Lucent Technologies
Bell Labs Innovations

Lucent Technologies Bell Laboratories

subject: **Optimizing MIMO Antenna Systems With Channel Covariance Feedback**

date: **May 1, 2002**

from: **Steven H. Simon**
Org. BL0111110
MH 1D-334
(908) 582-6006
shsimon@lucent.com

Aris L. Moustakas
Org. JW9130000
WH 1A-210
(973) 386-4432
arislm@lucent.com

Technical Memorandum

In this paper we consider a narrowband point-to-point communication system with $n_{\rm T}$ transmitters and $n_{\rm R}$ 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 $n_{\rm T} = 2$ 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.

1 Introduction

The idea of exploiting knowledge of channel covariance in MIMO (Multi-Input-Multi-Output) systems was proposed by Moustakas *et al.*[1] and Sengupta and Mitra[2] , and also by Visotsky and Madhow[3] , Jafar and Goldsmith[4] , and Narula *et al.*[5] for the MISO (many-input-single-output) 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 (TIMO, two-input-many-output) — where exact results can again be obtained analytically. It will turn out,

Lucent Technologies – PROPRIETARY
Use pursuant to Company Instructions

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) so as 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 work 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% outage" which is defined to be the minimum mutual information that occurs in all but x% of the instantiations of the channel. (In other words, if we measure the mutual information of the channel many times — in many instantiations of the random channel — we would find that a mutual information greater than the 5% outage would occur 95% of the time). Typically, system design aims to optimize either the ergodic average or some given outage. In this paper, we will aim to determine the input signal covariance at the transmitter that optimizes one of these quantities. In order to perform such an optimization, one needs to be able to calculate the ergodic or outage mutual information as a function of the input signal covariance. This can be done, for example, by integrating over a Gaussian channel with known, fixed nontrivial correlations. In the past, this has been done typically by Monte-Carlo methods (See for example, Refs. [3, 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 CLTD (closed loop transmit diversity[11]) where only a few bits describing the channel are fed back to the transmitter.

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 $n_{\rm T}n_{\rm R}$ complex random variables (with $n_{\rm T}$ = number of transmit antennas and $n_{\rm R}$ = 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

Lucent Technologies – PROPRIETARY
Use pursuant to Company Instructions

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 2 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 essentially 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 and arbitrary number of receivers.

1.1 Outline

After some general definitions (section 1.2 and 1.3) applicable for general MIMO systems, we present a number of analytic results in section 2.1 describing the probability distribution function, cumulative distribution function, and ergodic average of the mutual information for a given transmission covariance in TIMO systems. In section 2.2 we discuss how to optimize over transmission covariance so as to maximize capacity. In section 2.3 we then present simple analytic conditions for establishing whether beamforming is optimal in MIMO systems with an *arbitrary number* of transmit antennas. Finally in section 3 we return to the case of TIMO systems and present a number of results to demonstrate applications of these methods.

1.2 Definition of Channel

In the narrowband MIMO problem - n_{T} transmitters and n_{R} receivers - the channel is defined by

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

where $\bar{y}_\alpha$ is the n_{R} dimensional complex vector of received signals, $\bar{x}_i$ is the n_{T} dimensional complex vector of transmitted signal, $\bar{G}_{i\alpha}$ is the complex channel from the i^{th} transmitter to the α^{th} receiver, and $\bar{\eta}_\alpha$ is the complex noise at the α^{th} receiver. We will typically use the notation that roman indices $i, j, \ldots$ represent the transmitter antennas and take values $1 \ldots n_{\mathrm{T}}$ whereas Greek indices $\alpha, \beta, \ldots$ represent the receiver end and go from $1 \ldots n_{\mathrm{R}}$. The overbar notation here indicates that we are not yet working with rescaled quantities (which we will fix shortly). In this paper, boldface quantities denote matrices or vectors (hopefully usage will make clear which one is which) so we can also write $\bar{\mathbf{y}} = \bar{\mathbf{G}}^\dagger \bar{\mathbf{x}} + \bar{\boldsymbol{\eta}}$ We will assume that the noise is a gaussian random vector, $E\{\bar{\eta}^*_\alpha \bar{\eta}_\beta\} = \nu \delta_{\alpha\beta}$ where $E\{\cdot\}$ denotes the expectation value and ν is the noise power per receiver and δ is the Kronecker delta function. Here, 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.

The transmitted signal covariance matrix is given by $E\{\bar{x}^*_i \bar{x}_j\} = (P/n_{\mathrm{T}}) Q_{ij}$ where P is the total transmitted power and Q_{ij} is a non-negative definite hermitian matrix, with eigenvalues q_i, for $i = 1 \ldots n_{\mathrm{T}}$.

The constraint of fixed total transmitted power is expressed in terms of $\mathbf{Q}$ as $\mathrm{Tr}\{\mathbf{Q}\} = \sum_{i=1}^{n_\mathrm{T}} q_i = n_\mathrm{T}$ We note that the expectation of the total power $\mathcal{P}_\alpha$ that arrives at receiver α is given by $\mathcal{P}_\alpha = E\{y_\alpha^* y_\alpha\} = (P/n_\mathrm{T})[\bar{\mathbf{G}}^\dagger \mathbf{Q} \bar{\mathbf{G}}]_{\alpha\alpha} + \nu$.

In this paper we assume that the channel matrix $\bar{\mathbf{G}}$ is $\mathcal{N}(0, \bar{\mathbf{\Sigma}} \otimes \mathbf{1})$ i.e. a complex Gaussian random n_T by n_R dimensional matrix with zero mean and covariance matrix $\bar{\mathbf{\Sigma}} \otimes \mathbf{1}$. (Here $\bar{\mathbf{\Sigma}}$ is an n_T dimensional symmetric 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 $\langle G_{i\alpha}^* G_{j\beta} \rangle = \bar{\Sigma}_{ij} \delta_{\alpha\beta}$ where the brackets $\langle \cdot \rangle$ mean an ensemble average over instantiations of $\mathbf{G}$. 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, 13]). Again, 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).

The average power $\langle \mathcal{P}_\alpha \rangle$ arriving at any given receiver α given by $\langle \mathcal{P}_\alpha \rangle = \langle \mathcal{P}^s \rangle + \nu$ where $\langle \mathcal{P}^s \rangle = (P/n_\mathrm{T})\mathrm{Tr}[\mathbf{Q}\bar{\mathbf{\Sigma}}]$ is the average signal power received by each receiver, and ν is the noise power per receiver.

