User-Centric Virtual Sectorization for Millimeter-Wave 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*—Considering the joint design of analog beamformers when both link ends of a millimeter (mm)-wave massive multiple-input multiple-output (MIMO) system are equipped with hybrid digital/analog (HDA) structures, we aim to maximize the multi-user (MU) MIMO net average throughput of the downlink in a Frequency Division Duplex (FDD) system. To achieve this, we develop an optimization framework, namely *user-centric virtual sectorization* (UCVS), which explores the tradeoff of training overhead, beamforming gain, and spatial multiplexing gain. In the UCVS, both the channel-statistics-based analog beamforming design and a non-orthogonal downlink training scheme are investigated to reduce the necessary cost of instantaneous channel acquisition. By maximizing an approximate net average throughput, we devise efficient algorithms to realize the suboptimal UCVS. With generic mm-wave channel models, we demonstrate by simulations that our proposed scheme outperforms state-of-the-art methods in various scenarios typical for mm-wave communications.**

***Index Terms*—Massive MIMO, mm-wave, training overhead, hybrid beamforming, user-centric virtual sectorization.**

I. Introduction

Combining massive multiple-input-multiple-output (MIMO) [1] with a millimeter (mm)-wave system in a cost- and energy-effective way is not straightforward [2, 3]. One of the main difficulties for massive MIMO implementation is the prohibitive cost and high energy consumption to enable a complete radio frequency (RF) up (down) conversion chain for every antenna element, a problem that is exacerbated at mm-wave frequencies. A promising solution to these problems lies in the concept of hybrid transceivers [4,5], which uses analog beamformers in the RF domain together with a smaller number of RF chains (also see references in [6]).

The short coherence time at mm-wave frequencies constitutes another problem for massive MIMO. To combat the large pathloss, *both* link ends need to be equipped with multiple antenna elements to exploit beamforming gains, which creates significant burden of uplink/downlink training. In this paper, we focus on the approach to design analog beamformers at both link ends based on second-order (covariance) channel statistics. Within the stationarity time of the channel statistics, which can be equivalent to tens or hundreds of coherence times [7], the covariance-based analog beamforming reduces the effective channel dimension to the number of RF chains. Consequently, typical training schemes and digital beamformers, e.g., zero-forcing, for the multi-user (MU) MIMO system can be easily employed.

Designing the analog precoder at the base station (BS) as a function of the channel covariance matrices, *Joint spatial division multiplexing* (JSDM) [8] bears some formal resemblance to our investigations. Similarly, [9] extends JSDM to the scenario where each user equipment (UE) is equipped with a single RF chain and multiple antenna elements. However, their sector-specific designs, which enforce orthogonality between different groups of user equipment (UE), will null out signals from common scattereres, and thus may sacrifice significant beamforming and spatial multiplexing gains. In the current paper, from a perspective of *user-centric beam clustering* (UCBC), the BS forms a beam cluster for an individual UE, whereas the beam clusters of different UEs can overlap with each other. The overlapped part of beam clusters indicates the set of beams pointing toward common scatterers to serve corresponding UEs.

Meanwhile, the allocation of training resources will also be part of the optimization of our formulated problem. The inherent sparsity of mm-wave channels can be exploited by directional beams at both link ends [10]. With appropriately designed analog beamformers, the effective spatial channels of the UEs tend to be semi-orthogonal to each other, which creates the potential of *non-orthogonal beam training* (NOBT). Beam clusters of different UEs may complete their corresponding periods of downlink training at different time slots. Therefore, we may launch the downlink data transmission to UEs whose effective channel-state-information (CSI) is obtained by the BS before the completion of the training phase, namely *simultaneous training-data transmission* (STDT) phase. With both NOBT and potential STDT phase, we will utilize the spatial orthogonality to suppress the interference between training signals and payload data from the propagation perspective of the downlink.

The main contributions of this paper are summarized below. First, for the mm-wave massive MIMO downlink, we develop an optimization framework of *user-centric virtual sectorization* (UCVS), which explores the highly directional and sparse characteristics of mm-wave channels. Then, concerning the coupling effect of training resource allocation and analog beamformer optimization, we devise efficient algorithms to realize user-centric beamformers for maximization of the net average throughput.

Notations: $\mathcal{X} \cap \mathcal{Y}$, $\mathcal{X} \cup \mathcal{Y}$, and $\bar{\mathcal{X}}$ indicate the intersection and union of set $\mathcal{X}$ and $\mathcal{Y}$, and the complement of $\mathcal{X}$,

978-1-5090-5019-2/17/$31.00 ©2017 IEEE

respectively. $\mathcal{X} \setminus \mathcal{Y}$ indicates removing elements of $\mathcal{Y}$ from $\mathcal{X}$. $|\mathcal{X}|$ denotes the cardinality of $\mathcal{X}$. $(\cdot)^\dagger$ and $(\cdot)^T$ stand for Hermitian transpose and transpose, respectively. $\mathrm{tr}\,(\mathbf{X})$ and $|\mathbf{X}|$ denote the trace and determinant of $\mathbf{X}$, respectively. $\mathrm{diag}([x_k]_{k=1}^n) = \mathrm{diag}(x_1, ..., x_n)$, represents a diagonal matrix. $\mathrm{diag}(\mathbf{X})$ denotes a diagonal matrix with the diagonal elements of $\mathbf{X}$ on its diagonal line. $\mathbf{I}_n$ is the n-by-n identity matrix. $\mathbb{E}[\cdot]$ represents the expectation.

