

Rate of MIMO Systems with CSI at Transmitter and Receiver from Pilot-Aided Estimation

Neelesh B. Mehta, Fadel F. Digham, Andreas F. Molisch and Jin Zhang

TR2004-092 October 2004

Abstract

We consider realistic multiple input multiple output antenna systems operating over spatially colored channels with instantaneous, albeit imperfect, channel state information at the receiver and only covariance knowledge available at the transmitter. We focus on pilot-aided channel estimation in which the receiver uses minimum mean square error channel estimation. For such a setup, our goal is then to optimally design both the pilot and data sequences to maximize the information rates achievable over the channel. We first demonstrate that the estimation error cannot be modeled as additive white Gaussian noise. We adopt a lower bound on channel capacity with imperfect channel knowledge and show that an optimum design leads to a matching of the eigenspaces of the pilots and of the data to the eigenspace of the channel. Furthermore, the ranks of the pilot and data covariance matrices need to be equal, and the optimal training duration need only equal the above rank. The assignment of powers to the different modes of transmission can then be obtained numerically. This paper, in essence, extends the results in the literature that assume perfect channel knowledge at the receiver, and shows that fully exploiting covariance knowledge leads to a visible improvement in capacity when compared to schemes that assume no channel knowledge, whatsoever, at the transmitter.

VTC, Fall 2004

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Rate of MIMO Systems with CSI at Transmitter and Receiver from Pilot-Aided Estimation

Neelesh B. Mehta[†], *Member, IEEE*, Fadel F. Digham^{‡,+}, *Student Member, IEEE*,
Andreas F. Molisch^{†,*}, *Senior Member, IEEE*, and Jin Zhang[†], *Senior Member, IEEE*

[†]Mitsubishi Electric Research Labs,
201 Broadway, Cambridge, MA, USA.
{mehta, molisch, jzhang}@merl.com

[‡]Univ. of Minnesota,
Minneapolis, USA.
fadel.digham@ieee.org

Abstract— We consider realistic multiple input multiple output antenna systems operating over spatially colored channels with instantaneous, albeit imperfect, channel state information at the receiver and only covariance knowledge available at the transmitter. We focus on pilot-aided channel estimation in which the receiver uses minimum mean square error channel estimation. For such a setup, our goal is then to optimally design both the pilot and data sequences to maximize the information rates achievable over the channel. We first demonstrate that the estimation error cannot be modeled as additive white Gaussian noise. We adopt a lower bound on channel capacity with imperfect channel knowledge and show that an optimum design leads to a matching of the eigenspaces of the pilots and of the data to the eigenspace of the channel. Furthermore, the ranks of the pilot and data covariance matrices need to be equal, and the optimal training duration need only equal the above rank. The assignment of powers to the different modes of transmission can then be obtained numerically. This paper, in essence, extends the results in the literature that assume perfect channel knowledge at the receiver, and shows that fully exploiting covariance knowledge leads to a visible improvement in capacity when compared to schemes that assume no channel knowledge, whatsoever, at the transmitter.

Index Terms—MIMO systems, Antenna arrays, Receiving antennas, Transmitting antennas, Covariance matrices, Training, Fading channels, Time-varying channels, Least mean square methods, Information rates, Matrix decomposition, Estimation, Data communication, Scattering, Optimization methods, Array signal processing, Feedback communication, Gaussian channels, Gaussian noise, Transmitters, Radio communication

I. INTRODUCTION

The prediction of high data rates achievable under idealized conditions by multiple input multiple output (MIMO) antenna systems [1], [2] has spurred studies of the capacity achievable by them under various more realistic assumptions about the channel as well as the transmitter and receiver architectures. It is now widely understood that the spatial channel model and assumptions about the channel state information (CSI) at the transmitter (CSIT) and the receiver (CSIR) have a significant impact on the MIMO capacity [3].

For many systems, instantaneous CSIT is not feasible. For frequency division duplex (FDD) systems, in which the

forward and the reverse links operate at different frequencies, instantaneous CSIT requires fast feedback, which decreases the overall capacity of the system. For time division duplex (TDD) systems, in which the forward and reverse links operate at the same frequency, the use of instantaneous CSIT is impractical in channels with small coherence intervals as it necessitates short delays between these links.

Transmitter design based only on small-scale-averaged statistics such as covariance knowledge circumvents this problem because covariance varies slowly and is determined by parameters like angular spread, mean angles of arrival, etc. These parameters are the same for both links even in FDD or quickly-varying TDD systems. Therefore, systems with covariance knowledge at the transmitter, denoted henceforth by CovKT, have received great attention recently [4]–[7]. However, all of these papers assume perfect CSIR.

