# Maximum MIMO System Mutual Information with Antenna Selection and Interference

**Rick S. Blum**

*Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA 18015-3084, USA*
*Email: rblum@eecs.lehigh.edu*

Maximum system mutual information is considered for a group of interfering users employing single user detection and antenna selection of multiple transmit and receive antennas for flat Rayleigh fading channels with independent fading coefficients for each path. In the case considered, the only feedback of channel state information to the transmitter is that required for antenna selection, but channel state information is assumed at the receiver. The focus is on extreme cases with very weak interference or very strong interference. It is shown that the optimum signaling covariance matrix is sometimes different from the standard scaled identity matrix. In fact, this is true even for cases without interference if SNR is sufficiently weak. Further, the scaled identity matrix is actually that covariance matrix that yields worst performance if the interference is sufficiently strong.

**Keywords and phrases:** MIMO, antenna selection, interference, capacity.

## 1. INTRODUCTION

Multiple-input multiple-output (MIMO) channels formed using transmit and receive antenna arrays are capable of providing very high data rates [1, 2]. Implementation of such systems can require additional hardware to implement the multiple RF chains used in a standard multiple transmit and receive antenna array MIMO system. Employing antenna selection [3, 4] is one promising approach for reducing complexity while retaining a reasonably large fraction of the high potential data rate of a MIMO approach. One antenna is selected for each available RF chain. In this case, only the best set of antennas is used, while the remaining antennas are not employed, thus reducing the number of required RF chains. For cases with only a single transmit antenna where standard diversity reception is to be employed, this approach, known as "hybrid selection/maximum ratio combining," has been shown to lead to relatively small reductions in performance, as compared with using all receive antennas, for considerable complexity reduction [3, 4]. Clearly, antenna selection can be simultaneously employed at the transmitter and at the receiver in a MIMO system leading to larger reductions in complexity.

Employing antenna selection both at the transmitter and the receiver in a MIMO system has been studied very recently [5, 6, 7]. Cases with full and limited feedback of information from the receiver to the transmitter have been considered. The cases with limited feedback are especially attractive in that they allow antenna selection at the transmitter without requiring a full description of the channel or its eigenvector

decomposition to be fed back. In particular, the only information fed back is the selected subset of transmit antennas to be employed. While cases with this limited feedback of information from the receiver to the transmitter have been studied in these papers, each assume that the transmitter sends a different (independent) equal power signal out of each selected antenna. Transmitting a different equal power signal out of each antenna is the optimum approach for the case where selection is not employed [8] but it is not optimum if antenna selection is used. The purpose of this paper is to find the optimum signaling. This problem is still unsolved to date. For simplicity, we ignore any delay or error that might actually be present in the feedback signal. We assume the feedback signal is accurate and instantly follows any changes in the environment.

Consider a system where cochannel interference is present from $L - 1$ other users. We focus on the $L$th user and assume each user employs $n_t$ transmit antennas and $n_r$ receive antennas. In this case, the vector of received complex baseband samples after matched filtering becomes

$$\mathbf{y}_L = \sqrt{\rho_L}\, \mathbf{H}_{L,L}\mathbf{x}_L + \sum_{j=1}^{L-1} \sqrt{\eta_{L,j}}\, \mathbf{H}_{L,j}\mathbf{x}_j + \mathbf{n}, \tag{1}$$

where $\mathbf{H}_{L,j}$ and $\mathbf{x}_j$ represent the normalized channel matrix and the normalized transmitted signal of user $j$, respectively. The signal-to-noise ratio (SNR) of user $L$ is $\rho_L$ and the interference-to-noise ratio (INR) for user $L$ due to interference from user $j$ is $\eta_{L,j}$. For simplicity, we assume all

of the interfering signals $\mathbf{x}_j$, $j = 1, \ldots, L - 1$, are unknown to the receiver and we model each of them as being complex Gaussian distributed, the usual form of the optimum signal in MIMO problems. Then if we condition on $\mathbf{H}_{L,1}, \ldots, \mathbf{H}_{L,L}$, the interference-plus-noise from (1), $\sum_{j=1}^{L-1} \sqrt{\eta_{L,j}} \mathbf{H}_{L,j} \mathbf{x}_j + \mathbf{n}$, is complex Gaussian distributed with the covariance matrix $\mathbf{R}_L = \sum_{j=1}^{L-1} \eta_{L,j} \mathbf{H}_{L,j} \mathbf{S}_j \mathbf{H}_{L,j}^H + \mathbf{I}_{n_r}$, where $\mathbf{S}_j$ denotes the covariance matrix of $\mathbf{x}_j$ and $\mathbf{I}_{n_r}$ is the covariance matrix of $\mathbf{n}$. Under this conditioning, the interference-plus-noise is whitened by multiplying $\mathbf{y}_L$ by $\mathbf{R}_L^{-1/2}$. After performing this multiplication, we can use results from [2, 8, 9] (see also [10, pp. 12–23, pp. 250,256]) to express the ergodic mutual information between the input and output for the user of interest as in the following:

$$
\begin{aligned}
&I(\mathbf{x}_L; (\mathbf{y}_L, \mathcal{H})) \\
&= E\left\{ \log_2 \left[ \det \left( \mathbf{I}_{n_r} + \rho_L (\mathbf{R}_L^{-1/2} \mathbf{H}_{L,L}) \mathbf{S}_L (\mathbf{R}_L^{-1/2} \mathbf{H}_{L,L})^H \right) \right] \right\} \quad (2) \\
&= E\left\{ \log_2 \left[ \det \left( \mathbf{I}_{n_r} + \rho_L \mathbf{H}_{L,L} \mathbf{S}_L \mathbf{H}_{L,L}^H \mathbf{R}_L^{-1} \right) \right] \right\}
\end{aligned}
$$

($\mathcal{H}$ reminds us of the assumed model for $\mathbf{H}_{L,1}, \ldots, \mathbf{H}_{L,L}$). In (2), the identity $\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA})$ was used. If we wish to compute total system mutual information, we should find $\mathbf{S}_1, \ldots, \mathbf{S}_L$ to maximize

$$
\begin{aligned}
&\Psi(\mathbf{S}_1, \ldots, \mathbf{S}_L) \\
&= \sum_{i=1}^{L} I(\mathbf{x}_i; (\mathbf{y}_i, \mathcal{H})) \\
&= \sum_{i=1}^{L} E\left\{ \log_2 \left[ \det \left( \mathbf{I}_{n_r} + \rho_i \mathbf{H}_{i,i} \mathbf{S}_i \mathbf{H}_{i,i}^H \right. \right. \right. \\
&\qquad\qquad\qquad \left. \left. \left. \times \left( \mathbf{I}_{n_r} + \sum_{j=1, j \neq i}^{L} \eta_{i,j} \mathbf{H}_{i,j} \mathbf{S}_j \mathbf{H}_{i,j}^H \right)^{-1} \right) \right] \right\}.
\end{aligned}
$$
$$(3)$$

Now, assume that each receiver selects $n_{sr} < n_r$ receive antennas and $n_{st} < n_t$ transmit antennas based on the channel conditions and feeds back the information to the transmitter.[1] Then the observations from the selected antennas follow the model in (1) with $n_t$ and $n_r$ replaced by $n_{st}$ and $n_{sr}$, respectively, and $\mathbf{H}_{i,j}$ replaced by $\tilde{\mathbf{H}}_{i,j}$. The matrix $\tilde{\mathbf{H}}_{i,j}$ is obtained by eliminating those columns and rows of $\mathbf{H}_{i,j}$ corresponding to unselected transmit and receive antennas. Thus we can write $\tilde{\mathbf{H}}_{i,j} = g(\mathbf{H}_{i,j})$, where the function $g$ will choose $\tilde{\mathbf{H}}_{i,j}$ to maximize the instantaneous (and thus also the ergodic) mutual information (or some related quantity for the signaling approach employed). In order to promote brevity, we will restrict attention in the rest of this paper to the case where $n_{st} = n_{sr}$ so we will only use the notation for $n_{st}$. We note that the majority of the results given carry over immediately for the case of $n_{st} \neq n_{sr}$, and since this will be obvious in these cases, we will not discuss this further.