II. System and Spatial Channel Model

Consider a single cell downlink of a mm-wave system, where a BS equipped with M antenna elements and l_{BS} RF chains serves K UEs, each equipped with N antenna elements and a single RF chain, i.e., $l_{\text{UE}} = 1$. With hybrid digital/analog (HDA) structures at both ends, we have $M > l_{\text{BS}}$ and $N > l_{\text{UE}}$. In the data transmission of the downlink, the BS broadcasts the beamformed data streams to UEs. Specifically, the BS first projects the streams on digital beamforming vectors at baseband followed by an analog beamforming matrix in the RF domain. The received signal model at the UE$_k$ is $\hat{\mathbf{x}}_k = \mathbf{w}_{ak}^\dagger \mathbf{H}_k \mathbf{F}_{\text{a}} \mathbf{F}_{\text{d}} \mathbf{x} + \mathbf{w}_{ak}^\dagger \mathbf{n}_k$, where $\mathbf{x} \in \mathbb{C}^{K\times 1}$ is the sample symbol vector following the circularly symmetric complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$, $\mathbf{H}_k \in \mathbb{C}^{N\times M}$ denotes the transfer matrix of UE$_k$ whose modeling will be elaborated later, $\mathbf{F}_{\text{a}} \in \mathbb{C}^{M\times l_{\text{BS}}}$ and $\mathbf{F}_{\text{d}} \in \mathbb{C}^{l_{\text{BS}}\times K}$ denote the analog and digital precoder, respectively, $\mathbf{w}_{ak} \in \mathbb{C}^{N\times 1}$ is the analog combiner at UE$_k$, and $\mathbf{n}_k \in \mathbb{C}^{N\times 1}$ indicates the noise vector at UE$_k$ following $\mathcal{CN}(\mathbf{0}, \delta^2 \mathbf{I}_N)$. For ease of notation, we assume that the UEs have the same number of antenna elements, but the generalization to situations where UEs have different array sizes is straightforward.

For the radio propagation in the mm-wave band, we consider the double directional channel description $\mathbf{H}_k = \frac{1}{\sqrt{L_k}} \mathbf{A}_{\text{UE},k} \mathbf{\Sigma}_k \bar{\mathbf{G}}_k \mathbf{A}_{\text{BS},k}^\dagger$, where $\mathbf{A}_{\text{UE},k} \triangleq [\mathbf{a}_{\text{UE}}(\theta_{k1}), \mathbf{a}_{\text{UE}}(\theta_{k2}), ..., \mathbf{a}_{\text{UE}}(\theta_{kP_k})]$, $\mathbf{A}_{\text{BS},k} \triangleq [\mathbf{a}_{\text{BS}}(\phi_{k1}), \mathbf{a}_{\text{BS}}(\phi_{k2}), ..., \mathbf{a}_{\text{UE}}(\phi_{kP_k})]$, $\mathbf{\Sigma}_k \triangleq \mathrm{diag}([\sigma_{kp}]_{p=1}^{P_k})$, and $\bar{\mathbf{G}}_k \triangleq \mathrm{diag}([\bar{g}_{kp}]_{p=1}^{P_k})$, $\forall k, p$. P_k is the number of multipath components (MPCs) from the BS to UE$_k$, L_k is the large scale loss, including path loss and shadowing, σ_{kp} denotes the average power of p-th MPC after normalization by the large scale loss, and $\bar{g}_{kp} \sim \mathcal{CN}(0, 1)$ reflects the small scale fading of the p-th MPC, $\forall k, p$.[1] $\mathbf{a}_{\text{UE}} \in \mathbb{C}^{N\times 1}$ and $\mathbf{a}_{\text{BS}} \in \mathbb{C}^{M\times 1}$ indicate the steering vectors of direction-of-arrival (DOA) θ and direction-of-departure (DOD) ϕ, respectively. Assuming that each MPC exhibits independent fading,[2] we have $\sum_{p=1}^{P_k} \sigma_{kp}^2 = 1, \forall k$. With the block fading assumption, $[\bar{\mathbf{G}}_k]$ varies across coherence blocks, while $[\mathbf{\Sigma}_k]$, $[\mathbf{A}_{\text{UE},k}]$, and $[\mathbf{A}_{\text{BS},k}]$ remain the same within the stationarity time of the second order channel statistics.

Instead of treating each coherence block isotropically, we propose to design analog beamformers based on the knowledge of angular power spectra, including $[\mathbf{A}_{\text{UE},k}]$, $[\mathbf{A}_{\text{BS},k}]$, and $[\mathbf{\Sigma}_k]$. Averaging over the small scale fading, we can develop the closed-form expressions for channel covariance from the perspective of BS and UE, respectively, as $\mathbf{K}_{\text{BS},k} \triangleq \mathbb{E}[\mathbf{H}_k^\dagger \mathbf{H}_k] = \frac{N}{L_k} \mathbf{A}_{\text{BS},k} \mathbf{\Sigma}_k^2 \mathbf{A}_{\text{BS},k}^\dagger$ and $\mathbf{K}_{\text{UE},k} \triangleq \mathbb{E}[\mathbf{H}_k \mathbf{H}_k^\dagger] = \frac{M}{L_k} \mathbf{A}_{\text{UE},k} \mathbf{\Sigma}_k^2 \mathbf{A}_{\text{UE},k}^\dagger$.

Since analog beamformers, i.e., $[\mathbf{w}_{ak}]$ and $\mathbf{F}_{\text{a}}$, remain the same across multiple coherence blocks, we can view the instantaneous effective channel between BS and UE$_k$ as $\bar{\mathbf{h}}_k \triangleq \mathbf{F}_{\text{a}}^\dagger \mathbf{H}_k^\dagger \mathbf{w}_{ak}$, whose dimension is reduced to the number of RF chains, i.e. $l_{\text{BS}} \times 1$. Therefore, channel-statistics-based analog beamformers significantly alleviate the burden of instantaneous CSI acquisition. The covariance of the effective channel $\bar{\mathbf{h}}_k$ can be expressed as $\bar{\mathbf{K}}_{\text{BS},k} \triangleq \mathbb{E}[\bar{\mathbf{h}}_k \bar{\mathbf{h}}_k^\dagger] = \mathbf{F}_{\text{a}}^\dagger \tilde{\mathbf{K}}_{\text{BS},k} \mathbf{F}_{\text{a}}$, where we define the combiner-projected channel covariance as $\tilde{\mathbf{K}}_{\text{BS},k} \triangleq \mathbb{E}[\mathbf{H}_k^\dagger \mathbf{w}_{ak} \mathbf{w}_{ak}^\dagger \mathbf{H}_k] = \frac{1}{L_k} \mathbf{A}_{\text{BS},k} \mathbf{\Sigma}_k \,\mathrm{diag}(\mathbf{A}_{\text{UE},k}^\dagger \mathbf{w}_{ak} \mathbf{w}_{ak}^\dagger \mathbf{A}_{\text{UE},k}) \mathbf{\Sigma}_k \mathbf{A}_{\text{BS},k}^\dagger$.

