

A Rate-Splitting Strategy for Max-Min Fair Multigroup Multicasting

Hamdi Joudeh* and Bruno Clerckx*†

* Department of Electrical and Electronic Engineering, Imperial College London, United Kingdom

† School of Electrical Engineering, Korea University, Seoul, Korea

Email: {hamdi.joudeh10, b.clerckx}@imperial.ac.uk

Abstract—We consider the problem of transmit beamforming to multiple cochannel multicast groups. The conventional approach is to beamform a designated data stream to each group, while treating potential inter-group interference as noise at the receivers. In overloaded systems where the number of transmit antennas is insufficient to perform interference nulling, we show that inter-group interference dominates at high SNRs, leading to a saturating max-min fair performance. We propose a rather unconventional approach to cope with this issue based on the concept of Rate-Splitting (RS). In particular, part of the interference is broadcasted to all groups such that it is decoded and canceled before the designated beams are decoded. We show that the RS strategy achieves significant performance gains over the conventional multigroup multicast beamforming strategy.

Index Terms—Broadcasting, multicasting, downlink beamforming, degrees of freedom, WMMSE approach.

I. INTRODUCTION

Since the work of *Sidiropoulos et al.* [1], beamforming for physical-layer multicasting has received considerable research attention. In the most basic setup, the Base Station (BS) transmits a common data stream to all receivers. This was later generalized to multiple cochannel multicast groups, also known as multigroup multicasting [2]. The main problems considered in the multicasting literature are those of classical multiuser beamforming, namely the Quality of Service (QoS) constrained power minimization problem and the power constrained Max-Min Fair (MMF) problem. Such problems were shown to be NP-hard, and the solutions advised in [1], [2] are based on Semidefinite Relaxation (SDR) and Gaussian randomization techniques. Alternative solutions based on convex approximation methods were later proposed, exhibiting marginally improved performances under certain setups, and more importantly, lower complexities [3], [4]. The multigroup multicasting problem was also extended to incorporate per-antenna power constraints [5] and large-scale arrays [6]. In addition to the QoS and MMF problems, the sum-rate maximization problem was considered in [7].

The common transmission strategy adopted in multigroup multicasting is based on extending the multiuser beamforming paradigm, i.e. each message is first encoded into an independent data stream then transmitted through linear precoding (or beamforming). However, the multicast nature of each

stream results in different, and more difficult, design problems compared to their multiuser counterparts. In the beamforming strategy, each receiver decodes its desired stream while treating all interfering streams as noise. Hence, inter-group interference is inevitable under an insufficient number of BS antennas. Although rarely highlighted or treated in the multigroup multicasting literature, such interference can be highly detrimental.

We propose a beamforming strategy based on the concept of Rate-Splitting (RS), where the message intended to each group is split into a common part and a designated part. All common parts are packed into one super common message, broadcasted to all users in the system. Designated parts on the other hand are transmitted in the conventional beamforming manner. While the concept of RS is not particularly new (it appears in the interference channel literature), it has only been applied recently to multiuser beamforming, where it was shown to enhance the performance under residual interference arising from imperfect Channel State Information (CSI) at the BS [8], [9]. We show that RS brings significant performance gains to multigroup multicasting, particularly in inter-group interference limited scenarios. While the focus is on the MMF problem in this paper, RS can be extended to the QoS problem.

The rest of the paper is organized as follows. Section II presents the system model. The limitations of the conventional transmission strategy are analysed in Section III. In Section IV, the RS strategy is introduced and the performance gains over the conventional strategy are derived. The RS precoders are optimized using a Weighted Minimum Mean Square Error (WMMSE) algorithm in Section V. Simulation results are presented in Section VI, and Section VII concludes the paper.

