This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2767580, IEEE Transactions on Wireless Communications

User-Centric Virtual Sectorization for Millimeter-Wave Massive MIMO Downlink

Zheda Li, Shengqian Han, *Member, IEEE*, and Andreas F. Molisch, *Fellow, IEEE*

***Abstract*—The high training cost of massive multiple-input multiple-output (MIMO) systems motivates the use of hybrid digital/analog (HDA) beamforming structures. This paper considers the joint design of analog beamformers when both link ends of a millimeter (mm)-wave massive MIMO system are equipped with such HDA structures. We aim to maximize the multi-user (MU) MIMO net average throughput of the downlink in an Frequency Division Duplex (FDD) system. To achieve this, we develop an optimization framework, namely *user-centric virtual sectorization* (UCVS), to explore 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 donwlink 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 typical scenarios of mm-wave communications.**

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

I. Introduction

DUE to the large available bandwidth [1], the millimeter (mm)-wave spectrum will be an important component of fifth generation (5G) cellular communications. However, challenges brought by channel characteristics at such high frequencies, e.g., large pathloss, impede a direct extension of the legacy systems. The requirement of overcoming severe channel conditions for mm-wave systems ties seamlessly into another important candidate technology for 5G systems, namely massive multiple-input multiple-ouput (MIMO), which employs dozens or hundreds of antenna elements at the base station (BS) to enable high multiuser capacity, simplify signal processing, and enhance beamforming gain [2, 3]. Nevertheless, combining massive MIMO with a mm-wave system in a cost- and energy-effective way is not straightforward [4, 5].

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, especially at mm-wave frequencies. A promising solution to these problems lies in the concept of hybrid transceivers, which uses a combination of analog beamformers in the RF domain, together with a smaller number of RF chains. This concept was first introduced by one of the authors and collaborators in [6, 7]. While formulated originally for MIMO with arbitrary number of antenna elements, the approach is applicable in particular to massive MIMO, and in that context interest in hybrid transceivers has surged over the past years, e.g., [8–14] and the references in [15].

Many beamformer optimizations for massive MIMO with hybrid digial/analog (HDA) structure assume the full acquisition of, and adaptation to, instantaneous channel state information (CSI). However, it is nontrivial to obtain the full CSI with extremely large arrays, especially for mm-wave channels. Main challenges lie in the following aspects: 1) shorter coherence time at high carrier frequency caused by the larger Doppler spread, 2) for a single channel use of training, the number of sample measurements is less than that of the conventional fully digital system due to lack of RF chains. Achieving the same amount of measurements as a fully digital system requires extending the training duration, worsening the dilemma caused by 1).

Even in a system without hardware constraints, i.e., in a fully digital implementation, the short coherence time at mm-wave frequencies constitutes a problem for massive MIMO. Considering a large-array BS serving single-antenna user equipments (UEs), [2] suggests channel-reciprocity-based uplink training in a time-division-duplexing (TDD) mode to avoid the large overhead brought by the downlink training in frequency-division-duplexing (FDD) mode. However, the large pathloss at mm-wave frequencies necessitates *both* link ends to be equipped with multiple antenna elements in order to exploit beamforming gains. If the total number of antenna elements from all UEs is then the same order as that of the BS, the significant burden of uplink training at antenna level will also make massive MIMO based on instantaneous CSI infeasible. Therefore, analog beamforming has to be used at both link ends during the training phase to reduce the effective channel dimension without the full knowledge of instantaneous CSI.

Two major research directions dealing with the above challenges have been investigated in the past few years: 1) compressive-sensing-based channel estimation plus analog beamforming optimization [8, 9], 2) channel-statistics-based analog beamforming design [16]. In this paper, we focus on the latter 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 [17], 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-

Z. Li and A. F. Molisch are with the Department of Electrical Engineering, University of Southern California, Los Angeles, California 90089 (e-mail: {zhedali, molisch}@usc.edu). S. Han is with the School of Electronics and Information Engineering, Beihang University, Beijing, 100191, P. R. China (e-mail: sqhan@buaa.edu.cn).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2767580, IEEE Transactions on Wireless Communications

forcing, for the MU-MIMO system can be easily employed.

Joint spatial division multiplexing (JSDM) [16] designs the analog precoder at the BS as a function of the channel covariance matrices, which bears some formal resemblance to our investigations. However, its sector-specific design, which enforces orthogonality between different groups of UEs, will null out signals from common scattereres, and thus may sacrifice not only significant beamforming gain but also spatial multiplexing gain (see Section III-A for details). In this paper, we intend to design channel-statistics-based analog beamformers 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 [18]. 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). In [19], the tradeoff of training duration and achievable rate with HDA structure at the BS side is investigated, but retains the conventional orthogonal training scheme.

Our proposed user-centric virtual sectorization (UCVS) scheme exploits the UCBC to form exclusive or partially overlapped virtual sectors for different UEs, and the NOBT to save overall training overhead. Moreover, periods of downlink training for different UEs may end at different time slots in UCVS. Therefore, for a particular UE whose effective CSI is obtained by the BS before the completion of the training phase, we may launch the downlink data transmission to it. This *simultaneous training-data transmission* (STDT) phase is also considered in [20] for the uplink, where orthogonal training among UEs is assumed, and the interference between training signal and payload data is mitigated by using successive interference cancellation based on the orthogonality between the independent and identically distributed (i.i.d.) UE channels. The mm-wave channel with highly directional characteristics is generally not i.i.d. [21]. 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.

To the best of our knowledge, there is little work exploring the joint optimization of training resource allocation and channel-statistics-based analog beamformer design, and we are trying to close this gap. The main contributions of this paper are summarized below:

- We develop an optimization framework for the mm-wave massive MIMO downlink, where channel-statistics-based UCBC, NOBT, and implied STDT phase are introduced to combat the fast variation channel. A UCVS scheme is realized by exploring the highly directional and sparse characteristics of mm-wave channels.
- Given an analog beamforming design, we formulate the problem to optimize the training resource allocation from a graph theory perspective. An algorithmetic method is developed for an approximate solution of training resource allocation.
- We account for the coupling effect of training resource allocation and analog beamformer optimization to jointly maximize the overall net average throughput. We devise efficient algorithms to realize user-centric beamformers. Employing generic mm-wave channel models, simulations demonstrate the advantages of the proposed scheme over the state-of-the-art scheme under various typical parameter settings.

The rest of the paper is organized as follows. In Section II, the system and spatial channel model are presented. In Section III, we first review the concept of JSDM, then elaborate on the essential idea of UCBC. Section IV presents stepwise procedures of the UCBC scheme, and summarizes the developments of the problem formulation, based on which algorithm developments are exhibited in Section V. Simulations results are presented in Section VI before drawing the conclusions in Section VII.

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 $\mathcal{X}$, 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. $\mathrm{diag}([x_i]_{i=1}^n)$=$\mathrm{diag}(x_1, ..., x_n)$, represents a diagonal matrix, while $\mathrm{diag}([\mathbf{X}_i]_{i=1}^n)$=$\mathrm{diag}(\mathbf{X}_1, ..., \mathbf{X}_n)$ is a block diagonal matrix. $\mathrm{diag}(\mathbf{X})$ denotes a diagonal matrix with the diagonal elements of $\mathbf{X}$ on its diagonal line. $\mathbf{X}^{\frac{1}{2}}$ denotes the Cholesky decomposition. $\mathbf{I}_n$ is the n-by-n identity matrix. $\mathcal{CN}(\mathbf{m}, \mathbf{K})$ is the circularly symmetric complex Gaussian distribution with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{K}$. $\mathbb{E}[\cdot]$ 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., $l_{\mathrm{UE}} = 1$. With HDA structures at both ends, we have $M > l_{\mathrm{BS}}$ and $N > l_{\mathrm{UE}}$. In the data transmission of the downlink, the BS broadcasts the beamformed data streams to the 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_i is

$$\hat{\mathbf{x}}_i = \mathbf{w}_{\mathrm{a}i}^\dagger \mathbf{H}_i \mathbf{F}_{\mathrm{a}} \mathbf{F}_{\mathrm{d}} \mathbf{x} + \mathbf{w}_{\mathrm{a}i}^\dagger \mathbf{n}_i, \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^{K\times 1}$ is the sample symbol vector following the distribution $\mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$, $\mathbf{H}_i \in \mathbb{C}^{N\times M}$ denotes the transfer matrix of UE_i whose modeling will be elaborated later, $\mathbf{F}_{\mathrm{a}} \in \mathbb{C}^{M\times l_{\mathrm{BS}}}$ and $\mathbf{F}_{\mathrm{d}} \in \mathbb{C}^{l_{\mathrm{BS}}\times K}$ denote the analog and digital precoder, respectively, $\mathbf{w}_{\mathrm{a}i} \in \mathbb{C}^{N\times 1}$ is the analog combiner at UE_i, and $\mathbf{n}_i \in \mathbb{C}^{N\times 1}$ indicates the noise vector at UE_i following $\mathcal{CN}(\mathbf{0}, \delta^2\mathbf{I}_N)$. Note that we consider the fully-connected hybrid beamforming structure, where each RF chain has access to all antenna elements. For ease of notation,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2767580, IEEE Transactions on Wireless Communications

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 at mm-wave band, multipath components (MPC) suffering multiple diffractions have much lower power than those at cellular band, leading to limited scattering [22]. Due to this effect, and the existence of sparse dominant MPCs, the Kronecker channel model [23], which is popularly used for below 6 GHz channels, cannot effectively represent the coupling effect between the directions of departure (DOD) and directions of arrival (DOA) of the MPCs. Consequently, we consider the following double directional channel description

$$\mathbf{H}_i=\frac{1}{\sqrt{L_i}}\sum_{p=1}^{P_i} g_{ip}\mathbf{a}_{\mathrm{UE}}(\theta_{ip})\mathbf{a}_{\mathrm{BS}}^{\dagger}(\phi_{ip}), \tag{2}$$

where P_i is the number of MPCs from the BS to UE$_i$, L_i is the large scale loss, including path loss and shadowing, and $g_{ip} \sim \mathcal{CN}(0,\sigma_{ip}^2)$ reflects the small scale fading of the p-th MPC. 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 MPC often fulfills the condition of Rayleigh fading, which is also widely used in the mm-wave literature [8,14], as well as the 3GPP channel model [24] (which implicitly uses zero-mean Gaussian by using a large number of equal-powered subpaths per cluster). $\mathbf{a}_{\mathrm{UE}} \in \mathbb{C}^{N\times 1}$ and $\mathbf{a}_{\mathrm{BS}} \in \mathbb{C}^{M\times 1}$ indicate the steering vectors of DOA θ and DOD ϕ, respectively. If uniform linear arrays (ULA) are assumed at both link ends, the steering vector $\mathbf{a}_{\mathrm{UE}}(\theta)$ becomes

$$\mathbf{a}_{\mathrm{UE}}(\theta) = [1, \exp(j\tfrac{2\pi}{\lambda}d\sin\theta), \exp(j\tfrac{2\pi}{\lambda}2d\sin\theta), ..., \exp(j\tfrac{2\pi}{\lambda}(N-1)d\sin\theta)]^T, \tag{3}$$

where λ is the wavelength and d denotes the antenna spacing. The steering vector at the BS, $\mathbf{a}_{\mathrm{BS}}(\phi)$, can be written in a similar fashion.