Concerning the complexity of a practical massive MIMO system, we assume that the analog precoder at the BS consists of columns of the DFT matrix, which can be simply implemented by using a phase shifter network such as a Butler matrix at the BS. Therefore, $\mathbf{F}_{\text{a}}$ becomes a function of the combiner-projected channel covariance matrices and the DFT codebook, i.e. $\mathbf{F}_{\text{a}} = f_{\text{BS}}(\mathbf{\Omega}_M, [\tilde{\mathbf{K}}_{\text{BS},k}])$, where $\mathbf{\Omega}_M$ indicates an $M \times M$ normalized DFT matrix (each column has unit norm). In the massive MIMO regime, the BS antenna array is able to resolve infinitesimal angular differences and the DFT codebook can effectively approximate the eigenspace of the channel covariance if the uniform linear array (ULA) is equipped [8], which leads the codebook-based suboptimal solution to be close to the optimal one. For the analog combiner at the UE, on the other hand, we do not enforce this codebook constraint and directly treat it as a function of UE-side channel covariance, i.e. $\mathbf{w}_{ak} = f_{\text{UE}_k}(\mathbf{K}_{\text{UE},k}), \forall k$, since the number of UE antenna elements is typically smaller than that at the BS.

III. User-Centric Virtual Sectorization

We first define the UE-specific analog precoder as $\mathbf{B}_k \in \mathbb{C}^{M\times l_k}, \forall k$, where l_k is the number of BS RF chains used to serve UE$_k$. To better clarify the proposed UCVS, we define the following sets, which will be used in the remainder of the paper. The set of UEs whose beam cluster contains the j-th transmit beam as $\mathcal{K}_j$, i.e. $\mathcal{K}_j = \{k | \mathbf{b}_j \in \mathbf{B}_k\}$, where $\mathbf{b}_j$ is the j-th column of $\mathbf{\Omega}_M$. $\mathcal{K}_{\text{cc},t}$ is the set of UEs who have completed beam training at time slot t. $\mathcal{K}_{\text{tr},t}$ denotes the set of UEs awaiting the training signal at time slot t, and $\mathcal{K}_{\text{dd},t}$ is the set of UEs receiving a data signal at time slot t. $\mathcal{B}_t = \{j | \mathbf{b}_j \in \cup_{k\in\mathcal{K}_{\text{cc},t}} \mathbf{B}_k, \mathbf{b}_j \notin \cup_{k\in\bar{\mathcal{K}}_{\text{cc},t}} \mathbf{B}_k\}$, indicating the set of beams that are ready for data transmission at time slot t, while $\mathcal{T}_{\text{tr},t}$ is the set of beams trained at time slot t. We next use the example given in Fig. 1 to illustrate the above defined parameters.

Given analog combiners $[\mathbf{w}_{ak}]$, we can build up the beam measure vectors $[\mathbf{s}_k \in \mathbb{R}^{M\times 1}]$ by letting $\mathbf{s}_k(m) = \mathbf{b}_m^\dagger \tilde{\mathbf{K}}_{\text{BS},k} \mathbf{b}_m, \forall k, m$, implying the effective beam pair bipartite graph after appropriate thresholding. A toy example is illus-

[1] Note that MPCs occur in clusters in practice. If the large antenna array is capable of resolving between clusters, but not within them, then the effective channel fulfills the above conditions, which is also widely used in the mm-wave literature [11].

[2] This implies uncorrelated scattering, which is widely accepted in the assumption of channel modeling.

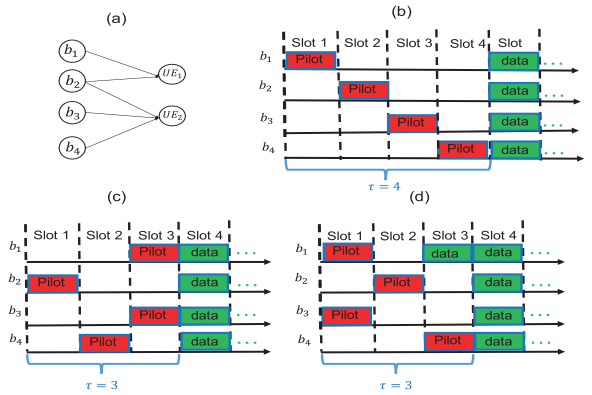

Fig. 1. Compare the training phase of JSDM and UCVS, where (a) is an example of reduced beam pair bipartite graph, (b) reflects the training process of the JSDM, while (c) and (d) represent the training periods of the UCVS with different training orders, respectively. τ is the duration of overall training window.

trated in Fig. 1(a), where we directly form the analog precoder by $\mathbf{F}_{\mathrm{a}} = [\mathbf{b}_k]_{k=1}^4$.

For JSDM, all UEs are placed in the same group with the common analog precoder $\mathbf{F}_{\mathrm{a}}$, and orthogonal beam training is implemented as Fig. 1(b) shows. However, with the partially overlapped beam clusters in UCVS shown in Fig. 1(c) and Fig. 1(d), UE-specific analog precoders are $\mathbf{B}_1 = [\mathbf{b}_1, \mathbf{b}_2]$ and $\mathbf{B}_2 = [\mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4]$. Since $\mathcal{K}_1 \cap \mathcal{K}_3 = \emptyset$, $\mathbf{b}_1$ and $\mathbf{b}_3$ can be trained simultaneously and we only need 3 orthogonal time slots to complete the training of 4 beams.

In Fig. 1(c), based on the association between transmit beams and UEs in Fig. 1(a), $\mathcal{K}_{\mathrm{tr},1} = \{1,2\}$, $\mathcal{K}_{\mathrm{tr},2} = \{2\}$, and $\mathcal{K}_{\mathrm{tr},3} = \{1,2\}$, while $\mathcal{T}_{\mathrm{tr},1} = \{2\}$, $\mathcal{T}_{\mathrm{tr},2} = \{4\}$, and $\mathcal{T}_{\mathrm{tr},3} = \{1,3\}$. $\mathcal{K}_{\mathrm{cc},t} = \emptyset, \forall t \leq 3$, and $\mathcal{K}_{\mathrm{cc},4} = \{1,2\}$, indicating that both UEs complete beam training after the whole training window. Therefore, $\mathcal{B}_t = \emptyset, \forall t \leq 3$. However, in Fig. 1(d), where we swap the order of training $\mathbf{b_1}$, $\mathbf{b}_3$, and $\mathbf{b}_4$, an interesting observation is that $\mathcal{K}_{\mathrm{cc},3} = \{1\}$ and $\mathcal{B}_3 = \{1\}$, which denotes that $\mathbf{b}_1$ can be used for payload transmission at time slot 3 to serve UE_1. Although $\mathbf{b}_2$ and $\mathbf{b}_3$ are also trained before time slot 3, scheduling them for data transmission will leak interference to the training signal of $\mathbf{b}_4$ at UE_2. We will optimize the training order of beams in Section V-A.

