ASYMPTOTIC CAPACITY GAIN OF TRANSMIT ANTENNA SELECTION

Shahab Sanayei and Aria Nosratinia

Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080
Email: shahab.sanayei@student.utdallas.edu,aria@utdallas.edu

ABSTRACT

Antenna selection provides a low-cost low complexity solution for MIMO systems, in particular transmit antenna selection exploits partial knowledge of the channel that can be made available without excessive feedback bandwidth. In this work we explore information theoretic limits of transmit antenna selection for high SNR and large number of transmit antennas. We define *the capacity gain* as the constant term in the asymptotic expansion of the ergodic capacity with respect to the SNR. We show that this value is directly related to the channel state information (CSI) at the transmitter. We compute and investigate the asymptotic behavior of the capacity gain for three cases: where the transmitter has complete CSI, no CSI, and partial CSI (antenna selection). We show that while water-filling provides a capacity gain that increases logarithmically in M (the number of transmit antennas), the capacity gain of transmit antenna selection behaves only like $\log(\log(M))$.

1. INTRODUCTION

Multiple antenna systems provide high spectral efficiency in rich scattering environments. However, with a large number of transmit or receive antennas, the complexity and cost of multiple RF-chains and low noise receivers become considerable. One solution to the cost/complexity problem is antenna subset selection [1, 2, 3]. It has been shown that the use of subset selection at transmit side, while requiring minimal feedback, can considerably improve the data rate in a MIMO channel [4, 5]. However, optimal subset selection itself has a computational complexity which grows exponentially with the number of selected antennas, therefore it is not suitable for practical purposes. Fast algorithms have been proposed to reduce the complexity while still capturing a large portion of the capacity [6, 7, 8]. Despite the importance of suboptimal algorithms, not much is known about their information information theoretic limits. In [9] and [7] an analytical framework for receive antenna selection has been purposed to prove the equivalence in diversity order between the full system and selected sub-system. In this paper we explore the capacity gain obtained by transmit antenna selection for high SNR and large number of transmit antennas.

A brief note on notation: $\mathbb{E}[\ ]$ refers to expected value of a random variable, I_N denotes the $N \times N$ identity matrix, $(x)^+ = \max\{x, 0\}$, and $\gamma \approx 0.57721566$ is the Euler-Mascheroni constant. We use $a_n \overset{\circ}{=} b_n$ to denote the asymptotic equivalence of a_n and b_n defined as: $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. We use the natural logarithm througout this paper so the capacity unit is in Nats/Sec/Hz.

This work was supported in part by a grant from NSF and in part by a grant from Nortel Networks

2. SYSTEM MODEL

We assume a frequency non-selective (flat) linear time invariant fading channel between M transmit and N receive antennas. The signal model is:

$$y(t) = \sqrt{\frac{\rho}{M}} Hx(t) + n(t) \tag{1}$$

where $y(t)$ represents the $N \times 1$ received vector sampled at time t, and $x(t)$ represents the $M \times 1$ vector transmitted by the antennas, ρ is the average SNR (per channel use), $n(t)$ is the $N \times 1$ additive circularly symmetric complex Gaussian noise vector with zero mean and covariance matrix equal to I_N (the $N \times N$ identity matrix) and H is the $N \times M$ channel matrix, whose ij-th element is the scalar channel between the i-th receive and j-th transmit antenna. We assume that the elements of H are independent and have complex Gaussian distribution with zero mean and unit variance. We also assume that H is perfectly known at the receiver but it is not necessarily known at the transmitter. For antenna selection, we assume there is a rate-limited feedback channel from receiver to transmitter so that a subset of transmit antennas can be selected by the receiver and furthermore we assume that the feedback channel is without error or delay.

3. TRANSMIT ANTENNA SELECTION

We consider a transmit antenna selection scheme where a subset of transmit antennas are used for transmission with equal power. This is made possible by a rate-limited feedback from the receiver. Optimal transmit antenna selection via exhaustive search among all $\binom{M}{L}$ combinations has complexity $O(M^L)$, which is impractical for large number of transmit antennas. One may reduce this complexity by employing a successive selection scheme, i.e., a greedy algorithm that at each step maximizes the capacity of the selected sub-channel. A very similar methodology was mentioned in [6] for receive antenna selection. In this work, starting from the original channel matrix, the algorithm removes antennas one after another in a way that the capacity loss is minimized. Gharavi-Alkhansari and Greshman [8] showed that an incremental successive selection leads to less computational complexity. Simulation results show that this successive selection captures almost all the capacity of optimal antenna selection in a wide range of SNRs. Therefore, we adopt the latter algorithm for our analysis.