Our paper considers a system in which the transmitter has access to only CovKT and the receiver uses a pilot-aided spatial minimum mean square error (MMSE) filter to generate instantaneous, albeit imperfect, CSIR. For brevity, we shall henceforth call this an MMSE receiver.¹ A novel and joint design of both the data and the pilots is proposed and optimizes a capacity bound on this system. For a given block fading length, we derive the necessary conditions for pilots and data to maximize capacity. The training duration and the power allocation ratio between the pilots and the data are also optimized. We show that covariance knowledge enables a more tailored pilot design resulting in a higher capacity than uniform pilot loading over all antennas that was previously assumed in the literature.

In the following, we briefly discuss related papers and explain how their assumptions differ from ours. A considerable body of work has been derived for *spatially white* channels. While this case gives valuable first insights, it does not correspond to the physical reality of most MIMO channels [8]. For the case of no CSIT and an MMSE receiver, Hassibi and Hochwald [9] considered pilot-aided channel estimation for a block fading wireless channel and derived a lower bound on the capacity given imperfect estimation at the receiver. They

⁺The author was with Mitsubishi Electric Research Labs during the course of this work.

^{*}A. F. Molisch is also at the Department of Electrosience, Lund University, Lund, Sweden.

¹Note that this does *not* refer to how the receiver processes the spatially multiplexed payload data.

then derived the optimal training sequences, training duration, and data and pilot power allocation ratio. The problem, including the case of a mismatched closed-loop system, has also been investigated in [10]–[12]. Orthogonal training sequences were shown to be optimal by Marzetta [13].

Data covariance for noisy channel estimates in spatially colored channels was looked at in [14]. However, the estimation noise was modeled in an ad hoc manner by adding a white noise to the channel state's spatially white component. In contrast to the aforementioned works, this paper takes into account channel correlation, training sequences, and MMSE channel estimation. Following a different track, lower and upper bounds on capacity were derived in [4], [15] without having to assume any a priori training schemes for generating CSIR, and serve as fundamental limits on capacity.

The rest of this paper is organized as follows. Section II sets up the basic model and notation. The joint pilot and data loading scheme is analyzed in Section III, and the numerical results are discussed in Section IV. We conclude in Section V. Only key mathematical details are highlighted in this paper; several details that are skipped are available in [16].

II. MIMO LINK MODEL

We consider a MIMO system with N_t transmit antennas and N_r receive antennas operating in a block fading frequency-flat channel model [15], in which the channel remains constant for T time instants and decorrelates thereafter. Of the T time instants, T_p are used for transmitting pilots, and the remaining $T_d = T - T_p$ for data. We shall use the subscripts p and d for symbols related to pilots and data, respectively. Let P_p and P_d denote the power allocated to pilots and data, respectively. Lower and upper case boldface letters shall be used to denote vectors and matrices, respectively.

A. Channel Model

The $N_r \times N_t$ matrix \mathbf{H} denotes the instantaneous channel state, where h_{ij} denotes the complex fading gain from transmit antenna j to receive antenna i . Experimental results have demonstrated that many channels allow a representation of their correlation matrix as a Kronecker product of the transmit and receive correlation matrices [17]. \mathbf{H} is then given by

$$\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{R}_t^{1/2}, \quad (1)$$

where \mathbf{R}_t and \mathbf{R}_r are the transmit and receive antenna correlation matrices, respectively. \mathbf{H}_w is spatially white, *i.e.*, its entries are zero-mean, independent, complex Gaussian random variables (RVs) with unit variance. Furthermore, we assume that $\mathbf{R}_r = \mathbf{I}_{N_r}$, which is fulfilled when the receiver is in a rich scattering environment, as is typically the case in the downlink of a cellular system, or a wireless LAN system from the access point to the receiving station. \mathbf{R}_t is taken to be full rank.

B. Training Phase

The signal received during the training phase of duration T_p instants, is an $N_r \times T_p$ matrix, $\mathbf{Y}_p = [y_{ij}]$, where y_{ij} is

the signal received at receive antenna i at time instant j . \mathbf{Y}_p is given by

$$\mathbf{Y}_p = \mathbf{H}\mathbf{X}_p + \mathbf{W}_p, \quad (2)$$

where $\mathbf{X}_p = [x_{ij}]$ is the transmitted pilot matrix of size $N_t \times T_p$ and is known a priori at the receiver. Here, x_{ij} is the signal transmitted from transmit antenna i at time j . \mathbf{W}_p is the spatially and temporally white noise matrix, defined in a similar manner; its entries have variance σ_w^2 .

