Tight Lower Bounds on the Ergodic Capacity of Rayleigh Fading MIMO Channels

Özgür Oyman[1)], Rohit U. Nabar[1)], Helmut Bölcskei[2)], and Arogyaswami J. Paulraj[1)]

[1)]Information Systems Laboratory, Stanford University
228 Packard, 350 Serra Mall, Stanford, CA 94305
Email: {oyman, nabar, apaulraj}@stanford.edu

[2)]Communication Technology Laboratory, ETH Zurich
ETH Zentrum, ETF E122, Sternwartstrasse 7, 8092 Zürich, Switzerland
Email: boelcskei@nari.ee.ethz.ch

Abstract— In this paper, we consider Gaussian multiple-input multiple-output (MIMO) fading channels assuming that the channel is unknown at the transmitter and perfectly known at the receiver. Using results from multivariate statistics, we derive a tight closed-form lower-bound for the ergodic capacity of such channels at any signal-to-noise ratio (SNR). Moreover, we provide an accurate closed-form analytical approximation of ergodic capacity in the high SNR regime. Our analysis incorporates the frequency-selective Rayleigh fading case and/or spatial fading correlation, and allows important insights into optimal (ergodic capacity maximizing) MIMO configurations. Finally, we verify our analytical expressions through comparison with numerical results.

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]. Analytical expressions for the resulting capacity gains are in general difficult to obtain.

Contributions. In this paper, we examine the ergodic capacity [6] of multiple-input multiple-output (MIMO) channels under the assumption that the channel is unknown at the transmitter and perfectly known at the receiver. Our detailed contributions are as follows:

- We derive a *closed-form lower-bound* for the ergodic capacity of MIMO channels experiencing *frequency-selective Rayleigh fading and/or spatial fading correlation.* Moreover, we provide an accurate closed-form approximation of ergodic capacity in the high signal-to-noise ratio (SNR) regime.
- We *quantify the loss in terms of ergodic capacity* due to spatial fading correlation *analytically.*
- For the i.i.d. Rayleigh fading case, given a *fixed total number of antennas* (transmit and receive), we determine *antenna configurations* that maximize ergodic capacity.

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 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 which may be evaluated using Monte Carlo methods. In [7], [8], closed-form lower bound expressions for the ergodic capacity of i.i.d. Rayleigh flat-fading channels with multiple antennas have been reported. While [7] provides closed-form ergodic capacity expressions for channels with multiple antennas at one end of the link (SIMO or MISO) and specifies the 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. Both lower bounds are derived assuming high SNR, which leads to poor accuracy at low SNR.

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], [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 evaluating the lower bound derived in [3] through Monte Carlo methods. Additionally, our

Ö. Oyman's work was supported by the Benchmark Stanford Graduate Fellowship.

R. U. Nabar's work was supported by the Dr. T. J. Rodgers Stanford Graduate Fellowship.

H. Bölcskei's work was supported by NSF grants CCR 99-79381 and ITR 00-85929.

results incorporate the frequency-selective case and the case of spatial fading correlation, and enable us to quantify the loss in ergodic capacity due to spatial fading correlation analytically.

Organization of the paper. The rest of this paper is organized as follows: In Section II, we derive a lower bound on the ergodic capacity of i.i.d. Rayleigh flat-fading MIMO channels. In Section III, we extend our results to incorporate the cases of frequency-selective fading and/or spatial fading correlation. Section IV examines optimal antenna allocation strategies for the i.i.d. Rayleigh flat-fading case. We present numerical results in Section V, and conclude in Section VI.

II. Ergodic Capacity Bound for the i.i.d. Case

Consider a narrow-band flat-fading MIMO system with M_T transmit and M_R receive antennas. The input-output relation for such a channel is characterized by the $M_R \times M_T$ channel transfer matrix $\mathbf{H}$ consisting of zero-mean uncorrelated circularly symmetric complex Gaussian elements with unit variance. We furthermore assume that the channel is unknown at the transmitter and perfectly known at the receiver. The mutual information of the corresponding MIMO system is given by [1] [2], [3]

$$I = \log_2 \det\left(\mathbf{I}_{M_R} + \frac{\rho}{M_T}\mathbf{H}\mathbf{H}^H\right) \text{ bps/Hz}, \quad (1)$$

where ρ is the SNR at each of the receive antennas and the input signal vector was assumed to be circularly symmetric complex Gaussian with covariance matrix $\frac{\rho}{M_T}\mathbf{I}_{M_T}$. Assuming that the fading process is ergodic, a Shannon capacity or ergodic capacity exists and is given by[2] $C = \mathcal{E}\{I\}$.

