MIMO Multiple Access Channels with Noisy Channel Estimation and Partial CSI Feedback

Alkan Soysal
Department of Electrical and Electronics Engineering
Bahçeşehir University, Istanbul, Turkey
alkan.soysal@bahcesehir.edu.tr

Sennur Ulukus
Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742
ulukus@umd.edu

***Abstract*—We consider correlated MIMO multiple access channels with block fading, where each block is divided into training and data transmission phases. We find the channel estimation and data transmission parameters that jointly optimize the achievable data rate of the system. Our results for the training phase are particularly interesting, where we show that the optimum training signals of the users should be non-overlapping in time. For the data transmission phase, we propose an iterative algorithm that updates the parameters of the users in a round-robin fashion. In particular, the algorithm updates the training and data transmission parameters of a user, when those of the rest of the users are fixed, in a way to maximize the achievable sum-rate in a multiple access channel; and iterates over users in a round-robin fashion.**

I. Introduction

In wireless communication scenarios, the achievable rate of a system depends crucially on the amount of channel state information (CSI) available at the receivers and the transmitters, especially when there are multiple antennas and multiple users in the system. With perfect CSI at the receiver and the transmitter, the optimum adaptation scheme is water-filling [1], [2]. However, in some cases, especially in multi-input multi-output (MIMO) links, feeding the instantaneous CSI back to the transmitter is not realistic. Therefore, some research assumes that there is perfect CSI at the receiver, but only partial CSI available at the transmitter [3]–[7].

Another line of research considers the actual estimation of the channel at the receiver, which is noisy. The capacity and the corresponding optimum signalling scheme for this case are not known. However, lower and upper bounds for the capacity can be obtained [8]–[10]. It is important to note that [8]–[10] assume the existence of a separate channel that does not consume system resources for channel estimation. For a MIMO system with no CSI available at the transmitter, [11] considers optimizing the achievable rate as a function of both the training and the data transmission phases.

In [12], [13], we studied the optimization of the achievable data rate jointly in terms of the channel estimation and data transmission parameters of a single-user, block-fading, correlated MIMO channel with noisy channel estimation at the receiver, and partial CSI available at the transmitter. We solved the trade-off between estimating the channel better and increasing the achievable data rate.

This work was supported by NSF Grants CCR 03-11311, CCF 04-47613 and CCF 05-14846; and ARL/CTA Grant DAAD 19-01-2-0011.

In a multi-user setting, the amount of resources required to measure the channel and to feed the estimated channel back to the transmitter increases substantially. When perfect CSI is available at the receiver and the transmitters at no cost, [14] finds the optimum transmission strategy, which is a multi-user water-filling scheme. Under a more practical assumption, when there is perfect CSI at the receiver but only partial CSI available at the transmitters, [6], [7] find the optimum transmit strategies for all users. When the channel estimation at the receiver is noisy, most of the research focuses on single-user systems [8], [10]–[13].

In this paper, we consider a multi-user setting. First, we consider the channel estimation process and find the optimum training signals for all users. Although all of the users are allowed to use the available training duration simultaneously, we find that the training signals of the users should be non-overlapping in time. Since the total block length, and therefore the total training duration is limited, each user can only train a fraction of its available channel dimensions, which might result in shorter individual training signal durations compared to the single-user case [12], [13].

Next, we move to the data transmission phase, and derive an achievable sum-rate expression that includes the channel estimation and data transmission parameters of all users. We first determine the optimum transmit directions for all users. Then, we develop an algorithm that maximizes the sum-rate jointly over the individual training durations of all users, the allocation of power of each user between training and data transmission phases, and also the allocation of the data transmission power of each user over its transmit directions.

II. System Model

We consider a multiple access channel (MAC) with multiple antennas at every user and the receiver. The channel between user k and the receiver is represented by a random matrix $\mathbf{H}_k$ with dimensions of $n_R \times n_T$, where n_R and n_T are the number of antennas at the receiver and at the transmitters, respectively. We consider a block fading scenario where the channel remains constant for a block (T symbols), and changes to an i.i.d. realization at the end of the block. In order to estimate the channels, the receiver performs a linear MMSE estimation for the channels of the users using training symbols over T_t symbols. During the remaining $T_d = T - T_t$ symbols, data transmission occurs. While the receiver has a noisy estimate of the realization of the fading channel, the transmitters have only the statistical model of the channel. Each transmitter sends a vector $\mathbf{x}_{kn}$, and the received vector at time n is

$$\mathbf{r}_n = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{x}_{kn} + \mathbf{n}_n, \quad n = 1, \ldots, T \tag{1}$$

where K is the number of users, $\mathbf{n}_n$ is a zero-mean, identity-covariance complex Gaussian vector, at time n, and the entries of $\mathbf{H}_k$ are complex Gaussian random variables. Each user has a power constraint of P_k, averaged over T symbols.

