The Effect of Channel Estimation on Transceiver Design for MIMO Systems with QoS Constraints

Semih Serbetli, Sandeep Bethanabhotla and Aylin Yener
Electrical Engineering Department
The Pennsylvania State University
University Park, PA 16802
serbetli@psu.edu *sandeepb@psu.edu* *yener@ee.psu.edu*

***Abstract*— We investigate the transceiver design problem for a multiple input multiple output (MIMO) link considering the effect of channel estimation. We work with the total mean squared error as the performance measure and develop transceiver structures considering the effect of maximum likelihood (ML) channel estimation. The proposed transceiver structures are optimum in the sense of minimizing the total MSE and distributing the MSE equally among the parallel data streams. Next, an upper bound for the number of data streams that can be transmitted for a given target MSE is derived. Motivated by the substantial effect the channel estimation process can have on the system performance, we next investigate the problem of how the MIMO link should distribute its total available power between power expended for channel estimation versus data transmission. The optimum power allocation between the training sequences for ML estimation of the channel and data transmission is derived given the total power budget. Numerical results supporting the analysis and the power allocation schemes are presented.**

I. Introduction

Improved spectral efficiency is needed to meet the rapidly growing demand for high data rate and reliable wireless communications. Recently, the use of multiple antennas at both the transmitter and the receiver side has attracted attention due to their potential to increase the spectral efficiency significantly [1]. There has been considerable research in exploiting the space dimension through transmit diversity, space-time coding and spatial multiplexing for multiple input multiple output (MIMO) systems that employ multiple transmit and/or receive antennas [2]–[5].

Performance of a MIMO system is highly dependent on the channel state information available at both the transmitter and the receiver side. Hence, estimation of the channel at the receiver side and feedback of this information to the transmitter side have significant impact on the performance. In the absence of channel state related feedback to the transmitter side, multiple antennas can be used for spatial multiplexing [6], or for space-time coding [2], [3], [6]. The effect of receiver side channel estimation on such schemes is analyzed in [3], [7]. Spatial multiplexing can significantly benefit from transmit precoding when channel information is available at the transmitter side. In such cases, designing the appropriate precoding strategy has been studied under a variety of system objectives [4], [5], [8], [9]. All of these studies have assumed exact channel state information at the transmitter side.

In order to fully realize the substantial capacity gain for a MIMO system, the MIMO channel has to be estimated at the receiver, and in turn should be fedback to the transmit side if precoding is to be employed. In this context, optimal transmission strategies with channel estimation are investigated for space-time coding and eigen beamforming transmission schemes [10], [11]. In practice, it is likely that the total transmission power budget would be limited for the MIMO system. When this is the case, it is meaningful to ask what fraction of the resources should be devoted to estimation versus actual data transmission. Related previous work in this area includes the design of optimum training sequences for ML estimation of MIMO channels and the effect of channel estimation on BLAST systems [12]. In the same context, considering the sum capacity as the performance metric, optimum training sequences and power allocation among the training sequences and data transmission are found in [7], [12].

In this paper, we consider a MIMO system where the transmitter and receiver have access to the same channel state information (CSI). CSI is assumed to be obtained at the receiver side by training and fedback to the transmitter side via an error-free feedback channel. Throughout the paper we will use the total mean squared error (MSE) as our performance metric [5], [13]. The contribution of the paper is three fold. We first investigate the effect of channel estimation process on the design of the precoder and the decoder for the MIMO system, with the objective of minimizing the total mean squared error (MSE). Next, we consider fairness constraints and aim at providing each symbol transmitted with an equal MSE performance. Using the precoder-decoder pair we construct and by allowing the data streams to be transmitted by the MIMO link to interfere with each other, we next find an upper bound for the number of independent data streams that can be transmitted through the channel for a given target MSE. The upper bound is achievable with the proposed structure, and can be larger than the rank of the channel matrix, an upper bound suggested by earlier work [5]. Third, motivated by the profound effect of the quality of channel estimation on the performance of the MIMO link, we consider the problem optimum sharing of resources between the process of channel estimation and data transmission. We consider the total power as the limited resource, and show that, given the coherence time of the channel, there is a unique solution to the optimum allocation problem between the training based ML channel estimation and data transmission.

The organization of the paper is as follows. In Section II, the system model is described. The MSE is formulated, the optimum transceiver structure and the upper bound on the number of independent data streams are derived in Sections III and IV. The effect of power allocation is investigated and a power allocation scheme minimizing the total MSE is given in Section V. Section VI provides the numerical results supporting the analysis. Section VII concludes the paper.

II. System Model and Performance Metric

We consider a communication link consisting of N_T transmitter antennas and N_R receiver antennas. The transmitter multiplexes a fixed number of data streams M through its N_T transmit antennas employing an $N_T \times M$ linear transmitter $\mathbf{F}$ in one symbol period. We assume that the number of data streams is given and fixed first, and analyze the effect of channel estimation on the design process of linear transmitter and receiver. We remove this assumption later in the paper, and determine the maximum number of data streams that can be transmitted with a certain MSE target. Similar to the notation in [5], the received vector is

$$\mathbf{r} = \mathbf{HFs} + \mathbf{n} \tag{1}$$

where, $\mathbf{s}$ is the $M \times 1$ symbol vector, $\mathbf{H}$ is the $N_R \times N_T$ matrix of complex channel gains and $\mathbf{n}$ is the zero mean complex Gaussian noise vector with $E\left[\mathbf{nn}^\dagger\right] = \sigma^2\mathbf{I}$, and $(\cdot)^\dagger$ denotes the hermitian of a vector or matrix.

