Beamspace Aware Adaptive Channel Estimation for Single-Carrier Time-varying Massive MIMO Channels

Gokhan M. Guvensen and Ender Ayanoglu

Center for Pervasive Communications and Computing (CPCC), Dept. of EECS, UC Irvine, CA, USA

E-mail: guvensen@metu.edu.tr and ayanoglu@uci.edu

Abstract—In this paper, the problem of sequential beam construction and adaptive channel estimation based on reduced rank (RR) Kalman filtering for frequency-selective massive multiple-input multiple-output (MIMO) systems employing single-carrier (SC) in time division duplex (TDD) mode are considered. In two-stage beamforming, a new algorithm for statistical pre-beamformer design is proposed for spatially correlated time-varying wideband MIMO channels under the assumption that the channel is a stationary Gauss-Markov random process. The proposed algorithm yields a nearly optimal pre-beamformer whose beam pattern is designed sequentially with low complexity by taking the user-grouping into account, and exploiting the properties of Kalman filtering and associated prediction error covariance matrices. The resulting design, based on the second order statistical properties of the channel, generates beamspace on which the RR Kalman estimator can be realized as accurately as possible. It is observed that the adaptive channel estimation technique together with the proposed sequential beamspace construction shows remarkable robustness to the pilot interference. This comes with significant reduction in both pilot overhead and dimension of the pre-beamformer lowering both hardware complexity and power consumption.

I. INTRODUCTION

It is anticipated that massive MIMO systems in the mm wave range form an important part of 5G systems expected to support much larger (e.g., 1000 times faster) data rates than the currently deployed standards [1]. In practice, channel state information (CSI) is typically obtained with the assistance of the periodically inserted pilot signals [2]. This brings the pilot overhead consumed by the training data to be proportional to the number of active users in the system for uplink training, and the number of BS antennas for downlink training respectively [3]. We assume CSI at the BS can be acquired by means of uplink training in time division duplex (TDD) mode, where the uplink pilots provide the BS with downlink as well as uplink channel estimates simultaneously via leveraging channel reciprocity [2], [3].

The processing of the signals with very large dimensionality, the pilot interference, and the pilot overhead are thought to be limiting factors for an accurate channel acquisition and throughput of massive MIMO transmission in mm wave especially in high mobility or in applications requiring low latency and short-packet duration. Recently, the *two-stage beamforming* concept under the name of Joint Spatial Division and Multiplexing (JSDM) [4], [5] has been proposed to reduce

the dimension of the MIMO channel effectively, and to enable massive MIMO gains and simplified system operations [6]. Even though JSDM is suggested as an effective reduced-complexity two-stage downlink precoding scheme for flat-fading multi-user MIMO systems in frequency division duplex (FDD) mode initially, the idea of *two-stage beamforming* (in [4]–[6]) can be applied to both downlink and uplink transmission in TDD. The key idea lies in *user-grouping*, i.e., partitioning the user population supported by the serving BS into multiple groups each with approximately the same channel covariance eigenspaces. Then, one can decompose the MIMO beamformer at the BS into two steps via the use of a spatial *pre-beamformer*, which distinguishes *intra-group* signals from other groups by suppressing the *inter-group interference* while reducing the signaling dimension. The major complexity reduction comes from the approach that the *pre-beamformer* is properly designed based only on the second-order statistics of the channel, and not on the instantaneous CSI (which varies on a much higher rate).

The wideband massive MIMO channel is expected to be sparse both in the angle and time (delay) domains. The *channel sparsity* [7], [8], which becomes particularly relevant at mm wave frequencies, is observed in practical cellular systems, where the channels are often characterized with limited scattering and hence correlated in the spatial domain; the BS sees the incoming multi-path components (MPCs) under a very constrained angular range, i.e., angle of arrival (AoA) support, and the MPCs occur in clusters in the *angle-delay* plane [7]. Thus, it is important to design statistical *pre-beamformer* yielding an efficient implementation of channel estimator taking this sparsity into account. Moreover, the subsequent stages preceded by the channel estimation can be realized in much lower complexity.

In this paper, reduced rank Kalman filter based CSI estimator together with the sequential design of statistical *pre-beamformer* is proposed for frequency-selective massive MIMO systems employing single-carrier (SC) modulation in TDD mode. With proposed pre-beamformer design, taking the Kalman prediction errors at each eigen-direction and *inter-group* interference in JSDM into account, significant reduction in beamspace dimension is obtained, which reduces the pilot overhead and hardware complexity considerably for hybrid beamforming.

II. System Model

We consider a cellular system based on massive MIMO transmission operating at mm wave bands in the TDD mode employing SC in which a BS, having N antennas, serves K single-antenna UTs. In order to reduce the overhead while acquiring the instantaneous CSI associated with massive MIMO, *two-stage beamforming* [4]–[6] is adopted throughout this paper. It is assumed that K users are partitioned into G groups, where the K_g users in group g have statistically independent but identically distributed (*i.i.d.*) channels [4], [5].

The frequency-selective massive MIMO channel is supposed to be time-varying in general. At the beginning of every coherence interval, all users of the intended group g transmit training sequences with length T. We assume a linear modulation (e.g., PSK or QAM) and a transmission over frequency-selective channel for all UTs with a slow evolution in time relative to the signaling interval (symbol duration). Under such conditions, the baseband equivalent received signal samples, taken at symbol rate (W) after pulse matched filtering, are expressed as[1]

$$\mathbf{y}_n = \underbrace{\sum_{\{k=1,\; g_k \in \Omega_g\}}^{K_g} \sum_{l=0}^{L_g-1} \mathbf{h}_{n,l}^{(g_k)} x_{n-l}^{(g_k)}}_{\text{Intra-Group Signal}}$$

$$+ \underbrace{\sum_{\{g'=1\,|\,g'\neq g\}}^{G} \left(\sum_{\{k=1,\; g_k' \in \Omega_{g'}\}}^{K_{g'}} \sum_{l=0}^{L_{g'}-1} \mathbf{h}_{n,l}^{(g_k')} x_{n-l}^{(g_k')} \right) + \mathbf{n}_n}_{\boldsymbol{\eta}_n^{(g)}:\text{Inter-Group Interference + AWGN}} \quad (1)$$

