Downlink Precoding for Massive MIMO Systems Exploiting Virtual Channel Model Sparsity

Thomas Ketseoglou, *Senior Member, IEEE,* and Ender Ayanoglu, *Fellow, IEEE*

***Abstract*—In this paper, the problem of designing a forward link linear precoder for Massive Multiple-Input Multiple-Output (MIMO) systems in conjunction with Quadrature Amplitude Modulation (QAM) is addressed. A challenge in such system design is to consider finite alphabet inputs, especially with larger constellation sizes such as $M \geq 16$. First, we employ a novel and efficient methodology that allows for a sparse representation of multiple users and groups in a fashion similar to Joint Spatial Division and Multiplexing (JSDM), thus offering an extension of JSDM to finite alphabet data symbols. We term the new approach JSDM for finite alphabets (JSDM-FA). We show that similarly to ordinary JSDM, the optimal pre-beamforming matrix for Massive MIMO in JSDM-FA becomes the DFT matrix. The proposed methodology is next applied jointly with the complexity-reducing Per-Group Processing within groups (PGP-WG) technique, on a per user group basis, in conjunction with QAM modulation and in simulations, for constellation size up to $M = 64$. We show by numerical results that the precoders developed offer significantly better performance than the configuration with no precoder or the plain beamformer and with $M \geq 16$.**

I. INTRODUCTION

Massive MIMO employs a very large number of antennas and enables very high spectral efficiency [1]–[3]. For Massive MIMO to be capable of offering its full benefits, accurate and instantaneous channel state information is required at the base station (BS). Within Massive MIMO research, the problem of designing an optimal linear precoder toward maximizing the mutual information between the input and output on the downlink in conjunction with a finite input alphabet modulation and multiple antennas per user has not been considered in the literature, due to its complexity. There are techniques proposed for downlink linear precoding in a multi-user MIMO scenario, e.g., Joint Spatial Division and Multiplexing (JSDM) [4]–[6], but their implementation has been challenging so far. On the other hand, the problem of finite-alphabet input MIMO linear precoding has been extensively studied in the literature. Globally optimal linear precoding techniques were presented [7], [8] for scenarios employing channel state information available at the transmitter (CSIT)[1] with finite-alphabet inputs, capable of achieving mutual information rates much higher than the previously presented Mercury Waterfilling (MWF) [9] techniques by introducing input symbol correlation through a unitary input transformation matrix in conjunction with channel weight adjustment (power allocation). In addition, more recently, [10] has presented an iterative algorithm for precoder optimization for sum rate maximization of Multiple Access Channels (MAC) with Kronecker MIMO channels. Furthermore, more recent work has shown that when only Statistical Channel State Information (SCSI)[2] is available at the transmitter, in asymptotic conditions when the number of transmitting and receiving antennas grows large, but with a constant transmitting to receiving antenna number ratio, one can design the optimal precoder by looking at an equivalent constant channel and its corresponding adjustments as per the pertinent theory [13], and applying a modified expression for the corresponding ergodic mutual information evaluation over all channel realizations. This development allows for a precoder optimization under SCSI in a much easier way [13]. Finally, [14], [15] present for the first time results for mutual information maximizing linear precoding with large size MIMO configurations and QAM constellations. Such systems are particularly difficult to analyze and design when the inputs are from a finite alphabet, especially with QAM constellation sizes, $M \geq 16$.

In this paper, we present optimal linear precoding techniques for Massive MIMO, suitable for QAM with constellation size $M \geq 16$ and CSIT. The type of antenna arrays are considered for the Base Station (BS) are Uniform Linear Arrays (ULA). We show that by projecting the per user antenna uplink channels on the DFT based angular domain, called virtual channel model[3] (VCM) herein, a sparse representation is possible for the channels. Then, by dividing spatially "distant" users into separate spatial sectors, we show that the spatial virtual channel representations between these users become approximately orthogonal. We then show that the concept of JSDM [4] can be easily applied in the sparse virtual channel model domain representation and show that linear precoding on the downlink using Per-Group Precoding within groups[4] (PGP-WG) in conjunction with the Gauss-Hermite approximation in MIMO [15], [16]. However, the

T. Ketseoglou is with the Electrical and Computer Engineering Department, California State Polytechnic University, Pomona, California (e-mail: tketseoglou@csupomona.edu). E. Ayanoglu is with the Center for Pervasive Communications and Computing, Department of Electrical Engineering and Computer Science, University of California, Irvine (e-mail: ayanoglu@uci.edu). This work was partially supported by NSF grant 1547155.

[1]Under CSIT the transmitter has perfect knowledge of the MIMO channel realization at each transmission.

[2]SCSI pertains to the case in which the transmitter has knowledge of only the MIMO channel correlation matrices [11], [12] and the thermal noise variance.

[3]We need to stress that this is not our propagation-related channel model, but it is a virtual angular orthonormal basis that we use to express the actual channel in an efficient way toward deriving a JSDM type of decomposition. Thus, the VCM model is used as a basis to which we project the actual channel and thus this representation is reversible.

[4]In this paper, we use PGP-BG to refer to the PGP approach between different groups, as proposed in [4], and PGP-WG to refer to PGP within the same group, as proposed in [16], respectively.

