Directional Training and Fast Sector-based Processing Schemes for mmWave Channels

Zheda Li, Nadisanka Rupasinghe, Ozgun Y. Bursalioglu, Chenwei Wang, Haralabos Papadopoulos, Giuseppe Caire

1Dept. of EE, University of Southern California, Los Angeles
2Dept. of ECE North Carolina State University, Raleigh, NC
3Docomo Innovations Inc, Palo Alto, CA
4Technische Universität Berlin, Germany
zhedali@usc.edu, rprupasi@ncsu.edu, {obursalioglu, cwang, hpapadopoulos}@docomoinnovations.com, caire@tu-berlin.de

Abstract—We consider a single-cell scenario involving a single base station (BS) with a massive array serving multi-antenna terminals in the downlink of a mmWave channel. We present a class of multiuser MIMO schemes, which rely on uplink training from the user terminals, and on uplink/downlink channel reciprocity. The BS employs virtual sector-based processing according to which, user-channel estimation and data transmission are performed in parallel over non-overlapping angular sectors.

The uplink training schemes we consider are non-orthogonal, that is, we allow multiple users to transmit pilots on the same pilot dimension (thereby potentially interfering with one another). Elementary processing allows each sector to determine the subset of user channels that can be resolved on the sector (effectively pilot contamination free) and, thus, the subset of users that can be served by the sector. This allows resolving multiple users on the same pilot dimension at different sectors, thereby increasing the overall multiplexing gains of the system. Our analysis and simulations reveal that, by using appropriately designed directional training beams at the user terminals, the sector-based transmission schemes we present can yield substantial spatial multiplexing and ergodic user-rates improvements with respect to their orthogonal-training counterparts.

I. INTRODUCTION

5G standardization efforts and deployments are projected to bring great performance gains with respect to their predecessors in a multitude of performance metrics, including user and cell throughput, end-to-end delay, and massive device connectivity. It is widely expected that these 5G requirements will be met by utilizing a combination of additional resources, including newly available licensed and unlicensed bands, network densification, large antenna arrays, and new PHY/network layer technologies. To meet the throughput/unit area requirements, for instance, 5G systems would need to provide much higher spatial multiplexing gains (e.g., number of users served simultaneously) than their 4G counterparts.

Large antenna arrays and massive MIMO are considered as key technologies for 5G and beyond. It is expected that new generation deployments would have to utilize the cm and mmWave bands where wide chunks of spectrum are readily available. Note that the spacing of antenna arrays is proportional to the wavelength, at mmWave, so large arrays can be packed even on small footprints. Such large-size arrays will be critical in combatting with the harsher propagation characteristics experienced at mmWave.

Massive MIMO, originally introduced in [1], [2], can yield large spectral efficiencies and spatial multiplexing gains through the use of a large number of antennas at the base stations (BSs). Large arrays enable focusing the radiated signal power and creating sharp beams to several users simultaneously, allowing a BS to serve them simultaneously at large spectral efficiencies.

In order to achieve large spectral efficiencies in the downlink (DL) via multiuser (MU) MIMO, channel state information at the transmitter (CSIT) is needed. Following the massive MIMO approach [2], CSIT can be obtained from the users’ uplink (UL) pilots via Time-Division Duplexing (TDD) and UL/DL radio-channel reciprocity. This allows training large antenna arrays by allocating as few UL pilot dimensions as the number of single-antenna users simultaneously served.

As is well known, with isotropic channels, the number of users that can be simultaneously trained (or the system multiplexing gains) is limited by the coherence time and bandwidth of the channel [3], [4]. Noting that the coherence time is inversely proportional to the carrier frequency, increasing the carrier frequency ten-fold, e.g., from 3 GHz to 30 GHz, results in a ten-fold decrease in coherence time, and, thereby, in the number of user channels that can be simultaneously trained within the coherence time of the channel.

In this paper we focus on single-cell DL transmission over a mmWave channel, enabled by UL training and UL/DL channel reciprocity [5]. We take advantage of the sparsity of mmWave channels in the angular domain to devise schemes that yield increases in the system spatial multiplexing gains. Indeed, typical mmWave channels are characterized by fewer multipath components than channels at lower frequencies [6], [7], [8], [9] resulting in a sparser angular support, both at the BS and the user terminal. This channel sparsity can be exploited to train multiple user channels simultaneously, that is, training multiple users using the same pilot dimension.

We consider a combination of non-orthogonal UL training from the user terminals based on pilot designs in [10] and sector-based processing and precoding from the BS with the goal to increase aggregate spatial multiplexing gains and user rates. The challenge with more than one user transmitting pilots on the same pilot dimension is pilot contamination.
which can substantially limit massive MIMO performance, as the beam used to send data (and therefore beamforming) to one user also beamforms unintentionally at the other (contaminating) user terminal.

In this work, multiple users, each equipped with many antenna elements and a single RF chain, are scheduled to transmit beamformed pilots on the same pilot dimension, thereby increasing the number of users simultaneously transmitting pilots for training. We exploit the presence of a massive Uniform Linear Array (ULA) at the BS and a form of pre-sectORIZATION in the Angle of Arrival (AoA) domain. Elementary processing at each sector allows determining the subset of user channels that can be resolved on the sector, effectively pilot contamination free. Each sector then serves only the subset of users whose channels it can resolve. This allows resolving multiple users on the same pilot dimension at different sectors, thereby increasing the overall multiplexing gains of the system.

