Multiple-Antenna Capacity in Correlated Rayleigh Fading with Channel Covariance Information

Syed Ali Jafar† and Andrea Goldsmith‡

† Electrical Engineering and Computer Science
University of California, Irvine
Irvine, CA 92697-2625, USA
Email: syed@uci.edu

‡Department of Electrical Engineering
Stanford University
Stanford, CA, 94305-9515, USA
Email: andrea@wsl.stanford.edu

Abstract

We analyze a mobile MIMO wireless link with M transmit and N receive antennas operating in a spatially correlated Rayleigh flat fading environment. Only the correlations between the channel coefficients are assumed to be known at the transmitter and the receiver. The channel coefficients are correlated in space and uncorrelated in time from one coherence interval to another. These coefficients remain constant for a coherence interval of T symbol periods after which they change to another independent realization according to the spatial correlation model. For this system we characterize the structure of the input signal that achieves capacity. The capacity achieving transmit signal is expressed as the product of an isotropically distributed unitary matrix, an independent non-negative diagonal matrix and a unitary matrix whose columns are the eigenvectors of the transmit fade covariance matrix. For the case where the number of transmit antennas M is larger than the channel coherence interval T, we show that the channel capacity is independent of the smallest $M - T$ eigenvalues of the transmit fade covariance matrix. In contrast to the previously reported results for the spatially white fading model where adding more transmit antennas beyond the coherence interval length ($M > T$) does not increase capacity, we find that additional transmit antennas always increase capacity as long as their channel fading coefficients are spatially correlated with the other antennas. We show that for fast hopping or fast fading systems ($T = 1$) with only channel covariance information available to the transmitter and receiver, transmit fade correlations are beneficial. Mathematically, we prove this by showing that capacity is a Schur-convex function of the vector of eigenvalues of the transmit fade correlation matrix. We also show that the maximum possible capacity gain due to transmitter fade correlations is $10 \log M$ dB.

Keywords

Multielement antenna arrays, wireless communications, antenna correlation, channel capacity, channel state information (CSI).

I. Introduction

The Shannon capacity of Multiple Input Multiple Output (MIMO) systems has attracted much research activity recently. Based on seminal works by Foschini and Gans [1] and Telatar [2], which show that the link capacity can grow linearly with the number of antennas, multiple element transmitter and receiver antenna arrays are widely perceived as the key technology for the next generation of wireless communications. However, there is a caveat in that these results assume perfect channel state information at the receiver (CSIR), and sometimes even at the transmitter (CSIT). The channel coefficients can fluctuate due to user mobility, frequency hopping etc. In practice, these fluctuations can be too rapid to allow reliable channel estimation. Multiple transmit and receive antennas make channel estimation especially difficult due to the large number of channel coefficients that need to be estimated. Since outdoor wireless systems strive to accommodate higher user mobility and indoor wireless communication systems such as HomeRF and BlueTooth rely on frequency hopping, these issues necessitate further research into MIMO link capacacity in the absence of CSIT and CSIR. As a first step in this direction, it was shown by Marzetta and Hochwald in [4] that increasing the number of transmit antennas does not increase channel capacity in a fast fading scenario. Specifically, they showed that with uncorrelated flat Rayleigh fading, increasing the number of transmit antennas (M) beyond the channel coherence interval length (T) does not increase capacity. Marzetta and Hochwald's work [4] underscores a fundamental property of multiple antenna systems. The capacity benefits of using multiple antennas depend dramatically on how well the channel variations can be tracked at the transmitter and the receiver. For users moving at high speeds or fast frequency hopping systems the coherence interval is quite small and the channel fluctuates rapidly. Based on the results of [4], it seems that there is no capacity advantage from using multiple antennas in such scenarios.

Before resigning ourselves to the less than optimistic results of [4] we note that these results assume spatially white fading. Uncorrelated fading is a valid assumption for systems with widely spaced antennas. In this paper we ask the following question: with no CSI is it beneficial to place the transmit antennas closer together? Our motivation for introducing spatial correlations comes from the observation that even if the channel variations are too fast, it is possible to track the spatial correlations because they vary much more slowly. Since the capacity advantages of multiple antennas are known to depend strongly upon how well the channel can be tracked, it is possible that by tracking the channel correlations we will be able to improve capacity. This general case forms the focus of this paper. Specifically, we extend the results of [4] to analyze MIMO capacity under correlated fading in the absence of channel state information (CSI).

The absence of CSIT or CSIR implies that the instantaneous channel realization is unknown to the transmitter and the receiver. However, we assume that the channel fade statistics are known to both the transmitter and the receiver. The channel statistics may change over time. However, we assume local stationarity and are interested in the channel capacity for each given, fixed channel distribution. Since the channel distribution changes much more slowly than the channel itself, reliable estimation of the channel fade distribution at the receiver is a much easier task than estimating the rapidly fluctuating instantaneous channel. The estimated distribution can be made available to the transmitter through a feedback channel. Note that because knowledge of the time-varying channel statistics constitutes some information about the channel that is available to the transmitter and the receiver, this assumption has also been termed *partial* CSI [5], *imperfect* CSI or *covariance feedback*[1] [6]. However, limiting our time horizon to the duration where the channel is described by a given, fixed distribution, the instantaneous channel state becomes independent of the *fixed* channel statistics. Thus, following the definitions in [7] we observe that covariance information falls under the no CSIT or CSIR assumption.

We use the following notation throughout the paper: $\mathrm{vec}(X)$ is the column vector obtained by stacking the columns of X on top of each other; δ_{mn} for integer m, n is defined to be one when $m = n$ and zero otherwise; $A_{i\cdot}$ and $A_{\cdot j}$ are the i^{th} row and the j^{th} column of the matrix A respectively; $A_{M\times N}$ is the $M \times N$ principal submatrix of A; a_{ij} is the $(i,j)^{th}$ element of the matrix A; $\mathrm{tr}(A)$ is the trace of the matrix A, and $\otimes$ denotes Kronecker product.

