Joint Channel Estimation and Resource Allocation for MIMO Systems–Part I: Single-User Analysis

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

***Abstract*—Multiple antenna systems are known to provide very large data rates, when the perfect channel state information (CSI) is available at the receiver. However, this requires the receiver to perform a noise-free, multi-dimensional channel estimation, without using communication resources. In practice, any channel estimation is noisy and uses system resources. We shall examine the trade-off between improving channel estimation and increasing the achievable data rate. We consider transmit-side correlated multi-input multi-output (MIMO) channels with block fading, where each block is divided into training and data transmission phases. The receiver has a noisy CSI that it obtains through a channel estimation process, while the transmitter has partial CSI in the form of covariance feedback. In Part I of this two-part paper, we consider the single-user case, and optimize the achievable rate jointly over parameters associated with the training phase and data transmission phase. In particular, we first choose the training signal to minimize the channel estimation error, and then, develop an iterative algorithm to solve for the optimum system resources such as time, power and spatial dimensions. Specifically, the algorithm finds the optimum training duration, the optimum allocation of power between training and data transmission phases, the optimum allocation of power over the antennas during the data transmission phase.**

***Index Terms*—MIMO, partial CSI, covariance feedback, optimum power allocation, channel estimation.**

I. INTRODUCTION

IN wireless communication scenarios, the achievable rate of a 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 feeding it back to the transmitter uses communication resources such as time, power and spatial dimensions, which could otherwise be used for useful information transmission. There have been several different assumptions in the literature on the availability of the CSI at the receiver and the transmitter. With perfect CSI at the receiver and the transmitter, the optimum adaptation scheme is water-filling [2], [3]. However, in some cases, especially in 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 [4]–[8]. 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 [9]–[11]. It is important to note that [9]–[11] do not consider optimizing the channel estimation process, because of the assumption of the existence of a separate channel that does not consume system resources for channel estimation. For a single-user multiple-antenna system with no CSI available at the transmitter, [12] considers optimizing the achievable rate as a function of both the training and the data transmission phases. Since there is no CSI feedback, the transmitter power allocation is constant over the channel states and the antennas.

Part I of this two-part paper considers a single-user, block-fading, transmit-side correlated MIMO channel with noisy channel estimation at the receiver, and partial CSI available at the transmitter. The CSI feedback that we consider lies somewhere between perfect CSI [11] and no CSI [12], and it is similar to [4]–[8]. We consider the fact that the training phase uses communication resources, and we optimize the achievable rate of the data transmission phase over the parameters of the training and data transmission processes.

The training phase is characterized by three parameters, namely, the training signal, the training sequence length and the training sequence power. Similarly, the data transmission phase is characterized by the data carrying input signal, data transmission length, and the data transmission power. Assuming that the receiver uses linear minimum mean square error (MMSE) detection to estimate the channel during the training phase, we first choose the training signal that minimizes the MMSE. This choice also increases the achievable rate of the data transmission phase [12]. However, unlike [12], our result does not necessarily allocate equal power over the antennas, and might not estimate all of the available channel variables. Then, we move to the data transmission phase, and maximize the achievable rate of the data transmission jointly over the rest of the training phase parameters, and data transmission phase parameters. Specifically, we first find the optimum partition of the total transmitter power and the block length between the training and the data transmission phases. Then, we find the optimum allocation of the data transmission power over the antennas during the data transmission phase.

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 Conference on Information Sciences and Systems, Baltimore, MD, March 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.2010.02.080099

TABLE I
SYMBOLS USED IN THIS PAPER IN ORDER OF APPEARANCE

n_R	number of receive antennas
n_T	number of transmit antennas
$\mathbf{H}$	$n_R \times n_T$ transmit-side correlated channel matrix
T	transmission block duration
T_t	training phase duration
T_d	data transmission phase duration
$\mathbf{x}_n$	$n_T \times 1$ channel input at time n
$\mathbf{n}_n$	$n_R \times 1$ channel noise
$\sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$	distributed as a zero-mean, identity covariance complex Gaussian vector
P	total average power constraint
$\mathbf{Z}$	$n_R \times n_T$ i.i.d. channel matrix
$\mathbf{\Sigma}$	$n_T \times n_T$ covariance matrix
P_t	power allocated to training phase
P_d	power allocated to data transmission phase
$\mathbf{S}$	$n_T \times T_t$ training signal
$\mathbf{N}_t$	$n_R \times T_t$ noise matrix in training phase
$\mathbf{R}_t$	$n_R \times T_t$ received matrix in training phase
$\hat{\mathbf{H}}$	$n_R \times n_T$ estimate of the channel matrix
$\hat{\mathbf{\Sigma}}$	$n_T \times n_T$ covariance of the channel estimate
$\hat{\lambda}_i^{\hat{\Sigma}}$	eigenvalues of $\hat{\mathbf{\Sigma}}$
$\tilde{\mathbf{H}}$	$n_R \times n_T$ channel estimation error
$\tilde{\mathbf{\Sigma}}$	$n_T \times n_T$ covariance of the channel estimation error
$\tilde{\lambda}_i^{\tilde{\Sigma}}$	eigenvalues of $\tilde{\mathbf{\Sigma}}$
$\mathbf{M}$	$n_T \times T_t$ MMSE estimator matrix
$\mathbf{U}_\Sigma$	$n_T \times n_T$ unitary matrix of the eigenvectors of $\mathbf{\Sigma}$
$\mathbf{\Lambda}_S$	diagonal eigenvalue matrix of $\mathbf{S}$
λ_{iS}	eigenvalue of $\mathbf{S}$
$\mathbf{\Lambda}_\Sigma$	diagonal eigenvalue matrix of $\mathbf{\Sigma}$
λ_i^Σ	eigenvalue of $\mathbf{\Sigma}$
μ_S	Lagrange multiplier associated with channel estimation minimization
$\mathbf{Q}$	$n_T \times n_T$ covariance matrix of the channel input
$\hat{\mathbf{Z}}$	$n_R \times n_T$ i.i.d. channel matrix
$\boldsymbol{\lambda}^Q$	$n_T \times 1$ vector of eigenvalues of $\mathbf{Q}$

Overall, we find the optimum allocation of resources such as power, time and spatial dimensions. Part II [13] extends the results of Part I to multiple-access channels. It also provides extensive numerical analysis for both single-user and multi-user scenarios. Symbols used in the paper are summarized in Table 1.

