Optimizing Multi-Transmitter–Single-Receiver (MISO) Antenna Systems With Partial Channel Knowledge

Aris L. Moustakas and Steven H. Simon

Abstract

In this paper we consider a narrowband point-to-point communication system with many (input) transmitters and a single (output) receiver (ie, a "MISO" system). We assume the receiver has perfect knowledge of the channel but the transmitter has only limited channel information. We focus on two canonical classes of channel models: (a) the channel has zero mean with a fixed covariance matrix and (b) the channel has nonzero mean with covariance matrix proportional to unity. In both cases we are able to derive simple analytic expressions for the ergodic average and the cumulative distribution function of the mutual information for arbitrary input (transmission) signal covariance. With minimal numerical effort, we then determine the ergodic and outage capacities and the input signal covariances which achieve these maximal capacities. We thus show how a transmitter with partial channel knowledge (either knowing the mean channel or the signal covariance) should correlate its transmissions to maximize throughput.

Keywords

Transmit Diversity; Multiple Antennas; Side Information.

I. Introduction

Multi-antenna arrays are known to perform better than their single antenna counterparts because they can more effectively counter the effects of multipath fading and interference. However, the enhanced performance depends on the amount of channel information that the receiver and the transmitter have. For example, it is well known [1] that in a Rayleigh fading environment, for a multi-antenna transmitter and a single antenna receiver (MISO – Multi-Input-Single-Output) the closed-loop (perfect channel knowledge at the transmitter) and open-loop (no channel knowledge at the transmitter) capacities differ by roughly a factor of $\log_2(n)$, where n is the number of transmitter antennas. The knowledge of the channel in the closed-loop case allows the transmitter to transmit with a signal covariance that maximizes the power at the receiver, thus increasing the mutual information.

In addition to the two extremes of closed and open-loop, there are realistic situations where the transmitter may have only partial channel information. Specifically, the transmitter may know the long term correlations between the transmitting antennas (the channel covariance). Alternatively, he may have a noisy estimate of the instantaneous channel (the mean channel), or know the coherent (constant) part of a Ricean channel. In these cases, extra throughput is gained by adapting the transmitter (i.e. adapting the transmitted signal covariance) to the *typical* channel. The idea of exploiting partial (or statistical)

information of the channel at the transmitter was proposed in the more general MIMO (Multi-Input-Multi-Output) context by Moustakas *et al* [2] and Sengupta and Mitra [3] and later by Visotsky [4], Jafar [5] and Narula [6] for the MISO case.

We would like to focus on how the mutual information changes as we change the input signal covariance (i.e. adapt the transmitter) so as to maximize the throughput. Of course, the mutual information of a random channel instantiation is also a random quantity. One can consider several measures of the typical mutual information for an ensemble of channels. One useful measure is the average mutual information (or "ergodic capacity" [1]). Alternatively, one can use the "x% outage mutual information" which is defined to be the minimum mutual information that occurs in all but x% of the instantiations of the channel. (In other words, if we measure the mutual information of the channel many times – in many instantiations of the random channel, we would find that a mutual information greater than the 5% outage would occur 95% of the time). Typically system design aims to optimize either the ergodic average or some given outage. In this paper, we will aim to determine the input signal covariance at the transmitter that optimizes one of these quantities. In order to perform such an optimization, one needs to be able to calculate the ergodic or outage mutual information as a function of the input signal covariance. This can be done, for example, by integrating over a Gaussian channel with known, fixed non-trivial correlations and/or a non-zero mean. In the past, this has been done typically by Monte-Carlo methods (see for example, [4], [5]). Given that the averaging over the channel realizations is performed mostly numerically, the optimization process has, in the past, been very tedious and slow, a clear drawback, especially when it is necessary to optimize dynamically. In addition, the brute-force optimization allows little room to gain intuition on the performance of a particular scheme.

An alternative approach, which we adopt in this paper for MISO systems, is to *analytically* (or mostly analytically) calculate the ergodic or the outage mutual information as a function of the input signal covariance. One can then simply optimize with respect to this covariance. Although this approach cannot be easily applied for general statistics of the channel, we show that for the cases of the channel having a known non-zero mean and/or a covariance matrix, the problem of calculating the ergodic or outage mutual in-

formation as a function of the input signal covariance reduces to the evaluation of a single finite integral of known functions. For a channel having a known covariance matrix and zero average channel (a case of obvious interest) the problem for both ergodic and outage capacities can be further reduced to a simple analytic function. Optimization is then quite straightforward. Even in the case of a channel with non-zero mean where the calculation requires a single numerical integration, this approach is far simpler than the Monte-Carlo approach discussed in Visotsky [4], being that in that paper the averages have to be performed over n random variables, where n is the number of transmitting antennas. We note that in the limit of large numbers of antennas certain calculations (such as average mutual information, or optimizing the covariance of transmissions) simplify substantially[2], [3] . The techniques used in such calculations are quite different from those used here and we defer a discussion of these simplifications to another paper[7].

A. Outline

After some general definitions (section I-B and I-C), we analyze the two canonical scenarios when dealing with partial channel information: In section II we consider the case of a channel with zero mean and a given covariance matrix and transmission with arbitrary covariance. The ergodic and outage mutual informations are derived analytically and optimized. In section III a non-zero mean channel with trivial covariance will be assumed, and similar calculations are performed. (The most general case of non-zero mean channel with nontrivial covariance is also discussed briefly in the Appendix B).

B. Definition of Channel

In the narrowband MISO problem - n transmitters and a single receiver - the channel is defined by

$$\bar{y} = \sum_{i=1}^{n} \bar{G}_i^* \bar{x}_i + \bar{\eta} = \bar{\mathbf{G}}^\dagger \bar{\mathbf{x}} + \bar{\eta} \tag{1}$$

where $\bar{y}$ is the complex received signal, $\bar{x}_i$ is the complex transmitted signal at transmitter i, $\bar{G}_i$ is the complex channel from the i^{th} transmitter, and $\bar{\eta}$ is the complex noise at the receiver. The overbar notation here indicates that we are not yet working with rescaled quantities (which we will fix shortly). In this paper, boldface quantities will always denote n-dimensional vectors.

We will assume that the noise is a gaussian random variable, $E\{\bar{\eta}^*\bar{\eta}\} = \nu$ where $E\{\cdot\}$ denotes the expectation value and ν is the noise power. The transmitted signal covariance matrix is given by $E\{\bar{x}_i^*\bar{x}_j\} = (P/n)Q_{ij}$ where P is the total transmitted power and Q_{ij} is a non-negative definite hermitian matrix, with eigenvalues q_i, for $i = 1 \ldots n$. The constraint of fixed total transmitted power is expressed in terms of Q as

$$\sum_{i=1}^{n} Q_{ii} = \sum_{i=1}^{n} q_i = \mathrm{Tr}\{Q\} = n \tag{2}$$

We now rescale some variables to simplify normalizations. We introduce the constant $\bar{s} = (1/n)\mathrm{Tr}\bar{\Sigma}$. We rewrite our channel equation (Eq. 1) in an equivalent form

$$y = \sum_{i=1}^{n} G_i^* x_i + \eta = \mathbf{G}^\dagger \mathbf{x} + \eta \tag{3}$$

where we have defined $y = \bar{y}/\sqrt{\nu}$, $\mathbf{x} = \bar{\mathbf{x}}\sqrt{\bar{s}/\nu}$, and $\eta = \bar{\eta}/\sqrt{\nu}$, and the new channel is defined in terms of the old channel as $\mathbf{G} = \bar{\mathbf{G}}/\sqrt{\bar{s}}$ such that the new channel $\mathbf{G}$ is $\mathcal{N}(\mathbf{g}, \Sigma)$ with $\Sigma = \bar{\Sigma}/\bar{s}$ and $\mathbf{g} = \bar{\mathbf{g}}/\sqrt{\bar{s}}$ and we now have conveniently $\mathrm{Tr}\{\Sigma\} = n$. We will frequently use the notation that s_i are the eigenvalues of Σ. In this new rescaled form of the channel equation, we also have $E\{\eta^*\eta\} = 1$ and $E\{x_i^* x_j\} = pQ_{ij}$ where

$$p = \frac{P\bar{s}}{\nu n} \tag{4}$$

The physical meaning of the parameter p is roughly that of a signal to noise ratio per antenna for the fluctuating part of the channel. For the case where $Q = 1$ and the average channel $\bar{\mathbf{g}}$ is zero, p is precisely the signal to noise ratio per transmitting antenna. It is also convenient to define the parameter

$$\gamma = \frac{\bar{\mathbf{g}}^\dagger \bar{\mathbf{g}}}{\bar{s}} = \mathbf{g}^\dagger \mathbf{g} \tag{5}$$

This quantity is roughly the ratio of the power from the constant part of the channel $\mathbf{g}$ to the power from the fluctuating part of the channel. More precisely, if transmissions are uncorrelated ($Q = 1$ again), the signal to noise ratio from the constant part of the channel is $p\gamma$.

Finally, we note that we will also frequently use the notation that a_i are the eigenvalues of $(pQ\Sigma)^{-1}$.

C. Definitions of Quantities to Calculate

For a given instantiation of $\mathbf{G}$ (or $\bar{\mathbf{G}}$), if the receiver knows the channel (which in practice is achieved by sending pilots), the mutual information $I((y,\mathbf{G});\mathbf{x}) = I((\bar{y},\bar{\mathbf{G}});\bar{\mathbf{x}})$ is given by

$$I((y,\mathbf{G});\mathbf{x}) = \log(1+\mathbf{G}^\dagger pQ\mathbf{G}) \tag{6}$$

Throughout this paper we will measure information in *nats* where 1 *nat* is equal to e bits per sec per Hz ($e = 2.718\ldots$).

The probability distribution function $\mathrm{PDF}(I)$ of the mutual information I in Eq. 6 over the ensemble of instantiations $\mathbf{G}$ (with Q fixed) can then be written as

$$\mathrm{PDF}(I) = \langle \delta(I - \log(1+\mathbf{G}^\dagger pQ\mathbf{G}))\rangle \tag{7}$$

where the brackets $\langle\cdot\rangle$ represent an ensemble average over realizations of $\mathbf{G}$ with $\delta()$ the Dirac delta function. Note that the PDF is implicitly a function of Q, p, $\mathbf{g}$, and Σ, as well as I. We will not usually make all of these dependencies explicit.

In terms of the PDF, we can define the so-called ergodic average (i.e., mean) of the mutual information to be the average of I over the ensemble

$$\langle I\rangle = \int_0^\infty dI\ I\ \mathrm{PDF}(I) \tag{8}$$

We also define the complementary cumulative distribution function

$$\mathrm{CCDF}(I) = \int_I^\infty dI'\ \mathrm{PDF}(I) \equiv 1 - \mathrm{CDF}(I) \tag{9}$$

(Here CDF is known as the cumulative distribution function).

It is convenient to define an inverse function of this CDF which we will call the outage mutual information $\mathcal{OUT}$. More specifically, for a fixed Q and Σ, we define $\mathcal{OUT}(P_{out})$ such that

$$I_{out} = \mathcal{OUT}(P_{out}) \tag{10}$$

when

$$P_{out} = \mathrm{CDF}(I_{out}) = 1 - \mathrm{CCDF}(I_{out}) \tag{11}$$

Since $\mathrm{PDF}(I) > 0$, the inversion is unique. The meaning of $I_{out} = \mathcal{OUT}(P_{out})$ is that there is a probability P_{out} that in any instantiation of $\mathbf{G}$ from the ensemble, we will obtain

a mutual information I less than I_{out}. This is the usual definition of outage (as we have already described it above).

