IEEE Wireless Communications Letters Volume: 8 , Issue: 4 , Aug 2019



# An Approach to Reduce the Overhead of Training Sequences in FDD Massive MIMO Downlink Systems

Haiquan Wang, *Member, IEEE*, Guobin Li, Sisi Zheng, Shaoshi Yang, *Member, IEEE*, and Peng Pan, *Member, IEEE*

***Abstract*—In frequency-division duplexing (FDD) massive multiple-input multiple-output (MIMO) systems, a large number of antennas are equipped at the base station (BS). In order to accurately estimate the downlink channel state information (CSI), a significant portion of time slots are usually invoked to send training sequences. To reduce the training time slots, we propose a dimension reduction method, which uses a matrix to divide and combine the transmit antennas into several groups. Noticed that the matrix is only related to the parameters of system and independent of CSI. As a beneficial result, a new equivalent system is created, where the training time slots can be decreased dramatically. Simulation results show that the new equivalent system can provide higher throughput than the original system because of the dimension reduction.**

***Index Terms*—Massive MIMO, downlink, overhead, training sequence, dimension reduction**

## I. Introduction

Compared to the conventional multiple-input multiple-output (MIMO) systems, a massive MIMO system is capable of providing services for more users simultaneously on the same time-frequency resource blocks, thereby greatly improving the spectrum efficiency [1]. Furthermore, in the context of massive MIMO, the simple linear precoding and linear detectors become near-optimal, and both noise and uncorrelated interference become minor.

The above benefits are based on the assumption that the accurate channel state information (CSI) is available to the transmitter or the receiver. For a frequency-division duplexing (FDD) massive MIMO system, in order to obtain the downlink CSI at the base station (BS), the BS needs to send training sequences to users, and then each user feeds the estimated CSI back to the BS. Existing research shows that in order to obtain accurate channel estimation, the length of the training sequences should be no less than the number of transmit antennas [2]- [5]. Thus, the overhead of training sequences is proportional to the number of antennas at the BS. Therefore, in massive MIMO systems, the BS has to expend a significant fraction of the limited coherent time resource to transmit the training sequences.

How to reduce the overhead of training sequences in massive MIMO systems is an important research topic. The authors of [6] first demonstrated the capacity limit of multi-antenna systems, that use training sequence based channel estimation. The authors of [7] proposed a design that uses time correlation to predict the channel, thereby reducing the overhead of training sequences. In [8], [9], under the assumption of no feedback, the concept of low-dimensional space-time code was proposed, and different training sequences along with different space-time block codes were studied. In [10], a generalized low-complexity beamspace approach is proposed, and the received signal vectors in the antenna-element space are transformed into the beamspace by employing beamforming vectors. In [11], a antenna group beamforming algorithm is proposed, which can reduce the overhead of CSI feedback. However, the method of [11] unable to reduce the cost of training sequences because the structure of grouping matrix is related to CSI.

In this paper, based on the basic parameters of the system and the statistic properties of the channel only, a dimension reduction method for reducing the overhead of training sequences is proposed. More specifically, we design a fixed transformation matrix, say $\mathbf{W}$, to divide and combine the $M$ transmit antennas into $N$ groups, and the training sequences can be designed for each group, not for each antenna. Therefore, on one hand, the signals of multiple antennas in each group are combined into a single beam, and the spatial gains of multiple individual antennas are effectively integrated. On the other hand, as a beneficial result of our scheme, the length of the training sequences can be reduced from $M$ to $N$, while the total throughput of the system can even be increased in contrast to that of the system without using the matrix $\mathbf{W}$, based on the optimal design of number $N$. The optimal $N$ can be obtained by numerical calculations to maximize the total throughput. The proposed method is also effective in limited feedback MIMO systems. It should be noticed that the selection of matrix $\mathbf{W}$ is independent of the CSI. It is only related to the number of the transmitting antenna and the statistical characteristics of the channel. Our simulation results show that compared with the method without dimension reduction, the proposed method can effectively reduce the overhead of training sequences and improve the performance of massive MIMO systems.

This work is supported by National Natural Science Foundations of China under grant no. U1709220, and 61372093. *(Corresponding author: Peng Pan.)*

H. Wang, G. Li, S. Zheng and P. Pan are with School of Communications Engineering, Hangzhou Dianzi University, Hangzhou 310018, China. (e-mail: *tx_wang*@hdu.edu.cn, *ligbgb*@163.com, *zhengss1994*@163.com, *panpeng*@hdu.edu.cn).

S. Yang is with Beijing University of Posts and Telecommunications, Beijing 100876, and was with Huawei Technologies Co., Ltd., China. (e-mail: *shaoshi.yang*@ieee.org)

This paper is organized as follows. In Section II, the system model is given. In Section III, the proposed dimensionality reduction method is presented. Our simulation results are shown in Section IV, and the conclusions are offered in Section V.

