On the characteristics of MIMO mutual information at high SNR

J. Salo, P. Suvikunnas, H.M. El-Sallabi and P. Vainikainen
Helsinki University of Technology/SMARAD, Radio Laboratory
Email: jari.salo@hut.fi

Abstract—We consider a Multiple-Input Multiple-Output (MIMO) communication system where the transmitted vector has a Gaussian distribution with (scaled) identity correlation matrix; this is the capacity-achieving channel input distribution for a block fading Rayleigh iid MIMO channel when the transmitter has no channel knowledge and the receiver knows the channel perfectly. We decompose a high-SNR approximation of the mutual information into three terms involving: i) the effect of eigenvalue dispersion of the channel; ii) the effect of fluctuation of SNR about its mean; iii) the effect of eigenvalue dispersion of the channel. The decomposition provides some insight on the mechanisms that affect the MIMO mutual information at high SNR. Further, we analyze the first and second order statistics of the terms in the decomposition under the assumption of frequency-flat Rayleigh iid fading.

I. INTRODUCTION

The enormous capacity of MIMO communication channels has spurred an avalanche of research on different aspects of multi-antenna radio communications [1] [2]. Some of the recent research has concentrated on analysis of first order statistics of MIMO mutual information (i.e., ergodic capacity) under Rayleigh fading channel statistics, see e.g. [3]–[8].

Our focus here is on the simple capacity lower bound analyzed by Grant and Gauthier et al [4] [7]. Our purpose is to extend the analysis therein to the SNR and eigenvalue dispersion dependent components of the mutual information.

The paper is organized as follows. In Section II we introduce some basic concepts and rewrite the Grant-Gauthier lower bound as a sum of three terms that form the mutual information “decomposition”. In Section III we compute the means and variances of the terms of the decomposition in the Rayleigh iid fading case. Some illustrative graphs are provided to accompany the analytical results. Section IV concludes the paper.

II. A DECOMPOSED LOWER BOUND FOR MIMO CAPACITY

A. Basic assumptions and definitions

We consider a MIMO system with \( n_t \) transmit and \( n_r \) receive antennas. Assume that the transmitter signal distribution is complex Gaussian with correlation matrix \( \frac{P}{n_r} I_{n_r} \), where \( P \) is the total transmitted signal power. Then, for a given \( n_r \times n_t \) channel matrix \( H \), the MIMO mutual information is [1], [2]

\[
C_H = \log_2 \left( \frac{P}{n_t} \| HH^H \| \right)
\]

where \( \rho \) is the average SNR at the output of each of the \( n_r \) receiver sensors. Complex additive white Gaussian noise, iid in spatial and time domains, is assumed throughout the paper.

Assuming that the receiver knows the fixed, nonrandom \( H \) perfectly and that the transmitter has no knowledge of it, \( C_H \) is the maximum of mutual information, i.e. the channel capacity associated with \( H \).

B. Decomposed Grant-Gauthier lower bound

Motivated by \( |1 + AB| = |1 + BA| \), and the inequality \( |1 + A| > |A| \) (for positive definite \( A \)), we define

\[
W = \begin{cases} 
HH^H, & \text{if } n_r \leq n_t \\
H^H H, & \text{if } n_r > n_t 
\end{cases}
\]

and consequently lower bound (1) as \( C_H > \log_2 \| \frac{\rho}{n_t} W \| \). This is the Grant-Gauthier lower bound [4] [7].

Note that the tighter Minkowski determinant inequality [10, p. 482] could also be applied here, as in [3]. However, at the high SNR region, this would lead us to the same end result in our upcoming development.

Consider an ergodic sequence of channel matrices, \( \{H(i), i = 1, 2, \ldots \} \). This is the block fading model. Denote \( K = \min(n_t, n_r) \). We adopt the usual power normalization \( E[\|H(i)\|^2_F] = E[\sum_{k=1}^K \lambda_k(i)] = n_t n_r \) and \( \{\lambda_k(i), k = 1, \ldots, K\} \) are the eigenvalues1 of \( W(i) \).