We now rescale our variables so as to make some of the normalizations simpler. We introduce the constant $\bar{s} = (1/n_\mathrm{T})\mathrm{Tr}\bar{\mathbf{\Sigma}}$. We rewrite our channel equation (Eq. 1) in an equivalent form $\mathbf{y} = \mathbf{G}^\dagger \mathbf{x} + \boldsymbol{\eta}$ where we have defined $\mathbf{y} = \bar{\mathbf{y}}/\sqrt{\nu}$, $\mathbf{x} = \bar{\mathbf{x}}\sqrt{\bar{s}/\nu}$, and $\boldsymbol{\eta} = \bar{\boldsymbol{\eta}}/\sqrt{\nu}$, and the new channel is defined in terms of the old channel as $\mathbf{G} = \bar{\mathbf{G}}/\sqrt{\bar{s}}$ such that the new channel $\mathbf{G}$ is $\mathcal{N}(0, \mathbf{\Sigma} \otimes \mathbf{1})$ (i.e. has zero mean and covariance $\mathbf{\Sigma} \otimes \mathbf{1}$) where $\mathbf{\Sigma} = \bar{\mathbf{\Sigma}}/\bar{s}$ and we now have conveniently $\mathrm{Tr}\{\mathbf{\Sigma}\} = n_\mathrm{T}$. We will frequently use the notation that s_i are the eigenvalues of $\mathbf{\Sigma}$. In this new rescaled form of the channel equation, we also have $E\{\eta_\alpha^* \eta_\beta\} = \delta_{\alpha\beta}$ and $E\{x_i^* x_j\} = pQ_{ij}$ where $p = \frac{P\bar{s}}{n_\mathrm{T}\nu}$. The physical meaning of the parameter p is roughly that of a signal to noise ratio per transmitting antenna. More precisely, if we consider transmission covariance with $\mathbf{Q} = \mathbf{1}$ (uncorrelated transmissions) then $p = \langle \mathcal{P}^s \rangle / (n_\mathrm{T}\nu)$. 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 $(p\mathbf{Q}\mathbf{\Sigma})^{-1}$ which are necessarily nonnegative since both $\mathbf{Q}$ and $\mathbf{\Sigma}$ are nonnegative definite.

1.3 Definitions of Quantities to Calculate

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

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

where $\mathbf{1}$ is an n_R dimensional unit matrix. Throughout this paper we will measure information in $nats/sec/Hz$ 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 $\mathrm{PDF}(I)$ of the mutual information I in Eq. 2 over the ensemble of instantiations $\mathbf{G}$ (with $\mathbf{Q}$ fixed) can then be written as

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

Lucent Technologies – PROPRIETARY
Use pursuant to Company Instructions

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 PDF is implicitly a function of $\mathbf{Q}$, p and $\mathbf{\Sigma}$, 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

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

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

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

when

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

Since $\mathrm{PDF}(I) > 0$, the inversion is unique. The meaning of $I_{out} = \mathcal{OUT}(P_{out})$ is that there is a probability P_{out} that in any instantiation of $\mathbf{G}$ from the ensemble, we will obtain a mutual information I less than I_{out}. This is the usual definition of outage (as we have already described it above). We note that the 50% outage is the median of the distribution.

2 Analytic Results

2.1 TIMO Analytic Results

We begin by stating a number of results for the case where there are 2 transmitters and an arbitrary number of receivers ($n_{\mathrm{T}} = 2$, i.e., TIMO systems).

2.1.1 PDF

In appendix A the probability distribution function is found to be

$$\mathrm{PDF}(I) = e^I(-ia_1a_2)^{n_{\mathrm{R}}} \int \frac{dk}{2\pi} \int_1^\infty dx \left(\frac{x-1}{x}\right)^{n_{\mathrm{R}}-1} \frac{\exp\left[ik\left(e^I - x\right) - (x-1)a_2)\right]}{(n_{\mathrm{R}}-1)![k - ia_1/x]^{n_{\mathrm{R}}-1}[k-ia_1]} \tag{8}$$

where a_i for $i = 1, 2$, are the eigenvalues of $(p\mathbf{Q}\ \mathbf{\Sigma})^{-1}$ and n_{R} is the number of receiving antennas. The k contour integral can be integrated analytically to yield

$$\mathrm{PDF}(I) = \frac{-e^I a_1(-a_2)^{n_{\mathrm{R}}}}{(n_{\mathrm{R}}-1)!} \int_1^{e^I} dx e^{-(x-1)a_2} \left[e^{-a_1(e^I-x)} - e^{-a_1(e^I-x)/x} \sum_{j=0}^{n_{\mathrm{R}}-2} \frac{1}{j!} \left((1/x-1)(e^I-x)a_1\right)^j \right] \tag{9}$$

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. Note, that we could perform the integral of the first term in brackets analytically, although we write in this form for simplicity of notation.

2.1.2 Mean (Ergodic) Mutual Information

The mean mutual information $\langle I \rangle$ for fixed $\mathbf{Q}$ can be derived by integrating the PDF (as shown in Eq. 4). Again, the calculation is complicated, but is outlined in in appendix A yielding

$$\langle I \rangle = \int_0^\infty dx \frac{e^{-x}}{x} \left[1 - \frac{1}{(n_{\rm R}-1)!} \left(\frac{a_1 a_2}{x} \right)^{n_{\rm R}} F_{n_{\rm R}} \left(\frac{(a_1+x)(a_2+x)}{x} \right) \right] \tag{10}$$

where

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

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

The integral defining F_n can be done analytically yielding

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

with Γ the incomplete gamma function and Ei the exponential integral. Although this form looks messy, for small n, the answer is quite simple. For example, for $n = 2$, we have

$$yF_n[y] = 1 + y e^y \mathrm{Ei}[-y] \tag{13}$$

For $n = 3$, we have

$$yF_n[y] = (1+y) + (2y+y^2) e^y \mathrm{Ei}[-y] \tag{14}$$

and so forth.

2.1.3 CDF

Analogous expressions for the cumulative distribution function (CDF) can be derived by directly integrating the PDF in the form of Eq. 8 to yield

$$\mathrm{CDF}(I) = (-i a_1 a_2)^{n_{\rm R}} \int \frac{dk}{2\pi} \int_1^\infty dx \left(\frac{x-1}{x} \right)^{n_{\rm R}-1} \frac{\exp\left[ik \left(e^I - x \right) - (x-1) a_2 \right]}{(n_{\rm R}-1)! (k - i0^+)[k - i a_1/x]^{n_{\rm R}-1} [k - i a_1]} \tag{15}$$

Again this contour integral can also be done explicitly, resulting in the expression

$$\begin{aligned} \mathrm{CDF}(I) &= \frac{a_2^{n_{\rm R}}}{(n_{\rm R}-1)!} \int_1^{e^I} dx e^{-(x-1)a_2} \Big[(x-1)^{n_{\rm R}-1} + (-1)^{n_{\rm R}} e^{-a_1(e^I - x)} \qquad (16) \\ &+ (-1)^{(n_{\rm R}+1)} e^{-a_1(e^I-x)/x} \sum_{j=0}^{n_{\rm R}-2} \frac{1}{j!} \left[a_1 (e^I - x)(1/x - 1) \right]^j \left[1 - (1-x)^{n_{\rm R}-1-j} \right] \Big] \end{aligned}$$

Again, a symmetry between a_1 and a_2 exists although it is not apparent in this form.

If one of the eigenvalues of $\mathbf{\Sigma Q}$ is zero, this corresponds to a beamforming solution where one is using only one effective antenna. Mathematically, this corresponds to the divergence of a particular a_k. We note in passing that one can take this limit analytically to recover the results (PDF, CDF, and ergodic average) for 2 to 1 transmission discussed previously in Ref. [6] . Finally, we note that the above stated analytic results have all been verified with Monte-Carlo simulations.

2.2 Optimization over 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 [13] for the ergodic capacity case. We give a similar proof in appendix B generalized to to apply to the case of outage capacity as well. We note that in the TIMO case of two transmitters, optimization over the two dimensional 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).

