Analysis of Multiple Antenna Wireless Links at Low SNR *

CHAITANYA RAO and BABAK HASSIBI
{rao, hassibi}@systems.caltech.edu

Department of Electrical Engineering
California Institute of Technology, Pasadena, CA 91125

Abstract

Wireless channels with multiple transmit/receive antennas are known to provide a high spectral efficiency both when the channel is known to the receiver, and when the channel is not known to the receiver if the signal-to-noise ratio (SNR) is high. Here we analyze such systems at low SNR, which may find application in sensor networks and other low power devices. The key point is that, since channel estimates are not reliable, it is usually not reasonable to assume that the channel is known at the receiver at low SNR. In this unknown channel case, we show that for reasonable input distributions, in particular all practical modulation schemes, the capacity is asymptotically quadratic in the SNR, ρ, and thus much less than the known channel case where it exhibits a linear growth in ρ. We show that under various signaling constraints, e.g., Gaussian modulation, unitary space-time modulation, and peak constraints, that mutual information is maximized by using a single transmit antenna. We also show that sending training symbols offer no advantage at low SNR and that the, per channel use, mutual information is linear in both the number of receive antennas and the channel coherence interval.

1 Introduction

Multiple antenna wireless systems have been shown to provide high capacity, exploiting the presence of fading in such channels. However, this is based on the premise that either the channel coefficients are known to the receiver, or that the signal-to-noise ratio (SNR) of the channel is high [1, 2, 3].

Wireless systems operating at low SNR (exhibiting weak signaling or in noisy environments) find increasing use in energy efficient devices such as sensor networks. Recent work on analyzing the capacity of low SNR multi-antenna links, assuming that the channel is known at the receiver, has appeared in [4]. However, at low SNR channel estimates are typically unreliable and so in such a regime it is sensible to assume that the channel is unknown. In the following analysis we therefore assume the channel is unknown to both transmitter and receiver. As shown below, this leads to results qualitatively different from the known channel case.

We use the block fading model of a wireless multiple antenna system proposed by Marzetta and Hochwald in [5], expressing the mutual information between input and output as a function of the model parameter ρ (proportional to the SNR) up to second order. This model is described in detail in the next section. Maximizing this expression gives us insight about desired signaling at low SNR as well as the optimal number of antennas to be used at the transmitter and receiver. It has been shown in [6] that the optimum signaling at low SNR achieves the same capacity as the known channel cases for single transmit antenna systems. We show that the on-off optimal signaling found in [6] also generalizes to the multi-antenna setting and attains the known channel capacity. However, this scheme requires increasingly peaky signals (indeed ones with unbounded fourth order moment) and so is not acceptable from a practical point of view. We therefore focus our attention on signaling schemes with bounded fourth order moment.

Recent work by Verdú [7] has shown that knowledge of the first and second derivatives of capacity at low SNR also tells us about bandwidth and energy efficiency for signal transmission. For example, these quantities are used to see how spectral efficiency grows with energy-per-bit. In [8], the low-SNR mutual information of the same block fading

*This work supported in part by the National Science Foundation under grant no. CCR-0133818, by the office of Naval Research under grant no. N00014-02-1-0578, and by Caltech's Lee Center for Advanced Networking.

multiple antenna channel is also calculated. Similar results to ours have also been obtained in [9], as a by-product of their study of the capacity of general communication channels under small-peak constraints. Our results differ from [8] and [9] in two ways. First, we require a weaker assumption on the input signals; essentially conditions on the fourth and sixth order moments, rather than an exponentially-decaying input distribution as in [8], or a peak constraint as in [9], both of which render all moments finite. Second, we study the optimal signaling structure derived in [5] and further optimize mutual information subject to various signaling constraints such as training.

There are two main parts to this paper. In the first part we expand the mutual information of the wireless link to second order in the SNR ρ. Secondly, we optimize this expression under both peak and fourth order moment signal constraints to determine what signaling should be applied to the input and how many transmit antennas should be employed. We also study Gaussian modulation, unitary space-time modulation and training-based schemes.

2 Model

We consider a discrete-time block-fading channel model [5] in which there are M transmit and N receive antennas. The channel is described by a propagation matrix H that is assumed constant for a coherence interval of length T symbols. For the next T symbols the propagation matrix changes to a new independent value, and so on. Signals are represented by complex valued matrices with T rows, where the i-th row is what is transmitted or received via each of the multiple antennas at the i-th time slot.

For each coherence interval the $T \times N$ received matrix X is related to the $T \times M$ transmitted matrix S by

$$X = \sqrt{\frac{\rho}{M}} SH + V, \tag{1}$$

where H is $M \times N$ and V is a $T \times N$ noise matrix, both comprised of zero-mean and unit variance circularly symmetric complex Gaussian entries. The matrices H, V and S are assumed to be independent and the values of H and V are unknown to both transmitter and receiver. S satisfies the power constraint $\mathbf{E}\mathrm{tr}\mathrm{SS}^* = \mathbf{E}\sum_{\mathrm{i=1}}^{\mathrm{T}}\sum_{\mathrm{j=1}}^{\mathrm{M}}|\mathrm{s_{ij}}|^2 \triangleq \eta_2 \leq \mathrm{P_{max}}$, where $\mathbf{E}$ and tr denote the expectation and trace operators respectively, and s_{ij} is the (i,j)-th entry of S. When $\eta_2 = TM$ the normalization factor $\sqrt{\frac{\rho}{M}}$ in (1) makes ρ equal to the signal-to-noise ratio at each receive antenna. Otherwise the SNR at each receive antenna is given by $\mathbf{E}[\mathrm{trXX}^*]/\mathbf{E}[\mathrm{trVV}^*] = \rho\eta_2/(\mathrm{TM})$. It is also known that there is no performance gain in having the number of transmit antennas greater than T [5]. Hence we will assume that $T \geq M$.

Computing the capacity of this multiple antenna system for generic ρ is an open problem. In [5], however, it is shown that capacity is achieved when the input signal S has the form

$$S = \Phi D, \tag{2}$$

where D is a diagonal $M \times M$ matrix with non-negative entries and Φ is a $T \times M$ isotropically distributed unitary random matrix. This means $\Phi^*\Phi = I_M$ and the distribution of Φ is unaltered when left multiplied by a deterministic $T \times T$ unitary matrix [5]. Φ and D are independently distributed.

