Optimizing Multiple-Input Single-Output (MISO) Communication Systems With General Gaussian Channels: Nontrivial Covariance and Nonzero Mean

Aris L. Moustakas, *Member, IEEE,* and Steven H. Simon

***Abstract*—In this correspondence, we consider a narrow-band point-to-point communication system with many (input) transmitters and a single (output) receiver (i.e., a multiple-input single output (MISO) system). We assume the receiver has perfect knowledge of the channel but the transmitter only knows the channel distribution. We focus on two canonical classes of Gaussian 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 the identity. In both cases, we are able to derive simple analytic expressions for the ergodic average and the cumulative distribution function (cdf) of the mutual information for arbitrary input (transmission) signal covariance. With minimal numerical effort, we then determine the ergodic and outage capacities and the corresponding capacity-achieving input signal covariances. Interestingly, we find that the optimal signal covariances for the ergodic and outage cases have very different behavior. In particular, under certain conditions, the outage capacity optimal covariance is a discontinuous function of the parameters describing the channel (such as strength of the correlations or the nonzero mean of the channel).**

***Index Terms*—Multiple antennas, side information, outage capacity, transmit diversity.**

I. Introduction

Multiple-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 at the transmitter and on whether the transmitter is able to take advantage of this information. For example, it is well known that in a spatially uncorrelated Rayleigh-fading environment, for a multiple-antenna transmitter with perfect channel knowledge and a single-antenna receiver (multiple-input single-output (MISO)), the gain in throughput due to the optimization of transmission is roughly $\log_2(n)$, where n is the number of transmitter antennas. The knowledge of the channel allows the transmitter to transmit with a signal covariance that maximizes the signal-to-noise ratio (SNR) at the receiver, thus increasing the mutual information.

In addition to the above so-called closed-loop case with instantaneous channel knowledge at the transmitter, the transmitter may have only statistical channel information, in particular, it may know the distribution of the channel. In this case, extra throughput is gained by the transmitter adapting the signal covariance to the full distribution of the channel. For example, the transmitter may know the long-term correlations between the transmitting antennas (the channel covariance), which could be modeled as a zero-mean Gaussian channel with nontrivial covariance. Alternatively, it may have knowledge of the mean channel, which could be modeled by a Gaussian channel with a nonzero mean and covariance proportional to the identity matrix. The idea of maximizing the mutual information based on statistical information of the channel at the transmitter was proposed in the more general multiple-input multiple-output (MIMO) context by Moustakas *et al.* [2] and Sengupta and Mitra [3] and also 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 amount of mutual information occurring in all but $x\%$ of the instantiations of the channel [7]. For example, 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 mutual information would occur 95% of the time. Typically, system design aims to optimize either the ergodic average or the outage mutual information for some given outage probability. In this correspondence, we will aim to determine the input signal covariances at the transmitter that optimizes each 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.

The approach 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 channel instantiations, we show that, for a Gaussian channel with a nonzero mean and a covariance matrix proportional to the identity, the problem of calculating the ergodic or outage mutual information as a function of the input signal covariance reduces to the evaluation of a single integral of known functions. Furthermore, for a Gaussian 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. Note that several of the analytic results we present in this correspondence have also been derived by other authors [4], [5], [12], [13]. We will mention these results briefly to make comparison with new cases that have not been previously studied analytically. We note that recently we generalized the methods described here to calculate the outage capacity for the two transmit antenna MISO case with 2-bits feedback information [21]. We also 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 [8].

A. Outline

After some general definitions (Sections 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 Gaussian 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 nonzero mean Gaussian channel with trivial covariance will be assumed, and similar calculations are performed. (The most general case of nonzero mean Gaussian channel with nontrivial covariance is also discussed in Appendix I.)

Manuscript received May 17, 2002; revised May 1, 2003. The material in this correspondence was presented in part at the DIMACS Workshop on Signal Processing for Wireless Communications, Rutgers University, Piscataway, NJ, October 2002.

A. L. Moustakas is with Bell Laboratories, Lucent Technologies, Whippany, NJ 07981 USA (e-mail: arislm@lucent.com).

S. H. Simon is with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA (e-mail: shsimon@lucent.com).

Communicated by G. Caire, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2003.817464

B. Definition of Channel

In the narrow-band MISO problem—n transmitters and a single receiver—the channel is defined by

$$y = \sqrt{p}\sum_{i=1}^{n} g_i x_i + \eta = \sqrt{p}\,\boldsymbol{g}^T \boldsymbol{x} + \eta \tag{1}$$

where y is the complex received signal, $\boldsymbol{x}$ is the complex transmitted signal vector, with the ith element x_i being the transmitted signal from antenna i, $\boldsymbol{g}$ is the complex channel vector, and η is the complex noise at the receiver. Here, p is a normalization constant signifying roughly the received SNR of the channel. In this correspondence, capital boldface quantities will denote $n \times n$ matrices, while lower case boldface letters will represent n-dimensional column vectors, and superscripts † and T will denote the Hermitian conjugate and transpose, respectively.

The channel statistics are given by the conditional probability

$$P(\boldsymbol{g},\, y|\boldsymbol{x}) = P(\boldsymbol{g})P(y|\boldsymbol{g},\, \boldsymbol{x})$$

where the equality holds due to the independence of y and $\boldsymbol{g}$. These distributions are assumed to be known at the transmitter. $P(y|\boldsymbol{g},\, \boldsymbol{x})$ describes the distribution of the noise η, which is assumed to be zero-mean Gaussian with unit variance $E\{\eta^*\eta\} = 1$, where $E\{\cdot\}$ denotes the expectation value (setting the noise power to unity defines our unit of power). The transmitted signal $\boldsymbol{x}$ is taken to be a Gaussian vector with covariance matrix given by $E\{\boldsymbol{x}\boldsymbol{x}^\dagger\} = \boldsymbol{Q}$ with $\boldsymbol{Q}$ a nonnegative definite Hermitian matrix. $\boldsymbol{x}$ has been normalized so that the constraint of the average total transmitted power can be expressed as $\mathrm{Tr}\{\boldsymbol{Q}\} = n$.

The channel $\boldsymbol{g}$ is defined to be $\mathcal{CN}(\boldsymbol{g}_0,\, \vec{\boldsymbol{\Sigma}})$, i.e., a complex Gaussian vector channel with mean $\boldsymbol{g}_0$ and covariance $\boldsymbol{\Sigma}$ (with $\boldsymbol{\Sigma}$ a Hermitian nonnegative definite matrix), normalized so that $\mathrm{Tr}\{\boldsymbol{\Sigma}\} = n$. The norm of $\boldsymbol{g}_0$ is given by

$$\gamma = \boldsymbol{g}_0^\dagger \boldsymbol{g}_0. \tag{2}$$

For the case of $\boldsymbol{g}_0 = \boldsymbol{0}$ and $\boldsymbol{Q} = \boldsymbol{I}_n$, p is precisely the SNR per transmitting antenna. Similarly, if transmissions are uncorrelated ($\boldsymbol{Q} = \boldsymbol{I}_n$ again), the SNR from the constant part of the channel is given by $p\gamma$. Finally, we note that we will also frequently use the notation that a_i are the eigenvalues of $(p\boldsymbol{Q}\boldsymbol{\Sigma})^{-1}$. Similarly, q_i are the eigenvalues of $\boldsymbol{Q}$ and s_i are the eigenvalues of $\boldsymbol{\Sigma}$.

C. Definitions of Quantities to Calculate

For a given instantiation of $\boldsymbol{g}$, if the receiver knows the channel [1] (which in practice is achieved by sending pilots), the mutual information $I(\boldsymbol{x};\, y|g = \boldsymbol{g})$, which is equal to $I((\boldsymbol{x},\, \boldsymbol{g});\, y)$ since $I(\boldsymbol{x};\, \boldsymbol{g}) = 0$, is given by

$$I(\boldsymbol{x};\, y|g = \boldsymbol{g}) = \log(1 + \boldsymbol{g}^\dagger p\boldsymbol{Q}\boldsymbol{g}). \tag{3}$$

Here the notation $g = \boldsymbol{g}$ implies that the channel is fixed at the value $\boldsymbol{g}$. Throughout this correspondence, we will measure information in nats where 1 nat is equal to e bits ($e = 2.718\cdots$).

One quantity of interest is the so-called ergodic average, i.e., the mean, of the mutual information I. We write

$$\langle I\rangle = \langle \log(1 + \boldsymbol{g}^\dagger p\boldsymbol{Q}\boldsymbol{g})\rangle \tag{4}$$

where the brackets $\langle\cdot\rangle$ represent an ensemble average over realizations of $\boldsymbol{g}$.

Another quantity of interest is the complementary cumulative distribution function (ccdf)

$$\mathrm{ccdf}(I) = \langle \Theta[I - \log(1 + \boldsymbol{g}^\dagger p\boldsymbol{Q}\boldsymbol{g})]\rangle \tag{5}$$

where $\Theta(x)$ is the step function ($\Theta(x) = 1$ for $x > 0$ and is 0 otherwise). We also define the cumulative distribution function (cdf) $\mathrm{cdf}(I) = 1 - \mathrm{ccdf}(I)$. Note that the ergodic average as well as the cdf and ccdf are implicit functions of $\boldsymbol{Q}$, p, $\boldsymbol{g}_0$, and $\boldsymbol{\Sigma}$. We will not usually make all of these dependencies explicit.