2.3 Optimality of Beamforming

We define beamforming to be the transmission of all of the power through the maximum eigenvalue and corresponding eigenvector of $\mathbf{\Sigma}$. When $\mathbf{\Sigma}$ has more than one non-zero 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 we can work in a basis with both of these matrices diagonal. Without loss of generality, we use a basis such that $s_1 \geq s_2 \geq s_3 \ldots \geq s_{n_\mathrm{T}}$. The beamforming mode is such that $q_1 = n_\mathrm{T}$ 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 so as to preserve the constraint $\sum_i q_i = n_\mathrm{T}$. In fact, it is easy to show that since $s_2 > s_3 > \ldots$, 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

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

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

The above condition (Eq. 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. Indeed, in cases where only q_1 and q_2 are nonzero, in calculating the mutual information (Eq. 2), only a 2 by $n_{\rm R}$ submatrix of the channel matrix $\mathbf{G}$ can contribute. Thus, we can obtain an analytic condition by appropriately differentiating our analytic expressions for TIMO ergodic and outage capacities.

2.3.1 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 $q_1 + q_2 = n_{\rm T}$ and $q_j = 0$ for $j > 2$ (as shown by Jafar in [4]) then Eq. 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 (Eq. 17) of the expression for ergodic average (Eq. 10) to yield (with a bit of algebra) the beamforming optimality condition

$$\Gamma\left(-n_{\rm R}+1, \frac{1}{\rho s_1}\right) \leq \frac{(\rho s_1)^{n_{\rm R}}[1-(n_{\rm R}-1)\rho s_2]}{1+\rho s_2}\exp\left(\frac{-1}{\rho s_1}\right) \tag{18}$$

with Γ the incomplete gamma function and $\rho = n_{\rm T}p$ the signal to noise ratio. When Eq. 18 is not satisfied, beamforming is no longer optimal. For the case of $n_{\rm R} = 1$ our result agrees with the result previous derived by Jafar[4] and Moustakas[6] in studies of MISO system. We note that in Ref. [9] two separate conditions were derived to bound the beamforming optimality condition. Here, however, we have derived a single result which is both necessary and sufficient and 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 using different methods[1, 12, 2] a simpler approximate form of Eq. 18 can be derived

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

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

In figure 1 we show the regions of parameters ρs_1 and ρs_2 for which beamforming optimizes the ergodic capacity for $n_{\rm R} = 1,..,5$ receivers where $\rho = pn_{\rm T}$ 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 Eq. 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 Eq. 19 which are extremely accurate for $n_{\rm R} > 1$ and for larger $n_{\rm R}$ become almost indistinguishable from the exact results.

2.3.2 Outage Capacity Beamforming Optimality

We now turn to the case of outage capacity. We can again use Eq. 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

$$\rho s_2\left[\frac{n_{\rm R}}{e^{I^b_{out}}-1} + n_{\rm R} - 1\right] < 1 \tag{20}$$

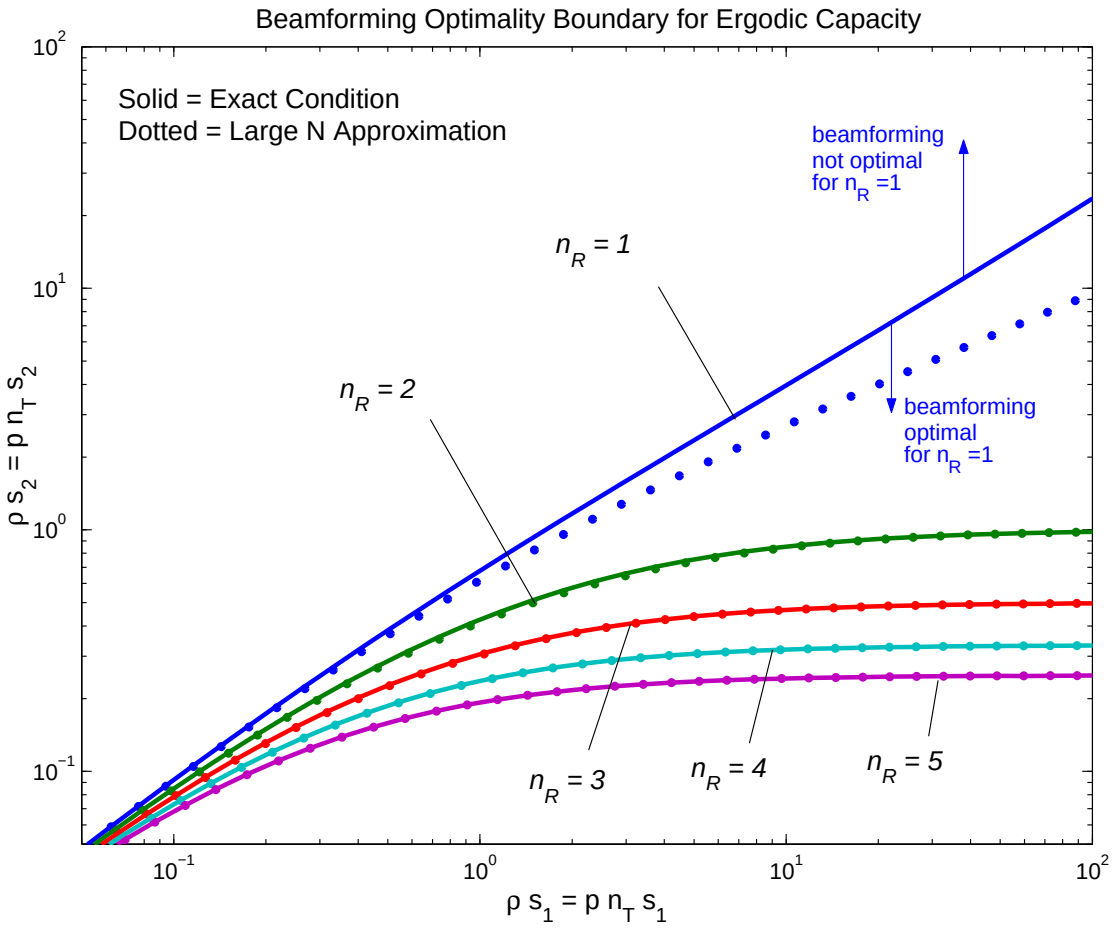


Figure 1: Beamforming Optimality Condition For Maximizing the Ergodic Capacity. The two axes represent ρs_1 and ρs_2 where $\rho = n_{\mathrm{T}} p$ is the signal to noise, and s_1 and s_2 are the two largest eigenvalues of the transmitter antenna array correlation matrix $\mathbf{\Sigma}$. Each curve separates the region of SNR below (and to the right of) which beamforming is optimal for n_{R} receive antennas ($n_{\mathrm{R}} = 1 \ldots 5$). Note that the curves are a both necessary and sufficient condition for beamforming optimality. The solid curves are exact results (Eq. 18) whereas the dotted curves are a simple approximation (Eq. 19).

Lucent Technologies – PROPRIETARY
Use pursuant to Company Instructions

where again $\rho = n_{\rm T} p$ is the signal to noise parameter, and I^b_{out} is the outage capacity for beamforming. Explicitly, we can write $P_{out} = C(I^b_{out})$ where C is the CDF of the beamforming solution[6]

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

The condition Eq. 20 turns out to be necessary but not sufficient in general. As shown in Ref. [6] the outage capacity can sometimes be a non-convex 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" non-convex. 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 Eq. 20 needs to be checked explicitly in every case numerically.

