Hybrid Beamforming Design for Millimeter-Wave Multi-User Massive MIMO Downlink

Zheda Li*, Shengqian Han*†, and Andreas F. Molisch*, *Fellow, IEEE*
*Department of Electrical Engineering, University of Southern California, Los Angeles, California 90089
†School of Electronics and Information Engineering, Beihang University, Beijing 100191, China
Email: zhedali@usc.edu, sqhan@buaa.edu.cn, molisch@usc.edu

Abstract—**In this paper, we consider the design of two-stage beamformers for the downlink of multi-user frequency-division-duplexing (FDD) massive multiple-input multiple-output (MIMO) systems. We consider the case that both link ends are equipped with hybrid analog/digital (HDA) beamforming structures. With analog beamforming and user grouping based on the second-order channel statistics, the user equipment (UE) only needs to feed back its intra-group effective channel. We first show that the strongest eigenbeams of the receive correlation matrix form the optimal analog combiner under the Kronecker channel model assumption. Then, with limited instantaneous channel state information, we jointly optimize the digital precoder and combiner for conditional average net sum-rate maximization by maximizing its lower bound. To initialize our algorithm efficiently, we present a digital precoder design to maximize the conditional average signal-to-leakage-plus-noise ratio (SLNR). Simulation results show significant performance improvements compared to state-of-the-art algorithms.**

I. Introduction

Massive multiple-input multiple-output (MIMO) systems in the millimeter(mm)-wave band will be an important component of the fifth-generation (5G) cellular systems, because they provide a tremendous increase of spectral efficiency by employing very large arrays at the base station (BS) [1, 2]. The performance advantages of massive MIMO systems have been established under the assumption of perfect channel state information at the BS, which can be retrieved from uplink training according to channel reciprocity [3] in time division duplexing (TDD) systems, though training overhead is a concern. Even more so, for frequency division duplexing (FDD) systems, the combined overhead of downlink training and uplink feedback might be prohibitive if no countermeasures are taken. This may hinder the implementation of massive MIMO systems, especially at mm-wave carrier frequencies where the coherence time is much shorter than in the microwave regime.

Furthermore, the inherent hardware constraint of mm-wave transceivers, i.e., high cost and power consumption of mixed-signal components, makes it impractical to build a complete radio frequency (RF) chain for each antenna element. This motivates the hybrid analog/digital (HDA) beamforming structure first proposed in [4, 5], where a reduced-dimensional baseband digital beamforming is concatenated with a phase-shifter network to reduce the necessary number of RF chains.

In order to implement FDD massive MIMO, a scheme called "Joint Spatial Division Multiplexing" (JSDM) was proposed in [6], which provides a two-stage precoder naturally fitting the HDA structure. The first-stage (analog) beamforming is based on the slowly-varying second order channel statistics, which significantly reduces the dimension of the effective channel that needs to be trained and fed back within each coherent fading block. To further alleviate the downlink training and uplink feedback burden, user equipments (UEs) with similar transmit channel covariance are grouped together and inter-group interference is suppressed by a block diagonalization (BD) based analog precoder, which creates multiple "virtual sectors". With this virtual sectorization, downlink training can be conducted in different virtual sectors in parallel, and each UE only needs to feed back the intra-group channels, leading to the reduction of both training and feedback overhead proportional to the number of formed virtual sectors.

In practice, however, to maintain the orthogonality between virtual sectors, JSDM often conservatively group UEs into few groups, because UEs' transmit channel covariances tend to be partially overlapped with each other. This limits the reduction of training and feedback overhead. Once grouping UEs into more virtual sectors violates the orthogonality condition, JSDM is not able to combat the inter-group interference. To solve the problem, [7] proposed to strike overlapped-eigenbeams of UEs in different groups, which however sacrifices some beamforming gain. Another limitation of JSDM is that it was designed only for the case that the UE has a single antenna and a single RF chain.

In this paper, we generalize the JSDM scheme to support non-orthogonal virtual sectorization and HDA structures at both BS and UEs, where analog precoders and combiners are based on the second order channel statistics, digital combiners are based on both intra and inter-group instant effective channels, while digital precoders are only based on intra-group channels at the BS and second order channel statistics.

Another set of problem that bears some formal resemblance to our task is multi-cell digital beamforming optimization, where each UE either needs to feed back the instantaneous channels from all cooperative BSs, e.g., in coordinated multi-point (CoMP) [8], or the instantaneous channel from its serving BS together with instant covariance of interference plus noise, e.g., in [9,10]. However, the problem we are solving is distinct, and actually more challenging, since we assume that the BS does not know the instantaneous information of inter-group interference. The contributions of this paper are thus as follows

978-1-4799-6664-6/16/$31.00 ©2016 IEEE

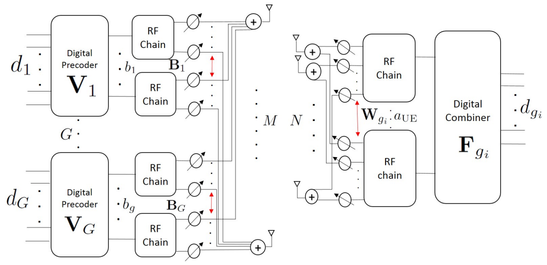

Fig. 1. Hybrid beamforming structure with virtual sectorization

- We analyze HDA structures for non-orthogonal virtual sectorization where the BS knows only instantaneous intra-group channels and second-order statistics.
- We generalize our approach (and JSDM) to the case that also the UE has an HDA structure.
- We investigate the design of analogue beamformers and show that under Kronecker channel model, the optimization of analog combiners and precoders that maximize the intra-group signal to inter-group interference plus noise ratio can be decoupled, where the optimal combiner of each UE consists of the dominant eigenvectors of its receive correlation matrix.
- Given the analog beamformers, the digital precoders are optimized, aimed at maximizing a lower bound of the net conditional average data rate. We develop a block descent algorithm to solve the problem by establishing the equivalence betweem the problem and a weighted average mean square error minimization (WAMMSE) problem.
- Simulations demonstrate the advantages of the proposed beamformers over the JSDM given the same user grouping and analog beanformers, and show that orthogonal user grouping is not necessarily optimal when taking the feedback overhead into account.

