Precoding Methods for Multi-User TDD MIMO Systems

(Invited Paper)

Jubin Jose*, Alexei Ashikhmin†, Phil Whiting†, Sriram Vishwanath*
*Laboratory for Informatics, Networks and Communications
Department of Electrical and Computer Engineering
The University of Texas at Austin, Austin, TX, USA 78712
†Bell Laboratories, Alcatel-Lucent Inc., Murray Hill, NJ, USA 07974

***Abstract*—We present linear precoding methods for downlink transmission in multi-user multiple-input multiple-output (MIMO) systems where channel state information (CSI) can be obtained through training. In this paper, channel training overhead and estimation error are rigorously accounted for while determining the net system throughput. First, we consider the case with training on the reverse link only. We study a precoding method which is a combination of two schemes: selection of users with largest estimated gains and zero-forcing to selected users. Next, we consider the case with training on both reverse and forward links. We obtain a lower bound on the weighted-sum capacity, and propose an algorithm to determine an efficient precoding matrix. This precoding method effectively utilizes the training on the reverse and forward links in improving net throughput.**

I. Introduction

We are interested in the downlink transmission from the base station to the users. This multi-user MIMO downlink transmission scenario is analyzed as the multi-antenna broadcast channel (BC) problem in information theory literature. The sum capacity of the multi-antenna Gaussian BC has been shown to be achieved by dirty paper coding (DPC) in [1]–[4]. It was shown in [5] that DPC characterizes the full capacity region of the multi-antenna Gaussian BC. In the multi-user setting, the existing results show that significant throughput gains can be obtained with multiple antennas at the base station and single antenna at the users. We are interested in this scenario where the the base station has M antennas, and the $K(\leq M)$ independent users have single antennas as shown in Figure 1.

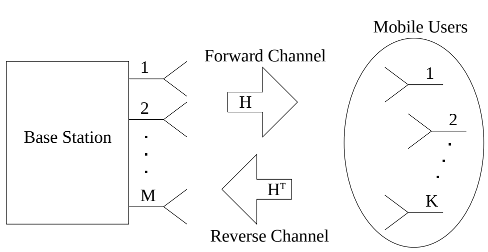

Fig. 1. Multi-User MIMO TDD System Model

The DPC technique [6] is fairly involved and is computationally challenging to implement in practice. Therefore, a problem that obtained significant research attention is throughput maximization with minimal complexity at the terminals. Different precoding methods have been studied [7]–[11]. The results suggest that sum rates close to sum capacity can be achieved with much lower computational complexity compared to DPC. However, these results assume channel state information (CSI) at the base station and the users. The techniques developed might require full CSI at the base station and the users, and can be sensitive to CSI accuracy. Motivated by this, different techniques have been developed when limited channel knowledge is available at the base station and prefect CSI is available at the users [12]–[16]. This limited feedback setting is motivated by the scenario where partial CSI is acquired by the base station through feedback.

In the limited feedback setting, the overhead associated with channel training and error due to channel estimation are not accounted for in the system throughput. In [17], [18], the effects of training and estimation in multi-user MIMO systems using time-division duplex (TDD) operation have been studied. Similar to this, we account for these factors in the net throughput and consider TDD operation. In TDD systems, the transmit channel can be obtained from the reverse link channel as both are very closely related [19]. Specifically, the channel model we study is the following. The channel is assumed to undergo block fading with a coherence interval of T symbols. We assume that the reverse channel matrix and forward channel matrix share a reciprocity relationship. In [18], we proposed scheduling and pre-conditioning schemes for multi-user MIMO TDD systems with training on reverse link only. In this paper, we focus on the case with training on both reverse and forward links.

We first look at the case with training on reverse link only. In this setting, we briefly describe the precoding method we developed in [18]. The base station obtains minimum-mean-square-error estimate (MMSE) of the channel matrix from the reverse link pilots. This channel estimate is used in the precoding method which consists of a scheduling strategy and zero-forcing. In the case of homogeneous users, the scheduling strategy is simply the selection of users with largest estimated channel gains. Next, we consider the case with training on both reverse and forward links. Recently, there has been similar work in [20]. The authors consider two-way training [21] and study two variants of linear MMSE precoders as alternatives to linear zero-forcing precoder used in [17]. We use a modified version of the precoder proposed in [11]. In the approach in [11], the precoding matrix is obtained using an iterative algorithm which tries (there is no proof of convergence) to find one of the local maxima of the sum rate maximization problem when CSI is available at both the base station and the users. Since the base station obtains CSI through training, we modify this algorithm to account for the error in the estimation process. We compare the performance of the different methods considered through numerical examples.

A. Notations

We use bold font variables to denote vectors and matrices. All vectors are column vectors. We use $(\cdot)^T$ to denote the transpose, $(\cdot)^*$ to denote the conjugate and $(\cdot)^\dagger$ to denote the Hermitian of vectors and matrices. Tr $(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$ and $\mathbf{A}^{-1}$ denotes the inverse of matrix $\mathbf{A}$. diag $\{\mathbf{a}\}$ denotes a diagonal matrix with diagonal entries equal to the components of $\mathbf{a}$. $\mathbb{E}[\cdot]$ and $\mathbf{var}\{\cdot\}$ stand for expectation and variance operations, respectively.

