Optimizing MIMO Antenna Systems With Channel Covariance Feedback

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

***Abstract*—In this paper, we consider a narrowband point-to-point communication system with n_T transmitters and n_R receivers. We assume the receiver has perfect knowledge of the channel, while the transmitter has no channel knowledge. We consider the case where the receiving antenna array has uncorrelated elements, while the elements of the transmitting array are arbitrarily correlated. Focusing on the case where $n_T = 2$, we derive simple analytic expressions for the ergodic average and the cumulative distribution function of the mutual information for arbitrary input (transmission) signal covariance. We then determine the ergodic and outage capacities and the associated optimal input signal covariances. We thus show how a transmitter with covariance knowledge should correlate its transmissions to maximize throughput. These results allow us to derive an exact condition (both necessary and sufficient) that determines when beamforming is optimal for systems with arbitrary number of transmitters and receivers.**

***Index Terms*—Information rates, information theory, multi-input–multi-output (MIMO) systems.**

I. Introduction

The idea of exploiting knowledge of channel covariance in multi-input–multi-output (MIMO) systems was proposed by Moustakas *et al.* [1], Sengupta and Mitra [2], and also by Visotsky and Madhow [3], Jafar and Goldsmith [4], and Narula *et al.* [5] for the many-input–single-output (MISO) case. It has been shown in these works that increases in throughput can be obtained by appropriately exploiting this knowledge. A very detailed analytic study of the MISO case has previously been performed by the current authors [6]. In this paper, we focus on the simplest MIMO situation—that of two transmitters and many receivers two-input–many-output (TIMO)—where exact results can again be obtained analytically. It will turn out, however, that by studying this case in detail, we will also be able to deduce some important results for the more general case of arbitrary number of transmitters.

As with previous works [1], [6], we focus on how the mutual information changes as we change the input signal covariance (i.e., adapt the transmitter) to maximize the throughput. Of course, the mutual information of a random channel instantiation is also a random quantity and must be described appropriately. In this paper, we consider two measures of the typical mutual information for an ensemble of channels. One such measure is the average mutual information (or "ergodic capacity" [7], [8]). Another useful measure is the "x percent outage," which is defined to be the minimum mutual information that occurs in all but x percent 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 nontrivial correlations. In the past, this has been done typically by Monte Carlo methods (see, for example, [3] and [4]). Given that the averaging over the channel realizations is performed mostly numerically, the optimization process has, in the past, been very tedious and slow. Perhaps more importantly, the brute-force optimization allows little room to obtain intuition on the performance of a particular scheme.

An alternative approach, which we adopt in this paper for TIMO 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 a particular case of interest, the problem is tractable. Specifically, the case we will address in this paper is that of a channel with zero mean and a fixed covariance—where the channel has known nontrivial correlations at the transmitter and trivial correlations at the receiver (a case also discussed by Jafar [9]). This situation is likely to be quite representative of a receiving base station with well-spaced multiple antenna elements (so that the correlations on the receiving end become trivial) and a mobile with antenna elements that are spatially correlated due to finite angle spread of incoming signals and/or close antenna spacing [10]. The idea that the covariance at the transmitter end of the channel would be known by the transmitter is quite realistic for many wireless systems where the channel is changing rapidly, but correlations in the channel change at a slower pace. (One could also easily imagine feeding back the covariance to the transmitter to enable optimization of transmission and updating the covariance only very occasionally.) Such a scheme is qualitatively similar to closed-loop transmit diversity (CLTD) [11], where only a few bits describing the channel are fed back to the transmitter. Alternatively the transmitter could determine

Manuscript received April 30, 2002; revised November 26, 2002.

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

A. L. Moustakas is with the Wireless Advanced Technology Laboratory, Lucent Technologies, Whippany, NJ 07981 USA (e-mail: arislm@lucent.com).

Digital Object Identifier 10.1109/JSAC.2003.809698

the covariance of the downlink channel through uplink channel measurements.

For this simplified case where the covariance has the above described simple form, the problem of calculating both ergodic and outage capacities can be further reduced to evaluation of a single integral of known functions. This analytic approach is far simpler than Monte Carlo approaches, being that Monte Carlo requires averaging over $n_T n_R$ complex random variables (with $n_T =$ number of transmit antennas and $n_R =$ number of receive 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 [1], [2], [12]. The techniques used in such calculations are quite different from those used here, and we will leave discussion of those methods to a different paper.

In this paper, we will study the case of two transmitters and an arbitrary number of receivers (TIMO) in detail. It will turn out, however, that what we learn in this case will allow us to address questions regarding optimality of beamforming in systems with an arbitrary number of transmitters. The reason for this generality is essentially that beamforming amounts to using the full antenna array as a single (directed) antenna. The optimality condition for beamforming roughly requires that using the array as a single effective transmit antenna should be better than using the array as two effective transmitting antennas. It thus turns out that understanding TIMO systems allows us to determine a beamforming optimality condition for more general MIMO systems with an arbitrary number of receivers. As would be expected, at high enough antenna cross correlation, beamforming is optimal. The interpretation of this is clear. If one has very high antenna correlation, then it is very difficult to send more than one independent signal, so beamforming is optimal. Whereas at low antenna correlation the antennas are completely independent, it is typically advantageous to send more than one data stream to take advantage of the usual MIMO diversity gains. Similarly, low signal strength always tends to favor beamforming since it is already hard to decode a single data stream, and splitting the stream into two streams just makes matters worse.

A. Outline and Summary of Contribution

We begin with some general definitions in Section I-B and C applicable for general MIMO systems. As discussed above, in this paper, we focus only on the case where there are no correlations at the receiver antennas, but the transmitter antennas are correlated. In other words, we assume the channel matrix has zero mean, correlated rows, and uncorrelated columns—a reasonable model for certain multiantenna wireless systems [9].

In Section II-A, we present a number of novel results giving explicit analytic forms for the probability distribution function, cumulative distribution function, and ergodic average of the mutual information for a given transmission covariance in TIMO systems (two transmitters, arbitrary numbers of receivers). Prior work has obtained analytic expressions only for the case of single-receive antennas [6]. Although one might consider generalizing the method to more than two transmitters, the algebra quickly becomes so horrible that it appears essentially intractable. The tractability of TIMO case seems to be a result of the simplicity of the expression for the determinant of two-dimensional (2-D) matrices. We do not completely rule out, however, the possibility that a very determined researcher might extend the method to more complex cases.

In Section II-B, we discuss how to use these new analytic expressions to optimize over transmission covariance to maximize capacities (both ergodic and outage). Detailed examples of using these methods and interpretation of results are given in Section III.

In Section II-C, we use the above analytic results to discuss the question of whether beamforming is optimal or whether sending multiple data streams is optimal in MIMO systems with an *arbitrary number* of transmit antennas (and also an arbitrary number of receive antennas). In Section II-C1, we focus on optimizing the ergodic capacity. For this case, we are able to present a new, simple analytic condition for establishing whether beamforming is optimal. Specifically, our analytic result determines for any given signal-to-noise level and any given antenna correlation at the transmitter, whether beamforming is optimal or whether multiple data streams should optimally be used. Prior work in this field has obtained such an exact analytic result only for the simpler case of one receive antenna [4], [6], [13]. For the case of more than one receiver, bounds have been previously established to give a range where the transition between beamforming and multiple data streams might occur [9]. However, these bounds were not particularly tight. In the current work, we have improved these results by determining *precisely* where the transition occurs rather than just bounding the possible locations of the transition. In addition, our result is analytically just as simple as the expressions previously derived for the bounds. In addition, we derive an even simpler approximation to this analytic condition, which is asymptotically exact for large numbers of receivers and seems to be numerically quite accurate for more than one receiver. In Section II-C2, we look at optimality of beamforming in the case where one is attempting to maximize some outage capacity. For this situation, we derive a bound (a necessary condition) for the transition, which we believe may be tight (i.e., also sufficient) in many circumstances, but is known not to be a tight result in certain cases.

Finally, in Section III, we return to the case of TIMO systems (two transmitters) and present a number of results to demonstrate more general applications of these methods. In Fig. 3, we show a typical case of how the transmission with optimal covariance improves capacity over either beamforming or fully independent transmission. However, we also see that the gain over the better of these two options is rather small (a few percent). This result tends to be generic—one can always stay very close to the optimal capacity (either outage or ergodic) by simply switching from independent transmission to beamforming at the appropriate antenna cross correlation. However, we will also see that having the covariance feedback leads to substantial capacity gains, which may be as high as a factor of two over an open-loop scenario.

Finally, in Figs. 4–7, we show how the transmission covariance should be optimized as a function of the number of receive antennas, the SNR, and the antenna cross correlation. We show results for optimizing both the ergodic capacity and the 10% outage capacity.

B. Definition of Channel

In the narrowband MIMO problem—n_T transmitters and n_R receivers—the channel is defined by

$$y_\alpha = \sqrt{p}\sum_{i=1}^{n} G_{\alpha i}^* x_i + \eta_\alpha \tag{1}$$

where y_α is the n_R-dimensional complex vector of received signals, x_i is the n_T-dimensional complex vector of transmitted signal, $G_{i\alpha}$ is the complex channel from the ith transmitter to the αth receiver, and η_α is the complex noise at the αth receiver. We will typically use the notation that roman indices $i, j, \ldots$ represent the transmitter antennas and take values $1 \ldots n_T$, whereas Greek indices $\alpha, \beta, \ldots$ represent the receiver end and go from $1 \ldots n_R$. In the above channel equation, to make our normalizations easier, we have conveniently chosen to include a factor of $\sqrt{p}$, where p is the SNR per transmit antenna, which we will define more precisely below. In this paper, boldface quantities denote matrices or vectors (hopefully, usage will make clear which one is which); so, for example, we can also write $\mathbf{y} = \sqrt{p}\mathbf{G}^\dagger \mathbf{x} + \boldsymbol{\eta}$.

