

Copyright © 2006 IEEE. Reprinted from:

J. Salo, P. Suvikunnas, H. M. El-Sallabi and P. Vainikainen, "Some Insights into MIMO Mutual Information: the High SNR Case," to appear in *IEEE Transactions on Wireless Communications*.

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of Helsinki University of Technology's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org.

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

Some Insights into MIMO Mutual Information: the High SNR Case

Jari Salo, Pasi Suvikunnas, Hassan M. El-Sallabi and Pertti Vainikainen

Abstract—We consider mutual information of Multiple-Input Multiple-Output (MIMO) wireless channels with complex isotropic Gaussian input in the case where the receiver has perfect channel knowledge. For arbitrary fading statistics, a mutual information lower bound is decomposed in a sum of three terms involving: a) average SNR; b) channel fading; c) a term characterizing the “effective rank”, or eigenvalue dispersion, of the channel matrix. The decomposition suggests that spatial multiplexing efficiency of a MIMO channel can be characterized by the so-called *ellipticity statistic*. Distribution functions, means and variances of the random terms in the decomposition for the case of Rayleigh fading are also derived.

Index Terms—MIMO, Rayleigh fading, mutual information.

I. INTRODUCTION

Mutual information (MI) defines the highest achievable rate of information transmission for a given channel input signal and channel realization. Consider a multiple-input multiple-output (MIMO) system where the $n_t \times 1$ channel input $\mathbf{x} \sim \mathcal{CN}(\frac{P}{n_t}, n_t)$, with P denoting the total power of the transmitted signal¹. An $n_r \times n_t$ matrix \mathbf{H} relates the channel input and output via $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where the $n_r \times 1$ noise vector $\mathbf{n} \sim \mathcal{CN}(\sigma^2, n_r)$. Assuming that the receiver knows a given realization of \mathbf{H} perfectly, the MIMO mutual information between \mathbf{x} and \mathbf{y} is [1]

$$I_{\mathbf{H}} = \log_2 \left| \mathbf{I}_K + \frac{\rho}{n_t} \mathbf{W} \right|, \quad (1)$$

where $\rho = \frac{P}{\sigma^2}$ is the average SNR at the output of each of the n_r receiver antennas, and $\mathbf{W} = \mathbf{H}\mathbf{H}^H$ if $n_r \leq n_t$, and $\mathbf{W} = \mathbf{H}^H\mathbf{H}$ otherwise. For \mathbf{H} a random variable, two definitions of mutual information arise: ergodic mutual information and outage mutual information. *Ergodic mutual information* is a meaningful channel measure in the case

This work was supported by the Academy of Finland, National Technology Agency of Finland, Nokia Foundation, Foundation of Commercial and Technical Sciences, TKK Foundation, and Graduate School of Electronics, Telecommunications, and Automation. The authors are with the SMARAD Center of Excellence, Radio Laboratory, Helsinki University of Technology, P.O. Box 3000, FI-02015-TKK, Finland (e-mail: jari.salo@tkk.fi). This work was presented in part at the Nordic Signal Processing Symposium, Espoo, Finland, 2004, and in part at the IEEE Globecom, Dallas, USA, 2004.

¹We denote determinant, trace, Frobenius norm, conjugate transpose, and rank of a matrix \mathbf{A} with $|\mathbf{A}|$, $\text{tr}(\mathbf{A})$, $\|\mathbf{A}\|_F$, \mathbf{A}^H , and $\text{rank}(\mathbf{A})$, respectively. An $n \times n$ identity matrix is denoted with \mathbf{I}_n . Symbols n_r and n_t denote the number of receive and transmit antennas, respectively, with $K = \min(n_r, n_t)$ and $L = \max(n_r, n_t)$. Expected value and variance of a random variable X are denoted with $E[X]$ and $\text{var}[X]$, and the probability of event A with $\Pr(A)$. Complex n -variate Gaussian random variable with a zero mean vector and covariance matrix $a\mathbf{I}_n$, for a some positive constant, is denoted with $\mathcal{CN}(a, n)$, and synonymously called a complex isotropic Gaussian random variable.

where the transmitted codeword spans over a large number (infinite, in theory) of channel realizations. It is defined as $E[I_{\mathbf{H}}]$, with the expectation taken over \mathbf{H} . If the columns of \mathbf{H} are distributed as $\mathcal{CN}(1, n_r)$ (“Rayleigh iid”), the complex Gaussian isotropic input achieves the ergodic channel capacity [1]. Some analytical expressions for the ergodic MI are derived in [1]–[3] among others. In the case where the infinite-length codeword spans over one random channel realization only, the *outage (non-ergodic) mutual information* t is a more suitable measure [4]. For an outage probability p , it is defined as $\Pr(I_{\mathbf{H}} < t) = p$, the evaluation of which requires the knowledge of the distribution of $I_{\mathbf{H}}$. For the Rayleigh fading case, exact and approximate results on the distribution can be found in [5]–[8].