It is important to note that we restrict attention to narrowband systems using single user detection, equal power

(constant over time) for each user, and fixed definitions of the transmitting and receiving users. Future extensions which remove some assumptions are of great interest. However, as we will show, these assumptions lead to interesting closed form results which we believe give insight into the fundamental properties of MIMO with antenna selection.

In Section 2, we give a general discussion and some useful relationships used to study the convexity and concavity properties of the system mutual information. In Section 3, we study cases with weak interference. We follow this, in Section 4, with our results for strong interference. The results in Sections 3 and 4 are general for any $n_{st} = n_{sr}, n_t, n_r$, and $L$. Section 5 is devoted to numerical studies for the particular case of $n_r = n_t = 8$, $n_{sr} = n_{st} = L = 2$ to illustrate the agreement with the theory from Sections 3 and 4. The results in Section 5 also show that our asymptotic results give useful information for nonasymptotic cases as well. The paper concludes with Section 6.

## 2. GENERAL ANALYSIS OF SYSTEM MUTUAL INFORMATION

Clearly, the nature of the functional[2] $\Psi(\mathbf{S}_1, \ldots, \mathbf{S}_L)$ will depend on the SNRs $\rho_i$, $i = 1, \ldots, L$, and the INRs $\eta_{i,j}$, $i, j = 1, \ldots, L, i \neq j$. This can be seen by considering the convexity and the concavity of $\Psi(\mathbf{S}_1, \ldots, \mathbf{S}_L)$ as a function of $\mathbf{S}_1, \ldots, \mathbf{S}_L$. Towards this goal, we define a general convex combination of two possible solutions $(\mathbf{S}_1, \ldots, \mathbf{S}_L)$ and $(\hat{\mathbf{S}}_1, \ldots, \hat{\mathbf{S}}_L)$ as follows:

$$
\begin{aligned}
(\bar{\mathbf{S}}_1, \ldots, \bar{\mathbf{S}}_L) &= (1 - t)(\mathbf{S}_1, \ldots, \mathbf{S}_L) + t(\hat{\mathbf{S}}_1, \ldots, \hat{\mathbf{S}}_L) \\
&= (\mathbf{S}_1, \ldots, \mathbf{S}_L) + t((\hat{\mathbf{S}}_1, \ldots, \hat{\mathbf{S}}_L) - (\mathbf{S}_1, \ldots, \mathbf{S}_L)) \\
&= (\mathbf{S}_1, \ldots, \mathbf{S}_L) + t(\mathbf{S}_1', \ldots, \mathbf{S}_L')
\end{aligned}
$$
$$(4)$$

for $0 \leq t \leq 1$ a scalar. Then $\Psi(\mathbf{S}_1, \ldots, \mathbf{S}_L)$ is a convex function of $(\mathbf{S}_1, \ldots, \mathbf{S}_L)$ if [12]

$$
\frac{d^2}{dt^2} \Psi(\bar{\mathbf{S}}_1, \ldots, \bar{\mathbf{S}}_L) \geq 0 \quad \forall \bar{\mathbf{S}}_1, \ldots, \bar{\mathbf{S}}_L. \quad (5)
$$

Similarly, $\Psi(\mathbf{S}_1, \ldots, \mathbf{S}_L)$ is a concave function of $(\mathbf{S}_1, \ldots, \mathbf{S}_L)$ if

$$
\frac{d^2}{dt^2} \Psi(\bar{\mathbf{S}}_1, \ldots, \bar{\mathbf{S}}_L) \leq 0 \quad \forall \bar{\mathbf{S}}_1, \ldots, \bar{\mathbf{S}}_L. \quad (6)
$$

There are several useful known relationships for the derivative of a function of a matrix $\boldsymbol{\Phi}$ with respect to a scalar parameter $t$. In particular, we note that [13, Appendix A, pp. 1342, 1345, 1349, 1351, 1359, 1401]

$$
\begin{aligned}
\frac{d}{dt} \ln[\det(\boldsymbol{\Phi})] &= \text{trace}\left[ \boldsymbol{\Phi}^{-1}\left( \frac{d}{dt} \boldsymbol{\Phi} \right) \right], \\
\frac{d}{dt} \boldsymbol{\Phi}^{-1} &= -\boldsymbol{\Phi}^{-1}\left( \frac{d}{dt} \boldsymbol{\Phi} \right) \boldsymbol{\Phi}^{-1}.
\end{aligned}
$$
$$(7)$$

---

[1] The case where each user employs a different $n_{st}$ and $n_{sr}$ is also easy to handle.

[2] In the case without antenna selection [11], it is possible to argue that each $\mathbf{S}_j$ can be taken as diagonal. These arguments are based on the joint Gaussianity of the $\mathbf{H}_{ij}$ which does not hold after selection.

Assuming selection is employed, we can use (3) and (7) to find (interchanging a derivative and an expected value)

$$\frac{d}{dt}\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L) = \frac{1}{\ln(2)}\sum_{i=1}^{L} E\left\{\text{trace}\left[\mathbf{Q}_i^{-1}\frac{d}{dt}\mathbf{Q}_i\right]\right\}, \quad (8)$$

where

$$\mathbf{Q}_i = \mathbf{I}_{n_{st}} + \rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\left(\mathbf{I}_{n_{st}} + \sum_{j=1, j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\bar{\mathbf{S}}_j\tilde{\mathbf{H}}_{i,j}^H\right)^{-1}$$
$$= \mathbf{I}_{n_{st}} + \rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\tilde{\mathbf{Q}}_i^{-1}, \quad (9)$$

$$\frac{d}{dt}\mathbf{Q}_i = \rho_i\tilde{\mathbf{H}}_{i,i}\mathbf{S}_i'\tilde{\mathbf{H}}_{i,i}^H\tilde{\mathbf{Q}}_i^{-1} - \rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\tilde{\mathbf{Q}}_i^{-1}\left(\frac{d}{dt}\tilde{\mathbf{Q}}_i\right)\tilde{\mathbf{Q}}_i^{-1}, \quad (10)$$

$$\frac{d}{dt}\tilde{\mathbf{Q}}_i = \sum_{j=1, j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\mathbf{S}_j'\tilde{\mathbf{H}}_{i,j}^H. \quad (11)$$

A second derivative yields

$$\frac{d^2}{dt^2}\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L)$$
$$= \frac{1}{\ln(2)}\sum_{i=1}^{L} E\left\{\text{trace}\left[\mathbf{Q}_i^{-1}\left(\frac{d^2}{dt^2}\mathbf{Q}_i\right) \right.\right. \quad (12)$$
$$\left.\left. - \mathbf{Q}_i^{-1}\left(\frac{d}{dt}\mathbf{Q}_i\right)\mathbf{Q}_i^{-1}\left(\frac{d}{dt}\mathbf{Q}_i\right)\right]\right\}$$

with

$$\frac{d^2}{dt^2}\mathbf{Q}_i = -2\rho_i\tilde{\mathbf{H}}_{i,i}\mathbf{S}_i'\tilde{\mathbf{H}}_{i,i}^H\tilde{\mathbf{Q}}_i^{-1}\left(\frac{d}{dt}\tilde{\mathbf{Q}}_i\right)\tilde{\mathbf{Q}}_i^{-1}$$
$$+ 2\rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\tilde{\mathbf{Q}}_i^{-1}\left(\frac{d}{dt}\tilde{\mathbf{Q}}_i\right)\tilde{\mathbf{Q}}_i^{-1}\left(\frac{d}{dt}\tilde{\mathbf{Q}}_i\right)\tilde{\mathbf{Q}}_i^{-1}. \quad (13)$$

## 3. OPTIMUM SIGNALING FOR WEAK INTERFERENCE

We can use (12) to investigate convexity and concavity for any particular set of SNRs $\rho_i$, $i = 1,\ldots,L$, and INRs $\eta_{i,j}$, $i,j = 1,\ldots,L$, $i \neq j$. We investigate extreme cases, weak or strong interference, to gain insight. The following lemma considers the case of very weak interference.

