in [10] and $\approx 1.5$ dB less than the one in [9] while using $N_{\mathrm{RF}} = 6$ antennas for transmission. Moreover, the proposed SCA based method outperforms the SDR approach in [4], [9] by 2 dB when $N_{\mathrm{RF}} = N_T$. On the contrary, the SDR technique in [10] performs the same as that of the SCA method in [20] while enabling all antennas.

In Fig. 2, we consider a fading scenario, where the channels are generated randomly from $\mathbf{h}_{k,n} \sim \mathcal{CN}(0, \mathbf{I}_{N_T})$. The figure compares the average transmit power required to provide the QoS of 3 bits to all users by the proposed and an existing SDR based schemes in [9], [10] for solving $P_2$. We consider a model with $N_G = 2$ multicast groups consisting of 10 users each. The SDP method proposed in [9] yields solution with rank greater than one, whereas the biconvex SDP formulation in [11] generates rank-one solution mostly. For comparison, we have also plotted the antenna selection based on the sorted power, which is nothing but solving (22) without $f(\mathbf{a})$ and selecting the elements corresponding to the indices of maximum $N_{\mathrm{RF}}$ entries from $\mathbf{a}$. It is evident from Fig. 2 that solving (22)

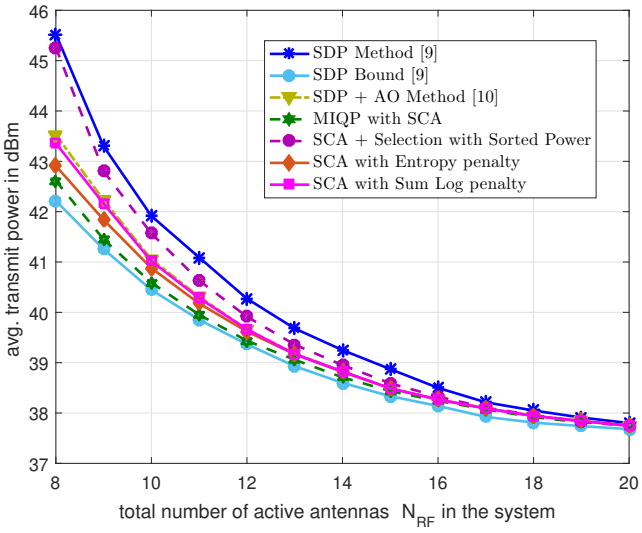
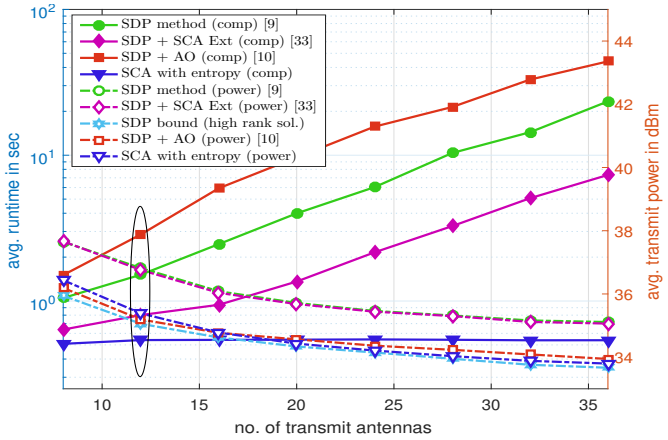


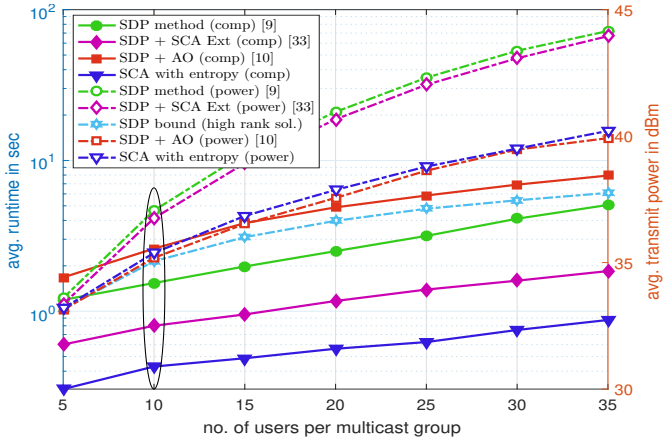

Fig. 2. Average transmit power requirement for model $\{N_T, N, K, N_G\} = \{20, 1, 20, 2\}$ vs. the number of active elements $N_{\mathrm{RF}}$ with $\bar{r}_g = 3$ bits.

without the penalty function requires more transmit power compared to that of the regularized problems. Fig. 2 also depicts that as we increase $N_{\mathrm{RF}}$ to $N_T$, the transmit powers required by the SDP methods in [9], [10] are comparable to that of the SCA based techniques. Additionally, the SDP scheme in [10] performs similar to that the penalty based SCA methods. The superior performance of the SCA technique is due to the inherent rank-one restriction on the solution unlike the SDP method wherein the problem is solved by relaxing the rank-one constraint, which requires randomization search to extract a rank-one vector if the solution has higher rank.

Fig. 3 compares both the computational time and the total transmit power required by [9], [10] and the proposed entropy based solution for solving $\widehat{P}_2$. The system model is as mentioned in Fig. 3 along with the QoS requirement of $\bar{r}_g = 3$ bits. In both figures of Fig. 3, the l.h.s denotes the computational complexity, whereas the r.h.s exhibits the transmit power requirement. The randomization search is performed over 100 samples if the solution has rank greater than one. In the case of [10], 10 iterations are used to determine a subset of antennas, which is followed by the alternating optimization (AO) procedure for finding the beamformers. Both [10] and the SCA scheme are iterated until two subsequent objective values differ by $10^{-3}$. Upon increasing $N_T$, the computational times of the SDR based techniques scales-up quickly in the order of $\mathcal{O}(N_T^6)$, whereas the proposed SCA algorithm increases modestly by $\mathcal{O}(N_T^3)$ as shown in Fig. 3(a). However, unlike varying $N_T$, the complexity of all the algorithms scales-up similarly while increasing $|\mathcal{G}_g|$ as shown in Fig. 3(b), since the complexity scales by $\mathcal{O}(K^3)$ for all schemes as shown in Appendix B. Regarding the algorithm performance, as $N_T$ increases, the transmit power required by all the schemes decreases monotonically due to the additional spatial degrees of freedom (DoF) provided by the increasing $N_T$ as shown in Fig. 3(a), whereas, the transmit power scales-up quickly when $|\mathcal{G}_g|$ increases as shown in Fig. 3(b). Furthermore, the performance difference between the SDP and the SCA schemes grows significantly with the number of users in each group, which is




(a). For variable $N_T$ with $|\mathcal{G}_g| = 10$ users.




(b). For variable $|\mathcal{G}_g|$ with $N_T = 12$ antennas.

Fig. 3. Average computational time for a system $\{N_G, N, N_{\mathrm{RF}}\} = \{1, 1, 8\}$ with $\bar{r}_g = 3$ bits using i.i.d channel. Ellipses denote the same configuration.

