

# Joint Channel Estimation and Resource Allocation for MIMO Systems—Part II: Multi-User and Numerical Analysis

Alkan Soysal, *Member, IEEE*, and Sennur Ulukus, *Member, IEEE*

**Abstract**—This is the second part of a two-part paper on the joint channel estimation and resource allocation problem in MIMO systems with noisy channel estimation at the receiver side and partial CSI, in the form of covariance feedback, available at the transmitter side. We consider transmit-side correlated MIMO channels with block fading, where each block is divided into training and data transmission phases. In this paper, we extend the single-user results of Part I to the multiple access channel. For the data transmission phase, we propose an iterative algorithm to solve for the optimum system resources such as time, power and spatial dimensions. This algorithm 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. Finally, we provide a detailed numerical analysis to support the analytical results of both parts of this two-part paper.

**Index Terms**—Multi-user MIMO, MIMO multiple access channel, partial CSI, covariance feedback, optimum power allocation, channel estimation.

## I. INTRODUCTION

**I**N Part I [2] of this paper, 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, transmit-side correlated MIMO channel with noisy channel estimation at the receiver, and partial channel state information (CSI) available at the transmitter. We concluded that there is a trade-off between estimating the channel better and increasing the achievable data rate [2].

In wireless communication scenarios, the achievable rate of the system depends crucially on the amount of CSI available at the receivers and the transmitters. The CSI is observed only by the receiver, which can estimate it and feed the estimated CSI back to the transmitter. However, measuring the CSI and providing CSI feedback to the transmitter uses communication resources, such as time, power and spatial dimensions which

could otherwise be used for useful information transmission. In a multi-user setting, the amount of resources required to measure the channel and provide CSI feedback to the transmitter increases substantially. When perfect channel information is assumed to be available at the receiver and the transmitters at no cost, [3] 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, [4], [5] finds the optimum transmit strategies for all users. There has been works such as [2], [6], [7], [8] that considered noisy channel estimates at the receiver, but they restricted attention to a single-user scenario.

Part II extends the results of Part I to multiple access channels. In a multi-user setting, we first 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 orthogonal 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. However, as a result of having shorter individual training signal duration and the conservation of energy, the training signal power that is used by a particular user in a multi-user case could be larger than the training signal power that the same user would use in a single-user case. This higher effective training power results in a better channel estimate, and hence lower channel estimation error. Therefore, although fewer dimensions of the channel are estimated, the channel estimation error corresponding to those estimated dimensions become smaller.

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. Finally, we provide detailed simulation results that investigates the effects of the power constraint, coherence interval (block length), and channel covariance matrices on our results.

Manuscript received January 23, 2008; revised August 21, 2009; accepted November 12, 2009. The associate editor coordinating the review of this paper and approving it for publication was M. Chiang.

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, and presented in part at the Globecom, New Orleans, LA, 2008 [1].

A. Soysal is with the Department of Electrical and Electronics Engineering, Bahçeşehir University, Istanbul, Turkey (e-mail: alkan.soysal@bahcesehir.edu.tr).

S. Ulukus is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA (e-mail: ulukus@umd.edu).  
Digital Object Identifier 10.1109/TWC.2010.02.080100

## II. SYSTEM MODEL

We consider a multiple access channel (MAC), which corresponds to the uplink of a cellular system, with multiple transmit antennas at every user and multiple receive antennas at 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

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

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  $P_k = \frac{1}{T} E[\sum_n \mathbf{x}_{kn}^\dagger \mathbf{x}_{kn}]$ , averaged over  $T$  symbols<sup>1</sup>.

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. By generalizing the single-user model of Part I [2], the channel can be written as

$$\mathbf{H}_k = \mathbf{Z}_k \boldsymbol{\Sigma}_k^{1/2} \quad (2)$$

where the entries of  $\mathbf{Z}_k$  are i.i.d., zero-mean, unit-variance complex Gaussian random variables. From this point on, we will refer to matrix  $\boldsymbol{\Sigma}_k$  as the channel covariance feedback matrix of user  $k$ . Similar covariance feedback models have been used in [4], [5], [9], [10].

## III. JOINT OPTIMIZATION FOR MULTI-USER MIMO

Note that in our model, a transmission block is divided into training and data transmission phases. During the training phase, each user has training signal  $\mathbf{S}_k$ , training signal power  $P_{t_k}$ , and training signal duration  $T_t$ . During the data transmission phase, each user has data transmission power  $P_{d_k}$ , which appears as a constraint on the trace of the transmit covariance matrix. Our goal in this paper is to find the optimum values of these training and the data transmission parameters.

<sup>1</sup>Note that since the noise power is assumed to be unity,  $P_k$  is in fact the relative power with respect to noise power. It can be regarded as an SNR value.

### A. Training and Channel Estimation Phase

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

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

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  $n_R \times T$  dimensional received signal and noise matrices, respectively. The  $n^{\text{th}}$  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  $\frac{1}{T_t} \text{tr}(\mathbf{S}_k \mathbf{S}_k^\dagger) \leq P_{t_k}$ .

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

$$\mathbf{R}_t = \bar{\mathbf{H}} \bar{\mathbf{S}} + \mathbf{N}_t \quad (4)$$

where  $\bar{\mathbf{H}} = [\mathbf{H}_1, \dots, \mathbf{H}_K]$  is an  $n_R \times Kn_T$  dimensional channel matrix, and  $\bar{\mathbf{S}} = [\mathbf{S}_1^\dagger, \dots, \mathbf{S}_K^\dagger]^\dagger$  is a  $Kn_T \times T_t$  dimensional training signal matrix. Note that, we put the channel matrices next to each other to form longer rows, and the training symbols on top of each other to form longer columns. 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  $i$  of the channel matrix is i.i.d. with row  $j$ . 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  $E[\mathbf{h}_{ki} \mathbf{h}_{ki}^\dagger] = \boldsymbol{\Sigma}_k, i = 1, \dots, n_R$ , and row  $i$  of  $\bar{\mathbf{H}}$  as  $\bar{\mathbf{h}}_i = [\mathbf{h}_{1i}^\dagger, \dots, \mathbf{h}_{Ki}^\dagger]^\dagger$ , where  $\bar{\boldsymbol{\Sigma}} = E[\bar{\mathbf{h}}_i \bar{\mathbf{h}}_i^\dagger] = \text{diag}\{\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_K\}$  is a block diagonal matrix, having  $\boldsymbol{\Sigma}_k$  on its diagonals.

Let the eigenvalue representation of the channel covariance matrix of user  $k$  be  $\boldsymbol{\Sigma}_k = \mathbf{U}_{\boldsymbol{\Sigma}_k} \boldsymbol{\Lambda}_{\boldsymbol{\Sigma}_k} \mathbf{U}_{\boldsymbol{\Sigma}_k}^\dagger$ , then the eigenvectors of the stacked channel covariance matrix  $\bar{\boldsymbol{\Sigma}} = \bar{\mathbf{U}}_{\bar{\boldsymbol{\Sigma}}} \bar{\boldsymbol{\Lambda}}_{\bar{\boldsymbol{\Sigma}}} \bar{\mathbf{U}}_{\bar{\boldsymbol{\Sigma}}}^\dagger$  can also be written as  $\bar{\mathbf{U}}_{\bar{\boldsymbol{\Sigma}}} = \text{diag}\{\mathbf{U}_{\boldsymbol{\Sigma}_1}, \dots, \mathbf{U}_{\boldsymbol{\Sigma}_K}\}$  [11, 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  $i^{\text{th}}$  row of (4), which represents the received signal at the  $i^{\text{th}}$  antenna of the receiver over the training duration, can be written as

$$\mathbf{r}_{ti} = \bar{\mathbf{S}}^\dagger \bar{\mathbf{h}}_i + \mathbf{n}_{ti}. \quad (5)$$