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 $\boldsymbol{Q}$ and $\boldsymbol{\Sigma}$, we define $\mathcal{OUT}(P_{\mathrm{out}})$ such that

$$I_{\mathrm{out}} = \mathcal{OUT}(P_{\mathrm{out}}) \tag{6}$$

when

$$P_{\mathrm{out}} = \mathrm{cdf}(I_{\mathrm{out}}) = 1 - \mathrm{ccdf}(I_{\mathrm{out}}). \tag{7}$$

Since the cdf is monotonic, the inversion is unique. The meaning of $I_{\mathrm{out}} = \mathcal{OUT}(P_{\mathrm{out}})$ is that there is a probability P_{out} that in any instantiation of $\boldsymbol{g}$ 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.

The ccdf (and cdf) as well as the outage $\mathcal{OUT}$ have an important "scaling" property. To see this, we note that, since exponentiation is monotonic, we can rewrite (5) as

$$\mathrm{ccdf}(I) = \langle \Theta[z - \boldsymbol{g}^\dagger \boldsymbol{Q}\boldsymbol{g}]\rangle \tag{8}$$

where

$$z = \frac{e^I - 1}{p}. \tag{9}$$

Thus, the ccdf is only a function of the combined quantity z and not of p and I separately. This can be seen directly from (3). Thus, for a given $\boldsymbol{Q}$ and $\boldsymbol{\Sigma}$, a single calculation of the ccdf applies for all power levels p. Since the outage mutual information can be found by inverting the calculated ccdf, we can use this scaling property to then write

$$I_{\mathrm{out}} = \log\left[1 + pZ(P_{\mathrm{out}},\, \boldsymbol{Q},\, \boldsymbol{\Sigma},\, \boldsymbol{g}_0)\right] \tag{10}$$

where the function $Z = (e^{I_{\mathrm{out}}} - 1)/p$ depends only on the target outage probability P_{out}, as well as on $\boldsymbol{Q}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{g}_0$, but does not depend on the power level p. This implies that for a given channel ensemble (defined by $\boldsymbol{\Sigma}$ and $\boldsymbol{g}_0$), the optimal transmission covariance $\boldsymbol{Q}$ for a given outage probability P_{out} (i.e., that maximizes $Z(P_{\mathrm{out}},\, \boldsymbol{Q},\, \boldsymbol{\Sigma},\, \boldsymbol{g}_0)$) is *independent* of the power p. 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 may be more technologically relevant) an *easier* quantity to work with than the mean mutual information (which is usually considered to be simpler).

To reiterate, the purpose of this correspondence, whether we are concerned with ergodic or outage mutual information, is to ask how the transmitter (assumed to know $P(\boldsymbol{g})$ and $P(y|\boldsymbol{g},\, \boldsymbol{x})$) should choose the transmission covariance $\boldsymbol{Q}$ so as to maximize the mutual information.

II. Zero-Mean Channel With Nontrivial Covariance

We start by analyzing the case where the channel $\boldsymbol{g}$ has zero mean $\boldsymbol{g}_0 = \boldsymbol{0}$ and arbitrary covariance $\boldsymbol{\Sigma}$. In other words, the channel is $\mathcal{CN}(\boldsymbol{0},\, \boldsymbol{\Sigma})$. This situation has been often used as a model of wireless systems where the channel is constantly changing, but correlations in the channel are only slowly varying. For example, one could think of a block-fading channel where the covariance $\boldsymbol{\Sigma}$ is fed back at each block.

The mean mutual information $\langle I \rangle$ is derived in Appendix I-A. With a_i being the eigenvalues of $(p\boldsymbol{Q}\boldsymbol{\Sigma})^{-1}$, and assuming all the eigenvalues are different, we have

$$\langle I \rangle = \left(\prod_{i=1}^{n} a_i\right) \sum_{j=1}^{n} \frac{m(a_j)}{a_j \prod\limits_{k \neq j} (a_k - a_j)} \tag{11}$$

where

$$m(a) = -e^a \mathrm{Ei}(-a) = e^a \Gamma(0,\, a) \tag{12}$$

with Ei the exponential integral and

$$\Gamma(0,\, x) = -\mathrm{Ei}(-x) = \int_x^\infty dy \, \frac{\exp(-y)}{y}$$

the zeroth incomplete gamma function. This result has been derived independently in [12].

Similarly, in Appendix I-A, we derive the ccdf

$$\mathrm{ccdf}(I) = \left(\prod_{i=1}^{n} a_i\right) \sum_{j=1}^{n} \frac{g(a_j,\, I)}{a_j \prod\limits_{m \neq j} (a_m - a_j)} \tag{13}$$

where

$$g(a,\, I) = \exp(-a(e^I - 1)) \tag{14}$$

which has been derived independently in [13]. Note that, since each a_i is proportional to $1/p$, the ccdf does indeed have the scaling property mentioned above that it is only a function of the parameter z in (9). This can be seen by noting that each $a_i \propto 1/p$. In the next sections, it will be convenient to think of (13) and (7) in terms of z, $\boldsymbol{Q}$, $\boldsymbol{\Sigma}$, i.e., as

$$P_{\mathrm{out}} = \mathrm{cdf}(z_{\mathrm{out}},\, \boldsymbol{Q}\boldsymbol{\Sigma}) = 1 - \mathrm{ccdf}(z_{\mathrm{out}},\, \boldsymbol{Q}\boldsymbol{\Sigma}) \tag{15}$$

where $0 < P_{\mathrm{out}} < 1$ is the outage probability (that $\boldsymbol{g}^\dagger \boldsymbol{Q} \boldsymbol{g}$ is less than z_{out}).

We note that both (11) and (13) are continuous and finite when any two (or more) a's become equal to each other (see also the comment on this in Appendix I-A).

A. Optimization Over $\boldsymbol{Q}$

To reach the full information capacity of the communication link, the transmitter has to optimize the transmitting signal covariance matrix $\boldsymbol{Q}$. 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 $I_{\mathrm{out}} = \log(1 + pz_{\mathrm{out}})$ for a fixed outage probability P_{out}. As we shall see, the resulting capacities and optimal $\boldsymbol{Q}$ matrices in these two cases will be quite different.

The optimization problem of $\boldsymbol{Q}$ for the ergodic capacity has been analyzed in the past by Visotsky [4] in the context of the MISO system. In addition, we are also interested in outage capacities, which in many cases are more useful (see [9], [10]), 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 $\boldsymbol{Q}$ in the basis of $\boldsymbol{\Sigma}$ (i.e., $\boldsymbol{Q}$ and $\boldsymbol{\Sigma}$ should be simultaneously diagonalizable). The proof of this statement is given by Visotsky [4] and Jafar [11] for the ergodic capacity case and has been generalized by the current authors [14] to apply to the case of outage capacity. This simplification is crucial as it allows us to write $a_i = 1/(q_i s_i p)$ (with q_i and s_i the eigenvalues of $\boldsymbol{Q}$ and $\boldsymbol{\Sigma}$, respectively). Since we have derived analytic expressions for both the ergodic information and outage mutual information, in terms of the a_i's, maximization of the capacity reduces to an n-parameter maximization over the q_i's (subject to the trace constraint on $\boldsymbol{Q}$, which makes it an $n-1$-dimensional problem) which can now be done with minimal numerical effort[17].

B. Examples for Zero-Mean Channels

In this subsection, we analyze (11) and (13) and their optimal $\boldsymbol{Q}$'s in greater detail. We assume (without loss of generality) that the eigenvalues of $\boldsymbol{\Sigma}$ and $\boldsymbol{Q}$ are ordered such that $s_1 \geq s_2 \geq \cdots$ and $q_1 \geq q_2 \geq \cdots$, respectively. We define beamforming to be the transmission of all power through the maximum eigenvalue (s_1) and corresponding eigenvector of $\boldsymbol{\Sigma}$ such that $q_1 = n$ and all other q_i's are zero. A question that we will keep in mind in the following is, if $\boldsymbol{\Sigma}$ has more than one nonzero eigenvalue, when is beamforming optimal?

1) Two Transmitting Antennas: We begin by looking at the case of two transmitting antennas and a single receiver antenna. We consider a model $\boldsymbol{\Sigma}$ given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & x \\ x^* & 1 \end{pmatrix}. \tag{16}$$

The form of $\boldsymbol{\Sigma}$ in (16) implies that the two transmitting antennas are equivalent. We call the parameter x the "antenna cross correlation" (since x represents $\langle g_1^* g_2 \rangle$) and, without loss of generality, we take x to be real and positive. Clearly, when $x = 0$, the antennas are completely independent (uncorrelated), and when $x = 1$ the antennas are fully correlated. We can thus express x in terms of the eigenvalues of $\boldsymbol{\Sigma}$ as

$$x = \frac{s_1 - s_2}{2} \tag{17}$$

and since $\mathrm{Tr}\{\boldsymbol{\Sigma}\} = s_1 + s_2 (= 2)$ is fixed, varying x from 0 to 1 gives us all of the possible values of s_1 and s_2 for a two-antenna MISO system.

In Fig. 1, we plot the optimal q_2 for the two-antenna case as a function of the antenna cross correlation x (Note that $q_1 = 2 - q_2$ due to the power constraint). We note that for the simple case of two antennas, the optimization of the capacity can be done partially analytically by setting $q_1 = 2 - q_2$ and differentiating our analytic expressions for the ergodic capacity (11) and the outage (7) and (13).

The solid lines in Fig. 1 are plots optimizing the outage mutual information for various outages while the dotted lines optimize the ergodic mutual information for various values of the power level p (thus, giving the outage capacities and ergodic capacities, respectively). Recall that when q_2 goes to 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 we call "independent transmission"). In both the outage and ergodic cases, we see a transition from independent transmission to beamforming as the antenna cross correlation is increased. However, the quantitative details of this crossover differ substantially in the two cases.

Looking at the outage curves (solid) we see that more independent transmission (q_2 closer to 1) becomes favored as the outage probability is reduced. Similarly, for lower outage probabilities, 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. Analogously, for the ergodic case (dashed) we see that the transition to beamforming occurs at lower antenna cross correlation at lower power levels. This is simply the well-known fact that beamforming is favored at low power.