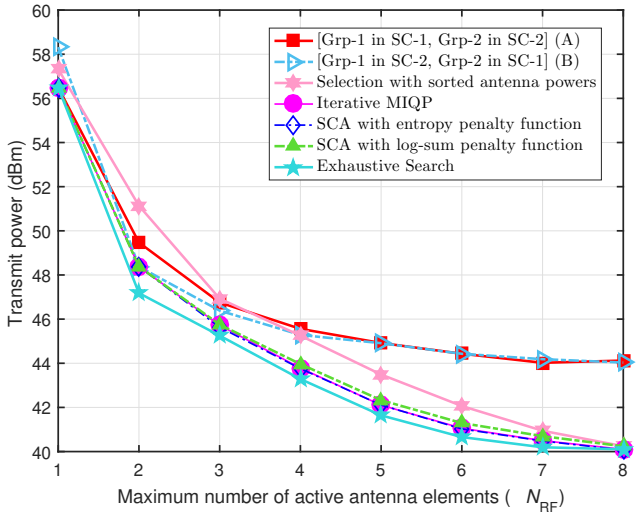



Fig. 4. Average transmit power requirement for model $\{N_T, N, K, N_G\} = \{8, 2, 10, 2\}$ vs. the number of active elements and $\bar{r}_g = 5$ bits.

due to high rank solutions produced by the SDP based schemes in [9], [33]. Even though beamformers are designed based on biconvex SDP formulation in [10], it performs similar to that of the proposed SCA based technique, since the solutions produced by the method in [10] are predominantly rank-one.

*2) Multi-Group with Multiple Sub-Channels:* We consider a scenario with $N_T = 8$ transmit antennas and $N = 2$ sub-channels, serving $N_G = 2$ multicast groups with $|\mathcal{G}_g| = 5$ users each. As there are two multicast groups that are to be served over two sub-channels, it requires at least $N_{\mathrm{RF}} = 1$ active antennas for an interference-free transmission by an orthogonal scheduling of multicast groups over the sub-channels. The QoS requirement of all users is fixed to 5 bits and the path loss seen by all users is set to $0\,\mathrm{dB}$. Fig. 4 compares the transmit power required by various schemes for a given QoS requirement. The total transmit power required by the techniques in Section IV are comparable to that of the exhaustive search, whereas the selection based on the sorted antenna powers is inferior compared to the proposed schemes with $\psi > 0$. Fig. 4 also includes two possible user allocations wherein the multicast groups are allocated orthogonally onto independent sub-channels as shown by legends the $A$ and $B$. For the subset selections with $N_{\mathrm{RF}} = \{1, 2\}$, the multicast groups are assigned orthogonally over non-overlapping sub-channels for an interference-free transmission. As shown in Fig. 4, when $N_{\mathrm{RF}} \geq N_G$, the orthogonal allocation schemes are not optimal when compared to the proposed techniques, due to under utilization of the available frequency resources.

In case of multiple sub-channels, we extended the SDP-SCA method in [33] to perform antenna selection by including a penalty term in the objective as in (22a). Fig. 5 compares both the total transmit power and the computational complexity among the entropy based SCA technique and an extension of [33] by varying each one of the system parameters by keeping the remaining ones fixed. The comparisons are made between the two schemes by varying the number of transmit antennas in Fig. 5(a), sub-channels in Fig. 5(b), multicast groups in Fig. 5(c), and the number of users in each multicast group in Fig. 5(d). Increasing either the number of transmit antennas $N_T$ or sub-channels $N$, the transmit power required to provide certain QoS decreases as shown in Figs. 5(a) and 5(b), respectively. This behavior is due to the additional spatial and frequency DoF provided by the increase in $N_T$ and $N$, respectively. Thus, by increasing $N_T$ for a fixed number of active elements $N_{\mathrm{RF}} = 8$, the transmit powers of both the schemes are minimized with a fixed gap as shown in Fig. 5(a). However, as $|\mathcal{G}_g|$ increases, the probability of producing high rank solutions by the SDP-SCA method increases, since the rank of SDP solution is bounded by the number of constraints, which is $|\mathcal{G}_g|$, thereby yielding inferior rank-one vectors compared to that of the SCA method for a given complexity as in Fig. 5(d). As $N_G$ increases, the transmit power required by the SDP method is higher compared to the SCA technique in Fig. 5(c). It is due to the increase in the randomness involved in finding $N_G$ rank-one beamformers as the multicast group increases.

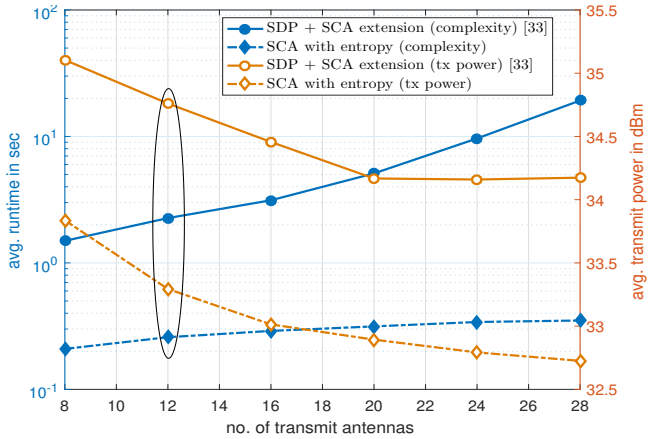
As shown in Appendix B, as $N_T$ increases, the complexity of the SDP approach scales up quadratically with respect to the SCA technique, since the complexities are in the order of $\mathcal{O}(N_T^6)$ and $\mathcal{O}(N_T^3)$, respectively. This behavior can be seen in Fig. 5(a). However, if either the number of sub-channels or multicast groups increases, the complexity scales up similarly for both the SCA and SDP schemes as shown in Figs. 5(b) and 5(c). On the contrary, if the number of users in each group increases as in Fig. 5(d), then the complexity of both the SDP and the SCA schemes scales up modestly and the gap between them diminishes as $K$ increases. It can be explained as follows. The complexity of both schemes grow as $\mathcal{O}(N^3K^3)$ with an additional fixed overhead, which depends on $N_T$, $N_G$ and $N$. Therefore, if $K \ll N_T$, the SDP complexity is dominated by the term $\mathcal{O}(N_T^6 N_G^3 N^3)$, which is greater than the SCA complexity of $\mathcal{O}(N_T^3 N_G^3 N^3)$. However, if $K \gg N_T$, both the SCA and the SDP complexities are comparable as the fixed overhead due to $\mathcal{O}(N_T^3 N_G^3 N^3)$ and $\mathcal{O}(N_T^6 N_G^3 N^3)$ is much lower than $\mathcal{O}(N^3K^3)$ while keeping $N_T, N_G$ and $N$ fixed. When $\mathcal{G}_g = 30$, the randomization search is unable to extract a rank-one beamformer for the SDP-SCA method in [33], hence, the corresponding data point is not shown in Fig. 5(d).