Assuming that each MPC exhibits independent fading,[1] we have $\sum_{p=1}^{P_i}\sigma_{ip}^2 = 1, \forall i$. With the block fading assumption, $[g_{ip}]$ varies across coherence blocks, while $[\sigma_{ip}]$ remains the same within the stationarity time of the second order channel statistics. Defining steering matrices $\mathbf{A}_{\mathrm{UE},i} \triangleq [\mathbf{a}_{\mathrm{UE}}(\theta_{i1}), \mathbf{a}_{\mathrm{UE}}(\theta_{i2}), ..., \mathbf{a}_{\mathrm{UE}}(\theta_{iP_i})]$ and $\mathbf{A}_{\mathrm{BS},i} \triangleq [\mathbf{a}_{\mathrm{BS}}(\phi_{i1}), \mathbf{a}_{\mathrm{BS}}(\phi_{i2}), ..., \mathbf{a}_{\mathrm{UE}}(\phi_{iP_i})]$, we can rewrite (2) as

$$\mathbf{H}_i = \frac{1}{\sqrt{L_i}}\mathbf{A}_{\mathrm{UE},i}\Sigma_i\bar{\mathbf{G}}_i\mathbf{A}_{\mathrm{BS},i}^{\dagger}, \tag{4}$$

where $\Sigma_i \triangleq \mathrm{diag}([\sigma_{ip}]_{p=1}^{P_i})$, $\bar{\mathbf{G}}_i \triangleq \mathrm{diag}([\bar{g}_{ip}]_{p=1}^{P_i})$, and $\bar{g}_{ip} \triangleq \frac{g_{ip}}{\sigma_{ip}}, \forall i,p$. Instead of treating each coherence block isotropically, we propose to design analog beamformers based on the knowledge of angular power spectra, including $[\mathbf{A}_{\mathrm{UE},i}]$, $[\mathbf{A}_{\mathrm{BS},i}]$, and $[\Sigma_i]$, which remain approximately the same within the stationarity region of channel statistics. Note that the acquisition of the long-term CSI does not require the geometry locations of terminals or scatteres, but rather efficient estimation algorithms: e.g., [25] utilizes coprime sampling method to track the channel subspace with a hybrid beamforming structure, which can be used to investigate the directional characteristics, either through Bartlett beamforming, or through various high-resolution techniques [26, 27]. Meanwhile, the cost of long-term CSI acquisition is negligible after normalization by the stationarity time of channel statistics. 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 $\mathbf{K}_{\mathrm{BS},i} \triangleq \mathbb{E}[\mathbf{H}_i^{\dagger}\mathbf{H}_i] = \frac{N}{L_i}\mathbf{A}_{\mathrm{BS},i}\Sigma_i^2\mathbf{A}_{\mathrm{BS},i}^{\dagger}$ and $\mathbf{K}_{\mathrm{UE},i} \triangleq \mathbb{E}[\mathbf{H}_i\mathbf{H}_i^{\dagger}] = \frac{M}{L_i}\mathbf{A}_{\mathrm{UE},i}\Sigma_i^2\mathbf{A}_{\mathrm{UE},i}^{\dagger}$.

Since analog beamformers, i.e. $[\mathbf{w}_{ai}]$ and $\mathbf{F}_{\mathrm{a}}$, remain the same across multiple coherence blocks, we can view the instantaneous effective channel between BS and UE$_i$ as $\bar{\mathbf{h}}_i \triangleq \mathbf{F}_{\mathrm{a}}^{\dagger}\mathbf{H}_i^{\dagger}\mathbf{w}_{ai}$, whose dimension is reduced to the number of RF chains, i.e. $l_{\mathrm{BS}} \times 1$. Therefore, channel-statistics-based analog beamformers significantly alleviate the burden of instantaneous CSI acquisition for both FDD and TDD systems. The covariance of the effective channel $\bar{\mathbf{h}}_i$ can be expressed as

$$\begin{aligned}\bar{\mathbf{K}}_{\mathrm{BS},i} &\triangleq \mathbb{E}[\bar{\mathbf{h}}_i\bar{\mathbf{h}}_i^{\dagger}] \\ &= \tfrac{1}{L_i}\mathbf{F}_{\mathrm{a}}^{\dagger}\mathbf{A}_{\mathrm{BS},i}\Sigma_i\,\mathrm{diag}(\mathbf{A}_{\mathrm{UE},i}^{\dagger}\mathbf{w}_{ai}\mathbf{w}_{ai}^{\dagger}\mathbf{A}_{\mathrm{UE},i})\Sigma_i\mathbf{A}_{\mathrm{BS},i}^{\dagger}\mathbf{F}_{\mathrm{a}} \\ &= \mathbf{F}_{\mathrm{a}}^{\dagger}\tilde{\mathbf{K}}_{\mathrm{BS},i}\mathbf{F}_{\mathrm{a}},\end{aligned} \tag{5}$$

where we define the combiner-projected channel covariance as $\tilde{\mathbf{K}}_{\mathrm{BS},i} \triangleq \mathbb{E}[\mathbf{H}_i^{\dagger}\mathbf{w}_{ai}\mathbf{w}_{ai}^{\dagger}\mathbf{H}_i]$.

Concerning the complexity of a practical massive MIMO system, we assume that 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, or can be implemented by means of lense antennas. Therefore, $\mathbf{F}_{\mathrm{a}}$ becomes a function of the combiner-projected channel covariance matrices and the DFT codebook, i.e. $\mathbf{F}_{\mathrm{a}} = f_{\mathrm{BS}}(\mathbf{\Omega}_M, [\tilde{\mathbf{K}}_{\mathrm{BS},i}])$, 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 [21, 28], which leads to 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. $\mathbf{w}_{ai} = f_{\mathrm{UE}_i}(\mathbf{K}_{\mathrm{UE},i}), \forall i$, since the number of UE antenna elements is typically smaller than that at the BS.

Due to the highly directional channel characteristics of mm-wave channels, the analog combiner (receive beam) at UE$_i$ may not be capable of collecting significant energy from all transmit beams of analog precoder $\mathbf{F}_{\mathrm{a}}$, which is designed to serve multiple UEs. To illustrate this concept, Fig. 1 exhibits a beam measure table, where scatter dots indicate the MPCs illuminated by different beam pairs, e.g., UE$_1$ collects most of its energy from transmit beam $\mathbf{b}_2$ (note that this table can be interpreted as the beam coupling matrix in the Weichselberger channel model [23]). Later, we will show how to acquire this table in Section III-B. We can observe that the effective channels of UEs are approximately orthogonal to each other, which motivates us to perform parallel beam training and further reduce the overhead. Similarly, TDD systems can also

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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2767580, IEEE Transactions on Wireless Communications

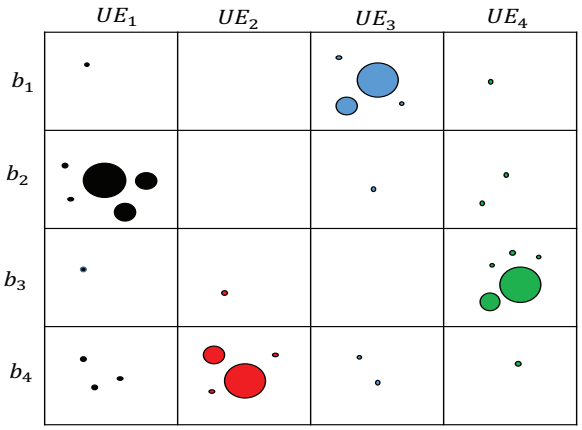


Fig. 1: Transmit/receive beam measure table between transmit beams $[\mathbf{b}_i]_{i=1}^4$ and 4 UEs, where each UE forms its own analog combiner. Scatter dots indicate MPCs, different sizes denote average power $[\sigma_{ip}]$, and different colors separately represent different UEs.

benefit from the parallel uplink training for different UEs, as Fig. 1 shows.

III. Overview of User-Centric Virtual Sectorization

The main objective of this paper is to provide a user-centric optimization framework that incorporates the concern of training overhead reduction. In this section, we will first give a recap of JSDM, which provides a sector-centric analog precoder design based on the channel statistics. Later, comparing JSDM and UCVS by illustrating some toy examples, we elaborate on the usefulness of our proposed idea in typical scenarios of mm-wave communications and also explain its working mechanism conceptually.

A. Recap of JSDM

The JSDM-based framework can be interpreted as a *sector-centric beam clustering*, where the BS individually forms covariance-based analog precoders to illuminate each "sector", while different UE groups tend to be semi-orthogonal to each other. Specifically, single-antenna UEs with similar channel covariance are grouped together and inter-group interference is suppressed by an analog precoder based on the approximate block diagonalization method, which creates multiple "virtual sectors".

Treating each RF chain at a BS with M antenna elements as an individual "BS", we can view JSDM as a coordinated multi-point (CoMP) transmission scheme [29] under particular constraints: an exclusive set of "BSs" serves its corresponding UE group in joint transmission (JT) mode, meanwhile, it also needs to work in coordinated beamforming (CB) mode with other groups, suppressing the leakage interference. However, the enforced constraint may lead to a solution that is away from net sum rate maximization. For example, Fig. 2 exhibits a 2-path channel model of three UEs, where both UE_1 and UE_2 have the line of sight (LOS) propagation to the BS.

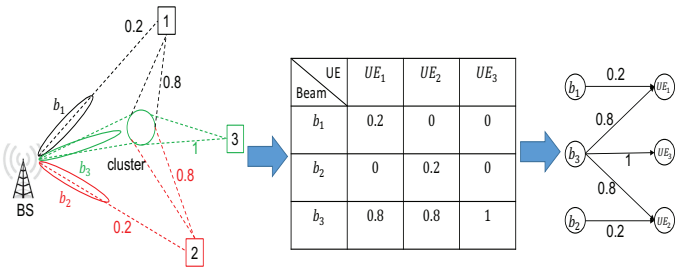


Fig. 2: Toy example of 3-UE channel: 1) both UE_1 and UE_2 have LOS propagation to the BS, all three UEs "see" a common cluster that couples them, and normalized average power of MPCs, i.e. $[\sigma_{ip}^2]$, is also labeled next to dashed lines; 2) generation of beam pair bipartite graph from beam measure table.

Additionally, all UEs share a common cluster. Assume three transmit beams illuminating all MPCs of this network: if we place UE_1 and UE_2 into separate groups, the BS has to null out $\mathbf{b}_3$ following the orthogonality principle of JSDM across different groups. Although parallel training can be implemented and simultaneously serve two UEs (channels of $\mathbf{b}_1$ to UE_1 and $\mathbf{b}_2$ to UE_2 tend to be quasi-optical, which are orthogonal to each other), we not only lose significant beamforming gain since the average power from $\mathbf{b}_3$ to UE_1 and UE_2 is 0.8, but also lose one degree of freedom (DoF) by generating a poor effective channel condition for UE_3, which lies in the sector edge between groups.

B. Basic Idea of User-Centric Virtual Sectorization

Maximizing the net sum rate of UEs necessitates the joint consideration of training costs, beamforming gains, and overall spatial multiplexing gains. We generalize the JSDM-like sector-centric beam clustering to a UE-centric one, where the BS forms a cluster of transmit beams for each scheduled UE individually. Unlike the constraint of JSDM that the common set of beams is assigned to UEs within the same group, while UEs in different groups exhibit exclusive beam clusters, we allow partially overlapped beam clusters among UEs.

Define the UE-specific analog precoder as $\mathbf{B}_i \in \mathbb{C}^{M\times l_i}, \forall i$, where l_i is the number of BS RF chains used to serve UE_i. In the toy example exhibited in Fig. 2, two interesting scenarios of UE grouping can be developed following the principle of JSDM [16]: 1) separate UE_1 and UE_2 into two groups, therefore $\mathbf{B}_1{=}\mathbf{b}_1$, $\mathbf{B}_2{=}\mathbf{b}_2$, and $\mathbf{B}_3{\approx}\mathbf{0}$; 2) Group all three UEs together, and let $\mathbf{B}_1{=}\mathbf{B}_2{=}\mathbf{B}_3{=}[\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3]$. Note that in scenario 1), since UE_3 lies in the sector edge as we mentioned before, its analog precoder is approximately zero. Comparing both scenarios, we can simultaneously serve UE_1 and UE_2 with one pilot dimension for parallel training at the expense of beamforming gains from the common cluster in scenario 1), while in scenario 2), all three UEs can be scheduled at the cost of three pilot dimensions for orthogonal beam training. However, there is no explicit conclusion as to which scenario, i.e., UE grouping, is optimal to maximize the net sum rate in [16, 21, 28, 30–32].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2767580, IEEE Transactions on Wireless Communications

Meanwhile, there is another scenario that is not covered by JSDM, say scenario 3), where we have $\mathbf{B}_1$=[$\mathbf{b}_1, \mathbf{b}_3$], $\mathbf{B}_2$=[$\mathbf{b}_2, \mathbf{b}_3$], and $\mathbf{B}_3$=$\mathbf{b}_3$. Although the overall analog precoder $\mathbf{F}_a$ remains the same for both scenario 2) and 3), orthogonal beam training is not necessary for scenario 3). Since the channels of $\mathbf{b}_1$ to UE_1 and $\mathbf{b}_2$ to UE_2 are approximately orthogonal to each other, we can assign the same pilot dimension to $\mathbf{b}_1$ and $\mathbf{b}_2$, which will not cause the problem of pilot contamination [2]. Therefore, we can use only two pilot dimensions to complete the training of three beams by utilizing the spatial orthogonality between effective channels.

Before proceeding to specific problem formulations in Section IV, we explain core concepts that are introduced by our scheme.