We define a transmitted signal covariance matrix to be given by $Q_{ij} = E\{x_i^* x_j\}$, where $E\{\cdot\}$ means expectation value (or temporal average), and $\mathbf{Q}$ is a 2-D non-negative definite hermitian matrix. Similarly, we define the noise $\boldsymbol{\eta}$ to be an independent and identically distributed (i.i.d.) complex Gaussian random vector. We have assumed that there are no notrivial correlation between the received noises. Note, however, that in Appendix A, the more general case of nontrivial correlations is considered briefly.

In this paper, we assume that the channel matrix $\mathbf{G}$ is $\mathcal{N}(0, \boldsymbol{\Sigma} \otimes \mathbf{1})$, i.e., a complex Gaussian random n_T by n_R-dimensional matrix with zero mean and covariance matrix $\boldsymbol{\Sigma} \otimes \mathbf{1}$. (Here, $\boldsymbol{\Sigma}$ is an n_T-dimensional hermitian matrix representing antenna correlations at the transmitter end, and $\mathbf{1}$ is an n_R-dimensional unit matrix representing trivial antenna correlations at the receiver end.) By this, we mean that $\langle G_{i\alpha}^* G_{j\beta}\rangle = \Sigma_{ij}\delta_{\alpha\beta}$, where $\delta_{\alpha\beta}$ is the Kronecker delta, and the brackets $\langle\cdot\rangle$ represent an ensemble average over instantiations of $\mathbf{G}$ (frequently, this ensemble average is written as $E\{\cdot\}$). In other words, the rows of the channel matrix are i.i.d., whereas the columns are correlated (a case also discussed by Jafar [9], [14]). One could have equivalently described this ensemble of channels as $\mathbf{G} = \sqrt{\boldsymbol{\Sigma}}\mathbf{Z}$, where $\mathbf{Z}$ is an n_T by n_R i.i.d. matrix. In Appendix A, the more general case of nontrivial correlations at the receiver are also considered briefly (i.e., where the $\delta_{\alpha\beta}$ is replaced by a more general matrix or, equivalently, where $\mathbf{Z}$ is also multiplied on the right by a second matrix as well).

Since we have separated out a factor of $\sqrt{p}$ (with p the signal-to-noise parameter) in the definition of the channel, we can choose a normalization $\mathrm{Tr}\{\boldsymbol{\Sigma}\} = \mathrm{Tr}\{\mathbf{Q}\} = n_T$. We can now more precisely define the signal-to-noise parameter p to be given as $p = P/\nu$, where ν is the (i.i.d.) noise power at each receive antenna, and P is the signal power received at each receive antenna *under conditions where* $\mathbf{Q} = \mathbf{1}$ (with $\mathbf{1}$ the n_T-dimensional unit matrix here), i.e., for uncorrelated transmissions. (For general transmission covariance $\mathbf{Q}$, the received signal power divided by the noise at each receive antenna is $p\mathrm{Tr}\{\mathbf{Q}\boldsymbol{\Sigma}\}/\mathrm{Tr}\{\boldsymbol{\Sigma}\}$). This normalization is chosen so that the total transmitted power stays fixed as the transmission covariance $\mathbf{Q}$ is changed subject to the normalization constraint $\mathrm{Tr}\{\mathbf{Q}\} = n_T$. We will frequently refer to the signal-to-noise parameter p as the "SNR per transmit antenna" even though this is only precisely correct for $\mathbf{Q} = \mathbf{1}$. We will leave the quotation marks around the phrase "SNR per transmit antenna" to remind ourselves that this interpretation is not precise.

Finally, we note that we will also frequently use the notation that a_i are the eigenvalues of $(p\mathbf{Q}\boldsymbol{\Sigma})^{-1}$ which are necessarily non-negative since both $\mathbf{Q}$ and $\boldsymbol{\Sigma}$ are non-negative definite.

C. Definitions of Quantities to Calculate

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

$$I((\mathbf{y}, \mathbf{G}); \mathbf{x}) = \log\det\left(\mathbf{1} + \mathbf{G}^\dagger p\mathbf{Q}\mathbf{G}\right) \tag{2}$$

where $\mathbf{1}$ is an n_R-dimensional unit matrix. Throughout this paper, we will measure information in nats/s/Hz, where 1 nat is equal to e bits ($e = 2.718\ldots$), i.e., the $\log$ in this equation is defined to be a natural log.

The probability distribution function PDF(I) of the mutual information I in (2) over the ensemble of instantiations $\mathbf{G}$ (with $\mathbf{Q}$ fixed) can then be written as

$$\mathrm{PDF}(I) = \left\langle \delta\left(I - \log\det\left(\mathbf{1} + \mathbf{G}^\dagger p\mathbf{Q}\mathbf{G}\right)\right)\right\rangle \tag{3}$$

where again, the brackets $\langle\cdot\rangle$ represent an ensemble average over realizations of $\mathbf{G}$ with $\delta()$ the Dirac delta function. Note that the probability density function (PDF) is implicitly a function of $\mathbf{Q}$, p, and $\boldsymbol{\Sigma}$, 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 of the mutual information to be the average (mean) of I over the ensemble

$$\langle I\rangle = \int_0^\infty dI I \mathrm{PDF}(I). \tag{4}$$

We also define the cumulative distribution function (CDF)

$$\mathrm{CDF}(I) = \int_0^I dI' \mathrm{PDF}\left(I'\right). \tag{5}$$

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

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

when

$$P_{\mathrm{out}} = \mathrm{CDF}\left(I_{\mathrm{out}}\right). \tag{7}$$

Since PDF$(I) > 0$, the inversion is unique. The meaning of $I_{\mathrm{out}} = \mathrm{OUT}(P_{\mathrm{out}})$ 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). We note that the 50% outage is the median of the distribution.

II. Analytic Results

A. TIMO Analytic Results

We begin by stating a number of results for the case where there are two transmitters ($n_T = 2$, i.e., TIMO systems) and an arbitrary number of receivers n_R. The transmission covariance is $\mathbf{Q}$, the channel covariance at the transmitter is $\mathbf{\Sigma}$, and we use the notation that a_i for $i = 1, 2$, are the eigenvalues of $(p\mathbf{Q}\mathbf{\Sigma})^{-1}$. If one of the two eigenvalues of $\mathbf{\Sigma}\mathbf{Q}$ is zero, this corresponds to a beamforming solution where one is using only one effective "eigen"-antenna. Mathematically, this corresponds to the divergence of either a_1 or a_2. We note in passing that one can take this limit analytically to recover the results (PDF, CDF, and ergodic average) for one to n_R transmission discussed previously in [6]. Finally, we note that the analytic results stated here have all been verified with Monte Carlo simulations.

1) PDF: In Appendix A, the PDF is found to be

$$\mathrm{PDF}\,(I) = e^I\left(-ia_1a_2\right)^{n_R}\int\frac{dk}{2\pi}\int_1^\infty dx\left(\frac{x-1}{x}\right)^{n_R-1} \cdot\frac{\exp\left[ik\left(e^I-x\right)-(x-1)a_2\right)\Big]}{(n_R-1)!\left[k-i\frac{a_1}{x}\right]^{n_R-1}\left[k-ia_1\right]}. \quad (8)$$

With some algebraic effort, the k contour integral can be integrated analytically to yield

$$\begin{aligned}\mathrm{PDF}\,(I) =& \frac{-e^I a_1\left(-a_2\right)^{n_R}}{(n_R-1)!} \cdot\int_1^{e^I} dx e^{-(x-1)a_2} \cdot\Bigg[e^{-a_1\left(e^I-x\right)} - e^{-a_1\left(e^I-x\right)/x}\\ &\cdot\sum_{j=0}^{n_R-2}\frac{1}{j!}\left(\left(\frac{1}{x}-1\right)\left(e^I-x\right)a_1\right)^j\Bigg]. \quad (9)\end{aligned}$$

We note that this expression is in fact symmetric under interchange of a_1 and a_2 although this is not obvious from these expressions. Also, we could perform the integral of the first term in brackets analytically, although we write in this form for simplicity of notation.

2) Mean (Ergodic) Mutual Information: The mean mutual information $\langle I\rangle$ can be derived by integrating the PDF [as shown in (4)]. Again, the calculation is complicated but is outlined in Appendix A, yielding

$$\langle I\rangle = \int_0^\infty dx\frac{e^{-x}}{x}\left[1-\frac{1}{(n_R-1)!}\left(\frac{a_1a_2}{x}\right)^{n_R} \cdot F_{n_R}\left(\frac{(a_1+x)(a_2+x)}{x}\right)\right] \quad (10)$$

where

$$F_n(y) = \frac{1}{y}\int_0^\infty dz\frac{e^{-z}z^{n-1}}{(z+y)^{n-1}}. \quad (11)$$

Note that the symmetry between a_1 and a_2 is now explicit again.