C.1 Comment on Closed-Loop Capacities

For a closed-loop MISO system, where the channel $\mathbf{G}$ is known perfectly to the transmitter, it is clear (see Eq. 6) that the mutual information is optimized for a signal covariance Q given by

$$Q_{ij} = n\frac{G_i G_j^*}{\mathbf{G}^\dagger \mathbf{G}} \tag{12}$$

(This Q yields the largest possible I subject to the power constraint $\mathrm{Tr}\{Q\} = n$). Using this Q, the mutual information for the given $\mathbf{G}$ is then given by

$$I = \log(1 + \mathbf{G}^\dagger pn\mathbf{G}) \tag{13}$$

We see that this is equivalent to the above expression for mutual information (Eq. 6) except that p has been replaced by pn and Q is replaced by the unit matrix. Calculating $\langle I \rangle$ for this equivalent problem then yields the closed-loop ergodic capacity and calculating (and inverting) the CDF(I) for this equivalent problem yields the closed-loop outage capacity. Appropriate results for this equivalent problem (with $Q = 1$) are given in Eqs. 17, 21 and 26 below.

We note again that if the transmitter does not have perfect knowledge of the channel, but knows the statistics of the channel (i.e., knows $\mathbf{g}$ and Σ) then the optimal Q is harder to determine. We reiterate that the optimization of Q when only the statistics of the channel are known is the main focus of this paper.

II. Channel with known Covariance

We start by analyzing the case where the channel $\mathbf{G}$ has zero mean $\mathbf{g} = 0$ and covariance Σ. This situation is quite realistic for many wireless systems where the channel is constantly changing, but correlations in the channel are only slowly varying. (One could easily imagine feeding back Σ to the transmitter to enable optimization of transmission, and updating Σ only very occasionally).

A. PDF

In appendix A the probability distribution function for a zero mean channel is found to be

$$\mathrm{PDF}(I) = e^I \prod_{i=1}^{n}(-ia_i) \int \frac{dk}{2\pi} \frac{e^{ik(e^I-1)}}{\prod_{i=1}^{n}(k-ia_i)} \tag{14}$$

where a_i, $i = 1 \ldots n$, are the eigenvalues of $(pQ\Sigma)^{-1}$ (which are necessarily non-negative). If all the eigenvalues are different, the contour integral can be performed to yield

$$\mathrm{PDF}(I) = \left(\prod_{i=1}^{n} a_i\right) \sum_{j=1}^{n} \frac{f(a_j, I)}{a_j \prod_{m\neq j}(a_m - a_j)} \tag{15}$$

where

$$f(a, I) = a \exp\Big(I - a(e^I - 1)\Big) \tag{16}$$

Despite its apparent singular form, $\mathrm{PDF}(I)$ is finite and continuous when some of the eigenvalues a_i become degenerate. This can be seen by taking a limit of Eq. 15 where the degenerate eigenvalues approach each other. It is typically easier, however, to back up to Eq. 14 and perform the contour integral with a multiple pole. An illustrative example is the PDF of the mutual information of a white Gaussian channel ($\Sigma = 1$) with uncorrelated transmissions ($Q = 1$). In this case all n eigenvalues are equal $a_i = 1/p$ and we obtain

$$\mathrm{PDF}(I) = \frac{1}{p}\left(\frac{e^I - 1}{p}\right)^{(n-1)} \frac{1}{(n-1)!} \exp\left(I - \frac{e^I - 1}{p}\right) \tag{17}$$

Here note that p is precisely the average signal to noise per antenna. If one of the eigenvalues of ΣQ is zero, this indicates that there is a redundancy in the transmission (i.e. the n channels are not all independent). Mathematically, this corresponds to the divergence of a particular a_k. We can see that in Eq. 15, divergence of a_k eliminates the k^{th} term of the sum and in each of the other terms of the sum, the k^{th} term of the products are reduced to unity. Thus, we see that having a zero eigenvalue is precisely equivalent to having 1 fewer transmitter.

B. CCDF and Outage Mutual Information

Analogous expressions for the complementary cumulative distribution function (CCDF) can be found by direct integration (See Eq. 9) of either Eq. 14 or Eq. 15. Most generally

we obtain

$$\mathrm{CCDF}(I) = \prod_{i=1}^{n}(-ia_i)\int \frac{dk}{2\pi}\frac{e^{ik(e^I-1)}}{i(k+i0^+)\prod_{i=1}^{n}(k-ia_i)} \tag{18}$$

Again, if all the eigenvalues are different, this reduces to

$$\mathrm{CCDF}(I) = \left(\prod_{i=1}^{n} a_i\right)\sum_{j=1}^{n}\frac{g(a_j,I)}{a_j\prod_{m\neq j}(a_m-a_j)} \tag{19}$$

where

$$g(a,I) = \int_I^\infty dI'\ f(a,I') = \exp\left(-a(e^I-1)\right) \tag{20}$$

In the case of a white Gaussian channel ($\Sigma = 1$) with uncorrelated transmissions ($Q = 1$) we obtain the simple form

$$\mathrm{CCDF}(I) = \exp\left(\frac{-(e^I-1)}{p}\right)\sum_{m=1}^{n-1}\frac{1}{m!}\left(\frac{e^I-1}{p}\right)^m = \frac{1}{(n-1)!}\ \Gamma\left[n,\frac{e^I-1}{p}\right] \tag{21}$$

with Γ here the incomplete gamma function (and again p is precisely the average signal to noise per antenna here).

Eqs. 18 and 19 have a very important scaling property. Since the eigenvalues a_i's (eigenvalues of $(pQ\Sigma)^{-1}$) are inversely proportional to the "signal power" p, the CCDF is a function only of the combined quantity

$$z = \frac{\exp(I)-1}{p} \tag{22}$$

and is not a function of p and I independently. Thus for a given Q and Σ, a single calculation of the CCDF applies for all power levels. As described above, the outage mutual information can be found by inverting the calculated CCDF. Due to this scaling property, the outage capacity can be written as

$$I_{out} = \log\left[1 + p\,F(P_{out},Q,\Sigma)\right] \tag{23}$$

where the function F depends only on the target outage probability P_{out}, as well as on Q and Σ, but does not depend on the power level (signal to noise per antenna) p. This implies that for a given channel covariance Σ, the optimal transmission covariance Q for a given outage probability P_{out} (i.e. that which maximizes F) is *independent* of the power. We note that this simple scaling property does not hold for the mean mutual information.

It is amusing that this scaling may make the outage mutual information (which is probably the more technologically relevant quantity) an *easier* quantity to work with than the mean mutual information (which is usually considered to be simpler).

C. Mean (Ergodic) Mutual Information

The mean mutual information $\langle I \rangle$ for fixed Q can be derived by integrating the PDF (as shown in Eq. 8) resulting in(when all eigenvalues a_i of $(pQ\Sigma)^{-1}$) are different)

$$\langle I \rangle = \left(\prod_{i=1}^{n} a_i \right) \sum_{j=1}^{n} \frac{m(a_j)}{a_j \prod_{k \neq j} (a_k - a_j)} \tag{24}$$

where

$$m(a) = a \int_0^\infty dI \; I \; e^{I - a(e^I - 1)} = -e^a \mathrm{Ei}(-a) = e^a \Gamma(0, a) \tag{25}$$

with Ei the exponential integral and Γ the incomplete gamma function. As in the case of Eqs. 15 and 19, $\langle I \rangle$ is continuous and finite when any two a's become equal. In particular, when all a_i's are equal (relevant for the case of uncorrelated transmissions $Q = 1$ and white channel covariance $\Sigma = 1$) the mean mutual information becomes

$$\langle I \rangle = \left(\frac{-1}{p} \right)^n \frac{-1}{(n-1)!} \left(\frac{d}{da} \right)^{(n-1)} \left[\frac{\exp(a)}{a} \Gamma(0, a) \right] \Bigg|_{a = \frac{1}{p}} \tag{26}$$

with p again the signal to noise per antenna.

D. Optimization over Q

To reach the full information capacity of the communication link, the transmitter has to optimize the transmitting signal covariance matrix Q. Of course, in the closed-loop case when the transmitter has perfect channel knowledge this is very straightforward (as discussed above in section I-C.1). However, the situation is more complicated when the transmitter has only partial (statistical) channel knowledge (i.e. the transmitter knows Σ–recall we have assumed $\mathbf{g} = 0$ for now).

The transmitter may either choose to maximize the average mutual information $\langle I \rangle$, yielding the ergodic capacity [1], or to maximize the outage mutual information I_{out} for a fixed outage probability P_{out}. As we shall see, the resulting capacities and optimal Q matrices in these two cases will be quite different. (This should be contrasted with the

closed-loop case where we obviously choose the same Q's independent of whether we are maximizing ergodic capacity or outage capacity).

The optimization problem of Q for the ergodic capacity has been analyzed in the past by Visotsky [4] in the context of the MISO system. However, the equations to be solved for the determination of Q were averages over n random variables, rather than analytic expressions as Eqs. 19 and 24. Therefore their approach is not very useful for more than two or three transmitting antennas. In addition, we are also interested in outage capacities, which in many cases are more useful (see references [8], [9]), and turn out to be simpler than the equations for the ergodic capacity due to the above mentioned scaling law.

In either case (outage capacity or ergodic capacity) it is sufficient to optimize over the eigenvalues of Q in the basis of Σ (i.e. Q and Σ should be simultaneously diagonalizable). The proof of this statement is given by Visotsky [4] and Jafar [10] for the ergodic capacity case and has been generalized by the current authors[11] to apply to the case of outage capacity. Since we have derived analytic expressions for both the ergodic information and outage mutual information, this n-parameter maximization (subject to the trace constraint on Q, which makes it an $n-1$ dimensional problem) can now be done with minimal numerical effort.

In the next subsections we analyze Eqs. 19 and 24 and their optimal Q's in greater detail.

We define beamforming to be the transmission of all power through the maximum eigenvalue and corresponding eigenvector of Σ. When Σ has more than one non-zero eigenvalue, we would like to know when beamforming remains optimal.

We write the eigenvalues of Q as q_i and the eigenvalues of Σ as s_i. As noted above, we also work in a basis where both Q and Σ are diagonal. Without loss of generality, we use a basis such that $s_1 \geq s_2 \geq s_3 \ldots \geq s_n$. The beamforming mode is such that $q_1 = n$ and all other q's are zero. We note also that if $s_1 = s_2$ then beamforming is never optimal, so we may assume that $s_1 > s_2$. We will also assume $s_2 > s_3$ although our result will hold even for $s_2 = s_3$.

We note that since s_2 is the second largest eigenvalue (the second best mode of trans-

mission), in order to find the optimality condition for beamforming we need only find the point where q_2 first becomes nonzero.

D.1 Optimality of Beamforming for the Ergodic Capacity

For the beamforming solution to be optimal for the ergodic capacity $\langle I \rangle$, we must have $\langle I \rangle$ decrease as q_1 is reduced and any other q_i is increased so as to preserve the constraint $\sum_i q_i = n$. Thus, the condition for beamforming is

$$\left.\frac{d\langle I\rangle}{dq_k}\right|_{q_1=n,\ q_j=0,\, j\geq 2,\, k\geq 2} \equiv \left[\frac{\partial\langle I\rangle}{\partial q_k} - \frac{\partial\langle I\rangle}{\partial q_1}\right]_{q_1=n,\ q_j=0,\, j\geq 2,\, k\geq 2} \leq 0 \tag{27}$$

An equivalent condition was also derived by Visotsky [4] and Jafar [5]. It is, in fact, easy to show that since $s_2 > s_3 > \ldots$, if this condition holds for $k = 2$, then it holds for $k > 2$ (In other words if it is not advantageous to move some power from the strongest mode s_1 to the next strongest mode s_2, then it is also not advantageous to move power to any of the weaker modes). In fact, since $\langle I \rangle$ as a function of q_2 is convex ($d^2\langle I\rangle/dq_2^2 < 0$) over [0,1] subject to $q_1 + q_2 = n$ and $q_j = 0$ for $j > 2$ (as shown by Jafar in [5]) then Eq. 27 for $k = 2$ is in fact a necessary and sufficient condition for beamforming to optimize the ergodic mutual information $\langle I \rangle$.

