MMSE Transmitter Design for Correlated MIMO Systems with Imperfect Channel Estimates: Power Allocation Trade-offs

Semih Serbetli, *Member, IEEE,* and Aylin Yener, *Member, IEEE*

***Abstract*— We investigate the transmit precoder design problem for a multiple input multiple output (MIMO) link with correlated receive antennas, considering the effect of channel estimation. We work with the total mean squared error (MSE) as the performance measure, and develop transceiver structures considering the effect of channel estimation and the correlation of the MIMO link. The proposed transceiver structures are optimum in the sense of minimizing the total MSE and distributing the total MSE equally among the parallel data streams. Motivated by the substantial effect the channel estimation process can have on the system performance, we next investigate the problem of how the correlated MIMO link should distribute its total available power between power expended for channel estimation versus data transmission. The optimum power allocation problem between the training sequences for channel estimation and data transmission for the correlated MIMO link is shown to have a unique solution, that is different than the uncorrelated case. It is observed that the proposed transceiver achieves near-minimum MSE values via a relatively wide range of power allocation parameters. This is in contrast to the transceiver that is oblivious to the estimation errors when a more precise power allocation strategy is needed to achieve the best performance. Our results demonstrate that the correlation structure of the MIMO link has a profound effect on the performance, and that the transceiver optimization should be done by taking both the correlation and the channel estimation process into account.**

***Index Terms*— Correlated MIMO system, linear precoding, MMSE receiver, channel estimation, power allocation.**

I. INTRODUCTION

CONSIDERING the rapidly increasing demand for high data rate and reliable wireless communications, spectrally efficient transmission schemes are of great importance for next generation wireless systems. Recent studies indicate that using multiple antennas at the transmitter and receiver can dramatically improve the performance of wireless communication systems [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 [2]–[4].

Manuscript received April 15, 2004; revised June 27, 2004 and November 14, 2004; accepted January 12, 2005. The associate editor coordinating the review of this paper and approving it for publication was V. Lau. This paper was presented in part in IEEE International Conference on Communications, ICC'05, Seoul, Korea, May 2005.

S. Serbetli is with Philips Research, 5656 AA, Eindhoven, The Netherlands (e-mail: semih.serbetli@philips.com).

A. Yener is with Wireless Communications and Networking Laboratory (WCAN), Electrical Engineering Department, The Pennsylvania State University, University Park, PA 16802 USA (e-mail: yener@ee.psu.edu).

Digital Object Identifier 10.1109/TWC.2006.04243.

Performance of a MIMO system is highly dependent on the channel state information (CSI) 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 [5], or for space-time coding [2], [3]. The effect of receiver side channel estimation on such schemes is analyzed in [3], [6]–[8]. 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], [9]–[11]. These studies assume exact CSI at both transmitter and receiver side.

In order to be able to employ precoding at the transmitter, the MIMO channel has to be estimated at the receiver, and in turn should be fedback to the transmit side. Towards that end, the design of optimum training sequences for estimation of uncorrelated and correlated MIMO channels are studied in [6], [7], [12]. Also considered in detail, in references [13]–[18], are optimal transmission strategies with imperfect CSI at the transmitter side. We must note that all of these approaches optimize the transmission strategy for a given channel estimation process, and assume perfect CSI at the receiver side. In contrast, in this paper, we will consider the existence of errors in the channel estimates both at the receiver and at the transmitter.

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. Towards that end, optimum training sequences and power allocation among the training sequences and data transmission are found for BLAST transmission in [7]. Reference [7] considered a lower bound of the sum capacity as the performance metric, and assumed uncorrelated MIMO links. In contrast, in this paper, we will consider a more practical performance metric, MSE, and address the more general case of correlated links.

In this paper, we consider the uplink of a MIMO system with correlated receive antennas. The existence of correlation between the receive antennas is a likely scenario, for example, when fewer local scatterers exist near the base station antennas for the uplink. We will address the case where the receiver estimates the channel, and that both the transmitter and receiver have access to the same imperfect CSI. Throughout the paper, we will use the total mean squared error (MSE) as our performance metric, a metric that is widely used for multi-symbol transmission [9], [19]. The contribution of this paper is two fold:

• We investigate the joint effect of channel estimation process and the correlation among the receiver antennas on the design of the precoder and the decoder for the MIMO system, with the objective of minimizing the MSE, such that each symbol is transmitted with an equal MSE performance. We show that knowing that we have an estimate of the channel, we can design the precoder and decoder more robust to channel estimation errors. We derive the optimal precoder and decoder structures, and show that the optimal linear transceivers proposed in [9] are not optimal for correlated MIMO systems with channel estimates. We show that both the correlation, and the channel estimation process should be taken into account for transceiver optimization. We also show, using the optimum transceiver we propose, that the number of independent data streams that can be transmitted through the channel for a given MSE target, can be larger than the rank of the channel matrix, a bound suggested by earlier work [9].

• Motivated by the profound effect of the quality of channel estimation on the MSE metric, we consider the problem of optimum sharing of resources between the process of channel estimation and data transmission. We consider the total power as the limited resource, similar to [7], [20], but consider the more general problem of power allocation between the channel estimation and data transmission for the *correlated* MIMO link. We show that, given the coherence time of the channel, there is a unique solution to the optimum power allocation problem. We observe that, unlike the uncorrelated case addressed in [20], the optimum power allocation derived for minimizing the MSE, is different than the optimum power allocation found in [7] that considers a lower bound of the sum capacity as the performance metric. Therefore, we once again observe the correlations and the channel estimation process should be explicitly taken into account in identifying the power allocation trade-offs. The results also demonstrate that the system performance is more sensitive to the transceiver design rather than the power allocation. That is, with the "right" transceiver design, we can afford to be more flexible in distributing the total power between data transmission and channel estimation.

The organization of the paper is as follows. In Section II, the system model is described, and the performance metric 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 problem of optimum power allocation between training and data transmission is formulated and solved 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 receive 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 (Fig. 1). We assume that the number of data streams is given and fixed first, and analyze the effect of channel estimation on the design of linear transmitter (precoder) and receiver (decoder). 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 [9], the received vector is

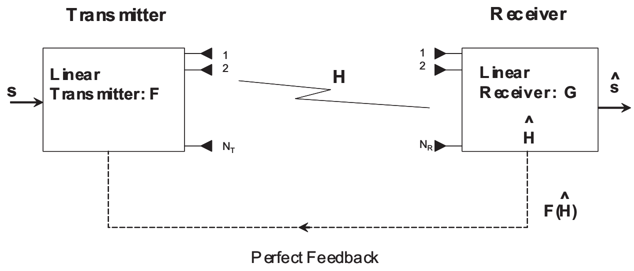

Fig. 1. System Model

$$\mathbf{r} = \mathbf{H}\mathbf{F}\mathbf{s} + \mathbf{n} \tag{1}$$

where, $\mathbf{s}$ is the $M \times 1$ symbol vector, $\mathbf{H}$ is a realization of the $N_R \times N_T$ random matrix of complex channel gains with arbitrarily correlated receive antennas, $\mathcal{H}$, with $E\{\mathcal{H}\mathcal{H}^\dagger\} = N_T\mathbf{C}_{RX}$, and uncorrelated transmit antennas, $E\{\mathcal{H}^\dagger\mathcal{H}\} = \gamma\mathbf{I}$, where $(\cdot)^\dagger$ denotes the hermitian of a vector or matrix. $\mathbf{n}$ is the zero mean complex Gaussian noise vector with $E\big[\mathbf{n}\mathbf{n}^\dagger\big] = \sigma^2\mathbf{I}$.

We assume that the channel is flat fading with coherence time of $(L_{tr} + L_d)$ symbols where L_{tr} symbol intervals are dedicated to training sequences, and the remaining L_d 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_d symbols. Thus, the precoder should be designed with the power constraint $tr\{\mathbf{F}\mathbf{F}^\dagger\} \le P_s = (P_{total} - P_{tr})/L_d$.