*3) Multicast Beamformer Design with Fairness Objective:* We study the beamformer design aiming at maximizing the minimum achievable rate of all users in a multi-group multiple sub-channel case. Algorithm 3 is performed for a model consisting of $N_T = 8$ transmit antennas with $N = 2$ sub-channels, serving $N_G = 2$ multicast groups with 5 users each. Due to the available DoF provided by $N = 2$ sub-channels, the minimum number of antennas required to achieve a nonzero rate to all users can be as low as one element. The total transmit power $P_{\mathrm{tot}}$ is fixed at $40\,\mathrm{dBm}$ in order to compare with Fig. 4, which has the same system settings.

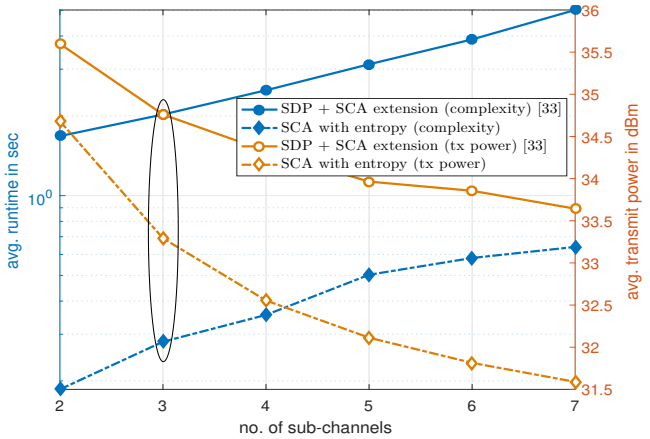
Fig. 6 plots the minimum rate achieved by various beamformer designs presented in Section V for the users belonging to two multicast groups by varying $N_{\mathrm{RF}}$. The average minimum rate achieved by the Algorithm 3 using both penalty functions are fairly comparable, whereas the selection scheme with $\psi = 0$ is inferior compared to that of the penalty based approaches. As we can see from Fig. 6 that the achievable rate between the proposed methods with $\psi > 0$ and that of the exhaustive search is quite small, meaning that Algorithm 3 is efficient in solving $P_3$. For the same reason as in Fig. 4, we also see in Fig. 6 that all the proposed beamformer designs are able to offer nonzero rate even for $N_{\mathrm{RF}} = 1$, using orthogonal sub-channel allocation for multiplexing multicast groups. However, for a multiple antenna transmission, allocating users over all the sub-channels proves to be more efficient than the orthogonal allocations in terms of the minimum rate. It ensures that the proposed multicast design for multiple sub-channels exploits both the spatial and frequency resources efficiently in order to maximize the minimum guaranteed rate of all users.

It is worth noting the similarity between Figs. 4 and 6, having the same system model, in terms of the performance metrics when $N_{\mathrm{RF}} = N_T$. In Fig. 4, a minimum transmit power of $40\,\mathrm{dBm}$ is required to provide a QoS of 5 bits to all users with $N_{\mathrm{RF}} = N_T = 8$. On the other hand, Fig. 6 shows that the minimum achievable rate is $\approx 5$ bits for a power budget of $P_{\mathrm{tot}} = 40\,\mathrm{dBm}$, thereby establishing an relation between problems $P_2$ and $P_3$ having different objectives.

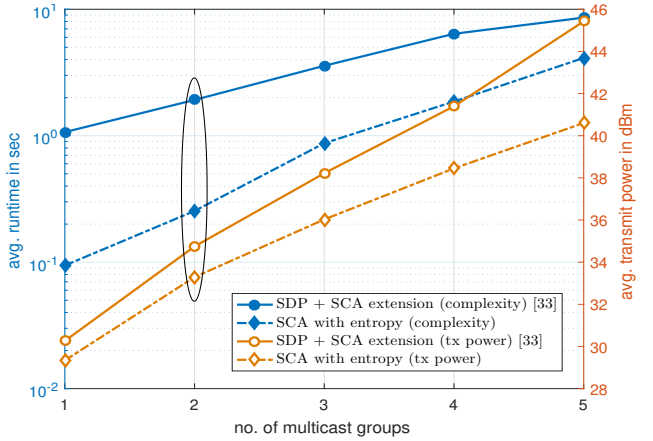
Finally, we conclude with Fig. 7 comparing both objectives, namely, power minimization and max-min fairness, by varying the number of sub-channels while fixing the remaining parameters. The l.h.s y-coordinate denotes the total transmit power and the r.h.s y-coordinate marks the minimum achievable rate for all users. The system model consists of a BS with $N_T = 12$



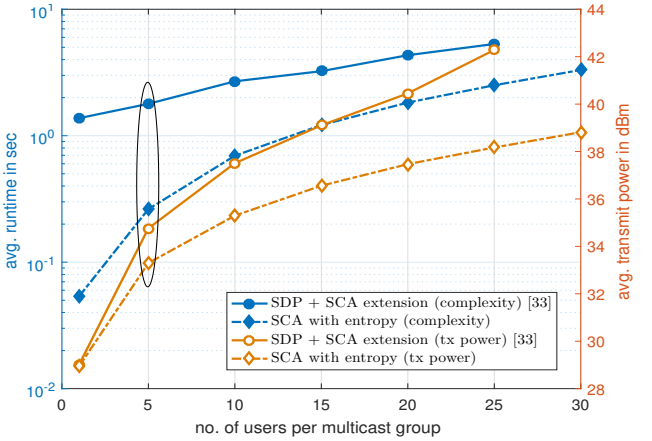

(a). For variable $N_T$ with $\{N_G, |\mathcal{G}_g|, N\} = \{2, 5, 3\}$.




(b). For variable $N$ with $\{N_T, N_G, |\mathcal{G}_g|\} = \{12, 2, 5\}$.




(c). For variable $N_G$ with $\{N_T, |\mathcal{G}_g|, N\} = \{12, 5, 3\}$.




(d). For variable $|\mathcal{G}_g|$ with $\{N_T, N_G, N\} = \{12, 2, 3\}$.

Fig. 5. Average computational time for system with $N_{\mathrm{RF}} = 8$ active elements for $\bar{r}_g = 3$ bits using an i.i.d channel model for evaluation. The ellipse markings in all figures are used to denote similar type of system configuration, which are used for cross-reference.

