Limited Feedback Scheme for Massive MIMO in Mobile Multiuser FDD Systems

Ernest Kurniawan, Jingon Joung, and Sumei Sun
Institute for Infocomm Research, Agency for Science Technology and Research (A*STAR)
1 Fusionopolis Way, #21-01 Connexis (South Tower), Singapore 138632
Email: {ekurniawan, jgjoung, sunsm}@i2r.a-star.edu.sg

Abstract—**Massive multiple input multiple output is a promising technology to keep up with the explosive demand of wireless data traffic. The benefits of having a large number of antennas, however, depend on the availability of channel state information (CSI), especially at the transmitter. In frequency division duplex systems, this CSI has to be sent back via the uplink channel, hence incurring a large overhead and degrading the spectral efficiency. Mobility of the users and the large number of antennas exacerbate the problem with frequent tracking of the many time-varying CSI coefficients. This paper presents one approach to address this issue. By tracking only the principal components of the channel gain, and exploiting the wide-sense stationarity of the channel, the amount of required feedback can be reduced significantly. Simulation study shows that the proposed technique is able to achieve high sum-rate with good tracking capability using only limited feedback.**

I. INTRODUCTION

The exponential growth of wireless data traffic and the limited available spectrum have made resource-efficient transmission strategy critical in future wireless systems. Multiple Input Multiple Output (MIMO) is a technique to increase the spectral efficiency, and recently, the idea of deploying very large number of antennas (massive MIMO) has attracted a lot of attention, both in the academia [1]-[2] and industry [3]. Advantages of massive MIMO include the channel hardening phenomena [4], where the channel between users are asymptotically orthogonal, hence eliminating the interference. Other advantages such as increased capacity [5] and better energy efficiency [6] have also been reported. Despite all the benefits of massive MIMO, as with other multi-antenna systems, their performance depends largely on the Channel State Information (CSI) availability, especially at the transmitter. There has been a lot of research on how to obtain this using only limited feedback (e.g. [7]-[8] and reference therein). For massive MIMO, however, the problem is more severe, and earlier methods for conventional MIMO systems cannot be directly applied due to the large dimension of the CSI. Furthermore, in a mobile multiuser environment, the time variation requires frequent channel tracking for each user at the base station, making it impractical for massive MIMO implementation.

In Time Division Duplex (TDD) systems, channel reciprocity can be used to alleviate this problem. In Frequency Division Duplex (FDD) systems however, this is not the case, the uplink and downlink CSI may be statistically independent of each other when the two bands are separated wider than the coherence bandwidth. This paper focuses on this FDD scenario

with mobile multiuser setting. Particularly, we develop a technique to track channel variation of multiple mobile users, while keeping the amount of feedback low.

The proposed scheme combines three strategies. Firstly, the scheme tracks only its *principal components* instead of tracking the actual channel vector. This idea of using a sparse representation of the channel vector has been considered before in the context of feedback reduction in massive MIMO [9]. Here, we extend its application for channel tracking.

Secondly, in order to perform Principal Component Analysis (PCA), the knowledge of channel spatial correlation matrix is needed. Due to the large number of elements, it is impractical for the receiver to feedback this information. To address this issue, we exploit the propagation property of the channel. Namely, for Wide-Sense-Stationary and Uncorrelated Scattering (WSSUS) channels, with the same dominant scatterers in both uplink and downlink such that the Angle of Arrival (AoA) and Angle of Departure (AoD) at and from the base station are similar, the spatial correlation matrix is frequency independent [10]. As such, the correlation matrix can be estimated in the uplink channel using *subspace learning* techniques.

Lastly, we consider using a *perturbation based channel tracking*. This idea is motivated by [11], whereby several perturbation vectors (each occupying orthogonal downlink resources) are added to the current tracking vector. The receiver can then feedback only the index of the most favorable vector.

Simulation study shows that under typical 3rd Generation Partnership Project (3GPP) Long Term Evolution (LTE) parameters, our proposed scheme is able to achieve approximately 8 to 10 b/s/Hz sum-rate improvement compared to the case without channel tracking. Moreover, the proposed user scheduling scheme is shown to prevent sum-rate saturation at high Signal to Noise Ratio (SNR) as experienced by the round robin scheme due to better interference management. All these are achieved using only limited amount of feedback bits.

The remaining part of this paper is organized as follows. Section II describes the system model used throughout the analysis. The detailed description on the proposed scheme is given in Section III, followed by simulation study in Section IV. Finally, Section V ends this paper with conclusions.

II. SYSTEM MODEL

A cellular system with one Base Station (BS) equipped with $M >> 1$ antennas communicating simultaneously with

K single-antenna[1] Mobile Stations (MS) is considered. The downlink transmission uses Orthogonal Frequency Division Multiple Access (OFDMA), whereby different user groups are allocated orthogonal resource blocks. The number of MS in each group and the scheduling process will be explained later. The uplink transmission, on the other hand, uses Single Carrier Frequency Division Multiple Access (SC-FDMA), and each user is assigned a non-overlapping subcarrier block. This combination of downlink OFDMA and uplink SC-FDMA has been used in the 3GPP LTE standard [12].

