On the Capacity of MIMO Wireless Channels with Dynamic CSIT

Mai Vu, *Member, IEEE,* and Arogyaswami Paulraj, *Fellow, IEEE*

***Abstract*— Transmit channel side information (CSIT) can significantly increase MIMO wireless capacity. Due to delay in acquiring this information, however, the time-selective fading wireless channel often induces incomplete, or partial, CSIT. In this paper, we first construct a dynamic CSIT model that takes into account channel temporal variation. It does so by using a potentially outdated channel measurement and the channel statistics, including the mean, covariance, and temporal correlation. The dynamic CSIT model consists of an effective channel mean and an effective channel covariance, derived as a channel estimate and its error covariance. Both parameters are functions of the temporal correlation factor, which indicates the CSIT quality. Depending on this quality, the model covers smoothly from perfect to statistical CSIT.**

We then summarize and further analyze the capacity gains and the optimal input with dynamic CSIT, asymptotically at low and high SNRs. At low SNRs, dynamic CSIT often multiplicatively increases the capacity for all multi-input systems. The optimal input is typically simple single-mode beamforming. At high SNRs, for systems with equal or fewer transmit than receive antennas, it is well-known that the capacity gain diminishes to zero because of equi-power optimal input. With more transmit than receive antennas, however, the capacity gain is additive. The optimal input then is highly dependent on the CSIT. In contrast to equi-power, it can drop modes for channels with a strong mean or strongly correlated transmit antennas. For such mode-dropping at high SNRs in special cases, simple conditions on the channel K factor or the transmit covariance condition number are subsequently quantified.

Next, using a convex optimization program, we study the MIMO capacity with dynamic CSIT non-asymptotically. Particularly, we numerically analyze effects on the capacity of the CSIT quality, the relative number of transmit and receive antennas, and the channel K factor. For example, the capacity gain based on dynamic CSIT is more sensitive to the CSIT quality at higher qualities. The program also helps to evaluate a simple, analytical capacity lower-bound based on the Jensen optimal input. The bound is tight at all SNRs for systems with equal or fewer transmit than receive antennas, and at low SNRs for others.

***Index Terms*— Capacity, MIMO wireless, partial CSIT, correlated Rician fading, temporal correlation.**

Manuscript received June 1, 2006; revised Dec 28, 2006. Parts of this paper were presented at the 37th and 39th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, Nov 2003 and 2005, and the 43rd Allerton Conference on Communication, Control, and Computing, Sept. 2005. This work was supported in part by 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.

Mai Vu is with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 (e-mail: maivu@seas.harvard.edu).

Arogyaswami Paulraj is with the Information System Lab, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 (e-mail: apaulraj@stanford.edu).

Digital Object Identifier 10.1109/JSAC.2007.070902.

I. INTRODUCTION

WITH PERFECT channel information at the receiver, channel side information at the transmitter (CSIT) can significantly increase MIMO channel capacity. Assuming frequency-flat fading, this benefit comes from the spatial channel dimension. In contrast to temporal CSIT, which provides diminishing capacity gain at medium-to-high SNR [1], spatial CSIT can enhance channel capacity in both low and high SNR regions, depending on the relative number of antennas. For example, in a 4-transmit 2-receive antenna system, perfect CSIT doubles the capacity at -5dB SNR and adds 2bps/Hz additional capacity at 15dB SNR [2]. Perfect spatial CSIT has been analyzed for fading channels in both ergodic and outage capacities, showing significant capacity gains [3], [4]. These gains promise valuable increase in the transmission rate of wireless systems.

The time-varying nature of the wireless channel, coupled with the inherent delay in CSIT acquisition, however, often induces partial CSIT. To account for the channel temporal variation, we explicitly formulate a dynamic CSIT model, by combining a potentially outdated channel measurement with the channel statistics. The formulation allows evaluating the CSIT based on the channel temporal correlation factor ρ, which is a function of the delay and the channel Doppler spread. When $\rho = 1$, the CSIT is perfect; when $\rho \to 0$, the CSIT approaches the actual channel statistics. Specifically, dynamic CSIT consists of an estimate of the channel at the transmit time and the associated error covariance, which function effectively as the channel mean and the channel covariance, respectively.

To establish the capacity with dynamic CSIT, it is necessary to find the optimal input signal. For memoryless channel with perfect channel information at the receiver, the optimal input is Gaussian distributed with zero-mean [5]. Therefore, the objective remains to find its optimal covariance. This covariance eigenvectors function as transmit beam directions, and the eigenvalues as the beam power allocation. For dynamic CSIT, involving a non-zero effective channel mean and a non-trivial effective channel covariance, the capacity optimization problem involves evaluating an expectation over the non-central Wishart distribution. Solutions exist partially only for special cases: covariance CSIT, when the channel covariance is non-trivial but the mean is zero [6], [7], and mean CSIT, when the channel mean is non-zero but the covariance is the identity matrix [8], [9]. In these cases, the optimal beam directions are known analytically (as the eigenvectors of the mean or covariance matrix), but not the power allocation. The latter usually requires numerical optimization, where efficient iterative algorithms exist for covariance CSIT [10] and mean CSIT [11].

After formulating the capacity with dynamic CSIT, we first analyze the optimal input covariance and the capacity gain from the CSIT asymptotically at low and high SNRs. Asymptotic MIMO capacity have been studied by various authors. We summarize some of the existing results in the context of capacity gain and develop new results on optimal mode-dropping at high SNRs. Specifically, at low SNRs, the optimal input typically becomes simple single-mode beamforming, and the capacity gain is multiplicative [12]. At high SNRs, the optimal solution depends on the relative numbers of transmit and receive antennas. For systems with equal or fewer transmit than receive antennas, it is well-known that the optimal input approaches equi-power and the capacity gain diminishes to zero. For others, however, both the optimal input and the capacity gain depend heavily on the CSIT. In contrast to equi-power, the optimal input may drop modes at high SNRs in systems with more transmit than receive antennas. We establish conditions for mode dropping for representative channels with a high K factor or strong transmit antenna correlation. These conditions provide intuition to when it is optimal to activate only a fraction of the available eigen-modes at all SNRs, given the CSIT. In such cases, CSIT provides an additive capacity gain at high SNRs.

Fortunately, capacity optimization with dynamic CSIT is a convex problem, hence allowing efficient numerical implementation [13]. Using the program in [14], we study the non-asymptotic capacity impacts of the channel mean, the transmit covariance, the CSIT quality, and the K factor. The program also helps evaluate a sub-optimal input covariance, which maximizes Jensen's bound on capacity, and establish conditions with which, using this covariance creates a tight lower-bound to the capacity, hence allowing a simple, analytical capacity approximation.

The paper is organized as follows. Section II discusses the wireless channel model. Section III then establishes the dynamic CSIT model. Section IV formulates the channel capacity with dynamic CSIT. Asymptotic capacity results are analyzed in the next two sections. Section V establishes the capacity gains at low and high SNRs, and Section VI characterizes the optimal input. Numerical capacity results follow in Section VII. Section VIII then provides the conclusion.

Notation used in this paper is as follows. $(.)^*$ for conjugate transpose, $E[.]$ for expectation, tr$(.)$ for trace, $||.||_F$ for the Frobenius norm, $\lambda(.)$ for eigenvalues, $(.)_+$ for a positive value inside the parenthesis or zero, and $\succcurlyeq$ for the matrix positive semi-definite relation. Matrices are denoted by boldface capital letters, and their vectorized version is denoted by the corresponding lower-case letter in bold-face, with any subscript carried through (for example, $\mathbf{h}_0 = \text{vec}(\mathbf{H}_0)$).

II. Channel Model

Consider a frequency flat, quasi-static block fading channel with N transmit and M receive antennas. The channel can be modeled as a complex Gaussian process, represented by a matrix $\mathbf{H}_s$ of size $M \times N$, with s indicating the time. Assuming that the channel is stationary, it can be specified by its time-invariant mean, covariance, and auto-covariance. Specifically, omitting the time subscript for brevity, the channel $\mathbf{H}$ can be decomposed as into fixed and variable parts as

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

where $\mathbf{H}_m$ is the complex channel mean, and $\tilde{\mathbf{H}}$ is a zero-mean complex Gaussian random matrix.

A. Channel covariance and antenna correlations

The channel covariance $\mathbf{R}_0$ captures the spatial correlation among all the transmit and receive antennas. In other words, it defines the correlation among all MN channel elements as a $MN \times MN$ matrix

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

where $\tilde{\mathbf{h}} = \text{vec}(\tilde{\mathbf{H}})$, and $(.)^*$ denotes a conjugate transpose. $\mathbf{R}_0$ is a positive semi-definite Hermitian matrix. Its diagonal elements represent the power gain of the MN scalar channels, and the off-diagonal elements are the cross-coupling between these scalar channels.

The covariance $\mathbf{R}_0$ sometimes assumes a simple Kronecker structure with separable transmit and receive antenna correlations [15]. The channel covariance can now be decomposed as

$$\mathbf{R}_0 = \mathbf{R}_t^T \otimes \mathbf{R}_r \ , \tag{3}$$

where $\otimes$ denotes the Kronecker product [16]. Both $\mathbf{R}_t$ and $\mathbf{R}_r$ are complex Hermitian positive semi-definite. The channel (1) can then be written as

$$\mathbf{H} = \mathbf{H}_m + \mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{R}_t^{1/2} \ , \tag{4}$$

where $\mathbf{H}_w$ is a $M \times N$ matrix, whose entries have the real and imaginary parts independent and identically distributed as zero-mean Gaussian with unit-variance. Here $\mathbf{R}_t^{1/2}$ is the unique square-root of $\mathbf{R}_t$, such that $\mathbf{R}_t^{1/2}\mathbf{R}_t^{1/2} = \mathbf{R}_t$; similarly for $\mathbf{R}_r^{1/2}$.

The Kronecker correlation model has been experimentally verified in indoor environments for up to 3×3 antenna configurations [17], [18], and in outdoor environments for up to 8×8 configurations [19]. Other more general covariance structures have been proposed in the literature [20], [21], where the transmit covariances corresponding to different reference receive antennas are assumed to have the same eigenvectors, but not necessarily the same eigenvalues; similarly for the receive covariances.

B. Channel mean and the Rician K factor

The channel mean $\mathbf{H}_m$ $(M \times N)$ is the fixed component of the channel, usually corresponding to a line-of-sight propagation path or a cluster of strong paths, obtained as

$$\mathbf{H}_m = E[\mathbf{H}] \ . \tag{5}$$

The elements of the mean can have different amplitudes and arbitrary phase. The strength of a channel mean can be loosely quantified by the Rician K factor. It defines the ratio of the power in the channel mean and the average power in the channel variable component as

$$K = \frac{||\mathbf{H}_m||_F^2}{\text{tr}(\mathbf{R}_0)} \ , \tag{6}$$

where $||.||_F$ is the matrix Frobenius norm, and tr(.) is the trace of a matrix. The K factor can take any real value between 0 and infinity. When $K = 0$, the channel has Rayleigh distribution, otherwise it is Rician fading. When $K \to \infty$, the channel becomes deterministic. Measurements of fixed broadband channels have shown that the K factor can have a wide range from 0 to up-to 30dB in practice, and it tends to decrease with the distance between the transmitter and the receiver [22].

C. Channel auto-covariance and the Doppler spread

The channel auto-covariance characterizes how rapidly the channel decorrelates with time. Because of the stationarity assumption, the covariance between two channels samples $\mathbf{H}_0$ and $\mathbf{H}_s$ depends only on the time difference but not the absolute time

$$\mathbf{R}_s = E\left[\tilde{\mathbf{h}}_0 \tilde{\mathbf{h}}_s^*\right] , \tag{7}$$

where $\tilde{\mathbf{h}}$ denotes the vectorized version of the zero-mean part of the corresponding channel matrix. Note that when $s = 0$, this auto-covariance coincides with the channel covariance $\mathbf{R}_0$ (2); when s becomes large, it eventually decays to zero.

