Capacity of MIMO Channels with Antenna Selection

Shahab Sanayei and Aria Nosratinia
Department of Electrical Engineering
The University of Texas at Dallas, Richardson, TX 75083-0688
E-mail: {sxs025500,aria}@utdallas.edu

Abstract

We explore the capacity of MIMO channels in the presence of antenna selection. Antenna selection reduces the complexity of the radio devices and requires only a small amount of channel state feedback to the transmit side. For high SNR, we define *the capacity gain* as the constant term in the expansion of the ergodic capacity in terms of SNR. We show that this value is representative of the channel state information (CSI) at the transmitter. We investigate the asymptotic behavior of the capacity gain for three cases: complete CSI, no CSI, and partial CSI at transmitter (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 $\log(\log M)$. For the low SNR case, we use the concept of *channel gain*, a measure introduced by Verdu [1]. We show that channel gain for antenna selection increases only logarithmically in M as opposed to water-filling channel gain which increases linearly in M. The methodology developed in this paper, although motivated by antenna selection, is fairly general and is useful whenever partial CSI is available at the transmitter. The same techniques are also applied to the receive selection, and corresponding results are noted in hight- and low-SNR regimes.

I. Introduction

In MIMO systems, the cost and complexity of multiple RF-chains, power amplifiers, and low noise amplifiers (LNA) is a serious practical issue. One solution to the cost/complexity problem is antenna subset selection [2], [3], [4], [5].

At the receive side, antenna subset selection reduces the complexity. At the transmit side, antenna subset selection not only reduces the complexity, but also improves the capacity of the MIMO system [6], [7],

[8], [9] at the cost of a minimal amount of feedback. Fast and efficient algorithms have been devised to determine the selected antenna subsets [10], [11], [12]. However, despite the practical importance of antenna selection, the information theoretic properties of the resulting channels remains a mostly unexplored territory. A few notable exceptions exist [13], [14], however, to date closed form expressions for capacity have not been available.

In this paper, we analyze the capacity of the antenna selection channel in the high-SNR and low-SNR regimes. In the high-SNR regime, we define the notion of *capacity gain* as the constant term in the high-SNR expansion of the capacity expression, and demonstrate that it is directly related to transmit-side channel state information. This concept was first introduced in [8], [9] and is closely related to a similar concept that was independently proposed by Lozano et al. [15]. We are able to draw conclusions about the behavior of the system in the asymptote of large number of transmit antennas, and draw comparisons between the antenna selection capacity and water-filling capacity.

Our results have interesting implications in the design and analysis of all MIMO systems (not just antenna selection). The waterfilling capacity (C_{wf}) has the same growth rate as the capacity of the uninformed transmitter (C) [16], but nevertheless $C_{wf} > C$ with non-vanishing difference at high SNR[1]. The difference is an *excess rate* that is due to channel state feedback. The excess rate is not limited to waterfilling; when partial CSI is available at the transmitter the excess rate is still there, but is smaller. Examples of partial CSI include transmit antenna selection and channel covariance feedback. The growth of this excess rate with the number of transmit antennas is a measure of how effectively CSI is being used by a given method.

The above developments, the reader may recall, were in the high-SNR regime. In the low-SNR regime we employ a similar methodology for analysis. In particular, we look at the power series expansion of the capacity around SNR=0, where the coefficient of the first-order term is called *channel gain*[2], a quantity that is related to the channel state information. Using this notion, we show that at low SNR, the optimal selection strategy at the transmitter is to select exactly one transmit antenna, regardless of other parameters. This is a new result that is reminiscent of, but distinct from, the well-known water-filling result at low SNR.

Finally, motivated by the symmetry inherent in the problem, we analyze the receive side antenna selection in a manner similar to transmit selection. Even if no multiplexing is lost due to selection,

[1] We assume there are more transmit than receive antennas.

