On-the-fly Large-scale Channel-Gain Estimation for Massive Antenna-Array Base Stations

Chenwei Wang*, Ozgun Y. Bursalioglu*, Haralabos Papadopoulos*, Giuseppe Caire†

*DOCOMO Innovations, Inc., Palo Alto, CA
†Technische Universitat Berlin, Germany
*{cwang, obursalioglu, hpapadopoulos}@docomoinnovations.com, †caire@tu-berlin.de

Abstract—We propose a novel scheme for estimating the large-scale gains of the channels between user terminals (UTs) and base stations (BSs) in a cellular system. The scheme leverages TDD operation, uplink (UL) training by means of properly designed non-orthogonal pilot codes, and massive antenna arrays at the BSs. Subject to Q resource elements allocated for UL training and using the new scheme, a BS is able to estimate the large-scale channel gains of K users transmitting UL pilots in its cell and in nearby cells, provided $K \leqslant Q^2$. Such knowledge of the large-scale channel gains of nearby out-of-cells users can be exploited at the BS to mitigate interference to the out-of-cell users that experience the highest levels of interference from the BS. We investigate the large-scale gain estimation performance provided by a variety of non-orthogonal pilot codebook designs. Our simulations suggest that among all the code designs considered, Grassmannian line-packing type codes yield the best large-scale channel gain estimation performance.

I. INTRODUCTION

Massive MIMO, originally introduced by Marzetta in [1], has been widely recognized as a key enabling technology of 5G/B5G. To harvest massive MIMO gains in the downlink (DL) with multiuser (MU)-MIMO, channel state information (CSI) of the users is needed at the BS, referred to as CSI acquisition. In the FDD systems, UTs first learn the channels based on DL reference signals and subsequently feed their CSI to BS in the UL. The overwhelming CSI-overhead trends of conventional FDD-based schemes, such as those in FDD-based 4G-LTE systems, have spurred a lot of activities and more efficient schemes design [2]. Alternatively, in TDD systems, DL CSI can be directly obtained at the BS from UL training and by capitalizing on the principle of UL/DL channel reciprocity [1]. Reciprocity-based schemes offer fast acquisition of channels between users and a massive BS array and with low overheads, as they require allocating as few UL pilot dimensions as the number of simultaneously served single-antenna users [3].

At higher carrier frequency (e.g., mmWave) bands where large chunks of bandwidth would be available, more massive antennas can be readily deployed on the same footprint due to the shorter wavelengths. Also, due to physics, channels decorrelate much faster at these bands than at the sub-6GHz bands used by LTE, which make the fast low-overhead CSI acquisition of TDD/reciprocity-based schemes even more attractive. In addition, at these higher frequencies, dynamic shadowing and intermittent signal blocking due to the appearance of obstacles lead to shorter intermittent coverage and dramatic pathloss swings, implying the need for much denser, inherently irregular BS/radio-head deployments.

In densely deployed inherently irregular networks, however, managing resource allocation and interference becomes very challenging. For example, the notion of traditional cell planning (along with traditional frequency reuse) is no longer practical, and new access techniques are needed (whether slotted or random access) that readily scale with the densification of the infrastructure. Recently, a class of non-orthogonal pilot codes were proposed in [4] for radio remote head (RRH) networks, called "Spotlight", where users using common pilot dimensions over a slotted-access system are opportunistically served by a subset of nearby RRHs. Spotlight relies on aggressive pilot reuse where many users are aligned on a UL pilot dimension, and exploits very fast user detection at each RRH based on a simple binary energy detection scheme. Although [4] focuses on slotted scheduled access, the same principles can be employed in the context of random access.

In this paper, we focus on a TDD/OFDM-based system which relies on reciprocity-based CSI acquisition, much like the ones considered in e.g., [1] and [4]. Assuming slotted transmission, we consider the OFDM plane is split into resource blocks (RBs), each of which consists of resource elements (REs) within the coherence time and bandwidth of the channel. Also, assuming Q REs in an RB are allocated for UL training, we consider the scenario where K users are in the vicinity of a massive-array BS and transmit pilots over these Q pilot REs using pre-assigned codewords. Focusing on a quasi-static channel model where the user-channels stay constant within any given RB, we study the problem of estimation of the large-scale gains of the user channels based on massive-array observations over the Q pilot REs. In particular, we first determine the conditions on the number of users and the number of pilot REs which ensure large-scale channel gain estimation is feasible. Then we develop and analyze high-performing pilot designs. Due to the space limitation, for more design options and analysis, including extensions to frequency-selective MIMO channel models, please see the extended version [5].

Designing the UL pilot frame to enable the BS to learn large-scale channel gains in addition to small-scale channels has several important application scenarios. For instance, it

can be used in the context of dense small-cell deployments to allow a BS to learn the large-scale channel gains of the strong (nearby) out-of-cell users simultaneously served by nearby BSs, so that it can subsequently exploit them for interference mitigation purposes, which we will investigate in detail in [5]. Another example where the need to learn these large-scale channel gains naturally arises, involves slotted random access, where the BS would also wish to know the subset of the K users that decided to access the medium during any given RB.

II. SYSTEM MODEL

We consider a setting where a single BS is equipped with an M-element antenna array and K single-antenna UTs are randomly distributed around the BS (e.g., in its cell and in nearby cells). We assume TDD operation over OFDM, and employ a quasi-static channel model where each user channel remains constant within an RB comprising a set of T OFDM REs within the user-channel coherence time and bandwidth.