Reference [12] showed that orthogonal training sequences that require $L_{tr} \ge N_T$ are optimum for estimating MIMO channels with correlated receive antennas. Hence, throughout this paper, we will assume that $L_{tr} \ge N_T$ and orthogonal training sequences are used for the channel estimation process. Following the channel estimation model for MIMO systems in [6], [7], [12], when orthogonal training sequences are transmitted from each transmit antenna, i.e., $\mathbf{t}_i^\dagger\mathbf{t}_j = \delta_{ij}$, the received signal at the ith receive antenna is given by

$$\mathbf{r}_i = \sum_{j=1}^{N_T} \sqrt{P_{tr}/N_T} h_{ij}\mathbf{t}_j + \mathbf{n}_i \tag{2}$$

where h_{ij} is the channel gain from the jth transmit antenna to the ith receive antenna, and $\mathbf{n}_i$ is the zero mean complex Gaussian noise vector at the ith receive antenna with $E\Big[\mathbf{n}_i\mathbf{n}_i^\dagger\Big] = \sigma^2\mathbf{I}$. We obtain a noisy version of h_{ij}, $\hat{h}_{ij}$, simply by

$$\hat{h}_{ij} = \sqrt{N_T/P_{tr}}\mathbf{t}_j^\dagger\mathbf{r}_i = h_{ji} + \sqrt{N_T/P_{tr}}\mathbf{t}_j^\dagger\mathbf{n}_i = h_{ij} + x_{ij} \tag{3}$$

where x_{ij} is the zero mean complex Gaussian noise with $E\left[x_{ij}x_{ij}^{\dagger}\right] = \sigma_e^2 = N_T\sigma^2/P_{tr}$. Note that x_{ij}s are independent of h_{ij} and i.i.d. $\forall i,j$. Thus, when orthogonal training sequences are used, the receiver observes a noisy version of the MIMO channel $\hat{\mathcal{H}} = \mathcal{H} + \mathcal{X}$. $\hat{\mathcal{H}}$ is an $N_R \times N_T$ matrix whose (i,j)th entry is $\hat{h}_{ij}$. $\mathcal{X}$ 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 $\mathcal{H}$. $\hat{\mathcal{H}}$ will be utilized for designing the linear precoder which is fedback to the transmitter via an error-free and low-delay feedback channel. Recall that $\mathcal{H}$ is a complex Gaussian random matrix with $E\{\mathcal{H}\mathcal{H}^{\dagger}\} = N_T\mathbf{C}_{RX} = N_T\mathbf{U}_{RX}\Lambda\mathbf{U}_{RX}^{\dagger}$. Hence, the channel estimate can be represented as

$$\hat{\mathcal{H}} = \mathbf{U}_{RX}\Lambda^{1/2}\mathcal{H}_{\mathcal{W}} + \mathcal{X} \tag{4}$$

where the elements of $\mathcal{H}_{\mathcal{W}}$ are i.i.d. with $CN(0,1)$. Using simple algebra and the Bayes' rule, it is easily seen that the distribution of the ith column of $\mathcal{H}$, $\mathcal{H}_i$, for a given estimate $\hat{\mathcal{H}} = \hat{\mathbf{H}}$ is $CN((\mathbf{H}_\mu)_i, \mathbf{C}_{\mathcal{H}|\hat{\mathbf{H}}})$ with

$$E\{\mathcal{H}_i|\hat{\mathcal{H}} = \hat{\mathbf{H}}\} = (\mathbf{H}_\mu)_i = \mathbf{U}_{RX}\Lambda(\Lambda+\sigma_e^2\mathbf{I})^{-1}\mathbf{U}_{RX}^{\dagger}\hat{\mathbf{H}}_i \tag{5}$$

$$E\{\mathcal{H}_i\mathcal{H}_i^{\dagger}|\hat{\mathcal{H}} = \hat{\mathbf{H}}\} = \mathbf{C}_{\mathcal{H}|\hat{\mathbf{H}}} = \sigma_e^2\mathbf{U}_{RX}\Lambda(\Lambda + \sigma_e^2\mathbf{I})^{-1}\mathbf{U}_{RX}^{\dagger} \tag{6}$$

where $\hat{\mathbf{H}}_i$ represents the ith column of $\hat{\mathbf{H}}$. Observe that from the receiver's perspective, the actual channel is a random MIMO channel with mean $\mathbf{H}_\mu = [(\mathbf{H}_\mu)_1, (\mathbf{H}_\mu)_2 ... (\mathbf{H}_\mu)_{N_T}]$ and; $E\{\mathcal{H}\mathcal{H}^{\dagger}|\hat{\mathcal{H}} = \hat{\mathbf{H}}\} = N_T\mathbf{C}_{\mathcal{H}|\hat{\mathbf{H}}}$ and $E\{\mathcal{H}^{\dagger}\mathcal{H}|\hat{\mathcal{H}} = \hat{\mathbf{H}}\} = \upsilon\mathbf{I}$.

Let us denote the $M \times N_R$ linear receiver by $\mathbf{G}$; the decision statistic $\mathbf{y}$, for a channel realization $\mathbf{H}$, is then given by

$$\mathbf{y} = \mathbf{GHFs} + \mathbf{Gn} \tag{7}$$

For a given channel estimate, $\hat{\mathbf{H}}$, the total MSE can be expressed as

$$\begin{aligned}\text{MSE}_{\hat{\mathbf{H}}} &= E\left[\|\mathbf{y} - \mathbf{s}\|^2\right] \\ &= E\left[\text{tr}\left\{\mathbf{F}^{\dagger}\mathcal{H}^{\dagger}\mathbf{G}^{\dagger}\mathbf{G}\mathcal{H}\mathbf{F} - \mathbf{F}^{\dagger}\mathcal{H}^{\dagger}\mathbf{G}^{\dagger} - \mathbf{G}\mathcal{H}\mathbf{F} + \mathbf{I} + \sigma^2\mathbf{G}\mathbf{G}^{\dagger}\right\}\right]\end{aligned} \tag{8}$$

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

Total MSE minimization by choosing the transmitters and receivers has been studied for uncorrelated MIMO links with exact channel state information in [9]. In practice, the channel state information available to the transmitter and receiver would not be perfect. In Section III, we pose the problem of minimizing the total MSE for a given noisy observation of the correlated MIMO channel, $\text{MSE}_{\hat{\mathbf{H}}}$, considering equal MSE among the parallel data streams, and construct the optimum transceiver structure.

Note that when the optimum linear transmitter and receiver are used at each realization, the total MSE over all channel realizations and estimates, MSE, can be expressed as

$$\text{MSE} = E\left[\min_{\{\mathbf{F},\mathbf{G}\}} \text{MSE}_{\hat{\mathbf{H}}}\right] \tag{9}$$

We will use MSE in (9) as the performance metric in Section V.

III. Transmitter Optimization with Channel Estimation

Our aim in this section is to find the MSE minimizing transceiver structure that takes the effect of channel estimation into account. We also impose the constraint of achieving equal MSE among the parallel data streams. The optimization problem is

$$\begin{aligned}&\min_{\{\mathbf{F},\mathbf{G}\}} \quad \text{MSE}_{\hat{\mathbf{H}}} = \text{tr}\{\mathbf{B}\} \\ &\text{s.t. tr}\{\mathbf{F}\mathbf{F}^{\dagger}\} \le P_s; \quad \text{MSE}_1 = \text{MSE}_2 = ... = \text{MSE}_M\end{aligned} \tag{10}$$

where

$$\mathbf{B} = E\left[\mathbf{F}^{\dagger}\mathcal{H}^{\dagger}\mathbf{G}^{\dagger}\mathbf{G}\mathcal{H}\mathbf{F} - \mathbf{F}^{\dagger}\mathcal{H}^{\dagger}\mathbf{G}^{\dagger} - \mathbf{G}\mathcal{H}\mathbf{F} + \mathbf{I} + \sigma^2\mathbf{G}\mathbf{G}^{\dagger}\right],$$