C. Data Transmission

The noise vectors at different time instants are independent and identically distributed. Therefore, considering the capacity for block transmissions is equivalent to considering the capacity for vector transmissions. For any given time instant, the received vector, \mathbf{y}_d , is related to the transmitted signal vector, \mathbf{x}_d , by

$$\mathbf{y}_d = \mathbf{H}\mathbf{x}_d + \mathbf{w}_d, \quad (3)$$

where \mathbf{w}_d is the spatially white noise vector. \mathbf{y}_d , \mathbf{x}_d , and \mathbf{w}_d are of dimensions $N_r \times 1$, $N_t \times 1$, and $N_r \times 1$, respectively.

D. Other Notation

In the paper, $\mathbb{E}_{\Gamma_1|\Gamma_2}$ denotes the expectation over the RV Γ_1 given Γ_2 , $(\cdot)^\dagger$ is the Hermitian transpose, $(\cdot)^T$ is the transpose, $(\cdot)^{(k)}$ is the $k \times k$ principal sub-matrix formed by including the first k rows and the first k columns, $\text{Tr}\{\cdot\}$ is the trace, $|\cdot|$ is the determinant, and \mathbf{I}_n denotes the $n \times n$ identity matrix. The matrices $\mathbf{Q}_d = \mathbb{E}_{\mathbf{x}_d} [\mathbf{x}_d \mathbf{x}_d^\dagger]$ and $\mathbf{Q}_p = \mathbf{X}_p \mathbf{X}_p^\dagger$ denote the data and pilot covariance matrices, respectively.² The singular value decompositions of \mathbf{Q}_d , \mathbf{X}_p , \mathbf{Q}_p , and \mathbf{R}_t are written as $\mathbf{Q}_d = \mathbf{U}_d \mathbf{\Lambda}_d \mathbf{U}_d^\dagger$, $\mathbf{X}_p = \mathbf{U}_p \mathbf{\Sigma}_p \mathbf{V}_p^\dagger$, $\mathbf{Q}_p = \mathbf{U}_p \mathbf{\Lambda}_p \mathbf{U}_p^\dagger$, and $\mathbf{R}_t = \mathbf{U}_t \mathbf{\Lambda}_t \mathbf{U}_t^\dagger$. Note that \mathbf{Q}_d , \mathbf{Q}_p , and \mathbf{R}_t are all Hermitian. Also, $\mathbf{\Lambda}_p = \mathbf{\Sigma}_p \mathbf{\Sigma}_p^\dagger$.

E. MMSE Channel Estimator

Given the covariance information and the pilots \mathbf{X}_p , the MMSE channel estimator passes the received \mathbf{Y}_p through a deterministic matrix filter to generate the channel estimate $\hat{\mathbf{H}}$. For $\mathbf{R}_r = \mathbf{I}_{N_r}$, it can be shown that,

$$\hat{\mathbf{H}} = \mathbf{Y}_p \mathbf{A} = (\mathbf{H}\mathbf{X}_p + \mathbf{W}_p) \mathbf{A},$$

where

$$\mathbf{A} = (\mathbf{X}_p^\dagger \mathbf{R}_t \mathbf{X}_p + \sigma_w^2 \mathbf{I}_{T_p})^{-1} \mathbf{X}_p^\dagger \mathbf{R}_t. \quad (4)$$

Furthermore, $\hat{\mathbf{H}}$ is statistically equivalent to

$$\hat{\mathbf{H}} = \tilde{\mathbf{H}}_w \tilde{\mathbf{R}}_t^{1/2}, \quad (5)$$

where $\tilde{\mathbf{H}}_w$ is spatially white with its entries having unit variance, and $\tilde{\mathbf{R}}_t$ is given by

$$\tilde{\mathbf{R}}_t = \mathbf{R}_t \mathbf{X}_p (\mathbf{X}_p^\dagger \mathbf{R}_t \mathbf{X}_p + \sigma_w^2 \mathbf{I}_{T_p})^{-1} \mathbf{X}_p^\dagger \mathbf{R}_t. \quad (6)$$

The above result shows that, for an MMSE estimator, the error due to noise during estimation also affects the transmit antenna

²Given that the pilot \mathbf{X}_p is a deterministic matrix, no expectation operator is used for defining \mathbf{Q}_p .

correlation matrix of the estimated channel $\widehat{\mathbf{H}}$, and cannot be modeled by mere addition of a spatially white noise to \mathbf{H}_w , as was done in other work.