## II. System Model

We consider a massive MIMO downlink system, where the BS is equipped with $M$ transmit antennas and serves $K$ randomly distributed single-antenna users. Then, the training signals and data signals received by the $k$-th user can be expressed as

$$\mathbf{y}_{kp} = \sqrt{\rho}\mathbf{h}_k^H\mathbf{X}_p + \mathbf{z}_{kp} \tag{1}$$

and

$$\mathbf{y}_{kd} = \sqrt{\rho}\mathbf{h}_k^H\mathbf{X}_d + \mathbf{z}_{kd}, \tag{2}$$

where $\mathbf{y}_{kp}$ is a $1 \times T_p$ received training signal vector and $\mathbf{y}_{kd}$ is a $1 \times T_d$ received data signal vector. $T_p$ and $T_d$ are the length of training signals and data signals, respectively, while $\rho$ is the normalized signal-to-noise ratio (SNR). By following the assumption of channel given in [1], $\mathbf{h}_k \in \mathbb{C}^{M\times 1}$ is the complex Gaussian channel vector, in which each element has a zero mean and unit variance. $\mathbf{X}_p \in \mathbb{C}^{M\times T_p}$ is a training sequence matrix, and $\mathbf{X}_d = \mathbf{GS} \in \mathbb{C}^{M\times T_d}$ is the signal matrix transmitted by the BS, where $\mathbf{G} \in \mathbb{C}^{M\times K}$ is typically a precoding matrix, and $\mathbf{S} \in \mathbb{C}^{K\times T_d}$ is the data matrix composed of zero-mean unit-variance entries. $\mathbf{z}_{kp} \in \mathbb{C}^{1\times T_p}$ and $\mathbf{z}_{kd} \in \mathbb{C}^{1\times T_d}$ are standard additive white Gaussian noise(AWGN) obeying $\mathcal{CN}(0,\,1)$, respectively.

When the CSI has no error, the total throughput (i.e., the achievable sum rate) can be expressed as

$$C = \mathbb{E}\left\{(1 - T_p/T)\sum\nolimits_{k=1}^{K} \log_2[1 + \rho\|\mathbf{h}_k^H\mathbf{G}\|^2]\right\}, \tag{3}$$

where $T$ is the coherent time, and $T = T_p + T_d$. When using traditional methods to estimate the CSI, $T_p$ should be no less than $M$.

It can be seen from (3) that the total throughput is related to the length of the training sequences. Under the condition that the coherence time remains unchanged, increasing the length of the training sequences can make the data transmission more reliable. On the other hand, as the length of the training sequences increases, more coherent time is occupied, and the time for transmitting data signals is shortened. Therefore, increasing the overhead of the training sequences also causes the total throughput to decrease. There is a trade-off between the overhead of training sequences and the total throughput.

## III. Dimension reduction method

In this section, we propose a dimension reduction scheme and optimize related parameters. Firstly, two cases based on Gaussian channel are considered: CSI without error and CSI with error. Then we propose a dimension reduction method based on the uniform planar array (UPA) model. Finally, We also apply the method into limited feedback MIMO systems.

### A. Dimension Reduction Matrix

Assume that $N$ is a positive integer, a power of 2 and a factor of $M$. Let $\mathbf{e}_{M/N}$ be an $M/N$-dimensional vector, whose each component is one. Define

$$\mathbf{W} \triangleq \left(\mathbf{e}_{M/N} \otimes \mathbf{F}_N\right)/\sqrt{M/N}, \tag{4}$$

where $\otimes$ is the Kronecker product and $\mathbf{F}_N$ is the $N$-point discrete Fourier transform (DFT) matrix with $\mathbf{F}_N^H\mathbf{F}_N = \mathbf{I}_N$. Here $\mathbf{F}_N$ can also be replaced by unit matrix $\mathbf{I}_N$. Thus, $\mathbf{W}$ is an $M \times N$ matrix, satisfying $\mathbf{W}^H\mathbf{W} = \mathbf{I}_N$. Note that for a given massive MIMO system, $\mathbf{W}$ is relative to $M$ and $N$ only.

We use $\mathbf{W}$ and precoding matrix $\mathbf{P}$ to process the data before sending the data signals $\mathbf{S} \in \mathbb{C}^{K\times T_d}$, where $\mathbf{P}$ is an $N \times K$ matrix. Thus, we have $\bar{\mathbf{G}} = \mathbf{WP}$ and $\bar{\mathbf{X}}_d = \mathbf{WPS}$.