978-1-5386-3180-5/18/$31.00 ©2018 IEEE

issue of group-based decoding is still present at the destination. We employ the method of decoding cooperation within groups to mitigate this problem.

The paper is organized as follows: Section II presents the system model and problem statement. Then, in Section III, we present a novel virtual channel approach which allows for efficient downlink precoding in a JSDM fashion for ULA in narrowband channels. In Section IV, we present numerical results for optimal downlink precoding on the system proposed that implements the Gauss-Hermite approximation in the block coordinate gradient ascent method in conjunction with the complexity reducing PGP methodology [15]. Finally, our conclusions are presented in Section V.

II. SYSTEM MODEL AND PROBLEM STATEMENT

Consider the downlink precoding equation on a narrowband (flat-fading) Massive MIMO system with a single cell and JSDM [4]

$$\mathbf{y}_d = \mathbf{H}_u^H \mathbf{P} \mathbf{x}_d + \mathbf{n}_d, \tag{1}$$

where $\mathbf{y}_d$ is the downlink received vector of size $\sum_{g=1}^{G} N_{d,g} \times 1$, $\mathbf{x}_d$ is the $N_u \times 1$ vector of transmitted symbols drawn independently from a QAM constellation, where the downlink channel matrix $\mathbf{H}_d = \mathbf{H}_u^H$, where $\mathbf{H}_u = [\mathbf{H}_1, \cdots, \mathbf{H}_G]$ is the $N_u \times K_{eff}$ uplink channel matrix from all K users, employing N_u receiving antennas at the base, with $K_{eff} = \sum_g N_{d,g}$, where $N_{d,g}$ is defined below. Users have been divided into G groups with K_g users in group g ($1 \leq g \leq G$), with user k of group g denoted as $k^{(g)}$ and employing $N_{d,k^{(g)}}$ transmitting antennas, with ($\sum_{g=1}^{G} K_g = K$), $\mathbf{H}_g = [\mathbf{H}_{g^{(1)}} \cdots \mathbf{H}_{g^{(K_g)}}]$ being group g's uplink channel matrix of size $N_u \times N_{d,g}$, with $N_{d,g}$ comprising the total number of antennas in the group, i.e., $N_{d,g} = \sum_{k^{(g)}} N_{d,k^{(g)}}$, where $\mathbf{n}_u$ represents the independent, identically distributed (i.i.d.) complex circularly symmetric Gaussian noise of variance per component $\sigma_u^2 = \frac{1}{\mathrm{SNR}_{s,u}}$, where $\mathrm{SNR}_{s,u}$ is the channel symbol signal-to-noise ratio (SNR). The uplink symbol vector of size $\mathbf{x}_u$ of size $\sum_g^G N_{d,g} \times 1$ has i.i.d. components drawn from a QAM constellation of order M. We assume that Time Division Duplexing (TDD) is employed in the system to be able to use reciprocity between the uplink and downlink channels. The optimal CSIT precoder $\mathbf{P}$ needs to satisfy

$$\begin{aligned} &\underset{\mathbf{P}}{\text{maximize}} \quad I(\mathbf{x}_d; \mathbf{y}_d) \\ &\text{subject to} \quad \mathrm{tr}(\mathbf{P}\mathbf{P}^H) = N_u, \end{aligned} \tag{2}$$

where the constraint is due to keeping the total power emitted from the N_u antennas constant.

The problem in (2) results in exponential complexity at both transmitter and receiver, and it becomes especially difficult for QAM with constellation size $M \geq 16$ or large MIMO configurations. There are two major difficulties in (2): a) There are N_u input symbols in (2) where N_u is very large, thus making the design of the precoder and its optimization practically impossible, and b) The decoding operation at the receiver needs to be performed by employing all elements of $\mathbf{y}_d$ simultaneously, another impossible demand due to the users being distributed over the entire cell. In order to circumvent these difficulties, the JSDM concept was proposed in [4]. JSDM divides users into approximately orthogonal groups assuming a Gaussian correlated channel [4], based on approximately equal channel covariance matrices in each group. Furthermore, JSDM employs Gaussian data inputs. Because of the orthogonality between different groups, $I(\mathbf{x}_d; \mathbf{y}_d) = \sum_{g=1}^{G} I(\mathbf{x}_g; \mathbf{y}_g)$, where $\mathbf{x}_g$, $\mathbf{y}_g$ represent the data symbols, and received data of group g, respectively (see (11), (12) below and [4]). Thus, the problem in (2) becomes equivalent to the one that maximizes the sum of the group information rates, i.e., the total sum-rate of the system. Under this premise, the downlink precoding problem is divided in two parts: a) A pre-beamforming matrix that comprises the square root of each group's channel covariance matrix, and b) a CSIT Multi-User MIMO (MU-MIMO) optimal precoder. In addition, [4] shows that when the number of base antennas grows to be very high, the (optimal) pre-beamforming matrix approaches a DFT matrix. JSDM helps reduce the complexity inherent in the downlink precoding design tremendously due to the two-stage design optimal approach, plus the introduction of PGP-BG technique. Thus, it represents a major breakthrough advancement toward downlink precoding optimization. However, a major impediment to JSDM in practice has been the lack of a simple way that identifies the different groups of users with ease. Furthermore, [4] has employed Gaussian input symbols, an assumption that can lead to discrepancies as far as the precoder performance is concerned, especially in high SNR [7], [17]. Finally, JSDM deals with the overall downlink precoding on a per group basis, i.e., no methodology for individual UE deriving their own information is presented so far. In this paper, a methodology that employs the virtual channel model decomposition, based on the DFT channel angular domain is employed in order to facilitate the group selection problem in JSDM and then the methodology of PGP-WG technique [15] is employed in order to allow for the design of an optimal overall precoder on a per group basis. In addition, we present different ways to distribute the group received information to the multiple UEs. Our system model allows for multiple receiving antennas per user and multiple data symbols per user with ease, as well as multiple data streams per user, with separate channel coding per stream, in a fashion similar to the methodology in, e.g., [7].

