Capacity of MIMO Systems: Impact of Spatial Correlation with Channel Estimation Errors

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*—We study the impact of channel spatial correlation on the ergodic capacity of multiple-input multiple-output (MIMO) systems with channel estimation errors in block-fading channels. We consider transmit antenna correlation and use an accurate capacity lower bound to carry out the theoretical analysis for MIMO systems with and without covariance feedback. We show that for non-feedback systems, the capacity increases with channel correlation at low signal-to-noise ratio (SNR). This finding is in contrast to the existing result for non-feedback systems with perfect channel estimation. For covariance feedback systems, we find that the capacity increases with correlation. We also show the robustness of equal power transmission and the optimality of beamforming at low SNR for non-feedback and covariance feedback systems, respectively. Our numerical results validate our analysis and show that the theoretical low SNR analysis can extend to moderate SNR values.**

I. Introduction

The use of multiple transmit and receive antennas provides high data rate transmission in wireless communication systems, especially when channel state information (CSI) is known at the receiver [1, 2]. To acquire CSI, training symbols (or pilots) are usually transmitted periodically to allow channel estimation at the receiver. For multiple-input multiple-output (MIMO) systems with spatially uncorrelated block fading channels, the authors in [3] studied the optimal pilot sequence, optimal pilot power allocation as well as optimal pilot length, which maximizes an ergodic capacity lower bound. For correlated MIMO systems, the authors in [4] studied the performance of the least square (LS) and the linear minimum mean-square-error (LMMSE) estimators. Furthermore, the optimal training sequence for the LMMSE estimator was found to follow a water-filling solution [4, 5].

Most studies on the effect of channel spatial correlation on the ergodic capacity do not consider channel estimation errors and assume perfect CSI at the receiver (CSIR) [6–8]. For multiple-input single-output (MISO) systems, the authors in [6] showed that the ergodic capacity decreases as correlation increases when no CSI is available at the transmitter. Similar studies on MIMO systems found that the channel spatial correlations among both the transmit antennas and the receive antennas can reduce the capacity at high signal to noise ratio (SNR) [7] or for large number of transmit and receive antennas [8].

Information-theoretic studies on spatially correlated MIMO channels have also considered covariance feedback systems. Since the statistics of the channel realizations change much slower than the instantaneous channel gains, it is practical for the receiver to accurately measure the channel covariance matrix and feed it back to the transmitter at negligible cost. With the covariance matrix known at the transmitter, the authors in [6] showed that the ergodic capacity increases as correlation increases. In addition, assuming perfect CSIR the optimal data transmission strategies for covariance feedback systems are studied in [9, 10].

Since the CSIR is never perfect in reality, it is necessary to take the channel estimation error into account in the capacity analysis, especially at low operating SNR. Also, it is known that the channel spatial correlation may have a significant effect on the channel estimation error [4, 11]. Therefore, the channel correlation and estimation error have a joint impact on the ergodic capacity. In this paper, we extend previous studies for systems with perfect CSIR and analyze the impact of channel spatial correlation on the ergodic capacity of MIMO systems with channel estimation errors. In particular, we consider both non-feedback and covariance feedback systems in block-fading channels and focus on the low SNR regime where the channel estimation error is non-negligible. The main contributions of this paper are:

- In Section III-A, we extend existing results for spatially independent channels in [12] and derive a tight lower and upper bound of the ergodic capacity for spatially correlated MIMO systems.
- In Section III-B and III-C, we show that for non-feedback systems, the capacity increases with channel correlation at low SNR, but decreases with correlation at high SNR. The finding at low SNR is in contrast to the existing result in [6] with perfect CSIR assumption. We also show that equal power transmission is a robust transmission scheme for non-feedback systems at low SNR.
- In Section III-B, we show that for covariance feedback systems at low SNR, beamforming transmission is the optimal transmission scheme and the capacity increases with channel spatial correlation.

Our numerical results in Section IV validate our theoretical analysis and provide insights into the impact of channel spatial correlation at moderate SNR values.

Throughout the paper, the following notations will be used: Boldface upper and lower cases denote matrices and column vectors, respectively. The matrix $\boldsymbol{I}_N$ is the $N \times N$ identity matrix. $[\cdot]^*$ denotes the complex conjugate operation, and $[\cdot]^\dagger$

denotes the conjugate transpose operation. The notation $E\{\cdot\}$ denotes the mathematical expectation. tr$\{\cdot\}$, $|\cdot|$ and rank$\{\cdot\}$ denote the matrix trace, determinant and rank, respectively.

II. System Model

A. Signal Model

We consider a MIMO block-flat-fading channel model. After matched filter, the $N_r \times 1$ received symbol vector is given by

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

where $\boldsymbol{x}$ is the $N_t \times 1$ transmitted symbol vector, $\boldsymbol{H}$ is the $N_r \times N_t$ channel gain matrix, and $\boldsymbol{n}$ is the $N_r \times 1$ noise vector having zero-mean circularly symmetric complex Gaussian (ZMCSCG) entries with variance $\sigma_{\boldsymbol{n}}^2$. The entries of $\boldsymbol{H}$ are also ZMCSCG with unit variance. We consider a typical downlink transmission scenario where there is insufficient scattering around the base station, resulting in spatial correlations at the base station transmitter. Therefore, $\boldsymbol{H} = \boldsymbol{H}_0 \boldsymbol{R}_{\boldsymbol{H}}^{1/2}$, where $\boldsymbol{H}_0$ has independent identically distributed (i.i.d.) ZMCSCG entries with unit variance. The spatial correlation is characterized by the covariance matrix $\boldsymbol{R}_{\boldsymbol{H}} = E\{\boldsymbol{H}^\dagger \boldsymbol{H}\}/N_r$. In the case where the channels are spatially independent, we have $\boldsymbol{R}_{\boldsymbol{H}} = \boldsymbol{I}_{N_t}$. We assume that $\boldsymbol{R}_{\boldsymbol{H}}$ is a positive semi-definite matrix with full rank, and denote the eigenvalues of $\boldsymbol{R}_{\boldsymbol{H}}$ by $\boldsymbol{g} = [g_1\ g_2\ ...\ g_{N_t}]^T$.