The statistical model that we consider is the "partial CSI with covariance feedback" model where each transmitter knows the channel covariance information of all transmitters, in addition to the distribution of the channel. In this paper, we will assume that the receiver does not have any physical restrictions and therefore, there is sufficient spacing between the antenna elements on the receiver such that the signals received at different antenna elements are uncorrelated. However, there exists correlation between the signals transmitted by different antenna elements. The channel is modeled as [15],

$$\mathbf{H}_k = \mathbf{Z}_k \mathbf{\Sigma}_k^{1/2} \tag{2}$$

where the entries of $\mathbf{Z}_k$ are i.i.d., zero-mean, unit-variance complex Gaussian random variables, and $\mathbf{\Sigma}_k$ is the channel covariance feedback matrix of user k. Similar covariance feedback models have been used in [4]–[7].

III. Joint Optimization for Multi-user MIMO

Our goal in this paper is to find the maximum achievable sum-rate which is optimized jointly over the training parameters (training signal, training duration, training signal power) and the data transmission parameters (transmit directions and allocation of power over the antennas).

A. Training and Channel Estimation Phase

The input-output relationship during the training phase in a multiple access channel is

$$\mathbf{R}_t = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{S}_k + \mathbf{N}_t \tag{3}$$

where $\mathbf{S}_k$ is an $n_T \times T_t$ dimensional training signal for user k that will be chosen and known at both ends, $\mathbf{R}_t$ and $\mathbf{N}_t$ are $n_R \times T_t$ dimensional received signal and noise matrices, respectively. The n^{th} column of the matrix equation in (3) represents the input-output relationship at time n. The power constraint for the training input signal for user k is $\frac{1}{T_t}\text{tr}(\mathbf{S}_k\mathbf{S}_k^\dagger) \leq P_{t_k}$.

The receiver is to estimate the channels of all users simultaneously, during the same training phase, with the knowledge of all training symbols. Therefore, it can regard the multi-user channel as a single-user channel, where the channel and the training signal matrices of all the users are stacked together. We can then write (3) equivalently as

$$\mathbf{R}_t = \bar{\mathbf{H}}\bar{\mathbf{S}} + \mathbf{N}_t \tag{4}$$

where $\bar{\mathbf{H}} = [\mathbf{H}_1, \ldots, \mathbf{H}_K]$ is an $n_R \times Kn_T$ dimensional channel matrix, and $\bar{\mathbf{S}} = [\mathbf{S}_1^\dagger, \ldots, \mathbf{S}_K^\dagger]^\dagger$ is a $Kn_T \times T_t$ dimensional training signal matrix. In this equivalent problem, the receiver will estimate $\bar{\mathbf{H}}$ using the output $\mathbf{R}_t$ and the training signal $\bar{\mathbf{S}}$.

Due to our channel model in (2), the entries in a row of $\mathbf{H}_k$ are correlated, and the entries in a column of $\mathbf{H}_k$ are uncorrelated. In other words, for each user, row i and row j of the channel matrix are i.i.d. This also holds for the stacked matrix $\bar{\mathbf{H}}$. Let us represent row i of $\mathbf{H}_k$ as $\mathbf{h}_{ki}^\dagger$, where $E[\mathbf{h}_{ki}\mathbf{h}_{ki}^\dagger] = \mathbf{\Sigma}_k, i = 1, \ldots n_R$, and row i of $\bar{\mathbf{H}}$ as $\bar{\mathbf{h}}_i = [\mathbf{h}_{1i}^\dagger, \ldots, \mathbf{h}_{Ki}^\dagger]^\dagger$, where $\bar{\mathbf{\Sigma}} = E[\bar{\mathbf{h}}_i\bar{\mathbf{h}}_i^\dagger] = \text{diag}\{\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_K\}$ is a block diagonal matrix, having $\mathbf{\Sigma}_k$ on its diagonals.

Let the eigenvalue representation of the channel covariance matrix of user k be $\mathbf{\Sigma}_k = \mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{\Sigma_k}\mathbf{U}_{\Sigma_k}^\dagger$, then the eigenvectors of the stacked channel covariance matrix $\bar{\mathbf{\Sigma}} = \bar{\mathbf{U}}_\Sigma\bar{\mathbf{\Lambda}}_\Sigma\bar{\mathbf{U}}_\Sigma^\dagger$ can also be written as $\bar{\mathbf{U}}_\Sigma = \text{diag}\{\mathbf{U}_{\Sigma_1}, \ldots, \mathbf{U}_{\Sigma_K}\}$ [16, Lemma 1.3.10], which is a block diagonal matrix as well.

