

Characterizing the Statistical Properties of Mutual Information in MIMO Channels

Özgür Oyman, *Student Member, IEEE*, Rohit U. Nabar, *Student Member, IEEE*, Helmut Bölcskei, *Senior Member, IEEE*, and Arogyaswami J. Paulraj, *Fellow, IEEE*

Abstract—We consider Gaussian multiple-input multiple-output (MIMO) frequency-selective spatially correlated fading channels, assuming that the channel is unknown at the transmitter and perfectly known at the receiver. For Gaussian codebooks, using results from multivariate statistics, we derive an analytical expression for a tight lower bound on the ergodic capacity of such channels at any signal-to-noise ratio (SNR). We show that our bound is tighter than previously reported analytical lower bounds, and we proceed to analytically quantify the impact of spatial fading correlation on ergodic capacity. Based on a closed-form approximation of the variance of mutual information in correlated flat-fading MIMO channels, we provide insights into the multiplexing-diversity tradeoff for Gaussian code books. Furthermore, for a given total number of antennas, we consider the problem of finding the optimal (ergodic capacity maximizing) number of transmit and receive antennas, and we reveal the SNR-dependent nature of the maximization strategy. Finally, we present numerical results and comparisons between our capacity bounds and previously reported bounds.

Index Terms—Channel capacity, diversity, MIMO, spatial multiplexing.

I. INTRODUCTION

THE use of multiple antennas at both ends of a wireless link enables the opening of multiple spatial data pipes between transmitter and receiver within the frequency band of operation for no additional power expenditure. This leads to a dramatic increase in spectral efficiency [1]–[5] known as spatial multiplexing gain. Analytical expressions for the resulting capacity gains are in general hard to obtain.

Contributions: In this paper, we examine the statistics of mutual information [6] of multiple-input multiple-output (MIMO) channels, assuming that Gaussian code books are employed and that the channel is unknown at the transmitter and perfectly known at the receiver. Our detailed contributions are as follows.

- We derive a *tight closed-form lower bound on ergodic capacity* of MIMO channels experiencing spatially corre-

lated frequency-selective Rayleigh fading. Moreover, we provide an accurate closed-form analytical approximation of the variance of mutual information in the high signal-to-noise ratio (SNR) regime under (correlated) frequency-flat fading.

- We analytically quantify the impact of spatial fading correlation both at the transmitter and the receiver on ergodic capacity and the variance of mutual information.
- Using our analytical expressions for ergodic capacity and for the variance of mutual information, we provide insights into the *tradeoff between multiplexing and diversity gains* in frequency-flat correlated MIMO channels.
- Given a *fixed total number of antennas*, we determine the *optimal* (ergodic capacity maximizing) number of transmit and receive antennas and show the *SNR dependent* nature of the optimal antenna allocation strategy.

Relation to Previous Work: Expressions for the ergodic capacity of i.i.d. Rayleigh flat-fading MIMO channels under the assumption that the channel is unknown at the transmitter and perfectly known at the receiver have been derived in [2], [3]. Specifically, [2] gives closed-form expressions for ergodic capacity in integral form involving Laguerre polynomials and provides a look-up table obtained by numerically evaluating the underlying integrals to find the associated values of ergodic capacity for different numbers of transmit and receive antennas. On the other hand, [3] derives a lower bound on ergodic capacity that can be evaluated numerically using Monte Carlo methods. In [7] and [8], closed-form lower bounds for the ergodic capacity of i.i.d. Rayleigh flat-fading channels with multiple antennas have been reported. While [7] treats the case of multiple antennas at one end of the link (SIMO or MISO) and specifies ergodic capacity for MIMO channels with the aid of a look-up table for only a few antenna configurations, [8] derives a more general expression that applies to any antenna configuration. In both cases, the bounds are derived assuming high SNR, which leads to poor accuracy at low SNR. Expressions for ergodic and outage capacities of MIMO channels based on the Gaussian approximation of mutual information have been reported in [9]–[11].

The analysis in this paper distinguishes itself from previous results in that it provides a *tighter closed-form lower bound* than the one reported in [7] and [8] at any SNR and for any number of transmit and receive antennas. Moreover, our analytical lower bound is as tight as the bound obtained by numerically evaluating (through Monte Carlo methods) the lower bound derived in [3]. Additionally, our results incorporate the frequency-selective case and the case of spatial fading correlation and enable us

Manuscript received December 13, 2002; revised June 3, 2003. The work of Ö. Oyman was supported by the Benchmark Stanford Graduate Fellowship. The work of H. Bölcskei was supported by the National Science Foundation under Grants CCR 99-79381 and ITR 00-85929. This paper was presented in part at IEEE GLOBECOM 2002, Taipei, Taiwan, R.O.C., November 2002, and in part at the Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 2002. The associate editor coordinating the review of this paper and approving it for publication was Dr. Rick S. Blum.

Ö. Oyman and A. J. Paulraj are with the Information Systems Laboratory, Stanford University, Stanford, CA 94305 USA (e-mail: oyman@stanford.edu; apaulraj@stanford.edu).

R. U. Nabar and H. Bölcskei are with the Communication Technology Laboratory, Swiss Federal Institute of Technology (ETH), CH-8092 Zürich, Switzerland (e-mail: nabar@nari.ee.ethz.ch; boelcskei@nari.ee.ethz.ch).

Digital Object Identifier 10.1109/TSP.2003.818153

to quantify the loss in ergodic capacity due to spatial fading correlation analytically. We note that we have previously reported some of our results in [12] and [13].

Organization of the Paper: The rest of this paper is organized as follows. In Section II, we introduce the channel model and state our assumptions. In Section III, we derive bounds on ergodic and outage capacities of i.i.d. Rayleigh flat-fading MIMO channels. Some of these results are then extended to incorporate spatially correlated frequency-selective fading channels. In Section IV, we provide a closed-form approximation of the variance of mutual information in the high-SNR regime. Based on these results, in Section V, we provide insights into the multiplexing-diversity tradeoff in spatially correlated frequency-flat fading MIMO channels assuming Gaussian code books. Section VI examines the ergodic capacity maximizing antenna allocation strategies for fixed total number of antennas in the uncorrelated Rayleigh flat-fading case. Our conclusions are provided in Section VII.

Notation: \mathcal{E} denotes the expectation operator, $\text{var}(X)$ stands for the variance of the random variable X , \mathbf{I}_M is the $M \times M$ identity matrix, $\mathbf{0}_{M \times N}$ stands for the $M \times N$ all zeros matrix, $r(\mathbf{A})$ is the rank of the matrix \mathbf{A} , $\lambda_i(\mathbf{A})$ is the i th eigenvalue of \mathbf{A} , $\mathbf{\Lambda}_A$ is the diagonal matrix containing the nonzero eigenvalues of \mathbf{A} , $\text{vec}(\mathbf{A}) = [\mathbf{a}_0^T \ \mathbf{a}_1^T \ \dots \ \mathbf{a}_{N-1}^T]^T$, where \mathbf{a}_i denotes the i th column of \mathbf{A} , $\text{Tr}(\mathbf{A})$ is the trace of \mathbf{A} , and $\det(\mathbf{A})$ denotes the determinant of \mathbf{A} . The superscripts $T, *, H$ stand for transpose, elementwise conjugation, and conjugate transpose, respectively. The squared Frobenius norm of \mathbf{A} is defined as $\|\mathbf{A}\|_F^2 = \sum_{i,j} |[\mathbf{A}]_{i,j}|^2$. A circularly symmetric complex Gaussian random variable is a random variable $z = (x + jy) \sim \mathcal{CN}(0, \sigma^2)$, where x and y are i.i.d. $\mathcal{N}(0, \sigma^2/2)$. The chi-squared distribution with N degrees of freedom, which is denoted as χ_N^2 , is defined as the distribution of the sum of the squares of N i.i.d. $\mathcal{N}(0, 1/2)$ random variables.

II. CHANNEL MODEL

In the following, M_T and M_R denote the number of transmit and receive antennas, respectively. We assume a single-user point-to-point link and employ the broadband Rayleigh MIMO channel model introduced in [14]. Denoting the discrete-time index by k , the input-output relation for this channel is given by

$$\mathbf{r}[k] = \sum_{l=0}^{P-1} \mathbf{H}_l \mathbf{s}[k-l] + \mathbf{n}[k] \quad (1)$$

where $\mathbf{r}[k]$ is the $M_R \times 1$ received vector sequence, \mathbf{H}_l ($l = 0, 1, \dots, P-1$) is the $M_R \times M_T$ matrix channel impulse response, $\mathbf{s}[k]$ is the $M_T \times 1$ transmit vector sequence, and $\mathbf{n}[k]$ is the $M_R \times 1$ zero-mean additive white Gaussian noise vector satisfying $\mathcal{E}\{\mathbf{n}[k]\mathbf{n}^H[k']\} = \mathbf{I}_{M_R} \delta[k-k']$. For the sake of simplicity, we assume a uniform linear array at both the transmitter and the receiver with identical antenna elements [15], [16]. The relative antenna spacing is denoted as $\Delta = d/\lambda$, where d stands for the absolute antenna spacing, and $\lambda = c/f_c$ is the wavelength of a signal with center frequency f_c . Assuming that the fading statistics are the same for all transmit antennas and denoting the n th column of \mathbf{H}_l by $\mathbf{h}_n^{(l)}$, we define the $M_R \times$

