Now

$$\begin{aligned} g(t) &= \log\det(X+Z+tN) - \log\det(Z+tN) \\ &= \sum_{i=1}^{n} \left\{ \log(1+t\lambda_i(NY^{-1})) \right. \\ &\quad \left. - \log(1+t\lambda_i(NZ^{-1})) \right\} + \log\det(ZY^{-1}) \end{aligned}$$

so that

$$\begin{aligned} \frac{d^2 g}{dt^2} &= \sum_{i=1}^{n} \left\{ \frac{1}{\left(t+\frac{1}{\lambda_i(NZ^{-1})}\right)^2} - \frac{1}{\left(t+\frac{1}{\lambda_i(NY^{-1})}\right)^2} \right\} \\ &\geq 0 \end{aligned}$$

where for the last inequality we have used (15) and the fact that $t \in \mathcal{T}$. Strict inequality holds for $X \succ 0$.

REFERENCES

[1] T. W. Anderson, "The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities," *Proc. Amer. Math. Soc.*, vol. 6, pp. 170–176, 1955.
[2] T. Başar, "The Gaussian test channel with an intelligent jammer," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 152–157, Jan. 1983.
[3] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. Philadelphia, PA: SIAM, 1999.
[4] J. M. Borden, D. M. Mason, and R. J. McEliece, "Some information theoretic saddlepoints," *SIAM J. Control Optimiz.*, vol. 23, no. 1, pp. 129–143, 1985.
[5] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Inform. Theory*, vol. 47, pp. 3072–3081, Nov. 2001.
[6] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K: Cambridge Univ. Press, 1985.
[7] A. Kashyap, T. Başar, and R. Srikant, "Correlated jamming on MIMO Gaussian fading channels," in *Proc. IEEE Int. Conf. Communications (ICC)*, Paris, France, 2004, pp. 458–462.
[8] Y. H. Kim and A. Lapidoth, "On the log-determinant of the noncentral wishart distribution," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Yokohama, Japan, 2003, p. 54.
[9] M. Medard, "Capacity of correlated jamming channels," in *Proc. 35th Allerton Conf.*, 1997, pp. 1043–1052.
[10] W. E. Stark and R. J. McEliece, "On the capacity of channels with block memory," *IEEE Trans. Inform. Theory*, vol. 34, pp. 322–324, Mar. 1988.
[11] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 585–589, Nov./Dec. 1999.

Analysis of Multiple-Antenna Wireless Links at Low SNR

Chaitanya Rao, *Student Member, IEEE,* and Babak Hassibi

***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 often not reasonable to assume that the channel is known at the receiver at low SNR. In this unknown channel case, we show that for sensible 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 at low SNR, sending training symbols leads to a rate reduction in proportion to the fraction of training duration time so that it is best not to perform training. Furthermore, we show that the per-channel use mutual information is linear in both the number of receive antennas and the channel coherence interval.**

***Index Terms*—Low-signal-to-noise ratio (SNR) regime, multiple-antenna systems, noncoherent channels, Rayleigh fading.**

I. 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]–[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 multiple-antenna links, assuming that the channel is known at the receiver, has appeared in [4]. However, at low SNR, channel estimates in some circumstances are unreliable and so 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 later, 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 minimum energy per bit 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 multiple-antenna setting (a result that also follows from [7, Theorems

Manuscript received March 10, 2003; revised February 1, 2004. This work was supported in part by the National Science Foundation under Grant CCR-0133818, by the Office of Naval Research under Grant N00014-02-1-0578, and by Caltech's Lee Center for Advanced Networking.
The authors are with the Department of Electrical Engineering, California Institute of Technology, Pasadena CA 91125 USA (e-mail: rao@systems.caltech.edu; hassibi@systems.caltech.edu).
Communicated by G. Caire, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2004.833369

1 and 5]). However, this scheme requires increasingly peaky signals (indeed, ones with unbounded fourth-order moment) and so may not be acceptable from a practical point of view in some situations. 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. More work on constrained signaling in the low-power regime for Rayleigh-fading channels is given in [8], [9] and [10], while [11] and [12] study the Rician case. In [13], among other things, the low-SNR mutual information of the same block-fading multiple-antenna channel of [5] is also calculated. Similar results to ours have also been obtained in [14], as a by-product of their study of the capacity of general communication channels under small-peak constraints. Our results differ from [13] and [14] 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 [13], or a peak constraint on the singular values of the transmitted signal as in [14], 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 correspondence. In the first part, we expand the mutual information of the wireless link to second order in the SNR ρ using an approach that may be applied to other channel models. 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.