II. SYSTEM MODEL

Consider a transmitter equipped with N antennas communicating with K single-antenna receivers grouped into the M multicast groups $\{\mathcal{G}_1, \dots, \mathcal{G}_M\}$, where $\mathcal{K} \triangleq \{1, \dots, K\}$ and $\mathcal{M} \triangleq \{1, \dots, M\}$. We assume that $\bigcup_{m \in \mathcal{M}} \mathcal{G}_m = \mathcal{K}$, and $\mathcal{G}_m \cap \mathcal{G}_j = \emptyset$, for all $m, j \in \mathcal{M}$ and $m \neq j$. Let $\mathbf{x} \in \mathbb{C}^N$ denote the signal vector transmitted by the BS in a given channel use, which is subject to an average power constraint $\mathbb{E}\{\mathbf{x}^H \mathbf{x}\} \leq P$. Denoting the corresponding signal received by the k th user as y_k , the input-output relationship writes as $y_k = \mathbf{h}_k^H \mathbf{x} + n_k$, where $\mathbf{h}_k \in \mathbb{C}^N$ is the narrow-band channel vector from the BS to the k th user, and $n_k \in \mathcal{CN}(0, \sigma_{n,k}^2)$ is the Additive White Gaussian Noise (AWGN) at the receiver.

This work has been partially supported by the UK EPSRC under grant number EP/N015312/1.

We assume, without loss of generality, that $\sigma_{n,1}^2, \dots, \sigma_{n,K}^2 = \sigma_n^2$, from which the transmit SNR is given by P/σ_n^2 . Moreover, the transmitter perfectly knows all K channel vectors, and each receiver knows its own channel vector.

In multigroup multicast transmission, the BS communicates the messages W_1, \dots, W_M to $\mathcal{G}_1, \dots, \mathcal{G}_M$ respectively. Consider a conventional linear precoding (beamforming) transmission model. Messages are first encoded into independent data streams, where the vector of coded data symbols in a given channel use writes as $\mathbf{s}_p \triangleq [s_1, \dots, s_M]^T \in \mathbb{C}^M$. We assume that $\mathbb{E}\{\mathbf{s}_p \mathbf{s}_p^H\} = \mathbf{I}$, where power allocation is considered part of the beamforming. Data streams are then mapped to the transmit antennas through a linear precoding matrix $\mathbf{P}_p \triangleq [\mathbf{p}_1, \dots, \mathbf{p}_M]$, where $\mathbf{p}_m \in \mathbb{C}^N$ is the m th group's precoding vector. The resulting transmit signal is

$$\mathbf{x} = \sum_{m=1}^M \mathbf{p}_m s_m \quad (1)$$

where the power constraint reduces to $\sum_{m=1}^M \|\mathbf{p}_m\|^2 \leq P$.

The k th user's average receive power (over multiple channel uses in which the channel is fixed) writes as

$$T_k = \underbrace{|\mathbf{h}_k^H \mathbf{p}_{\mu(k)}|^2}_{S_k} + \underbrace{\sum_{m \neq \mu(k)} |\mathbf{h}_k^H \mathbf{p}_m|^2 + \sigma_n^2}_{I_k}. \quad (2)$$

where $\mu : \mathcal{K} \mapsto \mathcal{M}$ maps a user-index to the corresponding group-index, i.e. $\mu(k) = m$ such that $k \in \mathcal{G}_m$. In the following, $\mu(k)$ is referred to as μ for brevity where the argument of the function is clear from the context. In (2), S_k and I_k denote the desired receive power and the interference plus noise power, respectively. Hence, the Signal to Interference plus Noise Ratio (SINR) experienced by the k th user is defined as $\gamma_k \triangleq S_k I_k^{-1}$. Under Gaussian signalling, the k th achievable user-rate is given by $R_k = \log_2(1 + \gamma_k)$. In multigroup multicasting, users belonging to the same group decode the same data stream. Therefore, to guarantee that all users in the m th group are able to recover W_m successfully, the corresponding code-rate should not exceed the group-rate defined as $r_m \triangleq \min_{i \in \mathcal{G}_m} R_i$.

III. MAX-MIN FAIRNESS AND INTER-GROUP INTERFERENCE

In the light of the conventional multi-stream beamforming model, the MMF problem is formulated as

$$\mathcal{R}(P) : \begin{cases} \max_{\mathbf{P}_p} & \min_{m \in \mathcal{M}} \min_{i \in \mathcal{G}_m} R_i \\ \text{s.t.} & \sum_{m=1}^M \|\mathbf{p}_m\|^2 \leq P \end{cases} \quad (3)$$

where the inner minimization in (3) accounts for the multi-cast nature within each group, while the outer minimization accounts for the fairness across groups. It is common practice to formulate the above problem in terms of the SINRs [2], [4]–[6]. Since each group receives a single stream, and due to the Rate-SINR monotonic relationship, the two formulations

are equivalent. The rate formulation is preferred in this work in order to compare the performance to the RS scheme.