In figure 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 Eq. 20 is satisfied for $n_{\rm R} = 1, .., 5$ receivers where $\rho = p n_{\rm T}$ 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 $n_{\rm R} > 1$ 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.

3 TIMO Results

In this section we focus on the case of $n_{\rm T} = 2$ transmitters (TIMO).

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 $n_{\rm R} = 3$ receive antennas ($n_{\rm T} = 2$) 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 PDF's are roughly symmetric. 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. 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 (at least in this particular case) switching from beamforming to independent transmission at an appropriate antenna cross correlation can come reasonably close to (within 5% of) the full capacity. Note

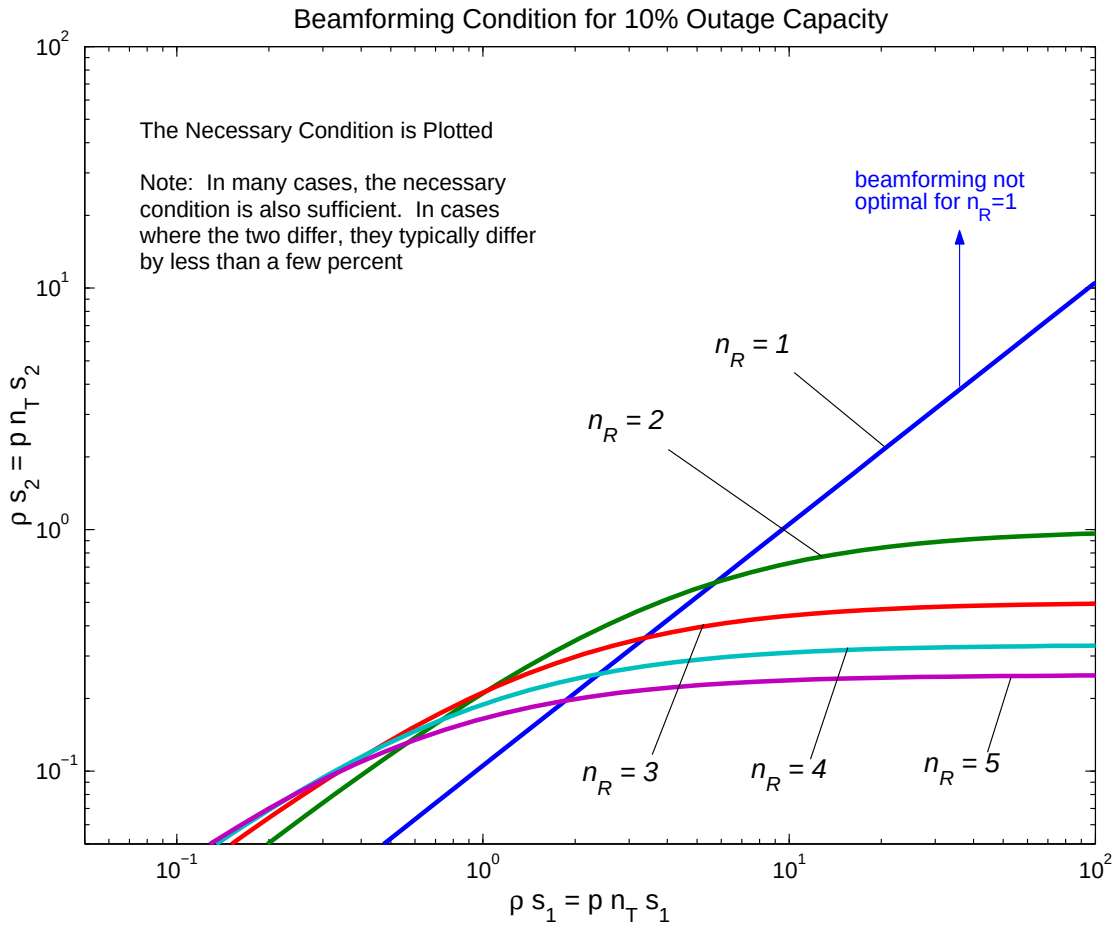


Figure 2: Beamforming Necessary Condition For 10% Outage Capacity. The axes represent ρs_1 and ρs_2 where $\rho = pn_{\mathrm{T}}$ is the signal to noise parameter and s_1,s_2 are the two largest eigenvalues of the transmit antenna array covariance matrix $\mathbf{\Sigma}$. Each curve separates the region of SNR below (and to the right of) which the beamforming condition (Eq. 20) is satisfied for n_{R} receive antennas ($n_{\mathrm{R}} = 1 \ldots 5$). 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.

Lucent Technologies – PROPRIETARY
Use pursuant to Company Instructions

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.