Notations: $(\cdot)^\dagger$, $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^{-1}$ stand for Hermitian transpose, transpose, conjugate, and pseudo inverse operators, respectively, $\mathbb{E}[\cdot]$ ($\mathbb{E}[\cdot|\cdot]$) represents expectation (conditional), $\mathrm{tr}(\cdot)$ and $|\cdot|$ denote matrix trace and determinant, respectively, $\otimes$ denotes Kronecker product, $\mathbf{I}_n$ is the n-by-n identity matrix, $\mathcal{CN}(\mathbf{m},\mathbf{K})$ indicates the circularly symmetric complex Gaussian distribution with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{K}$, $\mathrm{vec}(\cdot)$ vectorizes a matrix by stacking its columns, and $\|\cdot\|$ is the Frobenius norm of a matrix.

II. System Model

Consider the downlink transmission of a single-cell massive MIMO FDD system, where the BS is equipped with M antennas and b RF chains to serve K UEs, and each UE has N antennas and a_{UE} RF chains. To reduce the complexity of the digital precoder at the BS and also the channel feedback overhead at the UEs, the concept of virtual sectorization proposed in [6] is employed as shown in Fig. 1. We partition the RF chains at the BS into G groups, where the g-th group serves k_g UEs with $\mathbf{V}_g$. Let d_g and b_g denote the number of data streams and RF chains of the BS assigned to the g-th group, respectively, where $d_g \le b_g$ and $\sum_{g=1}^{G} b_g = b$. Then, for the g-th group, the digital precoder $\mathbf{V}_g$ has the dimension of $b_g \times d_g$ because data streams of each group are processed separately. Therefore, when given the total number of RF chains at the BS, the dimension of $\mathbf{V}_g$ decreases with G, leading to reduced computational complexity.

By assuming flat fading within the coherence time, the received signal of the i-th UE in the g-th group (denoted by UE_{g_i}) is expressed by

$$\hat{\mathbf{s}}_{g_i} = \underbrace{\mathbf{F}_{g_i}^\dagger \mathbf{W}_{g_i}^\dagger \mathbf{H}_{g_i} \mathbf{B}_g \mathbf{V}_{g_i} \mathbf{s}_{g_i}}_{\text{Desired signals}} + \underbrace{\mathbf{F}_{g_i}^\dagger \mathbf{W}_{g_i}^\dagger \mathbf{H}_{g_i} \mathbf{B}_g \sum_{j=1, j\neq i}^{k_g} \mathbf{V}_{g_j} \mathbf{s}_{g_j}}_{\text{Intra-group interference}}$$

$$+ \underbrace{\mathbf{F}_{g_i}^\dagger \mathbf{W}_{g_i}^\dagger \mathbf{H}_{g_i} \sum_{z=1, z\neq g}^{G} \sum_{l=1}^{k_z} \mathbf{B}_z \mathbf{V}_{z_l} \mathbf{s}_{z_l}}_{\text{Inter-group interference}} + \underbrace{\mathbf{F}_{g_i}^\dagger \mathbf{W}_{g_i}^\dagger \mathbf{n}_{g_i}}_{\text{Noise}}, \qquad (1)$$

where $\mathbf{V}_g = [\mathbf{V}_{g_1}, ..., \mathbf{V}_{g_{k_g}}] \in \mathbb{C}^{b_g \times d_g}$ is the digital precoder of the g-th group, $\mathbf{V}_{g_i} \in \mathbb{C}^{b_g \times d_{g_i}}$ is the digital precoder for UE_{g_i}, d_{g_i} is the number of data streams assigned to UE_{g_i} with $\sum_{i=1}^{k_g} d_{g_i} = d_g$, $\mathbf{F}_{g_i} \in \mathbb{C}^{a_{\mathrm{UE}} \times d_{g_i}}$ represents the digital combiner of UE_{g_i}, $\mathbf{B}_g \in \mathbb{C}^{M \times b_g}$ indicates the analog precoder for group g, $\mathbf{W}_{g_i} \in \mathbb{C}^{N \times a_{\mathrm{UE}}}$ is the analog combiner at UE_{g_i}, $\mathbf{H}_{g_i} \in \mathbb{C}^{N \times M}$ denotes the channel matrix of UE_{g_i}, and $\mathbf{n}_{g_i} \sim \mathcal{CN}(\mathbf{0}, \delta_{g_i}^2 \mathbf{I}_N)$ denotes the additive white Gaussian noise vector.

As analyzed in [6], to reduce the downlink training overhead and considering the practical constraints on analog hardware [5], the analog beamformers $\mathbf{B}_z$ and $\mathbf{W}_{g_i}$ need to be designed based on long-term channel information (second-order statistics), while the digital beamformers $\mathbf{V}_z$ and $\mathbf{F}_{g_i}$ can be designed based on reduced-dimensional instantaneous channels. Given the analog beamformers $\mathbf{B}_z$ and $\mathbf{W}_{g_i}$, we consider that orthogonal downlink training is employed for the estimation of the analog-precoded-combined effective channel at UEs. Define the effective channel from the z-th group to UE g_i as $\bar{\mathbf{H}}_{g_i,z} = \mathbf{W}_{g_i}^\dagger \mathbf{H}_{g_i} \mathbf{B}_z \in \mathbb{C}^{a_{\mathrm{UE}} \times b_z}$, where $\mathrm{vec}(\mathbf{H}_{g_i}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_{g_i})$ is the propagation channel of UE_{g_i} and $\mathbf{K}_{g_i} \in \mathbb{C}^{MN \times MN}$ denotes the channel covariance. We assume perfect effective channel estimation at UEs. The effective channels are required for the optimization of digital precoders at the BS, which can be fed back by the UEs in FDD systems. Under virtual sectorization the feedback overhead of each UE can be reduced by limiting that each UE only feeds back the intra-group effective channel, i.e., UE_{g_i} feeds back $\bar{\mathbf{H}}_{g_i,g}$. We also assume that each UE feeds back its channel covariance $\mathbf{K}_{g_i}$ to the BS with negligible overhead since this information changes slowly so that the feedback period can be large.