To provide context for the problem of interest, we consider reciprocity-based training for DL transmission over a generic RB, and the DL user channels are learned at the BS via UL user pilots broadcasted within that RB[1]. Moreover, we assume that Q REs are allocated for UL pilot transmission in the RB, and user k broadcasts a pre-assigned $Q \times 1$ pilot pattern, denoted by $\mathbf{p}_k$, over these Q REs. When K users simultaneously transmit their pilot patterns over an RB to train their channels, the received signal at the BS is given by:

$$\mathbf{Y} = \sum_{k=1}^{K} \mathbf{p}_k \mathbf{h}_k^T + \mathbf{W} \tag{1a}$$

where

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_{(1)} & \cdots & \mathbf{y}_{(M)} \end{bmatrix} \text{ and } \mathbf{W} = \begin{bmatrix} \mathbf{w}_{(1)} & \cdots & \mathbf{w}_{(M)} \end{bmatrix} \tag{1b}$$

are both $Q \times M$ matrices. The $M \times 1$ random vector $\mathbf{h}_k = [h_{k1} \cdots h_{kM}]^T \sim \mathcal{CN}(\mathbf{0}, g_k \mathbf{I}_M)$ represents the UL/DL channel between the BS and user k within the RB, with g_k representing the large-scale gain of the user k's channel. In addition, the $Q \times 1$ vectors $\mathbf{y}_{(m)}$, $\mathbf{w}_{(m)} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}_M)$ represent the received signal vector and the additive white Gaussian noise vector at the mth antenna over the Q REs respectively, and they satisfy

$$\mathbf{y}_{(m)} = \sum_{k=1}^{K} \mathbf{p}_k h_{km} + \mathbf{w}_{(m)} = \mathbf{P}\mathbf{h}_{(m)} + \mathbf{w}_{(m)} \tag{2}$$

where $\mathbf{h}_{(m)} \triangleq [h_{1m} \cdots h_{Km}]^T$ is a $K \times 1$ vector and $\mathbf{P} \triangleq [\mathbf{p}_1 \ \cdots \ \mathbf{p}_K] = (p_{qk})_{Q \times K}$ is a $Q \times K$ matrix.

In this paper, we focus on the problem of estimating the large-scale user-channel gains $\{g_k\}$ based on the observation of $\mathbf{Y}$ in (1a) and *a priori* knowledge of the pilot sequences $\mathbf{P}$ pre-assigned to the BS.

Instances of such a problem where the large-scale gains are unknown and need to be estimated naturally arise in cellular networks, whereby each BS independently serves UTs in its

---

[1]Due to to radio channel reciprocity, the UL and DL radio channels between the BS and a given user are the same within any given RB.

cell. In the context of slotted transmission, the K nearby users transmitting UL pilots also include all scheduled users in neighboring cells whose channel gains are unknown at the given BS. Learning the out-of-cell user's channel gains can be used for combating interference to the nearby (strong) out-of-cell users. The problem also naturally arises in the context of slotted random-access, where K represents the number of users in the system, and only a fraction of them are active (i.e., access the UL channel with pilot transmissions) within a given RB. In this case, g_k represents the "effective" channel gain of user k (which is zero if user k is not active) and can be thus treated as unknown at the BS. In either case, it is natural to estimate the large-scale channel gains of the K users first, and then proceed to estimation of the small-scale channel coefficients of the dominant/detected users.

In the following, we provide our solutions to the large-scale gain estimation problem in Sec. III. In the process we also determine the maximum number of users, K, that can be supported for a given number of pilot REs, Q, in the sense that all K large-scale gains can be estimated via the observation in (1a). In Sec. IV and V, we present theoretical and simulation-based performance analysis for a class of pilot designs.

III. PROPOSED SCHEME: EXTRACTING LARGE-SCALE CHANNEL GAINS WITH LARGE ARRAYS

Our proposed method estimates the large-scale channel gains $\{g_k\}$ using the sample covariance matrix of $\mathbf{Y}$:

$$\hat{\mathbf{R}}_{\mathbf{y}} \triangleq \frac{1}{M} \mathbf{Y}\mathbf{Y}^H = \frac{1}{M} \sum_{m=1}^{M} \mathbf{y}_{(m)} \mathbf{y}_{(m)}^H. \tag{3}$$

In particular, the method leverages the presence of a large array (i.e., large M) at the BS, where the sample covariance $\hat{\mathbf{R}}_{\mathbf{y}}$ in (3) converges to the covariance of $\mathbf{y}_{(m)}$. The next theorem gives conditions under which (accurate) estimation of the $\{g_k\}$ is possible with large arrays at the BS.

Theorem 1: Consider the matrix $\mathbf{P} = (p_{qk})_{Q \times K}$ where p_{qk} is complex and generic. When $M \to \infty$, the BS can estimate the large-scale channel gains g_k's of all the K users from (3) almost surely if $K \leqslant Q^2$.

Proof: Since $\mathbf{h}_{(m)}$ and $\mathbf{w}_{(m)}$ are both i.i.d. over m, (2) automatically implies that $\mathbf{y}_{(m)}$ is also i.i.d. over m. Hence, we can drop the foot script m of the relevant notations for brevity. Also, when $M \to \infty$, $\hat{\mathbf{R}}_{\mathbf{y}}$ converges to its mean $\mathbb{E}[\hat{\mathbf{R}}_{\mathbf{y}}]$ almost surely, where

$$\mathbb{E}[\hat{\mathbf{R}}_{\mathbf{y}}] = \mathbb{E}[\mathbf{y}_{(m)} \mathbf{y}_{(m)}^H]. \tag{4}$$

Letting $\mathbf{R}_{\mathbf{y}}$ denote the covariance of $\mathbf{y}_{(m)}$, we have

$$\lim_{M \to \infty} \frac{1}{M^2} \|\hat{\mathbf{R}}_{\mathbf{y}} - \mathbf{R}_{\mathbf{y}}\|_F^2 = 0, \tag{5}$$

which implies that, if $M \to \infty$, then $\mathbf{R}_{\mathbf{y}}$ is also available at the BS almost surely.