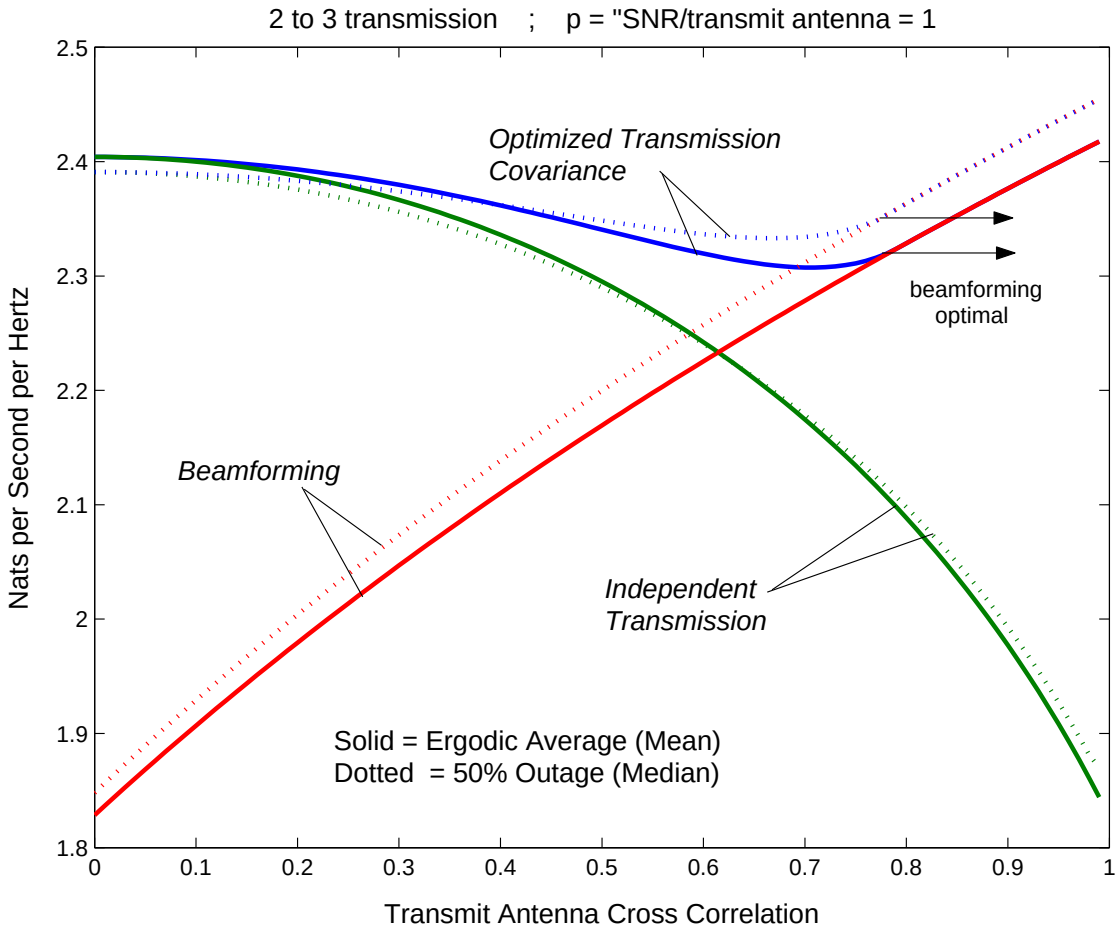


Figure 3: Comparison of Mutual Information for Different Modes of Transmission. Here we are considering the case of two transmitters and three receivers with a signal to noise parameter "SNR/antenna"=p=1. The abscissa is the antenna cross-correlation $x = (s_1 - s_2)/2$. Shown are curves for beamforming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

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 non-beamforming 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 ($n_{\rm R} > 2$) is quite similar to increasing the overall signal to noise as one might expect. Note in particular the similarities between the curves in Fig. 5 with $n_{\rm R} = 2$ receive antennas with signal to noise parameter $p = 1, 2, 3, 4$ to the curves in Fig. 4 for $p = 1$ with $n_{\rm R} = 2, 3, 4, 5$ receive antennas. In both these figures, it is clear that high antenna cross correlation $x = (s_1 - s_2)/2$ and low effective signal to noise ratio (low p or $n_{\rm R}$) favors beamforming. These trends are similar to those previously obtained in the MISO case[6] .

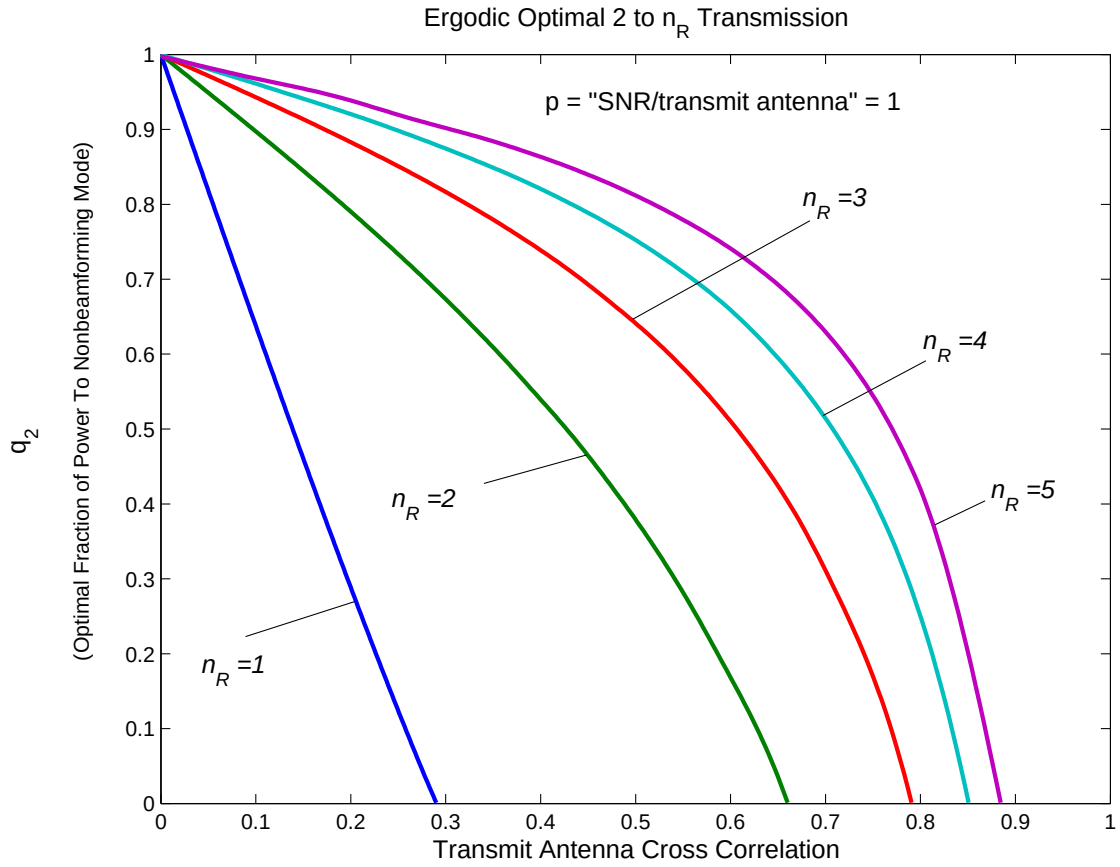


Figure 4: Optimal power to nonbeamforming mode q_2 with 2 transmit antennas and $n_{\rm R}$ receive antennas, as a function of the antenna cross-correlation $x = (s_1 - s_2)/2$ of the transmit antennas. Here we optimize q_2 so as to maximize the ergodic average. In this figure we fix the signal to noise 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.