Denoting the set of subcarriers allocated for downlink and uplink as $\mathcal{F}_D$ and $\mathcal{F}_U$, respectively (with $\mathcal{F}_D \cap \mathcal{F}_U = \emptyset$), the discrete time signal models at the MS and BS corresponding to user index k at its allocated subcarrier n ($n \in \mathcal{F}_D$ for downlink and $n \in \mathcal{F}_U^{(k)} \subset \mathcal{F}_U$ for uplink) and time slot t are

$$y_k^{(MS)}[n,t] = \boldsymbol{h}_k^H[n,t]\boldsymbol{x}^{(BS)}[n,t] + z_k[n,t], \quad n \in \mathcal{F}_D, \ (1)$$

$$\boldsymbol{y}^{(BS)}[n,t] = \boldsymbol{g}_k[n,t]x_k^{(MS)}[n,t] + \boldsymbol{z}[n,t], \quad n \in \mathcal{F}_U^{(k)}. (2)$$

We will drop the subcarrier index n and time index t whenever the indexing is either irrelevant or it is clear from the context to simplify the notation. The $M \times 1$ column vector $\boldsymbol{h}_k$ and $\boldsymbol{g}_k$ are the downlink and uplink channel gains between the BS and the kth MS, respectively. The $x_k^{(MS)}$ is the transmitted uplink symbol from the kth MS, and the $\boldsymbol{x}^{(BS)}$ is the downlink transmitted symbol vector from the BS which is given by

$$\boldsymbol{x}^{(BS)}[n,t] = \sum_{k \in \mathcal{B}_{n,t}} \boldsymbol{v}_k[n,t]s_k[n,t]. \quad (3)$$

We have used the notation $\mathcal{B}_{n,t}$ for the set of MS indexes that belong to the same user group assigned to subcarrier n during time slot t. $\boldsymbol{v}_k[n,t]$ and $s_k[n,t]$ are the beamforming vector and the information symbol for the kth MS, respectively.

In massive MIMO, as the antennas are placed close to one another, the channel gains tend to be correlated. Following the WSSUS property (frequency independent spatial correlation) and the channel reciprocity on each subcarrier, the spatial correlation matrix of MS k's channel can be calculated as

$$E\left[\boldsymbol{h}_k[n,t]\boldsymbol{h}_k^H[n,t]\right] = E\left[\boldsymbol{g}_k[n,t]\boldsymbol{g}_k^H[n,t]\right] = \boldsymbol{R}_k(t), \forall\{n,t\}. (4)$$

While it is reasonable to assume that the channel gain in the uplink and downlink on each subcarrier are conjugate to each other, we only require statistical reciprocity, which is less stringent and can be justified by the fact that the channel propagation path in the uplink and downlink directions are likely to go through the same set of dominant scatterers, hence the AoA and AoD (for up and downlink) at the BS are similar. For the temporal correlation, the channel of each user k is modeled by a block fading Gauss-Markov process with time correlation parameter $\tau_k \in [0,1]$ and coherence period $N_c T_s$ as

$$\boldsymbol{h}_k[n, t + N_c T_s] = \tau_k \boldsymbol{h}_k[n,t] + \sqrt{1 - \tau_k^2}\,\boldsymbol{\delta}_k[n,t], (5)$$

where $\boldsymbol{\delta}_k[n,t]$ contains independent and identically distributed

[1] The proposed scheme works for arbitrary number of MS antennas, although only single antenna is considered here for simplicity of exposition.

(i.i.d.) elements according to a complex normal distribution $\mathcal{CN}(\boldsymbol{0}, \boldsymbol{R}_k)$, and it is uncorrelated with $\boldsymbol{h}_k[n,t]$. The coefficient τ_k depends on the Doppler frequency $f_{d,k}$ (which in turn depends on the mobility speed of kth MS) and coherence interval $N_c T_s$ according to Jake's model [13] as follows

$$\tau_k = J_0\left(2\pi f_{d,k} N_c T_s\right), \quad (6)$$

where $J_0(\cdot)$ is the zero order Bessel function of the first kind.

III. Proposed Scheme

At every scheduling interval, the BS makes use of the estimated correlation matrix $\boldsymbol{R}_k$ to perform user scheduling. The idea is to group the users whose $\boldsymbol{R}_k$ dominant eigenspaces are most mutually orthogonal. Each group is then assigned to a disjoint set of subcarriers. Let $\boldsymbol{R}_k$ be decomposed into

$$\boldsymbol{R}_k = \boldsymbol{U}_k \boldsymbol{\lambda}_k \boldsymbol{U}_k^H,$$

where $\boldsymbol{\lambda}_k = \text{diag}\left[\{\lambda_{k,i}\}_{i=1}^{\text{rank}\{\boldsymbol{R}_k\}}\right]$ is the diagonal eigenvalues of $\boldsymbol{R}_k$ arranged in a decreasing order. Any channel realization $\boldsymbol{h}_k$ can then be expressed as

$$\boldsymbol{h}_k = \boldsymbol{R}_k^{1/2}\tilde{\boldsymbol{h}}_k = \boldsymbol{U}_k\boldsymbol{\lambda}_k^{1/2}\tilde{\boldsymbol{h}}_k = \sum_i \boldsymbol{u}_{k,i}\lambda_{k,i}^{1/2}\tilde{h}_{k,i}, \quad (7)$$

where $\boldsymbol{u}_{k,i}$ and $\tilde{h}_{k,i}$ denote the ith column of $\boldsymbol{U}_k$ and the ith element of $\tilde{\boldsymbol{h}}_k$, respectively. Here, $\tilde{\boldsymbol{h}}_k$ is distributed according to $\mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$. Using the D_k-dominant eigenvectors to construct the beamforming vector

$$\begin{aligned}\boldsymbol{v}_k &= [\boldsymbol{u}_{k,1}, \cdots, \boldsymbol{u}_{k,D_k}][w_{k,1}, \cdots, w_{k,D_k}]^T, \\ &\triangleq \underline{\boldsymbol{U}}_k \boldsymbol{w}_k\end{aligned}$$