*1) **Beam pair bipartite graph**:* With the assumption of DFT-codebook-based design, the optimization of analog precoder at BS becomes a selection problem, which falls into the realm of integer programming. Given the UE-side channel covariance $\mathbf{K}_{\text{UE},i}$, we can write its eigen decomposition as $\mathbf{K}_{\text{UE},i} = \mathbf{E}_{\text{UE},i}\mathbf{\Lambda}_{\text{UE},i}\mathbf{E}^{\dagger}_{\text{UE},i}$, where $\mathbf{E}_{\text{UE},i} = [\mathbf{r}_{i1}, \mathbf{r}_{i2}, ..., \mathbf{r}_{ir_i}]$ is a semi-unitary matrix with rank $r_i \le \min(N, P_i)$, $\mathbf{r}_{ij}$ denotes the j-th receive eigenmode of $\mathbf{E}_{\text{UE},i}$, and $\mathbf{\Lambda}_{\text{UE},i}$ aligns eigenvalues of $\mathbf{K}_{\text{UE},i}$ on its diagonal. Therefore, we can build up a measure matrix between DFT beam tones and receive eigenmodes as follows:

$$\begin{aligned}&\mathbf{S}(m, j + \textstyle\sum_{k=1}^{i-1} r_k)=\mathbf{b}_m^{\dagger}\tilde{\mathbf{K}}_{\text{BS},i,j}\mathbf{b}_m, \qquad (6)\\ &\tilde{\mathbf{K}}_{\text{BS},i,j}\triangleq\mathbb{E}[\mathbf{H}_i^{\dagger}\mathbf{r}_{ij}\mathbf{r}_{ij}^{\dagger}\mathbf{H}_i]\\ &= \tfrac{1}{L_i}\mathbf{A}_{\text{BS},i}\mathbf{\Sigma}_i\,\text{diag}(\mathbf{A}^{\dagger}_{\text{UE},i}\mathbf{r}_{ij}\mathbf{r}_{ij}^{\dagger}\mathbf{A}_{\text{UE},i})\mathbf{\Sigma}_i\mathbf{A}^{\dagger}_{\text{BS},i},\end{aligned}$$

where $\mathbf{b}_m$ denotes the m-th column of $\mathbf{\Omega}_M$, and $\tilde{\mathbf{K}}_{\text{BS,i,j}}$ represents the BS side channel covariance of UE_i projected by $\mathbf{r}_{ij}$. $\mathbf{S} \in \mathbb{R}_{>0}^{M\times\sum_{i=1}^{K} r_i}$ indicates the measure matrix between M DFT beams and receive eigenmodes of all UEs. The entry indexed by $(m, j + \sum_{k=1}^{i-1} r_k)$ denotes the average channel gain between m-th DFT beam tone and j-th receive eigenmode of UE_i, where j ranges from 1 to r_i.

For the toy example exhibited in Fig. 2, we simply let $L_i = 1, \forall i$, and $N = 1$, while the steering matrices consist of normalized DFT columns with $\mathbf{A}_{\text{BS},1}$=[$\mathbf{b}_1, \mathbf{b}_3$], $\mathbf{A}_{\text{BS},2}$=[$\mathbf{b}_2, \mathbf{b}_3$], and $\mathbf{A}_{\text{BS},3}$=$\mathbf{b}_3$. Substituting the above parameter set into (6) generates the beam measure table in Fig. 2, where we only exhibit the measure table with effective transmit beams: $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$. Equipped with a single antenna element, UEs in the toy example receive omnidirectional signals. Therefore, we only have one receive eigenmode for each UE. To build the beam pair bipartite graph, we place nodes of transmit beams and UEs at left and right side, respectively. If the entry between a beam and a UE is non-zero, we connect the two nodes by a weighted edge.

If UEs with multiple antenna elements are able to resolve different MPCs, UE_i will have P_i receive eigenmodes, $\forall i$, which leads to the development of beam pair bipartite graph as Fig. 3 exhibits. To display a toy example, we simply let [$\mathbf{A}_{\text{UE},i}$] also consist of normalized DFT columns, which then become receive eigenmodes. With directional beams at both ends, we can observe that the beam measure table becomes even sparser, based on which non-orthogonal beam training

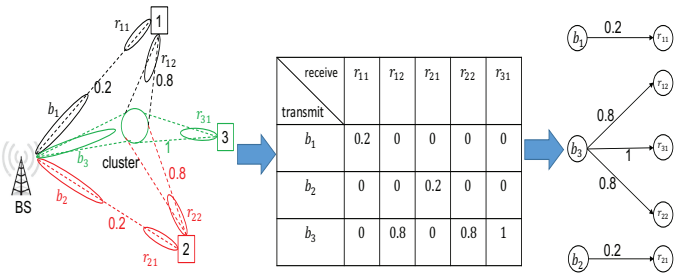


Fig. 3: Generation of beam pair bipartite graph when there are multiple receive eigenmodes. Both UE_1 and UE_2 exhibit two receive eigenmodes, while UE_3 has only one pointing to the common cluster.

can be utilized to reduce the overhead cost. On the other hand, [13,31] design analog combiners at the UEs by selecting its strongest eigenmode individually, which may be far away from the maximization of the net sum rate in a mm-wave channel. For example, for the beam pair bipartite graph shown in Fig. 3, if we let all three UEs point toward to the common cluster, which exhibits the largest weights for all of them, the DoF of the analog-combiner-projected MU-MIMO channel will be only one. Therefore, we will also investigate the joint optimization of analog combiners based on the beam pair bipartite graph.

In reality, there will be no entry with exact zero-value in the beam measure table, which implies a fully connected beam pair bipartite graph. However, after an appropriate thresholding, we strike out weak edges with weight below a threshold, say the noise floor, so that we obtain an effective bipartite graph as in Fig. 2 and Fig. 3. The threshold parameter plays an important role in the beam clustering, which will be elaborated in Section V. On one hand, striking weak beam pairs generates a sparser beam measure table, which needs less pilot dimensions for training. On the other hand, the effective bipartite graph should maintain dominant directional characteristics of the multi-user channel, or we will suffer severe pilot contamination and inter-user interference (see below).

*2) **Non-orthogonal Beam Training** (NOBT):* Given the analog beamformers at both ends, the beam measure table $\mathbf{S}$ with the dimension $M \times \sum_{i=1}^{K} r_i$ is reduced to an effective one, denoted by $\bar{\mathbf{S}}$, projected by analog beamformers, where $\bar{\mathbf{S}}$ has dimension $l_{\text{BS}} \times K$. Based on $\bar{\mathbf{S}}$, we can develop the beam cluster of an individual UE, containing all transmit beams connected to it. For example, let us revisit the toy example exhibited in Fig. 2. Considering a system with $l_{\text{BS}} = 3$ and N=l_{UE}=1, we assume that the optimized analog precoder $\mathbf{F}_a$ is [$\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$]. Therefore, Fig. 2 is equivalent to its reduced beam measure table $\bar{\mathbf{S}}$. The analog precoders (beam clusters) of the UEs are $\mathbf{B}_1$=[$\mathbf{b}_1, \mathbf{b}_3$], $\mathbf{B}_2$=[$\mathbf{b}_2, \mathbf{b}_3$], and $\mathbf{B}_3$=$\mathbf{b}_3$.

The training overhead cost depends on the minimum number of necessary orthogonal pilot dimensions. Define the set of UEs whose beam cluster contains the i-th transmit beam as $\mathcal{K}_i$, i.e. $\mathcal{K}_i = \{k|\mathbf{b}_i \in \mathbf{B}_k\}$. Therefore, if $\mathcal{K}_i \cap \mathcal{K}_j \neq \emptyset$, $\forall i \neq j$, we cannot schedule $\mathbf{b}_i$ and $\mathbf{b}_j$ for training on the same pilot dimension, since any UE lying in the intersection set will encounter severe pilot contamination. However, consider

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2767580, IEEE Transactions on Wireless Communications

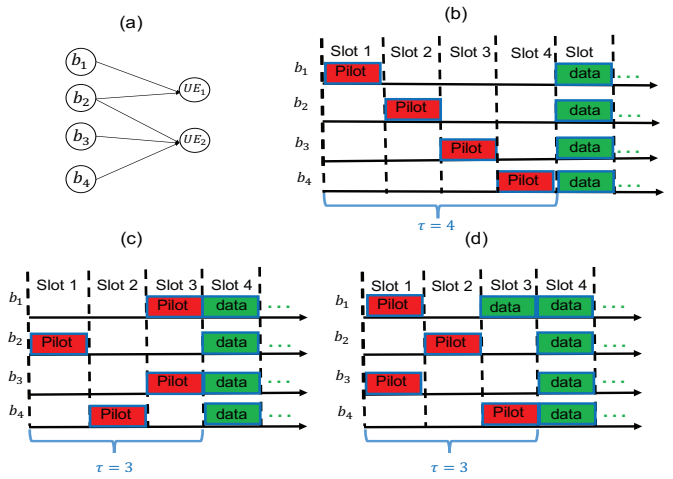


Fig. 4: Compare the training phase of JSDM and UCVS, where (a) is an example of reduced beam pair bipartite graph, (b) reflects the training process of the JSDM, while (c) and (d) represent the training periods of the UCVS with different training orders, respectively. τ is the duration of overall training window.

a set of beams $\mathcal{T}$, such that their served UE sets do not overlap: in that case, we can train them simultaneously, i.e. $\mathcal{T} = \{i | \mathcal{K}_i \cap \mathcal{K}_j = \emptyset, \forall j \in \mathcal{T} \setminus \{i\}\}$.[2] For example, in Fig. 2, we have $\mathcal{K}_1$={1}, $\mathcal{K}_2$={2}, and $\mathcal{K}_3$={1, 2, 3}. BS cannot train $\mathbf{b}_1$ and $\mathbf{b}_3$ (or $\mathbf{b}_2$ and $\mathbf{b}_3$) simultaneously, since $\mathcal{K}_1 \cap \mathcal{K}_3$={1} ($\mathcal{K}_2 \cap \mathcal{K}_3$={2}). On the other hand, $\mathbf{b}_1$ and $\mathbf{b}_2$ can be placed on the same pilot dimension, since $\mathcal{K}_1 \cap \mathcal{K}_2 = \emptyset$. The total number of orthogonal resource elements occupied by training can be reduced to 2 for the toy example, while JSDM suggested by [16, 21] will perform orthogonal training across $[\mathbf{b}_i]_{i=1}^3$, treated as intra-group transmit beams serving all three UEs. Detailed developments on the minimization of training cost can be found in Section V-A.

*3) **Simultaneous training-data transmission** (STDT):* Conventional cellular systems will start the data transmission phase after the completion of the training phase. However, in this paper, we propose a novel training scheme where the BS can "partially" launch the data transmission during the training window. We illustrate its mechanism conceptually by a toy example exhibited in Fig. 4.

To better clarify the STDT phase, we define the following sets, which will be used in the remainder of the paper. $\mathcal{K}_{\text{cc},t}$ 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. $\mathcal{B}_t = \{i | \mathbf{b}_i \in \cup_{k \in \mathcal{K}_{\text{cc},t}} \mathbf{B}_k, \mathbf{b}_i \notin \cup_{k \in \bar{\mathcal{K}}_{\text{cc},t}} \mathbf{B}_k\}$, 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.

With the reduced beam pair bipartite graph exhibited in Fig. 4a, JSDM will place UEs in the same group with the common analog precoder, i.e. $\mathbf{F}_{\text{a}} = [\mathbf{b}_1, ..., \mathbf{b}_4]$, and orthogonal beam training is implemented as Fig. 4b shows. However, with the partially overlapped beam clusters in UCVS shown in Fig. 4c and Fig. 4d, 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 $\mathcal{K}_1 \cap \mathcal{K}_3 = \emptyset$, $\mathbf{b}_1$, $\mathbf{b}_3$ can be trained simultaneously and we only need 3 orthogonal time slots to complete the training of 4 beams. For Fig. 4c, based on the association between transmit beams and UEs in Fig. 4, $\mathcal{K}_{\text{tr},1} = \{1, 2\}$, $\mathcal{K}_{\text{tr},2} = \{2\}$, and $\mathcal{K}_{\text{tr},3} = \{1, 2\}$, while $\mathcal{T}_{\text{tr},1} = \{2\}$, $\mathcal{T}_{\text{tr},2} = \{4\}$, and $\mathcal{T}_{\text{tr},3} = \{1, 3\}$. $\mathcal{K}_{\text{cc},t} = \emptyset, \forall t \leq 3$, and $\mathcal{K}_{\text{cc},4} = \{1, 2\}$, indicating both UEs complete beam training after the whole training window. Therefore, $\mathcal{B}_t = \emptyset, \forall t \leq 3$.

However, for Fig. 4d, where we swap the order of training $\mathbf{b}_1$, $\mathbf{b}_3$, and $\mathbf{b}_4$, an interesting observation is that $\mathcal{K}_{\text{cc},3} = \{1\}$, 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.

In summary, the NOBT phase exploits the directional characteristics to reduce the training cost, while the implied STDT phase utilizes additional DoFs in the training phase for data transmission. Individual gains from NOBT and STDT respectively depend on the topology of the beam pair bipartite graph. For example, if we maintain dominant entries of the measure table in Fig. 1 and build up its corresponding beam pair bipartite graph, parallel training can be implemented across different transmit beams. Although there is no STDT phase, the training cost is tremendously reduced by the NOBT phase. In Section VI, we investigate the individual contributions from NOBT and STDT, respectively, through simulations with random topology of the beam pair bipartite graph.

