Scalable Non-Orthogonal Pilot Design for Massive MIMO Systems with Massive Connectivity

Ruichen Deng, Sheng Zhou, and Zhisheng Niu
Tsinghua National Laboratory for Information Science and Technology
Department of Electronic Engineering, Tsinghua University
Beijing 100084, China
Email: drc13@mails.tsinghua.edu.cn, {sheng.zhou,niuzhs}@tsinghua.edu.cn

***Abstract*—With the growing need for Internet of things (IoT) applications, wireless communications are facing the challenge of providing massive connectivity. However, as the number of users increases, the capacity of massive MIMO systems will be severely affected due to the huge channel estimation overhead. Therefore it is necessary to introduce non-orthogonal pilots to massive MIMO systems in order to ease the burden on the resources for channel estimation. In this paper, we propose a KKT-based iterative non-orthogonal pilot design algorithm which maximizes the mutual information between the received signal and the channel for MMSE channel estimation. For the typical antenna deployment of uniform linear array (ULA), we also provide scalable estimation schemes with low complexity based on channel angular representation as alternatives. The simulation results show that our schemes significantly save the estimation overhead at the cost of slight decrease in the MSE performance of the channel estimation, and the saving is even more prominent when the system accommodates more users.**

I. Introduction

Massive multiple-input multiple-output (MIMO) has established an important role in the development of next generation cellular networks for its enormous enhancements in spectral efficiency and promising capacity to support the explosive growth of user traffic [1]. Besides traditional wireless applications such as voice calling, online videos and emails, machine to machine (M2M) communications are getting widely spread with the vigorous development of the Internet of Things (IoT) [2]. The substantial increase in the communication devices requires the networks to have the ability to handle massive connectivity.

On the other hand, the power consumption of the cellular system has increased to a tremendous amount and will continue growing with the densification of networks. To allow for efficient BS sleeping and load balancing mechanisms from the view point of the whole system, a novel framework of hyper-cellular system has been proposed, which separates the control signaling and data service by deploying two types of BSs, namely the control BS (CBS) and the traffic BSs (TBS) [3]. The CBS's function to provide large area coverage for mobile users also requires the system possessed with massive connectivity.

For uncorrelated MIMO channels, ref. [4] has shown that the optimal training pilots for different users are orthogonal vectors. The system should allocate timeslots with length no less than the number of users in training process to acquire channel state information (CSI). As a result, the overhead of channel estimation using orthogonal pilots increases with the growing number of users. In systems with massive connectivity, the estimation overhead may be too expensive for the limited channel coherence time to afford. Luckily, channel correlations exist in many practical situations due to not so rich scattering conditions, which provides the opportunity to design non-orthogonal pilots and reduce the overhead.

Recently a number of works have investigated the CSI acquisition issue in the case of correlated MIMO channels. Ref. [5] exploits the sparsity of massive MIMO channels in poor scattering environment and proposes a compressive sensing approach, but the performance relies highly on the sparsity of channels. Ref. [6] proposes the joint spatial division and multiplexing (JSDM) scheme, which divides users with similar channel correlation matrices into groups and then conducts a two-stage precoding, but the inter-group interference is the bottleneck of the scheme. The authors in [7] apply the MMSE technique to reduce inter-cell pilot contamination, and illustrate the fact that users with non-overlapping angular spreads will hardly interfere with each other. The work is extended to intra-cell case by [8], in which a greedy pilot reuse scheme is designed.

Different from [7], [8], we consider more general scenarios where the MIMO channels are composed of multiple clusters of multi-path components (MPCs). We focus on the design of scalable pilot schemes for massive MIMO systems with massive connectivity and exploit user diversity to further reduce the training overhead. In the paper, we propose a mutual-information-optimal iterative pilot design algorithm for MMSE estimation. Furthermore, we provide efficient channel estimation schemes based on angular representation for the case of uniform linear array (ULA).

The rest of the paper is organized as follows. Section II describes the signal and channel models adopted in this paper. The optimization of non-orthogonal pilots for MMSE channel estimation is discussed in Section III. Then we demonstrate heuristic estimation schemes for ULA case in Section IV. The simulation results are presented in Section V. Finally, we draw our conclusions in Section VI.

II. Signal and Channel Models

We consider a MU-MIMO system operating in TDD mode with a BS and K single-antenna users. The BS is equipped with an antenna array of M elements. We consider users located on the one side of the antenna array to avoid mirroring ambiguity. The channel reciprocity is exploited, so that we only need to estimate uplink CSI.