for $n = 0, 1, 2, \ldots$, where $\mathbf{h}_{n,l}^{(g_k)}$ is $N \times 1$ time-varying multi-path channel vector, namely, the array impulse response of the serving BS stemming from the lth multi-path component (MPC) of kth user in group g at the nth signaling interval. Here, $\left\{ x_n^{(g_k)}; \; n = -L_g + 1, \ldots \right\}$ are the training symbols for the kth user in group g[2], L_g is the channel memory of group g multi-path channels, Ω_g is the set of all UTs belonging to group g with cardinality $|\Omega_g| = K_g$, and $\{g_k\}_{k=1}^{K_g}$ are UT indices forming Ω_g. The $L_g - 1$ symbols at the start of the preamble, prior to the first observation at $n = 0$, are the precursors. Training symbols are selected from a signal constellation $S \in \mathbb{C}$ and $\mathbb{E}\left\{ |x_n^{(g_k)}|^2 \right\}$ is set to E_s for all g_k. In (1), $\mathbf{n}_n$ are the additive white Gaussian noise (AWGN) vectors during uplink pilot segment with spatially and temporarily *i.i.d.* as $\mathcal{CN}(\mathbf{0}, N_0 \mathbf{I}_N)$, and N_0 is the noise power. The first term of (1) is the received signal of the intended group g, named as the *intra-group* signal of group g users. The second term, $\boldsymbol{\eta}_n^{(g)}$, namely the *inter-group interference*, comprises of all the interfering signals, which stem from all inner or outer cell users belonging to different groups other than g.

In (1), we assume users come in groups, either by nature or by the application of proper *user grouping* algorithms in [5], [7], which are out of scope of this work. Finally, the average received signal-to-noise ratio (*snr*) can be defined as $snr \triangleq \frac{E_s}{N_0}$[3].

A. Fundamental Assumptions on Signal and Channel Model

Each resolvable MPC of the users, belonging to any group g, is assumed to span some particular angular sector in azimuth-elevation plane, capturing local scattering around the corresponding UTs' angle of arrival (AoA) (with respect to the BS). Then, their corresponding cross-covariance matrices can be expressed in the form of

$$\mathbb{E}\left\{ \mathbf{h}_{n,l}^{(g_k)} \left(\mathbf{h}_{n,l'}^{(g_{k'}')} \right)^H \right\} = \rho_l^{(g)} \mathbf{R}_l^{(g)} \delta_{gg'} \delta_{kk'} \delta_{ll'}, \quad (2)$$

for $\sum_{l=0}^{L_g-1} \rho_l^{(g)} = 1$ and $\text{Tr}\left\{ \mathbf{R}_l^{(g)} \right\} = 1$ by using the uncorrelated local scattering model where all MPCs are assumed to be mutually independent according to the well-known wide sense stationary uncorrelated scattering (WSSUS) model [7], [8], the multi-path channel vectors are uncorrelated with respect to l, and also mutually uncorrelated with that of the different users (independent of whether in the same group or not). In (2), $\rho_l^{(g)}$ is the power delay profile (pdp) of the group g multi-path channels, showing the average channel strength at each delay, and $\mathbf{R}_l^{(g)}$, with $r_{g,l}$ non-zero or dominant eigenvalues, can be considered as the common spatial covariance matrix of group g UTs at l^{th} delay. In this model, each antenna element at a BS is assumed to see the incoming MPCs at the same common support on the *angle-delay* plane (similar to the one in [7], [9]). Also, the effective rank of $\mathbf{R}_l^{(g)}$, namely, $r_{g,l}$ is expected to be much smaller than the number of array elements, N, due to the *channel sparsity* pronounced at mm wave [7], [10].

The channel *long-term* channel spatial correlation characteristic is assumed to be stationary. In fact, $\mathbf{R}_l^{(g)}$s are slowly varying in time as the AoA of each user signal evolves depending on the user mobility, variation rate of the scattering environment characteristics, etc. [7], [9], [11]. However, this rate of change is significantly lower than that of the small-scale fading (instantaneous CSI), and they can be estimated with guaranteed accuracy for all intended groups in practice with the help of fast initial acquisition techniques or by acquiring the AoA sector of each user group [9]. Here, our focus is on the tracking of instantaneous CSI with significantly reduced complexity and overhead.

When Rayleigh-correlated channel coefficients are assumed such as $\mathbf{h}_l^{(g_k)} \sim \mathcal{CN}\left(\mathbf{0}, \rho_l^{(g)} \mathbf{R}_l^{(g)} \right)$, mutually independent across the users for all g_k,

Spatio-temporal covariance matrix of the *inter-group interference* in (1) can be calculated by taking *long-term* expectation over all MPCs $\mathbf{h}_{n,l'}^{(g_{k'}')}$s other than the ones belonging to

---

[1] Only the UTs, belonging to same group, are assumed to be synchronized for coherent uplink SC transmission.

[2] Training sequences are assumed to be non-orthogonal for synchronized *intra-group* users for SC transmission in general.

[3] It shows the maximum achievable *snr* after beamforming when the beam is steered towards a point, i.e., angular location by assuming that the channel is normalized so that $\frac{1}{N}\frac{E_s}{N_0}$ can be seen as the average received *snr* at each antenna element before beamforming.

group (g) in spatial domain, and transmitted symbols $x_{n'}^{(g'_{k'})}$s in temporal domain. Considering the mutual independence across multi-path channel vectors (due to the WSSUS model) given by (2), and considering that the transmitted symbols of different users are uncorrelated (including the data transmission period), i.e., $\mathbb{E}\left\{x_n^{(g_k)}\left(x_{n'}^{(g'_{k'})}\right)^H\right\} = \gamma^{(g)}E_s\delta_{nn'}\delta_{gg'}\delta_{kk'}$, the following is obtained: $\mathbb{E}\left\{\boldsymbol{\eta}_n^{(g)}\left(\boldsymbol{\eta}_{n'}^{(g)}\right)^H\right\} = \mathbf{R}_{\boldsymbol{\eta}}^{(g)}\delta_{nn'}$, where

$$\mathbf{R}_{\boldsymbol{\eta}}^{(g)} \triangleq E_s\left(\sum_{g'\neq g}\gamma^{(g')}K_{g'}\sum_{l=0}^{L_{g'}-1}\rho_l^{(g')}\mathbf{R}_l^{(g')}\right) + N_0\mathbf{I}_N, \quad (3)$$

and $\gamma^{(g')}$ for $g' \neq g$ can be regarded as the relative average received power at BS of *inter-group* users normalized with that of the group g users. In (3), $\gamma^{(g')}$s are accountable for the *near-far* effect stemming from the fact that average received signal strength of different UTs may differ significantly depending on their distance to the BS. Moreover, It is important to note that $N \times N$ covariance matrix of the *inter-group interference* $\mathbf{R}_{\boldsymbol{\eta}}^{(g)}$ in (3) consists of all the statistical information of the CSI in spatial domain (i.e., AoA support) for all inner and outer cell users interfering with group g users.