II. 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 ith row is what is transmitted or received via each of the multiple antennas at the ith 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 an $M \times N$ matrix 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

$$\mathbf{E}\mathrm{tr}SS^* = \mathbf{E}\sum_{i=1}^{T}\sum_{j=1}^{M}|s_{ij}|^2 \triangleq \eta_2 \leq P_{\max}$$

where $\mathbf{E}$ and tr denote the expectation and trace operators, respectively, and s_{ij} is the (i,j)th entry of S. Throughout this work, the * operator denotes the conjugate transpose of a matrix. When $\eta_2 = TM$, the normalization factor $\sqrt{\frac{\rho}{M}}$ in (1) makes ρ equal to the SNR at each receive antenna. Otherwise, the SNR at each receive antenna is given by $\mathbf{E}[\mathrm{tr}XX^*]/\mathbf{E}[\mathrm{tr}VV^*] = \rho\eta_2/(TM)$. 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 nonnegative entries and Φ is a $T \times M$ isotropically distributed unitary random matrix. This means

- $\Phi^*\Phi = I_M$ (the $M \times M$ identity matrix) although $\Phi\Phi^* \neq I_T$ for $T > M$, and
- the distribution of Φ is unaltered when left multiplied by a deterministic $T \times T$ unitary matrix or when right multiplied by a deterministic $M \times M$ unitary matrix [5].

Φ and D are independently distributed.

III. Mutual Information to Second Order

In this section we prove the following result.

Theorem 1: Consider the model (1) and let $p(S)$ denote the proobability density function (pdf) of S.

1) **First-order result:** If i) $\partial p(S)/\partial\rho$ exists at $\rho = 0$ and ii) $\lim_{\rho\to 0} \rho\mathbf{E}\mathrm{tr}(SS^*)^2 = 0$, 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., $\mathbf{E}\mathrm{tr}(SS^*)^2 \triangleq \eta_4 < \infty$, and iii) $\lim_{\rho\to 0} \rho\mathbf{E}\mathrm{tr}(SS^*)^3 = 0$, then the mutual information between S and X up to second order in ρ is given by

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

The second-order part of the theorem is essentially a result in [13] and [14]. 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 III-B and V).

For the optimum signaling structure (2), (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 nonzero 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 multiple-antenna generalizations.

A. Conditional Entropy Approximation

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

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

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

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

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

The approximation step (7) is made assuming that the second-order approximation

$$\mathbf{E}\log\left(1+\frac{\rho}{M}d_i^2\right) \approx \mathbf{E}\left[\frac{\rho}{M}d_i^2 - \frac{\rho^2}{2M^2}d_i^4\right]$$

is valid for each i. Consider the inequality

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

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

$$\rho\mathbf{E}\mathrm{tr}(SS^*)^3 = \rho\mathbf{E}\sum(d_i^2)^3 \to 0$$

in the second-order statement of Theorem 1 ensures that this occurs. The first-order condition $\rho\mathbf{E}\mathrm{tr}(SS^*)^2$ 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.