In the training procedure with τ timeslots, User k sends a pilot sequence $\boldsymbol{s}_k = [s_{k1}, s_{k2}, \cdots, s_{k\tau}]^T$ to the BS. All users' pilot sequences forms the pilot matrix

$$\boldsymbol{S} = [\boldsymbol{s}_1 \ \boldsymbol{s}_2 \cdots \boldsymbol{s}_K].$$

As a result, the received signal of the BS is

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{S}^T + \boldsymbol{N}, \tag{1}$$

where $\boldsymbol{Y} = [\boldsymbol{y}_1 \ \boldsymbol{y}_2 \cdots \boldsymbol{y}_\tau]$ is the received signal vectors, $\boldsymbol{H} = [\boldsymbol{h}_1 \ \boldsymbol{h}_2 \cdots \boldsymbol{h}_K]$ consists of each user's channel vectors and $\boldsymbol{N} = [\boldsymbol{s}_1 \ \boldsymbol{s}_2 \cdots \boldsymbol{s}_\tau]$ is the white additive Gaussian noise (AWGN) with zero-mean and elementwise unit variance.

We extend the channel model in [8] and model User k's channel vector as the sum of L_k clusters of MPCs.

$$\boldsymbol{h}_k = \sum_{l=1}^{L_k} \int_{\mathcal{A}} \alpha_{kl}(\theta)\boldsymbol{v}(\theta)d\theta = \int_{\mathcal{A}} \alpha_k(\theta)\boldsymbol{v}(\theta)d\theta, \tag{2}$$

where $\boldsymbol{v}(\theta)$ is the steering vector of BS array with incidence angle θ, which is constrained in the angular region $\mathcal{A}$, and $\alpha_k(\theta) = \sum_{l=1}^{L_k} \alpha_{kl}(\theta)$ is the total channel gain function of User k.

We assume channel gain function $\alpha_k(\theta)$ is uncorrelated in angular domain. The channel power distribution is modeled as power azimuth spectrum (PAS) [9],

$$PAS_k(\theta) = \mathbb{E}\{|\alpha_k(\theta)|^2\}. \tag{3}$$

Then the channel correlation matrix of User k is given by

$$\boldsymbol{R}_k = \mathbb{E}\{\boldsymbol{h}_k\boldsymbol{h}_k^H\} = \int_{\mathcal{A}} \boldsymbol{v}(\theta)\boldsymbol{v}^H(\theta)PAS_k(\theta)d\theta. \tag{4}$$

When the BS is equipped with ULA with critical antenna spacing (i.e., half the wavelength), the steering vector is denoted by,

$$\boldsymbol{v}(\theta) = \frac{1}{\sqrt{M}}\left[1, e^{-j\pi\sin\theta}, \cdots, e^{-j\pi(M-1)\sin\theta}\right]^T. \tag{5}$$

The angular representation of the received signal is obtained by multiplexing the signal by the inverse descrete Fourier transformation (inverse DFT) matrix (up to some phase shift) [10],

$$\boldsymbol{Y}^a = \boldsymbol{U}^H\boldsymbol{Y}, \tag{6}$$

where the DFT matrix is represented by

$$\begin{aligned}\boldsymbol{U} = [&\boldsymbol{v}(\arcsin(\frac{0}{M}-1)), \boldsymbol{v}(\arcsin(\frac{2}{M}-1), \cdots, \\ &\boldsymbol{v}(\arcsin(\frac{2(M-1)}{M}-1)].\end{aligned} \tag{7}$$

III. Non-orthogonal Pilot Optimization for MMSE estimation

The channel correlation matrix varies much slower compared to the instantaneous CSI. So we assume the channel correlation matrix is known and stays constant over the whole training process.