where $\boldsymbol{w}_k$ is the $D_k \times 1$ dimensional *steering vector* with unit norm $\|\boldsymbol{w}_k\|^2 = 1$, the BS transmit vector and k^{th} MS received signal are given by

$$\boldsymbol{x}^{(BS)} = \sum_{k \in \mathcal{B}_{n,t}} \boldsymbol{v}_k s_k = \sum_{k \in \mathcal{B}_{n,t}} \underline{\boldsymbol{U}}_k \boldsymbol{w}_k s_k, \quad (8)$$

$$y_k^{(MS)} = \boldsymbol{h}_k^H \boldsymbol{x}^{(BS)} + z_k \quad (9)$$

$$= \underbrace{[\lambda_{k,1}^{1/2}\tilde{h}_{k,1}^H, \cdots, \lambda_{k,D_k}^{1/2}\tilde{h}_{k,D_k}^H]}_{\triangleq (\underline{\boldsymbol{\lambda}}_k^{1/2}\underline{\tilde{\boldsymbol{h}}}_k)^H}\boldsymbol{w}_k s_k + \sum_{k' \in \mathcal{B}_{n,t}, k' \neq k}\boldsymbol{h}_k^H \underline{\boldsymbol{U}}_{k'}\boldsymbol{w}_{k'}s_{k'} + z_k.$$

In the above, the first term is the desired signal, while the second term is the interference. It is assumed that the transmit power are distributed equally to all users, i.e., $E[|s_k|^2] = P/K, \forall k$. The motivation to choose the user group with the most orthogonal $\boldsymbol{R}_k$ eigenspaces is to limit the interference, as in this case $\boldsymbol{h}_k^H \underline{\boldsymbol{U}}_{k'}$ will be small for $k \neq k'$. The maximum transmission rate, which is defined by the mutual information between $y_k^{(MS)}$ and s_k is given by

$$\begin{aligned}r_k &\leq \mathcal{I}(y_k^{(MS)}; s_k) \\ &= \log_2\left(1 + \frac{|\underline{\tilde{\boldsymbol{h}}}_k^H \underline{\boldsymbol{\lambda}}_k^{1/2}\boldsymbol{w}_k|^2}{\sum\limits_{k' \in \mathcal{B}_{n,t}, k' \neq k}|\boldsymbol{h}_k^H \underline{\boldsymbol{U}}_{k'}\boldsymbol{w}_{k'}|^2 + \frac{N_0 K}{P}}\right). (10)\end{aligned}$$

The above maximum rate is achieved by using independent Gaussian signaling on each user.

In order to implement the above scheme, we need to overcome several challenges, e.g., obtaining the correlation matrix $\boldsymbol{R}_k$ information at the transmitter to perform the PCA, user scheduling and dominant subspace dimension D_k selection, tracking the variation of the principal channel coefficients over time, etc. The following three subsections address each of these issues separately.

A. Spatial Correlation Matrix Estimation

In order to estimate the spatial correlation matrix $\boldsymbol{R}_k$, we dedicate some of the uplink resources for this purpose. Since in SC-FDMA the uplink channel of different users occupy non-overlapping carriers, there is no inter-user interference within the cell, hence we will focus the discussion only on user k.

Denoting the time-frequency slots dedicated for correlation training of user k as $\mathcal{P}_k \subset \left(\mathcal{F}_U^{(k)}, \{t - N_p T_s, \cdots, t - T_s\} \right)$, and setting[2] $x_k^{(MS)}[n, t] = 1, \forall (n, t) \in \mathcal{P}_k$, equation (2) gives

$$\boldsymbol{y}^{(BS)}[n, t] = \boldsymbol{g}_k[n, t] + \boldsymbol{z}[n, t], \qquad (n, t) \in \mathcal{P}_k. \quad (11)$$

Since the spatial correlation is a second order statistics which changes slowly, we can allocate a larger time scale than the coherence period ($N_p > N_c$) for $\mathcal{P}_k$, as $\boldsymbol{R}_k$ is relatively constant during this period. The problem is now to determine $\boldsymbol{R}_k$ defined in (4) from the noisy observations in (11). The simplest way to do this is to use the sample covariance as

$$\hat{\boldsymbol{R}}_k = \frac{1}{|\mathcal{P}_k|} \sum_{(n, t) \in \mathcal{P}_k} \boldsymbol{y}^{(BS)}[n, t] \, (\boldsymbol{y}^{(BS)}[n, t])^H \; - \; N_0 \boldsymbol{I},$$

where $|\cdot|$ denotes the cardinality of the set. Although convenient and it provides an unbiased estimate of $\boldsymbol{R}_k$, there are several problems associated with the above approach. The first is the high computational complexity and memory requirement to derive from $\boldsymbol{R}_k$ the eigenvalues and the eigenvectors, due to its large dimension. The second problem lies with the well-known over dispersion in the sample covariance, whereby the obtained eigenvalues tend to be larger than that of the actual population covariance. These problems are well known in the field of subspace learning for pattern recognition, data mining, as well as machine learning. Several approaches such as shrinkage estimation [14], spectral regression [15], and Nyström method [16] have been proposed to deal with these issues. In this paper, we adopt the Nyström method due to its low complexity and desirable shrinkage property. The main idea is to approximate $\hat{\boldsymbol{R}}_k$ as

$$\hat{\boldsymbol{R}}_{k,NYS} = \begin{bmatrix} \hat{\boldsymbol{R}}_{k,11} & \hat{\boldsymbol{R}}_{k,12} \\ \hat{\boldsymbol{R}}_{k,12}^H & \hat{\boldsymbol{R}}_{k,12}^H \hat{\boldsymbol{R}}_{k,11}^{-1} \hat{\boldsymbol{R}}_{k,12} \end{bmatrix}, \quad (12)$$

which is a low-rank approximation of the sample covariance

$$\hat{\boldsymbol{R}}_k = \begin{bmatrix} \hat{\boldsymbol{R}}_{k,11} & \hat{\boldsymbol{R}}_{k,12} \\ \hat{\boldsymbol{R}}_{k,12}^H & \hat{\boldsymbol{R}}_{k,22} \end{bmatrix}. \quad (13)$$

[2]Setting the pilot symbols as 1 is sufficient in this case as there is no inter-user interference in the uplink SC-FDMA within the cell.