[1]The term *feedback* comes from the underlying model that the channel statistics are provided by the receiver to the transmitter through a feedback link

II. System Model and Problem Statement

We focus on a point-to-point wireless communication system using M transmit antennas and N receive antennas over a narrowband flat Rayleigh fading channel. Following the notation and system model of [4] we have the following mathematical representation:

$$x_{tn} = \sqrt{\frac{\rho}{M}} \sum_{m=1}^{M} h_{mn} s_{tm} + w_{tn}, \qquad t \in \{1, 2, \cdots, T\},\ n \in \{1, 2, \cdots, N\}. \tag{1}$$

where x_{tn} is the signal received at n^{th} receive antenna at time t, h_{mn} is the fade coefficient between transmit antenna m and receive antenna n, s_{tm} is the signal transmitted from the m^{th} transmit antenna at time t and w_{tn} is the additive noise at the n^{th} receive antenna at time t. The noise components are assumed to be i.i.d. (independent identically distributed) zero mean unit variance circularly symmetric complex Gaussian. The channel coefficients remain fixed for the coherence time T after which they change to a new set of values generated according to the spatial correlation model described in Section II-A. The channel coefficients are assumed to be independent from one coherence interval to another. The underlying assumptions behind this block-fading model and its applicability to fast fading or frequency hopping systems are discussed by Abou-Faycal, Shamai, and Trott for the single antenna case in [3].

In matrix notation, the system model admits the following representation:

$$X = \sqrt{\frac{\rho}{M}} SH + W, \tag{2}$$

where X is the $T \times N$ received signal matrix, S is the $T \times M$ transmitted signal matrix, H is the $M \times N$ channel matrix, and W is the $T \times N$ additive noise vector. The transmit power constraint at each instant can be expressed as follows:

$$\frac{1}{TM} \mathrm{E}\left[\mathrm{tr}\left(SS^{\dagger}\right)\right] = 1. \tag{3}$$

Next we proceed to describe the spatial correlation model.

A. Spatial Correlation Model

In general the spatially correlated channel can be modeled by

$$\mathrm{vec}(H) = R^{1/2} \mathrm{vec}(H_w) \tag{4}$$

where H_w is the spatially white $M \times N$ MIMO channel with i.i.d. zero mean unit variance circularly symmetric complex Gaussian components, and R is the $MN \times MN$ covariance matrix

$$R = \mathrm{E}\left[\mathrm{vec}(H)\mathrm{vec}(H)^{\dagger}\right]. \quad (5)$$

Though the model described above is capable of representing any correlation effects between the elements of H, we use a simpler model

$$H = (R^t)^{1/2} H_w (R^r)^{1/2}. \quad (6)$$

Although less general than (4), this model is a commonly used [8] [9] spatial correlation model. While the model has been found to be satisfactory based on field measurements [10], we would also like to point out that recent work by Ozcelik et. al. [11] has shown that the model may not render the multipath structure correctly, leading to pessimistic capacity estimates in some cases. R^t and R^r are called the transmit and receive fade covariance matrices respectively. It is easily seen that R, R^t and R^r are related by:

$$R = R^r \otimes R^t. \quad (7)$$

III. Conditional Probability Density

The starting point of the analysis is the conditional probability density $p(X|S)$. Conditioned on the transmitted signal, the received signal has zero mean complex jointly Gaussian components because the channel and the white noise are zero mean complex jointly Gaussian. The probability density $p(X|S)$ is completely described by the second moments, obtained as follows:

$$\begin{aligned} \mathrm{E}[x_{t_1n_1}x^{\star}_{t_2n_2}|S] &= \frac{\rho}{M}\sum_{m_1=1}^{M}\sum_{m_2=1}^{M}\mathrm{E}[h_{m_1n_1}h^{\star}_{m_2n_2}]s_{t_1m_1}s^{\star}_{t_2m_2} + \mathrm{E}[w_{t_1n_1}w^{\star}_{t_2n_2}] \\ &= \frac{\rho}{M}R^r_{n_1n_2}S_{t_1\cdot}R^tS^{\dagger}_{t_2\cdot} + \delta_{t_1t_2}\delta_{n_1n_2} \end{aligned} \quad (8)$$

which gives us the conditional probability density as:

$$p(X|S) = \frac{\exp\left\{-\mathrm{tr}\left[\left(I_{NT} + \frac{\rho}{M}R^r \otimes SR^tS^{\dagger}\right)^{-1}\mathrm{vec}(X)\mathrm{vec}(X)^{\dagger}\right]\right\}}{\pi^{TN}\det\left[I_{NT} + \frac{\rho}{M}R^r \otimes SR^tS^{\dagger}\right]} \quad (9)$$

Comparing this to the density expression in [4] for independent fades,

$$p(X|S) = \frac{\exp\left\{-\mathrm{tr}\left[\left(I_T + \frac{\rho}{M}SS^\dagger\right)^{-1} XX^\dagger\right]\right\}}{\pi^{TN}\det^N\left[I_T + \frac{\rho}{M}SS^\dagger\right]} \tag{10}$$

we notice that (9) is slightly more involved because of need for the Kronecker product form and the vec$(\cdot)$ operation.

A. Special Properties of the Conditional PDF

In this section we extend the previously determined [4] special properties of $p(X|S)$ under independent Rayleigh fading to the general case of correlated Rayleigh fading. These properties follow from (9) by inspection and will be used later to derive the structure of the capacity achieving input signals.

Property 1: The conditional probability density $p(X|S)$ depends on the transmitted signals S only through the $T \times T$ matrix $SR^tS^\dagger$.

Property 2: For any $T \times T$ unitary matrix Φ, $p(\Phi X|\Phi S) = p(X|S)$.

Property 3: For uncorrelated receive antennas, i.e., $R^r = I_N$, the pdf of (9) simplifies to

$$p(X|S) = \frac{\exp\left\{-\mathrm{tr}\left[\left(I_T + \frac{\rho}{M}SR^tS^\dagger\right)^{-1} XX^\dagger\right]\right\}}{\pi^{TN}\det^N\left[I_T + \frac{\rho}{M}SR^tS^\dagger\right]} \tag{11}$$

In this case the $T \times T$ matrix $XX^\dagger$ is a sufficient statistic.