M_R receive correlation matrix as $\mathbf{R}_l = \sigma_l^{-1} \mathcal{E}\{\mathbf{h}_n^{(l)} \mathbf{h}_n^{(l)H}\}$, which is independent of n , and where σ_l^2 is the average path gain of the l th tap (derived from the power delay profile). Letting $h_{m,n}^{(l)} = [\mathbf{H}_l]_{m,n}$ and denoting the fading correlation between two receive antenna elements spaced $s\Delta$ wavelengths apart as $\rho_l(s\Delta) = \sigma_l^{-2} \mathcal{E}\{h_{r,k}^{(l)} h_{r+s,k}^{(l)*}\}$, $l = 0, 1, \dots, P-1$, $k = 0, 1, \dots, M_T-1$, the receive correlation matrix \mathbf{R}_l can be written as

$$[\mathbf{R}_l]_{m,n} = \sigma_l \rho_l((n-m)\Delta). \quad (2)$$

We assume that spatial fading correlation can occur both at the transmitter and the receiver, the impact of which is modeled by decomposing the l th tap according to

$$\mathbf{H}_l = \mathbf{R}_l^{1/2} \mathbf{H}_{w,l} \mathbf{S}_l^{1/2}, \quad l = 0, 1, \dots, P-1 \quad (3)$$

with the $\mathbf{H}_{w,l}$ ($l = 0, 1, \dots, P-1$) being $M_R \times M_T$ matrices with i.i.d. $\mathcal{CN}(0, 1)$ entries. The transmit correlation matrix $\mathbf{S}_l = \mathbf{S}_l^{1/2} \mathbf{S}_l^{1/2}$ is defined similarly to the receive correlation matrix. The $\mathbf{H}_{w,l}$ are furthermore assumed to be uncorrelated, i.e., $\mathcal{E}\{\text{vec}(\mathbf{H}_{w,l}) \text{vec}(\mathbf{H}_{w,l'})^H\} = \mathbf{I}_{M_T M_R} \delta[l-l']$. We note that the decomposition (3) does not incorporate the most general case of spatial fading correlation but yields a reasonable compromise between analytical tractability and validity of the channel model. For a detailed discussion on the implications of this model, see [16].

In the following, we will allow for receive correlation only (i.e., $\mathbf{S}_l = \sigma_l \mathbf{I}_{M_T}$, $l = 0, 1, \dots, P-1$) in the frequency-selective fading case. Moreover, in the frequency-flat fading case, we will always assume $\sigma_0^2 = 1$, and use the notation $\mathbf{S}_0 = \mathbf{S}$ and $\mathbf{R}_0 = \mathbf{R}$.

III. BOUNDS ON ERGODIC CAPACITY AND OUTAGE CAPACITY

In this section, taking into account the channel model introduced in Section II, we derive bounds on ergodic and outage capacities. Throughout the paper, we assume that the channel is unknown at the transmitter and perfectly known at the receiver.

A. Ergodic Capacity Bounds

The i.i.d. Flat-Fading Case: We start by considering the i.i.d. frequency-flat fading case. The input-output relation is characterized by the $M_R \times M_T$ channel transfer matrix \mathbf{H} consisting of i.i.d. $\mathcal{CN}(0, 1)$ entries. The mutual information in bits per second per Hertz (bps/Hz) of the corresponding MIMO system is given by [2], [3]

$$I = \log_2 \det \left(\mathbf{I}_{M_R} + \frac{\rho}{M_T} \mathbf{H} \mathbf{H}^H \right) \quad (4)$$

where ρ is the average SNR at each of the receive antennas, and the input signal vector was assumed to be circularly symmetric complex Gaussian with covariance matrix $(\rho/M_T) \mathbf{I}_{M_T}$. Assuming that the fading process is ergodic, a Shannon capacity or ergodic capacity exists and is given by $C = \mathcal{E}\{I\}$.

Theorem 1: The ergodic capacity of an i.i.d. $M_R \times M_T$ frequency-flat fading MIMO channel can be lower bounded as

$$C \geq L \log_2 \left(1 + \frac{\rho}{M_T} \exp \left(\frac{1}{L} \sum_{j=1}^L \sum_{p=1}^{K-j} \frac{1}{p} - \gamma \right) \right) \quad (5)$$

where $L = \min(M_T, M_R)$, $K = \max(M_T, M_R)$, and $\gamma \approx 0.57721566$ is Euler's constant.

Proof: Applying Minkowski's inequality [17] to (4), we can lower bound mutual information in the case of $M_T \geq M_R$ as

$$I \geq \log_2 \left(1 + \rho \left[\det \left(\frac{1}{M_T} \mathbf{H} \mathbf{H}^H \right) \right]^{1/M_R} \right)^{M_R} \quad (6)$$

which can alternatively be expressed as

$$I \geq M_R \log_2 \left(1 + \rho \exp \left(\frac{1}{M_R} \ln \det \left(\frac{1}{M_T} \mathbf{H} \mathbf{H}^H \right) \right) \right). \quad (7)$$

Noting that $\log_2(1+ae^x)$ is a convex function in x for $a > 0$ and using Jensen's inequality [18], we can lower bound C according to

$$C \geq M_R \log_2 \left(1 + \rho \exp \left(\frac{1}{M_R} \times \mathcal{E} \left\{ \ln \det \left(\frac{1}{M_T} \mathbf{H} \mathbf{H}^H \right) \right\} \right) \right). \quad (8)$$

For $M_T \geq M_R$, we can infer from [19] that

$$\mathcal{E} \left\{ \ln \det \left(\frac{1}{M_T} \mathbf{H} \mathbf{H}^H \right) \right\} = \sum_{j=1}^{M_R} \mathcal{E} \{ \ln X_j \} - M_R \ln M_T \quad (9)$$

where the $X_j \sim \chi_{2(M_T-j+1)}^2$ ($j = 1, 2, \dots, M_R$) are independent. From [20], we know that

$$\mathcal{E} \{ \ln X_j \} = \psi(M_T - j + 1) \quad (10)$$

where $\psi(x)$ is the digamma function. For integer x , this function may be expressed as [21]

$$\psi(x) = -\gamma + \sum_{p=1}^{x-1} \frac{1}{p}. \quad (11)$$

Combining (8)–(11), we have a lower bound on the ergodic capacity when $M_T \geq M_R$. Using the identity $\det(\mathbf{I}_{M_R} + (\rho/M_T)\mathbf{H}\mathbf{H}^H) = \det(\mathbf{I}_{M_T} + (\rho/M_T)\mathbf{H}^H\mathbf{H})$, similar steps can be pursued to derive a lower bound for the case $M_T < M_R$. \square

Proposition 2: In the high-SNR regime, the ergodic capacity of an i.i.d. $M_R \times M_T$ frequency-flat fading MIMO channel can be approximated as

$$C \approx L \log_2 \left(\frac{\rho}{M_T} \right) + \frac{1}{\ln 2} \left(\sum_{j=1}^L \sum_{p=1}^{K-j} \frac{1}{p} - \gamma L \right). \quad (12)$$

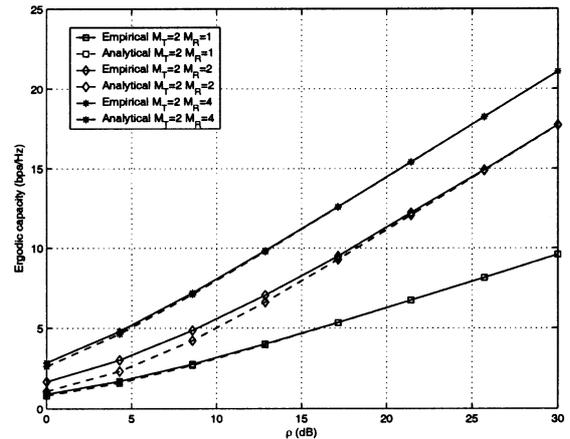


Fig. 1. Comparison of the empirically determined ergodic capacity and the analytical lower bound in (5) for several antenna configurations.

Proof: We start by noting that for ρ large, the mutual information I can be approximated as

$$I \approx \begin{cases} \log_2 \det \left(\frac{\rho}{M_T} \mathbf{H} \mathbf{H}^H \right), & M_T \geq M_R \\ \log_2 \det \left(\frac{\rho}{M_T} \mathbf{H}^H \mathbf{H} \right), & M_T < M_R. \end{cases} \quad (13)$$

Using $C = \mathcal{E} \{ I \}$ and applying steps similar to (9)–(11), the result follows. \square

Proposition 2 is intuitively appealing since it shows explicitly that ergodic capacity grows linearly with $L = \min(M_T, M_R)$. More specifically, C increases by $\min(M_T, M_R)$ bps/Hz for every 3-dB increase in SNR. Thus, the number of spatial data pipes that can be opened up between transmitter and receiver is constrained by the minimum of the number of antennas at the transmitter and the receiver. This is a well-known fact first observed in [22] without providing an explicit analytical expression for C . We note that (12) has been found independently in [8]; more precisely, (12) is the equivalent to the lower bound given by (9) in [8]. In (12), the digamma function in (9) of [8] is given in expanded form based on (11).

Numerical Example 1 (Tightness of the General Bound). Fig. 1 shows the empirically obtained (through Monte Carlo methods) ergodic capacity and the analytical lower bound (5) for several MIMO configurations as a function of average SNR per receive antenna. It is clearly seen that (5) becomes almost exact at high SNR. We can furthermore observe that in the low-SNR regime, the bound improves as the difference in the number of antennas on the two sides of the link increases. \square

We note that the lower bound in (5) can be improved by starting from the lower bound on mutual information given in (12) of [3], where it is assumed that $M_T \geq M_R$. We generalize this expression to apply for any M_T, M_R by expressing the lower bound as

$$I \geq \sum_{j=1}^L \log_2 \left(1 + \frac{\rho}{M_T} X_j \right) \quad (14)$$

where the $X_j \sim \chi_{2(K-j+1)}^2$ ($j = 1, 2, \dots, L$) are independent.