Next, we elaborate on how to estimate $\{g_k\}$ from $\mathbf{R}_{\mathbf{y}}$. First of all, we decompose $\mathbf{R}_{\mathbf{y}}$ in the following form:

$$\mathbf{R}_{\mathbf{y}} = \mathbf{P}\mathbb{E}[\mathbf{h}_{(m)} \mathbf{h}_{(m)}^H]\mathbf{P}^H + \mathbb{E}[\mathbf{w}_{(m)} \mathbf{w}_{(m)}^H] \tag{6}$$

$$= \mathbf{PGP}^H + \sigma_w^2 \mathbf{I} \qquad (7)$$

where $\mathbf{G} = \mathrm{diag}([g_1, g_2, \cdots g_K])$. Rewriting (7) as

$$\mathbf{R_y} - \sigma_w^2 \mathbf{I} = \sum_{k=1}^{K} g_k \mathbf{p}_k \mathbf{p}_k^H \qquad (8)$$

yields Q^2 linear equations with the g_k's as the only unknowns. Alternatively, (8) is a system of Q^2 linear equations and K unknowns. To see this, we rewrite (8) into a canonical form. Specifically, we first reshape $\mathbf{R_y}$ and $\sigma_w^2 \mathbf{I}$ to

$$\mathbf{r}_y \triangleq \mathrm{vec}(\mathbf{R_y}) = [\mathbf{r}_{y1}^T \cdots \mathbf{r}_{yQ}^T]^T, \qquad (9)$$

$$\mathbf{r}_w \triangleq \mathrm{vec}(\sigma_w^2 \mathbf{I}) = \sigma_w^2 [\mathbf{e}_1^T \cdots \mathbf{e}_Q^T]^T, \qquad (10)$$

where $\mathbf{r}_{yq}$ denotes the q-th column of $\mathbf{R_y}$, and $\mathbf{e}_q$ represents the unit column vector with $e_{ql} = \delta(q - l)$ as the l-th entry. Next, we define

$$\mathbf{D} = [\mathrm{vec}(\mathbf{p}_1 \mathbf{p}_1^H) \ \mathrm{vec}(\mathbf{p}_2 \mathbf{p}_2^H) \ \cdots \ \mathrm{vec}(\mathbf{p}_K \mathbf{p}_K^H)] \quad (11)$$

$$= [\mathbf{p}_1^* \otimes \mathbf{p}_1 \ \ \mathbf{p}_2^* \otimes \mathbf{p}_2 \ \ \cdots \ \ \mathbf{p}_K^* \otimes \mathbf{p}_K]_{Q^2 \times K} \quad (12)$$

$$\triangleq [\mathbf{d}_1 \ \mathbf{d}_2 \ \cdots \ \mathbf{d}_K] \qquad (13)$$

where "$\otimes$" is the Kronecker product operator, and its kth column is $\mathbf{d}_k = \mathbf{p}_k^* \otimes \mathbf{p}_k$. Then (8) can be rewritten as

$$\mathbf{r} \triangleq \mathbf{r}_y - \mathbf{r}_w = \mathbf{Dg} \qquad (14)$$

where $\mathbf{g} = \mathrm{diag}(\mathbf{G}) = [g_1, \cdots, g_K]^T$. Note that $\mathbf{r}$ is obtained from $\mathbf{R_y}$ and the noise power σ_w^2 only, and $\mathbf{D}$ depends on $\mathbf{P}$ only. Thus, given $\mathbf{R_y}$, σ_w^2 and $\mathbf{P}$, we can obtain a *unique* $\mathbf{g}$ satisfying the Q^2 linear equations in (14), as long as the fat $Q^2 \times K$ matrix $\mathbf{D}$ (due to $Q^2 \leqslant K$) has full rank K. As shown in Appendix A, since $\mathrm{rank}(\mathbf{D}) = K$, the solution is

$$\mathbf{g} = \mathbf{D}^\dagger \mathbf{r}. \qquad (15)$$

Therefore, we complete the proof of Theorem 1. ∎

Remark 1: Theorem 1 shows that Q pilot REs allow the BS to estimate the large-scale channel gains of up to Q^2 users. The key insight behind this fact is that a complex covariance matrix has Q^2 degrees of freedom, so that we can collect up to Q^2 linearly independent equations.

In the finite M case where $\widehat{\mathbf{R}}_\mathbf{y}$ differs from $\mathbf{R_y}$, the proof of Theorem 1 suggests a method for estimating $\mathbf{g}$ from the sample covariance $\widehat{\mathbf{R}}_\mathbf{y}$ in (3). Following the same vectorization operation as in (9), we first compute

$$\widehat{\mathbf{r}} = \mathrm{vec}(\widehat{\mathbf{R}}_\mathbf{y}) - \mathbf{r}_w, \qquad (16)$$

and then obtain $\widehat{\mathbf{g}}$ (the estimate of $\mathbf{g}$) as the solution to the following optimization problem:

$$\underset{\boldsymbol{\theta}}{\mathrm{minimize}} \quad \|\widehat{\mathbf{r}} - \mathbf{D}\boldsymbol{\theta}\|^2 \qquad (17a)$$

$$\mathrm{subject\ to} \quad \boldsymbol{\theta} \succeq \mathbf{0}. \qquad (17b)$$

Problem (17) is a non-negative least-squares (NNLS) problem, a special quadratic programming problem, which has been intensively studied recently [11], [12], and can be readily solved by general-purpose quadratic-programming solvers[2].