IV. Problem Formulation

Given the analog beamforming, the achievable rate of $\mathrm{UE}_{\pi(k)}$ at time slot t by using the dirty paper coding (DPC) scheme in digital baseband is given by [12]

$$C_{\pi(k),t} = \log \left| \frac{\delta^2 \mathbf{w}_{\mathrm{a}\pi(k)}^{\dagger} \mathbf{w}_{\mathrm{a}\pi(k)} + \bar{\mathbf{h}}_{\pi(k),\mathrm{dd},t}^{\dagger} \sum_{j \geq k} \mathbf{\Gamma}_{\pi(j),t} \bar{\mathbf{h}}_{\pi(k),\mathrm{dd},t}^{\dagger}}{\delta^2 \mathbf{w}_{\mathrm{a}\pi(k)}^{\dagger} \mathbf{w}_{\mathrm{a}\pi(k)} + \bar{\mathbf{h}}_{\pi(k),\mathrm{dd},t}^{\dagger} \sum_{j > k} \mathbf{\Gamma}_{\pi(j),t} \bar{\mathbf{h}}_{\pi(k),\mathrm{dd},t}^{\dagger}} \right|,$$

where $\pi(k) \in \mathcal{K}_{\mathrm{dd},t}$ and $[\pi(k)]_{k=1}^{|\mathcal{K}_{\mathrm{dd},t}|}$ is the ordered index set of UEs in DPC, and $\mathbf{\Gamma}_{\pi(k),t}$ is the input covariance of $\mathrm{UE}_{\pi(k)}$ at the time slot t. The DPC scheme optimizes $\mathbf{\Gamma}_{\pi(k)}$ in a sequential manner so that $\mathrm{UE}_{\pi(k)}$ will not be interfered by the streams for prior UEs, i.e., $[\mathrm{UE}_{\pi(j)}], j < k$. Therefore, the net average MU-MIMO downlink capacity within the coherence block is $C_{\mathrm{avg,DL}} = \frac{\sum_{t=1}^{T_{\mathrm{cor}}} \sum_{\pi(k) \in \mathcal{K}_{\mathrm{dd},t}} C_{\pi(k),t}}{T_{\mathrm{cor}}}$, where T_{cor} is the coherence time in units of channel use. If we do not consider the data transmission during the training window, $C_{\mathrm{avg,DL}}$ becomes $(1 - \frac{\tau}{T_{\mathrm{cor}}}) \sum_{k=1}^{K} C_{\pi(k)}$.

Considering the whole stationarity region of channel statistics, we intend to jointly optimize the analog beamformers and pilot assignment matrix $\mathbf{P}_{\mathrm{tr}}$, which leads to the maximization of the net average downlink capacity:

$$\max_{[\mathbf{B}_k, \mathbf{w}_{ak}]_{k=1}^{K}, [\rho_{\mathrm{p},t}]_{t=1}^{\tau}, \mathbf{P}_{\mathrm{tr}}} \mathbb{E}\Big[\max_{[\mathbf{\Gamma}_{\pi(k),t}, \pi(k) \in \mathcal{K}_{\mathrm{dd},t}]_{t=1}^{T_{\mathrm{cor}}}} C_{\mathrm{avg,DL}} \Big] \quad (1a)$$

$$s.t.\ \mathbf{B}_k \subset \mathbf{\Omega}_M, \forall k, l_{\mathrm{use}} = \mathrm{rank}([\mathbf{B}_k]_{k=1}^{K}) \leq l_{\mathrm{BS}}, \quad (1b)$$

$$\mathbf{P}_{\mathrm{tr}}(m,n) \in \{0,1\}, \forall m, n, \sum_{n=1}^{\tau} \mathbf{P}_{\mathrm{tr}}(m,n) = 1, \forall m, \quad (1c)$$

$$\rho_{\mathrm{p},t} |\mathcal{T}_{\mathrm{tr},t}| + \sum_{\pi(k) \in \mathcal{K}_{\mathrm{dd},t}} \mathrm{tr}(\mathbf{\Gamma}_{\pi(k),t}) \leq \rho_{\mathrm{d}}, \forall t, \quad (1d)$$

where the expectation of $C_{\mathrm{avg,DL}}$ is taken to average out the small scale fading, i.e. $[\bar{\mathbf{G}}_k]$, across multiple coherence blocks within the stationarity time of the channel statistics. Note that the CSI feedback can be realized by the dedicated uplink channel right after the training. Since we focus on the performance of the downlink, we assume ideal instantaneous channel acquisition from the uplink feedback channel, and do not incorporate the feedback cost in problem (1), an assumption that is widely used in the literature [8, 9]. Constraint (1b) indicates that an individual beam cluster consists of normalized DFT columns and the total number of used transmit beams, i.e. l_{use}, shall not surpass l_{BS}. Analog combiners at UEs, $[\mathbf{w}_{ak}]$, are functions of UE-side channel covariance matrices. $\mathbf{P}_{\mathrm{tr}} \in \mathbb{N}^{l_{\mathrm{use}} \times \tau}$ denotes the pilot assignment matrix, where each row has a single non-zero entry to indicate the assigned pilot for the beam. In (1d), the total transmit power is constrained by ρ_{d}, and $\mathrm{tr}(\mathbf{\Gamma}_{\pi(k),t}) = \mathrm{tr}(\mathbf{F}_{\mathrm{a},t} \mathbf{\Gamma}_{\pi(k),t} \mathbf{F}_{\mathrm{a},t}^{\dagger})$ is the power for data transmission to UE_k at time slot t, and $\rho_{\mathrm{p},t}|\mathcal{T}_{\mathrm{tr},t}|$ is the total power used for downlink training at time slot t.

