Proceedings of the 6th Nordic Signal Processing Symposium - NORSIG 2004
June 9 - 11, 2004
Espoo, Finland

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 as a sum of three terms involving: i) a supremum capacity; 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 not to derive new bounds but instead to decompose the Grant-Gauthier lower bound into a form that provides us with some insight into the structure of the mutual information at high SNR. Our results are closely related to those in [9], where the statistics of mutual information were characterized in terms of first and second order statistics. The contribution of this paper 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 $\frac{P}{n_t}\mathbf{I}_{n_t}$, where P is the total transmitted signal power. Then, for a given $n_r \times n_t$ channel matrix $\mathbf{H}$, the MIMO mutual information is [1], [2]

$$C_{\mathbf{H}} = \log_2 \left| \mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H \right|, \quad (1)$$

where ρ 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 $\mathbf{H}$ perfectly and that the transmitter has no knowledge of it, $C_{\mathbf{H}}$ is the maximum of mutual information, i.e. the channel capacity associated with $\mathbf{H}$.

B. Decomposed Grant-Gauthier lower bound

Motivated by $|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}|$, and the inequality $|\mathbf{I} + \mathbf{A}| > |\mathbf{A}|$ (for positive definite $\mathbf{A}$), we define

$$\mathbf{W} = \begin{cases} \mathbf{H}\mathbf{H}^H , & \text{if } n_r \le n_t \\ \mathbf{H}^H\mathbf{H} , & \text{if } n_r > n_t \, . \end{cases}$$

and consequently lower bound (1) as $C_{\mathbf{H}} > \log_2 |\frac{\rho}{n_t}\mathbf{W}|$. 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, $\{\mathbf{H}^{(i)}, i = 1, 2, \dots\}$. This is the block fading model. Denote $K = \min(n_r, n_t)$. We adopt the usual power normalization $E\big[\|\mathbf{H}^{(i)}\|_F^2\big] = E\big[\sum_{k=1}^{K} \lambda_k^{(i)}\big] = n_r n_t$, and $\{\lambda_k^{(i)}, k = 1, \dots, K\}$ are the eigenvalues[1] of $\mathbf{W}^{(i)}$.

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

$$\begin{aligned} C_{\mathbf{H}^{(i)}} &> \log_2 \left| \frac{\rho}{n_t}\mathbf{W}^{(i)} \right| \\ &= \log_2 \left[\left(\frac{\rho}{n_t} \frac{n_r n_t}{K m_a} \right)^K \prod_{k=1}^{K} \lambda_k^{(i)} \right] \\ &= \log_2 \left[\left(\frac{\rho}{n_t} \frac{n_r n_t}{K} \right)^K \left(\frac{m_a^{(i)}}{m_a} \right)^K \left(\frac{m_g^{(i)}}{m_a^{(i)}} \right)^K \right] \end{aligned}$$

[1] We neglect the pathological cases (e.g. keyhole) where $\mathrm{rank}[\mathbf{W}^{(i)}] < K$, or, rather, assume that probability of such an event is vanishingly small.

$$= K\log_2\left(\frac{\rho n_r}{K}\right) + K\log_2\left(\frac{m_a^{(i)}}{m_a}\right) + K\log_2(\gamma^{(i)}), \qquad (2)$$

where $m_g^{(i)} = \left(\prod_{k=1}^{K}\lambda_k^{(i)}\right)^{1/K}$ is the geometric mean of the eigenvalues of $\mathbf{W}^{(i)}$ and $\gamma^{(i)} = m_g^{(i)}/m_a^{(i)}$ is the ratio of their geometric and arithmetic means, which is a well-known measure of sphericity in multivariate statistic [11, p. 427]. Interestingly, the same parameter also arises in the likelihood term of several model order estimators used in array signal processing [12]. It provides a natural scale-invariant (wrt ρ) measure for the dispersion of the channel eigenvalues. Note that $0 \le \gamma^{(i)} \le 1$, where the maximum value is attained when all eigenvalues are equal, i.e. the case of K parallel AWGN channels.