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 Part I [2]. 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 optimum MMSE estimator can be found as  $\bar{\mathbf{M}}^* = \bar{\boldsymbol{\Sigma}} \bar{\mathbf{S}} (\bar{\mathbf{S}}^\dagger \bar{\boldsymbol{\Sigma}} \bar{\mathbf{S}} + \mathbf{I})^{-1}$

[2], and the mean square error becomes,

$$\min_{\mathbf{M}} E \left[ \tilde{\mathbf{h}}_i^{\dagger} \tilde{\mathbf{h}}_i \right] = \text{tr} \left( (\bar{\Sigma}^{-1} + \bar{\mathbf{S}}\bar{\mathbf{S}}^{\dagger})^{-1} \right) \quad (6)$$

Note that mean square error of the channel estimation process can be further decreased by choosing the optimum training signal  $\bar{\mathbf{S}}^*$  to minimize (6). The following theorem finds  $\bar{\mathbf{S}}^*$ , and training signals of individual users  $\mathbf{S}_k^*$ , for a given training power and training duration.

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

$$\lambda_{ki}^S = \left( \frac{1}{\mu_k^S} - \frac{1}{\lambda_{ki}^S} \right)^+, \quad i = 1, \dots, \min(n_T, T_{t_k}) \quad (7)$$

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

*Proof:* Let us have  $\bar{\mathbf{S}} = \bar{\mathbf{U}}_{\Sigma} \bar{\Lambda}_{\Sigma}^{1/2} \bar{\mathbf{V}}_{\Sigma}^{\dagger}$ . The expression in (6) is minimized when  $\bar{\Sigma}^{-1}$  and  $\bar{\mathbf{S}}\bar{\mathbf{S}}^{\dagger}$  have the same eigenvectors [12]. Therefore, we have  $\bar{\mathbf{U}}_{\Sigma} = \bar{\mathbf{U}}_{\Sigma}$ . Since,  $\bar{\mathbf{S}}\bar{\mathbf{S}}^{\dagger} = \bar{\mathbf{U}}_{\Sigma} \bar{\Lambda}_{\Sigma} \bar{\mathbf{U}}_{\Sigma}^{\dagger}$ , and the unitary matrix  $\bar{\mathbf{V}}_{\Sigma}$  does not appear in the objective function and the constraint, we can choose  $\bar{\mathbf{V}}_{\Sigma} = \mathbf{I}$ . Now, we have

$$\bar{\mathbf{S}} = \bar{\mathbf{U}}_{\Sigma} \bar{\Lambda}_{\Sigma}^{1/2} \quad (8)$$

$$\begin{bmatrix} \mathbf{S}_1 \\ \vdots \\ \mathbf{S}_K \end{bmatrix} = \begin{bmatrix} \mathbf{U}_{\Sigma_1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{U}_{\Sigma_K} \end{bmatrix} \begin{bmatrix} \Lambda_{\Sigma_1}^{1/2} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \Lambda_{\Sigma_K}^{1/2} \end{bmatrix} \quad (9)$$

where each user has  $\mathbf{S}_k = [\mathbf{0}, \dots, \mathbf{0}, \mathbf{U}_{\Sigma_k} \Lambda_{\Sigma_k}^{1/2}, \mathbf{0}, \dots, \mathbf{0}]$ . Note that  $\mathbf{S}_k$  is an  $n_T \times T_t$  dimensional matrix, and  $\mathbf{U}_{\Sigma_k}$  is an  $n_T \times n_T$  dimensional matrix. Let us denote the dimension of  $\Lambda_{\Sigma_k}$  as  $n_T \times T_{t_k}$  in such a way that  $\sum_{k=1}^K T_{t_k} = T_t$ .

Inserting  $\bar{\mathbf{S}}$  into (6), the optimization problem can be written as

$$\tilde{\sigma}_{sum} = \min_{\substack{\text{tr}(\Lambda_{\Sigma_k}) \leq P_{t_k} T_{t_k} \\ k=1, \dots, K}} \text{tr} \left( (\bar{\Lambda}_{\Sigma}^{-1} + \bar{\Lambda}_{\Sigma})^{-1} \right) \quad (10)$$

$$= \min_{\substack{\text{tr}(\Lambda_{\Sigma_k}) \leq P_{t_k} T_{t_k} \\ k=1, \dots, K}} \sum_{k=1}^K \text{tr} \left( (\Lambda_{\Sigma_k}^{-1} + \Lambda_{\Sigma_k})^{-1} \right). \quad (11)$$

The Lagrangian of the problem in (11) can be written as

$$\sum_{k=1}^K \sum_{i=1}^{n_T} \frac{1}{\frac{1}{\lambda_{ki}^S} + \lambda_{ki}^S} + \sum_{k=1}^K (\mu_k^S)^2 \left( \sum_{i=1}^{n_T} \lambda_{ki}^S - P_{t_k} T_{t_k} \right) \quad (12)$$

where  $(\mu_k^S)^2$  are the Lagrange multipliers. The Lagrangian is a convex function of  $\lambda_{ki}^S$ 's, therefore the solution that satisfies the KKT conditions is the unique optimum solution. This gives us (7), which is water-filling the available power of each user over the eigenvalues of its own channel covariance matrix. In order to calculate  $\mu_k^S$ , we sum both sides of (7) over all

antennas to get  $\mu_k^S = \frac{J_k}{P_{t_k} + \sum_{i=1}^{J_k} \frac{1}{\lambda_{ki}^S}}$ , where  $J_k$  is the largest index that has non-zero  $\lambda_{ki}^S$ .  $\square$

Similar to the single-user case, for any given  $P_{t_k}$ , and  $T_{t_k} > n_T$ , increasing  $T_{t_k}$  beyond  $n_T$  does not result in better channel estimates. However, larger  $T_{t_k}$  will result in smaller data transmission length, and will decrease the achievable rate of the data transmission phase. Therefore, it is sufficient to consider only  $T_{t_k} \leq n_T$ , which we will assume for the rest of this paper. Due to the constraint on the training duration, intuitively fewer dimensions of the individual channels will be estimated for each user. In a multi-user setting, training duration cannot be the number of users multiplied with the training duration of each user in a single-user environment, since the total training duration will need to be less than the block length. Therefore, we conclude that individual training duration of a user in a multi-user setting will be shorter than the training duration of that user in a single-user environment. Since the total energy is the same, training power will be higher for a shorter training duration. Higher training power in turn, will result in a better channel estimate, and hence lower channel estimation error. As a result, although fewer dimensions of the channel will be estimated, the channel estimation error corresponding to those estimated dimensions will be smaller. It was shown in other contexts as well that the degrees of freedom of a MAC does not increase by increasing the number of users. For example in [13], the degrees of freedom is limited by the number of receive antennas. In our case, it is limited by the duration of the training phase, which itself depends on several variables including the number of receive antennas.

Theorem 1 states that orthogonality in the time domain holds over the users in a multi-user setting as well. Since the receiver can stack up the channels for the channel estimation process, and the resulting stacked channel covariance matrix is block-diagonal, the problem is equivalent to a single-user problem where all transmit antennas are on the same unit, but antennas are put into  $K$  groups. Each group is correlated within the group, but groups are uncorrelated, which results in a block diagonal channel covariance matrix. Since this is an equivalent single-user problem, training signals of different users are orthogonal in time. This is achieved by transmitting the training signal of user  $k$  during its own time slot for  $T_{t_k}$  channel uses.