Applying Minkowski's inequality [9] to (1), we can lower-bound the ergodic capacity as

$$C \geq M_R\,\mathcal{E}\left\{\log_2\left(1+\rho\left[\det\left(\frac{1}{M_T}\mathbf{H}\mathbf{H}^H\right)\right]^{1/M_R}\right)\right\}, \quad (2)$$

which can alternatively be expressed as

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

Noting that $\log_2(1+ae^x)$ is a convex function in x for $a > 0$, and applying Jensen's inequality [10], we can further lower-bound C as

$$C \geq M_R\log_2\left(1+\rho\exp\left(\frac{1}{M_R}\times \mathcal{E}\left\{\ln\det\left(\frac{1}{M_T}\mathbf{H}\mathbf{H}^H\right)\right\}\right)\right). \quad (3)$$

For $M_T \geq M_R$, we can infer from [11] that

$$\mathcal{E}\left\{\ln\det\left(\frac{1}{M_T}\mathbf{H}\mathbf{H}^H\right)\right\} = \sum_{j=1}^{M_R}\mathcal{E}\{\ln X_j\} - M_R\ln 2M_T, \quad (4)$$

where X_j is a chi-squared random variable with $2(M_T - j + 1)$ degrees of freedom. From [12], we know that

$$\mathcal{E}\{\ln X_j\} = \ln 2 + \psi(M_T - j + 1), \quad (5)$$

where $\psi(x)$ is the *digamma* function. For integer x, $\psi(x)$ may be expressed as [13]

$$\psi(x) = -\gamma + \sum_{p=1}^{x-1}\frac{1}{p}, \quad (6)$$

where $\gamma \approx 0.57721566$ is Euler's constant. Combining (3), (4), (5), and (6) we have a lower bound for the ergodic capacity of MIMO channels when $M_T \geq M_R$. Using the identity $\det(\mathbf{I}_{M_R} + \frac{\rho}{M_T}\mathbf{H}\mathbf{H}^H) = \det(\mathbf{I}_{M_T} + \frac{\rho}{M_T}\mathbf{H}^H\mathbf{H})$, similar steps can be pursued to derive a lower bound on ergodic capacity for the case when $M_T < M_R$. We can now summarize our results as follows: The ergodic capacity of an $M_R \times M_T$ MIMO channel can be lower-bounded as

$$C \geq L\log_2\left(1+\frac{\rho}{M_T}\exp\left(\frac{1}{L}\sum_{j=1}^{L}\sum_{p=1}^{K-j}\frac{1}{p} - \gamma\right)\right), \quad (7)$$

where $K = \max(M_T, M_R)$ and $L = \min(M_T, M_R)$.

In the high SNR regime($\rho \gg 1$), the ergodic capacity can be approximated as

$$C \approx \begin{cases} \mathcal{E}\left\{\log_2\det\left(\frac{\rho}{M_T}\mathbf{H}\mathbf{H}^H\right)\right\}, & M_T \geq M_R \\ \mathcal{E}\left\{\log_2\det\left(\frac{\rho}{M_T}\mathbf{H}^H\mathbf{H}\right)\right\}, & M_T < M_R \end{cases}. \quad (8)$$

Starting from (8) and following similar steps as above, we find the following approximation for ergodic capacity at high SNR

$$C \approx L\log_2\left(\frac{\rho}{M_T}\right) + \frac{1}{\ln 2}\left(\sum_{j=1}^{L}\sum_{p=1}^{K-j}\frac{1}{p} - \gamma L\right). \quad (9)$$

[1] The superscript ^{H} stands for conjugate transpose. $\mathbf{I}_m$ is the $m \times m$ identity matrix

[2] $\mathcal{E}$ stands for the expectation operator.

This result is intuitively appealing since it shows explicitly that the ergodic capacity grows linearly with $\min(M_T, M_R)$. More specifically, C increases by $\min(M_T, M_R)$ for every 3 dB increase in SNR. Thus, the number of spatial data pipes that can be opened up between the transmitter and the receiver is constrained by the minimum of the number of antennas at the transmitter and receiver. Numerical results (obtained through Monte Carlo methods) in Sec. V reveal (7) to be a tight lower bound on ergodic capacity at any SNR and (9) to be an accurate expression in the high SNR regime.