In the above, $\hat{\boldsymbol{R}}_{k,11}$ is the $d \times d$ sub-matrix of $\hat{\boldsymbol{R}}_k$, where d is chosen to approximate the rank of $\boldsymbol{R}_k$. The complexity of calculating the eigenvectors and the corresponding non-zero eigenvalues using Nyström method is $\mathcal{O}(M^2)$, which is smaller than $\mathcal{O}(M^3)$ using the sample covariance [16].

Denoting all the observations in (11) available for this estimation as $\boldsymbol{Y}^{(k)} = \left[\boldsymbol{y}^{(BS)}[\mathcal{P}_k] \right]$, which is an $M \times |\mathcal{P}_k|$-dimensional matrix, the algorithm using Nyström approach is described as follows.

Algorithm 1 Spatial Correlation Estimation for User k

INPUT: $\boldsymbol{Y}^{(k)}$, N_0, and $d \approx \text{rank}\{\boldsymbol{R}_k\}$ from the earlier estimate.
OUTPUT: Eigenmatrix $\hat{\boldsymbol{U}}_k$ and eigenvalues $\hat{\boldsymbol{\lambda}}_k^{\frac{1}{2}}$ of $\hat{\boldsymbol{R}}_{k,NYS}$.

1: Randomly select d indexes $\mathcal{S} \subset \{1, \cdots, M\}$, and denote its complement by $\overline{\mathcal{S}}$.
2: Denote the row sub-matrices $\boldsymbol{A} = \boldsymbol{Y}_{\mathcal{S},:}^{(k)}$ and $\boldsymbol{B} = \boldsymbol{Y}_{\overline{\mathcal{S}},:}^{(k)}$.
3: Perform the thin Singular Value Decomposition (SVD) on $\boldsymbol{A}$ as $\boldsymbol{A} = \boldsymbol{U}_A \boldsymbol{D}_A \boldsymbol{V}_A^H$.
4: Construct the $M \times d$ matrix $\boldsymbol{W} = \begin{bmatrix} \frac{1}{\sqrt{|\mathcal{P}_k|}} \boldsymbol{U}_A \boldsymbol{D}_A \\ \frac{1}{\sqrt{|\mathcal{P}_k|}} \boldsymbol{B} \boldsymbol{V}_A \end{bmatrix}$.
5: Perform the thin SVD on $\boldsymbol{W}$ as $\boldsymbol{W} = \boldsymbol{U}_W \boldsymbol{D}_W \boldsymbol{V}_W^H$
6: Obtain the results $\hat{\boldsymbol{U}}_k = \boldsymbol{U}_W$ and $\hat{\boldsymbol{\lambda}}_k^{\frac{1}{2}} = \boldsymbol{D}_W - \sqrt{N_0} \boldsymbol{I}$.

The above algorithm uses random uniform selection of subset $\mathcal{S}$, which is shown to perform well in [16]. Since $\boldsymbol{R}_k$ is likely to be low rank, the parameter d and therefore the complexity of Algorithm 1 can be kept low.

B. Channel Dimension Reduction and User Scheduling

Considering that the antennas are placed close to one another in massive MIMO, the channel gain tends to be correlated. As a result, the spatial correlation matrix $\boldsymbol{R}_k$ will have low rank. Therefore, following the PCA decomposition in (7), and the fact that the eigenvalues $\boldsymbol{\lambda}_k$ are concentrated on its first few diagonal elements, only some elements of $\tilde{\boldsymbol{h}}_k$ is sufficient to describe the channel, hence reducing the effective channel dimension.

Using only D_k elements to describe $\boldsymbol{h}_k$ as $\underline{\boldsymbol{U}}_k \underline{\boldsymbol{\lambda}}_k^{\frac{1}{2}} \underline{\tilde{\boldsymbol{h}}}_k$, the Mean Square Error (MSE) due to this approximation is

$$E[|\boldsymbol{h}_k - \underline{\boldsymbol{U}}_k \underline{\boldsymbol{\lambda}}_k^{\frac{1}{2}} \underline{\tilde{\boldsymbol{h}}}_k|^2] = \begin{cases} 0, & \text{when } D_k \geq \text{rank}\{\boldsymbol{R}_k\}, \\ \sum_{i > D_k} \lambda_{k,i}, & \text{otherwise.} \end{cases} \quad (14)$$

While larger D_k is able to represent $\boldsymbol{h}_k$ accurately, smaller D_k is desirable for better tracking and multiuser scheduling performance. In this work, D_k is chosen based on the tolerable normalized MSE level μ, according to

$$D_k = \min \nu, \text{ s.t. } \sum_{j > \nu} \lambda_{k,j} < \mu \sum_i \lambda_{k,i}, \quad (15)$$

and the choice of μ is subject to optimization.

As far as the user scheduling is concerned, the optimal solution requires prohibitive complexity [5], hence we resort to propose a greedy method as described in Algorithm 2.