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

As 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. Using the distribution derived in the previous section, $\mathcal{H}$ can be modelled as $\mathcal{H} = \mathbf{H}_\mu + \mathbf{C}_{\mathcal{H}|\hat{\mathbf{H}}}^{1/2}\mathcal{Z}$ where $\mathcal{Z}$ is a random matrix with i.i.d. complex Gaussian entries of $CN(0,1)$. Reformulating the total MSE in terms of $\mathbf{H}_\mu$, $\mathbf{C}_{\mathcal{H}|\hat{\mathbf{H}}}$, and $\mathcal{Z}$ we have

$$\begin{aligned}\text{MSE}_{\hat{\mathbf{H}}} = E[\text{tr}\{&\mathbf{F}^{\dagger}\mathbf{H}_\mu^{\dagger}\mathbf{G}^{\dagger}\mathbf{G}\mathbf{H}_\mu\mathbf{F} + \mathbf{F}^{\dagger}\mathcal{Z}^{\dagger}\mathbf{C}_{\mathcal{H}|\hat{\mathbf{H}}}^{1/2}\mathbf{G}^{\dagger}\mathbf{G}\mathbf{C}_{\mathcal{H}|\hat{\mathbf{H}}}^{1/2}\mathcal{Z}\mathbf{F} \\ &+ \mathbf{F}^{\dagger}\mathcal{Z}^{\dagger}\mathbf{C}_{\mathcal{H}|\hat{\mathbf{H}}}^{1/2}\mathbf{G}^{\dagger}\mathbf{G}\mathbf{H}_\mu\mathbf{F} + \mathbf{F}^{\dagger}\mathbf{H}_\mu^{\dagger}\mathbf{G}^{\dagger}\mathbf{G}\mathbf{C}_{\mathcal{H}|\hat{\mathbf{H}}}^{1/2}\mathcal{Z}\mathbf{F} - \mathbf{F}^{\dagger}\mathbf{H}_\mu^{\dagger}\mathbf{G}^{\dagger} \\ &- \mathbf{F}^{\dagger}\mathcal{Z}^{\dagger}\mathbf{C}_{\mathcal{H}|\hat{\mathbf{H}}}^{1/2}\mathbf{G}^{\dagger} - \mathbf{G}\mathbf{H}_\mu\mathbf{F} - \mathbf{G}\mathbf{C}_{\mathcal{H}|\hat{\mathbf{H}}}^{1/2}\mathcal{Z}\mathbf{F} + \mathbf{I} + \sigma^2\mathbf{G}\mathbf{G}^{\dagger}\}]\end{aligned} \tag{11}$$

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

$$\begin{aligned}\text{MSE}_{\hat{\mathbf{H}}} = \text{tr}\{&\mathbf{F}^{\dagger}\mathbf{H}_\mu^{\dagger}\mathbf{G}^{\dagger}\mathbf{G}\mathbf{H}_\mu\mathbf{F} - \mathbf{F}^{\dagger}{\mathbf{H}_\mu}^{\dagger}\mathbf{G}^{\dagger} - \mathbf{G}\mathbf{H}_\mu\mathbf{F} \\ &+ \mathbf{I} + \mathbf{G}(\sigma^2\mathbf{I} + \text{tr}\{\mathbf{F}\mathbf{F}^{\dagger}\}\mathbf{C}_{\mathcal{H}|\hat{\mathbf{H}}})\mathbf{G}^{\dagger}\}\end{aligned} \tag{12}$$

Observe that the total MSE in (12) has the same form of the total MSE expression of a MIMO system with a channel matrix $\mathbf{H}_\mu$ and a colored noise factor with covariance $\mathbf{C}_{colored} = \sigma^2\mathbf{I} + \text{tr}\{\mathbf{F}\mathbf{F}^{\dagger}\}\mathbf{C}_{\mathcal{H}|\hat{\mathbf{H}}}$.

Let us now consider the minimization of the total 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

$$\mathbf{G} = \mathbf{F}^{\dagger}\mathbf{H}_\mu^{\dagger}\left(\mathbf{C}_{colored} + \mathbf{H}_\mu\mathbf{F}\mathbf{F}^{\dagger}{\mathbf{H}_\mu}^{\dagger}\right)^{-1} \tag{13}$$

Using (13), the total MSE function can be reformulated as

$$\begin{aligned}\text{MSE}_{\hat{\mathbf{H}}} &= \text{tr}\{\mathbf{I} - \mathbf{F}^{\dagger}{\mathbf{H}_\mu}^{\dagger}\left(\mathbf{C}_{colored} + \mathbf{H}_\mu\mathbf{F}\mathbf{F}^{\dagger}{\mathbf{H}_\mu}^{\dagger}\right)^{-1}\mathbf{H}_\mu\mathbf{F}\} \\ &= M - N_R + \text{tr}\{\mathbf{T}^{-1}\}\end{aligned} \tag{14}$$

where $\mathbf{T} = \mathbf{I} + \mathbf{H}_e\mathbf{F}\mathbf{F}^{\dagger}{\mathbf{H}_e}^{\dagger}$ with $\mathbf{H}_e = \mathbf{C}_{colored}^{-1/2}\mathbf{H}_\mu$. Notice that the linear transmitter and receiver pair that minimizes the total 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, any $\{\mathbf{F}_{opt}^{\star}\} = \mathbf{F}_{opt}\mathbf{U}^{\dagger}$ where $\mathbf{U}$ is an arbitrary $M \times M$ matrix satisfying $\mathbf{U}^{\dagger}\mathbf{U} = \mathbf{I}$, also attains the minimum MSE if $\mathbf{F}_{opt}$ does.

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

Defining the transmitter covariance matrix $\mathbf{R} = \mathbf{F}\mathbf{F}^{\dagger}$, the total MSE minimization problem without fairness constraints can be restated as

$$\begin{aligned} \min_{\mathbf{R}} \quad & \mathrm{MSE}_{\hat{\mathbf{H}}} = M - N_R + \mathrm{tr}\Big\{(\mathbf{I} + \mathbf{H}_e\mathbf{R}\mathbf{H}_e{}^{\dagger})^{-1}\Big\} \\ \text{s.t.} \quad & \mathrm{tr}\{\mathbf{R}\} \le P_s; \quad \mathbf{R} \ge 0; \qquad \mathrm{rank}(\mathbf{R}) \le M \end{aligned} \tag{15}$$

where $\mathbf{A} \ge 0$ refers to the positive semi-definiteness constraint on $\mathbf{A}$.

For $M \le \mathrm{rank}(\mathbf{H}_e)$, reference [9] suggests that one possible optimum linear transmitter is in the form of $\mathbf{F}_{opt} = \mathbf{V}_e\mathbf{Q}_f$ where $\mathbf{V}_e$ is an $N_T \times M$ orthogonal matrix that has columns as the eigenvectors of the largest M eigenvalues of $\mathbf{H}_e^{\dagger}\mathbf{H}_e = \begin{bmatrix}\mathbf{V}_e & \tilde{\mathbf{V}}_e\end{bmatrix}\begin{bmatrix}\Lambda_e & 0 \\ 0 & \tilde{\Lambda}_e\end{bmatrix}\begin{bmatrix}\mathbf{V}_e^{\dagger} \\ \tilde{\mathbf{V}}_e^{\dagger}\end{bmatrix}$ where Λ_e is a diagonal matrix containing the largest M eigenvalues arranged in a decreasing order from top-left to bottom-right, and $\mathbf{Q}_f = (\mu^{-1/2}\Lambda_e^{-1/2} - \Lambda_e^{-1})_+^{1/2}$ is a diagonal matrix with μ factor satisfying the power constraint, and $(.)_+ = max(0,.)$. If $\hat{\mathbf{H}}$ were the perfect CSI, then the optimum linear transmitter would be in the form of $\hat{\mathbf{F}}_{opt} = \hat{\mathbf{V}}_e\hat{\mathbf{Q}}_f$ with $\hat{\mathbf{Q}}_f = (\hat{\mu}^{-1/2}\hat{\Lambda}_e^{-1/2} - \hat{\Lambda}_e^{-1})_+^{1/2}$. $\hat{\mathbf{V}}_e$, $\hat{\Lambda}_e$, and $\hat{\mathbf{Q}}_f$ are the corresponding eigenvectors, eigenvalues and power allocation for $1/\sigma^2\hat{\mathbf{H}}^{\dagger}\hat{\mathbf{H}}$, respectively. Observe that, in this case, the resulting eigen modes and their corresponding power allocation would differ from the solution of (15). It is important to note that optimum linear transmitter uses only largest M eigen modes is due to the rank constraint in (16), and rank of the optimum $\mathbf{R}$ can not be larger than $\mathrm{rank}(\mathbf{H}_e)$. Notice that when $M \ge \mathrm{rank}(\mathbf{H}_e)$, the rank constraint in (16) is redundant for the optimization problem since $\mathrm{rank}(\mathbf{H}_e) \ge \mathrm{rank}(\mathbf{R})$, and optimum transmission uses all eigen modes of the channel. Thus, the optimum transmit covariance matrix for $M = \mathrm{rank}(\mathbf{H}_e)$ is also the optimum covariance matrix for $M > \mathrm{rank}(\mathbf{H}_e)$. Specifically, all optimum linear transmitters are in the form of $\mathbf{F}_{opt} = \mathbf{V}_e\mathbf{Q}_f\mathbf{U}^{\dagger}$ where $\mathbf{U}^{\dagger}\mathbf{U} = \mathbf{I}$. Thus, the total MSE of each data stream can be expressed by the diagonal entries of

$$\begin{aligned} \mathbf{B} &= \mathbf{I} - \mathbf{F}^{\dagger}\mathbf{H}_{\mu}{}^{\dagger}\left(\mathbf{C}_{colored} + \mathbf{H}_{\mu}\mathbf{F}\mathbf{F}^{\dagger}\mathbf{H}_{\mu}^{\dagger}\right)^{-1}\mathbf{H}_{\mu}\mathbf{F} \\ &= \mathbf{U}\mathbf{D}\mathbf{U}^{\dagger} \end{aligned} \tag{16}$$

