

Optimizing Training-based Transmission for Correlated MIMO Systems with Hybrid Feedback

Xiangyun Zhou, Tharaka A. Lamahewa, Parastoo Sadeghi and Salman Durrani

College of Engineering and Computer Science, The Australian National University, Canberra 0200 ACT, Australia

Email: {xiangyun.zhou, tharaka.lamahewa, parastoo.sadeghi, salman.durrani}@anu.edu.au

Abstract—In this paper, we consider multiple-input multiple-output (MIMO) communication systems with combined channel covariance feedback (CCF) and channel gain feedback (CGF), hereafter called hybrid CCF-CGF systems. Using an ergodic capacity lower bound as the figure of merit, we investigate the optimal training and data transmission strategies as well as the optimal transmit resource allocation. We prove that the optimal structure for data transmission follows a water-filling solution according to the estimated channel gains, rotated and truncated into the trained eigen-directions. We analytically find the range of the optimal training length. Through numerical evaluations we also show that a closed-form solution of the training power allocation achieves near optimal performance. Finally, we show that the capacity of hybrid CCF-CGF systems can be significantly increased by adding extra transmit antennas without increasing the training resources or feedback overhead.

I. INTRODUCTION

The use of multiple transmit and receive antennas dramatically increases the information capacity in wireless communications systems, especially when the channel state information (CSI) is perfectly known at the receiver [1, 2]. If partial or complete CSI is available at the transmitter, the information capacity can be further increased by adopting adaptive transmission strategies [2–6]. The transmitter CSI is often obtained in the form of channel covariance feedback (CCF) or channel gain feedback (CGF).

However, the assumption of perfect receiver CSI never holds in practical wireless communications due to the time-varying nature of the fading channel and the noise present in the receiver. Hence, channel estimation errors need to be taken into account in the communication system design. Pilot-symbol-assisted modulation (PSAM) is a widely-used technique to facilitate channel estimation, which periodically inserts training symbols or pilots into data transmission blocks [7]. Although the insertion of pilots provide reliable channel estimation, it directly reduces the amount of available resource for data transmission. Hence, it is important to optimize the resource allocation (in terms of time and power) to pilot transmission that maximizes the information capacity.

Using an ergodic capacity lower bound, the optimal pilot structure and resource allocation was investigated for non-feedback [8] and CGF-only systems [9] with independent and identically distributed (i.i.d.) channels. For spatially correlated channels with CCF only, the optimal pilot structure was studied in [9–11], and the optimal training resource allocation was also investigated in [9]. When both CCF and CGF are

available, the system can be considered as a hybrid CCF-CGF system. However, the optimal design for hybrid CCF-CGF systems is still an open research problem partially due to lack of an analytical solution for the optimal data transmission that takes both CCF and CGF into account. In addition, the complicated nature of the capacity bound expression, due to hybrid CCF-CGF, makes it nearly impossible to directly solve the problem of optimizing the training resource allocation in closed form.

In this paper, we study the optimal pilot and data transmission strategy as well as the training resource allocation for hybrid CCF-CGF systems. Utilizing the optimal pilot structure and the property of linear minimum mean square error (LMMSE) estimator, we are able to derive a closed-form solution for the optimal data transmission. We also analytically study the possible range of the optimal training length, *i.e.*, the duration of pilot transmission. In addition, we investigate the performance of a closed-form suboptimal solution for the training power allocation. The main contributions of this paper are summarized as follows.

- In Section III-B, we prove that the optimal data transmission for hybrid CCF-CGF systems is given by a water-filling solution according to the estimated channel gains, rotated and truncated into the trained eigen-directions.
- In Section IV-A, we prove that the optimal training length for hybrid CCF-CGF systems is at most as large as the number of transmit antennas for any given PSAM power factor¹ and hence, for the optimal PSAM power factor.
- In Section IV-B, we consider a closed-form solution of PSAM power factor and numerically show that this solution achieves near optimal performance for hybrid CCF-CGF systems.
- In Section V, we show that one can significantly increase the capacity of hybrid CCF-CGF systems by adding extra antennas at the transmitter without requiring additional training resources or feedback overhead.

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

¹We refer to the ratio of power allocation between pilot and data transmission as the PSAM power factor, which will be defined in Section II-A.

II. SYSTEM MODEL

We consider a flat fading MIMO system with N_t transmit antennas and N_r receive antennas with input-output relationship given by $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where \mathbf{x} and \mathbf{y} are the transmitted and received symbol vectors, respectively. \mathbf{H} is the $N_r \times N_t$ channel gain matrix. Both the noise vector \mathbf{n} and the channel matrix \mathbf{H} have zero-mean circularly symmetric complex Gaussian (ZMCSCG) entries with unit variance. The channel gains remain constant over a block of L symbol periods and change to independent realizations in the next block. We consider spatial correlation among the transmit antennas only, which 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 its eigenvalue decomposition (EVD) as $\mathbf{R}_H = \mathbf{U}\mathbf{G}\mathbf{U}^\dagger$, where the diagonal matrix $\mathbf{G} = \text{diag}\{g_1, g_2, \dots, g_{N_t}\}$. We assume that the eigenvalues are sorted in descending order, i.e., $g_i \geq g_j, i < j$. Each of the eigenvalues of \mathbf{R}_H indicates the strength of the corresponding eigen-direction.