Algorithm 2 Greedy User Scheduling

INPUT: Current estimates of $\boldsymbol{U}_k$, $\boldsymbol{\lambda}_k$, and $\tilde{\underline{\boldsymbol{h}}}_k$; D_k, $\mathcal{F}_D$.
OUTPUT: $\mathcal{B}_{n,t}, \forall n \in \mathcal{F}_D$ and $\boldsymbol{w}_k, \forall k$.

1: Initialize: $\mathcal{K} = \{1..K\}$, $\mathcal{B}_{n,t} = \emptyset$ for $n \in \mathcal{F}_D$, $\boldsymbol{w}_k = \frac{\boldsymbol{\lambda}_k^{\frac{1}{2}} \tilde{\underline{\boldsymbol{h}}}_k}{\|\boldsymbol{\lambda}_k^{\frac{1}{2}} \tilde{\underline{\boldsymbol{h}}}_k\|}$.

2: **while** $\mathcal{K} \neq \emptyset$ **do**

3:   For each $n \in \mathcal{F}_D$, search for the best user j_n^* as follows
$$j_n^* = \arg\max_{j \in \mathcal{K}} \Delta R_n(j),$$
   where:
$$\Delta R_n(j) \triangleq \sum_{k \in \{\mathcal{B}_{n,t}, j\}} \log_2(1 + \frac{|\tilde{\underline{\boldsymbol{h}}}_k^H \boldsymbol{\lambda}_k^{\frac{1}{2}} \boldsymbol{w}_k|^2}{\sum_{k' \in \{\mathcal{B}_{n,t}, j\}, k' \neq k} |\tilde{\underline{\boldsymbol{h}}}_k^H \boldsymbol{\lambda}_k^{\frac{1}{2}} \boldsymbol{U}_k \boldsymbol{U}_{k'} \boldsymbol{w}_{k'}|^2 + \frac{N_0 K}{P}}) -$$
$$\sum_{k \in \mathcal{B}_{n,t}} \log_2(1 + \frac{|\tilde{\underline{\boldsymbol{h}}}_k^H \boldsymbol{\lambda}_k^{\frac{1}{2}} \boldsymbol{w}_k|^2}{\sum_{k' \in \mathcal{B}_{n,t}, k' \neq k} |\tilde{\underline{\boldsymbol{h}}}_k^H \boldsymbol{\lambda}_k^{\frac{1}{2}} \boldsymbol{U}_k \boldsymbol{U}_{k'} \boldsymbol{w}_{k'}|^2 + \frac{N_0 K}{P}}).$$

4:   Find the best subcarrier n^* with the largest rate increase
$$n^* = \arg\max_{n \in \mathcal{F}_D} \Delta R_n(j_n^*).$$

5:   Update: $\mathcal{B}_{n^*,t} = \mathcal{B}_{n^*,t} \cup \{j_{n^*}^*\}$, $\mathcal{K} = \mathcal{K} \setminus \{j_{n^*}^*\}$.

6: **end while**

The estimate of $\tilde{\underline{\boldsymbol{h}}}_k$ above is initially set to any arbitrary value, and subsequently updated by the channel tracking algorithm described in the next subsection, via the relation $\boldsymbol{w}_k = \frac{\boldsymbol{\lambda}_k^{\frac{1}{2}} \tilde{\underline{\boldsymbol{h}}}_k}{\|\boldsymbol{\lambda}_k^{\frac{1}{2}} \tilde{\underline{\boldsymbol{h}}}_k\|}$. Compared to the approach in [5] which is based on user clustering, each step of the iteration in our greedy algorithm searches for the best user and subcarrier pair that directly maximizes the sum-rate. So far, we only consider conjugate beamforming for the steering vectors $\boldsymbol{w}_k$, which is optimal when only one user is allocated on each subcarrier. Furthermore, the sum-rate is calculated under the assumption that the receivers treat interference as noise. By allowing general steering vectors and more sophisticated multiuser decoding at the receivers, the overall sum-rate can be increased further, but this consideration is outside the scope of this work.

C. Principal Channel Coefficient Tracking

In order to perform beamforming, the knowledge of the actual channel realization is needed. Therefore, we still need to track the variation of the reduced-dimension channel vectors $\underline{\boldsymbol{h}}_k$ over time, and adapt the steering vector $\boldsymbol{w}_k$ accordingly.

In the absence of interference, the sum-rate maximizing beamforming design problem can be formulated as
$$\max_{\{\boldsymbol{w}_k\}_{k=1}^K} \sum_k \log_2\left(1 + \frac{P}{N_0 K}\left|\tilde{\underline{\boldsymbol{h}}}_k^H \boldsymbol{\lambda}_k^{\frac{1}{2}} \boldsymbol{w}_k\right|^2\right) \quad (16)$$
$$\text{s.t.} \quad \|\boldsymbol{w}_k\|^2 = 1, \forall k,$$

the solution of which is given by the conjugate beamforming, which corresponds to a steering vector of $\boldsymbol{w}_k = \frac{\boldsymbol{\lambda}_k^{\frac{1}{2}} \tilde{\underline{\boldsymbol{h}}}_k}{\|\boldsymbol{\lambda}_k^{\frac{1}{2}} \tilde{\underline{\boldsymbol{h}}}_k\|}$. This choice maximizes the desired signal power. In the presence of interference (for the case when more than one user is allocated to the same time-frequency slot), maximizing the desired signal power is not necessarily the best option, even though it often gives good result, especially when the interference link gain is small. Here, we propose to use a probing signal and simple energy comparator at the MS for channel tracking.