---

[2]In fact, seeking efficient and special-purpose solvers of (17) which exploit the sparsity promoting properties of NNLS are a topic and worth further investigation.

It is worth making a few remarks. First, massive arrays improve the estimation performance, as increasing the size of the BS antenna array improves the sample covariance estimate and thus the quality of the large-scale gain estimates provided by (17). In addition, the pilot design matrix $\mathbf{P}$'s, with random entries sufficing in principle systematic designs as we will show later in Sec. V, can provide $\mathbf{D}$'s with desirable properties and superior performance.

While Theorem 1 is stated for the entries of $\mathbf{P}$ being complex and generic, we also consider practical constraints such as restricting all pilot values to be real-valued, or have a constant amplitude but random phases. Based on Theorem 1, we have the following two corollaries.

Corollary 1: Consider $\mathbf{P} = (p_{qk})_{Q \times K}$ where p_{qk}'s are real and generic. When $M \to \infty$, the BS is able to identify all the K users almost surely if $K \leqslant Q(Q+1)/2$.

Proof: The proof is deferred to Appendix B. ∎

Corollary 2: Consider $\mathbf{P} = (p_{qk})_{Q \times K}$ where $p_{qk} = e^{\theta_{qk}}$ and $\theta_{qk} \sim \mathcal{U}[0, 2\pi]$ is i.i.d. generated. When $M \to \infty$, the BS is able to identify all the K users almost surely if $K \leqslant Q^2 - Q + 1$.

Proof: The proof is deferred to Appendix C. ∎

Remark 2: The intuition behind the corollaries above is that the degrees of freedom of their covariance matrices reduce to $Q(Q+1)/2$ and $Q^2 - Q + 1$, respectively.

In fact, in both motivating examples we described at the onset, only a fraction of users are expected to have significantly nonzero (or appreciable) large-scale channel gains with respect to the BS, say K' users. If $K' \ll K$, then (17) becomes a compressed sensing problem, such as in [10].

Finally, we briefly discuss the setting with $K > Q^2$ where Theorem 1 cannot be applied. As a matter of fact, as long as $K' \leqslant Q^2$, certain accurate solutions can still be found with high probability even with $K > Q^2$. In this case, (17) can be directly solved using an NNLS solver (without the need for regularization). If $K' \ll K$, then the solution can still be uniquely identified with high probability (see [11]).

IV. PERFORMANCE METRIC ANALYSIS

In this section, we investigate the estimation performance of large-scale gains. Using "$\circ$" as the Hadamard product operator and "$|\cdot|$" as the element-wise amplitude-taking operator, we first show a key property between the column inner products of $\mathbf{D}$ and $\mathbf{P}$ in the following proposition:

Proposition 1: $\mathbf{d}_i^H \mathbf{d}_j = |\mathbf{p}_i^H \mathbf{p}_j|^2$ for $\forall i$ and $\forall j$. In addition, $\mathbf{D}^H \mathbf{D} = (\mathbf{P}^H \mathbf{P}) \circ (\mathbf{P}^H \mathbf{P})^H = |\mathbf{P}^H \mathbf{P}|^{\circ 2}$.

Proof: We first prove the first equality in the following:

$$\mathbf{d}_i^H \mathbf{d}_j = \sum_{l=1}^{Q^2} d_{li}^* d_{lj} = \sum_{q=1}^{Q} \left(\sum_{q'=1}^{Q} (p_{qi}^* p_{q'i})^* (p_{qj}^* p_{q'j}) \right) \quad (18a)$$

$$= \sum_{q=1}^{Q} p_{qi}(p_{qj})^* \left(\sum_{q'=1}^{Q} p_{q'i}(p_{q'j})^* \right)^* \quad (18b)$$

$$= |\mathbf{p}_i^H \mathbf{p}_j|^2. \qquad (18c)$$

It can be seen that no matter if $\mathbf{D}$ is real-valued or complex-valued, the inner product of its any two columns is always real-valued and non-negative. From (18), we can directly obtain the other equations $(\mathbf{D}^H\mathbf{D})_{i,j} = \mathbf{d}_i^H\mathbf{d}_j = |\mathbf{p}_i^H\mathbf{p}_j|^2 = |\mathbf{P}_{:,i}^H\mathbf{P}_{:,j}|^{\circ 2} = (|\mathbf{P}^H\mathbf{P}|^{\circ 2})_{i,j}$ for any i, j. ∎

Remark 3: The proposition reveals that $\mathbf{D}^H\mathbf{D}$ is real, symmetric, and non-negative, and each of its diagonal entries dominates all the entries on the same row and column.

In the rest of this section, we restrict our attention to the case of $K \leqslant Q^2$, i.e., in the range where Theorem 1 is valid. Let us go back to problem (17) where we aim to minimize $\|\hat{\mathbf{r}} - \mathbf{D}\boldsymbol{\theta}\|^2$ subject to $\boldsymbol{\theta} \succeq \mathbf{0}$. Recall that in the finite M case $\hat{\mathbf{r}}$ differ from $\mathbf{r}$. Letting $\mathbf{r}_e = \hat{\mathbf{r}} - \mathbf{r}$ denote the large-scale gain estimation error and using (14), we obtain

$$\hat{\mathbf{r}} = \mathbf{D}\mathbf{g} + \mathbf{r}_e. \tag{19}$$

We ignore the non-negativity constraint on $\{g_k\}$ and thereby focus on the analysis of the ZF estimator:

$$\hat{\mathbf{g}} = \mathbf{D}^\dagger\hat{\mathbf{r}} = (\mathbf{D}^H\mathbf{D})^{-1}\mathbf{D}^H\hat{\mathbf{r}} = \mathbf{g} + \underbrace{(\mathbf{D}^H\mathbf{D})^{-1}\mathbf{D}^H\mathbf{r}_e}_{\triangleq \; \mathbf{z}: \text{ estimation noise}}. \tag{20}$$