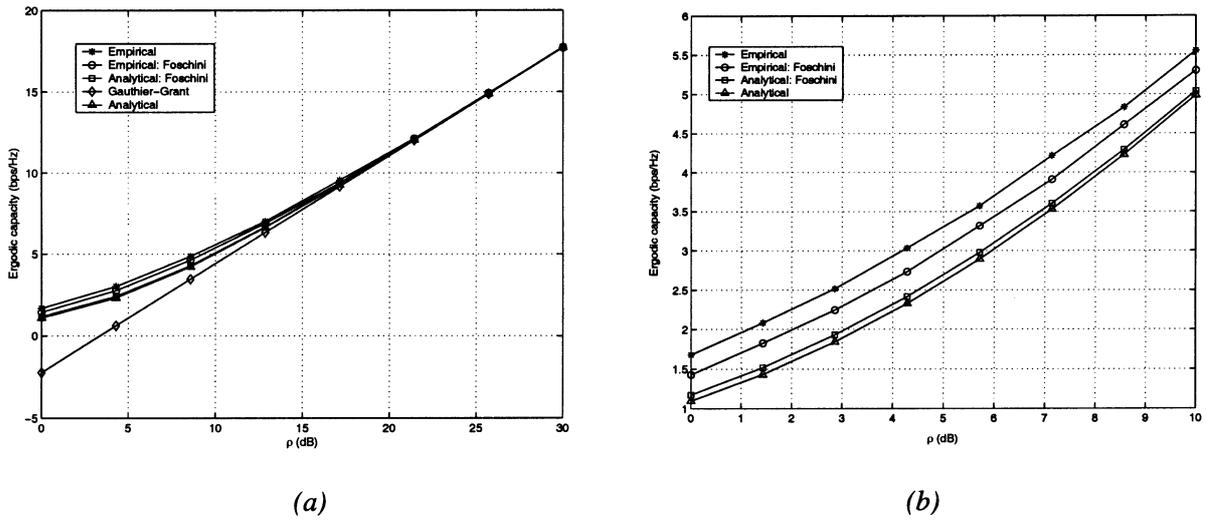


Fig. 2. (a) Comparison of lower bounds (5) and (15) with known lower bounds on ergodic capacity for the i.i.d. MIMO channel with $M_T = M_R = 2$. (b) Zoomed view at low SNR.

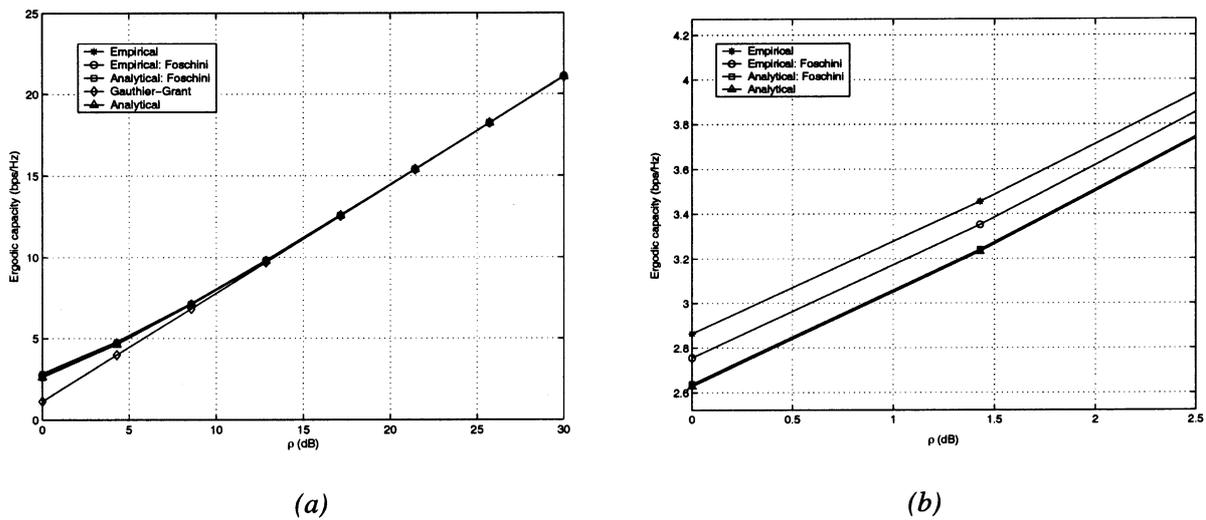


Fig. 3. (a) Comparison of lower bounds (5) and (15) with known lower bounds on ergodic capacity for the i.i.d. MIMO channel with $M_T = 2, M_R = 4$. (b) Zoomed view at low SNR.

Proposition 3: Compared with (5), an improved lower bound on the ergodic capacity of the i.i.d. frequency-flat fading $M_R \times M_T$ MIMO channel is given by

$$C \geq \sum_{j=1}^L \log_2 \left(1 + \frac{\rho}{M_T} \exp \left(\sum_{p=1}^{K-j} \frac{1}{p} - \gamma \right) \right). \quad (15)$$

Proof: We will first derive (15) and then show that the bound in (15) is strictly better than the bound in (5). Starting from (14) and using $C = \mathcal{E}\{I\}$, we can follow the proof of Theorem 1 and further lower bound C through Jensen’s inequality (based on the convexity of $\log_2(1 + ae^x)$ in x for $a > 0$) and apply similar steps as in (9)–(11) to obtain (15). The second part of the proposition can be shown by applying Jensen’s inequality to

$$\frac{1}{L} \sum_{j=1}^L \log_2 \left(1 + \frac{\rho}{M_T} \exp \left(\sum_{p=1}^{K-j} \frac{1}{p} - \gamma \right) \right)$$

which yields

$$\begin{aligned} & \sum_{j=1}^L \log_2 \left(1 + \frac{\rho}{M_T} \exp \left(\sum_{p=1}^{K-j} \frac{1}{p} - \gamma \right) \right) \\ & \geq L \log_2 \left(1 + \frac{\rho}{M_T} \exp \left(\frac{1}{L} \sum_{j=1}^L \sum_{p=1}^{K-j} \frac{1}{p} - \gamma \right) \right) \end{aligned}$$

and hence proves that the bound in (15) is strictly better than the bound in (5). \square

Numerical Example 2 (Comparison of the General Bounds with Known Bounds). Next, we compare the closed-form expressions (5) and (15) with previously reported lower bounds. For $M_T = M_R = 2$ and $M_T = 2, M_R = 4$, Figs. 2 and 3, respectively, show the lower bounds reported in [7] and [8] (Gauthier–Grant), the lower bound obtained by evaluating the results in [3] through Monte Carlo methods (Empirical: Foschini) and the analytical lower bounds (5) (Analytical) and (15)

(Analytical: Foschini). We observe that in the low-SNR regime (5) and (15) are equally tight [consistent with Proposition 3, the bound (15) is slightly better in the $M_T = M_R = 2$ case, as seen in Fig. 2(b)], and both bounds are close to the numerically evaluated lower bound of [3]. Furthermore, we note that (5) and (15) are much tighter than the lower bound specified in [7] and [8] (Gauthier–Grant). In the high-SNR regime, all bounds are equally tight. We also note that the accuracy of the analytical lower bounds is better in the asymmetric setting of $M_T = 2$, $M_R = 4$. \square

Correlated Flat-Fading Case: Based on the techniques established so far, we can easily lower bound the ergodic capacity of channels with correlated spatial fading for the following three cases.

- In the case of receive correlation only (i.e., $\mathbf{S} = \mathbf{I}_{M_T}$), assuming that \mathbf{R} has full rank (i.e., $r(\mathbf{R}) = M_R$) and $M_R \leq M_T$, the lower bound on ergodic capacity becomes

$$C \geq M_R \log_2 \left(1 + \frac{\rho}{M_T} (\det(\mathbf{\Lambda}_R))^{1/M_R} \times \exp \left(\frac{1}{M_R} \sum_{j=1}^{M_R} \sum_{p=1}^{M_T-j} \frac{1}{p} - \gamma \right) \right). \quad (16)$$

In order to generalize (16) to arbitrary rank of \mathbf{R} , we observe that the effective channel dimensions in this case reduce to $r(\mathbf{R}) \times M_T$. Assuming that $r(\mathbf{R}) \leq M_T$, the ergodic capacity is lower bounded as

$$C \geq r(\mathbf{R}) \log_2 \left(1 + \frac{\rho}{M_T} (\det(\mathbf{\Lambda}_R))^{1/r(\mathbf{R})} \times \exp \left(\frac{1}{r(\mathbf{R})} \sum_{j=1}^{r(\mathbf{R})} \sum_{p=1}^{M_T-j} \frac{1}{p} - \gamma \right) \right). \quad (17)$$

Comparing (17) with (5), we can see that the number of spatial data pipes opened up in this case is given by $r(\mathbf{R})$.

- Similarly, for the case of spatial fading correlation at the transmitter only (i.e., $\mathbf{R} = \mathbf{I}_{M_R}$), assuming that $r(\mathbf{S}) \leq M_R$, we obtain

$$C \geq r(\mathbf{S}) \log_2 \left(1 + \frac{\rho}{M_T} (\det(\mathbf{\Lambda}_S))^{1/r(\mathbf{S})} \times \exp \left(\frac{1}{r(\mathbf{S})} \sum_{j=1}^{r(\mathbf{S})} \sum_{p=1}^{M_R-j} \frac{1}{p} - \gamma \right) \right). \quad (18)$$

Again, we can see that the number of effective spatial data pipes is given by $r(\mathbf{S})$. We emphasize that in the case of transmit correlation only, choosing the transmit signal's covariance matrix to satisfy $\mathbf{Q} = \mathcal{E} \{ \mathbf{ss}^H \} = \rho/M_T \mathbf{I}_{M_T}$ will in general only yield a lower bound on ergodic capacity. Strategies for determining the transmit signal vector's optimal covariance matrix assuming knowledge of \mathbf{S} have been discussed in [23]–[28].

