Rate Splitting for General Multicast

Lingzhi Zhao, Ying Cui
Shanghai Jiao Tong Univ., CN

Sheng Yang
Paris-Saclay Univ., FR

Shlomo Shamai (Shitz)
Technion-Israel Inst. of Tech., IL

Yunbo Han, Yunfei Zhang
Tencent Tech., CN

ICC 2022 - IEEE International Conference on Communications | 978-1-5386-8347-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICC45855.2022.9838824

***Abstract*—Immersive video, such as virtual reality (VR) and multi-view videos, is growing in popularity. Its wireless streaming is an instance of general multicast, extending conventional unicast and multicast, whose effective design is still open. This paper investigates the optimization of general rate splitting with linear beamforming for general multicast. Specifically, we consider a multi-carrier single-cell wireless network where a multi-antenna base station (BS) communicates to multiple single-antenna users via general multicast. Linear beamforming is adopted at the BS, and joint decoding is adopted at each user. We consider the maximization of the weighted sum rate, which is a challenging nonconvex problem. Then, we propose an iterative algorithm for the problem to obtain a KKT point using the concave-convex procedure (CCCP). The proposed optimization framework generalizes the existing ones for rate splitting for various types of services. Finally, we numerically show substantial gains of the proposed solutions over existing schemes and reveal the design insights of general rate splitting for general multicast.**

***Index Terms*—General multicast, general rate splitting, linear beamforming, joint decoding, optimization, concave-convex procedure (CCCP).**

I. Introduction

Conventional mobile Internet services include (traditional) video, audio, web browsing, social networking, software downloading, etc. These services can be supported by unicast, single-group multicast, and multi-group multicast. Immersive video, such as 360 video and multi-view video is growing in popularity. When watching a tiled 360 video, the tiles in a user's current field-of-view (FoV) plus a safe margin are usually transmitted to the user in case of an FoV change. On the other hand, when watching a multi-view video, a user's current view and adjacent views are usually transmitted to the user in case of a view switch. When streaming a popular immersive video to multiple users simultaneously, multiple messages (e.g., tiles for 360 video and views for multi-view video) are transmitted to each user, and one message may be intended for multiple users [1], [2], as illustrated in Fig. 1. This emerging service plays an important role in online gaming, self-driving, and cloud meeting, etc. but cannot perfectly adapt to the conventional transmission schemes mentioned above. This motivates us to consider general multicast (also referred to as general connection [3] and general groupcast [4]) where one message can be intended for any user. Clearly, general multicast includes the three conventional transmission schemes as special cases.

This work was supported in part by the National Key R&D Program of China under Grant 2018YFB1801102 and Natural Science Foundation of Shanghai under Grant 20ZR1425300. *(Corresponding author: Ying Cui, e-mail: cuiying@sjtu.edu.cn)*

(a) 360 video (b) Multi-view video

Fig. 1. Applications of general multicast.

References [1], [2] are pioneer works for supporting wireless streaming of a 360 video [1] and wireless streaming of a multi-view video [2], which are instances of general multicast. Specifically, in [1], [2], Orthogonal Multiple Access (OMA) is adopted to convert general multicast to per resource block single-group multicast. While the OMA-based mechanisms are easy to implement, spatial multiplexing gain is not exploited. On the other hand, non-orthogonal transmission mechanisms achieve higher transmission efficiency but are also more challenging due to interference. Space Division Multiple Access (SDMA) and Non-Orthogonal Multiple Access (NOMA) are two solutions. The cost to suppress interference in SDMA can be high when the channels for some users are spatially aligned, while decoding interference in NOMA may not be possible when the interfering message rate is too high. Thus, SDMA and NOMA may also have unsatisfactory performance. Rate splitting, originally proposed to effectively support unicast services [5], can partially suppress interference and partially decode interference and hence may circumvent the limitations mentioned above.

In [5], [6], the authors investigate the simplest form of rate splitting for unicast, hereafter called 1-layer rate splitting, for the two-user interference channel [5] and two-user multi-antenna broadcast channel [6], respectively. In [8], the authors investigate the precoder optimization of 1-layer rate splitting for unicast for Gaussian multiple-input multiple-output channels. Later, 1-layer rate splitting for unicast is extended to general rate splitting for unicast [9], 1-layer rate splitting for unicast together with a multicast message intended for all users [10], and multi-group multicast [11], respectively. Specifically, [9]–[11] focus on the optimizations of rate splitting with linear beamforming. Optimization-based random linear network coding design for general multicast has been studied in [3] for wired networks. Besides general rate splitting for general multicast has been studied in [4] for discrete memoryless broadcast channels. Here, we are interested in Gaussian fading

Authorized licensed use limited to: Beijing Jiaotong University. Downloaded on October 13,2023 at 07:35:02 UTC from IEEE Xplore. Restrictions apply.

channels and specifically the linear beamforming design from the optimization perspective. Besides, the optimization of rate splitting with linear beamforming for unicast and its slight generalizations in [10], [11] cannot apply to general multicast. Therefore, for general multicast, the optimization of general rate splitting with linear beamforming remains an open problem.

This paper intends to shed some light on the above issue. Specifically, we consider a multi-carrier single-cell wireless network, where a multi-antenna base station (BS) communicates to multiple single-antenna users via general multicast. First, we present general rate splitting for general multicast and characterize the achievable rate regions under linear beamforming at the BS and joint decoding at each user. Then, we optimize the transmission beamforming vectors and rates of sub-message units to maximize the weighted sum rate subject to the achievable rate region constraints and power constraint. Note that the proposed problem formulation includes those in [9]–[12] as special cases. This problem is a challenging nonconvex problem. Next, we propose an iterative algorithm to obtain a KKT point using the concave-convex procedure (CCCP). Finally, we numerically demonstrate substantial gains of the proposed solutions over existing schemes and reveal the design insights of general rate splitting for general multicast.

II. System Model

In this section, we first introduce general multicast in a single-cell wireless network and briefly illustrate its connection with unicast, single-group multicast, and multi-group multicast. Then, we present general rate splitting. Finally, we illustrate the physical layer model and the implementation with linear beamforming and joint decoding.

A. General Multicast

We consider a single-cell wireless network consisting of one BS and K users. Let $\mathcal{K} \triangleq \{1, \ldots, K\}$ denote the set of user indices. The BS has I independent messages. Let $\mathcal{I} \triangleq \{1, \ldots, I\}$ denote the set of I messages. We consider general multicast. Specifically, each user $k \in \mathcal{K}$ can request arbitrary I_k messages in $\mathcal{I}$, denoted by $\mathcal{I}_k \subseteq \mathcal{I}$, from the BS. We do not have any assumptions on $\mathcal{I}_k,\ k \in \mathcal{K}$ except that each message in $\mathcal{I}$ is requested by at least one user, i.e., $\cup_{k\in\mathcal{K}}\mathcal{I}_k = \mathcal{I}$ [3].