Since a row of $\bar{\mathbf{H}}$ is formed by combining the rows of all $\mathbf{H}_k$ into a single, and longer row, we can conclude that the rows of $\bar{\mathbf{H}}$ are also i.i.d., and the receiver can estimate each of them independently using the same training symbols. The i^{th} row of (4) can be written as

$$\mathbf{r}_{ti} = \bar{\mathbf{S}}^\dagger\bar{\mathbf{h}}_i + \mathbf{n}_{ti} \tag{5}$$

Since this is equivalent to a single-user channel estimation problem with the exception of a block diagonal channel covariance matrix, we can use the MMSE estimation results of [12], [13]. Denoting the estimate of $\bar{\mathbf{h}}_i$ as $\hat{\mathbf{h}}_i = \bar{\mathbf{M}}\mathbf{r}_{ti}$, and the channel estimation error as $\tilde{\mathbf{h}}_i = \bar{\mathbf{h}}_i - \hat{\mathbf{h}}_i$, the MMSE estimation problem can be written as

$$\min_{\bar{\mathbf{M}}} E\left[\tilde{\mathbf{h}}_i^\dagger\tilde{\mathbf{h}}_i\right] = \min_{\bar{\mathbf{M}}} E\left[\text{tr}\left((\bar{\mathbf{h}}_i - \bar{\mathbf{M}}\mathbf{r}_{ti})(\bar{\mathbf{h}}_i - \bar{\mathbf{M}}\mathbf{r}_{ti})^\dagger\right)\right] \tag{6}$$

Using the orthogonality principle [17, page 91] as in the single-user case [12], we can find the optimum estimator as $\bar{\mathbf{M}}^* = \bar{\mathbf{\Sigma}}\bar{\mathbf{S}}\left(\bar{\mathbf{S}}^\dagger\bar{\mathbf{\Sigma}}\bar{\mathbf{S}} + \mathbf{I}\right)^{-1}$, and the mean square error in (6) becomes,

$$\min_{\bar{\mathbf{M}}} E\left[\tilde{\mathbf{h}}_i^\dagger\tilde{\mathbf{h}}_i\right] = \text{tr}\left(\left(\bar{\mathbf{\Sigma}}^{-1} + \bar{\mathbf{S}}\bar{\mathbf{S}}^\dagger\right)^{-1}\right) \tag{7}$$

where we used the matrix inversion lemma [16, page 19]. Note that the mean square error of the channel estimation process can be further decreased by choosing the training signal $\bar{\mathbf{S}}$ to minimize (7). The following theorem finds $\bar{\mathbf{S}}$, and the training signals of individual users $\mathbf{S}_k$, for a given training power and training duration.

Theorem 1: For given $\mathbf{\Sigma}_k = \mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{\Sigma_k}\mathbf{U}_{\Sigma_k}^\dagger$, P_{t_k}, T_t, and the power constraints $\text{tr}(\mathbf{S}_k\mathbf{S}_k^\dagger) \leq P_{t_k}T_t$, the $Kn_T \times T_t$ dimensional optimum stacked training signal $\bar{\mathbf{S}}$ that minimizes the total power of the channel estimation error vector is $\bar{\mathbf{S}} = \bar{\mathbf{U}}_\Sigma\bar{\mathbf{\Lambda}}_S^{1/2}$, and the $n_T \times K$ dimensional optimum training signal of user k is $\mathbf{S}_k = \left[\mathbf{0},\ldots,\mathbf{0},\mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{S_k}^{1/2},\mathbf{0},\ldots,\mathbf{0}\right]$ with

$$\lambda_{ki}^S = \left(\frac{1}{\mu_k^S} - \frac{1}{\lambda_{ki}^\Sigma}\right)^+, \qquad i = 1,\ldots,\min(n_T, T_{t_k}) \quad (8)$$

where T_{t_k} is the duration of the training signal of user k, $(\mu_k^S)^2$ is the Lagrange multiplier that satisfies the power constraint with $\mu_k^S = \frac{J_k}{P_{t_k}+\sum_{i=1}^{J_k}\frac{1}{\lambda_{ki}^\Sigma}}$, and J_k is the largest index that has non-zero λ_{ki}^S for user k.

A result of Theorem 1 is that it is sufficient to consider only $T_{t_k} \leq n_T$, which we will assume for the rest of this paper. In addition, Theorem 1 states that orthogonality in the time domain holds over the users in a multi-user setting as well. Although this might seem counter-intuitive at first, after the diagonalization of the channel, we are left with orthogonal channels. Therefore, in order to estimate orthogonal channels, sending orthogonal training signals is sufficient.

Due to the constraint on the training duration, fewer dimensions of the individual channels will be estimated for each user, which will result in shorter individual training durations compared to a single-user case. However, by the conservation of energy, the training signal power of a particular user will be larger compared to the training signal power of the same user in a single-user environment. Therefore, although fewer dimensions of the channel are estimated, the channel estimation error corresponding to those estimated dimensions will be smaller.

Note that μ_k^S is a function of only P_{t_k} and T_{t_k}, both of which will be chosen to maximize the sum-rate of the data transmission phase. The value of T_{t_k} determines the total number of available parallel channels for user k, and the value of P_{t_k} determines the number of channels that will be estimated. The parametric values of P_{t_k} and T_{t_k} will appear in the sum-rate formula in the next section.

