Characterizing the Capacity for MIMO Wireless Channels with Non-zero Mean and Transmit Covariance

Mai Vu and Arogyaswami Paulraj
Information Systems Laboratory, Department of Electrical Engineering
Stanford University, Stanford, CA 94305-9510, USA
E-mails: {mhv, apaulraj}@stanford.edu

Abstract

Transmit channel side information (Tx-CSI), particularly the mean and covariance, helps to increase MIMO wireless capacity. We formulate a dynamic Tx-CSI framework, using outdated channel measurements and channel statistics, to create an effective channel mean and covariance and establish the capacity expression given such Tx-CSI. We then derive the asymptotic capacity-optimal input signal at low SNRs and characterize the asymptotic capacity gains due to Tx-CSI at low and high SNRs. The optimal input signal and the capacity gain in general depend on the relative number of transmit and receive antennas. For systems with equal or fewer transmit than receive antennas, we study a lower bound on the capacity. For systems with more transmit than receive antennas, we characterize the conditions for mode dropping at high SNRs, using simplified channel models. Numerical examples are provided to show the impacts of channel parameters on the optimal input signal and the capacity gain with Tx-CSI.

I. Introduction

Transmit channel side information (Tx-CSI) can enhance MIMO wireless capacity. While Tx-CSI can come in many forms, the statistical Tx-CSI comprising of a mean and a covariance is particularly interesting. Not only that these statistical quantities can be reliably obtained, they also represent a broad class of Tx-CSI involving channel estimates and the associated error covariance, obtained from outdated channel measurements and the channel statistics. We establish an explicit formulation of such Tx-CSI as a function of an estimate quality, taking into account channel temporal variation, and formulate the channel capacity given this Tx-CSI.

The capacity optimal signaling for a MIMO channel with a non-zero mean and a non-identity transmit covariance, however, is still an open problem. Solutions exist for special cases: uncorrelated non-zero mean channels [1], [2] and correlated zero-mean channels [3], [4], where the eigenvectors of the input covariance are given by the eigenvectors of the associated channel parameter in each case. In this paper, we derive the asymptotic optimal input covariance for general statistical Tx-CSI at low SNRs, and characterize the asymptotic capacity gains at both low and high SNRs. For other SNRs, the optimal input signal and the capacity gain depend on the relative numbers of transmit and receive antennas. For systems with equal or fewer transmit than receive antennas, we utilize a tight lower bound on capacity based on the Jensen inequality. For systems with more transmit than receive antennas, we derive the conditions for mode dropping at all SNRs for representative channels with a high K factor or strong transmit correlation. These conditions provide intuition as to when it is optimal to activate only a fraction of the available eigen-modes at all SNRs, given the Tx-CSI. We then provide numerical examples of how the estimate quality affects the channel capacity in both system configurations.

The paper is organized as follows: We first discuss the MIMO channel model in Section II, then establish a robust Tx-CSI framework involving a channel estimate and its error covariance in Section III. In Section IV, we formulate the ergodic capacity expression given this Tx-CSI and review special cases optimal inputs. Section V presents asymptotic capacity results for low and high SNRs. Next, we characterize the capacity optimal input signal in Section VI, differentiating the two cases based on the relative number of antennas. We provide numerical examples in Section VII and conclude in Section VIII.

II. CHANNEL MODEL

We consider a frequency flat, quasi-static block fading MIMO channel with N transmit and M receive antennas, represented by a random matrix $\mathbf{H}$ of size $M \times N$. We assume that the channel is Gaussian distributed with a mean $\bar{\mathbf{H}}$ and a covariance $\mathbf{R}_0$, thus $\mathbf{H}$ can be decomposed as

$$\mathbf{H} = \bar{\mathbf{H}} + \tilde{\mathbf{H}} \ , \tag{1}$$

where $\tilde{\mathbf{H}}$ represents the zero-mean Gaussian component. The channel covariance $\mathbf{R}_0$ of size $MN \times MN$ is defined as

$$\mathbf{R}_0 = E[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^*], \tag{2}$$

where $\tilde{\mathbf{h}} = \text{vec}(\tilde{\mathbf{H}})$. In other words, $\mathbf{h} \sim \mathcal{CN}(\bar{\mathbf{h}}, \mathbf{R}_0)$, where $\mathbf{h} = \text{vec}(\mathbf{H})$ and $\bar{\mathbf{h}} = \text{vec}(\bar{\mathbf{H}})$. The ratio of the power in the channel mean to the average power in the channel variable component is the channel K factor, or the Rician factor, defined as

$$K = \frac{||\bar{\mathbf{H}}||_F^2}{\text{tr}(\mathbf{R}_0)} \ , \tag{3}$$

where $||.||_F$ is the matrix Frobenius norm, and $\text{tr}(.)$ is the trace of a matrix. Often, the channel covariance is assumed to have a Kronecker structure, representing separable transmit and receive antenna correlations:

$$\mathbf{R}_0 = \mathbf{R}_t^T \otimes \mathbf{R}_r \ , \tag{4}$$

where $\mathbf{R}_t$, size $N \times N$, and $\mathbf{R}_r$, size $M \times M$, are the transmit and receive covariance, respectively. The channel statistics, $\bar{\mathbf{H}}$ and $\mathbf{R}_0$ (or $\mathbf{R}_t$ and $\mathbf{R}_r$), can be obtained by averaging instantaneous channel measurements over tens of channel coherence times; they remain valid for periods of tens to hundreds coherence time, during which, the channel can be considered as (short-term) stationary [5].