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

Now, consider a MIMO system where equal MSE values for each symbol is required. 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. Reference [21] suggests that the discrete Fourier transform (DFT) matrix or the Hadamard matrix (when M is a power of 2) provides a simple construction of this special matrix. The same construction is also used in [18], [19]. Such a matrix with real entries can also be obtained using the majorization theory: the eigenvalues of $\mathbf{B}$ always majorize the equal diagonal entries which ensures the existence of such a matrix $\mathbf{U}$ [22]. An arbitrary distribution of the total MSE among the parallel data streams is possible if the eigenvalues of $\mathbf{B}$ majorize the diagonal entries, the MSE targets of each data stream [22]. Notice that $\mathbf{U}$ is an $M \times M$ unitary matrix for $M \le \mathrm{rank}(\mathbf{H}_e)$, and an $M \times \mathrm{rank}(\mathbf{H}_e)$ orthogonal matrix for $M > \mathrm{rank}(\mathbf{H}_e)$.

As a final note, consider the special case where the receive antennas are uncorrelated, $E\{\mathcal{H}\mathcal{H}^{\dagger}\} = N_T\sigma_H^2\mathbf{I}$. In this case, the conditional channel mean becomes a scaled version of the channel estimate, $\mathbf{H}_{\mu} = \rho\hat{\mathbf{H}} = \frac{\sigma_H^2}{\sigma_H^2+\sigma_e^2}\hat{\mathbf{H}}$ and the linear transceiver structure treats the residual channel estimation error as AWGN resulting a total AWGN with a variance $\rho_1 = \sigma^2 + \rho\sigma_e^2\mathrm{tr}\{\mathbf{F}\mathbf{F}^{\dagger}\}$ [20]. This property results $\mathbf{F}_{opt}$, the optimum linear transmitter, and $\hat{\mathbf{F}}_{opt}$, the optimum transmitter corresponding to channel $\hat{\mathbf{H}}$, to transmit in the same eigen modes, but with different power allocation among the eigen modes. $\mathbf{F}_{opt}$ is exactly the same as $\hat{\mathbf{F}}_{opt}$ only when $\mathrm{rank}(\hat{\mathbf{F}}_{opt}) = 1$, i.e., only when the optimum transmit precoder is a beamformer. Note that $\mathrm{rank}(\hat{\mathbf{F}}_{opt}) = 1$ when $P_s \le \frac{\hat{\lambda}_2^{-1/2} - \hat{\lambda}_1^{-1/2}}{\hat{\lambda}_1^{1/2}}$ where $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are the two largest eigenvalues of $\frac{1}{\sigma^2}\hat{\mathbf{H}}^{\dagger}\hat{\mathbf{H}}$. When $\mathrm{rank}(\hat{\mathbf{F}}_{opt}) = 1$, $\mathbf{F}_{opt}$ is guaranteed to be rank 1 due to the fact that the two largest eigenvalues of $\frac{\rho^2}{\rho_1}\hat{\mathbf{H}}^{\dagger}\hat{\mathbf{H}}$, λ_1 and λ_2 are scaled versions of $\hat{\lambda}_1$ and $\hat{\lambda}_2$ with $\frac{\lambda_1}{\hat{\lambda}_1} = \frac{\lambda_2}{\hat{\lambda}_2} = \frac{\sigma^2\rho^2}{\rho_1} < 1$, and $P_s \le \frac{\hat{\lambda}_2^{-1/2} - \hat{\lambda}_1^{-1/2}}{\hat{\lambda}_1^{1/2}} \le \frac{\rho_1}{\sigma^2\rho^2}\frac{\lambda_2^{-1/2} - \lambda_1^{-1/2}}{\lambda_1^{1/2}}$. Notice that $\frac{\rho^2}{\rho_1}$ is inversely proportional to the variance of the channel estimation error, σ_e^2. As the quality of the channel estimate decreases, σ_e^2 increases and, the two largest eigenvalues of $\frac{\rho^2}{\rho_1}\hat{\mathbf{H}}^{\dagger}\hat{\mathbf{H}}$ decrease. In other words, the effective SNRs of each eigen mode decrease. Thus, the probability of the optimum linear transmitter to be a beamformer increases when σ_e^2 increases, as in the case of a MIMO system with perfect CSI operating in low SNR.

IV. Rate Allocation

The previous section 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

$$\begin{aligned} &\max_{\{\mathbf{F},\mathbf{G}\}} \qquad \mathrm{M} \\ &\text{s.t. } \mathrm{tr}\{\mathbf{F}\mathbf{F}^{\dagger}\} \le P_s; \quad \mathrm{MSE}_1 = \mathrm{MSE}_2 = \ldots = \mathrm{MSE}_M \le \beta \end{aligned} \tag{17}$$

where β is the MSE target of each symbol with a range of $[0, 1]$ when MMSE receivers are used.

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:

$$\mathrm{MSE}_{\hat{\mathbf{H}}} = \sum_{i=1}^{M}\mathrm{MSE}_i \le M\beta \tag{18}$$

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

$$M \leq \frac{N_R - \text{tr}(\mathbf{T}_{opt}^{-1})}{1-\beta} \tag{19}$$

where $\mathbf{T}_{opt} = \mathbf{I} + \mathbf{H}_e \mathbf{F}_{opt} \mathbf{F}_{opt}{}^{\dagger} \mathbf{H}_e^{\dagger}$ with $\mathbf{F}_{opt} = \mathbf{V}_e \mathbf{Q}_f$. Recall that the optimum linear transmit covariance matrix $\mathbf{R}_{opt} = \mathbf{F}_{opt}\mathbf{F}_{opt}^{\dagger}$ can be found for $M \leq \text{rank}(\mathbf{H}_e)$ by using the appropriate $\mathbf{Q}_f$, and remains the same for $M \geq \text{rank}(\mathbf{H}_e)$. Thus, $\mathbf{T}_{opt}$ is *independent* of M for $M \geq \text{rank}(\mathbf{H}_e)$, and the right hand side of (19) forms an upper bound for the number data streams to be transmitted with an MSE target. Observe that this number can exceed the rank of $\mathbf{H}_e$ depending on the value of β. As β gets larger, that is if a larger MSE can be tolerated, the number of parallel data streams can be much higher than the rank of the channel matrix, an upper bound suggested in [9]. In essence, when the optimum transceiver is used we can increase the number of data streams to be transmitted at the expense of higher error rates. Such transmission schemes may be favorable for multimedia applications that require high data rates and have good error correction capabilities.

Lastly, we note that the right hand side of (19) is achievable via the appropriate choice of the precoder-decoder pair as explained in Section III, and thus serves as a feasibility constraint for the MIMO system. At the same time, another feasibility constraint can be obtained by observing the mathematical equivalence of the MIMO system to a CDMA system with certain channel and power constraints. In this case, the feasibility condition turns out to be $\sum_{i=1}^{M} \frac{SIR_i}{1+SIR_i} \leq N_R$ where SIR_i is the SIR target of user i [23]. 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 $M \leq \frac{N_R}{1-\beta}$ for the CDMA system. Observe however, that the attainable bound in (19) is tighter than $\frac{N_R}{1-\beta}$ 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, and corresponding $\mathbf{F}_{opt}$.

V. Power Allocation Trade-offs

A. *Optimum Power Allocation*

It is evident from the preceding discussion in this paper, as well as several other references, e.g., [6]–[8], [13], [14], 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. Reference [7] investigated this problem using a lower bound on the channel capacity as the performance metric. Reference [7] also assumed an uncorrelated MIMO link.

In this section, we investigate the optimum power allocation problem between training and data transmission using total MSE as the performance metric. That is, we consider a different performance metric than that of [7], and address the case of correlated receive antennas.