A. Inter-Group Interference and Degrees of Freedom

An optimum MMF design achieves balanced group rates, requiring a simultaneous increase in powers allocated to all streams as P increases. In scenarios where the number of transmit antennas is insufficient to place each beam in the null space of all its unintended groups, inter-group interference is expected to limit the MMF performance. To characterize this, we resort to high SNR analysis through the Degrees of Freedom (DoF). This regime is of particular interest as the effect of noise can be neglected, and inter-group interference is the main limiting factor. The DoF can be roughly interpreted as the number of interference-free streams that can be simultaneously communicated in a single channel use. To facilitate the definition of the DoF, we first define a precoding scheme $\{\mathbf{P}_p(P)\}_P$ as a family of feasible precoders with one precoding matrix for each power level. The corresponding achievable user-rates write as $\{R_1(P), \dots, R_K(P)\}_P$, and the k th user-DoF is defined as $D_k \triangleq \lim_{P \rightarrow \infty} \frac{R_k(P)}{\log_2(P)}$. It follows that the m th group-DoF, denoted by d_m , satisfies $0 \leq d_m \leq D_i$ for all $i \in \mathcal{G}_m$. The corresponding symmetric-DoF is given by $d = \min_{m \in \mathcal{M}} d_m$.

For a given setup, the optimum MMF precoding scheme is denoted by $\{\mathbf{P}_p^*(P)\}_P$. The corresponding MMF-DoF is given by $d^* = \lim_{P \rightarrow \infty} \frac{\mathcal{R}(P)}{\log_2(P)}$, which is the maximum symmetric-DoF. Since each user is equipped with a single antenna, then $D_1, \dots, D_K \leq 1$, and $d \leq 1$ for any precoding scheme. Hence, when $d = 1$ is achievable, it is also optimum. It should be noted that although a rate-optimal precoder is also optimum in a DoF sense, the converse is usually untrue.

B. MMF-DoF of Multi-Stream Beamforming

In the DoF analysis, we make the following assumptions.

Assumption 1. The channel vectors $\mathbf{h}_1, \dots, \mathbf{h}_K$ are independently drawn from a set of continuous distribution functions. Hence, for any $N \times K_{\text{sub}}$ matrix in which the K_{sub} column vectors constitute any subset of the K channel vectors, it holds with probability one that the rank is $\min\{N, K_{\text{sub}}\}$.

Assumption 2. We assume equal size groups. i.e. $|\mathcal{G}_1|, \dots, |\mathcal{G}_M| = G$, where $G = K/M$ is a positive integer.

Next, the MMF-DoF of the conventional multi-stream transmission scheme is characterized.

Proposition 1. Under Assumptions 1 and 2, the optimum MMF-DoF achieved by solving (3) is given by

$$\lim_{P \rightarrow \infty} \frac{\mathcal{R}(P)}{\log_2(P)} = \begin{cases} 1, & N \geq N_{\min} \\ 0, & N < N_{\min} \end{cases} \quad (4)$$

where $N_{\min} = 1 + K - G$.

To show this, let us define \mathbf{H}_m as the matrix with columns constituting channel vectors of all users in \mathcal{G}_m , and $\bar{\mathbf{H}}_m = [\mathbf{H}_1, \dots, \mathbf{H}_{m-1}, \mathbf{H}_{m+1}, \dots, \mathbf{H}_M]$ as the complementary set

of channel vectors. By Assumptions 1 and 2, $\text{null}(\bar{\mathbf{H}}_m)$ has a dimension of $\max\{N + G - K, 0\}$ for all $m \in \mathcal{M}$. Hence, $N \geq N_{\min}$ is sufficient to place each beamforming vector in the null space of all groups it is not intended to, i.e. $\mathbf{p}_m \in \text{null}(\bar{\mathbf{H}}_m)$ for all $m \in \mathcal{M}$. Each group sees no inter-group interference, and a DoF of 1 per group is achievable. Such DoF is optimum as it cannot be surpassed. On the contrary, when $N < N_{\min}$, this is not possible, and inter-group interference limits the MMF-DoF to 0 as shown in the Appendix. We refer to this case as an overloaded system.