III. THE NARROWBAND SYSTEM DESCRIPTION UNDER THE VIRTUAL CHANNEL MODEL REPRESENTATION

We begin with a ULA deployed at the BS along the z direction as depicted in Fig. 1 and for flat fading, i.e., $B < B_{COH}$, where B, B_{COH} are the RF signal bandwidth and the coherence bandwidth of the channel, respectively. Each user group on the uplink transmits from the same "cluster" of elevation angles $\theta_g \in [\bar{\theta}_g - \Delta\theta, \bar{\theta}_g + \Delta\theta]$, with $\bar{\theta}_g$ being the mean of θ_g, distributed uniformly in the support interval, thus each user's $k^{(g)}$ of group g, ($1 \leq k^{(g)} \leq K_g$ and $1 \leq g \leq G$) transmitting antenna n channel, $\mathbf{h}_{u,g,k,n} = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} \beta_{lgkn} \mathbf{a}(\theta_{lgkn})$, where $\mathbf{a}(\theta_{lgkn}) =$

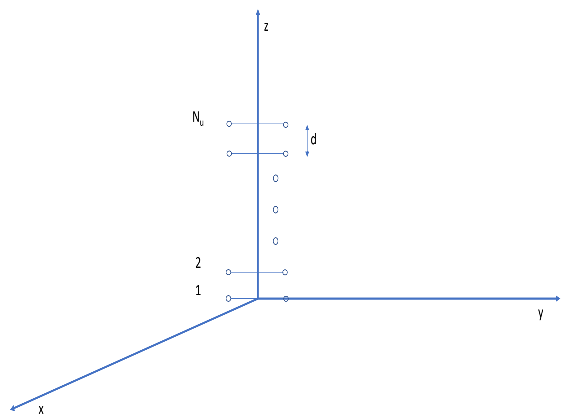

Fig. 1. A ULA deployed across the z axis, together with the projection of a tentative transmission point on the $x,\ y$ plane.

$[1, \exp(-j2\pi D\cos(\theta_{lgkn})), \cdots, \exp(-j2\pi D(N_u-1)\cos(\theta_{lgkn}))]^T$ is the array response vector, where each θ_{lgkn} is independently selected and uniformly distributed in its group's angular support $[\bar{\theta}_g-\Delta\theta, \bar{\theta}_g+\Delta\theta]$, with $D=d/\lambda$ representing the normalized distance of successive array elements, λ being the wavelength, θ_{lgkn} is the elevation (arrival) angle of the l path of group g k user's n receiving antenna, and the path gains β_{lgkn} are independent complex Gaussian random variables with zero mean and variance 1. The VCM representation, presented in [18], is formed by projecting the original channel $\mathbf{H}_u$ to the N_u dimensional space formed by the $N_u \times N_u$ DFT matrix $\mathbf{F}_{N_u}$, with row k, column l $(1 \le k,\ l \le N_u)$ element equal to $\exp(-j\frac{2\pi}{N_u}(k-1)(l-1))$. For Massive MIMO systems, i.e., when $N_u \gg 1$, the following Lemma 1 and 2 as well as Theorem 1 are true.

Lemma 1. *By employing VCM for a ULA at the BS and under flat fading, the number of non-zero components of the VCM representation is small, i.e., the number of non-zero or significant elements in the channels of each group g VCM representation, $|\mathcal{S}_g|$, satisfy $|\mathcal{S}_g| \ll N_u$. Thus, in the VCM domain, a sparse overall group channel representation results.*