For a MIMO channel, the covariance $\mathbf{R}_0$ captures the spatial correlation between all the transmit and receive antennas, while the auto-covariance at a non-zero delay $\mathbf{R}_s$ captures both channel spatial and temporal correlations. Based on the premise that the channel temporal statistics can be the same for all antenna pairs, it may be assumed that the temporal correlation is homogeneous and identical for any channel element. Then, the two correlation effects are separable, and the channel auto-covariance becomes their product as

$$\mathbf{R}_s = \rho_s \mathbf{R}_0 , \tag{8}$$

where ρ_s is the temporal correlation of a scalar channel. In other words, all the MN scalar channels between the M transmit and N receive antennas have the same temporal correlation function. This correlation is a function of the time difference s and the channel Doppler spread. For example, in Clark's model, $\rho_s = J_0(2\pi f_d s)$, where J_0 is the zero$^{\text{th}}$ order Bessel function of the first kind, and f_d is the Doppler spread [23]. Similar assumptions for MIMO temporal correlation have also been used in constructing channel models and verifying measurement data in [18], [21].

The channel statistics, $\mathbf{H}_m$, $\mathbf{R}_0$ and $\mathbf{R}_s$, can be obtained by averaging instantaneous channel measurements over tens of channel coherence times; they remain valid for a period of tens to hundreds coherence time, during which, the channel can be considered as stationary.

III. Dynamic CSIT

A. CSIT acquisition

While a receiver can directly estimate the channel from the received signal, the transmitter must obtain channel information indirectly by using the reverse-channel information, relying on the reciprocity principle, or by feedback from the receiver. Reciprocity applies to the "over-the-air" channel between the transmit and receive antennas and requires the forward and reverse links to occur at the same time, frequency, and space instance. In practice, it requires careful hardware calibration to make the transmit and receive RF chains identical and is usually applicable only in time-division-duplex (TDD) systems with small turn-around time between the reverse and forward transmissions [24]. Feedback can be used in both time and frequency division-duplex systems, provided that the feedback lag is relatively small compared to channel dynamics. In either case, there exists a delay between measuring the channel information at a receiver and using this information at the transmitter. This delay may affect the reliability of the obtained channel information, depending on the type of information.

Consider the channel statistics and instantaneous channel measurements. In practice, the channel statistics often remain unchanged for a relatively long time compared to the transmission intervals. Assuming stationarity, these statistics are not affected by the delay in channel acquisition, and hence become reliable CSIT. Channel statistics, however, are only partial information and therefore provide partial capacity gains. In contrast, instantaneous channel information provides the highest capacity gain, but is sensitive to the CSIT acquisition delay due to channel temporal variation. This delay leads to a potential mismatch between the instantaneous channel measurement and the current channel. For the purpose of this paper, the initial channel measurement is assumed to be accurate. The only cause of the mismatch then is channel varying over the delay period.

More reliable and complete CSIT provides more capacity gain. This principle suggests combining both the channel statistics and instantaneous measurements to create a CSIT model robust to channel variation, while optimally capturing the potential gain.

B. CSIT modeling

Consider CSIT at the transmit time s in the form of a channel estimate $\hat{\mathbf{H}}_s$ and its error covariance $\mathbf{R}_{e,s}$, which can be expressed as

$$\begin{aligned} \mathbf{H}_s &= \hat{\mathbf{H}}_s + \mathbf{E}_s , \\ \mathbf{R}_{e,s} &= E\big[\mathbf{e}_s \mathbf{e}_s^*\big] \end{aligned} \tag{9}$$

where $\mathbf{E}_s$ is the zero-mean estimation error and $\mathbf{e}_s = \text{vec}(\mathbf{E}_s)$. Assuming unbiased estimates, $\mathbf{E}_s$ can be modeled as a stationary Gaussian random process. The error covariance $\mathbf{R}_{e,s}$ is then dependent on s and the Doppler spread.

Now assume that the transmitter has an initial channel measurement $\mathbf{H}_0$ at time 0, together with the channel statistics $\mathbf{H}_m$, $\mathbf{R}_0$, and $\mathbf{R}_s$. Furthermore, the channel matrix coefficients are jointly Gaussian. Based on MMSE estimation theory [25], an optimal estimate of the channel at time s and the estimation error covariance can be established as

$$\begin{aligned} \hat{\mathbf{h}}_s &= E\big[\mathbf{h}_s|\mathbf{h}_0\big] &&= \mathbf{h}_m + \mathbf{R}_s^* \mathbf{R}_0^{-1}\big[\mathbf{h}_0 - \mathbf{h}_m\big] \\ \mathbf{R}_{e,s} &= \text{cov}\big[\mathbf{h}_s|\mathbf{h}_0\big] &&= \mathbf{R}_0 - \mathbf{R}_s^* \mathbf{R}_0^{-1} \mathbf{R}_s , \end{aligned} \tag{10}$$

where $\hat{\mathbf{h}}_s = \text{vec}\big(\hat{\mathbf{H}}_s\big)$ (again the lower-case letters $\mathbf{h}$ denote the vectorized version of the corresponding upper-case matrices $\mathbf{H}$). A similar model was proposed in [26] for estimating a scalar time-varying channel from a vector of out-dated estimates. CSIT formulations conditioned on noisy channel estimates were also studied in [27], [28].

The two quantities $\{\hat{\mathbf{H}}_s, \mathbf{R}_{e,s}\}$ constitute the CSIT. They effectively function as a new channel mean and covariance. Thus $\hat{\mathbf{H}}_s$ and $\mathbf{R}_{e,s}$ are also referred to as the effective mean and effective covariance, respectively, and the pair forms an effective statistics.

We next apply the simplified temporal correlation model (8). This model helps to isolate the effect of temporal channel variation on the CSIT; it has also been used to fit data in some channel measurements [18], [21]. The channel estimate and its error covariance then become

$$\begin{aligned} \hat{\mathbf{H}}_s &= \rho_s \mathbf{H}_0 + (1-\rho_s)\,\mathbf{H}_m\ , \\ \mathbf{R}_{e,s} &= \left(1-\rho_s^2\right)\mathbf{R}_0\ . \end{aligned} \tag{11}$$

The CSIT can now be simply characterized as a function of ρ_s, the initial channel measurement $\mathbf{H}_0$, and the channel mean $\mathbf{H}_m$ and covariance $\mathbf{R}_0$. The channel estimate becomes a linear combination between the initial measurement and the channel mean. The error covariance is a linear function of the channel covariance alone. With Kronecker antenna correlation (3), the transmit and receive covariances are separable. For the estimated channel, the effective antenna correlations can be decomposed as

$$\begin{aligned} \mathbf{R}_{t,s} &= \sqrt{1-\rho_s^2}\,\mathbf{R}_t\ , \\ \mathbf{R}_{r,s} &= \sqrt{1-\rho_s^2}\,\mathbf{R}_r\ , \end{aligned} \tag{12}$$

which again follow a Kronecker structure. Since the scaling factors ultimately affect the received power, this decomposition is for convenience from the analysis and signal processing perspectives. When the antennas at the receiver are uncorrelated, $\mathbf{R}_r = \mathbf{I}$, then the effective channel is also assumed to have no receive correlation ($\mathbf{R}_{r,s} = \mathbf{I}$), and the effective transmit correlation becomes

$$\mathbf{R}_{t,s} = \left(1-\rho_s^2\right)\mathbf{R}_t\ . \tag{13}$$

In the constructed CSIT models (11), (12) and (13), ρ_s acts as a channel estimate quality dependent on the time delay s. For a zero or short delay, ρ_s is close to one; the estimate depends heavily on the initial channel measurement, and the error covariance is small. As the delay increases, ρ_s decreases in magnitude to zero, reducing the impact of the initial measurement. The estimate then moves toward the channel mean $\mathbf{H}_m$, and the error covariance grows toward the channel covariance $\mathbf{R}_0$. Therefore, the estimate and its error covariance (11) constitute a form of CSIT, ranging between perfect channel knowledge (when $\rho = 1$) and the channel statistics (when $\rho = 0$). By taking into account channel time variation, this framework optimally captures the available channel information and creates a dynamic CSIT model.

C. Special CSIT cases

Here we define the terminology for several special cases of dynamic CSIT. When $\rho = 1$, it is *perfect CSIT*. When $0 \le |\rho| < 1$, it is *partial CSIT*, consisting of an effective channel mean and an effective covariance. For $\rho = 0$, the CSIT is referred to as *statistical CSIT*. If either the mean or covariance is trivial, then the CSIT collapses to a special case. When the covariance is arbitrary but the mean is zero ($\mathbf{H}_m = 0$), we have *covariance CSIT*. More specifically, transmit antenna correlation alone (without receive correlation) gives *transmit covariance CSIT*. When the mean $\mathbf{H}_m$ is arbitrary but the covariance is an identity matrix ($\mathbf{R}_0 = \mathbf{I}$), we have *mean CSIT*. Finally, when $\rho = 0$, $\mathbf{H}_m = 0$, and $\mathbf{R}_0 = \mathbf{I}$, we have *no CSIT*, equivalent to an i.i.d Rayleigh fading channel with no channel information at the transmitter.

IV. Channel Ergodic Capacity with Dynamic CSIT

Consider the ergodic capacity of a MIMO channel with a constant sum power across all transmit antennas at every time instance. Assume perfect channel state information at the receiver (CSIR) and dynamic CSIT (11) with a given estimate quality ρ. With perfect CSIR, the capacity is achieved by a zero-mean complex Gaussian input [5] with covariance dependent on the CSIT. With dynamic CSIT, this optimal input covariance and the ergodic capacity are obtained by two-stage averaging.

In the first stage, each initial channel measurement $\mathbf{H}_0$ with estimate quality ρ produces CSIT value $\{\hat{\mathbf{H}}, \mathbf{R}_e\}$ (11). The channel seen from the transmitter thus effectively has mean $\hat{\mathbf{H}}$ and covariance $\mathbf{R}_e$. Assuming that this effective statistics is valid for a reasonable duration, and using a zero-mean Gaussian input with covariance $\mathbf{Q}$, we can then calculate the average mutual information given $\mathbf{H}_0$ as $\mathcal{I}(\mathbf{H}_0) = E_{\mathbf{H}}\left[\text{logdet}(\mathbf{I} + \gamma \mathbf{H}\mathbf{Q}\mathbf{H}^*)\right]$. The signal covariance $\mathbf{Q}$ that maximizes $\mathcal{I}(\mathbf{H}_0)$ is the optimizer of the problem

$$\begin{aligned} \mathcal{I}_o(\mathbf{H}_0) \;=\; & \max_{\mathbf{Q}} \quad E_{\mathbf{H}}\left[\text{logdet}(\mathbf{I} + \gamma \mathbf{H}\mathbf{Q}\mathbf{H}^*)\right] \\ & \text{subject to} \quad \begin{array}{lcl} \text{tr}(\mathbf{Q}) & = & 1 \\ \mathbf{Q} & \succcurlyeq & \mathbf{0}\ , \end{array} \end{aligned} \tag{14}$$

where γ is the SNR. The equality constraint results from the constant sum transmit power, and the inequality from the positive semi-definite property of a covariance matrix. Note that the expectation is evaluated over the effective channel statistics with mean $\hat{\mathbf{H}}$ and covariance $\mathbf{R}_e$.