Finally, we conclude this section by highlighting the impact of a collapsing DoF on the rate performance. When $d = 0$, the MMF rate stops growing as SNR grows large, reaching a saturated performance¹. Although the DoF analysis is carried out as SNR goes to infinity, its results are highly visible in finite SNR regimes as we see in the simulation results.

IV. RATE-SPLITTING FOR MULTIGROUP MULTICASTING

The saturating performance can be avoided by single-stream multigroup transmission. In particular, the M messages are packed into one super message, encoded into a single data stream. This is broadcasted such that it is decoded by all groups, hence retrieving their corresponding messages. Since this interference-free transmission achieves a total DoF of 1, each group is guaranteed a non-saturating performance with a DoF of $1/M$. However, relying solely on this strategy jeopardizes partial gains potentially achieved by multi-stream beamforming. A simple example is the low-SNR regime, where interference is overwhelmed by noise, and beamforming each message to its corresponding group is a preferred strategy. Hence, we introduce the following unifying strategy.

A. The Rate-Splitting Strategy

Each group-message is split into a common part and a group-designated part, e.g. $W_m = \{W_{m0}, W_{m1}\}$, with W_{m0} and W_{m1} being the common and designated parts respectively. All common parts are packed into one super common message $W_c \triangleq \{W_{10}, \dots, W_{M0}\}$, encoded into the stream s_c , and then precoded using $\mathbf{p}_c \in \mathbb{C}^N$. On the other hand, the designated messages are encoded into s_1, \dots, s_M and precoded in the conventional multi-stream manner described in Section II. The transmit signal writes as a superposition of the common stream and the designated streams such that

$$\mathbf{x} = \mathbf{p}_c s_c + \sum_{m=1}^M \mathbf{p}_m s_m. \quad (5)$$

The power constraint writes as $\|\mathbf{p}_c\|^2 + \sum_{m=1}^M \|\mathbf{p}_m\|^2 \leq P$. The common stream can be interpreted as the part of the interference that is decoded (hence eliminated) by all groups, while interference from designated streams is treated as noise.

The k th user's average received power now writes as

$$T_{c,k} = |\mathbf{h}_k^H \mathbf{p}_c|^2 + T_k \quad (6)$$

¹To be more precise, this corresponds to a rate scaling as $o(\log_2(P))$, which either stops growing or grows extremely slow with P compared to the interference free scenario, reaching a flat or almost-flat performance.

where $S_{c,k} = |\mathbf{h}_k^H \mathbf{p}_c|^2$ denotes the common stream's receive power and $I_{c,k} = T_k$ is the interference plus noise power experienced by the common stream. By treating all designated streams as noise, the SINR of the common stream at the k th receiver is given by $\gamma_{c,k} \triangleq S_{c,k} I_{c,k}^{-1}$. Hence, transmitting W_c at a rate of $R_{c,k} = \log_2(1 + \gamma_{c,k})$ guarantees successful decoding by the k th receiver. To guarantee that W_c is successfully recovered by all receivers, the rate of the common data stream should not exceed the common-rate defined as $R_c = \min_{k \in \mathcal{K}} R_{c,k}$. After decoding the common stream, the receiver removes it from y_k using Successive Interference Cancellation (SIC). This is followed by decoding the designated stream in the presence of the remaining interference and noise, achieving the rate R_k defined in Section II.

The common rate writes as a sum of M portions: $R_c = \sum_{m=1}^M C_m$, where C_m is associated with W_{m0} . It follows that the m th group-rate is defined as $R_{g,m} \triangleq C_m + \min_{i \in \mathcal{G}_m} R_i$, consisting of a common-rate portion plus a designated-rate. It is easy to see that the RS group-rates reduce to the conventional group-rates defined in Section III when $|W_c| = 0$.