By vectorizing the matrices in (1), we get the vectorization form of the received signal,

$$\mathrm{vec}(\boldsymbol{Y}) = (\boldsymbol{S} \otimes \boldsymbol{I}_M)\mathrm{vec}(\boldsymbol{H}). \tag{8}$$

The optimal linear estimator to minimize mean square error (MMSE) is

$$\boldsymbol{G}_{opt} = \boldsymbol{R}(\boldsymbol{S}^H \otimes \boldsymbol{I}_M)((\boldsymbol{S}\otimes\boldsymbol{I}_M)\boldsymbol{R}(\boldsymbol{S}^H\otimes\boldsymbol{I}_M)+\boldsymbol{I}_{\boldsymbol{\tau M}})^{-1}, \tag{9}$$

where $\boldsymbol{R}$ is obtained by cascading all the channel spatial correlation matrices in the diagonal.

$$\boldsymbol{R} = \mathrm{diag}\{\boldsymbol{R}_1, \boldsymbol{R}_2, \cdots, \boldsymbol{R}_K\}. \tag{10}$$

The mean square error (MSE) matrix is

$$\boldsymbol{C}_e = (\boldsymbol{R}^\dagger + (\boldsymbol{S}^H\boldsymbol{S}) \otimes \boldsymbol{I}_M)^{-1}, \tag{11}$$

and the mutual information between the channel vector and the received signal conditioned on the training pilots is

$$\begin{aligned}\mathcal{I}(\boldsymbol{H};\boldsymbol{Y}|\boldsymbol{S}) &= \log\det(\boldsymbol{R}) - \log\det(\boldsymbol{C}_e) \\ &= \log\det(\boldsymbol{I} + (\boldsymbol{S}\otimes\boldsymbol{I}_M)\boldsymbol{R}(\boldsymbol{S}^H\otimes\boldsymbol{I}_M))).\end{aligned} \tag{12}$$

To design optimal training pilots, we consider two optimization criteria as in [11], namely minimizing the trace of MSE matrix $tr\{\boldsymbol{C}_e\}$, and maximizing the conditional mutual information $\mathcal{I}(\boldsymbol{H};\boldsymbol{Y}|\boldsymbol{S})$. The two optimization problems are formulated as follows, respectively:

$$\begin{aligned}OP_{MSE}: \min_{\boldsymbol{S}\in\mathcal{C}^{\tau\times K}} \quad & tr\{(\boldsymbol{R}^\dagger + (\boldsymbol{S}^H\boldsymbol{S})\otimes\boldsymbol{I}_M)^{-1}\} \\ s.t. \ & \boldsymbol{s}_k^H\boldsymbol{s}_k \le \beta, k = 1, 2, \cdots, K\end{aligned} \tag{13}$$

and

$$\begin{aligned}OP_{CMI}: \max_{\boldsymbol{S}\in\mathcal{C}^{\tau\times K}} \quad & \log\det(\boldsymbol{I} + (\boldsymbol{S}\otimes\boldsymbol{I}_M)\boldsymbol{R}(\boldsymbol{S}^H\otimes\boldsymbol{I}_M))) \\ s.t. \ & \boldsymbol{s}_k^H\boldsymbol{s}_k \le \beta, k = 1, 2, \cdots, K\end{aligned} \tag{14}$$

where the constraint is per user's power not larger than β.

The case that pilot length equals user number ($\tau = K$) has been investigated in [11]. It is proven that the optimal pilots satisfying either criteria are orthogonal vectors. However, for systems with massive connectivity and limited coherence time, using non-orthogonal pilots with length $\tau < K$ is better choice.

In fact, the two optimization criteria are closely related. Let $\lambda_i, i = 1, 2, \cdots, MK$ be the eigenvalues of the MSE matrix $\boldsymbol{C}_e$. Minimizing the trace of MSE matrix is equivalent to minimizing $\sum_{i=1}^{MK}\lambda_i$, and maximizing the conditional mutual information is equivalent to minimizing $\prod_{i=1}^{MK}\lambda_i$. For simplicity, we focus on the latter criterion, which can be regarded as an upper bounder of the former criterion, since $\sum_{i=1}^{MK}\lambda_i \ge MK \sqrt[MK]{\prod_{i=1}^{MK}\lambda_i}$.

To solve the optimization problem, we derive a KKT-based iterative algorithm inspired by [12]. The Lagrangian function of the optimization problem is

$$L(\boldsymbol{S},\boldsymbol{\mu}) = \log\det(\boldsymbol{I} + (\boldsymbol{S}\otimes\boldsymbol{I}_M)\boldsymbol{R}(\boldsymbol{S}^H\otimes\boldsymbol{I}_M))) + \sum_{k=1}^{K}\mu_k(\boldsymbol{s}_k^H\boldsymbol{s}_k - \beta). \tag{15}$$

It is obvious that the optimal pilots should use the maximum transmit power to fight against noise. Meanwhile, the optimal pilots should satisfy the KKT conditions

$$S_{ij} = \frac{1}{\mu_i}\sum_{m=1}^{M} T_{(i-1)M+m,(j-1)M+m}, \tag{16}$$

where $\boldsymbol{T} = (\boldsymbol{I} + (\boldsymbol{S}\otimes\boldsymbol{I}_M)R(\boldsymbol{S}^H\otimes\boldsymbol{I}_M))^{-1}(\boldsymbol{S}\otimes\boldsymbol{I}_M)R$.