Consider that the UEs employ an orthogonal analog feedback scheme [11] to feed back the effective channels. Then, the overhead of UE_{g_i} to feed back $\bar{\mathbf{H}}_{g_i,g}$ is b_g channel uses, and the total feedback overhead can be obtained as $\tau_{fb} = \sum_{g=1}^{G} k_g b_g$ channel uses [11]. We can see the impact of the number of groups G on the feedback overhead clearly under the special case where the b RF chains of the BS are evenly assigned to the G groups. In this case, we have $\tau_{fb} = Kb/G$, which is reduced by a factor of G.[1]

Based on (1), we can obtain the net data rate of UE_{g_i} as

$$\bar{R}_{g_i} = (1 - \frac{\tau_{tr} + \tau_{fb}}{T}) \cdot \log_2 |\mathbf{I}_{d_{g_i}} + \mathbf{F}_{g_i}^\dagger \bar{\mathbf{H}}_{g_i,g} \mathbf{V}_{g_i} \mathbf{V}_{g_i}^\dagger \bar{\mathbf{H}}_{g_i,g}^\dagger \mathbf{F}_{g_i} \overline{\mathbf{\Omega}}_{g_i}^{-1}|, \quad (2)$$

where $\tau_{tr} = b$ denotes the overhead of orthogonal downlink training in term of channel uses, T is the total number of channel uses in a coherence block, and $\overline{\mathbf{\Omega}}_{g_i}$ can be expressed as

$$\overline{\mathbf{\Omega}}_{g_i} = \mathbf{F}_{g_i}^\dagger \Big(\bar{\mathbf{H}}_{g_i,g} \sum_{j \neq i}^{k_g} \mathbf{V}_{g_j} \mathbf{V}_{g_j}^\dagger \bar{\mathbf{H}}_{g_i,g}^\dagger + \sum_{z \neq g}^{G} \sum_{l=1}^{k_z} \bar{\mathbf{H}}_{g_i,z} \mathbf{V}_{z_l} \cdot \mathbf{V}_{z_l}^\dagger \bar{\mathbf{H}}_{g_i,z}^\dagger + \mathbf{W}_{g_i}^\dagger \mathbf{W}_{g_i} \delta_{g_i}^2 \Big) \mathbf{F}_{g_i} \triangleq \mathbf{F}_{g_i}^\dagger \mathbf{\Omega}_{g_i} \mathbf{F}_{g_i}. \quad (3)$$

III. Analog Beamformers and Digital Combiner Optimization

In this section, we optimize the analog beamformers and digital combiners at UEs based on channel covariance and instant effective channels, respectively. Since different channel information is exploited for the design of analog and digital beamformers, joint optimization for them is very challenging. We resort to decoupled optimization by designing the analog beamformers to mitigate the inter-group interference while designing the digital combiner to maximize the net data rate of each UE.

A. Analog Beamformers

We begin with the design of the analog combiner of UE_{g_i} by maximizing the received "intra-group signal to inter-group interference plus noise" ratio, which is defined under the assumption of equal power allocation over groups as

$$\bar{\gamma}_{g_i} = \frac{\frac{P_t}{G\|\mathbf{B}_g\|^2} \mathbb{E}[\|\bar{\mathbf{H}}_{g_i,g}\|^2]}{\sum_{z=1,z\neq g}^{G} \frac{P_t}{G\|\mathbf{B}_z\|^2} \mathbb{E}[\|\bar{\mathbf{H}}_{g_i,z}\|^2] + \delta_{g_i}^2 \|\mathbf{W}_{g_i}\|^2}. \quad (4)$$

Given the analog precoders $[\mathbf{B}_g]_{g=1}^G$, the optimal analog combiner $\mathbf{W}_{g_i}$ that maximizes $\bar{\gamma}_{g_i}$ can be readily found by solving a generalized Rayleigh quotient problem, which consists of a_{UE} dominant eigenvectors of the matrix $\big(\sum_{z=1,z\neq g}^{G} \frac{P_t}{G\|\mathbf{B}_z\|^2} \mathbb{E}[\mathbf{H}_{g_i} \mathbf{B}_z \mathbf{B}_z^\dagger \mathbf{H}_{g_i}^\dagger] + \delta_{g_i}^2 \mathbf{I}_N\big)^{-1} \frac{P_t}{G\|\mathbf{B}_g\|^2} \mathbb{E}[\mathbf{H}_{g_i} \mathbf{B}_g \mathbf{B}_g^\dagger \mathbf{H}_{g_i}^\dagger]$.

Then, given the analog combiner of UE_{g_i}, we can obtain that the analog-combined effective channel $\mathbf{W}_{g_i}^\dagger \mathbf{H}_{g_i}$ follows $\text{vec}(\mathbf{W}_{g_i}^\dagger \mathbf{H}_{g_i}) \sim \mathcal{CN}(\mathbf{0}, (\mathbf{I}_M \otimes \mathbf{W}_{g_i}^\dagger) \mathbf{K}_{g_i} (\mathbf{I}_M \otimes \mathbf{W}_{g_i}))$, based on which existing analog precoder design methods, e.g., BD and eigen-beamforming (EB) schemes [6] can be employed.

Since the analog beamformers are coupled with each other, iterative updates of analog precoders and combiners are required in general. Nevertheless, we next show that the iteration between analog precoders and combiners can be avoided if the Kronecker channel model [12] is valid, i.e., the channel covariance $\mathbf{K}_{g_i}$ can be expressed as

$$\mathbf{K}_{g_i} = \mathbf{\Sigma}_{t,g_i} \otimes \mathbf{\Sigma}_{r,g_i}, \quad (5)$$

where $\mathbf{\Sigma}_{t,g_i}$ and $\mathbf{\Sigma}_{r,g_i}$ are transmit and receive correlation matrices, respectively.

Proposition 1: *Under Kronecker channel model (5), the optimal analog combiner that maximizes $\bar{\gamma}_{g_i}$ is independent from analog precoders $[\mathbf{B}_g]_{g=1}^G$ and consists of a_{UE} dominant eigenvectors of receive correlation matrix $\mathbf{\Sigma}_{r,g_i}$.*

The proof is omitted due to lack of space (for details refer to [13]). According to *Proposition 1*, as long as the Kronecker channel model is satisfied, each UE can optimize its own analog combiner individually, and then existing analog precoder design methods can be applied based on the analog-combined effective channel without requiring iterations.