B. MMF With Rate-Splitting

The MMF problem is formulated in terms of RS as follows

$$\mathcal{R}_{\text{RS}}(P) : \begin{cases} \max_{\mathbf{c}, \mathbf{P}} & \min_{m \in \mathcal{M}} \left(C_m + \min_{i \in \mathcal{G}_m} R_i \right) \\ \text{s.t.} & R_{c,k} \geq \sum_{m=1}^M C_m, \forall k \in \mathcal{K} \\ & C_m \geq 0, \forall m \in \mathcal{M} \\ & \|\mathbf{p}_c\|^2 + \sum_{m=1}^M \|\mathbf{p}_m\|^2 \leq P \end{cases} \quad (7)$$

where $\mathbf{c} \triangleq [C_1, \dots, C_M]^T$ is the vector of common-rate portions. The first set of constraints in (7) accounts for the global multicast nature of the common stream and guarantees that it can be decoded by all users. The second set of constraints guarantees that no user is allocated a negative common-rate portion. Solving (7) yields the optimum precoding matrix, in addition to the splitting ratio for each group-message. Next, the DoF performance of the RS scheme is characterized.

Proposition 2. The MMF-DoF achieved by solving the RS problem in (7) is lower-bounded by

$$\lim_{P \rightarrow \infty} \frac{\mathcal{R}_{\text{RS}}(P)}{\log_2(P)} \geq \begin{cases} 1, & N \geq N_{\min} \\ \frac{1}{M}, & N < N_{\min} \end{cases} \quad (8)$$

where the lower-bound is tight for $N \geq N_{\min}$.

This follows directly from Proposition 1, and the fact that the single-stream multigroup solution described at the beginning of this section is feasible for problem (7).

V. PRECODER OPTIMIZATION

In the RS scheme, each group-rate writes as a sum of two rate components. Hence, the MMF solutions in [2], [4] do not apply here, as the performance metric of each user cannot be expressed as a single SINR. Alternatively, we resort to the WMMSE approach [10], [11], which is particularly effective

in dealing with problems incorporating non-convex coupled sum-rate expressions, including RS problems [8], [9].

A. Rate-WMMSE Relationship

We start by defining the MSEs. The k th user's estimate of s_c , denoted by $\widehat{s}_{c,k}$, is obtained by applying the equalizer $g_{c,k}$ to the receive signal such that $\widehat{s}_{c,k} = g_{c,k}y_k$. After removing the common stream using SIC, the equalizer g_k is applied to the remaining signal to obtain an estimate of \widehat{s}_k given by $\widehat{s}_k = g_k(y_k - \mathbf{h}_k^H \mathbf{p}_c s_c)$. The common and private MSEs at the output of the k th receiver, defined as $\varepsilon_{c,k} \triangleq \mathbb{E}\{|\widehat{s}_{c,k} - s_c|^2\}$ and $\varepsilon_k \triangleq \mathbb{E}\{|\widehat{s}_k - s_k|^2\}$ respectively, write as:

$$\varepsilon_{c,k} = |g_{c,k}|^2 T_{c,k} - 2\Re\{g_{c,k} \mathbf{h}_k^H \mathbf{p}_c\} + 1 \quad (9a)$$

$$\varepsilon_k = |g_k|^2 T_k - 2\Re\{g_k \mathbf{h}_k^H \mathbf{p}_\mu\} + 1. \quad (9b)$$

The MMSEs are defined as $\varepsilon_{c,k}^{\text{MMSE}} \triangleq \min_{g_{c,k}} \varepsilon_{c,k} = T_{c,k}^{-1} I_{c,k}$ and $\varepsilon_k^{\text{MMSE}} \triangleq \min_{g_k} \varepsilon_k = T_k^{-1} I_k$, where the corresponding optimum equalizers are the well-known MMSE weights written as $g_{c,k}^{\text{MMSE}} = \mathbf{p}_c^H \mathbf{h}_k T_{c,k}^{-1}$ and $g_k^{\text{MMSE}} = \mathbf{p}_\mu^H \mathbf{h}_k T_k^{-1}$. The MMSEs are related to the SINRs such that $\gamma_{c,k} = (1/\varepsilon_{c,k}^{\text{MMSE}}) - 1$ and $\gamma_k = (1/\varepsilon_k^{\text{MMSE}}) - 1$, from which the achievable rates write as $R_{c,k} = -\log_2(\varepsilon_{c,k}^{\text{MMSE}})$ and $R_k = -\log_2(\varepsilon_k^{\text{MMSE}})$.