The relevant derivatives of $\langle I \rangle$ can be taken explicitly to give

$$\left.\frac{d\langle I\rangle}{dq_2}\right|_{q_j=0, j>2} = \frac{1}{q_1 s_1 - q_2 s_2}\left[\frac{m_1}{pq_1} + \frac{m_2}{pq_2} - (s_1 + s_2) + n s_1 s_2 \frac{m_1 - m_2}{q_1 s_1 - q_2 s_2}\right] \tag{28}$$

where $m_i = m\,(1/pq_i s_i)$ with the function m given by Eq. 25. Taking the limit $q_2 \to 0$ (with $q_1 + q_2 = 2$) and setting the above expression to zero we find that the optimality condition for beamforming becomes

$$\Gamma\left(0, \frac{1}{nps_1}\right) \leq \frac{s_1 np}{1 + s_2 np} \exp\left(-\frac{1}{nps_1}\right) \tag{29}$$

in agreement with Jafar's result in [5] and the more general result by the current authors in [11].

When the above expression (Eq. 29) is not satisfied, beamforming is no longer optimal. To find the optimal q_i's and the resulting ergodic capacity one needs to maximize $\langle I \rangle$ over q_i subject to the power constraint ($\sum q_i = n$). This optimization is a straightforward

problem given the convexity of $\langle I \rangle$ and our derived analytic form of $\langle I \rangle$ (see, for example [12]).

For the case of two antennas, finding q_2 is particularly simple – one need only set $n = 2$ and $q_1 = 2 - q_2$ and set Eq. 28 to zero. In Fig. 1 we plot the optimal q_2 for the two antenna case as a function of the antenna cross-correlation $x = (s_1 - s_2)/2$ for various values of the parameter p. Recall that p is proportional of the transmission power, and is roughly the signal to noise per antenna (it would be precisely the SNR per antenna if we set $q_1 = q_2 = 1$, see section I-B). When q_2 is zero (for high antenna cross-correlations) beamforming is optimal. At low cross-correlations, q_2 approaches unity, which means that it is optimal to distribute the power evenly between the two transmission eigenmodes (which is what we call "independent transmission"). Note that for large SNR the curves appear to converge to a universal result.

In Fig. 2 we compare the capacity obtained by optimizing the transmission covariance (i.e. using the optimal q_1 and q_2) to the mutual information attainable by using either independent transmission or beamforming. It is clear that using beamforming at low antenna cross-correlation, or using independent transmission at high antenna cross-correlation are both extremely inefficient. Furthermore we see that using the optimal transmission covariance allows more capacity than either beamforming or independent transmission for intermediate cross correlations (although the capacity gain is not very large).

D.2 Optimality of Beamforming for the Outage Capacity

We now turn to the case of outage capacity. Here, since the CCDF in Eq. 19 depends on I_{out} only through z, as discussed above, the results are independent of the signal to noise parameter p. However, determining the optimal Q is now more complicated because I_{out} is generally *not* a convex function of q_2. We demonstrate this nonconvexity by analyzing the case of 2 transmitting antennas in detail. After solving for z in Eq. 19 we set $q_1 = 2 - q_2$ and take the derivative of z with respect to q_2

$$\frac{dz}{dq_2} = \frac{1}{(h_1 - h_2)} \left[2P_{out} - h_1 \left(s_1 + \frac{z}{q_1} \right) - h_2 \left(s_2 + \frac{z}{q_2} \right) \right] \tag{30}$$

with

$$\begin{aligned} P_{out} &= \frac{q_1 s_1 h_1 - q_2 s_2 h_2}{q_1 s_1 - q_2 s_2} \\ h_i &= \exp\left(-\frac{z}{q_i s_i}\right) \\ z &= \frac{e^{I_{out}} - 1}{p} \end{aligned} \tag{31}$$

If $dz/dq_2 > 0$ for $q_2 = 0$, clearly $q_2 = 0$ is not optimal. Naively when $dz/dq_2 \leq 0$ at $q_2 = 0$ one would assume that $q_2 = 0$ would be the optimal solution. However this is only true if $d^2z/dq_2{}^2 < 0$ for all q_2 in $[0, 1]$, which in this case is not true. In fact, it is straightforward to show that at $q_2 = 0$, the opposite is always true $d^2z/dq_2^2 > 0$. This implies that the transition to beamforming always occurs via a finite jump of q_2 when $dz/dq_2 < 0$.

In Fig. 3, we show an example of this nonmonotonicity explicitly and show how beamforming becomes optimal in a discontinuous way. The optimal q_2 jumps discontinuously from close to 0.6 for $x < 0.225$ to 0 for $x > 0.225$. However, we also note that this figure shows that the size of the nonmonotonicity (differences of mutual informations) is rather small, and becomes even smaller at low outages.

In Fig. 4 we plot the optimal q_2 for the two antenna case as a function of the antenna cross-correlation $x = (s_1 - s_2)/2$ for various outages. When q_2 is zero (for high antenna cross-correlations) beamforming is optimal. At low cross-correlations, q_2 approaches unity, which means that it is optimal to distribute the power evenly between the two transmission eigenmodes. We note also that more independent transmission (q_2 closer to 1) becomes more favored as the outage is reduced. Note also that the discontinuity in the optimal q_2 as a function of antenna cross-correlations is evident in these plots, and is a more substantial discontinuity at high outage. When the discontinuity vanishes, the beamforming optimality condition is indeed defined by the above Eqs. 30.

Fig. 4 in the outage case is quite analogous to Fig. 1 for the ergodic case discussed above. However, the optimal distributions of transmitted power differ substantially in these two cases, although both describe a crossover from independent transmission to beamforming. For all outages, for large antenna cross-correlation, beamforming ($q_2 = 0$) becomes optimal, whereas for small cross-correlations, independent transmission ($q_1 = q_2 = 1$) is optimal. For lower outages, the crossover to beamforming occurs at higher

antenna cross-correlations. This is a reflection of the fact that beamforming is more susceptible to fading since it uses only one mode.

In Fig. 5 we show relative outage capacities for optimized transmission, beamforming, and independent transmission as a function of antenna cross-correlation and outage. Here we have plotted ratios of z parameters, where $z = (e^I - 1)/p$. (Note that the factors of p cancel in ratios of z's). At small I_{out} (low SNR), the ratio of z parameters is equal to the ratio of the outage capacities ($z_1/z_2 \approx I_1/I_2$). At high I_{out} (high SNR) the log of the z ratio becomes the difference in information capacities ($\log(z_1/z_2) \approx I_1 - I_2$).

From Fig. 5 we see that for high antenna cross correlation, beamforming becomes optimal $z_{optimal}/z_{beam} = 1$ whereas at low antenna cross-correlation independent transmission becomes optimal $z_{optimal}/z_{indep} \to 1$. We see that incorrectly choosing to use beamforming at low antenna cross correlations can lower the outage capacity by a very large factor (roughly a factor of 2.5 or 3.5 for 10% and 5% outage respectively at low SNR). Similarly using independent transmission at high antenna cross-correlation can carry a large penalty (a factor of 2 at low SNR). However, by switching between beamforming and independent transmission modes at an appropriate point (the point where $z_{beam} = z_{indep}$), one can stay very close to the optimal outage capacity. As can be seen from Fig. 5, a small amount of additional outage capacity can be obtained by always using the optimal transmission covariance (optimal q_2).

E. Example: Capacity for 4 Transmitting Antennas

In both the ergodic and outage case, the analytic form of the mutual information in Eqs. 19 and 24 reduces the optimization over q_j to a straightforward exercise. As a slightly more complicated case compared to the two antenna cases discussed above, Fig. 6 shows the outage and ergodic capacities for a 4x1 MISO system, as a function of angle spread Δ at the transmitter. The transmitting antenna array is a uniform linear array with elements separated by $\lambda/2$. The correlation matrix Σ is determined by the formula

$$\Sigma_{ij} = \exp\left[-\frac{(\pi\,(i-j)\,\Delta)^2}{2}\right] \tag{32}$$

which is a simple approximation for a planar, Gaussian-distributed ray model with the main path of departure from the array being at broadside. The capacities were calculated

by numerically optimizing I in Eqs. 19 and 24 with respect to q_1, q_2, q_3, and q_4 subject to the power constraint Eq. 2.

We see that the 50% outage capacity is very close to the ergodic capacity, reflecting the fact that the PDF is roughly symmetric around its average. Also, for low outages, one can obtain a substantial gain by using the optimal transmission covariance rather than beamforming or independent transmission. As in the above two antenna case, the crossover between independent transmission and beamforming is very dependent on outage. More importantly, if one is technologically interested in optimizing 5% outage, this is clearly quite different from optimizing the ergodic average (which is similar to 50% outage).

Optimizing a function (outage, for example) in a space of $n-1$ dimensions becomes a bit difficult numerically for larger n. One might think, therefore, that performing these calculations and finding the optimal transmission correlations (optimal q's) in real time for a system might be prohibitive. A common technological solution to this problem would be to implement a lookup table which, given measured values of s_i (eigenvalues of Σ) would produce the optimal values of q's. However, as n gets large, even a lookup table might be very hard to implement since it would require entries in a high-dimensional space. Even for $n = 4$ such a table may be a bit unwieldy. Instead, we propose that a system could be built to operate in several canonical modes such as beamforming ($q_1 = n$ and $q_i = 0$ for $i > 1$) or independent transmission (all $q_i = 1$) as well as several intermediate modes. An example of a intermediate mode is shown in Figure 5. In the figure we have considered an intermediate case (the dot-dashed curve) where only two modes are used ($q_1 = q_2 = 2$ and $q_3 = q_4 = 0$). This particular intermediate mode operates close to optimally where beamforming and independent transmission both fall short of the optimal capacity by about 20%. Since our analytic expressions for the capacity (outage or ergodic) are so simple, it would be quite easy for a system to perform calculations in real time to determine which of a number of predetermined modes will be best for a given antenna cross-correlation matrix.

III. Nonzero Mean Channel

In this section, we consider the case where the channel $\mathbf{G}$ has a nonzero mean $\mathbf{g}$ and a white covariance $\Sigma = 1$. In other words, we take the ensemble of $\mathbf{G}$ to be given by $\mathcal{N}(\mathbf{g}, 1)$. This case of constant average channel is applicable to situations where the channel estimate

$\mathbf{g}$ is fed back to the transmitter with an error which is assumed to be gaussian random iid. Indeed, any attempt at a closed-loop system would have some nonzero error and might thus be described by this case. More generally, one might consider a case where the channel has a nonzero mean and a nontrivial covariance (i.e., a $\mathcal{N}(\mathbf{g}, \Sigma)$ channel). This more general case is discussed in Appendix B. In the present section, however, we restrict our attention to $\Sigma = 1$.

