

Precoder Design For Multi-group Multicasting with a Common Message

Ahmet Zahid Yalcin and Melda Yuksel

Abstract

This paper considers precoding for multi-group multicasting with a common message. The multiple antenna base station communicates with K clusters, each with L users. There is a common message destined to all users and a private multicast message for each cluster. We study the weighted sum rate (WSR) maximization problem for two different schemes: (i) the base station transmits the superposition of common and multicast messages, (ii) the base station concatenates the multicast message vector with the common message. We also formulate a second problem, weighted minimum mean square error (WMMSE) minimization, and prove that WSR maximization and WMMSE minimization are equivalent at the optimal solution. Inspired by the WMMSE problem, we suggest a suboptimal algorithm, based on alternating optimization. We apply this algorithm to the two transmission schemes, and understand that there is a fundamental difference between the two. We compare the results with maximal ratio transmission (MRT), and zero-forcing (ZF) precoding, and investigate the effects of the number of base station antennas, the number of groups and the number of users in a group. Finally, we study imperfect successive interference cancellation (SIC) at the receivers and show that the first transmission scheme is more robust.

Index Terms

Common and private data transmission, multicast transmission, multiple input multiple output, physical layer precoding, superposition coding.

I. INTRODUCTION

With the introduction of the fifth generation (5G) mobile communication systems, mobile communications will diffuse into all areas of daily life. While the fourth generation (4G) is mainly about Internet access, video calls, and cloud computing, 5G will be about smart-health, smart-cities, 4K video, factory automation, self-driving cars, real-time cloud services and more.

All these new and different applications have different requirements, such as high data rates, ultra-low latency and ultra-high reliability, or ability to handle massive radio access.

The multiple-input multiple-output (MIMO) technology is one of the main enablers for high data rates required for enhanced mobile broadband. The capacity of MIMO channels are found in [1], [2]. These papers reveal that the degrees of freedom in a MIMO channel is limited by the minimum of the number of transmit and receive antennas. Then the degrees of freedom of a communication link between a multiple antenna base station and a single antenna user simply becomes equal to 1. This degrees of freedom limitation can be alleviated by multi-user MIMO (MU-MIMO) transmission [3].

In MU-MIMO systems, multiple users simultaneously receive their own unicast data streams [4]. The capacity of MU-MIMO channels is achieved by dirty paper coding (DPC) [5]. However, DPC is a highly complex, non-linear precoding scheme. Therefore, in the literature, simple linear precoding strategies have been investigated for MU-MIMO transmission [6], [7]. Although sub-optimal in general, superposition coding (SPC) is also used as a simpler alternative to DPC. It offers good interference management [8], [9], [10] and harvests some of the benefits of DPC in MU-MIMO systems with applications to non-orthogonal multiple access (NOMA) [11], [12].

While simultaneous multiple unicast data transmission is important, applications such as mobile application updates, mass advertisements and public group communications require multicasting [13], [14], [15]. In multicast transmission, the same data is sent to a group of users. Precoding for multicasting is studied in [16], [17] and [18]. Precoder design for max-min fairness for multicasting is studied in [19].

Unicast and multicast traffic can also coexist in cellular networks. Scheduling unicast and multicast traffic is considered in [20] and [21], while their simultaneous transmission via superposition coding for an enhanced sum rate is studied in [21]. While [21] is for a very limited system with two users, [22] studies precoder design problem for a system in which all users have to decode the common message and a subset of these users have to receive their own unicast messages.

In multi-group multicasting there are multiple groups, where each user in the same group seeks for the same message and different groups receive different messages [23]. Precoding for a guaranteed quality of service with minimum transmission power is studied in [24] and [25]. Max-min fair transmit precoding for multi-group multicasting under a total power constraint is also studied in [25]. Weighted fair multi-group multicasting precoding under per antenna

power constraints is investigated in [26]. A rate splitting approach is suggested for max-min fair transmit precoder design in [27]. The max-min fair transmit precoding problem for multi-group multicasting in massive MIMO systems is studied in [28] and [29]. Maximum sum rate for multigroup multicast precoding under per-antenna power constraints is studied in [30]. The authors in [31] also study sum rate maximization, yet under a total power constraint, and allow for coding over multiple blocks.

Apart from the above discussion on the objective function (sum rate maximization or max-min fairness) or the system topology (unicast/ multicast transmission etc.), another issue is whether successive interference cancellation (SIC) is perfect or not. In general imperfect channel state information (CSI) and/or hardware impairments result in imperfect SIC. While [9], [11], [12], [22], [27], [32] consider perfect CSI and perfect SIC, [10] considers imperfect CSI and perfect SIC, and [33], [34] consider imperfect SIC.

All the above multi-group multicasting papers, as well as the others we have encountered in the literature, consider non-overlapping multicasting groups. As an initial step in finding sum rate optimal precoders for the most general problem with overlapping multicast groups with unicast messages, in this work we study multi-group multicasting with common messages. In this system there are non-overlapping multicast groups, yet all groups are also interested in a common message. In this paper we follow the outline listed below:

- 1) To be able to send the common message along with the private multicast messages, we define two different transmission schemes: (i) the base station transmits the superposition of the common and multicast messages, (ii) the base station regards the common message as another message and concatenates the multicast message vector with the common message.
- 2) For these transmission schemes, we study the weighted sum rate (WSR) maximization problem. To be able to propose the iterative algorithm described in the next step, we also define the weighted minimum mean square error (WMMSE) problem. We then write the gradient expressions and the Karush-Kuhn-Tucker (KKT) conditions of the Lagrangian equations for both problems. Then, we prove that these two problems are equivalent at the optimum point. Although we cannot find a closed form expression, we state the equations the optimal precoders, receivers and the Lagrange multipliers satisfy.
- 3) The WSR maximization problem is non-convex and the optimal solution is hard to find. Inspired by the equivalency between WMMSE minimization and WSR maximization, we propose a low-complexity iterative algorithm for finding a locally optimum solution. The

algorithm is based on alternating optimization, which iterates between precoders, mean square error (MSE) weights and receiver structures written for the WMMSE problem. We also discuss and display convergence for this algorithm.

- 4) We apply this algorithm to both of the transmission schemes, and compare the results with maximal ratio transmission (MRT), and zero-forcing (ZF) precoding.
- 5) We also investigate the effect of imperfect SIC on achievable WSR.

As a result, we learn that there is a fundamental difference between transmitting the common message via the sum of the private multicast message precoders and via a separate precoder. We find that the two schemes favor common and private multicast messages differently. Also, the former transmission scheme is always better than the latter and is more robust against imperfect SIC. To the best of our knowledge, in the literature, there is no other paper which compares these two schemes to reveal their fundamental differences.

We present the system model, formulate the WSR and WMMSE problems and prove their equivalence in Section II. We propose the iterative algorithm in Section III. We present the simulation results in Section IV, and conclude the paper in Section V.

II. SYSTEM MODEL

We consider a single cell downlink communication system. The base station is equipped with M transmit antennas. The base station communicates with K clusters and there are L single antenna users in each cluster. Each user belongs to only one cluster. The base station has a common data s_c across all users, and private multicast data s_{u_k} ($k = 1, \dots, K$) destined to each of the K clusters.

In this work, our aim is to understand the conditions under which superposition coding is beneficial. For this purpose, we study two different signal models. In the first model, the common message is superposed onto the private multicast message vector, and in the second model, the common message is appended to the private multicast message vector.

Signal Model 1: The base station employs a 2-layer superposition coding scheme, in which the base and enhancement layers respectively carry common and private multicast data. The input data vector is denoted as $\mathbf{s}^{(1)} = [s_1, \dots, s_K]^T \in \mathbb{C}^{K \times 1}$, and each input data stream s_k , $k = 1, \dots, K$ is the sum of common and private multicast data, respectively denoted as s_c and s_{u_k} . Thus, the input data vector can be written as $\mathbf{s}^{(1)} = \mathbf{s}_c + \mathbf{s}_u$, where $\mathbf{s}_c = [s_c, \dots, s_c]^T \in \mathbb{C}^{K \times 1}$ and $\mathbf{s}_u = [s_{u_1}, \dots, s_{u_K}]^T \in \mathbb{C}^{K \times 1}$. We assume s_c and all s_{u_k} are independent and

$\mathbb{E}\{s_c s_c^*\} = \alpha$ and $\mathbb{E}\{s_{u_k} s_{u_k}^*\} = \bar{\alpha}$. Here, $\bar{\alpha} = 1 - \alpha$ and α is the ratio of power allocated to common data. The input data vector $\mathbf{s}^{(1)}$ is linearly processed by a precoder matrix $\mathbf{P}^{(1)} = [\mathbf{p}_1^{(1)}, \dots, \mathbf{p}_K^{(1)}] \in \mathbb{C}^{M \times K}$. Each precoding vector $\mathbf{p}_k^{(1)}$ is of size $M \times 1$.

Signal Model 2: The input data vector is defined as $\mathbf{s}^{(2)} = [s_c, s_{u_1}, \dots, s_{u_K}]^T \in \mathbb{C}^{(K+1) \times 1}$, where s_c and s_{u_k} , $k = 1, \dots, K$, are the same as in the above signal model 1. We assume s_c and all s_{u_k} are independent and $\mathbb{E}\{\mathbf{s}^{(2)} \mathbf{s}^{(2)H}\} = \mathbf{I}$. The input data vector $\mathbf{s}^{(2)}$ is linearly processed by a precoder matrix $\mathbf{P}^{(2)} = [\mathbf{p}_c, \mathbf{p}_1^{(2)}, \dots, \mathbf{p}_K^{(2)}] \in \mathbb{C}^{M \times (K+1)}$, where both the precoding vector $\mathbf{p}_k^{(2)}$ for each multicast data and \mathbf{p}_c for common data are of size $M \times 1$.

Then, for signal model $m = 1, 2$, the overall transmit data vector $\mathbf{x} \in \mathbb{C}^{M \times 1}$ at the base station can be written as

$$\mathbf{x}^{(m)} = \mathbf{P}^{(m)} \mathbf{s}^{(m)} = \mathbf{p}_A^{(m)} s_c + \sum_{k=1}^K \mathbf{p}_k^{(m)} s_{u_k}. \quad (1)$$

Here, $\mathbf{p}_A^{(1)} = \sum_{k=1}^K \mathbf{p}_k^{(1)}$ and $\mathbf{p}_A^{(2)} = \mathbf{p}_c$ for signal models 1 and 2, respectively.

As user-level precoding is assumed, there is an average total power constraint,

$$\mathbb{E}\{\mathbf{x}^{(m)H} \mathbf{x}^{(m)}\} = B^{(m)} \text{Tr}(\mathbf{p}_A \mathbf{p}_A^H) + C^{(m)} \sum_{k=1}^K \text{Tr}(\mathbf{p}_k^{(m)} \mathbf{p}_k^{(m)H}) \leq E_{tx}. \quad (2)$$

In (2), $(B^{(1)}, C^{(1)}) = (\alpha, \bar{\alpha})$ and $(B^{(2)}, C^{(2)}) = (1, 1)$ for signal models 1 and 2, respectively. Note that, signal model 1 seems to be more restrictive, as $\mathbf{p}_A^{(1)}$ has to be set to the sum of the private multicast data precoders. However, it introduces a new degrees of freedom due to the parameter α . The iterative precoder design algorithm we propose in Section III designs precoder directions and power levels jointly. The α parameter will serve as an external handle that allows for adjustments in precoder power levels. Because of this difference, in Section IV, we will observe that the two signal models are fundamentally different.

The received signal at the l -th user of the k -th cluster can be expressed as

$$y_{l,k}^{(m)} = \mathbf{h}_{l,k} \mathbf{p}_A^{(m)} s_c + \mathbf{h}_{l,k} \sum_{i=1}^K \mathbf{p}_i^{(m)} s_{u_i} + n_{l,k}. \quad (3)$$

In (3), $\mathbf{h}_{l,k} \in \mathbb{C}^{1 \times M}$ is the channel gain vector of the l -th user of the k -th cluster, $l = 1, \dots, L$, $k = 1, \dots, K$. The entries in $\mathbf{h}_{l,k}$ denote complex valued channel gains. The noise component $n_{l,k}$ is independent and circularly symmetric complex Gaussian random variable with zero mean and unit variance. The base station is assumed to know all $\mathbf{h}_{l,k}$, while the l -th user in the k -th cluster has to be informed about the composite channel gains $\mathbf{h}_{l,k} \mathbf{p}_A^{(m)}$ and $\mathbf{h}_{l,k} \mathbf{p}_k^{(m)}$. This information can be acquired by utilizing standard training techniques. For example, in time division duplex

mode, in the uplink, the users can send training data to the base station, and the base station learns all $\mathbf{h}_{l,k}$. Then, the base station can send training data in the downlink phase *twice*. This way, the users can learn the composite channel gains for the common and the private multicast messages respectively in the first and second downlink transmissions. We further assume that these steps are all error free.