Before moving on to the next section, we will state the eigenvalues of the covariance matrices of the estimated channel vector, and the channel estimation error vector for all users. This derivation is similar to the single-user case, and can be found in [12], [13]. The covariance matrix of the channel estimation error of user k can be found as $\tilde{\mathbf{\Sigma}}_k = \mathbf{U}_{\Sigma_k}\left(\mathbf{\Lambda}_{\Sigma_k}^{-1}+\mathbf{\Lambda}_{S_k}\right)^{-1}\mathbf{U}_{\Sigma_k}^\dagger$, where the eigenvalues are given as $\tilde{\lambda}_{ki}^\Sigma = \min\left(\lambda_{ki}^\Sigma, \mu_k^S\right)$. Similarly, the covariance matrix of the estimated channel of user k can be found using the orthogonality principle as $\hat{\mathbf{\Sigma}} = \bar{\mathbf{U}}_\Sigma\left(\bar{\mathbf{\Lambda}}_\Sigma - \tilde{\mathbf{\Lambda}}_\Sigma\right)\bar{\mathbf{U}}_\Sigma^\dagger$, where the eigenvalues are given as $\hat{\lambda}_{ki}^\Sigma = \min\left(\lambda_{ki}^\Sigma - \mu_k^S, 0\right)$.

B. Data Transmission Phase

The sum-rate of a multiple access channel can be derived using the stacked channel and input matrices,

$$\mathbf{r} = \sum_{k=1}^K \hat{\mathbf{H}}_k\mathbf{x}_k + \sum_{k=1}^K \tilde{\mathbf{H}}_k\mathbf{x}_k + \mathbf{n} = \hat{\mathbf{H}}\bar{\mathbf{x}} + \tilde{\mathbf{H}}\bar{\mathbf{x}} + \mathbf{n} \quad (9)$$

where $\hat{\mathbf{H}} = [\hat{\mathbf{H}}_1,\ldots,\hat{\mathbf{H}}_K]$, $\tilde{\mathbf{H}} = [\tilde{\mathbf{H}}_1,\ldots,\tilde{\mathbf{H}}_K]$ are $n_R \times Kn_T$ dimensional, and $\bar{\mathbf{x}} = [\mathbf{x}_1^\dagger,\ldots,\mathbf{x}_K^\dagger]^\dagger$ is $Kn_T \times 1$ dimensional. Although the optimum input distribution is not known, we can achieve the following lower bound with Gaussian $\bar{\mathbf{x}}$ [10],

$$C_{lb}^{sum} = I(\mathbf{r};\bar{\mathbf{x}}|\hat{\mathbf{H}}) = E\left[\log\left|\mathbf{I} + \mathbf{R}_{\tilde{\mathbf{H}}\bar{\mathbf{x}}+\mathbf{n}}^{-1}\hat{\mathbf{H}}\bar{\mathbf{Q}}\hat{\mathbf{H}}^\dagger\right|\right] \quad (10)$$

where $\mathbf{R}_{\tilde{\mathbf{H}}\bar{\mathbf{x}}+\mathbf{n}}$ is the covariance matrix of the effective noise, $\tilde{\mathbf{H}}\bar{\mathbf{x}}+\mathbf{n}$, and $\bar{\mathbf{Q}} = E[\bar{\mathbf{x}}\bar{\mathbf{x}}^\dagger]$. Since the inputs for different users are independent from each other, $\bar{\mathbf{Q}}$ is a block diagonal matrix, having $\mathbf{Q}_k$ in its diagonals with $\text{tr}(\mathbf{Q}_k) \leq P_{d_k}$. As a result, we have $\bar{\mathbf{H}}\bar{\mathbf{Q}}\bar{\mathbf{H}}^\dagger = \sum_{k=1}^K \mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^\dagger$. In addition, the covariance of the effective noise can be calculated as

$$\mathbf{R}_{\tilde{\mathbf{H}}\bar{\mathbf{x}}+\mathbf{n}} = \mathbf{I} + E\left[\tilde{\mathbf{H}}\bar{\mathbf{x}}\bar{\mathbf{x}}^\dagger\tilde{\mathbf{H}}^\dagger\right] = \mathbf{I} + \sum_{k=1}^K E\left[\tilde{\mathbf{H}}_k\mathbf{Q}_k\tilde{\mathbf{H}}_k^\dagger\right] \quad (11)$$