To see the estimation performance of $\hat{\mathbf{g}}$, we can analyze the noise enhancement, captured by the $K \times K$ covariance matrix of the estimation noise (error) vector $\mathbf{z}$:

$$\mathbf{R}_z \triangleq \mathbb{E}(\mathbf{z}\mathbf{z}^H) = (\mathbf{D}^H\mathbf{D})^{-1}\mathbf{D}^H\mathbb{E}(\mathbf{r}_e\mathbf{r}_e^H)\mathbf{D}(\mathbf{D}^H\mathbf{D})^{-1}. \tag{21}$$

As a closed-formed expression of $\mathbb{E}(\mathbf{r}_e\mathbf{r}_e^H)$ is cumbersome and $\mathbf{g}$ dependent, we opt to use $\mathcal{CN}(\mathbf{0}, \sigma_e^2\mathbf{I})$ in place of the true distribution[3] of $\mathbf{r}_e$, and $\sigma_e^2 = \max_q((\mathbb{E}(\mathbf{r}_e\mathbf{r}_e^H))_{q,q})$. Also, note that this distribution has the same asymptotic behavior as the true distribution, since $\sigma_e^2 \to 0$ when $M \to \infty$. Hence, we simply the expression of $\mathbf{R}_z$ to

$$\mathbf{R}_z = \sigma_e^2(\mathbf{D}^H\mathbf{D})^{-1}. \tag{22}$$

Inspection of the expression above reveals that $\mathbf{R}_z$ arises as the combined effect of $(\mathbf{D}^H\mathbf{D})^{-1}$ which clearly is a function of the pilot code design, $\mathbf{P}$, and of σ_e^2, which captures the accuracy of the sample covariance estimate and does *not* depend on $\mathbf{P}$. As a result, we opt to evaluate the impact in estimation performance of any given pilot code design, via the distribution of the eigenvalues[4] of $(\mathbf{D}^H\mathbf{D})^{-1}$ and their aggregate effect via the trace of $(\mathbf{D}^H\mathbf{D})^{-1}$. Given that the trace is minimized when $\mathbf{D}$ is unitary, we can view the trace of $(\mathbf{D}^H\mathbf{D})^{-1}$ as a measure of noice enhancement.

A. Noise Enhancement Analysis

In this section, we will consider Grassmannian line packing [9] to design the pilot code sequences, since it provides good properties of the inner product between arbitrary two vectors. Using algorithms such as in [7], we can find out a $Q \times K$

---

matrix $\mathbf{P}$, i.e., we can pack K 1-dimensional lines in the Q-dimensional space, so that for any two columns $\mathbf{p}_i$ and $\mathbf{p}_j$, the amplitude of their inner product $c_p \triangleq \max_{\mathbf{P}} \min_{i \neq j}(|\mathbf{p}_i^H\mathbf{p}_j|)$ is as large as possible. Given Q and K, [8] (Theorem 2.3 and Corollary 2.4) provided the following lower bound on c_p, also referred to as the Welch bound:

$$c_p \geqslant \sqrt{\frac{K-Q}{Q(K-1)}}. \tag{23}$$

It was shown in [8] that this lower bound is tight and thus $|\mathbf{p}_i^H\mathbf{p}_j| = \sqrt{\frac{K-Q}{Q(K-1)}}$ for $\forall i$ and $\forall j \neq i$, when $K \leqslant Q^2$ if $\mathbf{P} \in \mathbb{C}^{Q \times K}$, and when $K \leqslant Q(Q+1)/2$ if $\mathbf{P} \in \mathbb{R}^{Q \times K}$, i.e., satisfying the condition called the *full frame*. Note that the conditions $K \leqslant Q^2$ and $K \leqslant Q(Q+1)/2$ are consistent with our Theorem 1 and Corollary 1 in Sec. III, respectively.

Next, we consider $\mathbf{P} \in \mathbb{C}^{Q \times K}$, $K \leqslant Q^2$ and for any two columns of $\mathbf{P}$ their $c_p = \sqrt{\frac{K-Q}{Q(K-1)}}$ as an example. Because of Proposition 1, denoting by $c_d = c_p^2$, we can explicitly obtain

$$\mathbf{D}^H\mathbf{D} = \begin{bmatrix} 1 & c_d & \cdots & c_d \\ c_d & 1 & \cdots & c_d \\ \vdots & \vdots & \ddots & \vdots \\ c_d & c_d & \cdots & 1 \end{bmatrix}_{K \times K} \tag{24}$$

which can be further written into the following compact form:

$$\mathbf{D}^H\mathbf{D} = c_d\mathbf{1}_{K \times 1}(\mathbf{1}_{K \times 1})^T + (1 - c_d)\mathbf{I}_K, \tag{25}$$

where the diagonal entries are 1, and the off-diagonal entries are c_d. Hence, its eigenvalues are simply given by:

$$\lambda_k = \begin{cases} 1 + (K-1)c_d, & k = 1, \\ 1 - c_d, & k = 2, 3, \cdots, K. \end{cases} \tag{26}$$

Finally, we obtain the eigenvalues of $(\mathbf{D}^H\mathbf{D})^{-1}$ as follows:

$$\lambda_k' = \begin{cases} \frac{1}{1+(K-1)c_d}, & k = 1, \\ \frac{1}{1-c_d}, & k = 2, \cdots, K. \end{cases} \tag{27}$$

V. SIMULATION RESULTS

In this section, we will investigate the noise enhancement via simulation for $\mathbf{P} \in \mathbb{C}^{Q \times K}$ and $K \leqslant Q^2$.