Property 4: For uncorrelated transmit antennas, i.e., $R^t = I_M$, the pdf of (9) simplifies only slightly as

$$p(X|S) = \frac{\exp\left\{-\mathrm{tr}\left[\left(I_{NT} + \frac{\rho}{M}R^r \otimes SS^\dagger\right)^{-1} \mathrm{vec}(X)\mathrm{vec}(X)^\dagger\right]\right\}}{\pi^{TN}\det\left[I_{NT} + \frac{\rho}{M}R^r \otimes SS^\dagger\right]} \tag{12}$$

Also, in this case for any $M \times M$ unitary matrix Ψ, $p(X|S\Psi^\dagger) = p(X|S)$.

The special cases of uncorrelated transmit antennas ($R^t = I_M$) or uncorrelated receive antennas ($R^r = I_N$) are of interest because depending on the antenna separation and the density of scatterers around the antenna array the fades corresponding to either the transmitter or the receiver may be correlated or uncorrelated. It is a common scenario that the mobile unit is surrounded by a rich scattering environment while the base station is relatively unobstructed. This makes the decorrelating antenna separation much smaller at the mobile unit (centimeters) than at the base station (meters). But at the same time, the mobile unit is more size-constrained than the base station. For a particular system, whether the fades corresponding to the transmitter or the receiver are independent or correlated depends on the relative values of these parameters.

In the next section, we use these special properties of the conditional pdf $p(X|S)$ to derive some properties of the capacity-achieving input signal for the MIMO channel with correlated Rayleigh fading.

IV. Properties of Capacity-Achieving Transmitted Signals

Subject to the average power constraint (3) the link capacity is given by the following expression [4]:

$$\begin{aligned} C &= \sup_{p(S)} I(X;S) \\ &= \sup_{p(S)} \int dS p(S) \int dX p(X|S) \log \left\{ \frac{p(X|S)}{\int d\acute{S} p(\acute{S}) p(X|\acute{S})} \right\} \end{aligned} \tag{13}$$

The following sections characterize the capacity for our specific $p(X|S)$ in (9).

A. Capacity dependence on R^t and R^r

We begin this section with the following lemma.

Lemma 1: The link capacity (13) is independent of the eigenvectors of R^t and R^r.

Proof: Represent R^t by its singular value decomposition $U^t \Lambda^t U^{t\dagger}$ where U^t is the $M \times M$ unitary matrix of the eigenvectors of R^t and Λ^t is the diagonal matrix of the corresponding eigenvalues. Without loss of generality we assume that $\lambda^t_{11} \geq \lambda^t_{22} \geq \cdots \geq \lambda^t_{MM}$. Similarly, represent R^r by its singular value decomposition $U^r \Lambda^r U^{r\dagger}$ such that $\lambda^r_{11} \geq \lambda^r_{22} \geq \cdots \geq \lambda^r_{NN}$. Now (2) and (6) can be combined to give:

$$X = \frac{\rho}{M} S (R^t)^{1/2} H_w (R^r)^{1/2} + W \tag{14}$$

$$= \frac{\rho}{M} S U^t (\Lambda^t)^{1/2} U^{t\dagger} H_w U^r (\Lambda^r)^{1/2} U^{r\dagger} + W, \tag{15}$$

which, after a suitable change of variables, becomes:

$$\hat{X} = \frac{\rho}{M} \hat{S} (\Lambda^t)^{1/2} \hat{H}_w (\Lambda^r)^{1/2} + \hat{W}. \tag{16}$$

Here,

$$\hat{X} = XU^r, \tag{17}$$

$$\hat{S} = SU^t, \tag{18}$$

$$\hat{H}_w = U^{t\dagger} H_w U^r, \tag{19}$$

$$\hat{W} = WU^r. \tag{20}$$

It is easy to see that $\hat{H}_w$ and $\hat{W}$ have the same statistics as H_w and W, respectively, and $\hat{S}$ satisfies the power constraint (3) if and only if S satisfies it. Thus the capacity optimization problem with transmitter fade covariance matrix R^t and receiver fade covariance matrix R^r is equivalent to the capacity optimization problem with transmit fade covariance matrix Λ^t and receiver fade covariance matrix Λ^r. ■

While the eigenvectors of the transmitter and receiver fade covariance matrices do not affect capacity, they could still constitute relevant information required to achieve the capacity. In other words, the proof provided above does not imply that the transmitter does not need to know U^t or that the receiver does not need to know U^r. We explain this observation further as follows: Consider the system described by (14) and the system described by (16). The input, output, noise, transmit fade covariance matrix, and receive fade covariance matrix for system (14) are S, X, W, R^t and R^r respectively. For system (16), the corresponding quantities are $\hat{S}, \hat{X}, \hat{W}, \Lambda^t$, and Λ^r respectively. These systems are equivalent because starting from system (14) the transmitter can rotate the input by U^t and the receiver can rotate the output by U^r and the resulting system is (16). If the transmitter does not know U^t or the receiver does not know U^r then these transformations can not be made and therefore we can not claim that these two systems have the same capacity. Since knowledge of U^t and U^r can only increase capacity, without the knowledge of U^t and U^r the capacity of (16) will only be an upperbound on the capacity of (14). Thus, the knowledge of U^t is required at the transmitter in order to rotate S to yield $\hat{S}$. Similarly, the receiver needs to know U^r to obtain $\hat{X}$ from X. Interestingly enough, the transmitter does not need to know U^r and the receiver does not need to know U^t. The receiver may need to estimate U^t just to provide it to the transmitter through a feedback link, but other than that, we notice that the transmitter and receiver need to know only their corresponding fade covariance eigenvectors.

Lemma 2: The link capacity (13) obtained with $M > T$ antennas depends only on the T largest eigenvalues of R^t.