From [12], [13], we know that $E\left[\tilde{\mathbf{H}}_k\mathbf{Q}_k\tilde{\mathbf{H}}_k^\dagger\right] = \text{tr}(\mathbf{Q}_k\tilde{\mathbf{\Sigma}}_k)\mathbf{I}$. Since our goal is to find the largest lower bound, i.e., the largest achievable rate with Gaussian signaling, we maximize (10) over the entire block

$$R_s = \max_{\substack{(\mathbf{Q}_k,P_{t_k},T_{t_k})\in\mathcal{S}_k \\ \text{tr}(\mathbf{Q}_k)\leq P_{d_k},\forall k}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_{k=1}^K \hat{\mathbf{H}}_k\mathbf{Q}_k\hat{\mathbf{H}}_k^\dagger}{1+\sum_{k=1}^K \text{tr}(\mathbf{Q}_k\tilde{\mathbf{\Sigma}}_k)}\right|\right] \quad (12)$$

where $\mathcal{S}_k = \left\{(\mathbf{Q}_k, P_{t_k}, T_{t_k})\middle|\text{tr}(\mathbf{Q}_k)T_d + P_{t_k}T_{t_k} = P_kT\right\}$, and the coefficient $\frac{T-T_t}{T}$ reflects the amount of time that is spent during the training phase. Note that the maximization is over the parameters of all users, where user k has the training parameters P_{t_k}, and T_{t_k}, and the data transmission parameter $\mathbf{Q}_k$, which can be decomposed into its eigenvectors, i.e., the transmit directions, and eigenvalues, i.e., powers along the transmit directions.

1) Transmit Directions: When the CSI at the receiver is perfect, [7] showed that the eigenvectors of the transmit covariance matrix of each user must be equal to the eigenvectors of the channel covariance matrix of that user, i.e., $\mathbf{U}_{Q_k} = \mathbf{U}_{\Sigma_k}$. In other words, single-user transmit directions strategy is optimum in a multi-user setting as well. In the following theorem, we show that this is also true when there is channel estimation error at the receiver.

Theorem 2: Let $\mathbf{\Sigma}_k = \mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{\Sigma_k}\mathbf{U}_{\Sigma_k}^\dagger$ be the spectral decomposition of the covariance matrix of the channel of user k. Then, the optimum transmit covariance matrix $\mathbf{Q}_k$ of user k has the form $\mathbf{Q}_k = \mathbf{U}_{\Sigma_k}\mathbf{\Lambda}_{Q_k}\mathbf{U}_{\Sigma_k}^\dagger$.

Using Theorem 2, we can write the optimization problem in (12) as,

$$R_s = \max_{\substack{(\boldsymbol{\lambda}_k^Q,P_{t_k},T_{t_k})\in\mathcal{P}_k \\ k=1,\ldots,K}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_k\sum_{i=1}^{n_T}\lambda_{ki}^Q\hat{\lambda}_{ki}^\Sigma\hat{\mathbf{z}}_{ki}\hat{\mathbf{z}}_{ki}^\dagger}{1+\sum_k\sum_{i=1}^{n_T}\lambda_{ki}^Q\tilde{\lambda}_{ki}^\Sigma}\right|\right] \quad (13)$$

where $\mathcal{P}_k = \left\{\left(\boldsymbol{\lambda}_k^Q, P_{t_k}, T_{t_k}\right)\middle|\left(\sum_i\lambda_{ki}^Q\right)T_d + P_{t_k}T_{t_k} = P_kT\right\}$, $\boldsymbol{\lambda}_k^Q = [\lambda_{k1}^Q,\ldots,\lambda_{kn_T}^Q]$, and $\hat{\mathbf{z}}_{ki}$ is an $n_R \times 1$ dimensional i.i.d., zero-mean, identity-covariance Gaussian random vector.

2) Power Allocation Policy: For a MIMO-MAC system with perfect CSI at the receiver and partial CSI at the transmitters, [6] proposes an algorithm to find the optimum power allocation policy. However, the existence of P_{t_k} and T_{t_k} here, violates the symmetry in [6], and changes the form of the objective function. Therefore, in this paper, we modify the algorithm in [6] so that the new algorithm finds the optimum P_{t_k} and T_{t_k} as well as the powers along the transmit directions.

By plugging $\hat{\lambda}_{ki}^{\Sigma}$ and $\tilde{\lambda}_{ki}^{\Sigma}$ into (13), and choosing $\lambda_{ki}^{Q}=0$, for $i=J_k+1,\ldots,n_T$ since they do not contribute to the numerator, we get

$$R_s=\max\frac{T-T_t}{T}E\left[\log\left|\mathbf{I}+\frac{\sum_k\sum_{i=1}^{J_k}\lambda_{ki}^{Q}(\lambda_{ki}^{\Sigma}-\mu_k^S)\hat{\mathbf{z}}_{ki}\hat{\mathbf{z}}_{ki}^{\dagger}}{1+\sum_k\mu_k^S P_{d_k}}\right|\right] \quad (14)$$

