

# Design Guidelines for Training-Based MIMO Systems With Feedback

Xiangyun Zhou, *Student Member, IEEE*, Parastoo Sadeghi, *Senior Member, IEEE*,  
Tharaka A. Lamahewa, *Member, IEEE*, and Salman Durrani, *Member, IEEE*

**Abstract**—In this paper, we study the optimal training and data transmission strategies for block fading multiple-input multiple-output (MIMO) systems with feedback. We consider both the channel gain feedback (CGF) system and the channel covariance feedback (CCF) system. Using an accurate capacity lower bound as a figure of merit that takes channel estimation errors into account, we investigate the optimization problems on the temporal power allocation to training and data transmission as well as the training length. For CGF systems without feedback delay, we prove that the optimal solutions coincide with those for nonfeedback systems. Moreover, we show that these solutions stay nearly optimal even in the presence of feedback delay. This finding is important for practical MIMO training design. For CCF systems, the optimal training length can be less than the number of transmit antennas, which is verified through numerical analysis. Taking this fact into account, we propose a simple yet near optimal transmission strategy for CCF systems, and derive the optimal temporal power allocation over pilot and data transmission.

**Index Terms**—Channel covariance feedback, channel estimation, channel gain feedback, information capacity, multiple-input multiple-output.

## I. INTRODUCTION

THE study of multiple-input multiple-output (MIMO) communication systems can be broadly categorized based on the availability and accuracy of channel state information (CSI) at the receiver or the transmitter side. Under the perfect CSI assumption at the receiver, the MIMO channel information capacity and data transmission strategies often have elegantly simple forms and many classical results exist in the literature [1], [2]. From [2]–[8], we know that the MIMO information capacity with perfect receiver CSI can be further increased if some form of CSI is fed back to the transmitter. The transmitter CSI can be in the form of causal channel gain feedback (CGF) or channel covariance feedback (CCF).

In practical communication systems with coherent detection, the state of the MIMO channel needs to be estimated at the receiver and hence, the receiver CSI is never perfect due to

noise and time variations in the fading channel. Pilot-symbol-assisted modulation (PSAM) is a widely used technique to facilitate channel estimation at the receiver [9]. As pilot symbols are not information-bearing signals, an important design aspect of communication systems is the optimal allocation of resources (such as power and time) to pilot symbols that results in the best tradeoff between the quality of channel estimation and rate of information transfer. Three pilot parameters under a system designer's control are 1) spatial structure of pilot symbols, 2) temporal power allocation to pilot and data, and 3) the number of pilot symbols or simply training length.

The optimal pilot design has been studied from an information-theoretic viewpoint for nonfeedback multiantenna systems of practical interest [10]–[12]. For nonfeedback MIMO systems with independent and identically distributed (i.i.d.) channels, the authors in [10] formulated an ergodic capacity lower bound and used it as the performance metric to find optimal solutions for all the aforementioned pilot design parameters. For CCF systems with correlated MIMO channels, the optimal solution for the pilot's spatial structure was investigated in [13]–[15]. However, optimal solutions for the temporal pilot power allocation and training length are generally unknown for both i.i.d. and correlated MIMO systems with any form of feedback. Some results were reported in [16] for rank-deficient channel covariance matrix known at the transmitter, which are based on a relaxed capacity lower bound. However, this relaxed capacity bound is generally loose for moderately to highly correlated channels, which can render the provided solutions suboptimal.

### A. Approach and Contributions

In this paper, we are concerned with the optimal design of pilot parameters for MIMO systems with various forms of feedback at the transmitter. Our figure of merit is an accurate lower bound on the ergodic capacity of MIMO systems, which is an extension of those derived in [17] from i.i.d. channels to correlated channels. We address practical design questions such as: Are the simple solutions provided in [10] for nonfeedback MIMO systems also optimal for systems with feedback? In CGF systems, where feedback delay needs to be considered, we would like to know whether the optimal pilot design is significantly affected by the feedback delay. In CCF systems with correlated channels, the optimal training length may be shorter than the number of transmit antennas, which is difficult to solve analytically. In this case, we would like to know whether a near optimal, yet simple transmission and power allocation strategy exists. The main contributions of this paper are summarized as follows.

Manuscript received November 06, 2008; accepted April 16, 2009. First version published May 27, 2009; current version published September 16, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Kainam Thomas Wong. This work was supported under the Australian Research Council's Discovery Projects funding scheme (project no. DP0773898).

The authors are with the College of Engineering and Computer Science, Australian National University, Canberra, ACT 0200, Australia (e-mail: xiangyun.zhou@anu.edu.au; parastoo.sadeghi@anu.edu.au; tharaka.lamahewa@anu.edu.au; salman.durrani@anu.edu.au).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2009.2023930

- For delayless CGF systems with i.i.d. channels, we prove that the solutions to the optimal temporal power allocation to pilot and data transmission as well as the optimal training length coincide with the solutions for nonfeedback systems.
- For delayed CGF systems with i.i.d. channels, our numerical results show that evenly distributing the power over the entire data transmission (regardless of the delay time) gives near optimal performance at practical signal-to-noise ratio (SNR) conditions. As a result, the solutions to the optimal temporal power allocation to pilot and data transmission, as well as the optimal training length for the delayless system stay nearly optimal regardless of the delay time.
- For CCF systems with correlated channels, we propose a simple transmission scheme, taking into account the fact that the optimal training length  $L_p$  is at most as large as the number of transmit antennas. This scheme only requires numerical optimization of  $L_p$  and does not require numerical optimization over the spatial or temporal power allocation over pilot and data transmission. Our numerical results show that this scheme is very close to optimal. In addition, optimizing  $L_p$  can result in a significant capacity improvement for correlated channels.
- Using the proposed scheme for CCF systems, we find the solution to the optimal temporal power allocation to pilot and data transmission, which does not depend on the channel spatial correlation under a mild condition on block length or SNR. Therefore, the proposed transmission and power allocation schemes for CCF systems give near optimal performance while having very low computational complexity.

This paper is organized as follows. The PSAM transmission scheme, channel estimation method, as well as the capacity lower bound are presented in Section II. The optimal transmission and power allocation strategy for nonfeedback systems are summarized in Section III. The optimal transmission and power allocation strategy for CGF and CCF systems are studied in Section IV and Section V, respectively. Finally, the main contributions of this paper are summarized in Section VI.

The following notations will be used in the paper: Boldface upper and lower case letters denote matrices and column vectors, respectively. The matrix  $\mathbf{I}_N$  is the  $N \times N$  identity matrix.  $[\cdot]^*$  and  $[\cdot]^\dagger$  denotes the complex conjugate and conjugate transpose operation, respectively.  $E\{\cdot\}$  denotes the mathematical expectation.  $\text{tr}\{\cdot\}$ ,  $|\cdot|$  and  $\text{rank}\{\cdot\}$  denote the matrix trace, determinant and rank, respectively.

## II. SYSTEM MODEL

We consider a MIMO block-flat-fading channel model with input-output relationship given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (1)$$

where  $\mathbf{y}$  is the  $N_r \times 1$  received symbol vector,  $\mathbf{x}$  is the  $N_t \times 1$  transmitted symbol vector,  $\mathbf{H}$  is the  $N_r \times N_t$  channel gain matrix, and  $\mathbf{n}$  is the  $N_r \times 1$  noise vector having zero-mean circularly symmetric complex Gaussian (ZMCSCG) entries with unit variance. The entries of  $\mathbf{H}$  are also ZMCSCG with unit variance. We consider spatial correlations among the transmit antennas only.

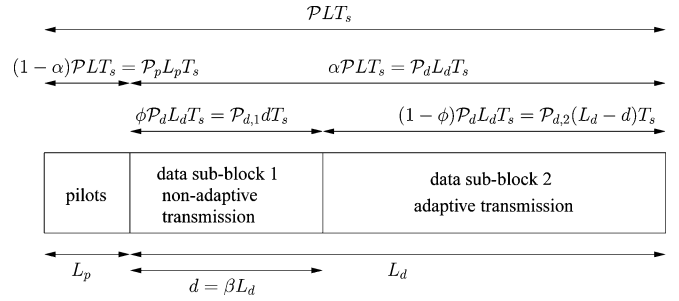


Fig. 1. An example of a transmission block of  $L$  symbols in a system with delayed feedback. Temporal power allocations are shown at the top and the length of each subblock is shown at the bottom.

The spatial correlation is characterized by the covariance matrix  $\mathbf{R}_H = E\{\mathbf{H}^\dagger \mathbf{H}\} / N_r$ . Therefore,  $\mathbf{H} = \mathbf{H}_0 \mathbf{R}_H^{1/2}$ , where  $\mathbf{H}_0$  has i.i.d. ZMCSCG entries with unit variance. We assume that  $\mathbf{R}_H$  is a positive-definite matrix and denote the eigenvalues of  $\mathbf{R}_H$  by  $\mathbf{g} = [g_1 \ g_2 \ \dots \ g_{N_r}]^T$ . Furthermore, we use the concept of majorization to characterize the degree of channel spatial correlation [18], [19], which is described in detail in [20].

### A. Transmission Scheme

Fig. 1 shows an example of a transmission block of  $L$  symbol periods in a PSAM scheme. The channel gains remain constant over one block and change to independent realizations in the next block. This is an appropriate channel model for time-division multiple access or frequency-hopping systems [10]. During each transmission block, each transmit antenna sends  $L_p$  pilot symbols, followed by  $L_d (= L - L_p)$  data symbols as shown in Fig. 1. The receiver performs channel estimation during the pilot transmission. For CGF systems, the receiver feeds the channel estimates back to the transmitter once per block to allow adaptive data transmission in the form of power control. In practical scenarios, there is a time delay of  $d$  symbol periods before the transmitter receives the feedback information as shown in Fig. 1. Therefore, the data transmission during the first  $d$  symbol periods is not adaptive to the channel, and adaptive transmission is only available for the remaining  $L_d - d$  symbol periods. We define  $\beta = d/L_d$  as the feedback delay factor. For CCF systems, less frequent feedback is required as the channel correlation changes much slower than the channel gains. Therefore, we do not consider feedback delay in CCF systems, i.e.,  $d = 0$ . Note that for nonfeedback systems,  $d = L_d$ .

The total transmission energy per block is given by  $\mathcal{P}LT_s$  as shown in Fig. 1, where  $\mathcal{P}$  is the average power per transmission and  $T_s$  is the symbol duration. We define the PSAM power factor as the ratio of the total energy allocated to the data transmission, denoted by  $\alpha$ . We also denote the power or SNR per pilot and data transmission by  $\mathcal{P}_p$  and  $\mathcal{P}_d$ , respectively. Therefore, we have the following relationships:

$$\begin{aligned} \mathcal{P}LT_s &= \mathcal{P}_p L_p T_s + \mathcal{P}_d L_d T_s, \\ \mathcal{P}_p &= (1 - \alpha) \frac{\mathcal{P}L}{L_p}, \quad \text{and} \quad \mathcal{P}_d = \alpha \frac{\mathcal{P}L}{L_d}. \end{aligned} \quad (2)$$

Ideally for CGF systems,  $\mathcal{P}_d$  should be larger for the transmission blocks over which the channel is strong and smaller for blocks over which the channel is weak. However, the results in [17] suggest that this temporal data power adaptation provides little capacity gain, hence it is not considered in this paper.

For CGF systems with delay of  $d$  symbol periods as shown in Fig. 1, we define the data power division factor as the ratio of the total data energy allocated to the nonadaptive subblock, denoted by  $\phi$ . Therefore, we have the following relationships:

$$\begin{aligned} \mathcal{P}_d L_d T_s &= \mathcal{P}_{d,1} d T_s + \mathcal{P}_{d,2} (L_d - d) T_s, \\ \mathcal{P}_{d,1} &= \frac{\phi}{\beta} \mathcal{P}_d, \quad \text{and} \quad \mathcal{P}_{d,2} = \frac{1 - \phi}{1 - \beta} \mathcal{P}_d \end{aligned} \quad (3)$$

where  $\mathcal{P}_{d,1}$  and  $\mathcal{P}_{d,2}$  are the power per transmission during the nonadaptive and adaptive subblocks.