3.1. Incremental Successive Subset Selection Algorithm

The input of the algorithm consists of ρ (the given SNR), L (the desired number of transmit antennas to be selected) and H (the original channel matrix). The output of the algorithm is $\widetilde{H}$ (the channel matrix associated with selected transmit antennas).

1. Let $\mathcal{S}_1 = \{\text{all columns of } H\}$ and $\widetilde{P}_1 = I_N$.
2. choose $\widetilde{h}_1 = \underset{h \in \mathcal{S}_1}{\arg\max} \ ||h||_2$, $\widetilde{H}_1 = \widetilde{h}_1$.
3. for $i = 2 : L$
   (a) $\mathcal{S}_i = \mathcal{S}_{i-1} \cap \{\widetilde{h}_{i-1}\}^c$
   (b) $\widetilde{P}_i = I - \frac{\rho}{L}\widetilde{H}_{i-1}(I + \frac{\rho}{L}\widetilde{H}_{i-1}^H\widetilde{H}_{i-1})^{-1}\widetilde{H}_{i-1}^H$
   (c) $\widetilde{h}_i = \underset{h \in \mathcal{S}_i}{\arg\max} \ h^H\widetilde{P}_i h$
   (d) $\widetilde{H}_i = [\widetilde{H}_{i-1} \ \widetilde{h}_i]$
4. $\widetilde{H} = \widetilde{H}_L$

3.2. A Framework for Analysis

Using the Sherman-Morisson formula for determinants, for the selected channel $\widetilde{H}$ we have:

$$det(I_N + \frac{\rho}{L}\widetilde{H}\widetilde{H}^H) = \prod_{i=1}^{L}(1 + \frac{\rho}{L}\widetilde{h}_i^H\widetilde{P}_i\widetilde{h}_i) \qquad (2)$$

As $\rho \to \infty$, $\widetilde{P}_i \to P_i$, where, $P_i = I_N - \widetilde{H}_i(\widetilde{H}_i^H\widetilde{H}_i)^{-1}\widetilde{H}_i$ is a projection matrix of rank $N - i + 1$.

When M is also large, at each selection step, the distribution of the remaining channel vectors can still be well approximated by a circularly symmetric Gaussian distribution. The actual proof of this claim does not seem to be tractable and even for very simple cases, there is no known result. Our simulation results show that for large M the Gaussianity assumption provides a good approximation for the actual distribution of the remaining columns. Using this assumption we can approximate the statistics of the right side of (**??**). We know that for an uncorrelated complex Gaussian vector x and a projection matrix P, x^HPx has χ^2 distribution with $rank(P)$ degrees of freedom. Hence for large ρ and large M, we have:

$$\det(\widetilde{H}\widetilde{H}^H) \sim \prod_{i=1}^{L}\widetilde{\chi}^2_{2(N-i+1),M-i+1} \qquad (3)$$

where $\widetilde{\chi}^2_{2p,n}$ stands for a random variable which is the maximum of n independent χ^2_{2p} random variables. The pdf of this random variable can be computed in closed form [10]:

$$f_{\widetilde{\chi}^2_{2p,n}}(x) = \frac{nx^{p-1}}{(p-1)!}e^{-x}(1 - e^{-x}e_p(x))^{n-1} \qquad (4)$$

where $e_p(x) = \sum_{k=0}^{p-1}\frac{x^k}{k!}$

4. CAPACITY GAIN OF MIMO SYSTEMS

We introduce the concept of capacity gain as a measure of effectiveness of channel state information at the transmitter. We start with the capacity expression for a general MIMO system. Under the flat fading assumption, given a general channel matrix, the ergodic capacity of the MIMO channel is calculated as follows [11]:

$$C = \mathbb{E}[\max_{tr(Q)\leq\rho} \log(\det(I_N + HQH^H))] \qquad (5)$$

where $Q = \mathbb{E}[xx^H]$ is the covariance of the transmitted vector x.

We consider three different cases: First, uninformed transmission in which CSI is perfectly known at receiver, but not at transmitter. Second, informed transmission in which CSI is perfectly known both at transmitter and receiver. Third, transmission using an optimal subset of transmit antennas selected by the receiver. The only information available at the transmitter is the indices of the selected transmit antennas.