A. PSAM Transmission and Feedback Schemes

Each transmission block of L symbol periods consists of L_p pilot transmissions and $L_d (= L - L_p)$ data transmissions. The receiver performs pilot-assisted channel estimation and feeds the estimated channel gains back to the transmitter once per block to allow adaptive data transmission. Similar to [12], we assume that the feedback link is delayless and error-free. In addition, the receiver also measures the channel spatial correlation. Since correlation is a long-term statistic which changes much more slowly than the channel gains, it can be accurately measured at the receiver and fed back to the transmitter at a much lower frequency. Therefore, we assume that both the transmitter and receiver have perfect knowledge of the channel spatial correlation.

The total transmission energy per block is given by $\mathcal{P}LT_s$, where \mathcal{P} is the average power or SNR 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 α . 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.

$$\mathcal{P}_p = (1 - \alpha) \frac{\mathcal{P}L}{L_p}, \quad \text{and} \quad \mathcal{P}_d = \alpha \frac{\mathcal{P}L}{L_d}. \quad (1)$$

B. Channel Estimation

In each transmission block, the receiver performs channel estimation during the pilot transmission. We denote the channel estimate and estimation error as $\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$, respectively. Using the LMMSE estimator, the channel estimates are given by [11]

$$\hat{\mathbf{H}} = \mathbf{Y}(\mathbf{X}_p^\dagger \mathbf{R}_H \mathbf{X}_p + \mathbf{I}_{L_p})^{-1} \mathbf{X}_p^\dagger \mathbf{R}_H, \quad (2)$$

where \mathbf{Y} is the $N_r \times L_p$ 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 [11]

$$\mathbf{R}_{\tilde{\mathbf{H}}} = E\{\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}}\}/N_r = (\mathbf{R}_H^{-1} + \mathbf{X}_p \mathbf{X}_p^\dagger)^{-1}, \quad (3)$$

The orthogonality property of LMMSE estimator implies that

$$\mathbf{R}_{\hat{\mathbf{H}}} = E\{\hat{\mathbf{H}}^\dagger \hat{\mathbf{H}}\}/N_r = \mathbf{R}_H - \mathbf{R}_{\tilde{\mathbf{H}}}. \quad (4)$$

C. A Capacity Lower Bound

The exact capacity expression under imperfect receiver CSI is still unavailable. Instead, we consider a lower bound on the ergodic capacity for systems using LMMSE channel estimation given by [9]

$$C_{\text{LB}} = \frac{L_d}{L} E_{\hat{\mathbf{H}}} \left\{ \log_2 \left| \mathbf{I}_{N_t} + (1 + \text{tr}\{\mathbf{R}_{\tilde{\mathbf{H}}}\mathbf{Q}\})^{-1} \hat{\mathbf{H}}^\dagger \hat{\mathbf{H}} \mathbf{Q} \right| \right\}, \quad (5)$$

where $\mathbf{Q} = E\{\mathbf{x}\mathbf{x}^\dagger\}$ is the input covariance matrix, which is positive semi-definite. As shown in [9], the bound in (5) is tight and therefore an accurate approximation of the true capacity. Hence, we will use it as the figure of merit in the rest of this paper.

III. OPTIMAL TRANSMISSION SCHEMES

In this section, we study the optimal spatial structures of both the pilots and data, which maximize the capacity lower bound. At the beginning of each transmission block, the transmitter only knows the statistical properties of the channels. Therefore, the pilot structure design is based on the knowledge of the channel covariance matrix. During the data transmission, the transmitter has received the estimated channel gains fed back from the receiver. Hence, the data structure design is based on the knowledge of both the channel covariance matrix and the estimated channel gains.

A. Optimal Pilot Transmission

The optimal pilot design should minimize the channel estimation errors by utilizing the knowledge of \mathbf{R}_H . As a result, one can use the optimal pilot structure for CCF systems in [9] outlined as follows: The optimal pilot structure is given by $\mathbf{X}_p \mathbf{X}_p^\dagger = \mathbf{U}\mathbf{P}\mathbf{U}^\dagger$, where $\mathbf{P} = \text{diag}\{p_1, p_2, \dots, p_{N_t}\}$. That is to say, the eigenvectors of $\mathbf{X}_p \mathbf{X}_p^\dagger$ are the same as those of \mathbf{R}_H . Let $n = \min\{L_p, N_t\}$, then the first n diagonal entries in \mathbf{P} are obtained from a water-filling solution as

$$p_i = [\mu - g_i^{-1}]^+, \quad i = 1, \dots, n, \quad \text{with} \quad \sum_{i=1}^n p_i = \mathcal{P}_p L_p, \quad (6)$$

where μ is the pilot water level and $[z]^+ = \max\{z, 0\}$. All the remaining p_i (if any) are set to zero. Note that we have not imposed any assumptions on the range of L_p . In practice, the transmitter can ensure the number of non-zero p_i equals n by reducing L_p when needed and hence, the water-filling solution in (6) need not include the $[\cdot]^+$ operator.