Therefore, we propose the KKT-based iterative pilot design (KIPD) scheme as in Algorithm 1. In the scheme, every iteration step generates a new pilot matrix according to the KKT conditions and the power constraint. The iteration stops until the pilot matrix converges.

Algorithm 1 KKT-based Iterative Pilot Design (KIPD) Scheme

Input: The channel correlation matrix $\boldsymbol{R}$, the power constraint β and the precision parameter ϵ_0

Output: The pilot matrix $\boldsymbol{S}$

1: Initialize the pilot matrix as $\boldsymbol{S}^{(0)}$ and $l=0$.
2: **repeat**
3: $l = l+1$.
4: Generate the new pilot matrix according to (16), $\boldsymbol{T}^{(l)} = (\boldsymbol{I} + (\boldsymbol{S}^{(l-1)}\otimes\boldsymbol{I}_M)R((\boldsymbol{S}^{(l-1)})^H\otimes\boldsymbol{I}_M))^{-1}(\boldsymbol{S}^{(l-1)}\otimes\boldsymbol{I}_M)R$, where $S_{ij}^{(l)} = \sum_{m=1}^{M} T_{(i-1)M+m,(j-1)M+m}^{(l)}$.
5: Normalize the pilots according to the power constraint, $S_{ij}^{(l)} \leftarrow \frac{\beta}{\sum_i (S_{ij}^{(l)})^2} S_{ij}^{(l)}$.
6: **until** $||\boldsymbol{S}^{(l)} - \boldsymbol{S}^{(l-1)}||_2 < \epsilon_0$

IV. Channel Estimation Schemes Based on Angular Representation for ULA

The KIPD scheme provides mutual-information-optimal pilots for MMSE channel estimation, which also has good MSE performance. Unfortunately, we can only guarantee the local optimality of the scheme, and the complexity of the scheme is rather high as a result of the iteration process. Therefore, for the typical antenna deployment of ULA case, we propose heuristic pilot design schemes based on the angular representation of user channels .

A. Truncation of PAS

We start by truncating the PASs of users, which reflect the channel power distribution in the angular domain and can be estimated by a series of techniques of array signal processing such as MUSIC and SAGE [13].

More specifically, we determine a threshold ξ and the corresponding superlevel set of the PAS

$$\mathcal{S}_k = \{\theta | PAS_k(\theta) \geq \xi\}, \tag{17}$$

such that the power loss due to ignoring the spectrum below the threshold is limited to a given parameter η,

$$\frac{\int_{\mathcal{A}} PAS_k(\theta)d\theta - \int_{\mathcal{S}_k} PAS_k(\theta)d\theta}{\int_{\mathcal{A}} PAS_k(\theta)d\theta} = \eta. \tag{18}$$

Thus we obtain a truncated angular spread $\mathcal{S}_k$ for User k, which is naturally a union of bounded intervals (*effective intervals*).

In the theorem below, we show that the non-zero elements of the channel angular representation $\boldsymbol{h}^a$ after the truncation of PAS are also restricted in intervals,

Theorem 1. *As $M \to \infty$, if $PAS(\theta) = 0$ for $\theta \in [\arcsin(\frac{2(m_1-1)}{M} - 1), \arcsin(\frac{2(m_2-1)}{M} - 1)]$, then $h_m^a = 0$ for index $m \in \{m_1, \cdots, m_2\}$, where $1 \leq m_1 < m_2 \leq M$.*

Proof. We know that as $M \to \infty$, the integral in (2) can be approximated by

$$\boldsymbol{h} = \sum_{m=0}^{M-1}(\phi_{m+1} - \phi_m)\alpha(\phi_m)\boldsymbol{v}_m(\phi_m), \tag{19}$$

where

$$\phi_m = \arcsin(\frac{2m}{M} - 1), m = 0, 1, \cdots, M. \tag{20}$$

As a result, the user's channel angular representation is

$$\boldsymbol{h}^a = \boldsymbol{U}^H\boldsymbol{h} = [(\phi_1 - \phi_0)\alpha(\phi_0), \cdots, (\phi_M - \phi_{M-1})\alpha(\phi_{M-1})]^T. \tag{21}$$