Proof: Suppose that a particular joint distribution of the elements of $\hat{S}\Lambda^t\hat{S}^\dagger$ achieves capacity. Let the singular value decomposition of $\hat{S}(\Lambda^t)^{1/2}$ be given by

$$\hat{S}(\Lambda^t)^{1/2} = FDG \tag{21}$$

where F is a $T \times T$ unitary matrix, D is a $T \times M$ oblong matrix containing the singular values along the main diagonal and zeros elsewhere, and G is an $M \times M$ unitary matrix. Without loss of generality we assume that $d_{11} \geq d_{22} \geq \cdots \geq d_{TT}$. Let D_T be the $T \times T$ leading principal submatrix of D. Also, let D_M be the $M \times M$ matrix obtained by padding D with trailing rows of zeros.

Note that $\hat{S}\Lambda^t\hat{S}^\dagger = FD_T^2F^\dagger$. Next, define a new $T \times M$ matrix

$$\tilde{S} \triangleq FD(\Lambda^t)^{-1/2} \tag{22}$$

Clearly,

$$\tilde{S}\Lambda^t\tilde{S}^\dagger = \hat{S}\Lambda^t\hat{S}^\dagger. \tag{23}$$

But the LHS of (23) depends only on the largest T eigenvalues of R^t. Thus if we use $\tilde{S}$ instead of $\hat{S}$ with the corresponding transformation of the input distribution we still get the same mutual information which is independent of the $M - T$ smallest eigenvalues of R^t. To complete the proof we need to show that if $\hat{S}$ satisfies the power constraint then so does $\tilde{S}$. Mathematically, we need to show that

$$\mathrm{tr}(\tilde{S}\tilde{S}^\dagger) \leq \mathrm{tr}(\hat{S}\hat{S}^\dagger). \tag{24}$$

But this can be proved proceeding along the lines of a similar proof presented in [12], albeit in a different context. The complete proof is provided in the Appendix. ■

We conclude this section by summarizing our results in the following Theorem.

Theorem 1: For any coherence interval T and any number of receiver antennas, the capacity obtained with $M > T$ transmitter antennas in spatially correlated fading depends on the $M \times M$ transmit fade correlation matrix R^t through only the T largest eigenvalues of R^t and is not a function of the eigenvectors of R^t or R^r.

Note that if the transmitter fades are uncorrelated ($R^t = I_M$), all eigenmodes of R^t are equally strong and using the T largest eigenmodes amounts to using any T of the transmit antennas, as found in [4]. This is true even with correlated receiver antennas ($R^r \neq I_N$).

B. Capacity dependence on the number of transmit antennas

One of the key results proved by Marzetta and Hochwald in [4] for spatially white fading is that regardless of the number of receiver antennas, increasing the number of transmit antennas beyond the coherence interval duration T does not increase capacity. This is in sharp contrast to the perfect CSIR case where capacity is known to increase linearly with $\min(M, N)$ even in the absence of CSIT [1] [2]. With perfect CSIR, the linear increase in channel capacity was shown in [9] to be true even with spatially correlated fading although the rate of growth is reduced relative to the spatially white fading case. The availability of CSIR seems to be crucial as it makes all the difference between a linear growth in capacity with transmit antennas (for $N > M$) and no growth at all. To explore this surprising result further, in this section we investigate the channel capacity dependence on the number of transmit antennas in spatially correlated fading.

The following theorem states our main result in this section:

Theorem 2: The capacity of a MIMO channel without CSI increases almost surely (with probability 1) with the number of antennas when the transmit antenna fades are spatially correlated.

Thus, Theorem 2 establishes that if we are able to introduce and track spatial correlations, then additional transmit antennas almostly surely increase capacity. An outline of the proof is as follows. First, we show that adding a correlated transmit antenna almost surely increases the principal eigenvalue of the transmit fade covariance matrix. Then we show that a larger principal eigenvalue implies a higher capacity. Putting these results together we obtain Theorem 2. The full proof is presented next.

Starting from a system with M transmit antennas, suppose we add another antenna at the transmitter, increasing the number of transmit antennas to $M + 1$. The new transmit fade correlation matrix $\hat{R}^t$ can be represented as the following bordered matrix

$$\hat{R}^t = \begin{bmatrix} R^t & y \\ y^\dagger & r \end{bmatrix},$$

where y is an $M \times 1$ vector representing the correlations between the added antenna and the previous M transmit antennas, and a is a positive real number representing the fading power from the added antenna. Also, let the singular value decomposition of the $(M+1) \times (M+1)$ transmit correlation matrix $\hat{R}^t$ be given by

$$\hat{R}^t = \hat{U}^t \hat{\Lambda}^t_{(M+1)\times(M+1)} \hat{U}^{t\dagger} = \hat{U}^t \begin{bmatrix} \hat{\Lambda}^t_{M\times M} & \mathbf{0}_{M\times 1} \\ \mathbf{0}_{1\times M} & \hat{\lambda}^t_{M+1,M+1} \end{bmatrix} \hat{U}^{t\dagger} \qquad (25)$$

where $\{\hat{\lambda}^t_{ii}\}$ are arranged in descending order. Then, as a consequence of Theorem 4.3.8 in [13] we have the following relationship between Λ^t and $\hat{\Lambda}^t_{M\times M}$.

$$\Lambda^t \leq \hat{\Lambda}^t_{M\times M}. \qquad (26)$$

Note that none of the inequalities are strict in general. However, the following lemma proves that the principal eigenvalue can remain unchanged only under very special conditions: i.e., when the correlation vector y is orthogonal to the principal eigenvector of the transmit fade correlation matrix R^t. Lemma 3 is used to establish our main result stated in Theorem 2.

Lemma 3: For $\hat{R}^t$ defined as in (25), if the principal eigenvalues of R^t and $\hat{R}^t$ are equal, then the principal eigenvector of R^t is orthogonal to y: i.e., $U^t_{\cdot 1} \perp y$.