### B. Channel Estimation

In each transmission block, the receiver performs channel estimation during the pilot transmission. Assuming the channel spatial correlation can be accurately measured at the receiver, the channel gain  $\mathbf{H}$  can be estimated using the linear minimum mean square error (LMMSE) estimator [21]. We denote the channel estimate and estimation error as  $\hat{\mathbf{H}} = \hat{\mathbf{H}}_0 \mathbf{R}_{\mathbf{H}}^{1/2}$  and  $\tilde{\mathbf{H}} = \tilde{\mathbf{H}}_0 \mathbf{R}_{\mathbf{H}}^{1/2}$ , respectively, where  $\hat{\mathbf{H}}_0$  and  $\tilde{\mathbf{H}}_0$  have i.i.d. ZM-CSCG entries with unit variance.  $\hat{\mathbf{H}}$  is given as [14]

$$\hat{\mathbf{H}} = \mathbf{Y} (\mathbf{X}_p^\dagger \mathbf{R}_{\mathbf{H}} \mathbf{X}_p + \mathbf{I}_{L_p})^{-1} \mathbf{X}_p^\dagger \mathbf{R}_{\mathbf{H}} \quad (4)$$

where  $\mathbf{Y}$  is the  $N_t \times L_p$  matrix combining the  $L_p$  received symbol vectors during pilot transmission and  $\mathbf{X}_p$  is the  $N_t \times L_p$  pilot matrix. The covariance matrix of the estimation error is given by [14]

$$\mathbf{R}_{\tilde{\mathbf{H}}} = \frac{E\{\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}}\}}{N_r} = (\mathbf{R}_{\mathbf{H}}^{-1} + \mathbf{X}_p \mathbf{X}_p^\dagger)^{-1}. \quad (5)$$

From the orthogonality property of LMMSE estimator, we have

$$\mathbf{R}_{\hat{\mathbf{H}}} = \frac{E\{\hat{\mathbf{H}}^\dagger \hat{\mathbf{H}}\}}{N_r} = \mathbf{R}_{\mathbf{H}} - \mathbf{R}_{\tilde{\mathbf{H}}}. \quad (6)$$

### C. Ergodic Capacity Bounds

The exact capacity expression under imperfect receiver CSI is still unavailable. We consider one lower and one upper bound on the ergodic capacity for systems using LMMSE channel estimation. In the following, we use the bounds derived in [20] for spatially correlated channels, which is an extension of the bounds derived in [17] for spatially i.i.d. channels. Denoting the input covariance matrix by  $\mathbf{Q} = E\{\mathbf{x}\mathbf{x}^\dagger\}$ , a lower bound on the ergodic capacity per channel use is given by

$$C_{\text{LB}} = E_{\hat{\mathbf{H}}} \left\{ \log_2 \left| \mathbf{I}_{N_t} + (1 + \text{tr}\{\mathbf{R}_{\hat{\mathbf{H}}}\mathbf{Q}\})^{-1} \hat{\mathbf{H}}^\dagger \hat{\mathbf{H}} \mathbf{Q} \right| \right\}. \quad (7)$$

and an upper bound on the ergodic capacity per channel use is given by

$$C_{\text{UB}} = C_{\text{LB}} + N_r E_{\mathbf{x}} \left\{ \log_2 \frac{1 + \text{tr}\{\mathbf{R}_{\tilde{\mathbf{H}}}\mathbf{Q}\}}{1 + \mathbf{x}^\dagger \mathbf{R}_{\tilde{\mathbf{H}}}\mathbf{x}} \right\}. \quad (8)$$

The authors in [17] studied the tightness of the bounds for i.i.d. channels with constant channel estimation errors. They observed that the gap between bounds is small for Gaussian inputs. Similarly, we numerically study the tightness of the above

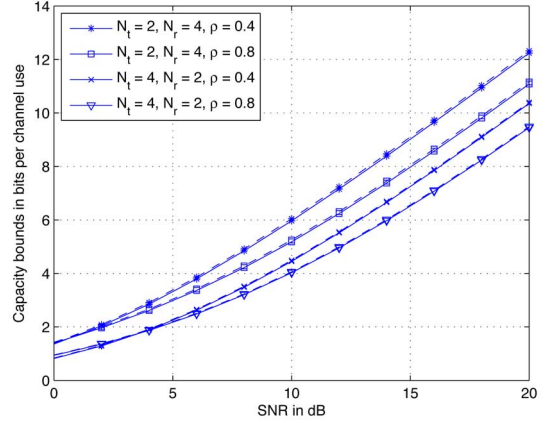


Fig. 2. Capacity bounds in (7) and (8) versus SNR  $\mathcal{P}$  for different correlation factor and antenna sizes.  $L_p = N_t$  and  $\mathcal{P}_p = \mathcal{P}_d = \mathcal{P}$  (i.e., fixed power transmission) is used. Dashed lines indicate the upper bound and solid lines indicate the lower bound.

two bounds for correlated channels with Gaussian inputs and LMMSE channel estimation. For numerical analysis, we choose the channel covariance matrix to be in the form of  $[\mathbf{R}_{\mathbf{H}}]_{ij} = \rho^{|i-j|}$ , where  $\rho$  is referred to as the spatial correlation factor [14]. We investigate the gap between the capacity bounds in (7) and (8) under different channel correlation conditions and antenna sizes. Examples are shown in Fig. 2 in which the upper bounds are plotted using dashed lines and the lower bounds are plotted using the solid lines. In general, we see that the gaps between the bounds are insignificant for any channel correlations. Therefore, the capacity lower bound per channel use in (7) is accurate enough to be used in our analysis assuming Gaussian inputs.

The average capacity lower bound per transmission block is therefore given by

$$\begin{aligned} \bar{C}_{\text{LB}} &= \frac{L_d}{L} C_{\text{LB}} \\ &= \frac{L_d}{L} E_{\hat{\mathbf{H}}} \left\{ \log_2 \left| \mathbf{I}_{N_t} + (1 + \text{tr}\{\mathbf{R}_{\hat{\mathbf{H}}}\mathbf{Q}\})^{-1} \hat{\mathbf{H}}^\dagger \hat{\mathbf{H}} \mathbf{Q} \right| \right\}. \end{aligned} \quad (9)$$

We will use “capacity lower bound” and “capacity” interchangeably throughout the rest of this paper. The average capacity lower bound in (9) will be used as the figure of merit to study the optimal transmission and resource allocation strategies in the following sections.

## III. NONFEEDBACK SYSTEMS

### A. Spatially i.i.d. Channels

The optimal pilot and data transmission scheme and optimal power allocation for nonfeedback systems with spatially i.i.d. channels were studied in [2] and [10]. For the sake of completeness, we briefly review their main results, which will also be referred to when providing new results for feedback systems in the next section. The optimal transmission strategy is to transmit orthogonal pilots and independent data among the transmit antennas with spatially equal power allocation to each

antenna during both pilot and data transmission. The optimal PSAM power factor  $\alpha^*$  is given by

$$\alpha^* = \begin{cases} \frac{\gamma - \sqrt{\gamma(\gamma-1)}}{\frac{1}{2}}, & \text{for } L_d > N_t \\ \frac{1}{2}, & \text{for } L_d = N_t \\ \frac{\gamma + \sqrt{\gamma(\gamma-1)}}{\gamma}, & \text{for } L_d < N_t \end{cases} \quad (10)$$

where  $\gamma = (N_t + \mathcal{P}L)/(\mathcal{P}L - \mathcal{P}LN_t/L_d)$ . With  $\alpha^*$ , the optimal training length is  $L_p^* = N_t$ .

### B. Spatially Correlated Channels

For nonfeedback systems with spatially correlated channels, it is difficult to find the optimal resource allocation and transmission strategies. Intuitively, the amount of training resource required should reduce as the channels become more correlated. Therefore, one good strategy may be to use the  $\alpha^*$  and  $L_p^*$  for i.i.d. channels so that sufficient training is ensured for all channel correlation conditions. Similarly, one may design the pilot and data transmission schemes to ensure a robust system performance for possibly correlated channels. Here, we define *robustness* to be the capability of achieving the best channel estimation and capacity performance for the least-favorable channel correlation. Note that the least or most favorable channel correlation condition depends on the choice of the training or data transmission strategy and does not necessarily imply i.i.d. or fully correlated channels.

*Theorem 1:* For nonfeedback systems, the transmission of orthogonal training sequences among the transmit antennas with spatially equal power allocation minimizes the channel estimation errors for the least-favorable channel correlation, i.e., using  $\mathbf{X}_p \mathbf{X}_p^\dagger = (\mathcal{P}_p L_p / N_t) \mathbf{I}_{N_t}$  is a robust training scheme.

*Proof:* see Appendix I.

*Theorem 2:* For nonfeedback systems, the transmission of i.i.d. data sequences among the transmit antennas with spatially equal power allocation, i.e.,  $\mathbf{Q} = (\mathcal{P}_d / N_t) \mathbf{I}_{N_t}$ , a) maximizes the capacity for the least-favorable channel correlation at sufficiently low SNR, and b) is the optimal transmission scheme at sufficiently high SNR.

*Proof:* see [20].

*Remark:* From Theorem 1 and Theorem 2, we see that the optimal transmission strategy for i.i.d. channels is also a robust choice for correlated channels in nonfeedback systems.

## IV. CHANNEL GAIN FEEDBACK SYSTEMS

CGF systems require the receiver to feed the channel estimates back to the transmitter once per transmission block. Once the transmitter receives the estimated channel gains, it performs spatial power adaptation accordingly. When noise is present in the feedback link, the capacity that can be achieved by adaptive transmission reduces as the noise increases. The capacity reduction due to corrupted channel gain estimates was studied in [22]. It was shown that the capacity reduction can increase quickly with the noise in the estimated channel gains. Therefore, a reliable feedback scheme which minimizes the noise in the estimated channel gains is essential for CGF systems. One solution is to use low rate feedback transmission with appropriate quantization scheme for the feedback information [23].

However, the design of digital feedback is beyond the scope of this paper. In this section, we assume that the feedback link is noiseless, which is reasonable for CGF systems with reliable feedback schemes.

We consider the channels to be spatially i.i.d. For a CGF system with correlated channels, it is fair to assume that the transmitter has the knowledge of the spatial channel correlations as well. In other words, a CGF system with correlated channels is effectively a hybrid CCF-CGF system. Therefore, we will briefly discuss this case in Section V-E after analyzing CCF-only systems. With the assumption of i.i.d. channels, the data transmission utilizes all the channels with equal probability. Hence, it is reasonable to have at least as many measurements as the number of channels for channel estimation, which implies that  $L_p \geq N_t$ . From [10], we know that the optimal training for i.i.d. channels consists of orthogonal pilots with equal power allocated to each antenna, i.e.,  $\mathbf{X}_p \mathbf{X}_p^\dagger = (\mathcal{P}_p L_p / N_t) \mathbf{I}_{N_t}$ .

For a given  $\mathcal{P}_d$ , the ergodic capacity lower bound per channel use in (7) can be rewritten as<sup>2</sup>

$$C_{\text{LB}} = E_{\hat{\mathbf{H}}_0} \left\{ \log_2 \left| \mathbf{I}_{N_t} + \frac{\sigma_{\hat{\mathbf{H}}}^2}{1 + \sigma_{\hat{\mathbf{H}}}^2 \mathcal{P}_d} \hat{\mathbf{H}}_0^\dagger \hat{\mathbf{H}}_0 \mathbf{Q} \right| \right\} \quad (11)$$

where  $\sigma_{\hat{\mathbf{H}}}^2 = (1 + \mathcal{P}_p L_p / N_t)^{-1}$  and  $\sigma_{\mathbf{H}}^2 = 1 - \sigma_{\hat{\mathbf{H}}}^2$ .