- Finally, in the case of spatial fading correlation both at the transmitter and the receiver, assuming that $r(\mathbf{S}) = r(\mathbf{R}) = N$, we obtain

$$C \geq N \log_2 \left(1 + \frac{\rho}{N} (\det(\mathbf{\Lambda}_S) \det(\mathbf{\Lambda}_R))^{1/N} \times \exp \left(\frac{1}{N} \sum_{j=1}^N \sum_{p=1}^{N-j} \frac{1}{p} - \gamma \right) \right) \quad (19)$$

which shows that in this case the number of spatial data pipes is given by N .

In the high-SNR regime, the lower bound in (5) approximates ergodic capacity for the i.i.d. case very accurately. Similarly (17)–(19) give very accurate approximations in the presence of spatial fading correlation. The bound reported in (15) has been generalized in [29] to the cases of spatial fading correlation only at the transmitter or only at the receiver. However, the resulting expressions are not given in analytic form and need to be evaluated via Monte Carlo methods. Based on the lower bounds established in (5) and (17)–(19), we can now analytically quantify the loss in ergodic capacity due to spatial fading correlation in the high-SNR regime. For the MIMO channel with joint transmit/receive correlation, at high SNR, assuming that \mathbf{S} and \mathbf{R} have full rank, the ergodic capacity loss is given by $\log_2(\det(\mathbf{\Lambda}_S)) + \log_2(\det(\mathbf{\Lambda}_R))$ bps/Hz. This loss is $\log_2(\det(\mathbf{\Lambda}_S))$ and $\log_2(\det(\mathbf{\Lambda}_R))$ bps/Hz, respectively, for the cases of transmit correlation only and receive correlation only. Furthermore, we note that the number of spatial data pipes opened up between transmitter and receiver is constrained by the rank of the correlation matrices. This observation can be interpreted as having $r(\mathbf{R})$ effective receive and $r(\mathbf{S})$ effective transmit antennas. We emphasize that the ergodic capacity loss was quantified in terms of the lower bounds (17)–(19). Since these bounds are very accurate in the high-SNR regime, the conclusions drawn above on the ergodic capacity loss can be expected to match well the exact behavior.

Numerical Example 3 (Impact of Spatial Fading Correlation on Ergodic Capacity). In this example, we investigate the ergodic capacity loss due to spatial fading correlation for a Rayleigh flat-fading MIMO channel with $M_T = M_R = 2$. We use the channel model specified in (3) with

$$\mathbf{S} = \begin{bmatrix} 1 & s \\ s^* & 1 \end{bmatrix} \text{ and } \mathbf{R} = \begin{bmatrix} 1 & r \\ r^* & 1 \end{bmatrix} \quad (20)$$

where s and r are the complex correlation coefficients between the two transmit and the two receive antennas, respectively. In Fig. 4, we compare the lower bound (19) with the empirically obtained ergodic capacity (through Monte Carlo methods) for three different levels of transmit and receive correlation, namely, $s = r = 0$ (i.i.d. channel), $s = r = 0.4$ (low transmit/receive correlation), and $s = 0.95, r = 0.4$ (high transmit, low receive correlation). As predicted by the analytical estimate $\log_2(\det(\mathbf{\Lambda}_S)) + \log_2(\det(\mathbf{\Lambda}_R)) = \log_2(\det(\mathbf{S})) + \log_2(\det(\mathbf{R})) = \log_2(1 - |s|^2) + \log_2(1 - |r|^2)$, we observe a very small ergodic capacity loss for the case of low transmit/receive correlation. In the case of high correlation at the transmitter, we observe an ergodic capacity loss of about 3.6 bps/Hz, which is consistent with the loss of 3.61 bps/Hz,

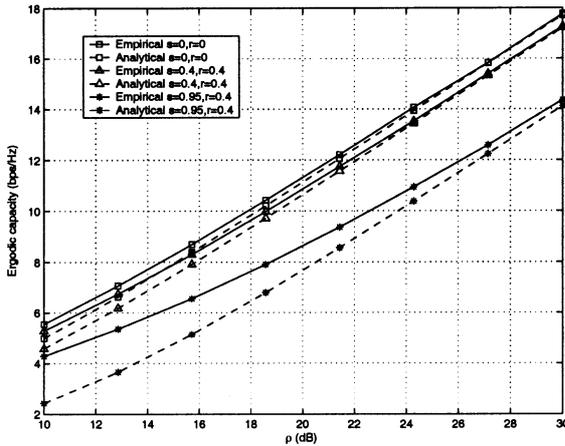


Fig. 4. Comparison of the empirically determined ergodic capacity and the analytical lower bound (19) for $M_T = M_R = 2$ and various levels of spatial fading correlation.

predicted by the analytical estimate. Again, we note that the analytical estimates become very accurate for high SNR. At low SNR, the analytical lower bounds (17)–(19) become more inaccurate with increasing correlation level. \square

Extension to Frequency-Selective Fading Channels: We will next assume frequency-selective fading with $\mathbf{S}_l = \sigma_l \mathbf{I}_{M_T}$ ($l = 0, 1, \dots, P-1$). The ergodic capacity for spatial multiplexing over this channel has been obtained in [5] as

$$C = \mathcal{E} \left\{ \log_2 \det \left(\mathbf{I}_{M_R} + \frac{\rho}{M_T} \mathbf{\Lambda} \mathbf{H}_w \mathbf{H}_w^H \right) \right\} \quad (21)$$

where \mathbf{H}_w is an $M_R \times M_T$ matrix consisting of i.i.d. $\mathcal{CN}(0, 1)$ elements, and $\mathbf{\Lambda} = \text{diag} \left\{ \lambda_i \left(\tilde{\mathbf{R}} \right) \right\}_{i=0}^{M_R-1}$ with $\tilde{\mathbf{R}} = \sum_{l=0}^{P-1} \sigma_l \mathbf{R}_l$. Following our previous analysis, it is easy to verify that the ergodic capacity in (21) for the case where $r(\tilde{\mathbf{R}}) \leq M_T$ may be conveniently lower bounded as

$$C \geq r(\tilde{\mathbf{R}}) \log_2 \left(1 + \frac{\rho}{M_T} \left(\det(\tilde{\mathbf{\Lambda}}_{\tilde{\mathbf{R}}}) \right)^{1/r(\tilde{\mathbf{R}})} \right) \times \exp \left(\frac{1}{r(\tilde{\mathbf{R}})} \sum_{j=1}^{r(\tilde{\mathbf{R}})} \sum_{p=1}^{M_T-j} \frac{1}{p} - \gamma \right). \quad (22)$$

Similar to the flat-fading case, we can establish that the ergodic capacity increases by $r(\tilde{\mathbf{R}})$ bps/Hz for every 3-dB increase in SNR or, equivalently, the number of spatial data pipes is given by $r(\tilde{\mathbf{R}})$. We conclude by noting that for full-rank $\tilde{\mathbf{R}}$, the loss in ergodic capacity in the high-SNR regime is quantified by $\log_2 \left(\det(\tilde{\mathbf{R}}) \right)$ bps/Hz.

Improved Lower Bound in the Low-SNR Regime: In the following, we focus on the i.i.d. flat-fading case and provide an improved lower bound on ergodic capacity in the low-SNR regime. We start with the following proposition.

Proposition 4: The mutual information defined in (4) can be lower bounded by

$$I \geq \log_2 \left(1 + \frac{\rho}{M_T} \text{Tr}(\mathbf{H}\mathbf{H}^H) \right) = \log_2 \left(1 + \frac{\rho}{M_T} \|\mathbf{H}\|_F^2 \right). \quad (23)$$

Moreover, in the low-SNR regime, this lower bound is strictly tighter than the bound (6).

Proof: We expand (4) by using

$$\det \left(\mathbf{I}_{M_R} + \frac{\rho}{M_T} \mathbf{H}\mathbf{H}^H \right) = 1 + \sum_{i=1}^{M_R} \left(\frac{\rho}{M_T} \right)^i \sum_{k_1 < \dots < k_i} \prod_{j=1}^i \lambda_{k_j}(\mathbf{H}\mathbf{H}^H)$$

where the notation $k_1 < \dots < k_i$ denotes all combinations of size i from the set $\{1, 2, \dots, M_R\}$. Noting that the terms $\lambda_{k_j}(\mathbf{H}\mathbf{H}^H)$ are positive with probability 1 and ignoring second and higher order terms (i.e., the cross-terms) in ρ , we obtain the result in (23). We can confirm that (23) is an improved bound on mutual information at low SNR over the bound obtained in (6) through Minkowski's inequality. Using the binomial theorem [30], we can rewrite (6) as

$$I \geq \log_2 \left(\sum_{i=0}^{M_R} \binom{M_R}{i} \left(\frac{\rho}{M_T} \right)^i \left[\det(\mathbf{H}\mathbf{H}^H) \right]^{i/M_R} \right).$$

For small ρ , second and higher order terms can be ignored. Thus, the generalized low-SNR bound on mutual information based on Minkowski's inequality is given by

$$I \geq \log_2 \left(1 + \frac{M_R \rho}{M_T} \left[\det(\mathbf{H}\mathbf{H}^H) \right]^{1/M_R} \right). \quad (24)$$

We observe, based on the arithmetic mean-geometric mean inequality, that (23) is a tighter lower bound on mutual information than (24). \square

Proposition 5: The ergodic capacity of an i.i.d. $M_R \times M_T$ frequency-flat fading MIMO channel at low SNR is lower bounded by

$$C \geq \log_2 \left(1 + \frac{\rho}{M_T} \exp \left(\sum_{p=1}^{M_T M_R - 1} \frac{1}{p} - \gamma \right) \right). \quad (25)$$

Proof: The result is a direct consequence of the mutual information bound (23) obtained in Proposition 4. Noting that $\text{Tr}(\mathbf{H}\mathbf{H}^H) = \|\mathbf{H}\|_F^2 \sim \chi_{2M_T M_R}^2$ and applying the steps used in the proof of Theorem 1, namely Jensen's inequality followed by the steps leading to (9)–(11), the desired result follows. \square

Numerical Example 4 (Comparison of the Generalized Bound and the Improved Low-SNR Bound). In Fig. 5, we focus on the low-SNR regime and compare the general lower bound in (5) to the improved low-SNR bound (25) for various antenna configurations. We observe, in particular, in the symmetric case $M_T = M_R = 2$ that (25) serves as a significantly more precise estimate of ergodic capacity at low SNR. \square

B. Bounds on Outage Capacity

Extending the analysis developed thus far, we will next establish a high-SNR lower bound on the cdf of mutual information and, hence, an upper bound on outage capacity.