Proof: Let x denote an $M \times 1$ vector and let $\hat{x}$ denote an $(M+1) \times 1$ vector. Also, we define an $(M+1) \times 1$ vector: $1_{M+1} \triangleq [0, 0, \cdots, 0, 1]$. Now we can write,

$$\begin{aligned} \hat{\lambda}^t_{11} &= \max_{\hat{x}\neq 0,} \frac{\hat{x}^\dagger \hat{R}^t \hat{x}}{\hat{x}^\dagger \hat{x}} \\ &\geq \max_{\substack{\hat{x}\neq 0, \\ \hat{x}\perp 1_{M+1}}} \frac{\hat{x}^\dagger \hat{R}^t \hat{x}}{\hat{x}^\dagger \hat{x}} & (27) \\ &= \max_{x\neq 0,} \frac{x^\dagger R^t x}{x^\dagger x} = \lambda^t_{11} & (28) \end{aligned}$$

For equality in (27), the principal eigenvector of $\hat{R}^t$ must be orthogonal to 1_{M+1}. That is, it must contain a zero in the $(M+1)^{th}$ component. From (28) the relationship between the principal eigenvectors can be written as $\hat{U}^t_{\cdot 1} = [(U^t_{\cdot 1})^\dagger \quad 0]^\dagger$. Using the definition of an eigenvector, i.e., $\hat{R}^t\hat{U}^t_{\cdot 1} = \hat{\lambda}^t_{11}\hat{U}^t_{\cdot 1}$, if the last component of $\hat{U}^t_{\cdot 1}$ is zero, then $\hat{U}^t_{\cdot 1}$ must be orthogonal to the last row of $\hat{R}^t$. This in turn implies that $U^t_{\cdot 1} \perp y$ and the proof is complete.

■

In light of Lemma 3, given that the vector y is determined by many random parameters associated with the physical propagation environment, this means that unless the additional antenna is uncorrelated with all the previous transmitter antennas ($y = 0$), the principal eigenvalue of the new transmit fade correlation matrix (with the addition of the new antenna) will almost surely be larger than the principal eigenvalue of the old transmit fade correlation matrix. The following lemma makes a connection between the eigenvalues and the capacity.

Lemma 4: For $\hat{R}^t$ defined as in (25), if the principal eigenvalue of $\hat{R}^t$ is strictly greater than the principal eigenvalue of R^t then the capacity with $M+1$ antennas is strictly greater than the capacity with M antennas.

Proof: As seen from Lemma 1 the channel capacity depends only on the eigenvalues of R^t and the dependence is only through the form $\hat{S}\Lambda^t\hat{S}^\dagger$, which can alternately be expressed as follows:

$$\hat{S}\Lambda^t\hat{S}^\dagger \quad = \quad \hat{S}(\Lambda^t)^{1/2}(\hat{\Lambda}^t_{M\times M})^{-1/2}\hat{\Lambda}^t_{M\times M}(\hat{\Lambda}^t_{M\times M})^{-1/2}(\Lambda^t)^{1/2}\hat{S}^\dagger. \tag{29}$$

So, with the additional transmit antenna, if we transmit

$$S_{T\times(M+1)} = \left[\begin{array}{cc} \hat{S}(\Lambda^t)^{1/2}(\hat{\Lambda}^t_{M\times M})^{-1/2} & \mathbf{0}_{T\times 1} \end{array}\right]_{T\times(M+1)} \hat{U}^{t\dagger}, \tag{30}$$

then we have

$$S\hat{R}^tS^\dagger = \hat{S}\Lambda^t\hat{S}^\dagger. \tag{31}$$

Thus, we can achieve the same mutual information as before. Now, without the additional antenna we achieved capacity using transmit power $\mathrm{E}\left[\mathrm{tr}(\hat{S}\hat{S}^\dagger)\right]$. However, with the additional transmit antenna we achieve the same mutual information with total transmit power $\mathrm{E}\left[\mathrm{tr}(SS^\dagger)\right] = \mathrm{E}\left[\mathrm{tr}(\hat{S}\Lambda^t(\hat{\Lambda}^t_{M\times M})^{-1}\hat{S}^\dagger)\right]$. If the principal eigenvalues of R^t and $\hat{R}^t$ are not strictly equal, then we achieve the same mutual information with *smaller* transmit power because

$$\mathrm{E}\left[\mathrm{tr}\left(\hat{S}\Lambda^t(\hat{\Lambda}^t_{M\times M})^{-1}\hat{S}^\dagger\right)\right] < \mathrm{E}\left[\mathrm{tr}(\hat{S}\hat{S}^\dagger)\right]. \tag{32}$$

The *strict* inequality (32) follows from the inequality (26) and the fact that $\hat{\lambda}^t_{11}$ is strictly larger than λ^t_{11}. Note that we stated this lemma in terms of the principal eigenvalue for the following reason: although it is not clear if the transmit signal uses all the eigenmodes, it must definitely use the principal eigenmode. Otherwise, as seen in the proof above, power can be saved by transmitting along the stronger eigenmode.

Thus with a larger principal eigenvalue we can achieve the same capacity as before with less transmit power than before. Since capacity is an increasing function of power, this implies that for the same transmit power, the larger principal eigenvalue corresponds to a larger capacity. ■

Lemmas 3 and 4 together give us the main result of this subsection, stated in Theorem 2.

Although Theorem 2 tells us that capacity increases almost surely with transmit antennas, it does not specify the size of the capacity increases. The actual size of the capacity increases will depend on the precise spatial correlation structure. For the extreme case of fast fading $T = 1$, Section V-B characterizes the actual size of the capacity increase. It is shown that the maximum achievable improvement as we go from M transmit antennas to $M + 1$ transmit antennas is a factor of $10 \log \left(\frac{M+1}{M}\right)$ db. Thus, the potential gains from using multiple spatially correlated antennas go to zero as the number of antennas increases.

C. Structure of the capacity-achieving signal

The structure of the capacity achieving input signal S is described by the following theorem.

Theorem 3: (Structure of signal that achieves capacity) The signal matrix that achieves capacity can be written as $S = \Phi V U^{t\dagger}$ where Φ is a $T \times T$ isotropically distributed unitary matrix, V is an independent $T \times M$ real, nonnegative, diagonal random matrix, and U^t is a deterministic unitary matrix with the eigenvectors of R^t as its columns.