If $PAS(\theta) = 0$ for $\theta \in [\phi_{m_1-1}, \phi_{m_2-1}]$, then $\alpha(\phi_m) = 0$ for index $m \in \{m_1 - 1, \cdots, m_2 - 1\}$. Therefore, the corresponding indexes of $\boldsymbol{h}^a$ are also 0 from (21). ☐

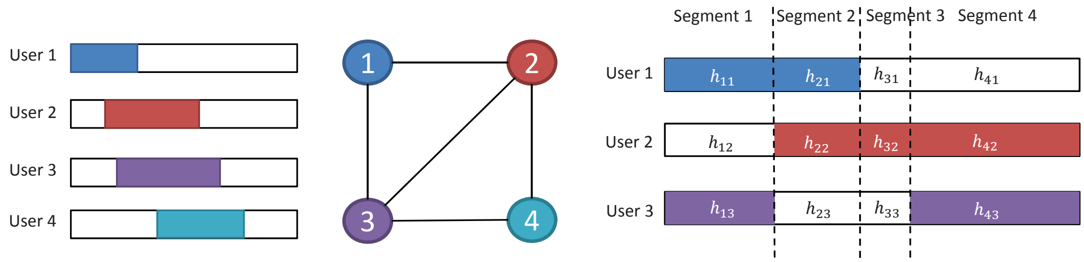
B. Non-orthogonal Pilot Schemes

We discuss non-orthogonal pilot schemes in two different cases, distinguished by the cluster number of channels.

1) Single-Interval Case: If the channel of each user consists of a single cluster of MPCs, the truncated PAS also has a single interval. We construct a user angular spread conflict graph $\mathcal{G}$, which is an interval graph, by using a vertex to represent each user and adding an edge whenever two users have overlapping effective intervals. An example of the conflict graph of 4 users is presented in Fig. 1(a).

We apply the pilot reuse scheme [8] to the single-interval case. The scheme adopts a set of orthogonal pilots and reuses them by assigning the same pilot to users with non-overlapping angular spreads. According to Theorem 1, the mixed receive signals can be separated in the angular domain due to the high spatial resolution in massive MIMO systems.

The pilot allocation scheme is equivalent to the coloring problem of $\mathcal{G}$. The minimum pilot length is the chromatic number of $\mathcal{G}$, i.e., the least color number to color $\mathcal{G}$. Apparently, $\mathcal{G}$ is an interval graph, whose chromatic number equals



(a) (b)

Fig. 1. Examples of truncated PAS. The effective intervals of users are in different colors. (a) Angular spreads of 4 users in the single-interval case and the corresponding conflict graph. (b) Angular spreads of 3 users in the multiple-interval case.

the size of its largest clique. Therefore, the minimum pilot length equals the cardinality the largest conflict user set. For example, the chromatic number and the size of largest clique of the conflict graph of Fig. 1(a) are both 3, so we need at least 3 timeslots to estimate the channels.

The pilot assigning problem can be solved by a greedy coloring algorithm [14], which is provided in Algorithm 2. By greedily coloring angular intervals in sorted order of their left endpoints, the GCPR scheme can find the minimum number of pilots in polynomial time.

Algorithm 2 Greedy Coloring Pilot Reuse (GCPR) Scheme

Input: The angular spreads of users $\mathcal{S}_k, k = 1, 2, \cdots, K$

Output: The pilot assignment $\boldsymbol{c}$

1: Initialization. Set the number of pilots $\tau = 0$, the pilot index set $\mathcal{C} = \emptyset$, the pilot assignment $\boldsymbol{c} = [0, \cdots, 0]_{K\times 1}$, generate conflict graph $\mathcal{G}$ according to the angular spreads of users.

2: Sorting. Arrange the users according to the ascending order of the left endpoints of their angular spreads.

3: Pilot allocation. For $i = 1, 2, \cdots, K$, orderly assign users with the first pilot which is not used by conflicting users. If there exists no such pilot, add a new pilot to $\mathcal{C}$ and assign it to the user.

2) Multiple-Interval Case: In this case, the truncated PAS of multiple-cluster channel is composed of multiple intervals, and the conflict graph $\mathcal{G}$ is a multiple-interval graph. The chromatic number of $\mathcal{G}$ may exceed the size of its largest clique, which means the minimum pilot length of the pilot reuse scheme may be much larger than the cardinality of the largest overlapping user set.