In the second stage, based on ergodicity, we can average $\mathcal{I}_o(\mathbf{H}_0)$ over the distribution of $\mathbf{H}_0$ to obtain the channel ergodic capacity. For a given CSIT quality ρ, therefore, the capacity is

$$\mathcal{C} = E_{\mathbf{H}_0}\left[\mathcal{I}_o(\mathbf{H}_0)\right]\ , \tag{15}$$

where $\mathbf{H}_0$ has the actual channel statistics, Gaussian distributed with mean $\mathbf{H}_m$ and covariance $\mathbf{R}_0$.

Establishing the capacity with dynamic CSIT thus essentially requires solving (14). This problem is to find the optimal input covariance and capacity for a channel with statistical CSIT, involving arbitrary channel mean and covariance. Note that the input covariance can be decomposed into its eigenvalues and eigenvectors as

$$\mathbf{Q} = \mathbf{U}_Q \mathbf{\Lambda}_Q \mathbf{U}_Q\ . \tag{16}$$

The columns of $\mathbf{U}_Q$ function as orthogonal eigen-beam directions (patterns), and $\mathbf{\Lambda}_Q$ represents the power allocation on these beams. The problem has analytical solution for the eigenvectors $\mathbf{U}_Q$ in special cases of mean CSIT and covariance CSIT. For mean CSIT, $\mathbf{U}_Q$ is given by the right singular-vectors of the mean matrix [8], [9]. For covariance CSIT, it is by the eigenvector of the transmit antenna correlation [6], [7]. For a non-Kronecker covariance, however, analytical result for $\mathbf{U}_Q$ is so far scarce. A general condition on the channel structure for isotropic input to achieve the capacity is given in [10]. In all cases of partial CSIT, the eigenvalue $\mathbf{\Lambda}_Q$ has no closed-form solution and usually requires numerical solution. Fortunately, the optimization problem is convex, allowing efficient numerical solutions for all CSIT cases [10], [11], [14].

A. The Jensen-optimal input covariance

Here we discuss a lower-bound on the capacity by using a sub-optimal input covariance. Consider $\mathbf{Q}_J$ that maximizes Jensen's bound on the mutual information. The Jensen bound on the average mutual information is

$$E\left[\log\det(\mathbf{I}+\mathbf{H}\mathbf{Q}\mathbf{H}^*)\right] \leq \log\det\left(\mathbf{I}+E[\mathbf{H}^*\mathbf{H}]\mathbf{Q}\right) \; .$$

Given CSIT, the transmitter can establish $\mathbf{G} = E[\mathbf{H}^*\mathbf{H}]$. The covariance $\mathbf{Q}_J$ is then obtained by standard water-filling [5] on $\mathbf{G}$. Using the Jensen covariance $\mathbf{Q}_J$ as the input covariance results in a *Jensen mutual information* value

$$\mathcal{I}_J = \mathcal{I}(\mathbf{Q}_J) \; , \tag{17}$$

satisfying $\mathcal{I}_J \leq \mathcal{I}_o$ in (14). For statistical CSIT, $\mathcal{I}_J$ can be used to lower-bound the capacity. For dynamic CSIT, averaging $\mathcal{I}_J$ over the initial channel measurement distribution (15), a lower bound to the channel ergodic capacity is obtained as

$$\mathcal{C}_J = E_{\mathbf{H}_0}[\mathcal{I}_J] \; . \tag{18}$$

The tightness of this capacity lower bound will be numerically evaluated in Section VII.

V. ASYMPTOTIC-SNR CAPACITY GAINS

This section summarizes the analytical results on the asymptotic capacity gains from dynamic CSIT at low and high SNRs. Most of these results are simple and have been derived in one form or another by various authors. Here we cast them in a common context of asymptotic capacity gains, thereby connecting these results and identifying the open problems, which we will provide a partial solution to in the next section.

In particular, the asymptotic capacity gain at low SNRs is multiplicative and is achieved typically by single-mode beamforming. The gain at high SNRs is additive and requires multi-mode transmission. These gains also depend on the CSIT.

A. Low-SNR optimal beamforming and capacity gain

1) Low-SNR optimal beamforming: The optimal signal at low SNRs is typically single-mode beamforming with the direction given by the CSIT.

Theorem 1: (Verdú [12]) *As the SNRs $\gamma \to 0$, the optimal input covariance of problem (14) converges to a rank-one matrix with a unit eigenvalue and the corresponding eigenvector given by the dominant eigenvector of*

$$\mathbf{G} = E[\mathbf{H}^*\mathbf{H}] \; ,$$

provided the dominant eigenvalue is unique. In other words, the optimal input becomes a single-mode beamforming signal along the dominant eigenvector of $\mathbf{G}$. If there are multiple dominant eigenvalues of $\mathbf{G}$, the transmit power must split equally among their eigenvector directions.

Note that for the channel model (3) without receive antenna correlation, $\mathbf{G} = E[\mathbf{H}^*\mathbf{H}] = \hat{\mathbf{H}}^*\hat{\mathbf{H}} + M\mathbf{R}_{t,s}$. With dynamic CSIT, the optimal beam direction changes with each update of $\mathbf{H}_0$ and ρ. This theorem can be proved by a simple application of the Taylor expansion to $\mathcal{I}(H_0)$ in (14).

When $\mathbf{G}$ has multiple dominant eigenvalues, considering wideband transmission, equally distributing power among them minimizes the wideband slope (the derivative of the capacity at the minimum bit energy point), hence minimizing the required bandwidth [12]. In the i.i.d. case when $\mathbf{G}$ is a scaled-identity matrix, distributing power equally is in fact capacity optimal for all SNRs [29], [30].

2) Low-SNR capacity ratio gain with statistical CSIT: With the optimal input at low SNRs, the capacity gain with statistical CSIT is multiplicative and can be quantified precisely.

Theorem 2: As the SNR $\gamma \to 0$, the ratio between the optimal mutual information in (14) and the value obtained by equi-power isotropic input approaches

$$r = \frac{N\lambda_{\max}(\mathbf{G})}{tr(\mathbf{G})} \; . \tag{19}$$

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

The derivation of this result is quite straightforward [12] (see Eqs. 52-53). The result implies that, at low SNRs, the transmitter has little power, and the CSIT allows it to focus all this power on the strongest known direction in the channel, rather than spreading it equally everywhere. More transmit antennas will increase the focusing and hence the capacity gain at low SNRs in (19). As an extreme example, when $\mathbf{G}$ is rank-one, the ratio equals the number of transmit antennas N.

For dynamic CSIT with a given quality ρ, each initial channel measurement $\mathbf{H}_0$ provides an effective channel correlation $\mathbf{G}$. The capacity gain is then obtained by averaging the ratio (19) over the distribution of $\mathbf{H}_0$.

3) Low-SNR capacity ratio gain with perfect CSIT: Perfect CSIT also multiplicatively increases the capacity at low SNRs. Moreover, the asymptotic gain can be quantified in the limit of a large number of antennas.

Theorem 3: As the SNR $\gamma \to 0$, the ratio of the ergodic capacity with perfect CSIT to that without CSIT equals

$$r = \frac{E[\lambda_{\max}(\mathbf{H}^*\mathbf{H})]}{\frac{1}{N}E[tr(\mathbf{H}^*\mathbf{H})]} \; , \tag{20}$$

where the expectations are performed over the actual channel distribution.

For an i.i.d channel, if the number of antennas increases to infinity, provided the transmit to receive antenna ratio N/M stays constant, this ratio approaches a fixed value as

$$r \overset{N\to\infty}{\longrightarrow} \left(1+\sqrt{\frac{N}{M}}\right)^2 \; . \tag{21}$$

This limit is always greater than 1.

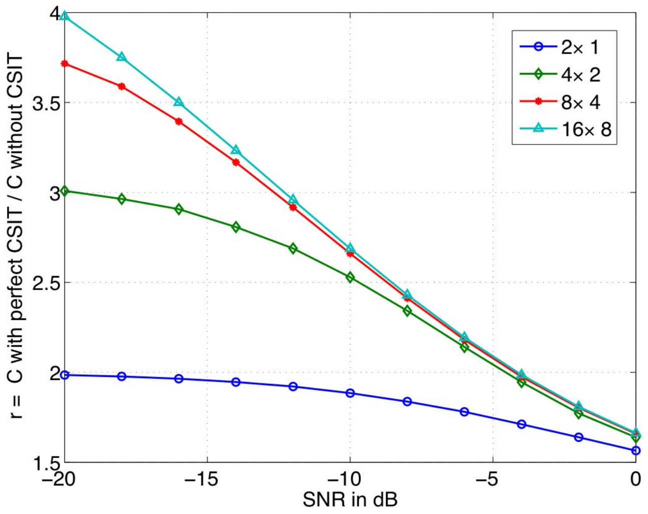

Fig. 1. Ratio of the capacity of i.i.d channels with perfect CSIT to that without CSIT. The legend denotes the numbers of transmit and receive antennas. The asymptotic capacity ratio, in the limit of large number of antennas, while keeping the number of transmit antennas twice the receive, is 5.83.

Expression (20) is straightforward from Theorem 1 and [29]. Expression (21) is implicitly derived in [12] (Eqs. 204-205) and requires a simple application of the largest eigenvalue result of large-dimension random matrices [31], [32].

Figure 1 shows examples of the capacity ratio for 4 channels with twice the number of transmit as receive antennas. The ratio increases as the SNRs decreases and as the number of antenna increases. Keeping the antenna proportion the same, as the number of antennas increases to infinity, this ratio will approach 5.83. CSIT at low SNRs thus can increase the capacity significantly.

B. High-SNR capacity gain

At high SNRs, the optimal input and the capacity gain depend on the channel rank and the relative antenna configuration. For full-rank channels, dynamic CSIT does not increase the capacity at high SNRs for systems with equal or fewer transmit than receive antennas ($N \leq M$), but does for systems with more transmit antennas ($N > M$). For rank-deficient channels with a non-full-rank $\mathbf{R}_t$, transmit covariance CSIT always helps increase the capacity. Each case is considered next.

1) Full-rank channels with equal or fewer transmit than receive antennas: When $N \leq M$, asymptotically at high SNRs, it is well-known that isotropic input is optimal for (14), independent of the CSIT (for example, the case $N = M$ is analyzed in [33]). For full-rank channel $\mathbf{H}$, the condition $N \leq M$ makes $\mathbf{H}^*\mathbf{H}$ full-rank, hence $\mathcal{I}(\mathbf{H}_0)$ at high SNRs can be decomposed as

$$\begin{aligned} \mathcal{I} &\overset{\gamma\to\infty}{\approx} E_{\mathbf{H}}[\log\det(\gamma\mathbf{H}^*\mathbf{H}\mathbf{Q})] \qquad (22)\\ &= E_{\mathbf{H}}\left[\log\det(\mathbf{H}^*\mathbf{H})\right] + \log\det(\gamma\mathbf{Q})\ . \end{aligned}$$

Maximizing this expression, subject to $\mathrm{tr}(\mathbf{Q}) = 1$, leads to $\mathbf{Q} = \mathbf{I}/N$. For these systems, the capacity gain from CSIT diminishes to 0 at high SNRs.

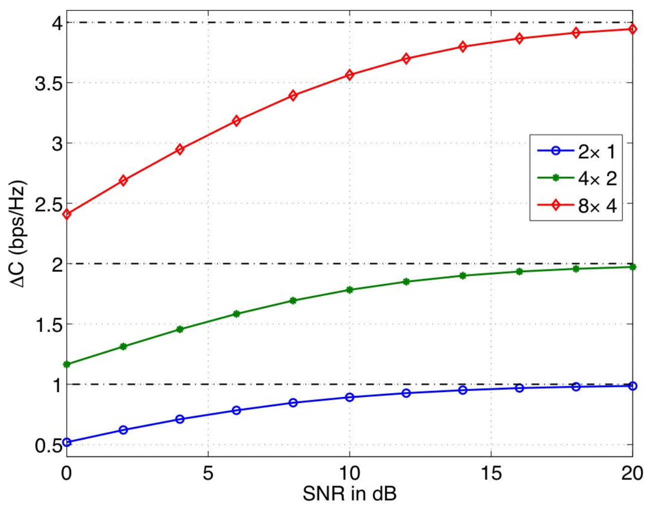

Fig. 2. Incremental capacity gain from perfect CSIT for i.i.d. channels. The legend denotes the numbers of transmit and receive antennas.