II. System Model

We consider a point-to-point channel with n_R antennas at the receiver and n_T antennas at the transmitter. The channel between the transmitter and the receiver is represented by an $n_R \times n_T$ dimensional random matrix $\mathbf{H}$. 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 channel, the receiver performs a linear MMSE estimation 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 transmitter has only the statistical model of the channel. At time n, the transmitter sends a vector $\mathbf{x}_n$, and the received vector is

$$\mathbf{r}_n = \mathbf{H}\mathbf{x}_n + \mathbf{n}_n, \quad n = 1, \ldots, T \tag{1}$$

where $\mathbf{n}_n \sim \mathcal{CN}(0, \mathbf{I})$, and the entries of $\mathbf{H}$ are complex Gaussian random variables. The transmitter has a power constraint $P = \frac{1}{T}E[\sum_n \mathbf{x}_n^H \mathbf{x}_n]$, averaged over T symbols[1].

The statistical model that we consider in this paper 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 model, there exists correlation between the signals transmitted by or received at different antenna elements. For each user, the channel is modeled as [14],

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

where the entries of $\mathbf{Z}$ are i.i.d., zero-mean, unit-variance complex Gaussian random variables, the receive antenna correlation matrix, $\mathbf{\Phi}$, is the correlation between the signals received at the n_R receive antennas of the receiver, and the transmit antenna correlation matrix, $\mathbf{\Sigma}$, is the correlation between the signals transmitted from the n_T transmit antennas of the user. 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. As a result, the receive antenna correlation matrix becomes the identity matrix[2], i.e., $\mathbf{\Phi} = \mathbf{I}$. Now, the channel is written as

$$\mathbf{H} = \mathbf{Z}\mathbf{\Sigma}^{1/2} \tag{3}$$

From this point on, we will refer to matrix $\mathbf{\Sigma}$ as the channel covariance feedback matrix. Similar covariance feedback models have been used in [4]–[8].

III. Joint Optimization for Single-user MIMO

In our model the channel is fixed over a coherence interval, which is divided into two phases: training phase and data transmission phase; see Figure 1. The transmitter uses P_t amount of power during the training phase and P_d amount of power during the data transmission phase. Due to the conservation of energy, we have $PT = P_t T_t + P_d T_d$.

The optimization criterion that we consider is the achievable rate of the data transmission phase. Unlike the case with perfect channel estimation, the data rate here depends on the estimation parameters: training signal $\mathbf{S}$, training signal power P_t, and training signal duration T_t. As a result, there is a trade-off between the training and data transmission parameters. A longer training phase will result in a better channel estimate (a lower channel estimation error). This in turn, results in a higher achievable rate, since the effective noise is lower. However, a longer training phase implies a shorter data transmission phase, as the block length (coherence time) is fixed. A shorter data transmission phase, in turn,

[1]Note that since the noise power is assumed to be unity, P is in fact the relative power with respect to noise power. It can be regarded as an SNR value.

[2]Extension of our results to arbitrary $\mathbf{\Phi}$, i.e., the case where the channel has double-sided correlation structure, can be carried out in a straight-forward manner as in [8].

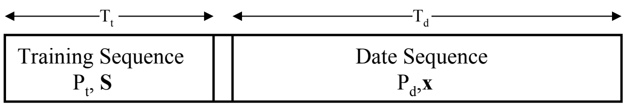


Fig. 1. Illustration of a single coherence time, over which the channel is fixed.

implies a smaller achievable rate. A similar trade-off is valid also for the training power. Here, we will solve this trade-off, and find the optimum training and data transmission parameters.

A. Training and Channel Estimation Phase

In practical communication scenarios, the channel is estimated at the receiver. One way of doing this is to use training symbols before the data transmission starts. The receiver estimates the channel using these training signals and the output of the channel. Since the channel stays the same during the entire block, we can write the input-output relationship during the training phase in a matrix form as

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

where $\mathbf{S}$ is an $n_T \times T_t$ dimensional training signal 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 (4) represents the input-output relationship at time n. The power constraint for the training input signal is $\frac{1}{T_t}\text{tr}(\mathbf{S}\mathbf{S}^\dagger) \leq P_t$.

Due to our channel model in (3), the entries in a row of $\mathbf{H}$ are correlated, and the entries in a column of $\mathbf{H}$ are uncorrelated, i.e., row i of the channel matrix is i.i.d. with row j. Let us represent row i of $\mathbf{H}$ as $\mathbf{h}_i^\dagger$, with $E[\mathbf{h}_i\mathbf{h}_i^\dagger] = \mathbf{\Sigma}, i = 1, \dots n_R$. Since rows are i.i.d., the receiver can estimate each of them independently using the same training signal. Row i of (4), which represents the received signal at the i^{th} antenna of the receiver over the training duration, can be written as

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

The receiver will estimate $\mathbf{h}_i$ using the received signal $\mathbf{r}_{ti}$, and the training signal $\mathbf{S}$. In general, the estimate $\hat{\mathbf{h}}_i$ can be set to any function of $\mathbf{S}$ and $\mathbf{r}_{ti}$. However, it is common to use and easier to implement linear MMSE estimation. Also, when the random variables involved in the estimation are Gaussian, as in Rayleigh fading channels, linear MMSE estimation is optimal. In order to find the linear MMSE estimator, we solve the following optimization problem with $\hat{\mathbf{h}}_i = \mathbf{M}\mathbf{r}_{ti}$ as the estimate of $\mathbf{h}_i$, and $\tilde{\mathbf{h}}_i = \mathbf{h}_i - \hat{\mathbf{h}}_i$ as the channel estimation error,

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

$$= \min_{\mathbf{M}} E\left[\text{tr}\left((\mathbf{h}_i - \mathbf{M}\mathbf{r}_{ti})(\mathbf{h}_i - \mathbf{M}\mathbf{r}_{ti})^\dagger\right)\right]. \tag{7}$$

Solving the optimum transformation matrix, $\mathbf{M}^*$ from (7) is equivalent to solving $\mathbf{M}^*$ from the orthogonality principle for vector random variables, which is given as [15, page 91],

$$E\left[(\mathbf{h}_i - \mathbf{M}^*\mathbf{r}_{ti})\mathbf{r}_{ti}^\dagger\right] = \mathbf{0} \tag{8}$$

where $\mathbf{0}$ is the $n_T \times T_t$ zero matrix. We can solve $\mathbf{M}^*$ from (8) as

$$\mathbf{M}^* = E\left[\mathbf{h}_i \mathbf{r}_{ti}^\dagger\right]\left(E\left[\mathbf{r}_{ti}\mathbf{r}_{ti}^\dagger\right]\right)^{-1}. \tag{9}$$

