

## **Spatial Multiplexing and Channel Statistics-Based RF Pre-Processing for Antenna Selection**

Pallav Sudarshan, Neelesh B. Mehta, Andreas F. Molisch, Jin Zhang

TR2004-112 November 2004

### **Abstract**

For a multiple input multiple output system, antenna selection reduces complexity at the expense of performance. In this paper, we propose two novel RF pre-processing architectures that significantly improve the performance of antenna selection, which marginally increasing the complexity. These architectures introduce a RF pre-processing matrix,  $M$ , that multiplies the vector of incoming signals prior to down-conversion. The elements of  $M$  use only the knowledge of the channel statistics. In the first architecture,  $M$  outputs a reduced number of streams - an explicit selection algorithm is therefore not required. In the second architecture, the number of output streams equals the number of input streams, and a selection switch that uses instantaneous channel state information achieves the reduction of the number of streams. We show that the optimal pre-processing receiver projects the received signal along the eigenvectors of the correlation matrix. In a correlated channel, both these architectures significantly outperform conventional antenna selection. We also develop a beam-pattern based intuition and compare the performance of our scheme to other RF pre-processing schemes previously proposed in the literature.

*Globecom 2004*

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.



# Spatial Multiplexing and Channel Statistics-Based RF Pre-Processing for Antenna Selection

Pallav Sudarshan, *Student Member, IEEE*,<sup>†</sup>  
Electrical and Computer Engineering Dept.,  
North Carolina State University,  
Raleigh, NC, USA.  
psudars@ncsu.edu

Neelesh B. Mehta, *Member, IEEE*,  
Andreas F. Molisch, *Senior Member, IEEE*<sup>‡</sup>,  
Jin Zhang, *Senior Member, IEEE*,  
Mitsubishi Electric Research Labs,  
201 Broadway, Cambridge, MA, USA.  
{mehta, molisch, jzhang}@merl.com

**Abstract**—For a multiple input multiple output system, antenna selection reduces complexity at the expense of performance. In this paper, we propose two novel RF pre-processing architectures that significantly improve the performance of antenna selection, which marginally increasing the complexity. These architectures introduce a RF pre-processing matrix,  $\mathbf{M}$ , that multiplies the vector of incoming signals prior to down-conversion. The elements of  $\mathbf{M}$  use only the knowledge of the channel statistics. In the first architecture,  $\mathbf{M}$  outputs a reduced number of streams – an explicit selection algorithm is therefore not required. In the second architecture, the number of output streams equals the number of input streams, and the reduction of the number of streams is achieved by a selection switch that uses instantaneous channel state information. We show that the optimal pre-processing receiver projects the received signal along the eigenvectors of the correlation matrix. In a correlated channel, both these architectures significantly outperform conventional antenna selection. We also develop a beam-pattern based intuition and compare the performance of our scheme to other RF pre-processing schemes previously proposed in the literature.

**Index Terms**—MIMO systems, Spatial multiplexing, Antenna arrays, Antenna selection, Information rates, Phase shifters.

## I. INTRODUCTION

It is well known that using multiple antennas at the transmitter and the receiver for spatial multiplexing can achieve high capacity in a wireless communication system [1]. However, those multiple input multiple output (MIMO) systems significantly increase the signal processing complexity and the hardware cost. Each antenna element requires a separate modulator (demodulator) chain at the transmitter (receiver) that consists of an up-converter (down-converter), a D/A (A/D), and a power amplifier (low noise amplifier), which can be expensive. Antenna selection [3]–[6] reduces the number of required chains in MIMO systems by adaptively choosing the signals from a subset of available antenna elements. Although antenna selection maintains the diversity order of the system [5], a penalty is paid in the form of a lower beamforming gain.

Recently [7]–[9], system architectures involving RF pre-processing have been shown to significantly increase the

beamforming gain of antenna selection systems for the same number of modulator/demodulator chains, while retaining full diversity. RF pre-processing at the receiver involves processing the received signal using a RF matrix,  $\mathbf{M}$ , prior to selection and down-conversion. At the transmitter, the order is reversed. Reference [8] considers the case when the entries of  $\mathbf{M}$  can tune to the instantaneous channel state information (CSI). In [7], an alternate approach in which  $\mathbf{M}$  – an FFT Butler matrix – is completely independent of the CSI is proposed. While the gains were promising, they were optimal for certain fixed angles of arrival/departure (AoA/AoD) only.