Next we introduce the main building blocks of the solution, the augmented WMSEs defined for the k th user as:

$$\xi_{c,k} \triangleq u_{c,k} \varepsilon_{c,k} - \log_2(u_{c,k}) \quad \text{and} \quad \xi_k \triangleq u_k \varepsilon_k - \log_2(u_k) \quad (10)$$

where $u_{c,k}, u_k > 0$ are the corresponding weights. In the following, $\xi_{c,k}$ and ξ_k are referred to as the WMSEs for brevity. The Rate-WMMSE relationship is established by optimizing (10) over the equalizers and weights such that:

$$\xi_{c,k}^{\text{MMSE}} \triangleq \min_{u_{c,k}, g_{c,k}} \xi_{c,k} = 1 - R_{c,k} \quad (11a)$$

$$\xi_k^{\text{MMSE}} \triangleq \min_{u_k, g_k} \xi_k = 1 - R_k \quad (11b)$$

where the optimum equalizers are given by: $g_{c,k}^* = g_{c,k}^{\text{MMSE}}$ and $g_k^* = g_k^{\text{MMSE}}$, and the optimum weights are given by: $u_{c,k}^* = u_{c,k}^{\text{MMSE}} \triangleq (\varepsilon_{c,k}^{\text{MMSE}})^{-1}$ and $u_k^* = u_k^{\text{MMSE}} \triangleq (\varepsilon_k^{\text{MMSE}})^{-1}$, obtained by checking the first order optimality conditions. By closely examining each WMSE, it can be seen that it is convex in each variable while fixing the other two.

B. WMSE Reformulation and Algorithm

Motivated by the relationship in (11), an equivalent WMSE reformulation of problem (7) writes as

$$\widehat{\mathcal{R}}_{\text{RS}}(P) : \begin{cases} \max_{r_g, \mathbf{r}, \mathbf{c}, \mathbf{P}, \mathbf{g}, \mathbf{u}} & r_g \\ \text{s.t.} & C_m + r_m \geq r_g, \forall m \in \mathcal{M} \\ & 1 - \xi_i \geq r_m, \forall i \in \mathcal{G}_m, \forall m \in \mathcal{M} \\ & 1 - \xi_{c,k} \geq \sum_{m=1}^M C_m, \forall k \in \mathcal{K} \\ & C_m \geq 0, \forall m \in \mathcal{M} \\ & \|\mathbf{p}_c\|^2 + \sum_{m=1}^M \|\mathbf{p}_m\|^2 \leq P \end{cases} \quad (12)$$

where r_g and $\mathbf{r} \triangleq \{r_1, \dots, r_M\}$ are auxiliary variables, $\mathbf{u} \triangleq \{u_{c,k}, u_k \mid k \in \mathcal{K}\}$ is the set of weights, and $\mathbf{g} \triangleq \{g_{c,k}, g_k \mid k \in \mathcal{K}\}$ is the set of equalizers. The equivalence between (12) and (7) is established by observing that the WMSEs are decoupled in their equalizers and weights. Hence, optimum \mathbf{g} and \mathbf{u} are obtained by minimizing each WMSE separately as shown in (11), yielding the MMSE solution. The equivalence follows by substituting (11) into (12).

The WMSE problem in (12) is solved using an Alternating Optimization (AO) algorithm, which exploits its block-wise convexity. In a given iteration of the algorithm, \mathbf{g} and \mathbf{u} are firstly updated using the optimum MMSE solution of (11). Next, the set of precoders \mathbf{P} alongside all auxiliary variables in (12) are updated by solving $\widehat{\mathcal{R}}_{\text{RS}}^{\text{MMSE}}(P)$, formulated by fixing \mathbf{g} and \mathbf{u} in (12). This is a convex problem which can be efficiently solved using interior-point methods [12]. The steps of the AO procedure are summarized in Algorithm 1.