Uninformed transmitter: As addressed in [11], when CSI is only available at receiver, the covariance matrix that maximizes the capacity is of the form $Q = \frac{1}{M}I_N$, hence, the ergodic capacity is:

$$C = \mathbb{E}[\log(\det(I_N + \frac{\rho}{M}HH^H))] = \mathbb{E}\left[\sum_{i=1}^{m}\log(1 + \frac{\rho}{M}\lambda_i)\right]$$

where where $\lambda_1, \ldots, \lambda_m$ are ordered nonzero eigenvalues of the Wishart matrix HH^H [11] and $m = rank(H) = \min\{M, N\}$ can be interpreted as the degrees of freedom of the MIMO channel. As shown in [12], $C = m \log \rho + O(1)$. So the ergodic capacity grows linearly with m. Now we notice that in the asymptotic expansion of C there is a constant term that does not vanish as $\rho \to \infty$. Thus we define the capacity gain as follows:

$$G \stackrel{\triangle}{=} \lim_{\rho \to \infty} (C - m \log \rho) \tag{6}$$

For uninformed transmission we have:

$$\begin{aligned} G &= \lim_{\rho\to\infty}\left(\mathbb{E}\left[\sum_{i=1}^{m}\log(1+\frac{\rho}{M}\lambda_i)\right] - m\log\rho\right) \\ &= \lim_{\rho\to\infty}\mathbb{E}\left[\sum_{i=1}^{m}\log(\frac{1}{\rho}+\frac{\lambda_i}{M})\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{m}\log\lambda_i\right] - m\log M \end{aligned} \tag{7}$$

in the last step, the exchange of expectation and limit is allowed by the monotone convergence theorem.

Informed transmitter (*water-filling capacity*): In this case the channel state information is available at transmitter. As addressed in [11] the so called water-filling capacity of the MIMO channel is:

$$C_{wf} = \mathbb{E}\left[\sum_{i=1}^{m}(\log(\mu\lambda_i))^+\right] \tag{8}$$

where μ should satisfy $\rho = \sum_{i=1}^{m}(\mu - \lambda_i^{-1})^+$. In large SNR scenario, all the eigenmodes of the channel are used by the beamformer, hence $\mu = \frac{\rho + \sum_{i=1}^{m}\lambda_i^{-1}}{m}$ and the water-filling capacity is equal to:

$$\begin{aligned} C_{wf} &= \mathbb{E}\left[\sum_{i=1}^{m}(\log(\mu\lambda_i))\right] = m\mathbb{E}\left[\log(\rho + \sum_{i=1}^{m}\lambda_i^{-1})\right] + \\ & \quad \mathbb{E}\left[\sum_{i=1}^{m}\log\lambda_i\right] - m\log m \end{aligned} \tag{9}$$

For large ρ we have: $C_{wf} \approx m \log \rho$. In other words, availability of CSI at transmitter side has no impact on the logarithmic growth rate of the ergodic capacity, because the growth rate only depends on the rank of the channel matrix. Now we similarly calculate the capacity gain for informed transmission:

$$G_{wf} = \mathbb{E}\left[\sum_{i=1}^{m}\log\lambda_i\right] - m\log m \tag{10}$$

$$\Delta G = G_{wf} - G = m\log(M/m) \tag{11}$$

In the asymptote of large SNR, this is *the maximum amount of excess rate obtainable by providing channel state information at the transmitter*. We note that if $M \leq N$ then $\Delta G = 0$, thus channel state information at transmitter cannot provide any excess rate asymptotically. This result agrees with one's intuition that beamforming is effective only when the number of transmit antennas is large. In the sequel, we only consider the interesting case of $M > N$. In particular, we are interested to understand the behavior of the capacity gain when $M \gg N$. In these cases, Equation (11) suggests that *at high SNR, the capacity gain can be used as an information-theoretic metric to evaluate any method that uses channel state information at the transmitter*.