At the BS, Q probing vectors $\{\boldsymbol{p}_i\}_{i=1}^Q$ are generated ran-

domly. The size Q is selected based on the amount of feedback ($\log_2 Q$ bits) available. The BS then allocates Q downlink time slots for user k to transmit the pilot symbols ($s_k = 1$) using the perturbed beamforming weight $\boldsymbol{v}_k = \underline{\boldsymbol{U}}_k \frac{\boldsymbol{w}_k + \boldsymbol{p}_i}{|\boldsymbol{w}_k + \boldsymbol{p}_i|}, \forall 1 \leq i \leq Q$, where $\boldsymbol{w}_k$ is the current steering vector obtained in the previous tracking period. The MS will then measure the received energy on these Q training slots, and select the one with the largest energy. Given that all other terms are fixed, it is apparent from the received signal $y_k^{(MS)}$ in (9) that this amounts to finding
$$i_k^* = \arg\max_{1 \leq i_k \leq Q} \left|\tilde{\underline{\boldsymbol{h}}}_k^H \boldsymbol{\lambda}_k^{\frac{1}{2}} \frac{\boldsymbol{w}_k + \boldsymbol{p}_{i_k}}{|\boldsymbol{w}_k + \boldsymbol{p}_{i_k}|}\right|^2, \quad (17)$$

and the best perturbation vector is the one that produces the steering vector $\frac{\boldsymbol{w}_k + \boldsymbol{p}_{i_k^*}}{|\boldsymbol{w}_k + \boldsymbol{p}_{i_k^*}|}$ closest to the $\frac{\boldsymbol{\lambda}_k^{\frac{1}{2}} \tilde{\underline{\boldsymbol{h}}}_k}{\|\boldsymbol{\lambda}_k^{\frac{1}{2}} \tilde{\underline{\boldsymbol{h}}}_k\|}$. This optimal index i_k^* will then be modulated and sent back to the BS via feedback link (assumed to be error free for simplicity), and the BS updates $\boldsymbol{w}_k^* = \frac{\boldsymbol{w}_k + \boldsymbol{p}_{i_k^*}}{|\boldsymbol{w}_k + \boldsymbol{p}_{i_k^*}|}$ accordingly. Using this approach, channel tracking is achieved without the actual channel estimation at either BS or MS. Furthermore, only energy comparator and index selection is needed at the MS, resulting in low implementation complexity.

Compared to the scheme in [11], the tracking in our approach is performed on the steering vector $\boldsymbol{w}_k \in \mathbb{C}^{D_k}$ rather than the beamforming vector $\boldsymbol{v}_k \in \mathbb{C}^M$ itself. This significantly reduces the dimension of the perturbation vector, which improves the tracking performance and makes it applicable for massive MIMO. Moreover, as the perturbation does not change the subspace spanned by the beamforming vector, we do not need to apply Gram-Schmidt orthonormalization process as in [11], hence reducing the overall complexity. In summary, the tracking algorithm is described as follows.

Algorithm 3 Channel Tracking for User k

INPUT: Q, D_k, and $\boldsymbol{w}_k$ from the earlier tracking period.
OUTPUT: New steering vector $\boldsymbol{w}_k'$.

1: **for** each tracking period **do**

2:   BS generates Q random perturbation vectors $\{\boldsymbol{p}_i\}_{i=1}^Q$, each having dimension $D_k \times 1$ according to $\mathcal{CN}(0, \boldsymbol{I})$.

3:   BS allocates Q downlink training slots and transmits the perturbed beamforming pilot $\underline{\boldsymbol{U}}_k \frac{\boldsymbol{w}_k + \boldsymbol{p}_i}{|\boldsymbol{w}_k + \boldsymbol{p}_i|}$ on each slot.

4:   At the MS k, measure the received signal power at the Q training slots $|y_k^{(MS)}|^2$, and find the largest one as
$$i_k^* = \arg\max_{1 \leq i_k \leq Q} |y_k^{(MS)}|^2.$$

5:   MS k sends the i_k^* information back to BS via feedback channel.

6:   BS updates the new steering vector as $\boldsymbol{w}_k' = \frac{\boldsymbol{w}_k + \boldsymbol{p}_{i_k^*}}{|\boldsymbol{w}_k + \boldsymbol{p}_{i_k^*}|}$.

7: **end for**

During the initial period when no prior estimate of $\boldsymbol{w}_k$ is available, such as the case where the user has just joined the network, an initial value of $\boldsymbol{w}_k = \frac{1}{\sqrt{D_k}} \boldsymbol{1}$ can be used.

Note: Since the received signal energy is used as the criteria for selecting the best perturbation vector, the probing on those subcarriers where more than one users are served

simultaneously (where $|\beta_{n,t}| > 1$) should be done sequentially on each of its members. As far as the tracking period is concerned, a good choice is to set it proportional to the channel coherence period. In the highly mobile scenario where the channel varies very rapidly, a limiting choice of $Q = 2$ shall be used. In this case, only one bit feedback per user is needed at the expense of tracking performance. Therefore, there is an interrelation between the feedback size, tracking performance, training overhead, and Doppler spread of the channel, which are subject to further investigation.