The integral defining F_n can be done analytically (simply defining a new integration variable $x = z + y$), yielding

$$\begin{aligned}yF_n(y) =& \sum_{j=0}^{n-1}\frac{(n-1)!(-y)^j e^y}{j!(n-1-j)!}\Gamma(-j+1,y)\\ =& 1+(n-1)ye^y\mathrm{Ei}[-y]\\ &+\sum_{j=2}^{n-1}\frac{(n-1)!y^j}{(n-1-j)!j!(j-1)!}\\ &\cdot\left[e^y\mathrm{Ei}[-y]-\sum_{k=1}^{j-1}(k-1)!(-y)^{-k}\right] \quad (12)\end{aligned}$$

with $\Gamma(a,b) = \int_b^\infty e^{-t}t^{a-1}dt$ the incomplete gamma function and $\mathrm{Ei}(b) = \Gamma(0,-b)$ the exponential integral. [The relation between the two expression in (12) is established by using successive integrations by parts.] Although this form looks messy, for small n, the answer is quite simple. For example, for $n = 2, 3, \ldots$, we have

$$\begin{aligned} yF_2[y] =& 1+ye^y\mathrm{Ei}[-y] \quad (13)\\ yF_3[y] =& (1+y)+\left(2y+y^2\right)e^y\mathrm{Ei}[-y]\\ &\vdots \quad (14)\end{aligned}$$

3) CDF: Analogous expressions for the CDF can be derived by directly integrating the PDF in the form of (8) to yield

$$\mathrm{CDF}\,(I) = \left(-ia_1a_2\right)^{n_R}\int\frac{dk}{2\pi}\int_1^\infty dx\left(\frac{x-1}{x}\right)^{n_R-1} \cdot\frac{\exp\left[ik\left(e^I-x\right)-(x-1)a_2\right)\Big]}{(n_R-1)!\left(k-i0^+\right)\left[k-\frac{ia_1}{x}\right]^{n_R-1}\left[k-ia_1\right]}. \quad (15)$$

Again this contour integral can also be done explicitly, resulting in (16), shown at the bottom of the next page. Again, a symmetry between a_1 and a_2 exists, although it is not apparent in this form.

B. Optimization Over $\mathbf{Q}$

To reach the full information capacity of the communication link, the transmitter has to optimize the transmitting signal covariance matrix $\mathbf{Q}$. Of course, in the closed-loop case when the transmitter has perfect channel knowledge, this is very straightforward (using so-called "waterpouring" on the eigenvalues of the channel matrix $\mathbf{G}$). However, the situation is more complicated when the transmitter has only partial (statistical) channel knowledge (i.e., the transmitter knows $\mathbf{\Sigma}$).

The transmitter may either choose to maximize the average mutual information $\langle I\rangle$, yielding the ergodic capacity [7], or the transmitter may choose to maximize the outage mutual information I_{out} for a fixed outage probability P_{out}. As seen for the MISO case [6], the resulting capacities and optimal $\mathbf{Q}$ matrices in these two cases can be quite different. This should be contrasted with the closed-loop case, where we obviously choose the same $\mathbf{Q}$'s independent of whether we are maximizing ergodic capacity or outage capacity.

In either case (outage capacity or ergodic capacity), it is sufficient to optimize over the eigenvalues of $\mathbf{Q}$ in the basis of $\mathbf{\Sigma}$

(i.e., $\mathbf{Q}$ and $\mathbf{\Sigma}$ should be simultaneously diagonalizable). The proof of this statement is given by Jafar [14] for the ergodic capacity case. We give a similar proof in Appendix B generalized to apply to the case of outage capacity as well. We note that in the TIMO case of two transmitters, optimization over the 2-D matrix $\mathbf{Q}$—either for maximizing ergodic or outage capacity—reduces to a trivial numerical optimization over a single parameter (recall that the trace of $\mathbf{Q}$ is fixed).

C. Optimality of Beamforming

We define beamforming to be the transmission of all of the power through the maximum eigenvalue and corresponding eigenvector of $\mathbf{\Sigma}$. (If the transmitter had full channel knowledge, one would prefer to beamform using the maximum eigenvalue and corresponding eigenvector of $\mathbf{GG}^\dagger$. However, we assume the transmitter has only knowledge of the channel covariance $\mathbf{\Sigma}$, which is more realistic since $\mathbf{\Sigma}$ changes very slowly compared to the channel itself and could, therefore, be fed back to the transmitter much more easily.) When $\mathbf{\Sigma}$ has more than one nonzero eigenvalue, we would like to know when beamforming remains optimal for maximizing either the ergodic average or some given outage capacity.

We write the eigenvalues of $\mathbf{Q}$ as q_i and the eigenvalues of $\mathbf{\Sigma}$ as s_i. As noted above, $\mathbf{Q}$ and $\mathbf{\Sigma}$ are simultaneously diagonalizable so that we can work in a basis with both of these matrices diagonal. Without loss of generality, we use a basis such that $s_1 \geq s_2 \geq s_3 \cdots \geq s_{n_T}$. The beamforming mode is such that $q_1 = n_T$, 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 can be trivially generalized for $s_2 = s_3$. For the beamforming solution to be optimal in maximizing 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 to preserve the constraint $\sum_i q_i = n_T$. In fact, it is easy to show that since $s_2 > s_3 > \cdots$, 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. Thus, the condition for beamforming can be written as

$$\left[\frac{\partial \mathcal{I}}{\partial q_2} - \frac{\partial \mathcal{I}}{\partial q_1}\right]_{q_1 = n_T, q_j = 0, j \geq 2} \leq 0 \tag{17}$$

where $\mathcal{I}$ represents either the ergodic mutual information $\langle I \rangle$ or an outage mutual information I_{out}. An equivalent condition has previously been derived by Visotsky [3] and Jafar [4], [9].

The above condition (17) requires us to think about the case where only two q's are nonzero (q_1 is nonzero and to differentiate with respect to q_2, we need to consider small but nonzero values of q_2). Since only two q's are nonzero, the beamforming criterion is equivalent to a TIMO problem. Since we know that the optimal transmission covariance $\mathbf{Q}$ is diagonal in the same basis as $\mathbf{\Sigma}$, we can work in a basis where $\mathbf{Q}$ and $\mathbf{\Sigma}$ are both diagonal (which can be done without loss of generality by a simple unitary transformation). Now if a given q_n is zero, this means that no power is transmitted from this "eigen"-transmitter, and we can eliminate that transmitter from the problem. Thus, if only q_1 and q_2 are nonzero, we have a TIMO problem. More specifically, examining (2) for the mutual information, again assuming $\mathbf{Q}$ and $\mathbf{\Sigma}$ are diagonal, when only q_1 and q_2 are nonzero, only a $2 \times n_R$ submatrix of the channel matrix $\mathbf{G}$ can contribute to the result. In fact, in that case the mutual information in (2) only depends on a $2 \times n_R$ submatrix of $\mathbf{G}$. If $\mathbf{Q}$ and $\mathbf{\Sigma}$ are both diagonal, then this $2 \times n_R$ matrix will be the tworows corresponding to the s_1 and s_2 eigenvalues values of Σ Thus, we can equivalently write a $2 \times n_R$ TIMO problem and obtain an analytic condition for beamforming optimality by appropriately differentiating our analytic expressions for TIMO capacities.

1) Ergodic Average Beamforming Optimality Condition: Since $\langle I \rangle$ as a function of q_2 is convex ($\partial^2 \langle I \rangle / \partial q_2^2 < 0$) over [0, 1] subject to $q_1 + q_2 = n_T$ and $q_j = 0$ for $j > 2$ (as shown by Jafar in [4]), then (17) is in fact a necessary and sufficient condition for beamforming to optimize the ergodic mutual information $\langle I \rangle$.

We take the relevant derivatives (17) of the expression for ergodic average (10) to yield (with a bit of algebra and some integrations by parts) the beamforming optimality condition

$$\Gamma\left(-n_R, \frac{1}{\rho s_1}\right) \geq \frac{(\rho s_1)^{n_R} \rho s_2}{1 + \rho s_2} \exp\left(\frac{-1}{\rho s_1}\right) \tag{18}$$

with Γ the incomplete gamma function and $\rho = n_T p$ the SNR. When (18) is not satisfied, beamforming is no longer optimal. For the case of $n_R = 1$ our result agrees with the result previously derived in [4], [6], and [13] in studies of the MISO system. For the more general case of arbitrary numbers of receivers, prior work in [9] has only been able to obtain two separate conditions for bounding the location of the optimality boundary. The bounds, however, were not particularly tight. Here, we have been able to derive a condition which gives the *exact* location of the boundary (being a both necessary and sufficient condition) which applies to systems with arbitrary numbers of transmitters and receivers [the only restriction being that the channel statistics are of the form $\mathcal{N}(0, \mathbf{\Sigma} \otimes \mathbf{1})$].

We note that by using different methods [1], [2], [12], a simpler approximate form of (18) can be derived

$$\rho s_2 < \frac{2\rho s_1}{1 + (n_R - 1)\rho s_1 + \sqrt{[(n_R - 1)\rho s_1 - 1]^2 + 4 n_R \rho s_1}} \tag{19}$$

$$\mathrm{CDF}(I) = \frac{a_2^{n_R}}{(n_R - 1)!} \int_1^{e^I} dx e^{-(x-1)a_2} \left[(x-1)^{n_R - 1} + (-1)^{n_R} e^{-a_1(e^I - x)} + (-1)^{(n_R+1)} e^{-a_1(e^I - x)/x} \sum_{j=0}^{n_R - 2} \frac{1}{j!} \left[a_1\left(e^I - x\right)\left(\frac{1}{x} - 1\right)\right]^j \left[1 - (1-x)^{n_R - 1 - j}\right]\right] \tag{16}$$

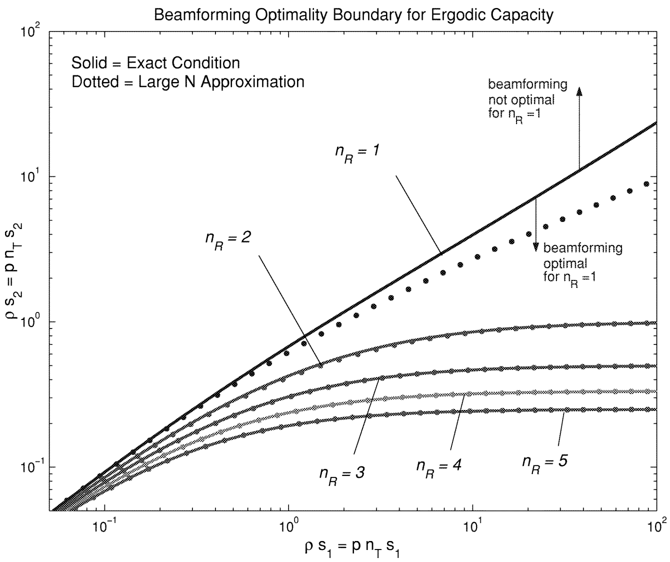

Fig. 1. Beamforming optimality condition for maximizing the ergodic capacity. The two axes represent ρs_1 and ρs_2, where $\rho = n_T p$ is the signal-to-noise and s_1 and s_2 are the two largest eigenvalues of the transmitter antenna array correlation matrix $\mathbf{\Sigma}$. Each curve separates the region of signal-to-noise ratio (SNR) below (and to the right of) which beamforming is optimal for n_R receive antennas ($n_R = 1 \ldots 5$). Note that the curves are both a necessary and sufficient condition for beamforming optimality. The solid curves are exact results (18), whereas the dotted curves are a simple approximation (19).