In this work, we introduce a novel RF pre-processing architecture that uses the knowledge of only the *channel statistics*, which depend on the large-scale parameters of the channel, such as mean AoA, mean AoD, angle spread etc. These parameters vary very slowly compared to the small-scale fading parameters. Using only the statistical knowledge ignores the instantaneous channel induced variations and alleviates tighter design constraints on the RF elements required in [8]. When pre-processing is employed at the transmitter, this also considerably reduces the feedback burden. In this paper, we focus on pre-processing and selection at the receiver; similar arguments hold of the transmitter as well. Hereafter, we refer to the channel statistics-based solution as the *time-invariant solution*. Variable phase-shifter-based approximations, suitable for implementation using today’s technology, are also proposed and evaluated.

For a system with  $N_r$  receive antennas and  $L \leq N_r$  demodulator chains, we propose and analyze two different but related pre-processing architectures that maximize the ergodic capacity of the system:

- 1) An  $L \times N_r$  time-invariant matrix,  $\mathbf{M}_{\text{TI}}$ , that outputs only  $L$  streams, thereby eliminating the need for subsequent selection.
- 2) An  $N_r \times N_r$  time-invariant matrix,  $\mathbf{M}_{\text{TI-S}}$ , followed by instantaneous CSI-based selection.

The output of the above set ups is down-converted and processed in baseband.

We show that for correlated channels, using  $\mathbf{M}_{\text{TI}}$  performs considerably better than conventional antenna selection algorithms, despite not requiring any selection. It thus obviates the

<sup>†</sup> This work was done when the author was at Mitsubishi Electric Research Labs.

<sup>‡</sup> A. F. Molisch is also at the Department of Electroscience, Lund University, Lund, Sweden.

need for fast antenna selection algorithms [5], [6]. However, this is not so for channels with low spatial correlation. For such cases,  $\mathbf{M}_{\text{TI-S}}$  is preferable as it outperforms both  $\mathbf{M}_{\text{TI}}$  and conventional antenna selection; performance parity occurs only for completely spatially uncorrelated channels.

The rest of the paper is organized as follows. Section II describes the spatial multiplexing system and the channel model used. In Section III, we derive the optimal time-invariant pre-processing matrices,  $\mathbf{M}_{\text{TI}}$  and  $\mathbf{M}_{\text{TI-S}}$ . Section IV presents the simulation results for the proposed solution and compares them with the conventional solutions. The conclusions follow in Section V.

## II. SYSTEM MODEL

The following notation is used in the paper:  $(\cdot)^\dagger$  denotes the Hermitian of a matrix,  $|\cdot|$  denotes the determinant of a matrix,  $C^{N,M}$  is the set of all  $N \times M$  matrices with complex entries,  $U^{N,M}$  is the set of all  $N \times M$  semi-unitary matrices,  $C^N$  is the set of all  $N \times 1$  vectors with complex entries,  $D^N$  is the set of all  $N \times N$  diagonal matrices with complex entries, and  $\mathbb{E}_X\{\cdot\}$  is the expectation with respect to  $X$ .

In spatial multiplexing, different data streams are transmitted simultaneously as shown in Fig. 1. The transmitter and the receiver have  $N_t$  and  $N_r$  antenna elements, respectively. No CSI is available at the transmitter, while perfect CSI is available at the receiver. The received vector,  $\mathbf{y}$ , for  $N_r$  receive antennas can be written as

$$\mathbf{y} = \sqrt{\frac{\rho}{N_t}} \mathbf{H} \mathbf{x} + \mathbf{n}, \quad (1)$$

where  $\rho$  is the signal to noise ratio (SNR),  $N_t$  is the number of transmit antennas,  $\mathbf{H}$  is the  $N_r \times N_t$  channel matrix, and  $\mathbf{x}$  is the  $N_t \times 1$  transmitted vector. The vector  $\mathbf{n}$  is additive white Gaussian noise and follows the complex Gaussian distribution  $\mathcal{N}_c(\mathbf{0}, \mathbf{I}_{N_r})$ ,  $\mathbf{0}$  is the all zeros mean vector and  $\mathbf{I}_{N_r}$  is the  $N_r \times N_r$  identity matrix. For such a system, the channel capacity of a full complexity receiver (that uses all receiver antennas) is given by

$$C_{\text{FC}} = \log_2 \left| \mathbf{I}_{N_t} + \frac{\rho}{N_t} \mathbf{H}^\dagger \mathbf{H} \right|.$$