B. Spatio-Temporal Domain Vector Definitions

Before elaborating on the details of the estimation technique, we give the following matrix and vector definitions that will be useful in the subsequent sections. First, the training vector (or convolution vector), comprising of the transmitted pilots for k^{th} user in group g at the nth signaling interval, is defined as $\mathbf{x}_{n,k}^{(g)} \triangleq \left[x_n^{(g_k)}\cdots x_{n-L_g+1}^{(g_k)}\right]^H$. In a similar manner, the extended multi-path channel vector of the kth user, belonging to the intended group g, is given as $\mathbf{f}_{n,k}^{(g)} \triangleq \left[\mathbf{h}_{n,0}^{(g_k)}\cdots\mathbf{h}_{n,L_g-1}^{(g_k)}\right]_{NL_g\times 1}^T$ by concatenating all MPCs of the kth user in group g. Then, by using these defined ones, it will be useful to construct the following vector that represents the concatenated channel vector (including all the related channel parameters of users in group g to be estimated simultaneously):

$$\mathbf{h}_n^{(g)} \triangleq \text{vec}\left\{\left[\begin{array}{cccc}\mathbf{f}_{n,1}^{(g)} & \mathbf{f}_{n,2}^{(g)} & \cdots & \mathbf{f}_{n,K_g}^{(g)}\end{array}\right]_{NL_g\times K_g}\right\}. \quad (4)$$

Finally, the complete training vector that consists of the training data of all users in group g at the nth signaling interval is given by

$$\mathbf{x}_n^{(g)} \triangleq \text{vec}\left\{\left[\begin{array}{cccc}\mathbf{x}_{n,1}^{(g)} & \mathbf{x}_{n,2}^{(g)} & \cdots & \mathbf{x}_{n,K_g}^{(g)}\end{array}\right]_{L_g\times K_g}\right\}. \quad (5)$$

Based on (4) and (5), the spatio-temporal domain signal model is obtained by expressing $\mathbf{y}_n$ in (1) as

$$\mathbf{y}_n = \left(\mathbf{x}_n^{(g)}\otimes\mathbf{I}_N\right)^H\mathbf{h}_n^{(g)} + \boldsymbol{\eta}_n^{(g)} \quad (6)$$

III. A KALMAN FILTER IMPLEMENTATION OF THE REDUCED-RANK MMSE CHANNEL ESTIMATOR

As in the case of the Wiener filter, our goal is to reach a dimension reduction of the estimator. The conventional spatial dimension for a MIMO system with N antennas is N, which can be large, especially for Massive MIMO Our goal is to reduce this dimensionality to D where D can be much smaller depending on the channel condition.

A. Pre-beamforming Stage

The pre-beamforming is applied in order to distinguish *intra-group* signal of group g users from other groups by suppressing the *inter-group interference* while reducing the signaling dimension of $\mathbf{y}_n$ in (6). At the pre-beamforming stage, a D-dimensional vector $\mathbf{y}_n^{(g)}$ can be formed for all *intra-cell* groups by a linear transformation through a matrix $\left(\mathbf{S}_D^{(g)}\right)^H$ as

$$\mathbf{y}_n^{(g)} = \left(\boldsymbol{\Psi}_n^{(g)}\right)^H\mathbf{h}_n^{(g)} + \left(\mathbf{S}_D^{(g)}\right)^H\boldsymbol{\eta}_n^{(g)} \quad (7)$$

where $\boldsymbol{\Psi}_n^{(g)} \triangleq \mathbf{x}_n^{(g)}\otimes\mathbf{S}_D^{(g)}$, and $\boldsymbol{\eta}_n^{(g)}$ is the *inter-group* interference. Here, $\mathbf{S}_D^{(g)}$ can be regarded as $N \times D$ statistical pre-beamforming matrix that projects the N-dimensional received signal samples $\{\mathbf{y}_n\}$ in (1) on a suitable D-dimensional subspace in spatial domain.

B. Reduced Rank Kalman Estimator

For channel variation in time, we adopt a state-space model (7) as a first-order stationary Gauss-Markov process [12]:

$$\mathbf{h}_n^{(g)} = \alpha\mathbf{h}_{n-1}^{(g)} + \sqrt{1-\alpha^2}\mathbf{b}_n^{(g)}. \quad (8)$$

This is an autoregressive first order (AR(1)) model, with $\mathbf{b}_n$ being the input, which is also known as the disturbance. As we do not want the covariance of $\mathbf{h}_n^{(g)}$ change with n, we set

$$\mathbb{E}\left\{\mathbf{h}_n^{(g)}\left(\mathbf{h}_n^{(g)}\right)^H\right\} = \mathbb{E}\left\{\mathbf{b}_n^{(g)}\left(\mathbf{b}_n^{(g)}\right)^H\right\} \triangleq \mathbf{R}_{\mathbf{h}}^{(g)} \quad (9)$$

We focus on minimum mean square error (MMSE) channel estimation based on the current and all previous received training signals. The following MMSE estimate and its corresponding estimation error covariance are given by

$$\hat{\mathbf{h}}_{n|m}^{(g)} \triangleq \mathbb{E}\left\{\mathbf{h}_n^{(g)}|\mathbf{y}_k^{(g)}, \; k = 0, 1, \ldots, m\right\}$$

$$\mathbf{P}_{n|m}^{(g)} \triangleq \mathbb{E}\left\{\left(\mathbf{h}_n^{(g)}-\hat{\mathbf{h}}_{n|m}^{(g)}\right)\left(\mathbf{h}_n^{(g)}-\hat{\mathbf{h}}_{n|m}^{(g)}\right)^H|\mathbf{y}_k^{(g)}, \; k = 0, 1, \ldots, m\right\}$$

$$(10)$$

Based on a standard derivation or a factor graph model, the MMSE estimates and the covariance matrices can be recursively computed based on Kalman filtering as shown in Table I. The Kalman filter description is standard. However, what we are implementing is a reduced-rank Kalman filter, that is to say, the rank of the Kalman gain matrix is significantly smaller than N due to the statistical *pre-beamformer*. To that end, the matrix $\boldsymbol{\Psi}_n^{(g)}$ is important. More specifically, we are seeking the subspace $\mathbf{S}_D^{(g)}$ that will achieve the rank reduction and the associated reduction in complexity.

The new subspace $\mathbf{S}_D^H$ will be different than the Wiener filter formulation earlier [13]. In particular, it will also be time-varying, or evolving in time. In what follows, we will develop this new filter.