B. Optimal Data Transmission

The data transmission strategy characterized by \mathbf{Q} shall be chosen to maximize the following objective function for given \mathbf{R}_H and $\hat{\mathbf{H}}$ as follows.

$$\max_{\mathbf{Q}: \text{tr}\{\mathbf{Q}\} \leq \mathcal{P}_d} \log_2 \left| \mathbf{I}_{N_t} + (1 + \text{tr}\{\mathbf{R}_H \mathbf{Q}\})^{-1} \hat{\mathbf{H}}^\dagger \hat{\mathbf{H}} \mathbf{Q} \right|. \quad (7)$$

Since \mathbf{Q} appears in two different places in (7), it is generally difficult to find a closed-form solution to optimal \mathbf{Q} and it is also not clear whether $\text{tr}\{\mathbf{Q}\} = \mathcal{P}_d$ maximizes the objective function. The constraint set of the above optimization problem is a convex cone and hence, the solution can be found using iterative algorithms such as cone programming [13]. However, in the case where the LMMSE channel estimation and the optimal pilot structure outlined in Section III-A are used, a closed-form solution for optimal \mathbf{Q} can be found as follows.

We denote the first n columns of $\hat{\mathbf{H}}\mathbf{U}$ as \mathbf{F} which is an $N_r \times n$ matrix, and let the EVD of $\mathbf{F}^\dagger \mathbf{F}$ be $\mathbf{F}^\dagger \mathbf{F} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^\dagger$, where the diagonal matrix $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$. The following theorem gives the optimal data transmission strategy.

Theorem 1: The optimal data transmission structure can be written as $\mathbf{Q} = \mathbf{U} \hat{\mathbf{Q}} \mathbf{U}^\dagger$ with

$$\hat{\mathbf{Q}} = \begin{pmatrix} \hat{\mathbf{Q}}_a & \mathbf{0}_{n \times (N_t - n)} \\ \mathbf{0}_{(N_t - n) \times n} & \mathbf{0}_{(N_t - n) \times (N_t - n)} \end{pmatrix}, \quad (8)$$

where $\hat{\mathbf{Q}}_a$ is an $n \times n$ matrix and $\mathbf{0}$ is the zero matrix. The optimal $\hat{\mathbf{Q}}_a$ satisfies $\text{tr}\{\hat{\mathbf{Q}}_a\} = \mathcal{P}_d$ and is given by $\hat{\mathbf{Q}}_a = \mathbf{V} \mathbf{A} \mathbf{V}^\dagger$, where $\mathbf{A} = \text{diag}\{a_1, a_2, \dots, a_{L_p}\}$ with

$$a_i = \left[\eta - \left(\frac{\lambda_i}{1 + \mu^{-1} \mathcal{P}_d} \right)^{-1} \right]^+, \quad \sum_{i=1}^n a_i = \mathcal{P}_d, \quad (9)$$

and η represents the data water level.

Proof: See Appendix A.

Remark: It can be seen from (8) that the optimal data transmission concentrates the transmit power into the n trained eigen-directions. $\text{tr}\{\hat{\mathbf{Q}}_a\} = \text{tr}\{\mathbf{Q}\} = \mathcal{P}_d$ implies that all available power for data transmission should be used to achieve the capacity. As shown in (9), the optimal $\hat{\mathbf{Q}}_a$ follows a water-filling solution according to the estimated channel gains that are rotated by \mathbf{U} and truncated into the n trained eigen-directions. These results agree with the intuitions obtained from CGF-only and CCF-only systems which are now rigorously proved in *Theorem 1*. Furthermore, *Theorem 1* contains an important message, that is, only the channel gains along the trained eigen-directions have effects on the optimal data transmission design. As a result, the receiver should right rotate $\hat{\mathbf{H}}$ by \mathbf{U} , and only feed back the first n columns of $\hat{\mathbf{H}}\mathbf{U}$, i.e., \mathbf{F} . Note that $n = L_p$ when $L_p < N_t$. This implies that only $N_r \times L_p$ out of $N_r \times N_t$ channel gains need to be fed back when $L_p < N_t$. Therefore, for a fixed training length, the amount of feedback overhead does not increase by adding extra transmit antennas.

With the optimal pilot structure given in Section III-A, and the optimal data structure given in *Theorem 1*, the average

capacity lower bound in (5) can be rewritten as (derivation is similar to that in Appendix A)

$$C_{\text{LB}} = \frac{L_d}{L} E_\lambda \left\{ \sum_{i=1}^m \log_2 \left(\frac{\mathcal{P}_d}{1 + \mu^{-1} \mathcal{P}_d} + \sum_{i=1}^m \frac{1}{\lambda_i} \right) + \sum_{i=1}^m \log_2 \frac{\lambda_i}{m} \right\}, \quad (10)$$

where m denotes the number of non-zero a_i in (9).

IV. OPTIMAL TRAINING RESOURCE ALLOCATION

Using the optimal transmission schemes in Section III, we now study the optimal training length and the optimal PSAM power factor, denoted by L_p^* and α^* , respectively, which maximize the average capacity lower bound in (10).

A. Optimal Training Length

First of all, we present results for L_p^* in the following theorem and corollary.

Theorem 2: With the optimal transmission strategy described in Section III, the optimal training length is at most as large as the number of transmit antennas for any given value of the PSAM power factor α , i.e., $L_p^ \leq N_t$.*

Proof: See Appendix B.