The data signals received by the $k$-th user can be expressed as

$$\bar{\mathbf{y}}_{kd} = \sqrt{\rho}\mathbf{h}_k^H\mathbf{WPS} + \mathbf{z}_{kd}. \tag{5}$$

By comparing $\mathbf{P}$ with the matrix $\mathbf{G} \in \mathbb{C}^{M\times K}$ used in Eq. (3), the row dimension of precoding matrix is reduced from $M$ to $N$ because of using $\mathbf{W}$. Let $\bar{\mathbf{h}}_k \triangleq \mathbf{h}_k^H\mathbf{W} \in \mathbb{C}^{1\times N}$ represent the equivalent channel. The dimension of the equivalent channel is reduced from $M$ to $N$, hence the length of the training sequence can be reduced from $M$ to $N$.

### B. CSI without Error

We first assume that the CSI is accurate and there is no error in the feedback process. Suppose that the BS adopts the zero-forcing (ZF) precoding. Then we have

$$\mathbf{P} = \bar{\mathbf{H}}^H\left(\bar{\mathbf{H}}\bar{\mathbf{H}}^H\right)^{-1}\mathbf{B}, \tag{6}$$

where $\bar{\mathbf{H}} = \begin{bmatrix}\bar{\mathbf{h}}_1^H & \bar{\mathbf{h}}_2^H & \cdots & \bar{\mathbf{h}}_K^H\end{bmatrix}^H$, and $\mathbf{B}$ is the normalized factor matrix. Furthermore, let $\mathbf{A} \triangleq \bar{\mathbf{H}}^H\left(\bar{\mathbf{H}}\bar{\mathbf{H}}^H\right)^{-1} \triangleq \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_K\end{bmatrix}$ and $b_k \triangleq 1/\|\mathbf{a}_k\|, k = 1, 2, \cdots, K$. Then $\mathbf{B} = \mathrm{diag}(b_1, b_2, \cdots, b_K)$, where $\mathrm{diag}(\cdot)$ represents a diagonal matrix.

Rewrite the data signals received by all $K$ users as

$$\mathbf{Y} = \sqrt{\rho}\bar{\mathbf{H}}\mathbf{PS} + \mathbf{Z} = \sqrt{\rho}\mathbf{BS} + \mathbf{Z}, \tag{7}$$

where $\mathbf{S} = \begin{bmatrix}\mathbf{s}_1^T & \mathbf{s}_2^T & \cdots & \mathbf{s}_K^T\end{bmatrix}^T, \mathbf{Z} = \begin{bmatrix}\mathbf{z}_{1d}^T & \mathbf{z}_{2d}^T & \cdots & \mathbf{z}_{Kd}^T\end{bmatrix}^T, \mathbf{Y} = \begin{bmatrix}\bar{\mathbf{y}}_{1d}^T & \bar{\mathbf{y}}_{2d}^T & \cdots & \bar{\mathbf{y}}_{Kd}^T\end{bmatrix}^T$.

Hence, the data signals received by the $k$-th user can be rewritten as

$$\bar{\mathbf{y}}_{kd} = \sqrt{\rho}b_k\mathbf{s}_k + \mathbf{z}_{kd}, \tag{8}$$

where $b_k = 1/\sqrt{[\mathbf{A}^H\mathbf{A}]_{k,k}} = 1/\sqrt{[(\bar{\mathbf{H}}\bar{\mathbf{H}}^H)^{-1}]_{k,k}}$.

Because $\mathbb{E}[\mathbf{H}^H\mathbf{H}] = M\mathbf{I}_K$, $\mathbb{E}[\mathbf{H}\mathbf{H}^H] = K\mathbf{I}_M$ and $\bar{\mathbf{H}} = \mathbf{H}^H\mathbf{W}$, we have $\mathbb{E}[\bar{\mathbf{H}}\bar{\mathbf{H}}^H] = N\mathbf{I}_K$ and $\mathbb{E}[\bar{\mathbf{H}}^H\bar{\mathbf{H}}] = K\mathbf{I}_N$.

Hence, we can see that all elements in $\mathbf{H}$ and $\bar{\mathbf{H}}$ are zero-mean complex Gaussian random variables with unit variance.

Define a random variable $X \triangleq 1/[(\bar{\mathbf{H}}\bar{\mathbf{H}}^H)^{-1}]_{k,k}$, where $[\mathbf{A}]_{k,k}$ represents the element in the $k$-th row and the $k$-th column of the matrix $\mathbf{A}$. Use the conclusion of [12], the probability density function of the random variable $X$ is

$$f(x) = x^{N-K}e^{-x}/(N-K)!. \tag{9}$$

Consequently, since the training time is $N$, the total throughput is

$$C = \mathbb{E}\left[(1-N/T)\sum_{k=1}^{K}\log_2\left(1+\rho/[(\bar{\mathbf{H}}\bar{\mathbf{H}}^H)^{-1}]_{k,k}\right)\right] = (1-N/T)\,K\int_0^\infty \log_2(1+\rho x)\frac{x^{N-K}e^{-x}}{(N-K)!}dx. \quad (10)$$

