# MMSE based Two-stage Beamforming for Large-Scale Multi-user MISO Systems

Younghyun Jeon[†⋆], Changick Song[∘], Seungjoo Maeng[⋆], Myonghee Park[⋆] and Inkyu Lee[†], *Fellow, IEEE*
[†] School of Electrical Eng., Korea University, Seoul, Korea
[⋆] Samsung Electronics, Suwon, Korea
[∘] Dept. of Inf. & Commun. Eng., Korea National University of Transportation, Chungju, Korea
Email: {young-hyun, inkyu}@korea.ac.kr, {younghyun.jeon, myongh.park, sjmaeng}@samsung.com and c.song@ut.ac.kr

***Abstract*—In this paper, we study a joint spatial division multiplexing (JSDM) beamforming scheme which enables large-scale spatial multiplexing gains for massive MIMO downlink systems. In contrast to the conventional JSDM which employs a block diagonalization (BD) method as a pre-beamformer, we aim to maximize sum-rate by applying minimum-mean-squared error (MMSE) approaches when designing a pre-beamformer and a multi-user precoder sequentially. First, to suppress inter-group interference, we design the pre-beamformer which minimizes an upper bound of the sum mean-squared-error in the large-scale array regime. Then, to mitigate same-group interference, we present the multi-user precoder based on the weighted MMSE (WMMSE) optimization method, which requires the same channel state information overhead as the conventional JSDM. Through simulation results, we confirm that the proposed two-step beamforming method brings substantial performance gains in terms of sum-rate over the conventional JSDM schemes especially in low to medium signal-to-noise ratio (SNR) regime.**

## I. INTRODUCTION

Multi-input multi-output (MIMO) systems have drawn a lot of attention due to their great potential to achieve high throughput in various wireless communication systems [1]–[4]. In response to the explosive increase in data traffic for 5G wireless systems, wireless standards such as 3GPP have focused on a large scale array system, often called "massive MIMO", which enables high spectral efficiency by adopting a large number of transmitting antennas.

To fully exploit the massive MIMO gain, however, proper channel state information at transmitter (CSIT) should be available. In time division duplex (TDD) systems, the CSIT can be obtained through uplink training by utilizing the channel reciprocity. In contrast, in frequency division duplex (FDD) systems, the downlink training resources required to collect the CSIT may become prohibitively large as the number of antennas at a base-station (BS) increases.

To reduce the training overhead while maximizing the throughput in the FDD massive MIMO downlink systems, joint spatial division and multiplexing (JSDM) has been recently proposed in [5]. The main idea of the JSDM is to partition users into multiple groups each of which experience the same transmit antenna correlation, and then exploit the reduced dimension of the effective channel matrix by eleminating inter-group interference (IGI) based on long-term channel state information (CSI).

This work was supported by the National Research Foundation of Korea (NRF) funded by the Korea Government (MSIP) under Grant 2014R1A2A1A10049769.

The JSDM adopts a two-step beamforming strategy: 1) a pre-beamformer which takes responsible for group-wise user separation using block diagonalization (BD) methods to reduce the dimension of the effective channel, based on only the second order channel statistics, and 2) a multi-user precoding which manages same-group interference (SGI) to improve the performance. For the JSDM strategy, the authors in [6] proposed several user selection and scheduling schemes. In addition, the JSDM algorithm was extended in [7] to some realistic propagation channels such as the partial overlap of the angular spectra from different users and the sparsity of mm-Wave channels.

Conventional JSDM schemes confront the limitation due to a noise enhancement issue caused by the BD scheme and thus the sum-rate of the conventional JSDM is degraded in low and medium signal-to-noise ratio (SNR) regime. Our goal is to improve the sum-rate performance while maintaining the required CSI feedback and the computational complexity comparable to the original JSDM. To this end, we first suppress IGI by adopting the regularized BD (RBD) scheme based on statistical CSI. We note that authors in [8] formulated the regularization problem for multi-user MIMO channels in order to handle multi-user interference (MUI). However, as the RBD precoder in [8] does not exploit the channel covariance $\mathbf{R}_g$ (long-term CSIT) and requires every users' short-term CSIT, it may not be suitable for FDD massive MIMO systems. In this paper, we propose a new RBD pre-beamforming scheme that can be implemented with the knowledge of only the second order channel statistics such as channel covariance by utilizing the results of large system limit.

Then, to mitigate SGI and thereby maximize the sum-rate performance, we present a multi-user precoder using the weighted MMSE (WMMSE) optimization method, which requires the same CSI overhead as the conventional JSDM. As a result, the proposed methods improve the received signal-to-interference-plus-noise ratio (SINR) at each user in low to medium SNR regime. Through simulation results, we demonstrate that the proposed two-step beamforming scheme shows a sum-rate gain over existing methods in various configurations.

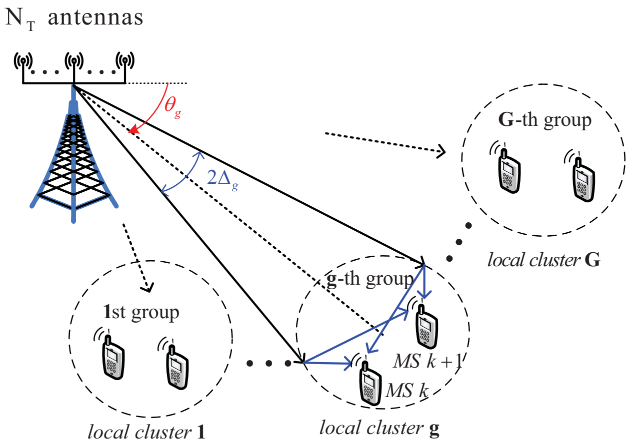


Fig. 1. Multi-group large-scale MISO downlink systems with a one-ring local scattering model

The rest of this paper is organized as follows: The system model is described in Section II. In Section III, the pre-beamformer is derived based on the MMSE approach in the large scale array regime. Section IV presents an iterative WMMSE algorithm with reduced CSI overhead. Section V confirms performance gains of the proposed schemes through simulation results. Finally, in Section VI, this paper is terminated with conclusions.