A. Achievable Data Rates

In this system, all users decode the common message. In addition to this, each user subtracts this common message from its received signal to decode its private multicast message using SIC. Then, the achievable rate for common and private multicast messages for the l -th user of the k -th cluster are respectively defined as $R_{c,lk}$, and $R_{u,lk}$, $l = 1, \dots, L, k = 1, \dots, K$ and are given as

$$R_{c,lk}^{(m)} = \log \left(1 + B^{(m)} \mathbf{p}_A^{(m)H} \mathbf{h}_{l,k}^H r_{c,lk}^{(m)-1} \mathbf{h}_{l,k} \mathbf{p}_A^{(m)} \right), \quad (4)$$

$$R_{u,lk}^{(m)} = \log \left(1 + C^{(m)} \mathbf{p}_k^{(m)H} \mathbf{h}_{l,k}^H r_{u,lk}^{(m)-1} \mathbf{h}_{l,k} \mathbf{p}_k^{(m)} \right). \quad (5)$$

Here $r_{c,lk}^{(m)}$ and $r_{u,lk}^{(m)}$ are the effective noise variances for common and multicast data at the l -th user of the k -th cluster for the m -th signal model. They can be calculated as

$$r_{c,lk}^{(m)} = C^{(m)} \mathbf{h}_{l,k} \left(\sum_{i=1}^K \mathbf{p}_i^{(m)} \mathbf{p}_i^{(m)H} \right) \mathbf{h}_{l,k}^H + 1, \quad (6)$$

$$r_{u,lk}^{(m)} = \delta^2 B^{(m)} \mathbf{h}_{l,k} \mathbf{p}_A^{(m)} \mathbf{p}_A^{(m)H} \mathbf{h}_{l,k}^H + C^{(m)} \mathbf{h}_{l,k} \left(\sum_{i=1, i \neq k}^K \mathbf{p}_i^{(m)} \mathbf{p}_i^{(m)H} \right) \mathbf{h}_{l,k}^H + 1, \quad (7)$$

where δ denotes the amount of residual self interference. If SIC is perfect, $\delta = 0$. Note that, overall the achievable rate for common data is determined by the minimum of all $R_{c,lk}^{(m)}$ and the achievable rate for multicast data for group k , $s_{u,k}$, is determined by the minimum of all $R_{u,lk}^{(m)}$ in group k . Thus, we also define

$$R_c^{(m)} = \min_{k=\{1, \dots, K\}} \min_{l=\{1, \dots, L\}} R_{c,lk}^{(m)}, \quad (8)$$

$$R_{u,k}^{(m)} = \min_{l=\{1, \dots, L\}} R_{u,lk}^{(m)}. \quad (9)$$

B. The Maximum WSR Problem

In this paper, our aim is to find the optimal precoders $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$ for signal models 1 and 2 respectively such that the WSR is maximized subject to a total power constraint. The WSR

of the system can be computed as

$$T_1 = \sum_{k=1}^K a_k R_{u,k}^{(m)} + b R_c^{(m)}, \quad (10)$$

where a_k and b respectively denote the rate weights that correspond to multicast data at cluster k and common data. The optimization problems \mathcal{P}_1 and \mathcal{P}_2 are defined as

$$\mathcal{P}_1 : [\mathbf{p}_1^{(1)SR}, \dots, \mathbf{p}_K^{(1)SR}] = \arg \max_{\mathbf{p}_k^{(1)}} T_1, \quad \text{s.t. (2)}, \quad (11)$$

$$\mathcal{P}_2 : [\mathbf{p}_c^{SR}, \mathbf{p}_1^{(2)SR}, \dots, \mathbf{p}_K^{(2)SR}] = \arg \max_{\mathbf{p}_c, \mathbf{p}_k^{(2)}} T_1, \quad \text{s.t. (2)}. \quad (12)$$

We need to convert these two problems to smooth constrained optimization problems because they are non-convex and difficult to solve. Thus, we introduce and constrain two new auxiliary variables $t_k, k = 1, \dots, K$ and z as

$$t_k \leq a_k R_{u,lk}^{(m)}, \quad \forall l, \forall k, \quad (13)$$

$$z \leq b R_{c,lk}^{(m)}, \quad \forall l, \forall k. \quad (14)$$

Then, we can reformulate \mathcal{P}_1 and \mathcal{P}_2 as:

$$\mathcal{P}'_1 : [\mathbf{p}_1^{(1)SR}, \dots, \mathbf{p}_K^{(1)SR}] = \arg \max_{\mathbf{p}_k^{(1)}, t_k, z} T_2, \quad \text{s.t. (2), (13), (14)}, \quad (15)$$

$$\mathcal{P}'_2 : [\mathbf{p}_c^{SR}, \mathbf{p}_1^{(2)SR}, \dots, \mathbf{p}_K^{(2)SR}] = \arg \max_{\mathbf{p}_c, \mathbf{p}_k^{(2)}, t_k, z} T_2, \quad \text{s.t. (2), (13), (14)}, \quad (16)$$

where $T_2 = \sum_{k=1}^K t_k + z$.

C. Error Variance Definitions and the Minimum WMMSE Problem

The one-to-one correspondence between mutual information and minimum mean square error (MMSE) for Gaussian channels is established in [35], and for MIMO broadcast channels in [7]. To establish such a correspondence in our system setup, we first write the MMSE expressions and then compare them with the achievable rates in (4) and (5).

The l -th user of the k -th cluster first processes its received signal $y_{l,k}^{(m)}$, with the common data receiver $W_{l,k}^{(m)}$ to form an estimate of s_c , denoted as $\hat{s}_{c,lk} = W_{l,k}^{(m)} y_{l,k}^{(m)}$. In the second stage, the l -th user of the k -th cluster forms an estimate for the multicast message u_k as $\hat{s}_{u_{lk}} = V_{l,k}^{(m)} \cdot \left(y_{l,k}^{(m)} - \mathbf{h}_{l,k} \mathbf{p}_A^{(m)} s_c + \delta \mathbf{h}_{l,k} \mathbf{p}_A^{(m)} s_c \right)$. In all the following analysis, we will assume perfect SIC, $\delta = 0$, and in Section IV we will investigate the effect of nonzero δ .

MSE expressions of common and multicast data for the l -th user of the k -th cluster are respectively defined as $\varepsilon_{c,lk}^{(m)} = \mathbb{E} \{ \|\hat{s}_{c,lk} - s_c\|^2 \}$, and $\varepsilon_{u,lk}^{(m)} = \mathbb{E} \{ \|\hat{s}_{u_{lk}} - s_{u_k}\|^2 \}$, and for perfect

SIC, their closed form expressions can be written as

$$\begin{aligned} \varepsilon_{c,lk}^{(m)} &= B^{(m)} \mathbf{h}_{l,k} \mathbf{p}_A^{(m)} \mathbf{p}_A^{(m)H} \mathbf{h}_{l,k}^H W_{l,k}^{*(m)} W_{l,k}^{(m)} + C^{(m)} \sum_{i=1}^K \mathbf{h}_{l,k} \mathbf{p}_i^{(m)} \mathbf{p}_i^{(m)H} \mathbf{h}_{l,k}^H W_{l,k}^{*(m)} W_{l,k}^{(m)} \\ &\quad + W_{l,k}^{*(m)} W_{l,k}^{(m)} - \mathbf{h}_{l,k} \mathbf{p}_A^{(m)} W_{l,k}^{(m)} B^{(m)} - B^{(m)} \mathbf{p}_A^{(m)H} \mathbf{h}_{l,k}^H W_{l,k}^{*(m)} + B^{(m)}, \end{aligned} \quad (17)$$

$$\begin{aligned} \varepsilon_{u,lk}^{(m)} &= C^{(m)} \sum_{i=1}^K \mathbf{h}_{l,k} \mathbf{p}_i^{(m)} \mathbf{p}_i^{(m)H} \mathbf{h}_{l,k}^H V_{l,k}^{*(m)} V_{l,k}^{(m)} + V_{l,k}^{*(m)} V_{l,k}^{(m)} - \mathbf{h}_{l,k} \mathbf{p}_k^{(m)} V_{l,k}^{(m)} C^{(m)} \\ &\quad - C^{(m)} \mathbf{p}_k^{(m)H} \mathbf{h}_{l,k}^H V_{l,k}^{*(m)} + C^{(m)}. \end{aligned} \quad (18)$$

The optimal MMSE receivers for common and multicast data are defined as $W_{l,k}^{(m)MMSE} = \arg \min_{W_{l,k}} \varepsilon_{c,lk}^{(m)}$ and $V_{l,k}^{(m)MMSE} = \arg \min_{V_{l,k}} \varepsilon_{u,lk}^{(m)}$. The closed form expressions for these MMSE receivers are then calculated as

$$W_{l,k}^{(m)MMSE} = B^{(m)} \mathbf{p}_A^{(m)H} \mathbf{h}_{l,k}^H \left(B^{(m)} \mathbf{h}_{l,k} \mathbf{p}_A^{(m)} \mathbf{p}_A^{(m)H} \mathbf{h}_{l,k}^H + r_{c,lk}^{(m)} \right)^{-1}, \quad (19)$$

$$V_{l,k}^{(m)MMSE} = C^{(m)} \mathbf{p}_k^{(m)H} \mathbf{h}_{l,k}^H \left(C^{(m)} \mathbf{h}_{l,k} \mathbf{p}_k^{(m)} \mathbf{p}_k^{(m)H} \mathbf{h}_{l,k}^H + r_{u,lk}^{(m)} \right)^{-1}. \quad (20)$$

Given that these MMSE receivers in (19) and (20) are employed, the resulting error variance expressions in (17) and (18) become

$$\varepsilon_{c,lk}^{(m)MMSE} = \left(\frac{1}{B^{(m)}} + \mathbf{p}_A^{(m)H} \mathbf{h}_{l,k}^H r_{c,lk}^{(m)-1} \mathbf{h}_{l,k} \mathbf{p}_A^{(m)} \right)^{-1}, \quad (21)$$

$$\varepsilon_{u,lk}^{(m)MMSE} = \left(\frac{1}{C^{(m)}} + \mathbf{p}_k^{(m)H} \mathbf{h}_{l,k}^H r_{u,lk}^{(m)-1} \mathbf{h}_{l,k} \mathbf{p}_k^{(m)} \right)^{-1}. \quad (22)$$

Comparing (4) and (5) with (21) and (22) we can write

$$R_{c,lk}^{(m)} = -\log \left(\frac{\varepsilon_{c,lk}^{(m)MMSE}}{B^{(m)}} \right), \quad (23)$$

$$R_{u,lk}^{(m)} = -\log \left(\frac{\varepsilon_{u,lk}^{(m)MMSE}}{C^{(m)}} \right). \quad (24)$$

We also define

$$\varepsilon_c^{(m)MMSE} = \max_{k=\{1,\dots,K\}} \max_{l=\{1,\dots,L\}} \varepsilon_{c,lk}^{(m)MMSE}, \quad (25)$$

$$\varepsilon_{u,k}^{(m)MMSE} = \max_{l=\{1,\dots,L\}} \varepsilon_{u,lk}^{(m)MMSE}. \quad (26)$$

The WMMSE minimization objective function is then

$$Q_1 = \sum_{k=1}^K v_k^{(m)} \varepsilon_{u,k}^{(m)MMSE} + w^{(m)} \varepsilon_c^{(m)MMSE}, \quad (27)$$

where $w^{(m)}$ and $v_k^{(m)}$ denote the MMSE weights for common data for all users and multicast data at cluster k respectively. The optimization problems \mathcal{P}_3 and \mathcal{P}_4 are defined as

$$\mathcal{P}_3 : [\mathbf{p}_1^{(1)MS}, \dots, \mathbf{p}_K^{(1)MS}] = \arg \min_{\mathbf{p}_k^{(1)}} Q_1, \quad \text{s.t. (2)}, \quad (28)$$

$$\mathcal{P}_4 : [\mathbf{p}_c^{MS}, \mathbf{p}_1^{(2)MS}, \dots, \mathbf{p}_K^{(2)MS}] = \arg \min_{\mathbf{p}_c, \mathbf{p}_k^{(2)}} Q_1, \quad \text{s.t. (2)}. \quad (29)$$

To solve these non-convex problems, we reformulate both \mathcal{P}_3 and \mathcal{P}_4 as smooth constrained optimization problems by introducing auxiliary variables. Without loss of generality, we choose the auxiliary variables as a function of the independent variables t_k and z defined in (13) and (14) such that

$$\varepsilon_{u,lk}^{(m)MMSE} \leq e^{-t_k/a_k}, \quad \forall l, \forall k, \quad (30)$$

$$\varepsilon_{c,lk}^{(m)MMSE} \leq e^{-z/b}, \quad \forall l, \forall k. \quad (31)$$

As stated above, this choice does not result in a loss of generality, but is necessary to establish the equivalence between WSR and WMMSE problems. Then, we reformulate \mathcal{P}_3 and \mathcal{P}_4 as

$$\mathcal{P}'_3 : [\mathbf{p}_1^{(1)MS}, \dots, \mathbf{p}_K^{(1)MS}] = \arg \min_{\mathbf{p}_k^{(1)}, t_k, z} Q_2, \quad \text{s.t. (2), (30), (31)}, \quad (32)$$

$$\mathcal{P}'_4 : [\mathbf{p}_c^{MS}, \mathbf{p}_1^{(2)MS}, \dots, \mathbf{p}_K^{(2)MS}] = \arg \min_{\mathbf{p}_c, \mathbf{p}_k^{(2)}, t_k, z} Q_2, \quad \text{s.t. (2), (30), (31)}, \quad (33)$$

where

$$Q_2 = \sum_{k=1}^K v_k^{(m)} e^{-t_k/a_k} + w^{(m)} e^{-z/b}. \quad (34)$$

In the following, we first prove that the precoders designed for maximum WSR and minimum WMMSE are equivalent at the optimal point. Then we propose an algorithm for precoder design.

D. Gradient Expressions and KKT Conditions for Maximum WSR

In this section, we study the gradients for the WSR maximization problem. To investigate the stationary points of the problems \mathcal{P}'_1 and \mathcal{P}'_2 , we formulate the Lagrangian expression as

$$\begin{aligned} f(\mathbf{P}^{(m)}, t_k, z) = & -T_2 + \sum_{k=1}^K \sum_{l=1}^L \mu_{l,k}^{(m)} (t_k - a_k R_{u,lk}^{(m)}) + \sum_{k=1}^K \sum_{l=1}^L \eta_{l,k}^{(m)} (z - b R_{c,lk}^{(m)}) \\ & + \lambda^{(m)} \left(B^{(m)} \left\| \mathbf{p}_A^{(m)} \right\|^2 + C^{(m)} \sum_{k=1}^K \left\| \mathbf{p}_k^{(m)} \right\|^2 - E_{tx} \right). \end{aligned} \quad (35)$$

Here $\lambda^{(m)}$, $\mu_{l,k}^{(m)}$ and $\eta_{l,k}^{(m)}$ are the Lagrange multipliers. We calculate $\nabla_{\mathbf{p}_k^{(m)}} f(\mathbf{P}^{(m)}, t_k, z)$ in Appendix A¹ only for signal model 1 ($m = 1$) due to the space limitations. We restate the result.