We adopt the widely used Kronecker correlation channel model for  $\mathbf{H}$  [2]. As per this model,  $\mathbf{H}$  is given by

$$\mathbf{H} = \mathbf{R}^{\frac{1}{2}} \mathbf{H}_w \mathbf{T}^{\frac{1}{2}}, \quad (2)$$

where the elements of  $\mathbf{H}_w$  are independent identically distributed complex Gaussians  $\sim \mathcal{N}_c(0, 1)$ ,  $\mathbf{R} = \mathbb{E}_{\mathbf{H}} \{\mathbf{H} \mathbf{H}^\dagger\}$  is the  $N_r \times N_r$  receiver correlation matrix, and  $\mathbf{T} = \mathbb{E}_{\mathbf{H}} \{\mathbf{H}^\dagger \mathbf{H}\}$  is the  $N_t \times N_t$  transmitter correlation matrix.

## III. OPTIMAL TIME-INVARIANT PRE-PROCESSING RECEIVER

In this section, we address two cases of time-invariant pre-processing. We first derive the optimal  $L \times N_r$  time-invariant pre-processing matrix,  $\mathbf{M}_{\text{TI}}$ , that maximizes the ergodic capacity. Since the pre-processor only outputs  $L$  streams, a

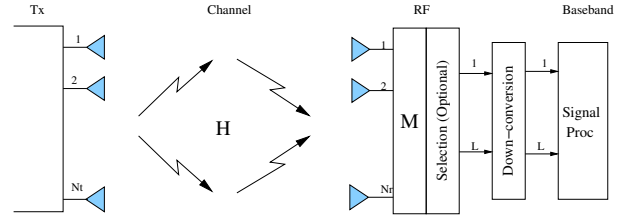


Fig. 1. Block Diagram for Spatial Multiplexing

subsequent selection algorithm is not required. Next, we derive a receiver architecture in which the time-invariant  $N_r \times N_r$  matrix  $\mathbf{M}_{\text{TI-S}}$  is followed by CSI-based selection.

### A. Optimal $L \times N_r$ Time-Invariant (TI) Pre-Processing

Let  $\mathbf{M}_L$  be an  $L \times N_r$  pre-processing matrix. The  $L \times 1$  vector,  $\tilde{\mathbf{y}}$ , at the output of  $\mathbf{M}_L$  is

$$\tilde{\mathbf{y}} = \sqrt{\frac{\rho}{N_t}} \mathbf{M}_L \mathbf{H} \mathbf{x} + \mathbf{M}_L \mathbf{n}.$$

For such a system, the ergodic capacity is

$$\bar{C} = \max_{\mathbf{M}_L} \mathbb{E}_{\mathbf{H}} \left\{ \log_2 \left| \mathbf{I}_{N_t} + \frac{\rho}{N_t} \mathbf{H}^\dagger \mathbf{M}_L^\dagger (\mathbf{M}_L \mathbf{M}_L^\dagger)^{-1} \mathbf{M}_L \mathbf{H} \right| \right\}. \quad (3)$$

The following lemmas, are required for finding the optimal  $\mathbf{M}_L$ .

*Lemma 1:* Let  $\mathbf{A} \in C^{N_t, N_r}$ ,  $\mathbf{\Lambda}_1 \in D^{N_r}$  and  $\mathbf{\Lambda}_2 \in D^{N_r}$  be given such that each element of  $\mathbf{\Lambda}_1$  is less than corresponding element of  $\mathbf{\Lambda}_2$ . Then,

$$|\mathbf{I}_{N_t} + \mathbf{A} \mathbf{\Lambda}_2 \mathbf{A}^\dagger| \geq |\mathbf{I}_{N_t} + \mathbf{A} \mathbf{\Lambda}_1 \mathbf{A}^\dagger|.$$

*Proof:* The proof is given in Appendix A-1.

*Lemma 2:* [13, Lemma 3.3.1] Let  $\mathbf{Q} \in U^{L, N_r}$  and  $\mathbf{R} \in C^{N_r, N_r}$  be given such that  $\mathbf{R}$  is full rank. Let  $e_i(\mathbf{X})$  represent the  $i^{\text{th}}$  largest eigenvalue of  $\mathbf{X} \mathbf{X}^\dagger$ . Then,

$$e_{N_r-L+i}(\mathbf{R}) \leq e_i(\mathbf{Q} \mathbf{R}) \leq e_i(\mathbf{R}). \quad (4)$$

Furthermore,  $e_i(\mathbf{Q} \mathbf{R}) = e_i(\mathbf{R})$ , ( $1 \leq i \leq L$ ) if  $\mathbf{Q}$  is the conjugate transpose of the  $L$  eigenvectors of  $\mathbf{R} \mathbf{R}^\dagger$  corresponding to its  $L$  largest eigenvalues.