The decomposition (2) admits an intuitively appealing interpretation. The first term of (2) is a lower bound of supremum capacity[2] for the set of channel matrices $S = \{\mathbf{H} : \|\mathbf{H}\|_F^2 \le n_t n_r\}$. Here we define $C_{sup} = \sup_{\mathbf{H}\in S} C_{\mathbf{H}}$. The second term is the effect of instantaneous SNR normalized by its mean. The third term measures the instantaneous capacity degradation from the supremum due to eigenvalue dispersion (note that this term is always non-positive); this may be interpreted as the efficiency of spatial multiplexing. Therefore, we interpret the mutual information as a sum of three terms and write (2) as

$$C_{\mathbf{H}^{(i)}} > C_{sup} + C_{fad}^{(i)} + C_{mux}^{(i)}. \qquad (3)$$

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 $\mathbf{H}$.

Note that the "fading gain" term may also be written as

$$C_{fad}^{(i)} = K\log_2\left(\frac{\|\mathbf{H}^{(i)}\|_F^2}{E\left[\|\mathbf{H}^{(i)}\|_F^2\right]}\right),$$

which perhaps better illustrates its role as the instantaneous capacity gain (or loss) due to instantaneous SNR fading defined as the sum of the power gains of the $n_t n_r$ paths.

Let $Y = \log_2|\frac{\rho}{n_t}\mathbf{W}|$. For ρ/n_t large, the lower bound (2) is tight in the sense that, for fixed n_r and n_t, $\lim_{\rho\to\infty}(C_{\mathbf{H}} - Y) = 0$ for any positive definite $\mathbf{W}$. Hence by selecting ρ large enough we can make (3) an accurate approximation of the mutual information. It can be shown that the smallest eigenvalue dominates the error of the bound and hence the bound is not tight if one or more of the eigenvalues are in the order of $\frac{n_t}{\rho}$ or less.

To give an example, we consider an iid sequence of Rayleigh iid channel matrices. The empirical distribution of the lower bound is very close to that of the true mutual information down to about 10% outage levels for $\rho = 20$ dB and $n_r = n_t = 2$, see Fig. 1. When $C_{fad}^{(i)}$ happens to be large and $C_{mux}^{(i)}$ is not too small, $C_{\mathbf{H}^{(i)}}$ can actually exceed C_{sup}. However, it holds that $E[C_{\mathbf{H}^{(i)}}] < C_{sup}$, which can be seen from Jensen's inequality and the concavity of $\log_2|\mathbf{A}|$ for positive definite $\mathbf{A}$. At high SNR, the decomposition (3) is a good approximation for $C_{\mathbf{H}^{(i)}}$ and can be used to examine the mechanisms that affect the mutual information in more detail.

[2]The exact supremum capacity for the given power constraint is $C_{sup} = K\log_2(1+\frac{\rho n_r}{K})$, which is the capacity of K parallel AWGN channels with power gains $\frac{n_r n_t}{K}$. For ρ large, $C_{sup} \approx K\log_2(\frac{\rho n_r}{K})$. Note that we could actually replace sup with max.

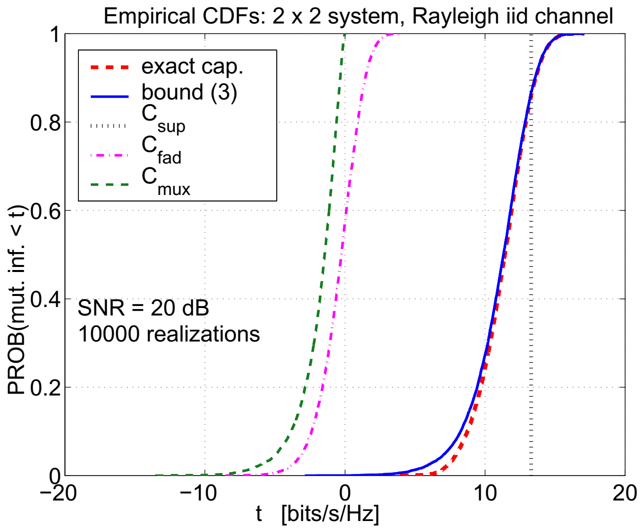

Fig. 1. Empirical cdfs of (3) and its individual terms.