Throughout the paper, we adopt uppercase boldface letters for matrices and lowercase boldface for vectors. The superscripts $(\cdot)^{\mathsf{T}}$ and $(\cdot)^{\mathsf{H}}$ stand for transpose and conjugate transpose, respectively. In addition, $\|\cdot\|$, $\mathsf{Tr}(\cdot)$, $[\cdot]_k$ and $[\cdot]_{ij}$ represent 2-norm, trace, the $k$-th element of a vector and the $(i,j)$-th entry of a matrix, respectively. Also, $\mathbf{I}_d$ denotes an identity matrix of size $d$. A set of $N$ dimensional complex column vectors is defined by $\mathbb{C}^N$.

## II. System Model

We consider a multiuser FDD MISO downlink systems where a BS equipped with $N_T$ large antennas serves $K$ single antenna users. Let $\mathbf{x} \in \mathbb{C}^{N_T \times 1}$ be the transmit signal vector at the BS and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ be the additive white Gaussian noise (AWGN) at the users. The transmit signal vector $\mathbf{x}$ is obtained from $\mathbf{x} = \mathbf{Fd}$ where $\mathbf{F} \in \mathbb{C}^{N_T \times S}$ and $\mathbf{d} = [d_1, \ldots, d_S]^T \in \mathbb{C}^{S \times 1}$ represent the linear precoding matrix and the data symbol vector with $E[|d_k|^2] = 1$ for $k = 1, \ldots, S$, respectively.

Then, the received signal vector $\mathbf{y} \in \mathbb{C}^{K \times 1}$ at $K$ users can be expressed by

$$\mathbf{y} = \mathbf{H}^{\mathsf{H}}\mathbf{x} + \mathbf{n} = \mathbf{H}^{\mathsf{H}}\mathbf{Fd} + \mathbf{n}, \tag{1}$$

where $\mathbf{H} \triangleq [\mathbf{h}_1, \ldots, \mathbf{h}_K] \in \mathbb{C}^{N_T \times K}$ with $\mathbf{h}_k \in \mathbb{C}^{N_T \times 1}$ being the channel vector from the BS to the $k$-th user. Also, the precoding matrix $\mathbf{F}$ must satisfy

$$\mathsf{Tr}\left(\mathbf{FF}^{\mathsf{H}}\right) \le P_T, \tag{2}$$

where $P_T$ denotes the total power budget at the BS. Assuming the Rayleigh correlated channel, we have $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_k)$ where $\mathbf{R}_k$ is a positive semi-definite channel covariance matrix whose eigenvalue decomposition is given by $\mathbf{R}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^{\mathsf{H}}$ with a tall unitary matrix $\mathbf{U}_k \in \mathbb{C}^{N_T \times r_k}$ and a diagonal matrix $\mathbf{\Lambda}_k \in \mathbb{R}^{r_k \times r_k}$ with $r_k$ non-zero positive eigenvalues. Thus, the $k$-th channel vector $\mathbf{h}_k$ can be written without loss of generality by

$$\mathbf{h}_k = \mathbf{U}_k \mathbf{\Lambda}_k^{1/2} \mathbf{w}_k, \tag{3}$$

where $\mathbf{w}_k \in \mathbb{C}^{r_k \times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{r_k})$.

As shown in Fig. 1, we consider the one-ring scattering model [9], i.e., each user in the $g$-th cluster with azimuth center angle $\theta_g$ and angular spread $\Delta_g$ experiences the channel covariance matrix whose element is given as

$$[\mathbf{R}_g]_{m,n} = \frac{1}{2\Delta_g} \int_{\theta_g - \Delta_g}^{\theta_g + \Delta_g} e^{-j2\pi \frac{D}{\lambda_c}(m-n)\sin\theta} d\theta, \tag{4}$$

where $\lambda_c$ is the carrier wavelength and $D$ represents the space between adjacent antennas. Thus, denoting $K_g$ by the number of users in group $g$, total $K = \sum_{g=1}^{G} K_g$ users in a cell can be separated into $G$ groups according to the channel covariance matrices $\{\mathbf{R}_g\}_{g=1}^{G}$. Note that $\mathbf{R}_g$ is assumed to be perfectly known to the BS and do not change frequently as a long-term CSI. For simplicity, we assume that all users in the same group $g$ have the same covariance matrix $\mathbf{R}_g = \mathbf{U}_g \mathbf{\Lambda}_g \mathbf{U}_g^{\mathsf{H}}$, with rank $r_g$ and $r_g^* \le r_g$ dominant eigenvalues.

We adopt the two-stage beamforming strategy with pre-beamformers $\{\mathbf{Q}_g \in \mathbb{C}^{N_T \times \bar{B}}\}$ for spatial division (group separation) and multi-user precoders $\{\mathbf{V}_g \in \mathbb{C}^{\bar{B} \times S_g}\}$ for spatial multiplexing in each group [5] for $g = 1, ..., G$. Here, $\bar{B}$ denotes a design parameter that determines the effective channel dimension seen by the multi-user precoder $\mathbf{V}_g$ with $S_g \le \bar{B} \le r_g^*$ and $B = \sum_{g=1}^{G} \bar{B}$, and $S_g$ stands for the number of data streams for group $g$ with $S = \sum_{g=1}^{G} S_g$. Thus, the effective precoder is given by $\mathbf{F} = \mathbf{QV}$ where $\mathbf{Q} = [\mathbf{Q}_1, \ldots, \mathbf{Q}_G] \in \mathbb{C}^{N_T \times B}$ and $\mathbf{V} = diag\{\mathbf{V}_1, \ldots, \mathbf{V}_G\}$.