*Lemma 1. Assuming sufficiently weak interference, the best $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ (that maximizes the ergodic system mutual information) must be of the form*

$$(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L)$$
$$= \alpha(\gamma_1\mathbf{I}_{n_{st}} + (1-\gamma_1)\mathbf{O}_{n_{st}},\ldots,\gamma_L\mathbf{I}_{n_{st}} + (1-\gamma_L)\mathbf{O}_{n_{st}}), \quad (14)$$

*where $\mathbf{O}_{n_{st}}$ is an $n_{st}$ by $n_{st}$ matrix of all ones, $\alpha = 1/n_{st}$, and $0 \leq \gamma_i \leq 1$, $i = 1,\ldots,L$.*

*Outline of the proof.* For the case of very weak interference, we ignore terms which are multiples of $\eta_{i,j}$ (essentially, we

set $\eta_{i,j} \to 0$ for $i = 1,\ldots,L$, $j = 1,\ldots,L$, and $j \neq i$) and we find $(d/dt)\tilde{\mathbf{Q}}_i = 0$ so that $(d^2/dt^2)\mathbf{Q}_i = 0$ which leads to

$$\frac{d^2}{dt^2}\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L)$$
$$= -\frac{1}{\ln(2)}\sum_{i=1}^{L} E\left\{\text{trace}\left[\left(\mathbf{I}_{n_{st}} + \rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\right)^{-1}\rho_i\tilde{\mathbf{H}}_{i,i}\mathbf{S}_i'\tilde{\mathbf{H}}_{i,i}^H \right.\right.$$
$$\left.\left. \times \left(\mathbf{I}_{n_{st}} + \rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\right)^{-1}\rho_i\tilde{\mathbf{H}}_{i,i}\mathbf{S}_i'\tilde{\mathbf{H}}_{i,i}^H\right]\right\}. \quad (15)$$

Since $\bar{\mathbf{S}}_i$ is a covariance matrix, $(\mathbf{I}_{n_{st}} + \rho_i\mathbf{H}_{i,i}\bar{\mathbf{S}}_i\mathbf{H}_{i,i}^H)^{-1} = (\mathbf{U}^H\mathbf{U} + \mathbf{U}^H\mathbf{\Lambda}\mathbf{U})^{-1} = (\mathbf{U}(\mathbf{I}_{n_{st}} + \mathbf{\Lambda})^{-1}\mathbf{U}^H) = \mathbf{U}(\mathbf{\Omega})^2\mathbf{U}^H = \mathbf{U}(\mathbf{\Omega})\mathbf{U}^H\mathbf{U}(\mathbf{\Omega})\mathbf{U}^H$, where $\mathbf{U}$ is unitary and $\mathbf{\Lambda}$ and $\mathbf{\Omega}$ are diagonal matrices with nonnegative entries. Define $\mathbf{A} = \rho_i\mathbf{H}_{i,i}\mathbf{S}_i'\mathbf{H}_{i,i}^H$ and note that $\mathbf{A}^H = \mathbf{A}$ due to $\mathbf{S}_i'$ being a difference of two covariance matrices (easy to see using $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ expansion for each covariance matrix). Thus the trace in (15) can be written as $\text{trace}[\mathbf{U}(\mathbf{\Omega})^2\mathbf{U}^H\mathbf{A}\mathbf{U}(\mathbf{\Omega})^2\mathbf{U}^H\mathbf{A}] = \text{trace}[\mathbf{U}\mathbf{\Omega}\mathbf{U}^H\mathbf{A}\mathbf{U}(\mathbf{\Omega})^2\mathbf{U}^H\mathbf{A}\mathbf{U}\mathbf{\Omega}\mathbf{U}^H] = \text{trace}[\mathbf{B}\mathbf{B}^H]$ since $\text{trace}[\mathbf{C}\mathbf{D}] = \text{trace}[\mathbf{D}\mathbf{C}]$ [13]. We see $\text{trace}[\mathbf{B}\mathbf{B}^H]$ must be nonnegative since the matrix inside the trace is nonnegative-definite so that (15) implies that $\Psi(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ is concave. This will be true for sufficiently small $\eta_{i,j}$, $i,j = 1,\ldots,L$, $i \neq j$, relative to $\rho_i$, $i = 1,\ldots,L$. To recognize the significance of the concavity, we note that given any permutation matrix $\mathbf{\Pi}$, we know [8] that $\tilde{\mathbf{H}}_{i,j}$ has the same distribution as $\tilde{\mathbf{H}}_{i,j}\mathbf{\Pi}$ (switching the ordering or names of selected antennas cannot change the physical problem), so $\Psi(\mathbf{\Pi}\mathbf{S}_1\mathbf{\Pi}^H,\ldots,\mathbf{\Pi}\mathbf{S}_L\mathbf{\Pi}^H) = \Psi(\mathbf{S}_1,\ldots,\mathbf{S}_L)$. Let $\sum_{\mathbf{\Pi}}$ denote the sum over all the different permutation matrices and let $N$ denote the number of terms in the sum. From concavity, $\Psi((1/N)\sum_{\mathbf{\Pi}}\mathbf{\Pi}\mathbf{S}_1\mathbf{\Pi}^H,\ldots,(1/N)\sum_{\mathbf{\Pi}}\mathbf{\Pi}\mathbf{S}_L\mathbf{\Pi}^H) \geq \Psi(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ [8] which implies that the optimum $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ must be of the form such that it is invariant to transforms by permutation matrices. This implies that the best $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ must be of the form given in (14). We refer the interested reader to [14] for a rigorous proof of this (taken from a single user case). □

Before considering specific assumptions on the SNR, we note the similarity of (14) to (4) with $(\mathbf{S}_1,\ldots,\mathbf{S}_L) = (1/n_{st})(\mathbf{O}_{n_{st}},\ldots,\mathbf{O}_{n_{st}})$, $(\hat{\mathbf{S}}_1,\ldots,\hat{\mathbf{S}}_L) = (1/n_{st})(\mathbf{I}_{n_{st}},\ldots,\mathbf{I}_{n_{st}})$, and $t = \gamma_1 = \cdots = \gamma_L$.

### Small SNR

Thus we have determined the best signaling except for the unknown scalar parameters $\gamma_1,\ldots,\gamma_L$ which we now investigate. Generally, the best approach will change with SNR. First, consider the case of weak SNR for which the following lemma applies (recall we have now already focused on very weak or no interference).

*Lemma 2. Let $\tilde{h}(p,p)_{i,j}$ denote the $(i,j)$th entry of the matrix $\tilde{\mathbf{H}}_{p,p}$ and define $\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L$ from (14). Assuming sufficiently*

*weak interference and sufficiently weak SNR,*

$$\frac{d}{d\gamma_p}\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L)$$

$$= -\frac{1}{n_{st}\ln(2)}\rho_p E\left\{\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}\sum_{j'=1,\,j'\neq j}^{n_{st}}\tilde{h}^*(p,p)_{i,j}\tilde{h}(p,p)_{i,j'}\right\}$$

$$\textit{for } p = 1,\ldots,L. \tag{16}$$

*Outline of the proof.* Using the similarity of (14) to (4), $(d/d\gamma_p)\Psi$ can be seen to be the $p$th component of the sum in (8) with $(\mathbf{S}_1,\ldots,\mathbf{S}_L) = (1/n_{st})(\mathbf{O}_{n_{st}},\ldots,\mathbf{O}_{n_{st}})$, $(\hat{\mathbf{S}}_1,\ldots,\hat{\mathbf{S}}_L) = (1/n_{st})(\mathbf{I}_{n_{st}},\ldots,\mathbf{I}_{n_{st}})$, and $t = \gamma_p$. To assert the weak signal and interference assumptions, we set $\eta_{i,j} \to 0$ for all $i, j$ and $\rho_i \to 0$ for all $i$ and in this case we find

$$\mathbf{Q}_i^{-1}\frac{d}{dt}\mathbf{Q}_i \longrightarrow \frac{d}{dt}\mathbf{Q}_i \longrightarrow \rho_i\tilde{\mathbf{H}}_{i,i}\mathbf{S}_i'\tilde{\mathbf{H}}_{i,i}^H \tag{17}$$

and using (8) gives

$$\frac{d}{d\gamma_p}\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L)$$

$$= \frac{1}{n_{st}\ln(2)}E\left\{\mathrm{trace}\left[\tilde{\mathbf{H}}_{p,p}(\mathbf{I}_{n_{st}} - \mathbf{O}_{n_{st}})\tilde{\mathbf{H}}_{p,p}^H\right]\right\}, \tag{18}$$

where the $n_{st} \times n_{st}$ matrix can be explicitly written as

$$\mathbf{I}_{n_{st}} - \mathbf{O}_{n_{st}} = \begin{pmatrix} 0 & -1 & \cdots & -1 & -1 & -1 \\ -1 & 0 & -1 & \cdots & -1 & -1 \\ -1 & -1 & 0 & -1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & -1 & \cdots & -1 & -1 & 0 \end{pmatrix}. \tag{19}$$