II. System Model

The base-station with M antennas communicates with the K independent users on both forward and reverse links as shown in Fig. 1. The forward channel is characterized by $K\times M$ propagation matrix $\mathbf{H}$. We assume independent Rayleigh fading channels, which remains constant over a duration of T symbols called the coherence interval. The entries of the channel matrix $\mathbf{H}$ are independent and identically distributed (i.i.d.) zero-mean, circularly-symmetric complex Gaussian $CN(0,1)$ random variables. Our model incorporates frequency selectivity of fading by using orthogonal frequency-division multiplexing (OFDM). Note that the duration of the coherence interval in symbols is chosen for the OFDM sub-band. Due to reciprocity, we assume that the reverse channel at any instant is the transpose of the forward channel.

Let the forward and reverse SINRs associated with k^{th} user be ρ_{fk} and ρ_{rk}, respectively. These forward and reverse SINRs remain fixed throughout the channel uses. On the forward link, the signal received by the k^{th} user is

$$x_{fk} = \sqrt{\rho_{fk}}\,\mathbf{h}_k^T\mathbf{s_f} + w_{fk} \tag{1}$$

where $\mathbf{h}_k^T$ is the k^{th} row of the channel matrix $\mathbf{H}$ and $\mathbf{s_f}$ is the $M\times 1$ vector in which information symbols to be communicated are embedded. The components of the additive noise vector $[w_{f1}\, w_{f2}\, \cdots\, w_{fK}]$ are i.i.d. $CN(0,1)$. The average power constraint at the base-station during transmission is $\mathbb{E}[\|\mathbf{s_f}\|^2]=1$ so that the total transmit power is fixed irrespective of its number of antennas. On the reverse link, the vector received at the base-station is

$$\mathbf{x_r} = \mathbf{H}^T\mathbf{E}_r\mathbf{s_r} + \mathbf{w_r} \tag{2}$$

where $\mathbf{s_r}$ is the signal-vector transmitted by the users and $\mathbf{E}_r = diag\{[\sqrt{\rho_{r1}}\,\sqrt{\rho_{r2}}\,\cdots\,\sqrt{\rho_{rK}}]^T\}$. The components of the additive noise $\mathbf{w_r}$ are i.i.d. $CN(0,1)$. There is average power constraint at every user during transmission given by $\mathbb{E}[\|s_{rk}\|^2]=1$ where s_{rk} is the k^{th} component of $\mathbf{s_r}$.

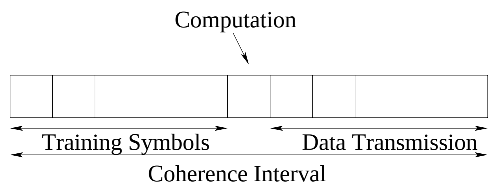

Fig. 2. Training on reverse link only

III. Training on Reverse Link Only

We consider the scenario in which the system is operated in three phases as shown in Fig. 2. In the training phase, the users transmit a training sequence to the base station on the reverse link. The base station performs the required computations including user selection and precoding in the computation phase. We assume that this takes one symbol. In the data transmission phase, the base station transmits data to the selected users.

A. Channel Estimation

A key advantage of TDD systems over frequency-division duplex (FDD) systems is channel reciprocity. We exploit this property to perform channel estimation by transmitting training sequences on the reverse link. Every user transmits a sequence of training signals of τ_{rp} symbols duration in every coherence interval. We assume that these training sequences are known a-priori to the base-station. The k^{th} user transmits the training sequence vector $\sqrt{\tau_{rp}}\,\psi_k^\dagger$. We use orthonormal sequences which implies $\psi_i^\dagger\psi_j = \delta_{ij}$ where δ_{ij} is the Kronecker delta. The use of orthogonal sequences restricts the maximum number of users to τ_{rp}, i.e., $K\le\tau_{rp}$.

The corrupted training signals received at the base-station is

$$\mathbf{Y}_r = \sqrt{\tau_{rp}}\,\mathbf{H}^T\mathbf{E}_r\mathbf{\Psi}^\dagger + \mathbf{V}_r \tag{3}$$

where $\mathbf{\Psi} = [\psi_1\,\psi_2\,\cdots\,\psi_K]$ and the components additive noise matrix $\mathbf{V}_r$ are i.i.d. $CN(0,1)$. The base-station obtains the LMMSE (linear minimum-mean-square-error) estimate of the channel

$$\hat{\mathbf{H}} = diag\left\{\left[\frac{\sqrt{\rho_{r1}\tau_{rp}}}{1+\rho_{r1}\tau_{rp}}\ \cdots\ \frac{\sqrt{\rho_{rK}\tau_{rp}}}{1+\rho_{rK}\tau_{rp}}\right]^T\right\}\mathbf{\Psi}^T\mathbf{Y}_r^T. \tag{4}$$

This estimate $\hat{\mathbf{H}}$ is the conditional mean of $\mathbf{H}$ and hence, the MMSE estimate as well. By the properties of conditional mean and joint Gaussian distribution, the estimate $\hat{\mathbf{H}}$ is independent of the estimation error $\tilde{\mathbf{H}} = \mathbf{H}-\hat{\mathbf{H}}$. The components of $\hat{\mathbf{H}}$ are independent and the elements of its k^{th} row are $CN\left(0,\frac{\rho_{rk}\tau_{rp}}{1+\rho_{rk}\tau_{rp}}\right)$. In addition, the components of $\tilde{\mathbf{H}}$ are independent and the elements of its k^{th} row are $CN\left(0,\frac{1}{1+\rho_{rk}\tau_{rp}}\right)$.