When the BS employs massive antenna arrays, the required feedback or the estimation overhead may become prohibitively large. To resolve the problem, we design the pre-beamforming matrix $\mathbf{Q}_g$ which depends only on the statistical CSI, i.e., the set of second order statistics $\{\mathbf{U}_g, \mathbf{\Lambda}_g,\ g = 1, \ldots, G\}$. In contrast, the multi-user precoder $\mathbf{V}_g$ is allowed to exploit the effective (instantaneous) CSIT.

Let us define the channel matrix of group $g$ as $\mathbf{H}_g = [\mathbf{h}_{g_1}, \mathbf{h}_{g_2}, \cdots, \mathbf{h}_{g_{K_g}}]$ where the index $g_k = \sum_{g'=1}^{g-1} K_{g'} + k$ indicates the $k$-th user in group $g$ for $k = 1, \ldots, K_g$. Then, the overall channel matrix $\mathbf{H}$ is expressed by $\mathbf{H} = \left[\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_G\right]$. For given channel correlation matrices $\{\mathbf{R}_g\}$, the dimension of the stacked channel matrix $\mathbf{H}$ in (1) can be reduced to $\bar{\mathbf{H}}^{\mathsf{H}} = \mathbf{H}^{\mathsf{H}}\mathbf{Q}$ as

$$\bar{\mathbf{H}}^{\mathsf{H}} = \begin{bmatrix} \mathbf{H}_1^{\mathsf{H}}\mathbf{Q}_1 & \mathbf{H}_1^{\mathsf{H}}\mathbf{Q}_2 & \cdots & \mathbf{H}_1^{\mathsf{H}}\mathbf{Q}_G \\ \mathbf{H}_2^{\mathsf{H}}\mathbf{Q}_1 & \mathbf{H}_2^{\mathsf{H}}\mathbf{Q}_2 & \cdots & \mathbf{H}_2^{\mathsf{H}}\mathbf{Q}_G \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_G^{\mathsf{H}}\mathbf{Q}_1 & \mathbf{H}_G^{\mathsf{H}}\mathbf{Q}_2 & \cdots & \mathbf{H}_G^{\mathsf{H}}\mathbf{Q}_G \end{bmatrix}. \tag{5}$$

The CSI of the whole effective channel $\bar{\mathbf{H}}$ is difficult to obtain at the BS in FDD systems due to the large feedback overhead. Hence, a more practical approach would be to estimate and feed back only $G$ diagonal blocks $\bar{\mathbf{H}}_g = \mathbf{H}_g^{\mathsf{H}}\mathbf{Q}_g$ of size $K_g \times \bar{B}$ for group $g = 1, \dots, G$ and treating each group separately. In this case, IGI caused by $\mathbf{H}_i^{\mathsf{H}}\mathbf{Q}_j$ with $i \neq j$ must be properly taken care of.

For given beamforming vectors, the received signal vector for users in group $g$ is expressed by

$$\begin{aligned} \mathbf{y}_g &= [y_{g_1}, \dots, y_{g_{K_g}}]^T \\ &= \mathbf{H}_g^{\mathsf{H}}\mathbf{Q}_g\mathbf{V}_g\mathbf{d}_g + \mathbf{H}_g^{\mathsf{H}}\sum_{c\neq g}\mathbf{Q}_c\mathbf{V}_c\mathbf{d}_c + \mathbf{n}_g, \end{aligned} \tag{6}$$

where $\mathbf{d}_g = [d_1, \dots, d_{g_{K_g}}]^T \in \mathbb{C}^{S_g \times 1}$ and $\mathbf{n}_g = [n_1, \dots, n_{g_{K_g}}]^T \in \mathbb{C}^{K_g \times 1}$ indicate the $g$-th subvector of $\mathbf{d} = [\mathbf{d}_1, \dots, \mathbf{d}_G]$ and $\mathbf{n} = [\mathbf{n}_1, \dots, \mathbf{n}_G]$, respectively. Denoting $\mathbf{V}_g = [\mathbf{v}_{g_1}, \dots, \mathbf{v}_{g_{K_g}}]$, the received signal of user $k$ in group $g$ is obtained by

$$\begin{aligned} y_{g_k} = \quad & \mathbf{h}_{g_k}^{\mathsf{H}}\mathbf{Q}_g\mathbf{v}_{g_k}d_{g_k} + \mathbf{h}_{g_k}^{\mathsf{H}}\sum_{j\neq k}\mathbf{Q}_g\mathbf{v}_{g_j}d_{g_j} \\ + \quad & \mathbf{h}_{g_k}^{\mathsf{H}}\sum_{c\neq g}\sum_{j}\mathbf{Q}_c\mathbf{v}_{c_j}d_{c_j} + n_{g_k} \end{aligned} \tag{7}$$

where the second and third terms in (7) stand for SGI and IGI, respectively.

Assuming an infinite length Gaussian codeword at the BS and single user detection at the receiver, the sum-rate $R_\Sigma$ is given by

$$R_\Sigma = \sum_{g=1}^{G}\sum_{k=1}^{K_g}\log_2\left(1 + \mathsf{sinr}_{g_k}(\{\mathbf{Q}_g\}, \{\mathbf{v}_{g_k}\})\right). \tag{8}$$

where $\mathsf{sinr}_{g_k}(\{\mathbf{Q}_g\}, \{\mathbf{v}_{g_k}\})$ denotes $\frac{|\mathbf{h}_{g_k}^{\mathsf{H}}\mathbf{Q}_g\mathbf{v}_{g_k}|^2}{\sum\limits_{(c,j)\neq(g,k)}|\mathbf{h}_{g_k}^{\mathsf{H}}\mathbf{Q}_c\mathbf{v}_{c_j}|^2+\sigma^2}$ which indicates the individual SINR for given pre-beamformer $\{\mathbf{Q}_g\}$ and multi-user precoder $\{\mathbf{v}_{g_k}\}$.