[2] Terminology due to Verdu [1]

receive antenna selection incurs a loss of recieve power due to de-selected antennas, thus, unlike transmit selection which may increase capacity (under conditions mentioned earlier), receive selection always incurs a capacity loss.

In each of the low- and high-SNR regimes, we focus on three important cases: full CSI at transmitter, no CSI at transmitter, and antenna-selection CSI at transmitter. In the latter case, capacity analysis naturally depends on the antenna selection algorithm. Optimal antenna selection is not only practically difficult, it also induces channel distributions that do not lead to a tractable formulation. Thankfully there exist antenna selection algorithms that deliver capacities almost indistinguishable from optimal selection [11], [8]. In this paper we undertake the analysis of antenna selection using these algorithms.

We use the following 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 throughout this paper so the capacity unit is in Nats/Sec/Hz. The chi-square distribution with $2p$ degrees of freedom is shown by χ^2_{2p}, and the maximum of n independent χ^2_{2p} distributions is denoted by $\widetilde{\chi}^2_{2p,n}$. With an abuse of notation, we show random variables following the latter distribution also with $\widetilde{\chi}^2_{2p,n}$.

II. 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:

$$y(t) = Hx(t) + n(t) \tag{1}$$

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 with power constrain $\mathbb{E}[x^H x] \leq \rho$, where 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.

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

Fig. 1. Antenna selection algorithm

III. Transmit Antenna Selection

We consider a transmit antenna selection scheme where a subset of transmit antennas are used for transmission with equal power. Optimal transmit antenna selection via exhaustive search among all $\begin{pmatrix} M \\ L \end{pmatrix}$ 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 [10] 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 [12] 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.

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). This algorithm, shown in Figure 1, is the basis for the following analysis.

A. A Framework for High SNR Analysis

Using the Sherman-Morisson formula for determinants [17], for the selected channel $\widetilde{H}$ we have:

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

As $\rho \to \infty$, $\widetilde{P}_i \to P_i$, where, $P_i = I_N - \widetilde{H}_{i-1}(\widetilde{H}_{i-1}^H \widetilde{H}_{i-1})^{-1}\widetilde{H}_{i-1}$ 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. Our simulations verify that for large M the Gaussianity assumption provides a good approximation for the actual distribution of the remaining columns.[3] Using this assumption we can approximate the statistics of the right side of (2). We know that for an uncorrelated complex Gaussian vector x and a projection matrix P, $x^H P x$ has χ^2 distribution with $rank(P)$ degrees of freedom. Hence for large ρ and large M, we have:

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

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 [18]:

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

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

B. 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 [19]:

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

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

[3] We have no formal proof for this, however, and even for very simple cases it is an open problem and there is no known analytical result.

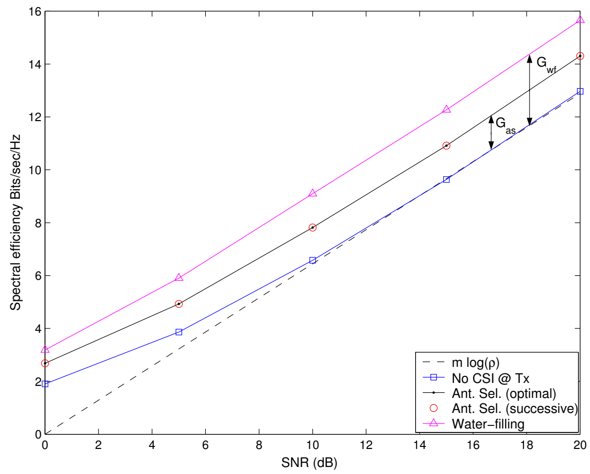

Fig. 2. Capacity gain of antenna selection: M=8 N=L=2

available at the transmitter is the indices of the selected transmit antennas. Figure 2 shows the ergodic capacity and the capacity gain for the above three cases.