Corollary 1: With the optimal transmission strategy described in Section III as well as the optimal PSAM power factor α^ , the optimal training length is given by $L_p^* \leq N_t$.*

Proof: We denote $\alpha^*(k)$ to be the optimal α which maximizes the capacity lower bound for $L_p = k$ and denote $C_{\text{LB}}(r, l)$ to be the capacity lower bound at $\alpha = r$ and $L_p = l$. From the definition of $\alpha^*(N_t)$ we have $C_{\text{LB}}(\alpha^*(N_t), N_t) \geq C_{\text{LB}}(\alpha, N_t) \forall \alpha \in [0, 1]$. From the proof of *Theorem 2*, we know that $C_{\text{LB}}(\alpha, N_t) \geq C_{\text{LB}}(\alpha, k) \forall \alpha \in [0, 1], k \geq N_t$. Combining the two inequalities and choosing $\alpha = \alpha^*(k)$, we have $C_{\text{LB}}(\alpha^*(N_t), N_t) \geq C_{\text{LB}}(\alpha^*(k), k) \forall k \geq N_t$, 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 α^* is used in both cases. Hence, $L_p^* \leq N_t$ holds when α^* is used. \square

Remark: In general, the optimal training length needs to be found numerically when $L_p^* < N_t$. Nevertheless, *Theorem 2* and *Corollary 1* suggest that the computational complexity of numerically optimizing L_p is low due to the fact that L_p only takes integer values ranging from 1 to N_t .

B. Optimal Training Power Allocation

According to *Corollary 1*, we only consider $L_p \leq N_t$ for investigating the optimal PSAM power factor α^* . Although numerical methods can be applied [13], it is generally difficult to find a closed-form solution for α^* due to the water-filling solution of $\hat{\mathbf{Q}}_a$. Here, we look for a near optimal closed-form solution, that is to say, a solution with near optimal performance as well as minimal computational complexity. To this end, we consider $\hat{\mathbf{Q}}_a$ to follow an equal power allocation given by $\hat{\mathbf{Q}}_a = (\mathcal{P}_d/n) \mathbf{I}_n$ instead of the water-filling solution given in (9). Indeed, the equal power allocation

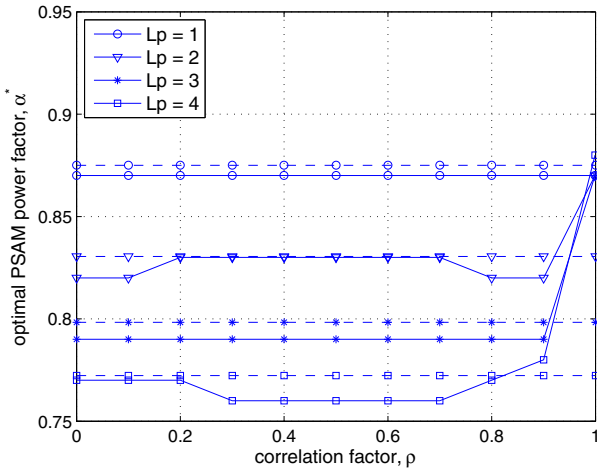


Fig. 1. Optimal PSAM power factor α^* vs. channel spatial correlation factor ρ for 4×4 systems with a block length of 50 and SNR at 10 dB. All values of α^* are found numerically. For comparison, the dashed lines indicate the values of α computed from (11).

for data transmission was also considered in [9] for CCF-only systems. Using, the result in [9], the optimal value of α for $\hat{\mathbf{Q}}_a = (\mathcal{P}_d/n)\mathbf{I}_n$ can be found as

$$\alpha = \begin{cases} \gamma - \sqrt{\gamma(\gamma - 1)}, & \text{for } L_d > N_t \\ \frac{1}{2}, & \text{for } L_d = N_t \\ \gamma + \sqrt{\gamma(\gamma - 1)}, & \text{for } L_d < N_t \end{cases} \quad (11)$$

where $\gamma = \frac{L_d}{L_d - L_p}$, provided that $\mathcal{P}L \gg 1$. The condition of $\mathcal{P}L \gg 1$ can be easily satisfied when the block length is not too small or the SNR is moderate. In the following, we numerically investigate the optimality of α given in (11) for hybrid CCF-CGF systems with data transmission following the water-filling solution. For numerical analysis, we choose the channel covariance matrix to be in the form of $[\mathbf{R}_H]_{ij} = \rho^{|i-j|}$, where ρ is referred to as the spatial correlation factor [11].

Fig. 1 shows α^* found numerically for 4×4 systems with $\mathcal{P} = 10$ dB and $L = 50$. We see that the values of α^* shown as solid lines are very close to the solutions computed from (11) shown as the dashed lines for a wide range of spatial correlation, *e.g.*, $\rho < 0.9$. This trend is also observed at other practical values of \mathcal{P} and L (plots omitted for brevity), provided that $\mathcal{P}L \gg 1$. Therefore, we can conclude that α given in (11) is a near optimal solution for hybrid CCF-CGF systems.