### A. CGF System With no Feedback Delay

Firstly, we study an ideal (delayless) scenario in which the transmitter receives the estimated channel gains at the start of the data transmission. The delayless case can be viewed as a reasonable approximation of the delayed case when  $d \ll L$ . Furthermore, the results in the delayless case will be used in Section IV-B for the delayed case. It was shown in [17] that the capacity is maximized when the matrix  $\mathbf{Q}$  has the same eigenvectors as  $\hat{\mathbf{H}}_0^\dagger \hat{\mathbf{H}}_0$ . Denoting the eigenvalues of  $\hat{\mathbf{H}}_0^\dagger \hat{\mathbf{H}}_0$  by  $\boldsymbol{\lambda} = [\lambda_1 \lambda_2 \cdots \lambda_{N_t}]^T$  sorted in descending order, the eigenvalues of  $\mathbf{Q}$  are found via the standard water-filling as

$$q_i = \left[ \eta - \left( \frac{\sigma_{\hat{\mathbf{H}}}^2}{1 + \sigma_{\hat{\mathbf{H}}}^2 \mathcal{P}_d} \lambda_i \right)^{-1} \right]^+ \quad \text{with} \quad \sum_{i=1}^{N_t} q_i = \mathcal{P}_d \quad (12)$$

where  $\eta$  represents the water level and  $[z]^+ \triangleq \max\{z, 0\}$ . Therefore, (11) can be rewritten as

$$C_{\text{LB}} = E_{\boldsymbol{\lambda}} \left\{ \sum_{i=1}^{N_t} \log_2 \left( 1 + \frac{\sigma_{\hat{\mathbf{H}}}^2}{1 + \sigma_{\hat{\mathbf{H}}}^2 \mathcal{P}_d} \lambda_i q_i \right) \right\} \quad (13)$$

$$= E_{\boldsymbol{\lambda}} \left\{ \sum_{i=1}^m \log_2 \left( \frac{\sigma_{\hat{\mathbf{H}}}^2}{1 + \sigma_{\hat{\mathbf{H}}}^2 \mathcal{P}_d} \lambda_i \eta \right) \right\} \quad (14)$$

$$= E_{\boldsymbol{\lambda}} \left\{ \sum_{i=1}^m \log_2 \left( \frac{\sigma_{\hat{\mathbf{H}}}^2 \mathcal{P}_d}{1 + \sigma_{\hat{\mathbf{H}}}^2 \mathcal{P}_d} + \sum_{i=1}^m \lambda_i^{-1} \right) + \sum_{i=1}^m \log_2 \frac{\lambda_i}{m} \right\} \quad (15)$$

<sup>2</sup>Note that the optimal pilot and data resource allocation should be found based on the *ergodic* capacity lower bound without having a prior knowledge of instantaneous channel realizations. Hence, the averaging over many transmission blocks is needed in (11).

where  $m$  denotes the number of nonzero  $q_i$ , and (15) is obtained by substituting  $\eta$  from (12) into (14). It should be noted that  $E_{\lambda}$  in (14) and (15) is the expectation over the  $m$  largest values in  $\lambda$ . Using (15), we now present the result for the optimal PSAM power factor  $\alpha^*$ .

*Theorem 3:* For delayless CGF systems with i.i.d. channels and the optimal pilot structure  $\mathbf{X}_p \mathbf{X}_p^\dagger = (\mathcal{P}_p L_p / N_t) \mathbf{I}_{N_t}$ , the optimal PSAM power factor  $\alpha^*$  is given by (10).

*Proof:* see Appendix II.

For nonfeedback systems, Hassibi *et al.* showed in [10] that the optimal training length  $L_p^*$  is equal to  $N_t$  when the optimal PSAM power factor  $\alpha^*$  given in (10) is used. In the next theorem and corollary, we show that  $L_p^* = N_t$  for any fixed value of  $\alpha$  as well as  $\alpha^*$  in delayless CGF systems.

*Theorem 4:* For delayless CGF systems with i.i.d. channels and the optimal pilot structure  $\mathbf{X}_p \mathbf{X}_p^\dagger = (\mathcal{P}_p L_p / N_t) \mathbf{I}_{N_t}$ , the optimal training length equals the number of transmit antennas for any given value of the PSAM power factor  $\alpha$ , i.e.,  $L_p^* = N_t$ .

*Proof:* see Appendix III.

*Remark:* In Theorem 4, the value of  $\alpha$  is required to be fixed. When  $\alpha$  is allowed to vary as  $L_p$  varies,  $L_p^*$  may or may not equal  $N_t$ . For example,  $\alpha$  varies according to  $\alpha = (L - L_p)/L$  when fixed power transmission ( $\mathcal{P}_d = \mathcal{P}_p = \mathcal{P}$ ) is used, in which case  $L_p^* = N_t$  does not hold in general.

*Corollary 1:* For delayless CGF systems with i.i.d. channels and the optimal pilot structure  $\mathbf{X}_p \mathbf{X}_p^\dagger = (\mathcal{P}_p L_p / N_t) \mathbf{I}_{N_t}$  as well as the optimal PSAM power factor  $\alpha^*$ , the optimal training length is given by  $L_p^* = N_t$ .

*Proof:* Among all possible values of  $\alpha \in [0, 1]$ , there exists an optimal  $\alpha$  that maximizes the capacity lower bound for any given  $L_p$ . We denote  $\alpha^*(k)$  to be the optimal  $\alpha$  for  $L_p = k$  and denote  $\bar{C}_{\text{LB}}(a, l)$  to be the capacity lower bound at  $\alpha = a$  and  $L_p = l$ . From the definition of  $\alpha^*(N_t)$  we have  $\bar{C}_{\text{LB}}(\alpha^*(N_t), N_t) \geq \bar{C}_{\text{LB}}(\alpha, N_t)$ ,  $\forall \alpha \in [0, 1]$ . From Theorem 4 we know that when  $k \geq N_t$ ,  $\bar{C}_{\text{LB}}(\alpha, N_t) \geq \bar{C}_{\text{LB}}(\alpha, k)$ ,  $\forall \alpha \in [0, 1]$ . Combining the two inequalities and choosing  $\alpha = \alpha^*(k)$ , we have  $\bar{C}_{\text{LB}}(\alpha^*(N_t), N_t) \geq \bar{C}_{\text{LB}}(\alpha^*(k), k)$ , that is, the capacity at  $L_p = N_t$  is greater than or equal to the capacity at  $L_p = k \geq N_t$  where the corresponding  $\alpha^*$  is used in both cases.  $\square$

Theorem 3 and Corollary 1 show that the optimal pilot design for delayless CGF systems coincide with that for nonfeedback systems in Section III-A. That is to say, one can use the same design to achieve optimal performance in both nonfeedback and delayless CGF systems.

## B. CGF System With Feedback Delay

For practical systems, a finite duration of  $d$  symbol periods is required before feedback comes into effect at the transmitter as shown in Fig. 1. Therefore, the transmitter has no knowledge about the channel during the first data subblock of  $d$  transmissions, which is equivalent to nonfeedback systems. From [2], we know that the transmitter should allocate equal power to each transmit antenna during the first data subblock (or the non-adaptive subblock). After receiving the estimated channel gains,

the transmitter performs spatial power water-filling similar to Section IV-A during the second data subblock (or the adaptive subblock) of length  $L_d - d$ . In order to optimize the PSAM power factor  $\alpha$ , we apply a two-stage optimization approach. Firstly, we optimize the data power division factor  $\phi$  for any given  $\mathcal{P}_d$  and  $L_d$ . Then, we optimize the PSAM power factor  $\alpha$ .

Using (3), the capacity lower bound per channel use in (11) becomes

$$\begin{aligned} C_{\text{LB}} &= E_{\lambda} \left\{ \beta \sum_{i=1}^{N_t} \log_2 \left( 1 + \frac{\sigma_{\mathbf{H}}^2 \lambda_i \mathcal{P}_d \phi}{1 + \frac{\sigma_{\mathbf{H}}^2 \mathcal{P}_d \phi}{\beta} N_t / \beta} \right) \right. \\ &\quad \left. + (1 - \beta) \sum_{i=1}^{N_t} \log_2 \left( 1 + \frac{\sigma_{\mathbf{H}}^2 \lambda_i q_i}{1 + \frac{\sigma_{\mathbf{H}}^2 \mathcal{P}_d (1 - \phi)}{(1 - \beta)}} \right) \right\} \quad (16) \\ &= E_{\lambda} \left\{ \beta \sum_{i=1}^{N_t} \log_2 \left( 1 + \frac{\sigma_{\mathbf{H}}^2 \lambda_i \mathcal{P}_d \phi}{1 + \frac{\sigma_{\mathbf{H}}^2 \mathcal{P}_d \phi}{\beta} N_t / \beta} \right) \right. \\ &\quad \left. + (1 - \beta) \sum_{i=1}^m \log_2 \left( \frac{\sigma_{\mathbf{H}}^2 \lambda_i \nu}{1 + \frac{\sigma_{\mathbf{H}}^2 \mathcal{P}_d (1 - \phi)}{(1 - \beta)}} \right) \right\} \quad (17) \end{aligned}$$

where the water-filling solution for  $q_i$  with water level  $\nu$  is given by

$$q_i = \left[ \nu - \left( \frac{\sigma_{\mathbf{H}}^2 \lambda_i}{1 + \frac{\sigma_{\mathbf{H}}^2 \mathcal{P}_d (1 - \phi)}{(1 - \beta)}} \right)^{-1} \right]^+ \quad \text{with} \quad \sum_{i=1}^{N_t} q_i = \frac{1 - \phi}{1 - \beta} \mathcal{P}_d. \quad (18)$$

*Lemma 1:* The capacity lower bound per channel use in (17) is concave on  $\phi \in [0, 1]$ .

*Proof:* From the property of water-filling solution [24], we know that  $q_i$  and  $\nu$  in (18) are continuous on  $\phi \in [0, 1]$ . As a result,  $C_{\text{LB}}$  in (17) is continuous on  $\phi \in [0, 1]$ . For any fixed  $m$ , one can show that  $d^2 C_{\text{LB}} / d\phi^2 < 0$  by directly computing the derivative using (17). This implies that  $dC_{\text{LB}} / d\phi$  given by

$$\begin{aligned} \frac{dC_{\text{LB}}}{d\phi} &= \frac{\mathcal{P}_d}{\ln 2} E_{\lambda} \left\{ \frac{\beta}{N_t} \sum_{i=1}^{N_t} \frac{\sigma_{\mathbf{H}}^2 \lambda_i - \frac{(\sigma_{\mathbf{H}}^2 \lambda_i \sigma_{\mathbf{H}}^2 \mathcal{P}_d \phi)}{(\beta + \sigma_{\mathbf{H}}^2 \mathcal{P}_d \phi)}}{\beta + \sigma_{\mathbf{H}}^2 \mathcal{P}_d \phi + \frac{\sigma_{\mathbf{H}}^2 \lambda_i \mathcal{P}_d \phi}{N_t}} \right. \\ &\quad \left. - \frac{1 + \sigma_{\mathbf{H}}^2 \sigma_{\mathbf{H}}^2 \sum_{i=1}^m \lambda_i^{-1}}{\nu} + \frac{m \sigma_{\mathbf{H}}^2 (1 - \beta)}{1 - \beta + \sigma_{\mathbf{H}}^2 \mathcal{P}_d (1 - \phi)} \right\} \quad (19) \end{aligned}$$

is a decreasing function of  $\phi$  for any given  $m$ . Furthermore, it can be shown that  $C_{\text{LB}}$  is differentiable on  $\phi \in [0, 1]$ , including the points where  $m$  changes its value. To obtain differentiability, we let  $\phi = \phi_0$  at which  $m$  changes between  $m_0$  and  $m_0 + 1$ . At this point, the water level is given by  $\nu_0 = \left( \sigma_{\mathbf{H}}^2 \lambda_{(m_0+1)} / [1 + \sigma_{\mathbf{H}}^2 \mathcal{P}_d (1 - \phi_0) / (1 - \beta)] \right)^{-1}$ . With  $\nu = \nu_0$ , one can show that the left and right derivatives of  $C_{\text{LB}}$  w.r.t.  $\phi$  equate, that is,  $dC_{\text{LB}} / d\phi (m = m_0) = dC_{\text{LB}} / d\phi (m = m_0 + 1)$  at  $\phi = \phi_0$ . Therefore, one can conclude that  $C_{\text{LB}}$  is differentiable and its derivative is decreasing on  $\phi \in [0, 1]$ , which implies concavity.  $\square$