Uninformed transmitter: As shown in [19], when CSI is available only at the 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 [19], and

$$m = \text{rank}(H) = \min\{M, N\}$$

is the degrees of freedom of the MIMO channel, assuming channel is full-rank. As shown in [20], $C = m \log \rho + O(1)$. 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:

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

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 [19] the 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:

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

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:

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

IV. 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 [20] that $\det(HH^H) \sim \prod_{i=1}^{N} \chi^2_{2(M-i+1)}$ therefore [21]:

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

where $\psi(n) = -\gamma + \sum_{k=1}^{n-1} \frac{1}{k}$ is the *di-gamma function*. We have [20]:

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

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

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

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

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

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 [18]:

Definition 1: 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

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

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:

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

where W is a random variable whose cdf is U. Moreover U can only be one of the following three distributions:

$$U_1(x) = \exp(-e^{-x}) \quad -\infty < x < \infty \ ,$$

$$U_2(x) = \begin{cases} \exp(-x^{\alpha}) & x > 0, \alpha > 0 \\ 0 & x \le 0 \end{cases},$$

$$U_3(x) = \begin{cases} 1 & x > 0 \\ \exp(-(-x)^{\alpha}) & x \le 0, \alpha > 0 \end{cases},$$

these distributions are also known as *Gumbel, Fréchet* and *Weilbull* distributions, repectively.

Theorem 1 is analogue to the *central limit theorem* which was for the normalized sum of iid random variables. Although unlink the central limit theorem, the normaliziation contrants a_n and b_n are not necessarily unique. One way to find constant a_n is to solve the following equation [18]:

$$\begin{aligned} a_n &= F^{-1}(1 - \frac{1}{n}) \\ b_n &= F^{-1}(1 - \frac{1}{ne}) - F^{-1}(1 - \frac{1}{n}) \end{aligned}$$

where F^{-1} is the inverse function. 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:

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

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.

Theorem 2: Let $\{X_n\}_{i=1}^n$ be positive random variables with finite mean and variance, $X_{(n)}$, W, $\{a_n\}$ and $\{b_n > 0\}$ are defined as in Theorem 1 and furthermore, $\frac{|a_n|}{b_n} \to \infty$, then as $n \to \infty$:

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

Proof: See Appendix.

So to calculate the logarithmic moment it is sufficient to only evaluate the mean of the above random variable. It is known that when the limiting distribution is of the first kind (Gumbel distribution) then Theorem 1 can also be used to evaluate the moments of $X_{(n)}$ (in fact for this case, convergence in distribution implies convergence in moments [22]). In particular we have

$$\mathbb{E}\left[\frac{X_{(n)} - a_n}{b_n}\right] \quad \to \quad \mathbb{E}[W] = \gamma \tag{21}$$

$$\mathbb{E}\left[\left(\frac{X_{(n)} - a_n}{b_n}\right)^2\right] \quad \to \quad \mathbb{E}[W^2] = \frac{\pi^2}{6} \tag{22}$$

Hence, the asymptotic growth of the logarithmic moment of the extreme order statistics of chi-square random variables is given by

$$\begin{aligned} \mathbb{E}[\log(X_{(n)})] &\overset{\circ}{=} \log(\mathbb{E}[X_{(n)}]) \\ &\overset{\circ}{=} \log\left(\log n + \log(\log(\frac{n^{p-1}}{(p-1)!})) + \gamma\right) \\ &\overset{\circ}{=} \log(\log n) \end{aligned} \tag{23}$$

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

$$\begin{aligned} \widetilde{G} &\overset{\circ}{=} \sum_{i=1}^{N} \mathbb{E}[\log(\tilde{\chi}^2_{2(N-i+1),M-i+1})] - N \log N \\ &\overset{\circ}{=} \sum_{i=1}^{N} \log\left(\frac{\log(M-i+1) + \log(\log(\frac{(M-i+1)^{N-i}}{(N-i)!})) + \gamma}{N}\right) \end{aligned} \tag{24}$$