In this paper, instead of new formulas or bounds for MI, we follow an alternative approach, and show that in the high SNR regime the MIMO MI can be, in both ergodic and outage formulations, decomposed as a sum of SNR and (soft)rank-dependent terms, thus shedding some insight into the structure of MI. Assuming that the elements of the channel matrix are iid Rayleigh flat fading, we then analyze the statistics of the individual terms in the decomposition, thereby providing a fairly complete statistical characterization of MIMO mutual information in the high SNR regime, complementing the results in [9].

II. A DECOMPOSED LOWER BOUND FOR MIMO MI

The Grant-Gauthier lower bound for the mutual information (1) is [10], [11]

$$I_{\mathbf{H}} > \log_2 \left| \frac{\rho}{n_t} \mathbf{W} \right|. \quad (2)$$

For a given realization of \mathbf{H} , the bounding error can be shown to be

$$I_{\mathbf{H}} - \log_2 \left| \frac{\rho}{n_t} \mathbf{W} \right| = \sum_{k=1}^K \log_2 \left(1 + \frac{n_t}{\rho \lambda_k} \right). \quad (3)$$

From (3) it is clear that, for fixed n_r and n_t , the bounding error can be made arbitrarily small by selecting ρ large enough, provided that all eigenvalues of \mathbf{W} (denoted $\{\lambda_k\}_{k=1}^K$) are nonzero. For the random channels considered in this paper, we assume that $\Pr[\text{rank}(\mathbf{W}) = K] = 1$. Thus, we can always select SNR “sufficiently” large, so that the bounding error is “sufficiently” small with high probability in the case of outage MI. Similarly, the error in the ergodic MI can be made arbitrarily small at high SNR. To give an idea of the practical applicability range of the Grant-Gauthier bound we

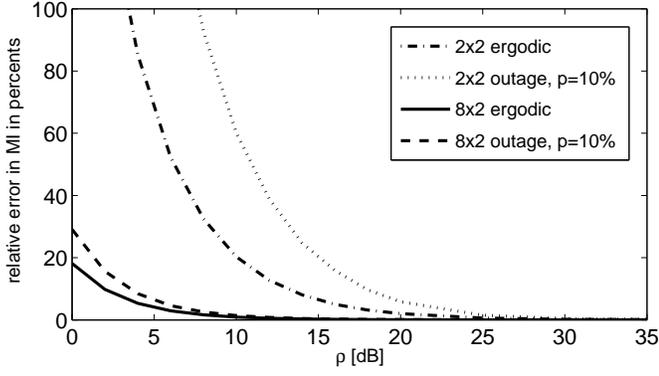


Fig. 1. Relative error of the Grant-Gauthier lower bound (2) for 2×2 and 8×2 MIMO systems ($n_r \times n_t$) over Rayleigh iid fading channel. Error is given for ergodic MI and outage MI ($p = 0.1$).

plot the relative bounding error in Fig. 1 for the Rayleigh iid fading channel. It can be seen that for the 2×2 and 8×2 systems the bound is useful for about $\rho > 20$ dB and $\rho > 5$ dB, respectively. In general, the bound becomes tighter as the ratio $\frac{K}{L}$ becomes smaller, and looser as the number of antennas increases (with $\frac{K}{L}$ held constant). For small outage probabilities (p) the bound may be loose at low SNR, particularly for square MIMO systems ($K = L$).

Definition 1: Consider the set of all channel probability distributions with the average power constraint $E[\|\mathbf{H}\|_F^2] \leq n_r n_t$. For a given ρ the supremum of ergodic mutual information maximized over all such *channel* probability laws is $I_{\text{sup}} = K \log_2(1 + \frac{\rho n_r}{K})$. We call I_{sup} the supremum mutual information. Note that I_{sup} is the mutual information of K parallel decoupled AWGN channels, each having constant power gain $\frac{n_r n_t}{K}$ [12]. In the sequel, we shall refer to this channel as the ‘ K -AWGN channel’, for short.