The following main theorem follows.

*Theorem 1:* Consider a general  $N_r \times N_t$  channel of the form  $\mathbf{H} = \mathbf{R}^{\frac{1}{2}} \mathbf{H}_w \mathbf{T}^{\frac{1}{2}}$ . The optimal  $L \times N_r$  time-invariant pre-processing matrix,  $\mathbf{M}_{\text{TI}}$ , that maximizes the ergodic capacity in (3), is of the form  $\mathbf{M}_{\text{TI}} = \mathbf{B} \mathbf{Q}_{\text{opt}}$ , where  $\mathbf{B}$  is any  $L \times L$  full rank matrix, and  $\mathbf{Q}_{\text{opt}}$  is given by

$$\mathbf{Q}_{\text{opt}} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_L]^\dagger.$$

Here  $\mathbf{r}_i$  is the singular vector of  $\mathbf{R}$  corresponding to its  $i^{\text{th}}$  largest singular value. The maximum ergodic capacity is then given by

$$C_{\text{TI}} = \mathbb{E}_{\mathbf{H}} \left\{ \log_2 \left| \mathbf{I}_{N_t} + \frac{\rho}{N_t} \mathbf{H}^\dagger \mathbf{Q}_{\text{opt}}^\dagger \mathbf{Q}_{\text{opt}} \mathbf{H} \right| \right\}. \quad (5)$$

*Proof:* Any  $L \times N_r$  matrix with row span  $L$  can be written as  $\mathbf{M}_L = \mathbf{B}\mathbf{Q}$ , where  $\mathbf{B} \in C^{L,L}$  and  $\mathbf{Q} \in U^{L,N_r}$ .<sup>1</sup> Since  $\mathbf{Q}\mathbf{Q}^\dagger = \mathbf{I}_L$ , (3) reduces to

$$C_{\text{TI}} = \max_{\mathbf{Q} \in U^{L,N_r}} \mathbb{E}_{\mathbf{H}} \left\{ \log_2 \left| \mathbf{I}_{N_t} + \frac{\rho}{N_t} \mathbf{H}^\dagger \mathbf{Q}^\dagger \mathbf{Q} \mathbf{H} \right| \right\}. \quad (6)$$

Recall that  $\mathbf{H} = \mathbf{R}^{\frac{1}{2}} \mathbf{H}_w \mathbf{T}^{\frac{1}{2}}$ . Define  $\mathbf{X}_L = \mathbf{Q}\mathbf{R}$ , and let its singular value decomposition (SVD) be  $\mathbf{X}_L = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger$ . Then

$$\begin{aligned} C_{\text{TI}} &= \max_{\mathbf{X}_L} \mathbb{E}_{\mathbf{H}_w} \left\{ \log_2 \left| \mathbf{I}_{N_t} + \frac{\rho}{N_t} \mathbf{T}^{\frac{1}{2}} \mathbf{H}_w^\dagger \mathbf{X}_L^\dagger \mathbf{X}_L \mathbf{H}_w \mathbf{T}^{\frac{1}{2}} \right| \right\}, \\ &= \max_{\mathbf{V}, \mathbf{\Sigma}} \mathbb{E}_{\mathbf{H}_w} \left\{ \log_2 \left| \mathbf{I}_{N_t} + \frac{\rho}{N_t} \mathbf{T}^{\frac{1}{2}} \mathbf{H}_w^\dagger \mathbf{V} \mathbf{\Sigma} \mathbf{V}^\dagger \mathbf{H}_w \mathbf{T}^{\frac{1}{2}} \right| \right\}, \\ &= \max_{\mathbf{\Sigma} \in D^{N_r}} \mathbb{E}_{\mathbf{H}_w} \left\{ \log_2 \left| \mathbf{I}_{N_t} + \frac{\rho}{N_t} \mathbf{T}^{\frac{1}{2}} \mathbf{H}_w^\dagger \mathbf{\Sigma} \mathbf{H}_w \mathbf{T}^{\frac{1}{2}} \right| \right\}. \end{aligned} \quad (7)$$