F. Capacity with Estimation Error

The channel estimation error is defined as $\Delta = \mathbf{H} - \widehat{\mathbf{H}}$. From (3), it follows that the data transmission phase is governed by the equation

$$\mathbf{y}_d = \widehat{\mathbf{H}}\mathbf{x}_d + \Delta\mathbf{x}_d + \mathbf{w}_d. \quad (7)$$

A capacity lower bound is obtained by considering a sub-optimal receiver that treats the term $\mathbf{e} = \Delta\mathbf{x}_d + \mathbf{w}_d$ as Gaussian noise [9], [18]. The channel capacity is therefore lower bounded by

$$C_\Delta = \left(1 - \frac{T_p}{T}\right) \mathbb{E}_{\widehat{\mathbf{H}}} \log_2 \left| \mathbf{I}_{N_t} + \widehat{\mathbf{H}}^\dagger (\mathbb{E}_{\mathbf{e}} [\mathbf{e}\mathbf{e}^\dagger])^{-1} \widehat{\mathbf{H}}\mathbf{Q}_d \right|.$$

The factor $\left(1 - \frac{T_p}{T}\right)$ is the training penalty resulting from pilot transmissions, which transfer no information.

Eqn. (5) implies that the distribution of $\widehat{\mathbf{H}}$ is left rotationally invariant³, *i.e.*, $\pi(\Theta\widehat{\mathbf{H}}) = \pi(\widehat{\mathbf{H}})$, where $\pi(\cdot)$ denotes the probability distribution function and Θ is any unitary matrix. It therefore follows from [9] that C_Δ is further lower bounded by,

$$C_\Delta \geq C_L = \left(1 - \frac{T_p}{T}\right) \mathbb{E}_{\widehat{\mathbf{H}}} \log_2 \left| \mathbf{I}_{N_t} + \frac{1}{\sigma_w^2 + \sigma_l^2} \widehat{\mathbf{H}}^\dagger \widehat{\mathbf{H}}\mathbf{Q}_d \right|, \quad (8)$$

where $\sigma_l^2 = \frac{1}{N_r} \text{Tr} \left\{ \mathbb{E}_{\Delta, \mathbf{x}_d} \left[\Delta\mathbf{x}_d\mathbf{x}_d^\dagger \Delta^\dagger \right] \right\}$. It can be shown that

$$\sigma_l^2 = \text{Tr} \left\{ \mathbf{Q}_d(\mathbf{R}_t - \widetilde{\mathbf{R}}_t) \right\}. \quad (9)$$

III. OPTIMAL JOINT PILOT AND DATA LOADING

We now seek solutions that maximize C_L , the lower bound on MIMO capacity with imperfect CSIR. The maximization problem can be stated as:

$$\max_{\substack{\mathbf{U}_d, \Lambda_d \\ \mathbf{X}_p, T_p}} \left(1 - \frac{T_p}{T}\right) \mathbb{E}_{\widehat{\mathbf{H}}} \log_2 \left| \mathbf{I}_{N_t} + \frac{\widehat{\mathbf{H}}^\dagger \widehat{\mathbf{H}}\mathbf{Q}_d}{\sigma_w^2 + \text{Tr} \left\{ \mathbf{Q}_d(\mathbf{R}_t - \widetilde{\mathbf{R}}_t) \right\}} \right|,$$

subject to the total energy constraint

$$P_p T_p + P_d T_d = PT, \quad (10)$$

where $\text{Tr} \{ \mathbf{Q}_d \} = P_d$, $\text{Tr} \{ \mathbf{X}_p \mathbf{X}_p^\dagger \} = P_p T_p$, and P is the total power budget. We first state the following lemma [19, Thm. 7.4.10] that shall come in handy.

Lemma 1: If AB and BA are positive semi-definite, there always exists a permutation τ such that $\text{Tr} \{ AB \} = \text{Tr} \{ BA \} = \sum_i \sigma_i(A) \sigma_{\tau(i)}(B)$, where $\sigma_i(\cdot)$ denotes the i th largest singular value. ■

The following theorem deals with just the self-interference term σ_l^2 .

³Due to notational differences, left rotational invariance in this paper is equivalent to right rotational invariance in [9].

Theorem 1:

$$\min_{\mathbf{U}_p, \mathbf{U}_d} \sigma_l^2 = \min_{\mathbf{U}_p, \mathbf{U}_d} \text{Tr} \left\{ \mathbf{Q}_d(\mathbf{R}_t - \widetilde{\mathbf{R}}_t) \right\} = \sigma_w^2 \sum_{i=1}^{N_t} \frac{\lambda_{d_i} \lambda_{t_i}}{\sigma_w^2 + \lambda_{t_i} \lambda_{p_i}},$$

where $\lambda_{t_1} \geq \lambda_{t_2} \geq \dots \geq \lambda_{t_{N_t}}$, $\lambda_{d_1} \geq \lambda_{d_2} \geq \dots \geq \lambda_{d_{N_t}}$, and $\lambda_{p_1} \geq \lambda_{p_2} \geq \dots \geq \lambda_{p_{N_t}}$.