Lucent Technologies – PROPRIETARY
Use pursuant to Company Instructions

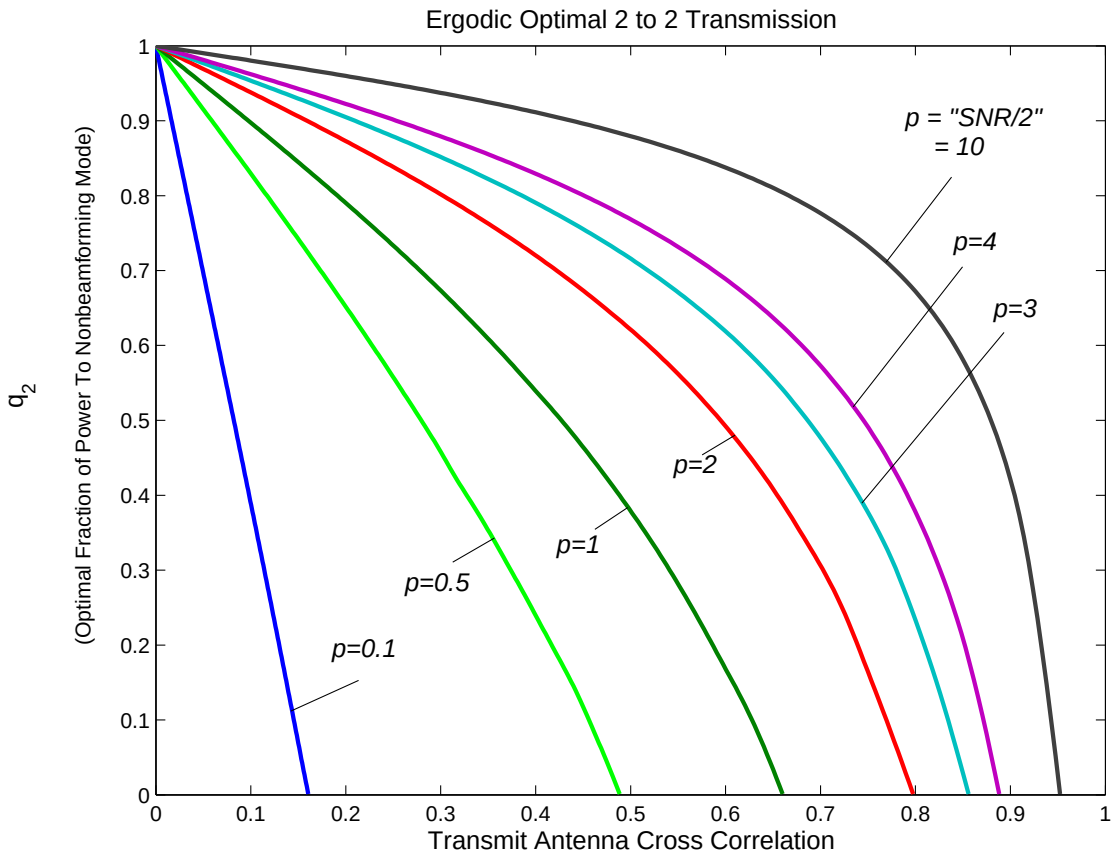


Figure 5: Optimal power to nonbeamforming mode q_2 with 2 transmit antennas and 2 receive antennas, as a function of the antenna cross-correlation $x = (s_1 - s_2)/2$ of the transmit antennas. Here, the optimization of q_2 is to maximize the ergodic capacity. In this figure we fix the various curves correspond to different signal to noise parameters "SNR/transmit antenna" $= p = 0.1, 0.5, 1, 2, 3, 4, 10$. 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 much higher antenna cross correlation.

Lucent Technologies – PROPRIETARY
Use pursuant to Company Instructions

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 above only 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 Fig. 6 and 7, the beamforming condition Eq. 20 above does indeed precisely define the optimality boundary (I.e., in these cases Eq. 20 is both necessary and sufficient).

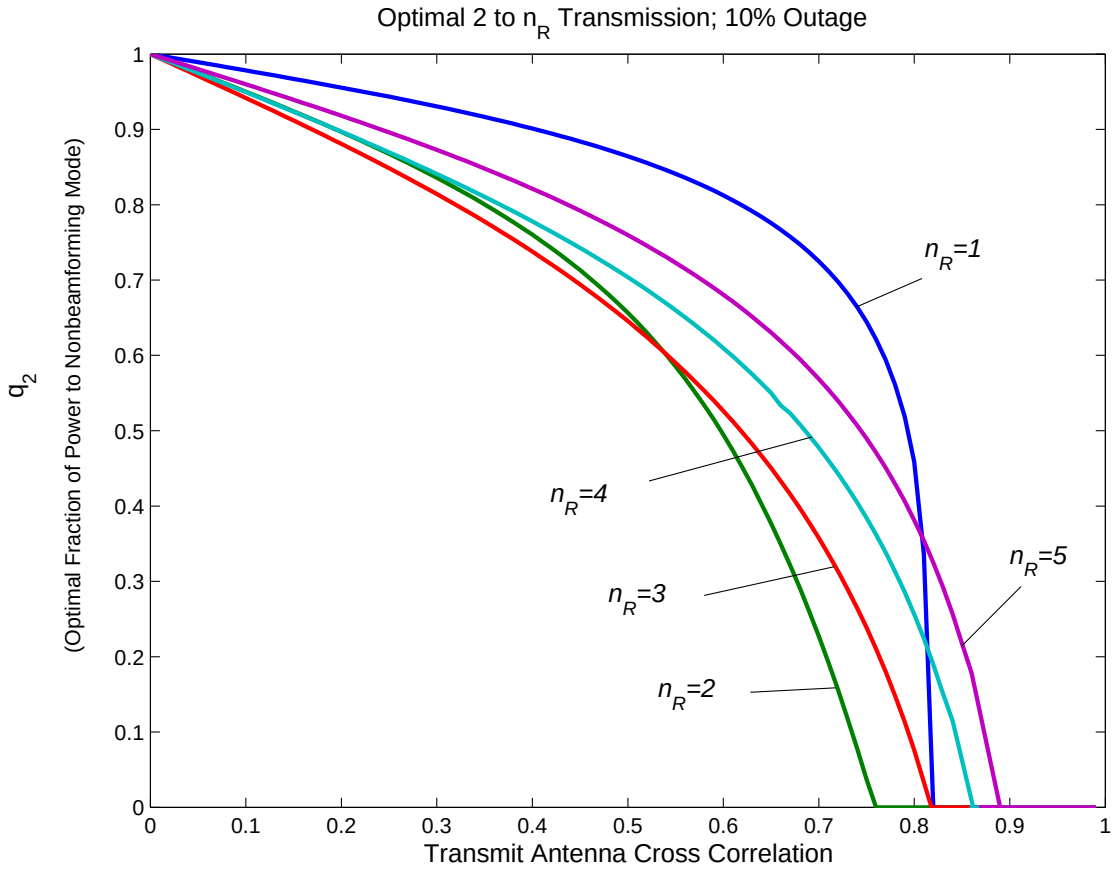


Figure 6: Optimal power to nonbeamforming mode q_2 with 2 transmit antennas and n receive antennas, as a function of the antenna cross-correlation $x = (s_1 - s_2)/2$ of the transmit antennas. Here, the optimization is to maximize the 10% outage capacity. In this figure we fix the signal to noise parameter "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. Because the width of the PDF is a strong function of the number of receive antennas, the curves can cross.