First, we consider $K = Q^2$ by employing the codebooks generated by Gaussian complex random variables, the theoretical Grassmannian line packing, and the packet of Grassmannian line packing provided in [9], respectively. In particular, we will look into the eigenvalues of corresponding $(\mathbf{D}^H\mathbf{D})^{-1}$.

1) Gaussian Random Codebook: The matrix $\mathbf{P}$ is constructed by first drawing each entry from $\mathcal{CN}(0, 1)$ and then normalizing every column to have unit norm. Once $(\mathbf{D}^H\mathbf{D})^{-1}$ is obtained, the average noise enhancement per dimension is given by $10\log_{10}(\sum_{k=1}^K \lambda_k'/Q^2)$ dB.

2) Theoretical Grassmannian-Line-Packing Codebook: The eigenvalues of $(\mathbf{D}^H\mathbf{D})^{-1}$ under $K = Q^2$ can be simplified to: $\lambda_1' = 1/Q$ and $\lambda_k' = 1 + 1/Q$ for $k = 2, \cdots, K$. Thus, the average noise enhancement per dimension is $10\log_{10}(1 + (Q-1)/Q^2)$ dB. Note that $\mathbf{P}$ archiving the equality in (23) can be explicitly expressed in the closed form for special cases

only (see discussion in [8]). In addition, when Q is large, the average noise enhancement per dimension is approximately $(10 \log_{10} e)/Q$ dB (close to 0 dB), where e is the natural logarithmic base number.

3) Simulated Grassmannian-Line-Packing Codebook: Owing to the existence of the Grassmannian line packing limit but the lack of general close-formed construction, a vast amount of iterative algorithms have been developed to construct a set of vectors approximating the theoretical limit. By running the package provided by [9] to generate $\mathbf{P}$ and then $(\mathbf{D}^H \mathbf{D})^{-1}$, the average noise enhancement per dimension is given by $10 \log_{10}(\sum_{k=1}^{K} \lambda'_k / Q^2)$ dB.

Based on the three types of codebook design introduced above, we will show their noise enhancement performance for $K = Q^2$ where $Q = 2, 3, \cdots, 10$. Fig. 1 shows the comparison of average noise enhancement per dimension. The noise enhancement of the Gaussian codebook could be even 50 dB higher than the Grassmannian approach at $Q = 10$, which means that Grassmannian approach is much better. Note that the red dashed curve (line packing algorithms) diverges from the black curve (the theoretical limit) when $Q \geqslant 7$. This is because when we ran the package provided by [9], we kept the output once we were asked whether to stop, which implies that when Q is large, the number of iterations is not large enough to produce more precise result. As we tested for several exam-
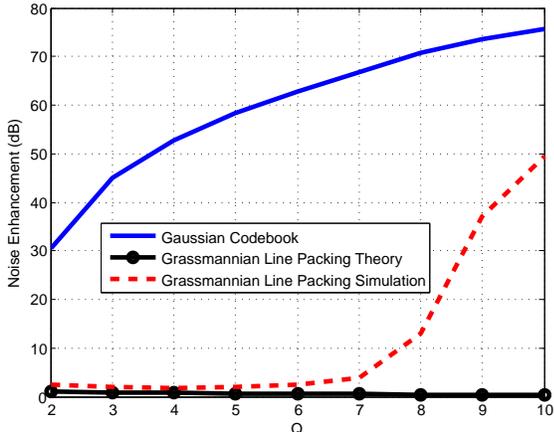
Fig. 1. Comparison of Average Noise Enhancement per Dimension
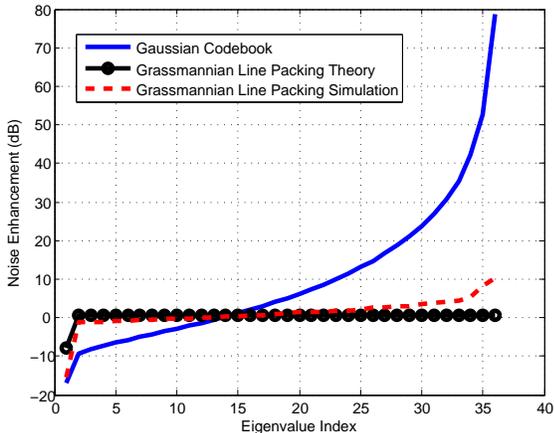
Fig. 2. Comparison of Noise Enhancement of Each Dimension
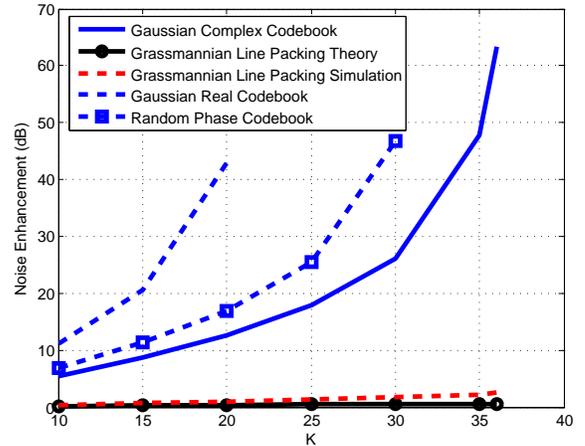
Fig. 3. Comparison of Noise Enhancement of Each Dimension

ples, running more iterations would make the simulated curve closer to the theoretical curve at cost of the time complexity[5]. In Fig. 2, we show the the noise enhancement of codebooks in each of the corresponding $Q^2 = 36$ dimensions. It can be seen that the noise enhancement of the Grassmannian Line Packing codebook is very close to 0 dB in all dimensions. In contrast, although the noise enhancement of the Gaussian codebook is even less in some dimensions, it could much higher (up to 78 dB) in the other dimensions than the Grassmannian approach.