We have also looked at the capacity gain in using hybrid CCF-CGF systems instead of CCF-only systems both with optimized design parameters (using results from [9] for CCF systems and results in this paper for hybrid CCF-CGF systems). For example, with the same system setup as in Fig. 1, the capacity gain increases from 0% at $L_p = 1$ to approximately 8% at $L_p = 4$. And this capacity gain generally increases as \mathcal{P} decreases. Since the capacity gain achieved by using hybrid CCF-CGF instead of CCF-only comes at the cost of extra feedback overhead, the tradeoff between performance

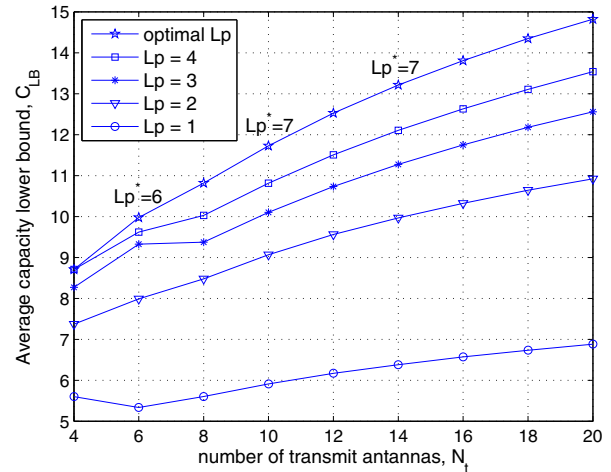


Fig. 2. Average capacity lower bound in (10) vs. number of transmit antennas. The transmit antennas are placed in a UCA with radius equal to half of the wavelength. The block length is 50, SNR is 10 dB and the number of receive antennas is 4. The value of α in (11) is used as a near optimal choice.

and complexity needs to be considered in practical designs.

V. CAPACITY GAIN FROM EXTRA TRANSMIT ANTENNAS

In practical scenarios such as the downlink of cellular systems, the cost of adding extra antennas at the base station is minimal compared with the capacity gain for every downlink transmissions. However, one may argue that adding extra transmit antennas increases the number of channels to be estimated, which results in an increase in the required training resources and feedback overhead for systems with CGF.

On the other hand, the spatial correlations between the transmit antennas usually increases as the number of antennas increases due to space constraints (at the base station). The increase in the channel spatial correlations makes the dominant eigen-directions stronger and the weak eigen-directions even weaker. Therefore, it is possible to design a system with a fixed training length, which only requires training a fixed number of strongest eigen-directions while still keeping the effect of channel estimation errors to a minimum. Note that for a fixed training length, the (near optimal) training power is also fixed according to (11). From the remark of *Theorem 1*, we also know that the amount of feedback transmission does not increase with the number of transmit antennas for a fixed training length. Hence, in this section we numerically investigate the capacity gain from adding extra transmit antennas with fixed training resources and feedback overhead.

Fig. 2 shows the average capacity lower bound C_{LB} versus the number of transmit antennas N_t for systems with $N_r = 4$ and $L_p = 1, 2, 3$ and 4. The transmit antennas are placed in a uniform circular array (UCA) with a fixed radius equal to half of the wavelength. The spatial correlation coefficients between the transmit antennas are calculated using the standard Jakes' model [14]. For each L_p , we see that the capacity lower bound increases as N_t increases from 4 to 20 (except for the case

where $L_p = 1$ and $N_t = 6^2$). For example, the capacity gain in using $N_t = 10$ compared to $N_t = 4$ is 24%. This gain increases to 55% when $N_t = 20$ is used. We also include the capacity lower bound for L_p^* which is found numerically, and show the value of L_p^* for $N_t = 6, 10$ and 14 in Fig. 2. It is clear that the optimal training length can be smaller than the number of transmit antennas for correlated MIMO systems, which agrees with *Corollary 1*.

VI. CONCLUSION

We have investigated the optimal transmission strategies as well as the optimal transmit resource allocation for hybrid CCF-CGF systems with correlated MIMO channels. The major findings in this paper are summarized as follows: (i) The optimal structure for data transmission follows a water-filling solution according to the estimated channel gains, rotated and truncated into the trained eigen-directions. (ii) The optimal training length is at most as large as the number of transmit antennas for any given PSAM power factor. (iii) We have found a simple closed-form solution for PSAM power factor which achieves near optimal capacity performance. (iv) The information capacity can be significantly increased by adding extra transmit antennas without increasing the training or feedback overhead.

APPENDIX A PROOF OF THEOREM 1

The proof consists of two steps.

Firstly, we show that the optimal data structure can be written as $\mathbf{Q} = \mathbf{U}\hat{\mathbf{Q}}\mathbf{U}^\dagger$ with (8). The proof is trivial when $L_p \geq N_t$ in which case we have $\hat{\mathbf{Q}} = \hat{\mathbf{Q}}_a$. In the following we consider $L_p < N_t$. In this case, we have $n = L_p$. From (3) and the optimal pilot structure in Section III-A, we have

$$\begin{aligned} \mathbf{R}_{\hat{\mathbf{H}}} &= \left((\mathbf{U}\mathbf{G}\mathbf{U}^\dagger)^{-1} + \mathbf{U}\mathbf{P}\mathbf{U}^\dagger \right)^{-1} \\ &= \mathbf{U}(\mathbf{G}^{-1} + \mathbf{P})^{-1}\mathbf{U}^\dagger \\ &= \mathbf{U}\tilde{\mathbf{G}}\mathbf{U}^\dagger, \end{aligned} \quad (12)$$

where (12) is indeed the EVD of $\mathbf{R}_{\hat{\mathbf{H}}}$. From (4) and (12), it is easy to show that the EVD of $\mathbf{R}_{\hat{\mathbf{H}}}$ is given by $\mathbf{R}_{\hat{\mathbf{H}}} = \mathbf{U}\hat{\mathbf{G}}\mathbf{U}^\dagger$. Substituting p_i from (6) into (12), we have

$$\hat{\mathbf{G}} = \mathbf{G} - \tilde{\mathbf{G}} = \begin{pmatrix} \hat{\mathbf{G}}_a & \mathbf{0}_{L_p \times (N_t - L_p)} \\ \mathbf{0}_{(N_t - L_p) \times L_p} & \mathbf{0}_{(N_t - L_p) \times (N_t - L_p)} \end{pmatrix}, \quad (13)$$

where $\hat{\mathbf{G}}_a = \text{diag}\{g_1 - \mu^{-1}, \dots, g_{L_p} - \mu^{-1}\}$, and $\mu > 0$.