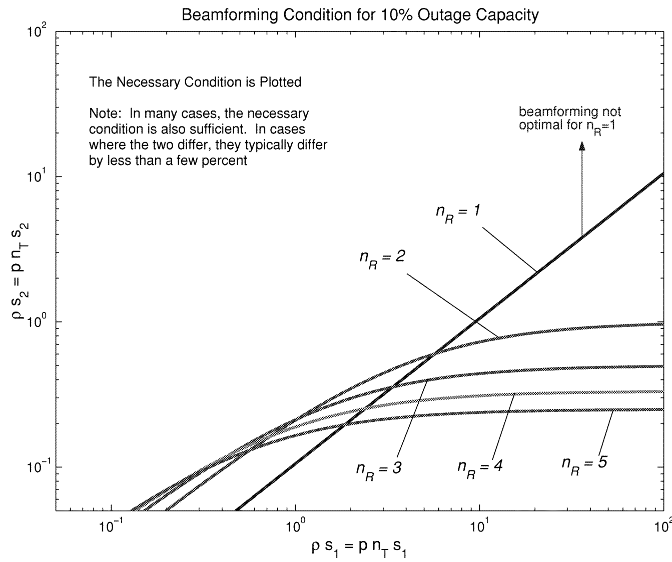

Fig. 2. Beamforming necessary condition for 10% outage capacity. The axes represent ρs_1 and ρs_2, where $\rho = p n_T$ is the signal-to-noise parameter and s_1, s_2 are the two largest eigenvalues of the transmit antenna array covariance matrix $\mathbf{\Sigma}$. Each curve separates the region of SNR below (and to the right of) which the beamforming condition (20) is satisfied for n_R receive antennas ($n_R = 1 \ldots 5$). Again, we emphasize that although these curves formally define a necessary, but not sufficient condition, over a large portion of the curves, the condition is also sufficient and the difference between the necessary and sufficient conditions are always numerically small.

which becomes rigorously valid in the limit of large numbers of antennas.

In Fig. 1, we show the regions of parameters ρs_1 and ρs_2 for which beamforming optimizes the ergodic capacity for $n_R = 1, \ldots, 5$ receivers where $\rho = p n_T$ is the signal-to-noise parameter, and s_1, s_2 are the two largest eigenvalues of the transmit antenna array covariance matrix $\mathbf{\Sigma}$. The exact results (solid curves) were calculated using (18). The region below (and to the right of) each curve is where beamforming is optimal, whereas in the region above (and to the left) beamforming is not optimal. We have also shown in this figure approximate results (dotted) obtained from the much simpler (19) which are extremely accurate for $n_R > 1$ and, for larger n_R, become almost indistinguishable from the exact results.

2) Outage Capacity Beamforming Optimality: We now turn to the case of outage capacity. We can again use (17) here, using the above analytic results for the outage capacity (CDF) of TIMO systems. Taking the derivatives explicitly, one obtains (with a bit of algebra) the beamforming condition

$$\rho s_2 \left[\frac{n_R}{e^{I_{\text{out}}^b} - 1} + n_R - 1 \right] < 1 \tag{20}$$

where again $\rho = n_T p$ is the total signal-to-noise parameter (p being the signal-to-noise per antenna) and I_{out}^b is the outage capacity for beamforming. Explicitly, we can write $P_{\text{out}} = C(I_{\text{out}}^b)$ where C is the CDF of the beamforming solution [6]

$$C(I) = \frac{1}{(n-1)!} \Gamma \left(n_R, \frac{e^I - 1}{\rho s_1} \right). \tag{21}$$

Equation (20) turns out to be necessary but not sufficient in general. As shown in [6], the outage capacity can sometimes be a nonconvex function (although the nonconvexity is typically very slight). Thus, the optimal q_2 can jump discontinuously from a finite value to zero. We note, however, that in many cases, there is no discontinuity in q_2, and the derived analytic condition does indeed define the optimality boundary. Furthermore, in all the cases we have looked at numerically—even when such a discontinuity exists—the above expression is not very different from the actual beamforming optimality boundary (typical deviations in the position of the boundary are on the order of a percent). This is just the statement that the outage capacity is not "too" nonconvex. We leave a detailed study of convexity (and determining when this necessary condition is also sufficient) to later work. For the time being, since we do not address convexity in any rigorous way here, the validity of (20) needs to be checked explicitly in every case numerically.

In Fig. 2, we focus on the case of 10% outage, and we show the regions of parameters ρs_1 and ρs_2 for which the beamforming condition (20) is satisfied for $n_R = 1, \ldots, 5$ receivers, where $\rho = p n_T$ is the signal-to-noise parameter, and s_1, s_2 are the two largest eigenvalues of the transmit antenna array covariance matrix $\mathbf{\Sigma}$. We emphasize again that although this condition is formally necessary but not sufficient, it differs only a very small amount (at most) from the sufficient condition. Note that at high SNR, the curves for $n_R > 1$ look very similar to the above curves in Fig. 1 for the ergodic capacity. This is simply the statement that the PDF is quite peaked so that the outage capacity and the ergodic capacity are not too different.

III. TIMO Results and Discussion

In this section, we focus on the case of $n_T = 2$ transmitters (TIMO).

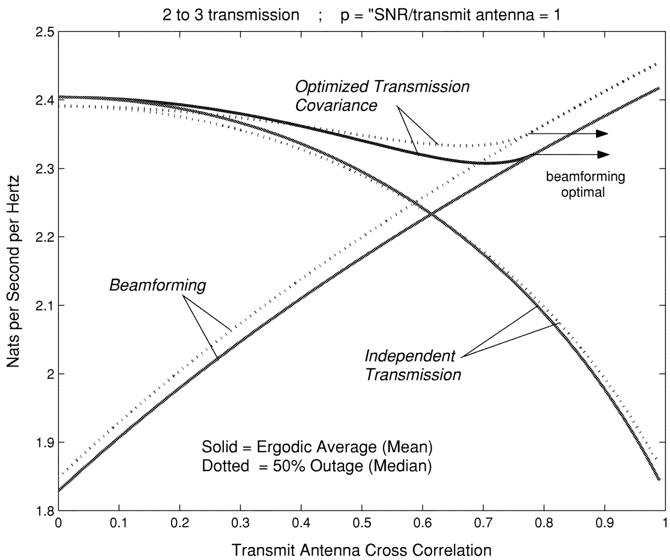

Fig. 3. Comparison of mutual information for different modes of transmission. Here, we are considering the case of two transmitters and three receivers with a signal-to-noise parameter "SNR/antenna" $= p = 1$. The abscissa is the antenna cross-correlation $x = (s_1 - s_2)/2$. Shown are curves for beamforming ($q_1 = 2$, $q_2 = 0$), uncorrelated transmission ($q_1 = q_2 = 1$), and the full capacity corresponding to the optimalvalues of q_1, q_2. The solid curves are ergodic (mean) capacities, whereas the dotted curves are 50% outage (median) capacities.

In Fig. 3, we show an example of calculated mutual information as a function of the transmit antenna cross correlation $x = (s_1 - s_2)/2$. (Antenna elements that are either very closely spaced or are receiving incoming waves from only a very narrow range of angles might be expected to have a large cross correlation.) Here, we have shown curves for three types of transmission covariance: beamforming ($q_1 = 2$ and $q_1 = 0$), independent transmission ($q_1 = q_2 = 1$), and optimal transmission covariance (optimized q_1 and q_2 subject to $q_1 + q_2 = 2$). Fig. 3 corresponds to $n_R = 3$ receive antennas ($n_T = 2$) with a signal-to-noise parameter of $p =$"SNR/transmit antenna" $= 1$. We have shown both the ergodic (mean) capacity (solid) and the 50% outage (median) capacity (dotted). The fact that the median and the mean are so similar indicates that the PDFs are roughly symmetric. It is interesting to note that the optimized capacity is nonmonotonic as a function of antenna cross correlation. This is clearly because at high cross-correlation beamforming is very effective, at low cross correlation, transmission of multiple data streams is very effective, but for intermediate cross correlations, both schemes are less effective.