3 Mutual Information to 2nd order

In this section we prove the following result.

Theorem 1. *Consider the model (1) and let $p(S)$ denote the pdf of S.*

1. **First order result:** *If (i) $\partial p(S)/\partial\rho$ exists at $\rho = 0$ and (ii) $\lim_{\rho\to 0}\rho\mathbf{E}\mathrm{tr}(\mathrm{SS}^*)^2 = 0$, the mutual information between the transmitted and received signals S and X for the multiple antenna system (1) is zero to first order in ρ, i.e., $I(X;S) = o(\rho)$.*

2. **Second order result:** *If in addition (i) $\partial^2 p(S)/\partial\rho^2$ exists at $\rho = 0$, (ii) the fourth order moment of S is finite, i.e., $\mathbf{E}\mathrm{tr}(\mathrm{SS}^*)^2 \triangleq \eta_4 < \infty$ and (iii) $\lim_{\rho\to 0}\rho\mathbf{E}\mathrm{tr}(\mathrm{SS}^*)^3 = 0$, then the mutual information between S and X up*

to second order in ρ is given by

$$I(X;S) = \frac{N\mathrm{tr}[\mathbf{E}(\mathrm{SS}^*)^2 - (\mathbf{E}\mathrm{SS}^*)^2]}{2M^2}\rho^2 + o(\rho^2). \tag{3}$$

The second order part of the theorem is essentially a result in [8] and [9]. However, we here require a much less stringent condition on the input distribution. Moreover, we shall optimize (4) for various signaling schemes.

The reason for the condition on $p(S)$ in Theorem 1, is that the choice of distribution may depend on the SNR ρ. Condition (ii) of the first order result limits the growth of the fourth order moment, whereas conditions (ii) and (iii) of the second order result, respectively bound and limit the growth of the fourth and sixth order moments. The regularity conditions (i) on $p(S)$ at $\rho = 0$ are required for reasons that will be seen shortly (see sections 3.2 and 5).

For the optimum signaling structure (2), Eq. (3) can be replaced by

$$I(X;S) = \frac{N(\eta_4 - \eta_2^2/T)}{2M^2}\rho^2 + o(\rho^2). \tag{4}$$

Note that under any reasonable input distribution (and certainly all practical modulation schemes) the mutual information has no linear term in ρ and so the capacity is *much less* than the known channel case where the low SNR expansion of the well known $\log\det$ formula has a non-zero first order term. Since η_4 and η_2 are independent of N, (4) suggests that the capacity increases linearly in the number of receive antennas. The dependence of the mutual information on M is more complicated since both the denominator (M^2), as well as the numerator (via η_2 and η_4) depend on M. However, careful analysis will show that for most practical signal constraints the optimal value is $M = 1$ transmit antenna.

Finally, note that the mutual information is affine linear in η_4 suggesting that it increases as the input becomes more peaky, in good agreement with the results of [6] and their multi-antenna generalizations.

3.1 Conditional entropy approximation

We compute $I(X;S) = h(X) - h(X|S)$ via the conditional probability density function $p(X|S)$. Given S, X is zero-mean complex Gaussian with covariance $\mathbf{E}(XX^*|S) = N(I_T + \frac{\rho}{M}SS^*)^{-1}$ and so as in [5],

$$p(X|S) \quad = \quad \frac{e^{-\mathrm{tr}\mathrm{X}^*(\mathrm{I_T}+\frac{\rho}{\mathrm{M}}\mathrm{SS}^*)^{-1}\mathrm{X}}}{\pi^{NT}\left[\det(I_T + \frac{\rho}{M}SS^*)\right]^N}. \tag{5}$$

Here I_T denotes a $T \times T$ identity matrix. From $p(X|S)$ it is possible to compute the conditioned entropy $h(X|S)$ directly:

$$
\begin{aligned}
h(X|S) &= -\mathbf{E}\log p(X|S) \\
&= NT\log\pi + \mathbf{E}\log\det(I_T + \frac{\rho}{M}SS^*)^N + \mathbf{E}\mathrm{tr}\mathrm{X}^*(\mathrm{I_T} + \frac{\rho}{\mathrm{M}}\mathrm{SS}^*)^{-1}\mathrm{X} \\
&= NT\log\pi + N\mathbf{E}\log\det(I_T + \frac{\rho}{M}SS^*) + \mathbf{E}\mathrm{tr}(\mathrm{I_T} + \frac{\rho}{\mathrm{M}}\mathrm{SS}^*)^{-1}\mathrm{XX}^* \\
&\qquad \text{(using } \mathrm{trAB} = \mathrm{trBA}) \\
&= NT\log\pi + N\mathbf{E}\log\det(I_T + \frac{\rho}{M}SS^*) + \mathbf{E}_S\mathrm{trI_T} \\
&= NT\log\pi e + N\mathbf{E}\log\prod_i(1 + \frac{\rho}{M}d_i^2) \\
&\qquad \text{(where } d_i^2 \text{ are the eigenvalues of } SS^*) \\
&= NT\log\pi e + N\mathbf{E}\sum_i\log(1 + \frac{\rho}{M}d_i^2) && (6) \\
&\approx NT\log\pi e + N\mathbf{E}\sum_i(\frac{\rho}{M}d_i^2 - \frac{\rho^2}{2M^2}d_i^4) && (7) \\
&= NT\log\pi e + N\frac{\rho}{M}\eta_2 - \frac{N\rho^2}{2M^2}\eta_4, && (8)
\end{aligned}
$$

since $\eta_2 = \mathbf{E}\mathrm{trSS}^* = \mathbf{E}\sum_{\mathrm{i}}\mathrm{d_i^2}$ and $\eta_4 = \mathbf{E}\mathrm{tr}(\mathrm{SS}^*)^2 = \mathbf{E}\sum_{\mathrm{i}}\mathrm{d_i^4}$.