We consider a segment-by-segment channel estimation (SCE) scheme for this case. The whole angular region $\mathcal{A}$ is divided by the endpoints of all users' effective intervals into I segments, $p_i, i = 1, 2, \cdots, I$. Then the angular spread AS_k of each user can be regarded as a subset of all these segments,

$$AS_k = \cup_{i=1}^{I} u_{ik} p_i, \tag{22}$$

where u_{ik} is an indication function that takes value 1 if AS_k contains Segment i, and 0 otherwise.

For each Segment i, the users with angular spread containing the segment form a conflict set $\mathcal{C}_i$,

$$\mathcal{C}_i = \{k | u_{ik} = 1\}, i = 1, 2, \cdots, I. \tag{23}$$

In the training process, the angular representation $\boldsymbol{y}_t^a$ of the received signal in Timeslot t $(t = 1, 2, \cdots, \tau)$ is divided into $\boldsymbol{y}_{it}^a$ by Segment p_i $(i = 1, 2, \cdots, I)$. For the i-th segment, the received signal is the superposition of the signals of users in set $\mathcal{C}_i$,

$$\boldsymbol{Y}_i^a = \sum_{k\in\mathcal{C}_i} \boldsymbol{h}_{ik}^a \boldsymbol{s}_k^T, i = 1, 2, \cdots, I, \tag{24}$$

where $\boldsymbol{Y}_i^a = [\boldsymbol{y}_{i1}^a, \boldsymbol{y}_{i2}^a, \cdots, \boldsymbol{y}_{i\tau}^a]$, and $\boldsymbol{h}_{ik}$ is the i-th segment of User k's channel in angular representation.

Fig. 1(b) depicts an example of the angular spreads of 3 users in the multiple-interval case. The whole angular region is divided into 4 segments. In Segment 1, $\boldsymbol{h}_{12} = \boldsymbol{0}$, therefore User 1 and User 3 form the conflict set $\mathcal{C}_1 = \{1, 3\}$. The conflict sets of Segment 2,3,4 are $\mathcal{C}_2 = \{1, 2\}, \mathcal{C}_3 = \{2\}, \mathcal{C}_4 = \{2, 3\}$, respectively.

As long as the pilots of users in $\mathcal{C}_i$ are linearly independent with each other, the i-th segment of the channels can be estimated by a simple transformation of (24),

$$\mathrm{cat}_{k\in\mathcal{C}_i}[\boldsymbol{h}_{ik}^a] = (\mathrm{cat}_{k\in\mathcal{C}_i}[\boldsymbol{s}_k^T])^{\dagger} \boldsymbol{Y}_i^a, \tag{25}$$

where $\mathrm{cat}[\cdot]$ denotes the concatenation of column vectors.

Finally, the channel is recovered by joining all the segments together,

$$\boldsymbol{h}_k = \boldsymbol{U} \begin{bmatrix} \boldsymbol{h}_{1k} \\ \vdots \\ \boldsymbol{h}_{Ik} \end{bmatrix}. \tag{26}$$

In the SCE scheme, we have the following theorem for the minimum pilot length,

Theorem 2. *The minimum pilot length for the proposed channel estimation scheme equals the cardinality of the largest user conflict set,*

$$\tau_{min} = \max_i |\mathcal{C}_i|. \tag{27}$$

Proof. We let $P = \max_i |\mathcal{C}_i|$. Firstly, the pilot length τ should be no less than P. Otherwise, for the largest user conflict set, we can not find P linearly independent pilot vectors in τ-dimension ($\tau < P$) space.

Then we show there exists a $P \times K$ pilot matrix $\boldsymbol{S}$ satisfying the condition that any P columns of the matrix are linearly independent.