To facilitate serving the K users, we partition the message set $\mathcal{I}$ according to the requests from the K users. For all $\mathcal{S} \subseteq \mathcal{K}, \mathcal{S} \neq \emptyset$, let

$$\mathcal{P}_{\mathcal{S}} \triangleq \left(\bigcap_{k\in\mathcal{S}} \mathcal{I}_k\right) \bigcap \left(\mathcal{I} - \bigcup_{k\in\mathcal{K}\backslash\mathcal{S}} \mathcal{I}_k\right) \tag{1}$$

denote the set of the messages that is requested by each user in $\mathcal{S}$ and not requested by any user in $\mathcal{K}\backslash\mathcal{S}$ [1]. Define

$$\boldsymbol{\mathcal{P}} \triangleq \{\mathcal{P}_{\mathcal{S}} | \mathcal{P}_{\mathcal{S}} \neq \emptyset, \mathcal{S} \subseteq \mathcal{K}, \mathcal{S} \neq \emptyset\},$$
$$\boldsymbol{\mathcal{S}} \triangleq \{\mathcal{S} | \mathcal{P}_{\mathcal{S}} \neq \emptyset, \mathcal{S} \subseteq \mathcal{K}, \mathcal{S} \neq \emptyset\}.$$

Thus, $\boldsymbol{\mathcal{P}}$ forms a partition of $\mathcal{I}$ and $\boldsymbol{\mathcal{S}}$ specifies the user groups corresponding to the partition. We refer to each element in $\boldsymbol{\mathcal{P}}$ as a message unit.[1] We can see that different message units in $\boldsymbol{\mathcal{P}}$ are requested by different user groups in $\boldsymbol{\mathcal{S}}$.

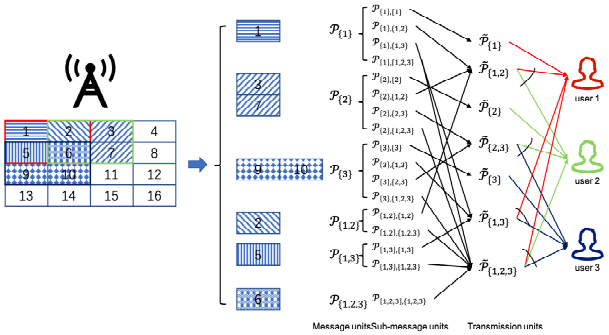

Fig. 2. Wireless streaming of a tiled 360 video to three users. The 360 video is divided into 4×4 tiles. The users have different FoVs which overlap to certain extent. $K = 3$, $I = 8$, $\mathcal{I}_1 = \{1, 2, 5, 6\}$, $\mathcal{I}_2 = \{2, 3, 6, 7\}$, $\mathcal{I}_3 = \{5, 6, 9, 10\}$.

Example 1 (Illustration of $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{S}}$): As illustrated in Fig. 2, we consider $K = 3$, $I = 8$, $\mathcal{I}_1 = \{1, 2, 5, 6\}$, $\mathcal{I}_2 = \{2, 3, 6, 7\}$, $\mathcal{I}_3 = \{5, 6, 9, 10\}$. Then, we have $\mathcal{P}_{\{1\}} = \{1\}$, $\mathcal{P}_{\{2\}} = \{3, 7\}$, $\mathcal{P}_{\{3\}} = \{9, 10\}$, $\mathcal{P}_{\{1,2\}} = \{2\}$, $\mathcal{P}_{\{1,3\}} = \{5\}$, $\mathcal{P}_{\{1,2,3\}} = \{6\}$, $\boldsymbol{\mathcal{P}} = \{\mathcal{P}_{\{1\}}, \mathcal{P}_{\{2\}}, \mathcal{P}_{\{3\}}, \mathcal{P}_{\{1,2\}}, \mathcal{P}_{\{1,3\}}, \mathcal{P}_{\{1,2,3\}}\}$, and $\boldsymbol{\mathcal{S}} = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{1, 2, 3\}\}$. There are 6 message units that are requested by 6 groups of users, respectively. For example, message unit $\mathcal{P}_{\{1\}}$ is requested only by user 1, message unit $\mathcal{P}_{\{1,2\}}$ is requested by user 1 and user 2, and message unit $\mathcal{P}_{\{1,2,3\}}$ is requested by user 1, user 2, and user 3.

Remark 1 (Connection with Unicast and Multicast): The considered general multicast includes conventional unicast, single-group multicast, and multi-group multicast as special cases. When $I = K, I_k = 1, k \in \mathcal{K}$, and $\mathcal{I}_k \neq \mathcal{I}_{k'}, k, k' \in \mathcal{K}, k \neq k'$, general multicast reduces to unicast. In this case, $\boldsymbol{\mathcal{P}} = \{\{1\}, \{2\}, \ldots, \{K\}\}$ and $\boldsymbol{\mathcal{S}} = \{\{1\}, \{2\}, \ldots, \{K\}\}$. When $I = 1$, implying $I_k = 1, k \in \mathcal{K}$, and $\mathcal{I}_k = \mathcal{I}_{k'}, k, k' \in \mathcal{K}, k \neq k'$, general multicast becomes single-group multicast. In this case, $\boldsymbol{\mathcal{P}} = \{\{1\}\}$ and $\boldsymbol{\mathcal{S}} = \{\mathcal{K}\}$. When $1 < I < K$ and $I_k = 1, k \in \mathcal{K}$, general multicast reduces to multi-group (I-group) multicast. In this case, $\boldsymbol{\mathcal{P}} = \{\{1\}, \{2\}, \ldots, \{I\}\}$ and $\boldsymbol{\mathcal{S}} = \{\{k \in \mathcal{K} | \mathcal{I}_k = \{1\}\}, \ldots, \{k \in \mathcal{K} | \mathcal{I}_k = \{I\}\}\}$. The general multicast considered in this paper, general connection in [3], and general groupcast considered in [4] mean the same.

B. General Rate Splitting

We consider rate splitting in the most general form for general multicast to serve the K users [4]. It allows each user group to decode not only the desired message unit $\mathcal{P}_{\mathcal{S}}$ but also part of the message unit of any other user group, $\mathcal{P}_{\mathcal{S}'}$ for all $\mathcal{S}' \neq \mathcal{S}, \mathcal{S}' \in \boldsymbol{\mathcal{S}}$, to flexibly reduce the interference level. For all $\mathcal{S} \in \boldsymbol{\mathcal{S}}$, let $\boldsymbol{\mathcal{G}}_{\mathcal{S}} \triangleq \{\mathcal{X} | \mathcal{S} \subseteq \mathcal{X} \subseteq \mathcal{K}\}$. Namely, $\boldsymbol{\mathcal{G}}_{\mathcal{S}}$ collects all $2^{K-|\mathcal{S}|}$ subsets of $\mathcal{K}$ that contain $\mathcal{S}$. Define $\boldsymbol{\mathcal{G}} \triangleq \bigcup_{\mathcal{S}\in\boldsymbol{\mathcal{S}}} \boldsymbol{\mathcal{G}}_{\mathcal{S}}$.