Our approach has strong connections but also important differences with respect to joint spatial division and multiplexing (JSDM), a two-stage method proposed in [11]. JSDM partitions users into groups with approximately similar channel covariances, and exploits two-stage downlink beamforming. In particular, precoding comprises a pre-beamformer, which depends on the user-channel covariances and minimizes interference across groups, in cascade with a MU MIMO precoder, which uses instantaneous CSI to multiplex users within a group. Using JSDM, two users with no overlapping AoA support in their channels can be trained and served simultaneously. JSDM has also been studied over mmWave band channels [12]; assuming full knowledge of the angular spectra of all the users, user scheduling algorithms were devised to maximize the spatial multiplexing, or received signal power. Our work similarly harvests spatial multiplexing gains, but the support of each user’s spectra are not a priori known and no special scheduling is employed.

We also study how varying the user beam width can affect these harvested multiplexing gains. Indeed, using a directional beam at a user terminal makes its channel sparser in terms of the number of sectors that are excited at the BS, thereby leaving more sectors available to resolve other users’ channels.

In this paper, we assume TDD operation and focus on DL data transmission enabled by UL pilot transmissions from the user terminals and reciprocity-based training [2]. As a result, in the case of uplink pilot (downlink data) transmission, $\theta_n$ and $\tilde{\theta}_n$ denote the $n$-th path angles of arrival (departure) and departure (arrival).

Spatial filtering can be applied at both the BS and the user terminal side. Given that each user terminal has a single RF chain, a user may transmit its pilot on an arbitrary physical angle $\phi$, while each terminal is equipped with an M-element ULA and a single RF chain. We assume OFDM and a quasistatic block fading channel model whereby the channel of the $k$-th user stays fixed within a fading block (within the coherence time and bandwidth of the channel). During a given fading block, the channel response between the BS and user $k$ is the $M \times M$ matrix $^1$ [6], [13]:

$$H_k(f) = \sum_{n=1}^{N_p} \beta_n a(\theta_n) a^H(\tilde{\theta}_n) e^{-j2\pi\tau_n f},$$

where $N_p$ is the number of paths, and $\beta_n$ and $\tau_n$ denote the complex gain and relative delay, respectively, associated with the $n$-th path$^2$. The $M \times 1$ vector $a(\theta)$ and the $M \times 1$ vector $a(\tilde{\theta})$ represent the array response and steering vectors, and are 1-periodic in $\theta$. The normalized angle $\theta$ is related to the physical angle $\phi$ (measured with respect to array broadside) as $\theta = D \sin(\phi)$, where $D$ is the antenna spacing between two antenna elements normalized by the carrier wavelength. Assuming a maximally spread channel in angular domain, the support of both $a(\theta)$ and $\tilde{a}(\theta)$ are $[-1/2, 1/2]$, as in [6].

In this paper, we assume JSDM and focus on DL data transmission enabled by UL pilot transmissions from the user terminals and reciprocity-based training [2]. As a result, in the case of uplink pilot (downlink data) transmission, $\theta_n$ and $\tilde{\theta}_n$ denote the $n$-th path angles of arrival (departure) and departure (arrival).

Spatial filtering can be applied at both the BS and the user terminal side. Given that each user terminal has a single RF chain, a user may transmit its pilot on an arbitrary physical angle $\phi$. Letting $\alpha_q(\theta) = \beta_q a(\theta) b$, and using $P(b)$ to denote the set of indices of paths that are excited with user’s UL transmission via beam $b$, the physical model for the vector channel can be written as follows:

$$h_k(f) = h_k(f; b) = \sum_{n\in P(b)} \alpha_n(b) a(\theta_n) e^{-j2\pi\tau_n f}. \tag{1}$$

We let $R_k \triangleq \mathbb{E}[h_k(f)h_k^*(f)]$ denote the $k$-th user channel covariance matrix and note that, due to uncorrelated scattering, $R_k$ is independent of the tone index, $f$. Given that our focus is on the large $M$ case, we will assume that the DFT matrix whitens $R_k$ and, as a result, $R_k$ is circulant$^3$. Hence, the eigendecomposition of $R_k$ is given by $R_k = F \Lambda_k F^H$, with $F$ denoting the $M \times M$ DFT matrix, and $\Lambda_k = \text{diag}(\lambda_{1,k}, \ldots, \lambda_{M,k})$ where $\lambda_{1,k}, \ldots, \lambda_{M,k}$ are the eigenvalues of $R_k$.

The MU MIMO schemes we consider in this paper combine a form of spatial division and multiplexing based on instantaneous CSI. The schemes rely on a form of pre-sectORIZATION in the AoA domain. First, $h_k(f)$ is projected onto $F$ to generate the $M \times 1$ vector of channel observations $g_k(f) \triangleq F h_k(f)$. Subsequently, the $M$ entries of $g_k(f)$ are split into $S$ non-overlapping “sector” groups. In particular, assuming without loss of generality, that $g = M/S$ is an integer, each sector comprises $g$ consecutive entries of $g_k(f)$.$^4$

$^1$We assume reciprocal uplink and downlink channels hence we use $H_k(f)$ for both. See [5].

$^2$For notational convenience, we have suppressed the dependence of $N_p$, $\beta_n$, $\tau_n$, $\theta_n$ and $\tilde{\theta}_n$ on the user index $k$.

$^3$Indeed, for ULAs with large $M$, the eigenvectors of the channel covariance matrix are accurately approximated by the columns of a DFT matrix [11].

$^4$If $M$ is not divisible by $S$, groups of different sizes can be arranged.

II. SYSTEM MODEL

We consider a single-cell scenario, involving a single BS serving $K_{\text{tot}}$ user terminals. The BS is equipped with an $M$-element ULA and $M$ RF chains (i.e., one RF chain per antenna), while each terminal is equipped with an $M$-element ULA and a single RF chain. We assume OFDM and a quasistatic block fading channel model whereby the channel of the $k$-th user stays fixed within a fading block (within the coherence time and bandwidth of the channel). During a given fading block, the channel response between the BS and user
We let $g_{s,k}(f)$ denote the $g \times 1$ vector associated with sector $s$ for $s \in \{1, 2, \ldots, S\}$: $g_{s,k}(f)$ can be expressed as
\begin{equation}
    g_{s,k}(f) = F_s h_{s,k}(f),
\end{equation}
where the $M \times g$ matrix $F_s$ comprises the $s$-th set of $g$ consecutive columns of $F$. It is worth remarking that, since the entries of $g_{k}(f)$ are uncorrelated ($\mathbb{E}[g_{k}(f)\bar{g}_{k}^H(f)] = \Lambda_{k}$); in this way the $M \times 1$ channel vector between a single BS and a user is turned into $S$ orthogonal $g \times 1$ sector channels without uncorrelated entries.