Hence

$$\widetilde{G} \overset{\circ}{=} N \log(\log M)$$

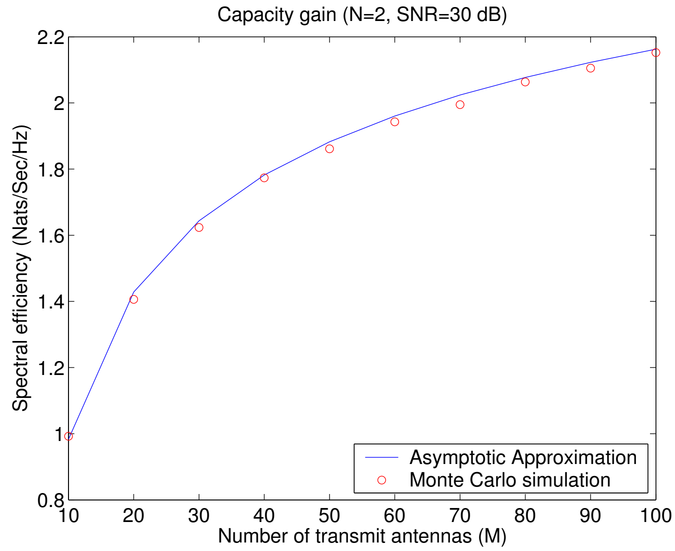

Fig. 3. Capacity gain of antenna selection (N=2 and SNR=30 dB)

Thus the capacity gain for transmit antenna selection behaves like $O\left(\log(\log M)\right)$. Also, for large number of transmit antennas Eq. 24 can be used as an approximate formula for ergodic capacity of transmit antenna selection at high SNR, in fact

$$\widetilde{C} \approx N \log \rho + \widetilde{G} \tag{25}$$

Figure 3 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 (24). Simulations match our analysis very well, thus the asymptotic formula proves to be a useful tool for the evaluating of the capacity of transmit antenna selection.

V. Transmit Antenna Selection: low SNR case

For low SNR case, we use the concept of *channel gain*, introduced by Verdu [1], which is essentially the slope of the linear term in the Taylor expansion of the ergodic capacity, namely,

$$\Gamma \triangleq \left. \frac{\partial C(\rho)}{\partial \rho} \right|_{\rho=0} \tag{26}$$

since $C = \Gamma\ \rho + O(\rho^2)$, Γ can be considered as an information theoretic metric for evaluating the spectral efficiency at low SNR [1].

Uninformed transmitter: In this case the channel gain [1] is:

$$\Gamma = \mathbb{E}\left[\frac{\|H\|_F^2}{M}\right] = N \tag{27}$$

we notice that in this case *the channel gain* does not depend on M, so increasing the number of transmit antenna will not affect the capacity.

Informed transmitter: When CSI is fully provided at the transmitter, at low SNR, the beamformer only uses the eigen mode of the channel associated with the largest eigen value of $H^H H$ hence the channel gain is [1]:

$$\Gamma_{wf} = \mathbb{E}[\lambda_{\max}(H^H H)] \tag{28}$$

It was first shown in [23] that $\lambda_{\max} \overset{\circ}{=} (\sqrt{M}+\sqrt{N})^2$ when M or N are large. Thus for the case $M \gg N$, we have $\Gamma_{wf} \overset{\circ}{=} M$.

Antenna selection: Suppose we are selecting L transmit antennas with equal power splitting among them. In low SNR scenario, the channel gain is:

$$\widetilde{\Gamma} = \mathbb{E}\left[\frac{\|\widetilde{H}\|_F^2}{L}\right] = \frac{\mathbb{E}[\sum_{i=1}^{L}\|\widetilde{h}_i\|_2^2]}{L}$$