By using (5), we calculate $E[\mathbf{h}_i\mathbf{r}_{ti}^\dagger] = \mathbf{\Sigma}\mathbf{S}$, and $E[\mathbf{r}_{ti}\mathbf{r}_{ti}^\dagger] = \mathbf{S}^\dagger\mathbf{\Sigma}\mathbf{S} + \mathbf{I}$. Then, $\mathbf{M}^*$ becomes $\mathbf{M}^* = \mathbf{\Sigma}\mathbf{S}(\mathbf{S}^\dagger\mathbf{\Sigma}\mathbf{S} + \mathbf{I})^{-1}$. By inserting $\mathbf{M}^*$ into (7), the mean square error becomes,

$$\min_{\mathbf{M}} E\left[\tilde{\mathbf{h}}_i^\dagger \tilde{\mathbf{h}}_i\right] = \text{tr}\left(\mathbf{\Sigma} - \mathbf{\Sigma}\mathbf{S}(\mathbf{S}^\dagger\mathbf{\Sigma}\mathbf{S} + \mathbf{I})^{-1}\mathbf{S}\mathbf{\Sigma}\right) \tag{10}$$

$$= \text{tr}\left(\left(\mathbf{\Sigma}^{-1} + \mathbf{S}\mathbf{S}^\dagger\right)^{-1}\right) \tag{11}$$

where the last line follows from 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 $\mathbf{S}$ to minimize (11). In addition, it is stated in [12] that the training signal $\mathbf{S}$ primarily affects the achievable rate through the so called *effective signal-to-noise ratio*, which is shown to be inversely proportional to the MMSE [12]. Therefore, choosing $\mathbf{S}$ to further minimize the MMSE, we also increase the achievable rate of the data transmission phase. The following theorem finds the optimal training signal for a given training power and training duration.

Theorem 1: For given $\mathbf{\Sigma} = \mathbf{U}_\Sigma \mathbf{\Lambda}_\Sigma \mathbf{U}_\Sigma^\dagger$, P_t, T_t, and the power constraint $\text{tr}(\mathbf{S}\mathbf{S}^\dagger) \leq P_t T_t$, the optimum training input that minimizes the power of the channel estimation error vector is $\mathbf{S}^* = \mathbf{U}_\Sigma \mathbf{\Lambda}_S^{1/2}$ with

$$\lambda_i^S = \left(\frac{1}{\mu_S} - \frac{1}{\lambda_i^\Sigma}\right)^+, \qquad i = 1, \dots, \min(n_T, T_t) \tag{12}$$

where μ_S^2 is the Lagrange multiplier that satisfies the power constraint with $\mu_S = \frac{J}{P_t + \sum_{i=1}^J \frac{1}{\lambda_i^\Sigma}}$, and J is the largest index that has non-zero λ_i^S.

Proof: Let us have $\mathbf{S} = \mathbf{U}_S \mathbf{\Lambda}_S^{1/2}\mathbf{V}_S^\dagger$. The expression in (11) is minimized when $\mathbf{\Sigma}^{-1}$ and $\mathbf{S}\mathbf{S}^\dagger$ have the same eigenvectors [17]. Therefore, we have $\mathbf{U}_S = \mathbf{U}_\Sigma$. Since, $\mathbf{S}\mathbf{S}^\dagger = \mathbf{U}_S\mathbf{\Lambda}_S\mathbf{U}_S^\dagger$, and the unitary matrix $\mathbf{V}_S$ does not appear in the objective function and the constraint, we can choose $\mathbf{V}_S = \mathbf{I}$. Inserting this into (11), the optimization can be written as

$$\tilde{\sigma} = \min_{\text{tr}(\mathbf{\Lambda}_S) \leq P_t T_t} \text{tr}\left(\left(\mathbf{\Lambda}_\Sigma^{-1} + \mathbf{\Lambda}_S\right)^{-1}\right). \tag{13}$$

The Langrangian of the problem in (13) can be written as

$$\sum_{i=1}^{n_T} \frac{1}{\frac{1}{\lambda_i^\Sigma} + \lambda_i^S} + \mu_S^2\left(\sum_{i=1}^{n_T}\lambda_i^S - P_t T_t\right) \tag{14}$$

where μ_S^2 is the Lagrange multiplier. The Lagrangian is a convex function of λ_i^S, therefore the solution that satisfies the Karush-Kuhn-Tucker (KKT) conditions is the unique optimum solution. This gives us (12), which is water-filling over the eigenvalues of the channel covariance matrix. In order to calculate μ_S, we sum both sides of (12) over all antennas to get $\mu_S = \frac{J}{P_t + \sum_{i=1}^J \frac{1}{\lambda_i^\Sigma}}$, where J is the largest index that has non-zero λ_i^S. □

It is important to note that for any given P_t, and $T_t > n_T$, the eigenvalues of $\mathbf{S}^*$ do not contain the training length parameter. Increasing T_t beyond n_T does not result in better channel estimates. On the other hand, larger T_t will result in smaller data transmission length, and decrease the achievable rate of the data transmission phase. Therefore, it is sufficient to consider only $T_t \leq n_T$, which we will assume through the rest of this paper.

Theorem 1 tells us that the optimum transmit directions of the training signal are the eigenvectors of the channel covariance matrix, and the right eigenvector matrix of the training signal is identity. As a result, the columns of $\mathbf{S}^*$ are the weighted columns of a unitary matrix, and they are orthogonal. Since each column of $\mathbf{S}^*$ is transmitted at a channel use during the training phase, vectors that are transmitted at each channel use during the training phase are orthogonal to each other. This means that, at each channel use, it is optimal to train only one dimension of the channel along one eigenvector. Moreover, the optimum power allocation policy for the training power is to water-fill over the eigenvalues of the channel covariance matrix using (12). Depending on the power constraint and the training signal duration, some of the eigenvalues of the training signal might turn out to be zero. This means that some of the channels along the directions corresponding to zero eigenvalues of the training signal, are not even trained.

Note that μ_S is a function of only P_t and T_t, which are given to the problem in Theorem 1, and will be picked as a result of the achievable rate maximization problem in the data transmission phase. The value of T_t determines the total number of available parallel channels in the channel estimation problem, and the value of P_t determines the number of channels that will be estimated. The parametric values of P_t and T_t will appear in the achievable rate formula in the data transmission phase. After the rate maximization is performed, the optimum P_t and T_t will be found, and these in turn, will give us $\mathbf{S}^*$ through Theorem 1.