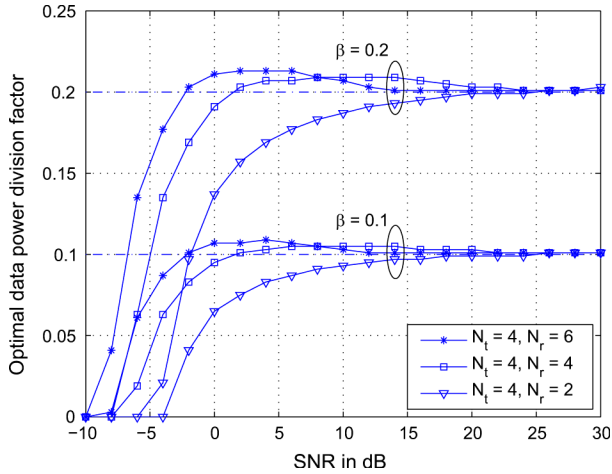


Fig. 3. Optimal data power division factor  $\phi^*$  versus SNR  $\mathcal{P}$  for different values of the delay factor  $\beta$  and antenna factors. In this example, a block length of  $L = 100$ , training length of  $L_p = N_t = 4$ , and PSAM power factor given in (10) are used.

With Lemma 1, the optimal data power division factor  $\phi^*$  can be found numerically. Using the Karush–Kuhn–Tucker (KKT) conditions [24], the result is given by

$$\begin{cases} \phi^* = 0, & \text{if } \frac{dC_{\text{LB}}}{d\phi}(\phi = 0) \leq 0 \\ \arg_{\phi} \frac{dC_{\text{LB}}}{d\phi} = 0, & \text{otherwise} \end{cases} \quad (20)$$

where  $dC_{\text{LB}}/d\phi$  was given in (19).

Fig. 3 shows the optimal data power division factor  $\phi^*$  given by (20) versus SNR  $\mathcal{P}$  for different delay factors  $\beta$ . It can be seen that  $\phi^*$  increases from 0 to  $\beta$  at low SNR. For moderate to high SNR,  $\phi^*$  stays above  $\beta$  and converges to  $\beta$  as  $\mathcal{P} \rightarrow \infty$ .<sup>3</sup> As we are concerned with practical design solutions, it is desirable to have a low complexity solution for  $\phi$  which still gives near optimal performance under practical transmission conditions. From Fig. 3, we see that  $\phi^*$  is usually close to  $\beta$  at practical SNR range, e.g.,  $\mathcal{P} > 0$  dB, and  $\phi^* \rightarrow \beta$  as  $\mathcal{P} \rightarrow \infty$ . Therefore,  $\phi = \beta$  is a simple solution which provides good system performance. We will also investigate the optimality of  $\phi = \beta$  using capacity results in Section IV-C.

From (3), we see that  $\phi = \beta$  is actually the simplest solution which allocates the same amount of power during each data transmission in both nonadaptive and adaptive subblocks, i.e.,  $\mathcal{P}_{d,1} = \mathcal{P}_{d,2} = \mathcal{P}_d$ . In addition, this simple solution does not require the knowledge of the feedback delay time. Furthermore, this choice of  $\phi$  leads to a simple closed-form solution for the optimal PSAM power factor  $\alpha^*$ , as well as the optimal training length  $L_p^*$  for delayed CGF system summarized in Corollary 2, which can be shown by combining the results in Theorem 3, Corollary 1 and those for the nonfeedback systems summarized in Section III-A.

**Corollary 2:** If  $\mathbf{X}_p \mathbf{X}_p^\dagger = (\mathcal{P}_p L_p / N_t) \mathbf{I}_{N_t}$  and  $\phi = \beta$ , the optimal PSAM power factor  $\alpha^*$  and the optimal training length  $L_p^*$  coincide with those in the delayless case given in Theorem 3 and Corollary 1.

<sup>3</sup> $\phi^*$  for the  $(N_t = 4, N_r = 2)$  system exceeds and converges back to  $\beta$  at a higher SNR, which is not shown in Fig. 3. This is because the use of spatial water-filling in data transmission gives a significant improvement in the capacity when  $N_t > N_r$ .

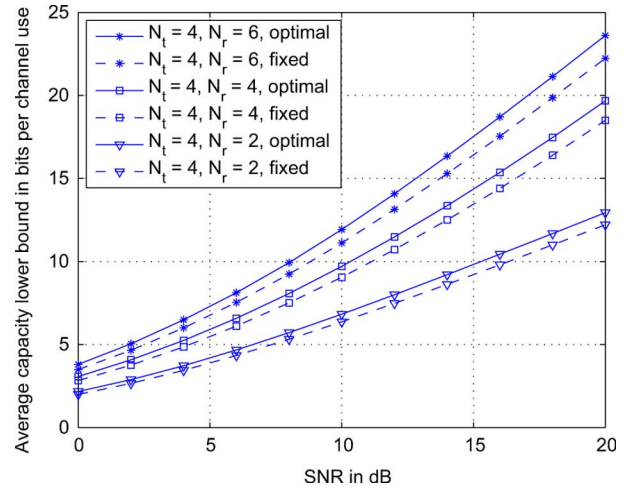


Fig. 4. Average capacity lower bound  $\bar{C}_{\text{LB}}$  versus SNR  $\mathcal{P}$  for delayless CGF systems ( $d = 0$  and  $\beta = 0$ ) with i.i.d. channels and different antenna sizes. Note that  $\bar{C}_{\text{LB}} = (L_d/L)C_{\text{LB}}$  where  $C_{\text{LB}}$  is given in (13). The block length is  $L = 100$ . Both optimal temporal power allocation to pilot and data as well as fixed power transmission are shown for comparison. For optimal temporal power allocation, the training length is  $L_p^* = 4$ ; while for fixed power transmission, the pilot length is optimized numerically.

### C. Numerical Results

Now, we present numerical results to illustrate the capacity gain from optimizing the PSAM power factor. Fig. 4 shows the average capacity lower bound  $\bar{C}_{\text{LB}}$  versus SNR  $\mathcal{P}$  for delayless CGF systems (i.e.,  $d = 0$  and  $\beta = 0$ ) with i.i.d. channels and different antenna sizes. The solid lines indicate systems using  $\alpha^*$  and  $L_p^*$  ( $L_p^* = 4$  in this case). The dashed lines indicate systems using equal temporal power allocation and  $L_p^*$  found numerically. Comparing the solid and dashed lines, we see that the capacity gain from optimal temporal power allocation is approximately 9% at 0 dB and 6% at 20 dB for all three systems. This range of capacity gain (5% to 10%) was also observed in [10] for nonfeedback systems which can be viewed as an extreme case of delayed CGF system with  $d = L_d$ . From the results for the extreme cases, i.e.,  $d = 0$  and  $d = L_d$ , we conclude that the capacity gain from optimizing the PSAM power factor is around 5% to 10% at practical SNR for delayed CGF systems with i.i.d. channels.

We now numerically verify the near optimality of  $\phi = \beta$ . Fig. 5 shows the average capacity lower bound  $\bar{C}_{\text{LB}}$  versus SNR  $\mathcal{P}$  for delayed CGF systems with i.i.d. channels and different antenna sizes. In this example, a transmission block of length  $L = 100$  consists of a training subblock of  $L_p = 4$  symbol periods, followed by a nonadaptive data subblock of  $d = 20$  symbol periods<sup>4</sup> and an adaptive data subblock of  $L_d - d = 76$  symbol periods. Therefore, the delay factor  $\beta = 0.208$ . The lines indicate the use of  $\phi = \beta$ , and the markers indicate optimal data power division found through numerical optimization using  $\bar{C}_{\text{LB}}$ . The values of  $\phi^*$  for SNR = 4, 10, and 16 dB are shown in the figure as well. We see that the capacity difference between the system using  $\phi = \beta$  and  $\phi = \phi^*$  is negligible. That is to say the use of temporal equal power transmission over the entire data block is near optimal for systems with channel estimation

<sup>4</sup>The delay length  $d$  takes into account the channel estimation and other processing time at the receiver and transmitter, as well as the time spent on the transmission of low-rate feedback.

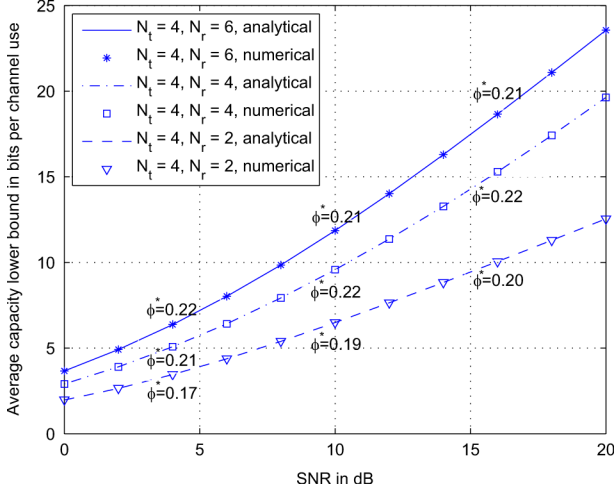


Fig. 5. Average capacity lower bound  $\bar{C}_{\text{LB}}$  versus SNR  $\mathcal{P}$  for delayed CGF systems with i.i.d. channels and different antenna sizes. Note that  $\bar{C}_{\text{LB}} = (L_d/L)C_{\text{LB}}$  where  $C_{\text{LB}}$  is given in (16). Within a block length of  $L = 100$ , the training length is  $L_p = 4$ , followed by a nonadaptive data transmission subblock of length  $d = 20$  and an adaptive data transmission subblock of length 76.  $\alpha$  in (10) is used. The lines indicate the use of  $\phi = \beta = 0.208$ , and the markers indicate optimal data power division factor found numerically.

errors. We have also confirmed that this trend is valid for a wide range of block lengths (results are omitted for brevity). These results validate the near optimality of  $\phi = \beta$ .

## V. CHANNEL COVARIANCE FEEDBACK SYSTEMS

As discussed in the previous section, CGF systems require frequent use of feedback due to the rapid change in the channel gains. This requires a significant amount of feedback overhead in the reverse link (from the receiver to the transmitter), which may cause a direct reduction in the overall information rate, especially when both the forward and the reverse links are operating at the same time, e.g., in cellular systems. Therefore, the CGF scheme may not be appropriate in fast fading environments where the block length is small. On the other hand, the statistics of the channel gains change much slower than the channel gains themselves. As a result, it is practical for the receiver to accurately measure the channel covariance matrix and feed it back to the transmitter at a much lower frequency with negligible feedback overhead and delay. Note that for completely i.i.d. channels, there is no need for CCF. In this section, we consider CCF systems with spatially correlated channels and investigate the optimal pilot and data transmission strategy, as well as the optimal power allocation. We assume that both the transmitter and the receiver have perfect knowledge of the channel spatial correlations.