III. A DYNAMIC TX-CSI FRAMEWORK

Transmit channel side information, or Tx-CSI, can be obtained by using the reverse-channel measurements, invoking the reciprocity principle, or via feedback from the receiver. In either case, there is usually a delay from when the channel information is obtained to when it is used by the transmitter; for example, a scheduling or a feedback delay. This delay may affect the reliability of the Tx-CSI obtained. Instantaneous channel measurements provide the most potential gain in system capacity; however, it is susceptible to channel variations because of the delay. Channel statistics, on the other hand, change much slower than the channel itself and can be obtained reliably, but they provide less gain. Our goal here is to utilize both of these forms to create a Tx-CSI framework that is robust to channel variation, while optimally capturing the potential gain.

Assume that we have at the transmitter a channel measurement $\mathbf{H}_0$ at time 0. We aim to establish an estimate of the current channel $\mathbf{H}_s$ at the transmit time s. The channel measurement is correlated with the current channel; this correlation is captured by the channel auto-covariance, defined as

$$\mathbf{R}_s = E[\tilde{\mathbf{h}}_0\tilde{\mathbf{h}}_s^*] \ , \tag{5}$$

where $\tilde{\mathbf{h}}_0 = \text{vec}(\mathbf{H}_0)$ and $\tilde{\mathbf{h}}_s = \text{vec}(\mathbf{H}_s)$. Due to stationarity, the auto-covariance $\mathbf{R}_s$ depends only on the time difference s, but not on the absolute time. Again, $\mathbf{R}_s$ can be obtained by an averaging operation over instantaneous channel measurements.

Given the channel measurement $\mathbf{H}_0$ and the statistics $\bar{\mathbf{H}}$, $\mathbf{R}_0$, and $\mathbf{R}_s$, an estimate of the channel at time s follows from MMSE estimation theory [6] as

$$\begin{aligned} \hat{\mathbf{h}}_s &= E\big[\mathbf{h}_s|\mathbf{h}_0\big] &&= \bar{\mathbf{h}} + \mathbf{R}_s^*\mathbf{R}_0^{-1}\big[\mathbf{h}_0 - \bar{\mathbf{h}}\big] \\ \mathbf{R}_{e,s} &= \text{cov}\big[\mathbf{h}_s|\mathbf{h}_0\big] &&= \mathbf{R}_0 - \mathbf{R}_s^*\mathbf{R}_0^{-1}\mathbf{R}_s \ , \end{aligned} \tag{6}$$

where $\hat{\mathbf{h}}_s = \text{vec}\big(\hat{\mathbf{H}}_s\big)$ is the estimated channel, and $\mathbf{R}_{e,s}$ is the estimation error covariance at time s. These become an effective mean and effective covariance of the channel, respectively.

The auto-covariance $\mathbf{R}_s$ captures both the channel antenna correlation and the temporal correlation effects, while $\mathbf{R}_0$ (2) represents the antenna correlation alone. If we assume that the MIMO temporal correlation is homogeneous, i.e., all the scalar channels between the N transmit and the M receive antennas have the same temporal correlation factor ρ_s, then we can separate the temporal correlation from the antenna correlation as

$$\mathbf{R}_s = \rho_s \mathbf{R}_0 \ . \tag{7}$$

The channel temporal correlation ρ_s is a function of the Doppler spread f_d and the delay s. For example, in Jake's model [7], $\rho_s = J_0(2\pi f_d s)$, where J_0 is the zeroth order Bessel function of the first kind. In general, we assume that $-1 \leq \rho \leq 1$, and $\rho = 1$ only at a zero delay. Using simplified auto-covariance model (7), the channel estimate and its error covariance become

$$\hat{\mathbf{H}}_s = \rho_s \mathbf{H}_0 + (1 - \rho_s)\,\bar{\mathbf{H}} \ , \qquad \mathbf{R}_{e,s} = \left(1 - \rho_s^2\right)\mathbf{R}_0 \ . \tag{8}$$

For the Kronecker antenna correlation model (4), the estimated channel has effective antenna correlations as:

$$\mathbf{R}_{t,s} = \left(1 - \rho_s^2\right)\mathbf{R}_t \ , \qquad \mathbf{R}_{r,s} = \left(1 - \rho_s^2\right)\mathbf{R}_r \ . \tag{9}$$

ρ_s here acts as the estimate quality dependent on the time delay s. For a short delay, ρ_s is close to one; the estimate thus depends heavily on the initial channel measurement, and the error covariance is small. As the delay increases, $|\rho_s|$ will decrease to zero, reducing the impact of the initial measurement. The estimate then moves toward the channel mean $\bar{\mathbf{H}}$, and the error covariance grows toward the channel covariance $\mathbf{R}_0$. Therefore, the estimate and its error covariance (8) constitute a form of Tx-CSI, ranging between perfect channel knowledge (when $\rho = 1$) and channel statistics (when $\rho = 0$). By taking into account the channel temporal variation, this framework dynamically captures the available channel information and creates robust Tx-CSI.

IV. Channel Ergodic Capacity Formulation

We study the ergodic capacity of a MIMO channel, assuming perfect receiver channel knowledge and the dynamic Tx-CSI framework (8), given an estimate quality ρ_s. The input signal is subject to an average sum power constraint, such that the input covariance $\mathbf{Q}$ must satisfy $\text{tr}(\mathbf{Q}) = 1$. For each initial channel measurement $\mathbf{H}_0$ with the corresponding Tx-CSI value $\{\hat{\mathbf{H}}_s, \mathbf{R}_s\}$, the average mutual information is maximized by a zero-mean complex Gaussian input signal [8] with the covariance matrix as the solution of the following optimization problem:

$$\begin{aligned} \mathcal{I}_s(\mathbf{H}_0) \ &= \ \max_{\mathbf{Q}} \ \ \mathcal{F}(\mathbf{Q}) \ = \ E_{\mathbf{H}}\big[\text{logdet}(\mathbf{I} + \gamma \mathbf{H}\mathbf{Q}\mathbf{H}^*)\big] \\ & \qquad\ \ \text{s.t.} \ \ \ \text{tr}(\mathbf{Q}) \ = \ 1 \ , \end{aligned} \tag{10}$$