Proof: From equation (22) we know that the optimal S can be represented as

$$\begin{aligned} S &= \tilde{S} U^{t\dagger} \\ &= F D(\Lambda^t)^{-1/2} U^{t\dagger}. \end{aligned} \tag{33}$$

We can combine $D(\Lambda^t)^{-1/2}$ into a diagonal matrix V. The unitary matrix F can be generalized to an isotropically distributed unitary matrix as in [4], by premultiplication with an isotropically distributed unitary matrix Θ to yield the new input signal $S_1 = \Theta S$ which generates the same mutual information. Note that a unitary transformation on S does not change capacity (by inspection from (9)). ■

Comparing with the results for uncorrelated fading, we notice that the structure of the capacity-achieving signal under correlated Rayleigh fading is similar, except for a rotation along the eigenvectors of the transmitter fade correlation matrix. While the receive correlations do not appear directly in the statement of Theorem 3, the distribution of the diagonal matrix V can depend on the eigenvalues of both R^t and R^r.

The structure of the capacity-achieving signal is relevant for several reasons. Capacity computation is a maximization of mutual information over the space of all valid input distributions. By identifying the structure of the capacity-achieving input distribution we limit the size of the optimization space. The only unknown in this case is the distribution of the diagonal matrix V. Since V is also independent of Φ the size of the search space is considerably reduced. The structure of capacity achieving codes also provides insights into the kind of practical codes that perform well on the channel. In this case, the diagonal matrix V can be interpreted as a random power allocation across the eigenvectors of the transmit fade covariance matrix.

V. Spatially correlated fading: good or bad?

In the perfect CSI case, it is well known that in order to achieve a higher capacity we want the channel gains associated with each transmit antenna to fade independently [9]. However, our results in the previous sections indicate that correlated fades at the transmitter may be desirable when CSI is not available. We saw that unlike the spatially white fading case where the capacity does not increase with the number of transmit antennas beyond the coherence interval duration T (as proved in [4]), for spatially correlated fading at the transmitter antennas the capacity increases with the number of transmit antennas. Thus, it seems that transmit fade correlations help capacity. To make this statement precise we need a mathematical notion that allows us to say when a correlation matrix is more correlated or less correlated than another correlation matrix. Such a notion is provided by the concept of *majorization* from order statistics.

A. Majorization, Correlations and Schur-Convexity

Recall that a real vector $\alpha = [\alpha_i] \in \mathbb{R}^n$ majorizes another real vector $\beta = [\beta_i] \in \mathbb{R}^n$ if and only if the sum of the k smallest entries of α is greater than or equal to the sum of the k smallest entries of β for $k = 1, 2, \cdots, n-1$ and the sums of the entries of α and β are equal. This is a mathematical way to capture the vague notion that the components of a vector α are "less spread out" or "more nearly equal" than are the components of a vector β. Now if the vector α consists of the eigenvalues of a covariance matrix R^t_α and the vector β consists of the eigenvalues of a covariance matrix R^t_β then majorization directly leads to a measure of correlation. If α majorizes β then the channel corresponding to the transmit covariance matrix R^t_α is *less* correlated than the channel corresponding to the transmit covariance matrix R^t_β. Majorization as a measure of correlations is also used in [9].

A function $f(x)$ of vector x is said to be Schur-convex if $f(\alpha) \leq f(\beta)$ whenever vector α majorizes vector β. For our purpose this can be interpreted as follows. If the capacity is a Schur-convex function of the eigenvalues of the channel transmit fade covariance matrix, then more spread out eigenvalues lead to higher capacity.

Next, we use these concepts of majorization and Schur-convexity, to mathematically state our previous observation that "*transmit fade correlations help capacity*".

B. Capacity is Schur-Convex

The following Corollary of Theorem 1 shows that transmit fade correlations help capacity.

Corollary 1: For fast Rayleigh fading (or fast frequency hopping), i.e. for $T = 1$, the MIMO channel capacity under correlated fading is Schur-convex in the eigenvalues of R^t.

Proof: Let Λ_A and Λ_B be the vectors containing eigenvalues of two different transmit fade correlation matrices R^t_A and R^t_B, respectively. If Λ_A majorizes Λ_B then the principal eigenvalue of R^t_A is at least as large as the principal eigenvalue of R^t_B. From the result of Theorem 1 we know that for $T = 1$, capacity depends only on the principal eigenvalue of the transmit fade correlation matrix. Thus, the larger principal eigenvalue leads to a larger capacity. ■

To estimate the benefits of transmitter fade correlations, we next present a simple example comparing the extreme cases of independent and perfectly correlated fading. This allows us to compute precisely the maximum possible capacity gains due to correlated transmitter fades.

Consider the case of fast Rayleigh fading (or fast frequency hopping), $T = 1$. Let us denote the capacity achieved with M transmit antennas and a total transmit power P as $C^i_M(P)$. The superscript $i = 0$ or 1 depending on whether the transmitter fades are perfectly uncorrelated ($R^t = I_M$) or perfectly correlated ($R^t = \text{ones}(M, M)$), respectively. First, let us assume that the channel components fade independently. As shown in [4], the capacity achieved with M transmit antennas is the same as the capacity achieved with only one transmit antenna. We denote this as:

$$C^0_M(P) = C_1(P). \tag{34}$$

Notice that $C_1(P)$ does not carry a superscript, because there is only one transmit antenna and therefore there are no correlations to be considered.

Now, let us consider the case where all the transmitter antenna fades are perfectly correlated, i.e., R^t consists of all ones. So the channel has only one non-zero eigenmode ($\lambda^t_{11} = M, \lambda^t_{22} = \cdots = \lambda^t_{MM} = 0$). By aligning the transmission along this mode, the channel becomes like a single antenna channel scaled by the factor M. Thus, we have:

$$C^1_M(P) = C_1(M * P). \tag{35}$$

Thus, with perfect correlation of the transmit fades there is a $10 \log M$ dB improvement in the capacity over the case of independent fading.