B. Digital Combiner

Given the analog beamformers $[\mathbf{B}_g]_{g=1}^G$ and $\mathbf{W}_{g_i}$ as well as the digital precoder $[\mathbf{V}_g]_{g=1}^G$, the optimal digital combiner $\mathbf{F}_{g_i}$ can be obtained by maximizing the net data rate $\bar{R}_{g_i}$ given in (2). It is not hard to find that the optimal $\mathbf{F}_{g_i}$ is the linear minimum mean square error (MMSE) combiner, which is

$$\mathbf{F}_{g_i} = (\bar{\mathbf{H}}_{g_i,g} \mathbf{V}_{g_i} \mathbf{V}_{g_i}^\dagger \bar{\mathbf{H}}_{g_i,g}^\dagger + \mathbf{\Omega}_{g_i})^{-1} \bar{\mathbf{H}}_{g_i,g} \mathbf{V}_{g_i}. \quad (6)$$

IV. Digital Precoder Optimization

Upon substituting the optimal linear MMSE combiner $\mathbf{F}_{g_i}$ into (2), the net data rate of UE_{g_i} can be rewritten as

$$R_{g_i} = \frac{T - \tau_{tr} - \tau_{fb}}{T} \log_2 |\mathbf{I}_{a_{\text{UE}}} + \bar{\mathbf{H}}_{g_i,g} \mathbf{V}_{g_i} \mathbf{V}_{g_i}^\dagger \bar{\mathbf{H}}_{g_i,g}^\dagger \mathbf{\Omega}_{g_i}^{-1}|, \quad (7)$$

where $\mathbf{\Omega}_{g_i}$ is defined in (3).

Since each UE, say UE_{g_i}, only feeds back the intra-group effective channel $\bar{\mathbf{H}}_{g_i,g}$ but not the inter-group effective channels $\bar{\mathbf{H}}_{g_i,z}$ for $z \neq g$, we need to optimize the digital precoders $[\mathbf{V}_g]_{g=1}^G$ by maximizing the sum rate of UEs averaged over the uncertainties. By noting the correlation between $\bar{\mathbf{H}}_{g_i,g}$ and $\bar{\mathbf{H}}_{g_i,z}$, both of which are determined by the same propagation channel $\mathbf{H}_{g_i}$, we formulate the digital precoder optimization problem aimed at maximizing the conditional-average net sum rate of UE as

$$\max_{[\mathbf{V}_{g_i}]} \sum_{g=1}^{G} \sum_{i=1}^{k_g} \mathbb{E}[R_{g_i} | \bar{\mathbf{H}}_{g_i,g}] \quad (8a)$$

$$s.t. \ \sum_{g=1}^{G} \sum_{i=1}^{k_g} \text{tr}(\mathbf{B}_g \mathbf{V}_{g_i} \mathbf{V}_{g_i}^\dagger \mathbf{B}_g^\dagger) \leq P_t, \quad (8b)$$

where P_t is the maximal transmit power of the BS.

Problem (8) is difficult to solve because it is hard to find an explicit expression for the conditional average data rate. To tackle this difficulty, we apply Jensen's inequality to obtain a lower bound of the conditional average data rate according to the convexity of $\log |\mathbf{I} + \mathbf{K}\mathbf{X}^{-1}|$ with respect to $\mathbf{X}$ as

$$\mathbb{E}[R_{g_i} | \bar{\mathbf{H}}_{g_i,g}] \geq \frac{T - \tau_{tr} - \tau_{fb}}{T} \cdot \log_2 |\mathbf{I}_{a_{\text{UE}}} + \bar{\mathbf{H}}_{g_i,g} \mathbf{V}_{g_i} \mathbf{V}_{g_i}^\dagger \bar{\mathbf{H}}_{g_i,g}^\dagger \mathbb{E}[\mathbf{\Omega}_{g_i} | \bar{\mathbf{H}}_{g_i,g}]^{-1}|. \quad (9)$$

[1]Herein error-free feedback is assumed. When imperfect feedback is considered, the feedback overhead will increase in order to improve the reliability of feedback channels, e.g., by error correction coding.

Despite that the obtained lower bound in (9) is still non-convex with respect to $[\mathbf{V}_{g_i}]$, we can show that the resultant problem maximizing the lower bound can be equivalently transformed into a weighted conditional average mean square error minimization (WAMMSE) problem, which can be solved by a block descent technique [9]. We omit the detailed derivations due to the lack of space, and directly give the equivalent WAMMSE problem as

$$\min_{[\mathbf{A}_{g_i}],[\tilde{\mathbf{F}}_{g_i}],[\mathbf{V}_{g_i}]} \sum_{g=1}^{G}\sum_{i=1}^{k_g} \mathrm{tr}(\mathbf{A}_{g_i}\tilde{\mathbf{E}}_{g_i}) - \log_2|\mathbf{A}_{g_i}| \quad (10a)$$

$$s.t. \quad \sum_{g=1}^{G}\sum_{i=1}^{k_g} \mathrm{tr}(\mathbf{B}_g\mathbf{V}_{g_i}\mathbf{V}_{g_i}^{\dagger}\mathbf{B}_g^{\dagger}) \le P_t, \quad (10b)$$

where $\mathbf{A}_{g_i} \succeq 0$ is the weight matrix for UE_{g_i}, $\tilde{\mathbf{F}}_{g_i}$ is an auxiliary variable representing the digital combiner of UE_{g_i} in the equivalent problem, and $\tilde{\mathbf{E}}_{g_i}$ is the conditional expectation of the MSE matrix $\mathbf{E}_{g_i}$ of the data streams of UE_{g_i} given $\bar{\mathbf{H}}_{g_i,g}$, i.e., $\tilde{\mathbf{E}}_{g_i} = \mathbb{E}[\mathbf{E}_{g_i}|\bar{\mathbf{H}}_{g_i,g}]$ with

$$\begin{aligned}\mathbf{E}_{g_i} =& \tilde{\mathbf{F}}_{g_i}^{\dagger}(\bar{\mathbf{H}}_{g_i,g}\mathbf{V}_{g_i}\mathbf{V}_{g_i}^{\dagger}\bar{\mathbf{H}}_{g_i,g}^{\dagger} + \mathbf{\Omega}_{g_i})\tilde{\mathbf{F}}_{g_i} \\ &+ \mathbf{I}_{d_{g_i}} - \mathbf{V}_{g_i}^{\dagger}\bar{\mathbf{H}}_{g_i,g}^{\dagger}\tilde{\mathbf{F}}_{g_i} - \tilde{\mathbf{F}}_{g_i}^{\dagger}\bar{\mathbf{H}}_{g_i,g}\mathbf{V}_{g_i}. \quad (11)\end{aligned}$$

A. Optimization Algorithm

The objective function (10a) is not jointly convex with respect to $[\mathbf{A}_{g_i}]$, $[\tilde{\mathbf{F}}_{g_i}]$, and $[\mathbf{V}_{g_i}]$, but it is respectively convex for every group of variables if the others are fixed. Based on this fact, we derive a block descent algorithm to find a suboptimal solution to problem (10).