Resorting to the uplink-downlink duality theory [13], we can develop an equivalent uplink problem of (1):

$$\max_{[\mathbf{B}_k, \mathbf{w}_{ak}]_{k=1}^{K}, [\rho_{\mathrm{p},t}]_{t=1}^{\tau}, \mathbf{P}_{\mathrm{tr}}} \mathbb{E}\Big[\max_{[\Gamma'_{k,t}, k \in \mathcal{K}_{\mathrm{dd},t}]_{t=1}^{T_{\mathrm{cor}}}} C_{\mathrm{avg,UL}} \Big], \quad (2a)$$

$$s.t.\ (1b), (1c), \rho_{\mathrm{p},t} |\mathcal{T}_{\mathrm{tr},t}| + \textstyle\sum_{k \in \mathcal{K}_{\mathrm{dd},t}} \Gamma'_{k,t} \mathbf{w}_{ak}^{\dagger} \mathbf{w}_{ak} \delta^2 \leq \rho_{\mathrm{d}}, \forall t,$$

where $C_{\mathrm{avg,UL}} = \frac{1}{T_{\mathrm{cor}}} \sum_{t=1}^{T_{\mathrm{cor}}} C_{t,\mathrm{UL}}$. $C_{t,\mathrm{UL}} = \log |\sum_{k=1}^{K} \bar{\mathbf{h}}_{k,\mathrm{dd},t} \Gamma'_{k,t} \bar{\mathbf{h}}_{k,\mathrm{dd},t}^{\dagger} + \mathbf{I}_{|\mathcal{B}_t|}|$, denoting the instantaneous uplink capacity at time slot t. $\Gamma'_{k,t}$ indicates the dual uplink transmit power coefficient of UE_k at the time slot t, $\forall k, t$. Detailed developments of the uplink-dual problem with HDA structure at both ends are discussed in [14].

*1) **Decoupled optimization with reduced complexity**:* Decoupling the interaction between instantaneous $[\Gamma'_{k,t}]$ and channel-statistics-based variables can significantly reduce the problem complexity. Therefore, rather than jointly optimizing power allocations $[\Gamma'_{k,t}]$, we stick with simple equal power allocation among training signals and payload data, i.e., $\Gamma'_{k,t} = \rho_{\mathrm{p},t}$, where $k \in \mathcal{K}_{\mathrm{dd},t}$ and $t = 1, \ldots, T_{\mathrm{cor}}$. With unit-norm combiners $[\mathbf{w}_{ak}]$, we have the following power allocation

equality: $\rho_{\mathrm{p},t} = \Gamma'_{k,t} = \frac{\rho_{\mathrm{d}}}{|\mathcal{T}_{\mathrm{tr},t}|+\delta^2|\mathcal{K}_{\mathrm{dd},t}|}, \forall k \in \mathcal{K}_{\mathrm{dd},t}$. At time slots dedicated for training, it is reduced to equal power allocation over trained beams, i.e. $\rho_{\mathrm{p},t} = \frac{\rho_{\mathrm{d}}}{|\mathcal{T}_{\mathrm{tr},t}|}$, while after the training window, we have equal power allocation among UEs, i.e. $\rho_{\mathrm{p},t} = \frac{\rho_{\mathrm{d}}}{\delta^2|\mathcal{K}_{\mathrm{dd},t}|}$. By introducing the power allocation equality, $C_{\mathrm{avg.,UL}}$ becomes an achievable net throughput rather than the net uplink capacity. However, we reduce the original downlink problem over different time scales to an uplink problem purely over the long-term CSI:

$$\max_{[\mathbf{B}_k,\mathbf{w}_{\mathrm{a}k}]_{k=1}^K,\mathbf{P}_{\mathrm{tr}}} \mathbb{E}[C_{\mathrm{avg,UL}}] \tag{3a}$$

$$s.t.\ (1b),(1c),$$

$$\|\mathbf{w}_{\mathrm{a}k}\| = 1, \forall k. \tag{3b}$$

*2) **Average throughput approximation**:* To avoid the computational burden in evaluation of the expectation at (3a), we consider the following upper bound of average uplink throughput based on Jensen's inequality: $\mathbb{E}[C_{t,\mathrm{UL}}] \leq C_{t,\mathrm{UL,upper}} \triangleq \log \mathbb{E}[|\rho_{\mathrm{p},t}\sum_{k=1}^K \bar{\mathbf{h}}_k\bar{\mathbf{h}}_k^\dagger + \mathbf{I}_{l_{\mathrm{BS}}}|]$. Without loss of generality, we ignore the time subscript in the following, and explore the uplink throughput bound approximation for the dedicated data transmission phase. The result is directly applicable for the STDT phase.

Proposition 1: Under a single-path channel model, i.e. $P_k = 1$, $\forall k$, we have the following equivalence: $\mathbb{E}[|\rho_{\mathrm{p}}\mathbf{F}_{\mathrm{a}}^\dagger\sum_{k=1}^K\mathbf{H}_k^\dagger\mathbf{w}_{\mathrm{a}k}\mathbf{w}_{\mathrm{a}k}^\dagger\mathbf{H}_k\mathbf{F}_{\mathrm{a}} + \mathbf{I}_{l_{\mathrm{BS}}}|] = |\rho_{\mathrm{p}}\mathbf{F}_{\mathrm{a}}^\dagger\sum_{k=1}^K\tilde{\mathbf{H}}_k^\dagger\mathbf{w}_{\mathrm{a}k}\mathbf{w}_{\mathrm{a}k}^\dagger\tilde{\mathbf{H}}_k\mathbf{F}_{\mathrm{a}} + \mathbf{I}_{l_{\mathrm{BS}}}|$, where $\tilde{\mathbf{H}}_k \triangleq \frac{1}{L_k}\mathbf{A}_{\mathrm{UE},k}\mathbf{\Sigma}_k\mathbf{A}_{\mathrm{BS},k}^\dagger$.