Thus with fast fading or fast frequency hopping ($T = 1$), it is beneficial to place transmit antennas close together to obtain correlated fades instead of the traditional approach that seeks to place transmit antennas separated by the decorrelating distance so that the fades corresponding to different transmit antennas are uncorrelated. In short, for fast fading channels transmitter fade correlation helps capacity, and the capacity gains from correlated transmitter fades are bounded above by $10 \log M$ dB.

Similar results have been found to apply to the case of Multiple Input Single Output (MISO) channels with perfect CSIR and no CSIT [14] [15]. Boche and Jorswieck show in [14] and [15] that if the transmitter can adapt to the channel spatial correlations then fade correlations improve capacity. Although their channel model is significantly different from our model in this paper - we assume no CSIR and multiple receive antennas while [14] [15] assume perfect CSIR and a single receive antenna - it is interesting that the findings are similar. In both cases, if we allow the transmitter to track the spatial correlations then capacity is improved.

VI. Conclusion

We analyzed a point to point mobile MIMO wireless link with M transmit and N receive antennas operating in a spatially correlated Rayleigh flat fading environment. While for our discussions in this paper we focused on mobile communications when users are travelling at high speeds, our results are equally applicable to frequency hopping systems. In either case the channel may not stay constant for long enough to allow enough time for reliable channel estimation. To address this case no channel state information was assumed at either the transmitter or the receiver. The channel coefficients were allowed to be correlated in space but constrained to be uncorrelated in time from one coherence interval to another. The coefficients were assumed to remain constant for a coherence interval of T symbol periods after which they change to another independent realization according to the spatial correlation model. For this system, we showed that the channel capacity is independent of the smallest $M - T$ eigenvalues of the transmit fade covariance matrix, as well as the eigenvectors of the transmit and receive fade covariance matrices. We observed that the eigenvectors of the transmit fade covariance matrix need to be known only to the transmitter while those of the receive fade covariance matrix need to be known only to the receiver. We characterized the structure of the input signal that achieves capacity. The capacity achieving transmit signal is expressed as the product of an isotropically distributed unitary matrix, an independent non-negative diagonal matrix and a unitary matrix whose columns are the eigenvectors of the transmit fade covariance matrix. Also, in contrast to the previously reported results for the spatially white fading model where adding more transmit antennas beyond the coherence interval length ($M > T$) does not increase capacity, we found that additional transmit antennas always increase capacity as long as the channel fading coefficients are spatially correlated. This leads to the interesting conclusion that for fast fading or fast frequency hopping channels ($T = 1$) where channel state information is hard to obtain, it is beneficial to place the transmitter antennas close to each other in order to generate highly correlated fades. Mathematically, we proved this by showing that capacity is a Schur-convex function of the vector of eigenvalues of the transmit fade correlation matrix. We also proved that the capacity gains due to transmitter fade correlations are bounded above by $10\log M$ dB. This is in sharp contrast to the well known capacity benefits of uncorrelated fades when perfect CSIR and perfect CSIT is assumed.

VII. Appendix

A. Proof of the inequality: $tr(\tilde{S}\tilde{S}^\dagger) \leq tr(\hat{S}\hat{S}^\dagger)$

We proceed as follows:

$$\text{tr}(\tilde{S}\tilde{S}^\dagger) = \text{tr}(FD(\Lambda^t)^{-1}DF^\dagger) = \text{tr}(D_M(\Lambda^t)^{-1}D_M) \quad (36)$$

$$\text{tr}(\hat{S}\hat{S}^\dagger) = \text{tr}(FDG(\Lambda^t)^{-1}G^\dagger DF^\dagger) = \text{tr}(DG(\Lambda^t)^{-1}G^\dagger D) = \text{tr}(D_M G(\Lambda^t)^{-1}G^\dagger D_M) \quad (37)$$

Define $Q \triangleq G(\Lambda^t)^{-1}G^\dagger$. So we need to prove that:

$$\sum_{i=1}^{T} \frac{d_{ii}^2}{\lambda_{ii}^t} \leq \sum_{i=1}^{T} q_{ii} d_{ii}^2 \quad (38)$$

From this point onwards, the proof follows along the lines of a similar proof in [12]. However, for the sake of completeness, we provide of the rest of the proof in this paper. We need the following three lemmas:

Lemma 5: For a Hermitian matrix Q the vector of diagonal entries $\{q_{ii}\}$ majorizes the vector of eigenvalues $\{\lambda_{ii}^Q\}$.

Lemma 6: For any two given positive real vectors $\alpha = [\alpha_i], \beta = [\beta_i] \in \mathbb{R}_+^n$ the permutation π^* that minimizes the sum $\sum_{i=1}^n \alpha_{\pi(i)}\beta_i$ is such that $\alpha_{\pi^*(i)}$ and β_i are in the opposite order: i.e., $\forall i, j \in \{1, 2, \cdots, n\}$, if $\alpha_{\pi^*(i)} \leq \alpha_{\pi^*(j)}$, then $\beta_i \geq \beta_j$.

Lemma 7: If $\alpha = [\alpha_i], \beta = [\beta_i]$ and $\gamma = [\gamma_i] \in \mathbb{R}_+^n$ are three vectors with components arranged in descending order, i.e., if $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$, $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$ and $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_n$, and if α majorizes β then the following is true:

$$\sum_{i=1}^{n} \frac{\alpha_i}{\gamma_i} \geq \sum_{i=1}^{n} \frac{\beta_i}{\gamma_i} \quad (39)$$

Lemma 5 is Theorem 4.3.26 in [13]. Lemma 6 is Theorem 2 in [12]. Lemma 7 is Theorem 3 in [12].