In this work, our goal is to find the beamforming matrices $\{\mathbf{Q}_g\}$ and $\{\mathbf{V}_g\}$ with sum transmit power constraint and thus the problem can be mathematically formulated as

$$\max_{\{\mathbf{Q}_g\},\{\mathbf{v}_{g_k}\}} R_\Sigma \quad \text{s.t.} \quad \sum_{g=1}^{G}\sum_{k=1}^{K_g}|\mathbf{Q}_g\mathbf{v}_{g_k}|^2 = P_T \tag{9}$$

Note that the pre-beamformers $\{\mathbf{Q}_g\}$ play a role of group-wise spatial separation and dimensional reduction of the channel matrix, while the multi-user precoders $\{\mathbf{V}_g\}$ enhance the throughput in each group. Thus, we aim to find efficient solutions for JSDM to maximize the sum rate, while maintaining the training and feedback overheads as low as possible.

## III. Proposed MMSE-based Pre-beamformer Designs

For group-wise user separation in JSDM, the authors in [5] adopted the BD method. However, the BD-based pre-beamformer, which focuses only on interference nulling, typically raises a noise enhancement problem. To tackle the problem, a regularized BD (RBD) scheme has been investigated in conventional multi-user MIMO systems to suppress the noise power as well as the interference power [8] [10] . In this section, we introduce a new RBD solution in the context of the JSDM and show that the proposed scheme requires only the second order statistics of channel, i.e., long-term CSI $\mathbf{R}_g$ in the large scale array regime.

Let $\mathbf{Q}_{\mathrm{BD},g}$ be the BD-based beamformer proposed in [5] for the $g$-th group users. Then, the proposed RBD beamformer $\mathbf{Q}_g$ can be achieved by solving the following MMSE optimization problem

$$\begin{aligned} &\min_{\mathbf{Q}_g,\beta}\sum_{c=1}^{G}\mathbb{E}[\|\mathbf{H}_c^{\mathsf{H}}\mathbf{Q}_{\mathrm{BD},c}\mathbf{V}_c\mathbf{d}_c - \frac{1}{\beta}\mathbf{y}_c\|^2] \\ &s.t. \sum_{c=1}^{G}\mathsf{Tr}(\mathbf{V}_c^{\mathsf{H}}\mathbf{Q}_c^{\mathsf{H}}\mathbf{Q}_c\mathbf{V}_c) = P_T \end{aligned} \tag{10}$$

where $\mathbf{H}_c^{\mathsf{H}}\mathbf{Q}_{\mathrm{BD},c}\mathbf{V}_c$ represents the target channel matrix for the $c$-th group users, $\beta$ indicates a positive scaling factor, and $\mathbf{y}_c$ is defined in (6). The resulting solution of problem (10) mitigates the interference plus noise power while making the desired signals as close as possible to the BD solution.

To solve this problem, we first convert it into an unconstrained minimization problem. Setting $\mathbf{Q}_g = \beta\bar{\mathbf{Q}}_g$, the scaling parameter $\beta$ equals

$$\beta = \sqrt{\frac{P_T}{\sum_{c=1}^{G}\mathsf{Tr}(\mathbf{V}_c^{\mathsf{H}}\bar{\mathbf{Q}}_c^{\mathsf{H}}\bar{\mathbf{Q}}_c\mathbf{V}_c)}}. \tag{11}$$

Applying this to the cost function in (10), the unconstrained MSE minimization problem on $\bar{\mathbf{Q}}_g$ is written by $\hat{\bar{\mathbf{Q}}}_g = \arg\min_{\bar{\mathbf{Q}}_g} f(\bar{\mathbf{Q}}_g)$ where

$$\begin{aligned} f(\bar{\mathbf{Q}}_g) \;=\; & \sum_{c=1}^{G}\mathbb{E}\Big[\big\|\mathbf{H}_c^{\mathsf{H}}\mathbf{Q}_{\mathrm{BD},c}\mathbf{V}_c\mathbf{d}_c - \sum_{j=1}^{G}\mathbf{H}_c^{\mathsf{H}}\bar{\mathbf{Q}}_j\mathbf{V}_j\mathbf{d}_j \\ & - \frac{1}{\sqrt{P_T}}\big(\sum_{j=1}^{G}\mathsf{Tr}(\mathbf{V}_j^{\mathsf{H}}\bar{\mathbf{Q}}_j^{\mathsf{H}}\bar{\mathbf{Q}}_j\mathbf{V}_j)\big)^{1/2}\mathbf{n}_c\big\|^2\Big]. \end{aligned}$$

Adopting the trace-norm transformation and ruling out the constant terms with respect to $\bar{\mathbf{Q}}_g$, it follows that $\hat{\bar{\mathbf{Q}}}_g = \arg\min_{\bar{\mathbf{Q}}_g} \bar{f}(\bar{\mathbf{Q}}_g)$ where

$$\begin{aligned} &\bar{f}(\bar{\mathbf{Q}}_g) = \\ &\mathsf{Tr}\big(\sum_{c=1}^{G}\mathbf{H}_c^{\mathsf{H}}\bar{\mathbf{Q}}_g\mathbf{V}_g\mathbf{V}_g^{\mathsf{H}}\bar{\mathbf{Q}}_g^{\mathsf{H}}\mathbf{H}_c - \mathbf{H}_g^{\mathsf{H}}\bar{\mathbf{Q}}_g\mathbf{V}_g\mathbf{V}_g^{\mathsf{H}}\mathbf{Q}_{\mathrm{BD},g}^{\mathsf{H}}\mathbf{H}_g \\ &- \mathbf{H}_g^{\mathsf{H}}\mathbf{Q}_{\mathrm{BD},g}\mathbf{V}_g\mathbf{V}_g^{\mathsf{H}}\bar{\mathbf{Q}}_g^{\mathsf{H}}\mathbf{H}_g + \frac{K\sigma^2}{P_T}\mathbf{V}_g^{\mathsf{H}}\bar{\mathbf{Q}}_g^{\mathsf{H}}\bar{\mathbf{Q}}_g\mathbf{V}_g\big). \end{aligned} \tag{12}$$

We now observe that the cost function (12) still contains the multi-user precoders $\mathbf{V}_g$. In fact, this may cause a difficulty in finding the optimal solution of $\bar{\mathbf{Q}}_g$, because the solutions for the pre-beamformer and subsequent precoder will eventually be inter-connected, which deviates from the goal of the JSDM strategy. To avoid this, we formulate the MSE upper-bound minimization problem whose solution is independent of $\mathbf{V}_g$. Before we proceed further, let us introduce the following two lemmas that will be useful for our derivations.