where γ is the SNR and $\mathbf{h} \sim \mathcal{CN}\left(\hat{\mathbf{h}}_s, \mathbf{R}_{e,s}\right)$; i.e. the expectation is evaluated over the effective channel statistics with mean $\hat{\mathbf{H}}_s$ and covariance $\mathbf{R}_{e,s}$. The channel ergodic capacity, given the estimate quality ρ_s (or $\mathbf{R}_s$ in the general case) is then

$$\mathcal{C} = E_{\mathbf{H}_0}\left[\mathcal{I}_s(\mathbf{H}_0)\right] \ , \tag{11}$$

where, denoting $\mathbf{h}_0 = \text{vec}(\mathbf{H}_0)$, $\mathbf{h}_0 \sim \mathcal{N}\left(\bar{\mathbf{h}}, \mathbf{R}_0\right)$. Establishing the channel capacity with the estimate Tx-CSI (8) thus essentially requires finding the optimal input signal and mutual information for a channel given the channel mean and covariance as formulated in (10).

Review of the optimal eigen-beam directions in special cases: The eigenvectors of the optimal input covariance solution of problem (10) is analytically known in several special Tx-CSI cases: (a) correlation Tx-CSI involving a zero effective mean ($\hat{\mathbf{h}}_s = \mathbf{0}$) with a Kronecker structured effective covariance $\mathbf{R}_{e,s} = \mathbf{R}_{t,s}^T \otimes \mathbf{R}_{r,s}$; and (b) mean Tx-CSI involving an arbitrary effective mean $\hat{\mathbf{H}}_s$ with an identity covariance ($\mathbf{R}_{e,s} = \mathbf{I}$). Let the eigenvalue decomposition of the input covariance be

$$\mathbf{Q} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^* \; , \tag{12}$$

then the columns of $\mathbf{U}$ are the orthogonal eigen-beam directions (patterns), and $\mathbf{\Lambda}$ represents the power allocation on these beams. In case (a), let $\mathbf{R}_{t,s} = \mathbf{U}_{t,s} \mathbf{\Lambda}_{t,s} \mathbf{U}_{t,s}^*$ be the eigenvalue decomposition of $\mathbf{R}_{t,s}$, then the optimal beam directions are $\mathbf{U} = \mathbf{U}_{t,s}$ [1], [2]. In case (b), let $\hat{\mathbf{H}}_s^* \hat{\mathbf{H}}_s = \hat{\mathbf{U}}_s \hat{\mathbf{\Lambda}}_s \hat{\mathbf{U}}_s^*$ be the eigenvalue decomposition of $\hat{\mathbf{H}}_s^* \hat{\mathbf{H}}_s$, then the optimal beam directions are $\mathbf{U} = \hat{\mathbf{U}}_s$ [3], [4]. The optimal power allocation $\mathbf{\Lambda}$ in both cases, however, requires a numerical solution. The optimal input covariance in the general case (a non-zero mean with a non-identity covariance) is so far not known analytically.

Using the Kronecker correlation structure (4), we assume that only transmit antenna correlation $\mathbf{R}_t$ exists and is full-rank (i.e. $\mathbf{R}_r = \mathbf{I}$). Consider problem (10) for an arbitrary effective mean $\hat{\mathbf{H}}_s$ and an effective transmit covariance $\mathbf{R}_{t,s}$ (9). We will examine the optimal input covariance and the capacity gain asymptotically at low and high SNRs. From these results, we characterize the capacity using a sub-optimal input solution based on the Jensen inequality, and derive the optimal conditions for mode dropping at all SNRs for simplified Tx-CSI cases.

V. Asymptotic SNR capacity results

A. Low SNR capacity results

Lemma 1: (Low SNR optimal input covariance) *As the SNR $\gamma \to 0$, the optimal input covariance of problem (10) converges to a unit-rank matrix with a unit eigenvalue and the corresponding eigenvector given by the dominant eigenvalue of $\mathbf{G}_s = \hat{\mathbf{H}}_s^* \hat{\mathbf{H}}_s + M \mathbf{R}_{t,s}$.*

In other words, the optimal input at low SNRs becomes a single-mode beamforming signal matched to the dominant eigenvector of $\mathbf{G}_s$.

Proof: Using the Taylor series, the function $f = \log\det(I + \gamma \mathbf{A})$, where $\mathbf{A}$ is a positive semi-definite matrix, can be expanded as a polynomial of γ as

$$f = \text{tr}(\mathbf{A})\gamma - \text{tr}(\mathbf{A}^2)\gamma^2 + \text{tr}(\mathbf{A}^3)\gamma^3 - \ldots$$

Noting that $E_{\mathbf{H}}\left[\text{logdet}(\mathbf{I} + \gamma \mathbf{H}\mathbf{Q}\mathbf{H}^*)\right] = E_{\mathbf{H}}\left[\text{logdet}(\mathbf{I} + \gamma \mathbf{H}^*\mathbf{H}\mathbf{Q})\right]$ and applying the above expansion, at low SNRs ($\gamma \to 0$), the mutual information $\mathcal{F}(\mathbf{Q})$ in (10) approaches

$$\mathcal{F}(\mathbf{Q}) \overset{\gamma \to 0}{\approx} E_{\mathbf{H}}[\text{tr}(\mathbf{H}^*\mathbf{H}\mathbf{Q})\gamma] = \gamma \, \text{tr}\left(E_{\mathbf{H}}[\mathbf{H}^*\mathbf{H}]\mathbf{Q}\right) = \gamma \, \text{tr}\left[\mathbf{G}_s \mathbf{Q}\right] \; .$$

Maximizing the above expression with the constraint $\text{tr}(\mathbf{Q}) = 1$ results in the optimal $\mathbf{Q}$ stated in Lemma 1. □

Lemma 2: (Low SNR capacity ratio gain) *As the SNR $\gamma \to 0$, the ratio between the optimal mutual information in (10) and the value obtained by equi-power isotropic input approaches*

$$r = \frac{N \lambda_{\max}(\mathbf{G}_s)}{tr(\mathbf{G}_s)} \; . \tag{13}$$

This ratio scales linearly with the number of transmit antennas and is related to the condition of the channel correlation matrix $\mathbf{G}_s = E[\mathbf{H}^*\mathbf{H}]$.