The approximation step (7) is made assuming that the second order approximation $\mathbf{E}\log(1 + \frac{\rho}{M}d_i^2) \approx \mathbf{E}[\frac{\rho}{M}d_i^2 - \frac{\rho^2}{2M^2}d_i^4]$ is valid for each i. Consider the inequality

$$
\begin{aligned}
\frac{\rho}{M}d_i^2 - \frac{\rho^2}{2M^2}d_i^4 \le \quad & \log(1 + \frac{\rho}{M}d_i^2) && \le \frac{\rho}{M}d_i^2 - \frac{\rho^2}{2M^2}d_i^4 + \frac{\rho^3}{3M^3}d_i^6 \\
\Rightarrow 0 \le \quad & \frac{\log(1 + \frac{\rho}{M}d_i^2) - (\frac{\rho}{M}d_i^2 - \frac{\rho^2}{2M^2}d_i^4)}{\rho^2} && \le \frac{\rho}{3M^3}d_i^6. \qquad (9)
\end{aligned}
$$

For the second order approximation to be a valid one the limit of the expression between the two inequalities in (9) should go to zero *in expectation* as $\rho \to 0$ for each i. The condition $\rho\mathbf{E}\mathrm{tr}(\mathrm{SS}^*)^3 = \rho\mathbf{E}\sum(\mathrm{d_i^2})^3 \to 0$ in the second order statement of Theorem 1 ensures that this occurs. The first order condition $\rho\mathbf{E}\mathrm{tr}(\mathrm{SS}^*)^2$ similarly ensures that $\rho d_i^2 \to 0$, making $\log(1 + \frac{\rho}{M}d_i^2) \approx \frac{\rho}{M}d_i^2$ a valid first order approximation.

3.2 Entropy approximation

The pdf $p(X)$ depends on the input distribution $p(S)$. Our regularity conditions (i) on $p(S)$ in Theorem 1 guarantee that the distribution can be expanded to second order around $\rho = 0$ as $p(S) = p(S,0) + \rho p'(S,0) + \frac{\rho^2}{2}p''(S,0) + o(\rho^2)$. Also $p(X|S)$ in (5) is a function whose derivatives with respect to ρ at $\rho = 0$ may be calculated. These two facts imply that

$$p(X) = \int p(S)p(X|S)\,dS$$

can also be expanded to second order as

$$p(X) = p(X,0) + \rho p'(X,0) + \frac{\rho^2}{2}p''(X,0) + o(\rho^2),$$

where $p'(X,0)$ and $p''(X,0)$ are used to denote the first and second partial derivatives of $p(X)$ with respect to ρ respectively, evaluated at $\rho = 0$. Also to second order

$$\log p(X) \quad \approx \quad \log p(X,0) + \rho\frac{p'(X,0)}{p(X,0)} + \frac{\rho^2}{2}\left[\frac{p''(X,0)}{p(X,0)} - \left(\frac{p'(X,0)}{p(X,0)}\right)^2\right].$$

This leads us to the following quadratic approximation

$$\begin{aligned} h(X) &= -\int p(X)\log p(X)\, dX \\ &\approx -\int p(X,0)\log p(X,0)\, dX - \rho\int p'(X,0)\log p(X,0) + p'(X,0)\, dX \\ &\quad -\frac{\rho^2}{2}\int p''(X,0) + (p'(X,0))^2/p(X,0) + p''(X,0)\log p(X,0)\, dX. \end{aligned} \tag{10}$$

We now claim that the integrals in (10) involving the second derivative $p''(X,0)$ are equal to zero. Firstly, note that

$$\int \left(p(X,0) + \rho p'(X,0) + \frac{\rho^2}{2}p''(X,0) + o(\rho^2)\right) dX = 1.$$

Comparing coefficients of ρ^n on both sides we conclude

$$\int p^{(n)}(X,0)\, dX = 0, \quad n = 1, 2. \tag{11}$$

Also, since S, H and V are independent (1) implies that

$$\begin{aligned} \mathbf{E}\mathrm{trXX^*} &= \mathbf{E}\mathrm{tr}\frac{\rho}{\mathrm{M}}\mathrm{SHH^*S^*} + \sqrt{\frac{\rho}{\mathrm{M}}}\mathbf{E}\mathrm{trSHV^*} + \sqrt{\frac{\rho}{\mathrm{M}}}\mathbf{E}\mathrm{trVH^*S^*} + \mathbf{E}\mathrm{trVV^*} \\ &= \frac{\rho}{M}\mathrm{tr}\mathbf{E}\mathrm{HH^*S^*S} + 0 + 0 + \mathrm{trNI_T} \\ &= \frac{\rho}{M}\mathrm{trNI_M}\mathbf{E}\mathrm{S^*S} + \mathrm{NT} \\ &= N(\rho\eta_2/M + T). \end{aligned} \tag{12}$$

Thus,

$$\int \mathrm{trXX^*}\left(\mathrm{p(X,0)} + \rho\mathrm{p'(X,0)} + \frac{\rho^2}{2}\mathrm{p''(X,0)} + \mathrm{o}(\rho^2)\right)\mathrm{dX} = \mathrm{N}(\eta_2/\mathrm{M}\rho + \mathrm{T}).$$

Comparing coefficients of ρ^n of on both sides

$$\begin{aligned} \int \mathrm{trX^*Xp(X,0)\,dX} &= NT \\ \int \mathrm{trX^*Xp'(X,0)\,dX} &= N\eta_2/M \\ \int \mathrm{trX^*Xp''(X,0)\,dX} &= 0. \end{aligned}$$

Now $p(X,0) = \frac{e^{-\mathrm{trX^*X}}}{\pi^{NT}}$ and so $\log p(X,0) = -\mathrm{trX^*X} - \mathrm{NT}\log\pi$. Hence the zero-th order term in (10) is

$$\begin{aligned} &\int p(X,0)\log p(X,0)\, dX \\ &= -\int p(X,0)\mathrm{trX^*X\,dX} - \mathrm{NT}\log\pi\int \mathrm{p(X,0)\,dX} \\ &= -NT - NT\log\pi \\ &= -NT\log\pi e. \end{aligned} \tag{13}$$

Similarly,

$$\int p'(X,0)\log p(X,0)\, dX = -N\eta_2/M$$

and

$$\int p''(X,0)\log p(X,0)\,dX = 0.$$

The above calculations combined with (10) lead to

$$h(X) \quad \approx \quad NT\log \pi e + N\eta_2/M - \frac{\rho^2}{2}\int (p'(X,0))^2/p(X,0)\,dX. \tag{14}$$

This shows that to express $h(X)$ to second order, it suffices to calculate only the first derivative of $p(X)$ at $\rho = 0$. We use the following result, proved in the appendix.