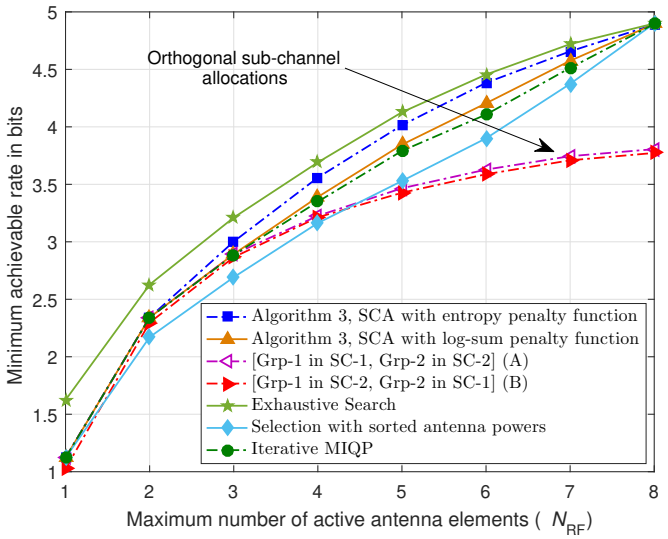



Fig. 6. Average minimum achievable user rate for model $\{N_T, N, K, N_G\} = \{8, 2, 10, 2\}$ vs. the number of active elements and $P_{\mathrm{tot}} = 40\,\mathrm{dBm}$.

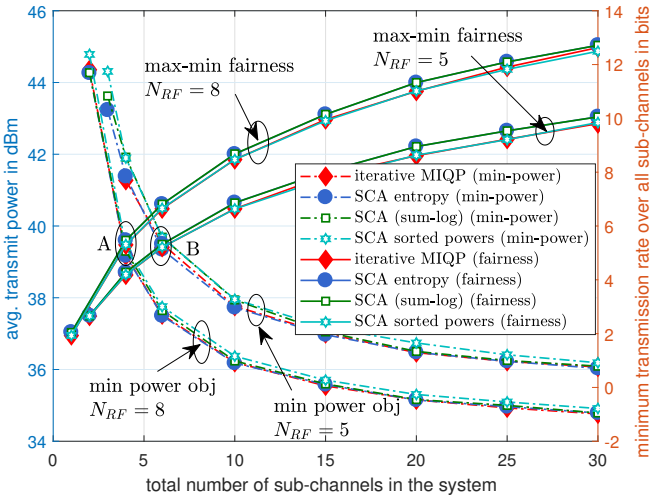



Fig. 7. Average transmit power and minimum achievable user rate for system with $\{N_T, |\mathcal{G}_g|, N_G\} = \{12, 5, 3\}$ vs. the number sub-channels $N$ and $N_{\mathrm{RF}} = 5, 8$. Additionally, $P_{\mathrm{tot}} = 40\,\mathrm{dBm}$ and $\bar{r}_g = 5$ bits.

transmit antennas and $N_G = 3$ multicast groups, constitutes of $|\mathcal{G}_g| = 5$ users each. Comparison is made by varying the number of sub-channels used for two different values of RF chains, namely, $N_{\mathrm{RF}} = 8$ and $N_{\mathrm{RF}} = 5$. In the case of power

minimization objective, the minimum guaranteed rate of all users are kept as $\bar{r}_g = 5$ bits, and for the max-min fairness objective, the total transmit power is restricted to $40\,\mathrm{dBm}$.

Two important observations can be made from Fig. 7 when the number of sub-channels $N$ increases. The total transmit power required for a power minimization objective decreases and the minimum achievable rate by all users for the max-min fairness objective increases. This behavior of the proposed SCA based solutions is attributed to an efficient utilization of the frequency resources provided by the increasing number of sub-channels used for transmission. In the case of $N_{\mathrm{RF}} = 8$, the solution obtained by solving the problem of max-min fairness $\widehat{P}_3$ can be used as an initial feasible point to solve the min-power design problem $\widehat{P}_2$ without solving (17) for the system setups with $N \geq 4$, which is marked as $(\mathrm{A})$ in Fig. 7. It is because of the fact that the total transmit power required to achieve $\bar{r}_g = 5$ bits is $\approx 40\,\mathrm{dBm}$ for $\widehat{P}_2$ and the minimum achievable rate for $\widehat{P}_3$ is $5$ bits when the transmit power is restricted to $40\,\mathrm{dBm}$. Similar discussion holds for the system setup with $N_{\mathrm{RF}} = 5$ and $N \geq 6$. This transition point is denoted in Fig. 7 with the help of marker $(\mathrm{B})$.

## VII. Conclusions

We considered the design of multicast beamformers for multiple groups with the min-power objective under certain quality-of-service requirements and multiple-input single-output orthogonal frequency division multiplexing. Due to nonconvexity of the problem formulation, we employed SCA technique and solved the resulting convex subproblems iteratively. We then extended our formulation to limit the number of antennas used for a multicast transmission. Thus, to select a subset of transmit antennas, we have introduced a binary variable to handle the active status of each antenna. Since the resulting problem is combinatorial, we have relaxed the binary variable to be continuous and solved the problem efficiently by employing the SCA technique. Moreover, we have augmented an additional penalty term in the objective to encourage sparsity in the relaxed binary variable upon convergence. Finally, we have also addressed the problem of providing fairness among multicast groups by maximizing the minimum achievable rate of users for a given transmit power. The performances of the proposed methods were demonstrated by the numerical simulations. Furthermore, we also compared the results of existing solutions based on semidefinite relaxation with the proposed schemes, which was based on the SCA technique. The proposed solutions outperformed the existing methods in all our numerical experiments.

## References


[1] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-advanced for mobile broadband*. Academic press, 2013.

[2] S. Sesia, M. Baker, and I. Toufik, *LTE-the UMTS long term evolution: from theory to practice*. John Wiley & Sons, 2011.

[3] M. Gruber and D. Zeller, "Multimedia broadcast multicast service: New transmission schemes and related challenges," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 176–181, December 2011.

[4] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, 2006.

[5] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Far-field multicast beamforming for uniform linear antenna arrays," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4916–4927, oct 2007.

[6] ——, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1268–1279, 2008.

[7] C. Suh and J. Mo, "Resource allocation for multicast services in multicarrier wireless communications," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 27–31, 2008.

[8] A. Schad and M. Pesavento, "Max-min fair transmit beamforming for multi-group multicasting," in *2012 International ITG Workshop on Smart Antennas (WSA)*, March 2012, pp. 115–118.

[9] O. Mehanna, N. Sidiropoulos, and G. Giannakis, "Joint multicast beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2660–2674, 2013.

[10] Ö. T. Demir and T. E. Tuncer, "Multicast beamforming with antenna selection using exact penalty approach," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 2489–2493.

[11] ——, "Alternating maximization algorithm for the broadcast beamforming," in *22nd European Signal Processing Conference (EUSIPCO)*. IEEE, 2014, pp. 1915–1919.

[12] B. Gopalakrishnan and N. D. Sidiropoulos, "High performance adaptive algorithms for single-group multicast beamforming," *IEEE Trans. Signal Process.*, vol. 63, no. 16, pp. 4373–4384, 2015.

[13] J. Joung, H. D. Nguyen, P. H. Tan, and S. Sun, "Multicast linear precoding for MIMO-OFDM systems," *IEEE Commun. Lett.*, vol. 19, no. 6, pp. 993–996, 2015.

[14] J. Xu, S.-J. Lee, W.-S. Kang, and J.-S. Seo, "Adaptive resource allocation for MIMO-OFDM based wireless multicast systems," *IEEE Transactions on Broadcasting*, vol. 56, no. 1, pp. 98–102, 2010.