Given the weight matrices $[\mathbf{A}_{g_i}]$ and digital precoders $[\mathbf{V}_{g_i}]$, the optimal digital combiners $[\tilde{\mathbf{F}}_{g_i}]$ can be obtained based on the first-order optimality condition as

$$\tilde{\mathbf{F}}_{g_i} = \tilde{\mathbf{J}}_{g_i}^{-1}\bar{\mathbf{H}}_{g_i,g}\mathbf{V}_{g_i},\ \forall\ i,g, \quad (12)$$

where $\tilde{\mathbf{J}}_{g_i}$ is the instantaneous covariance of the received signal conditionally averaged over the inter-group interference channels

$$\begin{aligned}\tilde{\mathbf{J}}_{g_i} =& \bar{\mathbf{H}}_{g_i,g}\textstyle\sum_{j=1}^{k_g}\mathbf{V}_{g_j}\mathbf{V}_{g_j}^{\dagger}\bar{\mathbf{H}}_{g_i,g}^{\dagger} \quad (13)\\ &+ \textstyle\sum_{z\neq g}^{G}\sum_{l=1}^{k_z}\mathbb{E}[\bar{\mathbf{H}}_{g_i,z}\mathbf{V}_{z_l}\mathbf{V}_{z_l}^{\dagger}\bar{\mathbf{H}}_{g_i,z}^{\dagger}|\bar{\mathbf{H}}_{g_i,g}] + \mathbf{W}_{g_i}^{\dagger}\mathbf{W}_{g_i}\delta_{g_i}^2.\end{aligned}$$

Similarly, we can obtain the optimal $[\mathbf{A}_{g_i}]$ and $[\mathbf{V}_{g_i}]$ as

$$\mathbf{A}_{g_i} = (\mathbf{I}_{d_{g_i}} - \mathbf{V}_{g_i}^{\dagger}\bar{\mathbf{H}}_{g_i,g}^{\dagger}\tilde{\mathbf{F}}_{g_i})^{-1},\ \forall\ i,g, \quad (14)$$

$$\begin{aligned}\mathbf{V}_{g_i} =& \big(\textstyle\sum_{j=1}^{k_g}\bar{\mathbf{H}}_{g_j,g}^{\dagger}\tilde{\mathbf{F}}_{g_j}\mathbf{A}_{g_j}\tilde{\mathbf{F}}_{g_j}^{\dagger}\bar{\mathbf{H}}_{g_j,g} \\ &+ \textstyle\sum_{z\neq g}^{G}\sum_{l=1}^{k_z}\mathbb{E}[\bar{\mathbf{H}}_{z_l,g}^{\dagger}\tilde{\mathbf{F}}_{z_l}\mathbf{A}_{z_l}\tilde{\mathbf{F}}_{z_l}^{\dagger}\bar{\mathbf{H}}_{z_l,g}|\bar{\mathbf{H}}_{z_l,z}] \\ &+ \mu\mathbf{B}_g^{\dagger}\mathbf{B}_g\big)^{-1}\bar{\mathbf{H}}_{g_i,g}^{\dagger}\tilde{\mathbf{F}}_{g_i}\mathbf{A}_{g_i}^{\dagger},\ \forall\ i,g, \quad (15)\end{aligned}$$

where μ is a Lagrange multiplier that can be found through bisection search to satisfy the power constrain (10b).

To obtain $\tilde{\mathbf{F}}_{g_i}$ and $\mathbf{V}_{g_i}$ from (12) and (15), the conditional expectation of the form $\mathbb{E}[\bar{\mathbf{H}}_{g_i,z}\mathbf{Q}\bar{\mathbf{H}}_{g_i,z}^{\dagger}|\bar{\mathbf{H}}_{g_i,g}]$ for $z \neq g$ needs to be computed, which can be done as follows.

By vectorizing $\mathbb{E}[\bar{\mathbf{H}}_{g_i,z}\mathbf{Q}\bar{\mathbf{H}}_{g_i,z}^{\dagger}|\bar{\mathbf{H}}_{g_i,g}]$ as

$$\mathrm{vec}(\mathbb{E}[\bar{\mathbf{H}}_{g_i,z}\mathbf{Q}\bar{\mathbf{H}}_{g_i,z}^{\dagger}|\bar{\mathbf{H}}_{g_i,g}]) = \mathbb{E}[\bar{\mathbf{H}}_{g_i,z}^{*}\otimes\bar{\mathbf{H}}_{g_i,z}|\bar{\mathbf{H}}_{g_i,g}]\mathrm{vec}(\mathbf{Q}),$$

we know that we only need to compute the conditional expectation $\mathbb{E}[\bar{\mathbf{H}}_{g_i,z}^{*}\otimes\bar{\mathbf{H}}_{g_i,z}|\bar{\mathbf{H}}_{g_i,g}]$. Further, defining $\mathbf{y} = \mathrm{vec}(\bar{\mathbf{H}}_{g_i,g})$ and $\mathbf{x} = \mathrm{vec}(\bar{\mathbf{H}}_{g_i,z})$, we can find that $\mathbb{E}[\bar{\mathbf{H}}_{g_i,z}^{*}\otimes\bar{\mathbf{H}}_{g_i,z}|\bar{\mathbf{H}}_{g_i,g}]$ is just a reshaped version of $\mathbb{E}[\mathbf{x}\mathbf{x}^{\dagger}|\mathbf{y}]$.

Since $\mathbf{x}$ and $\mathbf{y}$ are joint complex Gaussian vectors, we can find the conditional second moment of $\mathbf{x}$ as

$$\mathbb{E}[\mathbf{x}\mathbf{x}^{\dagger}|\mathbf{y}] = \mathbb{E}[\mathbf{x}|\mathbf{y}]\mathbb{E}[\mathbf{x}|\mathbf{y}]^{\dagger} + \mathbf{K}_{xx^*|y}, \quad (16)$$

where

$$\begin{aligned}\mathbb{E}[\mathbf{x}|\mathbf{y}] &= \mathbf{K}_{xy^*}\mathbf{K}_{yy^*}^{-1}\mathbf{y}, \\ \mathbf{K}_{xx^*|y} &= \mathbf{K}_{xx^*} - \mathbf{K}_{xy^*}\mathbf{K}_{yy^*}^{-1}\mathbf{K}_{xy^*}^{\dagger}.\end{aligned} \quad (17)$$