Consider a sequence of channel matrices, $\{\mathbf{H}^{(i)}\}_{i=1}^{\infty}$, where associated with the i th channel realization $\mathbf{H}^{(i)}$ are $\mathbf{W}^{(i)}$, its eigenvalues $\{\lambda_k^{(i)}\}_{k=1}^K$, geometric mean $m_g^{(i)} = |\mathbf{W}^{(i)}|^{\frac{1}{K}} = (\prod_{k=1}^K \lambda_k^{(i)})^{1/K}$, arithmetic mean $m_a^{(i)} = \frac{1}{K} \text{tr}(\mathbf{W}^{(i)}) = \frac{1}{K} \sum_{k=1}^K \lambda_k^{(i)}$, and the ratio $\gamma^{(i)} = m_g^{(i)}/m_a^{(i)}$. We assume that each member of the sequence is independently generated from a probability distribution that satisfies $E[\|\mathbf{H}^{(i)}\|_F^2] = n_r n_t = KL$, and $\Pr[\text{rank}(\mathbf{W}^{(i)}) = K] = 1$. By noting that $E[m_a^{(i)}] = L$ and $\text{tr}(\mathbf{W}^{(i)}) = \|\mathbf{H}^{(i)}\|_F^2$, we can approximate MI at high SNR as

$$\begin{aligned} I_{\mathbf{H}^{(i)}} &\approx \log_2 \left| \frac{\rho}{n_t} \mathbf{W}^{(i)} \right| \\ &= \log_2 \left[\left(\frac{\rho}{n_t} \frac{n_r n_t}{KL} \right)^K \prod_{k=1}^K \lambda_k^{(i)} \right] \\ &= \log_2 \left[\left(\frac{\rho}{n_t} \frac{n_r n_t}{K} \right)^K \left(\frac{m_a^{(i)}}{L} \right)^K \left(\frac{m_g^{(i)}}{m_a^{(i)}} \right)^K \right] \\ &= \underbrace{K \log_2 \left(\frac{\rho n_r}{K} \right)}_{\approx I_{\text{sup}}} + \underbrace{K \log_2 \left(\frac{\|\mathbf{H}^{(i)}\|_F^2}{E[\|\mathbf{H}^{(i)}\|_F^2]} \right)}_{I_{\text{fad}}^{(i)}} \end{aligned}$$

$$+ \underbrace{K \log_2 \left(\gamma^{(i)} \right)}_{I_{\text{mux}}^{(i)}} \quad (4)$$

$$\approx I_{\text{sup}} + I_{\text{fad}}^{(i)} + I_{\text{mux}}^{(i)}. \quad (5)$$

In the ergodic formulation, $\{\mathbf{H}^{(i)}\}_{i=1}^{\infty}$ should be interpreted as a temporal sequence of matrices, whereas in the non-ergodic one they are simply iid random variables from the channel distribution. Although the statistical analysis itself is the same, the cases differ in the assumptions about the coherence time of the channel. When the transmitted codeword spans over a large number of independent channel realizations, only the ergodic MI makes sense. Thus, one would be interested in the (time) averages $E[I_{\text{fad}}^{(i)}]$ and $E[I_{\text{mux}}^{(i)}]$. Similarly, one should bear in mind that in the case where the codeword spans over one channel realization only, the outage values, and possibly higher order moments, of $I_{\text{fad}}^{(i)}$ and $I_{\text{mux}}^{(i)}$ would be of interest. Some general remarks:

- The first term of (4) is a lower bound for the supremum MI and can be used to approximate I_{sup} at high SNR.
- $I_{\text{fad}}^{(i)}$ is the effect of channel fading on MI. By channel fading we mean the fluctuation of the sum of the $n_r n_t$ channel gains about its mean.
- In the third term, $\gamma^{(i)}$ is a well-known measure of ellipticity of the hyperellipsoid whose axis lengths correspond to the eigenvalues of $\mathbf{W}^{(i)}$ [13, p. 427]. In [14], it is called *ellipticity statistic*. It also arises in the formulation of several model order estimators used in array signal processing [15]. Geometrically, one can interpret MI as the log-volume of the hyperellipsoid, that can be decomposed as the product of a SNR dependent (fading) scaling factor and its ellipticity measured by the scalar $\gamma^{(i)}$.
- The ellipticity statistic provides a natural scale-invariant (wrt ρ) measure for dispersion of the channel eigenvalues. Note that $0 < \gamma^{(i)} \leq 1$; the maximum value is attained if and only if all eigenvalues are equal (i.e. $m_g^{(i)} = m_a^{(i)}$), that is, when there are K parallel equal-gain AWGN channels in operation. Therefore, $I_{\text{mux}}^{(i)}$ is always non-positive and can be interpreted as MI penalty due to eigenvalue dispersion from supremum MI.