Before moving on to the next section, we will calculate the eigenvalues of the covariance matrices of the estimated channel vector, and the channel estimation error vector. Plugging $\mathbf{S}^*$ into the covariance of the channel estimation error, $\tilde{\mathbf{\Sigma}} = E\left[\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^\dagger\right] = \left(\mathbf{\Sigma}^{-1} + \mathbf{S}\mathbf{S}^\dagger\right)^{-1}$, we find $\tilde{\mathbf{\Sigma}} = \mathbf{U}_\Sigma\left(\mathbf{\Lambda}_\Sigma^{-1} + \mathbf{\Lambda}_S\right)^{-1}\mathbf{U}_\Sigma^\dagger$, where the eigenvalues can be found using (12) as

$$\tilde{\lambda}_i^\Sigma = \begin{cases} \mu_S, & \mu_S < \lambda_i^\Sigma; \\ \lambda_i^\Sigma, & \mu_S > \lambda_i^\Sigma \end{cases} = \min\left(\lambda_i^\Sigma, \mu_S\right). \tag{15}$$

Note that along the directions that we send training signals, i.e., when the corresponding eigenvalues of the training signal are non-zero ($\mu_S < \lambda_i^\Sigma$), the variance of the channel estimation error is the same for all directions. Conversely, for the directions where no training signals are sent, the variance of the channel estimation error is equal to the variance of the channel along that direction. This is expected, since the channel is not estimated along that direction, the error in the channel estimation process is the same as the realization of the channel itself.

Next, we will calculate the eigenvalues of the covariance of the channel estimate. Using the orthogonality property of the MMSE estimation, $\hat{\mathbf{h}}_i$ and $\tilde{\mathbf{h}}_i$ are uncorrelated [15, page 91]. The covariance matrix of the channel estimate $\hat{\mathbf{\Sigma}} = E\left[\hat{\mathbf{h}}_i\hat{\mathbf{h}}_i^\dagger\right]$ becomes

$$\hat{\mathbf{\Sigma}} = \mathbf{U}_\Sigma\mathbf{\Lambda}_\Sigma\mathbf{U}_\Sigma^\dagger - \mathbf{U}_\Sigma\tilde{\mathbf{\Lambda}}_\Sigma\mathbf{U}_\Sigma^\dagger \tag{16}$$

$$= \mathbf{U}_\Sigma\left(\mathbf{\Lambda}_\Sigma - \tilde{\mathbf{\Lambda}}_\Sigma\right)\mathbf{U}_\Sigma^\dagger \triangleq \mathbf{U}_\Sigma\hat{\mathbf{\Lambda}}_\Sigma\mathbf{U}_\Sigma^\dagger \tag{17}$$

which has the same eigenvectors as the covariance matrix of the actual channel, however, their eigenvalues are different. We can write each eigenvalue of the covariance matrix of the estimated channel as $\hat{\lambda}_i^\Sigma = \max\left(0, \lambda_i^\Sigma - \mu_S\right)$. Along the directions that we do not send training signals, the value of the channel estimate itself is zero. Therefore, as expected, the power of the estimated channel is zero as well, along those channels with $\mu_S > \lambda_i^\Sigma$.

B. Data Transmission Phase

When the CSI at the receiver is noisy, the optimum input signaling that achieves the capacity is not known. Following [9]–[12], we derive a lower bound (i.e., an achievable rate) on the capacity for our model, and find the training and data transmission parameters that result in the largest such achievable rate. Using the channel estimation error, $\tilde{\mathbf{H}} = \mathbf{H} - \hat{\mathbf{H}}$, we can write (1) as

$$\mathbf{r} = \hat{\mathbf{H}}\mathbf{x} + \tilde{\mathbf{H}}\mathbf{x} + \mathbf{n}. \tag{18}$$

where $\mathbf{x}$ is the information carrying input, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. Let $\mathbf{Q} = E[\mathbf{x}\mathbf{x}^\dagger]$ be the transmit covariance matrix, which has an average power constraint[3] of P_d, $\text{tr}(\mathbf{Q}) \leq P_d$. Although the optimum input distribution is not known, we achieve the following rate with Gaussian $\mathbf{x}$ for a MIMO channel [11],

$$C_{lb} = I(\mathbf{r}; \mathbf{x}|\hat{\mathbf{H}}) \geq E_{\hat{\mathbf{H}}}\left[\log\left|\mathbf{I} + \mathbf{R}_{\tilde{\mathbf{H}}\mathbf{x}+\mathbf{n}}^{-1}\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^\dagger\right|\right] \tag{19}$$

where $\mathbf{R}_{\tilde{\mathbf{H}}\mathbf{x}+\mathbf{n}} = \mathbf{I} + E_{\tilde{\mathbf{H}}}\left[\tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^\dagger\right]$ is the covariance matrix of the effective noise, $\tilde{\mathbf{H}}\mathbf{x} + \mathbf{n}$. By denoting each row of $\tilde{\mathbf{H}}$ as $\tilde{\mathbf{h}}_i^\dagger$, we can write the $(i,j)^{th}$ entry of $E_{\tilde{\mathbf{H}}}\left[\tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^\dagger\right]$ as,

$$E\left[\tilde{\mathbf{h}}_i^\dagger\mathbf{Q}\tilde{\mathbf{h}}_j\right] = \text{tr}\left(\mathbf{Q}E\left[\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_j^\dagger\right]\right) = \begin{cases} \text{tr}(\mathbf{Q}\tilde{\mathbf{\Sigma}}), & \text{when} \quad i = j \\ 0, & \text{when} \quad i \neq j \end{cases} \tag{20}$$

which results in $E\left[\tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^\dagger\right] = \text{tr}(\mathbf{Q}\tilde{\mathbf{\Sigma}})\mathbf{I}$. Since our goal is to find the largest such achievable rate, the rate maximization problem over the entire block becomes

$$R = \max_{\substack{(\mathbf{Q}, P_t, T_t) \in \mathcal{S} \\ \text{tr}(\mathbf{Q}) \leq P_d}} \frac{T - T_t}{T} E_{\hat{\mathbf{H}}}\left[\log\left|\mathbf{I} + \frac{\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^\dagger}{1 + \text{tr}(\mathbf{Q}\tilde{\mathbf{\Sigma}})}\right|\right] \tag{21}$$

where $\mathcal{S} = \left\{(\mathbf{Q}, P_t, T_t)\middle|\text{tr}(\mathbf{Q})T_d + P_tT_t = PT\right\}$, and the coefficient $\frac{T-T_t}{T}$ reflects the amount of time spent during the training phase. The maximization is over the training parameters P_t, and T_t, and the data transmission parameter $\mathbf{Q}$, which can be decomposed into its eigenvectors (the transmit directions), and eigenvalues (powers along the transmit directions).

[3]Note that since the noise power is assumed to be unity, P_d and the eigenvalues of $\mathbf{Q}$ are relative power values with respect to the noise power. They can be regarded as SNR values.

1) Transmit Directions: Unlike the case with no-CSI at the transmitters [12], in this paper, the optimum transmit covariance matrix is not equal to the identity matrix. In this case, the problem becomes that of choosing the eigenvectors (the transmit directions), and the eigenvalues (the powers allocated to the transmit directions), of the transmit covariance matrix $\mathbf{Q} = \mathbf{U}_Q\mathbf{\Lambda}_Q\mathbf{U}_Q^\dagger$, to maximize (21).