Lemma 1. *For the model (1), the first derivative of the pdf of X evaluated at $\rho = 0$ is given by*

$$p'(X,0) = \frac{e^{-\mathrm{trX^*X}}}{M\pi^{NT}}(\mathrm{trX^*PX} - \mathrm{N}\eta_2),$$

where $P = \mathbf{E}SS^*$.

This gives us

$$\begin{aligned}\int \frac{(p'(X,0))^2}{p(X,0)}\,dX &= \int \frac{e^{-\mathrm{trX^*X}}}{M^2\pi^{NT}}(\mathrm{trX^*PX} - \mathrm{N}\eta_2)^2\,\mathrm{dX} \\ &= \frac{1}{M^2}\left[\mathbf{E}_G(\mathrm{trG^*PG})^2 - 2\mathrm{N}\eta_2\mathbf{E}_{\mathrm{G}}\mathrm{trG^*PG} + \mathrm{N}^2\eta_2^2\right], \end{aligned} \tag{15}$$

where G is a $T\times M$ random matrix having the pdf $p(G) = \frac{e^{-\mathrm{trG^*G}}}{\pi^{NT}}$ and $\mathbf{E}_G$ denotes expectation over the random variable G.

To proceed we use the following lemma proved in the appendix.

Lemma 2. *If G is a $T\times N$ matrix with independent zero-mean unit-variance complex Gaussian entries, then*

- $\mathbf{E}_G\mathrm{trG^*PG} = \mathrm{N}\eta_2$,
- $\mathbf{E}_G(\mathrm{trG^*PG})^2 = \mathrm{N}^2\eta_2^2 + \mathrm{NtrP}^2$,

for any $T\times T$ deterministic matrix P satisfying $\mathrm{trP} = \eta_2$.

For $P = \mathbf{E}SS^*$, from (15) and Lemma 2 we have

$$\begin{aligned}\int \frac{(p'(X,0))^2}{p(X,0)}\,dX &= \frac{1}{M^2}\left[N^2\eta_2^2 + N\mathrm{tr}\left[(\mathbf{E}\mathrm{SS^*})^2\right] - 2\mathrm{N}\eta_2\mathrm{N}\eta_2 + \mathrm{N}^2\eta_2^2\right] \\ &= \frac{N}{M^2}\mathrm{tr}\left[(\mathbf{E}\mathrm{SS^*})^2\right].\end{aligned}$$

Hence $h(X) = NT\log\pi e + N\eta_2/M\rho - \frac{N}{2M^2}\mathrm{tr}\left[(\mathbf{E}\mathrm{SS^*})^2\right]\rho^2$ from (14) and this together with (8) gives

$$\begin{aligned} I(X;S) &= h(X) - h(X|S) \\ &\approx (NT\log\pi e + N\eta_2/M\rho - \frac{N}{2M^2}\mathrm{tr}\left[(\mathbf{E}\mathrm{SS^*})^2\right]\rho^2) \\ &\quad - (NT\log\pi e + N\eta_2/M\rho - \frac{N\eta_4}{2M^2}\rho^2) \\ &= \frac{N\mathrm{tr}\left[\mathbf{E}(\mathrm{SS^*})^2 - (\mathbf{E}\mathrm{SS^*})^2\right]}{2M^2}\rho^2, \end{aligned} \tag{16}$$

as stated in Theorem 1.

We remark that to show the first order result $I(X;S) = o(\rho)$ we only require first order expansions of $p(X)$ and $h(X)$, and so the conditions stated in the first order result of Theorem 1 suffice.

In the special capacity-optimizing case of $S = \Phi D$, we have

$$\begin{aligned}
\mathbf{E}(SS^*)_{ij} &= \mathbf{E}(\Phi D^2 \Phi^*)_{ij} \\
&= \mathbf{E} \sum_k \phi_{ik} d_k^2 \overline{\phi_{jk}} \quad (d_k^2 \text{ are both the diagonal entries of } D^2 \text{ and the eigenvalues of } SS^*) \\
&= \sum_k \mathbf{E}[d_k^2] \mathbf{E}\left[\phi_{ik}\overline{\phi_{jk}}\right] \quad \text{(since } \Phi \text{ and } D \text{ are independent).}
\end{aligned}$$

The expectation $\mathbf{E}\left[\phi_{ik}\overline{\phi_{jk}}\right]$ is evaluated by noticing that the expectation is unchanged by adding $T - M$ orthonormal columns to Φ to make it a $T \times T$ unitary, denoted say by $\Psi = (\psi_{ij})$. Then using the relation $\Psi\Psi^* = I_M$ we have

$$\sum_{k=1}^{T} \psi_{ik}\overline{\psi_{jk}} = \delta_{ij}.$$

Taking expectations of both sides and using the fact that each entry of Ψ has the same distribution, gives us

$$\mathbf{E}\left[\psi_{ik}\overline{\psi_{jk}}\right] = \frac{\delta_{ij}}{T} \quad \text{for } k = 1 \text{ to } T.$$

This implies $\mathbf{E}\left[\phi_{ik}\overline{\phi_{jk}}\right] = \frac{\delta_{ij}}{T}$ for $k = 1$ to M.

Hence

$$\begin{aligned}
\mathbf{E}(SS^*)_{ij} &= \frac{\delta_{ij}}{T} \sum_k \mathbf{E}[d_k^2] \\
&= \frac{\delta_{ij}}{T} \eta_2
\end{aligned}$$

In other words $\mathbf{E}SS^* = \frac{\eta_2}{T} I_T$, and so $\mathrm{tr}(\mathbf{E}\mathrm{SS}^*)^2 = \eta_2^2/\mathrm{T}$. This in (16) gives

$$I(X;S) = \frac{N(\eta_4 - \eta_2^2/T)}{2M^2}\rho^2 + o(\rho^2).$$

4 Examples

We now compute the low SNR mutual information for some cases of interest.