If we let $\hat{q}$ be the vector of diagonal entries of Q arranged in ascending order, then based on Lemma 6 we have:

$$\sum_{i=1}^{T} \hat{q}_{ii} d_{ii}^2 \leq \sum_{i=1}^{T} q_{ii} d_{ii}^2. \tag{40}$$

Now from Lemma 5 we have that the vector of diagonal entries of Q, $\{\hat{q}_{ii}\}$ majorizes the vector of eigenvalues $\{\frac{1}{\lambda_{ii}^t}\}$. Using Lemma 7 we obtain:

$$\sum_{i=1}^{T} \frac{d_{ii}^2}{\lambda_{ii}^t} \leq \sum_{i=1}^{T} \hat{q}_{ii} d_{ii}^2 \tag{41}$$

Combining (40) and (41) we obtain the desired inequality (38). Thus we note that the capacity depends only on the T largest eigenvalues of the transmit covariance matrix.

References

[1] G. J. Foschini, M. J. Gans, "On Limits of Wireless Communications in a Fading Environment when using Multiple Antennas", Wireless Personal Communications, Vol. 6, pp. 311-335, March 1998.

[2] I. E. Telatar, "Capacity of Multi-antenna Gaussian Channels", unpublished Bell Labs Technical Memorandum, no. BL011217-950615-07TM, 1995.

[3] I. Abou-Faycal, S. Shamai, M. Trott, "The capacity of discrete-time memoryless Rayleigh-fading channels", IEEE Transactions on Information Theory, Vol. 47, Issue: 4, May 2001, page(s):1290-1301.

[4] T. Marzetta, B. Hochwald, "Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading", IEEE Transactions on Information Theory, Volume:45, Issue: 1, Jan. 1999, Page(s): 139-157.

[5] A. Narula, M. Trott, "Multiple-antenna transmission with partial side information", Proceedings of IEEE ISIT 1997, Page: 153.

[6] E. Visotsky, U. Madhow, "Space-Time Transmit Precoding with Imperfect Feedback", IEEE Transactions on Information Theory, Volume: 47, Issue: 6, Sep 2001, Page(s):2632-2639.

[7] Caire G., Shamai S., "On the Capacity of Some Channels with Channel State Information", IEEE Transactions on Information Theory, Volume:45, Issue: 6, Sept. 1999, Page(s): 2007-2019.

[8] D. Gesbert, H. Bolcskei, D. Gore, A. Paulraj, "MIMO Wireless Channels: Capacity and Performance Prediction", Proceedings of IEEE GLOBECOM, vol. 2, pp. 1083-1088, San Francisco, Nov./Dec. 2000.

[9] C Chuah, D. Tse, J. Kahn, R. Valenzuela, "Capacity Scaling in MIMO Wireless Systems under Correlated Fading", IEEE Transactions on Information Theory, Vol: 48, Issue: 3, March 2002, Page(s): 637-650.

[10] D. Chizhik, J. Ling, P. Wolniansky, R. Valenzuela, N. Costa, K. Huber, "Multiple Input Multiple Output Measurements and Modeling in Manhattan", Proceedings of IEEE Vehicular Technology Conference 2002, Vancouver B.C., Canada.

[11] H. Ozcelik, M. Herdin, W. Weichselberger, J. Wallace, E. Bonek, "Deficiencies of "Kronecker" MIMO radio channel model", Electronics Letters, Vol. 39, Issue: 16, 7 Aug. 2003. Pages:1209-1210.

[12] S. Jafar, S. Vishwanath, A. Goldsmith, "Channel Capacity and Beamforming for Multiple Transmit and Receive Antennas with Covariance Feedback", ICC2001, Helsinki, Finland.

[13] R. Horn, C. R. Johnson, "Matrix Analysis", Cambridge University Press.

[14] H. Boche, E. A. Jorswieck, "Optimal power allocation for MISO systems and complete characterization of the impact of correlation on the capacity", Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing 2003, Vol. 4, Pages 373-376.

[15] H. Boche, E. A. Jorswieck, "Optimum power allocation and complete characterization of the impact of correlation on the capacity of MISO systems with different CSI at the transmitter", Proceedings of the IEEE International Symposium on Information Theory 2003, Pages: 353-353.

Syed Ali Jafar (S'99-M'04) received the B. Tech. degree from the Indian Institute of Technology (IIT), Delhi, India in 1997, the M.S. degree from the California Institute of Technology (Caltech), Pasadena USA in 1999, and the Ph.D. degree from Stanford University, Stanford USA in 2003, all in Electrical Engineering. In 2004, he joined the Department of Electrical Engineering and Computer Science at the University of California Irvine as an assistant professor. He was a senior engineer at Qualcomm Incorporated, San Diego, CA, in 2003. He was a summer intern in the Wireless Communications Group of Lucent Bell Laboratories, Holmdel, NJ, in 2001. He was an engineer in the satellite networks division of Hughes Software Systems from 1997-1998.

His research interests include spread-spectrum communications, multiple antenna systems, ad-hoc networks and multi-user information theory.

Andrea Goldsmith (S'90-M'93-SM'99) received the B.S., M.S., and Ph.D. degrees in Electrical Engineering from U.C.Berkeley in 1986, 1991, and 1994, respectively. She was an Assistant Professor in the department of Electrical Engineering at Caltech from 1994-1999. In 1999 she joined the Electrical Engineering department at Stanford University, where she is currently an Associate Professor. Her industry experience includes affiliation with Maxim Technologies from 1986-1990, where she worked on packet radio and satellite communication systems, and with AT&T Bell Laboratories from 1991-1992, where she worked on microcell modeling and channel estimation. Her research includes work in capacity of wireless channels and networks, wireless information and communication theory, multiantenna systems, joint source and channel coding, cross-layer wireless network design, communications for distributed control, and adaptive resource allocation for cellular systems and ad-hoc wireless networks.

Dr. Goldsmith is a Terman Faculty Fellow at Stanford and a recipient of the Alfred P. Sloan Fellowship, the National Academy of Engineering Gilbreth Lectureship, a National Science Foundation CAREER Development Award, the Office of Naval Research Young Investigator Award, a National Semiconductor Faculty Development Award, an Okawa Foundation Award, and the David Griep Memorial Prize from U.C. Berkeley. She was an editor for the IEEE Transactions on Communications from 1995-2002, and has been an editor for the IEEE Wireless Communications Magazine since 1995. She is also an elected member of Stanford's faculty senate and the board of governers for the IEEE Information Theory Society.