Proposition 1 can be easily obtained from the result in [14], whose proof is omitted here due to lack of space. Based on Proposition 1, we obtain a closed-form expression to evaluate the net average uplink throughput under the single-path channel model, and develop the following problem:

$$\max_{[\mathbf{B}_k,\mathbf{w}_{\mathrm{a}k}]_{k=1}^K,\mathbf{P}_{\mathrm{tr}}} \tilde{C}_{\mathrm{avg,UL}} = \frac{1}{T_{\mathrm{cor}}}\sum_{t=1}^{T_{\mathrm{cor}}}\tilde{C}_{t,\mathrm{UL}}, s.t.\ (1b),(1c),(3b), \tag{4}$$

where $\tilde{C}_{t,\mathrm{UL}} = \log|\rho_{\mathrm{p},t}\mathbf{F}_{\mathrm{a},t}^\dagger\sum_{k\in\mathcal{K}_{\mathrm{dd},t}}\tilde{\mathbf{H}}_k^\dagger\mathbf{w}_{\mathrm{a}k}\mathbf{w}_{\mathrm{a}k}^\dagger\tilde{\mathbf{H}}_k\mathbf{F}_{\mathrm{a},t} + \mathbf{I}_{|\mathcal{B}_t|}|$. Without the assumption of $P_k = 1, \forall k$, Proposition 1 does not hold in general and problem (4) becomes an approximation of problem (3). Our simulation results in Section VI demonstrate that the approximation performs well, even with general settings of $[P_k]$.

V. Algorithm Development

Problem (4) is still generally non-convex, involving integer programming for designing $\mathbf{P}_{\mathrm{tr}}$ and $[\mathbf{B}_k]$. Meanwhile, given a topology of the beam pair bipartite graph as shown in Fig. 1, there is no closed-form expression for the minimum cost to complete the training, not to mention which training order we should apply to increase the opportunity of data transmission during the training window. In this section, we will first provide a graph-based algorithm to heuristically optimize the training order. Then, a greedy algorithm is proposed to achieve a suboptimal solution to problem (4).

A. Training Order Optimization

Given a beam pair bipartite graph, the minimum training cost can be evaluated by the algorithm proposed in [15], which provides a suboptimal solution to minimize an upper bound of the training cost: whereas [15] treats left side nodes as BSs, we view them as transmit beams. For the toy example in Fig. 1(a), the output of the algorithm will be $[t_1, t_2, t_3, t_2]$, corresponding to $[\mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4, \mathbf{b}_1]$, where t_i indicates the time slot index of the i-th pilot dimension, $\forall i$. However, the schedule order of pilot dimensions for training is not explored.

Considering that the purpose of optimizing the training order is to increase the transmission opportunity for payload data within the training window, we heuristically choose to maximize the total number of time slots for payload data transmission as the objective function, which is $\max\sum_{k=1}^K(T_{\mathrm{cor}} - T_{\mathrm{tr},k})$, where $T_{\mathrm{tr},k}$ indicates the time instance when the BS completes the training for UE_k. Apparently, it is equivalent to minimize the sum of training periods of all UEs, i.e. $\min\sum_{k=1}^K T_{\mathrm{tr},k}$. To approximate the optimal solution to this typical integer programming problem, we summarize the proposed algorithm below:

1) Define the degree of time slot t_i as $D(t_i)$, which is the number of transmit beams assigned to time slot t_i. Define the set $\mathcal{D} = \{D(t_1), ..., D(t_\tau)\}$, which includes all values of time slot degree, and sort the elements in a descending order.
2) For the i-th element in $\mathcal{D}$, i.e., $\mathcal{D}(i)$, extract the set of time slots $\mathcal{P}_i = \{t_m | D(t_m) = \mathcal{D}(i)\}$, and calculate their priority metrics $\sum_{j\in\mathcal{T}_{\mathrm{tr},t_m}}\sum_{k=1}^K\frac{I_k(\mathbf{b}_j)}{L_{\mathrm{tran},k}}, \forall m \in \mathcal{P}_i$, where $I_k(\mathbf{b}_j)$ is an indicator to denote whether $\mathbf{b}_j$ is associated with UE_k, $L_{\mathrm{tran},k}$ is the number of transmit beams connected to UE_k, and $\mathcal{T}_{\mathrm{tr},t_m}$ contains all transmit beams trained on pilot dimension t_m.
3) Sort the priority metrics of time slots belonging to $\mathcal{P}_i$ in a descending order and sequentially assign indices to them.
4) Repeat step 2) and step 3) for $i = 1, ..., |\mathcal{D}|$.

At step 2) and 3), for pilot dimensions with the same degree, say $\mathcal{D}(i)$, we introduce a metric $\sum_{j\in\mathcal{T}_{t_m}}\sum_{k=1}^K\frac{I_k(\mathbf{b}_j)}{L_k}$ to evaluate the priority order of the m-th pilot dimension, $\forall m \in \mathcal{P}_i$. $\sum_{k=1}^K\frac{I_k(\mathbf{b_j})}{L_k}$ can be interpreted as the relative significance of $\mathbf{b}_j$. If it is very large, $\mathbf{b}_j$ is connected to a lot of UEs associated with a few transmit beams, then scheduling $\mathbf{b}_j$ first increases the chance to finish training of many UEs earlier than τ.

B. Greedy User-Centric Beam Clustering

Thanks to the training order optimization in Section V-A, we can evaluate the performance of any given topology of beam bipartite graph, which lays the foundation of our proposed *greedy user-centric beam clustering* (GUCBC) algorithm. The implementation procedure is summarized as follows:

1) Initially, let $\mathbf{w}_{\mathrm{a}k} = \mathbf{0}, \forall k$ and $\mathbf{F}_{\mathrm{a}} = \emptyset$. Let $\mathbf{W}_{\mathrm{a}}$ be the ensemble of analog combiners as $\mathbf{W}_{\mathrm{a}} \triangleq [\mathbf{w}_{\mathrm{a}1}, ..., \mathbf{w}_{\mathrm{a}K}]$.
2) Extract the beam measure vectors $[\mathbf{s}_k]_{k=1}^K$, enforce small entries to be zero if a certain portion, i.e. γ, of total average energy can be maintained, and build up the beam

pair bipartite graph. Define a beam pair set $\mathcal{E}$ containing all edges, i.e. $(\mathbf{b},\mathbf{r}) \in \mathcal{E}$ if transmit beam $\mathbf{b}$ and receive eigenbeam $\mathbf{r}$ are connected.