Recall that for a given channel estimate, minimizing the total MSE, $\text{MSE}_{\hat{\mathbf{H}}}$, is equivalent to minimizing $\text{tr}\{\left(\mathbf{I} + \mathbf{C}_{colored}^{-1/2}\mathbf{H}_{\mu}\mathbf{F}\mathbf{F}^{\dagger}\mathbf{H}_{\mu}{}^{\dagger}\mathbf{C}_{colored}^{-1/2}\right)^{-1}\}$ where $\mathbf{H}_{\mu}$ is a realization of the random matrix $\tilde{\mathcal{H}} = \mathbf{U}_{RX}\Lambda(\Lambda + \sigma_e^2\mathbf{I})^{-1}\mathbf{U}_{RX}^{\dagger}\hat{\mathcal{H}}$ with $\hat{\mathcal{H}} = \mathbf{U}_{RX}\Lambda^{1/2}\mathcal{H}_W + \mathcal{X}$. Thus, the total MSE, given that we use the optimum precoder, over all channel realizations and estimates is

$$\text{MSE} = M - N_R + E\left[\min_{\mathbf{F}} \text{tr}\{(\mathbf{I} + \mathbf{C}_{colored}^{-1/2}\tilde{\mathcal{H}}\mathbf{F}\mathbf{F}^{\dagger}\tilde{\mathcal{H}}^{\dagger}\mathbf{C}_{colored}^{-1/2})^{-1}\}\right] \tag{20}$$

We can express $\hat{\mathcal{H}} = \mathbf{U}_{RX}(\Lambda + \sigma_e^2\mathbf{I})^{1/2}\Upsilon$ where Υ is a random matrix with complex Gaussian entries with $CN(0,1)$ and $\text{tr}\{\mathbf{F}\mathbf{F}^{\dagger}\} = P_s$. Inserting $\mathbf{C}_{colored} = \sigma^2\mathbf{I} + \text{tr}\{\mathbf{F}\mathbf{F}^{\dagger}\}\mathbf{C}_{\mathcal{H}|\hat{\mathbf{H}}}$, and the expressions for $\tilde{\mathcal{H}}$ and $\hat{\mathcal{H}}$ given above in (20), and normalizing the precoder matrices to unit trace matrices

$$\tilde{\mathbf{F}} = \sqrt{1/P_s}\mathbf{F} \Rightarrow \text{tr}\{\tilde{\mathbf{F}}\tilde{\mathbf{F}}^{\dagger}\} = 1 \tag{21}$$

we can show that

$$\min \text{MSE} \equiv \min E\left[\min_{\tilde{\mathbf{F}}} \text{tr}\{(\mathbf{I} + \mathbf{U}_{RX}\Delta^{1/2}\Upsilon\tilde{\mathbf{F}}\tilde{\mathbf{F}}^{\dagger}\Upsilon^{\dagger}\Delta^{1/2}\mathbf{U}_{RX}^{\dagger})^{-1}\}\right] \tag{22}$$

$$\text{where } \Delta = P_s\Lambda^2(\sigma^2(\Lambda + \sigma_e^2\mathbf{I}) + P_s\sigma_e^2\Lambda)^{-1} \tag{23}$$

Δ is a diagonal matrix with $\Delta_{ii} = \frac{\Lambda_i^2 P_s}{\sigma^2(\Lambda_i + \sigma_e^2) + \sigma_e^2\Lambda_i P_s}$ as the ith diagonal entry. Notice that $\mathbf{U}_{RX}$ does not have an effect on the total MSE. Thus, the expressions $\{\Delta_{ii}\}$ act as the *effective SNRs of each virtual receive antenna*.

Recall that we have the following relationship between data transmission power and power dedicated to the training sequences

$$P_s L_d + P_{tr} = P_{total} \tag{24}$$

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

$$\alpha = \frac{P_s L_d}{P_{total}} \quad 0 \leq \alpha \leq 1 \tag{25}$$

and

$$c_i = \frac{(N_T - L_d)P_{total}\Lambda_i}{L_d N_T \sigma^2 + L_d P_{total}\Lambda_i} \tag{26}$$

$$d_i = \frac{P_{total}^2 {\Lambda_i}^2}{\sigma^2 L_d(\Lambda_i P_{total} + N_T\sigma^2)} \tag{27}$$

the effective SNR of each virtual receive antenna can be expressed as

$$\Delta_{ii}(\alpha) = \frac{d_i\alpha(1-\alpha)}{c_i\alpha + 1} \tag{28}$$

Note that changing α corresponds to changing power dedicated to the training sequences, P_{tr} and the power dedicated to the transmission of the actual data at each symbol interval, P_s resulting different SNRs for each virtual receive antennas. Using (28), the optimum power allocation problem can be expressed as

$$\min_{0<\alpha<1} \quad \text{MSE} \equiv$$
$$\min_{0<\alpha<1} E\left[\min_{\tilde{\mathbf{F}}} \text{tr}\{(\mathbf{I} + \Delta^{1/2}(\alpha)\Upsilon\tilde{\mathbf{F}}\tilde{\mathbf{F}}^\dagger\Upsilon^\dagger\Delta^{1/2}(\alpha))^{-1}\}\right] \quad (29)$$

Let us define the set of diagonal SNR matrices, Ψ, such that $\Omega \in \Psi$ if $\Omega \preceq \Delta(\alpha)$ [1] for some $\alpha \in [0, 1]$. Then, optimum power allocation problem can be restated as

$$\min_{0<\alpha<1} \quad \text{MSE} \equiv$$
$$\min_{\Omega\in\Psi} E\left[\min_{\tilde{\mathbf{F}}} \text{tr}\{(\mathbf{I} + \Omega^{1/2}\Upsilon\tilde{\mathbf{F}}\tilde{\mathbf{F}}^\dagger\Upsilon^\dagger\Omega^{1/2})^{-1}\}\right] \quad (30)$$

It can be easily seen that the total MSE function is strictly convex over the SNR matrix, Ω. In addition, the convexity of the set Ψ is established by Lemma 2 which follows from Lemma 1.

Lemma 1: $\Delta_{ii}(\alpha)$ is concave over the interval $[0, 1]$, and the optimum α maximizing $\Delta_{ii}(\alpha)$ is given by

$$\alpha_i^\star = \begin{cases} \frac{-1+\sqrt{1+c_i}}{c_i}, & \text{for } N_T > L_d; \\ \frac{1}{2}, & \text{for } N_T = L_d; \\ \frac{-1+\sqrt{1+c_i}}{c_i}, & \text{for } N_T < L_d; \end{cases} \quad (31)$$

where c_i is given in (26).

Proof: See Appendix. □

Lemma 2: Ψ is a convex set. □

Proof: See Appendix. □

Next, we have the following theorem.

Theorem 1: The global minimizer α_{opt} of (29) is unique and $\alpha_{opt} \in [\alpha_{min}, \alpha_{max}]$.

Proof: It follows from above that (30) is a convex program, with a convex cost function and convex constraint set [24]. Hence a unique solution exists that is the global minimizer α. Notice that the effective SNR of each virtual receive antenna, $\Delta_{ii}(\alpha)$ is maximized by different $\alpha_i^\star$ values. However, as the effective SNRs of all receive antennas are increased, the total MSE function is guaranteed to decrease. Since $\Delta_{ii}(\alpha)$ is a concave function, an interval where all effective SNRs increase exists. Denote $\alpha_{min} = \min_i \alpha_i^\star$ and $\alpha_{max} = \max_i \alpha_i^\star$. In the intervals $[0, \alpha_{min}]$ and $[\alpha_{max}, 1]$, all effective SNRs of the virtual receive antennas have the same monotonic behavior: monotonically increasing in $[0, \alpha_{min}]$ and decreasing in $[\alpha_{max}, 1]$. Thus, the optimum power allocation, α_{opt} is guaranteed to lie in the interval $[\alpha_{min}, \alpha_{max}]$. □

The above theorem guarantees the existence of the unique global minimizer. The value that lies in $[\alpha_{min}, \alpha_{max}]$ can be easily found by any convex optimization method, e.g. iterative algorithms [24]. In this context, first, the optimum α that will allocate the total available transmit power among the training sequences and the actual data transmission should be found. Next, using the optimum P_{tr} that corresponds to the optimum α, orthogonal training sequences should be transmitted to obtain a channel estimate. Using the channel estimate, the channel estimation error statistics and the correlation among the receive antennas, the optimum precoder and decoder pair should be designed using the optimum P_s.

[1] $\mathbf{A} \preceq \mathbf{B}$ implies that $\mathbf{B} - \mathbf{A}$ is a positive semi-definite matrix.

B. Observations

The value of α_{opt} depends on the number of transmit 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, i.e., $N_T > L_d$, then all c_i's are positive resulting in all $\alpha_i^\star$ to be less than $1/2$. Thus, α_{opt} lies in the range of $[0, \frac{1}{2})$. This result suggests allocating more power to training for such systems with large number of transmit antennas.