### A. Proposed Transmission Scheme

Most studies on the optimal pilot design for correlated channels assume that the training length is at least as large as the number of transmit antennas, i.e.,  $L_p \geq N_t$  [13]–[15]. Intuitively, the amount of training resource can be significantly reduced as the channels become more correlated. As an extreme case where the channels are fully correlated, only one pilot

transmission is needed to train all the channels. Therefore, we relax this assumption by considering  $L_p \geq 1$ . It was shown in [14] that the optimal training strategy is to train along the eigenvectors of the channel covariance matrix with training power being water-filled according to the eigenvalues of the channel covariance matrix, provided that  $L_p \geq N_t$ . In the case where  $L_p < N_t$ , only  $L_p$  eigenchannels can be trained. Therefore, we propose that only the  $L_p$  strongest eigenchannels are to be trained when  $L_p < N_t$ .

We perform eigenvalue decomposition on  $\mathbf{R}_{\mathbf{H}}$  as  $\mathbf{R}_{\mathbf{H}} = \mathbf{U}\mathbf{G}\mathbf{U}^\dagger$ , and let the eigenvalues of  $\mathbf{R}_{\mathbf{H}}$  be sorted in descending order in  $\mathbf{g} = [g_1 \ g_2 \ \dots \ g_{N_t}]^T$ . The optimal training sequence which minimizes the channel estimation errors (i.e.,  $\text{tr}\{\mathbf{R}_{\hat{\mathbf{H}}}\}$ ) has the property that the eigenvalue decomposition of  $\mathbf{X}_p\mathbf{X}_p^\dagger$  is given by  $\mathbf{X}_p\mathbf{X}_p^\dagger = \mathbf{U}\mathbf{P}\mathbf{U}^\dagger$  [14], where  $\mathbf{P}$  is a diagonal matrix. We denote the diagonal entries of  $\mathbf{P}$  by  $\mathbf{p} = [p_1 \ p_2 \ \dots \ p_{N_t}]^T$ . Let  $n = \min\{L_p, N_t\}$ , then the first  $n$  entries in  $\mathbf{p}$  are given by

$$p_i = [\mu - g_i^{-1}]^+, \text{ with } \sum_{i=1}^n p_i = \mathcal{P}_p L_p \quad (21)$$

where  $\mu$  is the water level. All the remaining entries in  $\mathbf{p}$  (if any) are set to zero. It is not difficult to show that this choice of  $\mathbf{X}_p\mathbf{X}_p^\dagger$  is optimal. Note that in practice, the transmitter can ensure that the number of nonzero  $p_i$  equals  $n$  by reducing  $L_p$  when needed.

For data transmission, it was shown that the optimal strategy is to transmit along the eigenvectors of  $\mathbf{R}_{\mathbf{H}}$  under the perfect channel estimation [4]–[6]. With the proposed training sequence, it is easy to show from (5) and (6) that the eigenvectors of  $\mathbf{R}_{\hat{\mathbf{H}}}$  and  $\mathbf{R}_{\hat{\mathbf{H}}}$  are the same as those of  $\mathbf{R}_{\mathbf{H}}$ . Therefore, in the case of channel estimation errors, a reasonable strategy is to transmit data along the eigenvectors of  $\mathbf{R}_{\hat{\mathbf{H}}}$  which coincide with the eigenvectors of  $\mathbf{R}_{\mathbf{H}}$ . From (5) and (6), the eigenvalue decomposition of  $\mathbf{R}_{\hat{\mathbf{H}}}$  is given by  $\mathbf{R}_{\hat{\mathbf{H}}} = \mathbf{U}\hat{\mathbf{G}}\mathbf{U}^\dagger$ , and we set  $\mathbf{Q} = \mathbf{U}\hat{\mathbf{Q}}\mathbf{U}^\dagger$  where  $\hat{\mathbf{Q}}$  is a diagonal matrix with entries denoted by  $\hat{q}_i, \forall i = 1, \dots, N_t$ .

To the best of the authors' knowledge, there is no closed-form solution to the optimal spatial power allocation even with perfect channel estimation [4]–[6]. Instead, the optimal  $\hat{q}_i$  can be found via numerical methods. As we are concerned with practical design solutions, it is desirable to have a low complexity solution for  $\hat{q}_i$  which still gives near optimal performance. Following the proposed training scheme, we propose a simple strategy which transmits data through the  $n$  trained eigenchannels with equal power. That is

$$\hat{q}_i = \begin{cases} \frac{\mathcal{P}_d}{n}, & i = 1, \dots, n \\ 0, & \text{otherwise.} \end{cases} \quad (22)$$

It is easy to see that only the  $n$  trained eigenchannels should be used for data transmission, since the capacity is zero for untrained eigenchannels. We will investigate the optimality of (22) in Section V-D.

For the proposed training and data transmission scheme, the capacity lower bound per channel use in (7) reduces to

$$C_{\text{LB}} = E_{\hat{\mathbf{H}}_0} \left\{ \log_2 \left| \mathbf{I}_{N_t} + \hat{\mathbf{H}}_0^\dagger \hat{\mathbf{H}}_0 \hat{\mathbf{G}} \hat{\mathbf{Q}} (1 + \mu^{-1} \mathcal{P}_d)^{-1} \right| \right\} \quad (23)$$

where the nonzero (diagonal) entries of  $\hat{\mathbf{G}}$  are given by  $\hat{g}_i = g_i - \mu^{-1}$ ,  $\forall i = 1, \dots, n$ , which is derived from (5), (6), and (21).

### B. Optimal Training Resource Allocation

Now, we investigate the optimal training length  $L_p^*$  as well as the optimal PSAM power factor  $\alpha^*$  using the capacity lower bound given in (23). The results are summarized in the following theorems.

*Theorem 5:* For CCF systems in PSAM schemes with the transmission strategy proposed in Section V-A, the optimal training length is at most as large as the number of transmit antennas for any given value of the PSAM power factor  $\alpha$ , i.e.,  $L_p^* \leq N_t$ .

*Proof:* Assuming  $L_p \geq N_t$ , we have  $\hat{\mathbf{Q}} = (\mathcal{P}_d/N_t)\mathbf{I}_{N_t}$ . We let  $\boldsymbol{\theta} = [\theta_1\theta_2\cdots\theta_r]^T$  be the nonzero eigenvalues of  $\hat{\mathbf{H}}_0^\dagger\hat{\mathbf{H}}_0\hat{\mathbf{G}}/N_t$ , where  $r = \text{rank}\{\hat{\mathbf{H}}_0^\dagger\hat{\mathbf{H}}_0\hat{\mathbf{G}}\} = \min\{N_t, N_r\}$ .  $C_{\text{LB}}$  in (23) is reduced to  $C_{\text{LB}} = E_{\boldsymbol{\theta}}\{\sum_{i=1}^r \log_2(1 + \hat{\rho}_{\text{eff},i})\}$ , where  $\hat{\rho}_{\text{eff},i} = \theta_i\mathcal{P}_d(1 + \mu^{-1}\mathcal{P}_d)^{-1}$ . Note that  $\theta_i > 0 \forall i$  and they are independent of  $L_d$  for any fixed  $\alpha$ . Following the proof of Theorem 4 in Appendix III, one can show that  $d\bar{C}_{\text{LB}}/dL_d > 0$  for any given  $\alpha$ . Therefore,  $\bar{C}_{\text{LB}}$  reaches its maximum when  $L_d$  is maximized, which implies that  $L_p^* = N_t$  under the constraint of  $L_p \geq N_t$ . Hence, in general we have  $L_p^* \leq N_t$ .  $\square$

*Remark:* Theorem 5 implies that  $L_p^* \leq N_t$  when  $\alpha^*$  for the proposed transmission strategy is used (with proof similar to Corollary 1). Hence, we will only consider  $L_p \leq N_t$  in the analysis on  $\alpha^*$ . Although the optimal training length needs to be found numerically, the computational complexity of optimizing  $L_p$  is low due to the fact that  $L_p$  only takes integer values ranging from 1 to  $N_t$ .

*Theorem 6:* For CCF systems in PSAM schemes with the transmission strategy proposed in Section V-A, the optimal PSAM power factor  $\alpha^*$  is given by (10) with  $\gamma = L_d/(L_d - L_p)$ , provided that  $\mathcal{P}L \gg \sum_{i=1}^{L_p} g_i^{-1}$ .

*Proof:* see Appendix IV.

*Remark:* It is noted that  $\gamma$  in the optimal solution in Theorem 6 is essentially the same as the one given in Section III-A when  $\mathcal{P}L \gg 1$ . The condition of  $\mathcal{P}L \gg \sum_{i=1}^{L_p} g_i^{-1}$  can be easily satisfied when the block length is not too small or the SNR is moderate to high (i.e.,  $\mathcal{P}L \gg 1$ ), and the spatial correlation between any trained channels is not close to 1. Therefore, the result in Theorem 6 applies to many practical scenarios. It is important to note that the optimal PSAM power factor  $\alpha^*$  given in Theorem 6 does not depend on the channel spatial correlation, provided the condition above is met. In other words, this unique design is suitable for a relatively wide range of channel spatial correlations.

### C. A Special Case: Beamforming

From Theorem 5 we know that  $L_p^* \leq N_t$ . When only the strongest eigenchannel is used, i.e.,  $L_p = 1$ , this scheme is called beamforming, which may be the optimal strategy for highly correlated channels. Furthermore, the use of beamforming significantly reduces the complexity of the system as it allows the use of well-established scalar codec technology

and only requires the knowledge of the strongest eigenchannel (not the complete channel statistics) [5]. For beamforming transmission, the capacity lower bound in (23) reduces to

$$\begin{aligned} C_{\text{LB}} &= E_{\hat{\mathbf{h}}_0} \left\{ \log_2 \left( 1 + \hat{\mathbf{h}}_0^\dagger \hat{\mathbf{h}}_0 \frac{(g_{\text{max}} - \mu^{-1})\mathcal{P}_d}{1 + \mu^{-1}\mathcal{P}_d} \right) \right\} \\ &= E_{\hat{\mathbf{h}}_0} \left\{ \log_2 \left( 1 + \hat{\mathbf{h}}_0^\dagger \hat{\mathbf{h}}_0 \frac{g_{\text{max}}\mathcal{P}_p\mathcal{P}_d}{g_{\text{max}}^{-1} + \mathcal{P}_p + \mathcal{P}_d} \right) \right\} \end{aligned} \quad (24)$$

where  $\hat{\mathbf{h}}_0$  is a  $N_r \times 1$  vector with i.i.d. ZMCSG and unit variance entries,  $g_{\text{max}}$  is the largest eigenvalue in  $\mathbf{g}$ , and  $\mu = \mathcal{P}_p + g_{\text{max}}^{-1}$  which can be found by letting  $L_p = 1$  in (21).

*Theorem 7:* For CCF systems in PSAM schemes with beamforming, the optimal PSAM power factor  $\alpha^*$  is given in (10) with  $\gamma = (1 + g_{\text{max}}\mathcal{P}L)/(g_{\text{max}}\mathcal{P}L(L-2)/(L-1))$ .

*Proof:* The proof can be obtained by letting  $L_p = 1$  and  $g_i = g_{\text{max}}$  in the proof of Theorem 6.  $\square$

*Remark:* It can be shown for the beamforming case that  $d\alpha^*/dg_{\text{max}} > 0$ . Therefore, the optimal PSAM power factor  $\alpha^*$  increases as the channel spatial correlation increases, that is to say, more power should be allocated to data transmission when the channels become more correlated. When  $\mathcal{P}L \gg 1$ ,  $\gamma$  reduces to  $(L-1)/(L-2)$ , hence  $\alpha^*$  does not depend on the channel correlation.