In summary, the proposed transmission strategy can be described as follows

1) BS collects the noisy observations on the uplink time-frequency slots $\mathcal{P}_k$ for $k \in \{1, \cdots, K\}$. This update is performed on a long timescale of N_p symbol periods.
2) BS estimates the eigenmatrix $\boldsymbol{U}_k$ and the corresponding eigenvalues $\boldsymbol{\lambda}_k$ of the spatial correlation matrix for all users using Algorithm 1, and utilize them for user scheduling and beamforming.
3) For each scheduling interval, BS allocates resources according to Algorithm 2, and transmits using the current estimate of $\boldsymbol{U}_k$, $\boldsymbol{\lambda}_k$, and $\boldsymbol{w}_k$. At the same time, some time-frequency slots are dedicated for tracking purposes, where BS sends randomly perturbed vector $\frac{\boldsymbol{w}_k + \boldsymbol{p}_{i_k}}{|\boldsymbol{w}_k + \boldsymbol{p}_{i_k}|}$.
4) Each MS receives the intended messages on its allocated time-frequency resources, and also checks the received signal energy on the dedicated time-frequency slots. By feeding back the best index to the BS, the steering vector $\boldsymbol{w}_k$ can be updated following Algorithm 3.

IV. Simulation Results

To demonstrate the performance of the proposed scheme, we consider a cellular system with one base station at the center of the cell of radius 1 Km, equipped with $M = 100$ antennas arranged into a Uniform Linear Array (ULA) with half a wavelength antenna separation, serving $K = 20$ users simultaneously. The users are placed uniformly within the cell coverage area, and they are traveling on a straight line at 60 Km/hr, which is a typical vehicular mobility speed. Each user independently chooses the traveling direction at random, and their movements are confined within the cell coverage area.

The transmit spatial correlation is assumed to follow one ring model with scatterer radius of $r = 10$ meters, such that the correlation coefficient between two antenna elements at location (x_1, y_1) and (x_2, y_2) with respect to a user located at azimuth angle θ and distance s from the BS (having an angle spread of $\Delta \approx \arctan(r/s)$) is given by [17]

$$\frac{1}{2\Delta} \int_{-\Delta}^{\Delta} e^{-j\frac{2\pi}{\lambda}((x_2-x_1)\cos(\theta+\alpha)+(y_2-y_1)\sin(\theta+\alpha))} d\alpha. \quad (18)$$

The other simulation parameters are chosen following the 3GPP LTE specifications as follows. We consider 2.6 GHz carrier frequency with 5 MHz downlink bandwidth. The FFT size is set to 512 with 301 usable subcarriers (including one DC component), which are arranged into 25 resource blocks.

TABLE I
SIMULATION PARAMETERS

Coverage model	1 km-radius circular area
Tx/Rx antennas	100 Tx ULA/1 Rx antenna
Number of users	20 uniformly distributed
Delay profile	5-taps doubly exponential
FFT size	512 subcarriers
Correlation model	One-ring model
Bandwidth	5 MHz
Normalized MSE threshold	$\mu = 0.15$
Feedback bits	8 bits

Resource allocation to the users is done in terms of resource blocks, each of which spans 12 subcarriers over one subframe (spanning 14 OFDM symbols). In LTE, the Control Channel Element (CCE) limits the total number of users that can be scheduled on each subframe. In the ideal case, for 5 MHz bandwidth there can be up to 20 users, each allocated to different resource blocks (each user can have more than one resource blocks allocated). However, higher code rate is often required for encoding the CCE in order to protect it against channel impairments, hence reducing the total number of simultaneously supported user. In our proposed scheme, instead of exclusively assigning each resource block to a particular user, we allow multiple users to share the same group of resource blocks. Furthermore, instead of performing the user scheduling on a subcarrier basis, Algorithm 2 performs the allocation based on the groups of resource blocks, which is taken to be 10 in this simulation.

To model the frequency correlation, we use 5-taps doubly exponential power delay profile with exponentially distributed delay and power decay such as that used in [18]. For the time correlation, the coherence time is chosen to be equal to the subframe duration ($N_c = 14$), which is approximately 1 msec. For 60 Km/hr mobile speed with 2.6 GHz carrier frequency, this corresponds to a temporal correlation coefficient of 0.83 between subsequent resource blocks. For simplicity, the user scheduling is performed on every 20 subframes, although any scheduling period (e.g. 1 subframe in typical LTE systems) can be used. The update to the estimated spatial correlation matrix $\boldsymbol{R}_k$ is performed every $N_p = 300$ subframes. This choice can be justified since the location displacement for a mobile traveling at 60 Km/hr over 300 msec is only 5 meters, which is sufficiently small for the correlation matrix $\boldsymbol{R}_k$ to experience only minor variation. The simulation parameters are summarized in Table I. In order to evaluate the average performance, $50,000$ OFDM symbols are transmitted to all K users according to the proposed scheduling algorithm. The available feedback bits for channel tracking is set to 8 bits. System sum rate is used as the performance metric, and the accuracy of spatial correlation estimation, the effectiveness of user scheduling, as well as the channel tracking are analyzed.

Figure 1 shows the comparison between the average system sum rate achieved with channel tracking enabled and without channel tracking. It is apparent that the perturbation-based channel tracking as described in Algorithm 3 is able to achieve
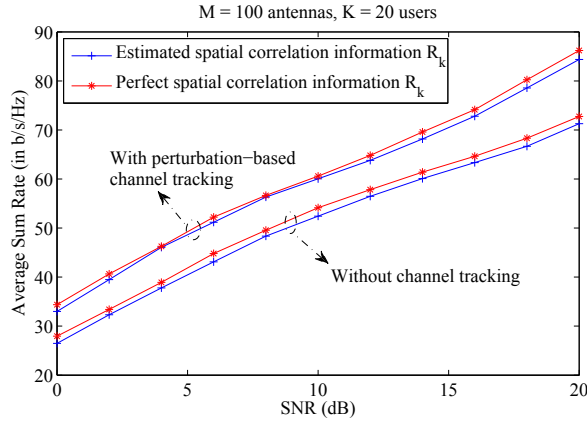
Fig. 1.   Channel tracking and correlation estimation performance