Note that the a posteriori and the a priori estimate covari-

TABLE I: Channel Estimation based on Reduced Rank Kalman Filtering after *pre-beamformer* $\mathbf{S}_D^{(g)}$ for user group g, $g = 1, \ldots, G$

Initialization: $\quad \hat{\mathbf{h}}_{0|-1}^{(g)} = \mathbf{0} \quad$ and $\quad \mathbf{P}_{0|-1}^{(g)} = \mathbf{R}_{\mathbf{h}}^{(g)} \qquad (11)$

while $n = 0, 1, \ldots$ **do**

$$\mathbf{\Psi}_n^{(g)} = \mathbf{x}_n^{(g)} \otimes \mathbf{S}_D^{(g)} \qquad (12)$$

Measurement Update:

Innovation: $\mathbf{z}_n^{(g)} = \mathbf{y}_n^{(g)} - \left(\mathbf{\Psi}_n^{(g)}\right)^H \hat{\mathbf{h}}_{n|n-1}^{(g)} \qquad (13)$

Innovation Covariance:
$$\mathbf{E}_n^{(g)} = \left(\mathbf{\Psi}_n^{(g)}\right)^H \mathbf{P}_{n|n-1}^{(g)} \mathbf{\Psi}_n^{(g)} + \left(\mathbf{S}_D^{(g)}\right)^H \mathbf{R}_{\boldsymbol{\eta}}^{(g)} \mathbf{S}_D^{(g)} \qquad (14)$$

Kalman Gain: $\mathbf{K}_n^{(g)} = \mathbf{P}_{n|n-1}^{(g)} \mathbf{\Psi}_n^{(g)} \left(\mathbf{E}_n^{(g)}\right)^{-1} \qquad (15)$

A Posteriori State Estimate: $\hat{\mathbf{h}}_{n|n}^{(g)} = \hat{\mathbf{h}}_{n|n-1}^{(g)} + \mathbf{K}_n^{(g)} \mathbf{z}_n^{(g)}$

(16)

A Posteriori Estimate Covariance:
$$\mathbf{P}_{n|n}^{(g)} = \left[\mathbf{I} - \mathbf{K}_n^{(g)} \left(\mathbf{\Psi}_n^{(g)}\right)^H\right] \mathbf{P}_{n|n-1}^{(g)} \qquad (17)$$

Prediction:

A Priori State Estimate: $\hat{\mathbf{h}}_{n+1|n}^{(g)} = \alpha \hat{\mathbf{h}}_{n|n}^{(g)} \qquad (18)$

A Priori Estimate Covariance:
$$\mathbf{P}_{n+1|n}^{(g)} = \alpha^2 \mathbf{P}_{n|n}^{(g)} + (1-\alpha^2) \mathbf{R}_{\mathbf{h}}^{(g)} \qquad (19)$$

end while

---

ances can be written as

$$\mathbf{P}_{n|n}^{(g)} = \mathbf{P}_{n|n-1}^{(g)} - \mathbf{P}_{n|n-1}^{(g)} \mathbf{\Psi}_n^{(g)} \left[\left(\mathbf{\Psi}_n^{(g)}\right)^H \mathbf{P}_{n|n-1}^{(g)} \mathbf{\Psi}_n^{(g)} \right.$$
$$\left. + \left(\mathbf{S}_D^{(g)}\right)^H \mathbf{R}_{\boldsymbol{\eta}}^{(g)} \mathbf{S}_D^{(g)}\right]^{-1} \left(\mathbf{\Psi}_n^{(g)}\right)^H \mathbf{P}_{n|n-1}^{(g)} \qquad (20)$$

$$\mathbf{P}_{n|n-1}^{(g)} = \alpha^2 \mathbf{P}_{n-1|n-1}^{(g)} + \left(1-\alpha^2\right) \mathbf{R}_{\mathbf{h}}^{(g)}, \quad \mathbf{P}_{0|-1}^{(g)} = \mathbf{R}_{\mathbf{h}}^{(g)} \qquad (21)$$

where we can express $\mathbf{R}_{\mathbf{h}}^{(g)}$ in (9) by using (2) and (4) in the following form

$$\mathbf{R}_{\mathbf{h}}^{(g)} = \sum_{l=0}^{L_g-1} \mathbf{I}_{K_g} \otimes \mathbf{E}_{L_g,l} \otimes \rho_l^{(g)} \mathbf{R}_l^{(g)}, \qquad (22)$$

where $\mathbf{E}_{L_g,l}$ is an $L_g \times L_g$ elementary diagonal matrix where all the entries except the $(l+1)^{th}$ diagonal one are zero. By applying the matrix inversion lemma on (20), we get

$$\left(\mathbf{P}_{n|n}^{(g)}\right)^{-1} = \left(\mathbf{P}_{n|n-1}^{(g)}\right)^{-1} + \mathbf{\Psi}_n^{(g)} \left[\left(\mathbf{S}_D^{(g)}\right)^H \mathbf{R}_{\boldsymbol{\eta}} \mathbf{S}_D^{(g)}\right]^{-1} \left(\mathbf{\Psi}_n^{(g)}\right)^H.$$
$$(23)$$

C. Sequential Beam Design

In this section, our goal is to find a good subspace (spanned by the columns of $\mathbf{S}_D^{(g)}$ matrix) on which the reduced dimensional instantaneous Kalman estimator can be realized as accurately as possible, so that a minimal performance compromise in the subsequent statistical signal processing operations after beamforming is provided. Different criteria can be used to design beam pattern as similar to one in [13]. In

this paper, we adopt error volume as the optimization criterion, namely, the minimization of $\det\left(\mathbf{P}_{n|n}^{(g)}\right)$, and $\mathbf{S}_D^{(g)}$ is designed to minimize the error volume at each signaling interval for a given prediction error covariance $\mathbf{P}_{n|n-1}^{(g)}$ that is a function of all previous training signals.

We consider to construct $\mathbf{S}_D^{(g)}$ optimally at beginning of each M consecutive signaling interval where the symbol time $n = mM + u$ for $m = 0, 1, \ldots$, and $0 \leq u \leq M - 1$. Here, M can be considered as the consecutive channel transmissions composed of a training period followed by a data transmission period as a block, and the *pre-beamformer* is supposed to be updated at each of this interval. Assuming that the channel is almost stationary for M consecutive symbol interval, one get the following equations in (25) and (26) together with (24)

$$\mathbf{P}_{n+M|n}^{(g)} = \alpha^{2M} \mathbf{P}_{n|n}^{(g)} + \left(1 - \alpha^{2M}\right) \mathbf{R}_h^{(g)} \qquad (24)$$

where $\left(1 - \alpha^{2M}\right)$ should be approximately equal to 0 so that stationarity is ensured.