In Appendix B-A the expressions for the PDF, CCDF, and ergodic average of the mutual information is derived for the case of $\Sigma = 1$ and $\mathbf{g}$ nonzero. In this scenario, the optimal Q that maximizes the mutual information for fixed $\mathbf{g}$, in both the ergodic capacity case and the outage capacity case, is shown in Appendix B, to be of the simple form

$$Q_{ij} = q_0 \frac{g_i g_j^*}{\mathbf{g}^\dagger \mathbf{g}} + q \left(\delta_{ij} - \frac{g_i g_j^*}{\mathbf{g}^\dagger \mathbf{g}} \right) \tag{33}$$

where

$$q_0 = n - (n-1)q \tag{34}$$

and $q_0 \geq q$. The first term of Eq. 33 is the projection of Q in the direction of $\mathbf{g}$, while the second is the projection into the basis perpendicular to $\mathbf{g}$. By symmetry, the directions perpendicular to $\mathbf{g}$ are equivalent. This form has also been found by Visotsky [4] for the case of ergodic capacity although their derivation is somewhat unclear.

With the Q of Eq. 33, the PDF of the mutual information I, Eq. 57, which was derived in Appendix B can be integrated (using the integration technique described there) to give

$$\mathrm{PDF}(I) = e^I \frac{e^{-\gamma} e^{-\frac{z}{q_0}}}{q_0 (n-2)!} \left(\frac{z}{q} \right)^{n-1} \int_0^1 dt\, t^{n-2} I_0 \left(2\sqrt{\frac{\gamma z}{q_0}} \sqrt{1-t} \right) \exp\left(-zt \left(\frac{1}{q} - \frac{1}{q_0} \right) \right) \tag{35}$$

where

$$z = \frac{e^I - 1}{p} \tag{36}$$

with p being proportional to the signal to noise per antenna for the fluctuating part of the signal (See section I-B for the definition of p) and $\gamma = \mathbf{g}^\dagger \mathbf{g}$ is the mean channel signal strength (divided by the strength of the fluctuating part of the channel per antenna, see section I-B).

One case of particular interest is when transmissions are uncorrelated ($Q = 1$, or $q_0 =$

$q = 1$) In this simple case we can integrate this result to give

$$\mathrm{PDF}(I) = \frac{1}{p} e^I e^{\gamma} e^{-z} \left(\frac{z}{\gamma}\right)^{(n-1)/2} I_{n-1}\left(2\sqrt{z\gamma}\right) \tag{37}$$

Similarly, for the above form of Q, Eq. 59 derived in Appendix B can be integrated to give the CCDF of the mutual information

$$\mathrm{CCDF}(I) = 1 - e^{-\gamma} \frac{z}{q_0} \int_0^1 dt e^{-\frac{zt}{q_0}} I_0\left(2\sqrt{\frac{\gamma z t}{q_0}}\right) \left[1 - \frac{\Gamma\left(n, \frac{z(1-t)}{q}\right)}{(n-1)!}\right] \tag{38}$$

where $\Gamma(n, x)$ is the incomplete Γ function. Finally, in Appendix B we show that the (ergodic) average mutual information can be written in terms of a single integral as

$$\langle I \rangle = \int_0^{\infty} \frac{dy \, \exp\left(-\frac{y}{p}\right)}{y} \left[1 - \frac{\exp\left(-\frac{\gamma q_0 y}{1+q_0 y}\right)}{(1+q_0 y)\,(1+qy)^{n-1}}\right] \tag{39}$$

Eqs. 38 and 39 constitute the main results of this section. Maximizing the outage mutual information in Eq. 38 or $\langle I \rangle$ in Eq. 39 will produce the optimal q as a function of γ and p and hence give the outage or ergodic capacity respectively. Although the optimization procedure involves maximizing an integral over one parameter, which has to be done numerically, it is a far simpler procedure than optimizing over an expectation value of a random vector as done by Visotsky in [4].

A. Beamforming criteria and Optimization over Q

Next we analyze the optimization over Q. In this section the word "capacity" means the capacity of a system with fixed $\mathbf{g}$. (To evaluate the capacity of an actual system one would typically want to average over all of the instantiations of $\mathbf{g}$ that might occur in an ensemble).

A.1 Ergodic Case

Since the ergodic mutual information is convex [5], the beamforming condition is that at $q = 0$

$$\left.\frac{d\langle I \rangle}{dq}\right|_{q=0} = \left[\frac{\partial \langle I \rangle}{\partial q} - \frac{1}{n-1}\frac{\partial \langle I \rangle}{\partial q_0}\right]_{q=0,\ q_0=n} \leq 0 \tag{40}$$

Taking the derivative in Eq. 39 yields

$$\int_0^{\infty} dt \frac{\exp\left(-\frac{t}{np}\right) \exp\left(-\frac{\gamma t}{1+t}\right)}{(1+t)^3} \left(t\,(t+1) - \gamma\right) \leq 0 \tag{41}$$

When beamforming is not optimal ($q > 0$), to find the optimal transmission covariance, one needs to maximize Eq. 39 with respect to q. This procedure is far simpler than the one proposed by Visotsky in [4] since it only involves one integral rather than an expectation value over n random Gaussian variables. In Fig. 7 we calculate the optimal q for the case of 4 transmit antennas. The optimal q from the maximization of the ergodic capacity is plotted versus γ for several values of the parameter p (see section I-B for the definition of p). As would be expected, when $\gamma = \mathbf{g}^\dagger\mathbf{g}$ is large with respect to the uncertainty of the channel, the beamforming solution ($q = 0$) is preferred. The crossover to the beamforming solution is faster when p is small.

A.2 Outage Case

As discussed above in section II-D.2, the outage capacity as a function of q is is not necessarily monotonic. We demonstrate this for the case of a non-zero mean channel by analyzing the behavior of $z(q, \gamma, P_{out})$ close to $q = 0$. The first derivative of z with respect to q at $q = 0$ is

$$\left.\frac{dz}{dq}\right|_{q=0} = (n-1)\left(1 - \frac{z}{n}\right) \tag{42}$$

where z is the solution of Eq. 38 at $q = 0$ which simplifies to

$$P_{out} = e^{-\gamma}\frac{z}{n}\int_0^1 dt e^{-\frac{zt}{n}} I_0\left(2\sqrt{\frac{\gamma z t}{n}}\right) \tag{43}$$

The second derivative at $q = 0$ is

$$\left.\frac{d^2z}{dq^2}\right|_{q=0} = \frac{n-1}{zI_0\left(2\sqrt{\frac{\gamma z}{n}}\right)}\left[I_0\left(2\sqrt{\gamma z/n}\right)\frac{z}{n} - \sqrt{\frac{\gamma z}{n}}I_1\left(2\sqrt{\frac{\gamma z}{n}}\right)\right]\left[1 + 2(n-1)\left(1 - \frac{z}{n}\right)^2\right] \tag{44}$$

where $I_n(x)$ is the n-th order modified Bessel function. For $q = 0$ to be a local maximum we need to have at $q = 0$, $z = n$ and $d^2z/dq^2 < 0$. This is only valid for $I_0(2\sqrt{\gamma}) < \sqrt{\gamma}I_1(2\sqrt{\gamma})$, i.e. $\gamma > 1.670$. This corresponds to

$$P_{out} < P_{crit} = 0.226 \tag{45}$$

in the sense that $z(0, \gamma > 1.67, P_{out}) = n$ in Eq. 43 has solution $P_{out} < 0.226$. By numerically analyzing Eq. 38 we find that for $P_{out} < 0.226$ the optimal $q(\gamma)$ decreases continuously to zero. This is seen in Fig. 8 for 10% and 5% outage probabilities.

In contrast, for $P_{out} > P_{crit}$ the optimal q jumps discontinuously to zero. This can be seen from the fact that for $\gamma < 1.670$ and $z(0, \gamma, P_{out}) = n$ the second derivative $d^2z/dq^2 > 0$ when $dz/dq = 0$ at $q = 0$. Thus the optimal q jumps from a finite value $q > 0$ to $q = 0$ at some γ and $z > n$ for which $dz/dq < 0$ at $q = 0$. This is clearly seen in Fig. 8 for $P_{out} = 30\%$ and $P_{out} = 50\%$.

In closing, from the comparison of Figs. 7 and 8 it becomes clear that the behavior in the outage cases is quite different from that of the ergodic case.

IV. Summary

In this paper we have shown how to analytically (or mostly analytically) calculate outage capacities and ergodic averages of the mutual information of MISO systems as a function of the transmission covariance Q, thus enabling us to determine which Q maximizes the mutual information. We have considered cases where the channel is described as being $\mathcal{N}(0, \Sigma)$ or being $\mathcal{N}(\mathbf{g}, 1)$, i.e., having mean zero and covariance Σ or having mean $\mathbf{g}$ and white covariance. (The more general case of $\mathcal{N}(\mathbf{g}, \Sigma)$ is discussed in Appendix B). These cases represent realistic situations where only partial channel information is fed back to the transmitter. While the channel itself may change quickly, these statistical quantities (covariance or mean channel) may change much more slowly, thus we think about adaptation of the transmitter to the slowly varying properties of the channel.

An important issue we have not discussed is coding for MISO transmission. For any case where beamforming is used, the problem reduces to that of simple 1×1 transmission. For cases of 2×1 transmission (or any case where only two eigenmodes are used, i.e. only 2 q's are nonzero), one may exploit a structure introduced by Alamouti[13] which consists of transmission blocks of two time slots. It has been shown that this transmission structure can, in principle, attain the full capacity of a 2×1 channel. This statement remains true even when the channel has nontrivial correlations. For cases when one wants to transmit n different modes with $n > 2$, simple transmission structures have not yet been devised that attain full capacity. However, several structures have been proposed that come close to the full capacity. For example, for 4×1 transmission, a simple scheme with a block of length 4 in time has been discussed in [9], and has been simulated for i.i.d. channels. We believe this structure should also be effective for correlated channels. Other

generalizations of these structures (so-called LD codes) could similarly be considered[14]. In general, however, the question of coding for nx1 MISO systems remains open.

Appendix

I. Details for Case $\mathbf{G} = \mathcal{N}(0, \Sigma)$

In this Appendix we derive the PDF and CCDF for the simple case when the ensemble of channels follows $\mathcal{N}(0, \Sigma)$.