The i.i.d. Case: Since \mathbf{H} is random, the mutual information I is random as well. The outage probability at rate R is defined as $P_{\text{out},R} = P(I \leq R)$ [2], [6], [31]. Equivalently, one can define the $q\%$ outage capacity $C_{\text{out},q}$ as the capacity that is guaranteed for $(100 - q)\%$ of the channel realizations, i.e.,

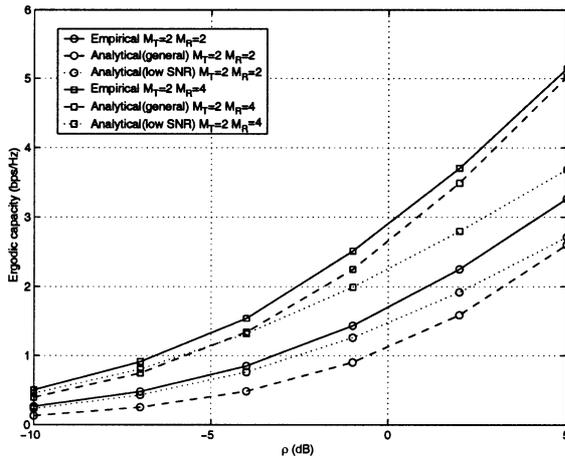


Fig. 5. Comparison of general lower bound (5) and low-SNR lower bound (25) for $M_T = M_R = 2$ and $M_T = 2, M_R = 4$.

$P(I \leq C_{out,q}) = q\%$ [2], [6], [31]. We denote the cdf of mutual information by $F_I(y) = P(I \leq y)$.

Theorem 6: For high SNR, the cdf of mutual information of the i.i.d. frequency-flat fading $M_R \times M_T$ MIMO channel is lower bounded as

$$F_I(y) \geq 1 - \min_{s \geq 0} e^{-sy} \mathcal{E}\{e^{sI}\} \quad (26)$$

where

$$\mathcal{E}\{e^{sI}\} = \left(\frac{\rho}{M_T}\right)^{Ls \log_2 e} \prod_{j=1}^L \frac{\Gamma(K+1-j+s \log_2 e)}{\Gamma(K+1-j)} \quad (27)$$

with $\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy$ denoting the Gamma function.

Proof: The expression in (26) follows from the Chernoff bound. We obtain (27) using the results of [19] and the high-SNR approximation for mutual information in (13). \square

We note that a similar bound for the frequency-selective case has been reported previously in [32].

Numerical Example 5 (Accuracy of the Bound on Outage Capacity). In Fig. 6, we compare the empirically obtained (via Monte Carlo methods) cdf of mutual information to the analytical expression in (26) and (27) for various antenna configurations at $\rho = 15$ dB. The Chernoff parameter s is selected to satisfy the minimization criterion in (26) for each outage level. While the bound is tight at high outage levels, the observed gap between empirical and analytical capacity is more than 2 bps/Hz at outage levels below 10%. \square

Correlated Case: The lower bound on the cdf of mutual information given by (26) still applies in the correlated case with (27) replaced as follows.

- In the case of receive correlation only (i.e., $\mathbf{S} = \mathbf{I}_{M_T}$), assuming that $r(\mathbf{R}) \leq M_T$, the moment-generating function for mutual information is given by

$$\mathcal{E}\{e^{sI}\} = \left(\det(\mathbf{\Lambda}_R) \left(\frac{\rho}{M_T}\right)^{r(\mathbf{R})}\right)^{s \log_2 e} \times \prod_{j=1}^{r(\mathbf{R})} \frac{\Gamma(M_T+1-j+s \log_2 e)}{\Gamma(M_T+1-j)}. \quad (28)$$

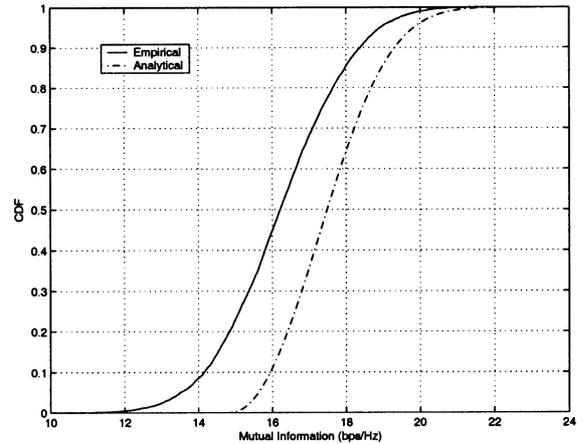


Fig. 6. Comparison of empirical and analytical cdfs of mutual information for the MIMO channel with $M_T = M_R = 4$ at $\rho = 15$ dB.

- Similarly, for the case of spatial fading correlation at the transmitter only (i.e., $\mathbf{R} = \mathbf{I}_{M_R}$), assuming that $r(\mathbf{S}) \leq M_R$, we obtain

$$\mathcal{E}\{e^{sI}\} = \left(\det(\mathbf{\Lambda}_S) \left(\frac{\rho}{M_T}\right)^{r(\mathbf{S})}\right)^{s \log_2 e} \times \prod_{j=1}^{r(\mathbf{S})} \frac{\Gamma(M_R+1-j+s \log_2 e)}{\Gamma(M_R+1-j)}. \quad (29)$$

- Finally, in the case of spatial fading correlation both at the transmitter and the receiver, assuming that $r(\mathbf{S}) = r(\mathbf{R}) = N$, we get

$$\mathcal{E}\{e^{sI}\} = \left(\det(\mathbf{\Lambda}_S) \det(\mathbf{\Lambda}_R) \left(\frac{\rho}{N}\right)^N\right)^{s \log_2 e} \times \prod_{j=1}^N \frac{\Gamma(N+1-j+s \log_2 e)}{\Gamma(N+1-j)}. \quad (30)$$

IV. VARIANCE OF MUTUAL INFORMATION

We will next compute the variance of mutual information for the frequency-flat spatially correlated fading case. In the following, the variance of I is denoted by σ_I^2 .

The i.i.d. Case: The variance result for the i.i.d. case can be summarized as follows:

Theorem 7: A general closed-form approximation of the variance of an i.i.d. frequency-flat $M_R \times M_T$ MIMO channel at high SNR is given by

$$\sigma_I^2 \approx (\log_2 e)^2 \sum_{j=1}^L \sum_{p=1}^{\infty} \frac{1}{(p+K-j)^2}. \quad (31)$$

A simpler approximate expression for σ_I^2 , which is often more amenable to analytical studies (but less accurate), is

$$\sigma_I^2 \approx (\log_2 e)^2 \sum_{j=1}^L \frac{1}{K-j+1}. \quad (32)$$

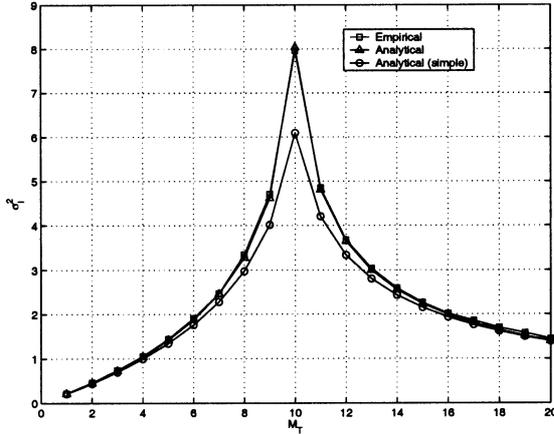


Fig. 7. Comparison of empirically determined σ_I^2 and analytical approximations (31) and (32) for various values of M_T with fixed $M_R = 10$ at high SNR.

Proof: We start with the high-SNR approximation of mutual information given in (13). From [19], we can infer that

$$\sigma_I^2 \approx (\log_2 e)^2 \sum_{j=1}^L \text{var}(\ln(X_j)) \quad (33)$$

where $X_j \sim \chi_{2(K-j+1)}^2$ ($j = 1, 2, \dots, L$) are independent, and K and L are as defined in Theorem 1. From [20], we know that

$$\text{var}(\ln(X_j)) = \psi'(K - j + 1) \quad (34)$$

where $\psi'(x)$ is the first derivative of the digamma function. By expanding (34) as [21]

$$\psi'(x) = \sum_{p=1}^{\infty} \frac{1}{(p+x-1)^2}$$

we arrive at (31). Alternatively, we can express $\psi'(x)$ in (34) in the form $(d^2/dx^2)(\ln \Gamma(x))$ and approximate $\Gamma(x)$ based on Stirling's formula as $\Gamma(x) \approx \sqrt{2\pi x} x^x e^{-x}$ to obtain $\psi'(x) \approx 1/x - 1/(2x^2)$, which letting $\psi'(x) \approx 1/x$, yields the simpler variance expression in (32). \square

Comparing (12) with (31), we can see that unlike ergodic capacity, the variance of I displays symmetry with respect to M_T and M_R at high SNR in the sense that σ_I^2 for a $k \times n$ system equals σ_I^2 for an $n \times k$ system.