B. 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

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

This leads us to the following quadratic approximation:

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

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

First, 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

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

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

$$\begin{aligned}
\mathbf{E}\mathrm{tr}XX^* &= \mathbf{E}\mathrm{tr}\frac{\rho}{M}SHH^*S^* + \sqrt{\frac{\rho}{M}}\mathbf{E}\mathrm{tr}SHV^*\\
&\quad + \sqrt{\frac{\rho}{M}}\mathbf{E}\mathrm{tr}VH^*S^* + \mathbf{E}\mathrm{tr}VV^*\\
&= \frac{\rho}{M}\mathrm{tr}\mathbf{E}HH^*S^*S + 0 + 0 + \mathrm{tr}NI_T\\
&= \frac{\rho}{M}\mathrm{tr}NI_M\mathbf{E}S^*S + NT\\
&= N(\rho\eta_2/M + T). \quad (12)
\end{aligned}$$

Thus,

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

Comparing coefficients of ρ^n of on both sides

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

Now $p(X,0) = \frac{e^{-\mathrm{tr}X^*X}}{\pi^{NT}}$ and so

$$\log p(X,0) = -\mathrm{tr}X^*X - NT\log\pi.$$

Hence the zeroth-order term in (10) is

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

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 preceding calculations combined with (10) lead to

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

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

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

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

This gives us

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

where G is a $T \times M$ random matrix having the pdf $p(G) = \frac{e^{-\mathrm{tr}G^*G}}{\pi^{NT}}$ 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

- $\mathbf{E}_G\mathrm{tr}G^*PG = N\eta_2$;
- $\mathbf{E}_G(\mathrm{tr}G^*PG)^2 = N^2\eta_2^2 + N\mathrm{tr}P^2$;

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

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

$$\begin{aligned}\int \frac{(p'(X,0))^2}{p(X,0)}dX &= \frac{1}{M^2}\left[N^2\eta_2^2 + N\,\mathrm{tr}\left[(\mathbf{E}SS^*)^2\right] - 2N\eta_2N\eta_2 + N^2\eta_2^2\right]\\ &= \frac{N}{M^2}\mathrm{tr}\left[(\mathbf{E}SS^*)^2\right].\end{aligned}$$

Hence,

$$h(X) \approx NT\log\pi e + N\eta_2/M\rho - \frac{N}{2M^2}\mathrm{tr}\left[(\mathbf{E}SS^*)^2\right]\rho^2$$

from (14) and this together with (8) gives

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

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

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

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

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

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 $\mathrm{tr}(\mathbf{E}SS^*)^2 = \eta_2^2/T$. This in (16) gives

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

IV. Examples

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

A. 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, $\eta_4 = \mathbf{E}(\mathrm{tr}(SS^*)^2) = MT(M+T)$ and so

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

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 $M = 1$.

B. 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 $\eta_4 = T^2M$. 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 $M = 1$.

C. Training-Based Schemes

In these schemes, we have

$$S = \begin{bmatrix} S_\tau \\ S_d \end{bmatrix}$$

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

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

Under these conditions it can be readily shown that

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

and

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

Therefore, using (3), we obtain

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

The latter equation shows that the mutual information is independent of S_τ. In fact, the right-hand side of (17) is just the mutual information of a system with coherence interval $T_d = T - T_\tau$. Thus, in the low-SNR regime, training actually contributes a rate reduction proportional to the fraction of time that one is sending the training symbols. One may contrast this with the result of [15] which shows that training-based schemes achieve capacity at high SNR.

V. 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, under the standard power (second-order) constraint, capacity approaches up to first order the capacity of a channel where the channel matrix is perfectly known to the receiver. This is achieved by a peaky input distribution [6].

We can show this is also the case for the multiple-antenna channel as follows. For any $\epsilon > 1$ and assuming $\rho < 1$, 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

$$A = T\rho^{-\epsilon}\begin{bmatrix} I_M & 0_{M\times(T-M)} \\ 0_{(T-M)\times M} & 0_{(T-M)\times(T-M)} \end{bmatrix}. \quad (18)$$

Then S satisfies the power constraint

$$\mathbf{E}\mathrm{tr}SS^* = \mathrm{tr}(T\rho^{-\epsilon}I_M)\times\rho^\epsilon = TM.$$

Note that the above distribution does not satisfy the regularity conditions i) of Theorem 1.

Then from (6)

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

Also, we have

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

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

$$\left(1+\frac{\rho}{M}T\rho^{-\epsilon}\right)^{MN} \approx \rho^{MN(1-\epsilon)}$$

which is large while the numerator is bounded above by 1. Therefore, the second term is much larger than the first and so

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

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

$$\mathbf{E}\log\det\left(I + \frac{\rho}{M}HH^*\right) = N\rho + o(\rho).$$

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.