IV. Problem Formulation

A. Training with STDT phase

*1) **Instantaneous Channel Estimation**:* To enable STDT, UEs need to feed back the instantaneous estimated effective channel to the BS at time slot t, $\forall t$. Then, the BS can extract available beams to form $\mathcal{B}_{t+1}$ for data transmission at time slot $t+1$. The received training signal at UE_i at time slot t can be expressed as

$$\begin{aligned}
\tilde{x}_{\text{tr},i,t} &= \sqrt{\rho_{\text{p},t}} \mathbf{w}_{\text{a}i}^{\dagger} \mathbf{H}_i \mathbf{F}_{\text{a}} \mathbf{p}_{\text{tr},t} + \mathbf{w}_{\text{a}i}^{\dagger} \mathbf{H}_i \mathbf{F}_{\text{a}} \mathbf{x}_{\text{d},t} + \mathbf{w}_{\text{a}i}^{\dagger} \mathbf{n}_{i,t} \\
&= \sqrt{\rho_{\text{p},t}} \bar{\mathbf{h}}_i^{\dagger} \mathbf{p}_{\text{tr},t} + \bar{\mathbf{h}}_i^{\dagger} \mathbf{x}_{\text{d},t} + \bar{n}_{i,t} \\
&= \underbrace{\sqrt{\rho_{\text{p},t}} \bar{h}_{i,G(i,t)}}_{\text{Desired training signal}} + \underbrace{\sum_{j \in \mathcal{T}_{\text{tr},t} \setminus \{G(i,t)\}} \sqrt{\rho_{\text{p},t}} \bar{h}_{i,j}}_{\text{Training contamination}} + \underbrace{\bar{\mathbf{h}}_{i,\text{dd},t}^{\dagger} \mathbf{F}_{\text{d},t} \mathbf{x}_t}_{\text{Payload interference}} + \underbrace{\bar{n}_{i,t}}_{\text{Noise}},
\end{aligned} \tag{7}$$

where $\rho_{\text{p},t}$ denotes the power used for training each beam in every time slot, $\forall t$. $\mathbf{p}_{\text{tr},t}$ is an $l_{\text{BS}} \times 1$ indicator vector to denote whether a transmit beam is scheduled for training at time slot t, e.g., if $\mathbf{p}_{\text{tr},t}(i) = 1$, the i-th transmit beam is trained at time slot t. $\mathbf{n}_{i,t} \in \mathbb{C}^{N \times \tau}$ indicates the i.i.d. complex Gaussian noise vector at UE_i, whose entries follow $\mathcal{CN}(0, \delta^2)$. The second term in (7) denotes the interference by data transmission,

[2]For a TDD system, a similar argument can be developed to utilize the directional characteristics of mm-wave channels for uplink training. Then, we need to investigate the set of UEs that can be trained together, whose set of receive beams at BS shall be orthogonal to each other.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2767580, IEEE Transactions on Wireless Communications

where $\mathbf{x}_{\mathrm{d},t} \in \mathbb{C}^{l_{\mathrm{BS}}\times 1}$ is the data symbol vector at time slot t. Since partial beams may be scheduled for data transmission instantly, $\mathbf{x}_{\mathrm{d},t}$ only has a few (or none) non-zero entries, which corresponds to beams in $\mathcal{B}_t$, $\forall t$. For UE_i belonging to $\mathcal{K}_{\mathrm{tr},t}$, we define the effective channel from the j-th transmit beam as $\bar{h}_{i,j}$, and $G(i,t)$ denotes the index of training beam associated with UE_i at time slot t.

The pilot suffers interference from two components: one from the pilot signal of other beams (*Training contamination*) and the other from beams scheduled for data transmission (*Payload interference*). In (7), $\bar{\mathbf{h}}_{i,\mathrm{dd},t} \in \mathbb{C}^{|\mathcal{B}_t|\times 1}$ denotes the effective channel from beams transmitting data symbols at time slot t. $\mathbf{F}_{\mathrm{d},t} \in \mathbb{C}^{|\mathcal{B}_t|\times|\mathcal{K}_{\mathrm{dd},t}|}$ denotes the digital precoder at time slot t for payload transmission to $|\mathcal{K}_{\mathrm{dd},t}|$ UEs. $\mathbf{x}_t \in \mathbb{C}^{|\mathcal{K}_{\mathrm{dd},t}|\times 1}$ denotes the data symbol vector, following $\mathcal{CN}(\mathbf{0}, \mathbf{I}_{|\mathcal{K}_{\mathrm{dd},t}|})$. From (7), we can estimate the effective channel $\bar{h}_{i,G(i,t)}$ by using existing channel estimation methods.

*2) **Partial data transmission**:* During the training window, we may launch the partial data transmission as Fig. 4d exhibits. Suppose UE_k is able to receive a data symbol at time slot t, where $t \le \tau$. The received signal model at UE_k is

$$\begin{aligned}\hat{x}_{\mathrm{d},k,t} =& \bar{\mathbf{h}}_{k,\mathrm{dd},t}^{\dagger}\mathbf{F}_{\mathrm{d},t}\mathbf{x}_t + \sqrt{\rho_{\mathrm{p},t}}\sum\nolimits_{j\in\mathcal{T}_{\mathrm{tr},t}} \bar{h}_{k,j} + \bar{n}_{k,t} \\ =& \underbrace{\bar{\mathbf{h}}_{k,\mathrm{dd},t}^{\dagger}\mathbf{f}_{\mathrm{d},t,k}x_{t,k}}_{\text{Desired signal}} + \underbrace{\bar{\mathbf{h}}_{k,\mathrm{dd},t}^{\dagger}\sum_{i\in\mathcal{K}_{\mathrm{dd},t}\setminus k}\mathbf{f}_{\mathrm{d},t,i}x_{t,i}}_{\text{Inter-user interference}} + \\ & \underbrace{\sqrt{\rho_{\mathrm{p},t}}\sum_{j\in\mathcal{T}_{\mathrm{tr},t}}\bar{h}_{k,j}}_{\text{Training interference}} + \underbrace{\bar{n}_{k,t}}_{\text{Noise}},\end{aligned} \tag{8}$$

where $\mathbf{F}_{\mathrm{d},t}$ consists of individual digital precoders serving UEs belonging to $\mathcal{K}_{\mathrm{dd},t}$, i.e., $\mathbf{F}_{\mathrm{d},t} = [\mathbf{f}_{\mathrm{d},t,k}]_{k\in\mathcal{K}_{\mathrm{dd},t}}$, and $x_{t,i}$ is the data symbol transmitted to UE_i at the time slot t. Similarly to (7), there exist two kinds of interference: the conventional inter-user interference and the interference from the simultaneously transmitted training signals.

B. Dedicated Data Transmission

After the period of downlink training, the BS can utilize all analog beams for data transmission and the received signal model at UE_k can be expressed as

$$\begin{aligned}\hat{x}_{d,k} =& \mathbf{w}_{\mathrm{a}k}^{\dagger}\mathbf{H}_k\mathbf{F}_{\mathrm{a}}\mathbf{f}_{\mathrm{d},k}x_k + \mathbf{w}_{\mathrm{a}k}^{\dagger}\mathbf{H}_k\mathbf{F}_{\mathrm{a}}\textstyle\sum_{i\neq k}\mathbf{f}_{\mathrm{d},i}x_i + \mathbf{w}_{\mathrm{a}k}^{\dagger}\mathbf{n}_k \\ =& \bar{\mathbf{h}}_k\mathbf{f}_{\mathrm{d},k}x_k + \bar{\mathbf{h}}_k\sum_{i\neq k}\mathbf{f}_{\mathrm{d},i}x_i + \bar{n}_k,\end{aligned} \tag{9}$$

where we ignore the subscript t since the receive signal model remains the same after the training window.

C. Beamformer Optimization

Given the analog beamforming, the achievable rate of $\mathrm{UE}_{\pi(i)}$ at time slot t by using the dirty paper coding (DPC) scheme in digital baseband is given by [33]

$$C_{\pi(i),t} = \log\left|\frac{\delta^2\mathbf{w}_{\mathrm{a}\pi(i)}^{\dagger}\mathbf{w}_{\mathrm{a}\pi(i)} + \bar{\mathbf{h}}_{\pi(i),\mathrm{dd},t}^{\dagger}\sum_{j\geq i}\mathbf{\Gamma}_{\pi(j),t}\bar{\mathbf{h}}_{\pi(i),\mathrm{dd},t}^{\dagger}}{\delta^2\mathbf{w}_{\mathrm{a}\pi(i)}^{\dagger}\mathbf{w}_{\mathrm{a}\pi(i)} + \bar{\mathbf{h}}_{\pi(i),\mathrm{dd},t}^{\dagger}\sum_{j>i}\mathbf{\Gamma}_{\pi(j),t}\bar{\mathbf{h}}_{\pi(i),\mathrm{dd},t}^{\dagger}}\right|, \tag{10}$$

where $\pi(i) \in \mathcal{K}_{\mathrm{dd},t}$ and $[\pi(i)]_{i=1}^{|\mathcal{K}_{\mathrm{dd},t}|}$ is the ordered index set of UEs in DPC, and $\mathbf{\Gamma}_{\pi(i),t}$ is the input covariance of $\mathrm{UE}_{\pi(i)}$ at the time slot t. Therefore, the net average MU-MIMO downlink capacity within the coherence block is

$$C_{\mathrm{avg,DL}} = \frac{\sum_{t=1}^{T_{\mathrm{cor}}}\sum_{\pi(i)\in\mathcal{K}_{\mathrm{dd},t}} C_{\pi(i),t}}{T_{\mathrm{cor}}}, \tag{11}$$

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 $(1-\frac{\tau}{T_{\mathrm{cor}}})\sum_{i=1}^{K} C_{\pi(i)}$, where $C_{\pi(i)}$, independent of t, remains the same within the data transmission phase of a coherence block.

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:

$$\max_{[\mathbf{B}_k,\mathbf{w}_{ak}]_{k=1}^{K},[\rho_{\mathrm{p},t}]_{t=1}^{\tau},\mathbf{P}_{\mathrm{tr}}} \mathbb{E}\Big[\max_{[\mathbf{\Gamma}_{\pi(i),t},\pi(i)\in\mathcal{K}_{\mathrm{dd},t}]_{t=1}^{T_{\mathrm{cor}}}} C_{\mathrm{avg,DL}}\Big] \tag{12a}$$

$$s.t.\ \mathbf{B}_k \subset \mathbf{\Omega}_M, \forall k, l_{\mathrm{use}} = \mathrm{rank}([\mathbf{B}_k]_{k=1}^{K}) \le l_{\mathrm{BS}}, \tag{12b}$$

$$\mathbf{P}_{\mathrm{tr}} \in \mathbb{N}^{l_{\mathrm{use}}\times\tau}, \mathbf{P}_{\mathrm{tr}}(i,j) = 1 \text{ or } 0, \forall i,j, \textstyle\sum_{j=1}^{\tau}\mathbf{P}_{\mathrm{tr}}(i,j) = 1, \forall i, \tag{12c}$$

$$\rho_{\mathrm{p},t}|\mathcal{T}_{\mathrm{tr},t}| + \textstyle\sum_{\pi(i)\in\mathcal{K}_{\mathrm{dd},t}}\mathrm{tr}(\mathbf{\Gamma}_{\pi(i),t}) \le \rho_{\mathrm{d}}, \forall t, \tag{12d}$$

where the expectation of $C_{\mathrm{avg,DL}}$ is taken to average out the small scale fading, i.e. $[\bar{\mathbf{G}}_i]$ in (4), 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 (12), an assumption that is widely used in the literature [16, 21, 31].

(12b) 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, $[\mathbf{w}_{ai}]$, are functions of UE-side channel covariance matrices. $\mathbf{P}_{\mathrm{tr}}$ denotes the pilot assignment matrix, where each row has a single non-zero entry to indicate the assigned pilot for the beam. In (12d), the total transmit power is constrained by ρ_{d}, and $\mathrm{tr}(\mathbf{\Gamma}_{\pi(i),t}) = \mathrm{tr}(\mathbf{F}_{\mathrm{a},t}\mathbf{\Gamma}_{\pi(i),t}\mathbf{F}_{\mathrm{a},t}^{\dagger})$ is the power for data transmission to UE_i 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.