where the maximization is over the same set as in (13). In (14), the parameters of the optimization problem are the powers of all users $\lambda_{k1}^{Q},\ldots,\lambda_{kT_{t_k}}^{Q}$ along the transmit directions, the training signal powers P_{t_k}, and the training durations T_{t_k} of all users. Solving for all these variables simultaneously seems intractable. Therefore, we propose a Gauss-Seidel type algorithm that solves (14) iteratively over the users as in [6]. When updating the parameters corresponding to user k, those of the rest of the users are fixed. The optimization problem corresponding to an update of each user becomes

$$R_s^k=\max\frac{T-T_t}{T}E\left[\log\left|\mathbf{\Phi}+\frac{\sum_{i=1}^{J_k}\lambda_{ki}^{Q}(\lambda_{ki}^{\Sigma}-\mu_k^S)\hat{\mathbf{z}}_{ki}\hat{\mathbf{z}}_{ki}^{\dagger}}{\phi+\mu_k^S P_{d_k}}\right|\right] \quad (15)$$

where $\mathbf{\Phi}=\mathbf{I}+\frac{\sum_{l\neq k}^{K}\sum_{i=1}^{J_l}\lambda_{li}^{Q}(\lambda_{li}^{\Sigma}-\mu_l^S)\hat{\mathbf{z}}_{li}\hat{\mathbf{z}}_{li}^{\dagger}}{1+\sum_{l=1}^{K}\mu_l^S P_{d_l}}$, and $\phi=1+\sum_{l\neq k}^{K}\mu_l^S P_{d_l}$. We note that for any pair (P_{t_k},T_{t_k}) that results in $J_k<T_{t_k}$, we can find another pair $(P_{t_k},T_{t_k}^{'})$ that results in a higher rate, and therefore it is sufficient to search over those (P_{t_k},T_{t_k}) pairs that result in $J_k=T_{t_k}$, with an additional constraint of $P_{t_k}>\sum_{i=1}^{T_{t_k}}\left(\frac{1}{\lambda_{T_{t_k}}^{\Sigma}}-\frac{1}{\lambda_i^{\Sigma}}\right)$, which guarantees that, using the pair (P_{t_k},T_{t_k}), all T_{t_k} channels are filled, i.e., $J_k=T_{t_k}$.

The parameters that we want to optimize (15) over are discrete valued T_{t_k}, and continuous valued P_{t_k}, and λ_{ki}^{Q}, for $i=1,\ldots,T_{t_k}$. Since, for every value of T_{t_k}, both the coefficient in front of the expectation, and the number of terms in the sum in the numerator of (15) are different, the form of the objective function is also different. Since T_{t_k} is discrete, and $1\leq T_{t_k}\leq n_T$, we can perform an exhaustive search over T_{t_k} and solve n_T reduced optimization problems with fixed T_{t_k} in each one. Then, we take the solution that results in the maximum rate, i.e.,

$$R_s^k=\max_{1\leq T_{t_k}\leq n_T}R_s^{kT_{t_k}} \quad (16)$$

where $R_s^{kT_{t_k}}$ is equal to (15) with a fixed T_{t_k}. While solving for the inner maximization problem, we define $f_{ki}(P_{t_k})=\frac{\lambda_{ki}^{\Sigma}-\mu_k^S}{\phi+\mu_k^S P_{d_k}}$, for $i=1,\ldots,T_{t_k}$. In this case, the inner optimization problem becomes

$$R_s^{kT_{t_k}}=\max_{(\boldsymbol{\lambda}^Q,P_t)\in\mathcal{R}_{kT_{t_k}}}\frac{T-T_t}{T}E\left[\log\left|\mathbf{\Phi}+\sum_{i=1}^{T_{t_k}}\lambda_{ki}^{Q}f_{ki}(P_{t_k})\hat{\mathbf{z}}_{ki}\hat{\mathbf{z}}_{ki}^{\dagger}\right|\right] \quad (17)$$

In the optimization problem in (17), we have $T_{t_k}+1$ optimization variables, $\lambda_{k1}^{Q},\ldots,\lambda_{T_{t_k}}^{Q}$, and P_{t_k}. The KKT conditions for the optimization problem in (17) can be written as

$$\frac{T_d}{T}f_{ki}(P_{t_k})E\left[\mathbf{z}_{ki}^{\dagger}\mathbf{B}^{-1}\mathbf{z}_{ki}\right]\leq\mu_k T_d,\ i=1,\ldots,T_{t_k} \quad (18)$$

$$\frac{T_d}{T}\sum_{i=1}^{T_{t_k}}\lambda_{ki}^{Q}E\left[\mathbf{z}_{ki}^{\dagger}\mathbf{B}^{-1}\mathbf{z}_{ki}\right]\frac{\partial f_{ki}(P_{t_k})}{\partial P_{t_k}}=\mu_k T_{t_k} \quad (19)$$