Algorithm 1 Alternating Optimization

- 1: Initialize: $n \leftarrow 0$, \mathbf{P} , $r_g^{(n)} \leftarrow 0$
 - 2: **repeat**
 - 3: $n \leftarrow n + 1$
 - 4: $(g_{c,k}, g_k) \leftarrow (g_{c,k}^{\text{MMSE}}, g_k^{\text{MMSE}}), \forall k \in \mathcal{K}$
 - 5: $(u_{c,k}, u_k) \leftarrow (u_{c,k}^{\text{MMSE}}, u_k^{\text{MMSE}}), \forall k \in \mathcal{K}$
 - 6: $(r_g^{(n)}, \mathbf{r}, \mathbf{c}, \mathbf{P}) \leftarrow \arg \widehat{\mathcal{R}}_{\text{RS}}^{\text{MMSE}}(P)$
 - 7: **until** $|r_g^{(n)} - r_g^{(n-1)}| < \epsilon$
-

Each iteration of Algorithm 1 increases the objective function, which is bounded above for a given power constraint, until convergence. The global optimality of the limit point cannot be guaranteed due to non-convexity. However, the stationarity (KKT optimality) of the solution can be argued based on the ideas in [13], avoided here due to space limitations.

VI. SIMULATION RESULTS

We consider i.i.d channels with entries drawn from $\mathcal{CN}(0, 1)$, and all results are averaged over 100 channel realizations. We compare the MMF rates for: 1) conventional beamforming (NoRS), 2) Single-Stream (SS) multigroup transmission described at the beginning of Section IV, and 3) the RS strategy. The results for NoRS and SS are obtained using the SDR method in [1], [2]. We plot the SDR upper-bounds (no randomization), hence presenting optimistic performances for NoRS and SS. On the other hand, the RS results are obtained by Algorithm 1, and represent the actual performances.

The MMF rates for a system with $N = 2$ transmit antennas and $M = 2$ groups with $G = 2$ users each are presented in Fig. 1. As predicted from the DoF result in Proposition 1, NoRS exhibits a saturating performance. Both SS and RS achieve non-saturating rates with DoFs of 1/2 and an improved rate performance for RS which comes from the designated beams. The gains of RS over NoRS are very pronounced. The results for a system with $N = 4$ transmit antennas and $M = 3$ groups with $G = 3$ users per group are shown in Fig. 2. The benefits of using designated beams over SS transmission at low SNRs are clearer in this scenario, as the performance

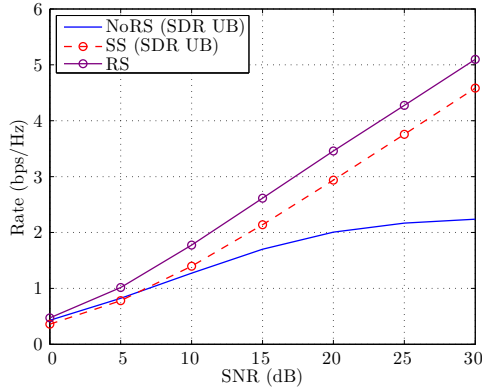


Fig. 1. $N = 2$ antennas, $K = 4$ users, $M = 2$ equal groups.

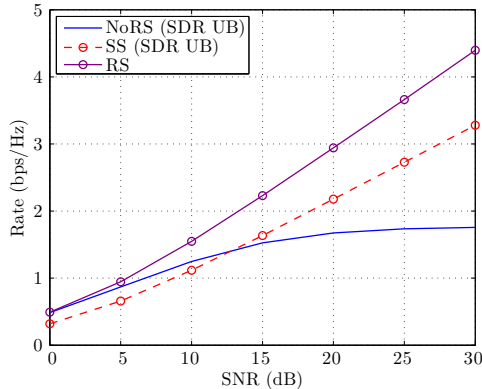


Fig. 2. $N = 4$ antennas, $K = 9$ users, $M = 3$ equal groups.

of the latter is constrained by the worst out of 9 users. While SS achieves a DoF of $1/3$, RS seems to surpass this DoF, which is evident from its slope at high SNR. This suggests that the trivial achievable lower-bound in Proposition 2 is in fact loose, and RS can achieve even higher MMF-DoF.