[15] H. Won, H. Cai, D. Y. Eun, K. Guo, A. Netravali, I. Rhee, and K. Sabnani, "Multicast scheduling in cellular data networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4540–4549, 2009.

[16] J. Chen, M. Chiang, J. Erman, G. Li, K. Ramakrishnan, and R. K. Sinha, "Fair and optimal resource allocation for lte multicast (embms): Group partitioning and dynamics," in *IEEE INFOCOM*, 2015.

[17] R. O. Afolabi, A. Dadlani, and K. Kim, "Multicast scheduling and resource allocation algorithms for OFDMA based systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 240–254, 2013.

[18] Z.-Q. Luo, W.-K. Ma, A.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[19] L.-N. Tran, M. F. Hanif, and M. Juntti, "A conic quadratic programming approach to physical layer multicasting for large-scale antenna arrays," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 114–117, 2014.

[20] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar, and N. D. Sidiropoulos, "Feasible point pursuit and successive approximation of non-convex QCQPs," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 804–808, 2015.

[21] G. Dartmann and G. Ascheid, "Equivalent quasi-convex form of the multicast max–min beamforming problem," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4643–4648, 2013.

[22] G. Scutari, F. Facchinei, L. Lampariello, S. Sardellitti, and P. Song, "Parallel and distributed methods for constrained nonconvex optimization-part II: Applications in communications and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 1945–1960, apr 2017.

[23] E. Matskani, N. D. Sidiropoulos, Z. Q. Luo, and L. Tassiulas, "Efficient batch and adaptive approximation algorithms for joint multicast beamforming and admission control," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4882–4894, Dec 2009.

[24] Z. Xiang, M. Tao, and X. Wang, "Coordinated multicast beamforming in multicell networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 12–21, 2013.

[25] N. Bornhorst, M. Pesavento, and A. B. Gershman, "Distributed beamforming for multi-group multicasting relay networks," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 221–232, 2012.

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

[27] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Weighted fair multicast multigroup beamforming under per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5132–5142, 2014.

[28] ——, "Multicast multigroup beamforming for per-antenna power constrained large-scale arrays," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2015, pp. 271–275.

[29] N. Jindal and Z.-Q. Luo, "Capacity limits of multiple antenna multicast," in *IEEE International Symposium on Information Theory*. IEEE, 2006, pp. 1841–1845.

[30] S. Y. Park and D. J. Love, "Capacity limits of multiple antenna multicasting using antenna subset selection," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2524–2534, 2008.

[31] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, ser. Wiley Series in Telecommunications. Cambridge University Pr., 2005.

[32] N. D. Sidiropoulos and Z.-Q. Luo, "A semidefinite relaxation approach to MIMO detection for high-order QAM constellations," *IEEE Signal Process. Lett.*, vol. 13, no. 9, p. 525, 2006.

[33] G. Venkatraman, A. Tölli, M. Juntti, and L.-N. Tran, "Multi-group multicast beamformer design for MIMO-OFDM transmission," in *22nd European Wireless Conference*, May 2016, pp. 1–6.

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

[35] A. Ben-Tal and A. Nemirovski, "On polyhedral approximations of the second-order cone," *Mathematics of Operations Research*, vol. 26, no. 2, pp. 193–205, 2001.

[36] MOSEK ApS, *The MOSEK Optimization Toolbox for MATLAB Manual. Version 7.1 (Revision 28).*, 2015.

[37] Gurobi Optimization, Inc., "Gurobi Optimizer Reference Manual," 2015. [Online]. Available: http://www.gurobi.com

[38] F. Rinaldi, "New results on the equivalence between zero-one programming and continuous concave programming," *Optimization Letters*, vol. 3, no. 3, pp. 377–386, 2009.

[39] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Structured sparsity through convex optimization," *Statistical Science*, vol. 27, no. 4, pp. 450–468, 2012.

[40] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted $\ell_1$ minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[41] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[42] S. Lucidi and F. Rinaldi, "Exact penalty functions for nonlinear integer programming problems," *Journal of optimization theory and applications*, vol. 145, no. 3, pp. 479–488, 2010.

[43] J. Löfberg, "YALMIP : A toolbox for modeling and optimization in MATLAB," in *Proceedings of the CACSD Conference*, Taiwan, 2004.

[44] W. Rudin *et al.*, *Principles of mathematical analysis*. McGraw-Hill New York, 1964, vol. 3.

[45] G. Scutari, F. Facchinei, and L. Lampariello, "Parallel and distributed methods for constrained nonconvex optimization—part i: Theory," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 1929–1944, apr 2017.

[46] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Athena Scientific, sep 1999.

[47] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*. Prentice hall Englewood Cliffs, NJ, 1989, vol. 23.

[48] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *Journal of Global Optimization*, vol. 47, no. 1, pp. 29–51, 2010.

[49] A. Nemirovskii, "Interior point polynomial time methods in convex programming," *Lecture Notes*, 2004.

[50] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. Siam, 2001, vol. 2.


## APPENDIX A
## CONVERGENCE ANALYSIS

In order to show the convergence of proposed iterative methods in Algorithms 1, 2 and 3, we need to ensure the following conditions.

(a) Objective sequence should be strictly decreasing and bounded from below

(b) Sequence of iterates (stacked solution vector) should be bounded or the feasible set of each subproblem is compact

Upon satisfying the conditions (a) and (b), we can ensure the convergence of objective sequence generated by Algorithms 1, 2 and 3 by using [44, Th. 3.14]. Additionally, by using the discussions in [34], [45], we can show that every limit point of bounded sequence of iterates generated by the iterative methods is a stationary point of the nonconvex problem.

Due to the presence of discrete constraints in $P_2$ and $P_3$, the existence of stationary points cannot be guaranteed. Therefore, we can only state that the solution obtained upon the convergence of iterative methods in Algorithms 2 and 3 is a stationary point for the respective relaxed nonconvex problems $\widehat{P}_2$ and $\widehat{P}_3$ and not for $P_2$ and $P_3$. With these assumptions, let us consider using a generalized formulation to represent nonconvex problem $P_1$ and the relaxed binary formulations $\widehat{P}_2$ and $\widehat{P}_3$ as