Proof. By projecting each group channel $\mathbf{H}_g$ on the DFT virtual channel space [18], we get

$$\tilde{\mathbf{H}}_g = \mathbf{F}_{N_u}^H \mathbf{H}_g, \tag{3}$$

where $\mathbf{F}_{N_u}$ is the DFT matrix of order N_u. Since each group attains the same angular behavior, over all users and antennas in the group, only a few, consecutive elements of $\tilde{\mathbf{H}}_g$ will be significant [19]. This comes as a result of the fact that significant angular components need to be in the main lobe of the response vector, i.e., the condition

$$|\cos(\theta_{lgkn}) - \frac{p}{DN_u}| \le \frac{1}{DN_u}, \tag{4}$$

with $D=\frac{d}{\lambda}$, needs to be satisfied for angular component in the VCM p $(1 \le p \le N_u)$ to be significant, i.e., with power >1. From (4), we can easily see that the corresponding condition over the significant components becomes

$$DN_u\cos(\theta_{lgkn}) - 1 \le p \le DN_u\cos(\theta_{lgkn}) + 1, \tag{5}$$

i.e., there are 3 significant non-zero components in the VCM representation for each channel's path. Since each path contains a different angle, due to the ULA model presented above, this number will be increased, but will be upper-bounded by $DN_u|\cos(\bar{\theta}_g+\Delta\theta) - \cos(\bar{\theta}_g-\Delta\theta)| + 3 = 3+2DN_u|\sin(\bar{\theta}_g)\sin(\Delta\theta)| \approx 3+2DN_u|\sin(\bar{\theta}_g)|(\Delta\theta)$, where $\Delta\theta$ is in radians. For a typical scenario, $N_u=100$, $D=1/2$, $\bar{\theta}_g=30°$, and $\Delta\theta=4°=0.0698$ radian, then the maximum number of non-zero (significant) paths is upper-bounded by 7. □

Lemma 2. *Within the premise of the previous Lemma, if $\cos(\bar{\theta}_g-\Delta\theta) < \cos(\bar{\theta}_{g'}+\Delta\theta) - \frac{2}{DN_u}$, where g and g' represent two different groups ($g \neq g'$) and with $\bar{\theta}_g > \bar{\theta}_{g'}$ and $0 \le \bar{\theta}_g,\ \bar{\theta}_{g'} \le 90°$, then their support sets for each group are mutually exclusive, thus their corresponding virtual channel model beams (VCMB) become orthogonal. A similar relationship holds in the remaining quadrants.*

Proof. When $\bar{\theta}_g > \bar{\theta}_{g'}$ and $0 \le \bar{\theta}_g,\ \bar{\theta}_{g'} \le 90°$, since the $\cos(\cdot)$ function is decreasing in this quadrant, we can easily see that the two support sets for the two groups, $\mathcal{S}_g,\ \mathcal{S}_{g'}$, will be disjoint. This comes from the fact that the assumed condition is equivalent to

$$\cos(\bar{\theta}_g-\Delta\theta) + \frac{1}{DN_u} < \cos(\bar{\theta}_{g'}+\Delta\theta) - \frac{1}{DN_u}, \tag{6}$$

which means that the two support sets are not overlapping, by virtue of (4). We can develop similar conditions for all remaining quadrants. Thus, by assuming adequate spatial separation between groups, we can ensure that the support sets of each group in the virtual channel representation do not overlap. Then, due to the non-overlapping of the support sets, there exists orthogonality between the components of each group in the virtual channel model, as it is next shown. □

Theorem 1. *By employing VCM for a ULA at the BS and under flat fading, provided user groups are sufficiently geographically apart, as per previous lemma, the channel model of the entire downlink channel can be expressed in a fashion that is fully suitable for JSDM type of processing where different groups become orthogonal and the downlink precoder is designed on a per group basis employing the virtual channel model representation alone. In the resulting JSDM type of decomposition, the corresponding group channel matrices are the virtual channel matrices of the group VCM projections and the group pre-beamforming matrices are the group's non-zero (significant) VCM beamforming directions.*

Proof. By employing a size $|\mathcal{S}_g| \times N_u$ selection matrix[5]

$$\mathbf{H}_{g,v} = \mathbf{S}_g^T \tilde{\mathbf{H}}_g = \mathbf{S}_g^T \mathbf{F}_{N_u}^H \mathbf{H}_g, \tag{7}$$

[5] A selection matrix $\mathbf{S}^T$ of size $k \times n$ with $k < n$ consists of rows equal to different unit row vectors $\mathbf{e}_i$ where the row vector element i is equal to 1 in the ith position and is equal to 0 in all other positions. Such a matrix has the property that $\mathbf{S}^T\mathbf{S} = \mathbf{I}$.

where the group g virtual channel matrix is a reduced size, $r_g \times N_{d,g}$, matrix, with $r_g = |\mathcal{S}_g|$ being the number of significant angular components in group g, due to the sparsity available in the angular domain. We can then write for the uplink group g channel matrix $\mathbf{H}_g$,

$$\mathbf{H}_g = \mathbf{F}_{N_u}\mathbf{S}_g\mathbf{S}_g^T\mathbf{F}_{N_u}^H\mathbf{H}_g = \mathbf{F}_{N_u,\mathcal{S}_g}\mathbf{H}_{g,v}, \tag{8}$$

where $\mathbf{F}_{N_u,\mathcal{S}_g}$ represents the selected columns of $\mathbf{F}_{N_u}$ due to its sparse representation in the angular domain. It is important to stress that the above equation is true, although $\mathbf{S}_g\mathbf{S}_g^T$ is not an identity matrix. The reason for the validity of (8) is due to the fact that because of sparsity, the columns of $\mathbf{H}_g$ only have components for the columns of $\mathbf{F}_{N_u}$ defined by $\mathcal{S}_g$, thus these columns only are needed in the representation of $\mathbf{H}_g$ over $\mathbf{F}_{N_u}$. We can then write that due to non-overlapping supports in groups g, g', $\mathcal{S}_g \cap_{g\neq g'} \mathcal{S}_{g'} = \emptyset$, that

$$\mathbf{H}_g^H\mathbf{F}_{N_u,\mathcal{S}_{g'}} = \mathbf{0}, \tag{9}$$

for $g \neq g'$.