Note that  $\mu_k^S$  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. After the sum-rate maximization is performed, the optimum  $P_{t_k}$  and  $T_{t_k}$  will be found, and this in turn, will give us the  $\mathbf{S}_k^*$  through Theorem 1.

Next, 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, which can be found in Part I [2]. The covariance matrix of the channel estimation error of user  $k$  can be found as  $\tilde{\Sigma}_k = \mathbf{U}_{\Sigma_k} (\Lambda_{\Sigma_k}^{-1} + \Lambda_{\Sigma_k})^{-1} \mathbf{U}_{\Sigma_k}^{\dagger}$ ,

where the eigenvalues are given as  $\hat{\lambda}_{ki}^\Sigma = \min(\lambda_{ki}^\Sigma, \mu_k^S)$ . Similarly, the covariance matrix of the estimated channel of user  $k$  can be found using the orthogonality principle as  $\hat{\Sigma} = \bar{\mathbf{U}}_\Sigma (\bar{\Lambda}_\Sigma - \tilde{\Lambda}_\Sigma) \bar{\mathbf{U}}_\Sigma^\dagger$ , where the eigenvalues are given as  $\hat{\lambda}_{ki}^\Sigma = \max(\lambda_{ki}^\Sigma - \mu_k^S, 0)$ .

### B. Data Transmission Phase

The sum-rate of a multiple access channel can be derived using the stacked channel and input matrices. We can write (1) as

$$\mathbf{r} = \sum_{k=1}^K \hat{\mathbf{H}}_k \mathbf{x}_k + \sum_{k=1}^K \tilde{\mathbf{H}}_k \mathbf{x}_k + \mathbf{n} = \hat{\mathbf{H}} \bar{\mathbf{x}} + \tilde{\mathbf{H}} \bar{\mathbf{x}} + \mathbf{n} \quad (13)$$

where  $\hat{\mathbf{H}} = [\hat{\mathbf{H}}_1, \dots, \hat{\mathbf{H}}_K]$ ,  $\tilde{\mathbf{H}} = [\tilde{\mathbf{H}}_1, \dots, \tilde{\mathbf{H}}_K]$  are  $n_R \times K n_T$  dimensional, and  $\bar{\mathbf{x}} = [\mathbf{x}_1^\dagger, \dots, \mathbf{x}_K^\dagger]^\dagger$  is  $K n_T \times 1$  dimensional. Although the optimum input distribution is not known, we achieve the following lower bound with Gaussian  $\bar{\mathbf{x}}$  [7],

$$C_{lb}^{sum} = I(\mathbf{r}; \bar{\mathbf{x}} | \hat{\mathbf{H}}) \geq E_{\hat{\mathbf{H}}} \left[ \log \left| \mathbf{I} + \mathbf{R}_{\hat{\mathbf{H}} \bar{\mathbf{x}} + \mathbf{n}}^{-1} \hat{\mathbf{H}} \hat{\mathbf{Q}} \hat{\mathbf{H}}^\dagger \right| \right] \quad (14)$$

where  $\mathbf{R}_{\hat{\mathbf{H}} \bar{\mathbf{x}} + \mathbf{n}}$  is the covariance matrix of the effective noise,  $\hat{\mathbf{H}} \bar{\mathbf{x}} + \mathbf{n}$ , and  $\hat{\mathbf{Q}} = E[\bar{\mathbf{x}} \bar{\mathbf{x}}^\dagger]$ . Since the inputs for different users are independent from each other,  $\hat{\mathbf{Q}}$  is a block diagonal matrix, having  $\mathbf{Q}_k$  in its diagonals with the power constraint<sup>2</sup>  $\text{tr}(\mathbf{Q}_k) \leq P_{d_k}$ . As a result, we have  $\hat{\mathbf{H}} \hat{\mathbf{Q}} \hat{\mathbf{H}}^\dagger = \sum_{k=1}^K \hat{\mathbf{H}}_k \mathbf{Q}_k \hat{\mathbf{H}}_k^\dagger$ . The covariance of the effective noise can be calculated as  $\mathbf{R}_{\hat{\mathbf{H}} \bar{\mathbf{x}} + \mathbf{n}} = \mathbf{I} + \sum_{k=1}^K E_{\tilde{\mathbf{H}}_k} [\tilde{\mathbf{H}}_k \mathbf{Q}_k \tilde{\mathbf{H}}_k^\dagger]$ . From Part I [2], we know that  $E[\tilde{\mathbf{H}}_k \mathbf{Q}_k \tilde{\mathbf{H}}_k^\dagger] = \text{tr}(\mathbf{Q}_k \tilde{\Sigma}_k) \mathbf{I}$ . Since our goal is to find the largest lower bound, i.e., the largest achievable rate with Gaussian signaling, maximization of (14) over the entire block becomes

$$R_{sum} = \max_{\substack{(\mathbf{Q}_k, P_{t_k}, T_{t_k}) \in \mathcal{S}_k \\ \text{tr}(\mathbf{Q}_k) \leq P_{d_k}, \forall k}} T_d E_{\hat{\mathbf{H}}} \left[ \log \left| \mathbf{I} + \frac{\sum_{k=1}^K \hat{\mathbf{H}}_k \mathbf{Q}_k \hat{\mathbf{H}}_k^\dagger}{1 + \sum_{k=1}^K \text{tr}(\mathbf{Q}_k \tilde{\Sigma}_k)} \right| \right] \quad (15)$$

where  $\mathcal{S}_k = \left\{ (\mathbf{Q}_k, P_{t_k}, T_{t_k}) \mid \text{tr}(\mathbf{Q}_k) T_d + P_{t_k} T_{t_k} = P_k T \right\}$ , and the coefficient  $T_d = \frac{T - T_t}{T}$  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 (the transmit directions) and eigenvalues (powers along the transmit directions).

1) *Transmit Directions*: When the CSI at the receiver is perfect, [5] 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 case 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  $\Sigma_k = \mathbf{U}_{\Sigma_k} \Lambda_{\Sigma_k} \mathbf{U}_{\Sigma_k}^\dagger$  be the spectral decomposition of the covariance matrix of the channel of user

<sup>2</sup>Note that since the noise power is assumed to be unity,  $p_{d_k}$  and the eigenvalues of  $Q_k$  are relative power values with respect to the noise power. They can be regarded as SNR values.

$k$ . Then, the optimum transmit covariance matrix  $\mathbf{Q}_k$  of user  $k$  has the form  $\mathbf{Q}_k = \mathbf{U}_{\Sigma_k} \Lambda_{Q_k} \mathbf{U}_{\Sigma_k}^\dagger$ .

The proof of Theorem 2 follows the proof of Theorem 2 in Part I of the paper [2] quite closely, and therefore is omitted here due to space limitations. It can be found in [14]. Using Theorem 2, we can write the optimization problem in (15) as,

$$R_{sum} = \max_{\substack{(\lambda_k^Q, P_{t_k}, T_{t_k}) \in \mathcal{P}_k \\ k=1, \dots, K}} T_d E_{\hat{\mathbf{z}}_{ki}} \left[ \log \left| \mathbf{I} + \frac{\sum_k \sum_{i=1}^{n_T} \lambda_{ki}^Q \hat{\lambda}_{ki}^\Sigma \hat{\mathbf{z}}_{ki} \hat{\mathbf{z}}_{ki}^\dagger}{1 + \sum_k \sum_{i=1}^{n_T} \lambda_{ki}^Q \hat{\lambda}_{ki}^\Sigma} \right| \right] \quad (16)$$

where  $\hat{\mathbf{z}}_{ki}$ , which is an  $n_R \times 1$  dimensional i.i.d., zero-mean, identity-covariance Gaussian random vector, is the  $i^{\text{th}}$  column of  $\hat{\mathbf{Z}}_{ki}$ ,  $\lambda_k^Q = [\lambda_{k1}^Q, \dots, \lambda_{kn_T}^Q]$ , and  $\mathcal{P}_k = \left\{ (\lambda_k^Q, P_{t_k}, T_{t_k}) \mid \left( \sum_{i=1}^{n_T} \lambda_{ki}^Q \right) T_d + P_{t_k} T_{t_k} = P_k T \right\}$ .