In the above equation, $T$ and $K$ are parameters of the system, while $N$ should be designed to maximize the $C$. However, it is difficult to get the optimal value of $N$ in the above equation. Thus, in order to make the problem tractable, the Jensen inequality is used to approximate the right side of the above equation. The calculation process is as follows:

$$\begin{aligned} C &= \mathbb{E}\left[(1-N/T)\sum_{k=1}^{K}\log_2(1+\rho X)\right] \\ &\leqslant \left[(1-N/T)\sum_{k=1}^{K}\log_2(1+\mathbb{E}(\rho X))\right] \\ &= (1-N/T)\,K\log_2[1+\rho(N-K+1)]. \end{aligned} \quad (11)$$

We try to find the optimal solution of $N$ based on the last formula of (11). Setting $\frac{\partial C}{\partial N}=0$, yields

$$\begin{aligned} &(T-N)\rho - [1+\rho(N-K+1)]\ln 2 \\ &\cdot\log_2[1+\rho(N-K+1)] = 0. \end{aligned} \quad (12)$$

It is still difficult to get a closed-form solution of $N$ from (12). But we can easily obtain numerical solutions. Some calculation results are shown in Table I.

TABLE I
OPTIMAL $N$ WHEN CSI HAS NO ERROR

| Coherent time $T$ | Number of users $K$ | optimal $N$ (By Eq. (10)) | optimal $N$ (By Eq. (12)) | values of other parameters |
|---|---|---|---|---|
| 256 | 1~10 | 32 | 32 | The number of antennas $M$ is 128; The SNR $\rho$ is 10dB; $K \leqslant N \leqslant M$; $N=2^n, n\in\mathbb{Z}$ |
| | 11 | 64 | 32 | |
| | 12~56 | 64 | 64 | |
| | 57 | 128 | 64 | |
| | others | 128 | 128 | |
| 512 | 1~26 | 64 | 64 | |
| | 27 | 128 | 64 | |
| | others | 128 | 128 | |
| 1024 | 1~128 | 128 | 128 | |

It should be noticed that the optimal points based on (10) and (12) are highly consistent. Hence, we believe that the search for the optimal value of $N$ based on (12) is reasonable. When the coherent time and the number of users are not so large, the optimal number of $N$ is less than $M$. It means that not only the length of training sequences can be reduced, but also the throughput can be increased.

### C. CSI with Error

Now let us consider the case that the CSI estimated by uesrs has error. First, the BS transmits the training signals $\bar{\mathbf{X}}_p$ to the users. $\bar{\mathbf{X}}_p$ can be expressed as $\bar{\mathbf{X}}_p = \mathbf{W}\Phi$, where $\Phi \in \mathbb{C}^{N\times N}$ is the training sequence matrix and satisfies the condition $\Phi^H\Phi = \mathbf{I}_N$.

The training signal received by the $k$-th user is

$$\bar{\mathbf{y}}_{kp} = \sqrt{\rho}\bar{\mathbf{h}}_k\Phi + \mathbf{z}_{kp}. \quad (13)$$

We assume the user estmates the channel by using the MMSE method [2], i.e.,

$$\hat{\bar{\mathbf{h}}}_k = \bar{\mathbf{y}}_{kp}\sqrt{1/\rho}\left(\mathbf{I}_N/\rho + \Phi^H\Phi\right)^{-1}\Phi^H. \quad (14)$$

The data signals received by the $k$-th user is then expressed as

$$\tilde{\mathbf{y}}_{kd} = \sqrt{\rho}\hat{\bar{\mathbf{h}}}_k\mathbf{P}_e\mathbf{S} + \sqrt{\rho}\left(\bar{\mathbf{h}}_k - \hat{\bar{\mathbf{h}}}_k\right)\mathbf{P}_e\mathbf{S} + \mathbf{z}_{kd}. \quad (15)$$

In the above equation, the second term is the interference caused by the channel estimation error, and the third term is the AWGN during the transmission. $\tilde{\bar{\mathbf{h}}}_k \triangleq \bar{\mathbf{h}}_k - \hat{\bar{\mathbf{h}}}_k$ is defined as the channel estimation error and its variance is $\mathbb{E}\left[\tilde{\bar{\mathbf{h}}}_k^H\tilde{\bar{\mathbf{h}}}_k\right] = \mathbf{I}_N/(1+\rho)$, $\mathbf{P}_e$ is the corresponding ZF precoding matrix, which is similar to (6).