A. Fourth-Order Moment Constraint

Suppose we enforce the fourth-order moment constraint $\eta_4 \le K\eta_2^2$ for some positive constant K. This may be a practical constraint to impose but as mentioned in [7] a bounded fourth-order moment will not lead to mutual information optimality at low SNR.

By the root mean square–arithmetic mean inequality we have

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

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

Then $\eta_4 - \eta_2^2/T \le (K - 1/T)\eta_2^2$ 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 $\eta_2 \le TM$ and $\eta_4 \le K\eta_2^2$. 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

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

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

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

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.

B. Peak Constraint

Due to the isotropic unitary matrix in (2) it is not possible to directly enforce a peak constraint on the transmitted signals. However, it is possible to force a maximum constraint on the diagonal entries of D (the singular values of S). Thus, assume that $d_i^2 \leq K$ 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 $\eta_2 \leq P_{\max}$. We also have $\eta_2 = \mathbf{E}\sum_{i=1}^{M} d_i^2 \leq MK$. Now

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

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

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

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

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

and

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

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

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

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

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

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.

VI. Conclusion

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 maximum constraint on the singular values of 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

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

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

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

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

To expand the denominator we use

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

so that

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

Putting everything together

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

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

$$p'(X|S, \rho = 0) = \frac{e^{-\text{tr}X^*X}}{M\pi^{NT}}(\text{tr}X^*SS^*X - N\text{tr}SS^*). \tag{25}$$

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

- $\mathbf{E}_G \text{tr}G^*PG = N\eta_2$
- $\mathbf{E}_G(\text{tr}G^*PG)^2 = N^2\eta_2^2 + N\,\text{tr}P^2$

for any $T \times T$ matrix P satisfying $\text{tr}P = \eta_2$.

Proof:

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

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

as required.

2) We have

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

$\mathbf{E}[\overline{g_{ji}} g_{ki} \overline{g_{ml}} g_{nl}]$ will be nonzero 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 nonzero terms when all its subscripts are equal we have

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

completing the proof. □

The following lemma was used when considering a Gaussian modulation scheme in Section IV-A.

Lemma 3: Let S be a $T \times M$ matrix with independent zero-mean unit-variance complex Gaussian entries. Then

$$\mathbf{E} \operatorname{tr}(SS^*)^2 = MT(M+T).$$

Proof: Denote the (i,j)th entry of S by s_{ij}. Then, as the *pdf* of the Gaussian matrix S is

$$p(S) = \frac{e^{-\operatorname{tr} S^* S}}{\pi^{MT}}$$

we have

$$\begin{aligned}\mathbf{E} \operatorname{tr}(S^*S)^2 &= \int \frac{e^{-\operatorname{tr} S^* S}}{\pi^{MT}} \operatorname{tr}(SS^*)^2 \, dS \\ &= \int \frac{e^{-\operatorname{tr} S^* S}}{\pi^{MT}} \sum_{i,j,k,l} s_{ij} \overline{s}_{kj} s_{kl} \overline{s}_{il} \, dS \\ &= \sum_{i,j,k,l} \int \frac{e^{-\operatorname{tr} S^* S}}{\pi^{MT}} s_{ij} \overline{s}_{kj} s_{kl} \overline{s}_{il} \, dS.\end{aligned}$$

In this summation, the indices i and k each range from 1 to T while the indices j and l each range from 1 to N. If both $i \neq k$ and $j \neq l$, the integral

$$\int \frac{e^{-\operatorname{tr} S^* S}}{\pi^{MT}} s_{ij} \overline{s}_{kj} s_{kl} \overline{s}_{il} \, dX = 0$$

as the integrand is an odd function of the variable s_{ij}.