Proof: In the sequence of inequalities that follow, we first arrive at a lower bound for σ_l^2 , without commenting at each step, on the conditions required to achieve equality. At the very end, we show that equality is indeed achievable.

First, define the following matrices,

$$\mathbf{S}_3 = \mathbf{U}^\dagger \begin{bmatrix} \left((\mathbf{U}\Lambda_t\mathbf{U}^\dagger)^{(k_p)} + \sigma_w^2 \Lambda_p^{(k_p)-1} \right)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{U},$$

$\mathbf{S}_2 = \Lambda_t(\mathbf{I}_{N_t} - \mathbf{S}_3\Lambda_t)$, and $\mathbf{S}_1 = \mathbf{V}\mathbf{S}_2\mathbf{V}^\dagger$, where $\mathbf{U} = \mathbf{U}_p^\dagger\mathbf{U}_t$ and $\mathbf{V} = \mathbf{U}_d^\dagger\mathbf{U}_t$.

It can be shown that $\sigma_l^2 = \text{Tr} \{ \Lambda_d \mathbf{S}_1 \}$. Lemma 1 enables the following key simplification that eliminates \mathbf{V} :

$$\begin{aligned} \min \text{Tr} \{ \Lambda_d \mathbf{S}_1 \} &= \min \text{Tr} \{ \Lambda_d \mathbf{S}_2 \}, \\ &= \text{Tr} \{ \Lambda_d \Lambda_t \} - \max \text{Tr} \{ \Lambda_d \Lambda_t^2 \mathbf{S}_3 \}. \end{aligned}$$

Only $\text{Tr} \{ \Lambda_d \Lambda_t^2 \mathbf{S}_3 \}$ now remains to be maximized. It can then be shown that

$$\text{Tr} \{ \Lambda_d \Lambda_t^2 \mathbf{S}_3 \} \leq \text{Tr} \{ \mathbf{S}_4^{-1} \}, \quad (11)$$

where

$$\mathbf{S}_4 = \Lambda_d^{(k_p)-1} \Lambda_t^{(k_p)-1} \left(\mathbf{I}_{k_p} + \sigma_w^2 \mathbf{U}^{(k_p)-1} \Lambda_p^{(k_p)-1} \mathbf{U}^{(k_p)-1} \Lambda_t^{(k_p)-1} \right).$$

Equality occurs when $\mathbf{U}^{(k_p)}$ is unitary.

Using results from matrix algebra [20] and Lemma 1, we know that the extrema of $\text{Tr} \{ \mathbf{S}_4^{-1} \}$ occur when $\mathbf{U}^{(k_p)}$ is a unitary, permutation matrix. It turns out that the identity permutation $\mathbf{U}^{(k_p)} = \mathbf{I}_{k_p}$ maximizes $\text{Tr} \{ \Lambda_d \Lambda_t^2 \mathbf{S}_3 \}$. Therefore,

$$\sigma_l^2 \geq \text{Tr} \{ \Lambda_d \Lambda_t \} - \sum_{i=1}^{k_p} \frac{\lambda_{t_i} \lambda_{d_i}}{1 + \sigma_w^2 \lambda_{p_i}^{-1} \lambda_{t_i}^{-1}} = \sigma_w^2 \sum_{i=1}^{N_t} \frac{\lambda_{d_i} \lambda_{t_i}}{\sigma_w^2 + \lambda_{t_i} \lambda_{p_i}},$$

where k_p is the rank of \mathbf{Q}_p . Finally, that equality is achievable is verified by substituting $\mathbf{U}_p = \mathbf{U}_d = \mathbf{U}_t$ in $\text{Tr} \left\{ \mathbf{Q}_d(\mathbf{R}_t - \widetilde{\mathbf{R}}_t) \right\}$ and simplifying its expression. ■

C_L is a function of $\mathbf{Q}_d = \mathbf{U}_d \Lambda_d \mathbf{U}_d^\dagger$ and $\widetilde{\mathbf{R}}_t = \widetilde{\mathbf{U}}_t \widetilde{\Lambda}_t \widetilde{\mathbf{U}}_t^\dagger$, which depends on the pilots $\mathbf{X}_p = \mathbf{U}_p \Sigma_p \mathbf{V}_p^\dagger$. The optimal joint pilot and data design that maximizes C_L now follows.