3) Let $\mathbf{W}'_{\mathrm{a}} = \mathbf{W}_{\mathrm{a}}$ and $\mathbf{F}'_{\mathrm{a}} = \mathbf{F}_{\mathrm{a}}$. For a candidate beam pair $\mathbf{e} = (\mathbf{b},\mathbf{r})$ in $\mathcal{E}$, we let $\mathbf{F}'_{\mathrm{a}} = [\mathbf{F}'_{\mathrm{a}}, \mathbf{b}]$ and assign $\mathbf{r}$ to its corresponding UE. Then, follow procedures in SectionV-A to optimize the pilot assignments $\mathbf{P}_{\mathrm{tr}}$, and evaluate the *net average uplink throughput approximation* (NAUTA) by (4).
4) Repeat step 3) for every candidate edge and find the optimal one $\mathbf{e}^\star = (\mathbf{b}^\star, \mathbf{r}^\star)$ that can enhance the NAUTA most.
5) Update $\mathbf{W}_{\mathrm{a}}$ by assigning $\mathbf{r}^\star$ to its corresponding $\mathrm{UE}_{k^\star}$, update $\mathbf{F}_{\mathrm{a}}$ by $\mathbf{F}_{\mathrm{a}} = [\mathbf{F}_{\mathrm{a}}, \mathbf{b}^\star]$, and remove the beam pairs starting with $\mathbf{b}^\star$ and beam pairs ended with all other receive eigenmodes of $\mathrm{UE}_{k^\star}$ from $\mathcal{E}$.
6) Repeat step 3) to step 5) until $\mathrm{rank}(\mathbf{F}_{\mathrm{a}}) = l_{\mathrm{BS}}$ or the NAUTA does not increase by adding additional beams.

Detailed discussions on the algorithm and its complexity analysis can be found in [16].

VI. Simulation Results

To evaluate the performance of the proposed scheme, our simulations compare the above algorithm with JSDM [8,9] in terms of net average sum rate. To have a fair comparison of different schemes, we always use a least squares (LS) channel estimation during the training phase, and a zero-forcing digital precoder for the payload transmission. Meanwhile, the BS performs a greedy UE scheduling algorithm based on the instantaneous reduced-dimensional CSI to achieve the approximate optimal performance with different analog beamformer designs, respectively. Following the dominant characteristics of mm-wave propagation, we mainly focus on the MPCs interacting with a single scatterer. Fig. 2 illustrates an example of how to generate the synthetic channel profiles. We independently place the scatterers and UEs in an angular range (as seen from the BS) that we call support interval of DOD.

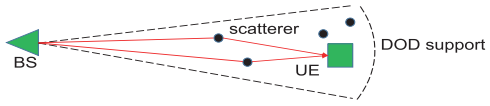

Fig. 2. Illustration of GSCM with UEs and scatterers

To activate scatterers for the channels between the BS and the UEs, we utilize the following probabilistic model: $P_{\mathrm{active}} = P_{\mathrm{UE,LOS}} \cdot P_{\mathrm{BS,LOS}}$, where P_{active} is the probability that a scatterer is active for an UE, which is the product of marginal probabilities that both link ends can "see" this scatterer. The marginal probability that a terminal has LOS propagation to the scatterer follows $P_{\mathrm{UE/BS,LOS}} = \min(d_1/d, 1)(1 - \exp(-d/d_2)) + \exp(-d/d_2)$, where d_1 and d_2 are modeling parameters, and d is the distance from the BS/UE to the scatterer. Settings of both d_1 and d_2 will be environment-dependent [17].[3] Unless otherwise specified, the parameter settings for channel model and system configuration are exhibited in Table I.

[3] [17] proposes the LOS probability model for mm-wave channels between BS and UE, whereas here we use this model to indicate the probability of LOS between a terminal and a scatterer.

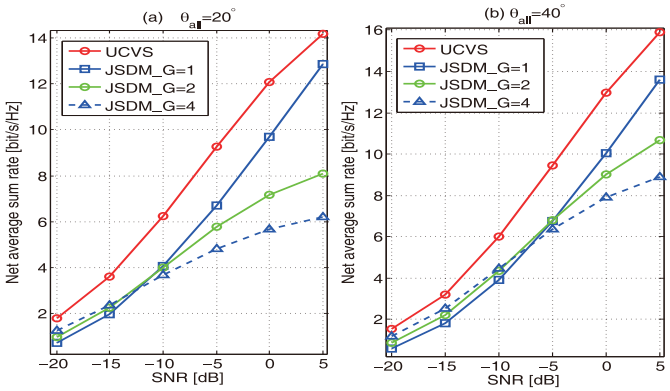

Fig. 3. Net average sum rate vs. ρ_{d} for $T_{\mathrm{cor}} = 20$ and $\gamma = 0.9$

TABLE I
Simulation Parameters

DOD support range	$\theta_{\mathrm{all}} = 20°$ or $40°$
LOS from BS to scatterer	$d_{1,\mathrm{BS2S}}$=24 m, $d_{2,\mathrm{BS2S}}$=45 m
LOS from UE to scatterer	$d_{1,\mathrm{UE2S}}$=2 m, $d_{2,\mathrm{UE2S}}$=10 m
Scatterer density	$\rho_{\mathrm{s}} = 0.01$
Energy threshold	$\gamma = 0.9$ or 1
Number of UEs	$K = 4$
No. of BS antennas and RF chains	$M = 64$, $l_{\mathrm{BS}} = 8$
No. of UE antennas and RF chains	$N = 8$, $l_{\mathrm{UE}} = 1$
Antenna spacing (in wavelength)	$D = \frac{1}{2}$

The coherence time ranges from 15 to 40 channel uses of the LTE standard, which approximately corresponds to mobile velocities from 5 m/s to 1.8 m/s at 60 GHz. For all simulation sets, we maintain the noise power and large scale loss to be unity, i.e. $\delta^2 = 1$, $L_k = 1, \forall k$. Therefore, transmit power ρ_{d} is equivalent to the signal-to-noise ratio (SNR) subsuming the impact of large scale loss. With the assumption of uncorrelated scattering, we independently generate $[\sigma_{kp}]$ following a uniform distribution within $[0,1]$, and then normalize them to satisfy $\sum_{p=1}^{P_k} \sigma_{kp}^2 = 1, \forall k$. Given locations of UEs and scatterers, we randomly generate UE-scatterer association graphs following the probabilistic model, based on which we can obtain double directional channel descriptions by assuming ULA at both link ends. The net average sum rates exhibited are all obtained by averaging over 100 UE drops, each of which consists of 20 independent realizations of UE-scatterer association graph and 50 independent realizations of small scale fading for each graph.