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

$$\begin{aligned}\int \frac{e^{-\operatorname{tr} SS^*}}{\pi^{MT}} \operatorname{tr}(S^*S)^2 \, dS &= \sum_{\substack{i,j,l \\ j \neq l}} \int \frac{e^{-\operatorname{tr} S^* S}}{\pi^{MT}} |s_{ij}|^2 |s_{il}|^2 \, dS \\ &\quad + \sum_{\substack{i,j,k \\ i \neq k}} \int \frac{e^{-\operatorname{tr} S^* S}}{\pi^{MT}} |s_{ij}|^2 |s_{kj}|^2 \, dS \\ &\quad + \sum_{i,j} \int \frac{e^{-\operatorname{tr} S^* S}}{\pi^{MT}} |s_{ij}|^4 \, dS \\ &= \sum_{\substack{i,j,l \\ j \neq l}} 1 + \sum_{\substack{i,j,k \\ i \neq k}} 1 + \sum_{i,j} 2 \\ &= TM(M-1) + TM(T-1) + 2TM \\ &= MT(M+T),\end{aligned}$$

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.*, vol. 1, no. 2, pp. 41–59, 1996.

[2] İ. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *Europ. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.

[3] L. Zheng and D. N. C. Tse, “Communication on the Grassman manifold: A geometric approach to the noncoherent multiple-antenna channel,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 359–383, Feb. 2002.

[4] A. Lozano, A. Tulino, and S. Verdú, “Multiple-antenna capacity in the low-power regime,” *IEEE Trans. Inform. Theory*, vol. 49, pp. 2527–2544, Oct. 2003.

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

[6] I. C. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz), “The capacity of discrete-time memoryless Rayleigh-fading channels,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 1290–1301, May 2001.

[7] S. Verdú, “Spectral efficiency in the wideband regime,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 1319–1343, June 2002.

[8] İ. E. Telatar and D. N. C. Tse, “Capacity and mutual information of wideband multipath fading channels,” *IEEE Trans. Inform. Theory*, vol. 46, pp. 1384–1400, July 2000.

[9] M. Médard and R. Gallager, “Bandwidth scaling for fading multipath channels,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 840–852, Apr, 2002.

[10] V. Subramaniam and B. Hajek, “Broad-band fading channels: Signal burstiness and capacity,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 809–827, Apr. 2002.

[11] M. Gursoy, H. V. Poor, and S. Verdú, "Efficient signaling for low-power Rician fading channels," in *Proc. 40th Annu. Allerton Conf. Communication, Control and Computing*, Oct. 2–4, 2002.

[12] M. Godavarti and A. O. Hero, III, "Multiple-antenna capacity in a deterministic Rician fading channel," *IEEE Trans. Inform. Theory*, submitted for publication.

[13] V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1567–1580, Aug. 2004.

[14] B. Hajek and V. Subramaniam, "Capacity and reliability function for small peak signal constraints," *IEEE Trans. Inform. Theory*, vol. 48, pp. 828–839, Apr. 2002.

[15] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, pp. 951–963, Apr. 2003.

Performance Bounds for Space–Time Block Codes With Receive Antenna Selection

Xiang Nian Zeng, *Student Member, IEEE,* and
Ali Ghrayeb, *Member, IEEE*

***Abstract*—In this correspondence, we present a comprehensive performance analysis of orthogonal space–time block codes (STBCs) with receive antenna selection. For a given number of receive antennas M, we assume that the receiver uses the best L of the available M antennas, where, typically, $L \leq M$. The selected antennas are those that maximize the instantaneous received signal-to-noise ratio (SNR). We derive explicit upper bounds on the bit-error rate (BER) performance of the above system for any M and L, and for any number of transmit antennas. We show that the diversity order, with antenna selection, is maintained as that of the full complexity system, whereas the deterioration in SNR is upper-bounded by $10 \log_{10}(M/L)$ decibels. Furthermore, we derive a tighter upper bound for the BER performance for any N and M when $L = 1$, and derive an expression for the *exact* BER performance for the Alamouti scheme when $L = 1$. We also present simulation results that validate our analysis.**

***Index Terms*—Antenna selection, multiple-input multiple-output (MIMO) systems, performance bounds, space–time block codes (STBCs).**