We let $\mathbf{Q} = \mathbf{U}\hat{\mathbf{Q}}\mathbf{U}^\dagger$. It can be shown that $\hat{\mathbf{Q}}$ is also a positive semi-definite matrix with $\text{tr}\{\hat{\mathbf{Q}}\} = \text{tr}\{\mathbf{Q}\}$. The optimization problem in (7) reduces to

$$\max_{\hat{\mathbf{Q}}: \text{tr}\{\hat{\mathbf{Q}}\} \leq \mathcal{P}_d} \log_2 \left| \mathbf{I}_{N_t} + (1 + \text{tr}\{\tilde{\mathbf{G}}\hat{\mathbf{Q}}\})^{-1} (\hat{\mathbf{H}}\mathbf{U})^\dagger \hat{\mathbf{H}}\mathbf{U}\hat{\mathbf{Q}} \right|. \quad (14)$$

²This is caused by the decrease in the largest eigenvalue of $\mathbf{R}_{\hat{\mathbf{H}}}$ from $N_t = 4$ to $N_t = 6$ due to the UCA geometry and Jakes' model.

Using $\hat{\mathbf{H}} = \hat{\mathbf{H}}_0 \mathbf{R}_{\hat{\mathbf{H}}}^{1/2} = \hat{\mathbf{H}}_0 \hat{\mathbf{G}}^{1/2} \mathbf{U}^\dagger$, where $\hat{\mathbf{H}}_0$ has i.i.d. ZMCSCG entries with unit variance, we have

$$\max_{\hat{\mathbf{Q}}: \text{tr}\{\hat{\mathbf{Q}}\} \leq \mathcal{P}_d} \log_2 \left| \mathbf{I}_{N_t} + (1 + \text{tr}\{\tilde{\mathbf{G}}\hat{\mathbf{Q}}\})^{-1} \hat{\mathbf{H}}_0^\dagger \hat{\mathbf{H}}_0 \hat{\mathbf{G}}^{1/2} \hat{\mathbf{Q}} \hat{\mathbf{G}}^{1/2} \right|. \quad (15)$$

We express $\hat{\mathbf{G}}^{1/2} \hat{\mathbf{Q}} \hat{\mathbf{G}}^{1/2}$ in block matrix form as

$$\begin{aligned} \hat{\mathbf{G}}^{1/2} \hat{\mathbf{Q}} \hat{\mathbf{G}}^{1/2} &= \begin{pmatrix} \hat{\mathbf{G}}_a^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{Q}}_a & \hat{\mathbf{Q}}_b \\ \hat{\mathbf{Q}}_c & \hat{\mathbf{Q}}_d \end{pmatrix} \begin{pmatrix} \hat{\mathbf{G}}_a^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} \hat{\mathbf{G}}_a^{1/2} \hat{\mathbf{Q}}_a \hat{\mathbf{G}}_a^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \end{aligned} \quad (16)$$

where $\hat{\mathbf{Q}}_a$ is an $L_p \times L_p$ matrix. From (16) we can see that $\hat{\mathbf{Q}}_b, \hat{\mathbf{Q}}_c$ and $\hat{\mathbf{Q}}_d$ have no effect on $\hat{\mathbf{G}}^{1/2} \hat{\mathbf{Q}} \hat{\mathbf{G}}^{1/2}$. Also, it can be shown from the property of trace that

$$\text{tr}\{\tilde{\mathbf{G}}\hat{\mathbf{Q}}\} = \sum_{i=1}^{L_p} \mu^{-1} \hat{q}_{ii} + \sum_{i=L_p+1}^{N_t} g_i \hat{q}_{ii}, \quad (17)$$

where $\hat{q}_{ii} \geq 0$ is the i th diagonal entry of $\hat{\mathbf{Q}}$. We see that (17) does not depend on $\hat{\mathbf{Q}}_b$ or $\hat{\mathbf{Q}}_c$. Therefore, (15) is independent of $\hat{\mathbf{Q}}_b$ and $\hat{\mathbf{Q}}_c$. For any given $\hat{\mathbf{Q}}_a$, (15) is achieved by minimizing (17), which can be done by letting $\hat{\mathbf{Q}}_d = \mathbf{0}$. We conclude that (15) can be achieved by letting $\hat{\mathbf{Q}}_b, \hat{\mathbf{Q}}_c$ and $\hat{\mathbf{Q}}_d$ be $\mathbf{0}$.