B. Channel Estimation

The channel gains remain constant for a block of T symbols and assume independent realizations in the next block. During each transmission block, each transmit antenna sends T_p pilot symbols (usually $T_p \geq N_t$), followed by $T - T_p$ data symbols. The receiver uses the T_p pilot symbols to perform channel estimation. Combining the first T_p receive symbols in a $N_r \times T_p$ matrix, we have

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X}_p + \boldsymbol{N}, \quad (2)$$

where $\boldsymbol{X}_p$ is the $N_t \times T_p$ pilot matrix and $\boldsymbol{N}$ is the $N_r \times T_p$ noise matrix.

Assuming the channel spatial correlation is known at the receiver, the LMMSE estimator can be used. We denote the channel estimate and estimation error as $\hat{\boldsymbol{H}} = \hat{\boldsymbol{H}}_0 \boldsymbol{R}_{\hat{\boldsymbol{H}}}^{1/2}$ and $\tilde{\boldsymbol{H}} = \tilde{\boldsymbol{H}}_0 \boldsymbol{R}_{\tilde{\boldsymbol{H}}}^{1/2}$ respectively, where $\hat{\boldsymbol{H}}_0$ and $\tilde{\boldsymbol{H}}_0$ have i.i.d. ZMCSCG entries. $\hat{\boldsymbol{H}}$ is given as [4]

$$\hat{\boldsymbol{H}} = \boldsymbol{Y}(\boldsymbol{X}_p{}^\dagger \boldsymbol{R}_{\boldsymbol{H}} \boldsymbol{X}_p + \sigma_{\boldsymbol{n}}^2 \boldsymbol{I}_{T_p})^{-1} \boldsymbol{X}_p{}^\dagger \boldsymbol{R}_{\boldsymbol{H}}. \quad (3)$$

The covariance matrix of the estimation error is given by [4]

$$\boldsymbol{R}_{\tilde{\boldsymbol{H}}} = E\{\tilde{\boldsymbol{H}}^\dagger \tilde{\boldsymbol{H}}\}/N_r = (\boldsymbol{R}_{\boldsymbol{H}}^{-1} + \frac{1}{\sigma_{\boldsymbol{n}}^2} \boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger)^{-1}, \quad (4)$$

where $\mathcal{P}_p$ is the total power per pilot transmission. From the orthogonality property of LMMSE estimator, we have $\boldsymbol{R}_{\hat{\boldsymbol{H}}} = E\{\hat{\boldsymbol{H}}^\dagger \hat{\boldsymbol{H}}\}/N_r = \boldsymbol{R}_{\boldsymbol{H}} - \boldsymbol{R}_{\tilde{\boldsymbol{H}}}$. We denote the eigenvalues of $\boldsymbol{R}_{\hat{\boldsymbol{H}}}$ by $\boldsymbol{\lambda} = [\lambda_1\ \lambda_2\ ...\ \lambda_{N_t}]^T$. It is noted that the LMMSE estimator is the MMSE estimator for channels with Gaussian statistics. We define the SNR during pilot transmissions as $\mathcal{P}_p/\sigma_{\boldsymbol{n}}^2$. Similarly, the SNR during data transmissions is given by $\mathcal{P}_d/\sigma_{\boldsymbol{n}}^2$, where $\mathcal{P}_d$ is the total power per data transmission.

C. A Measure of Channel Spatial Correlation

The concept of majorization has been used to characterize the degree of channel spatial correlation, e.g. [6, 8]. A vector $\boldsymbol{a} = [a_1\ a_2\ ...\ a_n]^T$ is said to be majorized by another vector $\boldsymbol{b} = [b_1\ b_2\ ...\ b_n]^T$ if

$$\sum_{i=1}^{k} a_i \leq \sum_{i=1}^{k} b_i, \quad k = 1, ..., n-1, \quad \text{and} \quad \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i, \quad (5)$$

where the elements in both vectors are sorted in descending order [13]. We denote the relationship as $\boldsymbol{a} \prec \boldsymbol{b}$. Any real-valued function Φ, defined on a vector subspace, is said to be Schur-convex, if $\boldsymbol{a} \prec \boldsymbol{b}$ implies $\Phi(\boldsymbol{a}) \leq \Phi(\boldsymbol{b})$ [13]. Similarly Φ is Schur-concave, if $\boldsymbol{a} \prec \boldsymbol{b}$ implies $\Phi(\boldsymbol{a}) \geq \Phi(\boldsymbol{b})$. Following [6], we have the following definition:

Definition 1: Let $\boldsymbol{a}$ contains the eigenvalues of the channel covariance matrix $\boldsymbol{R}_{\boldsymbol{a}}$, and $\boldsymbol{b}$ contains the eigenvalues of the channel covariance matrix $\boldsymbol{R}_{\boldsymbol{b}}$. The elements in both vectors are sorted in descending order. Then the $\boldsymbol{R}_{\boldsymbol{a}}$ is less correlated than $\boldsymbol{R}_{\boldsymbol{b}}$ if and only if $\boldsymbol{a} \prec \boldsymbol{b}$.

III. Theoretical Analysis on Ergodic Capacity

A. Ergodic Capacity Bounds

We consider a lower bound of the ergodic capacity per channel use for systems using LMMSE channel estimation, which has been used in information-theoretic studies, e.g. [3, 12]. This is due to the fact that the exact capacity expression is still unavailable for systems with imperfect CSIR. In particular, the authors in [12] derived a lower bound and an upper bound for spatially i.i.d. channels. Here we extend their results to spatially correlated channels as follows.