Further simplifying Eq. (15), yields

$$\begin{aligned} \tilde{\mathbf{y}}_{kd} &= \sqrt{\rho}\hat{\bar{\mathbf{h}}}_k\mathbf{P}_{ek}\mathbf{s}_k + \sum_{i\neq k}\sqrt{\rho}\hat{\bar{\mathbf{h}}}_k\mathbf{P}_{ei}\mathbf{s}_i + \sqrt{\rho}\left(\bar{\mathbf{h}}_k - \hat{\bar{\mathbf{h}}}_k\right) \\ &\cdot\mathbf{P}_{ek}\mathbf{s}_k + \sum_{i\neq k}\sqrt{\rho}\left(\bar{\mathbf{h}}_k - \hat{\bar{\mathbf{h}}}_k\right)\mathbf{P}_{ei}\mathbf{s}_i + \mathbf{z}_{kd} \\ &= \sqrt{\rho}\hat{\bar{\mathbf{h}}}_k\mathbf{P}_{ek}\mathbf{s}_k + \sum_{i=1}^{K}\sqrt{\rho}\tilde{\bar{\mathbf{h}}}_k\mathbf{P}_{ei}\mathbf{s}_i + \mathbf{z}_{kd}. \end{aligned} \quad (16)$$

In Eq. (16), the first term is related to the data signals that the $k$-th user expects to obtain, the second term represents the interference imposed by the other users on the $k$-th user and the impact imposed by the channel estimation error, while the third term denotes the AWGN. Thus, the total throughput of the system can be expressed as

$$C_e = \mathbb{E}\left[(1-N/T)\sum_{i=1}^{K}\log_2\left(1+\frac{\rho\|\hat{\bar{\mathbf{h}}}_k\mathbf{P}_{ek}\|^2}{1+\rho\sum_{i=1}^{K}\|\tilde{\bar{\mathbf{h}}}_k\mathbf{P}_{ei}\|^2}\right)\right]. \quad (17)$$

Thus, the optimal selection of the number $N$ can be expressed as the following optimization problem:

$$\max_N\ C_e, \quad \text{s.t.} \begin{cases} K \leqslant N \leqslant M \\ N = 2^n, n\in\mathbb{Z}. \end{cases} \quad (18)$$

The numerical solutions to this optimization problem are shown in Table II. It can be seen that, when there is channel estimation error and the number of users is not so large, the optimal value of $N$ is less than the number of antennas $M$. So the overhead caused by the training sequence can be reduced.

TABLE II
OPTIMAL $N$ WHEN CSI HAS ERROR

| Number of users $K$ | optimal $N$ | values of other parameters |
|---|---|---|
| 1~5 | 32 | The number of antennas $M$ is 128; The $\rho$ is 10dB; Coherent time $T$ is 256 $K \leqslant N \leqslant M$; $N=2^n, n\in\mathbb{Z}$ |
| 6~48 | 64 | |
| others | 128 | |

### D. Dimension Reduction Matrix Based on UPA Channel

Here we consider the dimension reduction matrix based on UPA channel. By following the assumption of channel given in [13], the UPA channel can be expressed as

$$\mathbf{h}_u = \sum_{p=1}^{P}\mathbf{d}_M(\psi_p^v,\psi_p^h)\alpha_p = \mathbf{D}\mathbf{a}, \quad (19)$$

where $\mathbf{D} = \left[\mathbf{d}_M(\psi_1^v,\psi_1^h)\ \cdots\ \mathbf{d}_M(\psi_P^v,\psi_P^h)\right] \in \mathbb{C}^{M\times P}$ is the set of radio paths and $\mathbf{a} = \left[\alpha_1\ \cdots\ \alpha_P\right] \in \mathbb{C}^{P\times 1}$ is the set of complex channel gains. More specifically, $\mathbf{d}_M(\psi^v,\psi^h)$ can be obtained by Kronecker product between radiation paths in vertical direction and horizontal direction, i.e., $\mathbf{d}_M(\psi^v,\psi^h) = \mathbf{d}_{M_v}(\psi^v)\otimes\mathbf{d}_{M_h}(\psi^h)$, where the array response vector $\mathbf{d}_{M_a}(\psi^a)$ is expressed as

$$\mathbf{d}_{M_a}(\psi^a) = \begin{bmatrix} 1 & e^{j\frac{2\pi d_a}{\lambda_c}\psi^a} & \cdots & e^{j\frac{2\pi d_a}{\lambda_c}(M_a-1)\psi^a} \end{bmatrix}^T \quad (20)$$

for $a \in \{v, h\}$, where $\psi^v = \sin\phi^v$ and $\psi^h = \sin\phi^h \cos\phi^v$. $d_a$ is the antennas spacing, and $\phi^a$ is the angle for array vector. $f_c$ and $\lambda_c$ is the carrier frequency and carrier wavelength, respectively, satisfying $c = f_c\lambda_c$ with the speed of light $c$.