When the number of symbols to be transmitted is greater than the number of transmit antennas, i.e., $N_T < L_d$, then the range of $\alpha_i^\star$ is $(\frac{1}{2}, 1]$ resulting α_{opt} to be in $(\frac{1}{2}, 1]$. This implies that when the data transmission interval is much larger than the number of transmit antennas, a significant portion of the system power should be allocated to symbol transmission rather than the estimation process.

For the case where the number of transmit antennas is equal to the number of symbols to be transmitted, i.e., $N_T = L_d$, all α_i's are $1/2$ resulting $\alpha_{opt} = 1/2$. That is, the available power should be allocated equally between training and data transmission.

For the special case when the receivers are uncorrelated, the effective SNR of each receive antenna is the same [20]. Thus, the optimum power allocation minimizing the total MSE can be found by using Lemma 1 directly, and corresponding optimal power allocation and effective SNRs are:

1) $L_d = N_T$:

$$\alpha = \frac{1}{2} \Rightarrow P_{tr} = L_d P_s = \frac{P_{total}}{2} \quad (32)$$
$$\rho_e = \frac{P_{total}^2 {\sigma_H}^4}{4\sigma^2 L_d({\sigma_H}^2 P_{total} + N_T\sigma^2)} \quad (33)$$

2) $N_T > L_d$ and $N_T < L_d$:

$$\alpha = \frac{-1 + \sqrt{1 + c_i}}{c_i} \quad (34)$$
$$\rho_e = \frac{P_{total}^2 {\sigma_H}^4}{\sigma^2 L_d({\sigma_H}^2 P_{total} + N_T\sigma^2)} \left(\frac{\sqrt{1 + c_i} - 1}{c_i}\right)^2 \quad (35)$$

When, we consider the optimum power allocation for high and low SNR cases, the following power distribution schemes and effective SNRs are observed:

1) For high SNR, i.e., $P_{total} \to \infty$, we have:

$$c_i = \frac{N_T - L_d}{L_d} \quad \forall i = 1, 2, ..., N_R \quad (36)$$

resulting the optimum power allocation and effective SNR

$$\alpha_{opt} = \frac{\sqrt{L_d}}{\sqrt{N_T} + \sqrt{L_d}}, \Delta_{ii} = \frac{P_{total}\Lambda_i}{\sigma^2}\frac{1}{(\sqrt{N_T} + \sqrt{L_d})^2} \quad (37)$$

2) For low SNR, i.e., $P_{total} \to 0$, we have:

$$c_i = 0 \quad \forall i = 1, 2, ..., N_R \quad (38)$$

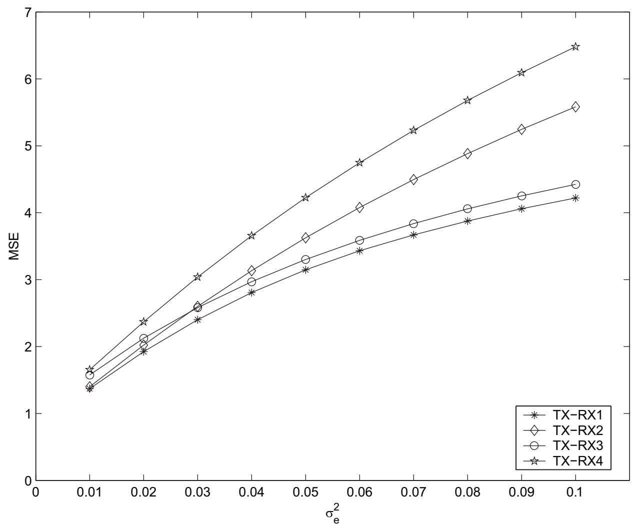

Fig. 2. Comparison of MSE vs σ_e^2 performance for 8×8 MIMO system with $P_s = 8$ and $L_d = 8$

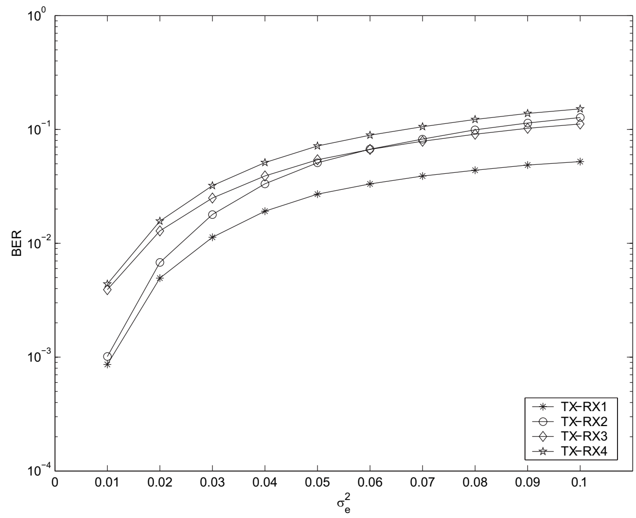

Fig. 3. Comparison of BER vs σ_e^2 performance for 8×8 MIMO system with $P_s = 8$ and $L_d = 8$

resulting the optimum power allocation and effective SNR

$$\alpha_{opt} = \frac{1}{2}, \qquad \rho_e = \frac{P_{total}^2 {\Lambda_i}^2}{4\sigma^4 N_T L_d} \tag{39}$$

At high SNR, the power is allocated according to the parameters, N_T and L_d, and the effective SNR is linear in the total power P_{total}. At low SNR, the power is distributed equally among the training sequences and the symbol transmission, and the effective SNR is quadratic in terms of total power P_{total}. We note that the results of the optimum power allocation for linear transceiver structures turn out to be identical to the optimum power allocation with capacity as the performance metric that are studied in [7] *when the receive antennas are uncorrelated*. This is due to the fact in the uncorrelated case, both problems reduce to maximizing the same effective SNR term [20]. When the transmit precoder structure is perfectly known at the receiver, and the MMSE receivers that consider the channel estimation error statistics is used at the receiver side, the effective SNR of the system, and hence, the optimum power allocation remains the same as the capacity maximizing case [20].

Lastly, observe that the optimum power allocation for both high and low SNR regimes does not depend on the correlation of the receive antennas, resulting in the same asymptotic optimal power allocation as the uncorrelated case.

VI. Numerical Results

In this section, we present numerical results related to the performance of the proposed transceiver structures and optimum power allocation for channel estimation and data transmission. The simulations are performed for a MIMO system where both the transmitter and the receiver is equipped with $N_T = N_R = 8$ antennas. The channel values are generated as realizations of a random matrix with complex Gaussian entries of $CN(0,1)$ with an exponential correlation matrix $\{(\mathbf{R})_{ik} = (0.9)^{|i-k|} e^{j2\pi(i-k)/12}\}$ [25]. The AWGN variance used in the simulations is 0.1.

First, we consider an 8×8 MIMO system transmitting $M = 8$ data streams with a power constraint $P_s \leq 8$. For the system considered, we have compared the performance of the linear transceiver structure we proposed (TX-RX1), the linear transceiver structure using the noisy estimate of the channel without considering the channel estimation error (TX-RX2), and their VBLAST versions [26]: VBLAST transmission with MMSE receivers considering the channel estimation error (TX-RX3), and VBLAST transmission with MMSE receiver using the noisy estimate of the channel (TX-RX4). For the sake of a fair comparison, linear MMSE receivers are used for VBLAST detection. We plot, in Fig. 2 and 3, the total MSE and the (uncoded) BER performances achieved by each linear transceiver structure versus the channel estimation error variance, σ_e^2. The results are evaluated over 10000 realizations of the MIMO channel with the same channel estimate. The linear transceiver structure we proposed, TX-RX1 performs the best in terms of both MSE and BER, and precoding and considering the channel estimation error provide robustness against the channel uncertainty. When the channel estimation error is low, the total MSE and BER performances of TX-RX1 and TX-RX2 pairs and TX-RX3 and TX-RX4 are very close as expected. However, as the accuracy of the channel estimate gets worse, the performances of the linear transceivers using the noisy channel estimate, TX-RX2 and TX-RX4 suffer dramatically, whereas the receiver structures considering the channel estimation errors, TX-RX1 and TX-RX3 provide robustness against the channel estimation errors.