The problem (12) is very challenging to solve, incorporating three tiers of optimization with different time scales, and also coupled together. In the first tier, we need to design the channel-statistics-based analog beamformers, where the codebook-based $\mathbf{F}_{\mathrm{a}}$ is coupled with $[\mathbf{w}_{ai}]$. Later, at the second tier, the pilot matrix $\mathbf{P}_{\mathrm{tr}}$ needs to be optimized based on the effective beam pair bipartite graph as Fig. 4 shows, which not only needs to minimize the training overhead but also optimize the training order to achieve additional spatial multiplexing gains in the STDT phase. For the first two tiers, our design is based on the long-term CSI, while in the third tier, the designs of input covariances $[\mathbf{\Gamma}_{\pi(i),t}]$ and permutation of index set $[\pi(i)]$ are based on the instantaneous CSI, which will eventually determine the performance of the first two-tier optimization.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2767580, IEEE Transactions on Wireless Communications

Resorting to the uplink-downlink duality theory [34], we can develop an equivalent uplink problem of (12):

$$\max_{[\mathbf{B}_k,\mathbf{w}_{ak}]_{k=1}^K,[\rho_{\mathrm{p},t}]_{t=1}^{\tau},\mathbf{P}_{\mathrm{tr}}} \mathbb{E}[\max_{[\Gamma'_{i,t},i\in\mathcal{K}_{\mathrm{dd},t}]_{t=1}^{T_{\mathrm{cor}}}} C_{\mathrm{avg,UL}}], \tag{13a}$$

$$s.t.\ (12b), (12c),$$

$$\rho_{\mathrm{p},t}|\mathcal{T}_{\mathrm{tr},t}| + \textstyle\sum_{i\in\mathcal{K}_{\mathrm{dd},t}} \Gamma'_{i,t}\mathbf{w}_{\mathrm{a}i}^{\dagger}\mathbf{w}_{\mathrm{a}i}\delta^2 \le \rho_{\mathrm{d}}, \forall t, \tag{13b}$$

where $C_{\mathrm{avg,UL}} = \frac{1}{T_{\mathrm{cor}}}\sum_{t=1}^{T_{\mathrm{cor}}} C_{t,\mathrm{UL}}$, and $C_{t,\mathrm{UL}} = \log|\sum_{i=1}^K \bar{\mathbf{h}}_{i,\mathrm{dd},t}\Gamma'_{i,t}\bar{\mathbf{h}}_{i,\mathrm{dd},t}^{\dagger} + \mathbf{I}_{|\mathcal{B}_t|}|$. $C_{t,\mathrm{UL}}$ denotes the instantaneous uplink capacity at time slot t. $\Gamma'_{i,t}$ indicates the uplink transmit power coefficient of UE_i at the time slot t, $\forall i,t$. Constraints $(12b)$ and $(12c)$ remain the same for the uplink dual problem, while the power constraint becomes (13b) instead of (12d).

Detailed developments of the uplink-dual problem with HDA structure at both ends are revealed in [35], which is briefly summarized as follows. Based on [34], the downlink channel has the same instantaneous sum rate as its dual uplink, which can be expressed as

$$\max_{[\Gamma'_{i,t},i\in\mathcal{K}_{\mathrm{dd},t}]} C_{t,\mathrm{UL}} = \log|(\textstyle\sum_{i=1}^K \bar{\mathbf{h}}_{i,\mathrm{dd},t}\Gamma'_{i,t}\bar{\mathbf{h}}_{i,\mathrm{dd},t}^{\dagger} + \mathbf{Q}_{1,t})\mathbf{Q}_{1,t}^{-1}| \tag{14}$$

$$s.t.\ \textstyle\sum_{i\in\mathcal{K}_{\mathrm{dd},t}} \Gamma'_{i,t}Q_{2i,t} \le \rho_{\mathrm{d}} - \rho_{\mathrm{p},t}|\mathcal{T}_{\mathrm{tr},t}|,$$

where $\mathbf{Q}_{1,t} = \mathbf{F}_{\mathrm{a},t}^{\dagger}\mathbf{F}_{\mathrm{a},t} = \mathbf{I}_{|\mathcal{B}_t|}$, since $\mathbf{F}_{\mathrm{a},t}$ consists of normalized DFT columns, and $Q_{2i,t} = \delta^2\mathbf{w}_{\mathrm{a}i}^{\dagger}\mathbf{w}_{\mathrm{a}i}, i \in \mathcal{K}_{\mathrm{dd},t}, \forall t$. Based on (14), we can obtain the optimization for the dual uplink channel as (13). Our goal is still focusing on the downlink problem, but we resort to its equivalent dual problem for mathematical convenience.

*1) **Decoupled optimization with reduced complexity**:* Although the uplink-dual problem (13) exhibits a more tractable objective function than that of (12a), it still incorporates joint multi-tier optimization with different time scales.

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

$$\rho_{\mathrm{p},t} = \Gamma'_{i,t} = \frac{\rho_{\mathrm{d}}}{|\mathcal{T}_{\mathrm{tr},t}|+\delta^2|\mathcal{K}_{\mathrm{dd},t}|}, \forall i \in \mathcal{K}_{\mathrm{dd},t}. \tag{15}$$

At time slots dedicated for training, (15) 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, (15) becomes 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 (15), $C_{\mathrm{avg.,UL}}$ 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:

$$\max_{[\mathbf{B}_k,\mathbf{w}_{ak}]_{k=1}^K,\mathbf{P}_{\mathrm{tr}}} \mathbb{E}[C_{\mathrm{avg,UL}}] \tag{16a}$$

$$s.t.\ (12b), (12c),$$

$$\|\mathbf{w}_{ak}\| = 1, \forall k, \tag{16b}$$

*2) **Average throughput approximation**:* To avoid the computational burden in evaluation of the expectation at (16a), we consider the following upper bound of average uplink throughput

$$\mathbb{E}[C_{t,\mathrm{UL}}] \underset{(a)}{\le} C_{\mathrm{UL,upper}} = \log\mathbb{E}[|\rho_{\mathrm{p},t}\textstyle\sum_{i=1}^K \bar{\mathbf{h}}_{i,\mathrm{dd},t}\bar{\mathbf{h}}_{i,\mathrm{dd},t}^{\dagger} + \mathbf{I}_{l_{\mathrm{BS}}}|], \tag{17}$$

where (a) follows from Jensen's inequality: $\mathbb{E}[\log|\mathbf{I}+\mathbf{X}|] \le \log\mathbb{E}[|\mathbf{I}+\mathbf{X}|]$. 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: By assuming a single-path channel model, i.e. $P_i = 1$ in (2), $\forall i$, we have the following equivalence: $\log\mathbb{E}[|\rho_{\mathrm{p}}\mathbf{F}_{\mathrm{a}}^{\dagger}\sum_{i=1}^K \mathbf{H}_i^{\dagger}\mathbf{w}_{\mathrm{a}i}\mathbf{w}_{\mathrm{a}i}^{\dagger}\mathbf{H}_i\mathbf{F}_{\mathrm{a}} + \mathbf{I}_{l_{\mathrm{BS}}}|] = \log|\rho_{\mathrm{p}}\mathbf{F}_{\mathrm{a}}^{\dagger}\sum_{i=1}^K \tilde{\mathbf{H}}_i^{\dagger}\mathbf{w}_{\mathrm{a}i}\mathbf{w}_{\mathrm{a}i}^{\dagger}\tilde{\mathbf{H}}_i\mathbf{F}_{\mathrm{a}} + \mathbf{I}_{l_{\mathrm{BS}}}|$, where $\tilde{\mathbf{H}}_i \triangleq \frac{1}{L_i}\mathbf{A}_{\mathrm{UE},i}\boldsymbol{\Sigma}_i\mathbf{A}_{\mathrm{BS},i}^{\dagger}$.

Proposition 1 can be easily obtained from the result in [35]. 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:

$$\max_{[\mathbf{B}_k,\mathbf{w}_{ak}]_{k=1}^K,\mathbf{P}_{\mathrm{tr}}} \tilde{C}_{\mathrm{avg,UL}} = \tfrac{1}{T_{\mathrm{cor}}}\textstyle\sum_{t=1}^{T_{\mathrm{cor}}} \tilde{C}_{t,\mathrm{UL}} \tag{18}$$

$$s.t.\ (12b), (12c), (16b),$$

where $\tilde{C}_{t,\mathrm{UL}} = \log|\rho_{\mathrm{p},t}\mathbf{F}_{\mathrm{a},t}^{\dagger}\sum_{i\in\mathcal{K}_{\mathrm{dd},t}} \tilde{\mathbf{H}}_i^{\dagger}\mathbf{w}_{\mathrm{a}i}\mathbf{w}_{\mathrm{a}i}^{\dagger}\tilde{\mathbf{H}}_i\mathbf{F}_{\mathrm{a},t} + \mathbf{I}_{|\mathcal{B}_t|}|$. Without the assumption of $P_i = 1, \forall i$, Proposition 1 does not hold in general and problem (18) becomes an approximation of problem (16). Our simulation results in Section VI-B demonstrate that the approximation performs well, even with general settings of $[P_i]$.

V. Algorithm Development

Problem (18) is still generally non-convex, involving integer programming for designing $\mathbf{P}_{\mathrm{tr}}$ and $[\mathbf{B}_k]$. Meanwhile, given a topology of beam pair bipartite graph as shown in Fig. 4, 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 (18).

A. Training Order Optimization

Given a beam pair bipartite graph, the minimum training cost can be evaluated by the algorithm proposed in [36], which provides a suboptimal solution to minimize an upper bound of the training cost: whereas [36] treats left side nodes as BSs, we view them as transmit beams. The algorithm is summarized below:

- Build up the conflict graph of transmit beams by treating them as vertices and connect any pair of them with which a common UE is associated as Fig. 5 illustrates.
- Sort the degree of vertex in descending order, which will be $[\mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4, \mathbf{b}_1]$ in Fig. 5.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2767580, IEEE Transactions on Wireless Communications

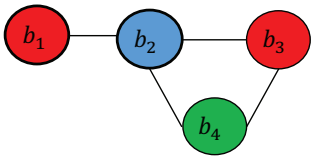


Fig. 5: Conflict graph of transmit beams for beam pair bipartite graph exhibited in Fig. 4, and different colors represent different pilot dimensions allocated to transmit beams.

- Allocate pilot dimensions to vertices (beams) in a sequential manner. For every vertex awaiting pilot assignment, if it is conflicted with all previous vertices, assign an orthogonal pilot dimension to it. Otherwise, assign a pilot dimension occupied by most transmit beams that have no conflict with the vertex.

For the toy example in Fig. 5, 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 dimension for training is not explored in above algorithm.

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 $\max \sum_{i=1}^{K}(T_{\text{cor}} - T_{\text{tr},i})$, where $T_{\text{tr},k}$ 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. $\min \sum_{i=1}^{K} T_{\text{tr},i}$. The aim of solving this problem is to complete as many as possible UEs' individual training earlier than τ by optimizing over all possible sequential orders of $[t_i]_{i=1}^{\tau}$. Minimizing the number of time slots used for training is not the same thing as making sure that we can send as many data slots as possible - there could be non-training slots for a UE before its training is finished (i.e., empty slots). However, the formulated problem is physically intuitive and tractable.

To approximate the optimal solution to this typical integer programming problem, we summarize our 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 $\sum_{j \in \mathcal{T}_{\text{tr},t_m}} \sum_{k=1}^{K} \frac{I_k(\mathbf{b}_j)}{L_{\text{tran},k}}, \forall m \in \mathcal{P}_i$, where $I_k(\mathbf{b}_j)$ is an indicator to denote whether $\mathbf{b}_j$ is associated with UE_k, $L_{\text{tran},k}$ is the number of transmit beams connected to UE_k, and $\mathcal{T}_{\text{tr},t_m}$ 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. Combining relative significance of trained beams on each pilot dimension, we obtain the priority orders, or we say the relative significance of pilot dimensions, and then we can schedule them sequentially. Based on the result of training allocation and order scheduling, we can build up the pilot assignment matrix $\mathbf{P}_{\text{tr}}$.