Numerical Example 6 (Accuracy of the Variance Estimate). In Fig. 7, we compare the empirically obtained variance of I (through Monte Carlo methods), its analytical estimate (31) (summed over p ranging from 1 to 500) (Analytical), and the simpler but less accurate analytical approximation (32) (Analytical (simple)) for a system with $M_R = 10$ and varying M_T at high SNR. We observe that the empirical and analytical results are in close agreement. Moreover, we note that the results depicted in Fig. 7 are consistent with the results to be described in Section V, i.e., σ_I^2 is maximum for $M_T = M_R$. \square

Correlated Case: In the presence of spatial fading correlation, the variance of mutual information can be analytically specified for the following three cases.

- For the case of spatial fading correlation at the receiver only (i.e., $\mathbf{S} = \mathbf{I}_{M_T}$), assuming that $r(\mathbf{R}) \leq M_T$, the

variance of mutual information in the high-SNR regime is obtained as

$$\sigma_I^2 \approx (\log_2 e)^2 \sum_{j=1}^{r(\mathbf{R})} \sum_{p=1}^{\infty} \frac{1}{(p + M_T - j)^2}. \quad (35)$$

- Similarly, for the case of spatial fading correlation at the transmitter only (i.e., $\mathbf{R} = \mathbf{I}_{M_R}$), assuming that $r(\mathbf{S}) \leq M_R$, we obtain

$$\sigma_I^2 \approx (\log_2 e)^2 \sum_{j=1}^{r(\mathbf{S})} \sum_{p=1}^{\infty} \frac{1}{(p + M_R - j)^2}. \quad (36)$$

- Finally, for the case of joint transmit-receive correlation, assuming that $r(\mathbf{S}) = r(\mathbf{R}) = N$, we find

$$\sigma_I^2 \approx (\log_2 e)^2 \sum_{j=1}^N \sum_{p=1}^{\infty} \frac{1}{(p + N - j)^2}. \quad (37)$$

We note that we have incorporated the spatially correlated channel model described in Section II into the result of Theorem 7 in deriving (35)–(37).

Comparing (35)–(37) with (31), we can infer that the variance of mutual information of an $M_R \times M_T$ spatially correlated channel is the same as that in an i.i.d. channel with dimensions $r(\mathbf{R}) \times M_T$ and $M_R \times r(\mathbf{S})$, respectively, for the cases of receive-only and transmit-only correlation and $N \times N$ for the case of joint transmit-receive correlation (assuming equal ranks). Hence, we see that spatial fading correlation in general may reduce the variance of mutual information as the effective number of transmit and receive antennas is constrained by the rank of the correlation matrices.

Low-SNR Case: We can also analytically characterize the variance of mutual information in the low-SNR regime based on the bound in Proposition 4. Applying the approximation $\log_2(1+x) \approx (\log_2 e)x$ valid for small x to (23) and using $\text{var}(\|\mathbf{H}\|_F^2) = M_T M_R$, we obtain

$$\sigma_I^2 \approx \text{var} \left((\log_2 e) \frac{\rho}{M_T} \|\mathbf{H}\|_F^2 \right) \approx (\rho \log_2 e)^2 \frac{M_R}{M_T}. \quad (38)$$

We can see from (38) that in the low-SNR regime, variance of mutual information increases linearly in the number of receive antennas and is inversely proportional to the number of transmit antennas. This trend is clearly different from its behavior in the high-SNR regime, where σ_I^2 is maximum when $M_T = M_R$ and decreases monotonically with increasing $K = \max(M_T, M_R)$ for fixed $L = \min(M_T, M_R)$.

V. TRADEOFF BETWEEN SPATIAL MULTIPLEXING AND DIVERSITY

MIMO systems offer *spatial diversity gain* to combat channel fading and *spatial multiplexing gain* resulting in increased spectral efficiency. These benefits are, in general, conflicting. It is therefore necessary to understand the tradeoff between multiplexing and diversity gains in designing MIMO systems. The variance analysis in Section IV provides a new framework to interpret this tradeoff when signaling with Gaussian codebooks over ergodic MIMO channels. An alternative framework for analyzing the diversity-multiplexing tradeoff has been proposed in [33]. The definitions of multiplexing and diversity gains used

in [33] are related to outage capacity and the associated outage probability (or equivalently packet error rate), respectively.

In this paper, we assume ergodicity of the MIMO channel and relate multiplexing gain to ergodic capacity and diversity gain to outage capacity. Due to the fading nature of the channel, mutual information is a random quantity. In a single-input single-output (SISO) scenario, with a sufficiently long coding horizon, it is possible to code over short-term channel fluctuations and achieve ergodic capacity through averaging in time. In the MIMO case, we can see how the extra spatial dimensions contribute to this averaging based on the results of our variance analysis. In particular, we can conclude that the channel hardening effect observed under certain antenna configurations (i.e., $M_T \gg M_R$ and $M_T \ll M_R$, where the variance of mutual information is small due to spatial averaging) reduces the amount of time-averaging required (i.e., the number of independent fading blocks that the codewords need to span) to stabilize rate at ergodic capacity.¹ The amount of temporal averaging necessary is maximal when σ_I^2 is maximal, which happens for $M_T = M_R$. Our notion of diversity shapes around this phenomenon as a measure on the *rate* of spatial averaging. We emphasize that our definition of diversity gain differs from that used in [33]. We assume that transmission takes place in pure spatial multiplexing mode and study the impact of the number of transmit and receive antennas on multiplexing and diversity gains. Stabilizing rate at ergodic capacity in the MIMO case requires sufficient averaging over *each* of the spatial data pipes. Motivated by this observation, we define *per-stream diversity order* as the metric that captures the rate of spatial averaging on a per-stream basis.

Definitions of Multiplexing Gain and Diversity Gain: Throughout this section, we focus on the high-SNR case. Let us start by defining the multiplexing gain as

$$m = \lim_{\rho \rightarrow \infty} \frac{C}{\log_2 \rho}$$

which, using (12), yields the intuitive result

$$m = L$$

or equivalently, the multiplexing gain is given by the number of parallel spatial data pipes opened up between transmitter and receiver. In order to motivate our definition of diversity gain, let us consider the SIMO case, where $m = 1$, and [34]

$$\sigma_I^2 \approx \frac{(\log_2 e)^2}{M_R}. \quad (39)$$

Thus, the variance of mutual information is inversely proportional to the number of degrees of freedom provided by the channel and can therefore be seen as a measure of the diversity order supported by the channel. Motivated by (39) and the fact that i.i.d. (across space) code books are employed, we define the per-stream diversity order as

$$d(m) = \frac{(\log_2 e)^2}{\frac{\sigma_I^2}{m}}.$$

Our definition of diversity order generally reflects the notion of the fading channel approaching the AWGN channel as the

¹We say that the rate stabilizes at ergodic capacity if the pdf of I concentrates around its mean $C = \mathcal{E}\{I\}$.

number of degrees of freedom approaches infinity [35]. In order to see this, note that in the SIMO case, as $M_R \rightarrow \infty$, the variance $\sigma_I^2 \rightarrow 0$, and hence, $d(1) \rightarrow \infty$. In the MIMO case, the behavior of σ_I^2 is complicated by the self-interference cancellation penalty across multiplexed streams, i.e., the M_R receive antennas have to separate the M_T independent data streams (cancel self interference) as well as provide spatial diversity gain. To see this, we consider the simplified expression for variance in (32). The right-hand side of (32) may be interpreted as the sum of the variances of L parallel SIMO channels with successively decreasing diversity orders (the maximum diversity order being K and the minimum being $K-L+1$). This decrease in diversity order can be attributed to the self-interference cancellation penalty across the multiplexed streams. In order to shed more light on this observation and our definition of diversity order, consider the case $M_T = 2$ and $M_R \geq 2$. Employing our definitions of multiplexing gain and per-stream diversity order, we obtain $m = 2$ and $d(2) = M_R(2M_R - 2)/(2M_R - 1) < M_R$, which basically tells us that each of the two streams gets a diversity order less than M_R . Recall that in the SIMO case, the diversity order seen by the single data stream was exactly M_R . Now, in the $M_T = 2$ case for M_R large, the self-interference cancellation penalty vanishes, and the per-stream diversity order approaches M_R . This result is intuitively appealing as M_R becoming large means that the dimensionality of the receive signal space becomes large, and hence, it becomes increasingly easier to separate the two transmitted data streams.

We will next discuss the multiplexing-diversity tradeoff for a general number of antennas. The discussion will be done separately for the i.i.d. and the correlated fading cases.