### D. Numerical Results

For numerical analysis, we choose the channel covariance matrix to be in the form of  $[\mathbf{R}_{\mathbf{H}}]_{ij} = \rho^{|i-j|}$ , where  $\rho$  is referred to as the spatial correlation factor [14]. Firstly we validate the solution to the optimal PSAM power factor given in Theorem 6 and Theorem 7. Fig. 6 shows the optimal PSAM power factor  $\alpha^*$  found numerically versus the channel correlation factor  $\rho$  for CCF  $4 \times 4$  systems with a block length of  $L = 20$  and SNR of 10 dB. We see that  $\alpha^*$  remains constant before the correlation factor gets close to 1 for  $L_p > 1$ , and this value of  $\alpha^*$  is the same as the analytical value given by Theorem 6. For the beamforming case where  $L_p = 1$ , we see that  $\alpha^*$  does not depend on the channel correlation, which agrees with our earlier observation from Theorem 7. Similar to CGF systems, we have also compared the capacity achieved using  $\alpha^*$  and that using fixed power transmission over pilot and data, and the same trend is observed (results are omitted for brevity), that is, capacity gain from optimizing PSAM power factor is around 5% to 10% at practical SNR.

In our proposed transmission scheme for CCF systems, spatially equal power allocation is used for data transmission. Here we illustrate the near optimality of this simple scheme in Fig. 7, which shows the average capacity lower bound  $\bar{C}_{\text{LB}}$  versus channel correlation factor  $\rho$  for CCF  $2 \times 2$  systems. We compute the capacity achieved using  $L_p = 1$ , and  $L_p = 2$  with spatially equal power allocation for data transmission (solid line) and optimal power allocation found numerically (dashed line) for a block length of  $L = 20$ .<sup>5</sup> We also indicate the critical  $\rho$  at which  $L_p^*$  changes from 2 to 1 in Fig. 7. It is clear that the capacity loss from spatially optimal power allocation to

<sup>5</sup>We see that the capacity increases with channel spatial correlation in the case of beamforming, while it is not monotonic for  $L_p = 2$ . These observations were explained in [20] using Schur-convexity of capacity in the channel correlation.



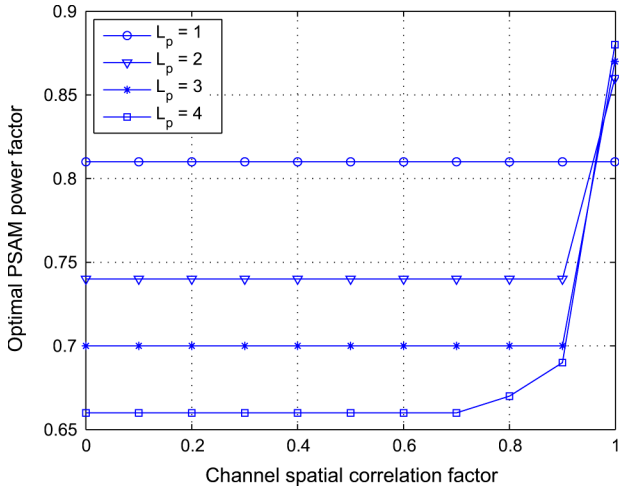


Fig. 6. Optimal PSAM power factor  $\alpha^*$  versus channel spatial correlation factor  $\rho$  for CCF  $4 \times 4$  systems with a block length of  $L = 20$  and SNR = 10 dB. All values of  $\alpha^*$  are found numerically.

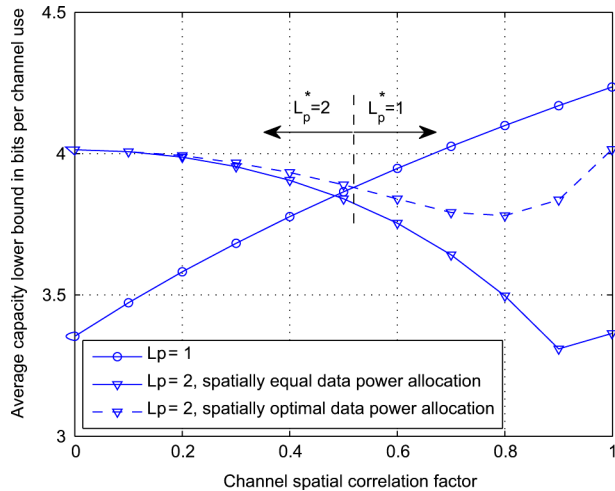


Fig. 7. Average capacity lower bound  $\bar{C}_{LB}$  in (9) versus channel spatial correlation factor  $\rho$  for CCF  $2 \times 2$  systems with a block length of  $L = 20$  and SNR of 10 dB. Training length of  $L_p = 1$  and  $L_p = 2$  are shown. For  $L_p = 2$ , both spatial equal data power allocation (dashed lines) and optimal data power allocation found numerically (solid lines) are shown.

spatially equal power allocation increases as  $\rho$  increases. At the critical  $\rho$ , this capacity loss is less than 1.5%. We also studied the results for different values of block lengths and the same trend was found (results are omitted for brevity). These results imply that our proposed transmission scheme is very close to optimal provided that the training length is optimized.

Fig. 8 shows the average capacity lower bound  $\bar{C}_{LB}$  versus the channel correlation factor  $\rho$  for CCF  $4 \times 4$  systems with a block length of  $L = 20$  and SNR of 10 dB. The optimal PSAM power factor  $\alpha^*$  shown in Fig. 6 is used in the capacity computation. We also include  $\bar{C}_{LB}$  for  $L_p = 5$  in Fig. 6 with  $\alpha^*$  found numerically. It is clear that  $\bar{C}_{LB}$  for  $L_p = 5$  is always smaller than  $\bar{C}_{LB}$  for  $L_p^* \leq 4$ , which agrees with Theorem 5. Comparing the capacity with different training lengths, we see that  $L_p^*$  decreases as the channel becomes more correlated. More importantly, the capacity gain from optimizing the training length according to the channel spatial correlation can be significant.

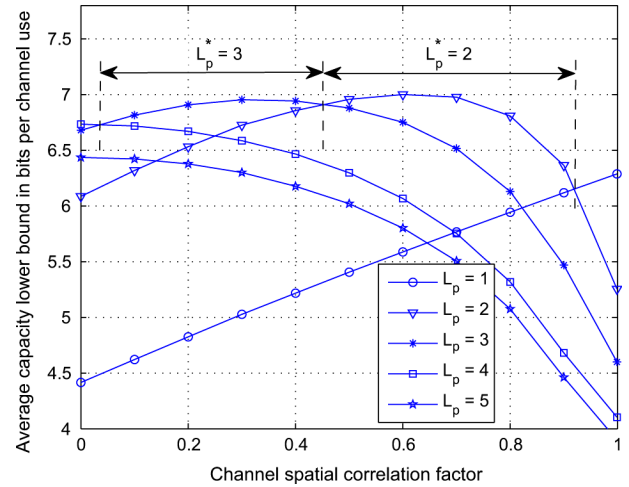


Fig. 8. Average capacity lower bound  $\bar{C}_{LB}$  versus channel spatial correlation factor  $\rho$  for CCF  $4 \times 4$  systems with a block length of  $L = 20$  and SNR = 10 dB. Note that  $\bar{C}_{LB} = (L_d/L)C_{LB}$  where  $C_{LB}$  is given in (23). The optimal PSAM power factor  $\alpha^*$  is used in all results.

For example, the capacity at  $\rho = 0.5$  using  $L_p = 4$  (which is optimal for i.i.d. channels) is approximately 6.3 bits per channel use, while the capacity at  $\rho = 0.5$  using  $L_p^* = 2$  is around 7 bits per channel use, that is to say, optimizing training length results in a capacity improvement of 11% at  $\rho = 0.5$ . Moreover, the capacity improvement increases as channel correlation increases. The same trends are found for different values of block lengths, although the capacity improvement by optimizing the training length reduces as the block length increases (results are omitted for brevity). Therefore, it is important to numerically optimize the training length for correlated channels at small to moderate block lengths.

Furthermore, one can record the range of  $\rho$  for each value of  $L_p^*$  from Fig. 8, and observe the value of  $\alpha^*$  in the corresponding range of  $\rho$  in Fig. 6. It can be seen that within the range of  $\rho$  where a given  $L_p$  is optimal, the value of  $\alpha^*$  for the given  $L_p$  is a constant given by Theorem 6 provided that  $\mathcal{P}L \gg 1$ . That is to say, the condition in Theorem 6 (i.e.,  $\mathcal{P}L \gg \sum_{i=1}^{L_p} g_i^{-1}$ ) can be simplified to  $\mathcal{P}L \gg 1$  provided that the training length is optimized. To summarize, our numerical results show that optimizing the training length can significantly improve the capacity, and the simple transmission scheme proposed in Section V-A gives near optimal performance.

### E. Hybrid CCF-CGF Systems

For systems with correlated channels, one can utilize both CCF and CGF to increase the capacity. We are currently investigating the optimal transmission and power allocation strategy for the hybrid CCF-CGF system. For spatially correlated channels, we expect that the optimal training follows a water-filling solution according to the channel covariance matrix, and the optimal data transmission follows a water-filling solution according to the estimated channel gains. The two different water-filling solutions make the problem of optimizing the PSAM power factor mathematically intractable. However, we expect that the solution given in Theorem 6 would still be a good choice for the hybrid system. Furthermore, we expect that the optimal

TABLE I  
 SUMMARY OF DESIGN GUIDELINES

System	Channel	Design Guidelines	Reference
Non-feedback	i.i.d.	<ul style="list-style-type: none"> <li>• Transmit orthogonal pilots among antennas with spatially equal power.</li> <li>• Transmit independent data among antennas with spatially equal power.</li> <li>• The optimal PSAM power factor <math>\alpha^*</math> is given by (10) with <math>\gamma = \frac{N_t + \mathcal{P}L}{\mathcal{P}L(1 - N_t/L_d)}</math>.</li> <li>• The optimal training length <math>L_p^*</math> equals the number of transmit antennas <math>N_t</math>.</li> </ul>	[2, 10]
	correlated	<ul style="list-style-type: none"> <li>• Use the designs for i.i.d. channels as a robust choice.</li> </ul>	Sec. III-B
CGF	i.i.d.	<ul style="list-style-type: none"> <li>• Transmit orthogonal pilots with spatially equal power.</li> <li>• Transmit independent data with spatially equal power in data sub-block 1 and spatial power water-filling in data sub-block 2 (see Fig. 1).</li> <li>• Distribute equal power per transmission throughout data sub-blocks 1 and 2.</li> <li>• <math>\alpha^*</math> and <math>L_p^*</math> for non-feedback system are (near) optimal for (delayed) CGF system.</li> </ul>	Sec. IV
CCF	correlated	<ul style="list-style-type: none"> <li>• For a given <math>L_p</math> (<math>L_p \leq N_t</math>), transmit pilots along the <math>L_p</math> strongest eigen-channels with spatial power water-filling according to (21).</li> <li>• Transmit data along the <math>L_p</math> trained eigen-channels with spatially equal power.</li> <li>• <math>\alpha^*</math> is given by (10) with <math>\gamma = \frac{L_d}{L_d - L_p}</math>, provided that <math>\mathcal{P}L \gg \sum_{i=1}^{L_p} g_i^{-1}</math>.</li> <li>• <math>L_p^* \leq N_t</math> and should be numerically optimized.</li> <li>• For beamforming (<i>i.e.</i>, <math>L_p = 1</math>), <math>\alpha^*</math> is given by (10) with <math>\gamma = \frac{1 + g_{\max} \mathcal{P}L}{g_{\max} \mathcal{P}L(L-2)/(L-1)}</math>.</li> </ul>	Sec. V

training length given in Theorem 5 would still hold for the hybrid system.

## VI. SUMMARY OF RESULTS

In this paper, we have studied block fading MIMO systems with feedback in PSAM transmission schemes. Two typical feedback systems are considered, namely the channel gain feedback and the channel covariance feedback systems. Using an accurate capacity lower bound as the figure of merit, we have provided the solutions for the (near) optimal power allocation to training and data transmission as well as the optimal training length. Table I summarizes the design guidelines for both nonfeedback systems and feedback systems.