In (17), $\mathbf{K}_{xx^*}$ and $\mathbf{K}_{yy^*}$ are the covariance matrices of $\mathbf{x}$ and $\mathbf{y}$, and $\mathbf{K}_{xy^*}$ is the cross covariance matrix of $\mathbf{x}$ and $\mathbf{y}$, all of which are functions of the channel covariance $\mathbf{K}_{g_i}$ and the analog beamformer and can be obtained as

$$\mathbf{K}_{xx^*} = (\mathbf{B}_z^T\otimes\mathbf{W}_{g_i}^{\dagger})\mathbf{K}_{g_i}(\mathbf{B}_z^*\otimes\mathbf{W}_{g_i}), \quad (18a)$$

$$\mathbf{K}_{yy^*} = (\mathbf{B}_g^T\otimes\mathbf{W}_{g_i}^{\dagger})\mathbf{K}_{g_i}(\mathbf{B}_g^*\otimes\mathbf{W}_{g_i}), \quad (18b)$$

$$\mathbf{K}_{xy^*} = (\mathbf{B}_z^T\otimes\mathbf{W}_{g_i}^{\dagger})\mathbf{K}_{g_i}(\mathbf{B}_g^*\otimes\mathbf{W}_{g_i}). \quad (18c)$$

Computation of (18) involves the multiplication of a large-dimensional matrix $\mathbf{K}_{g_i}$. However, this computation occurs only once within the stationarity of second order channel statistics, whose complexity is therefore affordable. Meanwhile, special covariance structures could be utilized to reduce the computational burden. For example, if the Kronecker channel model (5) is considered, (18a) can be simplified to

$$\mathbf{K}_{xx^*} = (\mathbf{B}_z^T\mathbf{\Sigma}_{t,g_i}\mathbf{B}_z^*)\otimes(\mathbf{W}_{g_i}^{\dagger}\mathbf{\Sigma}_{r,g_i}\mathbf{W}_{g_i}). \quad (19)$$

The full algorithm is summarized as follows.

Algorithm 1 WAMMSE algorithm for digital precoder

1: Initialize $[\mathbf{V}_{g_i}]$ such that constraint (10b) is satisfied.
2: **repeat**
3: update $[\tilde{\mathbf{F}}_{g_i}]$ according to (12) and (13),
4: update $[\mathbf{A}_{g_i}]$ according to (14),
5: update $[\mathbf{V}_{g_i}]$ according to (15),
6: **until** the required accuracy or the maximum number of iterations is reached.

B. Initialization Strategy

The performance and convergence speed of the WAMMSE algorithm largely depend on the selected initial value of $[\mathbf{V}_{g_i}]$. In order to achieve better performance, one may run the WAMMSE algorithm many times with different initializations and then keep the best result, which however leads to very high complexity. In the following we derive an initial $\mathbf{V}_{g_i}$ aimed at maximizing the signal to leakage plus noise ratio (SLNR) for UE_{g_i}, $\forall\ i,g$.

Since the BS does not know the inter-group effective interference channel, we define the SLNR in the following way

$$\mathrm{SLNR}_{g_i} = \frac{\mathrm{tr}(\mathbf{P}_{S,g_i})}{\mathrm{tr}(\tilde{\mathbf{P}}_{I,g_i})}, \quad (20)$$

where $\mathbf{P}_{S,g_i} = \hat{\mathbf{F}}_{g_i}^{\dagger}\bar{\mathbf{H}}_{g_i,g}\mathbf{V}_{g_i}\mathbf{V}_{g_i}^{\dagger}\bar{\mathbf{H}}_{g_i,g}^{\dagger}\hat{\mathbf{F}}_{g_i}$ is the instant covariance of received desired signal $\mathbf{s}_{g_i}$, $\hat{\mathbf{F}}_{g_i}$ is a preliminary digital combiner which is selected as the d_{g_i} left dominant singular vectors of the effective channel $\bar{\mathbf{H}}_{g_i,g}$, and $\tilde{\mathbf{P}}_{I,g_i}$ represents the covariance of leakage from the signal intended for UE_{g_i} plus noise conditionally averaged over the inter-group interference channels, which can be expressed as

$$\tilde{\mathbf{P}}_{I,g_i} = \sum_{z\neq g}^{G}\sum_{l=1}^{k_z}\hat{\mathbf{F}}_{z_l}^{\dagger}\mathbb{E}[\bar{\mathbf{H}}_{z_l,g}\mathbf{V}_{g_i}\mathbf{V}_{g_i}^{\dagger}\bar{\mathbf{H}}_{z_l,g}^{\dagger})|\bar{\mathbf{H}}_{z_l,z}]\hat{\mathbf{F}}_{z_l} \quad (21)$$
$$+\sum_{j\neq i}^{k_g}\hat{\mathbf{F}}_{g_j}^{\dagger}\bar{\mathbf{H}}_{g_j,g}\mathbf{V}_{g_i}\mathbf{V}_{g_i}^{\dagger}\bar{\mathbf{H}}_{g_j,g}^{\dagger}\hat{\mathbf{F}}_{g_j}+\delta_{g_i}^{2}\hat{\mathbf{F}}_{g_i}^{\dagger}\mathbf{W}_{g_i}^{\dagger}\mathbf{W}_{g_i}\hat{\mathbf{F}}_{g_i}.$$

Since the analog combiner $\mathbf{W}_{g_i}$ consists of eigenvectors as given after (4) and the preliminary digital combiner consists of singular vectors, we have $\hat{\mathbf{F}}_{g_i}^{\dagger}\mathbf{W}_{g_i}^{\dagger}\mathbf{W}_{g_i}\hat{\mathbf{F}}_{g_i} = \mathbf{I}_{d_{g_i}}$. As a result, to maximize (20) is equivalent to solving a generalized Rayleigh quotient problem. Let $P_{g_i} = \frac{P_t d_{g_i}}{\sum_{g=1}^{G} d_g}$ denote the power allocated to UE_{g_i} under the assumption of equal power allocation over data streams. Since existing BD and EB analog precoders [6] satisfy $\mathbf{B}_g^{\dagger}\mathbf{B}_g = \mathbf{I}_{b_g}$, we have $P_{g_i} = \mathrm{tr}(\mathbf{V}_{g_i}^{\dagger}\mathbf{B}_g^{\dagger}\mathbf{B}_g\mathbf{V}_{g_i}) = \mathrm{tr}(\mathbf{V}_{g_i}^{\dagger}\mathbf{V}_{g_i})$. Then, we can obtain the optimal initial $\mathbf{V}_{g_i} = \sqrt{\frac{P_{g_i}}{d_{g_i}}}\overline{\mathbf{V}}_{g_i}$, where $\overline{\mathbf{V}}_{g_i}$ consists of d_{g_i} dominant eigenvectors of the matrix $\tilde{\mathbf{U}}_{g_i}^{-1}\bar{\mathbf{H}}_{g_i,g}^{\dagger}\hat{\mathbf{F}}_{g_i}\hat{\mathbf{F}}_{g_i}^{\dagger}\bar{\mathbf{H}}_{g_i,g}$, where $\tilde{\mathbf{U}}_{g_i}$ is defined as