Besides $K = Q^2$, we also study the setting with $K < Q^2$ since we might not need to deal with fully loaded systems. Since $K \leqslant Q$ implies orthogonal pilot design, we consider $Q < K < Q^2$ only. Note that for "Theoretical Grassmannian Line Packing", we need to replace with the following result:

$$10 \log_{10} \left(\frac{Q}{K^2} + \frac{Q(K-1)^2}{(Q-1)K^2} \right) \text{ dB.} \tag{28}$$

Moreover, we also consider the Gaussian real codebook when $K \leqslant Q^2 - Q + 1$ and the random phase codebook when $K \leqslant Q(Q+1)/2$ specified in Sec. III. Fig. 3 shows the noise enhancement in each dimension in response to K by using the 5 methods of codebook construction for $Q = 6$. Clearly, the Grassmannian codebook outperforms all the others.

VI. CONCLUSION

In this paper, we propose a novel scheme for estimating large-scale gains of the channels between the users and the BSs on the fly when we apply non-orthogonal pilot codes for UL training in TDD/reciprocity-based systems with massive antenna arrays at the BSs. With Q REs allocated for UL training, the new scheme enables the BS to estimate the large-scale channel gains of up to all K users as long as $K \leqslant Q^2$. The key of the new scheme is the use of the massive antenna array incorporated with fully exploring the degrees of freedom of the covariance matrix and non-orthogonal pilot codes.

---

[5]It is worth noting that while [8] implies the theoretical Grassmannian line packing limit is achievable, the simulated curve differs from the theoretical curve due to precision allowance and numerical errors accumulated through algorithm iterations. The simulated curve can be made closer to the theoretical one by either running more iterations of existing algorithms or inventing more efficient algorithms, which are not the main focus of this paper.

Following our proposed scheme, several interesting problems need further investigation. For example, if channel dependencies exist among the co-located antennas, how to use the sample covariance matrix to better approximate the exact covariance matrix is interesting. For another example, while the Grassmannian line packing codebook is preferable for estimating large-scale gains, it is well know that orthogonal pilot code design is desired for estimating small-scale fading, because projecting the received signal into each spatial dimension does not incur power loss. Therefore, it would be of interest to investigate if there exist any other efficient and intermediate scheme to inherit both of their good features.

APPENDIX

A. The Proof of $\text{rank}(\mathbf{D}) = \mathbf{K}$ for Theorem 1

To show $\text{rank}(\mathbf{D}) = K$ almost surely for $K \leqslant Q^2$, it suffices to show $\det(\mathbf{D}) \neq 0$ almost surely when $K = Q^2$. Note that $\det(\mathbf{D})$ is a polynomial of the generic entries (or variables) p_{qk}'s of $\mathbf{P}$, defined in the continuous field. Thus, $\det(\mathbf{D}) = 0$ implies that either it always holds for arbitrary choices of p_{qk}'s or the finite number of solutions to satisfy $\det(\mathbf{D}) = 0$ constitute a subset with Lebesgue measure zero. Hence, to prove $\det(\mathbf{D}) \neq 0$ almost surely, we only need to find one specific choice of p_{qk}'s so that $\det(\mathbf{D}) \neq 0^6$. Specifically, we choose $\mathbf{P}$ to be a Vandermonde matrix:

$$\mathbf{P} = \begin{bmatrix} 1 & a_1 & (a_1)^2 & \cdots & (a_1)^{Q^2-1} \\ 1 & a_2 & (a_2)^2 & \cdots & (a_2)^{Q^2-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_Q & (a_Q)^2 & \cdots & (a_Q)^{Q^2-1} \end{bmatrix}_{Q \times Q^2} \quad (29)$$

where $a_i \neq a_j$ for $\forall i, \forall j \neq i$. We will show that the resulting $\mathbf{D}$ is also a Vandermonde matrix. To see this, we write the kth column of $\mathbf{D}$ for each $k = 1, \cdots, Q^2$ as follows:

$$\mathbf{d}_k \triangleq \mathbf{p}_k^* \otimes \mathbf{p}_k$$
$$= [(a_1)^{k-1} \cdots a_Q^{k-1}]^H \otimes [a_1^{k-1} \cdots (a_Q)^{k-1}]^T \quad (30)$$
$$= [((a_1)^{k-1})^* [a_1^{k-1} \cdots a_Q^{k-1}], \cdots, ((a_Q)^{k-1})^* [a_1^{k-1} \cdots a_Q^{k-1}]]^T \quad (31)$$
$$= [(a_1^*)^{k-1} [a_1^{k-1} \cdots a_Q^{k-1}], \cdots, (a_Q^*)^{k-1} [a_1^{k-1} \cdots a_Q^{k-1}]]^T \quad (32)$$
$$= [(a_1^* a_1)^{k-1}, \cdots, (a_1^* a_Q)^{k-1}, (a_2^* a_1)^{k-1}, \cdots, (a_2^* a_Q)^{k-1},$$
$$\cdots, (a_Q^* a_1)^{k-1}, \cdots, (a_Q^* a_Q)^{k-1}]^T. \quad (33)$$