B. Scheduling and precoding on Forward Link

In this work, we confine ourselves by the homogeneous case where forward SINRs from the base-station to all users are equal and also reverse SINRs from all users to the base-station are equal, i.e., $\rho_{f1} = \cdots = \rho_{fK} = \rho_f$ and $\rho_{r1} = \cdots = \rho_{rK} = \rho_r$. The case of heterogeneous users is discussed in another article [18].

The base station selects $N (\leq K)$ users among the K users and precode the information signals to be transmitted to these N users. The scheduling strategy used to select the users is explained in Section III-D. Let the set of users selected be $S_N \subseteq \{1, 2, \cdots, K\}$ with N distinct entries. The base station forms the $M \times 1$ transmission signal-vector $\mathbf{s}_f$ from the information symbol-vector $\mathbf{q} = [q_1 \; q_2 \; \cdots \; q_N]^T$ for the selected users by pre-multiplying it with a precoding matrix. We use the precoding matrix

$$\mathbf{A}_{S_N} = \frac{\hat{\mathbf{H}}^{\dagger}_{S_N} \left(\hat{\mathbf{H}}_{S_N} \hat{\mathbf{H}}^{\dagger}_{S_N} \right)^{-1}}{\sqrt{tr \left[\left(\hat{\mathbf{H}}_{S_N} \hat{\mathbf{H}}^{\dagger}_{S_N} \right)^{-1} \right]}} \quad (5)$$

which is proportional to the pseudo-inverse of the estimated channel. The $N \times M$ matrix $\hat{\mathbf{H}}_{S_N}$ is formed by the rows in set S_N of matrix $\hat{\mathbf{H}}$. We use this precoding matrix because of the lack of any channel knowledge at the users. The precoding matrix is normalized so that $tr(\mathbf{A}^{\dagger}_{S_N} \mathbf{A}_{S_N}) = 1$.

The transmission signal-vector is given by

$$\mathbf{s}_f = \mathbf{A}_{S_N} \mathbf{q} \quad (6)$$

and the power constraint at the base station is satisfied by imposing the condition $\mathbb{E}[\|q_n\|^2] = 1, \forall n \in \{1, \cdots, N\}$. From (1) and (6), we obtain the signal-vector received at the selected users to be

$$\mathbf{x}_f = \sqrt{\rho_f} \, \mathbf{H}_{S_N} \mathbf{A}_{S_N} \mathbf{q} + \mathbf{w}_f \quad (7)$$

where $\mathbf{H}_{S_N}$ is the matrix formed by the rows in set S_N of the matrix $\mathbf{H}$.

C. Lower Bound on Sum Capacity

In this section, we obtain a lower bound on the sum capacity of the system under consideration. The approach is similar to that in [17], [22]. The lower bound holds for any scheduling strategy used at the base station which selects a fixed number of users. Recall that the base station performs channel estimation as described in Section III-A.

Theorem 1: For the system under consideration, every selected user can achieve a downlink rate during data transmission of at least

$$C_{ind-lb} = \log_2 \left(1 + \frac{\rho_f \mathbb{E}^2 [\chi]}{1 + \rho_f \left(\frac{1}{1 + \rho_r \tau_{rp}} + \mathbf{var}\{\chi\} \right)} \right) \quad (8)$$

bits/transmission where χ is a scalar random variable given by $\chi = \left(tr \left[\left(\hat{\mathbf{H}}_{S_N} \hat{\mathbf{H}}^{\dagger}_{S_N} \right)^{-1} \right] \right)^{-\frac{1}{2}}$.

Corollary 1: For the system with homogeneous users considered, a lower bound on the sum capacity is

$$C_{sum-lb} = \max_{N \leq K, \, N \in \mathbb{I}^+} N \cdot C_{ind-lb}. \quad (9)$$

D. Scheduling Strategy

The need for explicit scheduling arises due the use of pseudo-inverse based precoding of the information symbols. With perfect channel knowledge at the base station ($\hat{\mathbf{H}} = \mathbf{H}$) and no scheduling ($N = K$), the pseudo-inverse based precoding diagonalizes the effective forward channel and every user sees statistically identical effective channel irrespective of its actual channel. The inability to vary the effective gains to the users depending on their channel states is due to lack of any channel knowledge at the users. This possibly causes a reduction in achievable sum rate. Motivated by this, we propose a scheduling strategy which explicitly selects $N \leq K$ users before precoding.