The eqn. (25) results in

$$\det\left(\mathbf{P}_{n|n}^{(g)}\right) =$$

$$\frac{\det\left(\mathbf{P}_{n|n-M}^{(g)}\right)}{\det\left\{\mathbf{I} + \left(\mathbf{I}_M \otimes \left[\left(\mathbf{S}_D^{(g)}\right)^H \mathbf{R}_{\boldsymbol{\eta}}^{(g)} \mathbf{S}_D^{(g)}\right]^{-1}\right) \left(\left(\mathbf{F}_n^{(g)}\right)^H \mathbf{P}_{n|n-M}^{(g)} \mathbf{F}_n^{(g)}\right)\right\}}.$$
$$(27)$$

where $\mathbf{F}_n^{(g)} \triangleq \mathbf{X}_n^{(g)} \otimes \mathbf{S}_D^{(g)}$. It is difficult to use (27) to find the optimal $\mathbf{S}_D^{(g)}$ since it depends on the training data. We will use an approximation of (26) to get to a solution. In an approximation similar to that performed to reach the stochastic gradient in the least mean square (LMS) algorithm, we will replace $\mathbf{X}_n^{(g)} \left(\mathbf{X}_n^{(g)}\right)^H$ with its expected value $\mathbb{E}\left\{\mathbf{X}_n^{(g)} \left(\mathbf{X}_n^{(g)}\right)^H\right\} = (E_s M) \mathbf{I}_{K_g L_g}$ to yield a block diagonal $\mathbf{P}_{n|n}^{(g)}$:

$$\mathbf{P}_{n|n}^{(g)} = \left[\left(\mathbf{P}_{n|n-M}^{(g)}\right)^{-1} + (E_s M) \left(\mathbf{I}_{K_g L_g} \otimes \mathbf{S}_D^{(g)}\right)\right.$$
$$\left.\left(\mathbf{I}_{K_g L_g} \otimes \left[\left(\mathbf{S}_D^{(g)}\right)^H \mathbf{R}_{\boldsymbol{\eta}}^{(g)} \mathbf{S}_D^{(g)}\right]^{-1}\right) \left(\mathbf{I}_{K_g L_g} \otimes \mathbf{S}_D^{(g)}\right)^H\right]^{-1}$$
$$= \mathbf{P}_{n|n-M}^{(g)} - (E_s M) \mathbf{P}_{n|n-M}^{(g)} \left(\mathbf{I}_{K_g L_g} \otimes \mathbf{S}_D^{(g)}\right)$$
$$\times \left[\mathbf{I}_{K_g L_g} \otimes \left(\mathbf{S}_D^{(g)}\right)^H \mathbf{R}_{\boldsymbol{\eta}}^{(g)} \mathbf{S}_D^{(g)} + (E_s M) \left(\mathbf{I}_{K_g L_g} \otimes \mathbf{S}_D^{(g)}\right)^H\right.$$
$$\left.\mathbf{P}_{n|n-M}^{(g)} \left(\mathbf{I}_{K_g L_g} \otimes \mathbf{S}_D^{(g)}\right)\right]^{-1} \left(\mathbf{I}_{K_g L_g} \otimes \mathbf{S}_D^{(g)}\right)^H \mathbf{P}_{n|n-M}^{(g)}$$
$$(28)$$

Since initially $\mathbf{P}_{0|-1}^{(g)} = \mathbf{R}_{\mathbf{h}}^{(g)}$ has block-diagonal structure given in (22), $\mathbf{P}_{n|n}^{(g)}$ and $\mathbf{P}_{n|n-M}^{(g)}$ can be expressed in the following form:

$$\mathbf{P}_{n|m}^{(g)} = \sum_{l=0}^{L_g-1} \mathbf{I}_{K_g} \otimes \mathbf{E}_{L_g,l} \otimes \mathbf{A}_{n|m}^{l,(g)}. \qquad (29)$$

$$\left(\mathbf{P}_{n|n}^{(g)}\right)^{-1} = \left(\mathbf{P}_{n|n-M}^{(g)}\right)^{-1} + \sum_{m=0}^{M-1} \mathbf{\Psi}_{n-m}^{(g)} \left[\left(\mathbf{S}_D^{(g)}\right)^H \mathbf{R}_{\boldsymbol{\eta}}^{(g)} \mathbf{S}_D^{(g)}\right]^{-1} \left(\mathbf{\Psi}_{n-m}^{(g)}\right)^H$$

$$= \left(\mathbf{P}_{n|n-M}^{(g)}\right)^{-1} + \Big(\underbrace{\left[\mathbf{x}_n^{(g)} \mathbf{x}_{n-1}^{(g)} \ldots \mathbf{x}_{n-M+1}^{(g)}\right]}_{\triangleq \, \mathbf{X}_n^{(g)}} \otimes \mathbf{S}_D^{(g)}\Big) \left(\mathbf{I}_M \otimes \left[\left(\mathbf{S}_D^{(g)}\right)^H \mathbf{R}_{\boldsymbol{\eta}}^{(g)} \mathbf{S}_D^{(g)}\right]^{-1}\right) \left(\left[\mathbf{x}_n^{(g)} \mathbf{x}_{n-1}^{(g)} \ldots \mathbf{x}_{n-M+1}^{(g)}\right] \otimes \mathbf{S}_D^{(g)}\right)^H \quad (25)$$

$$= \left(\mathbf{P}_{n|n-M}^{(g)}\right)^{-1} + \left(\mathbf{X}_n^{(g)} \left(\mathbf{X}_n^{(g)}\right)^H\right) \otimes \left(\mathbf{S}_D^{(g)} \left[\left(\mathbf{S}_D^{(g)}\right)^H \mathbf{R}_{\boldsymbol{\eta}}^{(g)} \mathbf{S}_D^{(g)}\right]^{-1} \left(\mathbf{S}_D^{(g)}\right)^H\right) \qquad (26)$$