¹The gradient of a function $f(\mathbf{x})$ with respect to its complex variable \mathbf{x} is denoted as $\nabla_{\mathbf{x}} f(\mathbf{x})$ and its m th element is defined as $[\nabla_{\mathbf{x}} f(\mathbf{x})]_m = \nabla_{[\mathbf{x}]_m} f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}^*]_m}$. For detailed derivation rules, we refer the reader to [36].

$$\begin{aligned}
\nabla_{\mathbf{p}_k^{(1)}} f(\mathbf{P}^{(1)}, t_k, z) = & - \sum_{l=1}^L \mu_{l,k}^{(1)} a_k \mathbf{h}_{l,k}^H r_{u,lk}^{(1)-1} \mathbf{h}_{l,k} \mathbf{p}_k^{(1)} \varepsilon_{u,lk}^{(1)MMSE} + \lambda^{(1)} \left(B^{(1)} \mathbf{p}_A^{(1)} + C^{(1)} \mathbf{p}_k^{(1)} \right) \\
& + \sum_{i=1, i \neq k}^K \sum_{l=1}^L C^{(1)} \mu_{l,i}^{(1)} a_i \mathbf{h}_{l,i}^H r_{u,li}^{(1)-1} \mathbf{h}_{l,i} \mathbf{p}_i^{(1)} \varepsilon_{u,li}^{(1)MMSE} \mathbf{p}_i^{(1)H} \mathbf{h}_{l,i}^H r_{u,li}^{(1)-1} \mathbf{h}_{l,i} \mathbf{p}_k^{(1)} \\
& + \sum_{i=1}^K \sum_{l=1}^L C^{(1)} \eta_{l,i}^{(1)} b \mathbf{h}_{l,i}^H r_{c,li}^{(1)-1} \mathbf{h}_{l,i} \mathbf{p}_A^{(1)} \varepsilon_{c,li}^{(1)MMSE} \mathbf{p}_A^{(1)H} \mathbf{h}_{l,i}^H r_{c,li}^{(1)-1} \mathbf{h}_{l,i} \mathbf{p}_k^{(1)} \\
& - \sum_{i=1}^K \sum_{l=1}^L \eta_{l,i}^{(1)} b \mathbf{h}_{l,i}^H r_{c,li}^{(1)-1} \mathbf{h}_{l,i} \mathbf{p}_A^{(1)} \varepsilon_{c,li}^{(1)MMSE}. \tag{36}
\end{aligned}$$

E. Gradient Expressions and KKT Conditions for Minimum WMMSE

To investigate the stationary points of the minimum WMMSE problems \mathcal{P}'_3 and \mathcal{P}'_4 , we formulate the Lagrangian expression as

$$\begin{aligned}
g(\mathbf{P}^{(m)}, t_k, z) = & Q_2 + \sum_{k=1}^K \sum_{l=1}^L \bar{\mu}_{l,k}^{(m)} v_k^{(m)} (\varepsilon_{u,lk}^{(m)MMSE} - e^{-t_k/a_k}) \\
& + \sum_{k=1}^K \sum_{l=1}^L \bar{\eta}_{l,k}^{(m)} w^{(m)} (\varepsilon_{c,lk}^{(m)MMSE} - e^{-z/b}) \\
& + \bar{\lambda}^{(m)} \left(B^{(m)} \left\| \mathbf{p}_A^{(m)} \right\|^2 + C^{(m)} \sum_{k=1}^K \left\| \mathbf{p}_k^{(m)} \right\|^2 - E_{tx} \right). \tag{37}
\end{aligned}$$

Here $\bar{\lambda}^{(m)}$, $\bar{\mu}_{l,k}^{(m)}$ and $\bar{\eta}_{l,k}^{(m)}$ are the Lagrange multipliers. The gradient $\nabla_{\mathbf{p}_k^{(1)}} g(\mathbf{P}^{(1)}, t_k, z)$ is computed in a similar manner as $\nabla_{\mathbf{p}_k^{(1)}} f(\mathbf{P}^{(1)}, t_k, z)$ and is written as in (38).

$$\begin{aligned}
& \nabla_{\mathbf{p}_k^{(1)}} g(\mathbf{P}^{(1)}, t_k, z) \\
= & - \sum_{l=1}^L \bar{\mu}_{l,k}^{(1)} \mathbf{h}_{l,k}^H r_{u,lk}^{(1)-1} \mathbf{h}_{l,k} \mathbf{p}_k^{(1)} \varepsilon_{u,lk}^{(1)MMSE} v_k^{(m)} \varepsilon_{u,lk}^{(1)MMSE} + \bar{\lambda}^{(1)} \left(B^{(1)} \mathbf{p}_A^{(1)} + C^{(1)} \mathbf{p}_k^{(1)} \right) \\
& - \sum_{i=1}^K \sum_{l=1}^L \bar{\eta}_{l,i}^{(1)} \mathbf{h}_{l,i}^H r_{c,li}^{(1)-1} \mathbf{h}_{l,i} \mathbf{p}_A^{(1)} \varepsilon_{c,li}^{(1)MMSE} w^{(m)} \varepsilon_{c,li}^{(1)MMSE} \\
& + \sum_{i=1, i \neq k}^K \sum_{l=1}^L C^{(1)} \bar{\mu}_{l,i}^{(1)} \mathbf{h}_{l,i}^H r_{u,li}^{(1)-1} \mathbf{h}_{l,i} \mathbf{p}_i^{(1)} \varepsilon_{u,li}^{(1)MMSE} v_i^{(m)} \varepsilon_{u,li}^{(1)MMSE} \mathbf{p}_i^{(1)H} \mathbf{h}_{l,i}^H r_{u,li}^{(1)-1} \mathbf{h}_{l,i} \mathbf{p}_k^{(1)} \\
& + \sum_{i=1}^K \sum_{l=1}^L C^{(1)} \bar{\eta}_{l,i}^{(1)} \mathbf{h}_{l,i}^H r_{c,li}^{(1)-1} \mathbf{h}_{l,i} \mathbf{p}_A^{(1)} \varepsilon_{c,li}^{(1)MMSE} w^{(m)} \varepsilon_{c,li}^{(1)MMSE} \mathbf{p}_A^{(1)H} \mathbf{h}_{l,i}^H r_{c,li}^{(1)-1} \mathbf{h}_{l,i} \mathbf{p}_k^{(1)} \tag{38}
\end{aligned}$$

F. Equivalence of WSR and WMMSE Problems

When (36) and (38) are compared, it is observed that for a given set of precoders $\mathbf{P}^{(m)}$ and corresponding error variances $\varepsilon_{u,lk}^{(m)MMSE}$ and $\varepsilon_{c,lk}^{(m)MMSE}$, $\nabla_{\mathbf{p}_k^{(m)}} f$ and $\nabla_{\mathbf{p}_k^{(m)}} g$ become equal, if

the weights a_k , b , v_k and w are chosen as

$$v_k^{(m)} = a_k \varepsilon_{u,lk}^{(m)MMSE^{-1}}, \quad (39)$$

for all k and l for which $\bar{\mu}_{l,k}^{(m)} > 0$, and

$$w^{(m)} = b \varepsilon_{c,lk}^{(m)MMSE^{-1}}, \quad (40)$$

for all k and l for which $\bar{\eta}_{l,k}^{(m)} > 0$. These relations could also be read as $\varepsilon_{u,lk}^{(m)MMSE} = a_k/v_k^{(m)}$, and $\varepsilon_{c,lk}^{(m)MMSE} = b/w^{(m)}$. In other words, at the optimal solution, the MMSE values for the multicast messages are the same within a group, the MMSE values for the common message are the same over the entire set of users, and $\varepsilon_{u,lk}^{(m)MMSE}$ and $\varepsilon_{c,lk}^{(m)MMSE}$ are equal to their own boundaries; i.e. e^{-t_k/a_k} , and $e^{-z/b}$ respectively. If this is not possible for any subset of users, then the corresponding Lagrange multipliers are zero. This fact also reveals that $\partial_{t_k} f(\mathbf{P}^{(m)}, t_k, z)$ ² and $\partial_{t_k} g(\mathbf{P}^{(m)}, t_k, z)$, and $\partial_z f(\mathbf{P}^{(m)}, t_k, z)$ and $\partial_z g(\mathbf{P}^{(m)}, t_k, z)$ are equivalent. In conclusion, the two problems have different variable names, but to find the optimal solution, we solve for exactly the same set of equations. Thus, the two optimization problems are equivalent with each other at the optimal solution.

III. ITERATIVE PRECODER DESIGN

In this section, we suggest a suboptimal and iterative precoder design algorithm to solve the WSR problem. The WSR problem is non-convex, and hard to solve. It requires efficient algorithms that perform well in practice. Although the WSR problem does not result in an intuitive algorithm, its equivalent WMMSE problem does. As the WMMSE problem is composed of two parts, the transmit precoders and MSE receivers, one can perform alternating optimization between the two.

To do so, we first define a new optimization problem, same as \mathcal{P}_3 (or \mathcal{P}_4) defined in (28) (or (29)), but instead of defining Q_1 in (27) in terms of $\varepsilon_{c,lk}^{(m)MMSE}$ and $\varepsilon_{u,lk}^{(m)MMSE}$, we use $\varepsilon_{c,lk}^{(m)}$ and $\varepsilon_{u,lk}^{(m)}$ defined in (17) and (18). In other words, instead of assuming MMSE receivers, we first allow for any receiver structure.

$$\mathcal{P}_5 : \arg \min_{\mathbf{P}^{(m)}, t_k, z, W_{l,k}^{(m)}, V_{l,k}^{(m)}, v_k^{(m)}, w^{(m)}} Q_2 \quad (41)$$

$$\text{s.t. } \varepsilon_{u,lk}^{(m)} \leq e^{-t_k/a_k}, \quad (42)$$

$$\varepsilon_{c,lk}^{(m)} \leq e^{-z/b} \text{ and (2)}. \quad (43)$$

²The notation, $\partial_x f(x)$ denotes $\frac{\partial f(x)}{\partial x}$.

Then, the new Lagrangian objective function becomes

$$h(\mathbf{P}^{(m)}, t_k, z) = Q_2 + \sum_{k=1}^K \sum_{l=1}^L \xi_{l,k}^{(m)} v_k^{(m)} (\varepsilon_{u,lk}^{(m)} - e^{-t_k/a_k}) + \sum_{k=1}^K \sum_{l=1}^L \psi_{l,k}^{(m)} w^{(m)} (\varepsilon_{c,lk}^{(m)} - e^{-z/b}) + \beta^{(m)} \left(B^{(m)} \left\| \mathbf{p}_A^{(m)} \right\|^2 + C^{(m)} \sum_{k=1}^K \left\| \mathbf{p}_k^{(m)} \right\|^2 - E_{tx} \right), \quad (44)$$

where $\beta^{(m)}$, $\xi_{l,k}^{(m)}$ and $\psi_{l,k}^{(m)}$ denote the Lagrange multipliers for the m -th signal model.

Studying the KKT conditions for this problem, similar to the analysis in the previous section, we can state the following theorem.

Theorem 1. *The common data receiver $W_{l,k}^{(m)}$ in (45), the multicast data receiver $V_{l,k}^{(m)}$ in (46) and the Lagrange multiplier $\beta^{(m)}$ in (47), transmit precoders $\mathbf{p}_k^{(1)}$ in (48), $\mathbf{p}_k^{(2)}$ in (49), and \mathbf{p}_c in (50) satisfy the KKT conditions for the optimization problem defined above with the Lagrangian function $h(\mathbf{P}^{(m)}, t_k, z)$ defined in (44).*

$$W_{l,k}^{(m)} = B^{(m)} \mathbf{p}_A^{(m)H} \mathbf{h}_{l,k}^H \left(B^{(m)} \mathbf{h}_{l,k} \mathbf{p}_A^{(m)} \mathbf{p}_A^{(m)H} \mathbf{h}_{l,k}^H + \sum_{i=1}^K C^{(m)} \mathbf{h}_{l,k} \mathbf{p}_i^{(m)} \mathbf{p}_i^{(m)H} \mathbf{h}_{l,k}^H + 1 \right)^{-1}, \quad (45)$$

$$V_{l,k}^{(m)} = C^{(m)} \mathbf{p}_k^{(m)H} \mathbf{h}_{l,k}^H \left(\sum_{i=1}^K C^{(m)} \mathbf{h}_{l,k} \mathbf{p}_i^{(m)} \mathbf{p}_i^{(m)H} \mathbf{h}_{l,k}^H + 1 \right)^{-1}, \quad (46)$$

$$\beta^{(m)} = \frac{1}{E_{tx}} \sum_{k=1}^K \sum_{l=1}^L \left[\xi_{l,k}^{(m)} v_k^{(m)} V_{l,k}^{(m)} V_{l,k}^{(m)*} + \psi_{l,k}^{(m)} w^{(m)} W_{l,k}^{(m)} W_{l,k}^{(m)*} \right], \quad (47)$$

$$\begin{aligned} \mathbf{p}_k^{(1)} &= \left(\beta^{(1)} \mathbf{I} + \sum_{i=1}^K \sum_{l=1}^L \left(\xi_{l,i}^{(1)} v_i^{(1)} C^{(1)} \mathbf{h}_{l,i}^H V_{l,i}^{(1)*} V_{l,i}^{(1)} \mathbf{h}_{l,i} + \psi_{l,i}^{(1)} w^{(1)} \mathbf{h}_{l,i}^H W_{l,i}^{(1)*} W_{l,i}^{(1)} \mathbf{h}_{l,i} \right) \right)^{-1} \\ &\times \left[\sum_{i=1}^K \sum_{l=1}^L \left(\psi_{l,i}^{(1)} w^{(1)} B^{(1)} \mathbf{h}_{l,i}^H W_{l,i}^{(1)*} - \psi_{l,i}^{(1)} w^{(1)} B^{(1)} \mathbf{h}_{l,i}^H W_{l,i}^{(1)*} W_{l,i}^{(1)} \mathbf{h}_{l,i} (\mathbf{p}_A^{(1)} - \mathbf{p}_k^{(1)}) \right) \right. \\ &\left. + \sum_{l=1}^L \xi_{l,k}^{(1)} v_k^{(1)} C^{(1)} \mathbf{h}_{l,k}^H V_{l,k}^{(1)*} - \beta^{(1)} B^{(1)} (\mathbf{p}_A^{(1)} - \mathbf{p}_k^{(1)}) \right], \end{aligned} \quad (48)$$

$$\begin{aligned} \mathbf{p}_k^{(2)} &= \left(\beta^{(2)} \mathbf{I} + \sum_{i=1}^K \sum_{l=1}^L \xi_{l,i}^{(2)} v_i^{(2)} \mathbf{h}_{l,i}^H V_{l,i}^{(2)*} V_{l,i}^{(2)} \mathbf{h}_{l,i} + \sum_{i=1}^K \sum_{l=1}^L \psi_{l,i}^{(2)} w^{(2)} \mathbf{h}_{l,i}^H W_{l,i}^{(2)*} W_{l,i}^{(2)} \mathbf{h}_{l,i} \right)^{-1} \\ &\times \left(\sum_{l=1}^L \xi_{l,k}^{(2)} v_k^{(2)} \mathbf{h}_{l,k}^H V_{l,k}^{(2)*} \right), \end{aligned} \quad (49)$$