It is clear from this figure that using beamforming at low antenna cross correlation or using independent transmission at high antenna cross correlation are both extremely inefficient. Note that as one might expect, at higher antenna cross-correlation, beamforming is favored and becomes optimal at a cross correlation of roughly 0.78 for this particular case. Furthermore, we see that using the optimal transmission covariance allows more capacity than either beamforming or independent transmission for intermediate cross correlations. However, it is also clear that switching from beamforming to independent transmission at an appropriate antenna cross correlation can come reasonably close to (within 5% of) the full capacity. This turns out to be generically true for any TIMO system we have examined: For any number of receivers and any signal-to-noise level and for either ergodic capacity or outage capacity maximization, we have found that either beamforming or fully independent transmission ($\mathbf{Q} = \mathbf{1}$) is always within roughly 5% of the maximum capacity that can be obtained with optimized transmission covariance (optimizing $\mathbf{Q}$). Thus, having a system that can only beamform or use fully independent transmission is always close to optimal. However, one must correctly choose when beamforming is better and when independent transmission is better. Were one to mistakenly choose beamforming at low antenna correlation at high SNR and with high number of receivers, one may fall short of the optimal capacity by up to 50% (in Fig. 3, the penalty is roughly 25%, but it increases as the signal strength increases). Similarly, incorrectly choosing independent transmission at high antenna correlation and low signal-to-noise could also cost roughly a factor of two (in Fig. 3, the penalty is again roughly 25%, but increases strongly at lower signal strength). The situation can be even more extreme when optimizing outage capacity for low outages (for example 10% outage), where the penalty for incorrectly choosing beamforming can be even more than a factor of two. This makes sense since beamforming is very susceptible to fading compared with independent transmission. Thus, it is clear that having covariance feedback of this sort can be quite advantageous, even if one is not fully optimizing, but rather just choosing between independent transmission and beamforming. We note that employing independent transmission does not require channel covariance knowledge at the transmitter. Indeed, in absence of transmission covariance knowledge (open loop), independent transmission is the optimal strategy. Thus, particularly at low signal strength, feeding back the channel covariance ($\mathbf{\Sigma}$) can result in substantial gains over open loop (up to factors of two or more at low enough signal strength or high antenna cross correlation) as it allows one to use beamforming when the there is high antenna cross correlation.

In Figs. 4 and 5, we focus on the transmission correlation that maximizes the ergodic capacity. In both figures, we plot the optimal q_2 (fraction of power to the nonbeamforming mode), where $q_2 = 1$ corresponds to independent transmission and $q_2 = 0$ corresponds to beamforming. As can be seen quite clearly from these figures, adding additional receive antennas ($n_R > 2$) is quite similar to increasing the overall signal-to-noise, as one might expect. Note in particular the similarities between the curves in Fig. 3 with $n_R = 2$ receive antennas with signal-to-noise parameter $p = 1, 2, 3, 4$ to the curves in Fig. 3 for $p = 1$ with $n_R = 2, 3, 4, 5$ receive antennas. In both of these figures, it is clear that high antenna cross correlation $x = (s_1 - s_2)/2$ and low effective SNR (low p or n_R) favors beamforming. This trend makes sense, as high antenna cross correlation makes transmission of independent data streams difficult.(Indeed, if the cross correlation were unity, then it would be impossible to transmit two independent data streams.) At such high antenna cross correlations, only at very high signal strengths does one want to "pour water" into the second, much weaker, eigen-antenna. Conversely, at low antenna cross correlation and high signal strength, one would like to make use of both transmission modes to get the full MIMO capacity enhancement so that

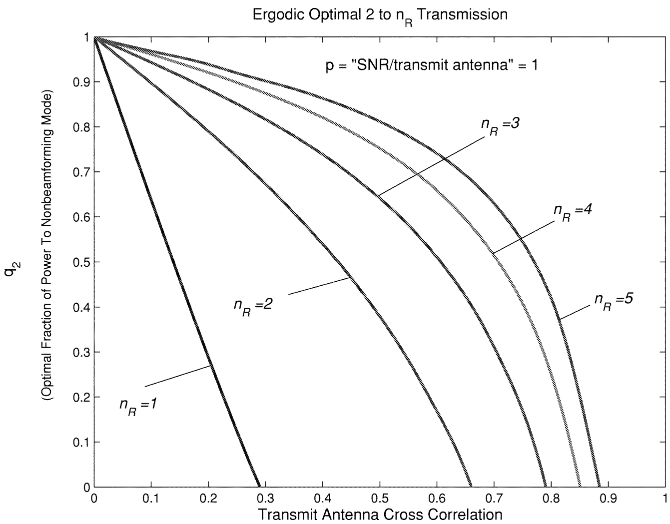

Fig. 4. Optimal power to nonbeamforming mode q_2 with two transmit antennas and n_R receive antennas, as a function of the antenna cross-correlation $x = (s_1 - s_2)/2$ of the transmit antennas. Here, we optimize q_2 to maximize the ergodic average. In this figure, we fix the signal-to-noise parameter "SNR/transmit antenna"$= p = 1$. Here, $q_2 = 0$ corresponds to beamforming, and $q_2 = 1$ corresponds to power being equally distributed between each of the two transmission eigenmodes. With increasing receive diversity, beamforming becomes optimal at much higher antenna cross correlation.

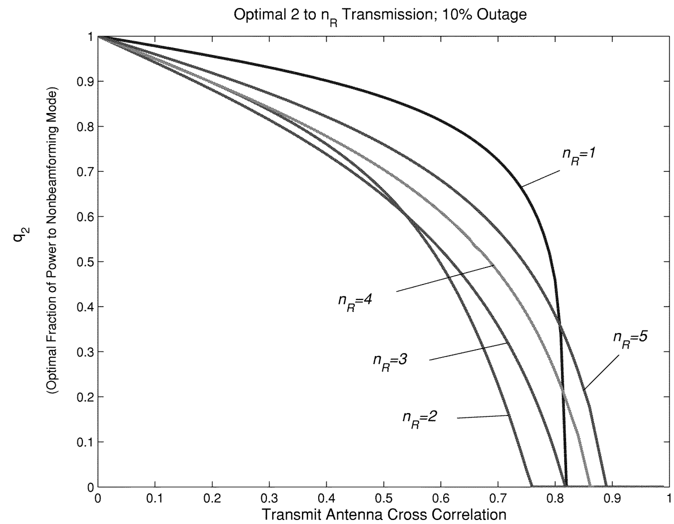

Fig. 6. Optimal power to nonbeamforming mode q_2 with two transmit antennas and n receive antennas, as a function of the antenna cross-correlation $x = (s_1 - s_2)/2$ of the transmit antennas. Here, the optimizationis to maximize the 10% outage capacity. In this figure, we fix the signal-to-noise parameter "SNR/transmit antenna" $= p = 1$. Here, $q_2 = 0$ corresponds to beamforming, and $q_2 = 1$ corresponds to power being equally distributed between each of the two transmission eigenmodes.

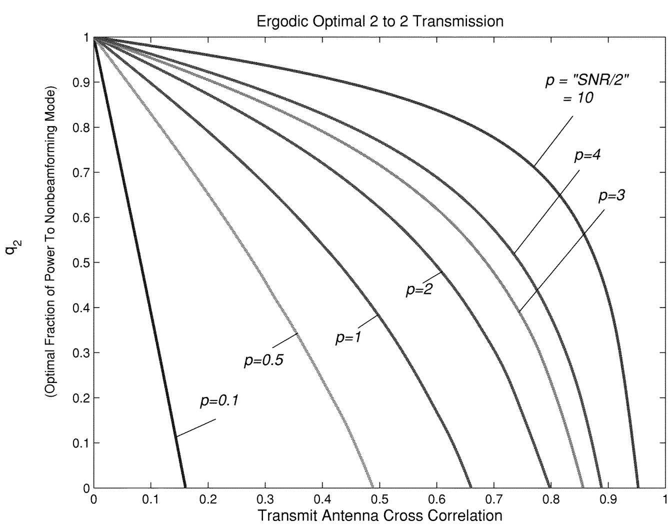

Fig. 5. Optimal power to nonbeamforming mode q_2 with two transmit antennas and two receive antennas, as a function of the antenna cross-correlation $x = (s_1 - s_2)/2$ of the transmit antennas. Here, the optimization of q_2 is to maximize the ergodic capacity. In this figure, we fix the various curves to correspond to different signal-to-noise parameters "SNR/transmit antenna" $= p = 0.1$, 0.5, 1, 2, 3, 4, and 10. Again, $q_2 = 0$ corresponds to beamforming, and $q_2 = 1$ corresponds to power being equally distributed between each of the two transmission eigenmodes. With increasing receive signal-to-noise, beamforming becomes optimal only at a much higher antenna cross correlation.

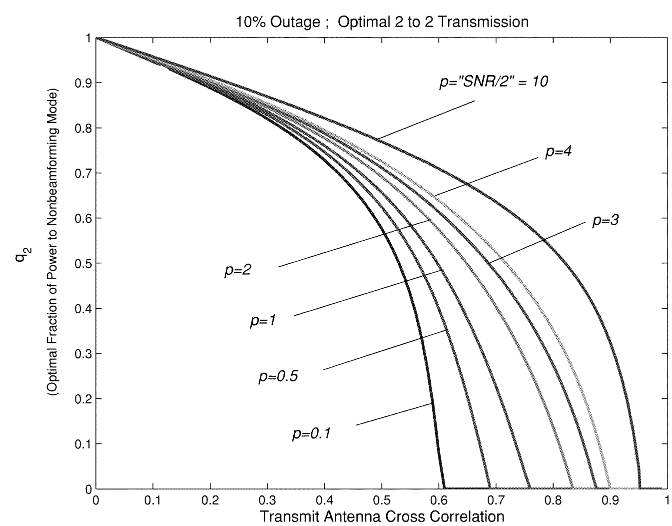

Fig. 7. Optimal power to nonbeamforming mode q_2 with two transmit antennas and two receive antennas, as a function of the antenna cross-correlation $x = (s_1 - s_2)/2$ of the transmit antennas. Here, the optimization is to maximize the 10% outage capacity. In this figure, the various curves correspond to different signal-to-noise parameters "SNR/transmit antenna" $= p = 0.1$, 0.5, 1, 2, 3, 4, and 10. Again, $q_2 = 0$ corresponds to beamforming and $q_2 = 1$ corresponds to power being equally distributed between each of the two transmission eigenmodes. With increasing receive signal-to-noise, beamforming becomes optimal at higher antenna cross correlation.

independent transmission is greatly favored. These trends are similar to those previously obtained in the MISO case [6].