III. Incorporating Frequency Selectivity and/or Spatial Fading Correlation

The analysis in Sec. II can easily be extended to more general channel models taking into account spatial fading correlation and frequency selectivity. In particular, we consider the broadband MIMO channel model introduced in [5], which is briefly reviewed in the following. Denoting the discrete-time index by n, the input-output relation for the channel model in [5] is given by

$$\mathbf{r}[n] = \sum_{l=0}^{P-1} \mathbf{H}_l \mathbf{s}[n-l], \tag{10}$$

where $\mathbf{r}[n]$ is the $M_R \times 1$ received signal, $\mathbf{H}_l \, (l = 0, 1, ..., P-1)$ is the $M_R \times M_T$ matrix channel impulse response, and $\mathbf{s}[n]$ is the $M_T \times 1$ transmit signal. Again the channel is assumed to be unknown at the transmitter and perfectly known at the receiver. Moreover, it is assumed that the transmit array is surrounded by local scatterers so that fading at the transmit antennas is spatially uncorrelated. The receive array is assumed to be high enough so that it is unobstructed and no local scattering occurs. Therefore, spatial fading at the receiver will be correlated. This correlation is captured through a set of $M_R \times M_R$ receive correlation matrices $\mathbf{R}_l \, (l = 0, 1, ..., P-1)$ such that

$$\mathbf{H}_l = \mathbf{R}_l^{1/2} \mathbf{H}_{w,l}, \quad l = 0, 1, ..., P-1, \tag{11}$$

where $\mathbf{H}_{w,l} \, (l = 0, 1, ..., P-1)$ is an $M_R \times M_T$ matrix with i.i.d. circularly symmetric complex Gaussian entries having zero mean and unit variance. The $\mathbf{H}_{w,l}$ are assumed to be uncorrelated. The receive correlation matrices $\mathbf{R}_l$ depend on the propagation environment and receive antenna spacing [5].

The ergodic capacity of the MIMO channel described by (10) and (11) is given by [5]

$$C = \mathcal{E}\left\{ \log_2 \det \left(\mathbf{I}_{M_R} + \frac{\rho}{M_T} \boldsymbol{\Lambda} \mathbf{H}_w \mathbf{H}_w^H \right) \right\}, \tag{12}$$

where $\mathbf{H}_w$ is an $M_R \times M_T$ matrix consisting of zero-mean uncorrelated circularly symmetric complex Gaussian elements with unit variance and[3] $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_i(\mathbf{R})\}_{i=0}^{M_R-1}$ with $\mathbf{R} = \sum_{l=0}^{P-1} \mathbf{R}_l = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$. Following our analysis in Sec. II, and assuming that $\mathbf{R}$ has rank $r \leq M_R$, it is easy to verify that the ergodic capacity in (12) for the case when $r \leq M_T$ may be conveniently lower-bounded as

$$\begin{aligned} C \quad \geq \quad & r \log_2 \left(1 + \frac{\rho}{M_T} \left(\det \boldsymbol{\Lambda}_r\right)^{1/r} \right. \\ & \left. \exp\left(\frac{1}{r} \sum_{j=1}^{r} \sum_{p=1}^{M_T-j} \frac{1}{p} - \gamma \right) \right), \end{aligned} \tag{13}$$

where $\boldsymbol{\Lambda}_r$ is the $r \times r$ diagonal matrix containing the non-zero eigenvalues of $\boldsymbol{\Lambda}$. Similar to the i.i.d. case, we can establish that the ergodic capacity increases by r bps/Hz for every 3 dB increase in SNR. We conclude by noting that for full-rank $\mathbf{R}$, the loss in ergodic capacity in the high SNR regime is quantified by $\log_2(\det(\boldsymbol{\Lambda}))$.

IV. Capacity Optimal Antenna Allocation

The problem addressed in this section is the following. 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? This is a relevant question in design of point-to-point MIMO wireless links with fixed number of antennas to be placed on transmit and receive sides. We restrict our analysis to the case of i.i.d. Rayleigh flat-fading and the high SNR regime, and use the approximation derived in (9) to quantify the *differential capacity gain* $\delta C(r \to t)$ as the ergodic capacity gain obtained by placing an extra antenna at the receiver instead of the transmitter. Denoting the approximation of ergodic capacity in (9) for an $M_R \times M_T$ antenna system by $C(M_R, M_T)$, the differential capacity gain $\delta C(r \to t)$ is given by

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

We now examine the behavior of $\delta C(r \to t)$ for three different scenarios:

Case 1: $M_T > M_R$

$$\begin{aligned} \delta C(r \to t) = \log_2 \frac{\rho}{M_T} - M_R \log_2 \frac{M_T}{M_T+1} + \\ \frac{1}{\ln 2} \left(\sum_{p=1}^{M_T-M_R-1} \frac{1}{p} - \sum_{j=1}^{M_R} \frac{1}{M_T+1-j} - \gamma \right), \end{aligned}$$