Algorithm 1: Iterative WMMSE

input: $m, a_k, b, \epsilon, \alpha, E_{tx}, \Upsilon$;

set $n = 0, [\mathbf{p}_k^{(m)}]^{(n)} = \mathbf{p}_k^{init} \forall k, \nu = \log_2(KL)/\epsilon$;

iterate;

1. update $n = n + 1$

2. compute $[W_{l,k}^{(m)}]^{(n)}$ using (45)

3. compute $[V_{l,k}^{(m)}]^{(n)}$ using (46)

4. compute $\varepsilon_{c,lk}^{(m)}, \varepsilon_{u,lk}^{(m)}$ using (17) and (18)

5. compute $w^{(m)}, v_k^{(m)}$ using (39) and (40)

6. compute $[\xi_{l,k}^{(m)}]^{(n)}, [\psi_{l,k}^{(m)}]^{(n)}$ using (51) and (52)

7. compute $[\beta^{(m)}]^{(n)}$ using (47)

8. compute $[\mathbf{p}_k^{(m)}]^{(n)}$ using (48) or (49), and $[\mathbf{p}_c]^{(n)}$ using (50)

9. scale $[\mathbf{p}_k^{(m)}]^{(n)}$ such that

$$B^{(m)} \left\| [\mathbf{P}_A^{(m)}]^{(n)} \right\|^2 + C^{(m)} \sum_{k=1}^{K_1} \left\| [\mathbf{P}_k^{(m)}]^{(n)} \right\|^2 = E_{tx}$$

10. If $\text{Tr} \left\{ \left([\mathbf{P}^{(m)}]^{(n)} - [\mathbf{P}^{(m)}]^{(n-1)} \right) \left([\mathbf{P}^{(m)}]^{(n)} - [\mathbf{P}^{(m)}]^{(n-1)} \right)^H \right\} < \Upsilon$ then
 terminate

else

go to Step 1

$$\mathbf{p}_c = \left(\beta^{(2)} \mathbf{I} + \sum_{i=1}^K \sum_{l=1}^L \psi_{l,i}^{(2)} w^{(2)} \mathbf{h}_{l,i}^H W_{l,i}^{(2)*} W_{l,i}^{(2)} \mathbf{h}_{l,i} \right)^{-1} \left(\sum_{i=1}^K \sum_{l=1}^L \psi_{l,i}^{(2)} w^{(2)} \mathbf{h}_{l,i}^H W_{l,i}^{(2)*} \right). \quad (50)$$

Proof. The proof is provided in Appendix B. \square

Remark. The receivers $W_{l,k}^{(m)}$ and $V_{l,k}^{(m)}$ in (45) and (46) are exactly equal to the MMSE receivers given in (19) and (20).

If it was possible to solve the equations stated in Theorem 1 in closed form, then the optimal solution would be obtained. As this is not possible, the iterative WMMSE algorithm, given in Algorithm 1, iterates between using the receiver structures (45), (46), MSE weights (39), (40) and transmit precoders (48), (49) and (50), until convergence to the local optimum; i.e. the error power of two consecutive precoders is small enough.

Calculating the Lagrange multipliers $\xi_{l,k}^{(m)}$ and $\psi_{l,k}^{(m)}$, $\forall l, k$ and m in Algorithm 1 is not trivial. Applying the exponential penalty method for solving min-max problems defined in [37], in the

algorithm in each iteration, we update $\xi_{l,k}^{(m)}$ and $\psi_{l,k}^{(m)}$ according to

$$\xi_{l,k}^{(m)} = \frac{\exp\{\nu(\varepsilon_{u,lk}^{(m)} - e^{-t_k/a_k})\}}{\sum_{l=1}^L \exp\{\nu(\varepsilon_{u,lk}^{(m)} - e^{-t_k/a_k})\}}, \quad (51)$$

$$\psi_{l,k}^{(m)} = \frac{\exp\{\nu(\varepsilon_{c,lk}^{(m)} - e^{-z/b})\}}{\sum_{k=1}^K \sum_{l=1}^L \exp\{\nu(\varepsilon_{c,lk}^{(m)} - e^{-z/b})\}}. \quad (52)$$

Here ν is a constant and as long as $\nu \geq \log(KL)/\epsilon$, the solution is ϵ -optimal. Note that, this choice satisfies the KKT conditions on $\xi_l^{(m)}$ and $\psi_{l,k}^{(m)}$ since $\sum_{l=1}^L \xi_{l,k}^{(m)} = 1$, $\sum_{k=1}^K \sum_{l=1}^L \psi_{l,k}^{(m)} = 1$ and $\xi_{l,k}^{(m)} \geq 0$, $\psi_{l,k}^{(m)} \geq 0$.

Note that, papers such as [10] and [27] utilize the CVX tool designed for MATLAB [38] and do not need to solve for the Lagrange multipliers or for the optimal structures for the transmit precoders, the common data receivers and the multicast data receivers defined in (45)-(50). When these optimal structures are used, the algorithm finishes in approximately 10 minutes, whereas the algorithm lasts for hours when the CVX tool is employed.

As there is a total power constraint, and each iteration of the algorithm increases the objective function, the proposed WMMSE algorithm converges to a limit value. Due to the non-convexity of the problem, this limit value is not guaranteed to be the global optimum. However, the algorithm employs the precoders and the MMSE receivers stated in Theorem 1 that satisfy the KKT conditions of the WMMSE problem, and results in a locally optimum solution. Following similar steps as in [7, Section IV-A] and [39], one can prove convergence in full detail.

IV. SIMULATION RESULTS

In this section we provide simulation results to compare the two new precoders Algorithm 1 generates. We denote these precoders WMMSE1 and WMMSE2 for signal models 1 and 2 respectively. In the following simulations we consider algorithm convergence, and sum rate for different settings. In the simulations, the entries in $\mathbf{h}_{l,k}$ are assumed to be circularly symmetric complex Gaussian distributed random variables with zero mean and unit variance, and are independent and identically distributed. The presented results are averaged over 10^3 channel realizations. Ideal Gaussian codebooks are used for transmission. In the algorithm, the maximum number of iterations is limited to 100, and both ϵ and Υ are set to 10^{-3} .

In the simulation results we will provide comparisons with ZF and MRT precoding. Thus, we first describe the two schemes briefly.

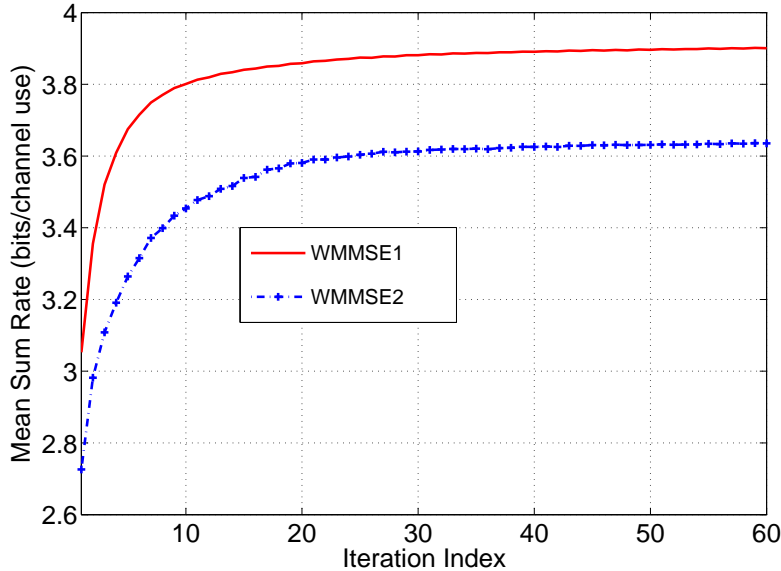


Fig. 1. Sum-rate convergence performance for $M = K = L = 2$. Transmit SNR is set to 15 dB, and $a_k = b = 1$. The optimal α is selected for WMMSE1.

ZF precoding: The ZF precoder aims to cancel all interference at all users. When, the number of users is less than or equal to the number of transmit antennas ($KL \leq M$), interference cancellation is easily achieved by the ZF precoder, $\mathbf{P}^{ZF} \in \mathbb{C}^{M \times KL}$, given by [40]

$$\mathbf{P}^{ZF} \triangleq \sqrt{\frac{E_{tx}}{\text{Tr}((\mathbf{H}\mathbf{H}^H)^{-1})}} \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1}, \quad (53)$$

where $\mathbf{H} = [\mathbf{h}_{1,1}, \mathbf{h}_{1,2}, \dots, \mathbf{h}_{1,L}, \dots, \mathbf{h}_{K,1}, \mathbf{h}_{K,2}, \dots, \mathbf{h}_{K,L}]^T$ is the composite channel gain matrix of all users, of size $KL \times M$. However, when $KL > M$, total interference cancellation at all users is no longer feasible. Instead, user selection is necessary before designing the ZF precoders.

In the simulations, we assume $K = M$. Thus, we select a single user in each group (denoted by l_k) and design the ZF precoder for this set of users only, where $\mathbf{H}' = [\mathbf{h}_{1,l_1}, \dots, \mathbf{h}_{K,l_K}]^T$ is the composite channel gain matrix of all selected users. In order to determine the selected set of users, we consider all possible KL channel gain matrices and choose the one, which maximizes the determinant $|\mathbf{H}'\mathbf{H}'^H|$. This method was shown to minimize MMSE in [41]. We would like to mention that this type of user selection ignores the presence of the common message. The common message is superposed onto the private multicast messages according to signal model 1 during transmission.

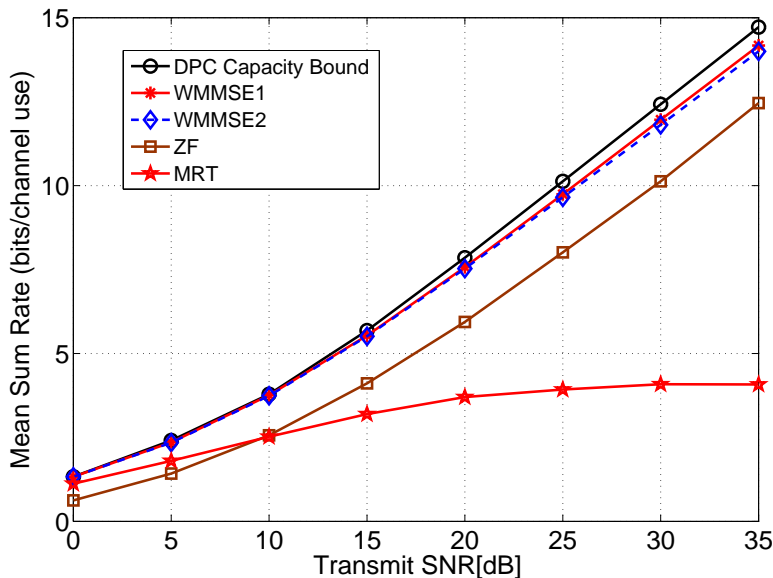


Fig. 2. The sum-rate $\left(\sum_{k=1}^K R_{u,k} + R_c\right)$ curves for $M = 2$, $K = 2$, $L = 1$, $a_k = 1$, $b = 1$. The optimal α is selected for WMMSE1, ZF and MRT.

Note that, this method designs ZF precoders according to a single user in each group. Then the achievable multicast rate for a group is determined by the minimum of all multicast rates achievable by each user in a group as in (9). Similarly, the achievable common data rate is determined according to (8).

MRT precoding: MRT aims to achieve the highest signal gain at the receivers, and ignores interference. The MRT precoder for cluster k is given in [29] as

$$\mathbf{p}_k^{MRT} = \Gamma \sum_{l=1}^L \mathbf{h}_{l,k}, \quad (54)$$

where Γ is set to

$$\Gamma = \frac{E_{tx}}{\sum_{k=1}^K \left| \sum_{l=1}^L h_{k,l} \right|^2}$$

to satisfy the total power constraint in (2). As in the ZF scheme, described above, we assume the common message is superposed onto the private multicast messages.