We define the average channel gain between a user $k$ and a sector $s$ as follows:
\begin{equation}
    \hat{\lambda}_{s,k} = \frac{1}{g} \sum_{i=(s-1)g+1}^{sg} \lambda_{i,k}.
\end{equation}

In the schemes we consider, the UL pilot transmissions by the user terminals allow each sector to detect the subset of the users it sees with sufficiently high pilot SINR, and subsequently serve the associated user streams with a form of zero-forced beamforming. In the baseline training schemes where each user is given a dedicated UL pilot dimension and is thus not interfered by the other users’ pilots, user $k$ is considered to have a high pilot SINR in sector $s$ as long as $\hat{\lambda}_{s,k}$ exceeds some predetermined threshold $\gamma$ as the user’s pilot is not interfered with other users’ pilots.

We also investigate the viability of non-orthogonal training schemes according to which multiple users are assigned on the same UL pilot dimension. In this case, if the pilots of multiple users using the same pilot dimension are received at sufficiently high power at a given sector (i.e., UL pilots collide at this sector), none of these user channels are resolvable. With non-orthogonal training, user $k$ is considered resolvable in sector $s$ if no collision is declared in its dedicated pilot dimension in sector $s$, and $\hat{\lambda}_{s,k}$ exceeds $\gamma$. Details of detecting high pilot SINR (or, resolvable) users are given in Sec. IV.

With non-orthogonal training, given $\tau$ dimensions per quasi-static fading block are used for UL training, each sector can at most serve simultaneously $\tau$ users per fading block. A user can be served by more than one sector at a time and each sector can serve more than one user at a time.\(^5\) The instantaneous multiplexing gain over a fading block is thus the number of users that are served (by at least one sector) in that block and can exceed the available pilot dimensions, $\tau$. With orthogonal training, the multiplexing gains are upper-bounded by the number of scheduled users $\tau$.

As (1) reveals, the choice of the beam $b$ employed by the user terminal to transmit its UL pilot affects $\mathcal{P}(b)$, the set of indices of paths that are excited. As explained in Appendix A of [14], different training beams may excite different paths but also different numbers of paths. In fact, the sparsity of the user channel in the AoA domain (as reflected by the number of $\hat{\lambda}_{s,k}$’s exceeding $\gamma$) can be controlled by the choice of the beam width. In Secs. III-IV we describe UL training and precoding schemes and tools for analyzing their performance for the case that each user employs a fixed but arbitrary beam for UL pilot transmission (and DL data reception). Subsequently, Sec. V studies the effect of the beam width choice on multiplexing gains and user rates.

It is worth noting that [12] also uses sparsity in AoA domain in the mmwave band to increase multiplexing gains. In [12], the $\Lambda_{k}$’s are assumed to be known prior to scheduling. Indeed, various user scheduling algorithms are designed to assign users to individual eigen directions based on knowledge of the $\Lambda_{k}$’s. In contrast, our work does not assume knowledge of the eigenvalues $\hat{\lambda}_{i,k}$’s or $\hat{\lambda}_{s,k}$’s to schedule transmission of user streams in each sector.

### III. DL MU-MIMO Precoding

In this section we describe the DL precoding schemes under consideration. We assume wideband scheduling, according to which a scheduling slot comprises $Q > 1$ concurrent fading blocks, an let $\tau$ denote the number of available orthogonal pilot dimensions per fading block. Each fading block can be viewed as spanning a contiguous set of time-frequency elements in the OFDM plane that are within the coherence bandwidth and time of the user channels. Since the fading blocks in a slot are concurrent (i.e., distinct fading blocks span distinct subbands over the same set of OFDM symbols), we index the fading blocks in a slot using a fading-block frequency index $f \in \{1, 2, \ldots, Q\}$. With this interpretation $g_{s,k}(f)$ in (2) denotes to the channel of user $k$ in sector $s$ and fading block $f$.

We assume $L$ users (out of the total of $K_{\text{tot}}$ users served by the BS) are scheduled (in round robin fashion) per slot by the BS. In the context of the baseline orthogonal training scheme, the BS schedules $L = \tau$ users per scheduling slot for UL pilot transmission. Thus $\tau$ users send orthogonal pilots on each fading block, i.e., one user per pilot dimension ($K = 1$).

With non-orthogonal UL training, as in [10], the BS schedules $L = K\tau$ users per slot for some $K > 1$. Hence, $K > 1$ users send pilots per pilot dimension. We use $\sigma_k$ to denote the pilot dimension used by user $k$, and $K_{\sigma}$ to denote the indices of users assigned to pilot dimension $\sigma$ for $1 \leq \sigma \leq \tau$.

We consider DL transmission over a generic slot, and assume without loss of generality that the scheduled users have indices from 1 to $L$. Assuming user $k$ uses the same beam $b = b_{k}$ for UL pilot transmission and as a receive front-end in the DL MIMO phase, the received signal at user $k$ over one channel use within fading block $f$ is given by
\begin{equation}
    r_k = \sqrt{\rho_d x^T(f)} h_k(f) + n_k
\end{equation}
where $x$ is the precoded signal, and where $n_k$ represents IID noise with $n_k \sim \mathcal{CN}(0, 1)$ and $\rho_d$ is the DL SNR.