Theorem 2: C_L satisfies the following upper bound:

$$C_L(\Lambda_d, \mathbf{U}_d, \mathbf{X}_p) \leq \left(1 - \frac{T_p}{T}\right) \mathbb{E}_{\widetilde{\mathbf{H}}_w} \log_2 \left| \mathbf{I}_{N_t} + \frac{\widetilde{\mathbf{H}}_w^\dagger \widetilde{\mathbf{H}}_w \widetilde{\Lambda}_t \Lambda_d}{\sigma_w^2 + \sigma_w^2 \sum_{i=1}^{N_t} \frac{\lambda_{t_i} \lambda_{d_i}}{\sigma_w^2 + \lambda_{p_i} \lambda_{t_i}}} \right|.$$

Furthermore, the upper bound is achieved when the eigenspaces match: $\mathbf{U}_d = \mathbf{U}_p = \mathbf{U}_t = \widetilde{\mathbf{U}}_t$, and, therefore, constitutes an optimal solution. ■

The proof for Thm. 2, not shown here, obtains consecutive upper bounds by first minimizing the denominator (using

Thm. 1) and then independently maximizing the numerator. In general, the maximizing arguments responsible for the two optimizations need not be the same, and the upper bound need not be achievable. However, the special structure inherent in our problem makes it a notable exception.

After eigenspace matching, $\tilde{\Lambda}_t$ simplifies to

$$\tilde{\Lambda}_t = \Lambda_t^2 \Lambda_p (\Lambda_t \Lambda_p + \sigma_w^2 \mathbf{I}_{T_p})^{-1}. \quad (12)$$

Let k_d and k_p denote the ranks of \mathbf{Q}_d and \mathbf{Q}_p , respectively. We now state the rank properties of the optimal \mathbf{Q}_d and \mathbf{Q}_p , and the optimal training duration T_p . The proofs are in [16].

Theorem 3: The data and pilot loading matrices \mathbf{Q}_d and \mathbf{Q}_p must be of the same rank to maximize C_L . ■

Theorem 4: C_L is maximized when $T_p = k_p = k$. ■

Given that [1] $k_d \leq \min(N_t, N_r)$, an important implication for transmit diversity systems with $N_r < N_t$ is the following.

Corollary 1: $T_p \leq \min(N_t, N_r)$. ■

The singular values of \mathbf{Q}_d and \mathbf{Q}_p , namely, Λ_d and Λ_p (and thereby P_d , P_p , and k) depend on P , T , and Λ_t , and are optimized numerically. Note that the conditions derived above, combined with the simple expressions for C_L and $\tilde{\Lambda}_t$, drastically reduce the search space to determine all the optimal parameters, and make the search feasible.

IV. RESULTS AND DISCUSSION

While the results above are applicable to arbitrary antenna arrangements, the numerical results presented in this section are for uniform linear arrays (ULA) at the transmitter and the receiver. Four transmit and receive antennas are assumed ($N_t = N_r = 4$), the mean angle of departure is 45° , and the antenna spacing is $d_t = 0.5\lambda$, where λ is the carrier wavelength. The block fading duration, T , is 10, and the noise variance, σ_w^2 , is normalized to 1, without loss of generality. The transmitter covariance matrix \mathbf{R}_t for a ULA is evaluated from [21]. The effects of varying system parameters such as N_r , N_t , T , etc., are investigated in detail in [16], [22].

Figure 1 plots C_L as a function of the total power budget P (in dB) for $\sigma_\theta = 5^\circ, 10^\circ, 15^\circ$, and 30° . We first note that for small P , C_L increases as the angular dispersion σ_θ decreases. This can be explained as follows. At lower P values, only one eigenmode is used. This corresponds to the region where beam-forming ($\text{rank}(\mathbf{Q}_d) = 1$) is optimal [23]. Reducing σ_θ increases the spread of the eigenvalues and concentrates more of the energy in the strongest eigenmode of \mathbf{R}_t that is used for transmission, thereby increasing the capacity. On the other hand, for higher values of P , it is optimal to allocate power to more, or even all, the eigenmodes of \mathbf{R}_t . Therefore, increasing σ_θ , which leads to a smaller spread in the eigen values for different modes, increases capacity. The points at which the number of eigenmodes used for transmission, k , changes are also illustrated in the figure.

Figure 2 investigates the optimal power ratio $\alpha = P_p/P_d$, as a function of P for $\sigma_\theta = 5^\circ, 10^\circ, 15^\circ$, and 30° . As P increases, α monotonically decreases from a value greater than 3.0 for $P = 0$ dB to 1.5 for large P . However, α is always greater than 1. The discontinuities in α occur when

more eigenmodes of \mathbf{R}_t begin to get used for transmission. k is the primary factor that determines the ratio; given $k = 1$, α marginally increases with σ_θ and is sensitive to P , while for the same $k > 1$, α is insensitive to both σ_θ and P .