We start with Eq. 7. The expectation of an arbitrary operator O can be written explicitly as

$$\langle O \rangle = \int d\mu(\mathbf{G}, \mathbf{G}^\dagger)\, e^{-\mathbf{G}^\dagger \Sigma^{-1} \mathbf{G}}\, O \tag{46}$$

where the integration measure is defined by

$$\int d\mu(\mathbf{G}, \mathbf{G}^\dagger) = \det \Sigma^{-1} \prod_{i=1}^{n} \frac{1}{\pi} \int d\mathrm{Re}G_i \int d\mathrm{Im}G_i \tag{47}$$

Thus we write $\mathrm{PDF}(I)$ as

$$\begin{aligned} \mathrm{PDF}(I) &= \int d\mu(\mathbf{G}, \mathbf{G}^\dagger)\, e^{-\mathbf{G}^\dagger \Sigma^{-1} \mathbf{G}}\, \delta(I - \log(1 + \mathbf{G}^\dagger p Q \mathbf{G})) & (48) \\ &= \int d\mu(\mathbf{G}, \mathbf{G}^\dagger)\, e^{-\mathbf{G}^\dagger \Sigma^{-1} \mathbf{G}}\, e^I\, \delta(e^I - 1 - \mathbf{G}^\dagger p Q \mathbf{G}) & (49) \end{aligned}$$

The delta function is now written as a Fourier transform to yield

$$\begin{aligned} \mathrm{PDF}(I) &= \int d\mu(\mathbf{G}, \mathbf{G}^\dagger)\, e^{-\mathbf{G}^\dagger \Sigma^{-1} \mathbf{G}}\, e^I \int \frac{dk}{2\pi} \exp\left(ik(e^I - 1 - \mathbf{G}^\dagger p Q \mathbf{G}) \right) & (50) \\ &= e^I \int \frac{dk}{2\pi} e^{ik(e^I - 1)} \int d\mu(\mathbf{G}, \mathbf{G}^\dagger)\, e^{-\mathbf{G}^\dagger [\Sigma^{-1} + ikpQ] \mathbf{G}} & (51) \end{aligned}$$

We can then do the integral over $\mathbf{G}$ to obtain

$$\begin{aligned} \mathrm{PDF}(I) &= \frac{e^I}{\det \Sigma} \int \frac{dk}{2\pi} \frac{e^{ik(e^I - 1)}}{\det[\Sigma^{-1} + ikpQ]} & (52) \\ &= \frac{e^I}{\det(i\Sigma Qp)} \int \frac{dk}{2\pi} \frac{e^{ik(e^I - 1)}}{\det[k - i(\Sigma Qp)^{-1}]} & (53) \end{aligned}$$

Let us define $1/a_i$ to be the eigenvalues of the matrix ΣQp. We then have

$$\mathrm{PDF}(I) = e^I \int \frac{dk}{2\pi} \frac{e^{ik(e^I - 1)}}{\prod_{i=1}^{n} (1 + ik/a_i)} \tag{54}$$

II. Details for Case $\mathbf{G} = \mathcal{N}(\mathbf{g}, \Sigma)$

A. Derivation of PDF, CCDF

For an ensemble of channels given by $\mathcal{N}(\mathbf{g}, \Sigma)$, it is convenient to separate out the constant piece of $\mathbf{G}$ and write

$$\mathbf{G} = \mathbf{g} + \mathbf{z} \tag{55}$$

such that $I = \log(1 + (\mathbf{g}^\dagger + \mathbf{z})pQ(\mathbf{g} + \mathbf{z}))$ where $\mathbf{g}$ is a given constant and $\mathbf{z}$ is a gaussian random vector with zero mean and covariance Σ, i.e. $\mathbf{z} = \mathcal{N}(0, \Sigma)$.

We write the probability distribution function in the form of Eq. 7. This can then be expressed similarly to Eq. 48 as

$$\mathrm{PDF}(I) = \int d\mu(\mathbf{z}, \mathbf{z}^\dagger)\, e^{-\mathbf{z}^\dagger \Sigma^{-1} \mathbf{z}}\, \delta(I - \log(1 + [\mathbf{g}^\dagger + \mathbf{z}^\dagger] pQ [\mathbf{g} + \mathbf{z}])) \tag{56}$$

Following the same steps as in Appendix A after Eq. 48 we reach to a form analogous to Eq. 54

$$\mathrm{PDF}(I) = e^{I - \mathbf{g}^\dagger \Sigma^{-1} \mathbf{g}} \int \frac{dk}{2\pi} \frac{e^{ik(e^I - 1)}}{\prod_j (1 + ik/a_j)} \exp\left[\sum_i \frac{b_i}{1 + ik/a_i}\right] \tag{57}$$

where

$$b_i = p\, a_i |\mathbf{g}^\dagger Q^{1/2} \mathbf{A}^i|^2 \tag{58}$$

In Eqs. 57 and 58 a_i and $\mathbf{A}^i$ are the eigenvalues and the respective normalized eigenvectors of $(pQ^{1/2}\Sigma Q^{1/2})^{-1}$. We note that since the eigenvectors form a complete set, we have $\sum_i b_i = \mathbf{g}^\dagger \Sigma^{-1} \mathbf{g}$. To obtain the cumulative distribution function of I we integrate over I to get

$$\begin{aligned} \mathrm{CCDF}(I) &= \int_I^\infty dI'\, \mathrm{PDF}(I') \\ &= \int \frac{dk}{2\pi} \frac{e^{ik(e^I - 1)}}{0^+ - ik} \frac{e^{-\mathbf{g}^\dagger \Sigma \mathbf{g}}}{\prod_j (1 + ik/a_j)} \exp\left[\sum_i \frac{b_i}{1 + ik/a_i}\right] \end{aligned} \tag{59}$$

Integrals of the type shown in Eq. 59 and 57 can be manipulated conveniently by using transformations of the form

$$\int \frac{dk}{2\pi} \frac{e^{ikz + B/(a+ik)}}{a + ik} = \frac{1}{\pi} \int dy dy^* \int \frac{dk}{2\pi} e^{ikz - |y|^2(a+ik) + \sqrt{B}(y + y^*)} \tag{60}$$

$$= \frac{1}{2\pi} \int_0^\infty d|y|^2 \int d\theta \delta(z - |y|^2) e^{-a|y|^2 + 2\sqrt{B}|y|\cos(\theta)} \tag{61}$$

$$= e^{-az} I_0(2\sqrt{Bz}) \tag{62}$$

with I_0 the modified Bessel function. (The case where there are multiple identical poles can be handled by taking successive derivatives with respect to B, and multiple different poles is handled with a different y_i integration for each different a_i). Transformations of this type can be used to obtain, for example, Eqs. 35 and 38 from the above Eq. 59. Most generally, for arbitrary Q and Σ we can use this approach to transform the above result into the useful form

$$\mathrm{CDF}(I) = e^{-\mathbf{g}^\dagger \Sigma \mathbf{g}} \left[\prod_{j=0}^{n} \int_0^\infty dz_j \right] \delta(e^I - 1 - \sum_{j=0}^{n} z_j) \prod_{j=1}^{n} a_j e^{-z_j a_j} I_0(2\sqrt{z_j a_j b_j}) \tag{63}$$

B. Derivation of $\langle I \rangle$

To get an expression for $\langle I \rangle$ we use the identity

$$\log(x) = \int_0^\infty \frac{dy}{y} [e^{-y} - e^{-xy}] \tag{64}$$

to express the log in the definition of $I = \log(1 + \mathbf{G}pQ\mathbf{G})$. Applying Eq. 46 for the case $G = \mathcal{N}(\mathbf{g}, 1)$ we get

$$\langle I \rangle = \int d\mu(\mathbf{z}, \mathbf{z}^\dagger)\, e^{-\mathbf{z}^\dagger \Sigma^{-1} \mathbf{z}} \int_0^\infty \frac{dy\, e^{-y}}{y} \left[1 - \exp\left(-y[\mathbf{g}^\dagger + \mathbf{z}^\dagger] pQ[\mathbf{g} + \mathbf{z}] \right) \right] \tag{65}$$

Integrating out the $\mathbf{z}$ variables we get

$$\langle I \rangle = \int_0^\infty \frac{dy\, e^{-y}}{y} \left[1 - \frac{\exp\left(-\mathbf{g}^\dagger \Sigma^{-1/2} \left[\frac{\Sigma^{1/2} Qp \Sigma^{1/2} y}{1 + \Sigma^{1/2} Qp \Sigma^{1/2} y} \right] \Sigma^{-1/2} \mathbf{g} \right)}{\det[1 + \Sigma^{1/2} Qp \Sigma^{1/2} y]} \right] \tag{66}$$

C. Optimization over Q *when* $\Sigma = 1$.

We first consider the case of the ergodic capacity $\langle I \rangle$ in Eq. 66 with $\Sigma = 1$. We see that the integrand depends on Q through its eigenvalues (in the determinant of the denominator) and the quantity $\zeta = \mathbf{g}^\dagger Qpy/(1 + Qpy)\mathbf{g}$. In addition the integrand is an *increasing* function of ζ. If we fix $\mathbf{g}$ and the eigenvalues of Q and allow the basis of eigenvectors of Q to vary (ie, consider a unitary rotation of Q), we find that ζ becomes a maximum when the eigenvector corresponding with the largest eigenvalue of Q coincides with $\mathbf{g}$. Therefore the optimal Q has $\mathbf{g}$ as the eigenvector with the largest eigenvalue. Furthermore, given that the determinant of a positive definite matrix with fixed trace is a maximum when the matrix is proportional to unity, we conclude that all eigenvalues of

Q with vectors perpendicular to $\mathbf{g}$, are equal. Thus we must have the solution Q in the form of Eq. 33

We now turn to the more difficult case of outage capacity with $\Sigma = 1$. First we note that since CDF(I) is a monotonically increasing function, and since the outage is the inverse function of CDF (see Eqs. 10 and 11) finding the Q that maximizes $\mathcal{OUT}(P_{out})$ is equivalent to finding the Q that minimizes CDF(I). Fixing the eigenvalues a_i of $(pQ)^{-1}$, we examine the CDF in the form of Eq. 63. As mentioned above $\sum_{j=1}^{n} b_j = \mathbf{g}^\dagger\mathbf{g}$ is fixed, however the individual b_j's can be varied by making a unitary rotation of Q. Thus, we need to ask how to distribute the "weight" of the b_j's so as to minimize the CDF. We claim all of the weight must be concentrated in a single b_j corresponding to the smallest eigenvalue a_j.

To prove this claim, we write consider two eigenvalues $a_1 < a_2$ and imagine varying b_1 and b_2 subject to the constraint $b_1 + b_2 = b$. We write the CDF in the form of Eq. 63. Focusing on the z_0, z_1, and z_2 integrals (which we choose to do first), we write these integrals

$$\begin{aligned} \text{Int} &= \int_0^\infty dz_0 dz_1 dz_2\, \delta(C - z_0 - z_1 - z_2)\, e^{-a_1 z_1 - a_2 z_2} I_0(2\sqrt{z_1 a_1 b_1}) I_0(2\sqrt{z_2 a_2 b_2}) \\ &= \int_0^C dz_2 \int_0^{C-z_2} dz_1\, e^{-a_1 z_1 - a_2 z_2} I_0(2\sqrt{z_1 a_1 b_1}) I_0(2\sqrt{z_2 a_2 b_2}) \end{aligned} \tag{67}$$

where $C = e^I - 1 - \sum_{j=3}^{n} z_j$. Now, since I_0 is a monotonic function, and $z_2 > z_1$ and $a_2 > a_1$, we have

$$\text{Int} > \int_0^C dz_2 \int_0^{C-z_2} dz_1 e^{-a_1 z_1 - a_2 z_2} I_0(2\sqrt{z_1 a_1 b_1}) I_0(2\sqrt{z_1 a_1 b_2}) \tag{68}$$

$$> \int_0^C dz_2 \int_0^{C-z_2} dz_1 e^{-a_1 z_1 - a_2 z_2} I_0(2\sqrt{z_1 a_1 (b_1 + b_2)}) \tag{69}$$

The final step follows from the fact that $I_0(\sqrt{x}) I_0(\sqrt{y}) > I_0(\sqrt{x+y})$ with $x, y > 0$ (which can be established by noting that $\log I_0(\sqrt{x})$ is a convex function). The last line here (Eq. 69) is exactly what we would get if we had put all of the weight ($b = b_1 + b_2$) in the direction corresponding to a_1, thus proving our claim.