When the CSI at the receiver is perfect, [5] showed that the eigenvectors of the transmit covariance and the channel covariance matrices must be equal, i.e., $\mathbf{U}_Q = \mathbf{U}_\Sigma$. In the next theorem, we show that this is also true when there is channel estimation error at the receiver.

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

Proof: We have shown in (15) and (17) that, when $\mathbf{\Sigma} = \mathbf{U}_\Sigma\mathbf{\Lambda}_\Sigma\mathbf{U}_\Sigma^\dagger$, we have $\hat{\mathbf{\Sigma}} = \mathbf{U}_\Sigma\hat{\mathbf{\Lambda}}_\Sigma\mathbf{U}_\Sigma^\dagger$, and $\tilde{\mathbf{\Sigma}} = \mathbf{U}_\Sigma\tilde{\mathbf{\Lambda}}_\Sigma\mathbf{U}_\Sigma^\dagger$. By using (3), we have $\hat{\mathbf{H}} = \hat{\mathbf{Z}}\mathbf{U}_\Sigma\hat{\mathbf{\Lambda}}_\Sigma^{1/2}\mathbf{U}_\Sigma^\dagger$. Inserting these into (21),

$$R = \max_{\substack{(\mathbf{Q},P_t,T_t)\in\mathcal{S}\\ \mathrm{tr}(\mathbf{Q})\le P_d}} \frac{T-T_t}{T} E_{\hat{\mathbf{Z}}}\left[\log\left|\mathbf{I} + \frac{\hat{\mathbf{Z}}\hat{\mathbf{\Lambda}}_\Sigma^{1/2}\mathbf{U}_\Sigma^\dagger\mathbf{Q}\mathbf{U}_\Sigma\hat{\mathbf{\Lambda}}_\Sigma^{1/2}\hat{\mathbf{Z}}^\dagger}{1+\mathrm{tr}\left(\mathbf{U}_\Sigma^\dagger\mathbf{Q}\mathbf{U}_\Sigma\tilde{\mathbf{\Lambda}}_\Sigma\right)}\right|\right] \tag{22}$$

where we used the fact that the random matrices $\hat{\mathbf{Z}}\mathbf{U}_\Sigma$ and $\hat{\mathbf{Z}}$ have the same distribution for zero-mean identity-covariance Gaussian $\hat{\mathbf{Z}}$ and unitary $\mathbf{U}_\Sigma$ [3]. We may spectrally decompose the expression sandwiched between $\hat{\mathbf{Z}}$ and its conjugate transpose in (22) as

$$\hat{\mathbf{\Lambda}}_\Sigma^{1/2}\mathbf{U}_\Sigma^\dagger\mathbf{Q}\mathbf{U}_\Sigma\hat{\mathbf{\Lambda}}_\Sigma^{1/2} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger. \tag{23}$$

Using (23), and the identity $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$, we can write the trace expression in the denominator of (22) as $\mathrm{tr}\left(\mathbf{U}_\Sigma^\dagger\mathbf{Q}\mathbf{U}_\Sigma\tilde{\mathbf{\Lambda}}_\Sigma\right) = \mathrm{tr}\left(\mathbf{U}^\dagger\hat{\mathbf{\Lambda}}_\Sigma^{-1}\tilde{\mathbf{\Lambda}}_\Sigma\mathbf{U}\mathbf{\Lambda}\right)$, and the optimization problem in (22) can be written as

$$R = \max_{\substack{(\mathbf{Q},P_t,T_t)\in\mathcal{S}\\ \mathrm{tr}(\mathbf{Q})\le P_d}} \frac{T-T_t}{T} E_{\hat{\mathbf{Z}}}\left[\log\left|\mathbf{I} + \frac{\hat{\mathbf{Z}}\mathbf{\Lambda}\hat{\mathbf{Z}}^\dagger}{1+\mathrm{tr}\left(\mathbf{U}^\dagger\hat{\mathbf{\Lambda}}_\Sigma^{-1}\tilde{\mathbf{\Lambda}}_\Sigma\mathbf{U}\mathbf{\Lambda}\right)}\right|\right] \tag{24}$$

where we again used the fact that the random matrices $\hat{\mathbf{Z}}\mathbf{U}$ and $\hat{\mathbf{Z}}$ have the same distribution. In (24), the numerator of the objective function does not involve $\mathbf{U}$, and using [18, Theorem 9.H.1.h, page 249], we know that $\mathrm{tr}(\hat{\mathbf{\Lambda}}_\Sigma^{-1}\tilde{\mathbf{\Lambda}}_\Sigma\mathbf{\Lambda}) \le \mathrm{tr}(\mathbf{U}^\dagger\hat{\mathbf{\Lambda}}_\Sigma^{-1}\tilde{\mathbf{\Lambda}}_\Sigma\mathbf{U}\mathbf{\Lambda})$, for all unitary $\mathbf{U}$. Therefore, we can choose $\mathbf{U}^* = \mathbf{I}$ to maximize the rate as long as this choice is feasible. In order to check for the feasibility, we write the trace constraint on $\mathbf{Q}$ using (23) as

$$\mathrm{tr}(\mathbf{Q}) = \mathrm{tr}(\mathbf{U}_\Sigma\hat{\mathbf{\Lambda}}_\Sigma^{-1/2}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger\hat{\mathbf{\Lambda}}_\Sigma^{-1/2}\mathbf{U}_\Sigma^\dagger) = \mathrm{tr}(\mathbf{U}^\dagger\hat{\mathbf{\Lambda}}_\Sigma^{-1}\mathbf{U}\mathbf{\Lambda}). \tag{25}$$

Again from [18, Theorem 9.H.1.h, page 249], $\mathrm{tr}(\hat{\mathbf{\Lambda}}_\Sigma^{-1}\mathbf{\Lambda}) \le \mathrm{tr}(\mathbf{U}^\dagger\hat{\mathbf{\Lambda}}_\Sigma^{-1}\mathbf{U}\mathbf{\Lambda}) \le P_d$, for all unitary $\mathbf{U}$. Therefore, we conclude that $\mathbf{U}^* = \mathbf{I}$ choice is feasible. Then, using $\mathbf{U}^* = \mathbf{I}$, from (23), we have the desired result, $\mathbf{Q}^* = \mathbf{U}_\Sigma\hat{\mathbf{\Lambda}}_\Sigma^{-1}\mathbf{\Lambda}\mathbf{U}_\Sigma$ with $\mathbf{\Lambda}_Q = \hat{\mathbf{\Lambda}}_\Sigma^{-1}\mathbf{\Lambda}$. □