In the non-ergodic case, the maximum spatial multiplexing gain afforded by a channel realization is defined in [16] as $\lim_{\rho \rightarrow \infty} \frac{I_{\mathbf{H}^{(i)}}}{\log_2(\rho)}$. From (4) it is clear that the spatial multiplexing gain of any full-rank channel is equal to that of the K -AWGN channel, that is, $\lim_{\rho \rightarrow \infty} \frac{I_{\mathbf{H}^{(i)}}}{\log_2(\rho)} = K$. As a large class of MIMO channels are equivalent in this sense, one can ask how can we classify spatial multiplexing properties of practical MIMO channels, a question that arises frequently in channel measurement studies where rank deficient channels are practically never encountered. The ellipticity statistic is one possible SNR independent measure for the spatial multiplexing capability of a MIMO channel, which, unlike the spatial multiplexing gain, is applicable also in the ergodic case². Its benefit over the condition number, defined as the ratio of

²As mentioned, in the ergodic case, one would compute $E[I_{\text{mux}}^{(i)}]$, or some normalized version of it.

maximum and minimum eigenvalues of \mathbf{W} , is that it depends on all eigenvalues, and also has the interpretation as the rate loss from the K -AWGN channel due to eigenvalue dispersion.

III. STATISTICAL ANALYSIS IN THE RAYLEIGH IID CASE

Throughout this section we assume, with slight simplification of terminology, that “ \mathbf{H} is Rayleigh iid”, i.e., the columns of $\mathbf{H} \sim \mathcal{CN}(1, n_r)$. Then the distribution of \mathbf{W} is complex Wishart with identity correlation matrix [10]. For brevity, we drop the index i ; it should be clear that all quantities depending on \mathbf{H} (i.e., \mathbf{W} , I_{fad} , I_{mux} , m_a , γ) are random variables.

A. Statistical independence of I_{fad} and I_{mux}

Result 1: Let \mathbf{H} be Rayleigh iid. Then I_{mux} and I_{fad} in (5) are statistically independent.

Proof: It can be shown that m_a and γ^K are statistically independent when \mathbf{W} is complex Wishart [17]. The result follows, since I_{fad} depends only on m_a , and I_{mux} depends only on γ^K . ■

Hence, MI is a sum of independent contributions from channel fading and eigenvalue dispersion. It is an open problem, whether the result will hold under more general assumptions on the statistics of \mathbf{H} , or what are its implications, if any, for information optimality of communication systems operating at high SNR.

B. Distributions of I_{fad} and I_{mux}

We report below the distribution functions of I_{fad} and I_{mux} as these may become handy, for example, when their empirical distributions (e.g. from simulations or channel measurements) need to be compared to the Rayleigh iid case. We remind the reader that the distributions of I_{fad} and I_{mux} are relevant quantities only with the outage formulation of MI, where the channel is constant during each transmitted code word (ideally of infinite length).

Result 2: Assume that \mathbf{H} is Rayleigh iid. The distributions of I_{fad} and I_{mux} in (5) are given by

$$f_{I_{\text{fad}}}(w) = \frac{(KL)^{KL} \ln 2}{K \Gamma(KL)} 2^{wL} [\exp(2\frac{w}{K})]^{-KL} \quad (6)$$

$$F_{I_{\text{fad}}}(t) = [\Gamma(KL)]^{-1} \Gamma(KL, KL 2^{\frac{t}{K}}) \quad (7)$$

with $\Gamma(\alpha, t) = \int_0^\alpha x^{t-1} e^{-x} dx$, $\Gamma(t) = \Gamma(\infty, t)$, and

$$f_{I_{\text{mux}}}(z) = A(K, L) 2^{(L-K)z} \times G_{K-1, K-1}^{K-1, 0} \left(2^z \middle| \begin{matrix} a_1, \dots, a_{K-1} \\ 1, \dots, K-1 \end{matrix} \right), \quad z \leq 0, \quad (8)$$

$$F_{I_{\text{mux}}}(t) = \frac{A(K, L)}{\ln 2} 2^{(L-K)t} \times G_{K, K}^{K, 1} \left(2^t \middle| \begin{matrix} K-L+1, a_1, \dots, a_{K-1} \\ 1, \dots, K-1, K-L \end{matrix} \right), \quad t \leq 0, \quad (9)$$

where $G_{p, q}^{m, n} \left(u \middle| \begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix} \right)$ is the Meijer G-function [18], $a_j = K + \frac{j}{K}$, and the normalization factor

$$A(K, L) = \frac{(2\pi)^{\frac{K-1}{2}} \Gamma(LK) \ln 2}{K^{LK} K^{-\frac{1}{2}} \prod_{i=1}^K \Gamma(L+1-i)}.$$

Proof: For proof of (6) and (8), see [17]. The proof of (7) is elementary, while that of (9) requires a fairly straightforward integration of (8). ■

It is worth remarking, that using various properties of the Meijer G-function [19], the pdf (8) and cdf (9) may be expressed using less general functions for the practically important special cases of $K = 2$ and $K = 3$.