In every coherence interval, the channel estimate at the base station is used to select the N users with largest estimated channel gains. Let $\hat{\mathbf{h}}^T_{(1)}, \hat{\mathbf{h}}^T_{(2)}, \cdots, \hat{\mathbf{h}}^T_{(K)}$ be the norm-ordered rows of the estimated channel matrix $\hat{\mathbf{H}}$. Then, the matrix $\hat{\mathbf{H}}_{S_N}$ is given by $\hat{\mathbf{H}}_{S_N} = [\hat{\mathbf{h}}_{(1)} \; \hat{\mathbf{h}}_{(2)} \; \cdots \; \hat{\mathbf{h}}_{(N)}]^T$ and the lower bound in (8) becomes $C_{ind-lb} =$

$$\log_2 \left(1 + \frac{\rho_f \left(\frac{\rho_r \tau_{rp}}{1 + \rho_r \tau_{rp}} \right) \mathbb{E}^2 [\eta]}{1 + \rho_f \left(\frac{1}{1 + \rho_r \tau_{rp}} + \frac{\rho_r \tau_{rp}}{1 + \rho_r \tau_{rp}} \mathbf{var}\{\eta\} \right)} \right). \quad (10)$$

Here, random variable $\eta = \left(tr \left[\left(\mathbf{U}\mathbf{U}^{\dagger} \right)^{-1} \right] \right)^{-\frac{1}{2}}$ where $\mathbf{U}$ is the $N \times M$ matrix formed by the N rows with largest norms of a $K \times M$ random matrix $\mathbf{Z}$ whose elements are i.i.d. $CN(0, 1)$. We provide numerical results showing the improvement obtained by using this strategy in Section V.

E. Net Achievable Sum Rate

Net achievable sum rate accounts for the reduction in achievable sum rate due to training. In every coherence interval of T symbols, first τ_{rp} symbols are used for training on reverse link, one symbol is used for computation (same assumption as in [17]) and the remaining $T - \tau_{rp} - 1$ symbols are used for transmitting information symbols. The number of users K and the training length τ_{rp} can be chosen such that net throughput of the system is maximized. Thus, net achievable sum rate is defined as

$$C_{net}(M, \rho_f, \rho_r) = \max_{K, \tau_{rp}} \frac{T - \tau_{rp} - 1}{T} C_{sum-lb}(\cdot) \quad (11)$$

subject to the constraints $\tau_{rp} \leq T$ and $K \leq \min(M, \tau_{rp})$. $C_{sum-lb}(\cdot)$ in (11) is given by (9).

IV. Training on Reverse and Forward Links

In the transmission scheme considered before, the users do not receive any knowledge about effective channel gains. Therefore, we used the expected value of the effective gains seen by the users to obtain a lower bound on weighted-sum

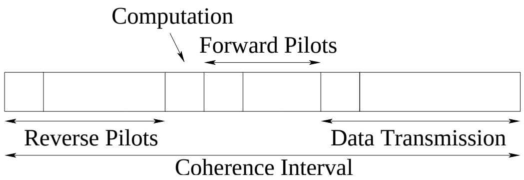

Fig. 3. Training on reverse and forward links

capacity. The base station can send forward pilots to the users so that the users can estimate their effective gains in every coherence interval. This gives a transmission scheme consisting of four phases - reverse pilots, computation phase, forward pilots and data transmission - as shown in Figure 3. In this scheme, the users can obtain effective channel gain estimates at the expense of increased training overhead.

A. Forward Training

The base station transmits τ_{fp} forward pilots so that the users can obtain estimates of their channel gains. Since we are interested in short coherence intervals, we consider the case with very few forward pilots. Note that τ_{fp} can be less than the number of users K. For this reason, we do not restrict to orthogonal pilots in forward training. The forward pilots are obtained by pre-multiplying the vectors $\mathbf{q}_p^{(1)}, \cdots, \mathbf{q}_p^{(\tau_{fp})}$ with the precoding matrix. In the case of one forward pilot ($\tau_{fp} = 1$), we consider the forward pilots obtained from the vector $\mathbf{q}_p^{(1)} = [1\,1 \cdots]^T$. In the case of $\tau_{fp} = 2$, we consider the forward pilots obtained from the vectors $\mathbf{q}_p^{(1)} = \sqrt{2}[1\,0\,1\,0 \cdots]^T$ and $\mathbf{q}_p^{(1)} = \sqrt{2}[0\,1\,0\,1 \cdots]^T$. It is straightforward to extend this to any number of forward pilots. We denote the vector of corrupted forward pilots received by the k^{th} user by $\mathbf{x}_{pk}$.

B. Net Achievable Weighted-Sum Rate

Let $\mathbf{A}$ be an arbitrary precoding matrix, not necessarily the pseudo-inverse of the estimated channel matrix. We use same lower bounding techniques as before to obtain net achievable weighted-sum rate for the transmission scheme with reverse and forward pilots. From (1), we obtain the signal-vector received at the users

$$\mathbf{x}_f = \mathbf{E}_f \mathbf{H} \mathbf{A} \mathbf{q} + \mathbf{w}_f \qquad (12)$$

where $\mathbf{E}_f = \text{diag}\left\{\left[\sqrt{\rho_f^1}\,\sqrt{\rho_f^2}\,\cdots\,\sqrt{\rho_f^N}\right]^T\right\}$. We denote the effective forward channel in (12) by $\mathbf{G} = \mathbf{E}_f \mathbf{H} \mathbf{A}$ with $(i,j)^{th}$ entry g_{ij}.