Using Theorem 2, we can write the optimization problem in (21) as,

$$R = \max_{(\boldsymbol{\lambda}^Q,P_t,T_t)\in\mathcal{P}} \frac{T-T_t}{T} E_{\hat{\mathbf{z}}_i}\left[\log\left|\mathbf{I} + \frac{\sum_{i=1}^{n_T}\lambda_i^Q\hat{\lambda}_i^\Sigma\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^\dagger}{1+\sum_{i=1}^{n_T}\lambda_i^Q\tilde{\lambda}_i^\Sigma}\right|\right] \tag{26}$$

where $\hat{\mathbf{z}}_i \sim \mathcal{CN}(0,\mathbf{I})$ is the i^{th} column of $\hat{\mathbf{Z}}$, the set of vectors $\{\hat{\mathbf{z}}_i\}$ are i.i.d, $\boldsymbol{\lambda}^Q = [\lambda_1^Q,\ldots,\lambda_{n_T}^Q]$, and $\mathcal{P} = \left\{(\boldsymbol{\lambda}^Q,P_t,T_t)\,\middle|\,\left(\sum_{i=1}^{n_T}\lambda_i^Q\right)T_d + P_tT_t = PT\right\}$.

2) Power Allocation Policy: In a MIMO system, a transmit strategy is a combination of a transmit direction strategy, and a transmit power allocation strategy, which is the set of optimum eigenvalues of the transmit covariance matrix, $\boldsymbol{\lambda}^Q$, that solves (26). Although Theorem 2 gives us a very simple closed form solution for the optimum transmit directions, solving (26) for $\boldsymbol{\lambda}^Q$ in a closed form does not seem to be feasible due to the expectation operation in the objective function. Therefore, we will develop an iterative algorithm that finds the optimum $\boldsymbol{\lambda}^Q$.

For a single-user MIMO system with perfect CSI at the receiver and partial CSI at the transmitter in the form of covariance feedback, an algorithm that finds the optimum power allocation policy is proposed in [7]. In this section, we extend the algorithm in [7] to the case when there is channel estimation error at the receiver, or in other words, when we have the training signal power and the training signal duration in the sum-rate expression. The algorithm in [7] cannot be trivially generalized to the model in this paper, since, here we have the training power P_t, and the training duration T_t as additional parameters.

By plugging $\hat{\lambda}_i^\Sigma$ and $\tilde{\lambda}_i^\Sigma$ into (26), we get

$$R = \max_{(\boldsymbol{\lambda}^Q,P_t,T_t)\in\mathcal{P}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_{i=1}^{J}\lambda_i^Q(\lambda_i^\Sigma-\mu_S)\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^\dagger}{1+\sum_{i=1}^{J}\lambda_i^Q\mu_S+\alpha}\right|\right] \tag{27}$$

where $\alpha = \sum_{i=J+1}^{n_T}\lambda_i^Q\lambda_i^\Sigma$. Note that J and μ_S are functions of P_t and T_t. Since λ_i^Q, for $i = J+1,\ldots,n_T$ does not contribute to the numerator, we should choose $\lambda_i^Q = 0$, for $i = J+1,\ldots,n_T$. This is to be expected, because we have trained only J transmit directions, and we should now solve for J power values along those directions. Consequently, we have

$$R = \max_{(\boldsymbol{\lambda}^Q,P_t,T_t)\in\mathcal{P}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_{i=1}^{J}\lambda_i^Q(\lambda_i^\Sigma-\mu_S)\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^\dagger}{1+\mu_SP_d}\right|\right]. \tag{28}$$

From Theorem 1, we know that $J \le T_t$. We further claim that while optimizing the rate, it is sufficient to search over those (P_t,T_t) pairs that result in $J = T_t$. In other words, for any pair (P_t,T_t) that results in $J < T_t$, we can find another pair $(P_t,T_t^{'})$ that results in a higher achievable rate. In order to see this, consider a pair (P_t,T_t) that results in $J < T_t$, then let us choose $T_t^{'} = J$. For this choice, the result of Theorem 1 is the same, since the available power can only fill J of the parallel channels. Therefore with $(P_t,T_t^{'}) = (P_t,J)$, the estimation process yields the same channel estimate. When we look at (28), we see that inside of the expectation is the same for both (P_t,T_t) and $(P_t,T_t^{'})$. However, the coefficient in front of the expectation is higher with $(P_t,T_t^{'})$, since $J = T_t^{'} < T_t$.

Therefore $(P_t, T_t^{'})$ yields a higher achievable rate and it is sufficient to search over those (P_t, T_t) pairs that result in $J = T_t$. We can now write (28) as

$$R = \max_{(\boldsymbol{\lambda}^Q, P_t, T_t)\in\mathcal{R}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_{i=1}^{T_t}\lambda_i^Q(\lambda_i^\Sigma - \mu_S)\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^\dagger}{1+\mu_S P_d}\right|\right] \quad (29)$$

where $\mathcal{R} = \left\{(\boldsymbol{\lambda}^Q, P_t, T_t) \,\middle|\, \left(\sum_{i=1}^{n_T}\lambda_i^Q\right)T_d + P_tT_t = PT, P_t > \sum_{i=1}^{T_t}\left(\frac{1}{\lambda_{T_t}^\Sigma} - \frac{1}{\lambda_i^\Sigma}\right)\right\}$, and the condition $P_t > \sum_{i=1}^{T_t}\left(\frac{1}{\lambda_{T_t}^\Sigma} - \frac{1}{\lambda_i^\Sigma}\right)$ guarantees that, using (P_t, T_t), all T_t channels are filled, i.e., $J = T_t$.

Note that the parameters that we want to optimize (29) over are discrete valued T_t, and continuous valued P_t, and $\boldsymbol{\lambda}^Q$. Since T_t is discrete, and $1 \le T_t \le n_T$, we can perform an exhaustive search over T_t and solve n_T reduced optimization problems with fixed T_t at each one. Then, we take the solution that results in the maximum rate, i.e.,

$$R = \max_{1\le T_t\le n_T} R_{T_t} \quad (30)$$

where

$$R_{T_t} = \max_{(\boldsymbol{\lambda}^Q, P_t)\in\mathcal{R}_{T_t}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \frac{\sum_{i=1}^{T_t}\lambda_i^Q(\lambda_i^\Sigma - \mu_S)\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^\dagger}{1+\mu_S P_d}\right|\right] \quad (31)$$

and $\mathcal{R}_{T_t} = \left\{(\boldsymbol{\lambda}^Q, P_t) \,\middle|\, \left(\sum_{i=1}^{n_T}\lambda_i^Q\right)T_d + P_tT_t = PT, \quad P_t > \sum_{i=1}^{T_t}\left(\frac{1}{\lambda_{T_t}^\Sigma} - \frac{1}{\lambda_i^\Sigma}\right)\right\}$. While solving the inner maximization problem, we define $f_i(P_t) = \frac{\lambda_i^\Sigma - \mu_S}{1+\mu_S P_d}$, for $i = 1, \ldots, T_t$. In this case, the inner optimization problem becomes

$$R_{T_t} = \max_{(\boldsymbol{\lambda}^Q, P_t)\in\mathcal{R}_{T_t}} \frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \sum_{i=1}^{T_t}\lambda_i^Q f_i(P_t)\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^\dagger\right|\right] \quad (32)$$

Note that, for the inner optimization problem, in addition to T_t, if P_t was fixed, $f_i(P_t)$ would also be fixed. In this case, the problem in (32) would become exactly the same as the corresponding convex optimization problem with perfect CSI assumption at the receiver [7], where here, $f_i(P_t)$ replaces λ_i^Σ in [7, equation (8)].