4.1 Gaussian modulation

Suppose the transmitted signal S has independent zero-mean unit-variance complex Gaussian entries. Then $(\mathbf{E}SS^*)_{ij} = \sum_k \mathbf{E}s_{ik}\overline{s_{jk}} = M\delta_{ij}$, so that $\mathbf{E}SS^* = MI_T$. In the appendix we show that for a Gaussian matrix, $\eta_4 = \mathbf{E}(\mathrm{tr}(\mathrm{SS}^*)^2) = \mathrm{MT}(\mathrm{M}+\mathrm{T})$ and so

$$\begin{aligned}
\frac{1}{T} I(X;S) &= \frac{N}{2M^2T}(MT(M+T) - (TM)^2/T) + o(\rho^2) \\
&= \frac{NT}{2M}\rho^2 + o(\rho^2).
\end{aligned}$$

This has two interesting ramifications. First the capacity per channel use increases linearly in T ($I(X;S)$ is actually quadratic in T) and, second, the optimal number of transmit antennas is $M = 1$.

4.2 Unitary space time modulation

In this scheme, we let $S = \Phi\sqrt{T}$ (where Φ has an isotropic unitary distribution), which gives $\eta_2 = TM$ and $\eta_4 = T^2 M$. Using this in (4) yields

$$\frac{1}{T}I(X;S) = \frac{N(T-M)}{2M}\rho^2 + o(\rho^2).$$

which is strictly less than the Gaussian case. Again, the optimal number of transmit antennas is $M = 1$.

4.3 Training-based schemes

In these schemes, we have

$$S = \left[\begin{array}{c} S_\tau \\ S_d \end{array}\right],$$

where S_τ is a $T_\tau \times M$ fixed training matrix and S_d is a $T_d \times M$ zero-mean random matrix. Furthermore

$$\text{tr}S_\tau^* S_\tau = \eta_{2,\tau},\ \mathbf{E}\text{tr}S_d S_d^* = \eta_{2,d},\ \eta_{2,\tau} + \eta_{2,d} = \eta_2,\ T_\tau + T_d = T.$$

Under these conditions it can be readily shown that

$$\text{tr}(\mathbf{E}SS^*)^2 = \text{tr}(S_\tau S_\tau^*)^2 + \text{tr}(\mathbf{E}S_d S_d^*)^2,$$

and

$$\text{tr}\mathbf{E}(SS^*)^2 = \text{tr}(S_\tau S_\tau^*)^2 + \text{tr}\mathbf{E}(S_d S_d^*)^2.$$

Therefore, using (3), we obtain

$$I(X;S) = \frac{N\text{tr}[\mathbf{E}(S_d S_d^*)^2 - (\mathbf{E}S_d S_d^*)^2]}{2M^2}\rho^2 + o(\rho^2). \tag{17}$$

The latter equation shows that the mutual information is independent of S_τ. Thus, in the low SNR regime, training does not help in increasing the mutual information. One may contrast this with the result of [10] which shows that training-based schemes achieve capacity at high SNR.

5 Optimal Signaling

In this section we shall optimize (4) to determine what kind of signaling should be applied to maximize the mutual information between the transmitted and received signals. It is known that the capacity for the known channel can be achieved to first order when the signal has arbitrarily large peaks with arbitrarily low probability ([6]).

We can show this is also the case for the multi-antenna channel as follows. For any $\epsilon > 1$ define our transmitted signal to satisfy

$$SS^* = \begin{cases} A & \text{w.p. } \rho^\epsilon \\ 0_{T\times T} & \text{w.p. } 1-\rho^\epsilon \end{cases}$$

where

$$A = T\rho^{-\epsilon}\left[\begin{array}{cc} I_M & 0_{M\times(T-M)} \\ 0_{(T-M)\times M} & 0_{(T-M)\times(T-M)} \end{array}\right] \tag{18}$$

Then S satisfies the power constraint $\mathbf{E}\mathrm{trSS^*} = \mathrm{tr(T\rho^{-\epsilon}I_M)} \times \rho^\epsilon = \mathrm{TM}$. Note that the above distribution does not satisfy the regularity conditions (i) of Theorem 1.

Then from (6),

$$
\begin{aligned}
h(X|S) &= NT\log\pi e + N\mathbf{E}\log\det(I_T + \frac{\rho}{M}SS^*) \\
&= NT\log\pi e + N\rho^\epsilon \log(1+\frac{\rho}{M}T\rho^{-\epsilon})^M \\
&\approx NT\log\pi e + N\rho^\epsilon M\log(\frac{T}{M}\rho^{1-\epsilon}) \quad \text{as } \rho^{1-\epsilon} \text{ is large} \\
&= NT\log\pi e + NM\rho^\epsilon[(1-\epsilon)\log\rho + \log(T/M)] \\
&= NT\log\pi e + o(\rho) \quad \text{since } \epsilon > 1.
\end{aligned}
$$

Also we have

$$
p(X) = \rho^\epsilon \frac{e^{-\mathrm{tr}\mathrm{X}^*\mathrm{A}^{-1}\mathrm{X}}}{\pi^{NT}\det(I_T+\frac{\rho}{M}A)^N} + (1-\rho^\epsilon)\frac{e^{-\mathrm{tr}\mathrm{X}^*\mathrm{X}}}{\pi^{NT}}. \tag{19}
$$