$$\mathbf{H}_{d,g} = \mathbf{H}_g^H = \mathbf{H}_{g,v}^H\mathbf{F}_{N_u,\mathcal{S}_{g'}}^H. \tag{10}$$

Since each group attains its non-zero virtual channel representation at non-overlapping positions, we can then use pre-beamforming matrices provided by the matrix $\mathbf{B} = [\mathbf{F}_{N_u,\mathcal{S}_1} \cdots \mathbf{F}_{N_u,\mathcal{S}_G}]$. As we show below these pre-beamforming matrices are optimal for the type of JSDM presented here. Finally, due to non-overlapping of the support sets, i.e., $\mathcal{S}_n \cap_{m\neq n} \mathcal{S}_m = \emptyset$, we see that the system becomes approximately orthogonal inter-group wise, i.e., $\sum_{m\neq g}\mathbf{H}_{d,g}\mathbf{H}_{d,m}^H \approx 0$.

$$\mathbf{y}_d = \begin{bmatrix} \mathbf{H}_{1,v}^H\mathbf{F}_{N_u,\mathcal{S}_1}^H \\ \mathbf{H}_{2,v}^H\mathbf{F}_{N_u,\mathcal{S}_2}^H \\ \vdots \\ \mathbf{H}_{G,v}^H\mathbf{F}_{N_u,\mathcal{S}_G}^H \end{bmatrix} \times \begin{bmatrix} \mathbf{F}_{N_u,\mathcal{S}_1} & \mathbf{F}_{N_u,\mathcal{S}_2} & \cdots & \mathbf{F}_{N_u,\mathcal{S}_G} \end{bmatrix} \times \begin{bmatrix} \mathbf{P}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_3 & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{P}_{G-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{P}_G \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_G \end{bmatrix} + \mathbf{n}, \tag{11}$$

where for $1 \leq g \leq G$, $\mathbf{H}_{g,v}^H$ is a size $N_{d,g} \times |\mathcal{S}_g|$ matrix, $\mathbf{F}_{N_u,\mathcal{S}_G}$ is a size $|\mathcal{S}_g| \times N_u$ matrix, $\mathbf{P}_g$ is a size $|\mathcal{S}_g| \times |\mathcal{S}_g|$ matrix, and $\mathbf{x}_g$ is the group g downlink symbol vector of size $|\mathcal{S}_g| \times 1$. At this point, we would like to stress a main difference between JSDM and JSDM-FA: In (12) we observe that the groups are formed based on the virtual channels $\mathbf{H}_{G,v}^H$ $(1 \leq g \leq G)$, while in the original JSDM, because Gaussian inputs are assumed, the groups are formed based on the matrix square root of the covariance matrix of each user's channel, $\mathbf{R}_h^{1/2}$ [4]. This is not the case in JSDM-FA because group formation is based on the projections of users' channels to the VCM DFT matrix basis. Now due to orthogonality, we can write equivalently

$$\mathbf{y}_d = \begin{bmatrix} \mathbf{H}_{1,v}^H \\ \mathbf{H}_{2,v}^H \\ \vdots \\ \mathbf{H}_{G,v}^H \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_3 & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{P}_{G-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{P}_G \end{bmatrix} \times \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_G \end{bmatrix} + \mathbf{n} = \begin{bmatrix} \mathbf{H}_{1,v}^H\mathbf{P_1}\mathbf{x}_1 \\ \mathbf{H}_{2,v}^H\mathbf{P_2}\mathbf{x}_2 \\ \vdots \\ \mathbf{H}_{G,v}^H\mathbf{P_G}\mathbf{x}_G \end{bmatrix} + \mathbf{n}, \tag{12}$$

where we can set $\mathbf{y} = [\mathbf{y}_1,\ \mathbf{y}_2, \cdots, \mathbf{y}_G]^T$, with $\mathbf{y}_g$ $(1 \leq g \leq G)$ representing the received downlink data of group g. Since each group's precoding becomes independent of other groups, the overall downlink precoding becomes much easier and less complex for both the transmitter and the receiver. In addition, the introduction of the pre-beamforming matrices in the form of VCM beamforming directions also simplifies the RF chains [4]. As in JSDM, in the JSDM-FA the maximization of (2) is equivalent to maximizing the sum of the group rates, i.e., the total sum-rate. The individual precoding of each group becomes now the optimization of a $|\mathcal{S}_g| \times |\mathcal{S}_g|$ precoding matrix $\mathbf{P}_g$, as per the next theorem. □