Based on matrix $\mathbf{W}$ defined by (4), we can obtain a $N$-dimensional equivalent channel $\bar{\mathbf{h}}_u = \mathbf{W}^H\mathbf{h}_u$ of the UPA model. In order to find the optimal $N$ in this case, we further calculate the equivalent channel $\bar{\mathbf{h}}_u$ based on the UPA model, yields

$$\begin{aligned} \bar{\mathbf{h}}_u &= \mathbf{W}^H\mathbf{h}_u \\ &= \sum_{p=1}^{P} \frac{\alpha_p}{\sqrt{M/N}} (\mathbf{e}_{M/N} \otimes \mathbf{I}_N)^H [\mathbf{d}_{M_v}(\psi_p^v) \otimes \mathbf{d}_{M_h}(\psi_p^h)] \\ &\overset{(a)}{=} \sum_{p=1}^{P} \frac{\alpha_p}{\sqrt{M/N}} [\mathbf{e}_{M/N}^H \cdot \mathbf{d}_{M_v}(\psi_p^v)] \otimes [\mathbf{I}_N \cdot \mathbf{d}_{M_h}(\psi_p^h)], \end{aligned} \quad (21)$$

where $(a)$ holds iff $M/N = M_v$. So we set the value of $N$ to

$$N = M/M_v. \quad (22)$$

Formulas above mean that, after the dimension reduction, all antennas in the vertical direction are combined into an antenna unit. In order to increase the gain after combining, the radiation angle of each vertical antenna component should be limited to a certain range, i.e.,

$$|2\pi d_v \psi^v l/\lambda_c| \le \Omega, l \in \{0, 1, \cdots, M_v - 1\}, \quad (23)$$

where $\Omega$ can be called *angular resolution*. Obviously, the smaller this angular resolution is, the larger the gain of the superimposed antenna components is. Thus, the value of $d_v$ should be as small as possible. Simulations given in next section will confirm this judgement.

### *E. Application in Limited Feedback MIMO Systems*

We apply the above method to a limited feedback MIMO system. For simplification, we assume $K = 1$. Thus, the signal received by the user can be written as

$$\bar{y} = \sqrt{\rho}\bar{\mathbf{h}}_u^H \mathbf{f} s + n, \quad (24)$$

where $\bar{\mathbf{h}}_u = \mathbf{W}^H\mathbf{h}_u$ is the $N$-dimensional equivalent channel and $\mathbf{h}_u \in \mathbb{C}^{M\times 1}$ is the UPA model. $\mathbf{f}$ is a precoding vector, $s$ is the data signal satisfying $\mathbb{E}[s] = 0$ and $\mathbb{E}[|s|^2] = 1$, and $n$ is the standard AWGN.

For the $N$-dimensional equivalent channel, the codebook generation and codeword selection in the limited feedback MIMO system can adopt the methods given in [14], which are detailed as follows.

**Step 1:** Let $\bar{\mathbf{H}}_u = \left[\bar{\mathbf{h}}_{u[1:N_h]} \ \bar{\mathbf{h}}_{u[N_h+1:2N_h]} \ \cdots \ \bar{\mathbf{h}}_{u[N_h(N_v-1)+1:N_vN_h]}\right]^T$, where $\bar{\mathbf{h}}_{u[a:b]}$ represents the vector that is extracted from the $a$-th element to the b-th element in $\bar{\mathbf{h}}_u$. Here the channel $\bar{\mathbf{h}}_u$ is rearranged to get an $N_v \times N_h$ channel matrix $\bar{\mathbf{H}}_u$ and $N = N_v \times N_h$. The singular value decomposition (SVD) of matrix $\bar{\mathbf{H}}_u$ is $\bar{\mathbf{H}}_u = \mathbf{U}\mathbf{D}\mathbf{V}^H$.

**Step 2:** Use the DFT codebook to quantize channels. The quantization criterion is as follows:
$\mathbf{f}_v = \underset{\mathbf{f}_{vi}\in F_{N_v}^{B_v}}{\operatorname{argmax}} \|\mathbf{u}_1^H\mathbf{f}_{vi}\|^2$, $\mathbf{f}_h = \underset{\mathbf{f}_{hi}\in F_{N_h}^{B_h}}{\operatorname{argmax}} \|\mathbf{v}_1^H\mathbf{f}_{hi}\|^2$,
where $\mathbf{u_1}$ and $\mathbf{v_1}$ are the first column of $\mathbf{U}$ and $\mathbf{V}$, respectively. $\mathcal{F}_{N_a}^{B_a}$ is the $B_a$-bit DFT codebook, and $a \in \{v, h\}$, where the $i$-th codeword is defined as
$\mathbf{f}_{ai} = \begin{bmatrix} 1 & e^{j2\pi\frac{i}{2^{B_a}}} & e^{j2\pi\frac{2i}{2^{B_a}}} & \cdots & e^{j2\pi\frac{(N_a-1)i}{2^{B_a}}} \end{bmatrix}^T / \sqrt{N_a}$,
$i = 1, 2, \cdots, 2^{B_a}$.