2) Full-rank channels with more transmit than receive antennas: On the contrary, when $N > M$, dynamic CSIT can provide capacity gain at high SNRs. The gain here is additive. Since $\mathbf{H}^*\mathbf{H}$ is rank-deficient in this case, the decomposition (22) does not apply. The optimal input covariance of (14) at high SNRs depends on the channel statistics, or the CSIT $\{\hat{\mathbf{H}}, \mathbf{R}_e\}$. An analytical optimal covariance for arbitrary $\hat{\mathbf{H}}$ and $\mathbf{R}_e$ is still an open problem. In the next section, we provide the solutions for some simplified cases.

The capacity gain in this case is maximum with perfect CSIT ($\rho = 1$, $\hat{\mathbf{H}} = \mathbf{H}_0$, and $\mathbf{R} = \mathbf{0}$), with which this gain can be accurately quantified.

Theorem 4: For $N > M$, at high SNRs, the incremental capacity gain from perfect CSIT ($\rho = 1$), over the mutual information obtained by equi-power isotropic input, equals

$$\Delta\mathcal{C} = M\log\left(\frac{N}{M}\right)\ . \qquad (23)$$

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

Intuitively, with $N > M$, the channel seen from the transmitter has a null-space. By knowing the channel, the transmitter can avoid sending any power into this null-space and therefore achieve a capacity gain. For example, for systems with twice the number of transmit as receive antennas, the capacity incremental gain approaches the number of receive antennas in bps/Hz and can be achieved at an SNR as low as 20dB, as shown in Figure 2.

The derivation of this result is straightforward. Although the ingredients used in proving the result have been used by different authors in different contexts (e.g. see [34] for large antenna analysis, and [35] for power offset analysis), we are not aware of a proof prior to [36] and hence provide one here for completeness.

Proof. With perfect CSIT, the solution for (14) is standard water-filling on $\mathbf{H}_0^*\mathbf{H}_0$ [5]. 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 ergodic capacity (15) 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. For full-rank $\mathbf{H}_0$, as $\gamma \to \infty$, $\mu \to \frac{1}{M}$, and the capacity approaches

$$\mathcal{C} \overset{\gamma\to\infty}{\approx} M\log\left(\frac{1}{M}\right) + M\log(\gamma) + \sum_{i=1}^{M} E\left[\log\left(\sigma_i^2\right)\right] . \quad (24)$$

Without CSIT, on the other hand, 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

$$\mathcal{C}_0 \overset{\gamma\to\infty}{\approx} M\log\left(\frac{1}{N}\right) + M\log(\gamma) + \sum_{i=1}^{M} E\left[\log\left(\sigma_i^2\right)\right] . \quad (25)$$

Subtracting (24) and (25) side-by-side yields the capacity gain in (23). □

3) Rank-deficient channels with a non-full-rank $\mathbf{R}_t$: We now consider channels with rank-deficient $\mathbf{R}_t$, zero mean and uncorrelated receive antennas. In this case, transmit covariance CSIT helps increase the capacity additively at high SNRs, regardless of the number of receive antennas. Let K_t be the rank of $\mathbf{R}_t$ ($K_t < N$), this capacity gain can be quantified in the case $K_t \le M$ as

$$\Delta\mathcal{C} = K_t \log\left(\frac{N}{K_t}\right) . \quad (26)$$

The derivation of this result is as follows. For transmit covariance CSIT, the optimal input beam directions are given by $\mathbf{R}_t$ eigenvectors. The mutual information can then be written as

$$\begin{aligned}\mathcal{I} &= \log\det\left(\mathbf{I}_M + \gamma\mathbf{H}_w\mathbf{\Lambda}_t\mathbf{\Lambda}_Q\mathbf{H}_w^*\right)\\ &= \log\det\left(\mathbf{I}_N + \gamma\mathbf{H}_w^*\mathbf{H}_w\mathbf{\Lambda}_t\mathbf{\Lambda}_Q\right) ,\end{aligned}$$

where $\mathbf{\Lambda}_t$ is the eigenvalue matrix of $\mathbf{R}_t$. With $K_t \le M$, the matrix $\mathbf{H}_w^*\mathbf{H}_w\mathbf{\Lambda}_t$ has rank K_t, hence at high SNRs, the optimal $\mathbf{\Lambda}_Q$ approaches equi-power on the K_t non-zero eigenmodes of $\mathbf{\Lambda}_t$. (Note that this equi-power input is not always optimal at high SNRs if $K_t > M$, but a capacity gain still exists.) Without the CSIT, however, the optimal input has equi-power on all N eigenmodes of $\mathbf{R}_t$. The difference in the corresponding mutual information then gives (26), similar to the proof of Theorem 4.

Figure 3 provides an example of the capacity with and without transmit covariance CSIT for rank-one correlated channels with various antenna configurations at 10dB SNR. The capacity without CSIT plus the gain (26) is also included. Note that for rank-one correlation, having more transmit antennas helps to increase the capacity with transmit covariance CSIT, but does not without the CSIT.

VI. Optimal Input Characterizations

The capacity-optimal input signal with dynamic CSIT can be analytically established in certain cases. At low SNRs, as specified in Theorem 1, it is single-mode beamforming with direction as a function of CSIT, implying mode-dropping. At high SNRs, the optimal input depends not only on the CSIT, but also on the antenna configurations. For systems with equal or fewer transmit than receive antennas, the optimal input approaches isotropic equi-power. For systems with more transmit than receive antennas, however, it may not approach equi-power at high SNRs, depending on the CSIT. Specifically, for statistical CSIT with a high-K mean or highly-conditioned transmit covariance, signifying a strong antenna correlation, mode-dropping may also occur at high SNRs. The intuition can be obtained by considering a 4×2 channel with different CSIT scenarios. With perfect CSIT, the optimal input has only 2 eigen-modes at high SNRs. Without CSIT, however, the optimal input is i.i.d isotropic, which has 4 modes. Therefore there exists partial CSIT, with which the optimal input has 3 modes at high SNRs, implying mode-dropping.

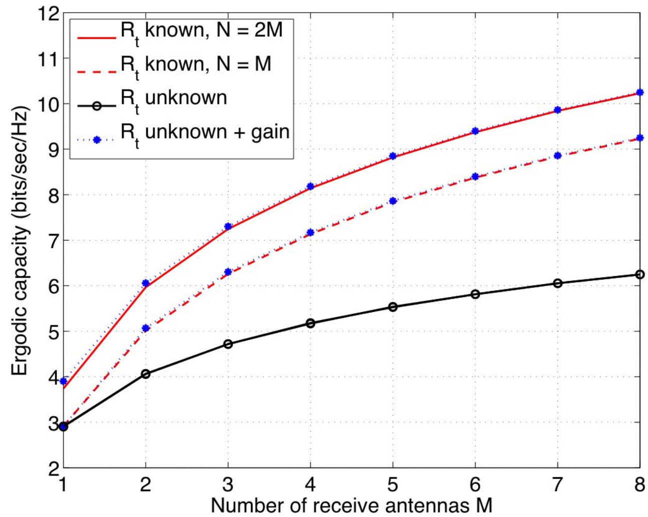

Fig. 3. Capacity of channels with a rank-one transmit correlation at SNR = 10dB, without and with transmit covariance CSIT.

In this section, we consider impacts of antenna configurations and the CSIT on the optimal input covariance. We first briefly discuss the optimal input for systems with $N \le M$. We then provide simplified analysis on the conditions for mode-dropping at high SNRs in systems with $N > M$. Two effects are considered: of the K factor and of the transmit antenna correlation. To isolate each effect, a simplified channel model is used in each case.

A. Systems with equal or fewer transmit than receive antennas

For $N \le M$ systems with dynamic CSIT, the optimal input signal is known asymptotically at both low and high SNRs. At low SNRs, it is single-mode beamforming, and at high SNRs, it approaches equi-power. Interestingly, it turns out that the optimal input covariance here can be closely approximated by the closed-form Jensen input covariance discussed in Section IV-A (see Section VII for numerical verification). This Jensen covariance becomes optimal at both low and high SNRs. At other SNRs, it produces a mutual information that is a tight lower-bound to the capacity. For the two special cases, transmit covariance CSIT and mean CSIT, the Jensen beam directions are optimal at all SNRs; only the power allocation is then approximated.

B. Systems with more transmit than receive antennas

For $N > M$ systems with dynamic CSIT, the capacity-optimal input, especially its power allocation, depends heavily on the channel effective mean and covariance matrices. If the channel is uncorrelated with zero mean, as with i.i.d Rayleigh fading, then the optimal input covariance is the identity matrix at all SNRs [29], implying equi-power allocation. However, if the mean is strong, characterized by a high K factor, or the transmit antennas are highly correlated, characterized by a large condition number of $\mathbf{R}_t$, the optimal input may drop modes even at high SNRs. A closed-form solution for the optimal input covariance, as a function of the channel mean and covariance, is still unknown. Furthermore, the Jensen covariance, which approaches equi-power at high SNRs, is no longer a good approximation.

This section provides some simple characterizations on effects of the K factor and the transmit antenna correlation on the optimal input. The analysis focuses on two simple channel models, belonging to the special mean CSIT and covariance CSIT cases. Of these, the optimal beam directions are known [6]-[9], thus only the power allocation needs to be specified. Each considered model results in the optimal allocation with only two distinct power levels. Conditions that lead to dropping the lower power level at high SNRs are analyzed.

1) Effects of the K factor: In an uncorrelated channel with $N > M$, given statistical CSIT, as K increases from 0 to infinity, the optimal number of input modes at high SNRs reduces from N to M. Thus, a sufficiently high K will result in less than N optimal modes. The mode-dropping effect, however, depends more broadly on the channel mean eigenvalues, of which K is a function. To isolate the impact of K alone, consider an uncorrelated channel, in which the channel mean has equal eigenvalues and hence is unitary. In practice, such a mean may be found in an isotropic channel as the measured channel mean, which changes slowly over time, or as a channel estimate. With this channel mean structure, the power allocation depends solely on K, but not the entire $\mathbf{H}_m$. The threshold for K, above which mode dropping occurs at high SNRs, can be obtained as follows.

Theorem 5: Consider a channel with $N > M$, uncorrelated antennas, and unitary channel mean. Specifically, the mean and transmit covariance are given as

$$\mathbf{H}_m\mathbf{H}_m^* = \frac{K}{K+1}\mathbf{I}_M \;, \quad \mathbf{R}_t = \frac{1}{K+1}\mathbf{I}_N \tag{27}$$

and the receive covariance is $\mathbf{R}_r = \mathbf{I}$.

With statistical CSIT, the condition on K, with which the optimal input activates only M out of the maximum N modes at all SNRs, is given as

$$tr\left(E\left[\left(\sum_{j=1}^{M}\left(\sqrt{K}\mathbf{e}_j + \mathbf{h}_{w,j}\right)\left(\sqrt{K}\mathbf{e}_j + \mathbf{h}_{w,j}^*\right)\right)^{-1}\right]\right) \leq 1, \tag{28}$$

where $\mathbf{e}_i$ is the M-vector with the i^{th} element equal to 1 and the rest zero, and $\mathbf{h}_{w,j} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_M)$ are i.i.d.