In the MU-MIMO schemes we consider, precoding is sector based. In particular, based on UL training, each sector resolves the channels of a subset of the $L$ users and serves them simultaneously. We let
\begin{equation}
    X_{s,k} = 1(\gamma, \infty)(\hat{\lambda}_{s,k})
\end{equation}
denote whether or not user $k$ is present on sector $s$ and
\[ X_s = \{k; X_{s,k} = 1\} \]  
(6)
denote the set of all users that are present in sector $s$. In the
preceding schemes we consider, user $k$ is resolved on sector $s$
(and thus will be served by sector $s$) if and only if $X_{s,k} = 1$
and there is no other user $k'$ sharing the same dimension as
user $k$ for which $X_{s,k'} = 1$. Specifically we let $D_{s,k}$ denote
whether or not user $k$’s channel can be resolved on sector $s$:
\[ D_{s,k} = X_{s,k} \left[ \prod_{k' \in X_{s,k} \setminus \{k\}} (1 - X_{s,k'}) \right], \]  
(7)
\[ D_s = \{k; D_{s,k} = 1\} \]  
(8)
be the subset of present users whose channels are resolvable in
sector $s$. Fig. 1 shows an example, involving two users using a
common pilot dimension, a BS and four of its sectors, and two
scatterers. As the figure reveals, user 1 is present in sectors 1
and 2, while user 2 is present in sectors 2 and 3. As a result,
the channel of user 1 is resolvable in sector 1, the channel of
user 2 is resolvable in sector 3, and neither user channel is
resolvable in sector 2 or 4.

In general, not all present users are resolvable and we have
$D_s \subseteq X_s$. Indeed, as inspection of (7) reveals if there are
are two users $k$ and $k'$ present in sector $s$ (i.e., $X_{s,k} = X_{s,k'} = 1$)
that use the same pilot dimension (i.e., with $\sigma_k = \sigma_{k'}$), we
have $D_{s,k} = D_{s,k'} = 0$. This is consistent with the fact that
neither channel can be resolved due to the pilot collision. The
number of sectors that can resolve (and thus will serve) user $k$ is
hence given by $N_k = \sum_{s=1}^{S} D_{s,k}$, while the number of
users that are actually served in the slot is given by
\[ L' = |\{k; N_k > 0\}| \]  
(9)
and, in general, $L' \leq L$.

In this paper we focus on a particular form of linear zero-
forced beam-forming. All $L'$ users are given equal power, that is, power $\rho_d/L'$. Furthermore for any served user $k$, the power
allocated to its stream is equally split across all sectors that
resolved the user’s channel. Hence, the $k$-th user’s channel
receives power $\rho_d/(L'N_k)$ from each sector that serves the
user. The precoded $1 \times M$ signal transmitted by the BS is
given by
\[ \mathbf{x}(f) = \sum_{s=1}^{S} \mathbf{u}(f) \mathbf{V}_s^H(f) \mathbf{F}_s^H \]  
(10)
where $\mathbf{u} = [u_1 \ u_2 \ldots \ u_L]$ is the information bearing
signal with $u_k \sim \mathcal{CN}(0,1)$, and
\[ \mathbf{V}_s(f) = [v_{s,1}(f) \ v_{s,2}(f) \ldots \ v_{s,L}(f)] \]  
(11)
denotes the $g \times L$ precoder at sector $s$ and fading block $f$.
In particular, $v_{s,k}(f) = 0$ for any $k \notin D_s$. For any $k \in D_s$,
$v_{s,k}(f)$ is in the direction of the unit-norm vector that is zero-
focused to all other resolvable user-sector channel estimates,
i.e., to $\{\mathbf{g}_{s,k}(f); k' \in D_s \setminus \{k\}\}$ where $\{\mathbf{g}_{s,k}(f)\}$’s denote
the estimates of $\{\mathbf{g}_{s,k}(f)\}$’s. Also $\|v_{s,k}(f)\|^2 = 1/(L'N_k)$. Note
that with this type of precoding, $\mathbf{V}_s(f)$ is invariant to any scalar
(and complex) scalings of any of the $\mathbf{g}_{s,k}(f)$, for $k \in D_s$.
Substituting the expression for $\mathbf{x}(f)$ from (10) in (4), and
using the fact that $\mathbf{h}_k(f) = \mathbf{F}\mathbf{g}_k(f)$ we obtain
\[ r_k = \sqrt{\rho_d} \sum_{s=1}^{S} \mathbf{u}(f)\mathbf{V}_s^H(f)\mathbf{g}_{s,k} + n_k. \]  
(12)

IV. TRAINING, RESOLVABLE CHANNELS AND PERFORMANCE METRICS

We next consider orthogonal and non-orthogonal UL training
and its implications on user channel resolvability. Each
scheduled user $k$ for $1 \leq k \leq L$ is scheduled to transmit pilots
on pilot dimension $\sigma_k$, that is, one of the $\tau$ pilot dimensions.
Each pilot dimension comprises $Q$ resource elements, one per
fading block. We let $p_k = [p_k(1) \ p_k(2) \ldots \ p_k(Q)]^T$
denote the UL pilot vector transmitted by user $k$, with $p_k(f)$
denoting the pilot value used by user $k$ on fading block $f$.