B. Greedy User-Centric Beam Clustering

We consider the case that the analog precoder and analog combiner are chosen from the DFT codebook and the eigenmode of UE-side channel covariance, respectively. Therefore, the beamformer optimization of (18) becomes to select the effective beam pairs from the bipartite graph implied by $\mathbf{S}$. Thanks to the training order optimization in Section V-A, we can evaluate the performance of any given topology of reduced $\bar{\mathbf{S}}$, which lays the foundation of our proposed *greedy user-centric beam clustering* (GUCBC) algorithm. The detailed implementation procedure can be summarized as follows:

1) Initially, let $\mathbf{w}_{\text{a}k} = \mathbf{0}, \forall k$ and $\mathbf{F}_{\text{a}} = \emptyset$. Let $\mathbf{W}_{\text{a}}$ be the ensemble of analog combiners as $\mathbf{W}_{\text{a}} \triangleq [\mathbf{w}_{\text{a}i}, ..., \mathbf{w}_{\text{a}K}]$.
2) Extract the $M \times \sum_{i=1}^{K} r_i$ measure matrix $\mathbf{S}$ following (6), 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 $\mathbf{W}'_{\text{a}} = \mathbf{W}_{\text{a}}$ and $\mathbf{F}'_{\text{a}} = \mathbf{F}_{\text{a}}$. For a candidate beam pair $\mathbf{e} = (\mathbf{b}, \mathbf{r})$ in $\mathcal{E}$, we let $\mathbf{F}'_{\text{a}} = [\mathbf{F}'_{\text{a}}, \mathbf{b}]$ and assign $\mathbf{r}$ to its corresponding UE, then run the *evaULthroughput* given in Algorithm 1 to return the *net average uplink throughput approximation* (NAUTA).
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 $\mathbf{W}_{\text{a}}$ by assigning $\mathbf{r}^\star$ to its corresponding $\text{UE}_{k^\star}$, update $\mathbf{F}_{\text{a}}$ by $\mathbf{F}_{\text{a}} = [\mathbf{F}_{\text{a}}, \mathbf{b}^\star]$, and remove the beam pairs starting with $\mathbf{b}^\star$ and beam pairs ended with all other receive eigenmodes of $\text{UE}_{k^\star}$ from $\mathcal{E}$.
6) Repeat step 3) to step 5) until $\text{rank}(\mathbf{F}_{\text{a}}) = l_{\text{BS}}$ or the NAUTA does not increase by adding additional beams.

The essential idea of the algorithm is to greedily add effective beam pairs from the bipartite graph. For every candidate beam pair, we need to utilize an inner function, so called *evaULthroughput*, to evaluate the NAUTA of the beamformed effective channel with this additional candidate, and then select the best beam pair to update $\mathbf{F}_{\text{a}}$ and $\mathbf{W}_{\text{a}}$. For example, let us investigate the toy example exhibited by Fig. 3, where initially we have total 5 edges in $\mathcal{E}$. The first step will select a beam pair that provides maximal NAUTA, which is $(\mathbf{b}_3, \mathbf{r}_{31})$ with the largest weight. Then, edges $(\mathbf{b}_3, \mathbf{r}_{12})$ and $(\mathbf{b}_3, \mathbf{r}_{22})$ have to be removed, since their transmit beam

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2767580, IEEE Transactions on Wireless Communications

$\mathbf{b}_3$ has already been selected. The remaining beam pair set will be $\{(\mathbf{b}_1, \mathbf{r}_{11}), (\mathbf{b}_2, \mathbf{r}_{21})\}$, and we sequentially assign them to $\mathbf{F}_{\mathrm{a}}$ and $\mathbf{W}_{\mathrm{a}}$ if the NAUTA can get enhancement. Detailed specifics of *evaULthroughput* can be found in Algorithm 1. Under the constraint of both channel rank and number of BS RF chains, the proposed GUCBC algorithm not only designs sub-optimal analog beamformers at both link ends, but also implicitly incorporates the functionality of UE scheduling in the sense that the users with $\mathbf{w}_{\mathrm{a}k} = \mathbf{0}$ are not scheduled.

Algorithm 1 *evaULthroughput*

1: Extract the effective beam pair bipartite graph projected by $\mathbf{F}'_{\mathrm{a}}$ and $\mathbf{W}'_{\mathrm{a}}$. Follow procedures in SectionV-A to optimize the pilot assignments $\mathbf{P}_{\mathrm{tr}}$.
2: **for** $t \leftarrow 1$ to T_{cor} **do**
3: Extract $\mathcal{K}_{\mathrm{cc},t}$ and its complement $\bar{\mathcal{K}}_{\mathrm{cc},t}$.
4: Build up $\mathbf{F}'_{\mathrm{a},t}$ by trained beams and make sure that no beam in $\mathbf{F}'_{\mathrm{a},t}$ is connected to UEs belonging to $\bar{\mathcal{K}}_{\mathrm{cc},t}$, then calculate $\tilde{C}_{t,\mathrm{UL}}$ in (18).
5: **end for**
6: Substitute $[\tilde{C}_{t,\mathrm{UL}}]$ into (18) and we can obtain the NAUTA $\tilde{C}_{\mathrm{avg,UL}}$.

To evaluate the complexity of the proposed algorithm, we compare it with the exhaustive beam search. Given $l_{\mathrm{sel.}}$ selected transmit eigenmodes and stream assignments $[y_i]_{i=1}^K$, where y_i represents whether or not assigning a stream to UE_i, and $\sum_{i=1}^K y_i \le l_{\mathrm{sel.}}$, there are total $\binom{r_i}{y_i}$ possible sets of receive eigenmodes which can be used to form the analog combiner at UE_i, $\forall i$. Therefore, for all K UEs, there will be total $\prod_{i=1}^K \binom{r_i}{y_i}$ possibilities for given stream assignments $[y_i]_{i=1}^K$ and transmit beams. The total number of combinations that the exhaustive search method needs to investigate is:

$$N_{\mathrm{comb.}} = \sum_{l_{\mathrm{sel.}}=1}^{l_{\mathrm{BS}}} \binom{M}{l_{\mathrm{sel.}}} \sum_{\sum_{i=1}^K y_i \le l_{\mathrm{sel.}}} \prod_{i=1}^K \binom{r_i}{y_i}, \tag{19}$$

where the second summation is over all possible realizations of stream assignments, and the first summation is over all possible numbers of transmit beams. The computational burden grows extremely fast with the increasing of variables M, r_i, and l_{BS}, which is prohibitive for implementation. For the proposed method, the iteration time is up to l_{BS}. At step 5) of the greedy user-centric beam clustering (GUCBC) algorithm, we will remove beam pairs that are not effective beam pair candidates for follow-up iterations. Considering the worst-case scenario, where we only remove the selected beam pair from the beam pair set $\mathcal{E}$ for every iteration, its complexity is upper bounded by $\sum_{t=1}^{l_{\mathrm{BS}}} (|\mathcal{E}| - t + 1) = l_{\mathrm{BS}}|\mathcal{E}| - \frac{l_{\mathrm{BS}}(l_{\mathrm{BS}}-1)}{2} \le l_{\mathrm{BS}}|\mathcal{E}|$, where t indicates the index of iteration. Therefore, the complexity order of the proposed scheme is roughly $O(l_{\mathrm{BS}}|\mathcal{E}|)$. If we conservatively remove weak beam pairs whose effective channel gains are below the noise power, the cardinality of $\mathcal{E}$ is approximately the ensemble of all MPCs, i.e., $\sum_{i=1}^K P_i$. Therefore, the complexity of the proposed scheme is upper bounded by $O(l_{\mathrm{BS}} \sum_{i=1}^K P_i)$, which is significantly less than $N_{\mathrm{comb.}}$ shown in (19), especially for mm-wave frequencies where dominant MPCs are much fewer than those at low frequencies. Considering a numerical example, we let $M = 64$, $l_{\mathrm{BS}} = 8$, and $K = 4$. Since the number of dominant MPCs is usually less than 10 in the mm-wave band [37], we temporally set $P_i = 10, \forall i$. According to (19), the size of search space for the exhaustive search is extremely large, e.g., $\binom{64}{8}$ is around 4.4×10^9. However, the complexity of the proposed method is upper bounded by $l_{\mathrm{BS}} \sum_{i=1}^K P_i = 320$.

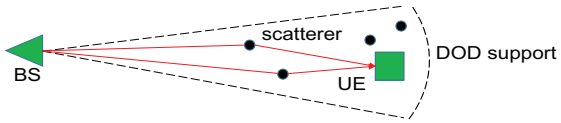


Fig. 6: Illustration of GSCM with UEs and scatterers in a range of DOD support.

VI. Simulation Results

In this section, we evaluate the performance of the proposed UCVS scheme via simulations. All simulation results exhibit the comparison with JSDM/BDMA [16, 31] with respect to net average throughput. 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.

A. Geometric Stochastic Channel Model (GSCM)

Following the dominant characteristics of mm-wave propagation, we mainly focus on the MPCs interacting with a single scatterer. Fig. 6 illustrates an example of how to generate the synthetic channel profiles. We place the scatterers and UEs in an angular range (as seen from the BS) that we call support interval of DOD. Therefore, different options of DOD support range can represent different scenarios. For example, for a crowded cafeteria, we may use a narrow DOD support, while for UEs separated far away from the BS perspective in the angular domain, we can use a wide DOD support range.

To activate scatterers for the channels between the BS and the UEs, we utilize the following probabilistic model:

$$P_{\mathrm{active}} = P_{\mathrm{UE,LOS}} \cdot P_{\mathrm{BS,LOS}}, \tag{20}$$

where P_{active} is the probability that a scatterer is active for a UE, which is the product of marginal probabilities that both ends can "see" this scatterer. The marginal probability that a terminal has LOS propagation to the scatterer follows

$$P_{\mathrm{UE/BS,LOS}} = \min(d_1/d, 1)(1 - \exp(-d/d_2)) + \exp(-d/d_2), \tag{21}$$

where d_1 and d_2 are modeling parameters, and d is the distance from the BS/UE to the scatterer. When $d < d_1$, we have $P_{\mathrm{UE/BS,LOS}} = 1$ indicating that the scatterer is deterministically visible by the UE/BS. For $d > d_1$, the probability is exponentially decreasing with increasing of d, where the decay rate is determined by d_2. Settings of both d_1 and d_2 will

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2767580, IEEE Transactions on Wireless Communications

be environment-dependent [37],[3] e.g., urban, rural, and the terminal heights will also make a difference. Note that this model provides an implementation of the "common scatterer" concept used, e.g., in COST 2100 [38]. Unless otherwise specified, the parameter settings for channel model and system configuration are exhibited in Table I.

TABLE I: Simulation parameters

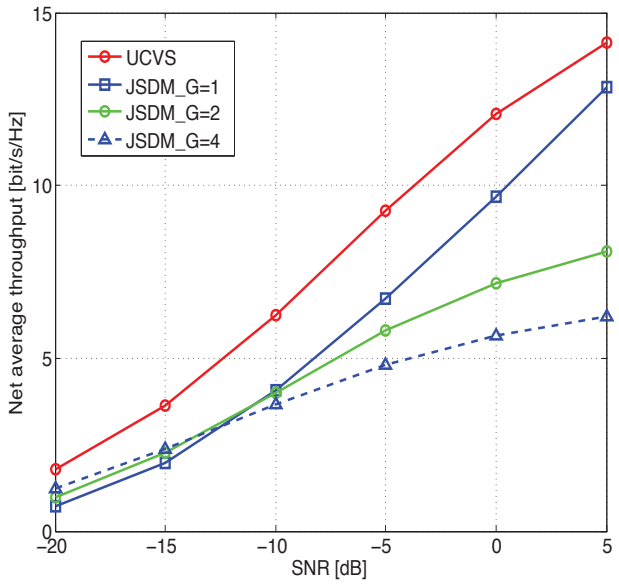
DOD support range	$\theta_{\text{all}} = 20°$ or $40°$
LOS from BS to scatterer	$d_{1,\text{BS2S}}$=24 m, $d_{2,\text{BS2S}}$=45 m
LOS from UE to scatterer	$d_{1,\text{UE2S}}$=2 m, $d_{2,\text{UE2S}}$=10 m
Scatterer density	$\rho_{\text{s}} = 0.01 \sim 0.09$
Energy threshold	$\gamma = 0.7 \sim 1$
Number of UEs	$K = 4$
No. of BS antennas and RF chains	$M = 64$, $l_{\text{BS}} = 8$
No. of UE antennas and RF chains	$N = 8$, $l_{\text{UE}} = 1$
Antenna spacing (in wavelength)	$D = \frac{1}{2}$

For modeling parameters in (21), since the BS is usually located higher, $d_{1,\text{BS2S}}$ and $d_{2,\text{BS2S}}$ are respectively larger than $d_{1,\text{UE2S}}$ and $d_{2,\text{UE2S}}$ (we use subscripts "BS2S" and "UE2S" to distinguish parameter sets for different terminals). The coherence time in the units of channel use can be evaluated by $\frac{1}{f_{\text{d}} T_{\text{s}}}$, where f_{d} is the Doppler spread, and T_{s} is the symbol duration, which is $66.7\mu s$ in the LTE standard. Substituting the mobility speed ranging from 6.5 km/h to 18 km/h and carrier frequency of 60 GHz, we can obtain coherence times approximately ranging from 40 to 15 channel uses of the LTE standard. For all simulation sets, we maintain the noise power and large scale loss to be unity, i.e. $\delta^2 = 1$, $L_i = 1, \forall i$. Therefore, transmit power ρ_{d} is equivalent to the signal-to-noise ratio (SNR) subsuming the impact of large scale loss.