We assume that the channel is flat fading with coherence time of $(N_T + L)$ symbols where N_T symbol intervals are dedicated to training sequences, and the remaining L to data transmission. The total power available to the system for the entire interval is P_{total} where P_{tr} portion of it is used for the transmission of the training sequences and the remaining portion is distributed equally among the L symbols. Thus, the precoder should be designed with the power constraint $tr\{\mathbf{FF}^\dagger\} \leq P_s = (P_{total} - P_{tr})/L$.

Throughout the paper, it is assumed that the receiver has the ML estimate of the MIMO channel $\overline{\mathbf{H}} = \mathbf{H} + \mathbf{X}$ that is perfectly fedback to the transmitter via an error-free and low-delay feedback channel. Following the ML estimate model of the MIMO channel in [7], [12], $\mathbf{X}$ is a random matrix with i.i.d. complex Gaussian entries having $CN(0, \sigma_e^2 = \frac{(\sigma^2 N_T)}{P_{tr}})$ and is independent of the MIMO channel $\mathbf{H}$. In this model, $\mathbf{H}$ is a realization of a random matrix with i.i.d. complex Gaussian entries having $CN(0, \sigma_H^2)$. Using Bayes' rule, the distribution of $(\mathbf{H})_{ij} = h_{ij}$ and $(\mathbf{X})_{ij} = x_{ij}$ for a given estimate $(\overline{\mathbf{H}})_{ij} = \overline{h_{ij}}$ can be expressed as

$$f_h(h_{ij}|\overline{h_{ij}}) = f_x(\overline{h_{ij}} - h_{ij}|\overline{h_{ij}}) \tag{2}$$

$$= \frac{f_h(h_{ij}) f_x(\overline{h_{ij}} - h_{ij})}{\int f_h(h_{ij}) f_x(\overline{h_{ij}} - h_{ij}) dh_{ij}} \tag{3}$$

resulting in

$$f(h_{ij}|\overline{h_{ij}}) \sim CN(\overline{h_{ij}} \frac{\sigma_H^2}{\sigma_H^2 + \sigma_e^2}, \frac{\sigma_H^2 \sigma_e^2}{\sigma_H^2 + \sigma_e^2}) \tag{4}$$

$$f(x_{ij}|\overline{h_{ij}}) \sim CN(\overline{h_{ij}} \frac{\sigma_e^2}{\sigma_H^2 + \sigma_e^2}, \frac{\sigma_H^2 \sigma_e^2}{\sigma_H^2 + \sigma_e^2}) \tag{5}$$

Observe that from the receiver's perspective, the actual channel is a random MIMO channel with i.i.d. complex Gaussian entries of $CN(\frac{\sigma_H^2}{\sigma_H^2 + \sigma_e^2}(\overline{\mathbf{H}})_{i,j}, \frac{\sigma_H^2 \sigma_e^2}{\sigma_H^2 + \sigma_e^2})$.

Let us denote the $M \times N_R$ linear receiver by $\mathbf{G}$; the decision statistic $\mathbf{y}$ is given by

$$\mathbf{y} = \mathbf{GHFs} + \mathbf{Gn} \tag{6}$$

Then, the total MSE is

$$\begin{aligned} \text{MSE} &= E\left[\|\mathbf{y} - \mathbf{s}\|^2\right] = \\ & E\Big[\text{tr}\Big\{\mathbf{F}^\dagger\mathbf{H}^\dagger\mathbf{G}^\dagger\mathbf{GHF} - \mathbf{F}^\dagger\mathbf{H}^\dagger\mathbf{G}^\dagger - \mathbf{GHF} + \mathbf{I} + \sigma^2\mathbf{GG}^\dagger\Big\}\Big] \end{aligned} \tag{7}$$

where tr$\{\mathbf{A}\}$ denotes the trace of matrix $\mathbf{A}$.

Total MSE minimization by choosing the transmitters and receivers has recently been studied for MIMO systems with exact channel state information [5]. In practice, the channel state information available to the transmitter and receiver may not be perfect. In addition, it is meaningful to consider a system where fairness is facilitated by ensuring that each symbol experiences equal MSE. In the following section, we pose the problem of minimizing the total MSE considering fairness among the parallel data streams, in the presence of a given ML estimate of the channel at both transmitter and receiver, and construct the optimum transceiver structure.