We also note that in the outage curves (solid) there is a discontinuity in the optimal q_2 as a function of antenna cross correlations where q_2 makes a jump from a finite value to zero (beamforming). Further we note that the discontinuity is substantial at high outage. This interesting property—a jump to beamforming—will be discussed in detail below in Section II-C.

In Fig. 2, we show relative outage mutual informations for optimized transmission, beamforming, and independent transmission as a func-

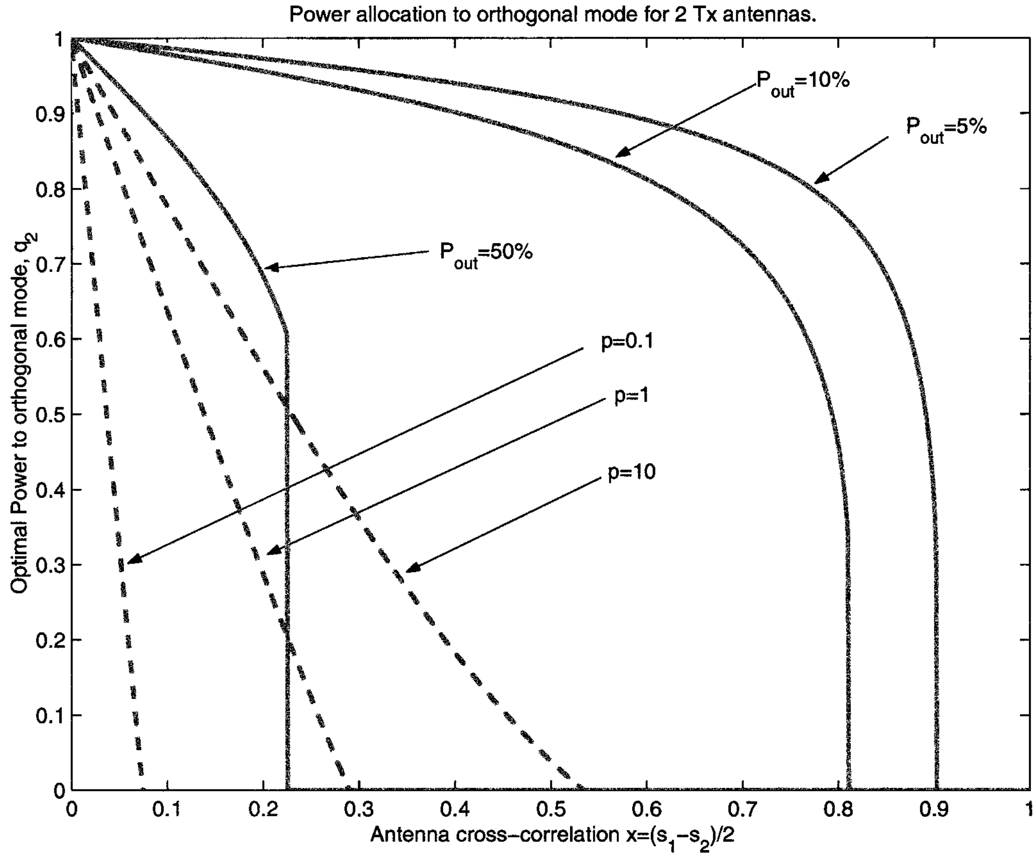

Fig. 1. Optimal power to nonbeamforming mode q_2 for the two transmit antenna case, as a function of the antenna cross correlation $x = (s_1 - s_2)/2$ for a Gaussian zero mean channel with nontrivial covariance (specified by x). Here $q_2 = 0$ corresponds to beamforming and $q_2 = 1$ corresponds to power being equally distributed between each of the two transmission eigenmodes. The dashed curves maximize the ergodic (mean) capacity for various values of the parameter p which is the signal-to-noise level per antenna here. The solid curves maximize the outage capacity for various different outage probabilities $P_{\rm out}$. Note that the solid curves jump discontinuously to zero, which can be seen most clearly for the $P_{\rm out} = 50\%$ case. (See Section II-C.)

tion 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 $I_{\rm out}$ (effectively 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 $I_{\rm out}$ (effectively 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. 2 we see that for high antenna cross correlation, beamforming becomes optimal, i.e., $z_{\rm optimal}/z_{\rm beam} = 1$ whereas at low antenna cross correlation, independent transmission becomes optimal, $z_{\rm optimal}/z_{\rm indep} \to 1$. We see that incorrectly choosing to use beamforming at low antenna cross correlations can lower the outage mutual information 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 $z_{\rm beam} = z_{\rm indep}$), one can stay very close to the optimal outage capacity. As can be seen from Fig. 2, only a small amount of additional outage capacity can be obtained by always using the optimal transmission covariance (optimal q_2).

2) Four Transmitting Antennas: In both the ergodic and outage case, the analytic form of the mutual information in (11) and (13) 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. 3 shows the outage and ergodic capacities for a 4×1 MISO system, as a function of angle spread Δ (in radians) at the transmitter. The transmitting antenna array is a uniform linear array with elements separated by $\lambda/2$. The correlation matrix $\mathbf{\Sigma}$ is determined by the formula

$$\Sigma_{ij} = \exp\left[-\frac{(\pi\,(i-j)\,\Delta)^2}{2}\right] \tag{18}$$

which is a simple approximation for a planar, Gaussian-distributed ray model with the main path of departure from the array being at broadside [15]. The capacities were calculated by numerically optimizing I in (11) and (13) with respect to q_1, q_2, q_3, and q_4 subject to the power constraint

$$\mathrm{Tr}\{\boldsymbol{Q}\} = q_1 + q_2 + q_3 + q_4 = n = 4.$$

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 the outage probability. More importantly, if, for example, 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 mutual information, 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

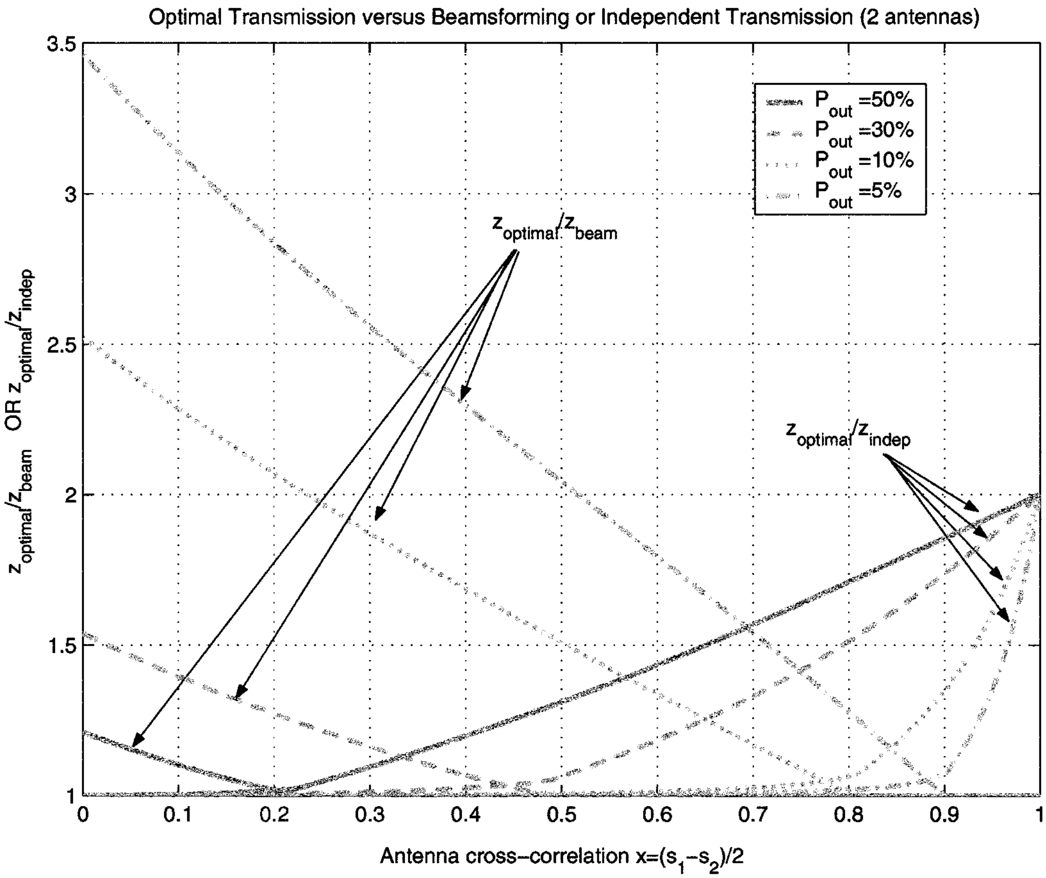

Fig. 2. Comparison of outage mutual information for optimized transmission, beamforming, and independent transmission as a function of antenna cross correlation $x = (s_1 - s_2)/2$ for different outage probabilities $P_{\rm out}$ for a Gaussian zero mean channel with nontrivial covariance (specified by x). 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 $I_{\rm out}$ (effectively low SNR), the ratio of z parameters is equal to the ratio of the outage capacities. In this case, we interpret the graph as showing $I_{\rm out}$ for optimal transmission divided by $I_{\rm out}$ for beamforming or independent transmission. At high $I_{\rm out}$ (effectively high SNR), the log of the z ratio becomes the difference in mutual informations. Thus, we see that for all antenna cross correlations and all outages, either beamforming is very close to 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 $x \approx 0.88$ only about 5% additional outage capacity (at low SNR) can be obtained by using optimized transmission correlations rather than beamforming or independent transmission.