Theorem 2. *For each group g in the VCM representation, the equivalent optimum precoder,* $\mathbf{P}_g$ *needs to satisfy*

$$\begin{aligned} &\underset{\mathbf{P}_g}{\text{maximize}} && I(\mathbf{x}_{d,g};\mathbf{y}_{d,g}) \\ &\text{subject to} && \operatorname{tr}(\mathbf{P}_g\mathbf{P}_g^H) = N_{d,g}, \end{aligned} \tag{13}$$

where the group g reception model becomes

$$\mathbf{y}_{d,g} = \mathbf{H}_{g,v}^H\mathbf{P}_g\mathbf{x}_g + \mathbf{n}_g, \tag{14}$$

$\mathbf{H}_{g,v}^H$ *is the VCM group's downlink matrix of size* $N_{d,g} \times |\mathcal{S}_g|$, $\mathbf{y}_{d,g}$ *is the group's size* $N_{d,g}$ *reception vector, and* $\mathbf{n}_g$ *is the corresponding noise. This per group precoding problem is equivalent to a precoding problem within the original group channel model, i.e., the VCM transformation does not result in mutual information gain loss in the precoding process.*

Proof. The only part of the theorem that needs proof is the one relating to no information loss. This is easy to prove, since the model in (11) relies equivalently on a $\mathbf{F}_{N_u,\mathcal{S}_g}\mathbf{P}_g$ precoder and the channel is ${\mathbf{H}_{g,v}}^H\mathbf{F}_{N_u,\mathcal{S}_g}^H$, the optimal precoder's left singular vector matrix has to be equal to the Hermitian matrix of the right singular vector matrix of $\mathbf{H}_{g,v}^H\mathbf{F}_{N_u,\mathcal{S}_g}^H$ [7]. Assume that the Singular Value Decomposition (SVD) of $\mathbf{H}_{g,v}^H = \mathbf{U}\mathbf{S}\mathbf{V}^H$. Then, it is easy to show that the right singular vector matrix of $\mathbf{H}_{g,v}^H\mathbf{F}_{N_u,\mathcal{S}_g}^H$ is equal to $\mathbf{F}_{N_u,\mathcal{S}_g}\mathbf{V}$, under the condition that $N_u > |\mathcal{S}_g|$, $N_u > N_{d,g}$, which is true in Massive MIMO. Thus, based on this theorem, the pre-beamforming matrices applied herein are optimal for the JSDM-FA system presented. □

IV. Numerical Results

In this section, we present our numerical results based on ULA Massive MIMO systems with $N_u = 100$ antennas at the base station. The systems employ QAM with size $M = 16,\ 64$. We have used an $L = 3$ Gauss-Hermite approximation [15] which results in 3^{2N_r} total nodes in the Gauss-Hermite approximation due to MIMO in order to facilitate results with optimal precoding in conjunction with QAM modulation. The implementation of the globally optimizing methodology is performed by employing two backtracking line searches, one for $\mathbf{W}$ and another one for $\mathbf{\Sigma}_G^2$ at each iteration, in a fashion similar to [13]. For the results presented, it is worth mentioning that only a few iterations (e.g., typically < 8) are required to converge to the optimal solution results as presented in this paper. We apply the complexity reducing method of PGP-WG [16] to each JSDM-FA formed group which offers semi-optimal results under exponentially lower transmitter and receiver complexity [16]. By employing PGP-WG, one can trade in higher values of $N_{t,v},\ N_{r,v}$ for higher overall throughput, albeit at a slightly increased complexity at the transmitter and receiver, as explained in detail in some of the examples below. Alternatively, one can employ a smaller number of $N_{t,v},\ N_{r,v}$, in order to achieve higher throughput, but at significantly lower complexity. In all cases, it is stressed that the actual number of transmission and reception antennas stays the same, while all physical antennas are employed always. The details of these techniques are omitted here due to space limitation. It is worthwhile mentioning that for precoding methods with finite inputs, two types of channels are regularly present in the literature [7], [13]–[16], [20]: a) Type-I channels in which the precoder offers gain in the lower SNR regime, and b) Type-II channels in which the precoder offers gain in the high SNR regime. Our results herein fully corroborate this type of behavior in all cases considered.

A. VCM Channel Sparsity for ULA Scenarios

First, we present results for the sparse behavior of the VCM representation in the ULA case. We randomly create 5 groups of channels as per the ULA model presented. The base ULA is deployed along the z direction with $N_u = 100$ elements spaced at a normalized distance $D = 0.5$. There are $L = 5$ paths in each channel (a smaller number of L results in sparser representations). The elevation angles for groups $G_1,\ G_2,\ G_3,\ G_4,\ G_5$ are at $5°,\ 33°,\ 61°,\ 89°$, and $117°$, respectively. In addition, the groups possess $16,\ 2,\ 4,\ 4$, and 6 antennas, respectively. The angular spread for all groups is taken to be $\pm\ 4°$ around the elevation angle of each group. The channels are projected to the VCM space, then only components greater than 1 in absolute square power are selected. In all cases considered, this selection process results in more than 94% of the total power of each channel selected. The corresponding, non-overlapping support sets are as follows (the numbers of each set correspond to the numbered components of the VCM representation vector, i.e., the significant VCMBs):

$\mathcal{S}_1 = [56,\ 57,\ 58,\ 59,\ 60,\ 61,\ 62,\ 63,\ 64,\ 65]$,
$\mathcal{S}_2 = [38,\ 39,\ 40,\ 41,\ 42,\ 43,\ 44]$,
$\mathcal{S}_3 = [27,\ 28,\ 29,\ 30,\ 31,\ 32,\ 33,\ 34]$,
$\mathcal{S}_4 = [1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 99,\ 100]$,
$\mathcal{S}_5 = [70,\ 71,\ 72,\ 73,\ 74,\ 75,\ 76,\ 77,\ 78,\ 79]$.