The received signal by the BS array that is based on the
pilots transmitted by the user set $\mathcal{K}_\sigma$ on the pilot dimension
$\sigma$ in fading block $f$ is given by
\[ \mathbf{y}_\sigma(f) = \sqrt{\rho_p} \sum_{k \in \mathcal{K}_\sigma} \mathbf{h}_k(f)p_k(f) + \mathbf{w}_\sigma(f), \]  
(12)
where $\mathbf{y}_\sigma(f)$ is the received vector of length $M$, $\rho_p$ is the UL
SNR, and the noise $\mathbf{w}_\sigma$ is $\mathcal{CN}(0,\mathbf{I})$. The corresponding $s$-th
sector observations are given by projecting $\mathbf{y}_\sigma(f)$ onto $\mathbf{F}_s$
\[ \tilde{\mathbf{y}}_{s,\sigma}(f) = \mathbf{F}_s^H \mathbf{y}_\sigma(f) = \sqrt{\rho_p} \sum_{k \in \mathcal{K}_\sigma} p_k(f)\mathbf{g}_{s,k}(f) + \tilde{\mathbf{w}}_{s,\sigma}, \]  
(13)
and where $\tilde{\mathbf{w}}_{s,\sigma} = \mathbf{F}_s^H \mathbf{w}_\sigma \sim \mathcal{CN}(0,\mathbf{I})$, since $\mathbf{F}_s^H \mathbf{F}_s = \mathbf{I}$.

A. Orthogonal training:

In the orthogonal training setting, user $k$ for $1 \leq k \leq \tau$
transmits pilots on the dedicated pilot dimension $\sigma_k = k$ (i.e.,
$\mathcal{K}_\sigma = \{\sigma\}$), and, as a result, there is no user $k' \neq k$
for which $\sigma_k = \sigma_{k'}$. Assuming also, without loss of
generality, that $p_k(f) = 1$, the associated received signal in fading block $f$ by the BS array based on the pilot transmitted by user $k$ on
the pilot dimension $k$ (since $\sigma_k = k$) from (12) is given by
\[ \mathbf{y}^u_k(f) = \sqrt{\rho_p} \mathbf{h}_k(f) + \mathbf{w}_u(f), \]  
(14)
while the corresponding $s$-th sector observations are given by
\[ \tilde{\mathbf{y}}_{s,k}(f) = \sqrt{\rho_p} \mathbf{g}_{s,k}(f) + \tilde{\mathbf{w}}^u_{s,k}. \]  
(15)
The precoder uses the following estimate of the $k$-th user’s instantaneous channel on fading block $f$ and sector $s$:

$$\tilde{g}_{s,k}(f) = \tilde{y}_{s,k}(f).$$

(16)

Note that this estimate does not make any use of the pilot SNR, and does not rely on knowledge of $\lambda_{s,k}$.

Inspection of (7) reveals, that in the orthogonal scheme, $X_{s,k} = D_{s,k}$, as there is no user $k'$ for which $\sigma_k = \sigma_{k'}$. That is, in the orthogonal scheme, a sector can resolve the channel of user $k$ if a user is present in sector $s$ (i.e., $\lambda_{s,k} \geq \gamma$), and thus $X_s = D_s$. Subsequently, the sector $s$ forms $V_s(f)$ for its resolvable user set $D_s$ according to (11) and using $\tilde{g}_{s,k}(f)$ from (16) for all $k \in D_s$.

Practical schemes for detecting the set of resolvable user channels can be devised by exploiting the key fact that

$$\mathbb{E}[\tilde{g}_{s,k}(f)\tilde{g}_{s,k}(f)'] = \text{diag}(\lambda_{s,0}, \lambda_{s,1}, \ldots, \lambda_{s,K})$$

(17)

for each fading block $f$ in the slot. Noting also that

$$\mathbb{E}[\|\tilde{y}_{s,k}(f)\|^2] = \text{tr}(\mathbb{E}[\tilde{y}_{s,k}(f)\tilde{y}_{s,k}(f)']) = (p_f \lambda_{s,k} + 1),$$

we can devise simple practical detection schemes that benefit from averaging over both the $g$ beams and the $Q$ tones, e.g.:

$$\frac{1}{Qp_fg} \sum_{f=1}^{Q} \|\tilde{y}_{s,k}(f)\|^2 - \frac{1}{\rho_p} \begin{cases} \frac{\lambda_{s,k}}{\lambda_{s,k}} \geq \gamma \\
\lambda_{s,k} = 0 \end{cases}.$$

(18)

If $\hat{D}_{s,k} = 1$, then user $k$’s channel on sector $s$ is considered as resolvable by the BS based on received UL signal.

B. Non-Orthogonal training

In the non-orthogonal training setting we consider, the pilots of $K > 1$ users pilots are aligned on a single pilot dimension. As a result, the non-orthogonal scheme splits the $L = K_T$ scheduled users uniformly across the $K_\sigma$ sets, so that $|K_\sigma| = K$ for each $\sigma \in \{1, 2, \ldots, \tau\}$. It is worth noting that, given $X_s$ (the set of present users in sector $s$) the set of detected and served users from sector $s$, $D_s$, is given by (8) and in general satisfies $D_s \subseteq X_s$. For example, if there is a $s$ for which multiple users are present in sector $s$, i.e., $|X_s \cap K_\sigma| > 1$ then $D_s \subset X_s$. Such situation would correspond to a collision, that is, two or more users using the same pilot dimension are present in sector $s$, in which case, neither one’s channel is resolvable for transmission. Given $D_s$, sector $s$ forms $V_s(f)$ according to (11) and using $\tilde{g}_{s,k}(f) = \tilde{y}_{s,k}(f)$ for all $k \in D_s$ where $\tilde{y}_{s,k}(f)$ is given by (13).

Practical detection schemes that detect which user channels are resolvable in a sector can be also readily devised. Noting

$$\mathbb{E}[\|\tilde{y}_{s,k}(f)\|^2] = g \left( \sum_{k \in K_\sigma} \lambda_{s,k} \right) \rho_p \mathbb{E}[|p_k(f)|^2 + 1],$$

(19)

and assuming a sufficiently large number of beams/sector, $g$, the RHS of (19) can be approximated by $\|\tilde{y}_{s,k}(f)\|^2$. This suggests that a system of $Q$ linear equations (one per fading block on pilot dimension $\sigma$) can be used to obtain $\{\lambda_{s,k}; k \in K_\sigma\}$’s.