A lower bound of the ergodic capacity per channel use is given by [12]

$$C_{\text{LB}} = E_{\hat{\boldsymbol{H}}}\left\{ \log_2 \left| \boldsymbol{I}_{N_t} + \hat{\boldsymbol{H}}^\dagger (\sigma_{\boldsymbol{n}}^2 \boldsymbol{I}_{N_t} + \Sigma_{\tilde{\boldsymbol{H}}\boldsymbol{x}})^{-1} \hat{\boldsymbol{H}} \boldsymbol{Q} \right| \right\}, \quad (6)$$

where $\boldsymbol{Q} = E\{\boldsymbol{x}\boldsymbol{x}^\dagger\}$ is the input covariance matrix, and

$$\Sigma_{\tilde{\boldsymbol{H}}\boldsymbol{x}} = E\{\tilde{\boldsymbol{H}}\boldsymbol{x}\boldsymbol{x}^\dagger \tilde{\boldsymbol{H}}^\dagger\} = E\{\tilde{\boldsymbol{H}}_0 \boldsymbol{R}_{\tilde{\boldsymbol{H}}}^{1/2} \boldsymbol{x}\boldsymbol{x}^\dagger (\boldsymbol{R}_{\tilde{\boldsymbol{H}}}^{1/2})^\dagger \tilde{\boldsymbol{H}}_0^\dagger\},$$
$$= E\left\{ \text{tr}\{\boldsymbol{R}_{\tilde{\boldsymbol{H}}}^{1/2} \boldsymbol{x}\boldsymbol{x}^\dagger (\boldsymbol{R}_{\tilde{\boldsymbol{H}}}^{1/2})^\dagger\} \right\} \boldsymbol{I}_{N_t} = \text{tr}\{\boldsymbol{R}_{\tilde{\boldsymbol{H}}} \boldsymbol{Q}\} \boldsymbol{I}_{N_t}.$$

Therefore, the ergodic capacity lower bound per channel use in (6) can be rewritten as

$$C_{\text{LB}} = E_{\hat{\boldsymbol{H}}}\left\{ \log_2 \left| \boldsymbol{I}_{N_t} + (\sigma_{\boldsymbol{n}}^2 + \text{tr}\{\boldsymbol{R}_{\tilde{\boldsymbol{H}}} \boldsymbol{Q}\})^{-1} \hat{\boldsymbol{H}}^\dagger \hat{\boldsymbol{H}} \boldsymbol{Q} \right| \right\}. \quad (7)$$

An upper bound of the ergodic capacity per channel use is given by [12]

$$C_{\text{UB}} = E_{\hat{\boldsymbol{H}}}\left\{ \log_2 \left| \pi e \Sigma_{\boldsymbol{y}|\hat{\boldsymbol{H}}} \right| \right\} - E_{\boldsymbol{x}}\left\{ \log_2 \left| \pi e (\Sigma_{\tilde{\boldsymbol{H}}\boldsymbol{x}|\boldsymbol{x}} + \sigma_{\boldsymbol{n}}^2 \boldsymbol{I}_{N_t}) \right| \right\},$$

where

$$\Sigma_{\boldsymbol{y}|\hat{\boldsymbol{H}}} = E\{\boldsymbol{y}\boldsymbol{y}^\dagger | \hat{\boldsymbol{H}}\} = \hat{\boldsymbol{H}} \boldsymbol{Q} \hat{\boldsymbol{H}}^\dagger + \text{tr}\{\boldsymbol{R}_{\tilde{\boldsymbol{H}}} \boldsymbol{Q}\} \boldsymbol{I}_{N_t} + \sigma_{\boldsymbol{n}}^2 \boldsymbol{I}_{N_t},$$

and

$$\Sigma_{\tilde{\boldsymbol{H}}\boldsymbol{x}|\boldsymbol{x}} = E\{\tilde{\boldsymbol{H}}\boldsymbol{x}\boldsymbol{x}^\dagger\tilde{\boldsymbol{H}}^\dagger|\boldsymbol{x}\} = E\{\tilde{\boldsymbol{H}}_0\boldsymbol{R}_{\tilde{\boldsymbol{H}}}^{1/2}\boldsymbol{x}\boldsymbol{x}^\dagger(\boldsymbol{R}_{\tilde{\boldsymbol{H}}}^{1/2})^\dagger\tilde{\boldsymbol{H}}_0^\dagger|\boldsymbol{x}\},$$
$$= \mathrm{tr}\{\boldsymbol{R}_{\tilde{\boldsymbol{H}}}^{1/2}\boldsymbol{x}\boldsymbol{x}^\dagger(\boldsymbol{R}_{\tilde{\boldsymbol{H}}}^{1/2})^\dagger\}\boldsymbol{I}_{N_t} = \boldsymbol{x}^\dagger\boldsymbol{R}_{\tilde{\boldsymbol{H}}}\boldsymbol{x}.$$

Therefore, the ergodic capacity upper bound per channel use can be rewritten as

$$\begin{aligned} C_{\mathrm{UB}} &= E_{\hat{\boldsymbol{H}}}\Big\{\log_2\Big|\boldsymbol{I}_{N_t} + (\sigma_{\boldsymbol{n}}^2 + \mathrm{tr}\{\boldsymbol{R}_{\tilde{\boldsymbol{H}}}\boldsymbol{Q}\})^{-1}\hat{\boldsymbol{H}}^\dagger\hat{\boldsymbol{H}}\boldsymbol{Q}\Big|\Big\} + \\ & E_{\boldsymbol{x}}\Big\{\log_2\Big|(\mathrm{tr}\{\boldsymbol{R}_{\tilde{\boldsymbol{H}}}\boldsymbol{Q}\} + \sigma_{\boldsymbol{n}}^2\boldsymbol{I}_{N_t})(\boldsymbol{x}^\dagger\boldsymbol{R}_{\tilde{\boldsymbol{H}}}\boldsymbol{x} + \sigma_{\boldsymbol{n}}^2\boldsymbol{I}_{N_t})^{-1}\Big|\Big\}, \\ &= C_{\mathrm{LB}} + C_{\mathrm{gap}}, \end{aligned} \quad (8)$$

where C_{gap} is the difference between the upper bound and the lower bound, which indicates the maximum error of the bounds. The authors in [12] studied the tightness of the bounds for i.i.d. channels with a simplified channel estimation model. They observed that C_{gap} is negligible compared with the bounds for Gaussian inputs, hence the bounds are tight. We find that this is also true for spatially correlated channels with LMMSE estimation. Therefore, the capacity lower bound is accurate enough to be used in our analysis assuming Gaussian inputs.