III. Statistical analysis in the Rayleigh iid case

In this section we analyze the case of Rayleigh iid $\mathbf{H}$, i.e. we assume the elements of the $n_r \times n_t$ matrix $\mathbf{H}$ are independently distributed according to the complex, zero-mean, circularly symmetric Gaussian distribution with unit variance. The distribution of $\mathbf{W}$ is then complex Wishart with identity covariance matrix, denoted here as $\mathbf{W} \sim \mathcal{CW}(K, L)$, where $L = \max(n_r, n_t)$. A collection of key results related to the complex Wishart distribution can be found in [4, Appendix] and in [13].

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

For Rayleigh iid fading we have the following interesting result.

Result 1: Let $\mathbf{H}$ be Rayleigh iid. Then $C_{mux} = K\log_2\gamma$ and $C_{fad} = K\log_2\left(\|\mathbf{H}\|_F^2\right) - K\log_2(KL)$ in (3) are statistically independent.

Before proving the result we need the following theorem.

Theorem 1: Let $\mathbf{W} \sim \mathcal{CW}(K, L)$, and m_g be the geometric mean of eigenvalues of $\mathbf{W}$. Then $\gamma^K = \left(\frac{m_g}{m_a}\right)^K$ and $m_a = \frac{1}{K}\operatorname{tr}(\mathbf{W})$ 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 γ^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

$$\begin{aligned}
E[C_{fad}^{(i)}] &= \frac{K}{\ln 2}\big[\Psi(KL) - \ln(KL)\big], && (4)\\
E[C_{mux}^{(i)}] &= \frac{K}{\ln 2}\Bigg[\frac{1}{K}\sum_{k=0}^{K-1}\Psi(L-k) \\
&\qquad -\Psi(KL) + \ln(K)\Bigg], && (5)\\
\mathrm{var}[C_{fad}^{(i)}] &= \left(\frac{K}{\ln 2}\right)^2 \Psi'(KL), && (6)\\
\mathrm{var}[C_{mux}^{(i)}] &= \left(\frac{K}{\ln 2}\right)^2 \Bigg[\frac{1}{K^2}\sum_{k=0}^{K-1}\Psi'(L-k) \\
&\qquad -\Psi'(KL)\Bigg]. && (7)
\end{aligned}$$

where $\Psi(x) = \frac{d}{dx}\{\ln[\Gamma(x)]\}$ 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]:

$$\begin{aligned}
E\big[\ln|\mathbf{W}|\big] &= \sum_{k=0}^{K-1}\Psi(L-k), && (8)\\
\mathrm{var}\big[\ln|\mathbf{W}|\big] &= \sum_{k=0}^{K-1}\Psi'(L-k)\,. && (9)
\end{aligned}$$

Eq. (4)*:* In the Rayleigh iid case the distribution of $\|\mathbf{H}\|_F^2$ is complex Wishart distribution with parameters $K = 1$ and $L = n_r n_t$. Hence, by (8),

$$E\big[\ln\|\mathbf{H}\|_F^2\big] = \Psi(n_r n_t), \tag{10}$$

and (4) follows.

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

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

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

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

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

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

$$\mathrm{var}\left[C_{mux}\right] = \mathrm{var}\left[\log_2\left|\frac{\rho}{n_t}\mathbf{W}\right|\right] - \mathrm{var}[C_{fad}]$$

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 $E[C_{\mathbf{H}}]$ [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.
- From Fig. 3 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

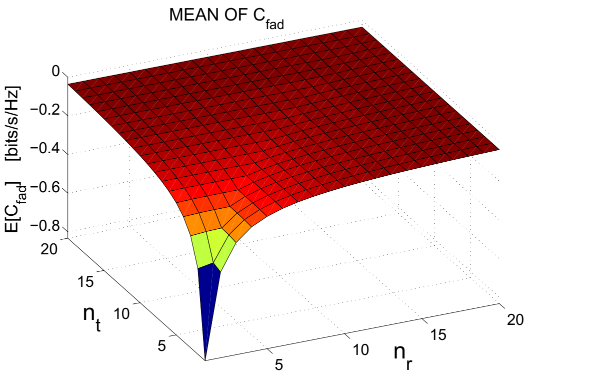

Fig. 2. Mean of the instantaneous “fading gain” under Rayleigh iid fading.