We choose the pilot matrix $\boldsymbol{S}$ to be

$$S_{tk} = e^{-j\frac{2\pi}{K}(t-1)(k-1)}, t \in 1, 2, \cdots, P, k \in 1, 2, \cdots, K. \tag{28}$$

Thus any P columns $\boldsymbol{s}_{c_1}, \boldsymbol{s}_{c_2}, \cdots, \boldsymbol{s}_{c_P}$ of the proposed matrix $\boldsymbol{S}$ form a Vandermonde matrix

$$\boldsymbol{V} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ e^{-j\frac{2\pi}{K}(P-1)(c_1-1)} & \cdots & e^{-j\frac{2\pi}{K}(P-1)(c_P-1)} \end{bmatrix} \tag{29}$$

with a determinant

$$\det V = \prod_{1 \le l < m \le P} (e^{-j\frac{2\pi}{K}(c_m-1)} - e^{-j\frac{2\pi}{K}(c_l-1)}) \neq 0. \tag{30}$$

So the condition is satisfied. We can estimate the channel in each segment by using the pilot matrix $\boldsymbol{S}$. □

The proof actually provides a pilot design scheme, i.e., choosing the first P rows of the $K \times K$ DFT matrix. Note that the estimation scheme does not utilize the precise second-order channel information of users (e.g. the channel correlation matrices). It only uses 1-bit information by applying a threshold ξ to the PASs of users, which makes the estimation scheme more robust to the error of the second-order channel information.

We also notice that the GCPR scheme and the SCE scheme have the same expression for the minimum pilot length, since they both estimate channels in the angular domain. In fact, we can apply the SCE scheme to the single-interval case and will get the same performance.

V. Simulation Results

In this section, we evaluate the performance of the proposed schemes under different system parameters.

We use the *modified Saleh-Valenzuelus model* [15] to generate PAS in our simulation. There is an exponential power decay between different clusters. The incidence angle within a cluster, θ_{kli}, satisfies the Laplacian distribution with mean $\overline{\theta}_{kl}$ uniformly distributed over a $120°$-sector $[-60°, 60°)$, and standard deviation σ_{kl}. The PDF of the incidence angle is

$$f_{kl}(\theta_{kli}) = \frac{1}{\sqrt{2}\sigma_{kl}} \exp(-\frac{\sqrt{2}|\theta_{kli} - \overline{\theta}_{kl}|}{\sigma_{kl}}). \tag{31}$$

The whole PAS is the superposition of Laplacian distributions in each cluster. We assume the same standard deviation $\sigma = \sigma_{kl}$ for every user and their total PAS is normalized to 1. The pilot power constraint for each user is set to 10 dB. In the KIPD scheme, the precision parameter ϵ_0 is set to 10^{-3}. For the multiple-interval case, the cluster number of each user takes value of 1,2,3 or 4 with equal probability. We obtained the results by averaging over 1000 simulation runs.

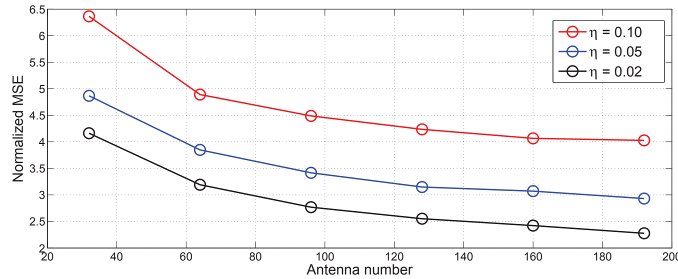


Fig. 2. Normalized MSE vs. antenna number of the GCPR scheme in single-interval case. $K = 20, \sigma = 10°$.

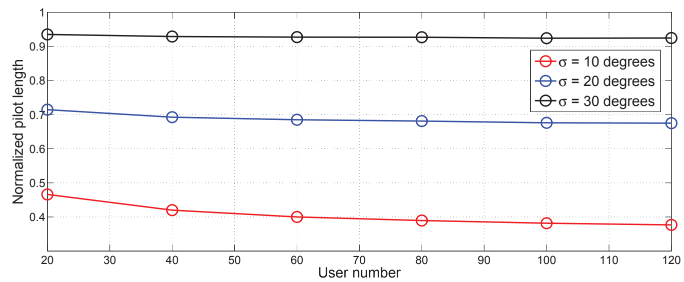


Fig. 3. Normalized pilot length vs. user number of the GCPR scheme in single-interval case. $M = 128, \eta = 0.10$.

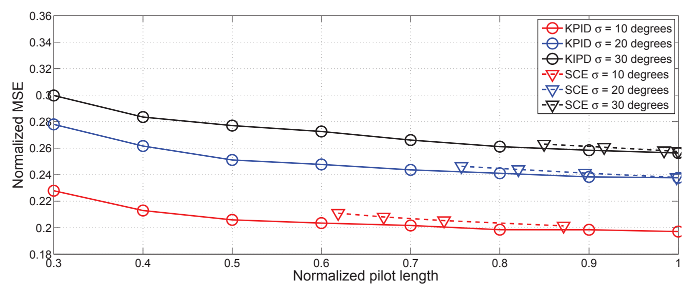


Fig. 4. Normalized MSE vs normalized pilot length of the KIPD and SCE schemes in multiple-interval case. $M = 64, K = 10$.