Using (7), we study the effect of channel spatial correlation on the ergodic capacity lower bound. We will use "capacity lower bound" and "capacity" interchangeably. Both non-feedback and covariance feedback systems are considered. For brevity we will only show the analysis using the orthogonal pilot matrix, i.e. $\boldsymbol{X}_p\boldsymbol{X}_p{}^\dagger = \mathcal{P}_pT_p/N_t\boldsymbol{I}_{N_t}$. One can show that for non-feedback systems, the use of orthogonal pilot matrix minimizes the channel estimation MSE, i.e. $\mathrm{tr}\{\boldsymbol{R}_{\tilde{\boldsymbol{H}}}\}$, for the least-favourable spatial correlation (with proof similar to *Theorem 2* in Section III-B). Hence, it is a robust choice for non-feedback systems. For covariance feedback systems, the optimal pilot matrix is discussed in [4, 5]. We find that the following theoretical analysis and results considering orthogonal pilot matrix would also apply to those using the optimal pilot matrix for covariance feedback systems.

When LMMSE estimation is used at the receiver, we find that the channel spatial correlation may have different effects on the capacity at low and high SNR.

B. Low SNR Analysis

At very low (data and pilot) SNR, the ergodic capacity lower bound per channel use in (7) can be approximated as

$$C_{\mathrm{LB}} = \frac{1}{\ln 2}E_{\hat{\boldsymbol{H}}}\mathrm{tr}\Big\{\ln\Big(\boldsymbol{I}_{N_t} + (\sigma_{\boldsymbol{n}}^2 + \mathrm{tr}\{\boldsymbol{R}_{\tilde{\boldsymbol{H}}}\boldsymbol{Q}\})^{-1}\hat{\boldsymbol{H}}^\dagger\hat{\boldsymbol{H}}\boldsymbol{Q}\Big)\Big\}, \quad (9)$$
$$\approx \frac{1}{\ln 2}E_{\hat{\boldsymbol{H}}}\,\mathrm{tr}\Big\{\ln\Big(\boldsymbol{I}_{N_t} + \frac{1}{\sigma_{\boldsymbol{n}}^2}\hat{\boldsymbol{H}}^\dagger\hat{\boldsymbol{H}}\boldsymbol{Q}\Big)\Big\}, \quad (10)$$
$$\approx \frac{1}{\ln 2}E_{\hat{\boldsymbol{H}}}\,\mathrm{tr}\Big\{\frac{1}{\sigma_{\boldsymbol{n}}^2}\hat{\boldsymbol{H}}^\dagger\hat{\boldsymbol{H}}\boldsymbol{Q}\Big\}, \quad (11)$$
$$= \frac{N_r}{\sigma_{\boldsymbol{n}}^2\ln 2}\mathrm{tr}\{\boldsymbol{R}_{\hat{\boldsymbol{H}}}\boldsymbol{Q}\}, \quad (12)$$

where (9) is obtained using $\ln|\cdot| = \mathrm{tr}\{\ln(\cdot)\}$, (10) uses low SNR approximation, and (11) is obtained using Taylor's series expansion of $\ln(\cdot)$ together with low SNR approximation.

For non-feedback systems, one may use equal power transmission, i.e. $\boldsymbol{Q} = \mathcal{P}_d/N_t\boldsymbol{I}_{N_t}$. The optimality of this transmission scheme will be discussed in *Theorem 2*, and the effect of channel spatial correlation on the capacity at low SNR will be summarized in *Theorem 1*. For covariance feedback systems, the optimal transmission scheme and the effect of spatial correlation will be presented in *Theorem 3* and *Theorem 4*, respectively.

Theorem 1: When LMMSE estimator is used for channel estimation in MIMO systems and the input covariance matrix is given by $\boldsymbol{Q} = \mathcal{P}_d/N_t\boldsymbol{I}_{N_t}$, the ergodic capacity lower bound is Schur-convex on the eigenvalues of channel covariance matrix at sufficiently low SNR. Therefore, capacity increases as channel spatial correlation increases at sufficiently low SNR.

Proof: With equal power transmission, (12) can be reduced to

$$\begin{aligned} C_{\mathrm{LB}} &\approx \frac{N_r\mathcal{P}_d}{N_t\sigma_{\boldsymbol{n}}^2\ln 2}\mathrm{tr}\{\boldsymbol{R}_{\hat{\boldsymbol{H}}}\}, \\ &= \frac{N_r\mathcal{P}_d}{N_t\sigma_{\boldsymbol{n}}^2\ln 2}(N_t - \mathrm{tr}\{\boldsymbol{R}_{\tilde{\boldsymbol{H}}}\}). \end{aligned} \quad (13)$$

From (4), we find that $\mathrm{tr}\{\boldsymbol{R}_{\tilde{\boldsymbol{H}}}\} = \sum_{i=1}^{N_t}(g_i^{-1} + \frac{\mathcal{P}_pT_p}{\sigma_{\boldsymbol{n}}^2N_t})^{-1}$. It can be shown that $f(g) = (g^{-1} + \frac{\mathcal{P}_pT_p}{\sigma_{\boldsymbol{n}}^2N_t})^{-1}$ is a concave function of g. Therefore, $\mathrm{tr}\{\boldsymbol{R}_{\tilde{\boldsymbol{H}}}\} = \sum_{i=1}^{N_t}f(g_i)$ is Schur-concave in $\boldsymbol{g}$ [13]. And the ergodic capacity lower bound in (13) is Schur-convex in $\boldsymbol{g}$. □

Remark: The Schur-concavity of $\mathrm{tr}\{\boldsymbol{R}_{\tilde{\boldsymbol{H}}}\}$ implies that the channel spatial correlation reduces the channel estimation error. Consequently, we see that the channel estimation error significantly affects the Schur-concavity of the ergodic capacity in non-feedback systems. For systems with perfect CSIR, the result in [6] shows that the ergodic capacity is Schur-concave at any SNR, whereas our analysis show that the capacity is Schur-convex at low SNR for imperfect CSIR systems.