The pdf of I_{fad} has been plotted for $K = 2$ and varying L in Fig. 2. It can be seen that as L increases I_{fad} concentrates about the origin. This makes sense intuitively, since increasing the number of antennas at one end of the link should decrease the effect of fading on MI.

In Fig. 3, the pdf of I_{mux} is plotted for the fixed $K = 3$ and varying L . We see how the probability mass concentrates, or channel ‘hardens’, as L increases. The result is quite natural and intuitive, since the eigenvalue dispersion should, of course, decrease as L becomes larger than K since $L^{-1}\mathbf{W}$ tends to identity matrix. We remark that from the results in [20], it follows that, for $L \gg K$, the random variable $-2L \ln 2 I_{\text{mux}}$ is approximately chi squared distributed with $K^2 - 1$ degrees of freedom.

Examining the large- KL behavior of (6) results in the following observation.

Result 3: For large KL , the random variable I_{fad} converges in distribution to a Gaussian random variable with zero mean and variance $\text{var}[I_{\text{fad}}] = \frac{K}{L(\ln 2)^2}$.

Proof: We use the following result from [21, Proposition 6.4.1]. Let X_n be an asymptotically Gaussian random variable with mean m_X and assume that the variance of X is asymptotically zero, i.e. $\lim_{n \rightarrow \infty} \text{var}[X_n] = 0$. If $y = g(x)$ is a continuously differentiable function, then $Y = g(X)$ is also asymptotically Gaussian with mean $g(m_X)$ and variance $\text{var}[X_n] g'(m_X)^2$. Result 3 follows, since it can be shown that $\frac{\|\mathbf{H}\|_F^2}{\mathbb{E}[\|\mathbf{H}\|_F^2]}$ is asymptotically Gaussian with $E\left[\frac{\|\mathbf{H}\|_F^2}{\mathbb{E}[\|\mathbf{H}\|_F^2]}\right] = 1$, $\text{var}\left[\frac{\|\mathbf{H}\|_F^2}{\mathbb{E}[\|\mathbf{H}\|_F^2]}\right] = (KL)^{-1}$ and $g'(x) = \frac{d}{dx} [K \log_2(x)] = \frac{K}{\ln 2} \frac{1}{x}$. The fact that $\frac{\|\mathbf{H}\|_F^2}{\mathbb{E}[\|\mathbf{H}\|_F^2]}$ is asymptotically Gaussian is immediate from the central limit theorem, see e.g. [21, Theorem 6.4.1]. ■

In practice the Gaussian approximation is good already for fairly small number of antennas, e.g. for $KL > 6$. In general, the variance of I_{fad} depends only on the ratio $\frac{K}{L}$, and the maximum variance, $\text{var}[I_{\text{fad}}] = (\log_2 e)^2 \approx 2.08$, is attained for $K = L$. Interestingly, in [22, Eq. (200)] the “amount of fading” is defined for MIMO channels as $\text{AF} = \mathbb{E}[\|\mathbf{H}\|_F^4] / \mathbb{E}[\|\mathbf{H}\|_F^2]^2 - 1$. AF is a general measure of channel fading; a large AF indicates more severe fading. We immediately note the relation $\text{AF} = \text{var}\left[\frac{\|\mathbf{H}\|_F^2}{\mathbb{E}[\|\mathbf{H}\|_F^2]}\right]$. As the proof of Result 3 is based on the central limit theorem, I_{fad} will be asymptotically Gaussian (with $\text{var}[I_{\text{fad}}] = \frac{\text{AF} \cdot K^2}{(\ln 2)^2}$, in

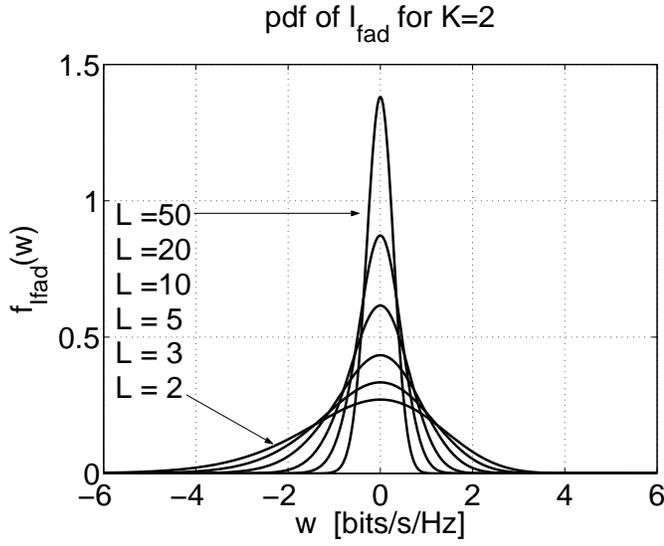


Fig. 2. Probability density functions of I_{fad} for MIMO systems with Rayleigh iid \mathbf{H} with $K = 2$ and $L = \{2, 3, 5, 10, 20, 50\}$.