where $\mathbf{B}=\mathbf{\Phi}+\sum_{i=1}^{T_{t_k}}\lambda_{ki}^{Q}f_{ki}(P_{t_k})\hat{\mathbf{z}}_{ki}\hat{\mathbf{z}}_{ki}^{\dagger}$, μ_k is the Lagrange multiplier, and the equality of the last equation follows from the complementary slackness condition. Note that when the optimum λ_{ki}^{Q} is non-zero, the corresponding inequality in (18) will be satisfied with equality. Therefore, we pull the expectation terms from (18) for those equations with non-zero λ_{ki}^{Q}'s, and insert them into (19). Since those indices with $\lambda_{ki}^{Q}=0$ do not contribute to (19), we have

$$\sum_{i=1}^{T_{t_k}}\lambda_{ki}^{Q}\frac{f_{ki}'(P_{t_k})}{f_{ki}(P_{t_k})}=\frac{T_{t_k}}{T_d}. \quad (20)$$

Now, we have a fixed-point equation which does not include any expectation terms. We can use this to solve P_{t_k} in terms of λ_{ki}^{Q}'s. Using our single-user results in [12], [13], we propose the following algorithm that first solves $P_{t_k}(n+1)$ from

$$\sum_{i=1}^{T_{t_k}}\lambda_{ki}^{Q}(n)\frac{f_{ki}'(P_{t_k}(n+1))}{f_{ki}(P_{t_k}(n+1))}=\frac{T_{t_k}}{T_d} \quad (21)$$

and then, updates $\lambda_{ki}^{Q}(n+1)$ using

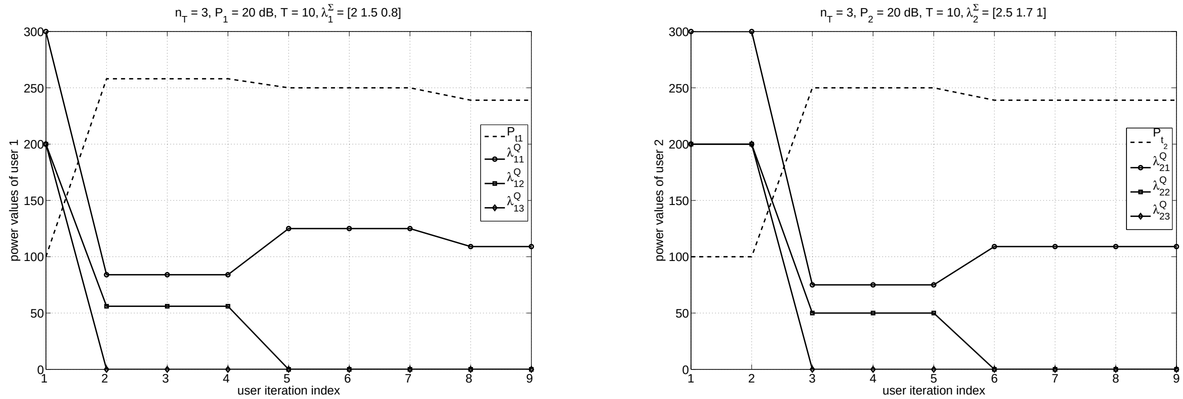
$$\lambda_{ki}^{Q}(n+1)=\frac{\lambda_{ki}^{Q}(n)f_{ki}(P_{t_k}(n+1))E\left[\mathbf{z}_{ki}^{\dagger}\mathbf{B}^{-1}\mathbf{z}_{ki}\right]P_{d_k}(n)}{\sum_{j=1}^{T_{t_k}}\lambda_{kj}^{Q}(n)f_{kj}(P_{t_k}(n+1))E\left[\mathbf{z}_{kj}^{\dagger}\mathbf{B}^{-1}\mathbf{z}_{kj}\right]} \quad (22)$$

where $P_{d_k}(n)=\frac{(P_k T-P_{t_k}(n+1)T_{t_k})}{T_d}$. This algorithm finds the solution of (17), when T_{t_k} and the parameters of the rest of the users are fixed. We run n_T such algorithms simultaneously for user k. The solution of (15) can be found by taking the one that results in the largest rate. Now, we know $\boldsymbol{\lambda}_k^{Q},P_{t_k},T_{t_k}$, that maximize (15). We then move to another user, and solve (17) for this user, and in this manner we iterate over the users in a round-robin fashion. Finally, round-robin algorithm gives us the optimum parameters of all users that maximize (14).

As a result, through P_{t_k}, we find the power allocation of user k over the training and data transmission phases. Through $T_t=\sum_{k=1}^{K}T_{t_k}$, we find the optimum allocation of time over the training and data transmission phases. Through Theorem 2, we find the optimum transmit directions of user k, and through $\lambda_{k1}^{Q},\ldots,\lambda_{T_{t_k}}^{Q}$, we find the allocation of data transmission power of user k over these transmit directions. Finally, the optimum training signal of user k, $\mathbf{S}_k$, is determined by the optimum T_{t_k} and P_{t_k} through Theorem 1.