In order to construct the nearly optimal $\mathbf{S}_D^{(g)}$, first the following matrices in (30) are constructed. Then, by using (28), (29) and (30), we get the following recursive relations in (32), which starts with (31)

$$\mathbf{SNR}_{0|-m}^{l,(g)} = \left[\left(\mathbf{S}_D^{(g)}\right)^H \mathbf{R}_{\boldsymbol{\eta}}^{(g)} \mathbf{S}_D^{(g)}\right]^{-1} \left[\rho_l^{(g)} \left(\mathbf{S}_D^{(g)}\right)^H \mathbf{R}_l^{(g)} \mathbf{S}_D^{(g)}\right] \tag{31}$$

for $m = 1, \ldots, M$. Finally, the following error volume expression is obtained after using (28) and the matrix inversion lemma:

$$\det\left(\mathbf{P}_{n|n}^{(g)}\right) = \frac{\det\left(\mathbf{P}_{n|n-M}^{(g)}\right)}{\det\left(\mathbf{I} + (E_s M) \sum_{l=0}^{L_g-1} \mathbf{I}_{K_g} \otimes \mathbf{E}_{L_g,l} \otimes \mathbf{SNR}_{n|n-M}^{l,(g)}\right)}. \tag{33}$$

Proposition: Given all previous training signals, the pre-beamformer $\mathbf{S}_D^{(g)}$ at n^{th} signaling interval minimizing $\det\left(\mathbf{P}_{n|n}^{(g)}\right)$ in (33) is given by a proper scaled version of the D dominant *generalized eigenvectors* of $\mathbf{R}_l^{(g)}$ and $\mathbf{R}_{\boldsymbol{\eta}}^{(g)}$ [13]. Under this pre-beamformer design, all three matrices $\mathbf{SNR}_{n|n}^{l,(g)}$, $\mathbf{SNR}_{n|n-M}^{l,(g)}$ and $\mathbf{SNR}_{0|-m}^{l,(g)}$ are simultaneously diagonalizable, i.e., given these *generalized eigenvectors* that diagonalize $\mathbf{SNR}_{0|-m}^{l,(g)}$ initially, the $\mathbf{SNR}_{n|n}^{l,(g)}$ and $\mathbf{SNR}_{n|n-M}^{l,(g)}$ are also diagonalizable for every n from the recursive relation in (32). This sequential beamspace construction algorithm is summarized in Algorithm 1.

IV. NUMERICAL RESULTS AND DISCUSSION

In this section, we provide some numerical results to evaluate the performance of the reduced rank Kalman Estimator for the proposed pattern design of the *pre-beamformer*. Throughout the demonstrations, we consider a massive MIMO system with uplink training in TDD mode where a BS is equipped with a uniform linear array (ULA) of $N = 100$ antenna elements along the y-axis[4], and each of K users has a single receive antenna.

In the studied scenario, K users were clustered into eight groups ($G = 8$), and each UT is assumed to be located at a specific azimuth angle θ along the ring centered at the origin in the x-y plane. The channel covariance matrix of each group is specified with the center azimuth angle θ (AoA), and can

[4]Although the proposed estimators are valid for an arbitrary array structure in this paper, ULA is considered for the ease of exposition only.

Algorithm 1 Sequential Pre-beamformer Construction

Require: Obtain the *generalized eigenvectors* of $\mathbf{R}_l^{(g)}$ and $\mathbf{R}_{\boldsymbol{\eta}}^{(g)}$ by solving $\rho_l^{(g)} \mathbf{R}_l^{(g)} \mathbf{V}^l = \mathbf{R}_{\boldsymbol{\eta}}^{(g)} \mathbf{V}^l \mathbf{\Lambda}^l$, $l = 0, \ldots, L_g - 1$,
$\mathbf{\Lambda}_0^l \leftarrow \mathbf{\Lambda}^l$,
while $m = 0, 1, 2, \ldots$ **do**
The pre-beamformer is in the form [13]:

$$\mathbf{S}_D^{(g)} \triangleq \left[\begin{array}{cccc} \mathbf{S}_D^{(g)}(0) & \mathbf{S}_D^{(g)}(1) & \cdots & \mathbf{S}_D^{(g)}(L_g - 1) \end{array}\right]_{N \times D} \tag{34}$$

$\mathbf{SNR}_{0|-M}^{l,(g)}$ is diagonalized with $N \times d_l$ $\mathbf{S}_D^{(g)}(l)$ constructed by the columns of $\mathbf{V}^l$ corresponding to the d_l dominant *generalized eigenvalues* for $\sum_{l=0}^{L_g-1} d_l = D$.
In order to find the set of column indices of $\mathbf{V}^l$ showing the d_l largest eigenvalues at each signaling interval m, namely, $\{\mathcal{I}_m^l\}$, solve the following

$$\{\mathcal{I}_m^l\} = \underset{\{\mathcal{I}'_m^l\}_{l=0}^{L_g-1}}{\operatorname{argmax}} \underbrace{\det\left(\mathbf{I} + (E_s M) \sum_{l=0}^{L_g-1} \mathbf{I}_{K_g} \otimes \mathbf{E}_{L_g,l} \otimes \mathbf{\Lambda}_{\{\mathcal{I}'_m^l\}}^l\right)}_{\mathfrak{F}\left(\{\mathcal{I}'_m^l\}_{l=0}^{L_g-1}\right)} \tag{35}$$

where $|\{\mathcal{I}_m^l\}| = d_l$, and

$$\max \; \mathfrak{F}\left(\{\mathcal{I}_m^l\}_{l=0}^{L_g-1}\right) = \prod_{l=0}^{L_g-1} \prod_{i=1}^{d_l} \left(1 + (E_s M)\lambda_i^l\right)^{K_g} \tag{36}$$

by assuming that $\mathbf{SNR}_{n|n-M}^{l,(g)}$ have nearly orthogonal eigenspaces for different l, and λ_i^l's are the diagonal entries of $\mathbf{\Lambda}^l$ whose indices are given by the set $\{\mathcal{I}_m^l\}$.
Update:

$$\mathbf{S}_D^{(g)}(l) \leftarrow \mathbf{V}_{\{\mathcal{I}_m^l\}}^l, \; l = 0, \ldots, L_g - 1 \tag{37}$$

$$\mathbf{\Lambda}_{\{\mathcal{I}_m^l\}}^l \leftarrow \mathbf{\Lambda}_{\{\mathcal{I}_m^l\}}^l - (E_s M) \left(\mathbf{\Lambda}_{\{\mathcal{I}_m^l\}}^l\right)^2 . \Big/ \left(\mathbf{I}_D + (E_s M)\mathbf{\Lambda}_{\{\mathcal{I}_m^l\}}^l\right)$$

$$\mathbf{\Lambda}^l \leftarrow \alpha^{2M} \mathbf{\Lambda}^l + \left(1 - \alpha^{2M}\right) \mathbf{\Lambda}_0^l. \tag{38}$$

end while Here, $. \Big/$ denotes the elementwise division and $\mathbf{\Lambda}_{\{\mathcal{I}_m^l\}}^l$ shows the diagonal elements of $\mathbf{\Lambda}^l$ whose indices given by the set $\{\mathcal{I}_m^l\}$ at the mth signaling interval.