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) = f_1(\mathbf{x}) - f_2(\mathbf{x}) && \text{(35a)} \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0,\, \forall i \in \mathcal{I} && \text{(35b)} \\ & \widetilde{g}_i(\mathbf{x}) \leq 0,\, \forall i \in \mathcal{I}' && \text{(35c)} \\ & h_j(\mathbf{x}) = 0,\, \forall j \in \mathcal{J} && \text{(35d)} \end{aligned}$$

where $\mathbf{x}$ is a vector formed by stacking all optimization variables, and $\mathcal{I}, \mathcal{I}', \mathcal{J}$ are all index sets. From the problem formulations, it is clear that $f_1(\mathbf{x}), f_2(\mathbf{x})$, and $\widetilde{g}_i(\mathbf{x}), \forall i \in \mathcal{I}'$ are all convex and differentiable functions, $h_j(\mathbf{x}), \forall j \in \mathcal{J}$ are affine, and $g_i(\mathbf{x}), \forall i \in \mathcal{I}$ are all concave functions. Due to the concavity of $g_i(\mathbf{x}), \forall i \in \mathcal{I}$, the corresponding constraints defined by (35b) becomes nonconvex. The feasible set of problem (35), denoted by $\mathcal{X}$, is defined as

$$\begin{aligned} \mathcal{X} = \{\mathbf{x} \,\big|\, & g_i(\mathbf{x}) \leq 0,\, \forall i \in \mathcal{I}, \\ & \widetilde{g}_i(\mathbf{x}) \leq 0, \forall i \in \mathcal{I}',\, h_j(\mathbf{x}) = 0,\, \forall j \in \mathcal{J}\}. \end{aligned} \quad \text{(36)}$$

In order to discuss the convergence of Algorithms 1, 2 and 3, we use the following notations. Let $\mathbf{x}^k$ and $f(\mathbf{x}^k)$ denote, respectively, the stacked vector of the solution and the objective value in the $(k-1)$th SCA iteration. Let us denote the sequence of iterates as $\{\mathbf{x}^k\}$, which is formed by collecting the solution vector from each SCA iteration. Similarly, $\{f(\mathbf{x}^k)\}$ represents the corresponding objective sequence.

### A. Boundedness of the Iterates

The feasible set of $P_1$,$P_2$, and $P_3$ with the SINR expression (2) is closed and bounded. Moreover, the objective function is coercive, i.e., $f(\mathbf{x}) \to \infty$ as $\|\mathbf{x}\| \to \infty$, and therefore the feasible set is also compact, following from [46, Prop. A.8]. Due to the lack of strict convexity in each SCA subproblem, the overall objective sequence generated by the proposed algorithms need not be strictly decreasing, i.e., $f(\mathbf{x}^k) \not> f(\mathbf{x}^{k+1}), \forall k > 0$, which is required to terminate the iterative procedure. In order to overcome this issue, a regularization term can be added to the objective function $f(\mathbf{x})$ without affecting the optimality as

$$f^k(\mathbf{x}) = f(\mathbf{x}) + \tau^k \left\|\mathbf{x} - \mathbf{x}^k\right\|_2^2 \quad \text{(37)}$$

where $\mathbf{x}^k$ is the solution obtained in the $(k-1)$th SCA iteration and the positive constant $\tau^k > 0$ ensures strong convexity in $f^k(\mathbf{x})$ for each step [47, Sec. 3.4.3].

### B. Boundedness of the Objective Sequence

This is obvious from the fact that the norm function is bounded from below as $\|\mathbf{x}\|_q \geq 0$, $\forall \mathbf{x}$ and $q \in \mathbb{Z}$.

### C. Strict Monotonicity of the Objective Sequence

The problem in (35) is not convex due to the nonconvexity of both the objective function (35a) and the constraint (35b). Therefore, we adopt the SCA technique by solving a series of convex subproblems instead of the original nonconvex problem. In order to do so, we need to find suitable convex approximations for both the objective function and the constraint, respectively.

We begin by addressing the nonconvexity of the feasible set $\mathcal{X}$, which is resulting from the constraint (35b). In order to find convex subset for (35b), we bound the l.h.s. of (35b), which is a concave function, by the respective first order Taylor approximation around an operating point, say, $\mathbf{x}^k$, $\forall i \in \mathcal{I}$ as

$$\hat{g}_i(\mathbf{x};\mathbf{x}^k) \triangleq g_i(\mathbf{x}^k) + \nabla g_i(\mathbf{x}^k)^{\mathrm{T}}(\mathbf{x}-\mathbf{x}^k) \geq g_i(\mathbf{x}) \qquad (38)$$

where the inequality is due to the concavity of $g_i(\mathbf{x})$, $\forall i \in \mathcal{I}$. As a result, (38) is a convex subset of (35b) and the approximate convex feasible set for the $k$th SCA iteration is given by

$$\mathcal{X}^k = \{\mathbf{x} \,|\, \hat{g}_i(\mathbf{x};\mathbf{x}^k) \leq 0,\ \forall i \in \mathcal{I}, \\ \widetilde{g}_i(\mathbf{x}) \leq 0,\ \forall i \in \mathcal{I}',\ h_j(\mathbf{x}) \leq 0,\ \forall j \in \mathcal{J}\} \qquad (39)$$

and by following (38), we have $\mathcal{X}^k \subset \mathcal{X}$, $\forall k$.

Secondly, we consider approximating the objective function $f(\mathbf{x}) = f_1(\mathbf{x}) - f_2(\mathbf{x})$, which is a DC function. To deal with the nonconvex part, i.e., $f_2(\mathbf{x})$, the proposed algorithms approximate it by the first order Taylor approximation around some fixed operating point $\mathbf{x}^k$ as

$$\hat{f}_2(\mathbf{x};\mathbf{x}^k) \triangleq f_2(\mathbf{x}^k) + \nabla f_2(\mathbf{x}^k)^{\mathrm{T}}(\mathbf{x}-\mathbf{x}^k) \leq f_2(\mathbf{x}). \qquad (40)$$

The approximate function $\hat{f}_2(\mathbf{x};\mathbf{x}^k) \leq f_2(\mathbf{x})$ is a global under-estimator for $f_2(\mathbf{x})$. Therefore, by using (40) instead of $f_2(\mathbf{x})$, the approximate convex objective for the $k$th SCA iteration is given by

$$\hat{f}^k(\mathbf{x};\mathbf{x}^k) = \hat{f}(\mathbf{x};\mathbf{x}^k) + \tau^k \|\mathbf{x}-\mathbf{x}^k\|_2^2 \geq f^k(\mathbf{x}) \qquad (41)$$

where $\hat{f}(\mathbf{x};\mathbf{x}^k) = f_1(\mathbf{x}) - \hat{f}_2(\mathbf{x};\mathbf{x}^k)$, and the inequality in (41) is due to $f_2(\mathbf{x})$.

In summary, we can write the approximate convex subproblem for each SCA iteration $k$ as

$$\underset{\mathbf{x}}{\text{minimize}} \quad \hat{f}^k(\mathbf{x};\mathbf{x}^k) \qquad (42a)$$

$$\text{subject to} \quad \mathbf{x} \in \mathcal{X}^k \qquad (42b)$$

The problem in (35) is convex, and therefore can be solved optimally by the appropriate convex solvers. Let $\mathbf{x}^{k+1}$ be the solution obtained by solving (42) in the $k$th SCA iteration. Due to the strong convexity of the objective function in (41), the optimal solution to (42) is unique. Due to (38), we have

$$g_i(\mathbf{x}^{k+1}) \leq \hat{g}_i(\mathbf{x}^{k+1};\mathbf{x}^k) \leq 0,\ \forall i \in \mathcal{I}. \qquad (43)$$

Therefore, the solution $\mathbf{x}^{k+1}$ obtained by solving (42) in the $k$th SCA iteration is feasible, i.e., $\mathbf{x}^{k+1} \in \mathcal{X}$ with the assumption that initial operating point $\mathbf{x}^k$ is feasible.