### APPENDIX I PROOF OF THEOREM 1

This is a max-min problem where the MSE of the channel estimates is to be minimized by  $\mathbf{X}_p \mathbf{X}_p^\dagger$  and to be maximized by  $\mathbf{R}_H$ . With the constraint of  $\text{tr}\{\mathbf{X}_p \mathbf{X}_p^\dagger\} = \mathcal{P}_p L_p$ , we need to show that  $\inf_{\mathbf{X}_p \mathbf{X}_p^\dagger} \sup_{\mathbf{R}_H} \text{tr}\{\mathbf{R}_{\hat{\mathbf{H}}}\}$  is achieved by orthogonal pilot sequence with equal power allocated among the transmit antennas, *i.e.*,  $\mathbf{X}_p \mathbf{X}_p^\dagger = (\mathcal{P}_p L_p / N_t) \mathbf{I}_{N_t}$ , assuming  $L_p \geq N_t$ .

From (5) we see that

$$\sup_{\mathbf{R}_H} \text{tr}\{\mathbf{R}_{\hat{\mathbf{H}}}\} \geq \text{tr}\{(\mathbf{I}_{N_t} + \mathbf{X}_p \mathbf{X}_p^\dagger)^{-1}\} = \sum_{i=1}^{N_t} (1 + p_i)^{-1} \quad (25)$$

where  $\mathbf{p} = [p_1 \ p_2 \ \dots \ p_{N_t}]^T$  are the eigenvalues of  $\mathbf{X}_p \mathbf{X}_p^\dagger$ . Since the sum of a convex function of  $p_i$  is Schur-convex in  $\mathbf{p}$  [25], we conclude that (25) is Schur-convex in  $\mathbf{p}$ . Consequently, we have

$$\sup_{\mathbf{R}_H} \text{tr}\{\mathbf{R}_{\hat{\mathbf{H}}}\} \geq \sum_{i=1}^{N_t} (1 + p_i)^{-1} \geq \sum_{i=1}^{N_t} \left(1 + \frac{\mathcal{P}_p L_p}{N_t}\right)^{-1} \quad (26)$$

where  $\sum_{i=1}^{N_t} (1 + p_i)^{-1} = \sum_{i=1}^{N_t} \left(1 + \mathcal{P}_p L_p / N_t\right)^{-1}$  when  $\mathbf{X}_p \mathbf{X}_p^\dagger = (\mathcal{P}_p L_p / N_t) \mathbf{I}_{N_t}$ . Note that (26) holds for any  $\mathbf{X}_p \mathbf{X}_p^\dagger$ . On the other hand

$$\begin{aligned} & \inf_{\mathbf{X}_p \mathbf{X}_p^\dagger} \sup_{\mathbf{R}_H} \text{tr}\{\mathbf{R}_{\hat{\mathbf{H}}}\} \\ & \leq \sup_{\mathbf{R}_H} \text{tr}\left\{\left(\mathbf{R}_H^{-1} + \frac{\mathcal{P}_p L_p}{N_t} \mathbf{I}_{N_t}\right)^{-1}\right\} \\ & = \sup_{\mathbf{R}_H} \sum_{i=1}^{N_t} \left(g_i^{-1} + \frac{\mathcal{P}_p L_p}{N_t}\right)^{-1} \\ & \leq \sum_{i=1}^{N_t} \left(1 + \frac{\mathcal{P}_p L_p}{N_t}\right)^{-1} \end{aligned} \quad (27)$$

where (27) is obtained using the Schur-concavity of  $\sum_{i=1}^{N_t} \left(g_i^{-1} + \mathcal{P}_p L_p / N_t\right)^{-1}$  in  $\mathbf{g}$ . The equality in (27) holds when  $\mathbf{R}_H = \mathbf{I}_{N_t}$ . From (26) and (27), we conclude that

$$\inf_{\mathbf{X}_p \mathbf{X}_p^\dagger} \sup_{\mathbf{R}_H} \text{tr}\{\mathbf{R}_{\hat{\mathbf{H}}}\} = \sum_{i=1}^{N_t} \left(1 + \frac{\mathcal{P}_p L_p}{N_t}\right)^{-1}$$

which can be obtained by choosing  $\mathbf{X}_p \mathbf{X}_p^\dagger = (\mathcal{P}_p L_p / N_t) \mathbf{I}_{N_t}$ .  $\square$

### APPENDIX II PROOF OF THEOREM 3

To prove Theorem 3, we begin with the following set of results.

- R1. From the property of water-filling solution [24],  $m$  is discrete and nondecreasing on  $\alpha \in [0, 1]$  as the number of nonzero  $q_i$  in (12) cannot decrease as the data transmission power increases.
- R2. With  $\sigma_{\hat{\mathbf{H}}}^2 = \left(1 + \mathcal{P}_p L_p / N_t\right)^{-1}$ ,  $\sigma_{\hat{\mathbf{H}}}^2 = 1 - \sigma_{\hat{\mathbf{H}}}^2$  and (2), it can be shown that  $\rho_{\text{eff}} \triangleq \sigma_{\hat{\mathbf{H}}}^2 \mathcal{P}_d / (1 + \sigma_{\hat{\mathbf{H}}}^2 \mathcal{P}_d)$  is a concave function of  $\alpha \in [0, 1]$ .
- R3. For any fixed  $m$ , we see from (15) that  $C_{\text{LB}}$  is maximized when  $\rho_{\text{eff}}$  reaches its maximum.

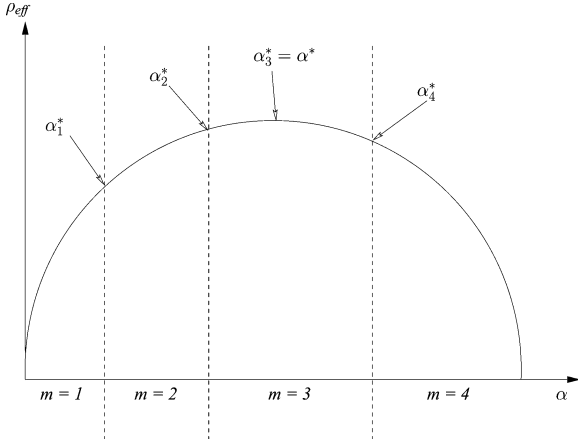


Fig. 9. Sketch example of  $\rho_{\text{eff}}$  versus  $\alpha$ . The vertical dashed lines indicates the values of  $\alpha$  at which  $m$  changes its value.  $\alpha_1^*$ ,  $\alpha_2^*$ ,  $\alpha_3^*$  and  $\alpha_4^*$  indicate the local optimal values of  $\alpha$  which gives local maximal  $\rho_{\text{eff}}$ .

- R4. From the property of water-filling solution [24], we know that  $q_i$  in (12) is continuous on  $\mathcal{P}_d$  and hence, is continuous on  $\alpha \in [0, 1]$ .

In Fig. 9, we show a sketch plot of  $\rho_{\text{eff}}$  versus  $\alpha$ . This figure helps to visualize the following proof. Basically, the main objective here is to show that the optimal  $\alpha$  is not affected by the number of active eigenchannels.

From R1, we can divide all values of  $\alpha$  into a finite number of regions in which  $m$  is constant.

From R2, there exist a unique local optimal point of  $\alpha$  in each of the aforementioned regions, denoted by  $\alpha_1^*$ ,  $\alpha_2^*$ ,  $\alpha_3^*$  and so on. And these optimal points of  $\alpha$  occur on the boundaries of these regions except for at most one of them which is the global optimal point of  $\alpha$ , denoted by  $\alpha^*$ . (Note that  $\alpha^*$  may occur on the boundary of some region as well.)

From R3, we know that the local optimal point of  $\alpha$  for  $\rho_{\text{eff}}$  is also the local optimal point for  $C_{\text{LB}}$  in each of the aforementioned regions. That is to say,  $\alpha_1^*$ ,  $\alpha_2^*$ , etc. maximize  $C_{\text{LB}}$  in the corresponding regions where  $m$  is fixed.

From R4, the continuity of  $q_i$  on  $\alpha \in [0, 1]$  implies the continuity of  $C_{\text{LB}}$  in (13) on  $\alpha \in [0, 1]$ . This implies that  $C_{\text{LB}}$  is continuous across the boundaries of two different regions of  $\alpha$ . For example in Fig. 9,  $C_{\text{LB}}(m=1) = C_{\text{LB}}(m=2)$  at  $\alpha = \alpha_1^*$  and  $C_{\text{LB}}(m=2) = C_{\text{LB}}(m=3)$  at  $\alpha = \alpha_2^*$ , etc.

Therefore, we can show that  $C_{\text{LB}}$  at  $\alpha^*$  is larger than  $C_{\text{LB}}$  at  $\alpha_1^*$ ,  $\alpha_2^*$ , etc. That is to say, the global optimal point of  $\alpha$  for  $\rho_{\text{eff}}$  is also the global optimal point for  $C_{\text{LB}}$ . Therefore, the objective function for optimizing  $\alpha$  is reduced from  $C_{\text{LB}}$  to  $\rho_{\text{eff}}$ . It is noted that the objective function  $\rho_{\text{eff}}$  is the same as that in nonfeedback systems given in [10]. Therefore, the solution of  $\alpha^*$  coincides with the solution for nonfeedback systems given in (10).  $\square$

### APPENDIX III PROOF OF THEOREM 4

We aim to show that  $d\bar{C}_{\text{LB}}/dL_d > 0$  (treating  $L_d$  as a positive real-valued variable) for any given  $\alpha$ . First of all, it can be shown that  $\bar{C}_{\text{LB}}$  is continuous on  $L_d$  regardless of the value of

$m$ , which implies that there is no discontinuity in  $\bar{C}_{\text{LB}}$ . Therefore, it suffices to show that  $d\bar{C}_{\text{LB}}/dL_d > 0$  for any fixed  $m$  and  $\alpha$ .

We let  $\rho_{\text{eff}} = \sigma_{\mathbf{H}}^2 \mathcal{P}_d / (1 + \sigma_{\mathbf{H}}^2 \mathcal{P}_d) = \sigma_{\mathbf{H}}^2 \alpha \mathcal{P} L / (L_d + \sigma_{\mathbf{H}}^2 \alpha \mathcal{P} L)$ ,  $y = \sum_{i=1}^m \ln(\lambda_i/m)$ , and  $z = \sum_{i=1}^m \lambda_i^{-1}$ . Then the average capacity lower bound in (9) can be rewritten using (15) as

$$\bar{C}_{\text{LB}} = \frac{L_d}{L} \frac{1}{\ln 2} E_{\lambda} \{ m \ln(\rho_{\text{eff}} + z) + y \}.$$

For any given  $\alpha$ , we know from (2) that  $\mathcal{P}_p L_p$  is constant. Therefore,  $\sigma_{\mathbf{H}}^2 = (1 + \mathcal{P}_p L_p / N_t)^{-1}$  and  $\sigma_{\mathbf{H}}^2 = 1 - \sigma_{\mathbf{H}}^2$  are also constant. Differentiating  $\bar{C}_{\text{LB}}$  w.r.t.  $L_d$  for any fixed  $m$  gives

$$\begin{aligned} \frac{d\bar{C}_{\text{LB}}}{dL_d} &= \frac{1}{\ln 2} \frac{m}{L} \left( E_{\lambda} \left\{ \ln(\rho_{\text{eff}} + z) + \frac{L_d}{\rho_{\text{eff}} + z} \frac{d\rho_{\text{eff}}}{dL_d} + \frac{y}{m} \right\} \right) \end{aligned} \quad (28)$$

where

$$\frac{d\rho_{\text{eff}}}{dL_d} = -\frac{\sigma_{\mathbf{H}}^2 \alpha \mathcal{P} L}{(L_d + \sigma_{\mathbf{H}}^2 \alpha \mathcal{P} L)^2} = -\frac{\rho_{\text{eff}}}{L_d + \sigma_{\mathbf{H}}^2 \alpha \mathcal{P} L}. \quad (29)$$