I. Introduction

In a wireless environment, unlike other applications, achieving reliable communication is much more challenging due to the possibility that received signals from multipaths may add destructively, which, consequently, results in a serious performance degradation. It has been shown that a key issue to achieving reliable wireless communication is to employ spatially separated antennas at the transmitter and/or at the receiver. To this end, several space–time coding schemes have been developed in recent years, including space–time trellis codes (STTCs) [1], a simple transmit diversity scheme developed by Alamouti [2], and space–time block codes (STBCs) developed by Tarokh *et al.* [3], which are essentially a generalization of the Alamouti scheme. However, among the implications of employing multiple antennas is the associated increase in the cost of the additional hardware required in the form of radio frequency (RF) chains, in addition to the constraints imposed by the physical limitation of wireless devices.

In an effort to overcome these problems, while utilizing the advantages of using multiple antennas, several papers have appeared recently in the literature in which the notion of *antenna selection* was introduced for both STTCs and STBCs [4]–[18]. The idea behind antenna selection is to use only a subset of the transmit and/or receive antennas in multiple-input multiple-output (MIMO) systems. In [4], the authors consider the joint transmit and receive antenna selection based on the second-order channel statistics, which is assumed to be available to the transmitter. The authors in [5] consider antenna selection for low-rank matrix channels where selection is done only at the transmitter. In [6], antenna selection is considered only at the transmitter with the assumption that the channel statistics are available to the transmitter. In [7], Ghrayeb and Duman show that, for full-rank STTCs over quasi-static fading channels, the diversity order is maintained as that of the full-complexity system. In [8], Molisch *et al.* studied the effect of antenna selection from a channel capacity perspective. It was shown that only a small loss in capacity is suffered when the receiver uses a good subset of the available receive antennas. Other work related to antenna selection for STTCs can be found in [9]–[15].

In [16], the authors consider antenna selection for STBCs at the transmitter. They demonstrate through computer simulations that the performance is improved by increasing the number of transmit antennas while keeping the number of selected antennas fixed. [17] considers antenna selection for MIMO systems employing three transmit and one receive antennas. They introduce an algorithm for selecting the best two transmit antennas based on the channel statistics provided by the receiver. In [18], the authors consider antenna selection at the transmitter (with full knowledge of the channel statistics) or at the receiver for orthogonal STBCs with particular emphasis on the Alamouti scheme [2]. In their analysis, they adopt a selection criterion that maximizes the channel Frobenius norm and, accordingly, derive expressions for the improvement in the average signal-to-noise ratio (SNR) and outage capacity. They use the outage probability analysis to argue that the spatial diversity, when the underlying space–time code is orthogonal, is maintained with antenna selection. They do not, however, investigate the direct impact of antenna selection on the bit-error rate (BER) performance, which is the focus of this work.

In [19], the authors propose a new scheme that involves using hybrid selection/maximal-ratio transmission where the transmitter uses a good subset of the available antennas and the receiver uses maximum-ratio combining. They investigate this scheme in terms of signal-to-noise ratio (SNR), BER, and capacity. They demonstrate the effectiveness of their scheme relative to already existing schemes. The same scheme was also treated in [20] but when the transmitter selects the best single antenna. Other schemes that use hybrid selection/maximal-ratio combining were also considered in [21]–[24]. A nice overview of antenna selection for MIMO systems can be found in [25].

In this correspondence, we present a comprehensive performance analysis of STBCs with receive antenna selection. We limit our analysis to orthogonal STBCs simply because they are easy to design and they achieve the maximum diversity order possible for a given number of transmit/receive antennas. In our analysis, we assume that, for a given number of receive antennas M, the receiver uses L out of the available M antennas where the selected antennas are those whose instantaneous SNRs are largest. This is achieved by comparing the sums of the magnitude squares of the fading coefficients at each receive antenna

Manuscript received December 15, 2003; revised April 27, 2004. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grant N00858, and by the National Science Foundation under Grant INT-0217549. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Chicago, IL, June/July 2004.

The authors are with the Department of Electrical and Computer Engineering, Concordia University, Montreal, QC H3G 1MB, Canada (e-mail: x_zeng@ece.concordia.ca; aghrayeb@ece.concordia.ca).

Communicated by R. R. Müller, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2004.833363