III. Optimum Transceiver Structure

Our aim in this section is to find the optimum transceiver structure in the sense of minimizing the total MSE and distributing the total MSE equally among the parallel data streams while considering the effect of ML channel estimation. Formally, the optimization problem is

$$\begin{aligned} \min_{\{\mathbf{F},\mathbf{G}\}} \quad & \text{MSE} = \text{tr}\{\mathbf{B}\} \\ \text{s.t. } \mathbf{F}^\dagger\mathbf{F} \leq P_s; \quad & \text{MSE}_1 = \text{MSE}_2 = ... = \text{MSE}_M \end{aligned} \tag{8}$$

$$\mathbf{B} = E\left[\mathbf{F}^\dagger\mathbf{H}^\dagger\mathbf{G}^\dagger\mathbf{GHF} - \mathbf{F}^\dagger\mathbf{H}^\dagger\mathbf{G}^\dagger - \mathbf{GHF} + \mathbf{I} + \sigma^2\mathbf{GG}^\dagger\right] \tag{9}$$

and MSE_i is the individual MSE of data stream i, and is the (i,i)th entry of $\mathbf{B}$.

As we mentioned in the previous section, the actual channel matrix is not known at the receiver side and from the receiver's perspective it is a random MIMO channel with a known distribution. Defining $\rho = \frac{\sigma_H^2}{\sigma_e^2 + \sigma_H^2}$ and using the distribution in (4), $\mathbf{H}$ can be modelled as $\mathbf{H} = \rho\overline{\mathbf{H}} + \mathbf{Y}$ where $\mathbf{Y}$ is a random matrix with i.i.d. complex Gaussian entries of $CN(0, \sigma_y^2 = \frac{\sigma_H^2 \sigma_e^2}{\sigma_H^2 + \sigma_e^2})$. Reformulating the MSE in terms of $\overline{\mathbf{H}}$ and $\mathbf{Y}$ we have

$$\begin{aligned} \text{MSE} = E\{\text{tr}\{&\rho^2\mathbf{F}^\dagger\overline{\mathbf{H}}^\dagger\mathbf{G}^\dagger\mathbf{G}\overline{\mathbf{H}}\mathbf{F} + \mathbf{F}^\dagger\mathbf{Y}^\dagger\mathbf{G}^\dagger\mathbf{GYF} \\ &+\rho\mathbf{F}^\dagger\mathbf{Y}^\dagger\mathbf{G}^\dagger\mathbf{G}\overline{\mathbf{H}}\mathbf{F} + \rho\mathbf{F}^\dagger\overline{\mathbf{H}}^\dagger\mathbf{G}^\dagger\mathbf{GYF} \\ &+\rho\mathbf{F}^\dagger\overline{\mathbf{H}}^\dagger\mathbf{G}^\dagger + \mathbf{F}^\dagger\mathbf{Y}^\dagger\mathbf{G}^\dagger \\ &-\rho\mathbf{G}\overline{\mathbf{H}}\mathbf{F} - \mathbf{GYF} \\ &\mathbf{I} + \sigma^2\mathbf{GG}^\dagger\}\} \end{aligned} \tag{10}$$

Using the properties of random matrices with zero mean i.i.d. entries, $E\{\mathbf{AY}\} = 0$ and $E\{\mathbf{YAA}^\dagger\mathbf{Y}^\dagger\} = \sigma_y^2 tr\{\mathbf{AA}^\dagger\}\mathbf{I}$ for an arbitrary matrix $\mathbf{A}$, and taking the expectation with respect to $\mathbf{Y}$, the MSE can be expressed as

$$\text{MSE} = \text{tr}\{\rho^2\mathbf{F}^\dagger\overline{\mathbf{H}}^\dagger\mathbf{G}^\dagger\mathbf{G}\overline{\mathbf{H}}\mathbf{F} + \rho(\mathbf{F}^\dagger\overline{\mathbf{H}}^\dagger\mathbf{G}^\dagger - \mathbf{G}\overline{\mathbf{H}}\mathbf{F}) + \mathbf{I} + \rho_1\mathbf{G}\mathbf{G}^\dagger\} \tag{11}$$

where $\rho_1 = \sigma^2 + \rho\sigma_e^2 P_s$. Observe that the MSE in (11) has the same form of the MSE expression of a MIMO system with a channel matrix $\rho\overline{\mathbf{H}}$ and an AWGN factor with variance ρ_1.

Let us now consider the minimization of the MSE in terms of the precoder and decoder. The first order condition with respect to the linear receiver (decoder) results in the well-known MMSE receiver

$$\mathbf{G} = \rho\mathbf{F}^\dagger\overline{\mathbf{H}}^\dagger\left(\rho_1\mathbf{I} + \rho^2\overline{\mathbf{H}}\mathbf{F}\mathbf{F}^\dagger\overline{\mathbf{H}}^\dagger\right)^{-1} \tag{12}$$

Using (12), the total MSE function can be reformulated as

$$\begin{aligned} \text{MSE} &= \text{tr}\{\mathbf{I} - \rho^2\mathbf{F}^\dagger\overline{\mathbf{H}}^\dagger\left(\rho_1\mathbf{I} + \rho^2\overline{\mathbf{H}}\mathbf{F}\mathbf{F}^\dagger\overline{\mathbf{H}}^\dagger\right)^{-1}\overline{\mathbf{H}}\mathbf{F}\} \quad (13)\\ &= M - N_R + \text{tr}(\mathbf{T}^{-1}) \quad (14)\end{aligned}$$

where $\mathbf{T} = \mathbf{I} + \frac{\rho^2}{\rho_1}\overline{\mathbf{H}}\mathbf{F}\mathbf{F}^\dagger\overline{\mathbf{H}}^\dagger$. Notice that the optimum linear transmitter and receiver set that minimizes the MSE is not unique and any linear transmitter that achieves the same covariance $\mathbf{F}\mathbf{F}^\dagger = \widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^\dagger$ achieves the same total MSE. Specifically, all optimum linear transmitters are in the form of $\{\mathbf{F}_{opt}^\star\} = \mathbf{F}_{opt}\mathbf{U}^\dagger$ where $\mathbf{U}$ is any arbitrary matrix satisfying $\mathbf{U}^\dagger\mathbf{U} = \mathbf{I}$.