2) *Power Allocation Policy*: For a MIMO-MAC system with perfect CSI at the receiver and partial CSI at the transmitters, [4] proposes an algorithm to find the optimum power allocation policy. However, the algorithm in [4] is not suitable to find the optimum values of  $P_{t_k}$  and  $T_{t_k}$ , if directly applied to the model in this paper. Existence of  $P_{t_k}$  and  $T_{t_k}$  violates the symmetry in [4], and changes the form of the objective function. Therefore, in this paper, we modify the algorithm in [4] 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 (16), choosing  $\lambda_{ki}^Q = 0$ , for  $i = J_k + 1, \dots, n_T$  due to the fact that they do not contribute to the numerator, we get

$$R_{sum} = \max_{\substack{(\lambda_k^Q, P_{t_k}, T_{t_k}) \in \mathcal{P}_k \\ k=1, \dots, K}} T_d E \left[ \log \left| \mathbf{I} + \frac{\sum_k \sum_{i=1}^{J_k} \lambda_{ki}^Q (\lambda_{ki}^\Sigma - \mu_k^S) \hat{\mathbf{z}}_{ki} \hat{\mathbf{z}}_{ki}^\dagger}{1 + \sum_k \mu_k^S P_{d_k}} \right| \right] \quad (17)$$

In (17), the parameters of the optimization problem are the powers of all users  $\lambda_{k1}^Q, \dots, \lambda_{kT_{t_k}}^Q$  along the transmit directions, the training signal powers  $P_{t_k}$  of all users, 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, round-robin algorithm that solves (17) iteratively over the users as in [4]. When updating the parameters corresponding to user  $k$ , we assume that the parameters of the rest of the users are fixed. The optimization problem corresponding to an update of each user becomes

$$R_{sum}^k = \max_{(\lambda_k^Q, P_{t_k}, T_{t_k}) \in \mathcal{P}_k} T_d E \left[ \log \left| \Phi + \frac{\sum_{i=1}^{J_k} \lambda_{ki}^Q (\lambda_{ki}^\Sigma - \mu_k^S) \hat{\mathbf{z}}_{ki} \hat{\mathbf{z}}_{ki}^\dagger}{\phi + \mu_k^S P_{d_k}} \right| \right] \quad (18)$$

where  $\Phi = \mathbf{I} + \frac{\sum_{l \neq k} \sum_{i=1}^{J_l} \lambda_{li}^Q (\lambda_{li}^\Sigma - \mu_l^S) \hat{\mathbf{z}}_{li} \hat{\mathbf{z}}_{li}^\dagger}{1 + \sum_{l=1}^K \mu_l^S P_{d_l}}$ , and  $\phi = 1 + \sum_{l \neq k} \mu_l^S P_{d_l}$ . Note that the optimization problem in (18) is now a single-user problem with fixed interference from the other users. Therefore, we can follow arguments similar to those in the single-user case in Part I [2]. Since 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, it is sufficient to search over those  $(P_{t_k}, T_{t_k})$  pairs that results in  $J_k = T_{t_k}$ . We can

now write (18) as

$$R_{sum}^k = \max_{(\lambda_k^Q, P_{t_k}, T_{t_k}) \in \mathcal{R}_k} T_d E \left[ \log \left| \Phi + \frac{\sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q (\lambda_{ki}^\Sigma - \mu_k^S) \hat{\mathbf{z}}_{ki} \hat{\mathbf{z}}_{ki}^\dagger}{\phi + \mu_k^S P_{d_k}} \right| \right] \quad (19)$$

where  $\mathcal{R}_k = \left\{ (\lambda_k^Q, P_{t_k}, T_{t_k}) \mid \left( \sum_{i=1}^{n_T} \lambda_{ki}^Q \right) T_d + P_{t_k} T_{t_k} = P_k T, P_{t_k} > \sum_{i=1}^{T_{t_k}} \left( \frac{1}{\lambda_{ki}^\Sigma} - \frac{1}{\lambda_i^\Sigma} \right) \right\}$ , and the condition  $P_{t_k} > \sum_{i=1}^{T_{t_k}} \left( \frac{1}{\lambda_{ki}^\Sigma} - \frac{1}{\lambda_i^\Sigma} \right)$  guarantees that, using the pair  $(P_{t_k}, T_{t_k})$ , all  $T_{t_k}$  channels are filled, i.e.,  $J_k = T_{t_k}$ . Note that the parameters that we want to optimize (19) over are discrete valued  $T_{t_k}$ , and continuous valued  $P_{t_k}$ , and  $\lambda_{ki}^Q$ , for  $i = 1, \dots, T_{t_k}$ . 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.,

$$R_{sum}^k = \max_{1 \leq T_{t_k} \leq n_T} R_{sum}^{k, T_{t_k}} \quad (20)$$

where

$$R_{sum}^{k, T_{t_k}} = \max_{(\lambda^Q, P_t) \in \mathcal{R}_{kT_{t_k}}} T_d E \left[ \log \left| \Phi + \frac{\sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q (\lambda_{ki}^\Sigma - \mu_k^S) \hat{\mathbf{z}}_{ki} \hat{\mathbf{z}}_{ki}^\dagger}{\phi + \mu_k^S P_{d_k}} \right| \right] \quad (21)$$

and  $\mathcal{R}_{kT_{t_k}} = \left\{ (\lambda_k^Q, P_{t_k}) \mid \left( \sum_{i=1}^{n_T} \lambda_{ki}^Q \right) T_d + P_{t_k} T_{t_k} = P_k T, P_{t_k} > \sum_{i=1}^{T_{t_k}} \left( \frac{1}{\lambda_{ki}^\Sigma} - \frac{1}{\lambda_i^\Sigma} \right) \right\}$ . 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  $i = 1, \dots, T_{t_k}$ . In this case, the inner optimization problem becomes

$$R_{sum}^{k, T_{t_k}} = \max_{(\lambda^Q, P_t) \in \mathcal{R}_{kT_{t_k}}} T_d E \left[ \log \left| \Phi + \sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q f_{ki}(P_{t_k}) \hat{\mathbf{z}}_{ki} \hat{\mathbf{z}}_{ki}^\dagger \right| \right] \quad (22)$$