In the optimization problem in (32), we have $T_t + 1$ optimization variables, $\lambda_1^Q, \ldots, \lambda_{T_t}^Q$, and P_t. In this case, the problem is not necessarily convex due to the existence of P_t. Equation (32) is concave when $T_t = 1$, which results in an affine $f_1(P_t)$. Therefore, in the most general case, the solution of the first order necessary conditions will give a local maximum. The Lagrangian for (32) can be written as

$$\frac{T-T_t}{T} E\left[\log\left|\mathbf{I} + \sum_{i=1}^{T_t}\lambda_i^Q f_i(P_t)\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^\dagger\right|\right] - \mu\left(\left(\sum_{i=1}^{T_t}\lambda_i^Q\right)T_d + P_tT_t - PT\right) \quad (33)$$

where μ is the Lagrange multiplier, and we omitted the complementary slackness conditions related to the positiveness of λ_i^Q and $P_t - \sum_{i=1}^{T_t}\left(\frac{1}{\lambda_{T_t}^\Sigma} - \frac{1}{\lambda_i^\Sigma}\right)$. By using the identity, $\frac{\partial}{\partial x}\log|\mathbf{A} + x\mathbf{B}| = \text{tr}\left[(\mathbf{A}+x\mathbf{B})^{-1}\mathbf{B}\right]$ which is proved in [5], the KKT conditions can be written by taking the derivative of the Lagrangian with respect to λ_i^Q's and P_t,

$$\frac{T_d}{T} f_i(P_t) E\left[\mathbf{z}_i^\dagger\mathbf{A}^{-1}\mathbf{z}_i\right] \le \mu T_d, \quad i = 1, \ldots, T_t \quad (34)$$

$$\frac{T_d}{T}\sum_{i=1}^{T_t}\lambda_i^Q E\left[\mathbf{z}_i^\dagger\mathbf{A}^{-1}\mathbf{z}_i\right]\frac{\partial f_i(P_t)}{\partial P_t} = \mu T_t \quad (35)$$

where $\mathbf{A} = \mathbf{I} + \sum_{i=1}^{T_t}\lambda_i^Q f_i(P_t)\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^\dagger$, and the equality of the last equation follows from the complementary slackness condition, which says $P_t > \sum_{i=1}^{T_t}\left(\frac{1}{\lambda_{T_t}^\Sigma} - \frac{1}{\lambda_i^\Sigma}\right)$. If the complementary slackness condition is not satisfied, i.e., if we had $P_t \le \sum_{i=1}^{T_t}\left(\frac{1}{\lambda_{T_t}^\Sigma} - \frac{1}{\lambda_i^\Sigma}\right)$, then at least one of the channels out of T_t channels could not be filled, i.e., $J < T_t$, and therefore this choice of (P_t, T_t) pair is not optimal. Therefore, the complementary slackness condition is always satisfied, resulting in the equality in (35).

The i^{th} inequality in (34) is satisfied with equality whenever the optimum λ_i^Q is non-zero, and with strict inequality whenever the optimum λ_i^Q is zero. Due to the expectation term, we cannot directly solve for λ_i^Q in (34). Instead, we multiply both sides of (34) by λ_i^Q,

$$\frac{\lambda_i^Q}{T} f_i(P_t) E\left[\mathbf{z}_i^\dagger\mathbf{A}^{-1}\mathbf{z}_i\right] = \lambda_i^Q\mu, \quad i = 1, \ldots, T_t \quad (36)$$

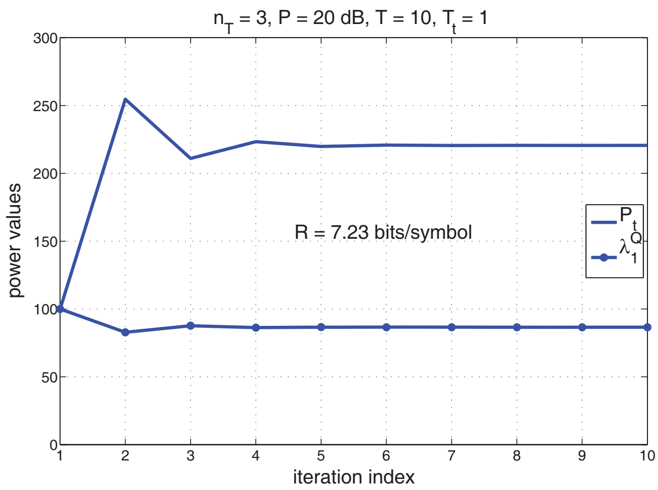
We note that when $\lambda_i^Q = 0$, both sides of (36) are equal to zero. Therefore, unlike (34), (36) is always satisfied with equality for optimum eigenvalues. By summing both sides over all antennas, we find μ, and by substituting this μ into (36), we find the fixed point equations which have to be satisfied by the optimum eigenvalues,

$$\lambda_i^Q = \frac{\lambda_i^Q f_i(P_t) E\left[\mathbf{z}_i^\dagger\mathbf{A}^{-1}\mathbf{z}_i\right]}{\sum_{j=1}^{n_T}\lambda_j^Q f_j(P_t) E\left[\mathbf{z}_j^\dagger\mathbf{A}^{-1}\mathbf{z}_j\right]} P_d, \quad i = 1, \ldots, T_t. \quad (37)$$

This gives a set of fixed point equations that can be used to update $\boldsymbol{\lambda}^Q$, however, we also need a fixed point equation to update the value of P_t. For this, we look at the last KKT equation. Note that when the optimum λ_i^Q is non-zero, the corresponding inequality in (34) will be satisfied with equality due to its corresponding complementary slackness condition. Therefore, we pull the expectation terms from (34) for those equations with non-zero λ_i^Q's, and insert them into (35). Since those indices with $\lambda_i^Q = 0$ do not contribute to (35), we have