Letting, $\mathbf{r}_{s,\sigma} = [\|\tilde{y}_{s,\sigma}(1)\|^2 |\tilde{y}_{s,\sigma}(2)|^2 \ldots |\tilde{y}_{s,\sigma}(Q)|^2]^T$, we have the following system of equations:

$$\mathbf{r}_{s,\sigma} = g \rho_p \mathbf{P}_\sigma \mathbf{\lambda}_{s,\sigma} + g \mathbf{1} + \text{noise},$$

(20)

where $\mathbf{\lambda}_{s,\sigma}$ is a $K \times 1$ vector whose entries comprise the set $\{\lambda_{s,k}; k \in K_\sigma\}$, and where row $f$ of the $Q \times K$ matrix $\mathbf{P}_\sigma$ contains the associated $|p_k(f)|^2$ values. For each $k$ in $K_\sigma$, let also $i_k$ denote the index $i$ for which $[\lambda_{s,k}]_i = \lambda_{s,k}$. For $Q \geq K$, the set $\{p_k(f); k \in K_\sigma, 1 \leq f \leq Q\}$ can be chosen a priori so that $\mathbf{P}_\sigma$ in (20) has full column rank, and hence the presence of each user can be individually detected. One such simple detector of the presence of user $k$ is given by

$$\frac{1}{\rho_p g} \mathbf{e}_{i_k}^T \mathbf{A}_\sigma \mathbf{r}_{s,\sigma} - \frac{1}{\rho_p} \mathbf{e}_{i_k}^T \delta_{s,k} \tilde{X}_{s,k} = 0 \gamma,$$

where $\mathbf{A}_\sigma = (\mathbf{P}_\sigma^H \mathbf{P}_\sigma)^{-1} \mathbf{P}_\sigma^H$, and $\delta_{s} = \mathbf{A}_s \mathbf{1}$, and where $e_n$ is the $n^{th}$ column of the $K \times K$ identity matrix. Note that both $\mathbf{A}_s$ and $\delta_s$ are independent of the user channels and can be computed offline. Subsequently user channel resolvability can be detected by substituting $\tilde{X}_{s,m}$ for $X_{s,m}$ in (7):

$$\hat{D}_{s,k} = \tilde{X}_{s,k} \left[ \prod_{k' \in K_\sigma \setminus \{k\}} \left(1 - \tilde{X}_{s,k'} \right) \right].$$

(21)

Various codes can be designed when $Q \geq K$ that yield $\mathbf{P}_\sigma$ having full column rank. A full column rank matrix $\mathbf{P}_\sigma$ allows estimating each user’s large-scale response on each sector, i.e., all the $\{\lambda_{s,k}\}$’s, thereby allowing to determine the presence of all users on all sectors. This together with (21) allows detecting the users with resolvable channels. To estimate the channels of any user $k$ that has been resolved, however, it is also necessary that $p_k(f) \neq 0, \forall f \in \{1, \ldots, Q\}$.

We remark that choosing a $\mathbf{P}_\sigma$ that is column rank allows detecting the presence of each user but is not necessary for detecting the resolvable user channels. Indeed, in the case that multiple active users (on a common pilot dimension $\sigma$) are present (i.e., collide) on a sector, the code design need only detect the collision event, and not the identities (and large-scale channel gains) of the users that are present in the sector. This fact was exploited in [10] to design ON-OFF codes that are capable of resolving user channels even in cases with $Q < K$ (where $\mathbf{P}_\sigma$ cannot be full rank). However, pilot resource elements where a user’s pilot is OFF (i.e., the $f$ values where where $p_k(f) = 0$) provide no information for channel estimation. As a result, these ON-OFF codes incur extra pilot overheads in order to enable the BS to estimate the resolved-user channels throughout the band [10].

C. Performance Metrics

We consider two types of metrics in evaluating the performance of the proposed schemes. The first metric we use is the slot-averaged multiplexing gains provided by orthogonal and non-orthogonal training schemes:

$$\bar{\overline{M}}G = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} L'(t),$$

(22)

where $L'(t)$ represents the instantaneous multiplexing gain over slot $t$ and is given by (9). The second performance metric is based on ergodic user-rate bounds. In Appendix B of [14],
we provide closed-form rate bound expressions assuming IID channels within each sector, that is, assuming
\[
\tilde{\lambda}_{s,k} = \lambda_{(s-1)M+1,k} = \lambda_{(s-1)M+2,k} = \ldots = \lambda_{sM,k}.
\] (23)
This abstraction is justified in Appendix C of [14].

D. Directional Training and Angular Spectra Sparsity

As inspection of (1) reveals, the number of excited paths and thus the extent to which a trained user channel is sparse in the AoA domain depends on the choice of the user beam. Similar to the AoA domain, where projecting onto the DFT basis F both whitens and sparsifies the channel (that is, \(g_k\) is both white and sparse), we consider creating the user-pilot beam as a linear combination of the AoD eigen directions, \(^6\) i.e., of the columns of the \(M \times M\) DFT matrix, \(F\). In particular, we consider training beams that arise from activating a subset of eigen directions. Letting \(\tilde{f}_i\) denote the \(i\)-th AoD eigen direction (i.e., the \(i\)-th column of \(F\)), we can describe such an \(M \times 1\) user training beam \(b\) in terms of an \(M \times 1\) vector \(c\) with zero-one entries. For each \(\tilde{n}\), with \(1 \leq \tilde{n} \leq M\), \([c]_{\tilde{n}}\) is 1 if \(\tilde{f}_{\tilde{n}}\) is activated (i.e., used as part of the training beam), while \([c]_{\tilde{n}} = 0\) otherwise. We also define the training beam width as the number of activated eigen directions, that is, \(w = 1^T c = \sum_{\tilde{n}=1}^{M} [c]_{\tilde{n}}\). Consequently given \(c\), the corresponding training beam \(b\) is given by \(b = b(c) = Fc/\sqrt{w}\). Note that \(w = 1\) corresponds to the user training on a beam with the narrowest possible width, while \(w = M\) corresponds to omni training.