Lucent Technologies – PROPRIETARY
Use pursuant to Company Instructions

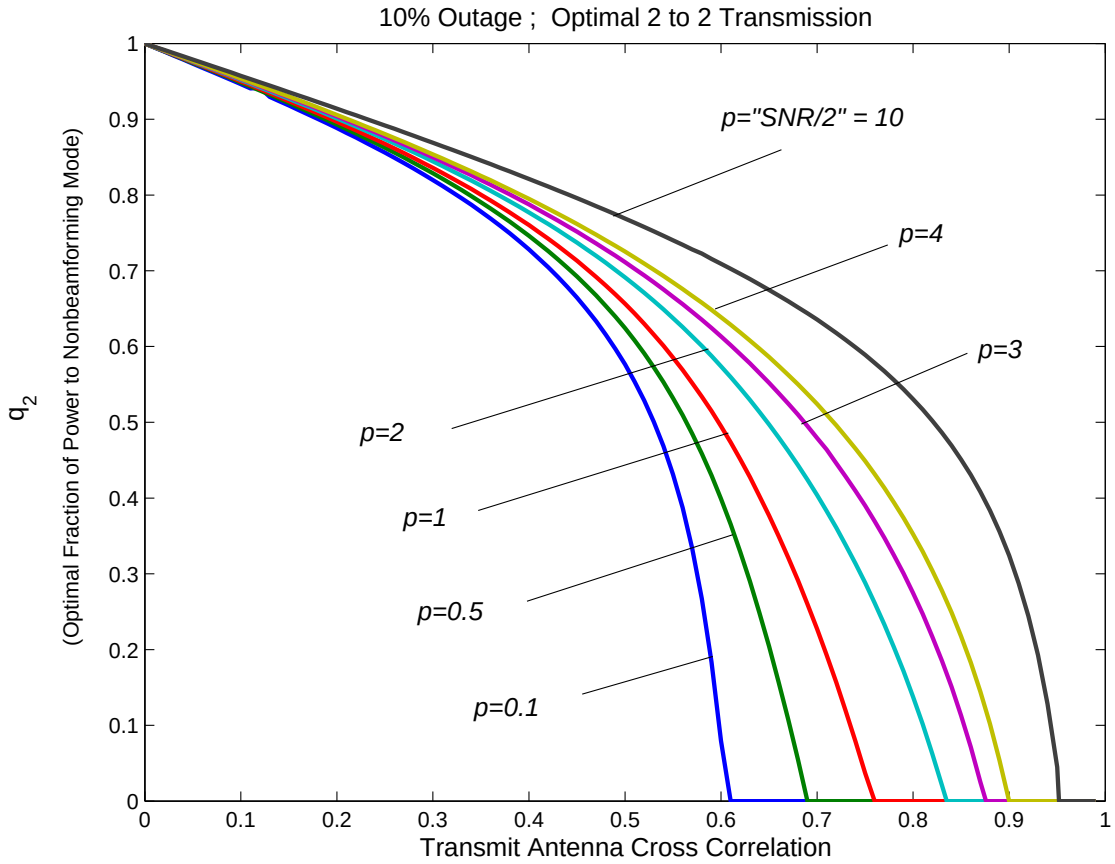


Figure 7: Optimal power to nonbeamforming mode q_2 with 2 transmit antennas and 2 receive antennas, as a function of the antenna cross-correlation $x = (s_1 - s_2)/2$ of the transmit antennas. Here, the optimization is to maximize the 10% outage capacity. In this figure we fix the various curves correspond to different signal to noise parameters "SNR/transmit antenna"$= p = .1, .5, 1, 2, 3, 4, 10$. 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.

Lucent Technologies – PROPRIETARY
Use pursuant to Company Instructions

4 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 $n_{\rm R}$ 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 $\mathcal{N}(0, \mathbf{\Sigma} \otimes \mathbf{1})$, 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.

A Details of Calculations

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

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

with $\mathbf{\Sigma}$ representing the correlations at the tranmitting antennas, and $\mathbf{\Xi}$ representing the correlations at the receiving antennas. The expectation of an arbitrary operator O can be written explicitly as

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

where the integration measure is defined by

$$\int d\mu(\mathbf{G}) \;= \det[\mathbf{\Sigma}^{-1} \otimes \mathbf{\Xi}^{-1}] \prod_{i=1}^{2} \prod_{\alpha=1}^{n} \frac{1}{\pi} \int d\mathrm{Re} G_{i\alpha} \int d\mathrm{Im} G_{i\alpha} \;\; e^{-\mathrm{Tr}[\mathbf{\Xi^{-1}} \mathbf{G}^\dagger \mathbf{\Sigma^{-1}} \mathbf{G}]}. \tag{24}$$

We write the mutual information as $I = \log\det(\mathbf{1} + \mathbf{QGPG}^\dagger)$ where $\mathbf{Q}$ is a fixed 2 by 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 thus have

$$\begin{aligned} \mathrm{PDF}(I) &= \langle \delta(I - \log\det(\mathbf{1} + \mathbf{QGPG}^\dagger)) \rangle & (25)\\ &= \int d\mu(\mathbf{G})\; \delta(I - \log\det(\mathbf{1} + \mathbf{QGPG}^\dagger)) & (26)\\ &= e^I \int d\mu(\mathbf{G})\; \delta(e^I - \det(\mathbf{1} + \mathbf{QG}^\dagger\mathbf{PG})). & (27) \end{aligned}$$

It is convenient to rewrite the determinant as

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

Lucent Technologies – PROPRIETARY
Use pursuant to Company Instructions

The subscripts here indicate an element of the 2 by 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 the full expression in a convenient form by inserting delta functions that are immediately integrated out

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

Fourier transforming all of the delta functions yields

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

We can now integrate out $\mathbf{G}$ yielding

$$\begin{aligned}\mathrm{PDF}(I) &= e^I \det[\mathbf{\Sigma^{-1}}\otimes\mathbf{\Xi^{-1}}]\int\frac{dk}{2\pi}\int dx_1dx_2dx_3\int\frac{dq_1}{2\pi}\frac{dq_2}{2\pi}\frac{dq_3}{2\pi} \\ &\quad \frac{\exp\left[ik\left(e^I - x_1 + \frac{1}{4}(x_2^2+x_3^2)\right) + iq_1(x_1-1) + iq_2x_2 + iq_3x_3\right]}{\det[\mathbf{\Sigma^{-1}}\otimes\mathbf{\Xi^{-1}} + \mathbf{MQ}\otimes\mathbf{P}]}\end{aligned} \tag{31}$$

where $\mathbf{M}$ is the matrix

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

It is trivial to integrate out x_2 and x_3 (and moving some factors of $\mathbf{Q}$ and $\mathbf{P}$ around between the determinants) yielding

$$\begin{aligned}\mathrm{PDF}(I) &= 4\pi e^I \det[\mathbf{\Sigma^{-1}Q^{-1}}\otimes\mathbf{\Xi^{-1}P^{-1}}]\int\frac{dk}{2\pi}\int dx_1\int\frac{dq_1}{2\pi}\frac{dq_2}{2\pi}\frac{dq_3}{2\pi} \\ &\quad \frac{\exp\left[ik\left(e^I - x_1\right) + iq_1(x_1-1) - \frac{i}{k}(q_2^2+q_3^2)\right]}{k\det[\mathbf{\Sigma^{-1}Q^{-1}}\otimes\mathbf{\Xi^{-1}P^{-1}} + \mathbf{M}\otimes\mathbf{1}]}\end{aligned} \tag{33}$$

Let us define the eigenvalues of $\mathbf{\Xi^{-1}P^{-1}}$ to be ϵ_m which all must be nonnegative. Let us also work in a basis where $\mathbf{\Sigma^{-1}Q^{-1}}$ is diagonal with eigenvalues a_1 and a_2. We then can rewrite the determinants as

$$\det[\mathbf{\Sigma^{-1}Q^{-1}}\otimes\mathbf{\Xi^{-1}P^{-1}} + \mathbf{M}\otimes\mathbf{1}] = \prod_{m=1}^{n}\left[(\epsilon_m a_1 + ikx_1)(\epsilon_m a_2 + iq_1) + (q_2^2+q_3^2)\right] \tag{34}$$

$$\det[\mathbf{\Sigma^{-1}Q^{-1}}\otimes\mathbf{\Xi^{-1}P^{-1}}] = (a_1a_2)^n\prod_{m=1}^{n}\epsilon_m^2 \tag{35}$$

We then notice that q_2 and q_3 only ever occur in the combination $q_2^2+q_3^3$. 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)

$$\mathrm{PDF}(I) = e^I\left[(a_1a_2)^n\prod_{m=1}^{n}\epsilon_m^2\right]\int\frac{dk}{2\pi}\int dx\int\frac{dq}{2\pi}\int_0^\infty dy\frac{\exp\left[ik\left(e^I-x\right) + iq(x-1) - \frac{iy}{k}\right]}{k\prod_m\left[(\epsilon_m a_1 + ikx)(\epsilon_m a_2 + iq) + y\right]} \tag{36}$$