Antenna selection: In the high SNR regime, one is interested in the case $L \geq N$, to maintain the degrees of freedom of the channel and prevent excessive rate loss. Suppose we have selected L ($L \geq N$) out of M transmit antennas ($M \gg N$) then if the selected channel is $\widetilde{H}$, the capacity gain is:

$$\begin{aligned} \widetilde{G} &= \lim_{\rho\to\infty}\mathbb{E}\left[\log\left(\det(I_N + \frac{\rho}{L}\widetilde{H}\widetilde{H}^H)\right)\right] - N\log\rho \\ &= \lim_{\rho\to\infty}\mathbb{E}\left[\log\left(\det(\frac{1}{\rho}I_N + \frac{1}{L}\widetilde{H}\widetilde{H}^H)\right)\right] \\ &= \mathbb{E}\left[\log\left(\det(\widetilde{H}\widetilde{H}^H)\right)\right] - N\log L \end{aligned} \tag{12}$$

5. ASYMPTOTIC BEHAVIOR OF CAPACITY GAIN

In this section we explore the behavior of capacity gain, for large M, in the case of informed, uninformed, and antenna selection transmitter.

Uninformed transmitter: in the case $M > N$, Equation (7) can be rewritten as:

$$G = \mathbb{E}\left[\log\det(HH^H)\right] - N\log M \tag{13}$$

It is known [12] that $\det(HH^H) \sim \prod_{i=1}^{N}\chi^2_{2(M-i+1)}$ therefore [13]:

$$G = \sum_{i=1}^{N}(\psi(M-i+1) - \log M) \tag{14}$$

where $\psi(n) = -\gamma + \sum_{k=1}^{n}\frac{1}{k}$ is the *di-gamma function*. We have [12]:

$$\lim_{M\to\infty} G = 0 \tag{15}$$

Informed transmitter: Using Equations (11) and (15) for large M we have:

$$G_{wf} \stackrel{\circ}{=} N\log(\frac{M}{N}) \stackrel{\circ}{=} N\log M \tag{16}$$

Antenna selection: Using the results of Section 3.2, we can evaluate the capacity gain for antenna selection. Using (12):

$$\begin{aligned}\widetilde{G} &= \mathbb{E}\left[\log\det(\widetilde{H}\widetilde{H}^H)\right] - N\log L \\ &\mathrel{\overset{\circ}{=}} \sum_{i=1}^{L}\mathbb{E}[\log(\tilde{\chi}^2_{2(N-i+1),M-i+1})] - N\log L \quad (17)\end{aligned}$$

Equation (17) suggests that for large M, selecting more than N antennas does not provide any further gain. In the previous section we argued that L cannot be less than N, hence for large M the optimal value for L is N. Henceforth we assume $L = N$. To evaluate the asymptotic behavior of $\widetilde{G}$, we only need to evaluate $\mathbb{E}[\log X]$, where $X \sim \tilde{\chi}^2_{p,n}$. We use the following result from order statistics [10]:

Definition: A cdf F is said to belong to *the domain of maximal attraction* of a nondegenerate cdf U if there exist sequences $\{a_n\}$ and $\{b_n > 0\}$ such that

$$\lim_{n\to\infty} F^n(a_n + b_n x) = U(x) \quad (18)$$

at all continuity points of $U(x)$.

Theorem 1 *: Let $X_{(n)}$ be the maximum of n i.i.d. random variables $\{X_i\}_{i=1}^n$, each with cdf F. If F belongs to the domain of maximal attraction of U, then:*

$$\frac{X_{(n)} - a_n}{b_n} \xrightarrow{d} W \quad (19)$$

where W is a random variable whose cdf is U.

It is known that the cdf of a χ^2_{2p} random variable is equal to $F(x) = 1 - e^{-x}e_p(x)$, where $e_p(x) = \sum_{k=0}^{p-1}\frac{x^k}{k!}$. Choosing $a_n = \log n + \log(\log(\frac{n^{p-1}}{(p-1)!}))$ and $b_n = 1$ we have:

$$\lim_{n\to\infty} F^n(a_n + b_n x) = e^{-e^{-x}} \quad (20)$$

We use (20) to evaluate the logarithmic moment of $\tilde{\chi}^2_{p,n}$ which is key to our analysis. In the above formulation we choose F to be the cdf of a chi-square random variable with p degrees of freedom. Jensen's inequality provides an upper bound on the logarithmic moment and furthermore the following theorem states that this bound is asymptotically tight (see appendix for the proof).