We first fix the coherence time T_{cor} to be 20 and the threshold parameter γ to be 0.9, then investigate the behavior of the net sum rate varying with SNR as Fig. 3 exhibits. Note that since there is no clear conclusion on the optimal UE grouping for JSDM in [8,18], we make comparisons with the JSDM scheme under different UE groupings, where K-means clustering to group UEs with similar channel covariance is applied [18]. We can observe that for both DOD support intervals, grouping all UEs together is optimal in the high SNR regime, since the channel-covariance-based analog precoder in JSDM cannot fully eliminate the inter-group interference, and forming more user groups will make the system operate in

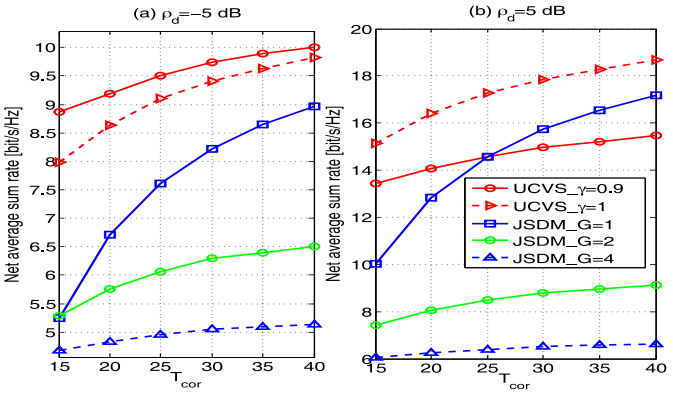

Fig. 4. Net average sum rate vs. T_{cor} for $\theta_{\text{all}} = 20°$

interference-limited mode. However, for the low SNR regime, the system tends to be noise-limited, and using more UE groups introduces additional gains from training cost reduction and thus obtains better performance. With $\theta_{\text{all}} = 40°$, we can observe that the impact of user grouping for JSDM is smaller, which is because dropping scatterers and UEs in a wider DOD support range leads more UE channels to be spatially orthogonal. Incorporating non-orthgonal training, STDT phase, and user-centric beamformer optimization, the proposed UCVS scheme outperforms JSDM with the optimal UE grouping setting in both cases. Fig. 4 shows the net average sum rate as a function of the coherence time under different settings of ρ_{d} for $\theta_{\text{all}} = 20°$. For the proposed UCVS scheme, we also investigate different settings of threshold parameter γ, i.e. 0.9 or 1. Comparing Fig. 4(a) and 4(b), we notice that appropriate settings of γ are scenario-dependent. Specifically, for the low SNR regime, the UCVS with a relatively smaller $\gamma = 0.9$ generates a sparser beam measure table and obtains more gains from the reduction of training overhead, while for the high SNR regime in Fig. 4(b), the interference-limited system is more sensitive to the threshold parameter, since striking out "weak" beam pair edges may generate nontrivial pilot contamination in the training phase and inter-user interference during data transmission, whose performance can be even worse than that of optimal JSDM as long as T_{cor} is large enough.

In conclusion, for the interesting range of parameter settings in the mm-wave systems, i.e., operating at SNR below 0 dB and coherence time below 50 channel uses, the proposed UCVS exhibits significant performance advantage over JSDM, e.g., more than 38% when $\rho_{\text{d}} = -5$ dB and $T_{\text{cor}} = 20$ as Fig. 4(a) shows. Meanwhile, for the large coherence time and high SNR regime, which is usually out of the scope of mm-wave systems, the proposed scheme with appropriate threshold setting still outperforms the state-of-the-art method in Fig. 4(b).

VII. Conclusions

In this paper, we built up an optimization framework based on the user-centric virtual sectorization for the implementation of massive MIMO systems in FDD mode, which incorporates three coupled optimization tiers with different time scales, including analog beamforming, training resource allocation, and digital beamforming, respectively. A UE-specified "virtual sectorization" employs the *simultaneous training-data transmission* (STDT) phase and *non-orthogonal beam training* (NOBT) to fully exploit the mm-wave channel characteristics. Heuristic low-complexity algorithms were devised to approach the suboptimal solution of analog beamformer design. Simulations revealed significant gains of the proposed scheme over state-of-the-art methods in typical mm-wave channels.

Acknowledgment

The authors would like to thank Prof. Giuseppe Caire, Dr. Shilpa Talwar, Dr. Nageen Himayat, and Dr. Roya Doostneyad for helpful discussions. Part of this work was financially supported by Intel, and by the National Science Foundation.

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, Nov. 2010.

[2] T. Rappaport, R. Heath Jr, R. Daniels, and J. Murdock, *Millimeter wave wireless communications*. Pearson Education, Sep. 2014.

[3] R. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Select. Topics Signal Processing*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[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, Nov. 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, Dec. 2006.

[6] A. Molisch, V. Ratnam, S. Han, Z. Li, S. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO-a survey," *arXiv preprint arXiv:1609.05078*, Sep. 2016.

[7] G. Matz, "Statistical characterization of non-WSSUS mobile radio channels," *e&i Elektrotechnik und Informationstechnik*, vol. 122, no. 3, pp. 80–84, Mar. 2005.

[8] 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, Oct. 2013.

[9] C. Sun, X. Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, "Beam division multiple access transmission for massive MIMO communications," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2170 –2184, Jun. 2015.

[10] Z. Li, N. Rupasinghe, O. Bursalioglu, C. Wang, H. Papadopoulos, and G. Caire, "Directional training and fast sector-based processing schemes for mmwave channels," *arXiv preprint arXiv:1611.00453*, Nov. 2016.

[11] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[12] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of gaussian MIMO broadcast channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[13] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 361–374, Feb. 2006.

[14] Z. Li, S. Han, and A. Molisch, "Optimizing channel-statistics-based analog beamforming for millimeter-wave multi-user massive MIMO downlink," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4288–4303, Jul. 2017.

[15] Z. Chen, X. Hou, and C. Yang, "Training resource allocation for user-centric base station cooperation networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2729–2735, Apr. 2016.

[16] Z. Li, S. Han, and A. Molisch, "User-centric virtual sectorization for millimeter-wave massive MIMO downlink," *to appear in IEEE Transactions on Wireless Communications*.

[17] S. Hur, S. Baek, B. Kim, Y. Chang, A. Molisch, T. Rappaport, K. Haneda, and J. Park, "Proposal on millimeter-wave channel modeling for 5G cellular system," *IEEE J. Select. Topics Signal Processing*, vol. 10, no. 3, pp. 454–469, Apr. 2016.

[18] 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*, Jun. 2014.