In the optimization problem in (22), we have  $T_{t_k} + 1$  optimization variables,  $\lambda_{k1}^Q, \dots, \lambda_{kT_{t_k}}^Q$ , and  $P_{t_k}$ . Unlike a similar problem in [4], which does not take channel estimation problem into consideration, the problem here is not necessarily convex due to the existence of  $P_{t_k}$ . Equation (22) is concave when  $T_{t_k} = 1$ , which results in an affine  $f_{k1}(P_t)$ . Therefore, in the most general case, the solution of the first order necessary conditions will give a local maximum. The Lagrangian for the optimization problem in (22) can be written as

$$\begin{aligned} & \frac{T - T_t}{T} E \left[ \log \left| \Phi + \sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q f_{ki}(P_{t_k}) \hat{\mathbf{z}}_{ki} \hat{\mathbf{z}}_{ki}^\dagger \right| \right] \\ & - \mu_k \left( \left( \sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q \right) T_d + P_{t_k} T_{t_k} - P_k T \right). \end{aligned} \quad (23)$$

where  $\mu_k$  is the Lagrange multiplier, and we omitted the complementary slackness conditions related to the positiveness

of  $\lambda_{ki}^Q$ , and  $P_{t_k} - \sum_{i=1}^{T_{t_k}} \left( \frac{1}{\lambda_{ki}^\Sigma} - \frac{1}{\lambda_i^\Sigma} \right)$ . The KKT conditions can be written as

$$\frac{T_d}{T} f_{ki}(P_{t_k}) E \left[ \mathbf{z}_{ki}^\dagger \mathbf{B}^{-1} \mathbf{z}_{ki} \right] \leq \mu_k T_d, \quad i = 1, \dots, T_{t_k} \quad (24)$$

$$\frac{T_d}{T} \sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q E \left[ \mathbf{z}_{ki}^\dagger \mathbf{B}^{-1} \mathbf{z}_{ki} \right] \frac{\partial f_{ki}(P_{t_k})}{\partial P_{t_k}} = \mu_k T_{t_k} \quad (25)$$

where  $\mathbf{B} = \Phi + \sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q f_{ki}(P_{t_k}) \hat{\mathbf{z}}_{ki} \hat{\mathbf{z}}_{ki}^\dagger$ , and the equality of the last equation again follows from the complementary slackness condition. In order to derive a set of fixed point equations to update the eigenvalues and the training power of user  $k$ , we follow [2] closely to propose the following fixed point algorithm that solves  $P_{t_k}(n+1)$  from

$$\sum_{i=1}^{T_{t_k}} \lambda_{ki}^Q(n) \frac{f'_{ki}(P_{t_k}(n+1))}{f_{ki}(P_{t_k}(n+1))} = \frac{T_{t_k}}{T_d} \quad (26)$$

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

$$\lambda_{ki}^Q(n+1) = \frac{\lambda_{ki}^Q(n) f_{ki}(P_{t_k}(n+1)) E \left[ \mathbf{z}_{ki}^\dagger \mathbf{B}^{-1} \mathbf{z}_{ki} \right]}{\sum_{j=1}^{T_{t_k}} \lambda_{kj}^Q(n) f_{kj}(P_{t_k}(n+1)) E \left[ \mathbf{z}_{kj}^\dagger \mathbf{B}^{-1} \mathbf{z}_{kj} \right]} P_d. \quad (27)$$

where  $P_d = \frac{(P_k T - P_{t_k}(n+1) T_{t_k})}{T_d}$ . This algorithm finds the solution of the inner optimization problem in (22) in terms of the training power  $P_{t_k}$ , and the eigenvalues of the transmit covariance matrix  $\lambda_{k1}^Q, \dots, \lambda_{kT_{t_k}}^Q$  of user  $k$ , 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 (19) can be found by taking the one that results in the largest rate, which gives us the value of  $T_{t_k}$  that maximizes (19). Now, we know the parameters  $\lambda_k^Q, P_{t_k}, T_{t_k}$ , that maximize (19), when the parameters of the rest of the users are fixed. We then move to another user, and perform the same inner maximization for this user keeping the parameters of the rest of the users fixed. 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 (17).

As a result, we solved the joint channel estimation and resource allocation problem in a MIMO multiple access channel with noisy channel estimation and partial CSI available at the transmitter. For user  $k$ , through  $P_{t_k}$ , we find the allocation of available power over the training and data transmission phases. Through  $T_{t_k}$ , we find the portion of the training duration that is allocated to user  $k$ , and through the sum of these portions  $T_t = \sum_{k=1}^K T_{t_k}$ , we find the allocation of available time over the training and data transmission phases. Through Theorem 2, we find the transmit directions of user  $k$ , and through  $\lambda_{k1}^Q, \dots, \lambda_{kT_{t_k}}^Q$ , we find the allocation of data transmission power of user  $k$  over these transmit directions. Finally, the training signal of user  $k$ ,  $\mathbf{S}_k$ , is determined by  $T_{t_k}$  and  $P_{t_k}$  through Theorem 1.

### C. Numerical Analysis

Analytical proof of the convergence of this algorithm seems to be more complicated than the proof in the case when there

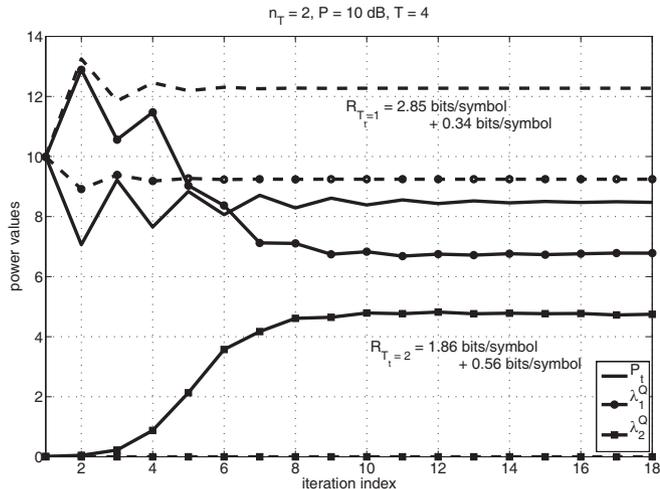


Fig. 1. The convergence of the single-user algorithm with  $n_T = n_R = 2$ , 10 dB total average power and  $T = 4$ . The dashed curves correspond to one symbol long training,  $T_t = 1$ , and solid curves correspond to two symbols long training,  $T_t = 2$ .

is no channel estimation error [4], and seems to be intractable for now. Moreover, other than the case when  $T_t = 1$ , the algorithm is not guaranteed to achieve the global optimum. However, in our extensive simulations, we observed that the algorithm always converges.

We start our numerical analysis with the single-user case of Part I [2] of this two-part paper. We first consider a system having  $n_T = n_R = 2$  with 10 dB total average SNR and block length  $T = 4$ . In Figure 1, we plot the eigenvalues of the data transmit matrix and the training power as a function of the iteration index for both possible values of the training signal duration<sup>3</sup>. We observe that when the training duration is one symbol period, we achieve a higher rate. Therefore, for this set of given system parameters, estimating only one dimension of the channel is optimum.

Next, we investigate the effect of total average power on the number of estimated channel dimensions. We observe that if we keep the block length small at  $T = 4$ , the amount of total power required in order to estimate the second channel dimension is very high. In Figure 2, for a 40 dB total average power, we plot the eigenvalues of the data transmit matrix and the training power as a function of the iteration index for both possible values of the training signal duration, and we see that achievable rate with two symbols of training is barely higher than the achievable rate with one symbol of training. We repeat this experiment with different numbers of antennas and channel eigenvalues, and we see that we need very high power levels in order to use more than one symbol of training. This suggests that the block length, i.e., the coherence interval,

<sup>3</sup>Our algorithm is proposed to maximize a lower bound to the capacity. By generalizing the results of [7] to transmit-side correlated channel fading, it is possible to derive an upper bound to the capacity as well. As in the case of [7], the difference between the upper and the lower bounds can be found in a closed form expression. In Figure 1,  $+ \dots$  bits/symbol, i.e., the expression under the achievable rate, denotes this difference between the upper and the lower bounds. We will use this notation in some of the other figures as well, specifically in Figures 1-5. We observe from these figures that our achievable rate scheme is very close to the upper bound, especially when the block length is larger. We also note that the difference does not depend on the SNR when the block length is short.