It can be seen that each entry has the same exponential factor $k - 1$ for the kth column of $\mathbf{D}$. Next, we denote by $\mathbf{d}_2 = [b_1, b_2, \cdots, b_{Q^2}]^T$, and its lth entry can be easily written as

$$b_l = a_{l_1}^* a_{l_2}, \qquad l = 1, \cdots, Q^2 \quad (34)$$

where $l_2 = \text{mod}(l - 1, Q) + 1$ and $l_1 = (l - l_2)/Q + 1$, i.e., $l = l_2 + (l_1 - 1)Q$. Thus, we directly obtain

$$\det(\mathbf{D}) = \prod_{1 \leqslant i < j \leqslant Q^2} (b_j - b_i) = \prod_{1 \leqslant i < j \leqslant Q^2} (a_{j_1}^* a_{j_2} - a_{i_1}^* a_{i_2}), \quad (35)$$

where $i = i_2 + (i_1 - 1)Q$, $j = j_2 + (j_1 - 1)Q$. Given any i, j and $i \neq j$, we must have $(i_1, i_2) \neq (j_1, j_2)$. Since each a_q, $q = 1, \cdots, Q$ is generic, we have $b_j - b_i \neq 0$ almost surely, which directly implies $\det(\mathbf{D}) \neq 0$ almost surely.

---

^{6}This approach has been widely used in network information theory to study linear independencies among the row/column vectors of a matrix.

B. The Proof of Corollary 1

Observations of (7) reveal that $\mathbf{R_y} - \sigma_w^2 \mathbf{I} = \mathbf{PGP}^H = \mathbf{PGP}^T$ is real symmetric, which means that the entries in its upper triangle are the same as those in the lower triangle. Thus, the number of equations that we need to consider are only those corresponding to the diagonal entries and the entries in one triangle, which is given by $Q + (Q^2 - Q)/2 = Q(Q+1)/2$. Translating this observation into examining each row of the resulting matrix $\mathbf{D}$ implies that its ith row and j^{th} row are identical, whenever $i = l_2 + (l_1 - 1)Q$, $j = l_1 + (l_2 - 1)Q$, for $\forall i, \forall j \neq i$. If we delete the repeated $Q(Q-1)/2$ rows of $\mathbf{D}$ and also the last $Q(Q-1)/2$ columns of $\mathbf{D}$, the resulting $(Q(Q+1)/2) \times (Q(Q+1)/2)$ matrix $\mathbf{D}'$ is still a Vandermonde matrix. Following the similar proof, we can show $\mathbf{D}'$ has full rank almost surely. Overall, if $K \leqslant Q(Q+1)/2$, (14) are under-constrained, and thus $\mathbf{g}$ can be uniquely determined.

C. The Proof of Corollary 2

Observations of (7) reveal that the diagonal entries of $\mathbf{R_y} - \sigma_w^2 \mathbf{I} = \mathbf{PGP}^H$ is a scaled identity matrix, which means that the lth row of $\mathbf{D}$ is identical to its first row, when $l = q + (q-1)Q$ for every $q = 2, \cdots, Q$. If we delete the repeated $Q - 1$ rows of $\mathbf{D}$ and also the last $Q - 1$ columns of $\mathbf{D}$, the resulting $(Q^2 - Q + 1) \times (Q^2 - Q + 1)$ matrix $\mathbf{D}'$ is again a Vandermonde matrix. Following the similar proof, we can show $\mathbf{D}'$ has full rank almost surely. Overall, if $K \leqslant Q^2 - Q + 1$, (14) is under-constrained, and $\mathbf{g}$ can be uniquely determined.

REFERENCES

[1] T. L. Marzetta, "How much training is required for multiuser MIMO?," *Proc. 2006 Fortieth Asilomar Conference on Signals, Systems and Computers,* pp.359 – 363, Nov. 2006.

[2] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint Spatial Division and Multiplexing – The Large-Scale Array Regime," *IEEE Transactions on Information Theory,* Vol. 59, Issue 10, pp.6441–6463, Oct. 2013.

[3] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. on Wireless Commun.,* vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[4] 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," *IEEE ICC 2016,* May 2016.

[5] C. Wang, O. Y. Bursalioglu, H. Papadopoulos, and G. Caire, "On-the-fly Large-scale Channel-Gain Estimation for Massive Antenna-Array Base Stations," *in Prepareation,* 2018.

[6] S. Kumar, and Z. Ahmed, "Spectral statistics for ensembles of various real random matrices," *arxiv: 1704.02715 [quant-ph],* April 2017.

[7] D. J. Love, "Grassmannian Subspace Packing," *[Available Online] https://engineering.purdue.edu/%7Edjlove/grass.html.*

[8] T. Strohmer, and R. W. Heath Jr., "Grassmannian Frames with Applications to Coding and Communication," 2002, *[Available online] https://www.math.ucdavis.edu/%7Estrohmer/papers/2002/grass.pdf.*

[9] A. Medra, and T. N. Davidson, "Flexible codebook design for limited feedback systems via sequential smooth optimization on the Grassmannian manifold," *IEEE Trans. Signal Processing,* Vol. 62, Issue 5, pp. 1305–1318, March 2014. The simnulation package is available at *http://www.ece.mcmaster.ca/~davidson/pubs/Flexible_codebook_design.html.*

[10] 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, Jun. 2010.

[11] V. Sharan, K. Tai, P. Bailis, and G. Valiant, "There and Back Again: A General Approach to Learning Sparse Models," *arXiv:1706.08146v1 [cs.LG],* June, 2017.

[12] S. Foucart, and D. Koslicki, "Sparse Recovery by Means of Nonnegative Least Squares," *IEEE Signal Processing Letters,* Vol. 21, Issue 4, pp.498–502, April 2014.