$$\tilde{\mathbf{U}}_{g_i} = \sum_{z\neq g}^{G}\sum_{l=1}^{k_z}\mathbb{E}[\bar{\mathbf{H}}_{z_l,g}^{\dagger}\hat{\mathbf{F}}_{z_l}\hat{\mathbf{F}}_{z_l}^{\dagger}\bar{\mathbf{H}}_{z_l,g}|\bar{\mathbf{H}}_{z_l,z}]$$
$$+\sum_{j\neq i}^{k_g}\bar{\mathbf{H}}_{g_j,g}^{\dagger}\hat{\mathbf{F}}_{g_j}\hat{\mathbf{F}}_{g_j}^{\dagger}\bar{\mathbf{H}}_{g_j,g}+\frac{\delta_{g_i}^{2}d_{g_i}}{P_{g_i}}\mathbf{I}_{b_g}. \quad (22)$$

The conditional expectation in (22) can be evaluated similarly by (16)∼(18).

V. Simulation Results

In this section we evaluate the proposed hybrid beamforming strategy via simulations. We consider a single-cluster channel model, where multipath components' (MPC) direction of departure (DOD)/direction of arrival (DOA) concentrates around a dominant direction with a certain angular spread. In every drop of multiple UEs, we independently generate a single-cluster with particular dominant DOD ϕ and DOA θ for each of them. Without loss of generality, we use the same angular spreads Δ_ϕ and Δ_θ for each UE. We consider uniform linear arrays (ULA) for both the BS and the UEs and assume the Kronecker model for the propagation channel. Then, the joint spatial correlation function becomes the product of transmit and receive correlation, which is

$$\rho(\bar{m},\bar{a},\bar{n},\bar{b}) = \frac{1}{2\Delta_\theta}\cdot\frac{1}{2\Delta_\phi}\int_{\phi-\Delta_\phi}^{\phi+\Delta_\phi}e^{-j2\pi D(\bar{m}-\bar{a})\sin(\phi)} \quad (23)$$
$$\int_{\theta-\Delta_\theta}^{\theta+\Delta_\theta}e^{-j2\pi D(\bar{n}-\bar{b})\sin(\theta)}d_\phi d_\theta \triangleq [\mathbf{\Sigma}_{t,g_i}]_{\bar{m},\bar{a}}\cdot[\mathbf{\Sigma}_{r,g_i}]_{\bar{n},\bar{b}},$$

where $\rho(\bar{m},\bar{a},\bar{n},\bar{b})$ indicates the spatial correlation coefficient between BS antenna $\bar{m}$ to UE antenna $\bar{n}$ and BS antenna $\bar{a}$ to UE antenna $\bar{b}$, $[\cdot]_{\bar{m},\bar{a}}$ represents the $(\bar{m},\bar{a})$-th entry of a matrix. Therefore, given the parameter set of a cluster, we can compute its corresponding channel covariance $\mathbf{K}\in\mathbb{C}^{MN\times MN}$ through (23). Ignoring the impact of large scale loss (path loss plus shadowing, which could of course be easily included, but would tend to obfuscate the effects of the overlap of angular spectra), we directly simulate the transfer channel realization through its covariance, i.e. $\mathrm{vec}(\mathbf{H}) = \mathbf{K}^{\frac{1}{2}}\mathbf{w}$, where $\mathbf{w}\sim\mathcal{CN}(\mathbf{0},\mathbf{I}_{MN})$. The detailed simulation configurations are listed in Table. I.

TABLE I
Simulation parameters

DOD range	$\phi_{\min} = -60°, \phi_{\max} = 60°$
DOA range	$\theta_{\min} = -180°, \theta_{\max} = 180°$
DOD/DOA spread	$\Delta_\phi = 15°$, $\Delta_\theta = 50°$
Number of UEs	$K = 16$
Number of BS antennas and RF chains	$M = 64$, $b = 16$
Number of UE antennas and RF chains	$N = 16$, $a_{\mathrm{UE}} = 1, 2, 4$
Antenna spacing (in wavelength)	$D = \frac{1}{2}$

For "virtual sectorization", we implement the K-means algorithm [14] to group UEs with similar transmit channel covariance. Considering the fact that increasing the number of clusters (and thus sectors) reduces the amount of feedback overhead but violates the orthogonality between UE groups, we investigate scenarios with different number of clusters by setting $G = 1, 2, 4$ and 8. For simplicity, we equally assign the BS's RF chains among UE groups, i.e. $b_g = \frac{b}{G}$. Under Kronecker channel model, the optimal analog combiner given in Section III-A is used. For analog precoders, both the BD and EB methods are considered [6]. Physically, the BD approach projects the desired signal of a UE onto the complementary space of its inter-group interference covariance, while the EB method concentrates energy on the strongest eigenbeams of group-averaged transmit covariance. We now assume that each UE is assigned a single data stream, and random user scheduling is employed to schedule b_g UEs if the number of UEs in a group is larger than b_g (the impact of stream assignment and different scheduling algorithms is beyond the scope of this paper, but will be treated in [13]). To implement JSDM in the scenario where each UE has multiple RF chains, we project the effective channel $\bar{\mathbf{H}}_{g_i,g}$ by its dominant left singular vector to turn the multiple RF chains as one effective RF chain. For both JSDM and the proposed method, the optimal combiner (6) at UE side is performed.

We evaluate the downlink training and uplink feedback overhead based on the model given in Section II. Considering a coherence bandwidth of 500 kHz [15] and a coherence time of 2 ms, corresponding approximately to the mobile speed of 1 m/s at 60 GHz, we can obtain that the coherence block includes around $T = 994$ channel uses based on long-term evolution (LTE) system configurations. All the results are averaged over 100 UE drops.