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 transmission schemes such as beamforming ($q_1 = n$ and $q_i = 0$ for $i > 1$) or independent transmission (all $q_i = 1$) as well as a few intermediate transmission schemes. An example of a intermediate scheme is shown in Fig. 3. 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 scheme 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 mutual information (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 schemes will be best for a given antenna cross-correlation matrix.

C. Beamforming Optimality Criterion and the Transition to Beamforming

Several authors [4], [5], [14], [16] have previously posed the question of when beamforming is optimal. In these works, the focus has been on optimizing the ergodic mutual information. Again, we write the eigenvalues of $\boldsymbol{Q}$ as q_i and the eigenvalues of $\boldsymbol{\Sigma}$ as s_i. As noted above, we also work in a basis where both $\boldsymbol{Q}$ and $\boldsymbol{\Sigma}$ are diagonal. Without loss of generality, we use a basis such that $s_1 \geq s_2 \geq s_3 \cdots \geq s_n$. The beamforming mode is such that $q_1 = n$ and all other q_i's are zero.

We note that since s_2 is the second largest eigenvalue (the second best mode of transmission), in order to find the optimality condition for beamforming we need only find the point where q_2 first becomes nonzero. 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

$$\left.\frac{d\langle I \rangle}{dq_k}\right|_{q_1=n,\, q_j=0,\, j\geq 2,\, k\geq 2} \equiv \left[\frac{\partial \langle I \rangle}{\partial q_k} - \frac{\partial \langle I \rangle}{\partial q_1}\right]_{q_1=n,\, q_j=0,\, j\geq 2,\, k\geq 2} \leq 0. \quad (19)$$

This condition has previously been derived by other authors [4], [5]. It is, in fact, easy to show that since $s_2 > s_3 > \cdots$, 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

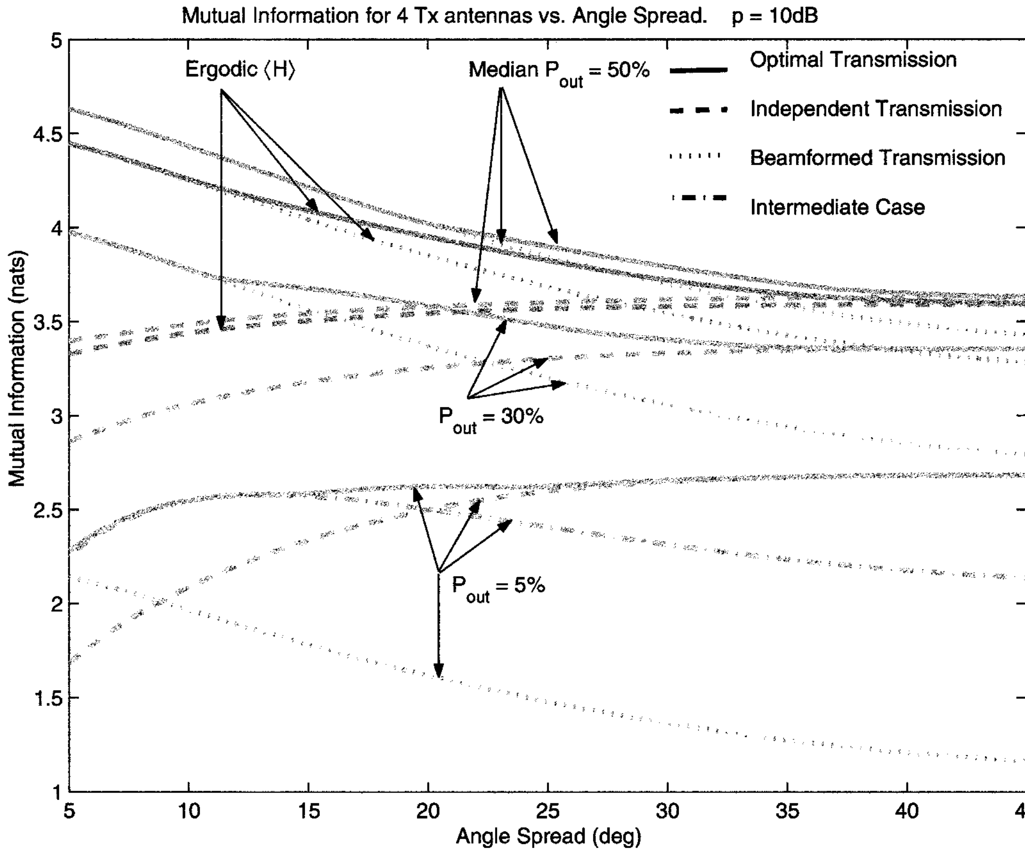

Fig. 3. Mutual information as a function of angle spread at the transmitter for a uniformly spaced four-antenna transmit array for a zero-mean Gaussian channel with covariance specified by the angle spread. (The antenna correlation matrix is given by (18).) The solid curves correspond to optimal transmission, 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 power level is $p = 10 =$ SNR/antenna. 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 probability density function (pdf) is roughly symmetric around its average. For low outage probabilities, one can obtain a substantial gain (roughly 20% near an angle spread of 10°) by using the optimal transmission rather than beamforming or independent transmission. However, the intermediate case of using only two modes is almost optimal in this range where both beamforming and independent transmission are poor.

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 (19) for $k = 2$ is, in fact, a necessary and sufficient condition for beamforming to optimize the ergodic mutual information $\langle I\rangle$.

Since we have an analytic expression (11) for $\langle I\rangle$, the relevant derivatives can be taken analytically to yield an analytic condition for beamforming optimality for the ergodic capacity in agreement with the prior result in [5] and the more general results by the current authors in [14], [16].

The question of when beamforming is optimal for maximizing the outage mutual information is more difficult to answer. Here, since the ccdf in (13) depends on $I_{\rm out}$ only through z, as discussed above, the results are independent of the signal-to-noise parameter p. However, determining the optimal $\boldsymbol{Q}$ is now more complicated because $I_{\rm out}$ is generally *not* a convex function of q_2.

We would like to maximize z in (13) by varying over the q_i's subject to the constraint $\sum_i q_i = n$. (Again, we take $\boldsymbol{Q}$ and $\boldsymbol{\Sigma}$ to be simultaneously diagonal). 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, using our analytic form for the ccdf (13) it is straightforward to show, at least for the case of two transmitting antennas, 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. 4, 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.

III. Nonzero-Mean Channel With Covariance $\boldsymbol{\Sigma} = \boldsymbol{I}_n$

In this section, we consider the case where the channel $\boldsymbol{g}$ has a nonzero mean $\boldsymbol{g}_0$ and covariance equal to the identity. In other words, we take the ensemble of $\boldsymbol{g}$ to be given by $\mathcal{CN}(\boldsymbol{g}_0, \boldsymbol{I}_n)$. More generally, one might consider a case where the channel has a nonzero mean and a nontrivial covariance (i.e., an $\mathcal{CN}(\boldsymbol{g}_0, \boldsymbol{\Sigma})$ channel). This more general case is discussed in Appendix I. In the present section, however, we restrict our attention to $\boldsymbol{\Sigma} = \boldsymbol{I}_n$.

It has been shown in [4] that the optimal $\boldsymbol{Q}$ that maximizes the ergodic mutual information for fixed $\boldsymbol{g}_0$ is given by

$$Q_{ij} = q_0 \frac{g_{0i}g_{0j}^*}{\boldsymbol{g}_0^\dagger \boldsymbol{g}_0} + q\left(\delta_{ij} - \frac{g_{0i}g_{0j}^*}{\boldsymbol{g}_0^\dagger \boldsymbol{g}_0}\right) \tag{20}$$

where

$$q_0 = n - (n-1)q \tag{21}$$

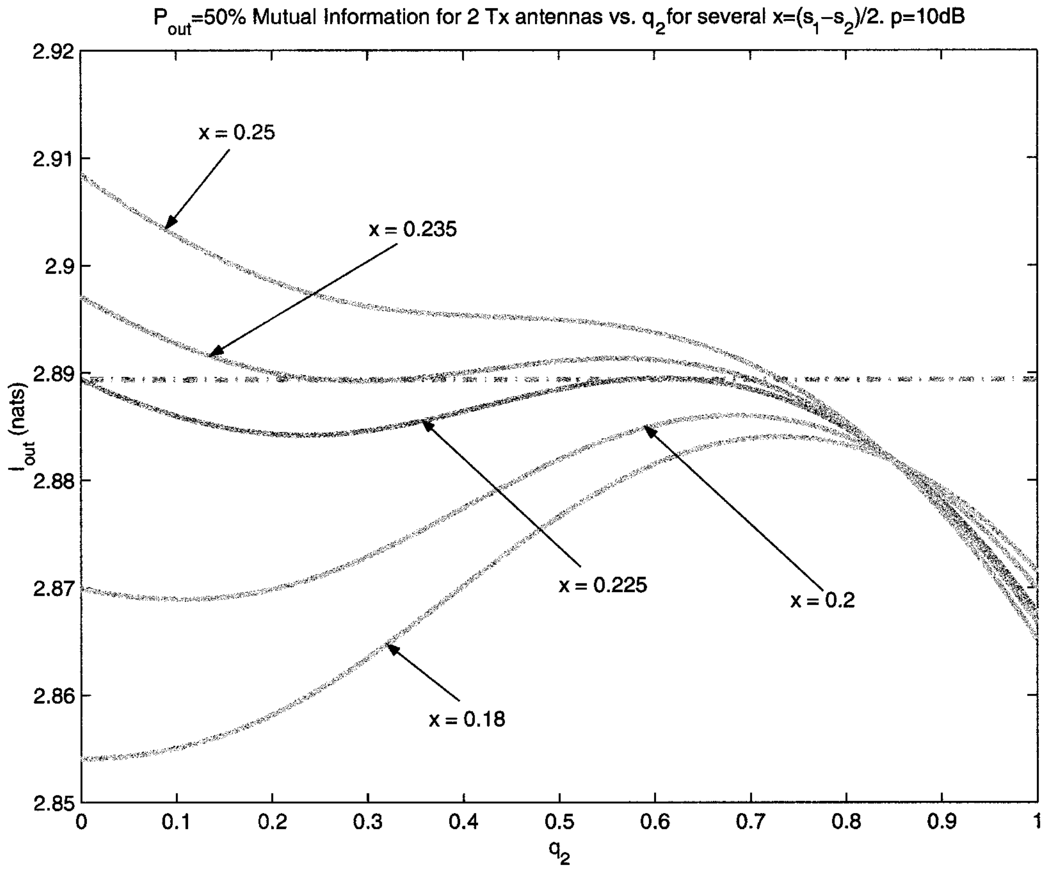

Fig. 4. 50% outage mutual information versus power q_2 to the orthogonal mode for a zero-mean Gaussian channel with different antenna cross correlations $x = (s_1 - s_2)/2$ with $p =$ SNR/antenna $= 10$. We see that the mutual information is a nonmonotonic function of q_2 and the optimal value of q_2 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.