V. SIMULATIONS AND CONCLUSION

In this section, we study the multiplexing gains and user-throughput performance of the proposed schemes with orthogonal training \((K = 1)\) and non-orthogonal training \((K > 1)\). In order to study the effects of user beam width on channel sparsity in the AoA domain and, subsequently, on multiplexing gains and user throughput, we consider a probabilistic connectivity channel model between each elemental training eigen direction in \(\{\tilde{f}_{\tilde{n}}; \ 1 \leq \tilde{n} \leq M\}\) and each of the \(S\) BS sectors. Specifically, we model the connection between BS sector \(s\) and a directional beam \(\tilde{f}_{\tilde{n}}\) from user \(k\) as a Bernoulli random variable \(X_{s,k}^{(\tilde{n})}\) with success probability \(p\). We also model the \(X_{s,k}^{(\tilde{n})}\)'s as IID\(^7\) in \(s, k, \text{and} \tilde{n}\).

In addition, we use \(\tilde{\lambda}_{s,k}^{(\tilde{n})}\) to denote the \(\tilde{\lambda}_{s,k}\) induced on sector \(s\) when user \(k\) training with elemental eigen direction \(b = \tilde{f}_{\tilde{n}}\), and model it as follows:

\[
\tilde{\lambda}_{s,k}^{(\tilde{n})} \sim \begin{cases} 
\Lambda[\lambda_L, \lambda_H] & \text{if } X_{s,k}^{(\tilde{n})} = 1, \\
0 & \text{if } X_{s,k}^{(\tilde{n})} = 0, 
\end{cases}
\]

\(^6\)For further details regarding how directional training sparsifies the channel in the AoA domain, see Appendix A in [14].

\(^7\)In general, \(Pr[X_{s,k}^{(\tilde{n})} = 1] = w, s, k, \text{and} \tilde{n}\) dependent. In addition, \(X_{s,k}^{(\tilde{n})}\)'s may be dependent random variables. Indeed, for two users \(k\) and \(k'\) nearby it is possible that \(X_{s,k}^{(\tilde{n})}\) and \(X_{s,k'}^{(\tilde{n}')}\) are strongly correlated for some specific training directions \(\tilde{n}\) and \(\tilde{n}'\). Although important in their own right, such spatial consistency investigations are beyond the scope of this paper. For spatial consistency investigations, see [15].

for some \(\lambda_L, \lambda_H\) with \(\lambda_H > \lambda_L > 0\) and where \(U[a, b]\) denotes a uniform distribution in \([a, b]\). Consequently, using a beam \(b_k = b(c_k)\) with beam width \(w_k = w(c_k)\) results in

\[
\tilde{\lambda}_{s,k} = \tilde{\lambda}_{s,k}(b_k) = \frac{1}{w_k} \sum_{\tilde{n}=1}^{M} \tilde{\lambda}_{s,k}^{(\tilde{n})} [c_k]_{\tilde{n}}.
\] (24)

It can be readily verified that, by choosing as a threshold in (5) a value of \(\gamma\) in the range \(0 < \gamma < \lambda_L/M\), and using \(\tilde{\lambda}_{s,k}(b_k)\) as in (24) the resulting \(X_{s,k}\) in (5) satisfies \(X_{s,k} = X_{s,k}(b_k) = \text{OR}_{m, [\tilde{e}_k]_{m=1}^{M} \{c_k\}_{m=1}^{M}} \tilde{\lambda}_{s,k}^{(\tilde{n})}\). Consequently, when user \(k\) uses a given training beam \(c_k\) of beam width \(w_k\), we have

\[
\mathbb{P}(X_{s,k} = 1) = q(w_k),
\] (25)
where \(q(w) \triangleq 1 - (1 - p)^w\). Also, the \(X_{s,k}\)'s are IID in \(s, k\).

We focus on the case where all users choose beams with the same beam width \(w\), and study the resulting multiplexing gains and user throughputs as a function of \(w\) in the range \(1 \leq w \leq M\). Note that, given a common beam width \(w\), the average number of sectors activated by a user is given by \(S(q(w))\). Also, the expected number of scheduled users that are not present at any of the sectors is given by \(L(1 - q(w))\).

As expected, wider beams result in broader angular support, but wider beams also make it less likely that a user is not present at any of the sectors. Using (7), (9), and (25) yields the following expression for \(\bar{M}G\) in (22):

\[
\bar{M}G(w, K) = E[L'] = E \left[ \sum_{k=1}^{K} (1 - (1 - q(w))^{K-1}) \right] 
\]

\[
= K \left[ 1 - \left(1 - q(w) \right)^{K-1} \right]^{S}. \] (26)

We next present a simulation-based study of the proposed schemes using the above model assuming \(K_{\text{tot}} = 100\) users terminals with \(M = 6\) antennas each, and a BS with \(M = 1000\) BS antennas, using sector-based processing over \(S = 25\) sectors (hence \(g = M/S = 40\) beams per sector). We assume \(p = 0.1\), and that \(\tau = 5\) pilot dimensions are available for training per fading block. With these parameters, we have \(q(1) = p = 0.1\) and \(q(M) = 0.47\), implying that the average AoA angular support for a user ranges from 2.5 sectors (achieved with the finest-directional training, i.e., \(w = 1\)) to about 11.72 sectors (achieved with omni-directional training i.e., \(w = M = 6\)). A single drop is created, i.e., a single set of \([X_{s,k}]^{(b)}\)'s are randomly created according to the model. For any given common beam width, \(w\), at any given scheduling instance, each scheduled user picks a \(c\) at random (out of those \(c\)'s yielding beams with beam width \(w\)).