In Figs. 6 and 7, we focus on the transmission correlation that maximizes the 10% outage capacity. These plots are analogous to Figs. 4 and 5 only, and it is the 10% outage we are maximizing rather than the ergodic (mean) capacity. For high receive antenna number, it is clear that the optimization of 10% outage is quite similar to optimizing the ergodic capacity. This is a reflection of the fact that the PDF becomes very peaked. However, for low receive antenna number—due to the relatively wider distribution of mutual informations—the optimizations yield substantially different results.

We note in passing that in all of the curves plotted in Figs. 6 and 7, the beamforming condition (20) above does indeed precisely define the optimality boundary [i.e., in these cases (20) is both necessary and sufficient].

One interesting feature that appears in Fig. 6 is the nonmonotonicity as a function of n_R. (Similar nonmonotonicity is observed in Fig. 2 at low signal strength.) We understand this as follows. With low numbers of receive antennas ($n_R = 1$), the system is highly susceptible to fades, so beamforming is extremely disfavored. This is more true for the case of outage capacity than for ergodic capacity since the PDF is very broad, and the fades can make the 10% outage very low. As we increase the number of antennas, the susceptibility to these fades drops, and beamforming is more favored. Further, averaging over the receivers makes the PDF more narrow, and optimizing the outage becomes the same as optimizing the ergodic capacity. Then, we have a similar trend as with the ergodic case where beamforming becomes less favored again as we go to higher numbers of receivers. In essence, the nonmonotonicity stems from change in the breadth of the PDF which is crucial for determining outage capacities but is irrelevant for ergodic capacities.

IV. Summary

In this paper, we have shown how to analytically calculate outages and ergodic averages of the mutual information of TIMO systems (two transmitting and n_R receiving antennas) as a function of the transmission covariance $\mathbf{Q}$, thus enabling us to determine which $\mathbf{Q}$ maximizes the mutual information. We have considered cases where the channel is described generally as being $\mathcal{N}(0, \mathbf{\Sigma} \otimes 1)$, i.e., having mean 0 and covariance $\mathbf{\Sigma}$ with respect to the transmitter and trivial correlations at the receiver. This case represents realistic situations where only partial channel information is fed back to the transmitter. While the channel itself may change quickly, the covariance may change much more slowly, thus allowing adaptation of the transmitter to the slowly varying properties of the channel. Using these TIMO results, we have also derived conditions for the optimality of beamforming applicable to MIMO systems with covariance feedback with arbitrary numbers of transmitters and receivers.

Appendix A
Details of Calculations

For a TIMO system with two transmitters and n receivers, the ensemble of $2 \times n$ channel matrices $\mathbf{G}$ is defined by

$$\left\langle G_{i\alpha} G^*_{j\beta} \right\rangle = \Xi_{\alpha\beta} \Sigma_{ij} \tag{22}$$

with $\mathbf{\Sigma}$ representing the correlations at the tranmitting antennas and $\mathbf{\Xi}$ representing the correlations at the receiving antennas. It is perhaps simplest conceptually to write $\mathbf{G} = \sqrt{\mathbf{\Sigma}}\mathbf{Z}\sqrt{\mathbf{\Xi}}$ with $\mathbf{Z}$ and i.i.d. matrix (i.e., $\langle Z_{i\alpha} Z^*_{j\beta}\rangle = \delta_{\alpha\beta}\delta_{ij}$) such that (22) is satisfied. The expectation of an arbitrary operator O can then be written explicitly as

$$\langle O \rangle = \int d\mu(\mathbf{Z}) O \tag{23}$$

where the integration measure is defined by

$$\int d\mu(\mathbf{Z}) = \prod_{i=1}^{2}\prod_{\alpha=1}^{n} \frac{1}{\pi} \int d\mathrm{Re} Z_{i\alpha} \int d\mathrm{Im} Z_{i\alpha} e^{-\mathrm{Tr}[\mathbf{Z}\mathbf{Z}^\dagger]} \tag{24}$$

which correctly defines $\mathbf{Z}$ as an i.i.d. matrix.

We write the mutual information as $I = \log\det(\mathbf{1} + \mathbf{QGPG}^\dagger)$, where $\mathbf{Q}$ is a fixed 2×2 matrix (representing the covariance of transmission), and $\mathbf{P}$ is a fixed n by n matrix (representing the covariance of the noises at the different receivers). We then rewrite this in terms of $\mathbf{Z}$ as $I = \log\det(\mathbf{1} + \tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger)$ with $\tilde{\mathbf{Q}} = \sqrt{\mathbf{\Sigma}}\mathbf{Q}\sqrt{\mathbf{\Sigma}}$ and $\tilde{\mathbf{P}} = \sqrt{\mathbf{\Xi}}\mathbf{P}\sqrt{\mathbf{\Xi}}$. We thus have

$$\begin{aligned}
\mathrm{PDF}(I) &= \left\langle \delta\left(I - \log\det\left(\mathbf{1} + \tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)\right)\right\rangle && (25)\\
&= \int d\mu(\mathbf{Z})\delta\left(I - \log\det\left(\mathbf{1} + \tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)\right) && (26)\\
&= e^I \int d\mu(\mathbf{Z})\delta\left(e^I - \det\left(\mathbf{1} + \tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)\right). && (27)
\end{aligned}$$

It is convenient to rewrite the 2×2 determinant in terms of its four elements as

$$\begin{aligned}
\det\left(\mathbf{1} + \tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right) = &\left(\mathbf{1} + \tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)_{11}\left(\mathbf{1} + \tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)_{22}\\
&+ \frac{1}{4}\left(\tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger{}_{12} + \tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger{}_{21}\right)^2\\
&- \frac{1}{4}\left(\tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger{}_{12} - \tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger{}_{21}\right)^2.
\end{aligned} \tag{28}$$

The subscripts here indicate an element of the 2×2 matrix. The 11 and 22 terms on the first line are real. The first term in parentheses on the second line is real, whereas the second term in parentheses is imaginary. We can then write

$$\begin{aligned}
\mathrm{PDF}(I) = &e^I \int d\mu(\mathbf{Z}) \int dx_1 dx_2 dx_3\\
&\cdot \delta\left(e^I - \left(\mathbf{1} + \tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)_{11} x_1 + \frac{1}{4}\left(x_2^2 + x_3^2\right)\right)\\
&\cdot \delta\left(x_1 - \left(\mathbf{1} + \tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)_{22}\right)\\
&\cdot \delta\left(x_2 - \left(\tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)_{12} + \left(\tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)_{21}\right)\\
&\cdot \delta\left(x_3 - i\left(\tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)_{12} + i\left(\tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)_{21}\right).
\end{aligned} \tag{29}$$

To see that this is the same as (27), one can perform the integral over x_1, x_2, and x_3 explicitly to leave only the first delta function remaining. Inside that delta function, one would then have the determinant written in the form of (28).

Using the Fourier representation of the delta function $\delta(x) = \int dk e^{ikx}/(2\pi)$ yields (30) and (31), shown at the bottom of the next page, where $\mathbf{M}$ is the matrix

$$\mathbf{M} = \begin{pmatrix} ikx_1 & iq_2 - q_3 \\ iq_2 + q_3 & iq_1 \end{pmatrix}. \tag{32}$$

We now aim to integrate out $\mathbf{Z}$. It is well-known that Gaussian integrals of this sort can always be done trivially. In our case, we have

$$\begin{aligned}
\int d\mu(\mathbf{Z}) e^{-\mathrm{Tr}[\mathbf{M}\tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger]} &= \prod_{i=1}^{2}\prod_{\alpha=1}^{n}\frac{1}{\pi}\int d\mathrm{Re}Z_{i\alpha}\\
&\quad \cdot \int d\mathrm{Im}Z_{i\alpha} e^{-\mathrm{Tr}[\mathbf{Z}\mathbf{Z}^\dagger + \mathbf{M}\tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger]}\\
&= \det\left[\mathbf{1} \otimes \mathbf{1} + \mathbf{M}\tilde{\mathbf{Q}} \otimes \tilde{\mathbf{P}}\right]^{-1}
\end{aligned} \tag{33}$$

where the outer product $\otimes$ is defined as follows: Since $\mathbf{M}\tilde{\mathbf{Q}}$ is a 2-D matrix (with indices $i,j = 1,2$) and $\tilde{\mathbf{P}}$ is an n-dimensional matrix (with indices $\alpha,\beta = 1,\ldots n$), then $\mathbf{M}\tilde{\mathbf{Q}} \otimes \tilde{\mathbf{P}}$ is a $2n$-dimensional matrix with indices $(i,\alpha) = (1,1),\ldots,(1,n),(2,1),\ldots(2,n)$. The elements of $\mathbf{M}\tilde{\mathbf{Q}} \otimes \tilde{\mathbf{P}}$ are given by $[\mathbf{M}\tilde{\mathbf{Q}} \otimes \tilde{\mathbf{P}}]_{(i\alpha),(j\beta)} = (\mathbf{M}\tilde{\mathbf{Q}})_{ij}\tilde{\mathbf{P}}_{\alpha\beta}$.

Thus, we can integrate out $\mathbf{Z}$ to obtain (34), shown at the bottom of the page. It is trivial to integrate out x_2 and x_3, shown in (35), at the bottom of the page. Here, we have also used the fact that $\det[\mathbf{A}]\det[\mathbf{B}] = \det[\mathbf{AB}]$ to move around factors of $\tilde{\mathbf{Q}}$ and $\tilde{\mathbf{P}}$ in rewriting the determinant.