[1] $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{S}}$ are assumed to be given in [4].

Authorized licensed use limited to: Beijing Jiaotong University. Downloaded on October 13,2023 at 07:35:02 UTC from IEEE Xplore. Restrictions apply.

Obviously, $\boldsymbol{\mathcal{S}} \subseteq \boldsymbol{\mathcal{G}}$. First, we split each message unit $\mathcal{P}_{\mathcal{S}}$ into $2^{K-|\mathcal{S}|}$ sub-message units, i.e.,

$$\mathcal{P}_{\mathcal{S}} = \prod\nolimits_{\mathcal{G}\in\boldsymbol{\mathcal{G}}_{\mathcal{S}}} \mathcal{P}_{\mathcal{S},\mathcal{G}},\ \mathcal{S} \in \boldsymbol{\mathcal{S}}, \tag{2}$$

where $\prod$ represents the Cartesian product. Accordingly, the rate of the message unit $\mathcal{P}_{\mathcal{S}}$, denoted by $R_{\mathcal{S}}$, is split into the rates of the $2^{K-|\mathcal{S}|}$ sub-message units $\mathcal{P}_{\mathcal{S},\mathcal{G}}, \mathcal{G} \in \boldsymbol{\mathcal{G}}_{\mathcal{S}}$,[2] denoted by $R_{\mathcal{S},\mathcal{G}}, \mathcal{G} \in \boldsymbol{\mathcal{G}}_{\mathcal{S}}$ i.e.,

$$R_{\mathcal{S}} = \sum\nolimits_{\mathcal{G}\in\boldsymbol{\mathcal{G}}_{\mathcal{S}}} R_{\mathcal{S},\mathcal{G}},\ \mathcal{S} \in \boldsymbol{\mathcal{S}}. \tag{3}$$

Let $\boldsymbol{\mathcal{S}}_{\mathcal{G}} \triangleq \{\mathcal{S} \in \boldsymbol{\mathcal{S}} | \mathcal{S} \subseteq \mathcal{G}\}$. Then, for all $\mathcal{G} \in \boldsymbol{\mathcal{G}}$, we re-assemble the sub-message units $\mathcal{P}_{\mathcal{S},\mathcal{G}}, \mathcal{S} \in \boldsymbol{\mathcal{S}}_{\mathcal{G}}$ to form a transmission unit $\widetilde{\mathcal{P}}_{\mathcal{G}}$ with rate:

$$\widetilde{R}_{\mathcal{G}} = \sum\nolimits_{\mathcal{S}\in\boldsymbol{\mathcal{S}}_{\mathcal{G}}} R_{\mathcal{S},\mathcal{G}},\ \mathcal{G} \in \boldsymbol{\mathcal{G}}. \tag{4}$$

That is, we first split $|\boldsymbol{\mathcal{S}}|$ message units, $\mathcal{P}_{\mathcal{S}}, \mathcal{S} \in \boldsymbol{\mathcal{S}}$, into $\sum_{\mathcal{S}\in\boldsymbol{\mathcal{S}}} 2^{K-|\mathcal{S}|}$ sub-message units, $\mathcal{P}_{\mathcal{S},\mathcal{G}}, \mathcal{G} \in \boldsymbol{\mathcal{G}}_{\mathcal{S}}, \mathcal{S} \in \boldsymbol{\mathcal{S}}$, and then we re-assemble these sub-message units to form $|\boldsymbol{\mathcal{G}}|$ transmission units, $\widetilde{\mathcal{P}}_{\mathcal{G}}, \mathcal{G} \in \boldsymbol{\mathcal{G}}$.

Example 2 (Illustration of $\boldsymbol{\mathcal{G}}$ and General Rate Splitting): For Example 1, we have $\mathcal{G}_{\{1\}} = \{\{1\},\{1,2\},\{1,3\},\{1,2,3\}\}$, $\mathcal{G}_{\{2\}} = \{\{2\},\{1,2\},\{2,3\},\{1,2,3\}\}$, $\mathcal{G}_{\{3\}} = \{\{3\},\{1,3\},\{2,3\},\{1,2,3\}\}$, $\mathcal{G}_{\{1,2\}} = \{\{1,2\},\{1,2,3\}\}$, $\mathcal{G}_{\{1,3\}} = \{\{1,3\},\{1,2,3\}\}$, $\boldsymbol{\mathcal{G}} = \{\{1\},\{2\},\{3\},\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}\}$. As shown in Fig. 2, we first split 6 message units into 17 sub-message units and then re-assemble the 17 sub-message units to form 7 transmission units.

Remark 2 (Connection with Rate Splitting for Unicast and Multicast): When general multicast degrades to unicast, the proposed general rate splitting reduces to the general rate splitting for unicast proposed in our previous work [9], which extends the one-layer rate splitting for unicast [5]. When general multicast degrades to single-group multicast, the proposed general rate splitting reduces to the conventional single-group multicast transmission as $\boldsymbol{\mathcal{G}}_{\mathcal{S}} = \boldsymbol{\mathcal{G}} = \{\mathcal{K}\}, \mathcal{S} \in \boldsymbol{\mathcal{S}}$. When general multicast degrades to multi-group multicast, the proposed general rate splitting reduces to the one-layer rate splitting for multi-group multicast [11].

C. Physical Layer Model and Implementation

The BS is equipped with M antennas, and each user has one antenna. We consider a multi-carrier system. Let N and $\mathcal{N} \triangleq \{1,2,\ldots,N\}$ denote the number of subcarriers and the set of subcarrier indices, respectively. The bandwidth of each subcarrier is B (in Hz). We consider a discrete-time system, i.e., time is divided into fixed-length slots. We adopt the block fading model, i.e., for each user and subcarrier, the channel remains constant within each slot and is independent and identically distributed (i.i.d.) over slots. We consider slow fading and study an arbitrary slot. Let $\mathbf{h} \triangleq (\mathbf{h}_{k,n})_{k\in\mathcal{K},n\in\mathcal{N}} \in \mathbb{C}^{M\times 1}$ denote the system channel state. Assume that user $k \in \mathcal{K}$ knows his channel state $\mathbf{h}_k \triangleq (\mathbf{h}_{k,n})_{n\in\mathcal{N}}$ and the system channel state $\mathbf{h}$ is known to the BS.