The matrix expression under expectation in (28) has the inverted non-central complex Wishart distribution. This expectation has no closed-form solution so far, but can be evaluated numerically.

Proof: Given mean and transmit covariance (27), let $\beta = \sqrt{K/(K+1)}$, and perform the SVD of the mean as

$$\mathbf{H}_m = \beta \mathbf{U}_m \mathbf{V}_m^* \;, \tag{29}$$

then the optimal beam directions are given by $\mathbf{V}_m$. The optimal power allocation can be completely characterized by the K factor, or β, and the SNRs. Because of 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 $N-M$ modes, where $\lambda_1 \geq \lambda_2$ [37]. Thus the optimal solution $\mathbf{Q}$ for problem (14) has the form

$$\mathbf{Q} = \mathbf{V}_m \mathbf{\Lambda}_Q \mathbf{V}_m^* \;, \tag{30}$$

where $\mathbf{\Lambda}_Q$ is a diagonal matrix with M diagonal entries as λ_1 and $N-M$ as λ_2. Let $\tilde{\mathbf{H}}$ be the zero-mean part of $\mathbf{U}_m^*\mathbf{H}\mathbf{V}_m$, then its N columns are i.i.d. with the distribution $\tilde{\mathbf{h}}_j \sim \mathcal{N}\left(\mathbf{0}, (1-\beta^2)\mathbf{I}_M\right)$. The first M columns of $\mathbf{U}_m^*\mathbf{H}\mathbf{V}_m$ can then be expressed as $\mathbf{g}_i = \beta\mathbf{e}_i + \tilde{\mathbf{h}}_i$, $1 \leq i \leq M$. Problem (14) can now be written as (31). Of interested is the condition on K (or β) that results in the optimal $\lambda_1^\star = 1/M$ and $\lambda_2^\star = 0$, implying mode-dropping. Based on the convexity of this problem, the sufficient and necessary condition for this optimality is

$$\text{tr}\left(E_{\tilde{\mathbf{h}}_j}\left[\left(\mathbf{I} + \frac{\gamma}{M}\sum_{j=1}^{M}(\beta\mathbf{e}_j + \tilde{\mathbf{h}}_j)(\beta\mathbf{e}_j + \tilde{\mathbf{h}}_j)^*\right)^{-1}\right]\right) \leq \frac{M}{1+\gamma\,(1-\beta^2)} \;, \tag{32}$$

where $\tilde{\mathbf{h}}_j \sim \mathcal{N}\left(\mathbf{0}, (1-\beta^2)\mathbf{I}_M\right)$. The derivation is given in Appendix A. This condition depends on M, β and γ and can be evaluated numerically. The condition, however, is independent of the number of transmit antennas N. From this condition, a threshold for K, above which mode-dropping occurs, can be established. As $\gamma \to \infty$, it becomes (28), which signifies mode-dropping at all SNRs. □

The threshold (32) is independent of the number of transmit antennas N. Thus if the channel mean is strong enough, the rank of this mean will dictate the number of active modes, regardless of the larger number of antennas. Figure 4 provides examples of this K factor threshold versus the SNR, derived from (32), for systems with 2 receive- and more than 2 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 drops the rest at all SNRs.

2) Effects of the transmit antenna correlation: In a zero-mean channel, the condition number of the transmit covariance matrix $\mathbf{R}_t$ can influence the number of optimal input modes. When the condition number is 1, corresponding to an identity covariance matrix, all N transmit modes are active. When the condition number is infinite, implying a rank-deficient

$$\begin{array}{ll} \max_{\lambda_1,\lambda_2} & E\left[\log\det\left(\mathbf{I}_M + \lambda_1\gamma\sum_{i=1}^{M}\left(\beta\mathbf{e}_i+\tilde{\mathbf{h}}_i\right)\left(\beta\mathbf{e}_i+\tilde{\mathbf{h}}_i\right)^* + \lambda_2\gamma\sum_{i=M+1}^{N}\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^*\right)\right] \\ \text{subject to} & \begin{array}{lcl} M\lambda_1 + (N-M)\lambda_2 & = & 1 \\ \lambda_1 \geq 0,\ \lambda_2 & \geq & 0 \end{array} \end{array} \quad (31)$$

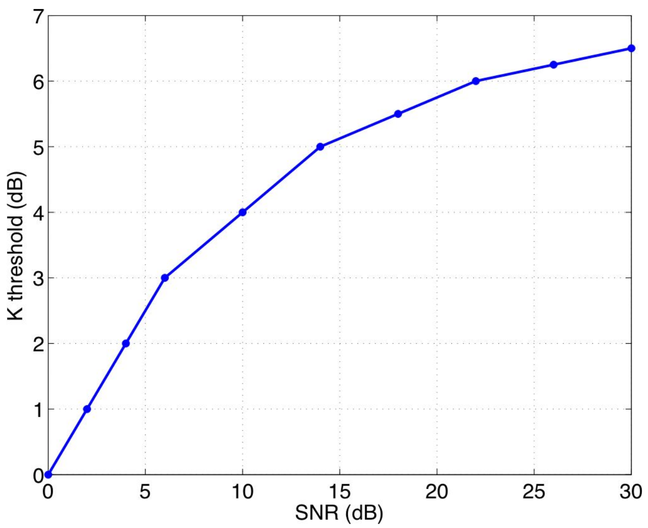

Fig. 4. K factor thresholds for systems with 2 receive and more than 2 transmit antennas, above which using 2 modes is capacity-optimal for the mean CSIT (27) at all SNRs.

covariance $\mathbf{R}_t$, the number of active modes must be less than N at all SNRs as no power should be allocated in $\mathbf{R}_t$ null-space. Thus there must be a finite threshold for $\mathbf{R}_t$ condition number, above which mode-dropping occurs at all SNRs. Similar to the K factor, however, $\mathbf{R}_t$ condition number is only a function of the eigenvalues, on which the optimal input power allocation depends. To isolate the impact of $\mathbf{R}_t$ condition number, consider a zero-mean channel with a transmit covariance matrix having only two distinct eigenvalues. The threshold on this matrix condition number for mode-dropping at all SNRs is given in the following theorem.

Theorem 6: Consider a zero-mean channel ($\mathbf{H}_m = \mathbf{0}$) *with correlated transmit antennas and uncorrelated receive antennas. Furthermore, the transmit covariance matrix has the eigen-value decomposition as*

$$\mathbf{R}_t = \mathbf{U}_t\ diag\left(\xi_1 \ldots \xi_1\ \xi_2 \ldots \xi_2\right)\ \mathbf{U}_t^*\ , \quad (33)$$

where L eigenvalues equal ξ_1 and $N-L$ equal ξ_2, provided $N > L > M$ and $\xi_1 > \xi_2 > 0$. With statistical CSIT, the threshold for $\mathbf{R}_t$ condition number, $\kappa = \xi_1/\xi_2$, above which mode-dropping occurs at all SNRs, is given as

$$\kappa \geq \frac{L}{L-M}\ . \quad (34)$$

This condition requires $N \geq M+2$.

Proof: Given zero channel mean and transmit covariance (33), the optimal input covariance $\mathbf{Q}$ has the eigenvectors given by $\mathbf{U}_t$ [6], [7]. Again because of symmetry, the optimal power allocation has only two levels: λ_1 for the L eigen-modes corresponding to the L larger eigenvalues of $\mathbf{R}_t$, and λ_2 for the rest $N-L$ modes. The optimal $\mathbf{Q}$ therefore has the eigenvalue decomposition as

$$\mathbf{Q} = \mathbf{U}_t\mathbf{\Lambda}_Q\mathbf{U}_t^*\ , \quad (35)$$

where $\mathbf{\Lambda}_Q$ is a diagonal matrix with L diagonal entries as λ_1 and $N-L$ as λ_2.

From (4), the channel can be written as $\mathbf{H} = \mathbf{H}_w\mathbf{R}_t$. Let $\tilde{\mathbf{H}} = \mathbf{H}_w\mathbf{U}_t$, then the columns of $\tilde{\mathbf{H}}$ are i.i.d with distribution $\tilde{\mathbf{h}}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_M)$ for $1 \leq i \leq N$. The mutual information optimization problem (14) is now equivalent to (36). Of interest is the condition that results in $\lambda_2 = 0$ and $\lambda_1 = 1/L$.

Based on the problem convexity, the sufficient and necessary condition for the optimal $\lambda_2 = 0$ is

$$\text{tr}\left(E_{\tilde{\mathbf{h}}_i}\left[\left(\mathbf{I}_M + \frac{\gamma\xi_1}{L}\sum_{j=1}^{L}\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^*\right)^{-1}(\gamma\xi_2+1)\right]\right) \leq M\ , \quad (37)$$

where $\tilde{\mathbf{h}}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_M)$. The derivation is given in Appendix B. At high SNRs ($\gamma \to \infty$), this condition becomes

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

Noting that the matrix under expectation is an inverted complex central Wishart matrix with rank M and L degrees of freedom, which has the first moment as $\mathbf{I}_M/(L-M)$ [38], [39], the above condition results in (34). This result requires $L > M$, and since $N > L$, this relation implies $N \geq M+2$. Thus, mode dropping at all SNRs occurs only if $N \geq M+2$. Consequently, as the SNR increases to infinity, the optimal power allocation for this transmit covariance CSIT always activate at least $M+1$ modes. □

From (34), noting that $L/(L-M) \leq N-1$, a looser bound on $\mathbf{R}_t$ condition number for dropping the weaker eigen-modes at all SNRs can be obtained as

$$\kappa = \frac{\xi_1}{\xi_2} \geq N-1\ . \quad (38)$$

This condition can be used as the first check for mode-dropping.

Figure 5 shows an example of the optimal power allocation for a 4×2 zero-mean channel with transmit covariance eigenvalues as $[1.25\ \ 1.25\ \ 1.25\ \ 0.25]$, using the optimization program in [14]. This covariance matrix has the condition number $\kappa = 5 > 3$, satisfying (34). The optimal power allocation therefore only activates 3 modes, dropping 1 mode, at all SNRs. Note also that the Jensen power as water-filling

$$\begin{aligned} \max_{\lambda_1,\lambda_2} \quad & E\left[\log\det\left(\mathbf{I}_M + \gamma\lambda_1\xi_1\sum_{j=1}^{L}\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^* + \gamma\lambda_2\xi_2\sum_{j=L+1}^{N}\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^*\right)\right] \\ \text{subject to} \quad & L\lambda_1 + (N-L)\lambda_2 = 1 \\ & \lambda_1 \geq 0\,, \quad \lambda_2 \geq 0\,. \end{aligned} \tag{36}$$

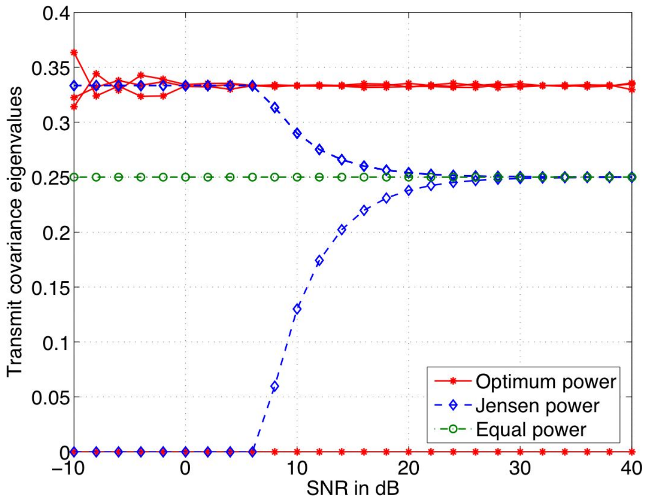

Fig. 5. Input power allocations for a 4×2 zero-mean channel with transmit covariance eigenvalues $[1.25 \quad 1.25 \quad 1.25 \quad 0.25]$. Each allocation scheme contains 4 power levels, corresponding to the 4 eigen-modes of the transmit covariance. The optimal allocation has 3 equal modes, as does the Jensen power at low SNRs. The fourth mode of the optimal scheme always has zero power.

on $E[\mathbf{H}^*\mathbf{H}]$ starts deviating from optimal at an SNR as low as 8dB.