We now rewrite the CDF using the fact that only a single b_j is nonzero (we will call this one b_n) separating out this integral

$$\text{CDF}(I) = a_n e^{-\mathbf{g}^\dagger\mathbf{g}} \int_0^{e^I - 1} dz_n e^{-z_n a_n} I_0(2\sqrt{z_n a_n b_n}) \tilde{C}(e^I - 1 - z_n) \tag{70}$$

where

$$\tilde{C}(x) = \left[\prod_{j=1}^{n-1} \int_0^\infty dz_j\right] \delta(x - \sum_{j=0}^{n-1} z_j) \prod_{j=1}^{n-1} a_j e^{-z_j a_j}. \quad (71)$$

Comparing to Eq. 63 we see that $\tilde{C}$ is nothing but the CDF for an $n-1$ dimensional system with $\mathbf{g} = 0$ i.e., with all b's set to zero (and with an effective mutual information $\tilde{I}$ given by $\tilde{I} = \log_2(1+x)$). We know that for such a system (with $\Sigma = 1$) the CDF is minimized (outage maximized) for Q being proportional to the unit matrix. Thus, the remaining $n-1$ eigenvalues of Q must all be equal to each other. Thus, we conclude that the optimal Q must have one eigenvector pointing in the direction $\mathbf{g}$ corresponding to the largest eigenvalue of Q and all other eigenvalues are equal to each other. This immediately implies that Q must have the form of Eq. 33.

Acknowledgments

One of the authors (ALM) would like to acknowledge illuminating discussions with S. Gollamudi and S. Nagaraj.

References

[1] E. Telatar, "Capacity of multi-antenna gaussian channels," *AT&T-Bell Labs, Internal Tech. Memo.*, pp. 287–290, June 1995.

[2] A. L. Moustakas *et al.*, "Communication through a diffusive medium: Coherence and capacity," *Science*, vol. 287, pp. 287–290, January 2000.

[3] A. M. Sengupta and P. P. Mitra, "Capacity of multivariate channels with multiplicative noise: I. random matrix techniques and large-n expansions for full transfer matrices," *LANL arXiv:physics/0010081*, October 31 2000.

[4] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2632–2639, Sept 2001.

[5] S. A. Jafar and A. Goldsmith, "On optimality of beamforming for multiple antenna systems," *Proceedings, IEEE International Symposium on Information Theory*, p. 321, 2001.

[6] A. Narula *et al.*, "Efficient use of side information in multple-antenna data transmission over fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1423–1436, October 1998.

[7] A. L. Moustakas *et al.*, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large N analysis," *submitted to IEEE J. Select. Areas Comm.* May, 2002.

[8] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1988.

[9] C. B. Papadias and G. J. Foschini, "A space-time coding approach for systems employing 4 transmit antennas," *Proceedings, IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, pp. 2481–2484, 2001.

[10] S. A. Jafar and A. Goldsmith, "Beamforming capacity and SNR maximization for muliple antenna systems with covariance feedback," *Vehicular Technology Conference (VTC), Rhodes Greece*, 2001.

[11] S. H. Simon and A. L. Moustakas, "Optimizing MIMO systems with channel covariance feedback," *submitted to IEEE J. Select. Areas Comm.*, May 2002.

[12] David G. Luenberger, *Optimization by Vector Space Methods*, Wiley-Interscience, New York, 1969.

[13] S. M. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct 1998.

[14] B. Hassibi and B. M. Hochwald, "High rate codes that are linear in space and time," *preprint*, 2000.

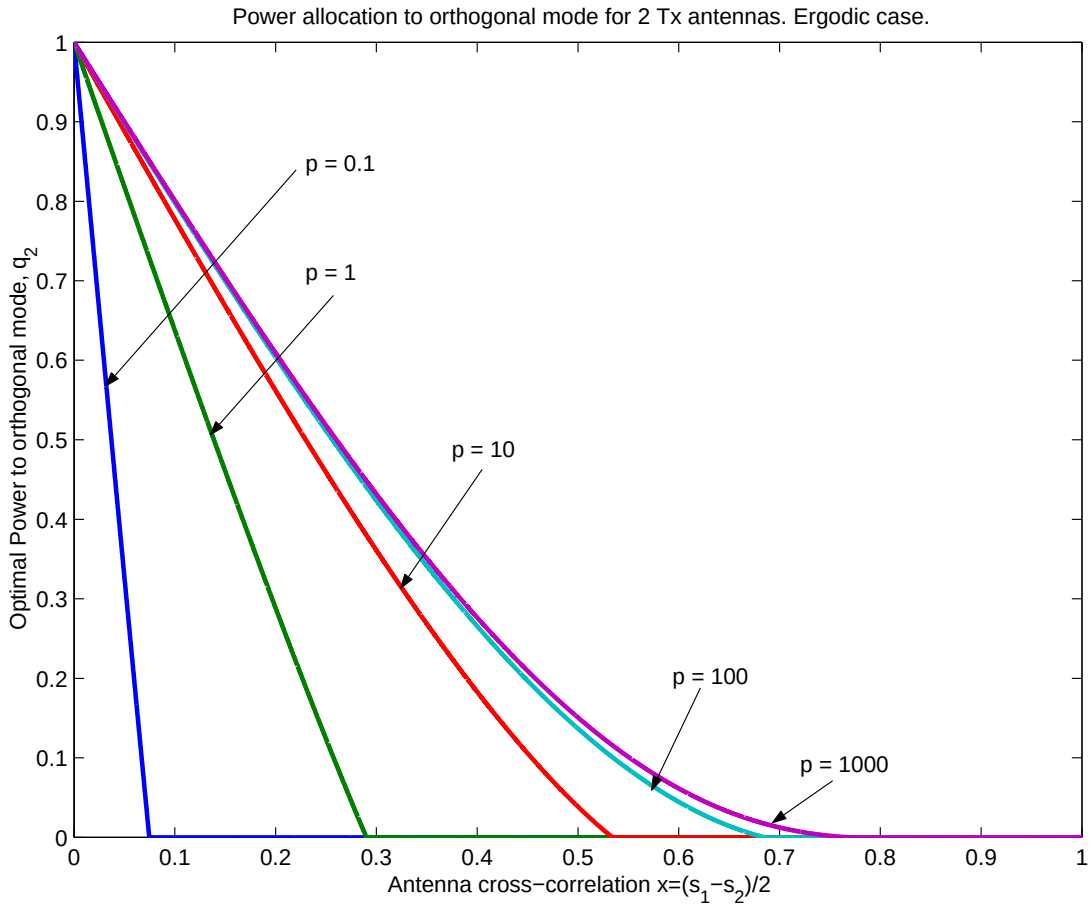


Fig. 1. Optimal power to nonbeamforming mode q_2 for the two transmit antenna case, as a function of the antenna cross-correlation $x = (s_1 - s_2)/2$. Here $q_2 = 0$ corresponds to beamforming and $q_2 = 1$ corresponds to power being equally distributed between each of the two transmission eigenmodes. We have shown curves for various values of the parameter p which is proportional to the signal to noise level per antenna (See section I-B for the precise definition of p). Note that at low SNR, beamforming becomes optimal at small antenna cross-correlation. These results were calculated using Eq. 28.

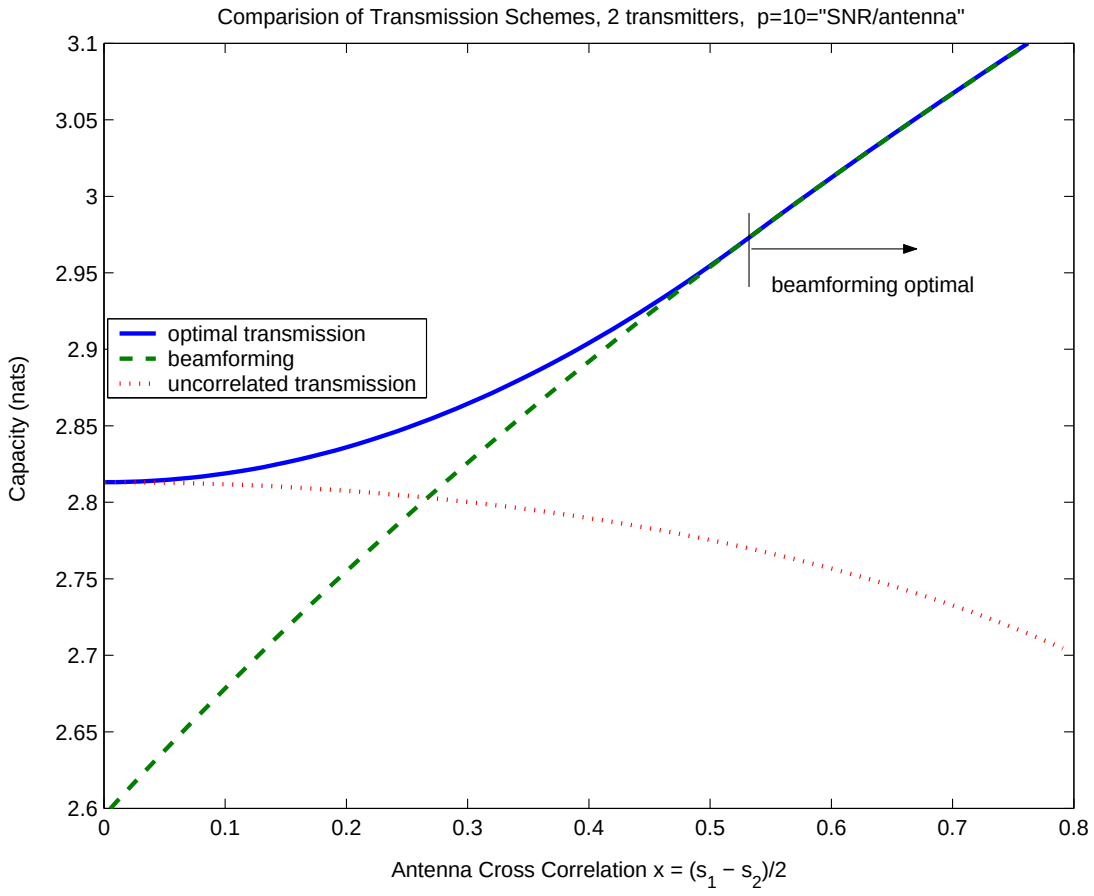


Fig. 2. Comparison of ergodic capacities (mutual informations) for different modes of transmission for two transmitters as a function of antenna cross-correlation $x = (s_1 - s_2)/2$. Shown are curves for beamforming ($q_1 = 2, q_2 = 0$), uncorrelated transmission ($q_1 = q_2 = 1$), and the result for optimal values of q_1, q_2 (as shown in Fig 1). This figure is for $p = 10$ which fixes the transmitted power level and corresponds precisely to SNR/antenna $= 10$ for the case of uncorrelated transmission. (See section I-B for the precise definition of p).