Explicitly carrying out the operations in (18) gives (16). $\square$

Notice that without selection (in this case $\tilde{\mathbf{H}}_{p,p} = \mathbf{H}_{p,p}$), the quantity in (16) becomes zero under the assumed model for $\mathbf{H}_{p,q}$ (i.i.d complex Gaussian entries). Thus selection turns out to be an important aspect in the analysis. The following lemmas will be used with the result in **Lemma 2** to develop the main result of this section.

**Lemma 3.** *Let $\tilde{h}(p,p)_{i,j}$ denote the $(i,j)$th entry of the matrix $\tilde{\mathbf{H}}_{p,p}$ and define $\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L$ from (14). Assuming sufficiently weak interference and sufficiently weak SNR,*

$$\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L)$$

$$= \frac{1}{n_{st}\ln(2)}\sum_{p=1}^{L}\rho_p$$

$$\times E\left\{\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}|\tilde{h}(p,p)_{i,j}|^2\right.$$

$$\left. + (1-\gamma_p)\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}\sum_{j'=1,\,j'\neq j}^{n_{st}}\tilde{h}^*(p,p)_{i,j}\tilde{h}(p,p)_{i,j'}\right\}. \tag{20}$$

*Outline of the proof.* Consider an $n_{st} \times n_{st}$ nonnegative definite matrix $\mathbf{A}$ and let $\lambda_1(\mathbf{A}),\ldots,\lambda_{n_{st}}(\mathbf{A})$ denote the eigenvalues of $\mathbf{A}$. For sufficiently weak SNR $\rho_i$, we can approximate $\ln[\det(\mathbf{I}+\rho_i\mathbf{A})] = \ln[\prod_{j=1}^{n_{st}}(1+\rho_i\lambda_j(\mathbf{A}))] = \sum_{j=1}^{n_{st}}\ln[1+\rho_i\lambda_j(\mathbf{A})] \approx \rho_i\sum_{j=1}^{n_{st}}\lambda_j(\mathbf{A}) = \rho_i\,\mathrm{trace}(\mathbf{A})$. Now, consider $\Psi$ itself, from (3), for the set of covariance matrices in (14) and assume that selection is employed. Thus we consider the resulting $\Psi$ as a function of $(\gamma_1,\ldots,\gamma_L)$ and we see

$$\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L)$$

$$= \frac{1}{n_{st}\ln(2)}\sum_{p=1}^{L}\rho_p \tag{21}$$

$$\times E\left\{\mathrm{trace}\left[\tilde{\mathbf{H}}_{p,p}[\gamma_p\mathbf{I}_{n_{st}} + (1-\gamma_p)\mathbf{O}_{n_{st}}]\tilde{\mathbf{H}}_{p,p}^H\right]\right\}.$$

Note that the $n_{st} \times n_{st}$ matrix can be explicitly written as

$$[\gamma_p\mathbf{I}_{n_{st}} + (1-\gamma_p)\mathbf{O}_{n_{st}}]$$

$$= \begin{pmatrix} 1 & 1-\gamma_p & \cdots & 1-\gamma_p & 1-\gamma_p & 1-\gamma_p \\ 1-\gamma_p & 1 & 1-\gamma_p & \cdots & 1-\gamma_p & 1-\gamma_p \\ 1-\gamma_p & 1-\gamma_p & 1 & 1-\gamma_p & \cdots & 1-\gamma_p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1-\gamma_p & 1-\gamma_p & \cdots & 1-\gamma_p & 1-\gamma_p & 1 \end{pmatrix}. \tag{22}$$

Using (22) in (21) with further simplification gives (20). $\square$

**Lemma 4.** *Assuming sufficiently weak interference and sufficiently weak SNR, the antenna selection that maximizes the ergodic system mutual information will make*

$$E\left\{\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}\sum_{j'=1,\,j'\neq j}^{n_{st}}\tilde{h}^*(p,p)_{i,j}\tilde{h}(p,p)_{i,j'}\right\} \tag{23}$$

*positive.*

*Outline of the proof.* First, consider the antenna selection approach for the $p$th link which maximizes the ergodic system mutual information in (20) when $\gamma_p = 1$ in (14). Thus the selection approach will maximize the quantity in the $p$th term in the first sum in (20) when $\gamma_p = 1$ by selecting antennas for each set of instantaneous channel matrices to make the terms inside the expected value as large as possible. It is important to note that the choice (if $\gamma_p = 1$) depends only on the squared magnitude of elements of the channel matrices.

If we use this selection approach when $\gamma_p \neq 1$, then the terms multiplied by $(1 - \gamma_p)$ in (20) will be averaged to zero due to the symmetry in the selection criterion. To see this, first note that the contribution to the ergodic mutual information due to the $p$th term is

$$\int\cdots\int_{h(p,p)_{1,1},\ldots,h(p,p)_{n_t,n_r}}\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}\sum_{j'=1,\,j'\neq j}^{n_{st}}\tilde{h}^*(p,p)_{i,j}\tilde{h}(p,p)_{i,j'}$$

$$\times f_{h(p,p)_{1,1},\ldots,h(p,p)_{n_t,n_r}}(h(p,p)_{1,1},\ldots,h(p,p)_{n_t,n_r})$$

$$\times dh(p,p)_{1,1}\cdots dh(p,p)_{n_t,n_r} \tag{24}$$

times the constant $\rho_p/n_{st} \ln(2)$. In (24),

$$f_{h(p,p)_{1,1},\ldots,h(p,p)_{n_t,n_r}}(h(p,p)_{1,1},\ldots,h(p,p)_{n_t,n_r}) \qquad (25)$$

is the probability density function of the channel coefficients prior to selection, the integral is over all values of the arguments and the selection rule $\tilde{\mathbf{H}} = g(\mathbf{H})$ is important in determining the integrand. If the optimum selection rule for (20) with $\gamma_p = 1$ will select a particular set of transmit and receive antennas for a particular instance of $h(p,p)_{1,1},\ldots,h(p,p)_{n_t,n_r}$, then due to symmetry, this same selection will also occur several more times as we run through all the possible values of $h(p,p)_{1,1},\ldots,h(p,p)_{n_t,n_r}$. Thus assume that terms with $|h(p,p)_{ij}|^2 = a$ and $|h(p,p)_{ij'}|^2 = b$ in (20) with $\gamma_p = 1$ are large enough to cause the corresponding antennas to be selected by the selection criterion trying to maximize (20) with $\gamma_p = 1$ for some set of $h(p,p)_{11},\ldots,h(p,p)_{n_{st},n_{st}}$. Then due to the symmetry,

$$\begin{aligned}
(\tilde{h}(p,p)_{ij}^*, \tilde{h}(p,p)_{ij'}) &= (\sqrt{a}e^{j\phi_a}, \sqrt{b}e^{j\phi_b}), \\
(\tilde{h}(p,p)_{ij}^*, \tilde{h}(p,p)_{ij'}) &= (\sqrt{a}e^{j\phi_a}, -\sqrt{b}e^{j\phi_b}), \\
(\tilde{h}(p,p)_{ij}^*, \tilde{h}(p,p)_{ij'}) &= (-\sqrt{a}e^{j\phi_a}, \sqrt{b}e^{j\phi_b}), \\
(h(p,p)_{ij}^*, h(p,p)_{ij'}) &= (-\sqrt{a}e^{j\phi_a}, -\sqrt{b}e^{j\phi_b})
\end{aligned} \qquad (26)$$

will all appear in (24). Since each of these four possible values appear for four equal area (actually probability) regions in channel coefficient space, a complete cancellation of these terms results in (24). In fact, this leads to (24) averaging to zero. Thus if we use the selection approach that will maximize (20) with $\gamma_p = 1$, this is the best we can do.