Theorem 2: For the transmission scheme considered, a lower bound on the downlink sum capacity during transmission is given by

$$C_{wsum-lb} = \sum_{k=1}^{K} \mathbb{E}\left[\log_2\left(1 + \frac{|\mathbb{E}\left[g_{kk}|\mathbf{x}_{pk}\right]|^2}{1 + \sum\limits_{i \neq k} \mathbb{E}\left[|g_{ki}|^2|\mathbf{x}_{pk}\right] + \mathbf{var}\{g_{kk}|\mathbf{x}_{pk}\}}\right)\right]. \qquad (13)$$

We define net achievable weighted-sum rate as

$$C_{net} = \max_{\tau_{rp}} \frac{T - \tau_{rp} - \tau_{fp} - 1}{T} C_{wsum-lb}(\cdot)$$

which is consistent with the earlier definition.

C. Efficient Precoding Matrix

In this section we suggest a different way for determining a precoding matrix $\mathbf{A}$.

As explained in Section III-A, the users transmit orthogonal training sequences on the reverse link. From the corrupted training sequences, the base station obtains the MMSE estimate of the channel. The base station uses this channel estimate $\hat{\mathbf{H}}$ to form a precoding matrix to perform linear precoding. Let $\mathbf{A}$ denote any precoding matrix which is a function of the channel estimate, i.e., $\mathbf{A} = f(\hat{\mathbf{H}})$. The precoding function $f(\cdot)$ usually depends on the system parameters such as forward SINRs, reverse SINRs and weights assigned to the users. We assume that the precoding matrix is normalized so that $\text{Tr}\left(\mathbf{A}^\dagger \mathbf{A}\right) = 1$. The transmission signal-vector is given by $\mathbf{s}_f = \mathbf{A}\mathbf{q}$ where $\mathbf{q} = [q_1\, q_2\, \cdots\, q_K]^T$ is the vector of information symbols for the users. The net achievable rate derived in Section IV-B is valid for any precoding function. In the remaining part of this section, we describe a particular precoding method.

Let us assume for the moment that the channel matrix $\mathbf{H}$ is known to the base station and mobiles. In [11], the following approach was suggested for finding a good precoding matrix $\mathbf{A}$ when $\mathbf{H}$ is known. Let $\mathbf{h}_i$ be the i-th row of the channel matrix $\mathbf{H}$ and let $\mathbf{a}_j$ be the j-th column of precoding matrix $\mathbf{A}$. The sum rate of the broadcast channel can be written in the form

$$R(\mathbf{H}, \mathbf{A}) = \sum_{j=1}^{M} \log\left(1 + \frac{|\mathbf{h}_j^T \mathbf{a}_j|^2}{\sigma^2 \text{Tr}\left(\mathbf{A}\mathbf{A}^\dagger\right) + \sum_{r \neq j} |\mathbf{h}_j^T \mathbf{a}_r|^2}\right).$$

Let

$$b_j = |\mathbf{h}_j \mathbf{a}_j|^2 \text{ and } c_j = \sigma^2 \text{Tr}\left(\mathbf{A}\mathbf{A}^\dagger\right) + \sum_{r \neq j} |\mathbf{h}_j \mathbf{a}_r|^2.$$

Let further $\boldsymbol{\Delta}$ and $\mathbf{D}$ be diagonal matrices with diagonals

$$\boldsymbol{\Delta} = \text{diag}\left(\frac{(\mathbf{HA})_{11}}{c_1}, \frac{(\mathbf{HA})_{22}}{c_2}, \ldots, \frac{(\mathbf{HA})_{MM}}{c_M}\right) \qquad (14)$$

and

$$\mathbf{D} = \text{diag}\left(\frac{b_1}{c_1(b_1 + c_1)}, \frac{b_2}{c_2(b_2 + c_2)}, \ldots, \frac{b_M}{c_M(b_M + c_M)}\right). \qquad (15)$$

In [11] it is shown that the equations $\frac{\partial R(\mathbf{H},\mathbf{A})}{\partial \mathbf{A}_{ij}} = 0$ imply

$$\mathbf{A} = ((\sigma^2 \text{Tr}(\mathbf{D})) I_M + \mathbf{H}^\dagger \mathbf{D} \mathbf{H})^{-1} \mathbf{H}^\dagger \boldsymbol{\Delta}. \qquad (16)$$

This equation allows one to use the following iterative algorithm for determining an efficient $\mathbf{A}$:

1) Assigning some initial values to matrices $\boldsymbol{\Delta}$ and $\mathbf{D}$, for instance $\boldsymbol{\Delta} = I_M, \mathbf{D} = I_M$
2) Repeat steps 3 and 4 several times

3) Compute $\mathbf{A}$ according to (16);
4) Compute $\mathbf{\Delta}$ and $\mathbf{D}$ according to (14) and (15).