For ρ small and $\epsilon > 1$, ρ^ϵ is small while $\rho^{1-\epsilon}$ is large. Hence in the first term of (19) the determinant in the denominator is $(1+\frac{\rho}{M}T\rho^{-\epsilon})^{MN} \approx \rho^{MN(1-\epsilon)}$ which is large while the numerator is bounded above by 1. Therefore the second term is much larger than the first and so

$$
\begin{aligned}
h(X) &= -\mathbf{E}\log p(X) \\
&\approx -\mathbf{E}\log(1-\rho^\epsilon)\frac{e^{-\mathrm{tr}\mathrm{X}^*\mathrm{X}}}{\pi^{NT}} \\
&= -\log(1-\rho^\epsilon) + \mathbf{E}\mathrm{tr}\mathrm{X}^*\mathrm{X} + \mathrm{NT}\log\pi \\
&\approx \rho^\epsilon + NT\log\pi + \mathbf{E}\mathrm{tr}\mathrm{X}^*\mathrm{X} \\
&= NT\log\pi + NT(\rho+1) + \rho^\epsilon \quad \text{using (12)} \\
&= NT(\log\pi e + \rho) + o(\rho).
\end{aligned}
$$

Then $I(X;S)/T = N\rho + o(\rho)$, so the first order term corresponds to that of the capacity when the channel is known, equal to $\mathbf{E}\log\det(I+\frac{\rho}{M}HH^*) = N\rho + o(\rho)$. However, such signals cannot be used in practice and so we shall consider signals that are peak constrained or have a fourth order moment constraint.

5.1 Fourth order moment constraint

Suppose we enforce the fourth order moment constraint $\eta_4 \leq K\eta_2^2$ for some positive constant K. By the RMS-AM inequality we have

$$
\sum_{i=1}^M d_i^4 \;\; \geq \;\; \frac{1}{M}\left(\sum_{i=1}^M d_i^2\right)^2,
$$

from which we conclude $\eta_4 \geq \eta_2^2/M$. Also $T \geq M$ as stated in Section 2 and so $\eta_4 \geq \eta_2^2/T$. Hence we require that $K \geq 1/M > 1/T$.

Then $\eta_4 - \eta_2^2/T \leq (K-1/T)\eta_2^2$ and as $K - 1/T > 0$, from (4) it follows that maximizing the mutual information is equivalent to maximizing η_2. We therefore have the following result.

Theorem 2. *Consider the model (1) and suppose that the input signal must satisfy the constraints $\eta_2 \leq TM$ and $\eta_4 \leq K\eta_2^2$. Then, to second order, the mutual information is maximized by any input distribution that simultaneously achieves $\eta_2 = TM$ and $\eta_4 = K\eta_2^2$ and is given by*

$$
\frac{1}{T}I(X;S) = \frac{N(K-1/T)T}{2}\rho^2 + o(\rho^2). \tag{20}
$$

One distribution that achieves this is given by (2) where

$$(d_1^2, d_2^2, \ldots, d_M^2) \quad = \quad \begin{cases} (TKM, TKM, \ldots, TKM) & \textit{w.p. } 1/(KM) \\ (0, 0, \ldots, 0) & \textit{w.p. } 1 - 1/(KM). \end{cases}$$

Note here that the optimal mutual information (per channel use) is independent of the number of transmit antennas and is proportional to both N and T.

5.2 Peak constraint

Due to the isotropic unitary matrix in (2) is not possible to directly enforce a peak constraint on the transmitted signals. However, it is possible to force a peak constraint on the diagonal entries of D. Thus, assume that $d_i^2 \le K$ for some positive constant K and for all i. For any fixed M, we wish to maximize $\eta_4 - \eta_2^2/T$ subject to the constraint $\eta_2 \le P_{\max}$. We also have $\eta_2 = \mathbf{E}\sum_{i=1}^M d_i^2 \le MK$. Now

$$\begin{aligned} \eta_4 - \eta_2^2/T &= \mathbf{E}\left[\sum_{i=1}^M d_i^4\right] - \eta_2^2/T \\ &\le K\mathbf{E}\left[\sum_{i=1}^M d_i^2\right] - \eta_2^2/T \qquad (21) \\ &= \eta_2(K - \eta_2/T), \end{aligned}$$

with equality iff all the d_i's are equal to 0 or K. This quantity is maximized at either $\eta_2 = TK/2$, $\eta_2 = P_{\max}$ or $\eta_2 = MK$, depending on which of the three quantities $TK/2$, $P_{\max}$ or MK is smallest. This leads to

$$\eta_4 - \eta_2^2/T \quad \le \quad \begin{cases} TK^2/4, & \text{if } L = TK/2 \\ P_{\max}(K - P_{\max}/T), & \text{if } L = P_{\max} \\ MK^2(T-M)/T, & \text{if } L = MK. \end{cases} \qquad (22)$$

where $L = \min\{TK/2, P_{\max}, MK\}$. Equality holds in (22) when η_2 is set to $\min\{TK/2, P_{\max}, MK\}$. Corresponding distributions that achieve equality are

$$(d_1^2, d_2^2, \ldots, d_M^2) \quad = \quad \begin{cases} (K, K, \ldots, K) & \text{w.p. } \min\{1, T/2M\} \\ (0, 0, \ldots, 0) & \text{w.p. 1-}\min\{1, T/2M\}, \end{cases} \qquad (23)$$

and

$$(d_1^2, d_2^2, \ldots, d_M^2) \quad = \quad \begin{cases} (K, K, \ldots, K) & \text{w.p. } \min\{1, P_{\max}/MK\} \\ (0, 0, \ldots, 0) & \text{w.p. 1 - }\min\{1, P_{\max}/MK\}, \end{cases} \qquad (24)$$

depending on whether $P_{\max} \ge TK/2$ or $P_{\max} < TK/2$ respectively. The mutual information bounds are

$$I(X;S) \quad \le \quad \begin{cases} NTK^2\rho^2/8M^2, & \text{if } L = TK/2 \\ NP_{\max}(K - P_{\max}/T)\rho^2/2M^2, & \text{if } L = P_{\max} \\ NK^2(T-M)\rho^2/2TM, & \text{if } L = MK. \end{cases}$$

Note that all the above bounds are decreasing functions of M. Therefore it is clear that the optimal choice of transmit antennas is $M = 1$. Since it is most likely that $K < P_{\max}$ (that is, $L = MK = K$ unless $T = 1$ in which case $L = TK/2 = K/2$), we have the following theorem.