Secondly, we derive the solution for optimal $\hat{\mathbf{Q}}_a$ as follows. Using the optimal structure of \mathbf{Q} , the objective function in (15) reduces to

$$\max_{\hat{\mathbf{Q}}: \text{tr}\{\hat{\mathbf{Q}}\} \leq \mathcal{P}_d} \log_2 \left| \mathbf{I}_{N_t} + \frac{1}{1 + \text{tr}\{\hat{\mathbf{Q}}\}/\mu} \hat{\mathbf{H}}_0^\dagger \hat{\mathbf{H}}_0 \hat{\mathbf{G}}^{1/2} \hat{\mathbf{Q}} \hat{\mathbf{G}}^{1/2} \right|. \quad (18)$$

We denote $\hat{\mathbf{H}}_0$ in a block matrix form as $\hat{\mathbf{H}}_0 = [\hat{\mathbf{H}}_a \ \hat{\mathbf{H}}_b]$, where $\hat{\mathbf{H}}_a$ is a $N_r \times n$ matrix. With some algebraic manipulations, the optimization problem in (18) reduces to

$$\max_{\hat{\mathbf{Q}}_a: \text{tr}\{\hat{\mathbf{Q}}_a\} \leq \mathcal{P}_d} \log_2 \left| \mathbf{I}_n + \frac{1}{1 + \text{tr}\{\hat{\mathbf{Q}}_a\}/\mu} (\hat{\mathbf{H}}_a \hat{\mathbf{G}}_a^{1/2})^\dagger \hat{\mathbf{H}}_a \hat{\mathbf{G}}_a^{1/2} \hat{\mathbf{Q}}_a \right|. \quad (19)$$

Since $\hat{\mathbf{H}} = \hat{\mathbf{H}}_0 \hat{\mathbf{G}}^{1/2} \mathbf{U}^\dagger$, it is easy to show that $\hat{\mathbf{H}}_a \hat{\mathbf{G}}_a^{1/2}$ contains the first n columns of $\hat{\mathbf{H}}\mathbf{U}$. We denote the first n columns of $\hat{\mathbf{H}}\mathbf{U}$ as \mathbf{F} , i.e., $\mathbf{F} = \hat{\mathbf{H}}_a \hat{\mathbf{G}}_a^{1/2}$, then the optimization problem in (19) can be rewritten as

$$\max_{\hat{\mathbf{Q}}_a: \text{tr}\{\hat{\mathbf{Q}}_a\} \leq \mathcal{P}_d} \log_2 \left| \mathbf{I}_n + \frac{1}{1 + \text{tr}\{\hat{\mathbf{Q}}_a\}/\mu} \mathbf{F}^\dagger \mathbf{F} \hat{\mathbf{Q}}_a \right|. \quad (20)$$

It is a well-known result that the optimal $\hat{\mathbf{Q}}_a$ in (20) has the same eigenvectors as $\mathbf{F}^\dagger \mathbf{F}$ [2]. For a given $\text{tr}\{\hat{\mathbf{Q}}_a\}$, the eigenvalues of $\hat{\mathbf{Q}}_a$ can be found via the standard water-filling algorithm given by

$$a_i = \left[\hat{\eta} - \left(\frac{\lambda_i}{1 + \mu^{-1} \hat{p}_d} \right)^{-1} \right]^+ \quad \text{with} \quad \sum_{i=1}^n a_i = \hat{p}_d, \quad (21)$$

where $\hat{\eta}$ represents the water level and $\hat{p}_d = \text{tr}\{\hat{\mathbf{Q}}_a\}$.

The remaining part is to show that $\hat{\mathcal{P}}_d = \mathcal{P}_d$. From (21), we can solve the water level as

$$\hat{\eta} = \frac{1}{m}\hat{\mathcal{P}}_d + (1 + \mu^{-1}\hat{\mathcal{P}}_d)\frac{1}{m}\sum_{i=1}^m \lambda_i^{-1}, \quad (22)$$

where m denotes the number of non-zero a_i in (21). Therefore, (20) reduces to

$$\max_{\hat{\mathcal{P}}_d: \hat{\mathcal{P}}_d \leq \mathcal{P}_d} \sum_{i=1}^n \log_2 \left(1 + \frac{\lambda_i a_i}{1 + \mu^{-1}\hat{\mathcal{P}}_d} \right) \quad (23)$$

$$= \max_{\hat{\mathcal{P}}_d: \hat{\mathcal{P}}_d \leq \mathcal{P}_d} \sum_{i=1}^m \log_2 \left(\frac{\lambda_i \hat{\eta}}{1 + \mu^{-1}\hat{\mathcal{P}}_d} \right) \quad (24)$$

$$= \max_{\hat{\mathcal{P}}_d: \hat{\mathcal{P}}_d \leq \mathcal{P}_d} \sum_{i=1}^m \log_2 \left(\frac{\hat{\mathcal{P}}_d}{1 + \mu^{-1}\hat{\mathcal{P}}_d} + \sum_{i=1}^m \lambda_i^{-1} \right) + \sum_{i=1}^m \log_2 \frac{\lambda_i}{m}, \quad (25)$$

where (25) is obtained by substituting $\hat{\eta}$ from (22) into (24). It is easy to show that the first derivative of $\frac{\hat{\mathcal{P}}_d}{1 + \mu^{-1}\hat{\mathcal{P}}_d}$ w.r.t. $\hat{\mathcal{P}}_d$ is positive. Therefore, the objective function in (25) is an increasing function of $\hat{\mathcal{P}}_d$ for any fixed m . From the property of water-filling solution [13], we know that $a_i, \forall i$ in (21) are continuous on $\hat{\mathcal{P}}_d$. Therefore, the objective function in (23) is also continuous on $\hat{\mathcal{P}}_d$. We can conclude that the objective function is an increasing function of $\hat{\mathcal{P}}_d$ regardless of the value of m . Hence, the objective function is maximized by letting $\hat{\mathcal{P}}_d = \mathcal{P}_d$. \square