Denote \( m_a(i) = \frac{1}{K} \sum_{k=1}^K \lambda_k(i) \) and \( m_a = E[m_a(i)] \). The lower bound may be written as

\[
C_H(i) > \log_2 \left( \frac{\rho}{n_t} W(i) \right) = \log_2 \left( \frac{\rho}{n_t} \frac{n_r n_t}{K m_a} \prod_{k=1}^K \lambda_k(i) \right) = \log_2 \left( \frac{\rho}{n_t} \frac{n_r n_t}{K m_a} \left( \frac{m_a(i)}{m_a} \right)^K \right)
\]

1We neglect the pathological cases (e.g. keyhole) where \( \text{rank}(W(i)) < K \), or, rather, assume that probability of such an event is vanishingly small.
The decomposition (2) admits an intuitively appealing interpretation. The first term of (2) is a lower bound of supremum power gains that actually replace \( \rho_n \) information down to about 10% outage levels for the lower bound is very close to that of the true mutual Rayleigh iid channel matrices. The empirical distribution of \( K \) processing [12]. It provides a natural scale-invariant (wrt \( \rho \)) term of several model order estimators used in array signal processing [12]. It provides a natural scale-invariant (wrt \( \rho \)) term of several model order estimators used in array signal processing [12]. Interestingly, the same parameter also arises in the likelihood
decomposition (2) admits an intuitively appealing interpretation. The first term of (2) is a lower bound of supremum power gains that actually replace \( \rho_n \) information down to about 10% outage levels for the lower bound is very close to that of the true mutual Rayleigh iid channel matrices. The empirical distribution of \( K \) processing [12]. It provides a natural scale-invariant (wrt \( \rho \)) term of several model order estimators used in array signal processing [12]. Interestingly, the same parameter also arises in the likelihood

\[ C_{\text{sup}} > C_{\text{sup}} + C_{\text{fad}} + C_{\text{mux}}. \]

We stress that in (3) we have separated the effects of SNR (both average and instantaneous) and eigenvalue dispersion on the mutual information. The only statistical assumption made so far is the average power gain constraint on \( H \).

Before proving the result we need the following theorem.

**Theorem 1:** Let \( W \sim \text{CW}(K, L) \), and \( m_a \) be the geometric mean of eigenvalues of \( W \). Then \( \gamma^K = \left( \frac{m_a}{m_a} \right)^K \) and \( m_a = \frac{1}{K} \text{tr}(W) \) are independent.

**Proof:** The case of real-valued Wishart matrices has been considered in [14] and [15]. The proof for the complex Wishart pdf is similar, and due to space limitations we omit details [16].

Using Theorem 1 the proof of Result 1 is easy.
Proof: (Result 1) Note that \( C_{fad} = K \log_2(m_a) - K \log_2(L) \) and \( C_{mux} = K \log_2(\gamma) \), respectively, are functions of \( m_a \) and \( \gamma^K \) only. Therefore, by Theorem 1, they are statistically independent.

We note that if Result 1 was not true, the decomposition (2) would be a sum of dependent terms, and as such, would be a less descriptive characterization of mutual information under Rayleigh iid fading.

B. Means and variances of \( C_{fad} \) and \( C_{mux} \)

Our main result is the following.

Result 2: Let \( \mathbf{H}^{(i)} \) be Rayleigh iid. Then the means and variances of \( C_{fad}^{(i)} \) and \( C_{mux}^{(i)} \) in (3) are given by

\[
E[C_{fad}^{(i)}] = \frac{K}{\ln 2} \left[ \Psi(KL) - \ln(KL) \right],
\]

\[
E[C_{mux}^{(i)}] = \frac{K}{\ln 2} \left[ \frac{1}{K} \sum_{k=0}^{K-1} \Psi(L - k) - \Psi(KL) + \ln(K) \right],
\]

\[
\text{var}[C_{fad}^{(i)}] = \left( \frac{K}{\ln 2} \right)^2 \Psi'(KL),
\]

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

where \( \Psi(x) = \frac{d}{dx} \left[ \ln[\Gamma(x)] \right] \) is the digamma function and \( \Psi'(x) = \frac{d^2}{dx^2} \ln[\Gamma(x)] \) is the trigamma function.