However, if $\gamma_p \neq 1$, we can do better. Due to the cross terms in (20) in the term multiplied by $(1 - \gamma_p)$, we can use selection to do better by modifying the selection approach. To understand the basic idea, let $\tilde{\mathbf{H}}'$ denote the matrix $\tilde{\mathbf{H}}_{p,p}$ for a particular selection of antennas and $\tilde{\mathbf{H}}''$ denote the same quantity for a different selection of antennas. Now consider two selection approaches which are the same except the second approach will choose $\tilde{\mathbf{H}}''$ in cases where

$$\sum_{i=1}^{n_{st}} \sum_{j=1}^{n_{st}} |\tilde{H}_{ij}''|^2 = \sum_{i=1}^{n_{st}} \sum_{j=1}^{n_{st}} |\tilde{H}_{ij}'|^2,$$

$$\sum_{i=1}^{n_{st}} \sum_{j=1}^{n_{st}} \sum_{j'=1, j'\neq j}^{n_{st}} \tilde{H}_{ij}'' \tilde{H}_{ij'}''^* > 0, \qquad (27)$$

and (in the sum, both a term and its conjugate appear, giving a real quantity)

$$\sum_{i=1}^{n_{st}} \sum_{j=1}^{n_{st}} \sum_{j'=1, j'\neq j}^{n_{st}} \tilde{H}_{ij}' \tilde{H}_{ij'}'^* < 0. \qquad (28)$$

Assume the first selection approach is the one trying to max-

imize (20) with $\gamma_p = 1$ so it will just select randomly if

$$\sum_{i=1}^{n_{st}} \sum_{j=1}^{n_{st}} |\tilde{H}_{ij}''|^2 = \sum_{i=1}^{n_{st}} \sum_{j=1}^{n_{st}} |\tilde{H}_{ij}'|^2, \qquad (29)$$

since it ignores the cross terms in its selection.

From (20), the second selection approach will give larger instantaneous mutual information for each event where the selection is different. Since the probability of the event that makes the two approaches different is greater than zero under our assumed model, then the second antenna selection approach will lead to improvement (if $\gamma_p \neq 1$) and it will do this by making the term multiplied by $(1 - \gamma_p)$ in (20) positive. Clearly the optimum selection scheme will be at least as good or better, so it must also give improvement by making the term multiplied by $(1 - \gamma_p)$ in (20) positive. $\quad\square$

We are now ready to give the main result of this section.

**Theorem 1.** *Assuming sufficiently weak interference, sufficiently weak SNRs, and optimum antenna selection, the best $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ (that maximizes the ergodic system mutual information) uses*

$$(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L) = \frac{1}{n_{st}}(\mathbf{O}_{n_{st}},\ldots,\mathbf{O}_{n_{st}}). \qquad (30)$$

*Outline of the proof.* The assumption of weak SNRs implies that $\rho_p$ is small for all $1 \leq p \leq L$. In this case, optimum selection will attempt to make $E\{\sum_{i=1}^{n_{st}} \sum_{j=1}^{n_{st}} \sum_{j'=1, j'\neq j}^{n_{st}} \tilde{h}^*(p,p)_{i,j}\tilde{h}(p,p)_{i,j'}\}$ as large as possible as shown in Lemma 3. Lemma 4 builds on Lemma 3 to show that optimum selection can always make $E\{\sum_{i=1}^{n_{st}} \sum_{j=1}^{n_{st}} \sum_{j'=1, j'\neq j}^{n_{st}} \tilde{h}^*(p,p)_{i,j}\tilde{h}(p,p)_{i,j'}\}$ positive. Lemma 2 shows that $(d/d\gamma_p)\Psi$ is directly proportional to the negative of $E\{\sum_{i=1}^{n_{st}} \sum_{j=1}^{n_{st}} \sum_{j'=1, j'\neq j}^{n_{st}} \tilde{h}^*(p,p)_{i,j}\tilde{h}(p,p)_{i,j'}\}$ which the selection is making positive and large. Thus it follows that $(d/d\gamma_p)\Psi$ is always negative which implies that the best solution employs $\gamma_p = 0$ since any increase in $\gamma_p$ away from $\gamma_p = 0$ causes a decrease in $\Psi$. Since $\rho_p$ is small for all $p$, the theorem follows. $\quad\square$

### Large SNR

Now consider the case of large SNR, where the following theorem applies.

**Theorem 2.** *Assuming sufficiently weak interference, sufficiently large SNRs, and optimum antenna selection, the best $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ (that maximizes the ergodic system mutual information) uses*

$$(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L) = \frac{1}{n_{st}}(\mathbf{I}_{n_{st}},\ldots,\mathbf{I}_{n_{st}}). \qquad (31)$$

*Outline of the proof.* Asserting the weak interference, large SNR assumption in (8) gives

$$\mathbf{Q}_i^{-1}\frac{d}{dt}\mathbf{Q}_i \longrightarrow (\rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H)^{-1}\rho_i\tilde{\mathbf{H}}_{i,i}\mathbf{S}_i'\tilde{\mathbf{H}}_{i,i}^H, \qquad (32)$$

so that

$$\begin{aligned}
&(d/d\gamma_p)\Psi(\bar{\mathbf{S}}_1,\dots,\bar{\mathbf{S}}_L)\\
&= \frac{1}{\ln(2)}E\Big\{\text{trace}\Big[\big(\tilde{\mathbf{H}}_{p,p}[\gamma_p\mathbf{I}_{n_{st}}+(1-\gamma_p)\mathbf{O}_{n_{st}}]\tilde{\mathbf{H}}_{p,p}^H\big)^{-1}\\
&\qquad\qquad \times\big(\tilde{\mathbf{H}}_{p,p}[\mathbf{I}_{n_{st}}-\mathbf{O}_{n_{st}}]\tilde{\mathbf{H}}_{p,p}^H\big)\Big]\Big\}\\
&= \frac{1}{\ln(2)}E\Big\{\text{trace}\Big[\big[\tilde{\mathbf{H}}_{p,p}^H\big]^{-1}[\gamma_p\mathbf{I}_{n_{st}}+(1-\gamma_p)\mathbf{O}_{n_{st}}]^{-1}\\
&\qquad\qquad \times[\mathbf{I}_{n_{st}}-\mathbf{O}_{n_{st}}]\tilde{\mathbf{H}}_{p,p}^H\Big]\Big\}\\
&= \frac{1}{\ln(2)}E\Big\{\text{trace}\Big[\big(\gamma_p\mathbf{I}_{n_{st}}+(1-\gamma_p)\mathbf{O}_{n_{st}}\big)^{-1}\big(\mathbf{I}_{n_{st}}-\mathbf{O}_{n_{st}}\big)\Big]\Big\}\\
&= \frac{n_{st}(n_{st}-1)(\gamma_p-1)}{\gamma_p((n_{st}-1)\gamma_p-n_{st})\ln(2)} \geq 0
\end{aligned}$$

$$(33)$$

which is positive for $0 < \gamma_p < 1$ (since $(n_{st}-1)\gamma_p < n_{st}$) and zero if $\gamma_p = 1$. In (33), we used trace $[\mathbf{CD}] = $ trace $[\mathbf{DC}]$ [13]. Thus for the large SNR case (large $\rho_p$ for all $p$) when the interference is very weak, the best signaling uses (14) with $\gamma_p = 1$. Since this is true for all $p$, the theorem follows. □

As a further comment on Theorem 2, we note that the proof makes it clear that if $\rho_p$ is large only for certain $p$, then $\gamma_p = 1$ for those $p$ only. Likewise, it is clear from Theorem 1 that if $\rho_p$ is small only for certain $p$, then $\gamma_p = 0$ for those $p$ only. Of course, this assumes weak interference. Thus we can image a case where the best signaling uses $\gamma_p = 1$ for some $p$ and $\gamma_{p'} = 0$ for some $p' \neq p$ with proper assumptions on the corresponding $\rho_p, \rho_{p'}$. One can construct similar cases where only some of the $\eta_{i,j}$ are small and easily extend the results given here in a straight forward way.