*Lemma 1:* ( [11]) For two positive semi-definite matrices $\mathbf{A}$ and $\mathbf{B}$, the following inequality holds

$$\mathsf{Tr}(\mathbf{AB})^m \leq \sqrt{\mathsf{Tr}(\mathbf{A})^{2m}\mathsf{Tr}(\mathbf{B})^{2m}}$$

where $m$ is a positive integer.

*Lemma 2 ( [12]):* Let $\mathbf{A}_1, \mathbf{A}_2, \cdots$, with $\mathbf{A}_N \in \mathbb{C}^{N\times N}$, be a series of matrices with uniformly bounded spectral norm. Let $\mathbf{x}_1, \mathbf{x}_2, \cdots$, with $\mathbf{x}_N \in \mathbb{C}^N$, be random vectors with independent and identically distributed (i.i.d.) entries of zero mean, variance $1/N$, and the eighth order moment of order $\mathcal{O}(1/N^4)$, independent of $\mathbf{A}_N$. Then, we almost surely have

$$\mathbf{x}_N^{\mathsf{H}}\mathbf{A}_N\mathbf{x}_N - \frac{1}{N}\mathsf{Tr}(\mathbf{A}_N) \overset{a.s.}{\to} 0, \tag{13}$$

as $N \to \infty$.

First, from Lemma 1, the MSE in (12) is upper-bounded by $\bar{f}(\bar{\mathbf{Q}}_g) \leq f_u(\bar{\mathbf{Q}}_g)$ where

$$\begin{aligned}
&f_u(\bar{\mathbf{Q}}_g)\\
&= \mathsf{Tr}\Big(\sum_{c=1}^{G}\mathbf{H}_c^{\mathsf{H}}\bar{\mathbf{Q}}_g\bar{\mathbf{Q}}_g^{\mathsf{H}}\mathbf{H}_c - \mathbf{H}_g^{\mathsf{H}}\bar{\mathbf{Q}}_g\mathbf{Q}_{\mathrm{BD},g}^{\mathsf{H}}\mathbf{H}_g\\
&\qquad - \mathbf{H}_g^{\mathsf{H}}\mathbf{Q}_{\mathrm{BD},g}\bar{\mathbf{Q}}_g^{\mathsf{H}}\mathbf{H}_g + \frac{K\sigma^2}{P_T}\bar{\mathbf{Q}}_g^{\mathsf{H}}\bar{\mathbf{Q}}_g\Big)\mathsf{Tr}(\mathbf{V}_g\mathbf{V}_g^{\mathsf{H}})\\
&= \zeta\mathsf{Tr}\Big(\sum_{c=1}^{G}\sum_{j=1}^{K_c}\mathbf{w}_{c_j}^{\mathsf{H}}\mathbf{\Lambda}_c^{1/2}\mathbf{U}_c^{\mathsf{H}}\bar{\mathbf{Q}}_g\bar{\mathbf{Q}}_g^{\mathsf{H}}\mathbf{U}_c\mathbf{\Lambda}_c^{1/2}\mathbf{w}_{c_j}\\
&\qquad - \sum_{j=1}^{K_g}\mathbf{w}_{g_j}^{\mathsf{H}}\mathbf{\Lambda}_g^{1/2}\mathbf{U}_g^{\mathsf{H}}\bar{\mathbf{Q}}_g\mathbf{Q}_{\mathrm{BD},g}^{\mathsf{H}}\mathbf{U}_g\mathbf{\Lambda}_g^{1/2}\mathbf{w}_{g_j}\\
&\qquad - \sum_{j=1}^{K_g}\mathbf{w}_{g_j}^{\mathsf{H}}\mathbf{\Lambda}_g^{1/2}\mathbf{U}_g^{\mathsf{H}}\mathbf{Q}_{\mathrm{BD},g}\bar{\mathbf{Q}}_g^{\mathsf{H}}\mathbf{U}_g\mathbf{\Lambda}_g^{1/2}\mathbf{w}_{g_j} + \frac{K\sigma^2}{P_T}\bar{\mathbf{Q}}_g^{\mathsf{H}}\bar{\mathbf{Q}}_g\Big)
\end{aligned}$$

where $\zeta$ is equal to $\mathsf{Tr}(\mathbf{V}_g\mathbf{V}_g^{\mathsf{H}})$.

Next, we apply Lemma 2 in the large scale array regime, i.e., $N_T \to \infty$. Then, it is easily shown that $f_u(\bar{\mathbf{Q}}_g) - g_u(\bar{\mathbf{Q}}_g) \overset{a.s.}{\to} 0$ where

$$\begin{aligned}
g_u(\bar{\mathbf{Q}}_g) = {}& \zeta\mathsf{Tr}\Big(\sum_{c=1}^{G}K_c\bar{\mathbf{Q}}_g^{\mathsf{H}}\mathbf{R}_c\bar{\mathbf{Q}}_g - K_g\mathbf{Q}_{\mathrm{BD},g}^{\mathsf{H}}\mathbf{R}_g\bar{\mathbf{Q}}_g\\
&- K_g\bar{\mathbf{Q}}_g^{\mathsf{H}}\mathbf{R}_g\mathbf{Q}_{\mathrm{BD},g} + \frac{K\sigma^2}{P_T}\bar{\mathbf{Q}}_g^{\mathsf{H}}\bar{\mathbf{Q}}_g\Big). 
\end{aligned}\tag{14}$$