Proof: From the proof of Lemma 1, the optimal mutual information with Tx-CSI in (10) at low SNRs approaches $\mathcal{I}_s \overset{\gamma \to 0}{=} \gamma \lambda_{\max}(\mathbf{G}_s)$. The mutual information with equi-power allocation, on the other hand, equals $\mathcal{I}_0 \overset{\gamma \to 0}{=} \frac{\gamma}{N} \text{tr}(\mathbf{G}_s)$. Taking the ratio between these two expressions, $r = \mathcal{I}_s / \mathcal{I}_0$, yields (13). □

B. *High SNR capacity results*

At high SNRs, we distinguish between two antenna configurations: when the number of transmit antennas is equal or fewer than the number of receive antennas ($N \leq M$), and when it is larger ($N > M$).

1) Systems with equal or fewer transmit than receive antennas: When $N \leq M$, the asymptotic high SNR optimal input covariance of (10) is $\frac{1}{N}\mathbf{I}$: Since $\mathbf{H}^*\mathbf{H}$ is full-rank, the mutual information $\mathcal{F}(\mathbf{Q})$ at high SNRs can be decomposed as

$$\mathcal{F} \overset{\gamma \to \infty}{\approx} E_{\mathbf{H}}[\log\det(\gamma \mathbf{H}^*\mathbf{H}\mathbf{Q})] = E_{\mathbf{H}}\left[\log\det(\mathbf{H}^*\mathbf{H})\right] + \log\det(\gamma \mathbf{Q}) \ . \tag{14}$$

Maximizing the above expression, subject to $\text{tr}(\mathbf{Q}) = 1$, leads to $\mathbf{Q} = \mathbf{I}/N$. In other words, when $N \leq M$, the optimal input covariance at high SNRs approaches equi-power in all directions and is independent of the Tx-CSI. This is a well-known result that, for these systems, the capacity gain due to Tx-CSI diminishes at high SNRs.

2) Systems with more transmit than receive antennas: When $N > M$, in contrary, Tx-CSI provides capacity gain at all SNRs. The decomposition (14) does not apply in this case, and the optimal input covariance of (10) at high SNRs depends on the channel statistics, or the Tx-CSI $\{\hat{\mathbf{H}}_s, \mathbf{R}_{t,s}\}$. While an analytical optimal covariance for arbitrary $\hat{\mathbf{H}}_s$ and $\mathbf{R}_{t,s}$ is still unknown, the capacity gain is maximum with perfect Tx-CSI ($\hat{\mathbf{H}}_s = \mathbf{H}_0,\ \mathbf{R}_{t,s} = \mathbf{0}$), in which case this gain can be accurately quantified.

Lemma 3: (High SNR incremental capacity gain) *At high SNRs, the incremental capacity gain due to perfect Tx-CSI ($\rho = 1$) over the mutual information obtained by equi-power isotropic input equals*

$$\Delta \mathcal{C} = M \log\left(\frac{N}{M}\right) \ . \tag{15}$$

This gain scales linearly with the number of receive antennas and depends on the ratio of the number of transmit to receive antennas.

Proof: With perfect Tx-CSI, the solution for (10) is standard water-filling [8] on $\mathbf{H}_0^*\mathbf{H}_0$. Let σ_i^2 be the eigenvalues of $\mathbf{H}_0^*\mathbf{H}_0$, then the optimal eigenvalues of $\mathbf{Q}$ are $\lambda_i = \left(\mu - \frac{1}{\gamma\sigma_i^2}\right)_+$, where μ is chosen to satisfy $\sum_i \lambda_i = 1$. The capacity (11) then becomes $\mathcal{C} = \sum_{i=1}^M E_{\sigma_i}\left[\log\left(\mu\gamma\sigma_i^2\right)\right]$, where σ_i^2 has the distribution of the underlying Wishart matrix eigenvalues. As $\gamma \to \infty$, $\mu \to \frac{1}{M}$, and the capacity approaches

$$\mathcal{C} \overset{\gamma \to \infty}{\approx} M \log\left(\frac{1}{M}\right) + M\log(\gamma) + \sum_{i=1}^M \log(\sigma_i^2) \ . \tag{16}$$

Assuming no Tx-CSI and using an equi-power isotropic input with covariance $\mathbf{Q} = \mathbf{I}/N$, the ergodic mutual information is given by $\mathcal{C}_0 = \sum_{i=1}^M E_{\sigma_i}\left[\log\left(1 + \frac{1}{N}\gamma\sigma_i^2\right)\right]$. At high SNRs, this expression approaches

$$\mathcal{C}_0 \overset{\gamma \to \infty}{\approx} M \log\left(\frac{1}{N}\right) + M\log(\gamma) + \sum_{i=1}^M \log(\sigma_i^2) \ . \tag{17}$$

Subtracting (16) and (17) side-by-side yields the capacity gain in (15). □

VI. CAPACITY CHARACTERIZATIONS

A. Systems with equal or fewer transmit than receive antennas

1) Jensen input covariance: To obtain a lower bound on the capacity, we explore the input covariance that is optimal for Jensen's bound [8] on the average mutual information $\mathcal{F}$ in (10). Due to the concavity of $\mathcal{F}$, this bound applies as

$$\mathcal{F}(\mathbf{Q}) = E_{\mathbf{H}}\big[\text{logdet}(\mathbf{I} + \gamma \mathbf{Q}\mathbf{H}^*\mathbf{H})\big] \leq \text{logdet}\left(\mathbf{I} + \gamma \mathbf{Q} E[\mathbf{H}^*\mathbf{H}]\right) \ .$$