Fig. 1 displays convergence properties for the two iterative WMMSE algorithms. In the figure, total transmit power, E_{tx} , is set to 15 dB and $a_k = b = 1$ and optimal α is chosen for WMMSE1. The initial precoder matrix, \mathbf{P}^{init} , is chosen as the zero forcing precoder, as described above. The figure confirms that both algorithms converge fast.

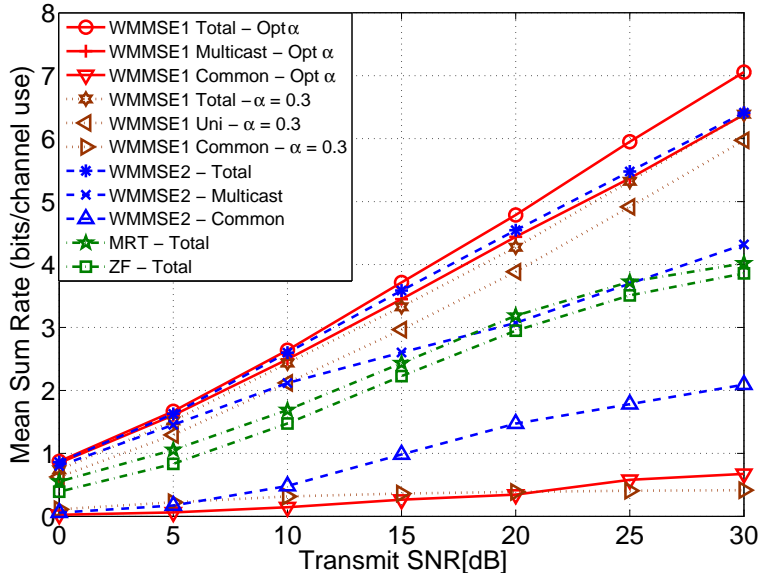


Fig. 3. The common data rate (R_c), the sum multicast data rate ($\sum_{k=1}^K R_{u,k}$), and the sum-rate ($\sum_{k=1}^K R_{u,k} + R_c$) curves for $M = K = L = 2$, $a_k = b = 1$.

Note that, the parameter α distributes power unequally among common and multicast data. This results in their average received signal to noise ratios (SNR) to be different. In order to be able to plot the common and multicast data rates on the same graph, in Figs. 2-4 and 6-7, the horizontal axis is defined as the total transmit SNR, namely E_{tx}/σ^2 . Here σ^2 is the noise variance and assumed to be 1.

Fig. 2 compares WMMSE1 and WMMSE2 with DPC, ZF and MRT, for $M = K = 2$, $L = 1$ and $a_k = b = 1$. Although this constitutes a limited setting in terms of number of antennas and users, in the literature, capacity region results for MIMO broadcast channels with a common message only exist for this configuration [32]. For WMMSE1, ZF and MRT, the optimal α is selected for each protocol for every channel realization. It is observed that both WMMSE1 and WMMSE2 achieve the optimal performance at low SNR, and close to optimal at high SNR. In this setting, ZF removes all undesired interference and is parallel to WMMSE1 and WMMSE2. However, MRT is quite suboptimal as it performs no interference mitigation.

Fig. 3 shows sum rate curves for $M = K = L = 2$, and $a_k = b = 1$. The results for WMMSE1 are presented for both optimal α values and for a fixed α value. It is observed that WMMSE1 for optimal α outperforms all protocols in weighted sum rate. The sum multicast rate of WMMSE1, both for optimal and fixed α , is larger than that of WMMSE2. On the other hand, WMMSE2

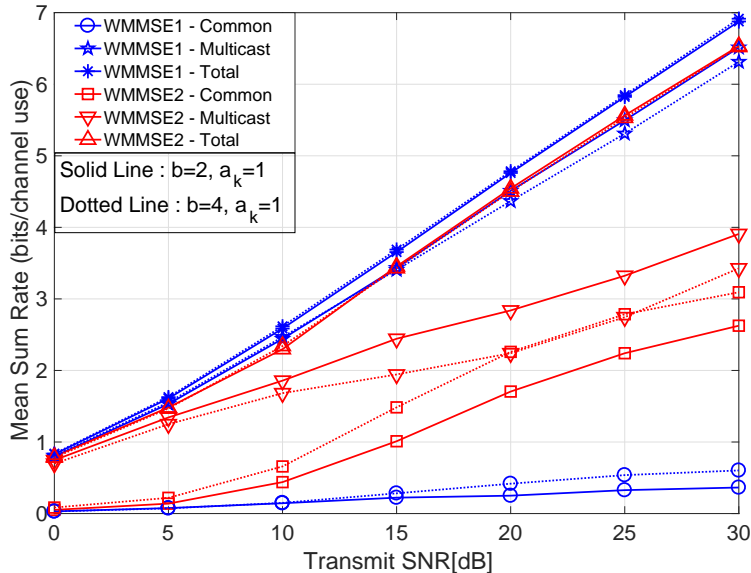


Fig. 4. The common data rate (R_c), the sum multicast data rate ($\sum_{k=1}^K R_{u,k}$), and the sum-rate ($\sum_{k=1}^K R_{u,k} + R_c$) curves for $M = K = L = 2$ and weights $a_k = 1$ and $b = \{2, 4\}$.

achieves a significantly larger common data rate, at the expense of sum multicast rate. This is because, WMMSE1 mainly designs multicast data precoders and sends the common data via the sum precoder $\mathbf{p}_A^{(1)} = \sum_{k=1}^K \mathbf{p}_k^{(1)}$. However, WMMSE2 assigns a separate precoder to common data \mathbf{p}_c . This way, it can achieve higher values for R_c , but the loss in multicast data rates is significant and performs worse than WMMSE1 for optimal α in terms of the total weighted sum rate. However, this is not the case if α is fixed. Note that, there is a tradeoff between WMMSE1 and WMMSE2. The precoder definition of WMMSE2 is more general and includes WMMSE1 definition as a special case. In other words, in WMMSE2, the common message precoder can be arbitrary, whereas in WMMSE1 it is constrained to be the sum of the private multicast data precoders. On the other hand, the WMMSE algorithm calculates the best directions and the power levels for all precoders \mathbf{p}_k and $\mathbf{p}_A^{(1)}$ jointly. It does not perform direction and power optimization steps separately. The WMMSE1 algorithm avoids this problem via superposition. WMMSE1 allows the designer to adjust the precoder power levels separately via the parameter α . If α is optimized, this new degrees of freedom (power optimization gain) becomes dominant over a more constrained precoder definition and WMMSE1 performs better than WMMSE2. Finally, we observe that MRT performs better than ZF, both MRT and ZF perform poorly in terms of weighted sum rate. When there are multiple users in a group, neither MRT, nor ZF

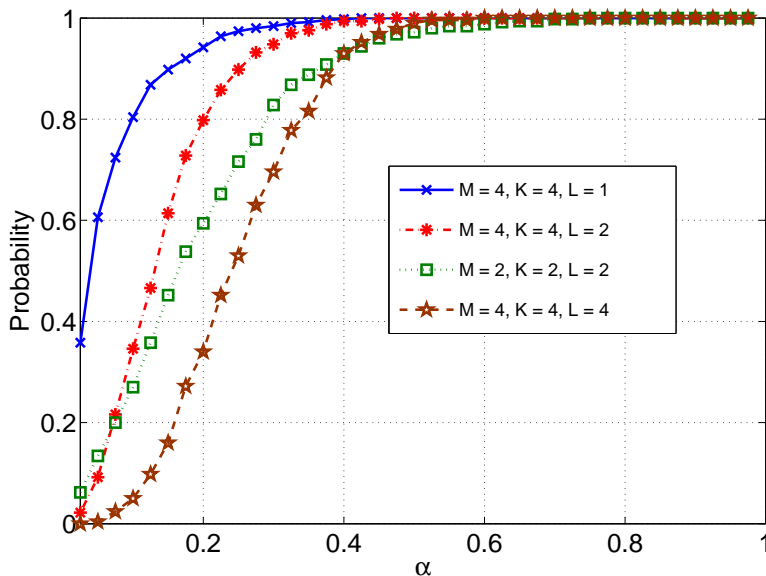


Fig. 5. Cumulative distribution function of optimal α for different scenarios.

can manage interference well. ZF cancels interference at the selected user in each group, but the interference at the other users in the group are not necessarily cancelled. MRT, on the other hand, behaves as if the group itself is one single user with an equivalent channel gain $\sum_{l=1}^L \mathbf{h}_{l,k}$, which does not provide any individual adaptation.

Fig. 4 shows sum rate curves for $M = K = L = 2$, $a_k = 1$ and $b = \{2, 4\}$. The results for WMMSE1 are presented for optimal α values. It is observed that when the weight for the common message, b , increases, the common data rate increases and the sum multicast data rate decreases for both protocols. However, the total data rate does not change.

Fig. 5 shows the cumulative distribution function of optimal α for $M = K = L = 2$ and for $M = K = 4$ and $L = 1, 2, 4$. The $M = K = L = 2$ case is shown to accompany Fig. 3 and the curves for $M = K = 4$ and $L = 1, 2, 4$ accompany Fig. 6. For $M = K = 4$ and $L = 1, 2, 4$, as L increases, optimal α values become larger. When L increases, interference management becomes harder and private multicast data rates decrease (as confirmed in Fig. 6). However, in order to sustain the common data rate, larger α values become beneficial. The comparison between $M = K = L = 2$ and $M = K = 4, L = 2$ is also in line with this observation. In the latter scenario, interference management is easier, and thus its cumulative distribution function is to the left $M = K = L = 2$.

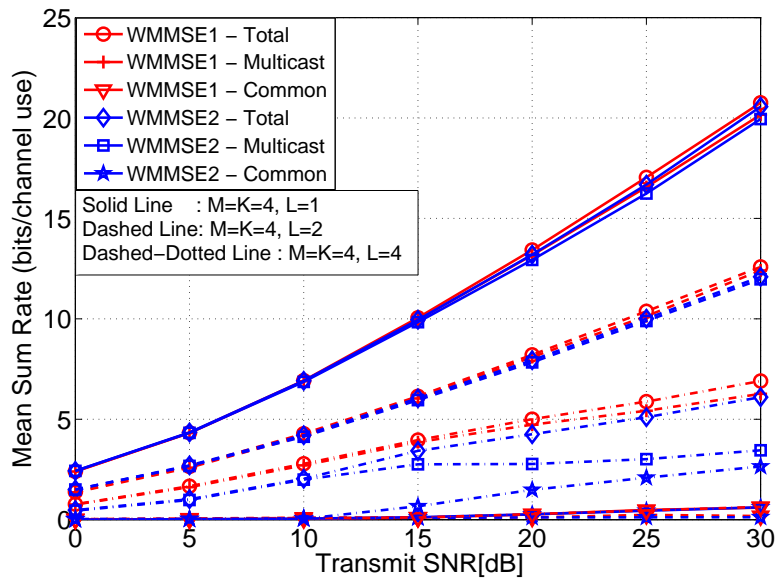


Fig. 6. The common data rate (R_c), the sum multicast data rate ($\sum_{k=1}^K R_{u,k}$), and the sum-rate ($\sum_{k=1}^K R_{u,k} + R_c$) curves for $\{M = K = 4, L = 1\}$, $\{M = K = 4, L = 2\}$ and $\{M = K = 4, L = 4\}$ for $a_k = b = 1$. For WMMSE1 optimal α values are used.

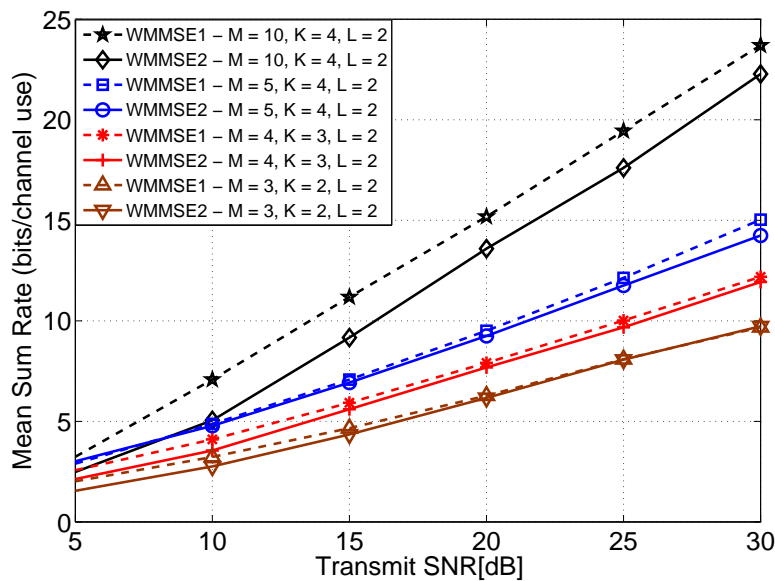


Fig. 7. The sum-rate ($\sum_{k=1}^K R_{u,k} + R_c$) curves for $\{M = 10, K = 4, L = 2\}$, $\{M = 5, K = 4, L = 2\}$, $\{M = 4, K = 3, L = 2\}$ and $\{M = 3, K = 2, L = 2\}$ for $a_k = b = 1$. For WMMSE1 optimal α values are used.