VII. CONCLUSION

In this paper, we proposed a RS multi-group multicast beamforming strategy. We showed through DoF analysis that the proposed RS strategy outperforms the conventional beamforming strategy in overloaded scenarios, i.e. when the number of transmit antennas is insufficient to cope with inter-group interference. An AO algorithm based on the WMMSE method was used to obtain the RS precoders. The effectiveness of the proposed algorithm and the significant gains associated with the RS strategy were demonstrated through simulations. Simulations also revealed that the trivial MMF-DoF lower-bound is in fact loose, which calls for a rigorous characterization of the optimum MMF-DoF achieved through RS.

APPENDIX

Proof of $d^ = 0$ for $N < N_{\min}$:* First, we write a precoding vector as $\mathbf{p}_m = \sqrt{q_m} \hat{\mathbf{p}}_m$, where q_m is the power and $\hat{\mathbf{p}}_m$ is the unit-norm beamforming direction. For a given precoding scheme characterized by one precoder for each power level, the m th power scales as $q_m = O(P^{a_m})$ with $a_m \leq 1$, further assumed to be non-negative as MMF necessitates non-vanishing powers allocated to all groups. Let

$\mathcal{I}_m \subset \mathcal{M}$ be the index set of groups that interfere with the m th group, depending on the precoder design. From the DoF definitions in Section III-A, it can be shown that

$$d_m \leq (a_m - \max_{j \in \mathcal{I}_m} a_j)^+. \quad (13)$$

The $(\cdot)^+$ can be omitted as when it is active, the MMF-DoF is limited to zero which is also achieved when all a_m are equal.

It is sufficient to show that the MMF-DoF is upper-bounded by 0 for $N = N_{\min} - 1 = (M-1)G$, as decreasing the number of antennas does not increase the DoF. For this case, \mathbf{p}_m can be placed in the null space of at most $M-2$ groups, i.e. each beam interferes with at least one group. It follows that $\bigcup_{m \in \mathcal{M}} \mathcal{I}_m = \mathcal{M}$. We assume that each beam interferes with exactly one group, as the contrary does not increase the DoF. It follows that at least two groups see non-zero interference. Let m_1 be the index of the group receiving the dominant interference, i.e. $\max\{a_m\}_{m \in \mathcal{M}} \in \mathcal{I}_{m_1}$, and m_2 be the index of the group receiving interference from m_1 , i.e. $a_{m_1} \in \mathcal{I}_{m_2}$. It follows from (13) that $d_{m_1} \leq a_{m_1} - a_{m_2}$ and $d_{m_2} \leq a_{m_2} - a_{m_1}$. Since the symmetric-DoF is upper-bounded by the average of any number of group DoFs, we write $d \leq \frac{d_{m_1} + d_{m_2}}{2} \leq 0$, which holds for any possible precoder design. ■

REFERENCES

- [1] N. Sidiropoulos, T. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.
- [2] E. Karipidis, N. Sidiropoulos, and Z.-Q. Luo, "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, Mar. 2008.
- [3] N. Bornhorst and M. Pesavento, "An iterative convex approximation approach for transmit beamforming in multi-group multicasting," in *Proc. IEEE SPAWC*, Jun. 2011, pp. 426–430.
- [4] A. Schach and M. Pesavento, "Max-min fair transmit beamforming for multi-group multicasting," in *Proc. Int. ITG WSA*, Mar. 2012, pp. 115–118.
- [5] 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, Oct. 2014.
- [6] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Multicast multi-group beamforming for per-antenna power constrained large-scale arrays," in *Proc. IEEE SPAWC*, Jun. 2015, pp. 271–275.
- [7] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Sum rate maximizing multigroup multicast beamforming under per-antenna power constraints," in *Proc. IEEE GLOBECOM*, Dec 2014, pp. 3354–3359.
- [8] H. Joudeh and B. Clerckx, "Robust transmission in downlink multiuser MISO systems: A rate-splitting approach," *arXiv preprint arXiv:1602.04345*, 2016.
- [9] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *arXiv preprint arXiv:1602.09028*, 2016.
- [10] S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.
- [11] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sept 2011.
- [12] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.
- [13] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.