Theorem 2: When LMMSE estimator is used for channel estimation in MIMO systems without feedback, equal power transmission maximizes the ergodic capacity lower bound for the least-favourable channel correlation at sufficiently low SNR.

Proof: see Appendix A.

Remark: Although the Schur-concavity of the capacity is opposite for perfect and imperfect CSIR systems, the robustness of equal power transmission is also found in [6] for systems with perfect CSIR.

When covariance feedback is available, we assume that the transmitter has full knowledge of $\boldsymbol{R}_{\boldsymbol{H}}$ or $\boldsymbol{R}_{\hat{\boldsymbol{H}}}$. Therefore, adaptive transmission is necessary to maximize the ergodic capacity lower bound.

Theorem 3: When LMMSE estimator is used for channel estimation in MIMO systems with covariance feedback, the optimal transmission scheme is beamforming along the strongest eigenvectors of $\boldsymbol{R}_{\boldsymbol{H}}$ or $\boldsymbol{R}_{\hat{\boldsymbol{H}}}$, that is to say, to transmit all *power in the direction of the eigenvector with the largest corresponding eigenvalue of $\boldsymbol{R}_{\boldsymbol{H}}$ or $\boldsymbol{R}_{\hat{\boldsymbol{H}}}$.*

Proof: see Appendix B.

Remark: We can see that beamforming is optimal at sufficiently low SNR for imperfect CSIR systems, regardless of the channel spatial correlation. This agrees with the results in [9] for systems with perfect CSIR, using which one can show that the necessary and sufficient conditions for optimality of beamforming are satisfied at sufficiently low SNR.

Theorem 4: When LMMSE estimator is used for channel estimation in MIMO systems with covariance feedback and the optimal beamforming transmission is adopted, the ergodic capacity lower bound is Schur-convex on the eigenvalues of channel covariance matrix at sufficiently low SNR. Therefore, capacity increases as channel spatial correlation increases at sufficiently low SNR.

Proof: With the optimal choice of $\boldsymbol{Q}$ according to *Theorem 3*, the ergodic capacity lower bound per channel use at low SNR in (12) is reduced to $C_{\text{LB}} = \frac{N_r \mathcal{P}_d}{\sigma_{\boldsymbol{n}}^2 \ln 2}\lambda_{\text{max}}$, where λ_{max} is the largest eigenvalue of $\boldsymbol{R}_{\hat{\boldsymbol{H}}}$. It is clear that λ_{max} is an increasing function of $\boldsymbol{\lambda}$ and is Schur-convex in $\boldsymbol{\lambda}$ [13]. Also, we have seen from the proof of *Theorem 3* that λ_i is a convex function of g_i. Since a composition of an increasing Schur-convex function and a convex function is Schur-convex [13], we find that λ_{max}, hence C_{LB} is Schur-convex in $\boldsymbol{g}$. □

C. High SNR Analysis

At very high (data and pilot) SNR, the covariance matrix of the channel estimation error in (4) can be approximated as $\boldsymbol{R}_{\tilde{\boldsymbol{H}}} \approx (\frac{\mathcal{P}_p T_p}{\sigma_{\boldsymbol{n}}^2 N_t})^{-1}\boldsymbol{I}_{N_t}$. Note that $\boldsymbol{R}_{\tilde{\boldsymbol{H}}}$ is independent of the channel spatial correlation at high SNR. Therefore, the ergodic capacity lower bound per channel use in (7) can be approximated as

$$\begin{aligned} C_{\text{LB}} &\approx E_{\hat{\boldsymbol{H}}}\Big\{ \log_2 \Big| \Big(\sigma_{\boldsymbol{n}}^2 + \frac{\sigma_{\boldsymbol{n}}^2 N_t \mathcal{P}_d}{T_p \mathcal{P}_p}\Big)^{-1} \hat{\boldsymbol{H}}^\dagger \hat{\boldsymbol{H}} \boldsymbol{Q} \Big| \Big\}, \\ &= E_{\hat{\boldsymbol{H}}}\Big\{ \log_2 \beta |\hat{\boldsymbol{H}}^\dagger \hat{\boldsymbol{H}} \boldsymbol{Q}| \Big\}, \\ &= E_{\hat{\boldsymbol{H}}_0}\Big\{ \log_2 \beta |\hat{\boldsymbol{H}}_0^\dagger \hat{\boldsymbol{H}}_0| \Big\} + \log_2 |\boldsymbol{R}_{\hat{\boldsymbol{H}}}| + \log_2 |\boldsymbol{Q}|, \quad (14) \end{aligned}$$

where $\beta = \left(\sigma_{\boldsymbol{n}}^2 + \frac{\sigma_{\boldsymbol{n}}^2 N_t \mathcal{P}_d}{T_p \mathcal{P}_p}\right)^{-N_t}$.