Let us define the eigenvalues of $\tilde{\mathbf{P}}^{-1}$ to be ϵ_m which all must be non-negative (these are also the eigenvalues of $\mathbf{\Xi}^{-1}\mathbf{P}^{-1}$). Let us also work in a basis where $\tilde{\mathbf{Q}}$ is diagonal with (again non-negative) eigenvalues a_1 and a_2 (these are also the eigenvalues of $\mathbf{\Sigma}^{-1}\mathbf{Q}^{-1}$). We can then can write the determinants as

$$\det\left[\tilde{\mathbf{Q}}^{-1} \otimes \tilde{\mathbf{P}}^{-1} + \mathbf{M} \otimes \mathbf{1}\right] = \prod_{m=1}^{n}\left[(\epsilon_m a_1 + ikx_1) \cdot (\epsilon_m a_2 + iq_1) + \left(q_2^2 + q_3^2\right)\right] \tag{36}$$

$$\det\left[\tilde{\mathbf{Q}}^{-1} \otimes \tilde{\mathbf{P}}^{-1}\right] = (a_1a_2)^n \prod_{m=1}^{n} \epsilon_m^2. \tag{37}$$

We then notice that q_2 and q_3 only ever occur in the combination $q_2^2 + q_3^2$. Thus, we define $y = q_2^2 + q_3^2$ so we have (dropping the index 1 from the remaining x and q for notational simplicity)

$$\mathrm{PDF}\,(I) = e^I \left[(a_1a_2)^n \prod_{m=1}^{n} \epsilon_m^2\right] \int \frac{dk}{2\pi} \int dx \int \frac{dq}{2\pi} \int_0^\infty dy \cdot \frac{\exp\left[ik\left(e^I - x\right) + iq(x-1) - \frac{iy}{k}\right]}{ik\prod_m \left[(\epsilon_m a_1 + ikx)(\epsilon_m a_2 + iq) + y\right]}. \tag{38}$$

We next aim to integrate out q. We note that all of the poles are in the upper half plane. Thus, the integration restricts $x > 1$. Very generally, we can perform this contour integral to obtain (assuming all of the eigenvalues ϵ_m to be different for now) in (39), shown at the bottom of the next page. Note that a similarly general expression for the cumulative distribution function can be found at this point by analytically integrating this with respect to I. Another thing to note about this expression is that (although it is not obvious) it must be symmetric under interchange of a_1 and a_2. In general, the integral over y can be performed analytically. For example, for the general case of 2 $\times$2 transmission, we obtain

$$\begin{aligned}\mathrm{PDF}(I) =& e^I\,(a_1a_2\epsilon_1\epsilon_2)^2 \int \frac{dk}{2\pi} \int_1^\infty dx \frac{\exp\left[ik\left(e^I - x\right)\right]}{k\,(\epsilon_1 - \epsilon_2)\,a_1} \\ &\times \left[e^{-(x-1)\epsilon_1 a_2} f\left[\left(\frac{a_2}{a_1}\right)(\epsilon_2 a_1 + ikx) \cdot \left(1 - \frac{i\epsilon_1 a_1}{k}\right)\right] - (\epsilon_1 \leftrightarrow \epsilon_2)\right]\end{aligned} \tag{40}$$

$$\begin{aligned}\mathrm{PDF}\,(I) =& e^I \int d\mu(\mathbf{Z}) \int \frac{dk}{2\pi} \int dx_1 dx_2 dx_3 \int \frac{dq_1}{2\pi}\frac{dq_2}{2\pi}\frac{dq_3}{2\pi} \\ &\cdot \exp\Bigg[ik\left(e^I - \left(\mathbf{1} + \tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)_{11} x_1 + \frac{1}{4}\left(x_2^2 + x_3^2\right)\right) + iq_1\left(x_1 - \left(\mathbf{1} + \left(\tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)_{22}\right)\right) \\ &+ iq_2\left(x_2 - \left(\tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)_{12} + \left(\tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)_{21}\right) + iq_3\left(x_3 - i\left(\tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)_{12} + i\left(\tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger\right)_{21}\right)\Bigg]\end{aligned} \tag{30}$$

$$\begin{aligned}=& e^I \int d\mu(\mathbf{Z}) \int \frac{dk}{2\pi} \int dx_1 dx_2 dx_3 \int \frac{dq_1}{2\pi}\frac{dq_2}{2\pi}\frac{dq_3}{2\pi} e^{-\mathrm{Tr}[\mathbf{M}\tilde{\mathbf{Q}}\mathbf{Z}\tilde{\mathbf{P}}\mathbf{Z}^\dagger]} \\ &\cdot \exp\left[ik\left(e^I - x_1 + \frac{1}{4}\left(x_2^2 + x_3^2\right)\right) + iq_1(x_1 - 1) + iq_2x_2 + iq_3x_3\right]\end{aligned} \tag{31}$$

$$\mathrm{PDF}\,(I) = e^I \int \frac{dk}{2\pi} \int dx_1 dx_2 dx_3 \int \frac{dq_1}{2\pi}\frac{dq_2}{2\pi}\frac{dq_3}{2\pi} \cdot \frac{\exp\left[ik\left(e^I - x_1 + \frac{1}{4}\left(x_2^2 + x_3^2\right)\right) + iq_1(x_1 - 1) + iq_2x_2 + iq_3x_3\right]}{\det\left[\mathbf{1} \otimes \mathbf{1} + \mathbf{M}\tilde{\mathbf{Q}} \otimes \tilde{\mathbf{P}}\right]} \tag{34}$$

$$\mathrm{PDF}(I) = 4\pi e^I \det\left[\tilde{\mathbf{Q}}^{-1} \otimes \tilde{\mathbf{P}}^{-1}\right] \int \frac{dk}{2\pi} \int dx_1 \int \frac{dq_1}{2\pi}\frac{dq_2}{2\pi}\frac{dq_3}{2\pi} \cdot \frac{\exp\left[ik\left(e^I - x_1\right) + iq_1(x_1 - 1) - \frac{i}{k}\left(q_2^2 + q_3^2\right)\right]}{ik\det\left[\tilde{\mathbf{Q}}^{-1} \otimes \tilde{\mathbf{P}}^{-1} + \mathbf{M} \otimes \mathbf{1}\right]} \tag{35}$$

where $f(x) = \exp(x)\mathrm{Ei}(-x)$, and $(\epsilon_1 \leftarrow \epsilon_2)$ represents the first term in brackets with ϵ_1 and ϵ_2 interchanged. Unfortunately, this expression is not easily simplified further. For this reason, we limit our attention to the case where there are trivial correlations at the receiver end; in other words, we consider the case where $\tilde{\mathbf{P}}$ (and, hence, $\mathbf{\Xi P}$) is proportional to the unit matrix. In this case, all of the ϵ's are the same (it is then convenient to absorb one factor of ϵ into each a_i). The contour integral over q then isolates a single nth-order pole yielding (41), shown at the bottom of the page. The integration over y is now trivially done to yield (8).

To obtain the ergodic average capacity, one starts with the form of the PDF in (8) and integrate explicitly as prescribed in (4) to yield

$$\langle I \rangle = \frac{(a_1 a_2)^n}{(n-1)!} \int \frac{dk}{2\pi} \int_1^\infty dx \left(\frac{x-1}{x}\right)^{n-1} \cdot \frac{\mathrm{Ei}(ik)\exp\left[-ikx-(x-1)a_2\right]}{(k+i0^+)\left[k-\frac{ia_1}{x}\right]^{n-1}\left[k-ia_1\right]} \tag{42}$$

with Ei exponential integral function which we then rewrite using the identity

$$\mathrm{Ei}(ik) = e^{ik}\int_0^\infty dy \frac{e^{-y}}{ik-y}. \tag{43}$$

The k contour is closed in the lower half plane and the result is simplified to give (10).

Appendix B
Optimal $\mathbf{Q}$ is Diagonal

We show that that the $\mathbf{Q}$ that maximizes the capacity (either outage or ergodic) is diagonal. We will need to use the fact that the capacity (outage or ergodic) is only a function of the eigenvalues of $p\mathbf{Q\Sigma}$ and that the capacity is a monotonic increasing function of the signal-to-noise parameter p. We will proceed by considering an arbitrary $\mathbf{Q}$ with off-diagonal components and explicitly constructing a diagonal $\tilde{\mathbf{Q}}$ with greater or equal capacity (both $\mathbf{Q}$'s must satisfy the trace condition $\mathrm{Tr}[\mathbf{Q}] = n_T$).

We work in a basis where $\mathbf{\Sigma}$ is diagonal with eigenvalues $s_1 \geq s_2 \geq \cdots \geq s_{n_T}$ without loss of generality. We construct the matrix $\mathbf{\Sigma Q}$ and make a unitary transformation such that we write it as $\mathbf{\Sigma Q} = \mathbf{U}^\dagger \mathbf{\Lambda U}$ with $\mathbf{U}$ an appropriate diagonalizing unitary matrix and $\mathbf{\Lambda}$ the diagonal matrix of eigenvalues λ_i. By choosing $\mathbf{U}$ appropriately, we can arrange such that the elements of $\mathbf{\Lambda}$ are ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_{n_T}$. We now construct a new matrix $\mathbf{Q}' = \mathbf{\Sigma}^{-1}\mathbf{\Lambda}$. (If $\mathbf{\Sigma}$ has zero eigenvalues, we define the corresponding elements of $\mathbf{Q}'$ to be also zero.) The matrix $\mathbf{Q}'$ is explicitly constructed such that the eigenvalues of $\mathbf{\Sigma Q}'$ are exactly the same as the eigenvalues of $\mathbf{\Sigma Q}$. Thus, transmission with covariance $\mathbf{Q}'$ has precisely the same capacity as transmission with covariance $\mathbf{Q}$. However, we can also show (see the Lemma below) that $\mathrm{Tr}[\mathbf{Q}'] \leq \mathrm{Tr}[\mathbf{Q}]$. This inequality means that the power transmitted with $\mathbf{Q}'$ is lower than that transmitted with $\mathbf{Q}$. Thus, we can bring the power up to the original strength by constructing $\tilde{\mathbf{Q}} = \mathbf{Q}'(\mathrm{Tr}[\mathbf{Q}]/\mathrm{Tr}[\mathbf{Q}'])$, effectively increasing the signal strength parameter p. This new matrix $\tilde{\mathbf{Q}}$ is thus diagonal, has the same trace as $\mathbf{Q}$, and, since we have now increased the power, must have more capacity than $\mathbf{Q}$. Thus, we conclude that the optimal $\mathbf{Q}$ is always diagonal.