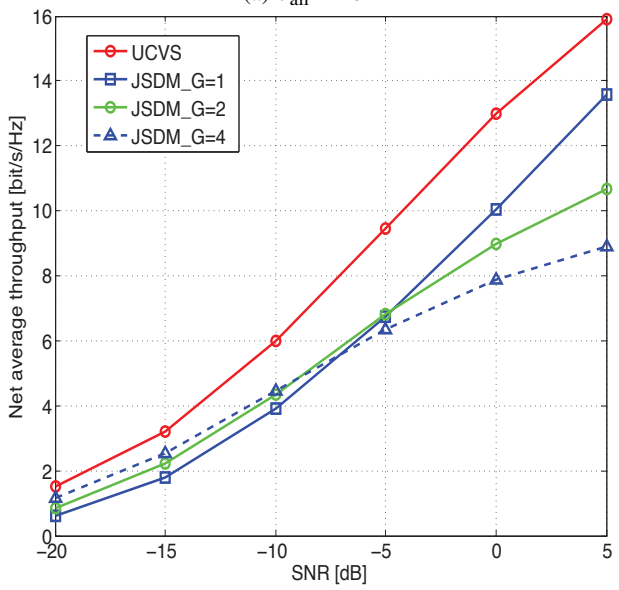
For every drop of UEs, multiple scatterers are independently generated following a Poisson process with parameter ρ_{s} in the sector-shape region as Fig. 6 shows. Meanwhile, random locations of UEs are constrained in the region, whose separation distances to the BS range from 50 to 60 m. With the assumption of uncorrelated scattering, we independently generate $[\sigma_{ip}]$ following a uniform distribution within $[0, 1]$, and then normalize them to satisfy $\sum_{p=1}^{P_i} \sigma_{ip}^2 = 1, \forall i$. 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 (4) of all UEs.

The net average throughputs 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 fading. We investigate the ensemble average over different realizations of beam pair bipartite graph to demonstrate the advantages of the proposed method in various propagation environments. Intuitively, if UE channels are fully spatially orthogonal to each other as Fig. 1 shows or their transmit eigenmodes are fully coupled, both schemes will achieve approximately the same performance, where in the former case each UE forms an independent UE group, while in the latter case, all UEs are grouped together.

[3] [37] proposes the LOS probability model (21) 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.



(a) $\theta_{\text{all}} = 20°$



(b) $\theta_{\text{all}} = 40°$

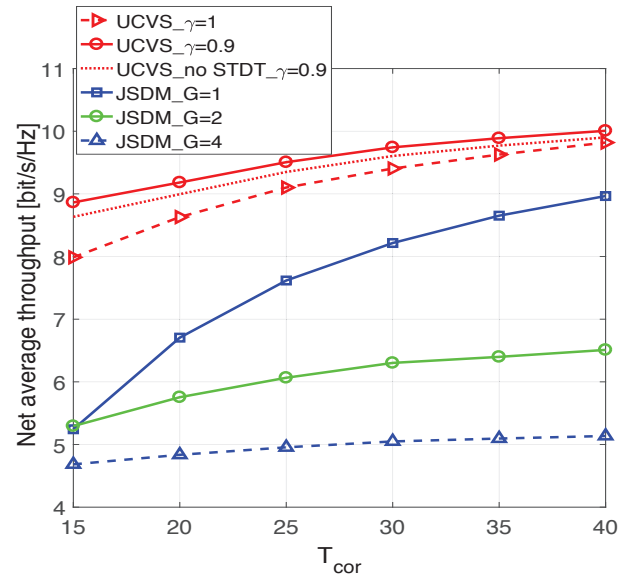
Fig. 7: Net average throughput vs. ρ_{d} for $T_{\text{cor}} = 20$ and $\gamma = 0.9$

In the Section VI-B, we investigate system performance under more realistic mm-wave channel models, which lies between above two extreme examples.

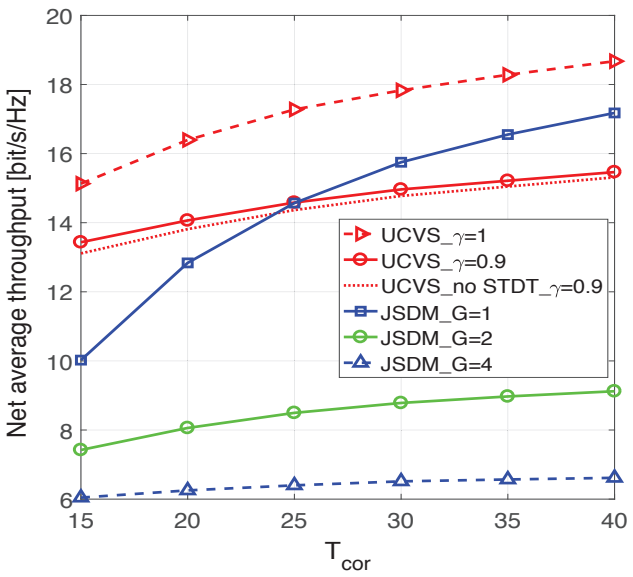
B. Results and Discussions

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. 7 exhibits. Note that since there is no clear conclusion on the optimal UE grouping for JSDM in [16, 30], we make comparisons with the JSDM scheme under different UE groupings, where K-means clustering to group UEs with similar channel covariance is applied [30]. 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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2767580, IEEE Transactions on Wireless Communications

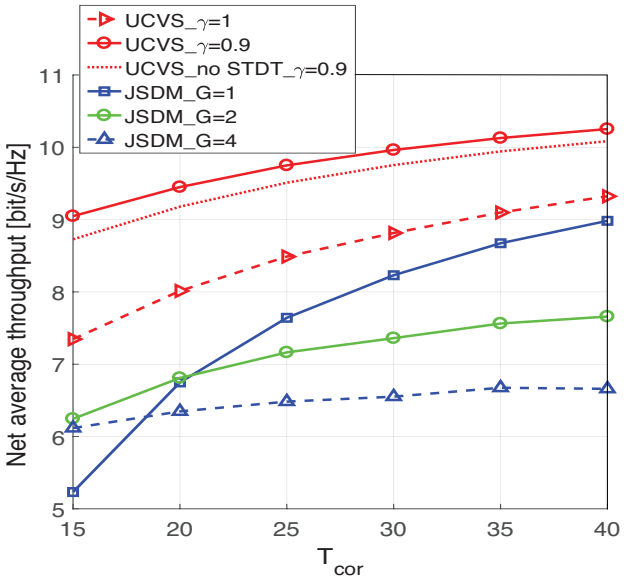


(a) $\rho_{\mathrm{d}} = -5$ dB

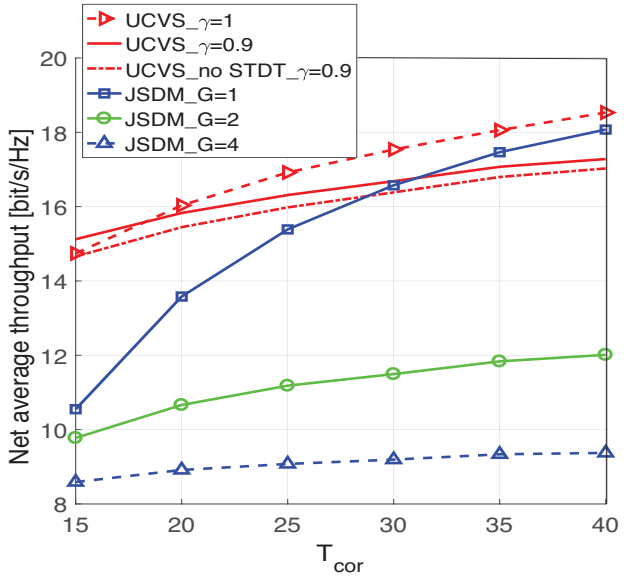


(b) $\rho_{\mathrm{d}} = 5$ dB

Fig. 8: Net average throughput vs. T_{cor} for $\theta_{\mathrm{all}} = 20°$



(a) $\rho_{\mathrm{d}} = -5$ dB



(b) $\rho_{\mathrm{d}} = 5$ dB

Fig. 9: Net average throughput vs. T_{cor} for $\theta_{\mathrm{all}} = 40°$

JSDM cannot fully eliminate the inter-group interference, and forming more user groups will make the system operate in 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 $\theta_{\mathrm{all}} = 40°$, 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. 8 and Fig. 9 show the net average throughput as a function of the coherence time under different DOD support ranges θ_{all} and SNRs ρ_{d}. For the proposed UCVS scheme, we also investigate different settings of threshold parameter γ, i.e. 0.9 or 1. Comparing Fig. 8a and 8b, or Fig. 9a and 9b, we notice that appropriate settings of γ are scenario-dependent. Specifically, for the low SNR regime, compared with the situation where $\gamma = 1$, 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. 8b and Fig. 9b, 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. However, for typical coherence times below

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2767580, IEEE Transactions on Wireless Communications

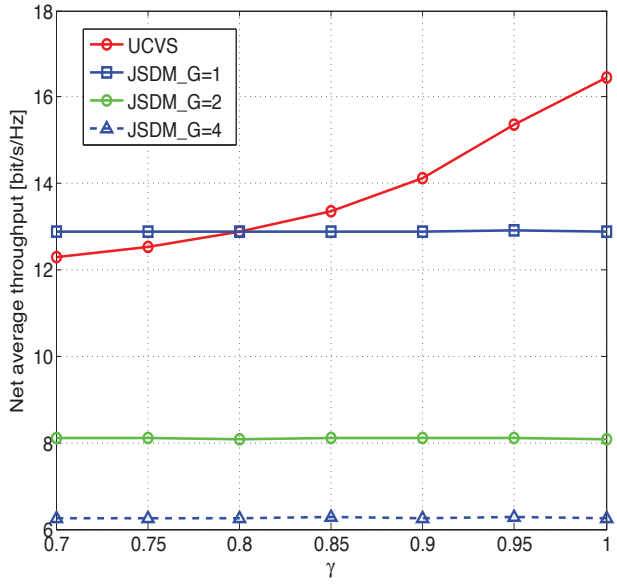
30, the proposed scheme at $\gamma = 0.9$ still outperforms the state-of-art method.

To evaluate the individual contributions from the NOBT and STDT respectively, we also investigate the performance of UCVS without consideration of the STDT phase, whose legend is "UCVS_no STDT" in Fig. 8 and Fig. 9. Although individual gains by STDT are not significant in the simulated scenarios with terminals and scatterers in a narrow DOD support range, they could be dominant in other typical scenarios. For example, with two UEs spatially orthogonal to each other in the angular domain, one has a much larger DOD spread than that of the other. The reduction of training cost by NOBT will be limited by training the UE channel with large DOD spread, while the system can start data transmission once the training for UE with narrower angular spread is completed, making STDT more advantageous.

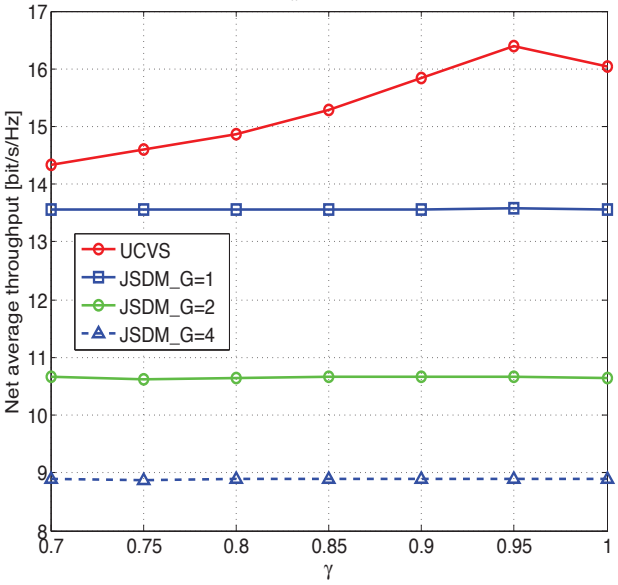
In conclusion, for the interesting range of parameter settings for mm-wave systems, i.e. operating at SNR below 0 dB and coherence time below 50 channel uses, the proposed UCVS exhibits significant performance advantage over JSDM, e.g., more than 38% when $\rho_d = -5$ dB and $T_{cor} = 20$ as Fig. 8a 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. 8b and Fig. 9b.

To further investigate the impact of the threshold γ, we fix $\rho_d = 5$ dB, $T_{cor} = 20$, and compare UCVS with JSDM by net average throughput varying with γ in Fig. 10. The optimal JSDM will still group all UEs together, and its net sum rate does not vary with parameter γ. For the proposed UCVS, we cannot adjust γ to be too small. Otherwise, the UCVS based on the reduced beam pair bipartite graph will cause training contamination and inter-user interference, whose performance may be even worse than that of JSDM, e.g., when $\gamma = 0.7 \sim 0.8$ in Fig. 10a. With $\theta_{all} = 20°$, the angular spectra of UEs tend to be largely overlapped. Therefore, with a relatively high SNR, the optimal γ is one, and we need to maintain the original fully-connected beam pair bipartite graph for orthogonal beam training. However, in Fig. 10b, because UE channels under a wider DOD support range tend to be more spatially orthogonal to each other, UCVS is more robust to small values of γ and it still performs better than that of JSDM when $\gamma = 0.7$. Meanwhile, with more spatial orthogonality of UE channels, the NOBT phase will introduce tolerable pilot contamination and interference, but the system performance can benefit more from the training cost reduction. Therefore, the optimal γ is less than one as Fig. 10b exhibits.