where $\widetilde{H}$ is the selected channel, and $\widetilde{h}_1, \dots, \widetilde{h}_L$ are the L columns of $\widetilde{H}$. Thus antenna selection in low SNR case leads to selecting L antennas with highest norm. This is also consistent with the successive antenna selection algorithm presented in Figure 1 for $\rho \approx 0$. Moreover, the channel gain is maximized when $L = 1$, because $\frac{\mathbb{E}[\sum_{i=1}^{L}\|\widetilde{h}_i\|_2^2]}{L} \leq \max_i\{\|\widetilde{h}_i\|_2^2\}$. This suggests that the optimal transmit antenna selection strategy in low SNR case is to select only one transmit antenna whose channel vector is of the highest norm. In other words $\widetilde{H} = h_j$ where $\|h_j\| = \max\{\|h_1\|, \dots, \|h_M\|\}$. Now in order to evaluate the channel gain for low SNR transmit antenna selection we need to evaluate $\mathbb{E}[\|\widetilde{H}\|_F^2]$. The random variable $\|\widetilde{H}\|_F^2$ is distributed according to $\widetilde{X}_{2N,M}$ defined in Section III-A. For $M \gg N$, using Theorem 1 we have:

$$\widetilde{\Gamma}_{opt} \overset{\circ}{=} \log M + \log(\log(\frac{M^{N-1}}{(N-1)!})) + \gamma \overset{\circ}{=} \log M \tag{29}$$

VI. Receive Antenna Selection

Up to this point, the emphasis of the paper has been on the transmitter side. However, it is not difficult to see that due to the reciprocity of electromagnetic propagation, the problem is highly symmetric, therefore we can use the framework developed thus far to also address receive antenna selection.

In particular, consider the following setup: A MIMO system with M transmit and N receive antennas, such that $N \gg M$. We wish to choose L out of N receive antennas in a way to maximize the retained capacity. The algorithm depicted in Figure 1 will perform the antenna selection, with the notable difference that we now must select *rows* and not columns of H. As before, we call the selected channel $\widetilde{H}$. Assuming receive antenna selection with $L = M$ and no CSI at transmitter, the capacity of the system is:

$$C = \log\det(I_L + \frac{\rho}{M}\widetilde{H}\widetilde{H}^H)$$

Similarly to the previous case, we can write:

$$\begin{aligned}\log\det(I + \frac{\rho}{M}\widetilde{H}\widetilde{H}^H) &\approx \log\prod_{i=1}^{M}(1 + \frac{\rho}{M}\widetilde{\chi}^2_{2M,N-i+1}) \\ &\approx M\log\rho - M\log M + \sum_{i=1}^{M}\log\widetilde{\chi}^2_{2M,N-i+1} \\ &= M\log\rho + \hat{G}\end{aligned}$$

where $\widetilde{\chi}^2_{2M,N-i+1}$ is as earlier defined, the maximum of $N - i + 1$ chi-square random variables.

Recall that in transmit selection the SNR scales by the factor ρ/L in Equation (2), i.e., the fewer the selected antennas, the more power can be sent through each antenna. The result was that the capacity actually increases through transmit antenna selection,[4] an increase that was characterized by $\widetilde{G}$.

In receive selection, selecting fewer antennas will result in smaller receive power, but antennas that are selected enjoy better channel distributions than the original MIMO channel. However, the loss of power cannot be made up by the improvement in the channel distributions. Therefore the capacity of the receive antenna selection is less than the capacity of the full-scale system. Unlike transmit selection, no additional capacity is obtained by selecting down to the best antennas, a result that is not surprising because information is being lost by receive selection. The above results are demonstrated on a 2×8 system in Figure 4.