[3] $\lambda_i(\mathbf{R})$ is the i-th eigenvalue of $\mathbf{R}$.

which is positive if

$$\rho > \frac{M_T^{M_T}}{(M_T+1)^{M_T-1}} \exp\left(\sum_{r=2}^{M_T} \frac{1}{r} + \gamma\right). \quad (15)$$

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 that by adding a receive antenna, the rank of the channel realizations increases or equivalently an 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 $L = \min(M_T, M_R) = M_R$. For a system with $M_T = 5$ and $M_R = 4$, the required SNR to satisfy (15) is 11.91 dB.

Case 2: $M_T = M_R$

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

which is clearly positive for all $M_T = M_R$, indicating that an additional antenna should be placed at the receiver. 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 assumption that the receiver knows the channel perfectly and can hence realize array gain.

Case 3: $M_T < M_R$

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

which is negative if

$$\rho > \frac{{M_R}^{M_R}}{(M_R-1)^{M_R-1}} \exp\left(\sum_{r=2}^{M_R} \frac{1}{r} + \gamma\right). \quad (17)$$

Hence, provided that ρ is sufficiently large, it is optimal to place an additional antenna at the transmitter rather than at the receiver. The explanation for this result is the same as in case 1. Adding an additional transmit antenna increases the rank of the individual channel realizations or equivalently an additional spatial data pipe can be opened up. For an antenna system with $M_T = 4$ and $M_R = 5$, the required SNR to satisfy (17) is 18.95 dB.

We note that using the results presented above, it is easy to verify that for a total of $2N$ antennas, a system with N antennas each at the transmitter and receiver (square system) maximizes the ergodic capacity.

V. Numerical Results

In this section, we demonstrate the accuracy of our analytical expressions. Furthermore, we compare our bounds with previously derived lower bounds on ergodic capacity and numerically analyze the loss in ergodic capacity due to spatial fading correlation.

A. Flat-fading i.i.d. Rayleigh channel

Fig. 1 shows the empirical (obtained through Monte Carlo methods) ergodic capacity and the analytical lower bound (7) for several MIMO configurations. It is clearly seen that (7) is almost exact at high SNR and that it gets tighter at low SNR as the difference in the number of antennas on the two sides of the link increases.

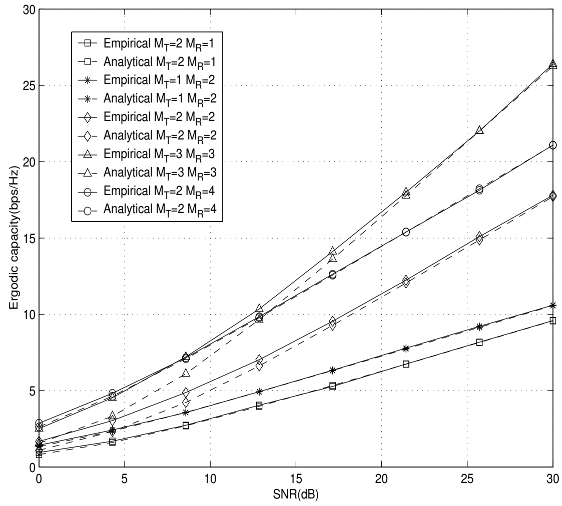

Fig. 1. *Comparison of the empirically determined ergodic capacity and the analytical lower bound for several antenna configurations.*

Next, we compare the closed-form expression (7) with previously published lower bounds. We consider a system with $M_T = M_R = 2$. Fig. 2 depicts the closed-form lower bounds reported in [7], [8] and the lower bound obtained by evaluating the results in [3] through Monte Carlo methods. We observe that in the low SNR regime our closed-form lower bound (7) is as tight as the numerically evaluated lower bound of [3] and much tighter than the lower bounds specified in [7], [8]. In the high SNR regime all bounds are equally tight.