Let us now assume that only an estimate $\hat{\mathbf{H}}$ of the channel matrix $\mathbf{H}$ and the statistics of the estimation error $\tilde{\mathbf{H}}$ are available. Let us also assume that the base station sends very powerful pilots so that each mobile can very precisely estimate its signal to noise ratio. In this case the sum capacity is equal to

$$R(\hat{\mathbf{H}}, \mathbf{A}) = \mathbb{E}_{\tilde{\mathbf{H}}}[R(\hat{\mathbf{H}} + \tilde{\mathbf{H}}, \mathbf{A})].$$

The approach suggested in [11] can be extended for this scenario as follows. Since the statistics of $\tilde{\mathbf{H}}$ is assumed to be known, we can generate L samples $\tilde{\mathbf{H}}^{(i)}, i = 1, \ldots, L$, according to the statistics. Define $\mathbf{H}^{(i)} = \hat{\mathbf{H}} + \tilde{\mathbf{H}}^{(i)}$. Then the average rate can be approximated as

$$\begin{aligned} & R(\hat{\mathbf{H}}, \mathbf{A}) \\ \approx \; & \frac{1}{L}\sum_{i=1}^{L}\sum_{j=1}^{M}\log\left(1 + \frac{|{\mathbf{h}_j^{(i)}}^T \mathbf{a}_j|^2}{\sigma^2 \mathrm{Tr}\,(\mathbf{A}\mathbf{A}^\dagger) + \sum_{r\neq j}|{\mathbf{h}_j^{(i)}}^T \mathbf{a}_r|^2}\right) \end{aligned}$$

We define $\mathbf{\Delta}^{(i)}$ and $\mathbf{D}^{(i)}$ as in (14) and (15) using the matrix $\mathbf{H}(i)$ instead of $\mathbf{H}$. Using arguments similar to ones used in [11] we obtain that the equations $\frac{\partial R}{\partial \mathbf{A}_{ij}} = 0$ imply

$$\sum_{i=1}^{L}\mathbf{H}^{(i)}\mathbf{\Delta}^{(i)} - \mathbf{H}^{(i)\dagger}\mathbf{D}^{(i)}\mathbf{H}^{(i)} - \sigma^2 \mathrm{Tr}\,(\mathbf{D}^{(i)})\mathbf{A} = 0. \quad (17)$$

Let

$$\mathbf{V} = \sum_{i=1}^{L}\mathbf{H}^{(i)\dagger}\mathbf{D}^{(i)}\mathbf{H}^{(i)} + \sigma^2 \mathrm{Tr}\,(\mathbf{D}^{(i)}) I_M, \mathbf{T} = \sum_{i=1}^{L}\mathbf{H}^{(i)}\mathbf{\Delta}^{(i)}.$$

From (17), we have that

$$\mathbf{A} = \mathbf{V}^{-1}\mathbf{T}. \quad (18)$$

This allows us to use the following iterative algorithm for determining a $\mathbf{A}$ that maximizes the average rate.

1) Assigning some initial values to matrices $\mathbf{\Delta}^{(i)}$ and $\mathbf{D}^{(i)}$, for instance $\mathbf{\Delta}^{(i)} = I_M, \mathbf{D}^{(i)} = I_M$
2) Repeat steps 3 and 4 several times
3) Compute $\mathbf{A}$ according to (18);
4) Compute $\mathbf{\Delta}^{(i)}$ and $\mathbf{D}^{(i)}$ according to (14) and (15) using $\mathbf{H}^{(i)}$ instead of $\mathbf{H}$.

V. Numerical Results

We provide numerical results in both reverse training only and reverse and forward training scenarios. We are particularly interested in high mobility users, i.e., short coherence intervals. We use FP(n) to denote a precoding method using n number of forward pilots. Note that FP(0) denotes training on reverse link only. We denote results obtained with zero-forcing (without scheduling) by ZF, zero-forcing with scheduling by ZF-Sch, the approach in [11] by SVH and the modified algorithm given in Section IV-C by Mod-SVH.