In the low-SNR regime, we once again use the concept of channel gain, which for the receive antenna selection leads to:

$$\widetilde{\Gamma} = \mathbb{E}\left[\frac{\|\widetilde{H}\|_F^2}{M}\right] = \mathbb{E}\left[\frac{\sum_{i=1}^{L}\|\widetilde{h}_i\|_2^2}{M}\right]$$

[4]Assuming the multiplexing gain of the system is unaffected by selection, which we ensured via our transmit-selection assumption of $N \geq L \geq M$

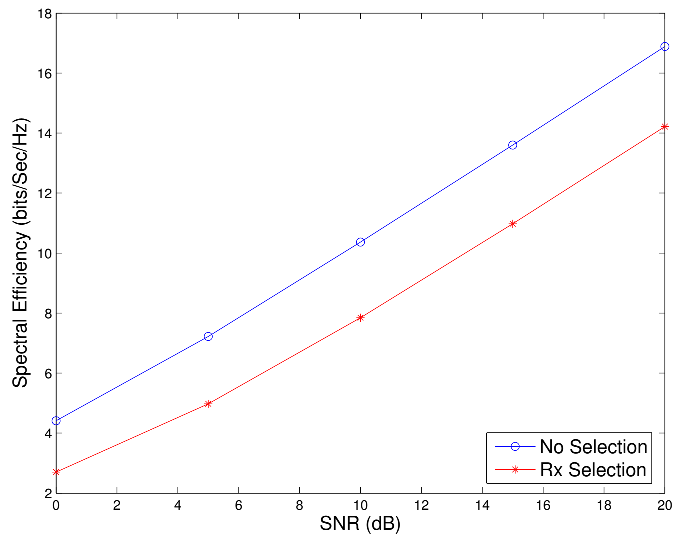

Fig. 4. Receive antenna selection in a 2×8 system

It is evident that, unlike the transmit-selection case, there is no penalty for selecting more antennas, in fact the more antennas selected, the higher the capacity. Calculating this value using Theorem 1 we have

$$\widetilde{\Gamma} \overset{\circ}{=} \frac{L}{M} \log N \tag{30}$$

VII. Conclusion

In this paper, we investigate the behavior of the capacity of antenna selection in the asymptote of large number of transmit antennas. For high SNR case, we introduce the concept of capacity gain, an information theoretic metric for schemes that use channel state information at the transmitter. It is shown that this quantity describes the advantage gained by having channel state information at the transmitter. We present new results in order statistics that are useful for asymptotic analysis of antenna selection problems. By exploring the behavior of the capacity gain, we show 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. For low SNR case, we first show that the optimal selection strategy is to select only one transmit antenna with the highest channel norm. Also, we evaluate the channel gain and show its logarithmic behavior in the asymptote of large number of antennas.

VIII. Appendix

Lemma 1: If $\frac{|a_n|}{b_n} \to \infty$, then $\log\left(\frac{X_{(n)}}{a_n}\right) \xrightarrow{d} 0$.

Proof: 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 $\frac{X_{(n)}}{a_n} \xrightarrow{i.p.} 1$, hence $\frac{X_{(n)}}{a_n} \xrightarrow{d} 1$. Now using the *continous mapping theorem* [24] we conclude that $\log\left(\frac{X_{(n)}}{a_n}\right) \xrightarrow{d} 0$.

Lemma 2: Let $\mu_n = \mathbb{E}[X_{(n)}]$, if $\frac{|a_n|}{b_n} \to \infty$, then $\log\left(\frac{\mu_n}{a_n}\right) \longrightarrow 0$.

Proof: From Eq. (21) we have $\frac{\mu_n - a_n}{b_n} \longrightarrow 0$, also we have $\frac{b_n}{a_n} \longrightarrow 0$. By multiplying two sides we get $\frac{\mu_n}{a_n} - 1 \longrightarrow 0$, hence $\log\left(\frac{a_n}{\mu_n}\right) \longrightarrow 0$.