Next, we consider a 8×8 MIMO system with a given channel. Power used for data transmission per symbol interval is $P_s = 8$, and the variance of the channel estimation error is $\sigma_e^2 = 0.01$. We plot, in Fig. 4, 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, [9] and [23]. 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 [9]. The CDMA user capacity bound [23] may not be achievable

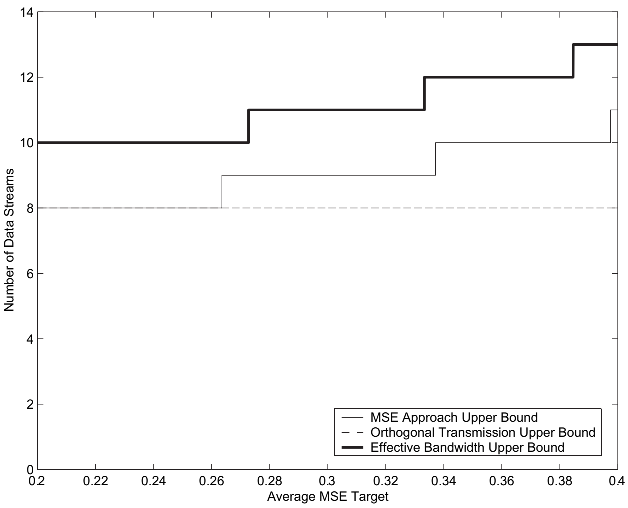

Fig. 4. Upper bound for the number of data streams vs MSE target for 8×8 MIMO system with $P_s = 8$, $\sigma_e^2 = 0.01$

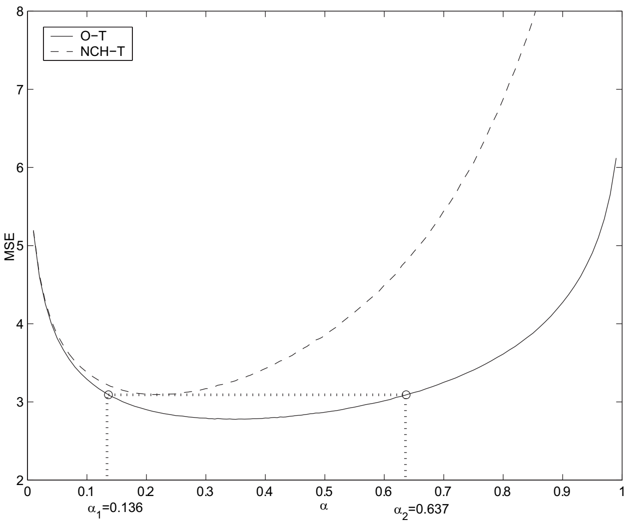

Fig. 5. MSE vs α for 8×8 MIMO system with $L_d = 2$ and $P_{total} = 50$

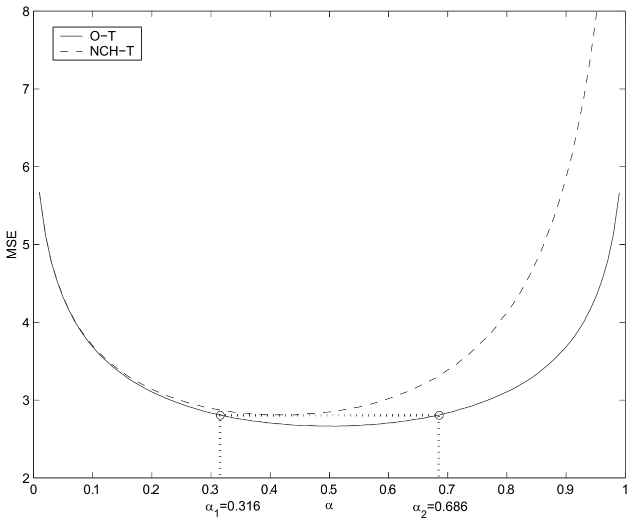

Fig. 6. MSE vs α for 8×8 MIMO system with $L_d = 8$ and $P_{total} = 100$

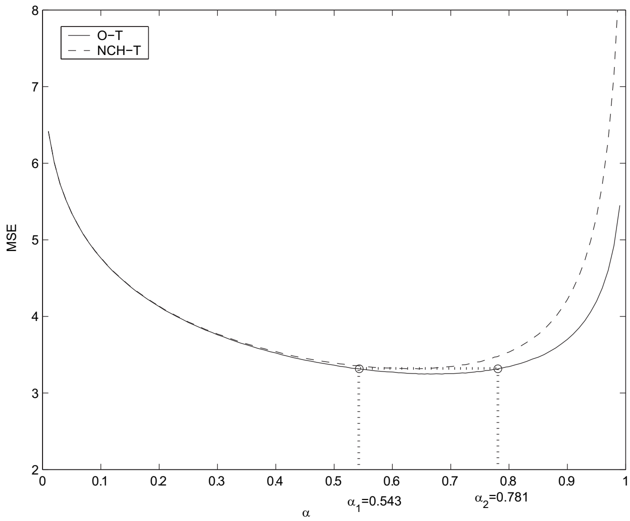

Fig. 7. MSE vs α for 8×8 MIMO system with $L_d = 40$ and $P_{total} = 150$

due to the constraints on the power and the channel structure.

To investigate the effect of the power allocation among the channel estimation process and data transmission, we consider an 8×8 MIMO system transmitting $M = 8$ data streams with three different values of L_d. First we consider a 8×8 MIMO system with $L_d = 2$ and $P_{total} = 50$, and evaluate the total MSE over 10000 realizations of the MIMO channel for different power allocation schemes. Fig. 5 shows the effect of power allocation on the total MSE as α changes for both the optimal linear transceiver we propose (O-T), and the linear transceiver structure using noisy channel estimate as the perfect CSI without considering the imperfection of the CSI (NCH-T). It is observed that the minimum total MSE is achieved at the optimal power allocation $\alpha_{opt} = 0.35$, and it is in the interval $[\alpha_{min}, \alpha_{max}]$=[0.334,0.354]. The optimum power allocation for uncorrelated case, $\tilde{\alpha}$ is 0.335. Observe that for a wide range of α, O-T achieves an MSE close to the minimum achievable MSE. Specifically, using an $\alpha \in [\alpha_1, \alpha_2] = [0.136, 0.637]$ with O-T still achieves a lower MSE than the minimum achievable MSE by NCH-T. However, the best performance is achieved when the system is optimized in terms of both the power allocation and linear transceiver structure. For $N_T = L_d = 8$ and $P_{total} = 100$, the effect of power allocation on the total MSE is presented in Fig. 6 where $\alpha_{opt} = 1/2$. For O-T, the range of α where it outperforms the best possible NCH-T transmission is $\alpha \in [\alpha_1, \alpha_2] = [0.316, 0.686]$. The third case, $N_T < L_d$ with $L_d = 40$ and $P_{total} = 150$ is investigated in Fig. 7, and $\alpha_{opt} = 0.67 \in [\alpha_{min}, \alpha_{max}] = [0.658, 0.691]$ with $\tilde{\alpha} = 0.689$. In this case, choosing $\alpha \in [\alpha_1, \alpha_2] = [0.543, 0.781]$ still outperforms NCH-T. For all cases, optimum power allocation with the optimal linear transceiver achieves the minimum total MSE.

Finally, in Fig. 8, we investigate the effect of correlation among the receive antennas on the performance of MIMO systems. We consider an 8×8 MIMO system with $L_d = 40$ and $P_{total} = 150$ with exponential correlations $R_1 = \{(\mathbf{R})_{ik} = (0.9)^{|i-k|} e^{j2\pi(i-k)/12}\}$, $R_2 = \{(\mathbf{R})_{ik} = (0.5)^{|i-k|} e^{j2\pi(i-k)/24}\}$, $R_3 = \{(\mathbf{R})_{ik} = (0.1)^{|i-k|} e^{j2\pi(i-k)/36}\}$, and evaluate the total MSE achieved by O-T with different α's. We observe that as the correla-

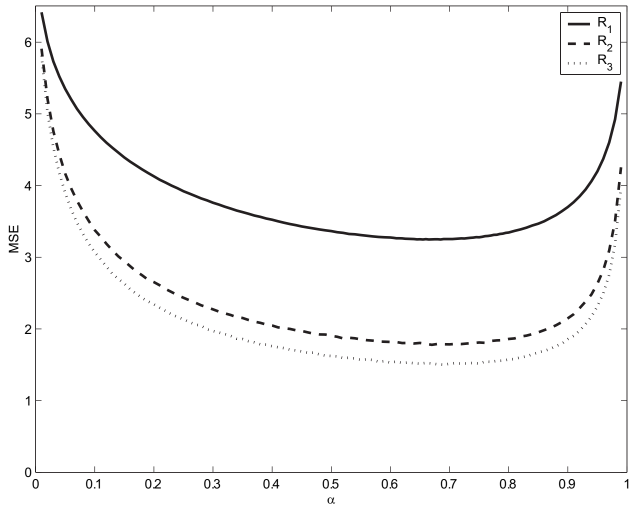

Fig. 8. MSE vs α for 8×8 MIMO system with $L_d = 40$ and $P_{total} = 150$ with different correlations among the receive antennas

tions among the receive antennas increase, the performance deteriorates due to the reduction in diversity.