## 4. STRONG INTERFERENCE

Now consider the other extreme of dominating interference where $\eta_{i,j}$, $i = 1,\dots,L$, $j = 1,\dots,L$, is large (compared to $\rho_1,\dots,\rho_L$). The following lemma addresses the worst signaling to use.

**Lemma 5.** *Assuming sufficiently strong interference, the worst* $(\mathbf{S}_1,\dots,\mathbf{S}_L)$ *(that minimizes the ergodic system mutual information) must be of the form*

$$\begin{aligned}
&(\bar{\mathbf{S}}_1,\dots,\bar{\mathbf{S}}_L)\\
&= \alpha(\gamma_1\mathbf{I}_{n_{st}}+(1-\gamma_1)\mathbf{O}_{n_{st}},\dots,\gamma_L\mathbf{I}_{n_{st}}+(1-\gamma_L)\mathbf{O}_{n_{st}}),
\end{aligned}$$

$$(34)$$

*where* $\mathbf{O}_{n_{st}}$ *is an* $n_{st}$ *by* $n_{st}$ *matrix of all ones,* $\alpha = 1/n_{st}$, *and* $0 \leq \gamma_i \leq 1$, $i = 1,\dots,L$.

*Outline of the proof.* Provided $\eta_{i,j}$ is sufficiently large, we can approximate (9) as

$$\mathbf{Q}_i = \mathbf{I}_{n_{st}} + \rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\Big(\mathbf{I}_{n_{st}} + \sum_{j=1,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\bar{\mathbf{S}}_j\tilde{\mathbf{H}}_{i,j}^H\Big)^{-1} \approx \mathbf{I}_{n_{st}}.$$

$$(35)$$

After applying this to (12) and using (13) for large $\eta_{i,j}$ so that $\tilde{\mathbf{Q}}_i^{-1} \approx (\sum_{j=1,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\bar{\mathbf{S}}_j\tilde{\mathbf{H}}_{i,j}^H)^{-1}$, we find the first term inside the trace in (12) depends inversely on $\eta_{i,j}$, while the second term inside the trace in (12) depends inversely on $\eta_{i,j}^2$ so that the first term dominates for large $\eta_{i,j}$. Further, we can interchange the expected value and the trace in (12) so we are concerned with the expected value of (13). Now note that the first term in (13) consists of the product of a term $\mathbf{A} = \tilde{\mathbf{H}}_{i,i}\mathbf{S}_i'\tilde{\mathbf{H}}_{i,i}^H$ and another term depending on $\tilde{\mathbf{H}}_{i,j}$ for $j \neq i$. Now consider the expected value of (13) computed first as an expected value conditioned on $\{\tilde{\mathbf{H}}_{i,j},\ j \neq i\}$ and then this expected value is averaged over $\{\tilde{\mathbf{H}}_{i,j},\ j \neq i\}$. Now note that the conditional expected value of $\mathbf{A}$ becomes the zero matrix.[3] Thus the contribution from the first term in (13) averages to zero so that

$$\begin{aligned}
&\frac{d^2}{dt^2}\Psi(\bar{\mathbf{S}}_1,\dots,\bar{\mathbf{S}}_L)\\
&\approx \frac{1}{\ln(2)}\sum_{i=1}^{L}\text{trace}\Big[E\Big\{\frac{d^2}{dt^2}\mathbf{Q}_i\Big\}\Big]\\
&\approx \frac{1}{\ln(2)}\sum_{i=1}^{L}E\Big\{\text{trace}\Big[2\rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\\
&\qquad\times\Big(\sum_{j=1,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\bar{\mathbf{S}}_j\tilde{\mathbf{H}}_{i,j}^H\Big)^{-1}\\
&\qquad\times\Big(\sum_{j=1,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\mathbf{S}_j'\tilde{\mathbf{H}}_{i,j}^H\Big)\\
&\qquad\times\Big(\sum_{j=1,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\bar{\mathbf{S}}_j\tilde{\mathbf{H}}_{i,j}^H\Big)^{-1}\\
&\qquad\times\Big(\sum_{j=1,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\mathbf{S}_j'\tilde{\mathbf{H}}_{i,j}^H\Big)\\
&\qquad\times\Big(\sum_{j=1,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\bar{\mathbf{S}}_j\tilde{\mathbf{H}}_{i,j}^H\Big)^{-1}\Big]\Big\}
\end{aligned}$$

$$(36)$$

which is nonnegative. To see this, we can use a few of the same simplifications used previously. Expand the nonnegative definite matrices $2\rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H$ and $(\sum_{j=1,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\bar{\mathbf{S}}_j\tilde{\mathbf{H}}_{i,j}^H)^{-1}$ using the unitary matrix/eigenvalue expansions as done after (15). Then the matrix inside the expected value in (36) can be factored

---

[3]Recall $\mathbf{S}_i' = \hat{\mathbf{S}}_i - \mathbf{S}_i$ and use the appropriate eigenvector expansions, problem symmetry, and constraints on trace $[\hat{\mathbf{S}}_i]$, trace $[\mathbf{S}_i]$.

into $\mathbf{BB}^H$ after manipulations similar to those used after (15). Thus $\Psi(\mathbf{S}_1, \ldots, \mathbf{S}_L)$ is convex. Thus using the same permutation argument as used for the weak interference case, the result stated in the theorem follows. $\qquad\square$

The following theorem builds on Lemma 5 to specify the exact $\gamma_1, \ldots, \gamma_L$ giving worst performance.

**Theorem 3.** *Assuming sufficiently strong interference and optimum antenna selection, the worst* $(\mathbf{S}_1, \ldots, \mathbf{S}_L)$ *(that minimizes the ergodic system mutual information) uses*

$$(\bar{\mathbf{S}}_1, \ldots, \bar{\mathbf{S}}_L) = \frac{1}{n_{st}}(\mathbf{I}_{n_{st}}, \ldots, \mathbf{I}_{n_{st}}). \tag{37}$$

*Outline of the proof.* Consider $\Psi(\mathbf{S}_1, \ldots, \mathbf{S}_L)$ for $(\mathbf{S}_1, \ldots, \mathbf{S}_L)$ of the form given by Lemma 5 which is (from (2) and (3))

$$
\begin{aligned}
&\Psi(\mathbf{S}_1, \ldots, \mathbf{S}_L) \\
&= \sum_{i=1}^{L} E\{\log_2[\det(\mathbf{I}_{n_{st}} + \rho_i \tilde{\mathbf{H}}_{i,i}\mathbf{S}_i\tilde{\mathbf{H}}_{i,i}^H\mathbf{R}_i^{-1})]\} \\
&\approx \sum_{i=1}^{L} \rho_i E\{\text{trace}[\tilde{\mathbf{H}}_{i,i}\mathbf{S}_i\tilde{\mathbf{H}}_{i,i}^H\mathbf{R}_i^{-1}]\} \\
&= \frac{1}{n_{st}\ln(2)}\sum_{p=1}^{L}\rho_p \\
&\quad \times E\Bigg\{ \sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}} |\hat{h}(p,p)_{i,j}|^2 \\
&\qquad + (1-\gamma_p)\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}\sum_{j'=1, j'\neq j}^{n_{st}} \hat{h}^*(p,p)_{i,j}\hat{h}(p,p)_{i,j'} \Bigg\},
\end{aligned}
\tag{38}
$$

where the first simplification follows from large $\eta_{i,j}$ and the same simplifications used in (21). The second simplification follows from those in (20) but now $\hat{h}(p,p)_{i,j}$ denotes the $(i, j)$th entry of the matrix $\mathbf{R}_p^{-1/2}\tilde{\mathbf{H}}_{p,p}$. Now note that antenna selection will attempt to make the second term in the last line of (38), which multiplies the positive constant $1 - \gamma_p$, as large and positive as it possibly can. In fact, it is easy to argue that antenna selection can always make this term positive as done previously for (20). We skip this since the problems are so similar. Thus we see that the best performance for $(\mathbf{S}_1, \ldots, \mathbf{S}_L)$ of the form given by Lemma 5 must be obtained for $\gamma_p = 0$ and the worst performance must occur at $\gamma_p = 1$. Since this is true for all $p$, the result in the theorem follows. $\qquad\square$