For all $\mathcal{G} \in \boldsymbol{\mathcal{G}}$, transmission unit $\widetilde{\mathcal{P}}_{\mathcal{G}}$ is encoded (channel coding) into codewords that span over the N subcarriers. Let $s_{\mathcal{G},n} \in \mathbb{C}$ denote a symbol for $\widetilde{\mathcal{P}}_{\mathcal{G}}$ that is transmitted on the n-th subcarrier. For all $n \in \mathcal{N}$, let $\mathbf{s}_n \triangleq (s_{\mathcal{G},n})_{\mathcal{G}\in\boldsymbol{\mathcal{G}}}$, and assume that $\mathbb{E}[\mathbf{s}_n\mathbf{s}_n^H] = \mathbf{I}$. We consider linear beamforming. For all $n \in \mathcal{N}$, let $\mathbf{w}_{\mathcal{G},n} \in \mathbb{C}^{M\times 1}$ denote the beamforming vector for transmitting $\widetilde{\mathcal{P}}_{\mathcal{G}}$ on subcarrier n. Using superposition coding, the transmitted signal on subcarrier n, denoted by $\mathbf{x}_n \in \mathbb{C}^{M\times 1}$, is given by:

$$\mathbf{x}_n = \sum\nolimits_{\mathcal{G}\in\boldsymbol{\mathcal{G}}} \mathbf{w}_{\mathcal{G},n} s_{\mathcal{G},n},\ n \in \mathcal{N}. \tag{5}$$

The transmission power on subcarrier $n \in \mathcal{N}$ is given by $\sum_{\mathcal{G}\in\boldsymbol{\mathcal{G}}} \|\mathbf{w}_{\mathcal{G},n}\|_2^2$, and the total transmission power is given by $\sum_{n\in\mathcal{N}}\sum_{\mathcal{G}\in\boldsymbol{\mathcal{G}}} \|\mathbf{w}_{\mathcal{G},n}\|_2^2$. The total transmission power constraint is given by:

$$\sum_{n\in\mathcal{N}}\sum_{\mathcal{G}\in\boldsymbol{\mathcal{G}}} \|\mathbf{w}_{\mathcal{G},n}\|_2^2 \le P. \tag{6}$$

Here, P denotes the transmission power budget. Define $\boldsymbol{\mathcal{G}}^{(k)} \triangleq \{\mathcal{G} \in \boldsymbol{\mathcal{G}} | k \in \mathcal{G}\}, k \in \mathcal{K}$. Then, the received signal at user $k \in \mathcal{K}$ on subcarrier $n \in \mathcal{N}$, denoted by $y_{k,n} \in \mathbb{C}$, is given by:

$$\begin{aligned} y_{k,n} = \mathbf{h}_{k,n}^H \mathbf{x}_n + z_{k,n} &= \mathbf{h}_{k,n}^H \sum\nolimits_{\mathcal{G}\in\boldsymbol{\mathcal{G}}^{(k)}} \mathbf{w}_{\mathcal{G},n} s_{\mathcal{G},n} \\ &+ \mathbf{h}_{k,n}^H \sum\nolimits_{\mathcal{G}'\in\boldsymbol{\mathcal{G}}\setminus\boldsymbol{\mathcal{G}}^{(k)}} \mathbf{w}_{\mathcal{G}',n} s_{\mathcal{G}',n} + z_{k,n}, \\ & k \in \mathcal{K},\ n \in \mathcal{N}, \end{aligned} \tag{7}$$

where the last equality is due to (5), and $z_{k,n} \sim \mathcal{CN}(0,\sigma^2)$ is the additive white gaussian noise (AWGN). In (7), the first term represents the desired signal, and the second represents the interference. It is noteworthy that the main idea of rate splitting is to make the undesired messages partially decodable in order to reduce interference [9]. To exploit the full potential of the general rate splitting for general multicast, we consider joint decoding at each user.[3] That is, each user $k \in \mathcal{K}$ jointly decodes the desired transmission units $\widetilde{\mathcal{P}}_{\mathcal{G}}, \mathcal{G} \in \boldsymbol{\mathcal{G}}^{(k)}$. Thus, the achievable rate region of the transmission units is described by the following constraints:

$$\begin{aligned} &\sum_{\mathcal{G}\in\boldsymbol{\mathcal{X}}} \widetilde{R}_{\mathcal{G}} \\ &\quad \le B\sum_{n\in\mathcal{N}} \log_2\left(1 + \frac{\sum_{\mathcal{G}\in\boldsymbol{\mathcal{X}}} |\mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G},n}|^2}{\sigma^2 + \sum_{\mathcal{G}'\in\boldsymbol{\mathcal{G}}\setminus\boldsymbol{\mathcal{G}}^{(k)}} |\mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G}',n}|^2}\right), \\ &\qquad \boldsymbol{\mathcal{X}} \subseteq \boldsymbol{\mathcal{G}}^{(k)}, k \in \mathcal{K}, \end{aligned} \tag{8}$$

where $\widetilde{R}_{\mathcal{G}}$ is given by (4).

III. Optimization Problem Formulation

In this section, we would like to optimize the transmission beamforming vectors $\mathbf{w} \triangleq (\mathbf{w}_{\mathcal{G},n})_{\mathcal{G}\in\boldsymbol{\mathcal{G}},n\in\mathcal{N}}$ and rates of

[2] When $\mathcal{S} = \mathcal{K}$, $\boldsymbol{\mathcal{G}}_S = \{\mathcal{S}\}$ and the message unit $\mathcal{P}_{\mathcal{S}}$ will not be split. For ease of exposition, we let $\mathcal{P}_{\mathcal{S}} = \mathcal{P}_{\mathcal{S},\mathcal{S}}$ and $R_{\mathcal{S}} = R_{\mathcal{S},\mathcal{S}}$ for $\mathcal{S} = \mathcal{K}$.

[3] We can easily extend it to successive decoding as in [9].

Authorized licensed use limited to: Beijing Jiaotong University. Downloaded on October 13,2023 at 07:35:02 UTC from IEEE Xplore. Restrictions apply.

the sub-message units $\mathbf{R} \triangleq (R_{\mathcal{S},\mathcal{G}})_{\mathcal{S}\in\boldsymbol{\mathcal{S}},\mathcal{G}\in\boldsymbol{\mathcal{G}}}$ to maximize the weighted sum rate,[4] $\sum_{\mathcal{S}\in\boldsymbol{\mathcal{S}}} \alpha_{\mathcal{S}} R_{\mathcal{S}}$, where the coefficient $\alpha_{\mathcal{S}} \geq 0$ denotes the weight for message unit $\mathcal{P}_{\mathcal{S}}$, subject to the total transmission power constraint in (6) and the achievable rate constraints in (8). Therefore, we formulate the following optimization problem.

Problem 1 (Weighted Sum Rate Maximization):

$$\max_{\mathbf{w},\mathbf{R}\succeq 0} \quad \sum_{\mathcal{S}\in\boldsymbol{\mathcal{S}}} \alpha_{\mathcal{S}} R_{\mathcal{S}}$$
$$\text{s.t.} \quad (6),\ (8).$$

Remark 3 (Connection with Rate Splitting for Unicast and Multicast): When general multicast degrades to unicast, Problem 1 reduces to the weighted sum rate maximization problem for general rate splitting for unicast in [9]. When general multicast degrades to single-group multicast, Problem 1 reduces to the rate maximization problem for single-group multicast in [12]. Finally, when general multicast degrades to multi-group multicast, Problem 1 can be viewed as a generalization of the weighted sum rate maximization for multi-group multicast in [11].