Eqn. (7) follows because the statistical properties of  $\mathbf{H}_w$  are not changed by multiplication by a unitary matrix. From Lemma 1, it follows that (7) is maximized when each non-zero element of  $\mathbf{\Sigma}$  is maximized. Since the non-zero elements of  $\mathbf{\Sigma}$  correspond to the eigenvalues of  $\mathbf{Q}\mathbf{R}\mathbf{Q}^\dagger$ , Lemma 2 implies that (7) is maximized when the eigenvalues of  $\mathbf{Q}\mathbf{R}\mathbf{Q}^\dagger$  reach the maximum possible value of  $e_i(\mathbf{R})$ . The optimal  $\mathbf{Q}$  is then the conjugate transpose of the eigenvectors corresponding to the  $L$  largest eigenvalues of  $\mathbf{R}$ . ■

We see from Theorem 1 that the optimal  $\mathbf{Q}$  matrix is obtained by performing a principal component analysis [11] on the columns of  $\mathbf{H}$ , and choosing the largest  $L$  principal components. The variance of the columns of  $\mathbf{H}$  along the eigenvectors of  $\mathbf{R}$  is equal to the corresponding eigenvalues of the correlation matrix. Thus, statistically, the first  $L$  eigenvectors of  $\mathbf{R}$  contribute the most to the variance.

### B. Time-Invariant $N_r \times N_r$ Pre-Processing Followed by Selection (TI-S)

In this section, we consider a system where the received vector  $\mathbf{y}$  is multiplied by an  $N_r \times N_r$  pre-processing matrix,  $\mathbf{M}_N$ , and  $L$  out of the  $N_r$  outputs of the pre-processor are selected for down-conversion. The vector,  $\check{\mathbf{y}}$ , at the output of the selection switch is

$$\check{\mathbf{y}} = \sqrt{\frac{\rho}{N_t}} \mathbf{S} \mathbf{M}_N \mathbf{H} \mathbf{x} + \mathbf{S} \mathbf{M}_N \mathbf{n}, \quad (8)$$

where  $\mathbf{S}$  is an  $L \times N_r$  selection matrix that selects  $L$  out of  $N_r$  signals. The selection matrix  $\mathbf{S}$  adapts to the instantaneous channel state. The capacity maximization problem now becomes

$$C_{\text{TI-S}} = \max_{\mathbf{M}_N} \mathbb{E}_{\mathbf{H}} \left\{ \max_{\mathbf{S}} \log_2 \left| \mathbf{I}_{N_t} + \frac{\rho}{N_t} \mathbf{H}^\dagger \mathbf{M}_N^\dagger \mathbf{S}^\dagger \right. \right. \\ \left. \left. \times (\mathbf{S} \mathbf{M}_N \mathbf{M}_N^\dagger \mathbf{S}^\dagger)^{-1} \mathbf{S} \mathbf{M}_N \mathbf{H} \right| \right\}. \quad (9)$$

It is difficult to analytically find  $\mathbf{M}_N$  that maximizes (9) because of the presence of  $\mathbf{S}$ , which depends on the channel

<sup>1</sup> $\mathbf{M}_L$  with row span less than  $L$  is sub-optimal, and is therefore not considered.

realization  $\mathbf{H}$ . By interchanging the order of  $\mathbb{E}_{\mathbf{H}}$  and  $\max$ , we get the analytically tractable lower bound

$$C_{\text{TI-S}} \geq \max_{\mathbf{S}} \max_{\mathbf{M}_N} \mathbb{E}_{\mathbf{H}} \left\{ \log_2 \left| \mathbf{I}_{N_t} + \frac{\rho}{N_t} \mathbf{H}^\dagger \mathbf{M}_N^\dagger \mathbf{S}^\dagger \right. \right. \\ \left. \left. \times (\mathbf{S} \mathbf{M}_N \mathbf{M}_N^\dagger \mathbf{S}^\dagger)^{-1} \mathbf{S} \mathbf{M}_N \mathbf{H} \right| \right\}. \quad (10)$$

*Theorem 2:* Given a selection matrix,  $\mathbf{S}_0$ , any  $N_r \times N_r$  matrix  $\mathbf{M}_{\text{TI-S}}$  of the form

$$\mathbf{M}_{\text{TI-S}} = \mathbf{P} [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_L, \mathbf{v}_1, \dots, \mathbf{v}_{N_r-L}]^\dagger \quad (11)$$

maximizes the lower bound in (10), where  $\mathbf{P}$  is an  $N_r \times N_r$  permutation matrix and  $\mathbf{v}_k$ ,  $k = 1, \dots, N_r - L$ , are any orthonormal vectors in  $N(U_L)$ . Here,  $U_L$  is the vector space spanned by  $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_L$  and  $N(U_L)$  is the null space of  $U_L$ . The vectors  $\mathbf{r}_i$  are as defined in Theorem 1.