Fig. 6 investigates the effect of number of users in a group for $M = K = 4$ and $L = 1, 2$ or 4. The figure shows sum rate curves for WMMSE1 for optimal α values. As the number

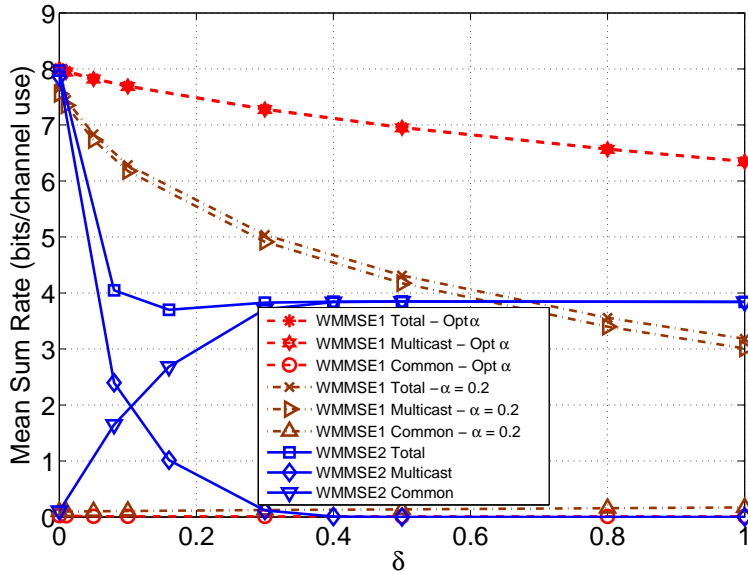


Fig. 8. The common data rate (R_c), the sum multicast data rate ($\sum_{k=1}^K R_{u,k}$), and the sum-rate ($\sum_{k=1}^K R_{u,k} + R_c$) curves for $\{M = K = 4, L = 2\}$ vs. different δ values for $a_k = b = 1$ and transmit SNR is set to 20 dB. For WMMSE1, both optimal α and fixed α is considered.

of users in a group increase, both $R_{u,k}$ and R_c become the minimum of a larger number of random variables, and become smaller. Thus, the weighted sum rate decreases for increasing number of users. As L increases, interference management at the receivers becomes harder. Moreover, with increasing number of users in a group, the difference between WMMSE1 with optimal α and WMMSE2 becomes more significant. WMMSE1 keeps multicasting rates as high as possible, whereas WMMSE2 prefers increasing the common data rate, which is not sufficient to compensate for the decrease in the sum multicast rate.

Fig. 7 compares WMMSE1 for optimal α and WMMSE2 performance for different number of base station antennas and clusters. In particular, the figure shows weighted sum rate curves for $\{M = 10, K = 4, L = 2\}$, $\{M = 5, K = 4, L = 2\}$, $\{M = 4, K = 3, L = 2\}$ and $\{M = 3, K = 2, L = 2\}$. We observe that WMMSE1 for optimal α performs better than WMMSE2 in all cases. Moreover, the gains become significant if $M > K \times L$.

Fig. 8 shows the effect of imperfect SIC on WMMSE1 and WMMSE2 performances. In the figure, $M = K = 4$, $L = 2$, $a_k = b = 1$, transmit SNR is 20 dB and WMMSE1 is displayed for both optimal α and for $\alpha = 0.2$. In obtaining the figure, we assume that δ is unknown at the transmitter and the receivers. Thus the precoders and the receivers designed for perfect SIC are

TABLE I
COMPLEXITY OF PROPOSED ALGORITHMS

Algorithm	Computational Complexity (Number of Matrix Operations)
WMMSE1	$M^3K + M^2(3K^2L + 2K) + M(24K^2L + 21KL + 3K - 1) + K^2(3L^2 + 24L) + K(3L^2 + 38L + 3) + 1$
WMMSE2	$M^3K + M^2(2K^2L + KL + 2K + 1) + M(20K^2L + 24KL + K) + K^2(3L^2 + 14L) + K(3L^2 + 43L) + 1$

continued to be used. The figure reveals that WMMSE1 (either for optimal or for fixed α) is quite robust against imperfect SIC. WMMSE1 with optimal α shows almost no degradation with increasing δ , and WMMSE1 for $\alpha = 0.2$ is better than WMMSE2 for $\delta \leq 0.6$. Note that, residual interference power is related with δ^2 , and practically δ is never as large as 0.6. The figure also emphasizes the fundamental difference between WMMSE1 and WMMSE2 once again. While WMMSE1 has a much higher sum multicast rate than of WMMSE2, WMMSE2 transmits at a higher common data rate than WMMSE1.

Finally, we find the algorithmic complexities for WMMSE1 and WMMSE2 in Table I using the techniques discussed in [42]. Note that, Table I does not show the complexity of searching for the optimal α for WMMSE1. Table I reveals that both WMMSE1 and WMMSE2 are cubic in M . For $M = K = L = 2$, WMMSE1 and WMMSE2 respectively require 1113 and 989 operations. For $M = K = 4, L = 2$, WMMSE1 and WMMSE2 respectively require 7033 and 5929 operations.

V. CONCLUSION

In this paper we study multi-group multicasting with a common message in a downlink MIMO broadcast channel. We assume the multicast groups are disjoint and we investigate the precoder design problem for maximum weighted sum rate. We first prove that weighted sum rate maximization problem is equivalent to the weighted minimum mean square error minimization problem. As both problems are non-convex and highly complex, we suggest a low-complexity, iterative precoder design algorithm inspired by the weighted minimum mean square error minimization problem. The algorithm iterates between receiver design for a given precoder, and then updates the precoder for a given receiver until convergence. We apply this algorithm on two transmission schemes. In the first scheme, the common message and private multicast messages are superposed. In the second scheme, the common message is appended to

the private multicast message vector. We show that the first scheme consistently performs better than the second transmission scheme in all settings. We understand that there is a fundamental difference between the two schemes. Although the second scheme is more general by definition, the first scheme introduces the freedom of power adaptation when employed within the proposed algorithm. Secondly, the first scheme favors multicast transmission, and the second scheme puts more emphasis on the common message. Future work includes studying the effect of successive cancellation order, and investigating the performance of the proposed schemes in a massive MIMO setting with imperfect channel state information at the transmitter.

APPENDIX A

Due to space limitations, in this appendix, we derive $\nabla_{\mathbf{p}_k^{(m)}} f(\mathbf{P}^{(m)}, t_k, z)$ for only $m = 1$. For notational convenience, we drop the upper indices (m). The Lagrangian objective function is given by

$$\begin{aligned}
 f(\mathbf{P}, t_k, z) = & - \sum_{k=1}^K t_k - z + \underbrace{\lambda \left(\alpha \text{Tr}(\mathbf{p}_A \mathbf{p}_A^H) + \bar{\alpha} \sum_{k=1}^K \text{Tr}(\mathbf{p}_k \mathbf{p}_k^H) - E_{tx} \right)}_C \\
 & + \underbrace{\sum_{k=1}^K \sum_{l=1}^L \mu_{l,k} (t_k - a_k R_{u,lk})}_A + \underbrace{\sum_{k=1}^K \sum_{l=1}^L \eta_{l,k} (z - b R_{c,lk})}_B. \quad (55)
 \end{aligned}$$

1) *Gradient of A:* To calculate the gradient of A, in (55) we need $\nabla_{\mathbf{p}_k} R_{u,lk}$ for both $i = k$ and $i \neq k$ for $l = 1, \dots, L$. Using (22) and (24), we have $\nabla_{\mathbf{p}_k} R_{u,lk} = (\nabla_{\mathbf{p}_k} \varepsilon_{u,lk}^{-1}) \varepsilon_{u,lk}$. Note that the noise variance $r_{u,lk}^{-1}$ is independent from \mathbf{p}_k , and $\nabla_{\mathbf{X}} (\mathbf{X}^H \mathbf{A} \mathbf{X}) = \mathbf{A} \mathbf{X}$ [43, ch E.3]. Then

$$\nabla_{[\mathbf{p}_k]_m} \varepsilon_{u,lk}^{-1} = \mathbf{e}_m^H \mathbf{h}_{l,k}^H r_{u,lk}^{-1} \mathbf{h}_{l,k} \mathbf{p}_k, \quad (56)$$

where \mathbf{e}_m is the unity column vector with 1 at the m^{th} element and zeros elsewhere and is size of $M \times 1$. As $[\nabla_{\mathbf{p}_k} R_{u,lk}]_m = \nabla_{[\mathbf{p}_k]_m} R_{u,lk} = \mathbf{e}_m^H \mathbf{h}_{l,k}^H r_{u,lk}^{-1} \mathbf{h}_{l,k} \mathbf{p}_k \varepsilon_{u,lk}$, we have

$$\nabla_{\mathbf{p}_k} R_{u,lk} = \mathbf{h}_{l,k}^H r_{u,lk}^{-1} \mathbf{h}_{l,k} \mathbf{p}_k \varepsilon_{u,lk}. \quad (57)$$

Next, we compute $\nabla_{\mathbf{p}_k} R_{u,li}$, for $i \neq k$ as

$$\nabla_{[\mathbf{p}_k]_m} R_{u,li} = \mathbf{p}_i^H \mathbf{h}_{l,i}^H \nabla_{[\mathbf{p}_k]_m} (r_{u,li}^{-1}) \mathbf{h}_{l,i} \mathbf{p}_i \varepsilon_{u,li}. \quad (58)$$

Using $\nabla_{\mathbf{X}} (\mathbf{X}^{-1}) = -\mathbf{X}^{-1} \nabla(\mathbf{X}) \mathbf{X}^{-1}$ [36], we can write

$$\nabla_{[\mathbf{p}_k]_m} (r_{u,li}^{-1}) = -r_{u,li}^{-1} \nabla_{[\mathbf{p}_k]_m} (r_{u,li}) r_{u,li}^{-1}. \quad (59)$$

Then we compute

$$\nabla_{[\mathbf{p}_k]_m}(r_{u,li}) = \mathbf{h}_{l,i}\mathbf{p}_k\bar{\alpha}\mathbf{e}_m^H\mathbf{h}_{l,i}^H. \quad (60)$$

By combining (58), (59) and (60) we have

$$\nabla_{[\mathbf{p}_k]_m}R_{u,li} = -\bar{\alpha}\mathbf{e}_m^H\mathbf{h}_{l,i}^H r_{u,li}^{-1}\mathbf{h}_{l,i}\mathbf{p}_i\varepsilon_{u,li}\mathbf{p}_i^H\mathbf{h}_{l,i}^H r_{u,li}^{-1}\mathbf{h}_{l,i}\mathbf{p}_k. \quad (61)$$

Overall, we have

$$\nabla_{\mathbf{p}_k}R_{u,li} = -\bar{\alpha}\mathbf{h}_{l,i}^H r_{u,li}^{-1}\mathbf{h}_{l,i}\mathbf{p}_i\varepsilon_{u,li}\mathbf{p}_i^H\mathbf{h}_{l,i}^H r_{u,li}^{-1}\mathbf{h}_{l,i}\mathbf{p}_k. \quad (62)$$

Using (57) and (62), we conclude that the gradient of A can be written as

$$\begin{aligned} \nabla_{\mathbf{p}_k}A &= -\sum_{l=1}^L \mu_{l,k}a_k\mathbf{h}_{l,k}^H r_{u,lk}^{-1}\mathbf{h}_{l,k}\mathbf{p}_k\varepsilon_{u,lk} \\ &\quad + \left(\sum_{i=1, i \neq k}^K \sum_{l=1}^L \bar{\alpha}\mu_{l,k}a_k\mathbf{h}_{l,i}^H r_{u,li}^{-1}\mathbf{h}_{l,i}\mathbf{p}_i\varepsilon_{u,li}\mathbf{p}_i^H\mathbf{h}_{l,i}^H r_{u,li}^{-1}\mathbf{h}_{l,i} \right) \mathbf{p}_k. \end{aligned} \quad (63)$$

2) *Gradient of B*: Secondly we compute the gradient of B defined in (55). Note that

$\nabla_{\mathbf{p}_k}R_{c,lk} = (\nabla\varepsilon_{c,lk}^{-1})\varepsilon_{c,lk}$. The gradient $\nabla_{[\mathbf{p}_k]_m}\varepsilon_{c,lk}^{-1}$ is calculated by applying the chain rule on $\nabla_{[\mathbf{p}_k]_m}\varepsilon_{c,lk}^{-1}$:

$$\begin{aligned} \nabla_{[\mathbf{p}_k]_m}\varepsilon_{c,lk}^{-1} &= \nabla_{[\mathbf{p}_k]_m} \left(\frac{1}{\alpha} + \mathbf{p}_A^H\mathbf{h}_{l,k}^H r_{c,lk}^{-1}\mathbf{h}_{l,k}\mathbf{p}_A \right) \\ &= \mathbf{p}_A^H\mathbf{h}_{l,k}^H \frac{\partial(r_{c,lk}^{-1})}{[\partial\mathbf{p}_k^*]_m}\mathbf{h}_{l,k}\mathbf{p}_A + \frac{\partial(\mathbf{p}_A^H\mathbf{h}_{l,k}^H)}{[\partial\mathbf{p}_k^*]_m}r_{c,lk}^{-1}\mathbf{h}_{l,k}\mathbf{p}_A + \mathbf{p}_A^H\mathbf{h}_{l,k}^H r_{c,lk}^{-1} \frac{\partial(\mathbf{h}_{l,k}\mathbf{p}_A)}{[\partial\mathbf{p}_k^*]_m} \\ &= -\mathbf{p}_A^H\mathbf{h}_{l,k}^H r_{c,lk}^{-1} \frac{\partial(r_{c,lk})}{[\partial\mathbf{p}_k^*]_m}r_{c,lk}^{-1}\mathbf{h}_{l,k}\mathbf{p}_A + \mathbf{e}_m^H\mathbf{h}_{l,k}^H r_{c,lk}^{-1}\mathbf{h}_{l,k}\mathbf{p}_A + 0. \end{aligned} \quad (64)$$