and $q_0 \geq q$. In Appendix II, we show that this form is also optimal for maximizing any outage capacity. The first term of (20) is the projection of $\boldsymbol{Q}$ in the direction of $\boldsymbol{g}_0$, while the second is the projection onto the basis perpendicular to $\boldsymbol{g}_0$.

With the $\boldsymbol{Q}$ of (20), and using the expression (64) for the ergodic average of the mutual information derived in Appendix I, with $\boldsymbol{\Sigma} = \boldsymbol{I}_n$ (as explained in Appendix I-C1) we obtain

$$\langle I \rangle = \int_0^\infty \frac{dy \exp\left(-\frac{y}{p}\right)}{y} \left[1 - \frac{\exp\left(-\frac{\gamma q_0 y}{1+q_0 y}\right)}{(1+q_0 y)(1+qy)^{n-1}}\right] \tag{22}$$

with p being proportional to the signal-to-noise per antenna for the fluctuating part of the signal and $\gamma = \boldsymbol{g}_0^\dagger \boldsymbol{g}_0$ is the mean channel signal strength (see Section I-B).

Similarly, for the above form of $\boldsymbol{Q}$, (46) derived in Appendix I-B gives the ccdf of the mutual information

$$\mathrm{ccdf}(I) = 1 - e^{-\gamma} \frac{z}{q_0} \int_0^1 dt e^{-\frac{zt}{q_0}} I_0\left(2\sqrt{\frac{\gamma z t}{q_0}}\right) \cdot \left[1 - \frac{\Gamma\left(n, \frac{z(1-t)}{q}\right)}{(n-1)!}\right] \tag{23}$$

where

$$\Gamma(n, x) = \int_x^\infty x^{n-1} e^{-x}$$

is the incomplete Γ function, and as above $z = \frac{e^I - 1}{p}$.

Equations (22) and (23) constitute the main results of this section. Maximizing $\langle I \rangle$ in (22) to produce the optimal q as a function of γ and p will give the ergodic capacity. Similarly, using (23) (and (7)) to maximize the outage mutual information over q will produce the outage capacity. In either case, the optimization procedure involves maximizing an integral over one parameter.

In Fig. 5, we show the optimal power to the orthogonal (nonbeamforming) direction (q) as a function of the constant part of the channel $\gamma = \boldsymbol{g}_0^\dagger \boldsymbol{g}_0$. As might be expected, beamforming $(q = 0)$ becomes optimal for large γ since in that case the transmitter can take better advantage of $\boldsymbol{g}_0$. Conversely, when $\gamma = 0$, independent transmission $(q = 1)$ becomes optimal. As might be expected, beamforming is also favored at lower power levels and higher outages.

The solid lines are again optimization of outage mutual information for various values of the outage probability P_{out} and the dashed curves are optimization of the ergodic mutual information at various levels of the power parameter p, yielding the outage and ergodic capacities, respectively. It is clear once again that maximization of ergodic mutual information is very different from maximization of any outage mutual information. Finally, we note that, analogous to the cases examined above, for the outage capacity, the transition to beamforming can occur via a jump—as is clear for the $P_{\mathrm{out}} = 50\%$ curve. Later, we will discuss this jump further.

A. Beamforming Criterion and Transition to Beamforming

Next we analyze the optimization over $\boldsymbol{Q}$. To be clear, we note that we are optimizing a capacity with fixed $\boldsymbol{g}_0$. (In a real system, $\boldsymbol{g}_0$ might change from time to time, and to evaluate the long-term capacity of

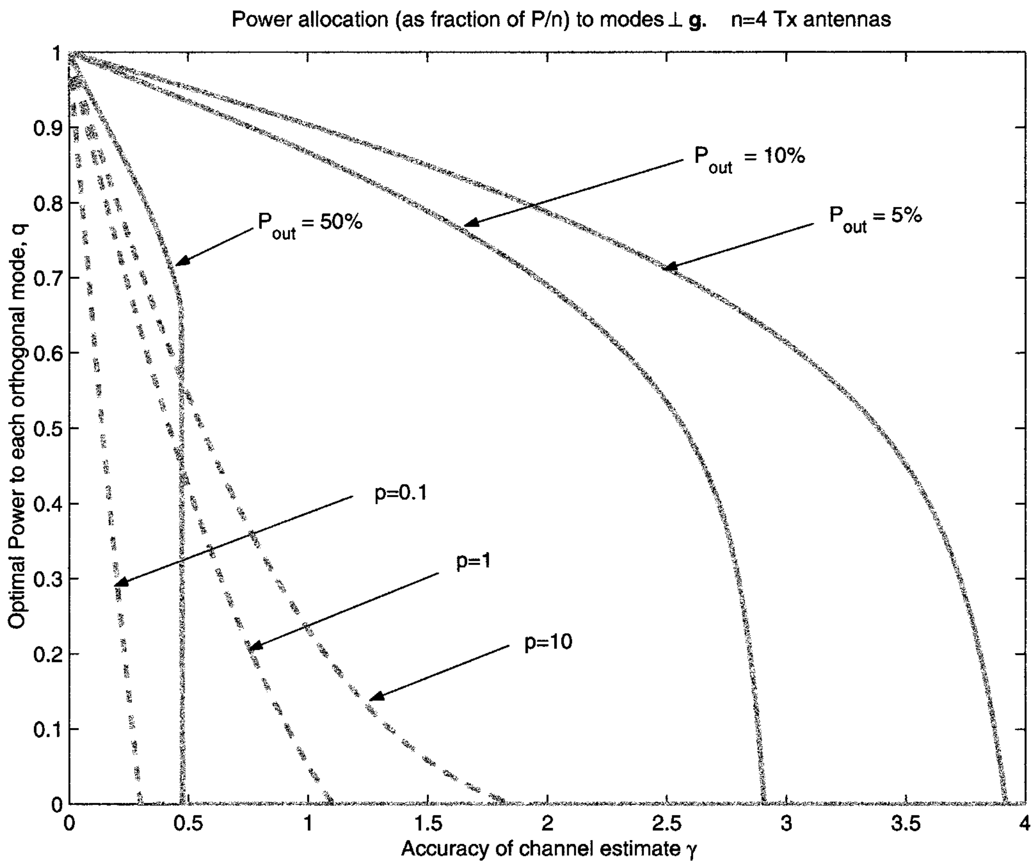

Fig. 5. Optimal power allocation for a nonzero-mean Gaussian channel and white covariance $\mathbf{\Sigma} = \boldsymbol{I}_n$. Here, q represents power to the "orthogonal modes," such that $q = 0$ is beamforming and $q = 1$ is independent transmission from all antennas. We have plotted the optimal q as a function of the channel magnitude $\gamma = \boldsymbol{g}_0^\dagger \boldsymbol{g}_0$, where $\boldsymbol{g}_0$ is the mean channel. Solid lines are the optimal q for maximizing outage mutual information (for various outage probabilities), whereas dashed lines are the optimal q for maximizing ergodic mutual information (at various power levels p). As one might expect, as the mean channel γ gets larger, beamforming becomes increasingly optimal. The crossover to beamforming occurs more quickly for smaller p and higher outage probability. Note that for $P_{\rm out} = 50\%$, q drops to zero discontinuously, signifying a nonmonotonic outage mutual information as a function of q. For smaller $P_{\rm out} < 0.226$ and larger γ ($\gamma > 1.670$), the discontinuity disappears.

an actual system one would typically want to average over all of the instantiations of $\boldsymbol{g}_0$ that might occur in an ensemble of $\boldsymbol{g}_0$'s).

Since the ergodic mutual information is convex [5], analogous to (19) above, the beamforming condition is that at $q = 0$

$$\left.\frac{d\langle I\rangle}{dq}\right|_{q=0} = \left[\frac{\partial\langle I\rangle}{\partial q} - \frac{1}{n-1}\frac{\partial\langle I\rangle}{\partial q_0}\right]_{q=0,\, q_0=n} \le 0. \tag{24}$$

Taking the derivative in (22) yields

$$\int_0^\infty dt\, \frac{\exp\left(-\frac{t}{np}\right)\exp\left(-\frac{\gamma t}{1+t}\right)}{(1+t)^3}\,(t\,(t+1)-\gamma) \le 0. \tag{25}$$

This condition, for optimizing the ergodic capacity, agrees with that derived more generally in [16].