3) Remarks: The two conditions (28) and (34), although specific to each respective channel and CSIT model, provide intuition on effects of the channel mean and the transmit antenna correlation on the optimal input power allocation. They can provide an initial check for mode-dropping with any CSIT by, for example, approximating the K factor using the minimum nonzero eigenvalue of the channel mean, or approximating R_t condition number by the ratio of its first and second largest eigenvalues. Furthermore, the condition for channels with both a non-zero mean and a transmit antenna correlation are likely to be more relaxed, such that mode dropping occurs at all SNRs for even a lower K factor and a lower transmit covariance condition number. Subsequently, channels with high K or strong transmit antenna correlation tends to result in mode dropping with statistical CSIT at all SNRs.

VII. Numerical Capacity Analysis

Establishing the capacity with dynamic CSIT for non-asymptotic SNRs usually requires numerical computation. For the special CSIT cases (mean CSIT or covariance CSIT), the optimal beam directions are known analytically, and transmit power optimization can be efficiently performed using an iterative algorithm involving the MMSE of the data streams transmitted on separate eigen-beams [10], [9]. For general dynamic CSIT, however, the optimal beam directions are still unknown, requiring an optimization for the whole covariance matrix $\mathbf{Q}$ [14].

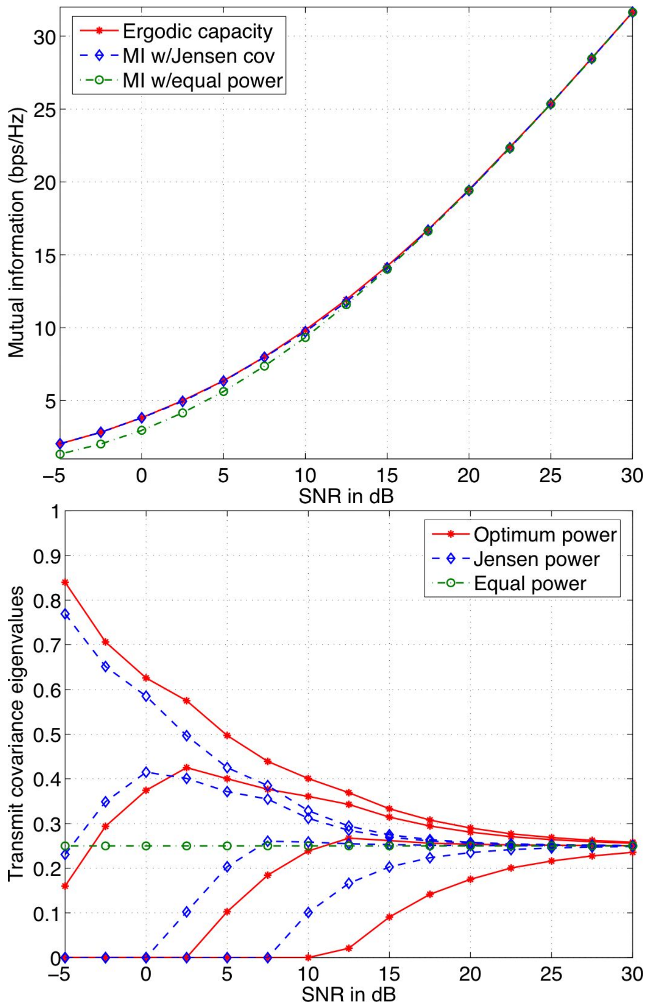

Fig. 6. Capacity and mutual information of a 4×4 system (above) and the corresponding power allocations (below). The channel mean and transmit covariance parameters are specified in Appendix C.

In this section, we use the program developed in [14] to study the non-asymptotic MIMO capacity. We numerically evaluate the Jensen input-covariance and the tightness of the capacity lower bound (18). Then using this bound and the optimization program, MIMO capacity is analyzed in terms of various parameters: relative transmit-receive antenna configuration, CSIT quality ρ, and the channel K factor.

A. Tightness of the capacity lower bound based on the Jensen input covariance

This section discusses the tightness of the Jensen mutual information (17) compared to the capacity, using statistical

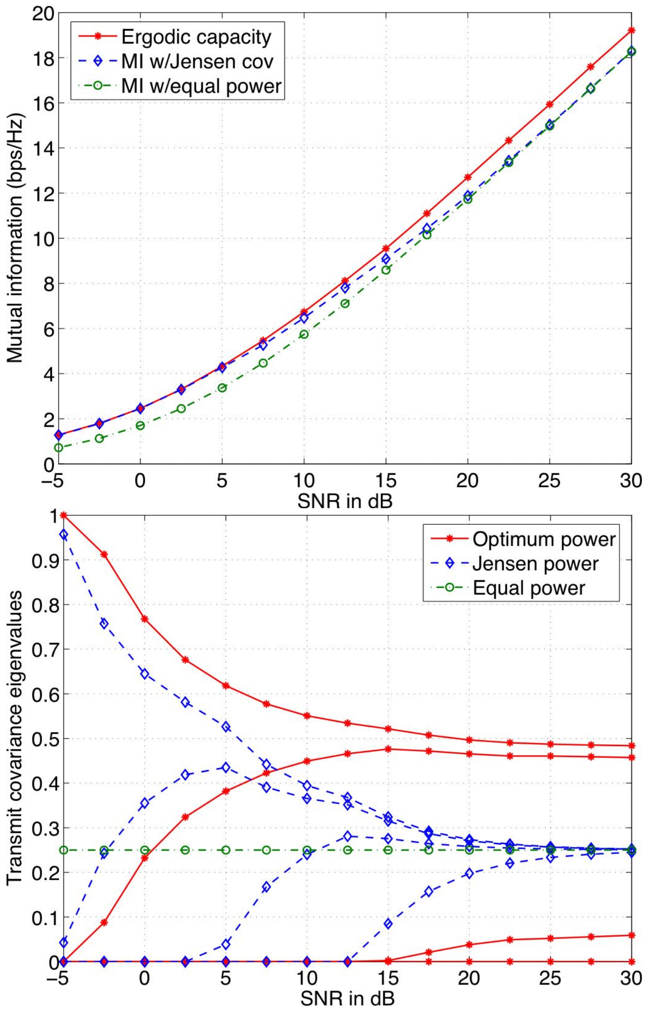

Fig. 7. Capacity and mutual information of a 4×2 system (above) and the corresponding power allocations (below). The channel mean and transmit covariance parameters are specified in Appendix C.

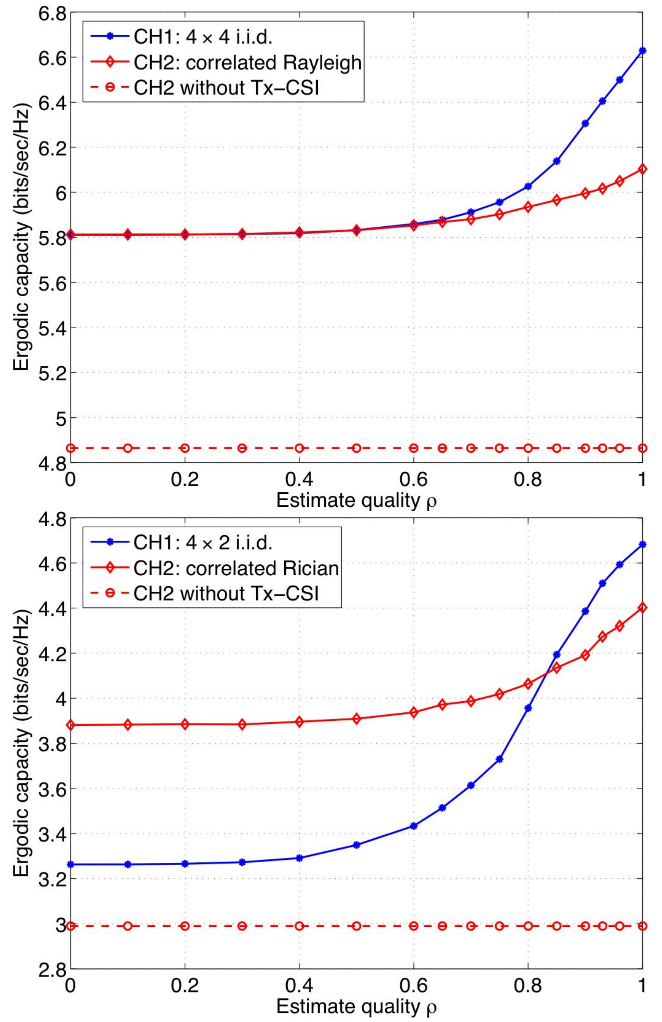

Fig. 8. Ergodic capacity versus the CSIT quality ρ at SNR = 4dB for 4×4 channels (above) and 4×2 channels (below).

CSIT as a representative. Based on simulating a wide range of channel parameters (number of antennas, mean and covariance matrices), we observed that this tightness depends on the relative transmit-receive antenna configuration and the SNR. For systems with equal or fewer transmit than receive antennas ($N \leq M$), simulation results show that the Jensen mutual information is a tight lower-bound to the channel capacity at all SNRs. Any minor difference between $\mathcal{I}_J$ and the capacity occurs only at mid-range SNRs, due to small difference in power allocation. Otherwise, the Jensen covariance approaches optimal at low and high SNRs. A similar observation is reported for channels with rank-one transmit covariance and uncorrelated receive antennas on both ergodic and outage capacities in [40]. Figure 6 shows a typical example with a 4×4 channel with the mean and transmit antenna correlation given in Appendix C. The plot includes the channel capacity and the Jensen mutual information (above), and the eigenvalues of $\mathbf{Q}^\star$ and $\mathbf{Q}_J$ (below). The mutual information with equal power allocation is also included for comparison.

For systems with more transmit than receive antenna ($N > M$), the Jensen mutual information is a tight lower-bound to the channel capacity at low SNRs. At high SNRs, however, it exhibits a gap to the capacity. This gap depends on the channel mean and the transmit antenna correlation. A higher K or more correlated channel (measured by, for example, a higher condition number of the correlation matrix) results in a bigger gap. The main reason for the gap at high SNRs is the difference in the power allocation. In contrast to the equi-power of the Jensen covariance, the capacity-optimal input can converge to non-equi-power at high SNRs. The optimal convergence values are still unknown analytically (Theorems 5 and 6 provide some partial results). Figure 7 provides an example of the mutual information and input power allocations for a 4×2 channel with the mean and correlation matrices given in Appendix C.

These comparisons also reveal that the value of CSIT depends on the antenna configuration and the SNRs. For $N \leq M$, CSIT helps increase the capacity only at low SNRs. For $N > M$, however, CSIT helps increase the capacity at all SNRs.