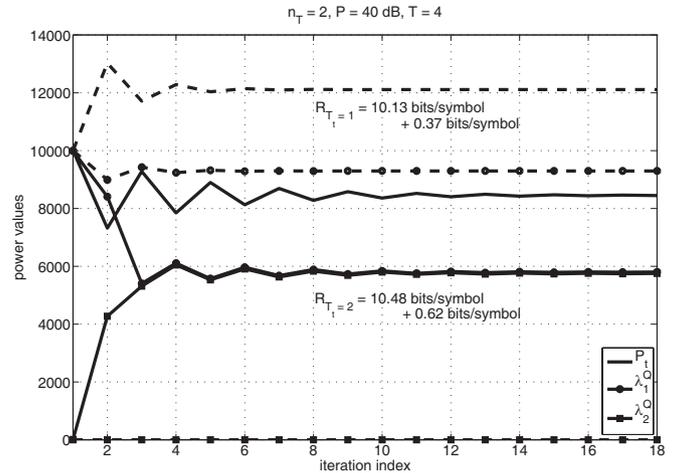


Fig. 2. The convergence of the single-user algorithm with  $n_T = n_R = 2$ , 40 dB total average power and  $T = 4$ . The dashed curves correspond to one symbol long training,  $T_t = 1$ , and solid curves correspond to two symbols long training,  $T_t = 2$ .

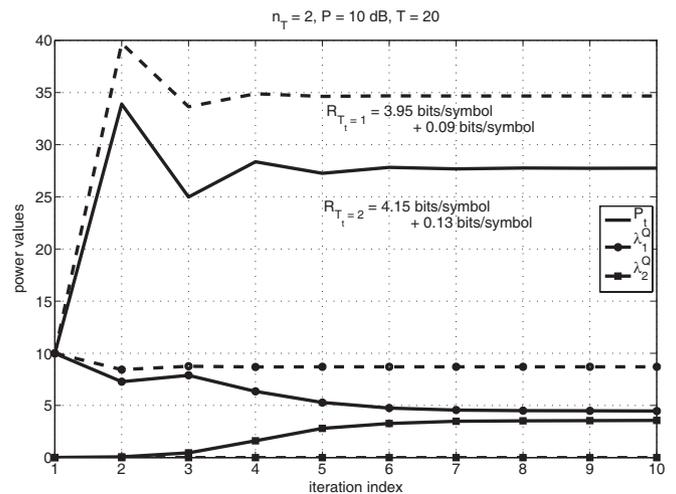


Fig. 3. The convergence of the single-user algorithm with  $n_T = n_R = 2$ , 10 dB total average power and  $T = 20$ . The dashed curves correspond to one symbol long training,  $T_t = 1$ , and solid curves correspond to two symbols long training,  $T_t = 2$ .

is more important for determining the duration of the training phase.

In order to investigate the effect of the block length, in Figure 3, we consider 10 dB total average power, and block length  $T = 20$ . We observe that similar to the high SNR case, in this case as well, having two symbols long training phase is optimum. We repeat this experiment with different numbers of antennas, and channel eigenvalues for long block lengths, and we see that moderate block lengths are sufficient in order to use more than one symbol of training. Therefore, we conclude that for very fast changing channels where the coherence interval and therefore the block length is short, and for low SNR systems, estimating only one dimension of the channel results in higher achievable rates. In this case, we cannot take advantage of the multiple dimensions that the MIMO channel provides, because the amount of time required to estimate those channels cancels the data rate advantage brought by having multiple channels.

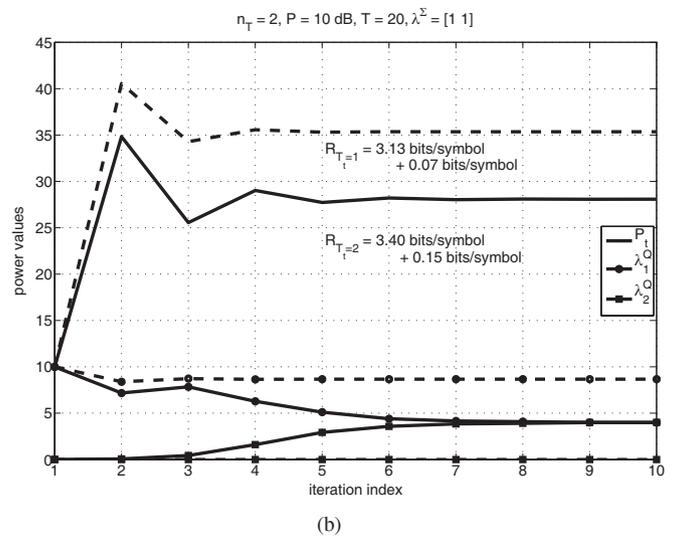
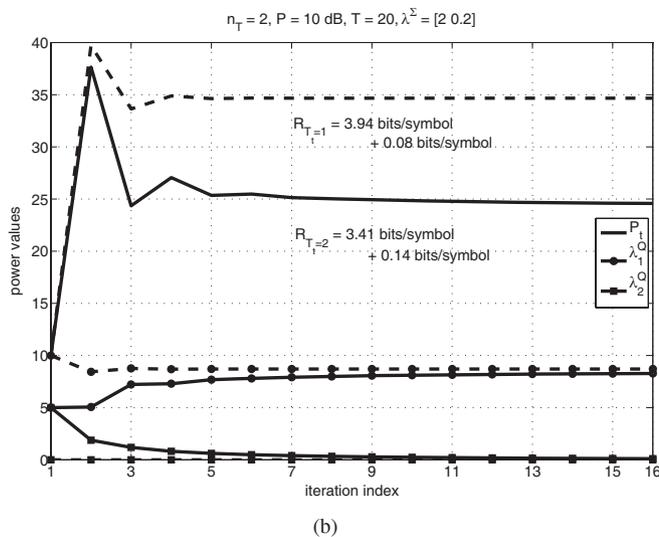
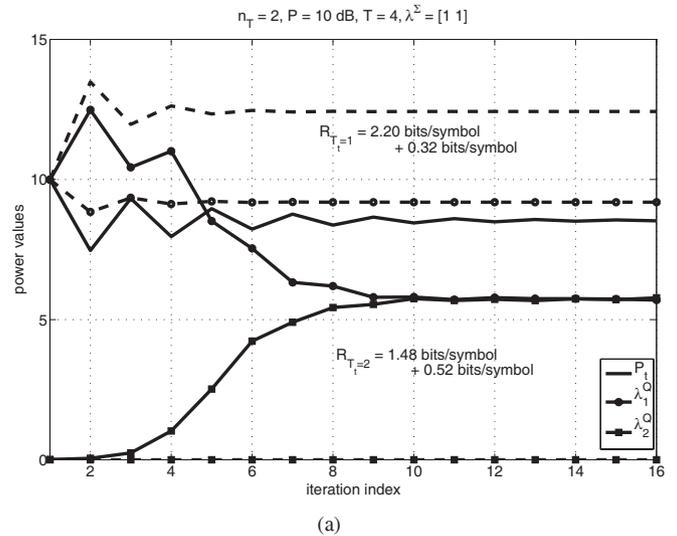
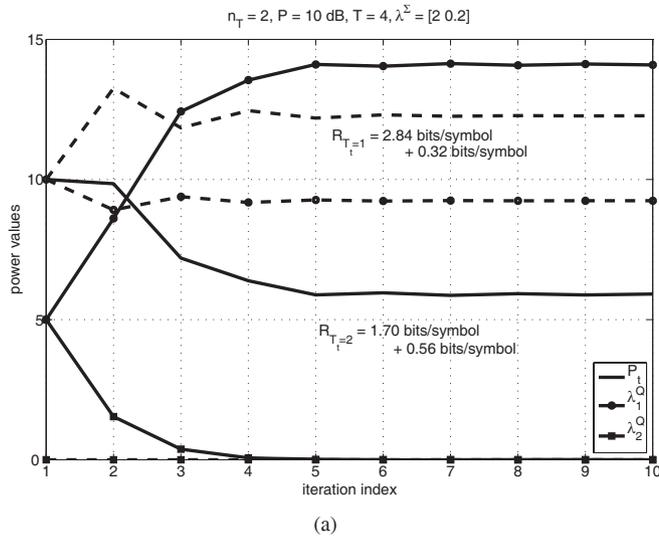


Fig. 4. The convergence of the single-user algorithm with  $n_T = n_R = 2$ , 10 dB total average power, and channel eigenvalues  $\lambda^\Sigma = [2, 0.2]$ , where dashed curves correspond to one symbol long training,  $T_t = 1$ , and solid curves correspond to two symbols long training,  $T_t = 2$ : (a)  $T = 4$ ; (b)  $T = 20$ .