B. Spatially correlated flat-fading Rayleigh channel

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 (10) with $P = 1$ (i.e. no delay spread). The level of spatial fading correlation is

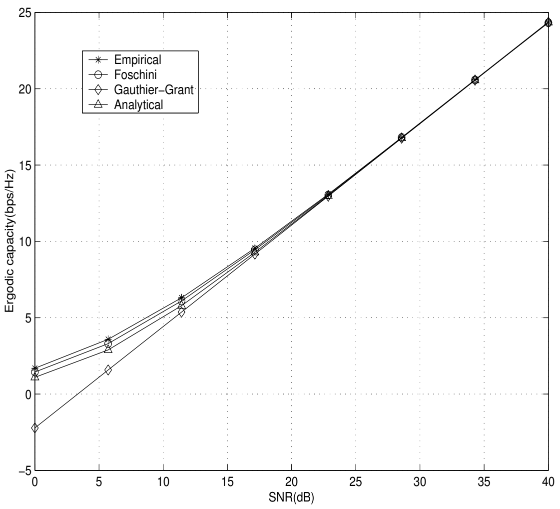

Fig. 2. *Comparison of lower bounds on ergodic capacity for* 2×2 *MIMO channel.*

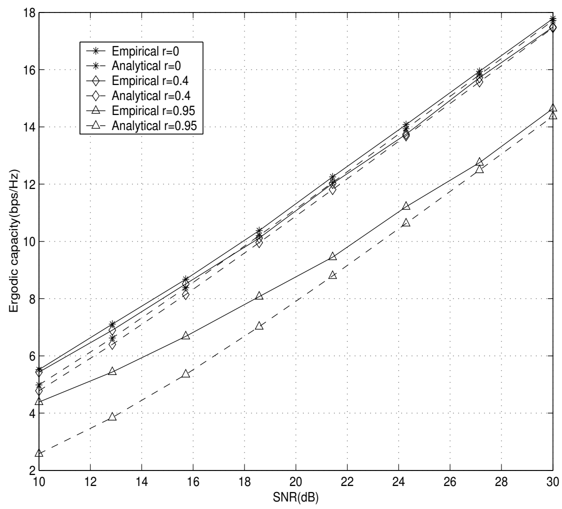

Fig. 3. *Comparison of the empirically determined ergodic capacity and the analytical lower bound for various levels of spatial fading correlation for* 2×2 *MIMO channel.*

determined through the receive correlation matrix $\mathbf{R}_0$. Specifically, we set

$$\mathbf{R}_0 = \begin{bmatrix} 1 & r \\ r^* & 1 \end{bmatrix},$$

where r is the complex correlation coefficient between the two receive antennas. In Fig. 3, we compare (13) with (12) evaluated by Monte Carlo methods for three different levels of correlation, namely $r = 0$ (i.i.d. channel), $r = 0.4$ (low correlation), and $r = 0.95$ (high correlation). As predicted by the analytical estimate $\log_2(\det(\mathbf{R}_0))$, we observe a very small ergodic capacity loss for the case of low correlation. In the case of high correlation, we observe an ergodic capacity loss of 3.35 bps/Hz again consistent with the loss predicted by the analytical estimate.

VI. Conclusions

We derived a tight closed-form analytical lower bound on the ergodic capacity of Rayleigh fading MIMO channels. Our analysis incorporates the frequency-selective case and/or spatial fading correlation. We demonstrated that our lower bound is tighter than previously developed analytical lower bounds and can be applied to a system with any number of transmit and receive antennas. For the high SNR case, we derived an almost exact approximation of ergodic capacity. Finally, using our results, we determined optimal (in the sense of ergodic capacity maximizing) MIMO antenna configurations for the high SNR regime.

References

[1] A. J. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception," *U. S. Patent,* no. *5,345,599*, 1994.

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecomm.*, 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 Personal Communications*, vol. 6, no. 3, pp. 311–335, March 1998.

[4] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Comm.*, vol. 46, no. 3, pp. 357–366, 1998.

[5] H. Bölcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Comm.*, vol. 50, no. 2, pp. 225–234, Feb. 2002.

[6] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[7] E. Gauthier, A. Yongacoglu, and J-Y. Chouinard, "Capacity of multiple antenna systems in Rayleigh fading channels," *Canadian Journal of Electrical and Computer Engineering*, vol. 25, no. 3, pp. 105–108, July 2000.

[8] A. Grant, "Rayleigh fading multiple-antenna channels," *EURASIP Journal on Applied Signal Processing, Special Issue on Space-Time Coding(Part I)*, vol. 2002, no. 3, pp. 316–329, March 2002.

[9] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge Press, New York, 1985.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[11] N. R. Goodman, "The distribution of the determinant of a complex Wishart distributed matrix," *Ann. Math. Stat.*, vol. 34, no. 1, pp. 178–180, March 1963.

[12] P. M. Lee, *Bayesian Statistics: An Introduction*, Arnold/John Wiley, London/New York, 2nd edition, 1997.

[13] X. Gourdon and P. Sebah, "The Gamma function," available online at http://numbers.computation.free.fr, Sept. 2000.