Theorem 2 *$X_{(n)}$, W, $\{a_n\}$ and $\{b_n > 0\}$ are defined as in Theorem 1 and furthermore, $\frac{|a_n|}{b_n} \to \infty$, then as $n \to \infty$:*

$$\log(\mathbb{E}[X_{(n)}]) - \mathbb{E}[\log X_{(n)}] \longrightarrow 0$$

So it is sufficient to evaluate the mean of the above random variable. Using Theorem 1, $\mathbb{E}[X_{(n)}] \mathrel{\overset{\circ}{=}} b_n\mathbb{E}[W] + a_n$, where W is a random variable with the cdf as in (20). It is easy to verify that $\mathbb{E}[W] = \gamma$. Hence, the asymptotic growth of the logarithmic moment of extreme order statistics of chi-square random variables is given by

$$\begin{aligned}\mathbb{E}[\log(X_{(n)})] &\mathrel{\overset{\circ}{=}} \log(\mathbb{E}[X_{(n)}]) \\ &\mathrel{\overset{\circ}{=}} \log\left(\log n + \log(\log(\frac{n^{p-1}}{(p-1)!})) + \gamma\right) \\ &\mathrel{\overset{\circ}{=}} \log(\log n) \quad (21)\end{aligned}$$

Now we can evaluate the behavior of $\widetilde{G}$ in (17):

$$\begin{aligned}\widetilde{G} &\mathrel{\overset{\circ}{=}} \sum_{i=1}^{N}\mathbb{E}[\log(\tilde{\chi}^2_{2(N-i+1),M-i+1})] - N\log N \\ &\mathrel{\overset{\circ}{=}} \sum_{i=1}^{N}\log\left(\frac{\log M + \log(\log(\frac{M^{N-i}}{(N-i)!})) + \gamma}{N}\right) \quad (22) \\ &\mathrel{\overset{\circ}{=}} N\log(\log M) \quad (23)\end{aligned}$$

Thus the capacity gain for transmit antenna selection behaves like $O(\log(\log M))$.

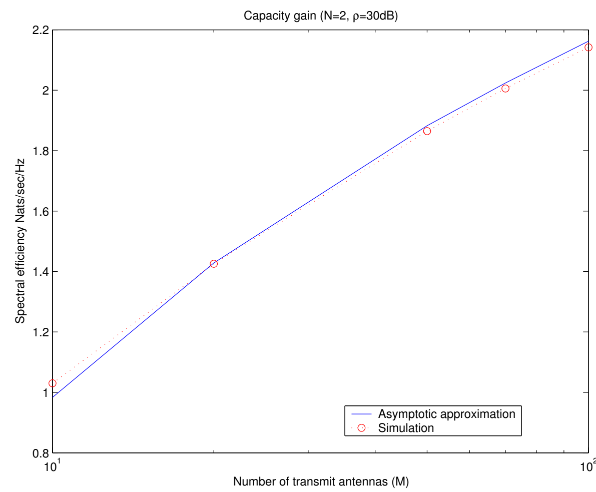

Fig. 1. Capacity gain of antenna selection (N=2 and SNR=30 dB)

6. SIMULATION RESULTS

Figure 1 compares our result with computer simulations. We run the simulation for SNR=30 dB and N=2. For each point on the plot, the capacity gain is calculated by averaging over 5000 different channel realizations and compared to the results obtained from equation (22). The results show a very good match between our formula and the simulation result. Thus the asymptotic formula is a useful tool for the approximation of the capacity of transmit antenna selection.

7. CONCLUSION

In this paper, we introduce the concept of capacity gain, an information theoretic metric for schemes that use various amounts of channel state information at the transmitter. We present some new results in order statistics that are useful for asymptotic analysis of antenna selection problems. We evaluate the capacity gain for transmit antenna selection in the asymptote of large number of transmit antennas. We show that this quantity describes the advantage gained by having channel state information at the transmitter. By exploring the behavior of the capacity gain, we showed that the optimal number of selected antennas for large M is exactly N. Simulations show that the analysis is accurate and can be used for approximation of the capacity of antenna selection.

8. APPENDIX

Lemma 1 *If* $X_n \stackrel{i.p.}{\longrightarrow} X$*, then:* $\mathbb{E}[X_n] \longrightarrow \mathbb{E}[X]$.

Proof: $X_n \stackrel{i.p.}{\to} X \implies X_n \stackrel{d}{\to} X \implies \mathbb{E}[X_n] \to \mathbb{E}[X]$

Lemma 2 *If* $X_n \stackrel{i.p.}{\to} X$ *and* $c_n \to c$ *then* $\frac{X_n}{c_n} \stackrel{i.p.}{\to} \frac{X}{c}$.