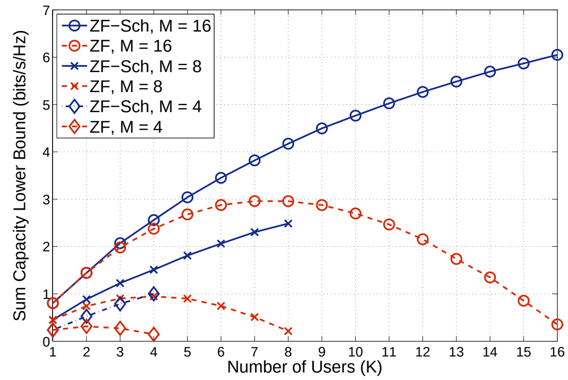

Fig. 4. Sum capacity lower bound

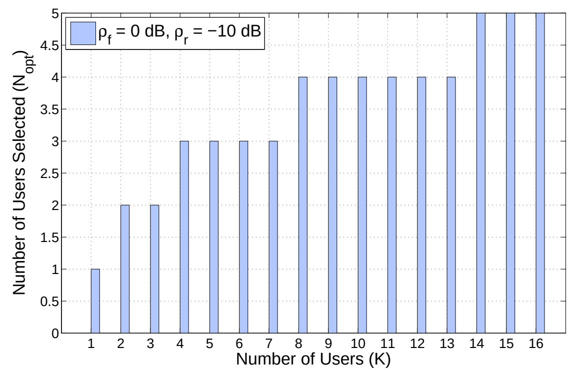

Fig. 5. Number of users selected

A. Reverse Training Only

First, we keep the training sequence length to the minimum possible, i.e., $\tau_{rp} = K$. In Figure 4, we plot sum capacity lower bound verses the number of users $K = \{1, 2, \cdots, M\}$ for $M = \{4, 8, 16\}$ when forward SINR $\rho_f = 0$ dB and reverse SINR $\rho_r = -10$ dB. We observe that the precoding method ZF-Sch gives significant improvement over ZF. In both the methods, the achievable sum rate increases with the number of base station antennas. In Figure 5, we plot the optimum number of users selected by ZF-Sch N_{opt} versus the number of users present K for the SINR considered above and $M = 16$.

Next, in Figure 6, we plot net achievable sum rate for coherence intervals $T = \{10, 20, 30\}$ symbols. We observe that the net achievable sum rate increases with M for both ZF and ZF-Sch, and the precoding method ZF-Sch outperforms ZF. We notice that the net achievable sum rate varies significantly with the coherence interval. This demonstrates the need to take the coherence interval into account while designing wireless systems.

B. Reverse and Forward Training

We compare the performance of different methods using numerical examples. We keep the value of reverse SINR 10

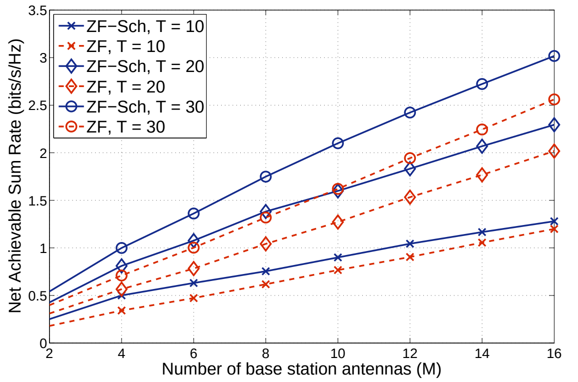

Fig. 6. Net achievable sum rate for different coherence intervals

TABLE I
MULTI-USER SYSTEM WITH $M = 8$ AND $K = 8$

ρ_f (dB)	10	15	20	25	30
ZF-Sch-FP(0)	7.56	11.56	15.40	17.87	18.86
ZF-Sch-FP(1)	6.58	10.33	15.21	21.83	22.79
ZF-Sch-FP(2)	7.69	11.67	16.67	22.08	28.99
Mod-SVH-FP(1)	7.11	11.17	16.56	24.14	30.43
Mod-SVH-FP(2)	7.77	11.93	17.22	24.04	30.04

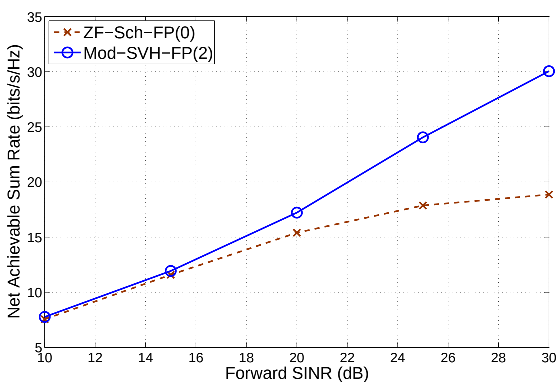

Fig. 7. Mod-SVH and ZF-Sch for $M = 8$

dB lower than the forward SINR in all the cases considered. For the algorithm Mod-SVH, we use the value $L = 50$ in the simulations. Since we expect different behavior for $M = K$ and $M \gg K$, we consider the following two examples. In the first example, we consider a system with $K = 8$ users and $M = 8$ antennas at the base station. For the different methods considered, we obtain the achievable sum rate for forward SINRs ranging from 10 dB to 30 dB. These sum rates are given in Table I. In this case, we observe that using forward pilots is advantageous at high SINRs. In particular, we observe significant improvement by using Mod-SVH algorithm over ZF-Sch at high SINRs. We plot the methods ZF-Sch-FP(0) and Mod-SVG-FP(2) in Figure 7.

In the second example, we consider a system with $K =$ 8 users and $M = 16$ antennas at the base station. Again, we obtain the achievable sum rate for different methods and tabulate these values in Table II. In this case, we observe that using forward pilots is not advantageous in the SINR range considered. We plot the methods ZF-Sch-FP(0) and Mod-SVG-FP(1) in Figure 8.