Finally, to ensure strictly nonincreasing behavior in the objective sequence, it is enough to show strict monotonic decrease in the objective between two consecutive SCA iterations, say, $k-1$ and $k$, respectively. Let $\mathbf{x}^k \in \mathcal{X}^{k-1}$ be the solution obtained by solving (42) in the $(k-1)$th SCA step. To define $\mathcal{X}^k$, the nonconvex constraints defined by (35b) are approximated by a convex first order Taylor approximation as in (38) around the fixed operating point $\mathbf{x}^k$. Due to this, the previous solution $\mathbf{x}^k$ is also included in the feasible set of the current SCA iteration $k$ as $\mathbf{x}^k \in \mathcal{X}^k$.

Now, by using the fact that $\mathbf{x}^k \in \{\mathcal{X}^{k-1} \cap \mathcal{X}^k\}$, the newly found solution $\mathbf{x}^{k+1}$ in the $k$th SCA iteration obeys

$$\hat{f}^k(\mathbf{x}^{k+1};\mathbf{x}^k) \leq \hat{f}^k(\mathbf{x};\mathbf{x}^k),\ \forall \mathbf{x} \in \mathcal{X}^k \qquad (44)$$

therefore, the objective between two subsequent iterations is non-increasing, thereby proving the monotonicity of the objective sequence. However, to ensure strict monotonic decrease in the objective sequence, we expand (44) using (37) at $\mathbf{x} = \mathbf{x}^k$ as

$$\hat{f}(\mathbf{x}^{k+1};\mathbf{x}^k) + \tau^k \|\mathbf{x}^{k+1}-\mathbf{x}^k\| \leq \hat{f}(\mathbf{x}^k;\mathbf{x}^k) \qquad (45)$$

where $\|\mathbf{x}^k - \mathbf{x}^k\| = 0$. Now, by the nonnegative quadratic term in the l.h.s of (45) and by the monotonic decrease of the objective value in each step, we can ensure that the objective sequence is strictly decreasing and the iterative procedure can be terminated upon reaching $\|\mathbf{x}^{k+1}-\mathbf{x}^k\| \to 0$ as $k \to \infty$.

### D. Stationarity of Limit Points

In order to discuss the convergence of sequence of beamformer iterates generated by the iterative algorithms, we first ensure the convergence of objective sequence $\{\hat{f}^k(\mathbf{x}^k;\mathbf{x}^{k-1})\}$. Function $\{\hat{f}^k(\mathbf{x}^k;\mathbf{x}^{k-1})\}$ is bounded and strictly decreasing by the discussions in Appendices A-B and A-C. Therefore, by using [46, Prop. A.3], we can ensure the convergence of $\{\hat{f}^k(\mathbf{x}^k;\mathbf{x}^{k-1})\}$. Unfortunately, such a claim cannot be made on the convergence of sequence of iterates $\{\mathbf{x}^k\}$. However, Appendix A-A shows that iterates generated by the iterative procedure are bounded. By [46, Prop. A.5], the sequence $\{\mathbf{x}^k\}$ has at least one limit such that there exists a subsequence of $\{\mathbf{x}^k\}$ converging to it.

We now show that every limit point of the bounded sequence $\{\mathbf{x}^k\}$ is a stationary point, *i.e.,* the limit point of every convergent subsequence is a stationary point. In order to prove the stationarity of limit points, let us consider a subsequence $\{\mathbf{x}^{k_j} | j = 0, 1, \dots\}$ of $\{\mathbf{x}^k\}$ that converges to $\bar{\mathbf{x}}$, which is a limit point. Now, by taking the limit as $j \to \infty$ along the subsequence $\{\mathbf{x}^{k_j}\}$, we have

$$\lim_{j\to\infty} \left\{\hat{f}(\mathbf{x}^{k_{j+1}};\mathbf{x}^{k_j}) + \tau^{k_j}\|\mathbf{x}^{k_{j+1}}-\mathbf{x}^{k_j}\|\right\} = \hat{f}(\bar{\mathbf{x}};\bar{\mathbf{x}}). \qquad (46)$$

The above relation in (46) can be shown by using the convergence of $\{\hat{f}^k(\mathbf{x}^{k_j};\mathbf{x}^{k_{j-1}})\}$, and by the continuity of functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x};\mathbf{x}^{k_{j-1}})$ along with the subsequence $\{\mathbf{x}^{k_j}\}$ as

$$\lim_{j\to\infty} \hat{f}^k(\mathbf{x}^{k_{j+1}};\mathbf{x}^{k_j}) = f_1(\lim_{j\to\infty} \mathbf{x}^{k_{j+1}}) - \hat{f}_2(\lim_{j\to\infty} \mathbf{x}^{k_{j+1}};\mathbf{x}^{k_j})$$

$$+ \lim_{j\to\infty} \tau^{k_j}\|\mathbf{x}^{k_{j+1}} - \mathbf{x}^{k_j}\| = f_1(\bar{\mathbf{x}}) - \hat{f}_2(\bar{\mathbf{x}};\bar{\mathbf{x}}) = \hat{f}(\bar{\mathbf{x}};\bar{\mathbf{x}}) \quad (47)$$

where $\lim_{j\to\infty} \|\mathbf{x}^{k_{j+1}} - \mathbf{x}^{k_j}\|$ leads to $\|\bar{\mathbf{x}} - \bar{\mathbf{x}}\| \to 0$.

Due to the fact that $\{\hat{f}^k(\mathbf{x}^k;\mathbf{x}^{k-1})\}$ is convergent, we can ensure that $\{\hat{f}^{k_j}(\mathbf{x}^{k_j};\mathbf{x}^{k_{j-1}})\}$ is also convergent along the subsequence $\{\mathbf{x}^{k_j}\}$, since $\{\mathbf{x}^{k_j}\} \subset \{\mathbf{x}^k\}$. Therefore, by taking the limit as $j \to \infty$ along the subsequence $\{\mathbf{x}^{k_j}\}$ and by using the relation from (46) and (47), we have

$$\lim_{k\to\infty} \hat{f}^k(\mathbf{x}^k;\mathbf{x}^{k-1}) = \lim_{j\to\infty} \hat{f}^{k_j}(\mathbf{x}^{k_j};\mathbf{x}^{k_{j-1}}) = \hat{f}(\bar{\mathbf{x}};\bar{\mathbf{x}}) \quad (48)$$

where the equality in (48) follows from the fact that $\bar{\mathbf{x}}$ is a limit point of $\{\mathbf{x}^k\}$.