Substituting (29) into (28), we get

$$\begin{aligned} \frac{d\bar{C}_{\text{LB}}}{dL_d} &= \frac{1}{\ln 2} \frac{m}{L} E_{\lambda} \left\{ \ln(\rho_{\text{eff}} + z) - \frac{\rho_{\text{eff}}}{\rho_{\text{eff}} + z} \frac{L_d}{L_d + \sigma_{\mathbf{H}}^2 \alpha \mathcal{P} L} + \frac{y}{m} \right\}. \end{aligned}$$

Since  $L_d / (L_d + \sigma_{\mathbf{H}}^2 \alpha \mathcal{P} L) < 1$ , it suffices to show that

$$E_{\lambda} \left\{ \ln(\rho_{\text{eff}} + z) - \frac{\rho_{\text{eff}}}{\rho_{\text{eff}} + z} + \frac{y}{m} \right\} \geq 0. \quad (30)$$

Furthermore, one can show that

$$\frac{d}{d\rho_{\text{eff}}} E_{\lambda} \left\{ \ln(\rho_{\text{eff}} + z) - \frac{\rho_{\text{eff}}}{\rho_{\text{eff}} + z} + \frac{y}{m} \right\} = \frac{\rho_{\text{eff}}}{(\rho_{\text{eff}} + z)^2} \geq 0$$

for any fixed  $m$ . Since  $E_{\lambda} \left\{ \ln(\rho_{\text{eff}} + z) - \rho_{\text{eff}} / (\rho_{\text{eff}} + z) + y/m \right\}$  is an increasing function of  $\rho_{\text{eff}}$ , we only need to show (30) holds at  $\rho_{\text{eff}} = 0$ , that is

$$\begin{aligned} E_{\lambda} \left\{ \ln z + \frac{y}{m} \right\} &= E_{\lambda} \left\{ \ln \sum_{i=1}^m \lambda_i^{-1} + \frac{1}{m} \sum_{i=1}^m \ln \frac{\lambda_i}{m} \right\} \\ &\geq E_{\lambda} \left\{ \frac{1}{m} \sum_{i=1}^m \ln \left( \frac{\lambda_i}{m} \right)^{-1} + \frac{1}{m} \sum_{i=1}^m \ln \frac{\lambda_i}{m} \right\} \\ &= E_{\lambda} \left\{ \frac{1}{m} \sum_{i=1}^m \ln \left( \frac{\lambda_i}{m} \right)^{-1} \frac{\lambda_i}{m} \right\} = 0 \end{aligned}$$

where we have used the concavity of  $\ln(\cdot)$ . Therefore, we conclude that  $d\bar{C}_{\text{LB}}/dL_d > 0$  for any given  $\alpha$ , which implies that  $L_d$  should be kept maximum. Hence,  $L_p$  should be kept minimum. With the assumption of  $L_p \geq N_t$  the optimal training length is given by  $L_p^* = N_t$ .  $\square$

APPENDIX IV  
PROOF OF THEOREM 6

For any positive definite matrix  $\mathbf{A}$ ,  $\log_2 |\mathbf{A}|$  is increasing in  $\mathbf{A}$  [25]. Also, for any positive semi-definite matrix  $\mathbf{B}$ ,  $\mathbf{I} + \hat{\mathbf{H}}_0^\dagger \hat{\mathbf{H}}_0 \mathbf{B}$  is a positive definite matrix [2]. Since  $\hat{\mathbf{G}}\hat{\mathbf{Q}}(1 + \mu^{-1}\mathcal{P}_d)^{-1}$  is a positive semi-definite matrix, the capacity lower bound in (23) is maximized when the diagonal entries of  $\hat{\mathbf{G}}\hat{\mathbf{Q}}(1 + \mu^{-1}\mathcal{P}_d)^{-1}$  are maximized.

The  $i$ th nonzero diagonal entry of  $\hat{\mathbf{G}}\hat{\mathbf{Q}}(1 + \mu^{-1}\mathcal{P}_d)^{-1}$  is given by

$$\begin{aligned} \rho_{\text{eff},i} &= \frac{(g_i - \mu^{-1})\mathcal{P}_d}{(1 + \mu^{-1}\mathcal{P}_d)L_p} \\ &= \frac{g_i \mathcal{P}_p \mathcal{P}_d + \mathcal{P}_d(y - g_i^{-1})}{L_p \mathcal{P}_p + \mathcal{P}_d + y} \end{aligned} \quad (32)$$

where we have used (21) and let  $y = \mu - \mathcal{P}_p = (1/L_p) \sum_{i=1}^{L_p} g_i^{-1}$ . Substituting  $\alpha$  from (2) into (32) with some algebraic manipulation, we get

$$\rho_{\text{eff},i} = \frac{g_i \mathcal{P}L}{L_p(L_d - L_p)} \frac{\alpha(1 - \alpha) + \alpha \frac{L_p}{\mathcal{P}L}(y - g_i^{-1})}{-\alpha + \frac{\mathcal{P}L + L_p y}{\mathcal{P}L(1 - L_d/L_p)}}. \quad (33)$$

Here we consider the case where  $L_d > L_p$  and omit the cases  $L_d = L_p$  and  $L_d < L_p$  which can be handled similarly. It can be shown that  $\rho_{\text{eff},i}$  in (33) is concave in  $\alpha \in (0, 1)$ . Therefore, the optimal  $\alpha$  occurs at  $d\rho_{\text{eff},i}/d\alpha = 0$ , which is the root to  $\alpha^2 - 2\alpha\gamma + \gamma + \gamma z = 0$ , where  $\gamma = (\mathcal{P}L + L_p y)/(\mathcal{P}L - \mathcal{P}L L_p/L_d)$  and  $z = L_p(y - g_i^{-1})/(\mathcal{P}L)$ . It is clear that  $\alpha$  depends on  $g_i$  through  $z$ . Therefore, there is no unique  $\alpha$  which maximizes all  $\rho_{\text{eff},i}$ . However, this dependence disappears when  $\mathcal{P}L \gg L_p y = \sum_{i=1}^{L_p} g_i^{-1}$ . Under this condition, one can show that  $\gamma \approx L_d/(L_d - L_p)$  and  $z \approx 0$ . And there exists a unique solution of  $\alpha^*$  which maximizes all the diagonal entries of  $\hat{\mathbf{G}}\hat{\mathbf{Q}}(1 + \mu^{-1}\mathcal{P}_d)^{-1}$ , given by

$$\alpha^* = \gamma - \sqrt{\gamma(\gamma - 1)}, \text{ where } \gamma = \frac{L_d}{L_d - L_p}.$$

□

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, no. 3, pp. 311–335, Mar. 1998.  
 [2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.  
 [3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.  
 [4] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2632–2639, Sep. 2001.  
 [5] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, Jul. 2004.  
 [6] E. A. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1543–1553, Sep. 2004.

[7] J. C. Roh and B. D. Rao, "Multiple antenna channels with partial channel state information at the transmitter," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 677–688, Mar. 2004.  
 [8] J. Li and Q. T. Zhang, "Transmitter optimization for correlated MISO fading channels with generic mean and covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3312–3317, Sep. 2008.  
 [9] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 686–693, Nov. 1991.  
 [10] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, pp. 951–963, Apr. 2003.  
 [11] V. Pohl, P. H. Nguyen, V. Jungnickel, and C. Helmolt, "Continuous flat-fading MIMO channels: Achievable rate and optimal length of the training and data phases," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1889–1900, Jul. 2005.  
 [12] X. Zhou, T. A. Lamahewa, P. Sadeghi, and S. Durrani, "Designing PSAM schemes: How optimal are SISO pilot parameters for spatially correlated SIMO?," in *Proc. IEEE Int. Symp. Personal, Indoor, Mobile Radio Commun. (PIMRC)*, Cannes, France, Sep. 2008.  
 [13] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.  
 [14] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.  
 [15] J. Pang, J. Li, L. Zhao, and Z. Lu, "Optimal training sequences for MIMO channel estimation with spatial correlation," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Baltimore, MD, Sep. 2007, pp. 651–655.  
 [16] J. Pang, J. Li, and L. Zhao, "Optimal training length for MIMO systems with transmit antenna correlation," in *Proc. IEEE Wireless Commun. Networking Conf. (WCNC)*, Las Vegas, NV, Mar. 2008, pp. 425–429.  
 [17] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2203–2214, May 2006.  
 [18] C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, pp. 637–650, Mar. 2002.  
 [19] E. A. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multi-antenna systems with different types of channel state information," *IEEE Trans. Signal Process.*, vol. 52, no. 12, pp. 3440–3453, Dec. 2004.  
 [20] X. Zhou, T. A. Lamahewa, P. Sadeghi, and S. Durrani, "Capacity of MIMO systems: Impact of spatial correlation with channel estimation errors," in *Proc. IEEE Int. Conf. Commun. Systems (ICCS)*, Guangzhou, China, Nov. 2008, pp. 817–822.  
 [21] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.  
 [22] E. Baccarelli, M. Biagi, and C. Pelizzoni, "On the information throughput and optimized power allocation for MIMO wireless systems with imperfect channel estimation," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2335–2347, Jul. 2005.  
 [23] D. J. Love, R. W. Heath, Jr, W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?," *IEEE Commun. Mag.*, pp. 54–59, Oct. 2004.  
 [24] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.  
 [25] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.



**Xiangyun Zhou** (S'08) received the B.E. (Hons.) degree in electronics and telecommunications engineering from the Australian National University, Australia, in 2007. He is currently working toward the Ph.D. degree in engineering and information technology at the Research School of Information Sciences and Engineering, Australian National University.

His research interests are in signal processing for wireless communications, including MIMO systems, ad hoc networks, relay and cooperative networks.



**Parastoo Sadeghi** (S'02–M'06–SM'07) received the B.E. and M.E. degrees in electrical engineering from Sharif University of Technology, Tehran, Iran, in 1995 and 1997, respectively, and the Ph.D. degree in electrical engineering from The University of New South Wales, Sydney, Australia, in 2006.

From 1997 to 2002, she worked as a Research Engineer and then as a Senior Research Engineer at Iran Communication Industries (ICI), Tehran, Iran, and at Deqx (Clarity Eq), Sydney, Australia. She is currently a Fellow at the Research School of Information Sciences and Engineering, the Australian National University, Canberra. She has visited various research institutes, including the Institute for Communications Engineering, Technical University of Munich, Germany, from April to June 2008. Her research interests include channel modeling and information theory of wireless communications systems.

Dr. Sadeghi is a Chief Investigator in two Australian Research Council Discovery Projects. In 2003 and 2005, she received two IEEE Region 10 student paper awards for her research in the information theory of time-varying fading channels.



**Tharaka A. Lamahewa** (M'06) received the B.E. degree from the University of Adelaide, South Australia, in 2000 and the Ph.D. from the Australian National University, Canberra, in 2007.

He worked two years for Motorola Electronics Pvt Ltd., Singapore, as a Software Design Engineer. From 2006 to 2007, he worked as an Algorithm Design Engineer at Nanoradio AB, Melbourne, Australia. He is now with the Department of Information Engineering, Research School of Information Sciences and Engineering at the Australian National

University. His research interests include information theory of time-varying fading channels, cooperative diversity, space–time coding, and space–time–frequency channel modeling.



**Salman Durrani** (S'00–M'05) received the B.Sc. (first-class hon.) degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2000 and the Ph.D. degree in electrical engineering from the University of Queensland, Brisbane, Australia, in 2004.

Since March 2005, he has been a Lecturer in the Department of Engineering, College of Engineering and Computer Science, The Australian National University, Canberra.

Dr. Durrani was awarded a University Gold Medal during his undergraduate studies. He was a recipient of an International Postgraduate Research Scholarship (IPRS), funded by the Australian government, for the duration of his Ph.D. His current research interests include wireless and digital communications, channel modelling and channel estimation, beamforming in wireless ad hoc networks and MIMO, and smart-antenna systems. He has 30 publications to date in refereed international journals and conferences. He is a Member of the Institution of Engineers, Australia.