Figure 2 depicts the sum rate achieved by the proposed WAMMSE scheme and JSDM, where Fig.2(a) and (b) use the BD analog precoder, Fig.2(c) and (d) use the EB analog precoder, and training plus feedback overhead is considered in Fig.2(b) and (d). In the legend we use "Gain" to denote the performance gain of the WAMMSE over JSDM, and given the number of clusters, different number of RF chains at each UE are simulated as marked on the bars. We can see that the proposed WAMMSE scheme exhibits significant performance gain over the JSDM in all considered scenarios. When training and feedback overhead is not considered as

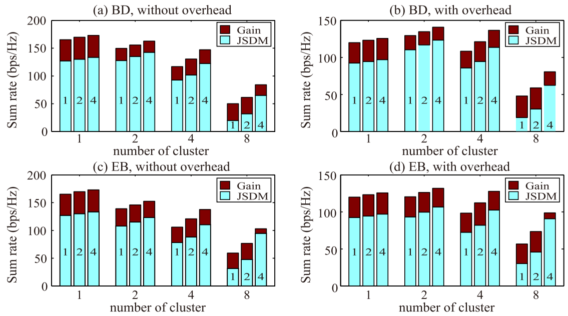

Fig. 2. Sum rate v.s. number of clusters with SNR = 40 dB.

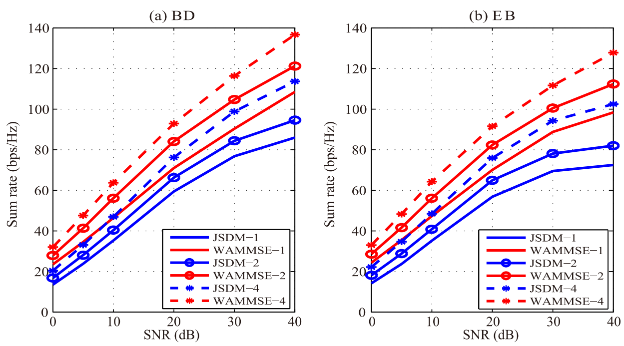

Fig. 3. Sum rate v.s. SNR with $G = 4$.

shown in Fig.2(a) and (c), grouping UEs in a single cluster achieves the best performance, in which case there is no inter-group interference. However, even conservatively considering the mobile speed of 1 m/s, we can see from Fig.2(b) and (d) that a single cluster is no longer optimal due to the high overhead. By comparing the upper and lower subfigures, we can find that the BD analog precoder outperforms the EB analog precoder when the number of clusters is $1, 2$ and 4. However, in the case of 8 clusters, the BD analog precoder sacrifices much beamforming gain to suppress inter-group interference, leading to lower data rate than the EB analog precoder that maximizes the beamforming gain.

Figure 3 compares the performance of the WAMMSE scheme and JSDM as a function of the SNRs, where the number of clusters is $G = 4$ and the numbers in the legends stand for the number of RF chains at each UE. Under the BD analog precoder as shown in Fig. 3(a), the sum rate of the JSDM tends to be flattened because of the residual inter-group interference after analog precoding. In contrast, the WAMMSE scheme is able to efficiently combat it. Under the EB analog precoder in Fig. 3(b), the performance gap between the two schemes is even larger, especially when each UE has only one RF chain. When more RF chains are available at the UEs, the UEs can partially mitigate the inter-group interference, which reduces performance gain of the WAMMSE over JSDM.

VI. Conclusions

In this paper, we studied the design of hybrid beamforming for massive MIMO FDD downlinks. The functionality of multiple RF chains at the UE is explored. We first prove the optimality of decoupling the design of analog precoder and combiner under the Kronecker channel model, which leads to an optimal analog combiner formed by selecting the strongest eigenbeams of the receive covariance matrix. Then, a WAMMSE algorithm is proposed to maximize a lower bound of the conditional average net sum rate. Simulation results demonstrate the necessity of utilizing conditional second order channel statistics for designing digital precoder to combat the inter-group interference. Compared with existing schemes, our algorithm provides better performance to support massive MIMO FDD downlink in a variety of scenarios.

Acknowledgment

Part of this work was supported by the Intel University Research Office under the 5G Higher Denser Wilder project and the National Science Foundation. We thank Dr. Shilpa Talwar, Dr. Nageen Himayat, and Dr. Roya Doostneyad, as well as Prof. Giuseppe Caire, for helpful discussions and valuable suggestions.

References

[1] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[2] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Mag.*, vol. 30, no. 1, pp. 40–60, 2013.

[3] A. Molisch, *Wireless communications*. John Wiley and Sons, 2007.

[4] X. Zhang, A. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Processing*, vol. 53, no. 11, pp. 4091–4103, 2005.

[5] P. Sudarshan, N. Mehta, A. Molisch, and J. Zhang, "Channel statistics-based rf pre-processing with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3501–3511, 2006.

[6] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—the large-scale array regime," *IEEE Trans. Inform. Theory*, vol. 59, no. 10, pp. 6441–6463, 2013.

[7] A. Adhikary, E. Al Safadi, M. Samimi, R. Wang, G. Caire, T. Rappaport, and A. Molisch, "Joint spatial division and multiplexing for mm-wave channels," *IEEE J. Select. Areas Commun.*, vol. 32, no. 6, pp. 1239–1255, 2014.

[8] C. Yang, S. Han, X. Hou, and A. Molisch, "How do we design CoMP to achieve its promised potential?" *IEEE Wireless Commun. Mag.*, vol. 20, no. 1, pp. 67–74, 2013.

[9] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Processing*, vol. 59, no. 9, pp. 4331–4340, 2011.

[10] S. Shim, J. S. Kwak, R. Heath, and J. Andrews, "Block diagonalization for multi-user MIMO with other-cell interference," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2671–2681, 2008.

[11] O. Ayach and R. Heath, "Interference alignment with analog channel state feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 626–636, 2012.

[12] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-Esposti, H. Hofstetter, P. Kyösti, D. Laurenson, G. Matz, *et al.*, "Survey of channel and radio propagation models for wireless mimo systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, no. 1, pp. 56–56, 2007.

[13] Z. Li, S. Han, and A. F. Molisch, "Joint optimization of hybrid beamforming for millimeter-wave multi-user massive MIMO downlink," *to be submitted*.

[14] Y. Xu, G. Yue, N. Prasad, S. Rangarajan, and S. Mao, "User grouping and scheduling for large scale MIMO systems with two-stage precoding," in *Proc. IEEE ICC*, 2014.

[15] M. Kobayashi, N. Jindal, and G. Caire, "Training and feedback optimization for multiuser MIMO downlink," *IEEE Trans. Commun.*, vol. 59, no. 8, pp. 2228–2240, 2011.