VII. Conclusions

In this paper, we have developed the optimum linear transceiver structure for a MIMO link with arbitrary correlations among the receive antennas, that minimizes the total MSE in the presence of channel estimation errors, and distributes the total MSE equally among the parallel data streams. We show that knowing that we will have only an estimate of the channel, better transceivers can be designed, by appropriately considering the imperfection of the CSI.

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 MSE target. 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. Motivated by the profound effect of the quality of channel estimation on the performance of the MIMO link, we have considered the problem of optimum sharing of resources between the process of channel estimation and data transmission. Considering the total power as the limited resource, we have shown that, given the coherence time of the channel, there is a unique solution to the optimum allocation problem between the training based channel estimation and data transmission. We observe that the optimum power allocation depends on the system parameters, and the correlation structure of the receiver antennas.

It is important to note that the optimum linear transceiver structure and power allocation we propose here is designed for the single user MIMO link where only the receive antennas are correlated. In a more general scenario, the effect of correlation among the transmit antennas on the transceiver design should be considered. In the case of a multiuser MIMO system, channel estimate of each user will affect the performance of other users due to inherent interfering structure of the system. The problem of jointly optimum power allocation and transceiver design in a multiuser MIMO system remains to be investigated.

Appendix A
Proof of Lemma 1

The optimization problem at hand is the maximization of $\Delta_{ii}(\alpha)$ over $0 \le \alpha \le 1$ with $\Delta_{ii}(\alpha)$ given by (28). Observe that $\Delta_{ii}(\alpha)$ has an asymptote at $\alpha = \frac{-1}{c_i}$. When $c_i > 0, \frac{-1}{c_i} < 0$ and when $c_i < 0, \frac{-1}{c_i} > 1$. Thus, the pole always lies outside the interval of interest. In order to find the maximum, we first analyze the derivatives of the function $\Delta_{ii}(\alpha)$ $\forall \alpha$. The first and the second derivatives of the function with respect to α are

$$\frac{\partial(\Delta_{ii}(\alpha))}{\partial \alpha} = d_i \frac{-(\alpha)^2 c_i - 2\alpha + 1}{(1 + c_i \alpha)^2} \tag{40}$$

$$\frac{\partial^2(\Delta_{ii}(\alpha))}{\partial \alpha^2} = d_i \frac{-2 - 2c_i}{(1 + c_i \alpha)^3} \tag{41}$$

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

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

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

The first order condition for finding the optimum α is

$$-(\alpha)^2 c_i - 2\alpha + 1 = 0 \tag{42}$$

Solving the equation for all cases we obtain the following roots,

$$\alpha_i^\star = \begin{cases} \frac{-1}{c_i} \pm \frac{\sqrt{1+c_i}}{c_i}, & c_i > 0 \ ; \\ \frac{-1}{c_i} \pm \frac{\sqrt{1+c_i}}{c_i}, & -1 < c_i < 0 \ ; \\ 1, & c_i < -1 \ ; \end{cases} \tag{43}$$

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

Appendix B
Proof of Lemma 2

Consider two SNR matrices, $\Omega_1 \in \Psi$ and $\Omega_2 \in \Psi$. By definition, we have, for some α_1 and $\alpha_2 \in [0, 1]$, $\Omega_1 \preceq \Delta(\alpha_1)$ and $\Omega_2 \preceq \Delta(\alpha_2)$. For Ψ be a convex set, any linear combination of any two matrices in the set must be in the set:

$$\lambda \Omega_1 + (1 - \lambda)\Omega_2 = \Omega_3 \in \Psi \tag{44}$$

(44) implies that there should exist an $\alpha_3 \in [0, 1]$ satisfying $\Omega_3 \preceq \Delta(\alpha_3)$. Using the concavity of Δ_{ii} in Lemma 1, it follows

$$\Omega_3 \preceq \lambda \Delta(\alpha_1) + (1 - \lambda)\Delta(\alpha_2) \preceq \Delta(\lambda \alpha_1 + (1 - \lambda)\alpha_2) \tag{45}$$

Thus, $\alpha_3 = \lambda\alpha_1 + (1-\lambda)\alpha_2 \in [0,1]$.

REFERENCES

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[2] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 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 Trans. Inform. Theory*, 44(2):744–765, March 1998.

[4] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, Mar. 1998.

[5] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, Autumn 1996.

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

[7] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 48, no. 4, pp. 951–963, Apr. 2003.

[8] D. Samardzija and N. Mandayam, "Pilot assisted estimation of MIMO fading channel response and achievable data rates," in *DIMACS Wksp. Signal Processing for Wireless Communications*, October 2002.

[9] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.

[10] A. Scaglione *et al.*,, "Optimal designs for space-time linear precoders and decoders, "*IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[11] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3221–3231, Dec. 1994.

[12] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[13] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space time block coding based on channel mean feedback," *IEEE Trans. Signal Processing*, vol.50, no. 10, pp. 2599–2613, October 2002.

[14] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space time block coding based on channel correlations," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1673–1690, July 2003.

[15] Hong Zhihong *et al.*, "Spatial multiplexing in correlated fading via the virtual channel representation," *IEEE Trans. Select. Areas Commun.*, vol. 21, no. 5, pp. 856–866, June 2003.

[16] H. Sampath and A. Paulraj, "Linear precoding for space-time coded systems with known fading correlations," *IEEE Commun. Letters*, vol. 6, pp. 239–241, June 2002.

[17] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1165–1175, July 2004.

[18] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Trans. Signal Processing*, vol. 52, no. 5, pp. 1179–1197, May 2004.

[19] S. Serbetli and A. Yener, "Transceiver optimization for multiuser MIMO systems," *IEEE Trans. Signal Processing*, vol. 52, no. 1, pp. 214–226, Jan. 2004.

[20] S. Serbetli, S. Bethanabhotla, and A. Yener, "The effect of channel estimation on transceiver design for MIMO systems with QoS constraints," in *Conference on Information Sciences and Systems, CISS'04*, Mar. 2004.

[21] T. A. Thomas and F. W. Vook, "MIMO strategies for equal-rate data streams," in *Vehicular Technology Conference, VTC 2001*, pp. 548–552, Oct. 2001.

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

[23] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inform. Theory*, vol. 455, no. 2, pp. 641–657, Mar. 1999.

[24] D. Bertsekas, *Nonlinear Programming*, Athena Scientific, 1995.

[25] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Letters*, vol. 5, no. 9, pp. 369–371, Sept. 2001.

[26] G. D.Golden *et al.*, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, no. 1, pp. 14–16, Jan. 1999.

Semih Serbetli received the B.S. degree in Electrical and Electronics Engineering from Boğaziçi University, Istanbul, Turkey, in 2000, and the Ph.D. degree in Electrical Engineering from the Pennsylvania State University in 2005. During his Ph.D studies, he focused on transceiver design problems for multiuser multiple antenna systems. He is currently with Philips Research Labs, Eindhoven, The Netherlands. His research interests include transceiver optimization for wireless communication systems with an emphasis on multiple antenna (MIMO) systems, cooperative communications and multicarrier communication systems.

Aylin Yener received the B.S. degrees in Electrical and Electronics Engineering, and in Physics, from Boğaziçi University, Istanbul, Turkey, in 1991, and the M.S. and Ph.D. degrees in Electrical and Computer Engineering from Rutgers University, NJ, in 1994 and 2000, respectively. During her Ph.D. studies, she was with Wireless Information Network Laboratory (WINLAB) in the Department of Electrical and Computer Engineering at Rutgers University, NJ. Between fall 2000 and fall 2001, she was with the Electrical Engineering and Computer Science Department at Lehigh University, PA, where she was a P.C. Rossin assistant professor. Currently, she is with the Electrical Engineering department at Penn State, University Park, PA, as an associate professor. Dr. Yener is a recipient of the NSF CAREER award in 2003. She is an associate editor of the IEEE Transactions on Wireless Communications. Dr. Yener's research interests include performance enhancement of multiuser systems, wireless communication theory and wireless networking in general.