Theorem 3. *In the model (1) with optimal signaling as in (2) suppose the diagonal entries d_i of D satisfy $d_i^2 < K$ for all i, where K is some constant less than $P_{\max}$. Then for asymptotically low SNR the optimal number of transmit* *antennas is $M = 1$ and the optimal mutual information is*

$$\frac{1}{T}I(X;S) \quad = \quad \begin{cases} \frac{NK^2(T-1)}{2T}\rho^2 + o(\rho^2) & \textit{if } T > 1, \\ \frac{NK^2}{8}\rho^2 + o(\rho^2) & \textit{if } T = 1. \end{cases}$$

One distribution that achieves this is given in (2) where the diagonal entries of D^2 are given by (23) or (24) depending on whether $P_{\max} \geq TK/2$ or $P_{\max} < TK/2$ respectively.

6 Conclusions

For the block fading multiple antenna channel model in which the channel is unknown to the transmitter and receiver, the low-SNR asymptotic mutual information has a quadratic leading term. This mutual information may be maximized by using one transmit antenna, many receive antennas, and with on-off signaling. When there is a peak constraint on the transmit signal, it is possible to obtain a higher capacity by lowering the signal power from its maximum allowed level.

Appendix

Here we prove the lemmas used in the main sections.

Lemma 1. *For the model (1), the first derivative of the pdf of X evaluated at $\rho = 0$ is given by*

$$p'(X,0) = \frac{e^{-\mathrm{trX^*X}}}{M\pi^{NT}}(\mathrm{trX^*PX - N}\eta_2),$$

where $P = \mathbf{E}SS^*$.

Proof. We first approximate $p(X|S)$ to first order in ρ. Expanding the numerator to first order:

$$\begin{aligned} e^{-\mathrm{trX^*(I_T+\frac{\rho}{M}SS^*)^{-1}X}} &\approx e^{-\mathrm{trX^*(I_T-\frac{\rho}{M}SS^*)X}} \\ &= e^{-\mathrm{trX^*X}}e^{\mathrm{tr(\frac{\rho}{M}X^*SS^*X)}} \\ &\approx e^{-\mathrm{trX^*X}}\left[1 + \frac{\rho}{M}\mathrm{trX^*SS^*X}\right]. \end{aligned}$$

To expand the denominator we use

$$\begin{aligned} \det(I + \rho A) &= e^{\log\det(\mathrm{I}+\rho\mathrm{A})} \\ &= e^{\mathrm{tr}\log(\mathrm{I}+\rho\mathrm{A})} \\ &\approx e^{\mathrm{tr}(\rho\mathrm{A})} \\ &\approx 1 + \mathrm{tr}(\rho\mathrm{A}) \end{aligned}$$

so that

$$\begin{aligned} \det(I_T + \frac{\rho}{M}SS^*)^{-N} &= \left[1 + \frac{\rho}{M}\mathrm{trSS^*}\right]^{-N} \\ &\approx 1 - \frac{\rho N}{M}\mathrm{trSS^*}. \end{aligned}$$

Putting everything together

$$p(X|S) \approx \frac{e^{-\mathrm{trX^*X}}}{\pi^{NT}}\left[1 + \frac{\rho}{M}\mathrm{trX^*SS^*X}\right]\left[1 - \frac{\rho N}{M}\mathrm{trSS^*}\right]$$

and by taking the coefficient of ρ of both sides it follows that

$$p'(X|S,\rho=0)=\frac{e^{-\mathrm{trX^*X}}}{M\pi^{NT}}(\mathrm{trX^*SS^*X}-\mathrm{NtrSS^*}) \tag{25}$$

To find $p'(X,0)$ we take the expectation of (25) over S, leading to the required result. □

Lemma 2. *If G is a $T\times N$ matrix with independent zero-mean unit-variance complex Gaussian entries, then*

- $\mathbf{E}_G\mathrm{trG^*PG}=\mathrm{N}\eta_2$
- $\mathbf{E}_G(\mathrm{trG^*PG})^2=\mathrm{N}^2\eta_2^2+\mathrm{NtrP}^2$,

for any $T\times T$ matrix P satisfying $\mathrm{trP}=\eta_2$.

Proof. 1. Denoting the (i,j)-th entries of G and P by g_{ij} and p_{ij} respectively we have

$$\begin{aligned}
E_G\mathrm{trG^*PG} &= \mathbf{E}_G\sum_{i,j,k}\overline{g_{ji}}p_{jk}g_{ki}\\
&= \sum_{i,j,k}p_{jk}\mathbf{E}_G[\overline{g_{ji}}g_{ki}]\\
&= \sum_{i,j,k}p_{jk}\delta_{jk}\\
&\quad \text{(since } g_{ki} \text{ and } g_{ji} \text{ are independent unless } j=k\text{, in which case the expectation is unity)}\\
&= N\sum_{j,k}p_{jk}\delta_{jk}\\
&= N\sum_{j}p_{jj}\\
&= N\mathrm{trP}
\end{aligned}$$

as required.

2. We have

$$\begin{aligned}
E_G(\mathrm{trG^*PG})^2 &= \mathbf{E}_G\left(\sum_{i,j,k}\overline{g_{ji}}p_{jk}g_{ki}\right)^2\\
&= \mathbf{E}_G\sum_{i,j,k,l,m,n}\overline{g_{ji}}p_{jk}g_{ki}\overline{g_{ml}}p_{mn}g_{nl}\\
&= \sum_{j,k,m,n}p_{jk}p_{mn}\sum_{i,l}\mathbf{E}_G\left[\overline{g_{ji}}g_{ki}\overline{g_{ml}}g_{nl}\right].
\end{aligned}$$

$\mathbf{E}[\overline{g_{ji}}g_{ki}\overline{g_{ml}}g_{nl}]$ will be non-zero only when terms pair up as $|g|^4$ or $|g_i|^2|g_j|^2$. This will occur in the following instances:

- $j=k=m=n, i=l$
- $j=k=m=n, i\neq l$
- $j=k\neq m=n$
- $j=n\neq k=m, i=l$

Using the Kronecker delta to indicate non-zero terms when all its subscripts are equal we have

$$
\begin{aligned}
\mathbf{E}_X(\mathrm{trG^*PG})^2 &= \sum_{j,k,m,n} p_{jk}p_{mn}\sum_{i,l}\mathbf{E}_X\left[\overline{g_{ji}}g_{ki}\overline{g_{ml}}g_{nl}\right] \\
&= \sum_{j,k,m,n} p_{jk}p_{mn}\sum_{i,l}\left(2\delta_{jkmn}\delta_{il}+\delta_{jkmn}(1-\delta_{il})+\delta_{jk}\delta_{mn}(1-\delta_{km})+\delta_{jn}\delta_{km}(1-\delta_{kn})\delta_{il}\right) \\
&= \sum_{j,k,m,n} p_{jk}p_{mn}\left[2\delta_{jkmn}N+\delta_{jkmn}(N^2-N)+\delta_{jk}\delta_{mn}(1-\delta_{km})N^2+\delta_{jn}\delta_{km}(1-\delta_{kn})N\right] \\
&= 2N\sum_j p_{jj}^2+(N^2-N)\sum_j p_{jj}^2+N^2\sum_{j,m}p_{jj}p_{mm}-N^2\sum_j p_{jj}^2+N\sum_{j,k}p_{jk}p_{kj}-N\sum_j p_{jj}^2 \\
&= N^2\sum_{j,m}p_{jj}p_{mm}+N\sum_{j,k}p_{jk}p_{kj} \\
&= N^2(\mathrm{trP})^2+\mathrm{NtrP}^2,
\end{aligned}
$$

completing the proof.

□

The following lemma was used when considering a Gaussian modulation scheme.

Lemma 3. *Let X be a $T\times M$ matrix with independent zero-mean unit-variance complex Gaussian entries. Then* $\mathbf{E}\mathrm{tr}(\mathrm{XX^*})^2 = \mathrm{MT(M+T)}$.

Proof. Denote the (i,j)-th entry of X by x_{ij}. Then as the *pdf* of the Gaussian matrix X is $p(X)=\frac{e^{-\mathrm{trX^*X}}}{\pi^{NT}}$ we have

$$
\begin{aligned}
\mathbf{E}\mathrm{tr}(\mathrm{X^*X})^2 &= \int \frac{e^{-\mathrm{trX^*X}}}{\pi^{MT}}\mathrm{tr}(\mathrm{XX^*})^2\,\mathrm{dX} \\
&= \int \frac{e^{-\mathrm{trX^*X}}}{\pi^{MT}}\sum_{i,j,k,l}x_{ij}\overline{x}_{kj}x_{kl}\overline{x}_{il}\,dX \\
&= \sum_{i,j,k,l}\int \frac{e^{-\mathrm{trX^*X}}}{\pi^{MT}}x_{ij}\overline{x}_{kj}x_{kl}\overline{x}_{il}\,dX.
\end{aligned}
$$

If both $i\neq k$ and $j\neq l$ the integral $\int\frac{e^{-\mathrm{trX^*X}}}{\pi^{MT}}x_{ij}\overline{x}_{kj}x_{kl}\overline{x}_{il}\,dX=0$ as the integrand is an odd function of the variable x_{ij}.

Using the elementary integrals $\int\frac{e^{-|x|^2}}{\pi}\,dx=\int\frac{e^{-|x|^2}}{\pi}|x|^2\,dx=1$, $\int\frac{e^{-|x|^2}}{\pi}|x|^4\,dx=2$ where the integrals are over $x\in\mathbb{C}$, we have

$$
\begin{aligned}
\int\frac{e^{-\mathrm{trXX^*}}}{\pi^{MT}}\mathrm{tr}(\mathrm{X^*X})^2\,\mathrm{dX} &= \sum_{\substack{i\\ j\neq l}}\int\frac{e^{-\mathrm{trX^*X}}}{\pi^{MT}}|x_{ij}|^2|x_{il}|^2\,dX+\sum_{\substack{j\\ i\neq k}}\int\frac{e^{-\mathrm{trX^*X}}}{\pi^{MT}}|x_{ij}|^2|x_{kj}|^2\,dX \\
&\quad+\sum_{i,j}\int\frac{e^{-\mathrm{trX^*X}}}{\pi^{MT}}|x_{ij}|^4\,dX \\
&= \sum_{\substack{i\\ j\neq l}}1+\sum_{\substack{j\\ i\neq k}}1+\sum_{i,j}2 \\
&= MT(T-1)+TM(M-1)+2MT \\
&= MT(M+T),
\end{aligned}
$$

as required. □

References

[1] G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs. Tech. J.*, 1(2):41–59, 1996.

[2] I. E. Telatar. Capacity of multi-antenna Gaussian channels. *Eur. Trans. Telecom.*, 10:585–595, Nov. 1999.

[3] L. Zheng and D. Tse. Communication on the Grassman manifold: a geometric approach to the noncoherent multiple-antenna channel. *submitted to IEEE Trans. Info. Theory*, 48:359–389, Feb. 2002.

[4] A. Lozano, A. Tullino, and S. Verdú. Multi-antenna capacity in the low-power regime. *Private communication*, 2002.

[5] T. L. Marzetta and B. M. Hochwald. Capacity of a mobile multiple-antenna communication link in flat fading. *IEEE Trans. Inform. Theory*, 45:139–157, 1999.

[6] I. C. Abou-Faycal, M. D. Trott, and S. Shamai. The capacity of discrete-time memoryless rayleigh-fading channels. *IEEE Trans. Inform. Theory*, 47:1290–1301, 2001.

[7] S. Verdú. Spectral efficiency in the wideband regime. *IEEE Trans. Inform. Theory*, 48:1319–1343, 2002.

[8] V. Prelov and S. Verdú. Second-order asymptotics of mutual information. *DIMACS Technical Report 2002-16*, April 2002.

[9] B. Hajek and V. Subramaniam. Capacity and reliability function for small peak signal constraints. *IEEE Trans. Info. Theory*, 48:828–839, Apr. 2002.

[10] B. Hassibi and B. Hochwald. How much training is needed in multiple-antenna wireless links? *to appear in the IEEE Trans. Info. Theory*, 2003. Download available at `http://mars.bell-labs.com`.