Perform the eigenvalue decompositions of $\mathbf{G}_s = E[\mathbf{H}^*\mathbf{H}] = \hat{\mathbf{H}}_s^*\hat{\mathbf{H}}_s + M\mathbf{R}_{t,s}$ as $\mathbf{G}_s = \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^*$, the optimal input eigen-directions and the power allocation that maximize the Jensen bound are

$$\mathbf{U} = \mathbf{U}_s \ \ , \quad \lambda_i = \left(\mu - \frac{1}{\gamma\lambda_{s,i}}\right)_+ \tag{18}$$

where $\lambda_{s,i}$ are diagonal values of $\mathbf{\Lambda}_s$. Let $\mathbf{\Lambda}_{J,s} = \text{diag}(\lambda_i)$, condition (18) results in the Jensen input covariance $\mathbf{Q}_{J,s} = \mathbf{U}_s\mathbf{\Lambda}_{J,s}\mathbf{U}_s^*$.

For all Tx-CSI, this Jensen input covariance $\mathbf{Q}_{J,s}$ approaches the optimal input covariance at low SNRs in Lemma 1. In the two special Tx-CSI cases, correlation Tx-CSI (zero effective mean) and mean Tx-CSI (identity covariance), the beam directions (18) coincide with the optimal directions in each respective case for all SNRs; only the optimal power allocation is then approximated.

2) A lower bound capacity approximation: Using the above Jensen optimal input covariance $\mathbf{Q}_{J,s}$ in the average mutual information expression results in a value $\mathcal{J}_s = \mathcal{F}(\mathbf{Q}_{J,s})$ such that $\mathcal{J}_s < \mathcal{I}_s$ in (10). Averaging $\mathcal{J}_s$ over the initial channel measurement distribution (11), we obtain a lower bound approximation of the channel ergodic capacity as

$$\mathcal{C}_J = E_{\mathbf{H}_0}[\mathcal{J}_s] \ . \tag{19}$$

The tightness of this bound depends on the antenna configurations.

For $N \leq M$, $\mathcal{J}_s$ closely approximates $\mathcal{I}_s$: in this case, the Jensen input covariance solution approaches the optimal solution at both low and high SNRs. Any difference between $\mathcal{J}_s$ and $\mathcal{I}_s$ thus occurs only at a mid-range SNR and, as confirmed by simulations [9], is usually negligible. Therefore, the Jensen input covariance $\mathbf{Q}_{J,s}$ can be used in deriving a tight lower bound on the channel capacity for $N \leq M$.

B. Systems with more transmit than receive antennas

For $N > M$, the optimal input covariance, particularly the optimal power allocation, depends heavily on the channel effective mean and covariance matrices. If the channel is uncorrelated with zero mean (i.e. an i.i.d Rayleigh fading channel), then the optimal input covariance is the identity matrix [10] (i.e. equi-power allocation). However, if the mean is strong, characterized by a high K factor, or if the transmit antenna is strongly correlated, the optimal power allocation may drop modes even at high SNRs. This optimal input covariance at high SNRs then differs from the Jensen solution, which approaches equi-power. To characterize effects of the K factor and the transmit antenna correlation on the optimal power allocation, we study two simple channel models. Each model results in the optimal allocation having only two distinct power levels, and we examine the conditions which lead to dropping the lower power level at all SNRs.

1) Effects of the K factor: Consider an uncorrelated channel with a mean $\mathbf{H}_m$ such that

$$\mathbf{H}_m\mathbf{H}_m^* = \frac{K}{K+1}\mathbf{I}_M \ , \tag{20}$$

and with transmit covariance $\mathbf{R}_t = \frac{1}{K+1}\mathbf{I}_N$ (assuming receive covariance $\mathbf{R}_r = \mathbf{I}$). For convenience, let $\beta = \sqrt{K/(K+1)}$, $(0 \leq \beta \leq 1)$. The optimal power allocation for this channel can be completely characterized by the K factor or β, representing the mean matrix, and the SNR. Due to symmetry, this optimal solution contains only two different power levels: λ_1 for the first M eigen-modes, corresponding to the non-zero eigen-modes of $\mathbf{H}_m^*\mathbf{H}_m$, and λ_2 for the rest. We are interested in the condition resulting in $\lambda_1 = \frac{1}{M}$ and $\lambda_2 = 0$. The mutual information optimization problem (10) becomes

$$\begin{aligned} \max_{\lambda_1,\lambda_2} \quad & E_{\mathbf{h}_i,\mathbf{h}_j}\left[\log\det\left(\mathbf{I}_M + \lambda_1\gamma\sum_{i=1}^{M}\mathbf{h}_i\mathbf{h}_i^* + \lambda_2\gamma\sum_{j=M+1}^{N}\mathbf{h}_j\mathbf{h}_j^*\right)\right] \\ \text{s.t.} \quad & M\lambda_1 + (N-M)\lambda_2 = 1 \\ & \lambda_1 \geq 0,\ \lambda_2 \geq 0\ , \end{aligned} \tag{21}$$

where $\mathbf{h}_i \sim \mathcal{N}\left(\sqrt{\beta}\mathbf{e}_i, (1-\beta)\mathbf{I}_M\right)$, $\mathbf{h}_j \sim \mathcal{N}\left(\mathbf{0}, (1-\beta)\mathbf{I}_M\right)$, and $\mathbf{e}_i$ is the vector with the i^{th} element equals to 1 and the rest are zero. Since (21) is a convex problem, to have the optimal λ_1 as $\frac{1}{M}$, it is sufficient and necessary that $\left.\frac{dg(\lambda_1)}{d\lambda_1}\right|_{\lambda_1=1/M} \geq 0$, which can be simplified to

$$\operatorname{tr}\left(E_{\mathbf{h}_j}\left[\left(\mathbf{I} + \frac{\gamma}{M}\sum_{j=1}^{M}(\beta\mathbf{e}_j + \mathbf{h}_j)(\beta\mathbf{e}_j + \mathbf{h}_j)^*\right)^{-1}\right]\right) \leq \frac{N}{1+\gamma}\ . \tag{22}$$