It is evident that the minimum MSE without any constraints lower bounds the minimum MSE with fairness constraints. However, as we explain below, the optimum transceiver structure with fairness constraints lies in this set of transmitters that yield the unconstrained minimum MSE for a special structure of $\mathbf{U}$.

Reference [5] suggests that one possible optimum linear transmitter is in the form of $\mathbf{VQ}_f$ where $\mathbf{V}$ is the unitary matrix that has columns as eigenvectors of $\overline{\mathbf{H}}^\dagger\overline{\mathbf{H}} = \mathbf{V}\Lambda\mathbf{V}^\dagger$ and $\mathbf{Q}_f = (\mu^{-1/2}\Lambda^{-1/2} - \Lambda^{-1/2})_+^{1/2}$ is a diagonal matrix with μ factor satisfying the power constraint, and $(.)_+ = max(0, .)$. Specifically, all optimum linear transmitters are in the form of $\mathbf{F}_{opt} = \mathbf{VQ}_f\mathbf{U}^\dagger$ where $\mathbf{U}^\dagger\mathbf{U} = \mathbf{I}$. Then, the MSE of each data stream can be expressed by the diagonal entries of

$$\begin{aligned} \mathbf{B} &= \mathbf{I} - \rho^2\mathbf{F}^\dagger\overline{\mathbf{H}}^\dagger\left(\rho_1\mathbf{I} + \rho^2\overline{\mathbf{H}}\mathbf{F}\mathbf{F}^\dagger\overline{\mathbf{H}}^\dagger\right)^{-1}\overline{\mathbf{H}}\mathbf{F} \quad (15)\\ &= \mathbf{UDU}^\dagger \quad (16)\end{aligned}$$

where $\mathbf{D}$ is a diagonal matrix.

Now, consider a MIMO system where equal MSE values for each symbol is required. The eigenvalues of $\mathbf{B}$ in (15) are the diagonal entries of $\mathbf{D}$, and are independent of $\mathbf{U}$. Thus, the sum of the achieved MSE's of the data streams is independent of the choice of $\mathbf{U}$. However, the individual MSE that each data stream achieves does depend on $\mathbf{U}$. Since the diagonal entries of $\mathbf{B}$, the achieved MSE of each data stream, are desired to be equal, a $\mathbf{U}$ that results in a $\mathbf{B}$ matrix with equal diagonal entries is needed. Observe that the eigenvalues of $\mathbf{B}$ always majorizes the equal diagonal entries which ensures the existence of such a matrix $\mathbf{U}$ [14]. Thus, for a given optimum linear transmitter, there always exists another optimum linear transmitter that provides equal MSE for each data stream. An example of such a transceiver structure is given in [13] where each entry of the unitary matrix $\mathbf{U}$ is given by $u_{kl} = e^{(\frac{j2\pi kl}{M})}$.

IV. Rate Allocation

In the previous section, we considered the transceiver optimization problem for a given number of symbols to be transmitted. In this section, we investigate how many parallel data streams the transmitter can send given that each symbol has to experience an MSE lower than or equal to a given MSE target.

The optimization problem is

$$\max_{\{\mathbf{F},\mathbf{G}\}} \quad M \tag{17}$$

s.t. $\text{tr}\{\mathbf{F}^\dagger\mathbf{F}\} \leq P_s$; $\text{MSE}_1 = \text{MSE}_2 = ... = \text{MSE}_M \leq \beta$

Using the transceiver structure proposed in the previous section, one can always distribute the total MSE to each parallel data streams equally. Thus, individual MSE constraints simply reduces to the total MSE constraint:

$$\text{MSE} = \sum_{i=1}^{M}\text{MSE}_i \leq M\beta \tag{18}$$

Using (14) an upper bound for the number of parallel data streams can be formulated as

$$\begin{aligned} M - N_R + \text{tr}(\mathbf{T}^{-1}) &\leq M\beta \\ M &\leq \frac{N_R - \text{tr}(\mathbf{T}^{-1})}{1-\beta} \quad (19)\end{aligned}$$

where $\mathbf{T} = \mathbf{I} + \frac{\rho^2}{\rho_1}\overline{\mathbf{H}}\mathbf{F}\mathbf{F}^\dagger\overline{\mathbf{H}}^\dagger$ with $\mathbf{F} = \mathbf{VQ}_f$. Observe that the bound can exceed the rank of $\mathbf{H}$ depending on the value of β. Especially for large values of β, that is if a larger MSE can be tolerated, the upper bound can be much higher than the rank of the channel matrix which is an upper bound for orthogonal transmissions [5].