The last term is 0 since $\frac{\partial\mathbf{p}_k}{\partial\mathbf{p}_k^*} = 0$. Now we continue to compute

$$\begin{aligned} \nabla_{[\mathbf{p}_k]_m}r_{c,lk} &= \frac{\partial \left(\mathbf{h}_{l,k} \left(\sum_{i=1}^K \mathbf{p}_i\bar{\alpha}\mathbf{p}_i^H \right) \mathbf{h}_{l,k}^H + 1 \right)}{[\partial\mathbf{p}_k^*]_m} \\ &= \frac{\partial \left(\mathbf{h}_{l,k}\mathbf{p}_k\bar{\alpha}\mathbf{p}_k^H\mathbf{h}_{l,k}^H + \mathbf{h}_{l,k} \left(\sum_{i=1, i \neq k}^K \mathbf{p}_i\bar{\alpha}\mathbf{p}_i^H \right) \mathbf{h}_{l,k}^H + 1 \right)}{[\partial\mathbf{p}_k^*]_m} \\ &= \mathbf{h}_{l,k}\mathbf{p}_k\bar{\alpha}\mathbf{e}_m^H\mathbf{h}_{l,k}^H, \end{aligned} \quad (65)$$

as $\nabla_{\mathbf{X}}(\mathbf{XAX}^H) = \mathbf{XA}$ [43]. Then, we have

$$\nabla_{[\mathbf{p}_k]_m}\varepsilon_{c,lk}^{-1} = -\mathbf{p}_A^H\mathbf{h}_{l,k}^H r_{c,lk}^{-1}\mathbf{h}_{l,k}\mathbf{p}_k\bar{\alpha}\mathbf{e}_m^H\mathbf{h}_{l,k}^H r_{c,lk}^{-1}\mathbf{h}_{l,k}\mathbf{p}_A + \mathbf{e}_m^H\mathbf{h}_{l,k}^H r_{c,lk}^{-1}\mathbf{h}_{l,k}\mathbf{p}_A, \quad (66)$$

$$\nabla_{[\mathbf{p}_k]_m}R_{c,lk} = -\mathbf{e}_m^H\mathbf{h}_{l,k}^H r_{c,lk}^{-1}\mathbf{h}_{l,k}\mathbf{p}_A\varepsilon_{c,lk}\mathbf{p}_A^H\mathbf{h}_{l,k}^H r_{c,lk}^{-1}\mathbf{h}_{l,k}\mathbf{p}_k\bar{\alpha} + \mathbf{e}_m^H\mathbf{h}_{l,k}^H r_{c,lk}^{-1}\mathbf{h}_{l,k}\mathbf{p}_A\varepsilon_{c,lk}. \quad (67)$$

Finally, we obtain

$$\nabla_{\mathbf{p}_k} R_{c,lk} = -\mathbf{h}_{l,k}^H r_{c,lk}^{-1} \mathbf{h}_{l,k} \mathbf{p}_A \varepsilon_{c,lk} \mathbf{p}_A^H \mathbf{h}_{l,k}^H r_{c,lk}^{-1} \mathbf{h}_{l,k} \mathbf{p}_k \bar{\alpha} + \mathbf{h}_{l,k}^H r_{c,lk}^{-1} \mathbf{h}_{l,k} \mathbf{p}_A \varepsilon_{c,lk}, \quad (68)$$

and the gradient of B becomes

$$\begin{aligned} \nabla_{\mathbf{p}_k} B = & \left(\sum_{i=1}^K \sum_{l=1}^L \bar{\alpha} \eta_{l,i} b \mathbf{h}_{l,i}^H r_{c,li}^{-1} \mathbf{h}_{l,i} \mathbf{p}_A \varepsilon_{c,li} \mathbf{p}_A^H \mathbf{h}_{l,i}^H r_{c,li}^{-1} \mathbf{h}_{l,i} \right) \mathbf{p}_k \\ & - \sum_{i=1}^K \sum_{l=1}^L \eta_{l,i} b \mathbf{h}_{l,k}^H r_{c,lk}^{-1} \mathbf{h}_{l,k} \mathbf{p}_A \varepsilon_{c,lk}. \end{aligned} \quad (69)$$

3) *Gradient of C*: Thirdly, we compute the gradient of C defined in (55) as

$$\nabla_{\mathbf{p}_k} C = \lambda (\alpha \mathbf{p}_A + \bar{\alpha} \mathbf{p}_k), \quad (70)$$

since $\nabla_{\mathbf{p}_k} \text{Tr}(\mathbf{p}_A \mathbf{p}_A^H) = \mathbf{p}_A$. Finally combining (63), (69) and (70), we have

$$\begin{aligned} \nabla_{\mathbf{p}_k} f(\mathbf{P}, t_k, z) = & - \sum_{l=1}^L \mu_{l,k} a_k \mathbf{h}_{l,k}^H r_{u,lk}^{-1} \mathbf{h}_{l,k} \mathbf{p}_k \varepsilon_{u,lk} \\ & + \left(\sum_{i=1, i \neq k}^K \sum_{l=1}^L \bar{\alpha} \mu_{l,i} a_i \mathbf{h}_{l,i}^H r_{u,li}^{-1} \mathbf{h}_{l,i} \mathbf{p}_i \varepsilon_{u,li} \mathbf{p}_i^H \mathbf{h}_{l,i}^H r_{u,li}^{-1} \mathbf{h}_{l,i} \right) \mathbf{p}_k \\ & + \left(\sum_{i=1}^K \sum_{l=1}^L \bar{\alpha} \eta_{l,i} b \mathbf{h}_{l,i}^H r_{c,li}^{-1} \mathbf{h}_{l,i} \mathbf{p}_A \varepsilon_{c,li} \mathbf{p}_A^H \mathbf{h}_{l,i}^H r_{c,li}^{-1} \mathbf{h}_{l,i} \right) \mathbf{p}_k \\ & - \sum_{i=1}^K \sum_{l=1}^L \eta_{l,i} b \mathbf{h}_{l,i}^H r_{c,li}^{-1} \mathbf{h}_{l,i} \mathbf{p}_A \varepsilon_{c,li} + \lambda (\alpha \mathbf{p}_A + \bar{\alpha} \mathbf{p}_k). \end{aligned} \quad (71)$$

APPENDIX B

In this appendix, we prove Theorem 1. In subsections A and B, we do the derivations respectively for signal models 1 and 2. For notational convenience, we drop the upper index (m).

A. Derivations for Signal Model 1 ($m = 1$)

Taking the derivative of the objective function h in (44) with respect to $W_{l,k}$, then equating it to zero, we obtain the following equation

$$\psi_{l,k} w \alpha \mathbf{p}_A^H \mathbf{h}_{l,k}^H = \sum_{i=1}^K \psi_{l,k} w \bar{\alpha} \mathbf{h}_{l,k} \mathbf{p}_i \mathbf{p}_i^H \mathbf{h}_{l,k}^H W_{l,k} + \psi_{l,k} w \alpha \mathbf{h}_{l,k} \mathbf{p}_A \mathbf{p}_A^H \mathbf{h}_{l,k}^H W_{l,k} + \psi_{l,k} w W_{l,k}. \quad (72)$$

This leads to

$$W_{l,k} = \alpha \mathbf{p}_A^H \mathbf{h}_{l,k}^H \left(\sum_{i=1}^K \bar{\alpha} \mathbf{h}_{l,k} \mathbf{p}_i \mathbf{p}_i^H \mathbf{h}_{l,k}^H + \alpha \mathbf{h}_{l,k} \mathbf{p}_A \mathbf{p}_A^H \mathbf{h}_{l,k}^H + 1 \right)^{-1}. \quad (73)$$

Similarly, we take the derivative of (44) with respect to $V_{l,k}$ and equating it to zero, we have

$$\xi_{l,k} v_k \bar{\alpha} \mathbf{p}_k^H \mathbf{h}_{l,k}^H = \sum_{i=1}^K \xi_{l,k} v_k \mathbf{h}_{l,k} \mathbf{p}_i \bar{\alpha} \mathbf{p}_i^H \mathbf{h}_{l,k}^H V_{l,k} + \xi_{l,k} v_k V_{l,k}, \quad (74)$$

and this leads to

$$V_{l,k} = \bar{\alpha} \mathbf{p}_k^H \mathbf{h}_{l,k}^H \left(\sum_{i=1}^K \bar{\alpha} \mathbf{h}_{l,k} \mathbf{p}_i \mathbf{p}_i^H \mathbf{h}_{l,k}^H + 1 \right)^{-1}. \quad (75)$$

Then, taking the gradient of (44) with respect to \mathbf{p}_k , and equating it to zero, we have the following equation

$$\begin{aligned} & \sum_{l=1}^L \xi_{l,k} v_k \bar{\alpha} \mathbf{h}_{l,k}^H V_{l,k}^* + \sum_{i=1}^K \sum_{l=1}^L \psi_{l,i} w \alpha \mathbf{h}_{l,i}^H W_{l,i}^* \\ &= \sum_{i=1}^K \sum_{l=1}^L \xi_{l,i} v_i \bar{\alpha} \mathbf{h}_{l,i}^H V_{l,i}^* V_{l,i} \mathbf{h}_{l,i} \mathbf{p}_k + \sum_{i=1}^K \sum_{l=1}^L \psi_{l,i} w \bar{\alpha} \mathbf{h}_{l,i}^H W_{l,i}^* W_{l,i} \mathbf{h}_{l,i} \mathbf{p}_k \\ &+ \sum_{i=1}^K \sum_{l=1}^L \psi_{l,i} w \alpha \mathbf{h}_{l,i}^H W_{l,i}^* W_{l,i} \mathbf{h}_{l,i} \mathbf{p}_A + \beta \left(\alpha \mathbf{p}_A + \bar{\alpha} \mathbf{p}_k \right). \end{aligned} \quad (76)$$

Then,

$$\begin{aligned} \mathbf{p}_k &= \left(\beta \mathbf{I} + \sum_{i=1}^K \sum_{l=1}^L \xi_{l,i} v_i \bar{\alpha} \mathbf{h}_{l,i}^H V_{l,i}^* V_{l,i} \mathbf{h}_{l,i} + \sum_{i=1}^K \sum_{l=1}^L \psi_{l,i} w \mathbf{h}_{l,i}^H W_{l,i}^* W_{l,i} \mathbf{h}_{l,i} \right)^{-1} \\ &\times \left[\sum_{l=1}^L \xi_{l,k} v_k \bar{\alpha} \mathbf{h}_{l,k}^H V_{l,k}^* + \sum_{i=1}^K \sum_{l=1}^L \psi_{l,i} w \alpha \mathbf{h}_{l,i}^H W_{l,i}^* - \beta \alpha (\mathbf{p}_A - \mathbf{p}_k) \right. \\ &\left. - \sum_{i=1}^K \sum_{l=1}^L \psi_{l,i} w \alpha \mathbf{h}_{l,i}^H W_{l,i}^* W_{l,i} \mathbf{h}_{l,i} (\mathbf{p}_A - \mathbf{p}_k) \right]. \end{aligned} \quad (77)$$

To calculate β , we post-multiply both sides of (72) by $W_{l,k}^*$ and (74) by $V_{l,k}^*$, and perform

$\sum_{k=1}^K \sum_{l=1}^L$ on (72) and (74) on both sides. Then we obtain

$$\begin{aligned} \sum_{k=1}^K \sum_{l=1}^L \psi_{l,k} w \alpha \mathbf{p}_A^H \mathbf{h}_{l,k}^H W_{l,k}^* &= \sum_{k=1}^K \sum_{l=1}^L \sum_{i=1}^K \psi_{l,k} w \bar{\alpha} \mathbf{h}_{l,k} \mathbf{p}_i \mathbf{p}_i^H \mathbf{h}_{l,k}^H W_{l,k} W_{l,k}^* \\ &+ \sum_{k=1}^K \sum_{l=1}^L \psi_{l,k} w \alpha \mathbf{h}_{l,k} \mathbf{p}_A \mathbf{p}_A^H \mathbf{h}_{l,k}^H W_{l,k} W_{l,k}^* + \sum_{k=1}^K \sum_{l=1}^L \psi_{l,k} w W_{l,k} W_{l,k}^* \end{aligned} \quad (78)$$

$$\sum_{k=1}^K \sum_{l=1}^L \xi_{l,k} v_k \bar{\alpha} \mathbf{p}_k^H \mathbf{h}_{l,k}^H V_{l,k}^* = \sum_{k=1}^K \sum_{l=1}^L \sum_{i=1}^K \xi_{l,k} v_k \mathbf{h}_{l,k} \mathbf{p}_i \bar{\alpha} \mathbf{p}_i^H \mathbf{h}_{l,k}^H V_{l,k} V_{l,k}^* + \sum_{k=1}^K \sum_{l=1}^L \xi_{l,k} v_k V_{l,k} V_{l,k}^*. \quad (79)$$

The summation of (78) and (79) leads to

$$\begin{aligned} &\sum_{k=1}^K \sum_{l=1}^L \psi_{l,k} w \alpha \mathbf{p}_A^H \mathbf{h}_{l,k}^H W_{l,k}^* + \sum_{k=1}^K \sum_{l=1}^L \xi_{l,k} v_k \bar{\alpha} \mathbf{p}_k^H \mathbf{h}_{l,k}^H V_{l,k}^* \\ &= \sum_{k=1}^K \sum_{l=1}^L \sum_{i=1}^K \psi_{l,k} w \bar{\alpha} \mathbf{h}_{l,k} \mathbf{p}_i \mathbf{p}_i^H \mathbf{h}_{l,k}^H W_{l,k} W_{l,k}^* + \sum_{k=1}^K \sum_{l=1}^L \psi_{l,k} w \alpha \mathbf{h}_{l,k} \mathbf{p}_A \mathbf{p}_A^H \mathbf{h}_{l,k}^H W_{l,k} W_{l,k}^* \\ &+ \sum_{k=1}^K \sum_{l=1}^L \psi_{l,k} w W_{l,k} W_{l,k}^* + \sum_{k=1}^K \sum_{l=1}^L \sum_{i=1}^K \xi_{l,k} v_k \mathbf{h}_{l,k} \mathbf{p}_i \bar{\alpha} \mathbf{p}_i^H \mathbf{h}_{l,k}^H V_{l,k} V_{l,k}^* \\ &+ \sum_{k=1}^K \sum_{l=1}^L \xi_{l,k} v_k V_{l,k} V_{l,k}^*. \end{aligned} \quad (80)$$