$$\left[\mathbf{I}_{K_g L_g} \otimes \left(\mathbf{S}_D^{(g)}\right)^H \mathbf{R}_{\boldsymbol{\eta}}^{(g)} \mathbf{S}_D^{(g)}\right]^{-1} \left[\left(\mathbf{I}_{K_g L_g} \otimes \mathbf{S}_D^{(g)}\right)^H \mathbf{P}_{n|m}^{(g)} \left(\mathbf{I}_{K_g L_g} \otimes \mathbf{S}_D^{(g)}\right)\right] = \sum_{l=0}^{L_g-1} \mathbf{I}_{K_g} \otimes \mathbf{E}_{L_g,l} \otimes \mathbf{SNR}_{n|m}^{l,(g)}, \quad \text{where we defined}$$

$$\mathbf{SNR}_{n|m}^{l,(g)} \triangleq \left[\left(\mathbf{S}_D^{(g)}\right)^H \mathbf{R}_{\boldsymbol{\eta}}^{(g)} \mathbf{S}_D^{(g)}\right]^{-1} \left[\left(\mathbf{S}_D^{(g)}\right)^H \mathbf{A}_{n|m}^{l,(g)} \mathbf{S}_D^{(g)}\right] \qquad (30)$$

$$\mathbf{SNR}_{n|n}^{l,(g)} = \mathbf{SNR}_{n|n-M}^{l,(g)} - (E_s M)\mathbf{SNR}_{n|n-M}^{l,(g)} \left[\mathbf{I}_D + (E_s M)\mathbf{SNR}_{n|n-M}^{l,(g)}\right]^{-1} \mathbf{SNR}_{n|n-M}^{l,(g)}$$

$$\mathbf{SNR}_{n+M|n}^{l,(g)} = \alpha^{2M} \mathbf{SNR}_{n|n}^{l,(g)} + \left(1 - \alpha^{2M}\right) \mathbf{SNR}_{0|-M}^{l,(g)}, \quad l = 0, \dots, L_g - 1. \qquad (32)$$

be calculated in a similar way to the ones in [4], [6]. In the simulations, our focus is on the channel estimation accuracy of the intended group g with 3 MPCs, i.e., $L_g = 3$. The first two MPCs of group g stem from a azimuth angular sector $[-1°, 1°]$ for delays at $l = 0, 1$, and the angular sector of the last MPC at $l = 2$ of g is given as $[5°, 7°]$ in azimuth. We assume two users served simultaneously for group g, i.e., $K_g = 2$. Each of the other 7 groups (interfering with the intended one) consists of three users, i.e., $K_{g'} = 3$, $g' \neq g$ and these users have 3 MPCs whose angular sectors have same supports of AoA ($L_{g'} = 3$, $g' \neq g$) given by $[-29, -26]$, $[-21, -19]$, $[-12, -9]$, $[-5.5, -3.5]$, $[9.5, 12.5]$, $[15, 17]$, $[24, 27]$ in azimuth respectively. The noise power is set as $N_0 = 1$ so that all dB power values are relative to 1. *Intra-group* users (of the intended group) use non-orthogonal training waveforms composed of T chips, and these are obtained by truncating length-63 *Kasami* codes by simply choosing the first T chips (training length) of last K_g *Kasami* sequences without any optimization.

The trace of the estimation error covariance matrix given by (17) for the extended channel vector of group g users in (4) to compare the performance of RR Kalman estimators based on different pre-beamformers. The covariance matrix of the *inter-group interference* is evaluated by (3) when the angular sector of each group is provided.

In this paper, we compare the performance of dimension reduction based on Generalized Eigendecomposition in Algorithm 1 (shown to be nearly optimal under some realistic assumptions) with that of the conventional subspace composed of the first D dominant eigenvectors of $\sum_{l=0}^{L_g-1} \rho_l^{(g)} \mathbf{R}_l^{(g)}$ in (2). We call this conventional beamspace as discrete Fourier transform (DFT) beamspace because the eigenvectors of the spatial correlation matrix of the ULA channel are well-approximated by the columns of the $N \times N$ unitary DFT matrix whose indices correspond to the support of the Fourier transform of the spatial correlation function (owing to the Szegö's asymptotic theory) [5] depending on the angular sector of group g UTs. This conventional beamspace is known to be information preserving for spatially white interference case, and thus, widely used in practical *hybrid beamforming* applications, where the beamforming in RF analog domain can be implemented by simple phase shifters [14].

In Figure 1, the beam patterns created by the generalized eigenvector beamspace (GEB) and DFT beamspaces are de-

picted for $D = 6$ at $snr = 30$ (dB). The GEB is designed based on the AoA support of the intended group g for $l = 0, 1, 2$ while taking the angular locations of the interfering groups into account. The *inter-group* users signal are assumed to have same power level with that of the intended group. As can be seen from the figure, the GEB tries to create deep nulls at the angular locations of interfering UTs, whereas the conventional pre-beamformer only tries to maximize the captured power of the intended group MPCs for a given dimension. It is expected that as the number of BS antennas increases, the eigenspaces of each group are approximately orthogonal. However, the number of transmit antennas is finite in practice, and there always exists some overlap among the virtual angular sectors of each group which leads to a leakage to the intended group signal. Therefore, as it will be shown later, the accuracy of the channel estimation realized on the reduced dimensional subspace, spanned by the conventional DFT beamspace, is considerably lost due to the residual *inter-group interference* after pre-beamforming. On the other hand, the GEB suppresses the *inter-group interference* while allowing the MPCs of the intended group to pass with a negligible distortion.
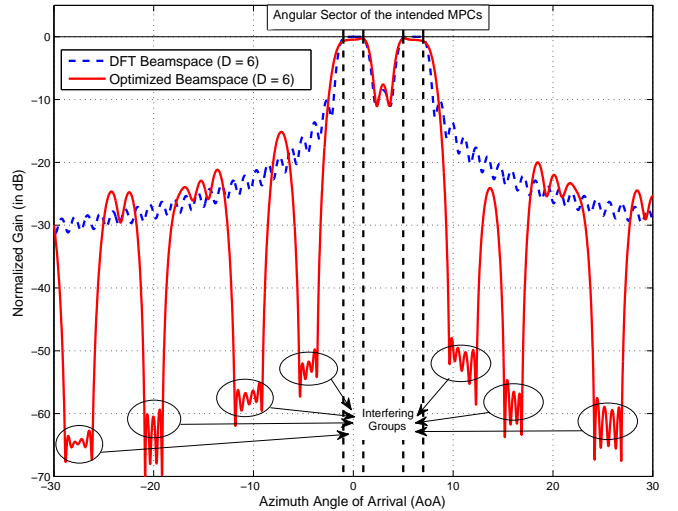