We note that the bound in (19) is achievable via the appropriate choice of the precoder-decoder pair as explained in Section III and serves as a feasibility constraint for the MIMO system. Observe that the MIMO system can be viewed as a CDMA system with channel and power constraints. For a CDMA system without any channel and power constraints a well-known upper bound exists on the user capacity [15]. The feasibility condition is

$$\sum_{i=1}^{M}\frac{SIR_i}{1+SIR_i} \leq N_R \tag{20}$$

where SIR_i is the SIR target of user i. Using $\frac{SIR_i}{1+SIR_i} = 1 - MSE_i$, and MSE target β, we can obtain an upper bound for the number of data streams that can be transmitted with an MSE target as

$$M \leq \frac{N_R}{1-\beta} \tag{21}$$

for the CDMA system. Observe however that the bound in (19) which is attainable is tighter than (21) for the MIMO system. This is due to the fact that the upper bound derived in (19) considers the constraints of the MIMO system, i.e., the channel constraints and power constraints.

After finding the maximum number of data streams that can be transmitted through the MIMO channel, one can easily construct a transceiver structure by changing the structure of the matrix $\mathbf{U}$ to satisfy the equal MSE constraints.

V. Optimal Power Allocation

It is evident from the preceding discussion in this paper as well as several other references, e.g., [7], [10]–[12], [16] that the availability of an accurate channel estimate has a substantial impact on the performance of a MIMO link. Therefore, it makes sense to devote some part of system resources to the channel estimation process if in turn the gain in performance is worth the effort. In practice, it is likely that, in a given interval, where the channel is likely to be static, the link would operate with a limited total budget. It is then meaningful to ask what fraction of this total power budget should be expended on the transmission of training sequences that are used in estimating the channel, versus the transmission of actual data. In this section, we investigate this optimum power allocation problem and show that it has a unique solution. The performance metric we consider for the system is total MSE. We note that minimizing the total MSE is equivalent to minimizing tr$(\mathbf{T}^{-1})$

$$\min \text{MSE} \equiv \min(\text{tr}\{\left(\mathbf{I}+\frac{\rho^2}{\rho_1}\overline{\mathbf{H}}\mathbf{F}\mathbf{F}^\dagger\overline{\mathbf{H}}^\dagger\right)^{-1}\}) \qquad (22)$$

Recall that $\overline{\mathbf{H}}$ is a random matrix with i.i.d. complex Gaussian entries with $CN(0,(\sigma_H^2+\sigma_e^2))$ and tr$\{\mathbf{F}\mathbf{F}^\dagger\}=P_s$. Normalizing the entries $\overline{\mathbf{H}}$ and $\mathbf{F}\mathbf{F}^\dagger$ in the MSE expression, we have

$$\min \text{MSE} \equiv \min(\text{tr}\{\left(\mathbf{I}+\frac{\rho^2 P_s(\sigma_H^2+\sigma_e^2)}{\rho_1}\tilde{\mathbf{H}}\tilde{\mathbf{R}}\tilde{\mathbf{H}}^\dagger\right)^{-1}\}) \qquad (23)$$

where $\tilde{\mathbf{R}}=(1/P_s)\mathbf{F}\mathbf{F}^\dagger$ and $\tilde{\mathbf{H}}=1/(\sigma_H^2+\sigma_e^2)^{1/2}\overline{\mathbf{H}}$. Due to normalization $\tilde{\mathbf{H}}$ has independent and identically distributed entries with unit variance and $\tilde{\mathbf{R}}$ has unit trace. Thus, the expression $\rho_e=\frac{\rho^2P_s(\sigma_H^2+\sigma_e^2)}{\rho_1}$ acts like the effective SNR of the system.

Note that as the effective SNR is increased the MSE function is decreased. Hence the minimization of MSE is equivalent to the maximization of ρ_e in terms of power allocation. Recall that we have the following relationship between data transmission power and training sequences power

$$P_sL+P_{tr}=P_{total} \qquad (24)$$

Defining α to be the fraction of the total power devoted to data transmission, i.e.,

$$\alpha=\frac{P_sL}{P_{total}} \quad 0\le\alpha\le1 \qquad (25)$$

and $c=\frac{(N_T-L)P_{total}{\sigma_H}^2}{LN_T\sigma^2+LP_{total}{\sigma_H}^2}$, the effective SNR ρ_e can be expressed as

$$\rho_e=\frac{P_{total}^2{\sigma_H}^4}{\sigma^2L({\sigma_H}^2P_{total}+N_T\sigma^2)}\frac{\alpha(1-\alpha)}{c\alpha+1} \qquad (26)$$

Let us define the function

$$f(\alpha)=\frac{\alpha(1-\alpha)}{c\alpha+1} \qquad (27)$$

The maximization of the effective SNR, ρ_e is equivalent to maximizing $f(\alpha)$ over $0\le\alpha\le1$. The result is given by the following theorem.