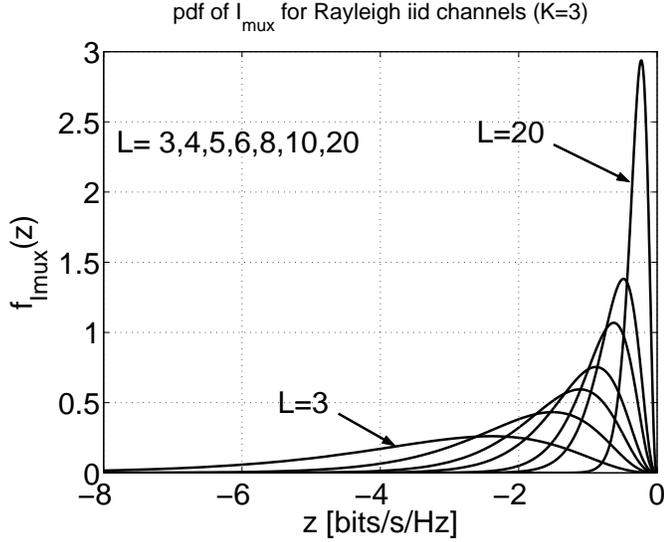


Fig. 3. Probability density functions of I_{mux} for MIMO systems with Rayleigh iid \mathbf{H} , $K = 3$. $L = \{3, 4, 5, 6, 8, 10, 20, 50\}$.

general) for any channel distribution, for which $\frac{\|\mathbf{H}\|_F^2}{\mathbb{E}[\|\mathbf{H}\|_F^2]}$ is asymptotically Gaussian with vanishing variance (or AF).

C. Means and variances of I_{fad} and I_{mux}

Denoting with $\Psi(x) = \frac{d}{dx} \{\ln[\Gamma(x)]\}$ the digamma function and with $\Psi'(x) = \frac{d^2}{dx^2} \{\ln[\Gamma(x)]\}$ the trigamma function, we provide the following result, that complements the analysis in [9].

Result 4: Let \mathbf{H} be Rayleigh iid. Then the means and variances of I_{fad} and I_{mux} in (5) are given by

$$\mathbb{E}[I_{\text{fad}}] = \frac{K}{\ln 2} [\Psi(KL) - \ln(KL)], \quad (10)$$

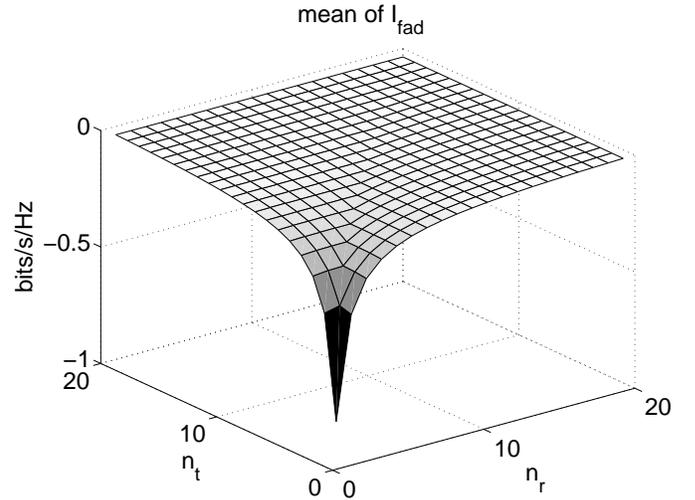


Fig. 4. Mean of mutual information due to channel fading.

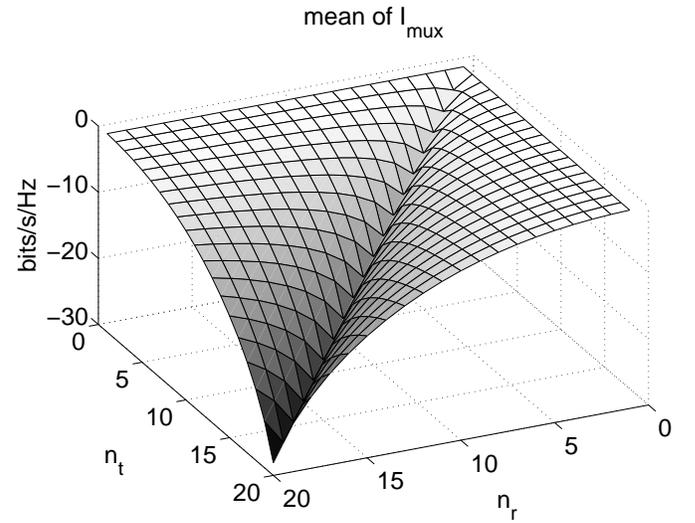


Fig. 5. Mean of mutual information due to eigenvalue dispersion.