B. Capacity versus dynamic CSIT quality

Since the capacity lower bound $\mathcal{I}_J$ is tight at low SNRs, it is used to plot the capacity versus CSIT quality ρ in Figure 8. The capacity increases with higher ρ. The increment,

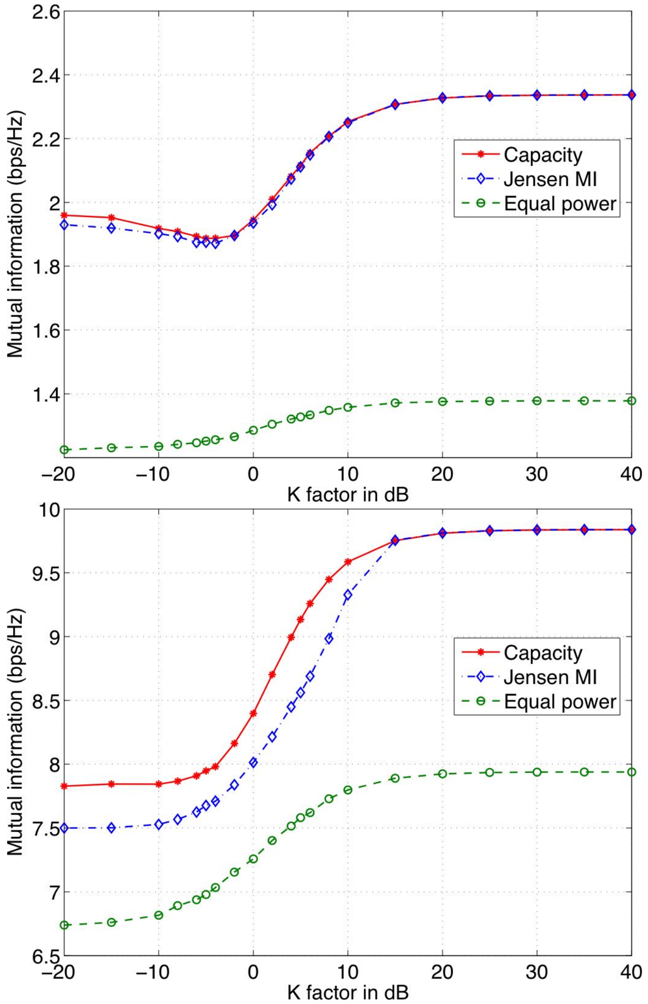

Fig. 9. Ergodic capacity and mutual information versus the K factor with SNR = -2dB (above) and SNR = 12dB (below).

however, is sensitive to ρ only when ρ is larger than about 0.6, corresponding to a relatively good channel estimate. This observation implies that in dynamic CSIT, the initial channel measurement adds value only when its correlation with the current channel is relatively strong; otherwise statistical CSIT provides most information.

Figure 8 examines two antenna configurations: 4×4 (above) and 4×2 (below). In each configuration, an i.i.d channel and a Rician correlated channel (with mean and covariance matrices given in Appendix C) are studied. Results show that the range of capacity gain for i.i.d channels is larger than for the correlated ones. Note that as the SNR increases, the capacity gain for the 4×4 channels decreases to 0, but for the 4×2 channels, it increases to up to 2 bps/Hz. For reference, the capacity of the non i.i.d channel without any CSIT is also included. In non i.i.d channels, knowing the channel statistics alone ($\rho = 0$) can enhance the capacity. Furthermore, at low SNRs, non i.i.d channels can have higher capacity than i.i.d ones, as seen in the second sub-figure.

C. Effects of the K factor

The channel Rician K factor affects the ergodic capacity differently depending on the SNR. Figure 9 shows the capacity versus K at two different SNRs for the 4×2 Rician correlation channels. Notice that at a low SNR (-2dB), the capacity is a non-monotonous function of the K factor, and a minimum exists. This minimum is partly caused by the transmit antenna correlation impact: at low K, the correlation effect becomes more dominant, and because of the low SNR, stronger correlation helps increase the capacity. At a higher SNR (12 dB), the correlation impact diminishes for full-rank correlation. Provided that the channel mean is also full-rank, the capacity then monotonically increases with the K factor. The increment, however, diminishes with higher K.

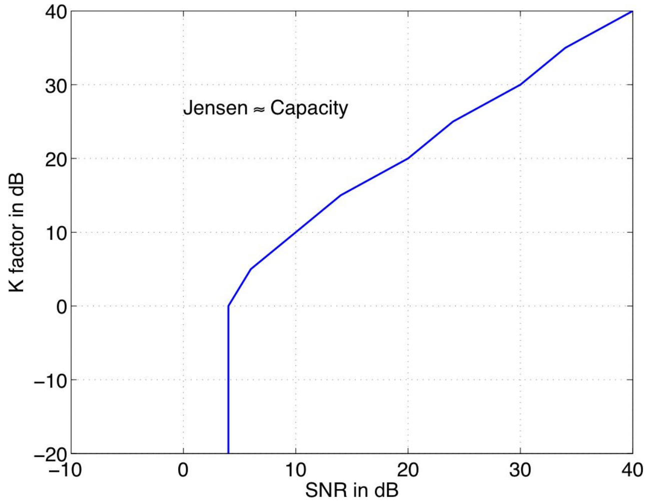

Fig. 10. K factor threshold for 4×2 channels for tight lower-bound to the ergodic capacity using the Jensen mutual information (difference < 0.03 bps/Hz).

For systems with more transmit than receive antennas, a higher K factor also causes the SNR point, at which the Jensen mutual information starts diverging from the channel capacity, to increase. This effect implies that with higher K factor, the bound is tight for larger range of SNRs. Figure 10 presents this K factor threshold versus the SNR for the 4×2 channels (keeping the same transmit antenna correlation but only varying its power with K). When K is above this threshold, the Jensen mutual information is a tight lower bound to the capacity. The difference then is less than 0.03 bps/Hz, which is within the numerical precision for optimizing the capacity.

VIII. Conclusion

In this paper, we propose a dynamic CSIT model and study the corresponding channel capacity. Dynamic CSIT is a model for transmit channel side information that takes into account the channel temporal variation. The model consists of a channel estimate and its error covariance, built on an initial, accurate channel measurement and the channel mean, covariance, and temporal correlation factor ρ. This factor functions as the CSIT quality, with 1 corresponding to perfect, and 0 to statistical information. Parameterized by ρ, the CSIT provides an effective channel mean and an effective covariance. This model can be applied to a general Rician correlated fading MIMO channel.

Asymptotic capacity analyses show that, at low SNRs, dynamic CSIT helps increase the capacity multiplicatively

$$\max_{\lambda_1} \quad g(\lambda_1) = E_{\tilde{\mathbf{h}}_j}\left[\log\det\left(\mathbf{I}_M + \lambda_1\gamma\sum_{j=1}^{M}\mathbf{g}_j\mathbf{g}_j^* + \frac{1-M\lambda_1}{N-M}\gamma\sum_{j=M+1}^{N}\tilde{\mathbf{h}}_j\tilde{\mathbf{h}}_j^*\right)\right] \tag{39}$$

$$\text{subject to} \quad \frac{1}{N} \le \lambda_1 \le \frac{1}{M},$$

$$\begin{aligned} 0 &\le E_{\tilde{\mathbf{h}}_j}\left[\mathrm{tr}\left[\left(\mathbf{I}_M + \frac{\gamma}{M}\sum_{j=1}^{M}\mathbf{g}_j\mathbf{g}_i^*\right)^{-1}\left(\gamma\sum_{j=1}^{M}\mathbf{g}_j\mathbf{g}_i^* - \frac{\gamma M}{N-M}\sum_{j=M+1}^{N}\tilde{\mathbf{h}}_j\tilde{\mathbf{h}}_j^*\right)\right]\right] \\ &\overset{(a)}{=} M^2 - ME_{\tilde{\mathbf{h}}_j}\left[\mathrm{tr}\left[\left(\mathbf{I}_M + \frac{\gamma}{M}\sum_{j=1}^{M}\mathbf{g}_j\mathbf{g}_i^*\right)^{-1}\left(\mathbf{I}_M + \frac{\gamma}{N-M}\sum_{j=M+1}^{N}\tilde{\mathbf{h}}_j\tilde{\mathbf{h}}_j^*\right)\right]\right], \end{aligned} \tag{40}$$

for all MIMO systems, and the optimal input is typically simple single-mode beamforming. At high SNRs, the capacity gain depends on the relative number of antennas. For systems with equal or fewer transmit than receive antennas, the gain diminishes to zero, since the optimal input approaches equi-power with increasing SNRs. For systems with more transmit than receive antennas, however, we show that the gain depends on the CSIT. Specifically, systems with strong transmit antenna correlation or strong mean can lead to an optimal input with mode-dropping at high SNRs, producing an additive capacity gain with the CSIT. For systems with rank-deficient transmit correlation, knowing this correlation at the transmitter can also additively increase the capacity at high SNRs.

A convex optimization program is then used to study the capacity non-asymptotically. We compare the capacity to a lower-bound based on the Jensen-optimal input. This simple lower-bound is often tight at all SNRs for systems with equal or fewer transmit than receive antennas. For systems with more transmit than receive antennas, however, the bound is tight at low SNRs but diverges at high SNRs. The divergence at high SNRs is caused by the difference in power allocation and depends on CSIT parameters – the channel mean and the transmit antenna correlation. A stronger mean or correlation causes a larger divergence. Furthermore, a higher channel K factor results in a higher SNR, at which the divergence begins.

Applied to dynamic CSIT, optimization results illustrate increasing capacity with better CSIT quality ρ. However, the capacity gain is sensitive to ρ for large ρ only, at roughly $\rho \ge 0.6$. Otherwise, the gain equals that with $\rho = 0$, corresponding to statistical CSIT. This observation suggests that, in dynamic CSIT, the initial channel measurement is useful only when its correlation with the current channel is relatively strong ($\rho \ge 0.6$); otherwise, using the channel statistics alone achieves most of the gain. Furthermore, the capacity gain depends not only on ρ but also on the channel statistics. Compared to a correlated Rician channel, the gain is higher for an i.i.d channel at high ρ, but becomes lower as ρ decreases.Thus evaluating the capacity gain requires knowing both the CSIT quality factor and the channel statistical parameters, the mean and covariance.

Appendix

A. K-factor threshold for mode-dropping at all SNRs

This section provides the derivation for (32). In problem (31), replacing λ_2 as a function of λ_1, noting that the optimal $\lambda_1 \ge \lambda_2$, the problem becomes equivalent to (39), where $\mathbf{g}_j = \left(\beta\mathbf{e}_j + \tilde{\mathbf{h}}_j\right)$. Since problem (31) is convex, this problem is convex. Thus, to have the optimal $\lambda_1^\star = 1/M$, it is sufficient and necessary that $\frac{dg(\lambda_1)}{d\lambda_1}\Big|_{\lambda_1=1/M} \ge 0$, which translates to (40), where (a) follows from adding and subtracting $M\mathbf{I}_M$ in the second parenthetic factor inside the trace expression, and $\tilde{\mathbf{h}}_j$ and $\mathbf{g}_j$ are independent. Due to this independence, and noting that $\tilde{\mathbf{h}}_j \sim \mathcal{N}\left(\mathbf{0}, (1-\beta^2)\mathbf{I}_M\right)$, the above inequality leads to (41). This expression results in (32).

B. $\mathbf{R}_t$ condition number threshold for mode-dropping at all SNRs

This section provides the derivation for (37). In problem (36), replacing λ_1 as a function of λ_2 and noting that $\lambda_1 \ge \lambda_2$, the problem becomes equivalent to (42), where $\tilde{\mathbf{h}}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_M)$. The condition for the optimal $\lambda_2 = 0$ is $\frac{\partial f}{\partial\lambda_2}\Big|_{\lambda_2=0} \le 0$, which translates to (43), where (a) follows from mutually exclusive, independent sums and that

$$E\left[\sum_{j=L+1}^{N}\tilde{\mathbf{h}}_j\tilde{\mathbf{h}}_j^*\right] = (N-L)\mathbf{I}_M.$$

Inequality (43) then leads to (37).