Fig. 2 shows the multiplexing gains per pilot dimension as a function of \(K\) for different beam width values in the range \(1 \leq w \leq M\). If orthogonal training is used \((K = 1)\), for any beam width the multiplexing gain is approximately equal to (and always upper-bounded by) 1. Considering all \(K \geq 1\) options, for any given training-beam width \(w\), there is an optimal value of \(K\), \(K_{\text{opt}}(w)\), which maximizes \(\bar{M}G(w, K)\). The best combination of beam width, \(w\), and
number of scheduled users per pilot dimension, $K$, is given by $(w_{\text{opt}}, K_{\text{opt}}) = \arg\max_{(w, K)} \bar{MG}(w, K) = (1, 13)$ and results in a more than 6-fold increase in multiplexing gains with respect to the orthogonal training scheme. Fig. 3a and Fig. 3b, respectively, display the arithmetic and geometric mean of user throughput as a function of $K$ for beam widths in the range $1 \leq w \leq \bar{M}$. Inspection of the two figures reveals similar trends between the arithmetic and geometric mean of the user rates as a function of $K$ and $w$. Also both the arithmetic mean and the geometric mean are maximized with $(w, K) = (1, 10)$, exhibiting in each case more than a 3-fold increase with respect to the orthogonal training scheme.

Fig. 4 shows the empirical user-rate CDFs for all beam widths in the range $1 \leq w \leq M$. For any $w$, the value of $K$ that maximizes the user-rate arithmetic mean is used. We also plot the empirical user-rate CDF with orthogonal training, for $w = 1$ and $w = 6$ (omni-training). Inspection reveals that optimized non-orthogonal training uniformly outperforms orthogonal training in terms of individual user rates. Furthermore, the minimum pilot beam width is best in this example, as its user-rate CDF dominates all the others.

In conclusion, non-orthogonal UL pilots and simple large-array BS processing are jointly exploited to significantly increase cell and cell-edge throughputs over sparse mmWave channels. The proposed method leverages scheduling multiple users randomly on each available pilot dimension with random (user-chosen) training directions, and coupled with low-complexity spatial processing at the BS to resolve user channels at each BS virtual sector. Although the focus of the paper is cellular transmission and, in particular, single-cell performance, the proposed methods are directly applicable to CRAN scenarios and cell-free type operation [10], [15]. Indeed, large improvements in multiplexing gains are also reported in the context of cell-free type networks in [15], based on simple, albeit spatially consistent channel models that include the effects of common scatterers and blockers.

REFERENCES

[1] T. L. Marzetta, “How much training is required for multiuser MIMO?” in Proc. of the Fortieth Asilomar Conf. on Signals, Systems and Computers, Nov. 2006.
[2] ——, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” IEEE Trans. on Wireless Commun., vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
[3] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, “Multiuser MIMO achievable rates with downlink training and channel state feedback,” IEEE Trans. Inf. Theory, vol. 48, no. 2, pp. 359–383, 2002.
[4] L. Zheng and D. N. C. Tse, “Communication on the Grassmann Manifold: A geometric approach to the noncoherent multiple-antenna channel,” IEEE Trans. Inf. Theory, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.
[5] R. Rogalin, O. Y. Bursalioglu, H. Papadopoulos, G. Caire, A. F. Molisch, A. Michaloliakos, V. Balan, and K. Psounis, “Scalable synchronization and reciprocity calibration for distributed multiuser MIMO,” IEEE Trans. on Wireless Commun., vol. 13, no. 14, Apr. 2014.
[6] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, “Compressed channel sensing: A new approach to estimating sparse multipath channels,” Proc. of the IEEE, vol. 98, no. 6, pp. 1058–1076, June 2010.
[7] W. C. Ao, C. Wang, O. Y. Bursalioglu, and H. Papadopoulos, “Compressed sensing-based pilot assignment and reuse for mobile UEs in mmWave cellular systems,” in IEEE Int. Conf. on Commun., May 2016.
[8] M. K. Samimi and T. S. Rapport, “3-D statistical channel model for millimeter-wave outdoor mobile broadband communications,” in IEEE Int. Conf. on Commun., June 2015.
[9] K. Taejon and D. J. Love, “Virtual AoA and AoD estimation for sparse mmWave MIMO channels,” in IEEE Int. Works. on SPAWC, June 2015.
[10] O. Y. Bursalioglu, C. Wang, H. Papadopoulos, and C. Caire, “RRH based massive MIMO with “on the fly” pilot contamination control,” in IEEE Int. Conf. on Commun., May 2016.
[11] J. Nam, A. Adhikary, J. Y. Ahn, and G. Caire, “Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling,” IEEE Journal of Selected Topics in Signal Proc., vol. 8, no. 5, pp. 876–890, Oct. 2014.
[12] A. Adhikary, E. Al Safadi, M. K. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, “Joint spatial division and multiplexing for mm-wave channels,” IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp. 1239–1255, June 2014.
[13] A. M. Sayeed, “Deconstructing multiantenna fading channels,” IEEE Trans. on Sign Proc., vol. 50, no. 10, pp. 2563–2579, Oct. 2002.
[14] Z. Li, N. Rupasinghe, O. Y. Bursalioglu, C. Wang, H. Papadopoulos, and G. Caire, “Directional training and fast sector-based processing schemes for mmwave channels,” arXiv preprint, 2016. [Online]. Available: http://arxiv.org/pdf/1611.00453.pdf
[15] O. Y. Bursalioglu, C. Wang, H. Papadopoulos, and C. Caire, “A novel alternative to cloud RAN for throughput densification: Coded pilots and fast user-packet scheduling at remote radio heads,” in Proc. of the Fiftieth Asilomar Conf. on Signals, Systems and Computers, Nov. 2016.