From (14), the optimal transmission scheme and the Schur-concavity of the ergodic capacity are easy to obtain. Firstly, it is clear that the optimal transmission scheme at high SNR is equal power transmission, i.e. $\boldsymbol{Q} = \mathcal{P}_d / N_t \boldsymbol{I}_{N_t}$, for both non-feedback and covariance feedback systems. This result is also observed in [10] for covariance feedback systems with perfect CSIR as well as in [12] for systems with instantaneous channel gain feedback. It is noted that the optimal transmission scheme at high SNR may be different when different channel estimation schemes are used. For example, the authors in [14] found that the optimal transmission is $\boldsymbol{Q} = (\mathcal{P}_d / \sum_{i=1}^{N_t} g_i^{-1})\boldsymbol{R}_{\boldsymbol{H}}^{-1}$ under a slightly different channel estimation setup in which $\boldsymbol{R}_{\tilde{\boldsymbol{H}}}$ always equals $\boldsymbol{R}_{\boldsymbol{H}}$ multiplied by a constant for a given SNR.

Furthermore, one can show from (14) that the ergodic capacity lower bound is Schur-concave in $\boldsymbol{\lambda}$ due to the Schur-concavity of $\Pi_{i=1}^{N_t}\lambda_i$ [13]. Also with $\boldsymbol{R}_{\tilde{\boldsymbol{H}}} \approx (\frac{\mathcal{P}_p T_p}{\sigma_{\boldsymbol{n}}^2 N_t})^{-1}\boldsymbol{I}_{N_t}$ at high SNR, it can be shown that $\boldsymbol{\lambda}$ is an affine function of $\boldsymbol{g}$, hence the concavity preserves. Therefore, the ergodic capacity lower bound is Schur-concave on the eigenvalues of channel covariance matrix, that is to say, the capacity decreases as channel spatial correlation increases at sufficiently high SNR. This agrees with the result in [6] for systems with the perfect CSIR. However, the Schur-concavity may be different when different channel estimation scheme is used. For example, with the optimal transmission scheme one can show that the ergodic capacity is independent of the channel spatial correlation at high SNR under the channel estimation setup in [14].

Finally, we have seen that that the ergodic capacity lower bound is Schur-convex at sufficiently low SNR and Schur-concave at sufficiently high SNR for systems with imperfect CSIR. This suggests that the capacity is neither Schur-concave nor Schur-convex at moderate SNR values.

IV. Numerical Results

To validate our theoretical low and high SNR analysis in Section III, we carry out numerical studies using the ergodic capacity lower bound in (7), not its approximations in (12) or (14). These numerical results also provide insights into the effects of channel spatial correlation on the capacity at moderate SNR values. It is noted that the ratio of power and time allocated to pilot transmission affects capacity results. Intuitively, we expect the optimal amount of training resource required for spatially i.i.d. channels to be higher than that for correlated channels. Here we do not optimize the ratio of power and time allocated to training. We use $T_p = N_t$ and equal power allocation to pilot and data symbols which is a practical power allocation scheme. For numerical analysis, we choose the channel covariance matrix to be in the form of $(\boldsymbol{R}_{\boldsymbol{H}})_{ij} = \rho^{|i-j|}$, where ρ is the spatial correlation factor [4, 14]. For systems having two transmit antennas, ρ is referred to as the channel correlation coefficient.

Fig. 1 shows the ergodic capacity lower bound per transmission in (7) vs. SNR for 2×1 MISO systems adopting equal power transmission. In general, we see that fully correlated channels result in the highest capacity for a wide range of SNR values, even at 18 dB. The spatially independent channels give the lowest capacity for SNR below 6 dB. Comparing the two extreme cases, we see a significant improvement from independent channels to fully correlated channels at low to moderate SNR values, which agrees with *Theorem 1*. For example, the capacity improvement from i.i.d. channels to fully correlated channels is approximately 33% at 0 dB and 13% at 5 dB.

Fig. 2 shows the ergodic capacity lower bound per transmission in (7) vs. SNR for 2×2 MIMO systems adopting equal power transmission. In general, we see that fully correlated channels result in the highest capacity for SNR values below 5 dB, while independent channels maximize the capacity SNRs above 5 dB. In particular, the capacity difference between

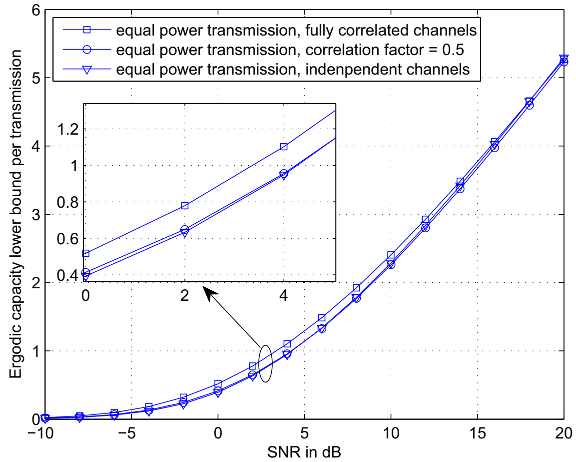

Fig. 1. The ergodic capacity lower bound per transmission in (7) vs. SNR, for 2×1 MISO systems with different levels of channel spatial correlation. Equal power transmission is used.

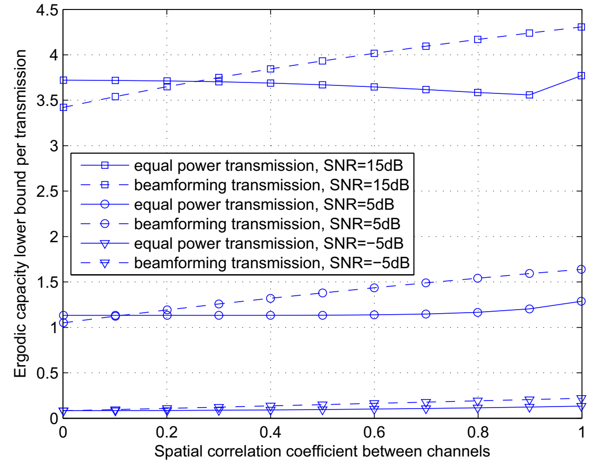

Fig. 3. The ergodic capacity lower bound per transmission in (7) vs. channel spatial correlation for 2×1 MISO systems at different SNR. Both equal power transmission (solid lines) and beamforming (dashed lines) are used.