On the other hand, pre-multiplying (76) with \mathbf{p}_k^H and summing over k from 1 to K , we have

$$\begin{aligned} &\sum_{k=1}^K \sum_{l=1}^L \xi_{l,k} v_k \bar{\alpha} \mathbf{p}_k^H \mathbf{h}_{l,k}^H V_{l,k}^* + \sum_{k=1}^K \sum_{i=1}^K \sum_{l=1}^L \psi_{l,i} w \alpha \mathbf{p}_k^H \mathbf{h}_{l,i}^H W_{l,i}^* \\ &= \sum_{k=1}^K \sum_{i=1}^K \sum_{l=1}^L \xi_{l,i} v_i \bar{\alpha} \mathbf{p}_k^H \mathbf{h}_{l,i}^H V_{l,i}^* V_{l,i} \mathbf{h}_{l,i} \mathbf{p}_k + \sum_{k=1}^K \sum_{i=1}^K \sum_{l=1}^L \psi_{l,i} w \bar{\alpha} \mathbf{p}_k^H \mathbf{h}_{l,i}^H W_{l,i}^* W_{l,i} \mathbf{h}_{l,i} \mathbf{p}_k \\ &+ \sum_{k=1}^K \sum_{i=1}^K \sum_{l=1}^L \psi_{l,i} w \alpha \mathbf{p}_k^H \mathbf{h}_{l,i}^H W_{l,i}^* W_{l,i} \mathbf{h}_{l,i} \mathbf{p}_A + \sum_{k=1}^K \mathbf{p}_k^H \beta \left(\alpha \mathbf{p}_A + \bar{\alpha} \mathbf{p}_k \right). \end{aligned} \quad (81)$$

Comparing the left sides of (80) and (81), we observe that they are equal. Then, the right sides are also equal to each other. As we assume that the power constraint in (2) is satisfied with equality, (80) = (81) leads to

$$\beta = \frac{1}{E_{tx}} \left[\sum_{k=1}^K \sum_{l=1}^L \xi_{l,k} v_k V_{l,k} V_{l,k}^* + \sum_{k=1}^K \sum_{l=1}^L \psi_{l,k} w W_{l,k} W_{l,k}^* \right]. \quad (82)$$

B. Derivations for Signal Model 2 ($m = 2$)

First, we would like to remind that in the second signal model, $\mathbf{p}_A = \mathbf{p}_c$. Taking the derivative of the objective function h in (44) with respect to $W_{l,k}$, then equating it to zero, we obtain the following equation

$$\psi_{l,k} w \mathbf{p}_c^H \mathbf{h}_{l,k}^H = \sum_{i=1}^K \psi_{l,k} w \mathbf{h}_{l,k} \mathbf{p}_i \mathbf{p}_i^H \mathbf{h}_{l,k}^H W_{l,k} + \psi_{l,k} w \mathbf{h}_{l,k} \mathbf{p}_c \mathbf{p}_c^H \mathbf{h}_{l,k}^H W_{l,k} + \psi_{l,k} w W_{l,k}. \quad (83)$$

This leads to

$$W_{l,k} = \mathbf{p}_c^H \mathbf{h}_{l,k}^H \left(\sum_{i=1}^K \mathbf{h}_{l,k} \mathbf{p}_i \mathbf{p}_i^H \mathbf{h}_{l,k}^H + \mathbf{h}_{l,k} \mathbf{p}_c \mathbf{p}_c^H \mathbf{h}_{l,k}^H + 1 \right)^{-1}. \quad (84)$$

Similarly, we take the derivative of (44) with respect to $V_{l,k}$ and equating it to zero, we have

$$\xi_{l,k} v_k \mathbf{p}_k^H \mathbf{h}_{l,k}^H = \sum_{i=1}^K \xi_{l,k} v_k \mathbf{h}_{l,k} \mathbf{p}_i \mathbf{p}_i^H \mathbf{h}_{l,k}^H V_{l,k} + \xi_{l,k} v_k V_{l,k}, \quad (85)$$

and this leads to

$$V_{l,k} = \mathbf{p}_k^H \mathbf{h}_{l,k}^H \left(\sum_{i=1}^K \mathbf{h}_{l,k} \mathbf{p}_i \mathbf{p}_i^H \mathbf{h}_{l,k}^H + 1 \right)^{-1}. \quad (86)$$

Taking the gradient of (44) with respect to \mathbf{p}_k and \mathbf{p}_c , and equating it to zero, we have the following equation

$$\sum_{l=1}^L \xi_{l,k} v_k \mathbf{h}_{l,k}^H V_{l,k}^* = \sum_{i=1}^K \sum_{l=1}^L \xi_{l,i} v_i \mathbf{h}_{l,i}^H V_{l,i}^* V_{l,i} \mathbf{h}_{l,i} \mathbf{p}_k + \sum_{i=1}^K \sum_{l=1}^L \psi_{l,i} w \mathbf{h}_{l,i}^H W_{l,i}^* W_{l,i} \mathbf{h}_{l,i} \mathbf{p}_k + \beta \mathbf{p}_k. \quad (87)$$

$$\sum_{k=1}^K \sum_{l=1}^L \psi_{l,k} w \mathbf{h}_{l,k}^H W_{l,k}^* = \sum_{k=1}^K \sum_{l=1}^L \psi_{l,k} w \mathbf{h}_{l,k}^H W_{l,k}^* W_{l,k} \mathbf{h}_{l,k} \mathbf{p}_c + \beta \mathbf{p}_c. \quad (88)$$

Then,

$$\mathbf{p}_k = \left(\beta \mathbf{I} + \sum_{i=1}^K \sum_{l=1}^L \xi_{l,i} v_i \mathbf{h}_{l,i}^H V_{l,i}^* V_{l,i} \mathbf{h}_{l,i} + \sum_{i=1}^K \sum_{l=1}^L \psi_{l,i} w \mathbf{h}_{l,i}^H W_{l,i}^* W_{l,i} \mathbf{h}_{l,i} \right)^{-1} \left(\sum_{l=1}^L \xi_{l,k} v_k \mathbf{h}_{l,k}^H V_{l,k}^* \right), \quad (89)$$

$$\mathbf{p}_c = \left(\beta \mathbf{I} + \sum_{i=1}^K \sum_{l=1}^L \psi_{l,i} w \mathbf{h}_{l,i}^H W_{l,i}^* W_{l,i} \mathbf{h}_{l,i} \right)^{-1} \left(\sum_{i=1}^K \sum_{l=1}^L \psi_{l,i} w \mathbf{h}_{l,i}^H W_{l,i}^* \right). \quad (90)$$

To calculate β for signal model 2, we followed the similar steps in the previous subsection and obtained the following expression,

$$\beta = \frac{1}{E_{tx}} \left[\sum_{k=1}^K \sum_{l=1}^L \xi_{l,k} v_k V_{l,k} V_{l,k}^* + \sum_{k=1}^K \sum_{l=1}^L \psi_{l,k} w W_{l,k} W_{l,k}^* \right]. \quad (91)$$

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments that improved the results.

REFERENCES

- [1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar 1998.
- [2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.
- [3] C. S. Vaze and M. K. Varanasi, "The degree-of-freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5354–5374, Aug 2012.
- [4] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, March 2006.
- [5] H. Weingarten, Y. Steinberg, and S. S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, Sept 2006.
- [6] M. Sharif and B. Hassibi, "A comparison of time-sharing, DPC, and beamforming for MIMO broadcast channels with many users," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 11–15, Jan 2007.
- [7] S. S. Christensen and R. Agarwal and E. D. Carvalho and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, December 2008.
- [8] R. Zhang and L. Hanzo, "A unified treatment of superposition coding aided communications: Theory and practice," *IEEE Communications Surveys Tutorials*, vol. 13, no. 3, pp. 503–520, Third 2011.
- [9] S. Vanka, S. Srinivasa, Z. Gong, P. Vizi, K. Stamatiou, and M. Haenggi, "Superposition coding strategies: Design and experimental evaluation," *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 2628–2639, July 2012.
- [10] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4847–4861, Nov 2016.
- [11] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *arXiv preprint arXiv:1406.1516*, 2014.
- [12] J. Seo, Y. Sung, and H. Jafarkhani, "A high-diversity transceiver design for K-user MISO broadcast channels," *arXiv preprint arXiv:1807.00114*, 2018.
- [13] A. Lee, "EMBMS delivers mobile video to the mass audience," 2015. [Online]. Available: <https://www.itu.int/en/ITU-D/Regional-Presence/AsiaPacific/Documents/Events/2015/August-MTV/S3B-Allan-Lee.pdf>
- [14] V. K. Thorsten Lohmar, Michael Sllsingar and S. Puustinen, "Delivering content with LTE broadcast," *Ericsson Review*, vol. 1, no. 11, February 2013.
- [15] 3GPP, "Multimedia broadcast/multicast service (MBMS); architecture and functional description," Tech. Rep. TS 23.246, V8.0.0, June 2007.
- [16] N. Jindal and Z. Q. Luo, "Capacity limits of multiple antenna multicast," in *2006 IEEE International Symposium on Information Theory*, July 2006, pp. 1841–1845.
- [17] A. Abdelkader and A. B. Gershman and N. D. Sidiropoulos, "Multiple-antenna multicasting using channel orthogonalization and local refinement," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3922–3927, July 2010.

- [18] E. Chiu and V. Lau, "Transmit precoding design for multi-antenna multicast broadcast services with limited feedback," in *2009 IEEE Wireless Communications and Networking Conference*, April 2009, pp. 1–6.
- [19] N. D. Sidiropoulos and T. N. Davidson and Zhi-Quan Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2239–2251, June 2006.
- [20] Y. C. B. Silva and A. Klein, "Adaptive beamforming and spatial multiplexing of unicast and multicast services," in *2006 IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2006, pp. 1–5.
- [21] S. Y. Baek, Y. J. Hong, and D. K. Sung, "Adaptive transmission scheme for mixed multicast and unicast traffic in cellular systems," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 6, pp. 2899–2908, July 2009.
- [22] A. Z. Yalcin, M. Yuksel, and I. Bahceci, "Downlink MU-MIMO with QoS aware transmission: Precoder design and performance analysis," *IEEE Transactions on Wireless Communications*, vol. 18, no. 2, pp. 969–982, Feb 2019.
- [23] Y. C. B. Silva and A. Klein, "Linear transmit beamforming techniques for the multigroup multicast scenario," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 8, pp. 4353–4367, Oct 2009.
- [24] Y. Gao and M. Schubert, "Group-oriented beamforming for multi-stream multicasting based on quality-of-service requirements," in *1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2005.*, Dec 2005, pp. 193–196.
- [25] E. Karipidis, N. D. Sidiropoulos, and Z. Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1268–1279, March 2008.
- [26] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Weighted fair multicast multigroup beamforming under per-antenna power constraints," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5132–5142, Oct 2014.
- [27] H. Joudeh and B. Clerckx, "Rate-splitting for max-min fair multigroup multicast beamforming in overloaded systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7276–7289, Nov 2017.
- [28] H. Zhou and M. Tao, "Joint multicast beamforming and user grouping in massive MIMO systems," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 1770–1775.
- [29] M. Sadeghi, E. Björnson, E. G. Larsson, C. Yuen, and T. L. Marzetta, "Max-min fair transmit precoding for multi-group multicasting in massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 1358–1373, Feb 2018.
- [30] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Multicast multigroup precoding and user scheduling for frame-based satellite communications," *IEEE Transactions on Wireless Communications*, vol. 14, no. 9, pp. 4695–4707, 2015.
- [31] M. Kaliszán, E. Pollakis, and S. Stańczak, "Multigroup multicast with application-layer coding: Beamforming for maximum weighted sum rate," in *Wireless Communications and Networking Conference (WCNC)*. IEEE, 2012, pp. 2270–2275.
- [32] E. Ekrem and S. Ulukus, "An outer bound for the Gaussian MIMO broadcast channel with common and private messages," *IEEE Transactions on Information Theory*, vol. 58, no. 11, pp. 6766–6772, 2012.
- [33] J. Kim, J. Koh, J. Kang, K. Lee, and J. Kang, "Design of user clustering and precoding for downlink non-orthogonal multiple access (NOMA)," in *MILCOM 2015 - 2015 IEEE Military Communications Conference*, Oct 2015, pp. 1170–1175.
- [34] K. Senel and S. Tekinay, "Optimal power allocation in NOMA systems with imperfect channel estimation," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–7.
- [35] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.
- [36] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," nov 2012, version 20121115. [Online]. Available: <http://www2.imm.dtu.dk/pubdb/p.php?3274>
- [37] S. P. X.S. Li, "Solving the finite min-max problem via an exponential penalty method," *Comput. Tech.*, vol. 8, pp. 3–15, 2003.

- [38] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” <http://cvxr.com/cvx>, Mar. 2014.
- [39] J. Kaleva, A. Tölli, and M. Juntti, “Decentralized sum rate maximization with QoS constraints for interfering broadcast channel via successive convex approximation,” *IEEE Transactions on Signal Processing*, vol. 64, no. 11, pp. 2788–2802, June 2016.
- [40] M. Joham, W. Utschick, and J. A. Nossek, “Linear transmit processing in MIMO communications systems,” *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2700–2712, Aug 2005.
- [41] N. Dao and Y. Sun, “User-selection algorithms for multiuser precoding,” *IEEE Transactions on Vehicular Technology*, vol. 59, no. 7, pp. 3617–3622, Sept 2010.
- [42] R. Hunger, *Floating Point Operations in Matrix-Vector Calculus*. Munich University of Technology, Inst. for Circuit Theory and Signal Processing, 2005.
- [43] T. Moon and W. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, 2000.