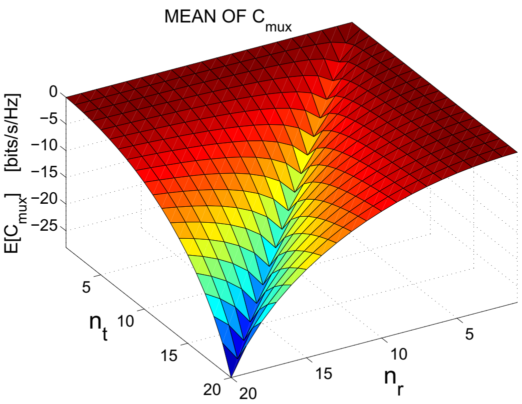

Fig. 3. Mean of C_{mux} under Rayleigh iid fading.

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

[1] E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–596, November 1999.

[2] G. J. Foschini and M. J. Gans, “On the limits of wireless communications in a fading environment when using multiple antennas,” *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, March 1998.

[3] Ö. Oyman, R. Nabar, H. Bölcskei, and A. Paulraj, “Tight lower bounds on the ergodic capacity of Rayleigh fading MIMO channels,” in *Proc. of GLOBECOM’02*, vol. 2, 2002, pp. 1172 –1176.

[4] A. Grant, “Rayleigh fading multiple-antenna channels,” *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 3, pp. 316–329, Mar. 2002.

[5] S. Loyka and A. Kouki, “New compound upper bound on MIMO channel capacity,” *IEEE Commun. Lett.*, vol. 6, no. 3, pp. 96–98, March 2002.

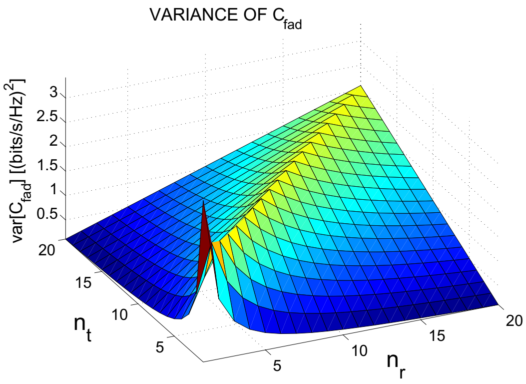

Fig. 4. Variance of instantaneous “fading gain” under Rayleigh iid fading.

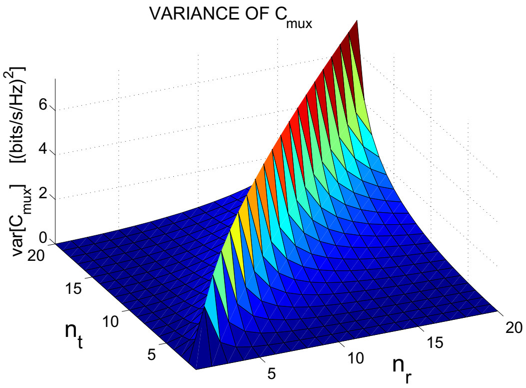

Fig. 5. Variance of C_{mux} under Rayleigh iid fading.

[6] M. Kiessling, J. Speidel, I. Viering, and M. Reinhardt, “A closed-form bound on correlated MIMO channel capacity,” in *Proc. of VTC-2002-Fall*, vol. 2, 2002, pp. 24–28.

[7] 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.

[8] 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.

[9] Ö. Oyman, R. Nabar, H. Bölcskei, and A. 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] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[11] T. Anderson, *Introduction to Multivariate Statistical Analysis*, 2nd ed. John Wiley Inc, 1984.

[12] 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, April 1985.

[13] T. Ratnarajah, R. Vaillancourt, and M. Alvo, “Complex random matrices and Rayleigh channel capacity,” *Communications in Information and Systems*, vol. 3, no. 2, pp. 119–138, Oct. 2003.

[14] R. Muirhead, *Aspects of Multivariate Statistical Theory*. John Wiley and Sons, Inc., 1982.

[15] A. Gupta and D. Nagar, *Matrix Variate Distributions*. Chapman and Hall/CRC, 2000.

[16] J. Salo, P. Suvikunnas, H. El-Sallabi, and P. Vainikainen, “Some results on MIMO capacity: the high SNR case,” 2004, to be submitted.