Fig. 1: Beam pattern of different pre-beamformers

In Figure 2, the average mean square error (MSE) values given by $\text{Tr}\left\{\mathbf{P}_{n|n}^{(g)}\right\}/K_g$ as a function of the dimension of the spatial domain pre-beamformer (D) are depicted for both fixed and time-varying pre-beamformer cases. In fixed GEB

case, the pre-beamformer is optimized based on the initial uncertainties, i.e., initial channel covariance, and kept fixed during whole transmission period while realizing RR Kalman estimator. Whereas, in time-varying beamspace construction, the proposed sequential design algorithm (Algorithm 1) is used. This algorithm, taking the varying Kalman prediction errors in each eigen directions into account so as to reduce error volume, results in much better dimension reduction capability, i.e., for a given dimension, time-varying beamspace construction yields much lower channel estimation errors compared to fixed design alternatives.
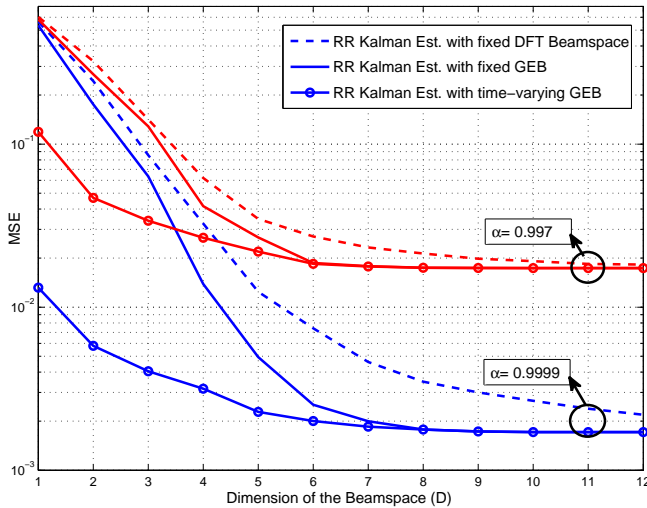

Fig. 2: MSE vs dimension for time-varying channels after 50 training symbol transmission at *snr* = 30 dB, *M* = 5

In Figure 3, the average mean square error (MSE) values as a function of training length are demonstrated for different beamspace dimensions. Fixed statistical pre-beamformer case show much inferior performance to time-varying case even with the use of GEB. Even though the convergence is slower for lower dimensional pre-beamformers in sequential pre-beamformer design, it achieves a much lower steady state estimation error.

V. CONCLUSIONS

In two-stage beamforming, a new algorithm for statistical pre-beamformer design together with Kalman filtering based channel estimator is proposed for massive MIMO transmission employing SC in frequency-selective fading. The proposed algorithm yields a nearly optimal pre-beamformer whose beam pattern is designed sequentially with low complexity by taking the user-grouping into account, and exploiting the properties of Kalman filtering and associated prediction error covariance matrices. The resulting design, based on the second order statistical properties of the channel, generates much lower mean square error values compared to conventional beamforming techniques.
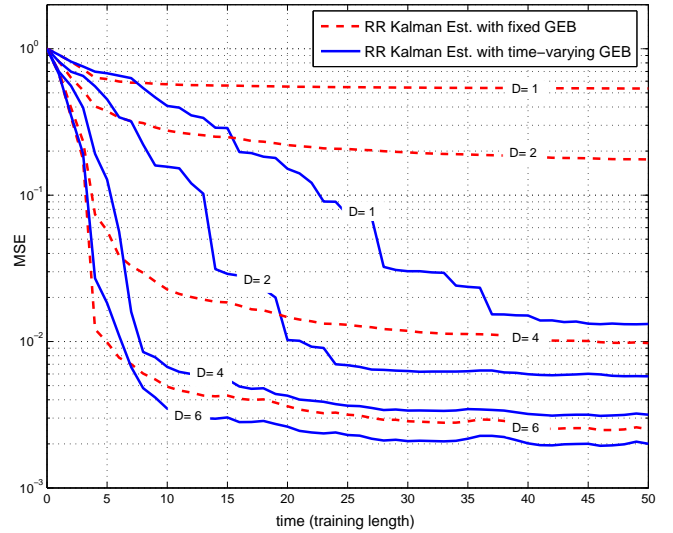
Fig. 3: MSE vs training length for time-varying channel with $\alpha = 0.9999$ at *snr* = 30 dB, *M* = 5

REFERENCES

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?," *IEEE J. Sel. Areas Commun.*, vol. 32, pp. 1065–1082, Jun. 2014.
[2] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Areas Commun.*, vol. 8, pp. 742–758, Oct. 2014.
[3] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 2640–2651, Aug. 2011.
[4] A. Adhikary, J. Nam, J. Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, pp. 6441–6463, Oct. 2013.
[5] J. Nam, A. Adhikary, J. Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, pp. 876–890, Oct. 2014.
[6] D. Kim, G. Lee, and Y. Sung, "Two-stage beamformer design for massive MIMO downlink by trace quotient formulation," *IEEE Trans. Commun.*, vol. 63, pp. 2200–2211, Jun. 2015.
[7] A. Adhikary, E. A. Safadi, M. K. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, "Joint spatial division and multiplexing for mm-wave channels," *IEEE J. Sel. Areas Commun.*, vol. 32, pp. 1239–1255, Jun. 2014.
[8] L. You, X. Gao, A. L. Swindlehurst, and W. Zhong, "Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots," *IEEE Trans. Signal Process.*, vol. 64, pp. 1461–1476, Mar. 2016.
[9] L. You, X. Gao, X. G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 3352–3366, Jun. 2015.
[10] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, May 2013.
[11] F. A. Dietrich and W. Utschick, "Pilot-assisted channel estimation based on second-order statistics," *IEEE Trans. Signal Process.*, vol. 53, pp. 1178–1193, Mar. 2005.
[12] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive mimo systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787–801, 2014.
[13] G. M. Guvensen and E. Ayanoglu, "A generalized framework on beamformer design and CSI acquisition for single-carrier massive MIMO systems in millimeter wave channels," *arXiv preprint arXiv:1607.01436*, 2016.
[14] A. Liu and V. Lau, "Phase only RF precoding for massive MIMO systems with limited RF chains," *IEEE Trans. Signal Process.*, vol. 62, pp. 4505–4515, Sep. 2014.