Definition 2: The sequence of random varibales $\{X_n\}$ is called *uniformly integrable* if [24]:

$$\lim_{c\to\infty} \limsup_n \int_{|X_n|>c} |x| dF_{X_n}(x) < \infty \tag{31}$$

Lemma 3: If the random variables $\{X_n\}$ have finite mean, then the uniform integrability is equivalent to the following condition:

$$\lim_{c\to\infty} \limsup_n \int_{|X_n|>c} \Pr[|X_n| > x] dx < \infty \tag{32}$$

Proof: Using integration by part we have

$$\int_{|X_n|>c} |x| dF_{X_n}(x) = -x(1 - F_{X_n}(x)) \,|_c^\infty + \int_c^\infty \Pr[|X_n| > x] dx \tag{33}$$

Since $\mathbb{E}[X_n] < \infty$, we have $\lim_{c\to\infty} c(1 - F_{X_n}(c)) = 0$ and this proves the lemma.

Theorem 3: If the sequance of random variables $\{X_n\}$ is uniformly integrable, then $X_n \xrightarrow{d} X$ implies $\mathbb{E}[X_n] \longrightarrow \mathbb{E}[X]$.

Proof: See [24].

Proof of Theorem 2: Let $Y_n = \log\left(\frac{X_n}{a_n}\right)$, then by Lemma 1 we have $Y_n \xrightarrow{d} 0$, we show that Y_n is uniformly integrable, and to do so we use the alternative form of uniform integrability provided by Lemma 2:

$$\begin{aligned} \int_c^\infty \Pr[|Y_n| > y] dy &= \int_c^\infty \Pr[Y_n > y] dy + \int_c^\infty \Pr[Y_n < -y] dy \\ &= I_1 + I_2 \end{aligned} \tag{34}$$

We evaluate each integral,

$$\begin{aligned} I_1 &= \int_c^\infty \Pr[Y_n > y] dy \\ &= \int_c^\infty \Pr[X_n > a_n e^y] dy \\ &= \int_c^\infty \Pr\left[\frac{X_n - a_n}{b_n} > \frac{a_n}{b_n}(e^y - 1)\right] dy \\ &= \int_\alpha^\infty \Pr\left[\frac{X_n - a_n}{b_n} > \frac{a_n}{b_n} t\right] \frac{dt}{t+1} \qquad , \quad t := e^y - 1 \ , \ \alpha = e^c - 1 \end{aligned} \tag{35}$$

Let $\eta_n \triangleq \frac{a_n}{b_n}$ therefore $\eta_n \to \infty$ as $n \to \infty$. From Eq. (22) we have $\mathbb{E}\left[\left(\frac{X_n - a_n}{b_n}\right)^2\right] \longrightarrow \frac{\pi^2}{6}$. Thus using Chebyshev's inequality, for every $\delta > 0$ we have:

$$\begin{aligned} I_1 &\leq \frac{1}{\eta_n^2} \int_\alpha^\infty \frac{\mathbb{E}\left[\left(\frac{X_n - a_n}{b_n}\right)^2\right]}{t^2(t+1)} dt \\ &\leq \frac{1}{\eta_n^2} \int_\alpha^\infty \frac{\pi^2/6 + \delta}{t^2(t+1)} dt \\ &\leq \frac{1}{\eta_n^2(\alpha+1)} \int_\alpha^\infty \frac{\pi^2/6 + \delta}{t^2} dt \\ &= \frac{\pi^2/6 + \delta}{\eta_n^2 \alpha(\alpha+1)} \longrightarrow 0 \end{aligned} \tag{36}$$

Also in a similar way for all $\delta > 0$ we have,

$$\begin{aligned} I_2 &= \int_c^\infty \Pr[Y_n < -y] dy \\ &= \int_c^\infty \Pr[X_n < a_n e^{-y}] dy \\ &= \int_c^\infty \Pr\left[\frac{a_n - X_n}{b_n} > \frac{a_n}{b_n}(1 - e^{-y})\right] dy \\ &\leq \left(\frac{b_n}{a_n}\right)^2 \int_c^\infty \frac{\mathbb{E}[\left(\frac{a_n - X_n}{b_n}\right)^2]}{(1 - e^{-y})^2} dy \\ &\leq \left(\frac{b_n}{a_n}\right)^2 \int_c^\infty \frac{\frac{\pi^2}{6} + \delta}{(1 - e^{-y})^2} dy \\ &= \left(\frac{b_n}{a_n}\right)^2 (\frac{\pi^2}{6} + \delta) \left(\frac{e^{-c}}{1 - e^{-c}} + \log(1 - e^{-c})\right) \longrightarrow 0 \end{aligned} \tag{37}$$