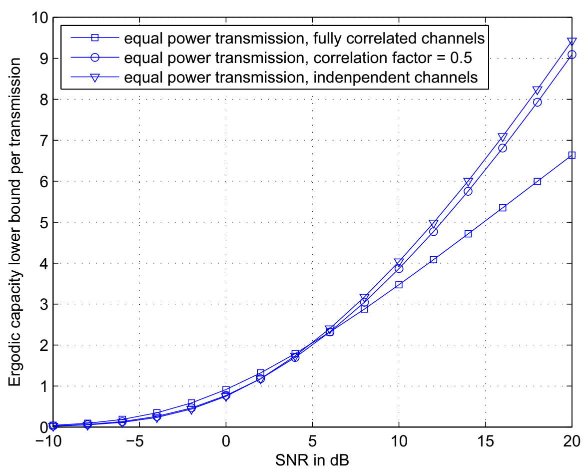

Fig. 2. The ergodic capacity lower bound per transmission in (7) vs. SNR, for 2×2 MIMO systems with different levels of channel spatial correlation. Equal power transmission is used.

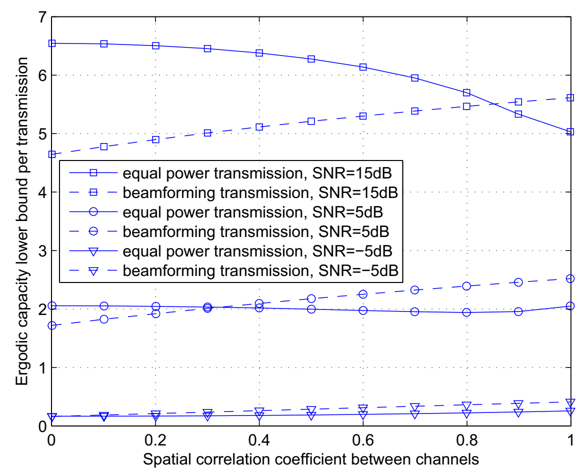

Fig. 4. The ergodic capacity lower bound per transmission in (7) vs. channel spatial correlation for 2×2 MIMO systems at different SNR. Both equal power transmission (solid lines) and beamforming (dashed lines) are used.

i.i.d. channels and fully correlated channels is significant at high SNR. This is because the benefit from receiver diversity outweighs the effect of channel estimation error for i.i.d. channels at moderate to high SNR.

Fig. 3 shows the ergodic capacity lower bound per transmission in (7) against the channel spatial correlation coefficient for 2×1 MISO systems adopting both equal power transmission and beamforming. For equal power transmission, we see that the capacity is an increasing function of the channel correlation coefficient at low SNR such as -5 dB and 5 dB, while the capacity lower bound is not a monotonic function at moderate SNR such as 15 dB. This again agrees with our analysis that the ergodic capacity is Schur-convex at sufficiently low SNR (i.e. *Theorem 1*) and neither Schur-convex nor Schur-concave at moderate SNR. Fig. 4 also shows the same trend for 2×2 MIMO systems, however, the SNR range for the capacity to be Schur-convex or Schur-concave is changed.

For beamforming transmission, our analysis in *Theorem 4* implies that the ergodic capacity lower bound increases with channel correlation coefficient at sufficiently low SNR. This is validated in Fig. 3 and Fig. 4. Furthermore, we also see that the capacity increases with channel correlation coefficient at moderate to high SNR values as well. This is intuitively true, since beamforming becomes more beneficial as the dominant eigenvector of the channel covariance matrix becomes stronger which happens when the channels become more correlated.

From Fig. 3 and Fig. 4 we also see that beamforming outperforms equal power transmission at moderate to high channel spatial correlation. And the correlation coefficient value at crossover reduces as SNR decreases. This agrees with our analysis on the optimality of beamforming at low SNR (i.e. *Theorem 3*) as well as the optimality of equal power transmission at high SNR.

V. Conclusion

We have studied the effects of channel spatial correlation on the ergodic capacity of MIMO systems taking into account the channel estimation errors at the receiver. We have considered transmit antenna correlation and used an accurate capacity lower bound to perform our theoretical analysis. For non-feedback systems, we have shown that the capacity increases with channel correlation at low SNR, which had not been revealed in previous studies assuming perfect channel estimation. In addition, we have shown that equal power transmission maximizes the capacity lower bound for the least favourable channel correlation, i.e. it is a robust transmission scheme. For covariance feedback systems, we have shown that beamforming transmission maximizes the ergodic capacity at low SNR and the capacity increases with channel spatial correlation. We are currently investigating the optimal training length and training power allocation schemes for covariance feedback systems.

Appendix A
Proof of Theorem 2

This is a max-min problem where the capacity lower bound is to be maximized by $\boldsymbol{Q}$ and to be minimized by $\boldsymbol{R_H}$ or effectively $\boldsymbol{R}_{\hat{\boldsymbol{H}}}$. We need to show that $\sup_{\boldsymbol{Q}} \inf_{\boldsymbol{R}_{\hat{\boldsymbol{H}}}} C_{\rm LB}$ is achieved by $\boldsymbol{Q} = \mathcal{P}_d/N_t \boldsymbol{I}_{N_t}$.