Note that the objective function is linear, the constraint in (6) is convex, and the constraints in (8) are nonconvex. Thus, Problem 1 is nonconvex.[5]

IV. Solution

In this section, we propose an iterative algorithm to obtain a KKT point of Problem 1 using CCCP. First, we transform Problem 1 into the following equivalent problem by introducing auxiliary variables $\mathbf{e} \triangleq (e_{k,n,\boldsymbol{\mathcal{X}}})_{\boldsymbol{\mathcal{X}}\subseteq\boldsymbol{\mathcal{G}}^{(k)},k\in\mathcal{K},n\in\mathcal{N}}$ and $\mathbf{u} \triangleq (u_{k,n,\boldsymbol{\mathcal{X}}})_{\boldsymbol{\mathcal{X}}\subseteq\boldsymbol{\mathcal{G}}^{(k)},k\in\mathcal{K},n\in\mathcal{N}}$ and extra constraints:

$$\sum_{\mathcal{G}'\in\boldsymbol{\mathcal{G}}\setminus\boldsymbol{\mathcal{G}}^{(k)}} |\mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G}',n}|^2 + \sigma^2 - \frac{\sum_{\mathcal{G}\in\boldsymbol{\mathcal{X}}} |\mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G},n}|^2 + \sum_{\mathcal{G}'\in\boldsymbol{\mathcal{G}}\setminus\boldsymbol{\mathcal{G}}^{(k)}} |\mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G}',n}|^2 + \sigma^2}{u_{k,n,\boldsymbol{\mathcal{X}}}(\mathbf{h})} \leq 0,$$
$$\boldsymbol{\mathcal{X}} \subseteq \boldsymbol{\mathcal{G}}^{(k)}, k\in\mathcal{K}, n\in\mathcal{N}, \quad (9)$$

$$\sum_{\mathcal{G}\in\boldsymbol{\mathcal{X}}} \sum_{\mathcal{S}\in\boldsymbol{\mathcal{S}}_{\mathcal{G}}} R_{\mathcal{S},\mathcal{G}} = \sum_{n\in\mathcal{N}} e_{k,n,\boldsymbol{\mathcal{X}}},\ \boldsymbol{\mathcal{X}} \subseteq \boldsymbol{\mathcal{G}}^{(k)}, k\in\mathcal{K}, \quad (10)$$

$$2^{\frac{e_{k,n,\boldsymbol{\mathcal{X}}}}{B}} \leq u_{k,n,\boldsymbol{\mathcal{X}}},\ \boldsymbol{\mathcal{X}} \subseteq \boldsymbol{\mathcal{G}}^{(k)}, k\in\mathcal{K}, n\in\mathcal{N}. \quad (11)$$

Problem 2 (Equivalent Problem of Problem 1):

$$\max_{\mathbf{w},\mathbf{R}\succeq 0,\mathbf{e},\mathbf{u}} \quad \sum_{\mathcal{S}\in\boldsymbol{\mathcal{S}}} \alpha_{\mathcal{S}} \sum_{\mathcal{G}\in\boldsymbol{\mathcal{G}}_{\mathcal{S}}} R_{\mathcal{S},\mathcal{G}}$$
$$\text{s.t.} \quad (6),\ (9),\ (10),\ (11).$$

Let $(\mathbf{w}^\star, \mathbf{R}^\star, \mathbf{e}^\star, \mathbf{u}^\star)$ denote an optimal solution of Problem 2.

Theorem 1 (Equivalence Between Problem 1 and Problem 2): $(\mathbf{w}^\star, \mathbf{R}^\star, \mathbf{e}^\star, \mathbf{u}^\star)$ satisfies $2^{\frac{e^\star_{k,n,\boldsymbol{\mathcal{X}}}}{B}} = u^\star_{k,n,\boldsymbol{\mathcal{X}}},\ \boldsymbol{\mathcal{X}} \subseteq \boldsymbol{\mathcal{G}}^{(k)}, k\in\mathcal{K}, n\in\mathcal{N}$. Furthermore, Problem 1 and Problem 2 are equivalent.

Proof 1: First, by introducing auxiliary variables $\mathbf{e}$ and $\mathbf{u}$ and extra constraints:

$$2^{\frac{e_{k,n,\boldsymbol{\mathcal{X}}}}{B}} = u_{k,n,\boldsymbol{\mathcal{X}}},\ \boldsymbol{\mathcal{X}} \subseteq \boldsymbol{\mathcal{G}}^{(k)}, k\in\mathcal{K}, n\in\mathcal{N}, \quad (12)$$

we can equivalently transform Problem 1 into the following problem:

$$\max_{\mathbf{w},\mathbf{R}\succeq 0,\mathbf{e},\mathbf{u}} \quad \sum_{\mathcal{S}\in\boldsymbol{\mathcal{S}}} \alpha_{\mathcal{S}} \sum_{\mathcal{G}\in\boldsymbol{\mathcal{G}}_{\mathcal{S}}} R_{\mathcal{S},\mathcal{G}}$$
$$\text{s.t.} \quad (6),\ (9),\ (10),\ (12).$$

Let $(\mathbf{w}^\ddagger, \mathbf{R}^\ddagger, \mathbf{e}^\ddagger(\mathbf{h}), \mathbf{u}^\ddagger)$ denote an optimal solution. It is obvious that $(\mathbf{w}^\ddagger, \mathbf{R}^\ddagger)$ is an optimal solution of Problem 1. Next, we transform the above problem to Problem 2 by relaxing the constrains in (12) to the constraints in (11). By contradiction and the monotonicity of the objective function with respect to (w.r.t.) $\mathbf{R}$ in Problem 2, we can show that the constraints in (11) are active at the optimal solution. Thus, $(\mathbf{w}^\ddagger, \mathbf{R}^\ddagger, \mathbf{e}^\ddagger, \mathbf{u}^\ddagger)$ is an optimal solution of Problem 2. Therefore, we can show Theorem 1. ■

Based on Theorem 1, solving Problem 1 is equivalent to solving Problem 2. Problem 2 is a difference of convex functions (DC) programming (one type of nonconvex problems) and a KKT point can be obtained by CCCP [9].[6] The main idea of CCCP is to solve a sequence of successively refined approximate convex problems, each of which is obtained by linearizing the concave part and preserving the remaining convex part in the DC problem. Specifically, at the i-th iteration, the approximate convex problem of Problem 2 is given as follows. Let $(\mathbf{w}^{(i)}, \mathbf{R}^{(i)}, \mathbf{e}^{(i)}, \mathbf{u}^{(i)})$ denote an optimal solution of the following problem.