**Step 3:** The precoding codeword is $\mathbf{f}_o = \mathbf{f}_v \otimes \mathbf{f}_h^*$.

As a result, the throughput is given as

$$\bar{C} = \mathbb{E}\left\{(1 - N/T)\log_2[1 + \rho\|\bar{\mathbf{h}}^H\mathbf{f}_o\|^2]\right\}. \quad (25)$$

Obviously, when $N = M$, this formula gives the throughput of the scheme without using dimension reduction method , and when $N < M$, (25) represents the throughput of our proposed scheme. It is further pointed out that the dimension of equivalent channel decreases, which can reduce the complexity of codebook construction and the overhead of feedback.

## IV. SIMULATIONS RESULTS AND DISCUSSIONS

In this section, three simulations are implemented to confirm the analytical results given in the previous sections.

**Simulation 1:** Here we assume $K$ is 2. We aim to show the optimal value of $N$ and the benefits of using the proposed method under the Gaussian channel environment.

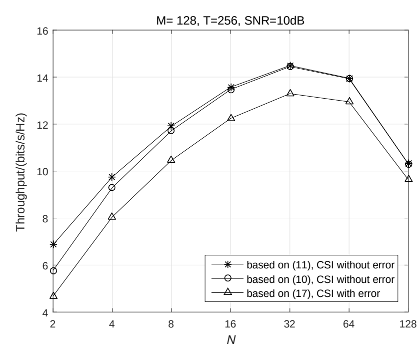



Fig. 1. The influences of $N$ on the total throughput

We observe from Fig. 1 that the upper bound based on (11) is close to the numerical calculation of (10). In particular, the optimal values of $N$ are the same in the three cases considered, which confirms that the method of calculating the optimal value of $N$ based on (12) is reliable. It also shows that the optimal value of $N$ is not equal to $M$. In this specific configuration, the optimal value of $N$ is 32, which is much smaller than 128. Therefore, the length of training sequences can be reduced from 128 to 32, while the corresponding throughput increases about 5 bits per channel use (pcu). Similarily, we also simulated the cases of $K = 4, 16$, and the optimal values of $N$ are 32 and 64, respectively.

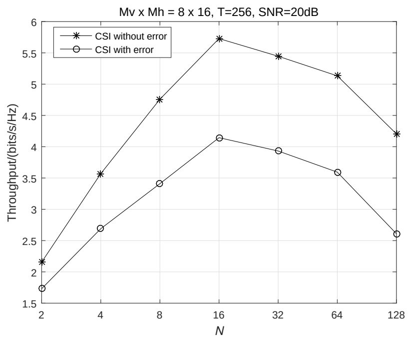
**Simulation 2:** We try to confirm the results given in III.*D*. The parameters of UPA model are shown in Table III.

Fig. 2 shows the influences of $N$ on the total throughput based on UPA model. Here the SNR $\rho$ is 20dB. When CSI has error, we use MMSE method to estmate the channel, i.e.,

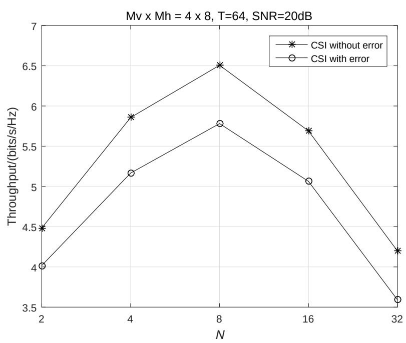
$$\hat{\mathbf{h}}_{ue} = \sum\nolimits_{p=1}^{P} \mathbf{d}_M(\hat{\psi}_p^v, \hat{\psi}_p^h)\hat{\alpha}_p, \quad (26)$$

TABLE III
SIMULATION PARAMETERS

| Parameter | Assumotion |
|---|---|
| Transmitting antennas | $8 \times 16, 4 \times 8$ co-polarized |
| Receiving antennas | $1 \times 1$ co-polarized |
| Carrier frequency | 2GHz |
| $d_v$ | $0.2\lambda_c$ |
| $d_h$ | $0.5\lambda_c$ |




(a) $M_v \times M_h = 8 \times 16$




(b) $M_v \times M_h = 4 \times 8$

Fig. 2. The influences of $N$ on the total throughput based on UPA model

where $\hat{\psi}_p^v = \psi_p^v + \tilde{\psi}_p^v$, $\hat{\psi}_p^h = \psi_p^h + \tilde{\psi}_p^h$ and $\hat{\alpha}_p = \alpha_p + \tilde{\alpha}_p$. Here $\tilde{\psi}_p^v$, $\tilde{\psi}_p^h$ and $\tilde{\alpha}_p$ all are the complex Gaussian distribution according to $\mathcal{CN}(0, 1/(1+\rho))$. In fig. 2(a), we can find that the optimal numbers of $N$ in both cases are 16, which is equal to the number calculated by (22). The same conclusion can also deduced form Fig. 2(b), where the optimal number of $N$ obtained by the figure and the calculation of (22) are the same.