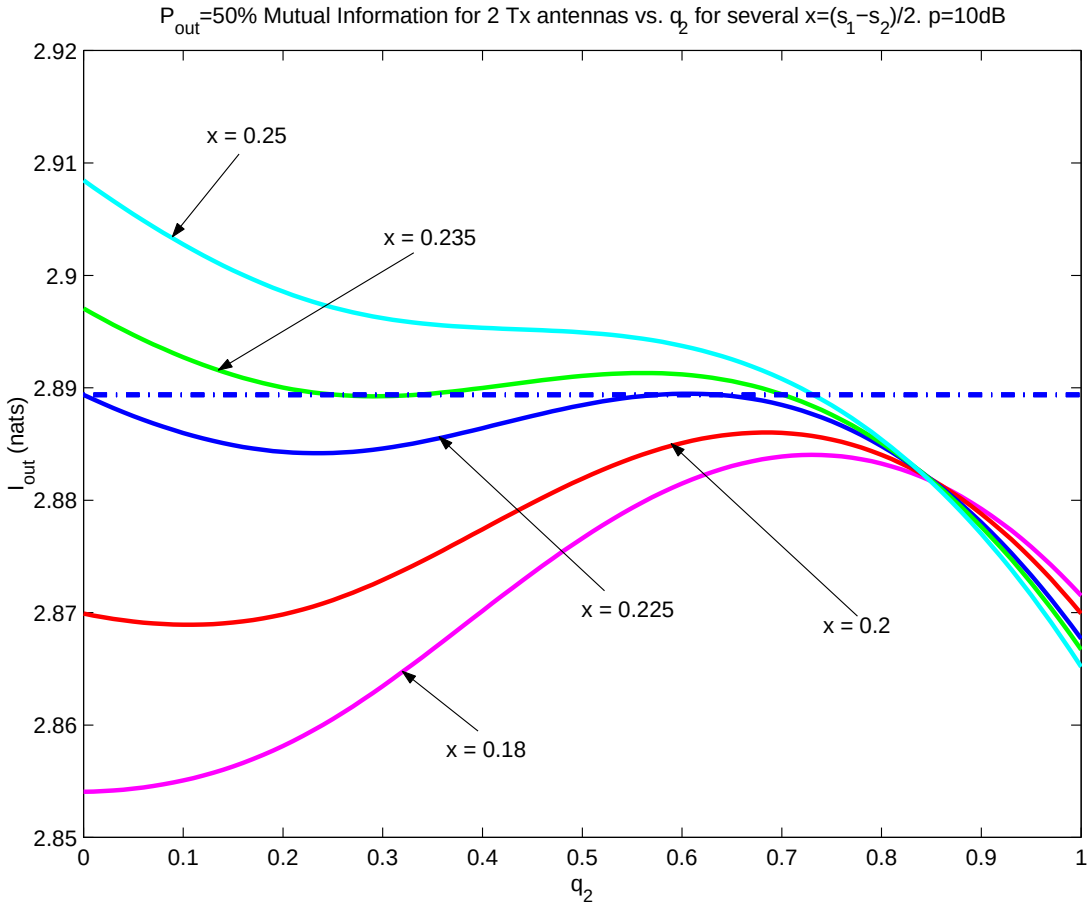


Fig. 3. 50% Outage Capacity Plot: versus power q_2 to the orthogonal mode for different antenna cross-correlations $x = (s_1 - s_2)/2$ with p =SNR/antenna $= 10$ (See section I-B for the precise definition of p). We see that the mutual information is a nonmonotonic function of q_2 and the optimal value of q_2 jumps (at roughly the dotted line) from a finite value to zero. For $x > 0.225$ beamforming $q_2 = 0$ is optimal. However, for $x < 0.225$ the optimal value of q_2 is greater than 0.6.

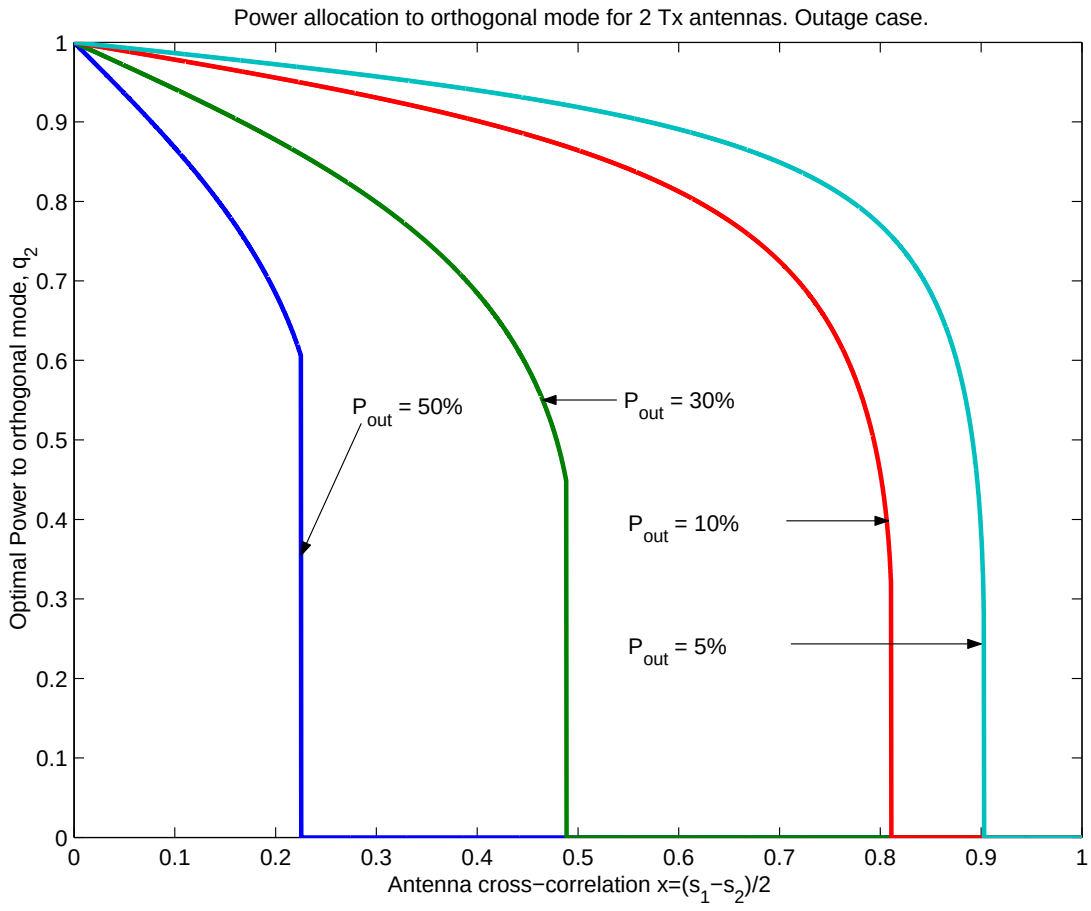


Fig. 4. Outage Capacity: Plot of optimal q_2 for the two transmit antenna case, as a function of the antenna cross-correlation $x = (s_1 - s_2)/2$. Here q_2 was calculated by setting Eq. 30 to zero for different outage probabilities P_{out} (probability that the mutual information is greater than the given value). For all outages, for large antenna cross-correlation, beamforming ($q_2 = 0$) becomes optimal, whereas for small cross-correlations, independent transmission ($q_1 = q_2 = 1$) is optimal. For lower outages, the crossover to beamforming occurs at higher antenna cross correlations. This is just a reflection of diversity protecting the link from fading.

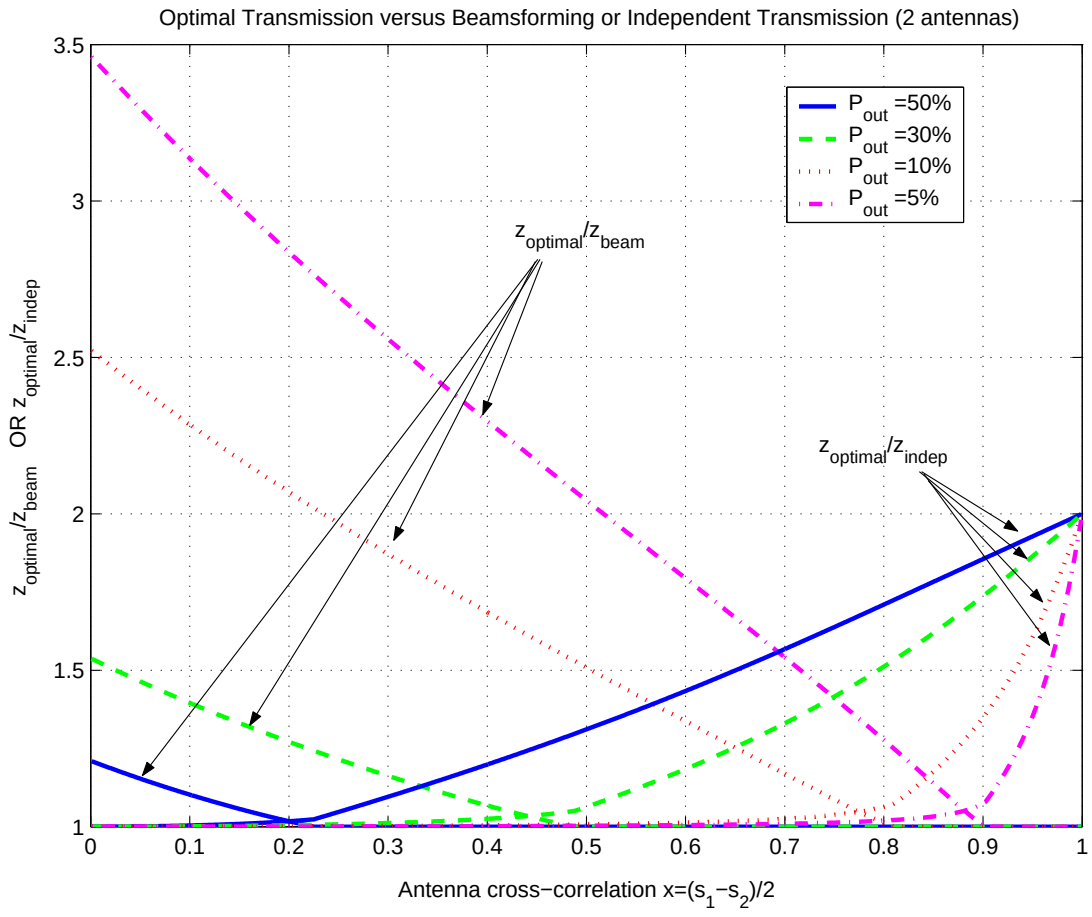


Fig. 5. Comparison outage capacity for optimized transmission, beamforming, and independent transmission as a function of antenna cross-correlation $x = (s_1 - s_2)/2$ and outage P_{out}. Here we have plotted ratios of z parameters ($z = (e^I - 1)/p$ note that the factors of p cancel in the z ratios). At small I_{out} (low SNR), the ratio of z parameters is equal to the ratio of the outage capacities. Thus, we interpret the graph as showing I_{out} for optimal transmission divided by I_{out} for beamforming or independent transmission. At high I_{out} (high SNR) the log of the z ratio becomes the difference in information capacities. Thus we see that for all antenna cross-correlations and all outages, either beamforming is very close optimal or independent transmission is very close to optimal. A small amount of additional capacity can be obtained by using fully optimized transmission covariance. For example, from the figure we see that for 5% outage and cross-correlation $x \approx 0.88$, roughly 5% additional outage capacity (at low SNR) can be obtained by using optimize transmission correlations rather than beamforming or independent transmission.