Thus, by taking a derivative of (14) with respect to $\bar{\mathbf{Q}}_g$ and setting it to zero, we have

$$\hat{\bar{\mathbf{Q}}}_g = K_g\Big(\sum_{c=1}^{G}K_c\mathbf{R}_c + \frac{K\sigma^2}{P_T}\mathbf{I}_{N_T}\Big)^{-1}\mathbf{R}_g\mathbf{Q}_{\mathrm{BD},g}. \tag{15}$$

Finally, a unitary RBD solution $\mathbf{Q}_g$ is obtained through the QR decomposition

$$\hat{\bar{\mathbf{Q}}}_g = \mathbf{Q}_g\tilde{\mathbf{R}}_g \tag{16}$$

where $\tilde{\mathbf{R}}_g$ is a $\bar{B}\times\bar{B}$ upper triangular matrix and the matrix $\mathbf{Q}_g$ is composed of $\bar{B}$ orthonormal basis vectors of $\hat{\bar{\mathbf{Q}}}_g$. In comparison to the BD methods, each orthonormal vector of $\mathbf{Q}_g$ takes the noise into account, and thus the proposed RBD pre-beamformer is able to overcome the noise enhancement problem, which gives rise to a performance gain in the low to medium SNR region.

## IV. WMMSE Multi-user Precoding Schemes

In this section, we design the multi-user precoders $\{\mathbf{V}_g\}_{g=1}^G$ based on the WMMSE criteria to maximize the sum-rate performance for given the pre-beamformers $\{\mathbf{Q}_g\}_{g=1}^G$. We introduce a multi-user preocder based on WMMSE designs that can be applied to the JSDM, namely, per-group processing based WMMSE (PGP-WMMSE). The scheme is studied because both the approximate BD pre-beamformer in [5] and the proposed RBD pre-beamformer generate residual IGI in addition to SGI and noise power. Thus, the RZF post-beamformer in [5] is not an optimal multi-user precoder in terms of sum-rate, because the RZF post-beamformer optimizes the normalization and regularization factors with respect to only SGI and noise power under the assumption of perfect nulling of IGI. However, based on the the equivalent structure between sum-rate maximization and weighted MMSE minimization, the proposed PGP-WMMSE scheme considers residual IGI as well as SGI and noise and optimize the beamforming parameters $s_{g_k}$ and $u_{g_k}$, relevant to beam direction and power, to achieve a locally optimal sum-rate in an iterative manner.

For the derivation of PGP-WMMSE scheme, using the expression in (7), we define the MSE at user $g_k$ as

$$\begin{aligned}
e_{g_k} =& \mathbb{E}_{d,n}\left[\left(\frac{u_{g_k}}{\gamma}y_{g_k} - d_{g_k}\right)\left(\frac{u_{g_k}}{\gamma}y_{g_k} - d_{g_k}\right)^{\mathsf{H}}\right]\\
=& \left|\frac{u_{g_k}}{\gamma}\mathbf{h}_{g_k}^{\mathsf{H}}\mathbf{Q}_g\mathbf{v}_{g_k} - 1\right|^2 + \sum_{j=1,j\neq k}^{K_g}\left|\frac{u_{g_k}}{\gamma}\mathbf{h}_{g_k}^{\mathsf{H}}\mathbf{Q}_g\mathbf{v}_{g_j}\right|^2\\
&+ \sum_{c\neq g}\sum_{j=1}^{K_c}\left|\frac{u_{g_k}}{\gamma}\mathbf{h}_{g_k}^{\mathsf{H}}\mathbf{Q}_c\mathbf{v}_{c_j}\right|^2 + \left|\frac{u_{g_k}}{\gamma}\right|^2\sigma^2,
\end{aligned}\tag{17}$$

where $\gamma$ indicates a scaling factor and $u_{g_k}$ denotes a scalar MMSE weight for user $g_k$.

Note that since the RBD pre-beamformers $\{\mathbf{Q}_g\}$ have already mitigated IGI, the residual IGI term $\sum_{c\neq g}\sum_{j=1}^{K_c}\left|\mathbf{h}_{g_k}^{\mathsf{H}}\mathbf{Q}_c\mathbf{v}_{c_j}\right|^2$ in (17) may not be significant especially at high SNR. Therefore, in the average sense, we can replace with its expected value $\mathbb{E}_{\mathbf{h}_{g_k}}\left[\sum_{c\neq g}\sum_{j-1}^{K_c}\left|\mathbf{h}_{g_k}^{\mathsf{H}}\mathbf{Q}_c\mathbf{v}_{c_j}\right|^2\right] = \sum_{c\neq g}\sum_{j-1}^{K_c}\mathbf{v}_{c_j}^{\mathsf{H}}\mathbf{Q}_c^{\mathsf{H}}\mathbf{R}_g\mathbf{Q}_c\mathbf{v}_{c_j}$

such that

$$
\begin{aligned}
e_{g_k} \simeq & \left|\frac{u_{g_k}}{\gamma}\mathbf{h}_{g_k}^{\mathsf{H}}\mathbf{Q}_g\mathbf{v}_{g_k}-1\right|^2 + \sum_{j=1,j\neq k}^{K_g}\left|\frac{u_{g_k}}{\gamma}\mathbf{h}_{g_k}^{\mathsf{H}}\mathbf{Q}_g\mathbf{v}_{g_j}\right|^2 \\
& + \sum_{c\neq g}\sum_{j=1}^{K_c}\left|\frac{u_{g_k}}{\gamma}\right|^2 \mathbf{v}_{c_j}^{\mathsf{H}}\mathbf{Q}_c^{\mathsf{H}}\mathbf{R}_g\mathbf{Q}_c\mathbf{v}_{c_j} + \left|\frac{u_{g_k}}{\gamma}\right|^2\sigma^2. \quad (18)
\end{aligned}
$$

This approximation allows us to reduce the feedback overhead of our scheme to the level of the conventional JSDM schemes [5].