Proof: see [14].

Lemma 3 *If* $X_n \stackrel{i.p.}{\to} X$*, then for any continuous measurable function* $g()$ *we have:* $g(X_n) \stackrel{i.p.}{\to} g(X)$

Proof: see [14].

Proof of Theorem 2: For every $\epsilon > 0$ and $\delta > 0$

$$\begin{aligned} Pr\left[\left|\frac{X_{(n)}}{a_n} - 1\right| > \epsilon\right] &= Pr\left[\left|\frac{X_{(n)} - a_n}{b_n}\right| > \epsilon\frac{|a_n|}{b_n}\right] \\ \leq \frac{\mathbb{E}\left[\left(\frac{X_{(n)} - a_n}{b_n}\right)^2\right]}{\epsilon^2(\frac{a_n}{b_n})^2} &< \frac{\mathbb{E}[W^2] + \delta}{\epsilon^2(\frac{a_n}{b_n})^2} \longrightarrow 0 \end{aligned}$$

as $n \to \infty$. Thus $\frac{X_{(n)}}{a_n} \stackrel{i.p.}{\to} 1$. Now using Lemma 1 we have $\frac{\mathbb{E}[X_{(n)}]}{a_n} \to 1$, hence using Lemma 2 $\frac{X_{(n)}}{\mathbb{E}[X_{(n)}]} \stackrel{i.p.}{\to} 1$ and by Lemma 3 (for $g(x) = \log x$) we have $\log(\frac{X_{(n)}}{\mathbb{E}[X_{(n)}]}) \stackrel{i.p.}{\to} 0$. Finally by Lemma 1, $\mathbb{E}[\log(X_{(n)})] - \log(\mathbb{E}[X_{(n)}]) \to 0$, as $n \to \infty$.

9. REFERENCES

[1] M. Z. Win and J. H. Winters, "Analysis of hybrid selection/maximal ratio combining in rayleigh fading," *IEEE Trans. on Communications*, vol. 47, pp. 1773–1776, Dec. 1999.

[2] R. Heath and A. Paulraj, "Antenna selection for spatial multiplexing systems based on minimum error rate," in *Proc. International Conference on Communications*, Helsinki, Finland, June 2003, pp. 2276–2280.

[3] D. Gore, R. U. Nabar, and A. Paulraj, "Selecting an optimal set of transmit antennas for a low rank matrix channel," in *Proc. IEEE ICASSP*, Istanbul, Turkey, May 2000, pp. 2785–2788.

[4] R. S. Blum and J. H. Winters, "On optimum mimo with antenna selection," *IEEE Communications Letters*, vol. 6, pp. 322–324, Aug. 2002.

[5] D. Gore, R. Heath, and A. Paulraj, "Transmit selection in spatial multiplexing systems," *IEEE Communications Letters*, vol. 6, no. 1, pp. 491–493, Nov. 2002.

[6] A. Gorokhov, "Antenna selection algorithms for mea transmission systems," in *Proc. IEEE ICASSP*, Orlando, FL, May 2002, pp. 2875–2860.

[7] A. Gorokhov, "Receive antenna selection for MIMO-spatial multiplexing: Theory and algorithms," *submitted to IEEE Trans. Sig. Proc.*, 2003.

[8] M. Gharavi-Alkhansari and A. Greshman, "Fast antenna selection in MIMO systems," *to appear in the IEEE Trans. on Signal Proc. special issue on space-time communication*, 2003.

[9] A. Gorokhov, "Performance bounds for antenna selection in MIMO systems," in *icc*, Anchorage, AK, May 2003, pp. 3021–3025.

[10] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A first course in order statistics*, John Wiley and Sons, 1992.

[11] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecommunication*, vol. 10, pp. 585–595, Nov. 1999.

[12] G. J. Foschini and M. J. Gans, "On limits of wireless communication in fading environment when using multiple antennas," *Wireless Personal Communication*, vol. 6, pp. 311–335, March 1998.

[13] O. Oyman, R. U. Nabar, H. Bolcskei, and A. J. Paulraj, "Characterizing the statistical properties of mutual information in mimo channels: Insights into diversity-multiplexing tradeoff," *submitted to IEEE Transactions on Signal Processing*, Dec. 2002.

[14] T. S. Ferguson, *A Course in Large Sample Theory*, CRC Press, 1996.