The i.i.d. Case: Using (32), fixing K , and considering L variable (to reflect this we use the notation $d(L)$), we obtain

$$d(L) = \frac{L}{L + \frac{d(L-1)}{K-L+1} - 1} d(L-1). \quad (40)$$

Now, since

$$d(L-1) = \left(\frac{1}{L-1} \sum_{j=1}^{L-1} \frac{1}{K-j+1} \right)^{-1}$$

it follows that

$$d(L-1) \geq K - L + 2$$

and hence

$$\frac{d(L-1)}{K-L+1} > 1$$

which implies that $d(L)$ in (40) is a strictly decreasing function of L , which is minimized for every L if $K = L$, and hence, $M_T = M_R$. This result is somewhat surprising since it says that fixing the number of receive antennas $M_R > M_T$ and increasing M_T leads to a reduction in the per-stream diversity order. There is a physically appealing interpretation of this phenomenon. As we increase the number of transmit antennas (for fixed $M_R > M_T$), we also increase the number of independent data streams that are spatially multiplexed. Thus, the additional degrees of freedom (obtained by increasing the number of transmit antennas) are exploited to increase the throughput

rather than exploiting them to increase the diversity order. Once the number of transmit antennas becomes larger than the number of receive antennas, the number of parallel spatial data pipes that can be opened up is constrained by the number of receive antennas [cf. (12)], and the transmit antennas in excess of this number are used to increase the diversity order experienced by the independent data streams, thus increasing the per-stream diversity order. We have thus exhibited a fundamental tradeoff between multiplexing and diversity gains. We emphasize that our conclusions are a consequence of the very nature of the transmit signal vector's statistics (i.i.d. complex Gaussian), which implies a pure spatial multiplexing mode.

Correlated Case: In Section III, we observed a loss in ergodic capacity due to spatial fading correlation. Interestingly, however, in the presence of spatial fading correlation at the receiver only, assuming that $r(\mathbf{R}) \leq M_T$, in the high-SNR regime, we concluded in (35) that the variance of mutual information decreases compared with that of the i.i.d. channel. This suggests that the per-stream diversity order would increase. Again, this observation has a physically appealing interpretation. High spatial fading correlation amounts to a reduced number of effective receive antennas, thus reducing the effective number of parallel spatial data pipes given by $r(\mathbf{R})$ and leading to a higher per-stream diversity order.

VI. CAPACITY OPTIMAL ANTENNA ALLOCATION

The problem addressed in this section is as follows: Given a MIMO system with M_T transmit and M_R receive antennas, is it better (from the point of view of maximizing ergodic capacity) to allocate an extra antenna, if available, to the transmitter or to the receiver? Besides being a relevant question in the design of point-to-point MIMO wireless links with fixed total number of antennas to be placed on transmit and receive sides, this problem also offers insight into the SNR-dependent impact of array gain and multiplexing gain on ergodic capacity. We restrict ourselves to the case of i.i.d. Rayleigh flat-fading and denote the ergodic capacity of an $M_R \times M_T$ system by $C(M_R, M_T)$. The differential capacity gain obtained by placing an additional antenna at the receiver rather than at the transmitter is hence given by

$$\delta C(r \rightarrow t) = C(M_R + 1, M_T) - C(M_R, M_T + 1).$$

We now examine the behavior of $\delta C(r \rightarrow t)$ in the cases of high and low SNR, respectively.

High-SNR Case: In the following, we will use the high-SNR ergodic capacity approximation (12) for $C(M_R, M_T)$. Our discussion is organized into three different cases.

Case 1)

$$M_T > M_R$$

$$\delta C(r \rightarrow t) = \log_2 \frac{\rho}{M_T} - M_R \log_2 \frac{M_T}{M_T + 1} + \frac{1}{\ln 2} \left(\sum_{p=1}^{M_T - M_R - 1} \frac{1}{p} - \sum_{j=1}^{M_R} \frac{1}{M_T + 1 - j} - \gamma \right) \quad (41)$$

which is positive if

$$\rho > \frac{M_T^{M_T}}{(M_T + 1)^{M_T - 1}} \exp \left(\sum_{p=2}^{M_T} \frac{1}{p} + \gamma \right). \quad (42)$$

Thus, for sufficiently high SNR, placing an additional antenna at the receiver yields higher ergodic capacity than placing an additional antenna at the transmitter. This result is intuitively appealing since it reflects the fact that by adding an extra receive antenna, the spatial multiplexing gain is increased by one (or equivalently one additional spatial data pipe can be opened up). On the other hand, placing the antenna at the transmitter does not improve the spatial multiplexing gain given by $m = L = \min(M_T, M_R)$. For a system with, e.g., $M_T = 5$ and $M_R = 4$, the required SNR to satisfy (42) is 11.91 dB.

Case 2)

$$M_T = M_R$$

$$\delta C(r \rightarrow t) = M_R \log_2 \frac{M_R + 1}{M_R}$$

which is clearly positive for all $M_T = M_R$, indicating that an additional antenna should always be placed at the receiver, irrespectively of the SNR [provided the SNR is high enough for (12) to give a good approximation]. Again, we can give a physically appealing interpretation of this result. While the number of spatial data pipes that can be opened up between transmitter and receiver remains the same whether an antenna is added at the transmitter or the receiver, placing an additional antenna at the receiver is more beneficial due to the resulting array gain. We note that this asymmetry can be attributed to the assumption that the transmitter does not know the channel and is therefore not able to realize array gain.

Case 3)

$$M_T < M_R$$

$$\delta C(r \rightarrow t) = -\log_2 \frac{\rho}{M_T + 1} - M_T \log_2 \frac{M_T}{M_T + 1} + \frac{1}{\ln 2} \left(\sum_{j=1}^{M_T} \frac{1}{M_R + 1 - j} - \sum_{p=1}^{M_R - M_T - 1} \frac{1}{p} + \gamma \right) \quad (43)$$

which is negative if

$$\rho > \frac{M_R^{M_R}}{(M_R - 1)^{M_R - 1}} \exp \left(\sum_{p=2}^{M_R} \frac{1}{p} + \gamma \right). \quad (44)$$

Hence, provided that ρ is sufficiently large, it is better to place an additional antenna at the transmitter rather than at the receiver, despite the fact that array gain can be realized only at the receive side. Adding an additional transmit antenna increases the spatial multiplexing gain by one. For a system with, e.g., $M_T = 4$ and $M_R = 5$, the required SNR to satisfy (44) is 18.95 dB. In Case 1, the required SNR was smaller, which is due to the fact that adding an antenna at the receiver also provides array gain.

We note that using the results presented above, it is easy to verify that in the high-SNR case for a total of $2N$ antennas, a system with N antennas each at the transmitter and receiver (square system) maximizes ergodic capacity. On the other hand, for a total of $2N + 1$ antennas, a system with N antennas at the transmitter and $N + 1$ antennas at the receiver maximizes ergodic capacity. Again, this is due to the receiver's ability to realize array gain. We note that [36] reports similar conclusions for the number of transmit and receive antennas to use in the high-SNR regime based on an asymptotic capacity analysis.

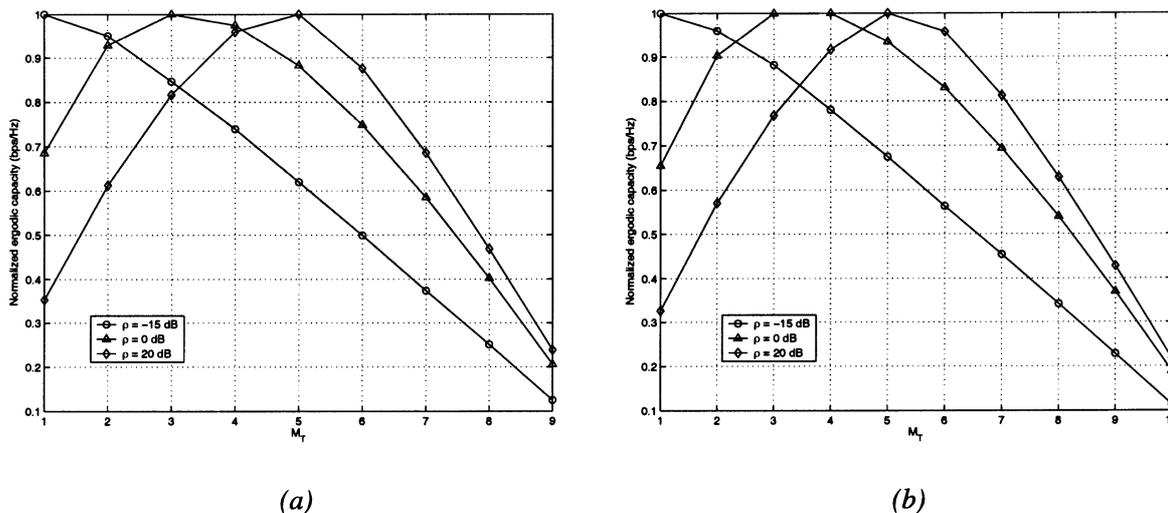


Fig. 8. Variation of empirically determined normalized C as a function of M_T at different values of ρ for a total number of (a) 10 and (b) 11 antennas.

Low-SNR Case: In the low-SNR regime, we employ the lower bound (25), along with the approximation $\log_2(1 + x) \approx (\log_2 e)x$ that is valid for small x to obtain

$$C(M_R, M_T) \approx (\log_2 e) \frac{\rho}{M_T} \exp \left(\sum_{p=1}^{M_T M_R - 1} \frac{1}{p} - \gamma \right).$$

Thus, the differential capacity gain $\delta C(r \rightarrow t)$ is given by

$$\delta C(r \rightarrow t) = \left(\frac{\rho \log_2 e}{M_T} \exp \left(\sum_{p=1}^{M_T(M_R+1)-1} \frac{1}{p} - \gamma \right) - \frac{\rho \log_2 e}{M_T + 1} \exp \left(\sum_{p=1}^{(M_T+1)M_R-1} \frac{1}{p} - \gamma \right) \right). \quad (45)$$

Simple manipulations reveal that (45) is positive for all M_T, M_R values, showing that in the low-SNR regime all available extra antennas should be placed at the receive side to maximize ergodic capacity. This result is intuitive since in the low-SNR regime, there is no spatial multiplexing gain and capacity optimizing configurations must therefore follow an array gain maximization strategy.