Proof: We drop the index \( i \) for convenience of notation. We use the following results from [4]:

\[
E[\ln |\mathbf{W}|] = \sum_{k=0}^{K-1} \Psi(L - k),
\]

\[
\text{var}[\ln |\mathbf{W}|] = \sum_{k=0}^{K-1} \Psi'(L - k).
\]

\[\text{Eq. (4): In the Rayleigh iid case the distribution of } \|\mathbf{H}\|_F^2 \text{ is complex Wishart distribution with parameters } K = 1 \text{ and } L = n_t n_r. \]

\[\text{Hence, by (8),} \]

\[E[\ln \|\mathbf{H}\|_F^2] = \Psi(n_t n_r),\]

\[\text{and (4) follows.} \]

\[\text{Eq. (5): Since the geometric mean of the eigenvalues of } \mathbf{W} \text{ is } m_g = |\mathbf{W}|^{-\frac{1}{2}} \text{ and the arithmetic mean } m_a = \frac{1}{K} \|\mathbf{H}\|_F^2, \text{ we can write} \]

\[E[C_{mux}] = E\left[K \log_2 \left( \frac{m_g}{m_a} \right) \right] = \frac{K}{\ln 2} \left( \frac{1}{K} E[\ln |\mathbf{W}|] - E[\ln \|\mathbf{H}\|_F^2] + \ln K \right).\]

The first expectation is given directly by (8) and the second was given in (10).

\[\text{Eq. (6): Since the latter term in } C_{fad} = K \log_2 \left( \|\mathbf{H}\|_F^2 \right) - K \log_2(n_t n_r) \text{ is constant, the variance of } C_{fad} \text{ is obtained from (9) as} \]

\[\text{var} \left[ \frac{K}{\ln 2} \ln \|\mathbf{H}\|_F^2 \right] = \left( \frac{K}{\ln 2} \right)^2 \Psi'(KL). \]

\[\text{Eq. (7): Note that the variance of the whole lower bound (2) must be equal to the variance of } \log_2 \left( \frac{n_r}{n_t} \mathbf{W} \right) \text{ which can be obtained directly by using (9). Since from Result 1 it follows that } C_{fad} \text{ and } C_{mux} \text{ are independent, and hence uncorrelated, we must have} \]

\[\text{var} [C_{mux}] = \text{var} \left[ \log_2 \left( \frac{n_r}{n_t} \mathbf{W} \right) \right] - \text{var}[C_{fad}] \]

\[\text{and, after some manipulations, (7) results.} \]

The means and variances in (4)-(7) have been plotted in Figs. 2-5 as a function of \( n_t \) and \( n_r \). We make some observations:

- When the transmitter has no channel information and the receiver knows \( \mathbf{H} \) perfectly the ergodic capacity is defined as \( E[C_{\mu}] \) [1]. In Result 2 we have decomposed the effects of SNR fading and eigenvalue dispersion in the ergodic capacity. From Fig. 2 we note that the mean value of \( C_{fad} \) is practically zero for even moderate \( KL \). Comparing to Fig. 3 confirms that the average eigenvalue dispersion practically determines the ergodic capacity.

- We note that the capacity degradation due to eigenvalue dispersion is largest when \( K = L \). This is natural since the eigenvalues are most dispersed in this case. However, already by increasing \( L \) from \( K \) to \( K + 1 \) results in considerable decrease in average eigenvalue dispersion.

- In general, the variance of \( C_{mux} \) increases with \( K \) extremely slowly for \( K \) large. However, for fixed \( K \), the variance decreases very sharply as \( L \) is made larger than \( K \). We also note that the variances of \( C_{mux} \) and \( C_{fad} \) are of the same order for small values of \( K \) and \( L \).

We emphasize that while many of the results just discussed may appear intuitively clear, we have provided a formal quantitative analysis of the issue at hand. In other words, we have separated the effects of SNR and eigenvalue dispersion on ergodic channel capacity and determines the degradation caused by each of them, in bits/s/Hz.

IV. CONCLUSION

We have decomposed the mutual information at high SNR as a sum of three terms that were interpreted as supremum capacity, “fading gain”, and spatial multiplexing degradation. We proved that the effects of SNR fading and eigenvalue dispersion (parameterized by the ratio of geometric and arithmetic means of channel eigenvalues) on mutual information are statistically independent under Rayleigh iid fading at the high
SNR regime. Further, for the Rayleigh iid channel statistics we computed the first and second order statistics of the terms in the decomposition and provided some insights into the characteristics of ergodic capacity and mutual information.

ACKNOWLEDGMENT

The authors would like to thank Nokia Research Center and Academy of Finland for supporting this research. The first and second author would also like to thank the Foundation of Commercial and Technical Sciences and the Nokia foundation for financial support.

REFERENCES