Then, utilizing the equivalence relationship between the sum-rate maximization and the WMMSE problem [13] [14], we can find a solution of (9) by solving the following problem

$$
\begin{aligned}
\min_{\gamma,\{\mathbf{v}_{g_k}\},\{u_{g_k}\},\{s_{g_k}\}} \quad & \sum_{g=1}^{G}\sum_{k=1}^{K_g}(s_{g_k}e_{g_k}-\log_2 s_{g_k}) \quad (19) \\
\text{s.t.} \quad & \sum_{g=1}^{G}\sum_{k=1}^{K_G}\|\mathbf{Q}_g\mathbf{v}_{g_k}\|^2 = P_T, \qquad \forall j,
\end{aligned}
$$

where $\{s_{g_k}\}$ is an auxiliary variable related to the MSE weight.

Since the cost function in (19) is convex with respect to each of the optimization variables $\{\mathbf{v}_{g_k}\}, \{s_{g_k}\}$ and $\{u_{g_k}\}$, we can solve the problem by updating one parameter after fixing the other two. For example, for given $\gamma$ and $\{\mathbf{v}_{g_k}\}$, the optimal $u_{g_k}$ and $s_{g_k}$ are calculated by

$$
u_{g_k} = \frac{\gamma\mathbf{h}_{g_k}^{\mathsf{H}}\mathbf{Q}_g\mathbf{v}_{g_k}}{\sum_{j=1}^{K_g}\left|\mathbf{h}_{g_k}^{\mathsf{H}}\mathbf{Q}_g\mathbf{v}_{g_j}\right|^2 + \sum_{c\neq g}\sum_{j=1}^{K_c}\mathbf{v}_{c_j}^{\mathsf{H}}\mathbf{Q}_c^{\mathsf{H}}\mathbf{R}_g\mathbf{Q}_c\mathbf{v}_{c_j} + \sigma^2} \quad (20)
$$

$$
s_{g_k} = \frac{1}{e_{g_k}} = \frac{1}{\left(1-\frac{u_{g_k}}{\gamma}\mathbf{v}_{g_k}^{\mathsf{H}}\mathbf{Q}_g^{\mathsf{H}}\mathbf{h}_{g_k}\right)}. \quad (21)
$$

Then, we compute $\gamma$ and $\mathbf{v}_{g_k}$ for given $s_{g_k}$ and $u_{g_k}$ as

$$
\begin{aligned}
\mathbf{v}_{g_k} &= \gamma\bar{\mathbf{v}}_{g_k} \quad (22) \\
&= \gamma s_{g_k}u_{g_k}\left(\sum_{j=1}^{K_g}\beta_{g_j}\mathbf{Q}_g^{\mathsf{H}}\mathbf{h}_{g_j}\mathbf{h}_{g_j}^{\mathsf{H}}\mathbf{Q}_g + \mathbf{C}_g + \lambda\mathbf{I}\right)^{-1}\mathbf{Q}_g\mathbf{h}_{g_k}
\end{aligned}
$$

where $\beta_{g_j} = |u_{g_j}|^2 s_{g_j}$ and $\mathbf{C}_g = \sum_{c\neq g}\sum_{j=1}^{K_c}\beta_{c_j}\mathbf{Q}_g^{\mathsf{H}}\mathbf{R}_c\mathbf{Q}_g$, and $\gamma$ and $\lambda$ are, respectively, expressed by

$$
\gamma = \sqrt{\frac{P_T}{\sum_{g=1}^{G}\sum_{k=1}^{K_g}\|\mathbf{Q}_g\bar{\mathbf{v}}_{g_k}\|^2}} \quad (23)
$$

$$
\lambda = \frac{1}{P_T}\sum_{g=1}^{G}\sum_{k=1}^{K_g}|u_{g_k}|^2 s_{g_k}. \quad (24)
$$

As a result, we determine $\{\mathbf{v}_{g_k}\}$, $\{s_{g_k}\}$ and $\{u_{g_k}\}$ in an alternating fashion until convergence. Although the PGP-WMMSE requires the same CSI feedback overhead as the per-group processing RZF in [5], the iterative beamforming computation for every short-term channel realization could be a burden especially for the BS with a large scale array antennas.

**Algorithm 1** PGP WMMSE scheme

Initialize $n=0$, $s_{g_k}=1$, $u_{g_k}=1$ and $\lambda = \frac{K}{P_T}$ for $\forall g,k$.
**repeat**
  Set $n \leftarrow n+1$
  Compute $\mathbf{v}_{g_k}$ for $\forall g,k$.
  Update $u_{g_k}, s_{g_k}$ for $\forall g,k$.
  Update $\beta_{g_k} = s_{g_k}|u_{g_k}|^2$ and $\lambda = \frac{\sum_{g=1}^{G}\sum_{k=1}^{K_g}s_{g_k}|u_{g_k}|^2}{P_T}$ for $\forall g,k$.
  Compute $R_{\Sigma}^{(n)} = \sum_{g,k}\log_2 s_{g_k}$ for $\forall g,k$.
**until** convergence of $|R_{\Sigma}^{(n)} - R_{\Sigma}^{(n-1)}|$

## V. Simulation Results

In this section, we demonstrate the efficiency of the proposed beamforming methods through simulation results. For all simulations, we assume a single-cell massive MISO system where the BS is equipped with $N_T = 100$ uniform linear array (ULA) antennas. All $K = 64$ users are equipped with a single antenna and are separated into $G$ groups. Then, all $K_g$ users in the $g$-th group experience the same transmit antenna correlation $\mathbf{R}_g$ as in (4). Unless specified otherwise, we set the azimuth angle of the $g$-th group as $\theta_g = -\frac{\pi}{3}+\frac{15}{180}\pi(g-1)$ and assume the same angular spread for all groups as $\Delta = \frac{5}{180}\pi$. With this setting, the correlation matrix $\mathbf{R}_g$ has the effective rank $r_g^* = 12$, $\forall g$, and we set $\bar{B} = 10$ and $S_g = \bar{S} = 8$, $\forall g$ for ease of presentation.