From this expression, a threshold for K, above which mode dropping occurs, can be derived. At high SNRs ($\gamma \to \infty$), condition (22) becomes

$$\operatorname{tr}\left(E_{\mathbf{h}_j}\left[\left(\sum_{j=1}^{M}(\beta\mathbf{e}_j + \mathbf{h}_j)(\beta\mathbf{e}_j + \mathbf{h}_j)^*\right)^{-1}\right]\right) \leq \frac{N}{M}\ . \tag{23}$$

The matrix expression under the above expectation has the inverted non-central complex Wishart distribution. Since at high K ($\beta \to 1$), the left-hand-side of (23) approaches M, mode-dropping at all SNRs occurs for this Tx-CSI only if $N \geq M^2$.

2) *Effects of the transmit antenna correlation:* Consider a zero-mean channel with a transmit covariance matrix having only two distinct eigenvalues: $\lambda_1(\mathbf{R}_t) = \ldots = \lambda_L(\mathbf{R}_t) = \xi_1$, and $\lambda_{L+1}(\mathbf{R}_t) = \ldots = \lambda_N(\mathbf{R}_t) = \xi_2$, where $N > L > M$ (note that $N \geq M+2$ here, for a reason that will become clear later), and $\xi_1 > \xi_2$. The eigenvectors of $\mathbf{R}_t$ have no effect on the power allocation and hence are not considered here. Similarly due to symmetry, the optimal power allocation has only two levels: λ_1 for the first L eigen-modes corresponding to the L larger eigenvalues of $\mathbf{R}_t$, and λ_2 for the rest. The mutual information optimization problem (10) is now equivalent to

$$\begin{aligned} \max_{\lambda_1,\lambda_2} \quad & E_{\mathbf{h}_i}\left[\log\det\left(\mathbf{I}_M + \gamma\lambda_1\xi_1\sum_{j=1}^{L}\mathbf{h}_i\mathbf{h}_i^* + \gamma\lambda_2\xi_2\sum_{j=L+1}^{N}\mathbf{h}_i\mathbf{h}_i^*\right)\right] \\ \text{s.t.} \quad & L\lambda_1 + (N-L)\lambda_2 = 1 \\ & \lambda_1 \geq 0\ ,\quad \lambda_2 \geq 0\ , \end{aligned} \tag{24}$$

where $\mathbf{h}_i \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}_M\right)$ are i.i.d. We are interested in the condition that results in $\lambda_2 = 0$, which clearly depends on the condition number of $\mathbf{R}_t$, i.e. the ratio $\kappa = \xi_1/\xi_2$.

Due to the concavity of the mutual information expression, the optimal $\lambda_2 = 0$ is achieved iff $\left.\frac{\partial f}{\partial\lambda_2}\right|_{\lambda_2=0} \leq 0$, which can be simplified to

$$\operatorname{tr}\left(E_{\mathbf{h}_i}\left[\left(\mathbf{I}_M + \frac{\gamma\xi_1}{L}\sum_{j=1}^{L}\mathbf{h}_i\mathbf{h}_i^*\right)^{-1}(\gamma\xi_2 + 1)\right]\right) \leq M\ .$$

At high SNRs ($\gamma \to \infty$), the above expression becomes

$$\frac{L\xi_2}{\xi_1}\mathrm{tr}\left(E_{\mathbf{h}_i}\left[\left(\sum_{j=1}^{L}\mathbf{h}_i\mathbf{h}_i^*\right)^{-1}\right]\right) \leq M \ . \tag{25}$$

Let $\mathbf{B} = \sum_{j=1}^{L}\mathbf{h}_i\mathbf{h}_i^*$, then $\mathbf{B}$ is a complex central Wishart matrix with rank M and L degrees of freedom: $\mathbf{B} \sim \mathcal{W}_M^C(L, \mathbf{I}_M)$. Using the first moment of an inverted Wishart matrix, $E[\mathbf{B}^{-1}] = \frac{1}{L-M}\mathbf{I}_M$ [11] [12], condition (25) translates to

$$\frac{\xi_1}{\xi_2} \geq \frac{L}{L-M} \ . \tag{26}$$

Since $N > L > M$, we have $L/(L-M) \leq N-1$, and a looser bound on $\mathbf{R}_t$ condition number for dropping the weaker eigen-modes at all SNRs can be obtained as

$$\kappa \geq N - 1 \ . \tag{27}$$

Note that when $L = M$, the minimum eigenvalue of $\mathbf{B}$ has the distribution $f_{\lambda_{\min}}(\lambda) = \frac{M}{2}\exp\left(-\frac{\lambda M}{2}\right)$ [13], thus $E[1/\lambda_{\min}]$ is infinity for all M. This fact implies that (25) can not hold for $L = M$; thus in that case, the optimal power allocation will activate all modes at high SNRs. In other words, as the SNR approaches infinity, the optimal power allocation for this correlation Tx-CSI will always use more than M (at least $M+1$) modes; hence mode dropping at high SNRs occurs only if $N \geq M+2$.

3) Remarks: The two conditions (23) and (26), although specific to each respective Tx-CSI model, provide intuition on effects of the channel mean and the transmit antenna correlation on the optimal input power allocation. These conditions for channels with both a non-zero mean and a transmit antenna correlation are likely to be further relaxed, such that mode dropping occurs at all SNRs for even a lower K factor or a lower correlation condition number. Subsequently, channels with high K or strong correlation tends to result in mode dropping with statistical Tx-CSI at all SNRs.