Problem 3 (Approximation of Problem 2 at Iteration i):

$$\max_{\mathbf{w},\mathbf{R}\succeq 0,\mathbf{e},\mathbf{u}} \quad \sum_{\mathcal{S}\in\boldsymbol{\mathcal{S}}} \alpha_{\mathcal{S}} \sum_{\mathcal{G}\in\boldsymbol{\mathcal{G}}_{\mathcal{S}}} R_{\mathcal{S},\mathcal{G}}$$
$$\text{s.t.} \quad (6),\ (10),\ (11),$$
$$L_{k,n,\boldsymbol{\mathcal{X}}}(\mathbf{w}_n, u_{k,n,\boldsymbol{\mathcal{X}}}; \mathbf{w}_n^{(i-1)}, u_{k,n,\boldsymbol{\mathcal{X}}}^{(i-1)}) \leq 0,$$
$$\boldsymbol{\mathcal{X}} \subseteq \boldsymbol{\mathcal{G}}^{(k)}, k\in\mathcal{K}, n\in\mathcal{N}, \quad (14)$$

where $\mathbf{w}_n^{(i-1)} \triangleq (\mathbf{w}_{\mathcal{G},n}^{(i-1)})_{\mathcal{G}\in\boldsymbol{\mathcal{X}}\cup(\boldsymbol{\mathcal{G}}\setminus\boldsymbol{\mathcal{G}}^{(k)})}$, $\mathbf{w}_n \triangleq (\mathbf{w}_{\mathcal{G},n})_{\mathcal{G}\in\boldsymbol{\mathcal{X}}\cup(\boldsymbol{\mathcal{G}}\setminus\boldsymbol{\mathcal{G}}^{(k)})}$, and $L_{k,n,\boldsymbol{\mathcal{X}}}(\mathbf{w}_n, u_{k,n,\boldsymbol{\mathcal{X}}}; \mathbf{w}_n^{(i-1)}, u_{k,n,\boldsymbol{\mathcal{X}}}^{(i-1)})$ is given by (13), as shown at the top of next page.

Problem 3 is convex and can be solved efficiently using standard convex optimization methods. Problem 3 has $MN|\boldsymbol{\mathcal{G}}| + \sum_{\mathcal{S}\in\boldsymbol{\mathcal{S}}} 2^{K-|\mathcal{S}|} + 2N\sum_{k\in\mathcal{K}}(2^{|\boldsymbol{\mathcal{G}}^{(k)}|}-1)$ variables and $1+(2N+1)\sum_{k\in\mathcal{K}}(2^{|\boldsymbol{\mathcal{G}}^{(k)}|}-1)$ constraints. Thus, when an interior point method is applied, the computational complexity for solving Problem 3 is $\mathcal{O}(K^{3.5}2^{1.75\times 2^K})$ as $K\to\infty$. The

[4]The proposed problem formulation and solution method can be readily extended to maximize the sum rate and worst-case rate as in [9].

[5]There are generally no effective methods for solving a nonconvex problem optimally. The goal of solving a nonconvex problem is usually to design an iterative algorithm to obtain a stationary point or a KKT point (which satisfies necessary conditions for optimality if strong duality holds).

[6]CCCP can exploit the partial convexity and usually converges faster to a KKT point than conventional gradient methods.

Authorized licensed use limited to: Beijing Jiaotong University. Downloaded on October 13,2023 at 07:35:02 UTC from IEEE Xplore. Restrictions apply.

$$L_{k,n,\boldsymbol{\mathcal{X}}}(\mathbf{w}_n, u_{k,n,\boldsymbol{\mathcal{X}}}; \mathbf{w}_n^{(i-1)}, u_{k,n,\boldsymbol{\mathcal{X}}}^{(i-1)}) \triangleq \sum_{\mathcal{G}' \in \boldsymbol{\mathcal{G}} \setminus \boldsymbol{\mathcal{G}}^{(k)}} |\mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G}',n}|^2 + \sigma^2 + \frac{\left(\sum\limits_{\mathcal{G} \in \boldsymbol{\mathcal{X}}} |\mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G},n}^{(i-1)}|^2 + \sum\limits_{\mathcal{G}' \in \boldsymbol{\mathcal{G}} \setminus \boldsymbol{\mathcal{G}}^{(k)}} |\mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G}',n}^{(i-1)}|^2 + \sigma^2 \right) u_{k,n,\boldsymbol{\mathcal{X}}}}{\left(u_{k,n,\boldsymbol{\mathcal{X}}}^{(i-1)} \right)^2}$$
$$- \frac{2\Re \left\{ \sum\limits_{\mathcal{G} \in \boldsymbol{\mathcal{X}}} \mathbf{w}_{\mathcal{G},n}^{(i-1)H} \mathbf{h}_{k,n} \mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G},n} + \sum\limits_{\mathcal{G}' \in \boldsymbol{\mathcal{G}} \setminus \boldsymbol{\mathcal{G}}^{(k)}} \mathbf{w}_{\mathcal{G}',n}^{(i-1)H} \mathbf{h}_{k,n} \mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G}',n} \right\} + 2\sigma^2}{u_{k,n,\boldsymbol{\mathcal{X}}}^{(i-1)}}, \ \boldsymbol{\mathcal{X}} \subseteq \boldsymbol{\mathcal{G}}^{(k)}, k \in \mathcal{K}, n \in \mathcal{N}. \quad (13)$$

details of CCCP for obtaining a KKT point of Problem 2 are summarized in Algorithm 1.[7] As the number of iterations of Algorithm 1 does not scale with the problem size [14], the computational complexity for Algorithm 1 is the same as that for solving Problem 3, i.e., $\mathcal{O}(K^{3.5}2^{1.75\times 2^K})$ as $K \to \infty$.

Theorem 2 (Convergence of Algorithm 1): As $i \to \infty$, $\left(\mathbf{w}^{(i)}, \mathbf{R}^{(i)}, \mathbf{e}^{(i)}, \mathbf{u}^{(i)}\right)$ obtained by Algorithm 1 converges to a KKT point of Problem 2 [13].