Once again, the outage capacity will prove more difficult to analyze. As previously discussed in Section II-C, the outage capacity as a function of q is not necessarily monotonic. We demonstrate this for the case of a nonzero-mean channel by analyzing the behavior of $z(q,\, \gamma,\, P_{\rm out})$ close to $q = 0$. The first derivative of z with respect to q at $q = 0$ is

$$\left.\frac{dz}{dq}\right|_{q=0} = (n-1)\left(1 - \frac{z}{n}\right) \tag{26}$$

where z is the solution of (23) at $q = 0$ which simplifies to

$$P_{\rm out} = e^{-\gamma}\,\frac{z}{n}\int_0^1 dt e^{-\frac{zt}{n}} I_0\left(2\sqrt{\frac{\gamma z t}{n}}\right). \tag{27}$$

Thus, setting $z = n$ along with solving (27) at $z(q = 0,\, \gamma,\, P_{\rm out}) = n$ establishes the existence of an extremum (equivalent to (19)). However, here, we need to check convexity. The second derivative at $q = 0$ is

$$\begin{aligned}\left.\frac{d^2 z}{dq^2}\right|_{q=0} = &\frac{n-1}{zI_0\left(2\sqrt{\frac{\gamma z}{n}}\right)} \\ &\cdot\left[I_0\left(2\sqrt{\gamma z/n}\right)\frac{z}{n} - \sqrt{\frac{\gamma z}{n}}\,I_1\left(2\sqrt{\frac{\gamma z}{n}}\right)\right] \\ &\cdot\left[1 + 2\,(n-1)\left(1-\frac{z}{n}\right)^2\right]\end{aligned} \tag{28}$$

where $I_n(x)$ is the nth-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 true for $I_0(2\sqrt{\gamma}) < \sqrt{\gamma} I_1(2\sqrt{\gamma})$, i.e., for $\gamma > 1.670$. Which corresponds to

$$P_{\rm out} < P_{\rm crit} = 0.226 \tag{29}$$

in the sense that $z(0,\, \gamma > 1.67,\, P_{\rm out}) = n$ in (27) has solution $P_{\rm out} < 0.226$. By numerically analyzing (23) we find that for $P_{\rm out} < 0.226$ the optimal $q(\gamma)$ decreases continuously to zero, and indeed in this region, (27) evaluated at $z = n$ gives the proper beamforming optimality condition. The continuous transition to beamforming is seen in Fig. 5 for 10% and 5% outage probabilities.

In contrast, for $P_{\rm out} > P_{\rm crit}$, the optimal q jumps discontinuously to zero. This can be seen from the fact that for $\gamma < 1.670$ and $z(q = 0,\, \gamma,\, P_{\rm out}) = n$, the second derivative of z is positive, i.e., $d^2z/dq^2 >$

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. 5 for $P_{\text{out}} = 50\%$.

IV. Summary

In this correspondence, we have shown how to analytically (or mostly analytically) calculate the outage mutual information and the ergodic average mutual information of MISO systems as a function of the transmission covariance $\boldsymbol{Q}$, thus enabling us to determine which $\boldsymbol{Q}$ maximizes the mutual information. We have considered cases where the channel is described as being $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ or being $\mathcal{CN}(\boldsymbol{g}_0, \boldsymbol{I}_n)$, i.e., having mean zero and covariance $\boldsymbol{\Sigma}$ or having mean $\boldsymbol{g}_0$ and white covariance. (The more general case of $\mathcal{CN}(\boldsymbol{g}_0, \boldsymbol{\Sigma})$ is discussed in Appendix I.) These cases represent realistic situations where the transmitter has only statistical channel information. While the channel itself may change rapidly, 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, one may exploit a structure introduced by Alamouti [18], 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 [10], and has been simulated for independent and identically distributed (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 [19]. In general, however, the question of coding for $n \times 1$ MISO systems remains open.

Appendix I
Calculational Details for General Case: $\mathcal{N}(\boldsymbol{g}_,\, \boldsymbol{\Sigma})$

Here we look at the general case where the channel has nonzero mean $\boldsymbol{g}_0$ and arbitrary covariance $\boldsymbol{\Sigma}$ (i.e., the channel is $\mathcal{CN}(\boldsymbol{g}_0, \boldsymbol{\Sigma})$). We will aim to first calculate the pdf which is defined as

$$\text{pdf}(I) = \frac{d\,\text{cdf}\,(I)}{dI} = \langle \delta[I - \log(1 + \boldsymbol{g}^\dagger p \boldsymbol{Q} \boldsymbol{g})] \rangle \tag{30}$$

where $\delta(\cdot)$ is the Dirac delta function. In terms of this quantity, we have

$$\text{ccdf}(I) = \int_I^\infty dI'\, \text{pdf}(I') \tag{31}$$

$$\langle I \rangle = \int_0^\infty dI\, I\, \text{pdf}(I). \tag{32}$$

It is convenient to separate out the constant piece of $\boldsymbol{g}$ and write $\boldsymbol{g} = \boldsymbol{g}_0 + \boldsymbol{z}$ such that $I = \log(1 + (\boldsymbol{g}_0^\dagger + \boldsymbol{z}) p \boldsymbol{Q} (\boldsymbol{g}_0 + \boldsymbol{z}))$ where $\boldsymbol{g}_0$ is a constant and $\boldsymbol{z}$ is a Gaussian random vector with zero mean and covariance $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{z} = \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

The average of any quantity O over the ensemble of $\boldsymbol{z}$ is given by [8]

$$\langle O \rangle = \int d\mu(\boldsymbol{z}, \boldsymbol{z}^\dagger) e^{-\boldsymbol{z}^\dagger \boldsymbol{\Sigma}^{-1} \boldsymbol{z}} O \tag{33}$$

where the integration measure is defined by

$$\int d\mu(\boldsymbol{z}, \boldsymbol{z}^\dagger) = \det \boldsymbol{\Sigma}^{-1} \prod_{i=1}^n \frac{1}{\pi} \int d\,\text{Re}\, z_i \int d\,\text{Im}\, z_i. \tag{34}$$

Thus, we can write

$$\begin{aligned}\text{pdf}(I) &= \int d\mu(\boldsymbol{z}, \boldsymbol{z}^\dagger) e^{-\boldsymbol{z}^\dagger \boldsymbol{\Sigma}^{-1} \boldsymbol{z}} \\ &\quad \cdot \delta[I - \log(1 + [\boldsymbol{g}_0^\dagger + \boldsymbol{z}^\dagger] p \boldsymbol{Q} [\boldsymbol{g}_0 + \boldsymbol{z}])] & (35)\\ &= \int d\mu(\boldsymbol{z}, \boldsymbol{z}^\dagger) e^{-\boldsymbol{z}^\dagger \boldsymbol{\Sigma}^{-1} \boldsymbol{z}} e^I \\ &\quad \cdot \delta(e^I - 1 - [\boldsymbol{g}_0^\dagger + \boldsymbol{z}^\dagger] p \boldsymbol{Q} [\boldsymbol{g}_0 + \boldsymbol{z}]). & (36)\end{aligned}$$

The delta function is now written as a Fourier transform (using $\delta(x) = \frac{1}{2\pi} \int dk e^{ikx}$) to yield

$$\begin{aligned}\text{pdf}(I) &= \int d\mu(\boldsymbol{z}, \boldsymbol{z}^\dagger) e^{-\boldsymbol{z}^\dagger \boldsymbol{\Sigma}^{-1} \boldsymbol{z}} \\ &\quad \cdot e^I \int \frac{dk}{2\pi} e^{ik(e^I - 1 - [\boldsymbol{g}_0^\dagger + \boldsymbol{z}^\dagger] p \boldsymbol{Q} [\boldsymbol{g}_0 + \boldsymbol{z}])} & (37)\\ &= e^I \int \frac{dk}{2\pi} e^{ik(e^I - 1) - ik \boldsymbol{g}_0^\dagger p \boldsymbol{Q} \boldsymbol{g}_0} \int d\mu(\boldsymbol{z}, \boldsymbol{z}^\dagger) \\ &\quad \cdot e^{-\boldsymbol{z}^\dagger [\boldsymbol{\Sigma}^{-1} + ikp\boldsymbol{Q}] \boldsymbol{z} - \boldsymbol{z}^\dagger ikp \boldsymbol{Q} \boldsymbol{g}_0 - \boldsymbol{g}_0^\dagger ikp \boldsymbol{Q} \boldsymbol{z}}. & (38)\end{aligned}$$

We next aim to do the integral over $\boldsymbol{z}$. These types of Gaussian integrals can be performed by using the general rule [8]

$$\prod_{i=1}^n \left[\int d\text{Re} x_i \int d\text{Im} x_i \right] e^{-\boldsymbol{x}^\dagger \boldsymbol{M} \boldsymbol{x} + \boldsymbol{n}^\dagger \boldsymbol{x} + \boldsymbol{x}^\dagger \boldsymbol{n}'} = \frac{\pi^n e^{-\boldsymbol{n}^\dagger \boldsymbol{M}^{-1} \boldsymbol{n}'}}{\det \boldsymbol{M}}. \tag{39}$$

Thus, we have

$$\text{pdf}(I) = e^I \int \frac{dk}{2\pi} \frac{e^{ik(e^I - 1) - ik \boldsymbol{g}_0^\dagger p \boldsymbol{Q} \boldsymbol{g}_0 - \boldsymbol{g}_0^\dagger kp \boldsymbol{Q} [\boldsymbol{\Sigma}^{-1} + ikp\boldsymbol{Q}]^{-1} kp \boldsymbol{Q} \boldsymbol{g}_0}}{\det[\boldsymbol{I}_n + ikp\boldsymbol{Q}\boldsymbol{\Sigma}]}. \tag{40}$$