Let us prove the stationarity of $\bar{\mathbf{x}}$ by contradiction. Assume that $\bar{\mathbf{x}}$ is not a stationary point. In such a case, there exists some other point $\mathbf{x}' \in \{\mathbf{x}^k\}$, such that $\nabla\hat{f}(\bar{\mathbf{x}};\bar{\mathbf{x}})^{\mathrm{T}}(\mathbf{x}' - \bar{\mathbf{x}}) < 0$ and $\hat{f}(\bar{\mathbf{x}};\bar{\mathbf{x}}) > \hat{f}(\mathbf{x}';\bar{\mathbf{x}})$. Since $\mathbf{x}'$ is a point in $\{\mathbf{x}^k\}$, there exists a subsequence that converges to $\mathbf{x}'$, for which $\mathbf{x}'$ is a limit point. However, by using (48), we have $\hat{f}(\bar{\mathbf{x}};\bar{\mathbf{x}}) = \hat{f}(\mathbf{x}';\bar{\mathbf{x}})$ for all limit points of $\{\mathbf{x}^k\}$, which is a contradiction. Additionally, due to the monotonic decrease in the objective sequence $\{f^{k_j}(\mathbf{x}^{k_j};\mathbf{x}^{k_{j-1}})\}$, we have $\mathbf{x}^{k_j} \to \bar{\mathbf{x}}$ and $f^{k_j}(\mathbf{x}^{k_j};\mathbf{x}^{k_{j-1}}) - f^{k_{j+1}}(\mathbf{x}^{k_{j+1}};\mathbf{x}^{k_j}) \to 0$ as $j \to \infty$. Furthermore, note that $\bar{\mathbf{x}}$ is the solution of (42), therefore, it satisfies (44) over $\mathcal{X}^k$ as $k \to \infty$. Using the above statements, we can show that $\bar{\mathbf{x}}$ satisfies the optimality condition in [46, Prop. 2.1.2] as

$$\nabla f(\bar{\mathbf{x}};\bar{\mathbf{x}})^{\mathrm{T}}(\mathbf{x} - \bar{\mathbf{x}}) \geq 0,\ \forall \mathbf{x} \in \mathcal{X}^k,\ k \to \infty. \quad (49)$$

Thus, $\bar{\mathbf{x}}$ is a stationary point, and by using (48), we can show that every limit point of $\{\mathbf{x}^k\}$ is a stationary point of (42). Using (38) and (41), we can show that as $i \to \infty$, $\hat{g}_i(\mathbf{x};\mathbf{x}^k) \to g_i(\bar{\mathbf{x}})$ and $\hat{f}^k(\mathbf{x};\mathbf{x}^k) \to f^k(\bar{\mathbf{x}})$. Therefore, by following [45, Thm. 2] and [48, Prop. 3.2], we can show that every limit point of bounded sequence of iterates $\{\mathbf{x}^k\}$ generated by the iterative method is a stationary point of problem (35).

## APPENDIX B
## COMPLEXITY ANALYSIS

Using the discussions presented in [49], [50], the computational complexity of the proposed algorithms and the existing techniques are discussed and compared here. In this analysis, we ignored the guaranteed rate constraint (15b) from the consideration. It creates a fixed overhead as it can be approximated by a system of SOC constrains as in [35] and it is scaled by the number of users, which is constant for both the SDP and the SCA based schemes for a multiple sub-channel scenario. Due to the existence of multiple sub-channels, the total number of variables used in (15) is given by $N \times (N_T N_G + K)$, which is attributed to $\mathbf{m}_{g,n}, \gamma_{k,n}$. The number of constraints in (15) corresponds to the total count of multicast users and the sub-channels, which adds up to $NK$.

The total number of iterations required by interior point methods, which is proportional to the number of constraints, is given by $\mathcal{O}((NK)^{0.5}\log(\epsilon^{-1}))$ to obtain an $\epsilon$ optimal solution. The arithmetic complexity of both assembling the Newtons system and to solve them in each SCA iteration requires utmost $\mathcal{O}(\,NK(N_T N_G N + KN)^2 + N_T N_G N + KN)^3\,) \approx \mathcal{O}((N_T N_G N + KN)^3)$ number of computations as the worst case complexity. It can be easily verified that as we increase the number of sub-channels, the overall complexity scales up to the power of three while fixing the remaining parameters. Similarly, if we design beamformers for multiple sub-channels based on the SDP together with the SCA technique proposed in [33], then the total number of interior point iterations required to obtain an $\epsilon$ optimal solution is given by $\mathcal{O}((N_T N_G N + KN)^{0.5}\log(\epsilon^{-1}))$ and the arithmetic complexity required for a Newton iteration is given by $\mathcal{O}((N_T^2 N_G N + KN)^3)$. It is worth noting that both the SDP technique in [33] and the SCA based schemes presented in Section III-B require an outer loop to update the SCA operating point when $N \geq 1$. Hence, the overall complexity is multiplied by a constant $I_{\max}$, which captures the total number of SCA iterations necessary to obtain a solution with reasonable accuracy. On the contrary, for a single sub-channel case, only the SCA technique proposed in Section III-B requires an outer loop to update the operating point whereas the SDP technique proposed in [4], [6], [9] yields a set of transmit beamformers by solving a SDP problem. In terms of computational complexity, the SDP requires utmost computations of $\mathcal{O}((N_T^2 N_G)^3 + KN_G N_T^2)$ and the SCA complexity is in the order of $\mathcal{O}(K(N_T N_G)^2 + (N_T N_G)^3)$, using the discussions in [9], [20], [49].

Furthermore, the complexity of antenna subset selection problems in (30) and (34) also follows similar complexity figures as those in (15). However, due to the introduction of new variables such as $\mathbf{a}, \mathbf{v}$, the variable size will be increased to $N \times (N_T N_G + K) + 2 \times N_T$, thereby leading to a marginal increase in the Newton complexity as it is determined by the dominant terms. It is evident from the complexity figures that the SDP requires more arithmetic computations compared to the SCA based techniques for a multiple sub-channel case. The arithmetic complexity required to find an initial feasible point for the SCA based scheme is given by same factor as $\mathcal{O}((N_T N_G N + KN)^3)$, which is similar to the main SCA algorithm itself, since it is formulated by relaxing the SINR constraint [20]. Similarly, the randomization search required to find an efficient rank-one beamformer for the SDP method requires only $\mathcal{O}(N(N_G^3 + N_G K))$, since each sub-channel can be solved independently upon finding an optimal $\gamma_{k,n}$ [6]. With a single sub-channel and antenna selection, the SDP based solution in [9] requires utmost $\mathcal{O}(N_T^6 N_G^3)$ arithmetic operations per Newton iteration. On the contrary, for a multi-group multicast beamformer design based on the proposed SCA based method requires only $\mathcal{O}(N_T^3 N_G^3)$ operations, which is significantly less when compared the SDP method while keeping the number of users fixed. In addition to the arithmetic complexity and the number of Newton iterations, the SCA also includes a multiplicative factor $I_{\max}$ due to the SCA update procedure. For $N = 1$ scenario, the proposed SCA based formulations can be cast as a QCQP, which has much lower computational complexity compared to the SDP based approach [19], [20].