C. Parameters for capacity optimization

This section lists the CSIT parameters for capacity optimization in Sections VII. All simulated channels have 4 transmit and either 2 or 4 receive antennas. The normalized transmit covariance matrix (such at $\mathrm{tr}(\dot{\mathbf{R}}_t) = MN$) is given in (44). This matrix has the eigenvalues $[2.717,\ 0.997,\ 0.237,\ 0.049]$ and a condition number of 55.5, representing strong antenna correlation. The normalized mean (such at $\mathrm{tr}(\dot{\mathbf{H}}_m\dot{\mathbf{H}}_m^*) = NM$) for the 4×2 channel is given in (45). The normalized mean for the 4×4 channel is given in (46). These parameters are then scaled according to the channel K factor in the

$$\begin{aligned} M &\geq \operatorname{tr}\left(E_{\tilde{\mathbf{h}}_j}\left[\left(\mathbf{I}_M + \frac{\gamma}{M}\sum_{j=1}^{M}\mathbf{g}_j\mathbf{g}_i^*\right)^{-1}\right]E_{\tilde{\mathbf{h}}_j}\left[\mathbf{I}_M + \frac{\gamma}{N-M}\sum_{j=M+1}^{N}\tilde{\mathbf{h}}_j\tilde{\mathbf{h}}_j^*\right]\right) \\ &= \operatorname{tr}\left(E_{\tilde{\mathbf{h}}_j}\left[\left(\mathbf{I}_M + \frac{\gamma}{M}\sum_{j=1}^{M}\mathbf{g}_j\mathbf{g}_i^*\right)^{-1}\right]\left[1+\gamma\left(1-\beta^2\right)\right]\right) . \end{aligned} \quad (41)$$

$$\begin{aligned} \max_{\lambda_2} \quad & f(\lambda_2) = E_{\tilde{\mathbf{h}}_i}\left[\log\det\left(\mathbf{I}_M + \gamma\frac{1-(N-L)\lambda_2}{L}\xi_1\sum_{i=1}^{L}\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^* + \gamma\lambda_2\xi_2\sum_{j=L+1}^{N}\tilde{\mathbf{h}}_j\tilde{\mathbf{h}}_j^*\right)\right] \\ \text{subject to} \quad & 0 \leq \lambda_2 \leq \frac{1}{N} , \end{aligned} \quad (42)$$

$$\begin{aligned} 0 &\geq E_{\tilde{\mathbf{h}}_i}\left[\operatorname{tr}\left[\left(\mathbf{I}_M + \frac{\gamma\xi_1}{L}\sum_{j=1}^{L}\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^*\right)^{-1}\left(\gamma\xi_2\sum_{j=L+1}^{N}\tilde{\mathbf{h}}_j\tilde{\mathbf{h}}_j^* - \gamma\xi_1\frac{N-L}{L}\sum_{i=1}^{L}\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^*\right)\right]\right] \\ &\overset{(a)}{=} \operatorname{tr}\left(E_{\tilde{\mathbf{h}}_i}\left[\left(\mathbf{I}_M + \frac{\gamma\xi_1}{L}\sum_{j=1}^{L}\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^*\right)^{-1}\left(\gamma\xi_2\mathbf{I}_M - \frac{\gamma\xi_1}{L}\sum_{j=1}^{L}\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^*\right)(N-L)\right]\right) \\ &= \operatorname{tr}\left(E_{\tilde{\mathbf{h}}_i}\left[\left(\mathbf{I}_M + \frac{\gamma\xi_1}{L}\sum_{j=1}^{L}\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^*\right)^{-1}\left[(\gamma\xi_2+1)\mathbf{I}_M - \left(\mathbf{I}_M + \frac{\gamma\xi_1}{L}\sum_{j=1}^{L}\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^*\right)\right]\right]\right)(N-L) \\ &= \left[\operatorname{tr}\left(E_{\tilde{\mathbf{h}}_i}\left[\left(\mathbf{I}_M + \frac{\gamma\xi_1}{L}\sum_{j=1}^{L}\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^*\right)^{-1}(\gamma\xi_2+1)\right]\right) - M\right](N-L) , \end{aligned} \quad (43)$$

$$\dot{\mathbf{R}}_t = \begin{bmatrix} 0.8758 & -0.0993-0.0877i & -0.6648-0.0087i & 0.5256-0.4355i \\ -0.0993+0.0877i & 0.9318 & 0.0926+0.3776i & -0.5061-0.3478i \\ -0.6648+0.0087i & 0.0926-0.3776i & 1.0544 & -0.6219+0.5966i \\ 0.5256+0.4355i & -0.5061+0.3478i & -0.6219-0.5966i & 1.1379 \end{bmatrix} . \quad (44)$$

simulations. The simulated channels have $K = 0.1$, except for the studies of the capacity versus the K factor.

REFERENCES

[1] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[2] M. Vu and A. Paulraj, "MIMO wireless linear precoding," accepted to *IEEE Signal Processing Mag.*.

[3] G. Raleigh and J. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar 1998.

[4] A. Grant, "Rayleigh fading multi-antenna channels," *EURASIP Journal on Applied Signal Processing*, pp. 316–329, Mar 2002.

[5] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[6] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.

[7] S. 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.

[8] 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.

[9] 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.

[10] A. Tulino, A. Lozano, and S. Verdú, "Capacity-achieving input covariance for single-user multi-antenna channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 3, pp. 662–671, Mar. 2006.

[11] D. Hösli and A. Lapidoth, "How good is an isotropic gaussian input on a MIMO Ricean channel?" *Proc. Int'l Symp. on Info. Theory (ISIT)*, July 2004.

[12] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2003. [Online]. Available: http://www.stanford.edu/~boyd/cvxbook.html

[14] M. Vu and A. Paulraj, "Capacity optimization for Rician correlated MIMO wireless channels," *Proc. 39th Asilomar Conf. Sig., Sys. and Comp.*, pp. 133–138, Nov. 2005.

[15] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[16] A. Graham, *Kronecker Products and Matrix Calculus with Application*. Ellis Horwood Ltd., 1981.

$$\dot{\mathbf{H}}_m = \sqrt{10} \begin{bmatrix} 0.0749 - 0.1438i & 0.0208 + 0.3040i & -0.3356 + 0.0489i & 0.2573 - 0.0792i \\ 0.0173 - 0.2796i & -0.2336 - 0.2586i & 0.3157 + 0.4079i & 0.1183 + 0.1158i \end{bmatrix}. \quad (45)$$

$$\dot{\mathbf{H}}_m = \sqrt{10} \begin{bmatrix} 0.2976 + 0.1177i & 0.1423 + 0.4518i & -0.0190 + 0.1650i & -0.0029 + 0.0634i \\ -0.1688 - 0.0012i & -0.0609 - 0.1267i & 0.2156 - 0.5733i & 0.2214 + 0.2942i \\ 0.0018 - 0.0670i & 0.1164 + 0.0251i & 0.5599 + 0.2400i & 0.0136 - 0.0666i \\ -0.1898 + 0.3095i & 0.1620 - 0.1958i & 0.1272 + 0.0531i & -0.2684 - 0.0323i \end{bmatrix}. \quad (46)$$

[17] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Second order statistics of NLOS indoor MIMO channels based on 5.2 GHz measurements," *Proc. IEEE Global Telecomm. Conf.*, vol. 1, pp. 25–29, Nov. 2001.

[18] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Select. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[19] D. Bliss, A. Chan, and N. Chang, "MIMO wireless communication channel phenomenology," *IEEE Trans. Antennas Propagat.*, vol. 52, no. 8, pp. 2073–2082, Aug. 2004.

[20] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[21] W. Weichselberger, M. Herdin, H. Özcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 90–100, Jan. 2006.

[22] D. Baum, D. Gore, R. Nabar, S. Panchanathan, K. Hari, V. Erceg, and A. Paulraj, "Measurement and characterization of broadband MIMO fixed wireless channels at 2.5ghz," *Proc. Int'l Conf. Per. Wireless Comm.*, pp. 203–206, Dec. 2000.

[23] W. Jakes, *Microwave Mobile Communications*. IEEE Press, 1994.

[24] A. Paulraj and C. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Mag.*, vol. 14, no. 6, pp. 49–83, Nov. 1997.

[25] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.

[26] D. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 844–855, June 1999.

[27] A. Narula, M. Lopez, M. Trott, and G. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.

[28] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. on Info. Theory*, vol. 48, no. 3, pp. 611–627, Mar. 2002.

[29] 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/

[30] A. Lozano, A. Tulino, and S. Verdú, *From Array Processing to MIMO Communications*. Cambridge University Press, 2006, ch. Multiantenna capacity: Myths and Realities.

[31] J. Silverstein, "The smallest eigenvalue of a large dimensional Wishart matrix," *Annals of Probability*, vol. 13, no. 4, pp. 1364–1368, Nov. 1985.

[32] S. Geman, "A limit theorem for the norm of random matrices," *Annals of Probability*, vol. 8, no. 2, pp. 252–261, Apr. 1980.

[33] C.-N. 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.

[34] A. Tulino, A. Lozano, and S. Verdú, "MIMO capacity with channel state information at the transmitter," *Proc. IEEE Int'l Symp. on Spread Spectrum Tech. & Apps.*, pp. 22–26, Sept 2004.

[35] A. Lozano, A. Tulino, and S. Verdú, "High-SNR power offset in multiantenna communication," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4134–4151, Dec 2005.

[36] M. Vu and A. Paulraj, "Some asymptotic capacity results for MIMO wireless with and without channel knowledge at the transmitter," *Proc. 37th Asilomar Conf. Sig., Sys. and Comp.*, vol. 1, pp. 258–262, Nov. 2003.

[37] D. Hösli, Y.-H. Kim, and A. Lapidoth, "Monotonicity results for coherent MIMO Rician channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4334–4339, Dec. 2005.

[38] 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.

[39] 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.

[40] M. Ivrlac, W. Utschick, and J. Nossek, "Fading correlations in wireless MIMO communication systems," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 819–828, June 2003.

Mai Vu was born in Vietnam. She received a BE degree in Computer Systems Engineering from RMIT, Australia in 1997, an MSE degree in Electrical Engineering from the University of Melbourne, Australia in 1999, an MS and a PhD degrees both in Electrical Engineering from Stanford University, CA in 2006. She is currently a Lecturer in Engineering Sciences at the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA.

Her research interests span the areas of signal processing for communications, wireless networks, information theory, and convex optimization. Her PhD focus has been on the impact of and techniques to exploit partial channel knowledge at the transmitter in multiple-input multiple-output (MIMO) wireless systems.

Ms. Vu has been a recipient of several awards including the Australian Institute of Engineers Award at RMIT, the Rambus Corporation Stanford Graduate Fellowship and the Intel Foundation PhD Fellowship at Stanford University.

Arogyaswami Paulraj received the Ph.D. degree from the Indian Institute of Technology, New Delhi in 1973. He is currently a professor with the Department of Electrical Engineering, Stanford University, where he supervises the Smart Antennas Research Group, working on applications of space-time techniques for wireless communications. His non-academic positions have included Head, Sonar Division, Naval Oceanographic Laboratory, Cochin, India; Director, Center for Artificial Intelligence and Robotics, India; Director, Center for Development of Advanced Computing, India; Chief Scientist, Bharat Electronics,India; CTO and Founder, Iospan Wireless Inc., Co-Founder and CTO of Beceem Communications Inc.. His research has spanned several disciplines, emphasizing estimation theory, sensor signal processing, parallel computer architectures/algorithms, and space-time wireless communications. His engineering experience has included development of sonar systems, massively parallel computers, and broadband wireless systems. Dr. Paulraj has won several awards for his research and engineering contributions, including the IEEE Signal Processing Society's Technical Achievement Award. He is the author of over 300 research papers and holds twenty patents. He is a Member of both the National Academy of Engineering (NAE) and the Indian National Academy of Engineering and a Fellow of the IEEE.