Proof 2: The constraints in (6), (10), (11) are convex, and the constraint function in (9) can be regarded as a difference between two convex functions, i.e., $\sum_{\mathcal{G}' \in \boldsymbol{\mathcal{G}} \setminus \boldsymbol{\mathcal{G}}^{(k)}} |\mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G}',n}|^2 + \sigma^2$ and $\frac{\sum_{\mathcal{G} \in \boldsymbol{\mathcal{X}}} |\mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G},n}|^2 + \sum_{\mathcal{G}' \in \boldsymbol{\mathcal{G}} \setminus \boldsymbol{\mathcal{G}}^{(k)}} |\mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G}',n}|^2 + \sigma^2}{u_{k,n,\boldsymbol{\mathcal{X}}}}$. Therefore, Problem 2 is a DC programming. Linearizing $\frac{\sum_{\mathcal{G} \in \boldsymbol{\mathcal{X}}} |\mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G},n}|^2 + \sum_{\mathcal{G}' \in \boldsymbol{\mathcal{G}} \setminus \boldsymbol{\mathcal{G}}^{(k)}} |\mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G}',n}|^2 + \sigma^2}{u_{k,n,\boldsymbol{\mathcal{X}}}}$ at $(\mathbf{w}_n^{(i-1)}, u_{k,n,\boldsymbol{\mathcal{X}}}^{(i-1)})$ and preserving $\sum_{\mathcal{G}' \in \boldsymbol{\mathcal{G}} \setminus \boldsymbol{\mathcal{G}}^{(k)}} |\mathbf{h}_{k,n}^H \mathbf{w}_{\mathcal{G}',n}|^2 + \sigma^2$ give $L_{k,n,\boldsymbol{\mathcal{X}}}(\mathbf{w}_n, u_{k,n,\boldsymbol{\mathcal{X}}}; \mathbf{w}_n^{(i-1)}, u_{k,n,\boldsymbol{\mathcal{X}}}^{(i-1)})$ in (13). Thus, Algorithm 1 implements CCCP. It has been validated in [13] that solving DC programming through CCCP always returns a KKT point. Therefore, we can show Theorem 2. ■

This paper focuses mainly on exploiting the full potential of general rate splitting for general multicast. The computational complexity of Algorithm 1 can be formidable with a large K. We can use successive decoding to reduce the computational complexity to $\mathcal{O}(K^{1.5}|\boldsymbol{\mathcal{S}}|^{1.5}2^{2K})$ as in [9].[8] We can also apply rate splitting with a reduced number of layers together with successive decoding to further reduce the computational complexity to $\mathcal{O}(K^{1.5}|\boldsymbol{\mathcal{S}}|^{1.5}(|\boldsymbol{\mathcal{G}}_{\text{lb}}|^2 + |\boldsymbol{\mathcal{G}}_{\text{lb}}||\boldsymbol{\mathcal{S}}|K + |\boldsymbol{\mathcal{S}}|^2K^2))$ as in [9], for some $\boldsymbol{\mathcal{G}}_{\text{lb}}$ satisfying $\boldsymbol{\mathcal{S}} \subseteq \boldsymbol{\mathcal{G}}_{\text{lb}} \subseteq \boldsymbol{\mathcal{G}}$. Note that $|\boldsymbol{\mathcal{G}}_{\text{lb}}|$ represents the reduced number of layers and satisfies $|\boldsymbol{\mathcal{S}}| \leq |\boldsymbol{\mathcal{G}}_{\text{lb}}| \leq 2^K - 1$.[9] Low-complexity optimization methods are beyond the scope of this paper.

V. Numerical Results

In this section, we numerically evaluate the proposed solution obtained by Algorithm 1, namely Prop-RS. We consider

[7]In practice, we can run Algorithm 1 multiple times with different feasible initial points to obtain multiple KKT points and choose the KKT point with the best objective value.

[8]Note that $|\boldsymbol{\mathcal{S}}|$ may not scale with K and how $|\boldsymbol{\mathcal{S}}|$ scales with K relies on the user requests. In the case of $|\boldsymbol{\mathcal{S}}| = \mathcal{O}(1)$, the reduced computational complexity is $\mathcal{O}(K^{1.5}2^{2K})$, as $K \to \infty$.

[9]In the case of $|\boldsymbol{\mathcal{S}}| = \mathcal{O}(1)$ and $|\boldsymbol{\mathcal{G}}_{\text{lb}}| = \mathcal{O}(1)$, the reduced computational complexity is $\mathcal{O}(N^{3.5}K^{3.5})$ as $K \to \infty$.

Algorithm 1 Obtaining a KKT Point of Problem 2

1: Initialization: Choose any feasible point of Problem 2 $(\mathbf{w}^{(0)}, \mathbf{R}^{(0)}, \mathbf{e}^{(0)}, \mathbf{u}^{(0)})$ and set $i = 0$.
2: **repeat**
3: Obtain an optimal solution $(\mathbf{w}^{(i)}, \mathbf{R}^{(i)}, \mathbf{e}^{(i)}, \mathbf{u}^{(i)})$ of Problem 3 with an interior point method.
4: Set $i = i + 1$.
5: **until** the convergence criterion $\|(\mathbf{w}^{(i)}, \mathbf{R}^{(i)}, \mathbf{e}^{(i)}, \mathbf{u}^{(i)}) - (\mathbf{w}^{(i-1)}, \mathbf{R}^{(i-1)}, \mathbf{e}^{(i-1)}, \mathbf{u}^{(i-1)})\|_2 \leq \epsilon$ is met.

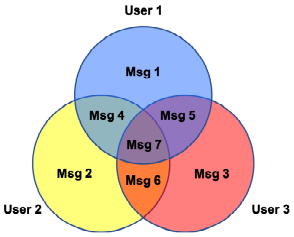

Fig. 3. General multicast setup in Section V.

three baseline schemes, namely 1L-RS, NoRS, and OFDMA. 1L-RS and NoRS extend 1-layer rate splitting [8] and SDMA [15], both for unicast, to general multicast. More specifically, 1L-RS and NoRS implement Algorithm 1 to obtain KKT points of Problem 1 with $\mathcal{G}_{\mathcal{S}} = \{\mathcal{S}, \mathcal{K}\}, \mathcal{S} \in \boldsymbol{\mathcal{S}}$ and with $\mathcal{G}_{\mathcal{S}} = \{\mathcal{S}\}, \mathcal{S} \in \boldsymbol{\mathcal{S}}$, respectively. OFDMA considers the maximum ratio transmission (MRT) on each subcarrier and optimizes the subcarrier and power allocation [1].

In the simulation, we set $K = 3$, $I = 7$, $I_1 = \{1, 4, 5, 7\}$, $I_2 = \{2, 4, 6, 7\}$, and $I_3 = \{3, 5, 6, 7\}$, as illustrated in Fig. 3. As a result, we have $\mathcal{P}_{\{1\}} = \{1\}$, $\mathcal{P}_{\{2\}} = \{2\}$, $\mathcal{P}_{\{3\}} = \{3\}$, $\mathcal{P}_{\{1,2\}} = \{4\}$, $\mathcal{P}_{\{1,3\}} = \{5\}$, $\mathcal{P}_{\{2,3\}} = \{6\}$, and $\mathcal{P}_{\{1,2,3\}} = \{7\}$. Additionally, we set $\alpha_{\mathcal{S}} = 1/7, \mathcal{S} \in \boldsymbol{\mathcal{S}}$, B = 30 kHz, N = 72, and $\sigma^2 = 10^{-9}$ W. We consider spatially correlated channel with the correlation following the one-ring scattering model as in [9]. When applying the one-ring scattering model, let G denote the number of user groups. We set the same angular spreads for the G groups and the same azimuth angle for the users in each group as in [9]. Note that G is related to the channel correlation among users. Specifically, the correlation increases as G decreases. When $G = 1$, all users belong to one group and have the same channel covariance matrix. When $G = 3$, all users are in different groups and have different channel covariance matrices. We

Authorized licensed use limited to: Beijing Jiaotong University. Downloaded on October 13,2023 at 07:35:02 UTC from IEEE Xplore. Restrictions apply.