TABLE II
MULTI-USER SYSTEM WITH $M = 16$ AND $K = 8$

ρ_f (dB)	10	15	20	25	30
ZF-Sch-FP(0)	20.99	29.79	37.90	44.43	48.71
ZF-Sch-FP(1)	12.73	20.45	28.38	36.38	44.41
ZF-Sch-FP(2)	12.18	19.59	27.37	35.40	43.64
Mod-SVH-FP(1)	13.59	20.73	28.52	36.60	44.40
Mod-SVH-FP(2)	12.96	20.01	27.72	35.66	43.74

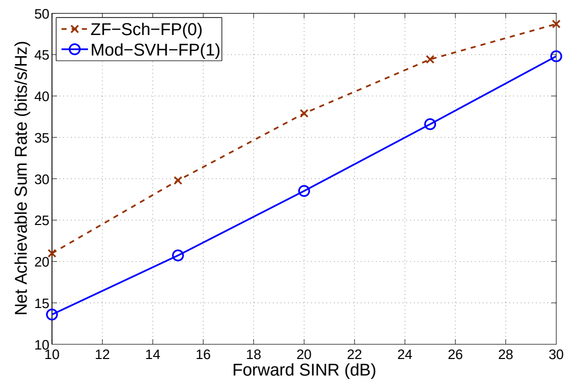

Fig. 8. Mod-SVH and ZF-Sch for $M = 16$

VI. CONCLUSION

In multi-user MIMO downlink, training is very important to obtain CSI at the terminals. Reciprocal training made feasible by time-division duplex (TDD) operation is key to obtaining CSI at the base station. When coherence intervals are short, the overhead associated with training can significantly affect the net throughput. Therefore, it is important to choose the number of reverse pilots, the number of forward pilots and the precoding method based on the coherence interval length. From the numerical examples, we observed that it is advantageous to introduce forward pilots when SINRs are high. We conclude that linear precoding methods can be used at the base station to take advantage of the multiple antennas. This significantly improve the net downlink throughput.

ACKNOWLEDGMENT

The authors would like to thank T. L. Marzetta for helpful discussions on this topic.

REFERENCES

[1] G. Caire and S. Shamai, "On the Achievable Throughput of a Multiantenna Gaussian Broadcast Channel," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1691–1707, Jul. 2003.

[2] P. Viswanath and D. N. C. Tse, "Sum Capacity of the Vector Gaussian Broadcast Channel and Uplink-Downlink Duality," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1912–1921, Aug. 2003.

[3] S. Vishwanath, N. Jindal, and A. J. Goldsmith, "Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2658–2668, Oct. 2003.

[4] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1875–1892, 2004.

[5] H. Weingarten, Y. Steinberg, and S. Shamai, "The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3936–3964, Sep. 2006.

[6] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, pp. 439–441, May 1983.

[7] F. Boccardi, F. Tosato, and G. Caire, "Precoding Schemes for the MIMO-GBC," in *Int. Zurich Seminar on Communications*, Feb. 2006.

[8] M. Airy, S. Bhadra, R. W. Heath, Jr., and S. Shakkottai, "Transmit Precoding for the Multiple Antenna Broadcast Channel," in *Vehicular Technology Conference*, vol. 3, 2006, pp. 1396–1400.

[9] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser CommunicationPart I: Channel Inversion and Regularization," *IEEE Trans. Commun.*, vol. 53, pp. 195–202, Jan. 2005.

[10] ——, "A Vector-Perturbation Technique for Near-Capacity Multi-antenna Multiuser CommunicationPart II: Perturbation," *IEEE Trans. Commun.*, vol. 53, pp. 537–544, Jan. 2005.

[11] M. Stojnic, H. Vikalo, and B. Hassibi, "Rate Maximization in Multi-Antenna Broadcast Channels with Linear Preprocessing," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 2338–2342, Sep. 2006.

[12] M. Sharif and B. Hassibi, "On the Capacity of MIMO Broadcast Channels with Partial Side Information," *IEEE Trans. Inf. Theory*, vol. 51, pp. 506–522, Feb. 2005.

[13] P. Ding, D. J. Love, and M. D. Zoltowski, "On the sum rate of channel subspace feedback for multi-antenna broadcast channels," in *Proc. IEEE Globecom*, Nov. 2005, pp. 2699–2703.

[14] N. Jindal, "MIMO Broadcast Channels with Finite-Rate Feedback," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5045–5060, Nov. 2006.

[15] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna broadcast channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 1478–1491, 2007.

[16] K. Huang, R. W. Heath, Jr., and J. G. Andrews, "Space Division Multiple Access with a Sum Feedback Rate Constraint," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3879–3891, 2007.

[17] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, Oct./Nov. 2006, pp. 359–363.

[18] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Scheduling and Pre-conditioning in Multi-User MIMO TDD Systems," in *Proc. IEEE International Conference on Communications(ICC'08)*, Beijing, China, May 2008, pp. 4100 – 4105.

[19] T. L. Marzetta and B. M. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. Signal Process.*, vol. 54, pp. 1268–1278, 2006.

[20] K. S. Gomadam, H. C. Papadopoulos, and C.-E. W. Sundberg, "Techniques for Multi-User MIMO with Two-Way Training," in *Proc. IEEE International Conference on Communications(ICC'08)*, Beijing, China, May 2008, pp. 3360–3366.

[21] S. Steger and A. Sabharwal, "Single-input two-way SIMO channel: diversity-multiplexing tradeoff with two-way training," *to appear in IEEE Trans. Wireless Commun.*, 2008.

[22] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, pp. 951–963, Apr. 2003.