Firstly, we observe the performance of the GCPR scheme in the single-interval case. Fig. 2 shows the MSE (normalized by antenna number) under different settings of BS antenna number M and PAS truncation loss η. With increasing antenna numbers, the MSE drops as a result of the improved spatial resolution ability of the array to eliminate interferences. On the other hand, setting PAS truncation loss to a higher level will reduce the pilot length at the cost of increasing the MSE. Fig. 3 shows the relationship between the pilot length (normalized by user number) and the user number K, and the impact of the standard deviation σ of angular spread. The angular spreads become more dispersive with larger σ, which leads to the widening of truncated intervals and the increase of the pilot length. We also observe that there is a decease of the normalized pilot length with increasing user number, which means more proportion of training overhead can be saved for massive connectivity scenarios where larger user diversity exists.

Then we turn to the performance of KIPD and SCE schemes in the multiple-interval case, which is depicted in Fig. 4. The results of SCE scheme is obtained by taking different values of η from the set $\{0.20, 0.14, 0.08, 0.02\}$. There is a slight performance difference between the two schemes, and the difference gradually vanishes with the decreasing of η. In addition, we can see that the estimation error of adopting non-orthogonal pilots ($\tau = (0.6 - 0.9)K$) is close to the error performance of orthogonal pilots ($\tau = K$), especially in the system with less dispersive angular spreads (smaller σ). By applying KIPD schemes, we can save 30% of the estimation overhead ($\tau = 0.7K$) with the cost of increasing only 3.3%, 2.5%, 2.3% of the MSE for $\sigma = 10, 20, 30$ degrees, respectively.

VI. Conclusion

In this article, we have investigated non-orthogonal pilot design problem to reduce training overhead for massive MIMO systems with massive connectivity. The pilot matrix is optimized by a KKT-based iterative pilot design (KIPD) algorithm to maximize the mutual information between the received signal and the channel. Furthermore, for the ULA case, we propose a greedy coloring pilot reuse (GCPR) scheme and a segment-by segment channel estimation (SCE) scheme for the single-interval and multiple-interval case of truncated PAS, respectively. The simulation results verify the performance gain of the proposed schemes, and show that non-orthogonal pilot schemes can reduce the pilot length by 30% at the cost of increasing the estimation error by as small as 2.3%.

Acknowledgment

This work is sponsored in part by the National Basic Research Program of China (973 Program: No.2012CB316001), the Nature Science Foundation of China (61571265, 61461136004, 61321061), and Hitachi Ltd.

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 Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[2] S. Y. Lien, K. C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3gpp machine-to-machine communications," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 66–74, April 2011.

[3] Z. Niu, S. Zhou, S. Zhou, X. Zhong, and J. Wang, "Energy efficiency and resource optimized hyper-cellular mobile communication system architecture and its technical challenges," *SCIENTIA SINICA Informationis*, vol. 42, no. 10, pp. 1191–1203, 2012.

[4] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[5] S. L. H. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser mimo systems," in *IEEE Wireless Communications and Networking Conference (WCNC)*, April 2013, pp. 2890–2895.

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

[7] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 264–273, February 2013.

[8] 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 Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3352–3366, June 2015.

[9] K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, "A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 2, pp. 437–447, Mar 2000.

[10] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[11] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated mimo channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 546–557, Feb 2004.

[12] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of fdd massive mimo systems with spatial channel correlation," *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2868–2882, May 2015.

[13] 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, Jul 1996.

[14] U. I. Gupta, D. T. Lee, and J. Y.-T. Leung, "Efficient algorithms for interval graphs and circular-arc graphs," *Networks*, vol. 12, no. 4, pp. 459–467, 1982. [Online]. Available: http://dx.doi.org/10.1002/net.3230120410

[15] Q. Spencer, M. Rice, B. Jeffs, and M. Jensen, "A statistical model for angle of arrival in indoor multipath propagation," in *IEEE 47th Vehicular Technology Conference*, vol. 3, May 1997, pp. 1415–1419.