The terms containing $\boldsymbol{g}_0$ in the exponent can be rewritten as

$$-\boldsymbol{g}_0^\dagger \boldsymbol{\Sigma}^{-1} \boldsymbol{g}_0 + \boldsymbol{g}_0^\dagger \boldsymbol{Q}^{1/2} \boldsymbol{A} [ik\boldsymbol{I} + \boldsymbol{A}]^{-1} \boldsymbol{A} \boldsymbol{Q}^{1/2} \boldsymbol{g}_0 \tag{41}$$

where $\boldsymbol{A} = (\boldsymbol{Q}^{1/2} \boldsymbol{\Sigma} \boldsymbol{Q}^{1/2} p)^{-1}$. Making an eigenvector expansion

$$A_{jk} = \sum_{i=1}^n a_i v_j^{i*} v_k^i \tag{42}$$

where a_i and $\boldsymbol{v}^i$ are the eigenvalues and the respective normalized eigenvectors of $\boldsymbol{A}$ (note that a_i are also the eigenvalues of $(p\boldsymbol{Q}\boldsymbol{\Sigma})^{-1}$ as defined above), the second term in (41) can be written as

$$\begin{aligned}&\boldsymbol{g}_0^\dagger \boldsymbol{Q}^{1/2} \boldsymbol{A} [ik\boldsymbol{I} + \boldsymbol{A}]^{-1} \boldsymbol{A} \boldsymbol{Q}^{1/2} \boldsymbol{g}_0 \\ &\qquad = \sum_{i=1}^n (\boldsymbol{g}_0^\dagger \boldsymbol{Q}^{1/2} \boldsymbol{v}^i a_i)(ik + a_i)^{-1} (a_i \boldsymbol{v}_i^* \boldsymbol{Q}^{1/2} \boldsymbol{g}_0) \qquad (43)\end{aligned}$$

where we have used the fact that eigenvectors form a complete set, i.e., $\sum_i v_j^i v_k^{i*} = \delta_{jk}$. Thus, we can rewrite

$$\text{pdf}(I) = e^{I - \boldsymbol{g}_0^\dagger \boldsymbol{\Sigma}^{-1} \boldsymbol{g}_0} \int \frac{dk}{2\pi} \frac{e^{ik(e^I - 1)}}{\prod_j (1 + ik/a_j)} \exp\left[\sum_i \frac{b_i}{1 + ik/a_i} \right] \tag{44}$$

where

$$b_i = p a_i |\boldsymbol{g}_0^\dagger \boldsymbol{Q}^{1/2} \boldsymbol{v}^i|^2. \tag{45}$$

Note that using (42) we can show $\sum_i b_i = \boldsymbol{g}_0^\dagger \boldsymbol{\Sigma}^{-1} \boldsymbol{g}_0$, which we will need below.

To obtain the cdf of I, we integrate over I (as shown in (31)) to yield

$$\mathrm{ccdf}(I) = \int \frac{dk}{2\pi} \frac{e^{ik(e^I-1)}}{0^+ - ik} \frac{e^{-\boldsymbol{g}_0^\dagger \boldsymbol{\Sigma} \boldsymbol{g}_0}}{\prod_j (1+ik/a_j)} \exp\left[\sum_i \frac{b_i}{1+ik/a_i}\right]. \tag{46}$$

Integrals of the type shown in (46) and (44) can be manipulated conveniently by using transformations like the following:

$$\begin{aligned}\text{Integral} &= \int \frac{dk}{2\pi} \frac{e^{ikx + B/(a+ik)}}{a+ik} \\ &= \frac{1}{\pi}\int dy\, dy^* \int \frac{dk}{2\pi} e^{ikx - |y|^2(a+ik) + \sqrt{B}(y+y^*)} \end{aligned} \tag{47}$$

where $\int dy\, dy^*$ means that y is integrated over the complex plane. We then define a radial variable $z = |y|^2$ and angular variable θ and rewrite the integral over the plane as

$$\text{Integral} = \frac{1}{2\pi}\int d\theta \int_0^\infty dz \int \frac{dk}{2\pi} e^{ik(x-z) - az + 2\sqrt{Bz}\cos(\theta)} \tag{48}$$

$$= \frac{1}{2\pi}\int_0^\infty dz \int d\theta \delta(x-z) e^{-az + 2\sqrt{Bz}\cos(\theta)} \tag{49}$$

$$= e^{-ax} I_0(2\sqrt{Bx}) \tag{50}$$

with I_0 the modified Bessel function. To handle more complicated integrals such as

$$\int \frac{dk}{2\pi} \frac{e^{ikx + B_1/(a_1+ik) + B_2/(a_2+ik)}}{(a_1+ik)(a_2+ik)} \tag{51}$$

one would introduces two complex variables y_1 and y_2. The integral over k then gives $\delta(x - z_1 - z_2)$ and the angular integrals can still be performed. We can apply this approach to (46) for arbitrary $\boldsymbol{Q}$ and $\boldsymbol{\Sigma}$ by using $n+1$ different y's to take care of all $n+1$ factors in the denominator, resulting in the useful form

$$\begin{aligned}\mathrm{cdf}(I) = e^{-\boldsymbol{g}_0^\dagger \boldsymbol{\Sigma} \boldsymbol{g}_0} &\left[\prod_{j=0}^n \int_0^\infty dz_j\right] \\ &\cdot \delta\left(e^I - 1 - \sum_{j=1}^n z_j\right) \prod_{j=1}^n a_j e^{-z_j a_j} I_0(2\sqrt{z_j a_j b_j}). \end{aligned} \tag{52}$$

A. cdf and $\langle I \rangle$ for Zero-Mean Channel $\mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma})$

In the case where $\boldsymbol{g}_0 = \boldsymbol{0}$ expressions (44) and (46) simplify to

$$\mathrm{pdf}(I) = e^I \prod_{i=1}^n (-ia_i) \int \frac{dk}{2\pi} \frac{e^{ik(e^I-1)}}{\prod_{i=1}^n (k - ia_i)} \tag{53}$$

$$\mathrm{ccdf}(I) = \prod_{i=1}^n (-ia_i) \int \frac{dk}{2\pi} \frac{e^{ik(e^I-1)}}{i(k-i0^-)\prod_{i=1}^n (k - ia_i)}. \tag{54}$$

Since the a_i are all nonnegative and real, we can close the integration contour in the upper half plane [20]. (In the case where one or more a_i are precisely zero, a limit should be taken where they approach zero from the positive side.) In addition, we see from (53) and (54) that in the limit of a particular a_i going to infinity $a_i \to \infty$, that term cancels, and the product is over the remaining $n-1$ degrees of freedom. In the case where all of the a_i's are different, this immediately yields

$$\mathrm{pdf}(I) = \left(\prod_{i=1}^n a_i\right) \sum_{j=1}^n \frac{f(a_j, I)}{a_j \prod_{m \neq j} (a_m - a_j)} \tag{55}$$

where

$$f(a, I) = a \exp\left(I - a(e^I - 1)\right) \tag{56}$$

as well as (13). If the a_i's are not all different, a limit can be taken of the same expression as the appropriate a_i's approach each other. More simply, perhaps, one can perform the above contour integral with appropriate multiple poles [20]. To obtain the ergodic capacity, one directly integrates the pdf in (32) using

$$a\int_0^\infty dI\, I\, e^{I - a(e^I-1)} = e^a \mathrm{Ei}(-a) \tag{57}$$

and we immediately obtain (11).

B. cdf for Nonzero-Mean Channel: $\mathcal{CN}(\boldsymbol{g}_0, \boldsymbol{I})$

For the case considered in Section III, we use $\boldsymbol{\Sigma} = \boldsymbol{I}$ and $\boldsymbol{Q}$ of the form (20). From this form we see that q_0 is an eigenvalue of $\boldsymbol{Q}$ corresponding to the eigenvector $\boldsymbol{g}_0/|\boldsymbol{g}_0|$, and q is an eigenvalue of $\boldsymbol{Q}$ with a multiplicity of $n-1$. Thus, we write $a_i = 1/(qp)$ for $i = 1, \ldots, n-1$ and $a_n = 1/(q_0 p)$. Similarly, we have $b_i = 0$ for $i = 2, \ldots, n-1$ and $b_n = 1$. Using (52), noting that $I_0(0) = 1$, we obtain (see also (68) and related comments)

$$\begin{aligned}\mathrm{cdf}(I) = a_n e^{-\boldsymbol{g}_0^\dagger \boldsymbol{g}_0} &\int_0^{e^I - 1} dz_n \\ &\cdot e^{-z_n/(q_0 p)} I_0\left(2\sqrt{\frac{z_n}{q_0 p}}\right) \tilde{C}(e^I - 1 - z_n) \end{aligned} \tag{58}$$

where

$$\tilde{C}(x) = \left[\prod_{j=0}^{n-1} \int_0^\infty dz_j\right] (qp)^{-n+1} \delta\left(x - \sum_{j=0}^{n-1} z_j\right) e^{-\frac{1}{qp}\sum_{j=1}^{n-1} z_j} \tag{59}$$

$$= (qp)^{-n+1} \int_0^\infty dz_0 \int_0^\infty \frac{dw\, w^{n-1}}{(n-1)!} \delta(x - z_0 - w) e^{-w/(qp)} \tag{60}$$

where we have made the substitution $w = \sum_{i=1}^{n-1} z_i$. Thus, we obtain

$$\begin{aligned}\tilde{C}(x) &= (qp)^{-n+1} \int_0^x \frac{dw\, w^{n-1}}{(n-1)!} e^{-w/(qp)} \\ &= qp\left[1 - \frac{1}{(n-1)!}\Gamma\left(n, \frac{x}{qp}\right)\right]. \end{aligned} \tag{61}$$

Substitution into (58) yields (23).