Fig. 5. The convergence of the single-user algorithm with  $n_T = n_R = 2$ , 10 dB total average power, and channel eigenvalues  $\lambda^\Sigma = [1, 1]$ , where dashed curves correspond to one symbol long training,  $T_t = 1$ , and solid curves correspond to two symbols long training,  $T_t = 2$ : (a)  $T = 4$ ; (b)  $T = 20$ .

The trade-off between estimating the channel better and increasing the data rate can also be seen from Figures 1-3. We see in Figure 1 that when we increase the training duration to two symbols, time remaining for data transmission becomes very short, and therefore the data rate decreases. On the other hand, in Figure 3, when the training duration is increased to two symbols, there is still enough time for increasing the rate of the data transmission. Therefore, increasing the training duration is not always helpful. The result crucially depends on the total block length, available SNR and the number of antennas in the system.

We next analyze the effects of different channel covariance matrices. In Figure 4, we consider 10 dB average power, and a channel covariance matrix that has a first eigen-direction much stronger than the second eigen-direction, i.e., the largest eigenvalue of the channel covariance matrix is much larger than the second largest eigenvalue. In such scenarios, even if the block length is large, beamforming turns out to be the optimal strategy for the data transmission period [5], [9].

Therefore, estimating the second dimension is a waste of resources, because no power will be allocated to that direction in the data transmission phase. Confirming this intuition, in Figure 4, for the cases when  $T_t = 2$ , the power allocated to the second eigen-direction is zero, although the training power is large enough to estimate both channels.

Another extreme for the eigenvalues of the channel covariance matrix is the case when both eigenvalues are equal. Note that this case is exactly the case considered [8]. However, in this paper, we do not assume the restriction that  $T_t \geq n_T$  as it was assumed in [8] by reasoning that one needs at least  $T_t \geq n_T$  measurements in order to estimate  $n_T$  variables. Although this reasoning is valid, we relax this restriction by pointing out that in some cases, we might not want to estimate  $n_T$  variables. If the resources are limited, estimating some of the variables and saving the resources for data transmission is more useful. As a result, in this paper, we find that the duration of the training signal is equal to the number of variables to be estimated rather than the total number of variables. Figure 5

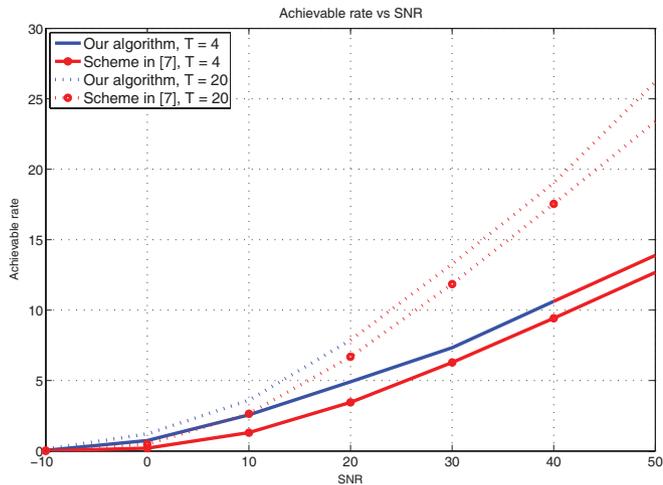


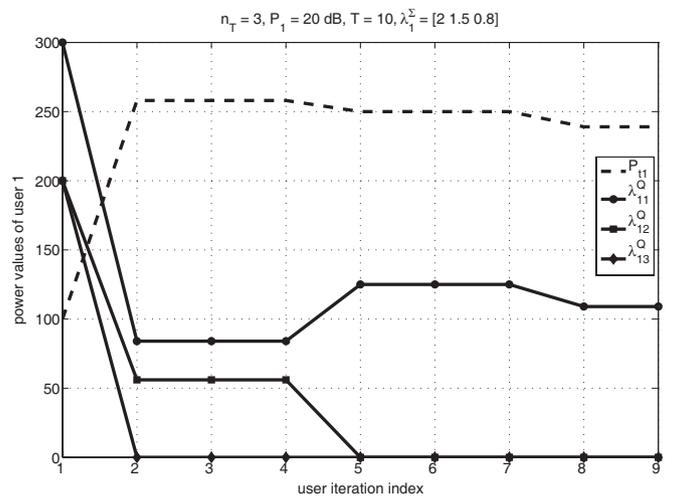
Fig. 6. The performance improvement of our algorithm compared to [8]. The color change from blue to red indicates that the number of spatial dimensions to be estimated is increased from one to two.

supports our findings, by showing that, for a short block length  $T = 4$  with 10 dB total power, not estimating one of the dimensions results in a higher data transmission rate. This advantage disappears when the block length is long enough. In order to see the performance improvement of our algorithm compared to [8], we plot the rates achieved by our algorithm and by [8] with respect to the SNR. In Figure 6, we see that especially at longer block lengths, there is a non-vanishing rate difference between our algorithm and allocating power equally over the spatial dimensions.

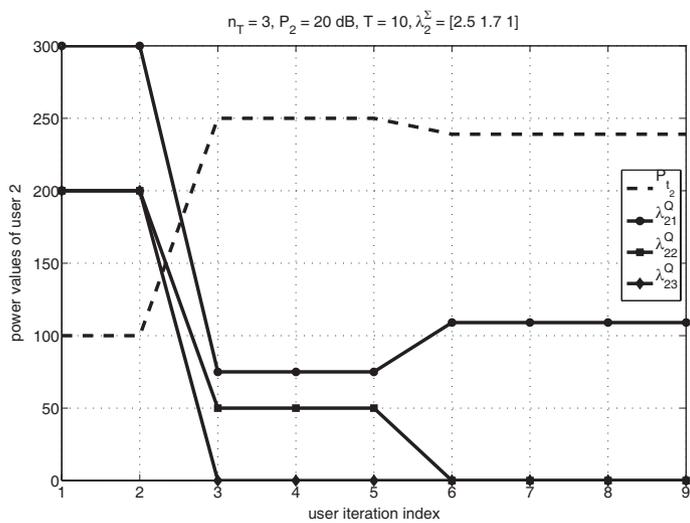
In a MIMO-MAC case, proving the convergence of our algorithm becomes even harder. However, again, we observed through extensive simulations that the proposed algorithm always converges. In Figure 7, we considered a system of  $K = 3$  users with  $n_T = n_R = 3$ , all having moderate power,  $P = 20$  dB, and moderate block length,  $T = 10$ . Each iteration in Figure 7, corresponds to solving (19) for one of the users, while the parameters of the rest of the users are fixed. Although we observe in Part I [2, Figures 2-4] that in the same system with a single user, it is optimal to estimate two dimensions of the channel, in this multi-user case, we observe in Figure 7 that, all users estimate only one dimension of the channel.

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. For example, we see in Figure 8 that, for a 3-user system, one of the users start estimating the second dimension, when  $T$  gets large enough, i.e., when  $T = 50$ . When the number of users increases, total number of channels estimated by all users also increases, since each user has to spend its power.