We observe that a ULA allows for easy sparse non-overlapping support sets for multiple groups.

B. Precoding Results

As a first example, we present results for a ULA with 5 groups formed, shown as $G_1,\ G_2, \cdots, G_5$, respectively. Groups $1,\ 2,\ 3,\ 4,\ 5$ occupy the following groups of non-overlapping, i.e., disjoint VCMBs

$\mathcal{S}_1 = [57,\ 58,\ 59,\ 60,\ 61,\ 62,\ 63,\ 64]$,
$\mathcal{S}_2 = [39,\ 41,\ 42, 43,\ 44,\ 45,\ 46,\ 47]$,
$\mathcal{S}_3 = [25,\ 26,\ 27,\ 28,\ 29,\ 30,\ 31,\ 32,\ 33, 34,\ 35]$,
$\mathcal{S}_4 = [1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 98,\ 99, 100]$, and
$\mathcal{S}_5 = [68,\ 69,\ 70,\ 71,\ 72,\ 73,\ 74,\ 75,\ 76,\ 77]$,

respectively. The groups include $4,\ 2,\ 4,\ 4,\ 6$ antennas at the User Equipment (UE), respectively. Users within groups need to co-ordinate their downlink. Thus, the number of users within the group becomes irrelevant and only the number of antennas becomes essential. In Fig. 3 we present results for G_4. We observe that high gains in throughput are available for low SNR, i.e., a Type-I channel behavior. For example, at $\mathrm{SNR}_b = -7\ dB$ there is a 33% throughput increase by using PGP-WG over the no precoding case. In addition, there is a precoding gain of $4 - 5\ dB$ over the low SNR regime. As far as complexity is concerned, based on the analysis of [15], the PGP-WG precoding example presented with $N_{t,v} = 6$ require a complexity (both at the transmitter and receiver) on the order of $3M^4$, while the no precoding example requires a complexity at the receiver on the order of M^{18}, thus PGP-WG needs $(1/3)M^{14}$ less complexity. For the $N_{t,v} = 8$ case the complexity reduction with PGP-WG over the no PGP-WG case becomes $(1/4)M^{14}$. Thus, we see that PGP-WG helps keep the UE complexity low, while it gives significant gains in throughput and SNR. In Fig. 3 we present results for

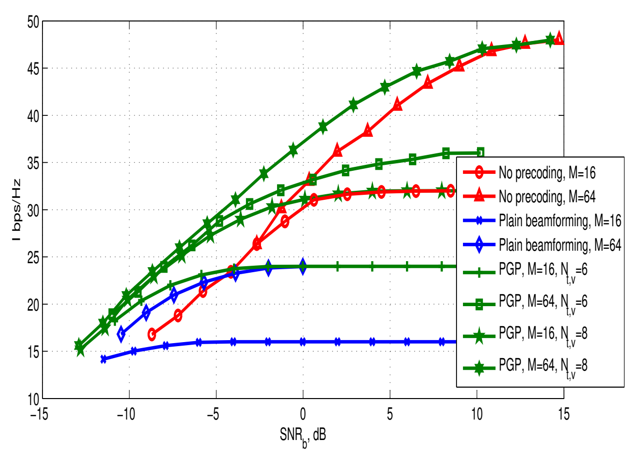

Fig. 2. $I(\mathbf{x};\mathbf{y})$ results for PGP-WG, plain beamforming, and no-precoding cases for the channel in G_4 in conjunction with QAM $M = 16,\ 64$ modulation.

G_5. Here, we observe high gains in throughput in high SNR regime. Here we employ $N_{t,v} = 6$. We observe that this is a

Type-II channel behavior. At $\mathsf{SNR}_b > 0$, the no precoding case throughput saturates at $40\ bps/Hz$. However, with PGP-WG we get significantly higher throughput, e.g., at $\mathsf{SNR}_b = 10\ dB$ the throughput is $48\ bps/Hz$. Further, it takes PGP-WG a factor of $(1/6)M^{16}$ less UE complexity than the no precoding one in order to achieve this additional throughput at the UE.