VII. Numerical Examples

A. Systems with equal or fewer transmit than receive antennas

For these systems, Tx-CSI helps to increase the capacity only at low SNRs. Using the lower bound approximation (19) at SNR = 4dB, we plot in Figure 1 the capacity (11) as a function of the estimate quality ρ for two 4×4 channels: an i.i.d channel and a zero-mean correlated channel with the covariance matrix given in the Appendix. We observe that the capacity increases with higher ρ, but this increment due to channel estimates is significant only with relatively good estimates ($\rho \geq 0.6$ in this case). Moreover, the range of capacity increase due to ρ for an i.i.d channel is larger than that for a correlated channel. For reference, capacity of the correlated channel without Tx-CSI is also included, illustrating that knowing the channel statistics alone ($\rho = 0$) can enhance the capacity over no Tx-CSI.

B. Systems with more transmit than receive antennas

For systems with excess transmit antennas, mode dropping can occur at high SNRs for mean and correlation Tx-CSI, depending on the K factor or transmit antenna correlation. We plot in Figure 2 the K factor threshold versus the SNR for mean Tx-CSI in (22) for systems with 2 receive antennas and from 4 to 6 transmit antennas. When K is above this threshold, signifying a strong channel mean or a good channel estimate, the optimal power allocation activates only two modes and dropping the rest at all SNRs. The threshold depends on the number of transmit antennas N; it decreases with larger N.

We then examine the correlation Tx-CSI case in (24) and plot in Figure 3 the optimal power allocation, using convex optimization numerical programs, for a 4×2 zero-mean channel

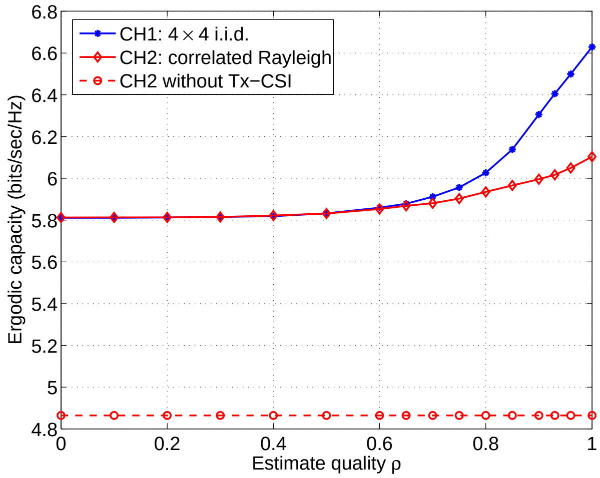

Fig. 1. The capacity (11) versus the estimate quality ρ for two 4×4 channels at SNR = 4dB.

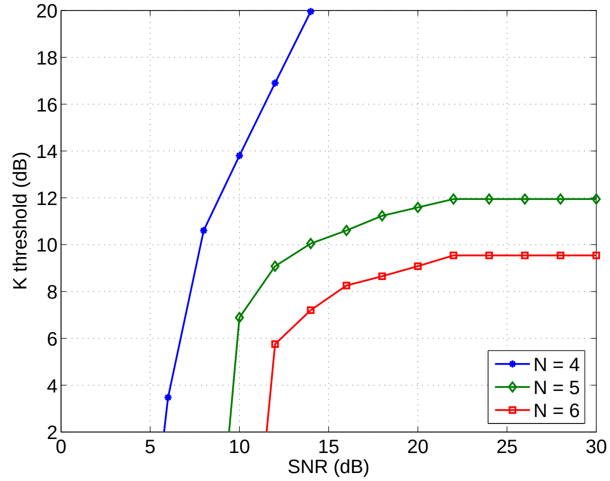

Fig. 2. K factor thresholds for systems with 2 receive and N transmit antennas, above which using 2 modes is capacity optimal for mean Tx-CSI (20).

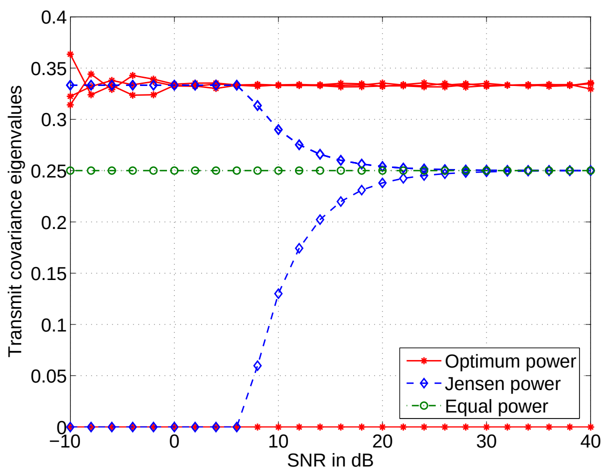

Fig. 3. Input power allocations for a 4×2 zero-mean channel with transmit covariance eigenvalues $[1.25 \;\; 1.25 \;\; 1.25 \;\; 0.25]$.

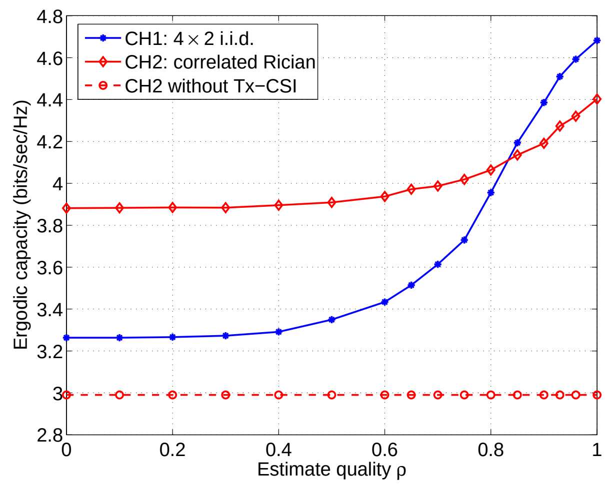

Fig. 4. The capacity (11) versus the estimate quality ρ for two 4×2 channels at SNR = 4dB.

with transmit covariance eigenvalues as $[1.25 \;\; 1.25 \;\; 1.25 \;\; 0.25]$. This covariance matrix has the condition number $\kappa = 5 > 3$, satisfying (26). The optimal power allocation therefore only activates 3 modes, dropping 1 mode, at all SNRs.