In order to verify that our algorithm also converges with larger numbers of users and with uneven number of antennas on users, we analyze a case with  $K = 10$ , and  $T = 50$ . In Figure 9, half of the users have two transmit antennas, and the other half have three transmit antennas. In this case as well, we observed the convergence of our algorithm, and the highest



(a)



(b)

Fig. 7. The convergence of the multi-user algorithm with  $n_T = n_R = 3$ , 20 dB total average power and  $T = 10$ : (a) convergence of user 1; (b) convergence of user 2; convergence of user 3 is similar and omitted here due to space limitations.

achievable sum rate is obtained when each user estimated only one dimension of its channel. In Figure 9, we plot the power values of the first user. The algorithm goes over the users five times in a round-robin fashion, and for each user it performs ten iterations.

#### IV. CONCLUSIONS

We considered a block-fading MIMO multiple access channel, where the receiver has a noisy estimate of the channel and the transmitters have the partial CSI in the form of covariance feedback. Each transmission block is divided into a training phase and a data transmission phase. During the training phase, we showed that the users send time-orthogonal training signals. During the data transmission phase, we formulated an optimization problem that maximizes 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. The proposed multi-user algorithm solves the problem iteratively over the users in a round-robin fashion,

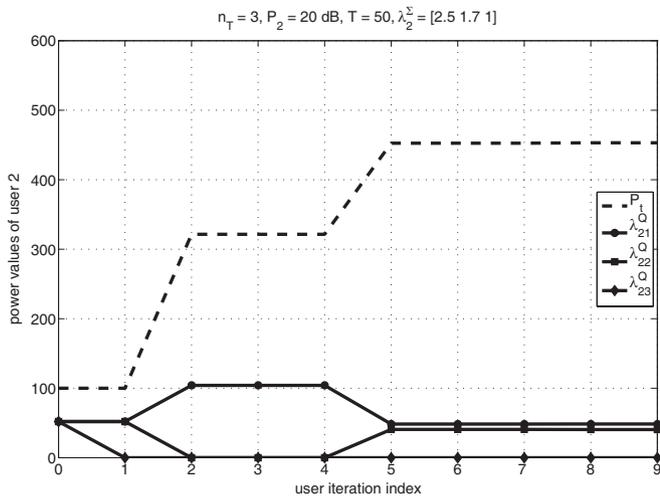


Fig. 8. The convergence of the multi-user algorithm with  $n_T = n_R = 3$ , 20 dB total average power and  $T = 50$ . Only the convergence of user 2 is given to show that two dimensions are estimated.

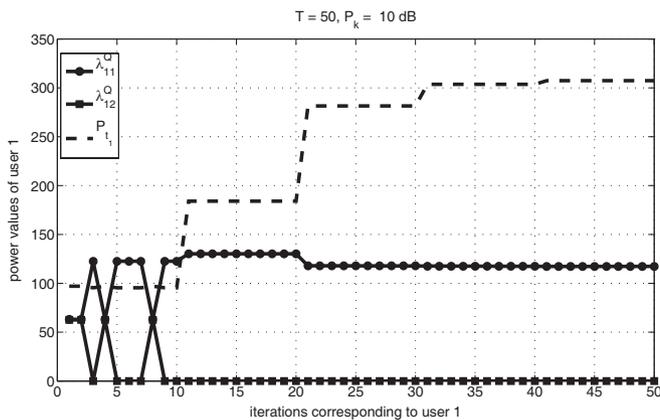
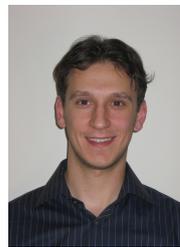


Fig. 9. The convergence of the multi-user algorithm with  $K = 10$ , 10 dB total average power and  $T = 50$ . The first five users have  $n_T = 2$ , other users have  $n_T = 3$ . Only the convergence of user 1 is given.

while utilizing a single-user algorithm similar to the one proposed in Part I, for an update of each user. Although theoretical convergence proofs of these algorithms remain as open problems, through extensive simulations, we observed that both the single-user algorithm of Part I and the multi-user algorithm of Part II converge, and they converge to the same point regardless of the initial point of the iterations.

## REFERENCES

- [1] A. Soysal and S. Ulukus, "MIMO multiple access channels with noisy channel estimation and partial CSI feedback," in *Proc. IEEE Globecom*, Nov. 2008.
- [2] A. Soysal and S. Ulukus, "Joint channel estimation and resource allocation for MIMO systems—part I: single-user analysis," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 624–631, Feb. 2009.
- [3] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 145–151, Jan. 2004.
- [4] A. Soysal and S. Ulukus, "Optimum power allocation for single-user MIMO and multi-user MIMO-MAC with partial CSI," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1402–1412, Sept. 2007.
- [5] A. Soysal and S. Ulukus, "Optimality of beamforming in fading MIMO multiple access channels," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 1171–1183, Apr. 2009.
- [6] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [7] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.
- [8] 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.
- [9] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, July 2004.
- [10] H. Boche and E. Jorswieck, "On the optimality range of beamforming for MIMO systems with covariance feedback," *IEICE Trans. Commun.*, vol. E85-A, no. 11, pp. 2521–2528, Nov. 2002.
- [11] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
- [12] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.
- [13] W. Yu and W. Rhee, "Degrees of freedom in wireless multiuser spatial multiplex systems with multiple antennas," *IEEE Trans. Commun.*, vol. 54, no. 10, pp. 1747–1753, Oct. 2006.
- [14] A. Soysal, "Optimum transmit strategies for Gaussian multi-user MIMO networks with partial CSI," Ph.D. thesis, University of Maryland, 2008.



**Alkan Soysal** received the B.S. degree in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey in 2002, and the M.S. and Ph.D. degrees in electrical and computer engineering from University of Maryland, College Park, MD in 2006 and 2008 respectively.

He joined Bahçeşehir University, Istanbul, Turkey in February 2008 as an Assistant Professor in the Department of Electrical and Electronics Engineering. His research interests are in wireless communication theory, information theory and signal

processing for wireless communications with particular focus on MIMO networks.



**Sennur Ulukus** received the B.S. and M.S. degrees in electrical and electronics engineering from Bilkent University, Ankara, Turkey, in 1991 and 1993, respectively, and the Ph.D. degree in electrical and computer engineering from Rutgers University, NJ, in 1998.

During her Ph.D. studies, she was with the Wireless Information Network Laboratory (WINLAB), Rutgers University. From 1998 to 2001, she was a Senior Technical Staff Member at AT&T Labs-Research in NJ. In 2001, she joined the University of Maryland at College Park, where she is currently an Associate Professor in the Department of Electrical and Computer Engineering, with a joint appointment at the Institute for Systems Research (ISR). Her research interests are in wireless communication theory and networking, network information theory for wireless networks, signal processing for wireless communications and security for multi-user wireless communications.

Sennur Ulukus is a recipient of the 2005 NSF CAREER Award, and a co-recipient of the 2003 IEEE Marconi Prize Paper Award in Wireless Communications. She serves/served as an Associate Editor for the IEEE Transactions on Information Theory since 2007, as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS between 2003–2007, as a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS in 2006–2008, as the co-chair of the Communication Theory Symposium at the 2007 IEEE Global Telecommunications Conference, as the co-chair of the Medium Access Control (MAC) Track at the 2008 IEEE Wireless Communications and Networking Conference, as the co-chair of the Wireless Communications Symposium at the 2010 IEEE International Conference on Communications, and as the Secretary of the IEEE Communication Theory Technical Committee (CTTC) in 2007–2009.