Numerical Example 7 (Empirical Selection of Ergodic Capacity Maximizing M_T and M_R): In this example, we corroborate the validity of our analytical results on ergodic capacity maximizing allocation of antennas. In Fig. 8, we plot the empirically determined (through Monte Carlo methods) normalized (such that the highest capacity achieved at a given value of ρ is 1 bps/Hz) ergodic capacity as a function of M_T with total number of transmit and receive antennas fixed at 10 [see Fig. 8(a)] and 11 [see Fig. 8(b)]. We observe that as ρ increases, the optimum (M_T, M_R) point shifts toward higher M_T favoring spatial multiplexing gain over array gain to maximize ergodic capacity and finally stabilizes at $M_T = 5$ (maximum possible spatial multiplexing gain in both cases is 5) in the case of high SNR. These results match our findings for the cases of low SNR (maximize array gain) and high SNR (maximize multiplexing gain). The intermediate SNR values place the optimal (M_T, M_R) operating point somewhere in between the two extremes. We have therefore revealed a tradeoff between array gain and spatial multiplexing gain dictated by the SNR level. \square

VII. CONCLUSIONS

In this paper, we derived tight closed-form lower bounds on ergodic and outage capacities of spatial multiplexing over spatially correlated (frequency-selective) MIMO Rayleigh fading channels. We demonstrated that our bounds are tighter than previously known analytical bounds. Furthermore, we derived an analytical approximation of the variance of mutual information, based on which we provided a new framework for studying the multiplexing-diversity tradeoff for correlated flat-fading MIMO channels when Gaussian code books are used. Finally, for a fixed total number of antennas, we studied ergodic capacity maximizing antenna allocation strategies and revealed the SNR-dependent impact of array gain and multiplexing gain on ergodic capacity.

REFERENCES

- [1] A. J. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception," U. S. Patent 5 345 599, 1994.
- [2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999.
- [3] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
- [4] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.
- [5] H. Bölcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, pp. 225–234, Feb. 2002.
- [6] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.
- [7] E. Gauthier, A. Yongacoglu, and J.-Y. Chouinard, "Capacity of multiple antenna systems in Rayleigh fading channels," *Can. J. Elect. Comput. Eng.*, vol. 25, no. 3, pp. 105–108, July 2000.
- [8] A. Grant, "Rayleigh fading multiple-antenna channels," *EURASIP J. Appl. Signal Process., Special Issue on Space-Time Coding (Part I)*, vol. 2002, pp. 316–329, Mar. 2002.
- [9] P. J. Smith and M. Shafi, "On a Gaussian approximation to the capacity of MIMO wireless systems," in *Proc. IEEE Int. Conf. Commun.*, vol. 1, Apr. 2002, pp. 406–410.
- [10] Z. Wang and G. B. Giannakis, "Outage mutual information of space-time MIMO channels," in *Proc. 40th Allerton Conf.*, Monticello, IL, 2002, pp. 885–894.
- [11] B. Hochwald, T. Marzetta, and V. Tarokh, "Multi-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inform. Theory*, submitted for publication.

- [12] Ö Oyman, R. U. Nabar, H. Bölcskei, and A. J. Paulraj, "Tight lower bounds on the ergodic capacity of Rayleigh fading MIMO channels," in *Proc. IEEE GLOBECOM*, Taipei, Taiwan, R.O.C., Nov. 2002, pp. 1172–1176.
- [13] —, "Characterizing the statistical properties of mutual information in MIMO channels: Insights into diversity-multiplexing tradeoff," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2002, pp. 521–525.
- [14] H. Bölcskei, M. Borgmann, and A. J. Paulraj, "Impact of the propagation environment on the performance of space-frequency coded MIMO-OFDM," *IEEE J. Selected Areas Commun.*, vol. 21, no. 3, pp. 427–439, Apr. 2003.
- [15] H. Bölcskei and A. J. Paulraj, "Performance analysis of space-time codes in correlated Rayleigh fading environments," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2000, pp. 687–693.
- [16] C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inform. Theory*, vol. 48, pp. 637–650, Mar. 2002.
- [17] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge Press, 1985.
- [18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [19] N. R. Goodman, "The distribution of the determinant of a complex Wishart distributed matrix," *Ann. Math. Stat.*, vol. 34, no. 1, pp. 178–180, Mar. 1963.
- [20] P. M. Lee, *Bayesian Statistics: An Introduction*, 2nd ed. London, U.K.: Arnold/Wiley, 1997.
- [21] X. Gourdon and P. Sebah. (2000) The gamma function. [Online]. Available: <http://numbers.computation.free.fr>
- [22] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
- [23] 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, pp. 1423–1436, Oct. 1998.
- [24] A. Narula, M. Trott, and G. Wornell, "Performance limits of coded diversity methods for transmitter antenna arrays," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2418–2433, Nov. 1999.
- [25] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2632–2639, Sept. 2001.
- [26] E. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Commun.*, to be published.
- [27] S. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems with imperfect feedback," *IEEE Trans. Wireless Commun.*, submitted for publication.
- [28] S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 406–417, Apr. 2003.
- [29] D. Shiu, G. J. Foschini, M. J. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502–513, Mar. 2000.
- [30] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd ed. New York: Addison Wesley, 1994.
- [31] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.
- [32] A. Scaglione, "Statistical analysis of the capacity of MIMO frequency selective Rayleigh fading channels with arbitrary number of inputs and outputs," in *Proc. IEEE Int. Symp. Inform. Theory*, Lausanne, Switzerland, 2002, p. 278.
- [33] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073–1096, May 2003.
- [34] R. U. Nabar, H. Bölcskei, and A. J. Paulraj, "Influence of propagation conditions on the outage capacity of space-time block codes," in *Proc. Eur. Wireless Conf.*, Florence, Italy, Feb. 2002, pp. 629–634.
- [35] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.
- [36] A. Lozano and A. M. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Trans. Inform. Theory*, vol. 48, pp. 3117–3128, Dec. 2002.
- Özgür Oyman** (S'02) was born in Istanbul, Turkey, on March 5, 1979. He received the B.S. degree (*summa cum laude*) in electrical engineering from Cornell University, Ithaca, NY, in 2000 and the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 2002. He is currently pursuing the Ph.D. degree in electrical engineering at Stanford University.
- He is a member of the Smart Antennas Research Group, Information Systems Laboratory, Stanford University. He is also working closely with the Communication Theory Group, Swiss Federal Institute of Technology (ETH), Zürich, Switzerland. His research interests include communication and information theory for wireless communications and multi-input multi-output (MIMO) antenna systems.
- Mr. Oyman received the Benchmark Stanford Graduate Fellowship. He is a member of Tau Beta Pi and Eta Kappa Nu.
- Rohit U. Nabar** (S'02) was born in Bombay, India, on December 18, 1976. He received the B.S. degree (*summa cum laude*) in electrical engineering from Cornell University, Ithaca, NY, in 1998 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 2000, and 2003, respectively. His doctoral research focused on signaling for general MIMO channels.
- He is currently a postdoctoral research assistant with the Communication Theory Group, Swiss Federal Institute of Technology (ETH), Zürich, Switzerland. His current research interests include signal processing and information theory for wireless communications, MIMO wireless systems, and *ad hoc* wireless networks.
- Dr. Nabar was a recipient of the Dr. T. J. Rodgers Stanford Graduate Fellowship.
- Helmut Bölcskei** (M'98–SM'02) was born in Austria on May 29, 1970. He received the Dr. techn. degree in electrical engineering from Vienna University of Technology, Vienna, Austria, in 1997.
- In 1998, he was with Vienna University of Technology. From 1999 to 2001, he was a postdoctoral researcher with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA. During that period, he was also a consultant for Iospan Wireless Inc., Palo Alto, CA. From 2001 to 2002, he was an Assistant Professor of electrical engineering at the University of Illinois at Urbana-Champaign. Since February 2002, he has been an assistant professor of communication theory with the Swiss Federal Institute of Technology (ETH), Zürich, Switzerland. He was a visiting researcher at Philips Research Laboratories Eindhoven, The Netherlands; ENST, Paris, France; and the Heinrich Hertz Institute, Berlin, Germany. His research interests include communication and information theory with special emphasis on wireless communications.
- Dr. Bölcskei received a 2001 IEEE Signal Processing Society Young Author Best Paper Award and was an Erwin Schrödinger Fellow of the Austrian National Science Foundation (FWF) from 1999 to 2001. He serves as an associate editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the *EURASIP Journal on Applied Signal Processing*.
- Arogyaswami J. Paulraj** (F'91) received the Ph.D. degree from the Naval Engineering College and the Indian Institute of Technology, New Delhi, India.
- He is 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 nonacademic positions included Head, Sonar Division, Naval Oceanographic Laboratory, Cochin, India; Director, Center for Artificial Intelligence and Robotics, Bangalore, India; Director, Center for Development of Advanced Computing, Pune, India; Chief Scientist, Bharat Electronics, Bangalore, India; Chief Technical Officer and Founder, Iospan Wireless Inc., Palo Alto, CA. His research has spanned several disciplines, emphasizing estimation theory, sensor signal processing, parallel computer architectures/algorithms and space-time wireless communications. His engineering experience included development of sonar systems, massively parallel computers, and more recently broadband wireless systems.
- Dr. Paulraj has won several awards for his engineering and research contributions. He is the author of over 280 research papers and holds eight patents. He is a Member of the Indian National Academy of Engineering.