A. Combined Impact of Estimation Error and Covariance Knowledge

Figure 3 compares C_L , the capacity achieved by the proposed system that jointly designs the pilot and data given CovKT and pilot-aided imperfect estimation at the receiver, with systems in which covariance knowledge at the transmitter is not exploited or is unavailable and perfect or imperfect estimation at the receiver. The parameter values are $\sigma_\theta = 10^\circ$, $T = 10$, and $N_r = N_t = 4$. Note that in the idealized systems with perfect CSIR, resources need not be wasted on pilots ($P_p = 0$ and $T_p = 0$). For reference, the performance of a system with CovKT that has perfect CSIR and, yet, incurs a training penalty equal to the number of transmit modes in use is also plotted. Without any CSIT, the training duration T_p needs to be N_t [9].

It can be seen that CovKT with perfect CSIR yields the same capacity as instantaneous and perfect CSIT and CSIR. Compared to the worst case (imperfect CSIR and no CSIT), perfect estimation of the CSIR (without exploiting CovKT), improves the capacity by 0.6 bits/sec/Hz at 0 dB and 1.7 bits/sec/Hz at 30 dB. Exploiting CovKT increases the capacity by another 1.1 bits/sec/Hz at 0 dB and 1.7 bits/sec/Hz at 30 dB, even though the CSIR is imperfect.

V. CONCLUSIONS

We considered a novel system that tailors not only the data, but also the pilot sequences, to the covariance knowledge that is available at the transmitter. Unlike instantaneous channel state information, covariance knowledge at the transmitter varies slowly and can be obtained, without explicit feedback, even in FDD systems and quickly-varying TDD systems. Mirroring reality, the CSIR was not assumed to be perfect – it was acquired at the receiver using pilot-aided MMSE channel estimation. Based on an analytically tractable lower bound of the ergodic capacity, we found a jointly optimal solution for pilots and data. The capacity is maximized when the eigenspaces of the covariance matrices of the pilots and data match that of the transmit covariance matrix \mathbf{R}_t . Furthermore, it is optimal to transmit the data over only those eigenmodes of \mathbf{R}_t that are allocated power during training and vice versa. It was shown that the optimal training duration can indeed be less than the number of transmit antennas, and equals the number of eigenmodes used for data transmission. For low angular spreads, imperfect CSIR with covariance knowledge outperforms perfect CSIR that does not exploit covariance knowledge.

These results are in contrast to those without CovKT in [9], [13], where the optimal training duration was N_t and the optimal pilot sequences were orthogonal. The results demonstrate that fully exploiting covariance knowledge at the transmitter increases the ergodic capacity achievable by realistic MIMO

systems. Our results are also a generalization of the results in the literature that assume perfect CSIR in their derivations.

REFERENCES

- [1] G. J. Foschini and M. J. Gans, "On the limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, 1998.
- [2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, pp. 585–595, 1999.
- [3] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 684–702, June 2003.
- [4] S. A. Jafar and A. Goldsmith, "Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information," *To appear in IEEE Trans. Wireless Commun.*, 2004.
- [5] S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 406–417, April 2003.
- [6] E. Jorswieck and H. Boche, "Optimal transmission with imperfect channel state information at the transmit antenna array," *Wireless Pers. Commun.*, pp. 33–56, October 2003.
- [7] A. M. Tulino, S. Verdú, and A. Lozano, "Capacity of antenna arrays with space, polarization and pattern diversity," in *ITW*, pp. 324–327, 2003.
- [8] A. F. Molisch and F. Tufvesson, *Multipath propagation models for broadband wireless systems*. Digital Signal Processing for Wireless Communications Handbook, M. Ibnkahla (ed.), CRC Press, 2004.
- [9] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, pp. 951–963, 2003.
- [10] D. Samardzija and N. Mandyam, "Pilot-assisted estimation of MIMO fading channel response and achievable data rates," *IEEE Trans. Sig. Proc.*, pp. 2882–2890, 2003.
- [11] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE Trans. Commun.*, pp. 2358–2368, 2001.
- [12] T. Yoo and A. Goldsmith, "Capacity of fading MIMO channels with channel estimation error," *Allerton*, 2002.
- [13] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annual Allerton Conf. Commun., Control, and Computing*, 1999.
- [14] T. Yoo, E. Yoon, and A. Goldsmith, "MIMO capacity with channel uncertainty: Does feedback help?," in *Submitted to Globecom*, 2004.
- [15] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, January 1999.
- [16] F. F. Digham, N. B. Mehta, A. F. Molisch, and J. Zhang, "Joint pilot and data loading for MIMO systems with covariance knowledge," *Submitted to IEEE Trans. Wireless Commun.*, 2004.
- [17] J. P. Kermaol *et al.*, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Select. Areas Commun.*, pp. 1211–1226, 2002.
- [18] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, pp. 933–946, 2000.
- [19] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1996.
- [20] T. P. Minka, "Old and new matrix algebra useful for statistics." MIT Media Lab note, 2000. Also available on the web at <http://www.stat.cmu.edu/~minka/papers/matrix/minka-matrix.pdf>.
- [21] D. Asztely, "On antenna arrays in mobile communication systems: Fast fading and GSM base station receiver algorithms," Tech. Rep. IR-S3-SB-9611, Royal Institute of Technology, 1996.
- [22] F. F. Digham, N. B. Mehta, A. F. Molisch, and J. Zhang, "Joint pilot and data loading technique for mimo systems operating with covariance feedback," in *3G Mobile Communication Technologies*, 2004.
- [23] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *To appear in IEEE Trans. Wireless Commun.*, 2004.