Fig. 2.   User scheduling performance

significant improvement (approximately 8 to 10 b/s/Hz in total sum-rate). Figure 1 also shows the small gap between the average sum rate achieved assuming perfect knowledge of spatial correlation matrix $\boldsymbol{R}_k$ and the average sum rate when the correlation matrix has to be estimated from the noisy observation. The small gap between these two shows that the proposed spatial correlation estimation using Nyström method as described in Algorithm 1 is able to get a good estimate of $\boldsymbol{R}_k$. This is achieved despite the significant complexity reduction compared to the sample covariance approach

Figure 2 shows the performance of the greedy scheduling algorithm. Without scheduling, the users are allocated a group of resource blocks in a round-robin manner. As such, the performance is interference limited, especially when the channels of the users sharing the same group of resource blocks spans similar subspace. This causes the average sum-rate to saturate in high SNR. With greedy scheduling, each user is allocated a group of resource block as long as the interference it introduce is minimal. Implicitly, this scheme allocates resource block to a group of users only when their channels are as mutually orthogonal as possible. Therefore, the saturation effect in high SNR can be avoided.

V. CONCLUSIONS

A transmission strategy for Massive MIMO downlink FDD systems in mobile multiusers environment is proposed. The scheme reduces the dimensionality of the channel by considering only its principal components, and the perturbation based tracking is used to track its variation. By exploiting the WSSUS property of the channel, low complexity algorithm to estimate the eigenmatrix and the eigenvalues of the channel statistics, which are necessary for the PCA, is used. Simulation results show that significant sum-rate improvement is achieved by the tracking mechanism, and that good scheduling performance is able to avoid the sum-rate saturation at high SNR. Future works include consideration of a more sophisticated beamforming design and the corresponding scheduling strategy, together with the trade-off analysis of the feedback reduction and system performance. Optimization of the MSE threshold and the design of structured perturbation vectors are also to be further analyzed.
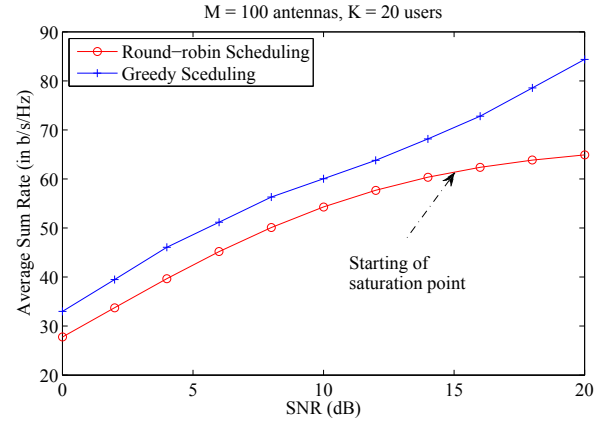
REFERENCES

[1] T. L. 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.

[2] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprashad, "Achieving massive MIMO spectral efficiency with a not-so-large number of antennas," *IEEE Trans. on Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, Sept. 2012.

[3] "Greentouch," [online], http://www.greentouch.org/.

[4] B. Hochwald and S. Vishwanath, "Space-time multiple access: Linear growth in the sum rate," *IEEE Annu. Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, USA, pp. 387–398, Oct. 2002.

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

[6] E. Bjornson, M. Kountouris, and M. Debbah, "Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination," *20th International Conf. on Telecommun.*, Casablanca, Morocco, pp. 1–5, May 2013.

[7] K. Kim, T. Kim, D. J. Love, and I. H. Kim, "Differential feedback in codebook-based multiuser MIMO systems in slowly varying channels," *IEEE Trans. on Commun.*, vol. 60, no. 2, pp. 578–588, Feb. 2012.

[8] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas in Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[9] J. Joung and S. Sumei, "SCF: Sparse channel-state-information feedback using Karhunen-Loeve transform," *IEEE Global Communication Conference 2014*, Austin, Texas, USA, to appear, Dec. 2014.

[10] A. Goldsmith, *Wireless Communications*, Cambridge, NY, Cambridge University Press, 2005.

[11] B. C. Banister and J. R. Zeidler, "Feedback assisted transmission subspace tracking for MIMO systems," *IEEE Journal on Selected Areas in Commun.*, vol. 21, no. 3, pp. 452–463, April 2003.

[12] 3GPP TS 36.211, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation," [online], http://www. 3gpp. org/DynaReport/36211.htm.

[13] W. C. Jakes, *Microwave Mobile Communications*, Wiley, NY, 1994.

[14] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high dimensional covariance matrices," *IEEE Trans. on Signal Processing*, vol. 59, no. 9, pp. 4097–4107, Sept. 2011.

[15] D. Cai, X. He, and J. Han, "Spectral regression: A unified approach for sparse subspace learning," *IEEE Int. Conf on Data Mining*, Omaha, Nebraska, USA, pp. 73–82, Oct. 2007.

[16] N. Arcolano and P. J. Wolfe, "Estimating principal components of large covariance matrices using the Nyström method," *IEEE ICASSP 2011*, Prague, Czech Republic, pp. 3784–3787, May 2011. [online extended version at]: http://arxiv.org/abs/1111.6926v2.

[17] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. on Commun.* vol. 48, no. 3, pp. 502–513, March 2000.

[18] 3GPP TS 36.873, "3D-Channel Model for LTE," [online], http:// www.3gpp.org/dynareport/36873.htm.