The result in Theorem 3 tells us that the best signaling for cases without interference and selection is the worst for strong interference and selection. It appears that the best signaling for $(\mathbf{S}_1, \ldots, \mathbf{S}_L)$ of the form given by Lemma 5 (see the discussion in Theorem 3) may be the best signaling overall. However, it appears difficult to show this generally.

The following intuitive discussion gives some further insight. Due to convexity, the best performance will occur at a point as far away from the point giving worst performance

$(\mathbf{S}_1, \ldots, \mathbf{S}_L) = (1/n_{st})(\mathbf{I}_{n_{st}}, \ldots, \mathbf{I}_{n_{st}})$ as possible (recall that the $\gamma_1 = \cdots = \gamma_L = 1$ point gives the worst performance). Thus the best performance occurs for a point on the boundary of our space of feasible $(\mathbf{S}_1, \ldots, \mathbf{S}_L)$ and this point must be as far away from the point giving the worst performance as possible. One such point is $(\mathbf{S}_1, \ldots, \mathbf{S}_L) = (1/n_{st})(\mathbf{O}_{n_{st}}, \ldots, \mathbf{O}_{n_{st}})$. It can be shown generally (for any $n_{st}$) that this solution is the farthest from $(\mathbf{S}_1, \ldots, \mathbf{S}_L) = (1/n_{st})(\mathbf{I}_{n_{st}}, \ldots, \mathbf{I}_{n_{st}})$ (Frobenius norm). This follows because $(1/n_{st})\mathbf{O}_{n_{st}}$ is the farthest from $(1/n_{st})\mathbf{I}_{n_{st}}$. Note that $\mathbf{S}$ with one entry of 1 and the rest zero is equally far from $(1/n_{st})\mathbf{I}_{n_{st}}$ but numerical results in some specific cases indicate that the rate of increase in this direction is not as great as the rate of increase experienced by moving along the line $(\mathbf{S}_1, \ldots, \mathbf{S}_L) = \gamma(1/n_{st})(\mathbf{I}_{n_{st}}, \ldots, \mathbf{I}_{n_{st}}) + (1 - \gamma)(1/n_{st})(\mathbf{O}_{n_{st}}, \ldots, \mathbf{O}_{n_{st}})$ away from $\gamma = 1$ towards $\gamma = 0$.

## 5. NUMERICAL RESULTS FOR $n_{st} = n_{sr} = L = 2$, $n_t = n_r = 8$

Consider the case of $n_{st} = n_{sr} = L = 2$, $n_t = n_r = 8$, $\eta_{1,2} = \eta_{2,1} = \eta$, and $\rho_1 = \rho_2 = \rho$ and assume that the optimum antenna selection (to optimize system mutual information) is employed. First consider the case of no interference and assume a set of covariance matrices of the form $(\mathbf{S}_1, \mathbf{S}_2) = \gamma(1/2)(\mathbf{I}_2, \mathbf{I}_2) + (1 - \gamma)(1/2)(\mathbf{O}_2, \mathbf{O}_2)$. Thus since $\rho_1 = \rho_2 = \rho$ and $\eta_{1,2} = \eta_{2,1} = \eta$, we set $\gamma_1 = \gamma_2 = \gamma$. Figure 1 shows a plot of the $\gamma$ giving the largest mutual information versus SNR, for SNR ($\rho$) ranging from $-10$ dB to $+10$ dB. We see that the best performance for very small $\rho$ is obtained for $\gamma = 0$ which is in agreement with our analytical results given previously. For large $\rho$, the best signaling uses $\gamma = 1$ which is also in agreement with our analytical results given previously. Figure 1 shows that the switch from where $\gamma = 0$ is optimum to where $\gamma = 1$ is optimum is very rapid and occurs near $\rho = -3$ dB.

Now consider cases with possible interference. Again consider the case of $n_{st} = n_{sr} = L = 2$, $n_t = n_r = 8$, $\eta_{1,2} = \eta_{2,1} = \eta$, and $\rho_1 = \rho_2 = \rho$ and assume that the optimum antenna selection (to optimize system mutual information) is employed. To simplify matters, we constrain $\mathbf{S}_1 = \mathbf{S}_2$ in all cases shown. First we considered three specific signaling covariance matrices which are

$$
\begin{aligned}
\mathbf{S}_1 = \mathbf{S}_2 &= \begin{pmatrix} \dfrac{1}{2} & 0 \\[2mm] 0 & \dfrac{1}{2} \end{pmatrix}, \\[2mm]
\mathbf{S}_1 = \mathbf{S}_2 &= \begin{pmatrix} \dfrac{1}{2} & \dfrac{1}{2} \\[2mm] \dfrac{1}{2} & \dfrac{1}{2} \end{pmatrix}, \\[2mm]
\mathbf{S}_1 = \mathbf{S}_2 &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.
\end{aligned}
\tag{39}
$$

We tried each of these for SNRs and INRs between $-10$ dB and $+10$ dB. Then we recorded which of the approaches provided the smallest and the largest system mutual information. These results can be compared with the analytical
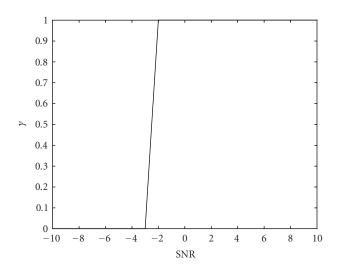
FIGURE 1: Optimum $\gamma$ versus $\rho_1 = \rho_2 = $ SNR for cases with no interference and $n_{st} = n_{sr} = 2$, $n_t = n_r = 8$. Note that $\gamma = 0$ is the best for $-10\,\mathrm{dB} < \mathrm{SNR} < -3\,\mathrm{dB}$ and $\gamma = 1$ is the best for $-2\,\mathrm{dB} < \mathrm{SNR} < 10\,\mathrm{dB}$.
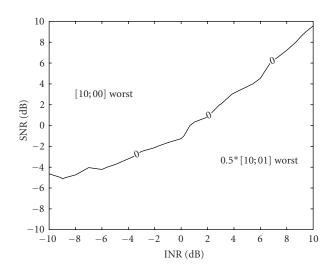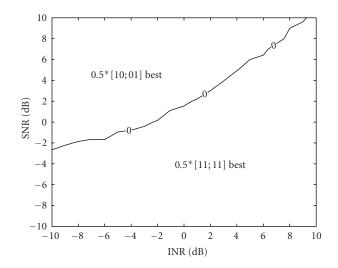


FIGURE 3: The best signaling (of the three choices) versus SNR and INR for $n_{st} = n_{sr} = L = 2$, $n_t = n_r = 8$, $\rho_1 = \rho_2 = $ SNR, and $\eta_{1,2} = \eta_{2,1} = $ INR.