Theorem 1: The maximizer of $f(\alpha)$ always lies in $[0,1]$. The optimum fraction of power allocated to data transmission, α_{opt}, is given by

$$\alpha_{opt}=\begin{cases}\frac{-1+\sqrt{1+c}}{c}, & \text{for } N_T>L;\\ \frac{1}{2}, & \text{for } N_T=L;\\ \frac{-1+\sqrt{1+c}}{c}, & \text{for } N_T<L;\end{cases} \qquad (28)$$

Proof: The optimization problem is at hand the maximization of $f(\alpha)$ over $0\le\alpha\le1$ with $f(\alpha)$ given by (27). Observe that $f(\alpha)$ has an asymptote at $\alpha=\frac{-1}{c}$. When $c>0, \frac{-1}{c}<0$ and when $c<0, \frac{-1}{c}>1$. Thus, the pole always lies outside the interval of interest. In order to find the maximum, we first analyze the derivatives of the function $f(\alpha)$ $\forall\alpha$. The first and the second derivatives of the function with respect to α are

$$\frac{\partial(f(\alpha))}{\partial\alpha} = \frac{-(\alpha)^2c-2\alpha+1}{(1+c\alpha)^2} \qquad (29)$$

$$\frac{\partial^2(f(\alpha))}{\partial\alpha^2} = \frac{-2-2c}{(1+c\alpha)^3} \qquad (30)$$

The behavior of the derivatives of the function is as shown below:

1. When $c>0$,
   $\frac{\partial^2(f(\alpha))}{\partial\alpha^2}>0$ for $\alpha<\frac{-1}{c}$
   $\frac{\partial^2(f(\alpha))}{\partial\alpha^2}<0$ for $\alpha>\frac{-1}{c}$.
2. When $c<0$,
   $\frac{\partial^2(f(\alpha))}{\partial\alpha^2}>0$ for $\alpha>\frac{-1}{c}$
   $\frac{\partial^2(f(\alpha))}{\partial\alpha^2}<0$ for $\alpha<\frac{-1}{c}$.
3. When $c=0$, $\frac{\partial^2(f(\alpha))}{\partial\alpha^2}<0$ $\forall\alpha$.

Observe that the first derivative has two zeroes. When $c>0$ as $\alpha\to(\frac{-1}{c})^+$ and as $\alpha\to\infty, f(\alpha)\to-\infty$. Also as $\alpha\to(\frac{-1}{c})^-$ and as $\alpha\to-\infty, f(\alpha)\to\infty$. These properties along with the fact that $\frac{\partial^2(f(\alpha))}{\partial\alpha^2}<0$ in $\alpha\in[0,1]$ implies that a maximum exists in this interval. A similar analysis for when $c<0$ enables us to say that a maximum exists in the interval [0,1].

The first order condition for finding the optimum α is

$$-(\alpha)^2c-2\alpha+1=0 \qquad (31)$$

Solving the equation for all cases we obtain the following

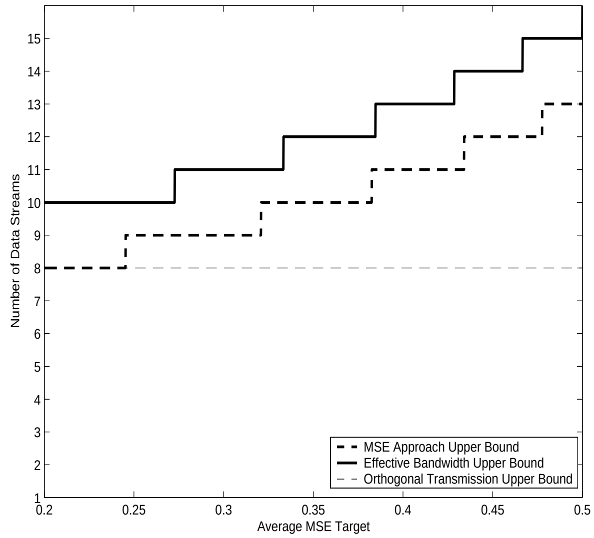

Fig. 1. MSE target vs upper bound for the number of data streams for 8×8 MIMO system with $Ps = 8$, $P_{tr} = 64$

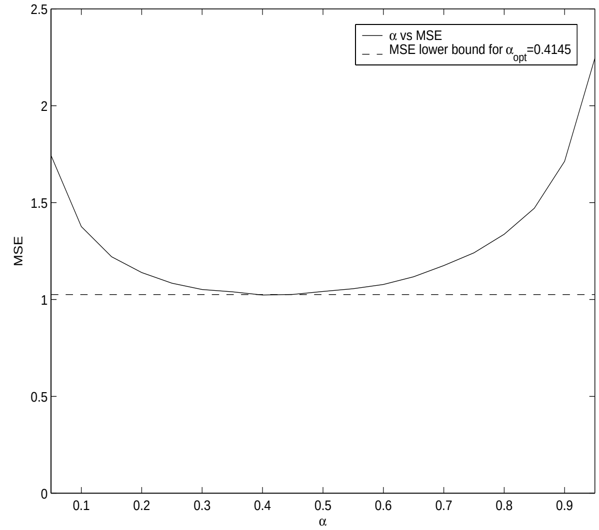

Fig. 2. α vs MSE for 8×8 MIMO system with $L = 4$ and $P_{total} = 100$

roots,

$$\alpha = \begin{cases} \frac{-1}{c} \pm \frac{\sqrt{1+c}}{c}, & c > 0 \ ; \\ \frac{-1}{c} \pm \frac{\sqrt{1+c}}{c}, & -1 < c < 0 \ ; \\ 1, & c < -1 \ ; \end{cases} \quad (32)$$