$$\mathbb{E}[I_{\text{mux}}] = \frac{K}{\ln 2} \left[\frac{1}{K} \sum_{k=0}^{K-1} \Psi(L-k) - \Psi(KL) + \ln(K) \right], \quad (11)$$

$$\text{var}[I_{\text{fad}}] = \left(\frac{K}{\ln 2} \right)^2 \Psi'(KL), \quad (12)$$

$$\text{var}[I_{\text{mux}}] = \left(\frac{K}{\ln 2} \right)^2 \left[\frac{1}{K^2} \sum_{k=0}^{K-1} \Psi'(L-k) - \Psi'(KL) \right]. \quad (13)$$

Proof: Straightforward by using [10, Lemma A.2] and, with $\text{var}[I_{\text{mux}}]$, Result 1. The basic approach is a modification of the proofs in [9].

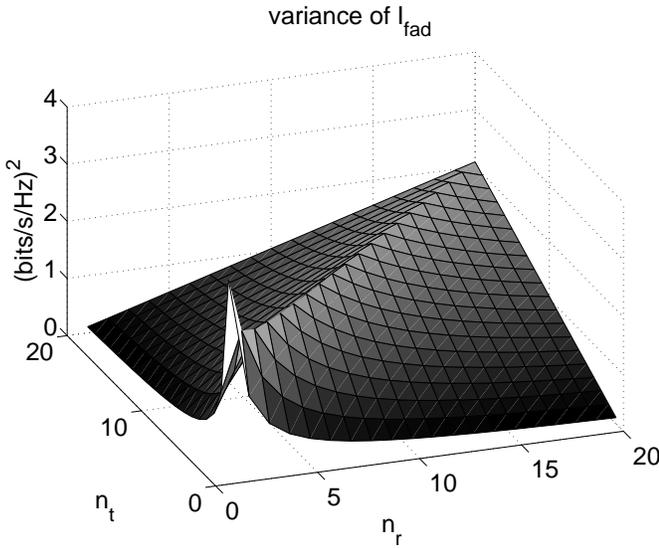


Fig. 6. Variance of mutual information due to channel fading.

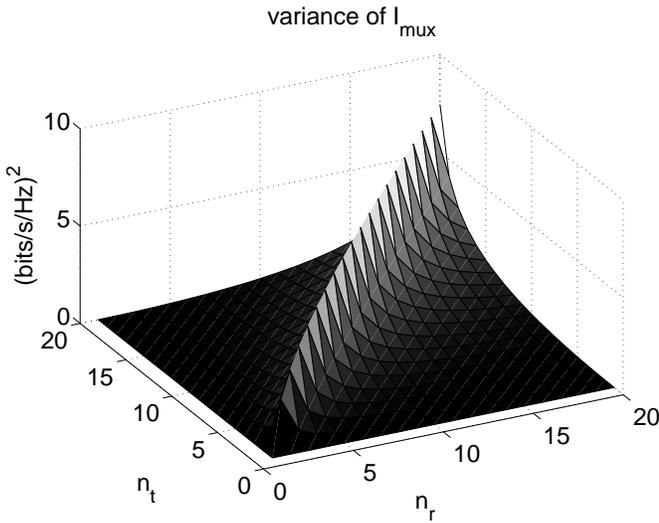


Fig. 7. Variance of mutual information due to eigenvalue dispersion.

The means and variances of I_{fad} and I_{mux} have been plotted in Figs. 4–7 for varying n_r and n_t . We make some observations:

- From Fig. 4 it is obvious that the channel fading has a negligible effect on ergodic mutual information. This can be interpreted as a MIMO equivalent of a general result for single-input single-output channels [4], and is intuitively justified by the fact that by transmitting a codeword over a large number of channel realizations the optimum transceiver should be able to average over channel fading.
- Unique to MIMO systems is the effect of eigenvalue dispersion on the ergodic mutual information. For SIMO/MISO systems, the ergodic mutual information approaches that of the AWGN channel as the number of antennas increases. For MIMO channels, MI does not, however, approach that of the K -AWGN channel,

since there will be a rate loss defined by the average eigenvalue dispersion loss, $E[I_{\text{mux}}]$. However, as I_{mux} does not depend on SNR, the relative rate loss will be negligible at the high-SNR limit, i.e., the ratio of I_{sup} and the ergodic MI, $E[I_{\text{H}}]$, will approach one.