From (12) we see that

$$\inf_{\boldsymbol{R}_{\hat{\boldsymbol{H}}}} C_{\rm LB} \leq \frac{N_r \bar{\lambda} \mathcal{P}_d}{\sigma_{\boldsymbol{n}}^2 \ln 2}, \tag{15}$$

where we have chosen $\boldsymbol{R}_{\hat{\boldsymbol{H}}} = \bar{\lambda} \boldsymbol{I}_{N_t}$ for some $\bar{\lambda}$. Therefore,

$$\sup_{\boldsymbol{Q}} \inf_{\boldsymbol{R}_{\hat{\boldsymbol{H}}}} C_{\rm LB} \leq \frac{N_r \bar{\lambda} \mathcal{P}_d}{\sigma_{\boldsymbol{n}}^2 \ln 2}. \tag{16}$$

On the other hand,

$$\sup_{\boldsymbol{Q}} \inf_{\boldsymbol{R}_{\hat{\boldsymbol{H}}}} C_{\rm LB} \geq \inf_{\boldsymbol{R}_{\hat{\boldsymbol{H}}}} C_{\rm LB}(\boldsymbol{Q} = \mathcal{P}_d/N_t \boldsymbol{I}_{N_t}). \tag{17}$$

From *Theorem 1* we know that $C_{\rm LB}(\boldsymbol{Q} = \mathcal{P}_d/N_t \boldsymbol{I}_{N_t})$ is Schur-convex in $\boldsymbol{g}$. Therefore, $\inf_{\boldsymbol{R}_{\hat{\boldsymbol{H}}}} C_{\rm LB}(\boldsymbol{Q} = \mathcal{P}_d/N_t \boldsymbol{I}_{N_t})$ is achieved by $\boldsymbol{R_H} = \bar{g} \boldsymbol{I}_{N_t}$, or effectively $\boldsymbol{R}_{\hat{\boldsymbol{H}}} = \bar{\lambda} \boldsymbol{I}_{N_t}$. Hence,

$$\sup_{\boldsymbol{Q}} \inf_{\boldsymbol{R}_{\hat{\boldsymbol{H}}}} C_{\rm LB} \geq \frac{N_r \bar{\lambda} \mathcal{P}_d}{\sigma_{\boldsymbol{n}}^2 \ln 2}. \tag{18}$$

From (16) and (18), we have $\sup_{\boldsymbol{Q}} \inf_{\boldsymbol{R}_{\hat{\boldsymbol{H}}}} C_{\rm LB} = \frac{N_r \bar{\lambda} \mathcal{P}_d}{\sigma_{\boldsymbol{n}}^2 \ln 2}$ which is achieved by equal power transmission. □

Appendix B
Proof of Theorem 3

We perform the eigenvalue decomposition on $\boldsymbol{R_H}$ as $\boldsymbol{R_H} = \boldsymbol{U G U}^\dagger$. It can be shown from (4) that $\boldsymbol{R}_{\hat{\boldsymbol{H}}} = \boldsymbol{U \Lambda U}^\dagger$, where $\boldsymbol{\Lambda} = \boldsymbol{G} - (\boldsymbol{G}^{-1} + \mu)^{-1}$ and $\mu = \frac{\mathcal{P}_p T_p}{\sigma_{\boldsymbol{n}}^2 N_t}$. Hence, $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{R}_{\hat{\boldsymbol{H}}}$ on its main diagonal, denoted by $\boldsymbol{\lambda} = [\lambda_1 \ \lambda_2 \ ... \ \lambda_{N_t}]^T$. Also, we have

$$\lambda_i = g_i - (g_i^{-1} + \mu)^{-1}, \qquad \forall i = 1, ..., N_t. \tag{19}$$

By examining the first and second derivatives of (19) with respect to g_i, it is easy to see that λ_i is an increasing and convex function of g_i.

Let $\boldsymbol{Q} = \boldsymbol{U} \hat{\boldsymbol{Q}} \boldsymbol{U}^\dagger$, and note that $\mathrm{tr}\{\boldsymbol{Q}\} = \mathrm{tr}\{\hat{\boldsymbol{Q}}\} = \mathcal{P}_d$. Therefore, the capacity lower bound in (12) is reduced to

$$C_{\rm LB} = \frac{N_r}{\sigma_{\boldsymbol{n}}^2 \ln 2} \sum_{i=1}^{k} \lambda_i \hat{q}_i, \tag{20}$$

where k is the rank of $\hat{\boldsymbol{Q}}$ and $\hat{q}_i$ is the ith diagonal entry of $\hat{\boldsymbol{Q}}$. We see that the off-diagonal entries of $\hat{\boldsymbol{Q}}$ have no effect on $C_{\rm LB}$. Hence, we may choose $\hat{\boldsymbol{Q}}$ to be a diagonal matrix. Therefore, $\boldsymbol{Q}$ has the same eigenvectors as $\boldsymbol{R_H}$ and $\boldsymbol{R}_{\hat{\boldsymbol{H}}}$. For any given $\boldsymbol{\lambda}$ with the power constraint $\mathrm{tr}\{\hat{\boldsymbol{Q}}\} = \mathcal{P}_d$, $C_{\rm LB}$ in (20) is maximized if $\hat{q}_i = \mathcal{P}_d$ and $\hat{q}_j = 0, \forall j \neq i$, where i is the index of the largest eigenvalue of $\boldsymbol{R}_{\hat{\boldsymbol{H}}}$, which is also the index of the largest eigenvalue of $\boldsymbol{R_H}$. □

Acknowledgements

This work was supported by the Australian Research Council Discovery Grant 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. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov 1999.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] O. Oyman, R. U. Nabar, H. Bolsckei, and A. J. Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels," *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 2784–2795, Nov. 2003.

[8] C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[9] S. A. Jafar, S. Vishwanath, 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.

[10] 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, Sept. 2004.

[11] 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 PIMRC*, Sept. 2008.

[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] A. W. Marshall and I. Olkin, *Theory of majorization and its application*. Academic Press, 1979.

[14] T. Yoo, E. Yoon, and A. Goldsmith, "MIMO capacity with channel uncertainty: Does feedback help?" in *Proc. IEEE Globecom.*, vol. 1, May 2004, pp. 96–100.