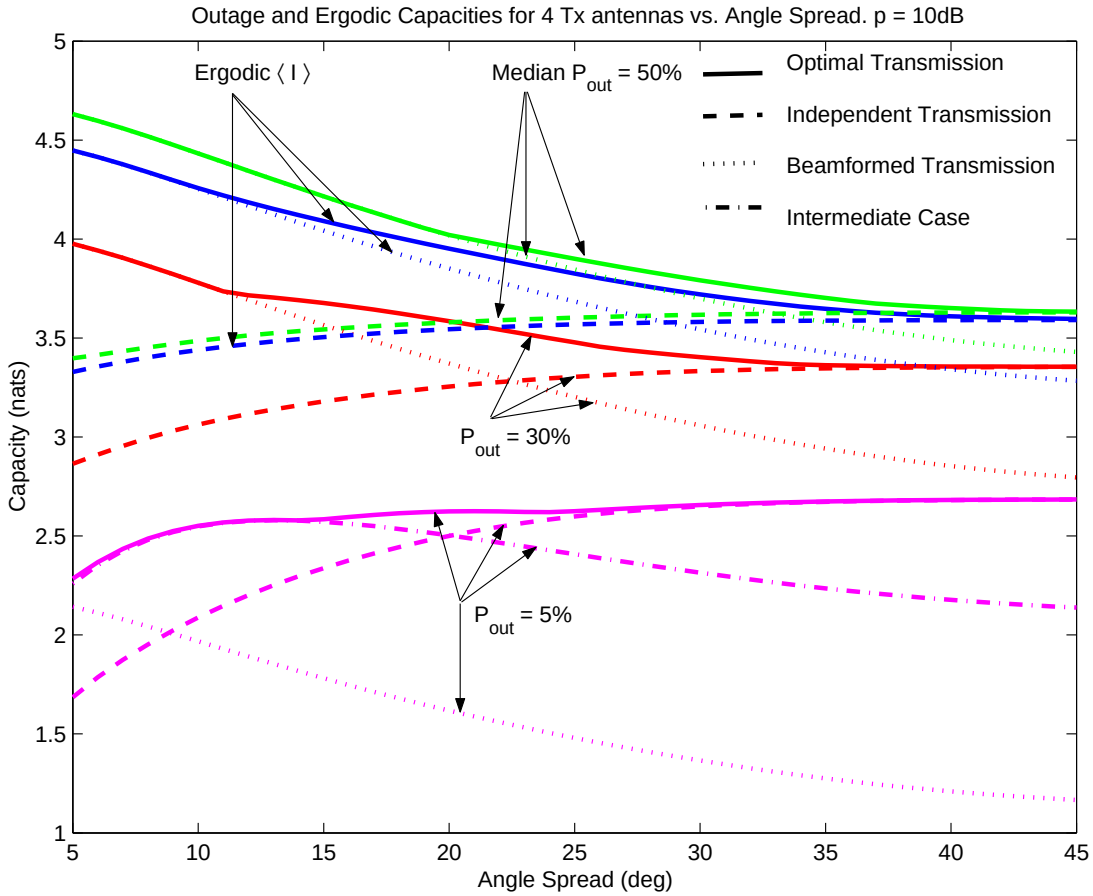


Fig. 6. Capacity as a function of angle spread at the transmitter for a uniformly spaced 4-antenna transmit array. The solid curves correspond to optimal transmission, the dotted curves correspond to beamforming ($q_1 = 4,\ \ q_2 = q_3 = q_4 = 0$), and the dashed curves correspond to independent transmission ($q_1 = q_2 = q_3 = q_4$). The dot-dashed curve (shown only for 5% outage) is an intermediate case using two modes only ($q_1 = q_2 = 2,\ \ q_3 = q_4 = 0$). The antenna correlation matrix is given by Eq. 32. The power level is $p = 10 \approx$ SNR/antenna. (See section I-B for the precise definition of p) We have shown here curves corresponding to the ergodic capacity as well as several outage capacities. We note that the 50% (median) outage capacity is very close to the ergodic capacity, reflecting the fact that the PDF is roughly symmetric around its average. Note that for low outages, one can obtain a substantial gain (roughly 20% near an angle spread of 10 degrees) by using the optimal transmission rather than beamforming or independent transmission. Note however, that the intermediate case of using only two modes is almost optimal in this range where both beamforming and independent transmission are both poor.

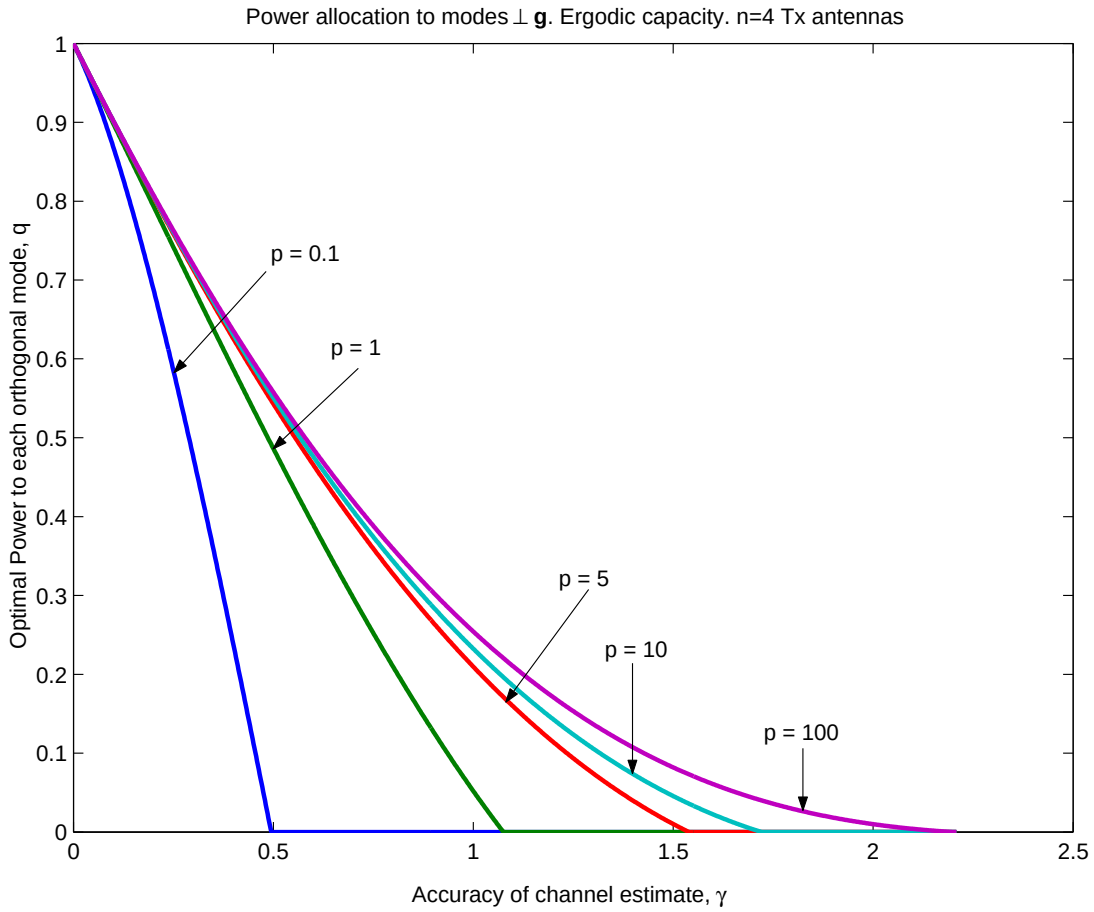


Fig. 7. Ergodic case: Optimal power allocation to the orthogonal modes q in the case of channel with nonzero average $\mathbf{g}$ and covariance $\Sigma = 1$. $\mathbf{g}$ can be thought of as the channel estimate in a closed-loop system with some error. Here, q is plotted as a function of the channel estimate quality γ for several fluctuation-to-noise strengths p. (Here γ is roughly n times the ratio of the power from the constant part of the channel to the power from the fluctuating part of the channel; and p is roughly the ratio of the power per antenna from the fluctuating part of the channel to the noise power. See section I-B for the precise definitions of p and γ). In this figure q is optimized by maximizing the average mutual information for fixed γ. $q = 0$ corresponds to the beamforming solution and $q = 1$ corresponds to n independent transmissions. As one might expect, as the channel estimate gets better, beamforming becomes increasingly optimal. Interestingly, the crossover to beamforming occurs more quickly for smaller p.

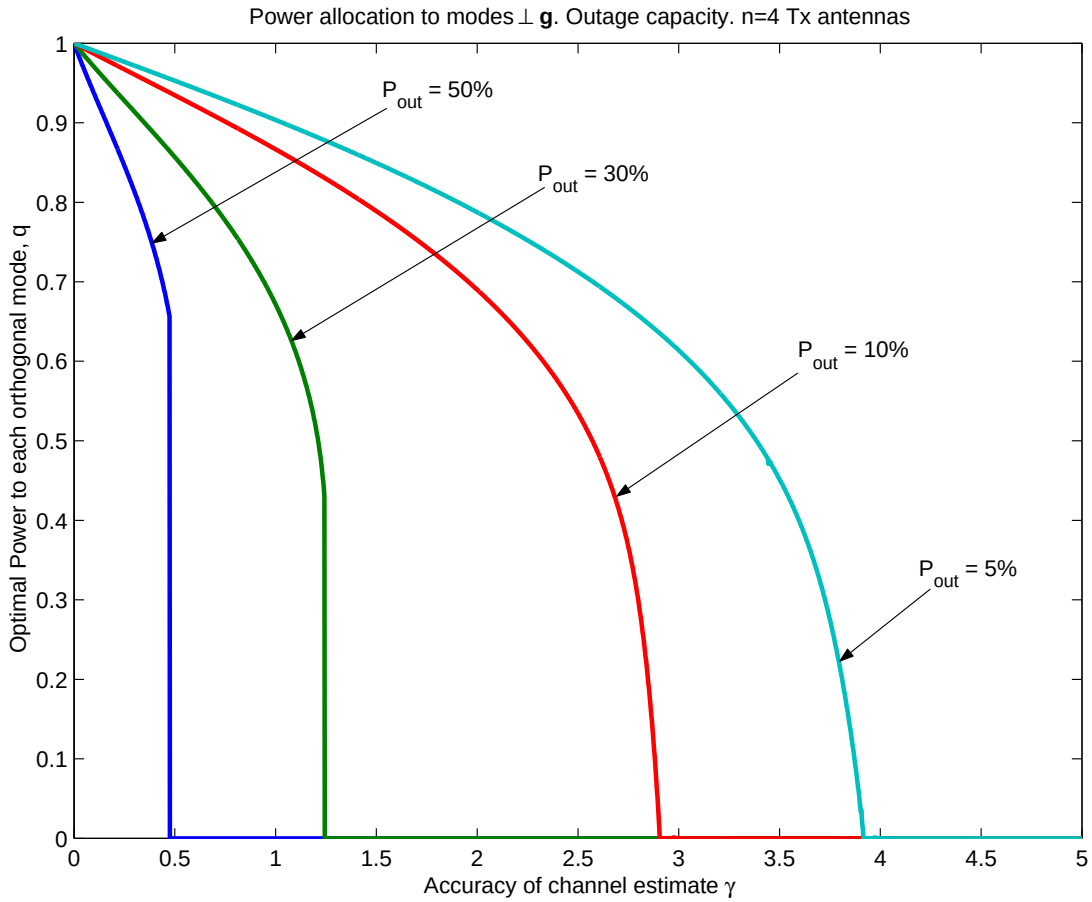


Fig. 8. Outage case: Optimal power allocation to the orthogonal modes q in the case of a non-zero mean channel $\mathbf{g}$ with white covariance. We can think of $\mathbf{g}$ as the channel estimate in a closed-loop system with some error. We have plotted q as a function of the channel estimate quality γ for several outage probabilities P_{out} (Here γ is roughly n times the ratio of the power from the constant part of the channel to the power from the fluctuating part of the channel. See section I-B for the precise definitions of γ). Note that for $P_{out} = 0.5$ (50% outage) and $P_{out} = 0.3$ (30% outage), q drops to zero discontinuously, signifying a non-monotonic outage mutual information as a function of q. For smaller P_{out} ($P_{out} < 0.226$) and larger γ ($\gamma > 1.670$) the mutual information becomes strictly convex. As one would expect, for better channel estimate, beamforming becomes more favored. On the other hand, for lower outage, beamforming is disfavored (since it is more susceptible to fading).