Finally we plot in Figure 4 the capacity (11) at SNR = 4dB as the function of ρ for two 4×2 channels: an i.i.d channel and a correlated Rician channel with the mean and covariance given in the Appendix. Due to the low SNR, the lower bound (19) is tight and is again used to approximate the optimal mutual information $\mathcal{I}_s(\mathbf{H}_0)$ in (11). Besides similar observations as those for Figure 1, this figure also shows that Rician correlated channels can have higher capacity than an i.i.d channel at low SNRs. Note that at higher SNRs, Tx-CSI results in more capacity gain for these systems, of up to 2 bps/Hz (15).

VIII. Conclusion

We have formulated a dynamic Tx-CSI framework for MIMO wireless in the form of an effective channel mean and covariance, based on a channel estimate and its error covariance obtained from an outdated channel measurement and channel statistics, and established the capacity expression given such Tx-CSI. We derive the asymptotic capacity-optimal input signal at low SNRs and characterize the asymptotic capacity gains with CSIT at both low and high SNRs. For systems with equal or fewer transmit than receive antennas, the capacity can be tightly approximated by a lower bound based on the Jensen's inequality. For systems with more transmit than receive antennas, we characterize the conditions for mode dropping at all SNRs, using simplified channel models. With N transmit antennas, a strong channel mean (with a high K factor) or a strong transmit antenna correlation (with a large condition number) can lead to an optimal power allocation that activates fewer than the maximum N modes at all SNRs. Utilizing the dynamic Tx-CSI model, we demonstrate numerically that the capacity given the Tx-CSI increases with a better channel estimate quality, and the increment depends on the channel statistics – the means and the transmit covariance. Significant capacity increase due to the channel estimate requires good estimate quality (roughly $\rho \geq 0.6$), and the increase is larger for i.i.d channels than for correlated channels. In all cases, however, Tx-CSI usually helps to increase the MIMO capacity over no Tx-CSI.

APPENDIX

The channel parameters used in the simulations of Figures 1 and 4 are listed below.

The transmit covariance matrix is

$$\mathbf{R}_t = \begin{bmatrix} 0.8758 & -0.0993-0.0877i & -0.6648-0.0087i & 0.5256-0.4355i \\ -0.0993+0.0877i & 0.9318 & 0.0926+0.3776i & -0.5061-0.3478i \\ -0.6648+0.0087i & 0.0926-0.3776i & 1.0544 & -0.6219+0.5966i \\ 0.5256+0.4355i & -0.5061+0.3478i & -0.6219-0.5966i & 1.1379 \end{bmatrix}.$$

This matrix has the eigenvalues $[2.717\ ,\ 0.997\ ,\ 0.237\ ,\ 0.049]$ and a condition number of 55.5. The transmit antennas therefore are quite strongly correlated.

The mean for the 4×2 channel is

$$\bar{\mathbf{H}} = \begin{bmatrix} 0.0749-0.1438i & 0.0208+0.3040i & -0.3356+0.0489i & 0.2573-0.0792i \\ 0.0173-0.2796i & -0.2336-0.2586i & 0.3157+0.4079i & 0.1183+0.1158i \end{bmatrix}.$$

The K factor here is 0.1.

ACKNOWLEDGMENT

Mai Vu would like to acknowledge the support of the Rambus Stanford Graduate Fellowship and the Intel Foundation PhD Fellowship. This work was also supported in part by NSF Contract DMS-0354674-001 and ONR Contract N00014-02-0088.

REFERENCES

[1] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. on Info. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.
[2] S. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. on Wireless Comm.*, vol. 3, no. 4, pp. 1165–1175, July 2004.
[3] S. Venkatesan, S. Simon, and R. Valenzuela, "Capacity of a gaussian MIMO channel with nonzero mean," *Proc. IEEE Vehicular Tech. Conf.*, vol. 3, pp. 1767–1771, Oct. 2003.
[4] D. Hösli and A. Lapidoth, "The capacity of a MIMO Ricean channel is monotonic in the singular values of the mean," *Proc. 5th Int'l ITG Conf. on Source and Channel Coding*, Jan. 2004.
[5] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, UK: Cambridge University Press, 2003.
[6] T. Kailath, A. Sayed, and H. Hassibi, *Linear Estimation*. Prentice Hall, 2000.
[7] W. Jakes, *Microwave Mobile Communications*. IEEE Press, 1994.
[8] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley & Sons, Inc., 1991.
[9] M. Vu and A. Paulraj, "Capacity optimization for Rician correlated MIMO wireless channels," *Proc. 39th Asilomar Conf. Sig., Sys. and Comp.*, Nov. 2005.
[10] I. Telatar, "Capacity of multi-antenna gaussian channels," *Bell Laboratories Technical Memorandum*, http://mars.bell-labs.com/papers/proof/, Oct. 1995. [Online]. Available: http://mars.bell-labs.com/papers/proof/
[11] D. Maiwald and D. Kraus, "Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices," *IEE Proceedings Radar, Sonar and Navigation*, no. 4, pp. 162–168, Aug. 2000.
[12] P. Graczyk, G. Letac, and H. Massam, "The complex Wishart distribution and the symmetric group," *The Annals of Statistics*, vol. 31, no. 1, pp. 287–309, Feb. 2003.
[13] A. Edelman, *Eigenvalues and Condition Numbers of Random Matrices*. MIT PhD Dissertation, 1989.