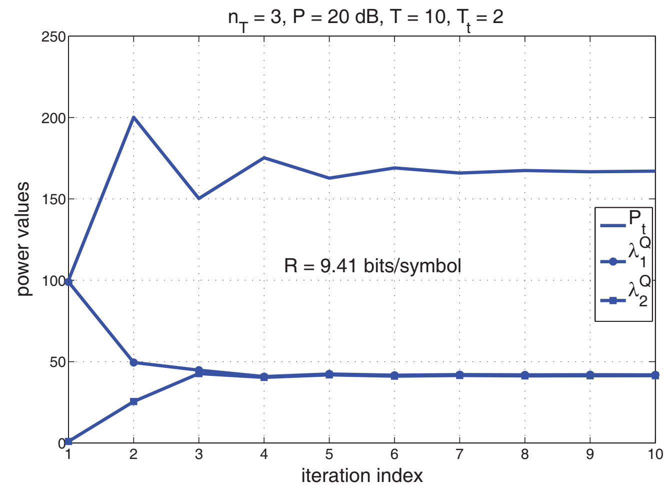
$$\sum_{i=1}^{T_t}\lambda_i^Q\frac{f_i'(P_t)}{f_i(P_t)} = \frac{T_t}{T_d} \quad (38)$$

where we canceled μ's out on both sides. Now, we can use this fixed-point equation to solve P_t in terms of λ_i^Q's. We propose the following fixed-point algorithm: at any given iteration, the



(a)

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



(a)

Fig. 3. The convergence of the single-user algorithm with $n_T = n_R = 3$, 20 dB total average power and $T = 10$, and two symbols long training, $T_t = 2$.

algorithm first solves (38) for the next update of P_t using

$$\sum_{i=1}^{T_t} \lambda_i^Q(n) \frac{f_i'(P_t(n+1))}{f_i(P_t(n+1))} = \frac{T_t}{T_d} \tag{39}$$

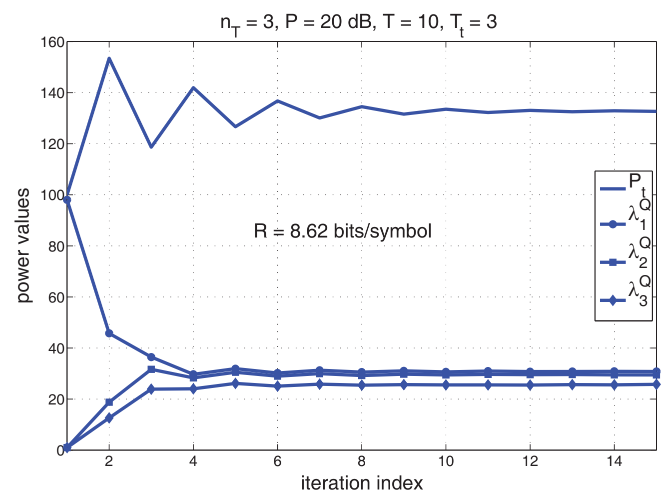
and then, updates $\lambda_i^Q(n+1)$ using (37) for $i = 1, \dots, T_t$

$$\lambda_i^Q(n+1) = \frac{\lambda_i^Q(n) f_i(P_t(n+1)) E\left[\mathbf{z}_i^\dagger \mathbf{A}^{-1} \mathbf{z}_i\right]}{\sum_{j=1}^{n_T} \lambda_j^Q(n) f_j(P_t(n+1)) E\left[\mathbf{z}_j^\dagger \mathbf{A}^{-1} \mathbf{z}_j\right]} P_d \tag{40}$$

where $P_d = \frac{(PT - P_t(n+1)T_t)}{T_d}$. This algorithm finds the solution for the training power P_t, and the eigenvalues of the transmit covariance matrix $\lambda_1^Q, \dots, \lambda_{T_t}^Q$, for a fixed T_t, for $1 \leq T_t \leq n_T$. We run n_T such algorithms, and the solution of (29) is found by taking the one that results in the largest rate, which gives us T_t.

As a result, we solved the joint channel estimation and resource allocation problem that we considered in this paper. Through T_t and P_t, we find the allocation of available time and power over the training and data transmission phases, since total block length and power is fixed. Through Theorem 2, we find the optimum transmit directions, and through $\lambda_1^Q, \dots, \lambda_{T_t}^Q$, we find the allocation of data transmission power over these transmit directions. Finally, the training signal $\mathbf{S}$ is determined by T_t and P_t through Theorem 1.

Analytical proof of the convergence of this algorithm seems to be more complicated than the proof in the case when there is no channel estimation error [7], and seems to be intractable for now. However, in our extensive simulations, we observed that the algorithm always converged. As an example, in Figures 2-4, we consider a system with $n_T = n_R = 3$ having SNR, $P = 20$ dB, and block length, $T = 10$. For this system, we run our algorithm for all three possible values of the training symbol duration, i.e., $T_t = 1, 2, 3$. We observe in Figures 2-4 that estimating two of the three dimensions of the channel is optimum for this setting. Since the power is relatively high with respect to the eigenvalues of the channel



(a)

Fig. 4. The convergence of the single-user algorithm with $n_T = n_R = 3$, 20 dBtotal average power and $T = 10$, and three symbols long training, $T_t = 3$.

covariance, it should be allocated almost equally among the spatial dimensions when there is perfect CSI at the receiver. We see that this is also the case in our model as well. The only difference is that the power is allocated almost equally to the spatial dimensions that are trained. We refer the reader to the Part II [13] of this two-part paper for a more detailed numerical analysis.

IV. Conclusions

We analyzed the joint optimization of the channel estimation and data transmission parameters of a a single-user MIMO block-fading channel where the receiver has a noisy estimate of the channel and the transmitter has the partial CSI in the form of covariance feedback. First the optimum training signal to minimize the MMSE is found, and then, we formulated the joint optimization problem over the eigenvalues of the transmit covariance matrix and the channel estimation parameters. This is solved by introducing a number of reduced optimization problems, each of which can be solved efficiently using the proposed iterative algorithm. Through simulations, it is observed that the proposed iterative algorithm converges to the same point regardless of the initial point of the iterations.

REFERENCES

[1] A. Soysal and S. Ulukus, "Optimizing the rate of a correlated MIMO link jointly over channel estimation and data transmission parameters," in *Proc. Conference on Information Sciences and Systems*, Mar. 2008.

[2] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[3] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[4] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.

[5] 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.

[6] 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.

[7] 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, Sep. 2007.

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.

[13] A. Soysal and S. Ulukus, "Joint channel estimation and resource allocation for MIMO systems–part II: multi-user and numerical analysis," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 632–640, Feb. 2010.

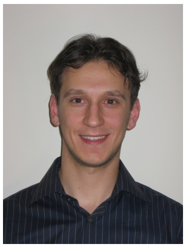
[14] C. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

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

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

[17] 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.

[18] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.



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.