For performance comparison, we consider the following schemes. The *"Full-CSI RZF"* indicates the RZF scheme in [1] based on the feedback of every short-term CSI without any pre-beamforming process, and requires the largest amount of feedback overhead among all reference schemes, and hence will serve as a performance upper bound. The *"BD RZF [5]"* schemes represent the conventional JSDM designs where the pre-beamformers are computed by the BD method, while adopting the RZF scheme for the multiuser precoding stage. On the other hand, the proposed *"RBD"* pre-beamformer is determined based on equations (15) and (16) sequentially. Also, the proposed *"PGP-WMMSE"* determines the beamformer's direction and power based on instantaneous parameters $s_{g_k}$ and $u_{g_k}$ that should be calculated for every short-term CSI.

Figure 2 shows the average sum-rate performance of various schemes as a function of SNR. First, we confirm that the proposed RBD scheme achieves improved rate performance than the conventional BD based pre-beamformer. Moreover, it is worthwhile noting that the proposed RBD prebeamformer combined with the proposed WMMSE multiuser precoder achieves the performance close to the *"Full-CSI RZF"* at low SNR with significantly reduced feedback overhead, although they may experience some performance loss as SNR goes to high. Also, we can check that the proposed JSDM "*RBD*

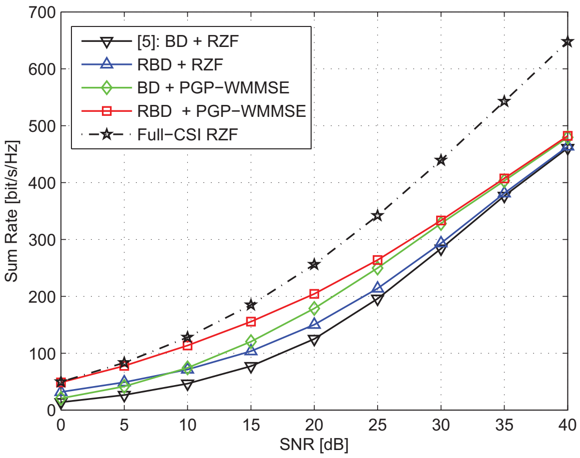


Fig. 2. Average sum-rate comparison in multiuser MISO systems with $N_T = 100$, $K = 64$, $G = 8$, and $K_g = 8$

+ *PGP-WMMSE*" substantially improves the performance over the conventional JSDM. Also, this result shows the performance of the proposed PGP-WMMSE depending on the BD-based pre-beamformer, i.e., $\mathbf{Q}_{\mathrm{BD},g}$ in [5]. It is seen that the RBD pre-beamformer achieves better performance than the BD based scheme in case of the PGP-WMMSE post-beamformer.

## VI. Conclusions

To further enhance the performance of the conventional JSDM, we have studied a new JSDM beamforming method for MISO downlink systems by applying MMSE approaches. First, to suppress IGI, we have designed the RBD based pre-beamformer which takes into account both noise and the statistical channel information, by minimizing an upper bound of the sum-MSE in the large-scale array regime. Then, we have presented the WMMSE based multi-user precoder which requires the same CSI overhead as the conventional JSDM. Through simulations, we have verified that the proposed beamforming schemes outperform the conventional JSDM scheme in low to medium SNR regime. As a future work, we plan to study a joint user-scheduling-and-beamforming schemes based on the proposed framework.

## References

[1] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser Communication −Part I: Channel Inversion and Regularization," *IEEE Transactions on Communications*, vol. 53, pp. 195–202, January 2005.

[2] C. Song, K.-J. Lee, and I. Lee, "MMSE Based Transceiver Designs in Closed-Loop Non-Regenerative MIMO Relaying Systems," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 2310–2319, July 2010.

[3] ——, "MMSE-Based MIMO Cooperative Relaying Systems: Closed-Form Designs and Outage Behavior," *IEEE Journal on Selected Areas in Communications*, vol. 30, pp. 1390–1401, Sep. 2012.

[4] H. Sung, S.-H. Park, K.-J. Lee, and I. Lee, "Linear Precoder Designs for K-user Interference Channels," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 291–301, Jan. 2010.

[5] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint Spatial division and multiplexing - The large-scale array regime," *IEEE Transactions on Information Theory*, vol. 59, pp. 6441–6463, October 2013.

[6] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint Spatial Division and Multiplexing: Opportunistic Beamforming, User Grouping and Simplified Downlink Scheduling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, pp. 876–890, Oct. 2014.

[7] A. Adhikary, E. A. Safadi, M. K. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, "Joint Spatial Division and Multiplexing for mm-Wave Channels," *IEEE Journal on Selected Areas In Communications*, vol. 32, pp. 1239–1255, June 2014.

[8] V. Stankovic and M. Haardt, "Generalized design of multi-user MIMO precoding matrices," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 953–961, Mar 2008.

[9] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading Correlation and Its Effect on the Capacity of Multielement Antenna Systems," *IEEE Transactions on Communications*, vol. 48, pp. 502–513, March 2000.

[10] H. Sung, S.-R. Lee, and I. Lee, "Generalized Channel Inversion Methods for Multiuser MIMO Systems," *IEEE Transactions on Communications*, vol. 57, pp. 3489–3499, November 2009.

[11] X. M. Yang, X. Q. Yang, and K. L. Teo, "A matrix trace inequality," *Journal of Mathematical Analysis and Applications*, vol. 263, pp. 327–331, 2001.

[12] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. 1st ed. Cambridge University Press, 2011.

[13] S. S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted Sum-Rate Maximization Using Weighted MMSE for MIMO-BC Beamforming Design," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 4792–4799, December 2008.

[14] 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 Transactions on Signal Processing*, vol. 59, pp. 4331–4340, September 2011.