- It can be verified that by summing the means of I_{fad} and I_{mux} we arrive at the lower bound reported in [9, Eq. (12)]. Similarly, adding their variances results in [9, Eq. (31)].
- Comparing Figs. 6 and 7 we note that while for $K = L$ most of the variance of MI is due to eigenvalue dispersion, the situation is reversed as $\frac{K}{L} \rightarrow 0$, where channel fading dominates the variance.
- Due to Result 1, the variance of mutual information is a sum of variances of I_{fad} and I_{mux} . This clarifies the two-fold nature of diversity in MIMO channels. The diversity, defined as the system's resilience towards signal fading, comes in two forms in MIMO systems. First, as reduced channel fading and, second, as reduced eigenvalue dispersion. The latter form is unique to MIMO systems.

IV. CONCLUSION

We showed that, at high SNR, mutual information can be approximated as a sum of terms incorporating the effects of average SNR, channel fading, and eigenvalue dispersion. The decomposition is independent of channel distribution and illustrates the two-fold effect of diversity in MIMO systems; the resilience towards fading of received power, and the decrease in the spread of eigenvalues captured by the so-called ellipticity statistic from multivariate analysis. We derived means, variances and distribution functions for the random terms in the mutual information decomposition under Rayleigh iid fading channel statistics.

ACKNOWLEDGMENT

Critical comments from Kalle Ruttik are gratefully acknowledged. We also thank the associate editor and the anonymous reviewers whose comments significantly improved the presentation of the paper.

REFERENCES

- [1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–596, Nov. 1999.
- [2] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: spatial fading correlation, double scattering, and keyhole," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2636–2647, Oct. 2003.
- [3] M. Kießling, "Unifying analysis of ergodic MIMO capacity in correlated Rayleigh fading environments," *European Transactions on Telecommunications*, vol. 16, no. 1, pp. 17–35, Jan. 2005.
- [4] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 2003.
- [5] P. J. Smith and M. Shafi, "On a Gaussian approximation to the capacity of wireless MIMO systems," in *Proc. IEEE Intern. Conf. Comm.*, New York, USA, Apr. 2002, pp. 406–410.
- [6] M. Chiani, M. Z. Win, and A. Zanella, "On the capacity of spatially correlated MIMO Rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2363–2371, Oct. 2003.

- [7] P. J. Smith, S. Roy, and M. Shafi, "Capacity of MIMO systems with semicorrelated flat fading," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2781–2788, Oct. 2003.
- [8] P. J. Smith, L. M. Garth, and S. Loyka, "Exact capacity distributions for MIMO systems with small numbers of antennas," *IEEE Commun. Lett.*, vol. 7, no. 10, pp. 481–483, Oct. 2003.
- [9] Ö. Oyman, R. U. Nabar, H. Bölcskei, and A. J. Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels," *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 2784–2795, Nov. 2003.
- [10] A. Grant, "Rayleigh fading multiple-antenna channels," *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 3, pp. 316–329, Mar. 2002.
- [11] E. Gauthier, A. Yongacoglu, and J.-Y. Chouinard, "Capacity of multiple antenna systems in Rayleigh fading channels," in *Proc. Canadian Conference on Electrical and Computer Engineering*, vol. 1, Halifax, NS, Canada, Jul. 2000, pp. 275–279.
- [12] H. Xu, M. Gans, N. Amitay, R. A. Valenzuela, T. Sizer, R. Storz, D. Taylor, M. McDonald, and C. Tran, "MIMO channel capacity for fixed wireless: Measurements and models," in *Proc. IEEE Veh. Tech. Conf. (VTC-Fall)*, vol. 2, Atlantic City, NJ, USA, Oct. 2001, pp. 1068–1072.
- [13] T. W. Anderson, *Introduction to Multivariate Statistical Analysis*, 2nd ed. John Wiley Inc, 1984.
- [14] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. John Wiley and Sons, Inc., 1982.
- [15] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, no. 2, pp. 387–392, Apr. 1985.
- [16] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
- [17] J. Salo, P. Suvikunnas, H. M. El-Sallabi, and P. Vainikainen, "Some results on MIMO mutual information: the high SNR case," in *Proc. IEEE Globecom'04*, vol. 2, Dallas, USA, Nov./Dec. 2004, pp. 943–947.
- [18] Y. L. Luke, *Special functions and their approximations*, ser. Mathematics in Science and Engineering. Academic Press, 1969, vol. 1.
- [19] "The Wolfram Functions Site," <http://functions.wolfram.com/>.
- [20] B. Nagarsenker and P. Nagarsenker, "Distribution of the likelihood ratio statistic for testing sphericity structure for a complex normal covariance matrix," *Sankhya B: The Indian Journal of Statistics*, vol. 48, no. 3, pp. 352–359, 1981.
- [21] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. Springer-Verlag, Inc., 1991.
- [22] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.