Lemma: $\mathrm{Tr}[\mathbf{Q}'] \leq \mathrm{Tr}[\mathbf{Q}]$ or equivalently $\mathrm{Tr}[\mathbf{\Sigma}^{-1}\mathbf{\Lambda}] \leq \mathrm{Tr}[\mathbf{\Sigma}^{-1}\mathbf{U}^\dagger\mathbf{\Lambda U}]$.

Consider two sets of numbers $s_1^{-1}, \ldots s_{n_T}^{-1}$ and $\lambda_1, \ldots \lambda_{n_T}$. Construct a sum of products $s_k^{-1}\lambda_l$ such that each number is used in exactly one product. As shown by Jafar [9], the permutation which minimizes the sum is such that the largest s^{-1} is paired with the smallest λ, the second largest s^{-1} is paired with the second smallest λ, and so forth, Thus, of all possible pairings of the eigenvalues s_i^{-1} and λ_j, the one constructed by $\mathrm{Tr}[\mathbf{\Sigma}^{-1}\Lambda]$ yields the minimal value in its sum. In other words, if we permute the order of the eigenvalues of $\mathbf{\Sigma}$ or $\mathbf{\Lambda}$ so that they are not in increasing order, we will obtain a larger trace. We can state this mathematically by considering permutation matrices $\mathbf{M}$ with the property that all elements are either zero or one and that there is exactly one nonzero element in any row or any column, and we have $\mathrm{Tr}[\mathbf{\Sigma}^{-1}\mathbf{\Lambda}] \leq \mathrm{Tr}[\mathbf{\Sigma}^{-1}\mathbf{M\Lambda}]$ for any such permutation matrix $\mathbf{M}$. We now have

$$\mathrm{Tr}\left[\mathbf{\Sigma}^{-1}\mathbf{U}^\dagger\mathbf{\Lambda U}\right] = \sum_{ij} s_i^{-1}\left|U_{ij}\right|^2\lambda_j. \tag{44}$$

$$\mathrm{PDF}(I) = e^I \left[(a_1a_2)^n \prod_{m=1}^n \epsilon_m^2\right] \int \frac{dk}{2\pi}\int_1^\infty dx \int_0^\infty dy \frac{\exp\left[ik\left(e^I - x\right) - \frac{iy}{k}\right]}{k\prod_m\left(\epsilon_m a_1 + ikx\right)} \times \sum_p \frac{\exp\left[(x-1)\left(-\epsilon_p a_2 - \frac{y}{\epsilon_p a_1 + ikx}\right)\right]}{\prod_{l\neq p}\left(\epsilon_l a_2 + \frac{y}{\epsilon_l a_1 + ikx} - \epsilon_p a_2 + \frac{y}{\epsilon_p a_1 + ikx}\right)} \tag{39}$$

$$\mathrm{PDF}(I) = e^I (a_1a_2)^n \int \frac{dk}{2\pi}\int_1^\infty dx\int_0^\infty dy \cdot \frac{(x-1)^{n-1}\exp\left[ik\left(e^I-x\right) - \frac{iy}{k} + (x-1)\left(-a_2 - \frac{y}{a_1+ikx}\right)\right]}{(n-1)!k\left[a_1+ikx\right]^n} \tag{41}$$

We now claim (see below for the proof) that we can make the following decomposition:

$$|U_{ij}|^2 = \sum_k \alpha_k M_{ij}^k \tag{45}$$

where each $\mathbf{M^k}$ is a permutation matrix, each $\alpha_k \geq 0$ and $\sum_k \alpha_k = 1$. Thus, we have

$$\begin{aligned} \text{Tr}\left[\mathbf{\Sigma^{-1}U^{\dagger}\Lambda U}\right] &= \sum_k \alpha_k \text{Tr}\left[\mathbf{\Sigma^{-1}M^k\Lambda}\right] \\ \geq \sum_k \alpha_k \text{Tr}\left[\mathbf{\Sigma^{-1}\Lambda}\right] &= \text{Tr}\left[\mathbf{\Sigma^{-1}\Lambda}\right] \end{aligned} \tag{46}$$

proving the Lemma.

Proof of Claim: We will explicitly construct the desired decomposition. (Note that the decomposition is not unique in general.) For simplicity of notation, write $N_{ij}^0 = |U_{ij}|^2$. All elements of $\mathbf{N}^0$ are non-negative definite. Since $\mathbf{U}$ is unitary, we have $\sum_i N_{ij}^0 = 1$ for any j, and similarly, $\sum_j N_{ij}^0 = 1$ for any i. Find the smallest nonzero element of $\mathbf{N}^0$, let us call this N_{i_0,j_0}^0. Assume there are m elements of $\mathbf{N}^0$ that are nonzero with $n_T \leq m \leq {n_T}^2$ (the sum conditions require there to be at least one nonzero element per row and one nonzero element per column). Choose any of the permutation matrices which have a 1 in the i_0, j_0 slot and have a zero in any of the slots where N^0 is already zero (it is guaranteed that we can do this since there is at least one nonzero element per row and per column). Call the chosen permutation matrix M^0. Now, construct $\mathbf{N}^1 = \mathbf{N}^0 - \alpha_0\mathbf{M}^0$, where $\alpha_0 = N_{i_0,j_0}^0$. We still have all of the elements of $\mathbf{N}^1$ being non-negative (since we have subtracted the magnitude of the smallest nonzero element, and we have not subtracted anything from any of the zero elements). Now, we have $\sum_i N_{ij}^1 = 1-\alpha^0$ for any j, and similarly, $\sum_j N_{ij}^1 = 1 - \alpha^0$ for any i. If the resulting N^1 is zero (i.e., if $1 - \alpha^0$ is zero), then we are done, and we have completely decomposed the matrix $\mathbf{N}^0$. Otherwise, we repeat the process: Find the smallest nonzero element of $\mathbf{N}^1$, call it N_{i_1,j_1}^1, and choose a permutation matrix $\mathbf{M}^1$ with 1 in the i_1, j_1 slot and zeros any place that N^1 is already zero. Now, construct $\mathbf{N}^2 = \mathbf{N}^1 - \alpha_1\mathbf{M}^1$ with $\alpha_1 = N_{i1,j1}^1$ and so forth until we have fully decomposed $\mathbf{N}^0$ (it is obvious that this process ends in no more than ${n_T}^2$ steps). Thus, we have proved the claim.

REFERENCES

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

[2] 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/0 010 081*, Oct. 2000.

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

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

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

[6] S. H. Simon and A. L. Moustakas, "Optimizing multitransmitter single receiver antenna systems with partial channel knowledge," *IEEE Trans. Inform. Theory*, submitted for publication.

[7] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun. Related Technol.*, vol. 10, no. 6, pp. 585–596, Nov. 1995.

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

[9] 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.

[10] A. L. Moustakas *et al.*, "Wideband characteristics of the urban PCS channel," presented at the Int. Symp. 3G Infrastructure Service, Athens, Greece, July 2001.

[11] "3GPP Tech. Spec. "Physical Layer Procedures (FDD)",", Release 5, 3GPP TS 25.214 V5.3.0, Dec. 2001.

[12] S. H. Simon, A. L. Moustakas, and A. M. Sengupta, *MIMO Capacity Through Channel With Correlated Interference and Noise: (Not So) Large N Analysis*, 2002, submitted for publication.

[13] E. Jorswieck and H. Boche, "On transmit diversity with imperfect channel state information," in *Proc. IEEE Conf. Acoustics, Speech and Signal Processing*, vol. 3, 2002, pp. 2181–2184.

[14] S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *Proc. Int. Conf. Communications (ICC2001)*, vol. 7, Helsinki, Finland, 2001, pp. 2266–2270.

Steven H. Simon received the B.Sc. degree in physics and mathematics from Brown University, Providence, RI, in 1989, and the M.S. and Ph.D. degrees in theoretical condensed matter physics from Harvard University, Cambridge, MA, in 1992 and 1995, respectively.

From 1995 to 1997, he was a Postdoctoral Researcher in the Physics Department, Massachusettes Institute of Technology, Cambridge. In 1997, he joined Bell Labs, Lucent Technologies, Murray Hill, NJ, as a Member of Technical Staff in the Theoretical Physics Department. In January of 2000, he was promoted to Director of theoretical physics research. His current research interests include multiple antenna communications, information theory, materials physics, quantum phenomena, and nanotechnology.

Aris L. Moustakes (M'02) received the B.S. degree in physics from the California Institute of Technology, Pasadena, in 1990 and the M.S. and Ph.D. degrees in theoretical solid-state physics from Harvard University, Cambridge, MA, in 1992 and 1996, respectively.

In 1998, after a two-year stint in the Greek Navy, he joined Bell Labs, Lucent Technologies, first in the Physical Sciences Division, Murray Hill, NJ, and then in the Wireless Advanced Technology Laboratory, Whippany, NJ. His current research interests lie in the areas of multiple antenna systems, signal processing for smart antennas, and wideband channel modeling. He has authored several papers and patents on these topics.