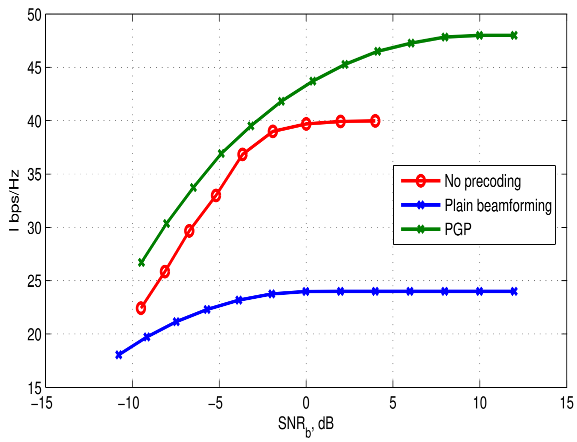

Fig. 3. $I(\mathbf{x};\mathbf{y})$ results for PGP-WG, plain beamforming, and no-precoding cases for the channel in G_5 in conjunction with QAM $M = 16$ modulation.

V. Conclusions

In this paper, a novel methodology for Massive MIMO systems is presented, allowing for optimal downlink linear precoding with finite-alphabet inputs, e.g., QAM and multiple antennas per user. The methodology is based on a sparse VMC decomposition of the downlink channels, which then allows for orthogonality between different user groups, due to non-overlapping sets of VCMBs. The presented methodology extends JSDM to finite alphabet data symbols for a general family of channels. We show that similarly to JSDM with massive MIMO, JSDM-FA resorts to a DFT matrix type of prebeamformer. The methodology is applied in systems with ULA antenna configurations. By employing the PGP-WG technique to the proposed system, we show very high gains are available on the downlink. However, the users in each group need to co-ordinate their detection processes in order to achieve precoding gains. Our numerical results for intragroup decoding coordination show high gains, e.g., typically around 50% in throughput while the incurred precoding complexity is exponentially lower at both the transmitter and receiver sites.

References

[1] T. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3590–3600, November 2010.

[2] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot Contamination and Precoding in Multi-Cell TDD Systems," *IEEE Transactions on Wireless Communications*, vol. 10, pp. 2640–2651, August 2011.

[3] H. Ngo, E. Larsson, and T. Marzetta, "Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems," *IEEE Transactions on Communications*, vol. 61, pp. 1436–1449, April 2013.

[4] A. Adhikary, J. Nam, J. Y. Ahn, and G. Caire, "Joint Spatial Division and Multiplexing: The Large-Scale Array Regime," *IEEE Trans. Inf. Theory*, vol. 59, pp. 6441–6463, October 2013.

[5] 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, no. 5, pp. 876–890, October 2014.

[6] 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, no. 6, June 2014.

[7] C. Xiao, Y. Zheng, and Z. Ding, "Globally Optimal Linear Precoders for Finite Alphabet Signals Over Complex Vector Gaussian Channels," *IEEE Transactions on Signal Processing*, vol. 59, pp. 3301–3314, July 2011.

[8] M. Lamarca, "Linear Precoding for Mutual Information Maximization in MIMO Systems," in *Proceedings International Symposium of Wireless Communication Systems 2009*, 2009, pp. 26–30.

[9] F. Perez-Cruz, M. Rodriguez, and S. Verdu, "MIMO Gaussian Channels with Arbitrary Inputs: Optimal Precoding and Power Allocation," *IEEE Transactions on Information Theory*, vol. 56, pp. 1070–1086, March 2010.

[10] M. Girnyk, M. Vehkapera, and L. K. Rasmussen, "Large System Analysis of Correlated MIMO Multiple Access Channels with Arbitrary Signaling in the Presence of Interference," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 2060–2073, April 2014.

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

[12] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "A Stochastic MIMO Channel Model with Joint Correlation of Both Links," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 90–100, January 2006.

[13] Y. Wu, C.-K. Wen, C. Xiao, X. Gao, and R. Schober, "Linear Precoding for the MIMO Multiple Access Channel With Finite Alphabet Inputs and Statistical CSI," *IEEE Transactions on Wireless Communications*, pp. 983–997, February 2015.

[14] Y. Wu, C.-K. Wen, D. Ng, R. Schober, and A. Lozano, "Low-Complexity MIMO Precoding with Discrete Signals and Statistical CSI," in *Proceedings ICC*, 2016.

[15] T. Ketseoglou and E. Ayanoglu, "Linear Precoding Gain for Large MIMO Configurations with QAM and Reduced Complexity," *IEEE Transactions on Communications*, vol. 64, pp. 4196–4208, October 2016.

[16] ——, "Linear Precoding for MIMO with LDPC Coding and Reduced Complexity," *IEEE Transactions on Wireless Communications*, pp. 2192–2204, April 2015.

[17] ——, "Linear Precoding Gain for Large MIMO Configurations with QAM and Reduced Complexity," in *2016 IEEE Global Communications Conference (GLOBECOM)*, 2016.

[18] A. M. Sayeed, "Deconstructing Multiantenna Fading Channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, October 2002.

[19] H. Xie, F. Gao, S. Zhang, and S. Jin, "Spatial-Temporal BEM and Channel Estimation Strategy for Massive MIMO Time-Varying Systems," in *2016 IEEE Global Communications Conference (GLOBECOM)*, December 2016, pp. 1–6.

[20] W. Zeng, C. Xiao, and J. Lu, "A Low Complexity Design of Linear Precoding for MIMO Channels with Finite Alphabet Inputs," *IEEE Wireless Communications Letters*, vol. 1, pp. 38–42, February 2012.