It can be seen that, for all cases, there is only one root which lies in the interval $[0, 1]$ and $\frac{\partial^2(f(\alpha))}{\partial \alpha^2} < 0$ in this interval. Recalling the fact that the pole always lies outside the interval $[0, 1]$ we conclude, there exists a unique α_{opt} in the interval $[0, 1]$ which maximizes the effective SNR. □

The value of α_{opt} depends upon the number of transmitter antennas and the length of the time interval used for symbol transmission. We observe that when the number of transmit antennas is larger than the time interval used for data transmission interval i.e $N_T > L$ then α_{opt} lies in the range of $[0, \frac{1}{2})$. This result suggests allocating more power to training process for such systems due to the large number of transmitter antennas. As the number of antennas is increased more power is allocated to the estimation process. When $N_T \to \infty$, then $c \to \frac{P_{total}\sigma_H{}^2}{\sigma^2 L}$. This implies that even when the number of transmitter antennas is very large, the power that should be allocated to the training sequences has a limit.

Similarly, when the number of symbols to be transmitted is greater than the number of transmit antennas, i.e., $N_T < L$, then the range of α_{opt} is $(\frac{1}{2}, 1]$. This implies that when the data transmission interval is much larger than the number of transmitter antennas, significant portion of the system power should be allocated to symbol transmission rather than the estimation process. When $L \to \infty$, then $c \to \frac{-P_{total}\sigma_H{}^2}{P_{total}\sigma_H{}^2 + N_T\sigma^2}$. Thus, however long the channel coherence interval is, a nonzero portion of the total power should be allocated to training.

For the case where the number of transmit antennas is equal to the number of symbols to be transmitted, i.e., $N_T = L$, α_{opt} is $\frac{1}{2}$. That is, the available power should be allocated equally between training and data transmission.

VI. Numerical Results

In this section, we present numerical results related to the performance of the proposed transceiver structures and power allocation for channel estimation and data transmission. For numerical results, we consider a 8×8 MIMO system transmitting 8 data streams. The channel values are generated as realizations of a random matrix with i.i.d. complex Gaussian entries of $CN(0, 1)$ and the variance of AWGN is 0.05.

We consider a 8×8 MIMO system with a given channel and estimation procedure. Power used for training is $P_{tr} = 64$ and the power used for data transmission per symbol interval is $P_s = 8$. We plot, in Figure 1, the number of data streams that can be transmitted with a given MSE target, to highlight the differences between the bounds suggested by this paper, [5] and [15]. The MSE approach bound is achievable by the proposed precoder and decoder structure in Section IV, and is larger than what is achievable by orthogonal transmissions [5]. The CDMA user capacity bound [15] is not achievable due to the constraints on the power and the MIMO channel structure.

To investigate the effect of the power allocation among the channel estimation process and data transmission, we consider three different values of L. First we consider a 8×8 MIMO System with $L = 4$ and $P_{total} = 100$, and evaluate the total MSE over 10000 realizations of the MIMO channel for different power allocation schemes. Figure 2 shows the effect of power allocation on the MSE as α changes. It is observed that the optimal power allocation achieves the minimum MSE. For $N_T = L = 8$, the effect of power allocation on the MSE is presented in Figure 3 where $\alpha_{opt} = 1/2$. The third case, $N_T < L$ with $L = 20$ is investigated in Figure 4. In Fig 5, α vs the scaled version of the effective SNR, $f(\alpha)$ is presented where α_{opt} maximizes the effective SNR. Observe that as L increases the power allocation favors to the data transmission rather than the estimation process.