FIGURE 2: The worst signaling (of the three approaches) versus SNR and INR for $n_{st} = n_{sr} = L = 2$, $n_t = n_r = 8$, $\rho_1 = \rho_2 = $ SNR, and $\eta_{1,2} = \eta_{2,1} = $ INR.

results given in Sections 3 and 4 of this paper for weak and strong interference and SNR. Figure 2 shows the worst signaling we found versus SNR and INR for $\rho_1 = \rho_2 = $ SNR and $\eta_{1,2} = \eta_{2,1} = $ INR. For large INR, Figure 2 indicates that $\mathbf{S}_1 = \mathbf{S}_2 = (1/2)\mathbf{I}_2$ leads to worst performance which is in agreement with our analytical results given previously. Figure 2 also shows that either $\mathbf{S}_1 = \mathbf{S}_2 = (1/2)\mathbf{I}_2$ (for weak SNR) or (for large SNR) $\mathbf{S}_1 = \mathbf{S}_2$ with only one nonzero entry (a one which must be along the diagonal) will lead to worst performance for weak interference.

For weak interference, Figure 3 shows that the best performance is achieved by either $\mathbf{S}_1 = \mathbf{S}_2 = (1/2)\mathbf{O}_2$ (for weak

SNR) or $\mathbf{S}_1 = \mathbf{S}_2 = (1/2)\mathbf{I}_2$ (for large SNR). This agrees with our analytical results presented previously. Figure 3 shows that the best performance is achieved by $\mathbf{S}_1 = \mathbf{S}_2 = (1/2)\mathbf{O}_2$ for large interference and this also agrees with our analytical results presented previously. We note that in the cases of interest (those for which we give analytical results), the difference in mutual information between the best and the worst approach in Figures 2 and 3 was about 1 to 3 bits/s/Hz.

We selected a few SNR-INR points sufficiently (greater than 2 dB) far from the dividing curves in Figures 2 and 3. For these points, we attempted to obtain further information on whether the approaches shown to be the best and worst in Figures 2 and 3 are actually the best and the worst of all valid approaches under the assumption that $\mathbf{S}_1 = \mathbf{S}_2$. We did this by evaluating the system mutual information for

$$\mathbf{S}_1 = \mathbf{S}_2 = \begin{pmatrix} a & b \\ b^* & \rho - a \end{pmatrix} \qquad (40)$$

for various values of the real constant $a$ and the complex constant $b$ on a grid. When we evaluated (40) for all real $a$ and $b$ an a grid for a range of values consistent with the trace (power) and nonnegative definite enforcing constraints on $\mathbf{S}_1 = \mathbf{S}_2$, we did find the approaches in Figures 2 and 3 did indicate the overall best and worst approaches for the few cases we tried. Limited investigations involving complex $b$ (here the extra dimension complicated matters, making strong conclusions difficult) indicated that these conclusions appeared to generalize to complex $b$ also.

### Partitioning the SNR-INR Plane

Based on Sections 3 and 4, we see that generally the space of all SNRs $\rho_i$, $i = 1, \ldots, L$, and INRs $\eta_{i,j}$, $i, j = 1, \ldots, L$, $i \neq j$, can be divided into three regions: one where the interference is considered weak (where Figure 1 and its generalization

apply), one where the interference is considered to dominate (where Figure 3 and its generalization apply), and a transition region between the two.

For the case with $n_{st} = n_{sr} = L = 2$, $n_t = n_r = 8$, $\eta_{1,2} = \eta_{2,1} = \eta$, and $\rho_1 = \rho_2 = \rho$, we have used (12) to study the three regions. We first evaluated (12) numerically using Monte Carlo simulations for a grid of points in SNR and INR space. The Monte Carlo simulations just described were calculated over a very fine grid over the region $-10\,\text{dB} \le \rho \le 10\,\text{dB}$ and $-10\,\text{dB} \le \eta \le 10\,\text{dB}$. For each given point in SNR and INR space, we evaluated (12) for many different choices of $(\mathbf{S}_1,\dots,\mathbf{S}_L)$, $(\hat{\mathbf{S}}_1,\dots,\hat{\mathbf{S}}_L)$, and the scalar $t$. We checked for a consistent positive or negative value for (12) for all $(\mathbf{S}_1,\dots,\mathbf{S}_L)$, $(\hat{\mathbf{S}}_1,\dots,\hat{\mathbf{S}}_L)$, and the scalar $t$ on the discrete grid (quantize each scalar variable, including those in each entry of each matrix). In this way, we have viewed the approximate form of these three regions. We found that generally for points sufficiently far (more than 2 dB from closest curve) from the two dividing curves in Figures 2 and 3, the convexity and concavity follows that for the asymptotic case (strong or weak INR) in the given region. Thus the asymptotic results appear to give valuable conclusions about finite SNR and INR cases. Limited numerical investigations suggest this is true in other cases but the high dimensionality of the problem (especially for $n_{st}, n_{sr}, L > 2$) makes strong conclusions difficult.

## 6. CONCLUSIONS

We have analyzed the (mutual information) optimum signaling for cases where multiple users interfere while using single user detection and antenna selection. We concentrate on extreme cases with very weak interference or very strong interference. We have found that the best signaling is sometimes different from the scaled identity matrix that is best for no interference and no antenna selection. In fact, this is true even for cases without interference if SNR is sufficiently weak. Further, the scaled identity matrix is actually the covariance matrix that yields worst performance if the interference is sufficiently strong.

## ACKNOWLEDGMENT

## REFERENCES

[1] J. H. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 5, pp. 871–878, 1987.

[2] 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, no. 3, pp. 311–335, 1998.

[3] N. Kong and L. B. Milstein, "Combined average SNR of a generalized diversity selection combining scheme," in *Proc. IEEE International Conference on Communications*, vol. 3, pp. 1556–1560, Atlanta, Ga, USA, June 1998.

[4] M. Z. Win and J. H. Winters, "Analysis of hybrid selection/maximal-ratio combining in Rayleigh fading," *IEEE Trans. Communications*, vol. 47, no. 12, pp. 1773–1776, 1999.

[5] R. Nabar, D. Gore, and A. Paulraj, "Optimal selection and use of transmit antennas in wireless systems," in *Proc. International Conference on Telecommunications*, Acapulco, Mexico, May 2000.

[6] D. Gore, R. Nabar, and A. Paulraj, "Selection of an optimal set of transmit antennas for a low rank matrix channel," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 2785–2788, Istanbul, Turkey, June 2000.

[7] A. F. Molisch, M. Z. Win, and J. H. Winters, "Capacity of MIMO systems with antenna selection," in *Proc. IEEE International Conference on Communications*, vol. 2, pp. 570–574, Helsinki, Finland, June 2001.

[8] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[9] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Communications*, vol. 46, no. 3, pp. 357–366, 1998.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.

[11] R. S. Blum, "MIMO capacity with interference," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 793–801, 2003.

[12] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[13] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation and Modulation Theory*, John Wiley & Sons, New York, NY, USA, 2002.

[14] P. J. Voltz, "Characterization of the optimum transmitter correlation matrix for MIMO with antenna subset selection," submitted to IEEE Trans. Communications.

**Rick S. Blum** received his B.S. degree in electrical engineering from the Pennsylvania State University in 1984 and his M.S. and Ph.D. degrees in electrical engineering from the University of Pennsylvania in 1987 and 1991. From 1984 to 1991, he was a member of technical staff at General Electric Aerospace in Valley Forge, Pennsylvania, and he graduated from GE's Advanced Course in Engineering. Since 1991, he has been with the Electrical and Computer Engineering Department at Lehigh University in Bethlehem, Pennsylvania, where he is currently a Professor and holds the Robert W. Wieseman Chair in electrical engineering. His research interests include signal detection and estimation and related topics in the areas of signal processing and communications. He is currently an Associate Editor for the IEEE Transactions on Signal Processing and for IEEE Communications Letters. He was a member of the Signal Processing for Communications Technical Committee of the IEEE Signal Processing Society. Dr. Blum is a member of Eta Kappa Nu and Sigma Xi, and holds a patent for a parallel signal and image processor architecture. He was awarded an Office of Naval Research (ONR) Young Investigator Award in 1997 and a National Science Foundation (NSF) Research Initiation Award in 1992.