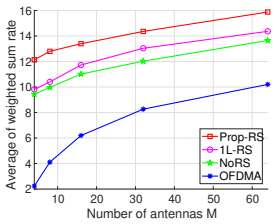

Fig. 4. Weighted sum rate versus M.

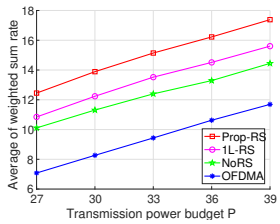

Fig. 5. Weighted sum rate versus P.

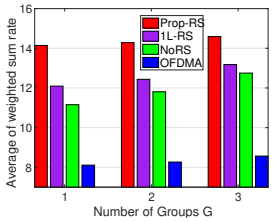

Fig. 6. Weighted sum rate versus G.

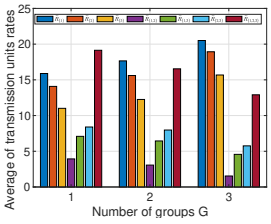

Fig. 7. Rates of transmission units of Prop-RS versus G.

generate 100 realizations of random system channel state, solve the weighted sum rate maximization problem for each realization, and evaluate the average of the weighted sum rate of each scheme over the 100 random realizations.

Fig. 4, Fig. 5, and Fig. 6 illustrate the average of the weighted sum rate versus the number of transmit antennas M, the total transmission power budget P, and the number of user groups G, respectively. From the three figures, we have the following observations. Firstly, the weighted sum average rate of each scheme increases with M, P, and G. Secondly, Prop-RS outperforms the baseline schemes. The gain of Prop-RS over 1L-RS is because the proposed solution unleashes the full potential of the flexibility of rate splitting. The gain of Prop-RS over NoRS arises because the cost for NoRS to suppress interference is high. In contrast, rate splitting together with joint decoding partially decodes interference and partially treats interference as noise. The gain of Prop-RS over OFDMA comes from an effective nonorthogonal transmission design. Additionally, Fig. 6 shows that the gains of Prop-RS over 1L-RS and NoRS increase as G decreases, demonstrating the advantage of flexibly dealing with interference in the presence of channel correlation among users. Fig. 7 shows the rates of the transmission units in the proposed solution versus the number of user groups G. We can see that $\widetilde{R}_{\{1\}}, \widetilde{R}_{\{2\}}$, and $\widetilde{R}_{\{3\}}$ increase with G, whereas $\widetilde{R}_{\{1,2\}}, \widetilde{R}_{\{1,3\}}, \widetilde{R}_{\{2,3\}}$, and $\widetilde{R}_{\{1,2,3\}}$ decrease with G. This is because as channel correlation among the users decreases, it is efficient to decode less interference and treat more interference as noise.

VI. Conclusion

While applications such as content delivery are responsible for a large and increasing fraction of Internet traffic, general multicast communication will play a central role for future 6G and beyond networks. This paper investigated the optimization of general rate splitting for general multicast. We adopted linear beamforming at the BS and joint decoding at each user. We maximized the weighted sum rate under the achievable rate region constraints and power constraint. We proposed an iterative algorithm to obtain a KKT point. The proposed optimization framework generalizes the existing ones for rate splitting for unicast, single-group multicast, and multi-group multicast. Numerical results demonstrate notable gains of the proposed solution over existing schemes and reveal the impact of channel correlation among users on the performance of general rate splitting for general multicast. There are still some key aspects that we leave for future investigations. One direction is to go beyond linear approaches and investigate nonlinear precoders such as binning. Another interesting perspective is general multicast with partial channel state information at the transmitter side.

References

[1] C. Guo, L. Zhao, Y. Cui, Z. Liu, and D. Ng "Power-efficient transmission of multi-quality tiled 360 VR video in MIMO-OFDMA systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 5408-5422, Aug. 2021.

[2] W. Xu, Y. Cui, and Z. Liu, "Optimal multi-view video transmission in multiuser wireless networks by exploiting natural and view synthesisenabled multicast opportunities," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1494-1507, Mar. 2020.

[3] Y. Cui, M. Médard, E. Yeh, D. Leith, F. Lai, and K. R. Duffy, "A linear network code construction for general integer connections based on the constraint satisfaction problem," *IEEE/ACM Trans. Netw.*, vol. 25, no. 6, pp. 3441-3454, Dec. 2017.

[4] H. P. Romero and M. K. Varanasi, "Rate splitting, superposition coding and binning for groupcasting over the broadcast channel: A general framework," *arXiv preprint arXiv:2011.04745*, Nov. 2020.

[5] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49-60, Jan. 1981.

[6] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, "Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 315-328, Jan. 2013.

[7] J. Park, J. Choi, N. Lee, W. Shin, and H. V. Poor, "Rate-splitting multiple access for downlink MIMO: A generalized power iteration approach," *arXiv preprint arXiv:2108.06844*, Aug. 2021.

[8] H. Joudeh and B. Clerckx, "Robust transmission in downlink multiuser MISO systems: A rate-splitting approach," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6227–6242, Dec. 2016.

[9] Z. Li, C. Ye, Y. Cui, S. Yang, and S. Shamai, "Rate splitting for multi-antenna downlink: Precoder design and practical implementation," *IEEE J. Select. Areas Commun.*, vol. 38, no. 8, pp. 1910–1924, Jun. 2020.

[10] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission: spectral and energy efficiency analysis," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8754-8770, Dec. 2019.

[11] H. Chen, D. Mi, B. Clerckx, Z. Chu, J. Shi, and P. Xiao, "Joint power and subcarrier allocation optimization for multigroup multicast systems with rate splitting," *IEEE Trans on Veh. Technol.*, vol. 69, no. 2, pp. 2306-2310, Feb. 2020.

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

[13] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[14] F. Facchinei, V. Kungurtsev, L. Lampariello, and G. Scutari, "Ghost penalties in nonconvex constrained optimization: Diminishing stepsizes and iteration complexity," *Math. Oper. Res.*, vol. 46, no. 2, pp. 595-627, Feb. 2021.

[15] W. Choi, A. Forenza, J. G. Andrews, and R. W. Heath, "Opportunistic space-division multiple access with beam selection," *IEEE Trans. Commun.*, vol. 55, no. 12, pp. 2371-2380, Dec. 2007.

Authorized licensed use limited to: Beijing Jiaotong University. Downloaded on October 13,2023 at 07:35:02 UTC from IEEE Xplore. Restrictions apply.