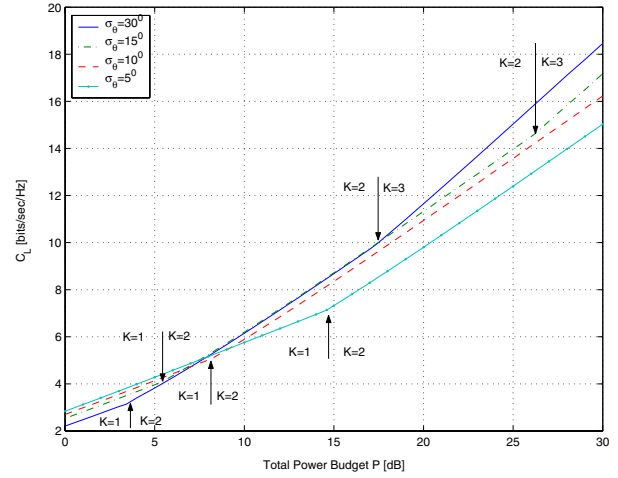


Fig. 1. Ergodic capacity, C_L , and eigenmode transitions as a function of P (dB) for small angular dispersions: $\sigma_\theta = 5^\circ, 10^\circ, 15^\circ$, and 30° ($N_t = N_r = 4, T = 10$).

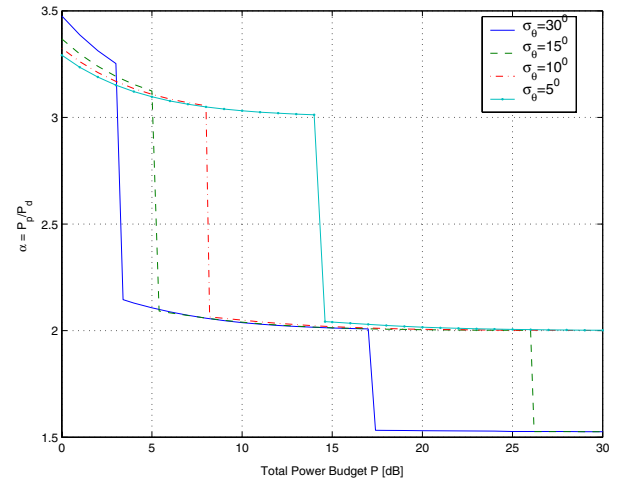


Fig. 2. Effect of P and σ_θ on the optimal power ratio α ($N_t = N_r = 4, T = 10$).

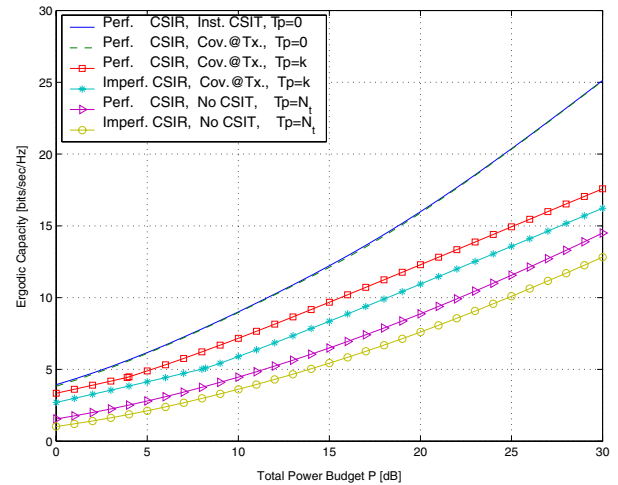


Fig. 3. Ergodic capacity comparison of systems with and without covariance knowledge at transmitter and with perfect or imperfect CSIR for $\sigma_\theta = 10^\circ$ ($N_t = N_r = 4, T = 10$).