C. Ergodic Average $\langle I \rangle$ for General Case: $\mathcal{CN}(\boldsymbol{g}_0, \boldsymbol{\Sigma})$

A useful general expression for $\langle I \rangle$ can be derived using the identity

$$\log(x) = \int_0^\infty \frac{dy}{y}\left[e^{-y} - e^{-xy}\right] \tag{62}$$

to express the log in the definition of $I = \log(1 + \boldsymbol{g}^\dagger p \boldsymbol{Q} \boldsymbol{g})$. Applying (33) for the case $G = \mathcal{CN}(\boldsymbol{g}_0, \boldsymbol{\Sigma})$ we get

$$\begin{aligned}\langle I \rangle = \int d\mu(\boldsymbol{z}, \boldsymbol{z}^\dagger) e^{-\boldsymbol{z}^\dagger \boldsymbol{\Sigma}^{-1}\boldsymbol{z}} &\int_0^\infty \\ &\cdot \frac{dy\, e^{-y}}{y}\left[1 - \exp\left(-y[\boldsymbol{g}_0^\dagger + \boldsymbol{z}^\dagger] p \boldsymbol{Q} [\boldsymbol{g}_0 + \boldsymbol{z}]\right)\right]. \end{aligned} \tag{63}$$

Integrating out the $\boldsymbol{z}$ variables (see (39)) we get

$$\begin{aligned}\langle I \rangle = \int_0^\infty &\frac{dy e^{-y}}{y} \\ &\cdot \left[1 - \frac{\exp\left(-\boldsymbol{g}_0^\dagger \boldsymbol{\Sigma}^{-1/2}\left[\frac{\boldsymbol{\Sigma}^{1/2}\boldsymbol{Q}p\boldsymbol{\Sigma}^{1/2}y}{1+\boldsymbol{\Sigma}^{1/2}\boldsymbol{Q}p\boldsymbol{\Sigma}^{1/2}y}\right]\boldsymbol{\Sigma}^{-1/2}\boldsymbol{g}_0\right)}{\det[1 + \boldsymbol{\Sigma}^{1/2}\boldsymbol{Q}p\boldsymbol{\Sigma}^{1/2}y]}\right]. \end{aligned} \tag{64}$$

1) Ergodic Average $\langle I \rangle$ for White Covariance: $\mathcal{CN}(\boldsymbol{g}_0, \boldsymbol{I})$: For the case considered in Section III, we use $\boldsymbol{\Sigma} = \boldsymbol{I}$ and $\boldsymbol{Q}$ of the form (20). Again, this gives us q_0 as an eigenvalue of $\boldsymbol{Q}$, and q as an eigenvalue of $\boldsymbol{Q}$ with a multiplicity of $n-1$. Thus, the determinant in (64) is

$(1+q_0py)(1+qpy)^{n-1}$. Furthermore, only the eigenvector corresponding to q_0 contributes in the exponent, giving $-\gamma q_0 py/(1+q_0py)$. Plugging these in to (64) and rescaling $y \to y/p$ gives (22).

APPENDIX II
OPTIMAL FORM OF $\boldsymbol{Q}$ WHEN $\boldsymbol{G} = \mathcal{CN}(\boldsymbol{g}_0, \boldsymbol{I}_n)$

In the case where $\boldsymbol{\Sigma} = \boldsymbol{I}_n$ it has been shown in [4] that to optimize the ergodic capacity, one must choose a $\boldsymbol{Q}$ of the form of (20). (Indeed, this can be seen relatively simply by analyzing the form of (64)). In this appendix, we prove that to optimize a given outage capacity, one must also choose $\boldsymbol{Q}$ of this form.

First we note that since cdf(I) is a monotonically increasing function, and since the outage is the inverse function of cdf (see (6) and (7)) finding the $\boldsymbol{Q}$ that maximizes $\mathcal{OUT}(P_{\rm out})$ is equivalent to finding the $\boldsymbol{Q}$ that minimizes cdf(I). Fixing the eigenvalues a_i of $(p\boldsymbol{Q})^{-1}$, we examine the cdf in the form of (52). As mentioned above $\sum_{j=1}^n b_j = \boldsymbol{g}_0^\dagger \boldsymbol{g}_0$ is fixed, however, the individual b_js can be varied by making a unitary rotation of $\boldsymbol{Q}$. 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 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 (52). Focusing on the z_0, z_1, and z_2 integrals (which we choose to do first), we write these integrals

$$\begin{aligned}\text{Int} &= \int_0^\infty dz_0\, dz_1\, dz_2 \delta(C - z_0 - z_1 - z_2) \\ &\quad \cdot e^{-a_1 z_1 - a_2 z_2} I_0(2\sqrt{z_1 a_1 b_1}) I_0(2\sqrt{z_2 a_2 b_2}) \\ &= \int_0^C dz_2 \int_0^{C - z_2} dz_1 e^{-a_1 z_1 - a_2 z_2} \\ &\quad \cdot I_0(2\sqrt{z_1 a_1 b_1}) I_0(2\sqrt{z_2 a_2 b_2}) \end{aligned} \tag{65}$$

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

$$\text{Int} > \int_0^C dz_2 \int_0^{C-z_2} dz_1 e^{-a_1 z_1 - a_2 z_2} I_0(2\sqrt{z_1 a_1 b_1}) I_0(2\sqrt{z_1 a_1 b_2}) \tag{66}$$

$$> \int_0^C dz_2 \int_0^{C-z_2} dz_1 e^{-a_1 z_1 - a_2 z_2} I_0(2\sqrt{z_1 a_1 (b_1 + b_2)}). \tag{67}$$

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 (67) 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

$$\begin{aligned}\text{cdf}(I) = a_n e^{-\boldsymbol{g}_0^\dagger \boldsymbol{g}_0} \int_0^{e^I - 1} dz_n \\ \cdot e^{-z_n a_n} I_0(2\sqrt{z_n a_n b_n}) \tilde{C}(e^I - 1 - z_n)\end{aligned} \tag{68}$$

where

$$\tilde{C}(x) = \left[\prod_{j=0}^{n-1} \int_0^\infty dz_j\right] \delta(x - \sum_{j=0}^{n-1} z_j) \prod_{j=1}^{n-1} a_j e^{-z_j a_j}. \tag{69}$$

Comparing to (52), we see that $\tilde{C}$ is nothing but the cdf for an $n-1$-dimensional system with $\boldsymbol{g}_0 = 0$, 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 $\boldsymbol{\Sigma} = \boldsymbol{I}_{n-1}$) the cdf is minimized (outage maximized) for $\boldsymbol{Q}$ being proportional to the identity matrix. Thus, the remaining $n-1$ eigenvalues of $\boldsymbol{Q}$ must all be equal to each other. Thus, we conclude that the optimal $\boldsymbol{Q}$ must have one eigenvector pointing in the direction $\boldsymbol{g}_0$ corresponding to the largest eigenvalue of $\boldsymbol{Q}$ and all other eigenvalues are equal to each other. This immediately implies that $\boldsymbol{Q}$ must have the form of (20).

ACKNOWLEDGMENT

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, June 1995.

[2] A. L. Moustakas *et al.*, "Communication through a diffusive medium: Coherence and capacity," *Science*, vol. 287, pp. 287–290, Jan. 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*, Oct. 31, 2000.

[4] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2632–2639, Sept. 2001.

[5] S. A. Jafar and A. Goldsmith, "On optimality of beamforming for multiple antenna systems," in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, June 2001, p. 321.

[6] A. Narula *et al.*, "Efficient use of side information in multple-antenna data transmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1423–1436, Oct. 1998.

[7] L. H. Ozarow, S. Shamai (Shitz), and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.

[8] A. L. Moustakas *et al.*, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large N analysis," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2545–2561, Oct. 2003.

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

[10] C. B. Papadias and G. J. Foschini, "A space-time coding approach for systems employing 4 transmit antennas," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, vol. 4, 2001, pp. 2481–2484.

[11] S. A. Jafar and A. Goldsmith, "Beamforming capacity and SNR maximization for muliple antenna systems with covariance feedback," in *Proc. Vehicular Technology Conf. (VTC)*, Rhodes, Greece, 2001, pp. 43–47.

[12] E. Jorswieck and H. Boche, "On transmit diversity with imperfect channel state information," in *Proc. IEEE Conf. Acoustics, Speech, and Signal Processing*, 2002, pp. 2181–2184.

[13] H. Boche and E. Jorswieck, "Analysis of diversity and multiplexing tradeoff for multi-antenna systems with covariance feedback," in *Proc. Vehicular Technology Conf. (VTC)*, 2002, pp. 864–868.

[14] S. H. Simon and A. L. Moustakas, "Optimizing MIMO systems with channel covariance feedback," *IEEE J. Select. Areas Commun.*, pp. 406–417, Apr. 2003.

[15] D. Chizhik *et al.*, "Effect of antenna separation on the capacity of BLAST in correlated channels," *IEEE Commun. Lett.*, vol. 4, pp. 337–339, Nov. 2000.

[16] S. H. Simon and A. L. Moustakas, "Optimality of beamforming in multiple transmit multiple receieve communications systems with partial channel knowledge," in *Proc. 2002 DIMACS Workshop on Multiantenna Channels: Capacity, Coding and Signal Processing Communications*, to be published.

[17] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley-Interscience, 1969.

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

[19] B. Hassibi and B. M. Hochwald, "High rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1804–1824, July 2002.

[20] S. D. Fisher, *Complex Variables*. Monterey, CA: Brook/Cole, 1986.

[21] A. L. Moustakas and S. H. Simon, "Outage capacity with 2–bit channel feedback for two-transmit and singel receive antenna system," in *Proc. IEEE GLOBECOM*, San Francisco, CA, Dec. 2003, to be published.