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)

$$\begin{aligned} \mathrm{PDF}(I) &= e^I \left[(a_1 a_2)^n \prod_{m=1}^{n} \epsilon_m^2 \right] \int \frac{dk}{2\pi} \int_1^\infty dx \int_0^\infty dy \frac{\exp\left[ik\left(e^I - x\right) - \frac{iy}{k}\right]}{k \prod_m (\epsilon_m a_1 + ikx)} \\ &\times \sum_p \frac{\exp\left[(x-1)\left(-\epsilon_p a_2 - \frac{y}{\epsilon_p a_1 + ikx}\right)\right]}{\prod_{l \neq p}\left(\epsilon_m a_2 + \frac{y}{\epsilon_m a_1 + ikx} - \epsilon_p a_2 + \frac{y}{\epsilon_p a_1 + ikx}\right)} \end{aligned} \tag{37}$$

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 to 2 transmission, we obtain

$$\begin{aligned} \mathrm{PDF}(I) &= e^I (a_1 a_2 \epsilon_1 \epsilon_2)^2 \int \frac{dk}{2\pi} \int_1^\infty dx \frac{\exp\left[ik\left(e^I - x\right)\right]}{k(\epsilon_1 - \epsilon_2) a_1} \\ &\times \left[e^{-(x-1)\epsilon_1 a_2} f\left[(a_2/a_1)(\epsilon_2 a_1 + ikx)(1 - i\epsilon_1 a_1/k)\right] - (\epsilon_1 \leftrightarrow \epsilon_2) \right] \end{aligned} \tag{38}$$

where $f(x) = \exp(x)\mathrm{Ei}(-\mathrm{x})$ 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 $\mathbf{\Xi P}$ 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 n^{th} order pole yielding

$$\begin{aligned} \mathrm{PDF}(I) &= e^I \ (a_1 a_2)^n \int \frac{dk}{2\pi} \int_1^\infty dx \int_0^\infty dy \\ &\qquad \frac{(x-1)^{n-1} \exp\left[ik\left(e^I - x\right) - \frac{iy}{k} + (x-1)\left(-a_2 - \frac{y}{a_1 + ikx}\right)\right]}{(n-1)!\, k\, [a_1 + ikx]^n} \end{aligned} \tag{39}$$

The integration over y is now trivially done to yield Eq. 8.

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

$$\langle I \rangle = \frac{(a_1 a_2)^n}{(n-1)!} \int \frac{dk}{2\pi} \int_1^\infty dx \left(\frac{x-1}{x}\right)^{n-1} \frac{\mathrm{Ei}(ik)\, \exp\left[-ikx - (x-1)a_2\right]}{(k + i0^+)[k - ia_1/x]^{n-1}[k - ia_1]} \tag{40}$$

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

$$\mathrm{Ei}(ik) = e^{ik} \int_0^\infty dy \frac{e^{-y}}{ik - y} \tag{41}$$

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

B Optimal 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 $\mathrm{Tr}[\mathbf{Q}]=n_{\mathrm{T}}$).

We work in a basis where $\mathbf{\Sigma}$ is diagonal with eigenvalues $s_1 \geq s_2 \geq \ldots \geq s_{n_{\mathrm{T}}}$ 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 $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_{n_{\mathrm{T}}}$. 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 $\tilde{\mathbf{Q}} = \mathbf{Q'}(\mathrm{Tr}[\mathbf{Q}]/\mathbf{Tr}[\mathbf{Q'}])$ 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 $\mathrm{Tr}[\mathbf{\Sigma^{-1}\Lambda}] \leq \mathrm{Tr}[\mathbf{\Sigma^{-1}U^\dagger \Lambda U}]$.

Consider two sets of numbers $s_1^{-1}, \ldots s_{n_{\mathrm{T}}}^{-1}$ and $\lambda_1, \ldots \lambda_{n_{\mathrm{T}}}$. Construct a sum of products $s_k^{-1}\lambda_l^{-1}$ 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 there is exactly one nonzero element in any row or any column, we have $\mathrm{Tr}[\mathbf{\Sigma^{-1}\Lambda}] \leq \mathrm{Tr}[\mathbf{\Sigma^{-1}M\Lambda}]$ for any such permutation matrix $\mathbf{M}$. We now have

$$\mathrm{Tr}[\mathbf{\Sigma^{-1}U^\dagger \Lambda U}] = \sum_{ij} s_i^{-1}|U_{ij}|^2 \lambda_j \tag{42}$$

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

$$|U_{ij}|^2 = \sum_k \alpha_k M_{ij}^k \tag{43}$$

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

$$\mathrm{Tr}[\mathbf{\Sigma^{-1}U^\dagger \Lambda U}] = \sum_k \alpha_k \mathrm{Tr}[\mathbf{\Sigma^{-1}M^k\Lambda}] \geq \sum_k \alpha_k \mathrm{Tr}[\mathbf{\Sigma^{-1}\Lambda}] = \mathrm{Tr}[\mathbf{\Sigma^{-1}\Lambda}] \tag{44}$$

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 nonnegative

definite. Since $\mathbf{U}$ is unitary, we have $\sum_i N^0_{ij} = 1$ for any j and similarly $\sum_j N^0_{ij} = 1$ for any i. Find the smallest nonzero element of $\mathbf{N}^0$, let us call this $N^0_{i_0,j_0}$. Assume there are m elements of $\mathbf{N}^0$ that are nonzero with $n_{\mathrm{T}} \leq m \leq {n_{\mathrm{T}}}^2$ (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 $\alpha_0 = N^0_{i_0,j_0}$. We still have all of the elements of $\mathbf{N}^1$ being nonnegative (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 $\sum_i N^1_{ij} = 1 - \alpha^0$ for any j and similarly $\sum_j N^1_{ij} = 1 - \alpha^0$ 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 $N^1_{i_1,j_1}$, 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 $\alpha_1 = N^1_{i1,j1}$ and so forth until we have fully decomposed $\mathbf{N}^0$ (it is obvious that this process ends in no more than ${n_{\mathrm{T}}}^2$ steps). Thus we have proved the claim.

Lucent Technologies – PROPRIETARY
Use pursuant to Company Instructions

References

[1] A. L. Moustakas *et al.*, "Communication through a diffusive medium: Coherence and capacity," *Science*, vol. 287, pp. 287–290, January 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/0010081*, October 31 2000.

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

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

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

[6] S. H. Simon and A. L. Moustakas, "Optimizing multitransmitter single receiver antenna systems with partial channel knowledge," in *to be published*, 2002.

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

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

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

[10] A. L. Moustakas *et al*, "Wideband characteristics of the urban pcs channel," *Proceedings, International Symposium on 3G Infrastructure and Service, Athens Greece*, July 2001.

[11] "Physical layer procedures (FDD)," TS25.214 v4.3.1, 3GPP, 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," in *preprint*, 2002.

[13] S. Vishwanath S. A. Jafar and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," *ICC2001, Helsinki Finland*, 2001.

Lucent Technologies – PROPRIETARY
Use pursuant to Company Instructions