Fig. 11 exhibits results of net average throughput varying with the scatterer density ρ_s. Due to lack of space, we only present results with $\theta_{all} = 20°$, while behavior of the average rate is similar when $\theta_{all} = 40°$. With the increase of scatterer density, the UE channel is less sparse and the performance gap between UCVS and JSDM becomes smaller. Consider a scenario with dense scatterers in a narrow DOD support interval; in that case many close-by scatterers act as a common scatterer, which can not be distinguished from the perspective of either BS or UE. The eigenspace of UE channel covariances will probably largely overlap. Therefore, by grouping all UEs



(a) $\theta_{all} = 20°$



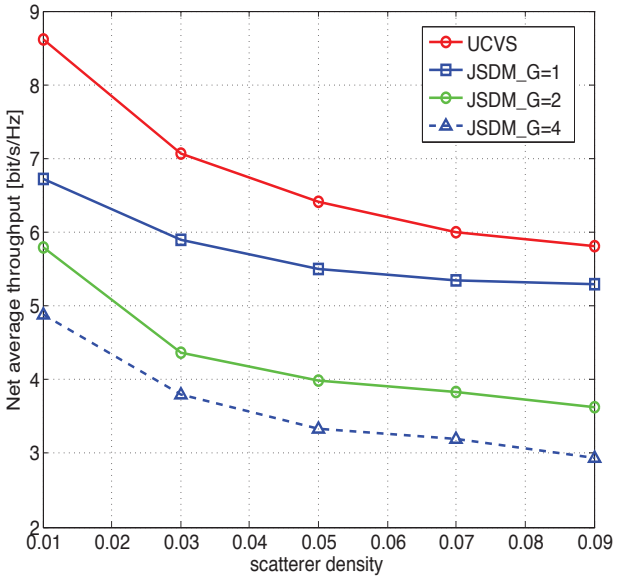
(b) $\theta_{all} = 40°$

Fig. 10: Net average throughput vs. γ for $\rho_d = 5$ dB

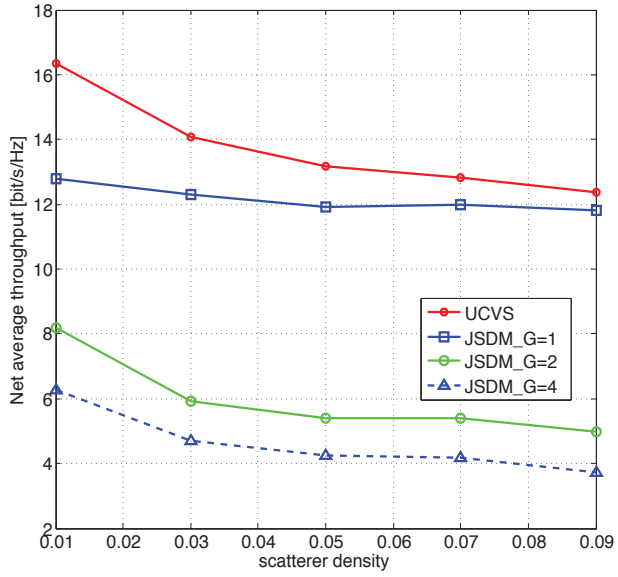
together and forming the JSDM-like analog beamformer approaches the optimal covariance-based solution, which aligns with our proposed method eventually for large ρ_s. However, for a typical sparse mm-wave channel, e.g., when $\rho_s = 0.01$, the number of MPCs is typically less than 10, $\forall i$, and the proposed scheme shows significant performance advantage for a typical mm-wave system operating at -5 dB: it outperforms JSDM by 35%.

In summary, with short coherence times at mm-wave frequencies, the UCVS benefits from NOBT and STDT, while with large coherence time at the high SNR regime, orthogonal training is necessary to avoid pilot contamination, and both UCVS and JSDM scheme achieve almost the same performance as Fig. 8 and Fig. 9 show. From Fig. 10b, we can observe an optimal trade-off between the reduction of training

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2767580, IEEE Transactions on Wireless Communications

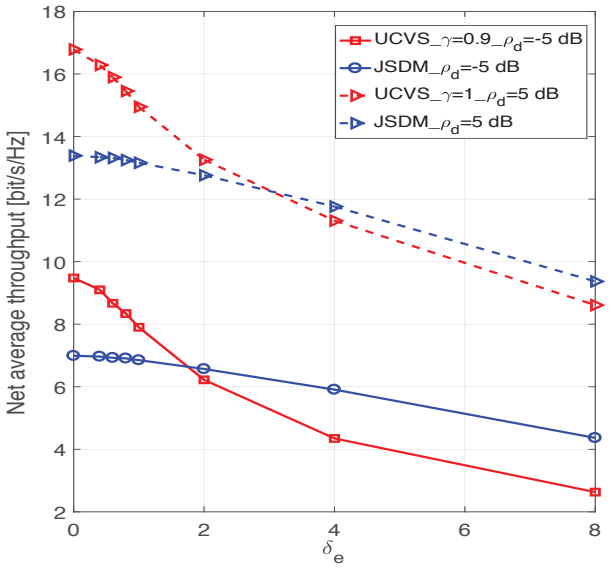


(a) $\rho_{\mathrm{d}} = -5$ dB

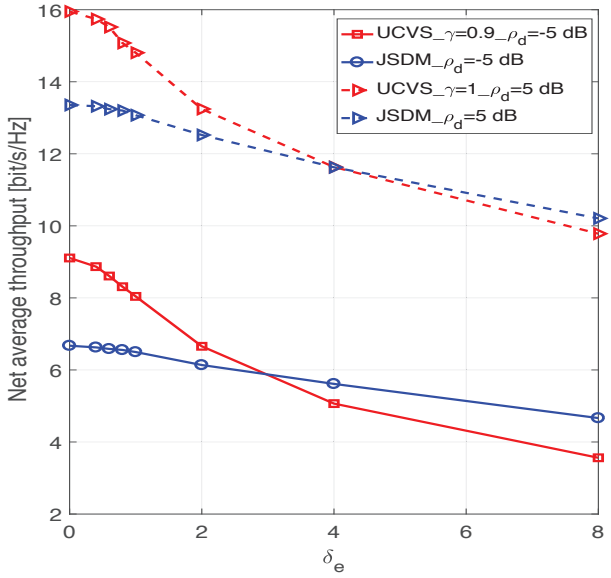


(b) $\rho_{\mathrm{d}} = 5$ dB

Fig. 11: Net average throughput vs. ρ_{s} for $\theta_{\mathrm{all}} = 20°$



(a) $\theta_{\mathrm{all}} = 20°$



(b) $\theta_{\mathrm{all}} = 40°$

Fig. 12: Net average throughput vs. δ_{e} for $T_{\mathrm{cor}} = 20$

cost and interference suppression, which illuminates the future work to explore the optimal threshold setting dependent on different propagation environments. For the impacts by the scatterer densities, we can find that sum rates of both schemes monotonically decrease with the channel sparsity until the saturation from Fig. 11. The performance gap compared with JSDM is much larger at low scatterer density than that at large one, since the proposed scheme is able to exploit the channel sparsity to improve the system peroformance.

To incorporate the impacts of non-ideal knowledge of channel statistics, we consider the following estimation model of MPC directions $\tilde{\phi}_{ip}$=$\phi_{ip} + e_{\mathrm{BS},ip}, \tilde{\theta}_{ip}$=$\theta_{ip} + e_{\mathrm{UE},ip}, \forall i, p$, where $e_{\mathrm{BS},ip} \sim \mathcal{N}(0, \delta_{\mathrm{e}}^2)$ and $e_{\mathrm{BS},ip} \sim \mathcal{N}(0, \delta_{\mathrm{e}}^2)$ represent the estimation error of DOD and DOA, respectively. $\tilde{\phi}_{ip}$ and $\tilde{\theta}_{ip}$ are estimated DODs and DOAs, respectively. For simplicity, we let estimation error variables follow i.i.d. Gaussian distribution with mean zero and variance δ_{e}^2. Therefore, by adjusting different values of δ_{e}, we can investigate impacts of direction misalignment with different extents.

Fig. 12 exhibits sum rates varying with δ_{e} for different parameter settings at $T_{\mathrm{cor}} = 20$, where channel model and system configurations are maintained the same as in Section VI-A. For the JSDM at $T_{\mathrm{cor}} = 20$, we only exhibit results corresponding to one UE group, which is optimal as exhibited in Fig. 8 and Fig. 9. For the proposed UCVS, we consider different settings of energy threshold γ for different SNRs: at $\rho_{\mathrm{d}} = 5$ dB, we let γ to be 1, and maintain the full beam pair bipartite graph, while γ is adjusted to be 0.9 at $\rho_{\mathrm{d}} = -5$ dB. We can observe that the proposed scheme requires more accurate directional characteristics, while JSDM

without UE grouping is more robust to imperfect channel statistics. However, with the help of a large array and sufficient SNR, the angle estimation deviation can be made small, e.g., less than 1 degree, with which the proposed scheme can outperform the JSDM scheme.

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 beamformer design, training resource allocation, and digital beamformer design, respectively. A UE-specified "virtual sectorization" employs the STDT phase and 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. For future work, we will consider the optimization of the threshold in the proposed scheme to strike weak paths, which is dependent on different propagation environments. Meanwhile, simulations over real mm-wave channel data will be investigated to check the practical applicability of the proposed scheme.

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, by the National Science Foundation, and by the National High Technology Research and Development Program of China (No. 2014AA01A705).

References

[1] "FCC 16-89," Federal Communications Commission, Tech. Rep., Jul. 2016.

[2] 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.

[3] 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, Jan. 2013.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Select. Topics Signal Processing*, vol. 8, no. 5, pp. 831 – 846, Oct. 2014.

[10] F. Sohrabi and Y. Wei, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Select. Topics Signal Processing*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[11] T. Bogale, L. Le, A. Haghighat, and L. Vandendorpe, "On the number of RF chains and phase shifters, and scheduling design with hybrid analogdigital beamforming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311–3326, May. 2016.

[12] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, Jan. 2015.

[13] A. Alkhateeb, G. Leus, and R. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[14] J. Li, L. Xiao, X. Xu, and S. Zhou, "Robust and low complexity hybrid beamforming for uplink multiuser mmwave MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1140–1143, Jun. 2016.

[15] 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.

[16] 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.

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

[18] 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.

[19] T. Bogale, L. Le, and X. Wang, "Hybrid analog-digital channel estimation and beamforming: Training-throughput tradeoff," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 5235 – 5249, Dec. 2015.

[20] X. Zheng, H. Zhang, W. Xu, and X. You, "Semi-orthogonal pilot design for massive MIMO systems using successive interference cancellation," in *Proc. IEEE GLOBECOM*, Dec. 2014.

[21] 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, Jun. 2014.

[22] A. Molisch, *Wireless communications*, 2nd ed. IEEE Press-Wiley, 2011.

[23] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 90–100, Jan. 2006.

[24] "Study on channel model for frequency spectrum above 6 ghz," 3rd Generation Partnership Project (3GPP), Tech. Rep. 38.900.

[25] S. Haghighatshoar and G. Caire, "Massive MIMO channel subspace estimation from low-dimensional projections," *IEEE Trans. Signal Processing*, vol. 65, no. 2, pp. 303 – 318, Jan. 2017.

[26] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE signal processing magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[27] B. H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. I. Pedersen, "Channel parameter estimation in mobile radio environments using the sage algorithm," *IEEE Journal on selected areas in communications*, vol. 17, no. 3, pp. 434–450, 1999.

[28] J. Nam, A. Adhikary, J. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Select. Topics Signal Processing*, vol. 8, no. 5, pp. 876–890, Oct. 2014.

[29] 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, Feb. 2013.

[30] 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.

[31] 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.

[32] Z. Li, S. Han, and A. Molisch, "Hybrid beamforming design for millimeter-wave multi-user massive MIMO downlink," in *Proc. IEEE ICC*, May. 2016.

[33] 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.

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

[35] 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.

[36] 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.

[37] 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.

[38] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, and P. Doncker, "The COST 2100 MIMO channel model," *IEEE Wireless Commun. Mag.*, vol. 19, no. 6, pp. 92–99, Dec. 2012.