IV. Numerical Analysis

Analytical proof of the convergence of this algorithm seems to be intractable for now. However, in our extensive sim-



(a) (b)

Fig. 1. The convergence of the multi-user algorithm with $n_T = n_R = 3$, 20 dB total average power and $T = 10$: (a) convergence for user 1; (b) convergence for user 2; convergence for user 3 is similar and omitted here due to space limitations.

ulations, we observed that the algorithm always converges. In Figure 1, we considered a system of $K = 3$ users with $n_T = n_R = 3$, all having moderate power, $P = 20$ dB, and moderate block length, $T = 10$. Each iteration in Figure 1, corresponds to solving (15) for one of the users, while the parameters of the rest of the users are fixed. We observe that, all users estimate only one dimension of the channel. Therefore, total training duration is $T_t = 3$ symbols.

We observed through extensive simulations that for a large set of channel eigenvalues, total available power and the block length, all users estimate only one dimension of the channel. In order to estimate a second dimension, either very large levels of power or a long enough coherence time is needed. We refer the reader to [18] for further simulation results, which we cannot show here due to space limitations.

V. CONCLUSIONS

We considered a block-fading MIMO-MAC, where the receiver has a noisy estimate of the channel and the transmitters have partial CSI. Each transmission block is divided into a training phase and a data transmission phase. During the training phase, we showed that the users should send non-overlapping training signals. During the data transmission phase, we formulated an optimization problem to maximize the achievable sum-rate jointly over the training signal durations of all users, the training powers of all users, and the transmit covariance matrices of all users. We proposed a multi-user algorithm that solves the problem iteratively over the users in a round-robin fashion, by utilizing a single-user algorithm similar to the one proposed in [12], [13], for an update of each user. Although the theoretical convergence proof of the proposed algorithm remains as open problem, through extensive simulations, we observed that the proposed algorithm always converged.

REFERENCES

[1] A. J. Goldsmith and P. P. Varaiya. Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory*, 43(6):1986–1992, November 1997.

[2] İ. E. Telatar. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunication*, 10(6):585–596, November 1999.

[3] E. Visotsky and U. Madhow. Space-time transmit precoding with imperfect feedback. *IEEE Transactions on Information Theory*, 47(6):2632–2639, September 2001.

[4] S. A. Jafar and A. Goldsmith. Transmitter optimization and optimality of beamforming for multiple antenna systems. *IEEE Transactions on Wireless Communications*, 3(4):1165–1175, July 2004.

[5] H. Boche and E. Jorswieck. On the optimality range of beamforming for MIMO systems with covariance feedback. *IEICE Trans. Commun.*, E85-A(11):2521–2528, November 2002.

[6] A. Soysal and S. Ulukus. Optimum power allocation for single-user MIMO and multi-user MIMO-MAC with partial CSI. *IEEE Journal on Selected Areas in Communications*, 25(7):1402–1412, September 2007.

[7] A. Soysal and S. Ulukus. Optimality of beamforming in fading MIMO multiple access channels. *IEEE Transactions on Communications*, 2008. Under review.

[8] M. Médard. The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. *IEEE Transactions on Information Theory*, 46(3):933–946, May 2000.

[9] T. E. Klein and R. G. Gallager. Power control for the additive white Gaussian noise channel under channel estimation errors. In *ISIT*, June 2001.

[10] T. Yoo and A. Goldsmith. Capacity and power allocation for fading MIMO channels with channel estimation error. *IEEE Transactions on Information Theory*, 52(5):2203–2214, May 2006.

[11] B. Hassibi and B. M. Hochwald. How much training is needed in multiple-antenna wireless links? *IEEE Transactions on Information Theory*, 49(4):951–963, April 2003.

[12] A. Soysal and S. Ulukus. Optimizing the rate of a correlated MIMO link jointly over channel estimation and data transmission parameters. In *Conference on Information Sciences and Systems*, March 2008.

[13] A. Soysal and S. Ulukus. Joint channel estimation and resource allocation for MIMO systems–Part I: Single-user analysis. *IEEE Transactions on Wireless Communications*. Submitted, available at http://staff.eng.bahcesehir.edu.tr/˜asoysal/.

[14] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi. Iterative water-filling for Gaussian vector multiple access channels. *IEEE Transactions on Information Theory*, 50(1):145–151, January 2004.

[15] C. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela. Capacity scaling in MIMO wireless systems under correlated fading. *IEEE Transactions on Information Theory*, 48(3):637–650, March 2002.

[16] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[17] E. W. Kamen and J. Su. *Introduction to Optimal Estimation*. Springer, 1999.

[18] A. Soysal and S. Ulukus. Joint channel estimation and resource allocation for MIMO systems–Part II: Multi-user and numerical analysis. *IEEE Transactions on Wireless Communications*. Submitted, available at http://staff.eng.bahcesehir.edu.tr/˜asoysal/.