from Eq. (36) and (37) we conclude that $I_1 + I_2 \longrightarrow 0$ thus, Y_n is uniformly integrable hence Y_n converges in mean, namely, $\mathbb{E}[\log X_{(n)}] - \log a_n \longrightarrow 0$. On the other hand, by Lemma 2 we have $\log(\mathbb{E}[X_{(n)}]) - \log a_n \longrightarrow 0$ thus we have $\mathbb{E}[\log X_{(n)}] - \log(\mathbb{E}[X_{(n)}]) \longrightarrow 0$.

■

REFERENCES

[1] S. Verdu, “Spectral efficiency in wideband regime,” *IEEE Trans. on Information Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[2] S. Sanayei and A. Nosratinia, “Antenna selection in MIMO systems,” *IEEE Communications Magazine*, vol. 42, no. 10, pp. 68–73, Oct. 2004.

[3] 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.

[4] 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.

[5] 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.

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

[7] 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.

[8] S. Sanayei and A. Nosratinia, “Asymptotic capacity gain of transmit antenna selection,” in *Proc. WNCG Symposium*, Austin, TX, Oct. 2003.

[9] S. Sanayei and A. Nosratinia, “Asymptotic capacity analysis of transmit antenna selection,” in *Proc. of IEEE International Symposium on Information Theory*, Chicago, IL, June 2004, p. 241.

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

[11] A. Gorokhov, D. A. Gore, and A. J. Paulraj, “Receive antenna selection for MIMO spatial multiplexing: Theory and algorithms,” *IEEE Trans. on Signal Processing*, pp. 2796 –2807, Nov. 2003.

[12] M. Gharavi-Alkhansari and A. Greshman, “Fast antenna selection in MIMO systems,” *IEEE Trans. on Signal Processing*, vol. 52, no. 2, pp. 339 – 347, Feb. 2003.

[13] A. Gorokhov, D. Gore, and A. Paulraj, “Receive antenna selection for MIMO flat fading channels: theory and algorithms,” *IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2687–2696, Oct. 2003.

[14] M. Win, A. Molisch, and Winters, “Capacity of MIMO systems with antenna selection,” in *Proc. International Conference on Communications*, 2001, vol. 2, pp. 570–574.

[15] A. Lozano, A. Tulino, and S. Verdu, “High-SNR power offset in multi-antenna communication,” in *Proc. of IEEE International Symposium on Information Theory*, Chicago, IL, June 2004, p. 288.

[16] D. Bliss, K. Forsythe, A. Hero, and A. Yegulalp, “Environmental issues for MIMO capacity,” *IEEE Trans. on Signal Processing*, vol. 50, no. 9, pp. 2128–2142, September 2002.

[17] G. Kéri, “The Sherman-Morisson formula for the determinant and its application for optimizing the quadratic functions on conditinos sets given by extreme generators,” in *Optimization Theory: Recent Developments from Mátraháza*, F. Giannessi, P. Pardalos, and T. Rapcsàk, Eds. Kluwer, 2001.

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

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

[20] 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.

[21] 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," *IEEE Trans. on Signal Processing*, vol. 51, no. 11, pp. 2784–2795, Nov. 2003.

[22] J. Galambos, *Asymptotic Theory of Order Statistics*, Krieger Pub. Co., 1987.

[23] S. Geman, "A limit theorem for the norm of the random matrices," *Annals of probability*, vol. 8, pp. 252–261, 1980.

[24] P. Billingsley, *Probability and Measure*, Wiley, 1995.