**Simulation 3:** We show the benefits of using the proposed method in limited feedback MIMO system. The parameters of UPA model are shown in Table III.

Fig. 3 shows the throughput comparison between proposed method with dimension reduction ($N = 16 < M$) and the method without dimension reduction ($N = 128 = M$). Additionally, we assume $B_v = B_h = 4$. It can be seen that, when $d_v = 0.5\lambda_c$, our method does not have advantages, since the value of $d_v$ is too large for the method. However, $d_v = 0.2\lambda_c$ is a suitable value, and the throughput of proposed method is apparently higher than that of the method without dimension reduction. As the SNR increases, the gap between them is getting larger. Thus, our dimension reduction method is also valuable in the limited feedback MIMO systems based on UPA model when the number of $d_v$ is appropriate.

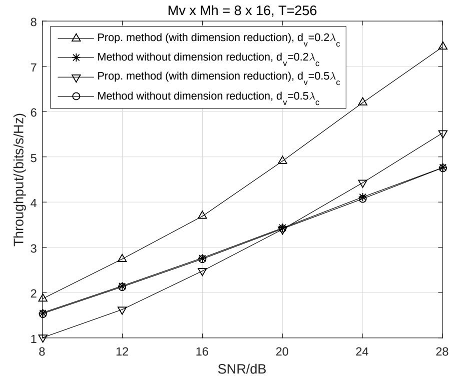



Fig. 3. Throughput comparison in limited feedback MIMO systems

## V. CONCLUSIONS

For the FDD massive MIMO downlink systems, the training sequence imposes a large overhead. We propose a dimension reduction method to divide and combine the transmit antennas into several groups. The number of groups is smaller than the original number of antennas, hence the proposed method is capable of reducing the length of the training sequences while improving the system performance. Our method is also applicable to limited feedback MIMO systems.

## REFERENCES

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

[2] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884-893, Feb. 2006.

[3] H. Wang, W. Zhang, Y. Liu, Q. Xu, and P. Pan, "On design of non-orthogonal pilot signals for a multi-cell massive MIMO system," *IEEE Wireless Commun. Lett.*, vol. 4, no. 2, pp. 129-132, Apr. 2015.

[4] M. Nekovee, Y. Wang, M. Tesanovic, S. Wu, and Y. M. Al-Imari, "Overview of 5G modulation and waveforms candidates," *J. Commun. Inf. Networks*, vol. 1, no. 1, pp: 44-60, 2016.

[5] L. Li, X. Niu, Y. Chai, L. Chen, T. Zhang, D. Cheng, H. Xia, J. Wang, T. Cui, and X. You, "The path to 5G:mmWave aspects," *J. Commun. Inf. Networks,* vol. 1, no. 2, pp: 1-18, 2016.

[6] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951-963, Apr. 2003.

[7] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802-814, Oct. 2014.

[8] X. Meng, X. Gao, and X. G. Xia, "Omnidirectional precoding based transmission in massive MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 174-186, Jan. 2016.

[9] X. G. Xia and X. Gao, "A space-time code design for omnidirectional transmission in massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 5, no. 5, pp. 512-515, Oct. 2016.

[10] T. Lv, F. Tan, G. Hui, and S. Yang, "A beamspace approach for 2-D localization of incoherently distributed sources in massive MIMO systems," *Signal Process.*, vol. 121, pp. 30-45, Apr. 2016.

[11] B. Lee, J. Choi, J. Seol, D. J. Love and B. Shim, "Antenna grouping based feedback compression for FDD-based massive MIMO systems," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3261-3274, Sept. 2015.

[12] Y. Jiang, M. K. Varanasi, and J. Li, "Performance analysis of ZF and MMSE equalizers for MIMO systems: An in-depth study of the high SNR regime," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 2008-2026, Apr. 2011.

[13] J. Song, J. Choi, T. Kim and D. J. Love, "Advanced quantizer designs for FDD-based FD-MIMO systems using uniform planar arrays," *IEEE Trans. Signal Process.*, vol. 66, no. 14, pp. 3891-3905, Jul. 2018.

[14] J. Choi, K. Lee, D. J. Love, T. Kim, and R. W. Heath, "Advanced limited feedback designs for FD-MIMO using uniform planar arrays," in *Proc. IEEE GLOBECOM*, Dec. 2016.