APPENDIX B PROOF OF THEOREM 2

Assuming $L_p \geq N_t$, we aim to show that L_p (or L_d) should be kept to a minimum (or maximum) for any given α . Equivalently, we need to show that $\frac{dC_{LB}}{dL_d} > 0$ (treating L_d as a positive real-valued variable) for any fixed α . First of all, it can be shown that C_{LB} is continuous on L_d (regardless of the value of m), which implies that there is no discontinuity in C_{LB} . Therefore, it suffices to show that $\frac{dC_{LB}}{dL_d} > 0$ for any fixed m and α .

We let $\rho_{\text{eff}} = \frac{\mathcal{P}_d}{1 + \mu^{-1}\mathcal{P}_d} = \frac{\alpha PL}{L_d + \mu^{-1}\alpha PL}$, $y = \sum_{i=1}^m \ln \frac{\lambda_i}{m}$, and $z = \sum_{i=1}^m \lambda_i^{-1}$. Then the average capacity lower bound in (10) can be rewritten as

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

For a given α , we know from (1) that $\mathcal{P}_p L_p$ is constant. Therefore, it can be shown that μ and $\lambda_i \forall i$ are independent of L_d . Differentiating C_{LB} w.r.t. L_d for any fixed m gives

$$\frac{dC_{LB}}{dL_d} = \frac{1}{\ln 2} \frac{m}{L} \left(E_{\lambda} \left\{ \ln(\rho_{\text{eff}} + z) + \frac{L_d}{\rho_{\text{eff}} + z} \frac{d\rho_{\text{eff}}}{dL_d} + \frac{y}{m} \right\} \right), \quad (26)$$

where

$$\frac{d\rho_{\text{eff}}}{dL_d} = -\frac{\alpha PL}{(L_d + \mu^{-1}\alpha PL)^2} = -\frac{\rho_{\text{eff}}}{L_d + \mu^{-1}\alpha PL}. \quad (27)$$

Substituting (27) into (26), we get

$$\frac{dC_{LB}}{dL_d} = \frac{1}{\ln 2} \frac{m}{L} E_{\lambda} \left\{ \ln(\rho_{\text{eff}} + z) - \frac{\rho_{\text{eff}}}{\rho_{\text{eff}} + z} \frac{L_d}{L_d + \mu^{-1}\alpha PL} + \frac{y}{m} \right\}.$$

Since $\frac{L_d}{L_d + \mu^{-1}\alpha PL} < 1$, it suffices to show that

$$f(\rho_{\text{eff}}) \triangleq E_{\lambda} \left\{ \ln(\rho_{\text{eff}} + z) - \frac{\rho_{\text{eff}}}{\rho_{\text{eff}} + z} + \frac{y}{m} \right\} \geq 0. \quad (28)$$

Furthermore, $f(\rho_{\text{eff}})$ is an increasing function of ρ_{eff} . Hence, we only need to show that (28) holds at $\rho_{\text{eff}} = 0$, that is

$$\begin{aligned} E_{\lambda} \left\{ \ln z + \frac{y}{m} \right\} &= E_{\lambda} \left\{ \ln \sum_{i=1}^m \lambda_i^{-1} + \frac{1}{m} \sum_{i=1}^m \ln \frac{\lambda_i}{m} \right\} \\ &\geq E_{\lambda} \left\{ \frac{1}{m} \sum_{i=1}^m \ln \left(\frac{\lambda_i}{m} \right)^{-1} \frac{\lambda_i}{m} \right\} \\ &= 0, \end{aligned} \quad (29)$$

where (29) is obtained using the concavity of $\ln(\cdot)$. Therefore, we conclude that $\frac{dC_{LB}}{dL_d} > 0$ for any given α . Therefore, with the assumption of $L_p \geq N_t$ the optimal training length is given by $L_p^* = N_t$. Hence, in general we have $L_p^* \leq N_t$. \square

ACKNOWLEDGEMENTS

This work was supported under the Australian Research Council's Discovery Projects funding scheme (project no. DP0773898).

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] S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. on Sel. Areas Commun.*, vol. 21, no. 3, pp. 406–417, Apr. 2003.
- [4] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, July 2004.
- [5] E. A. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information," *IEEE Trans. Signal Processing*, vol. 52, no. 12, pp. 3440–3453, Dec. 2004.
- [6] 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, Sept. 2008.
- [7] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.
- [8] B. Hassibi and M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [9] X. Zhou, P. Sadeghi, T. A. Lamahewa, and S. Durrani, "Design guidelines for training-based MIMO systems with feedback," *to appear in IEEE Trans. Signal Processing*.
- [10] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 546–557, Feb. 2004.
- [11] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Processing*, vol. 54, no. 3, pp. 884–893, Mar. 2006.
- [12] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.
- [13] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. Cambridge: Cambridge University Press, 2004.
- [14] T. A. Lamahewa, R. A. Kennedy, T. D. Abhayapala, and T. Betlehem, "MIMO channel correlation in general scattering environments," in *Proc. Australian Communication Theory Workshop*, Perth, Australia, Feb. 2006, pp. 93–98.