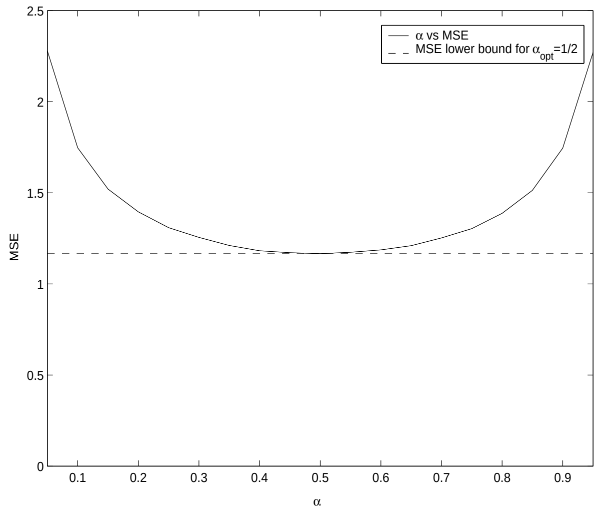

Fig. 3. α vs MSE for 8×8 MIMO system with $L = 8$ and $P_{total} = 100$

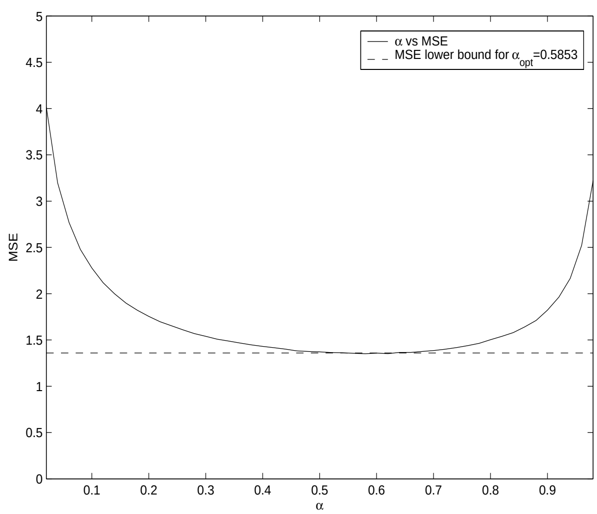

Fig. 4. α vs MSE for 8×8 MIMO system with $L = 20$ and $P_{total} = 100$

VII. CONCLUSIONS

In this paper we have considered the linear transceiver structure for a MIMO link which minimizes the total MSE in the presence of channel estimation errors, and distributes the MSE equally among the parallel data streams. Using the proposed optimum precoder and decoder, we have derived an upper bound on the maximum number of data streams that can be transmitted by a MIMO system for a given target MSE. This upper bound is achievable with the appropriate choice of the precoder and decoder and can be larger than the rank of the channel matrix. This paper also considers the resource trade-off between channel estimation and data transmission for a MIMO link with limited total power. Specifically, the optimal power allocation scheme between the training sequences and data transmission is derived. We observe that the power allocation depends on the system parameters, namely the number of transmit antennas and data transmission interval.

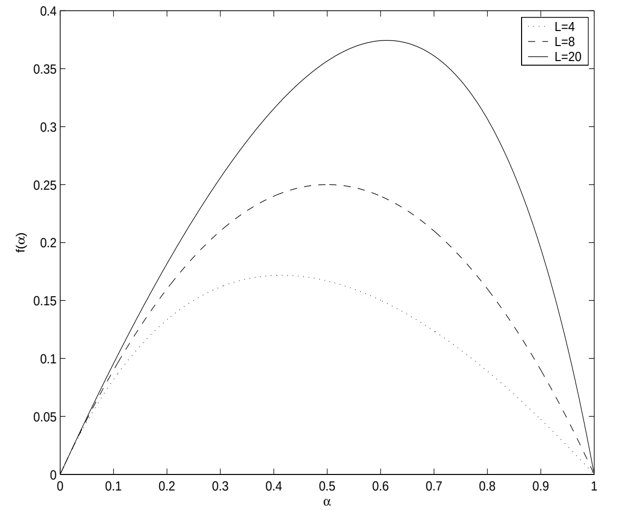

Fig. 5. α vs $f(\alpha)$ for 8×8 MIMO system with $L = 4, 8, 20$ and $P_{total} = 100$

REFERENCES

[1] E. Telatar. Capacity of multi-antenna gaussian channels. 1995. AT&T Bell Labs Memorandum.

[2] S. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications*, 16(8):1451–1458, Oct. 1998.

[3] V. Tarokh, N. Seshadri, and R. Calderbank. Space-time codes for high data rate wireless communication: performance criterion and code construction. *IEEE Transactions on Information Theory*, 44(2):744–765, March 1998.

[4] G. G. Raleigh and J. M. Cioffi. Spatio-temporal coding for wireless communication. *IEEE Transactions on Communications*, 46(3):357–366, March 1998.

[5] H. Sampath, P. Stoica, and A. Paulraj. Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion. *IEEE Transactions on Communications*, 49(12):2198–2206, Dec. 2001.

[6] G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Technical Journal*, 1(2):41–59, Autumn 1996.

[7] B. Hassibi and B. M. Hochwald. How much training is needed in multiple-antenna wireless links? *IEEE Transactions on Information Theory*, 48(4):951 – 963, April 2003.

[8] A. Scaglione, P. Stoica, S. Barbarossa, G. B.Giannakis, and H. Sampath. Optimal designs for space-Time linear precoders and decoders. *IEEE Transactions on Signal Processing*, 50(5):1051–1064, May 2002.

[9] J. Yang and S. Roy. On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems. *IEEE Transactions on Communications*, 42(12):3221–3231, Dec. 1994.

[10] S. Zhou and G. B. Giannakis. Optimal transmitter eigen-beamforming and space time block coding based on channel mean feedback. *IEEE Transactions on Signal Processing*, 50(10):2599–2613, Oct. 2002.

[11] S. Zhou and G. B. Giannakis. Optimal transmitter eigen-beamforming and space time block coding based on channel correlations. *IEEE Transactions on Information Theory*, 49(7):1673–1690, July 2003.

[12] T. L. Marzetta. BLAST training: Estimating channel characteristics for high capacity space-time wireless. In *37th Annual Allerton Conference on Communication, Control and Computing*, pages 958–966, Sept. 1999.

[13] S. Serbetli and A. Yener. Transceiver optimization for multiuser MIMO systems. *IEEE Transactions on Signal Processing*, 52(1):214–226, Jan. 2004.

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

[15] D. N. C. Tse and S. V. Hanly. Linear multiuser receivers: Effective interference, effective bandwidth and user capacity. *IEEE Transactions on Information Theory*, 455(2):641–657, March 1999.

[16] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of MIMO channels. *IEEE Journal on Selected Areas of Communications*, 21(5):684–702, June 2003.