*Proof:* We restrict  $\mathbf{M}_{\text{TI-S}}$  to be unitary.<sup>2</sup> Then, for a given selection matrix  $\mathbf{S}_0$ , the problem in (10) is similar to the one we solved in Section III-A (with  $\mathbf{B} = \mathbf{I}$ ). It can be seen that the  $\mathbf{M}_N$  that maximizes (10) has the following property: the rows of  $\mathbf{M}_N$  that are selected by  $\mathbf{S}_0$  are the eigenvectors corresponding to the  $L$  largest eigenvalues of  $\mathbf{R}$ . For example, if  $\mathbf{S}_0 = [\mathbf{I}_L \ \mathbf{0}]$ , then  $\mathbf{M}_N^{(0)} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_L, \mathbf{v}_1, \dots, \mathbf{v}_{N_r-L}]^\dagger$ . Given that  $\mathbf{M}_{\text{TI-S}}$  is unitary,  $\mathbf{v}_1, \dots, \mathbf{v}_{N_r-L}$  are orthonormal vectors in the null space of  $U_L$ . Any other selection matrix is a permutation of  $\mathbf{S}_0$ . Thus, in general,  $\mathbf{M}_{\text{TI-S}}$  is a permutation of  $\mathbf{M}_N^{(0)}$ . ■

In order to completely characterize  $\mathbf{M}_{\text{TI-S}}$ , we need to find the orthonormal vectors  $\mathbf{v}_k$ ,  $k = 1, \dots, N_r - L$ . Clearly, if the selection matrix selects the first  $L$  rows of  $\mathbf{M}_{\text{TI-S}}$ , i. e.,  $\mathbf{S} = \mathbf{S}_0$ , then  $\mathbf{M}_N^{(0)}$  maximizes the lower bound in (10). ( $\mathbf{S}_0$  and  $\mathbf{M}_N^{(0)}$  are defined in the proof above.) We fix the first  $L$  rows of  $\mathbf{M}_{\text{TI-S}}$  to be  $\mathbf{r}_i$  as this ensures that any subsequent choice for  $\mathbf{v}_i$  does not affect the performance for  $\mathbf{S}_0$ . This necessarily improves the lower bound in (10). If the selection matrix selects the 2<sup>nd</sup> row to the  $(L+1)$ <sup>th</sup> row of  $\mathbf{M}_{\text{TI-S}}$ , then Theorem 1 implies that the ergodic capacity is maximized when  $\mathbf{v}_1 = \mathbf{r}_{L+1}$ . Note that  $\mathbf{r}_{L+1} \in N(U_L)$ . Following a similar procedure, we get  $\mathbf{v}_2 = \mathbf{r}_{L+2}, \dots, \mathbf{v}_{N_r-L} = \mathbf{r}_{N_r}$ .

Thus the time-invariant matrix  $\mathbf{M}_{\text{TI-S}}$  is given by

$$\mathbf{M}_{\text{TI-S}} = \mathbf{P} \mathbf{\Upsilon}^\dagger, \quad (12)$$

where  $\mathbf{\Upsilon}$  is the eigenspace of  $\mathbf{R}$ . Note that this is not a rigorous proof of optimality of  $\mathbf{M}_{\text{TI-S}}$ , but we conjecture that the  $\mathbf{M}_{\text{TI-S}}$  thus obtained is indeed optimal.

### C. Phase-Only Approximation

While the elements of  $\mathbf{M}$ , derived above, can have an arbitrary amplitude, hardware considerations make the implementation of the RF pre-processing matrix using variable-phase shifters an attractive option [12]. A phase-only approximation to  $\mathbf{M}_{\text{TI}}$  can be obtained using the algorithm proposed in [9]. We study its performance in Section IV.

<sup>2</sup> $\mathbf{M}_{\text{TI-S}}$  can be written as product of a full-rank matrix and a unitary matrix. The full-rank matrix does not affect the capacity, and can therefore be chosen to be the identity matrix.

TABLE I  
ERGODIC CAPACITY (IN BITS/S/Hz) FOR DIFFERENT RECEIVERS.

Parameters			FC	TV	TI-S	TI	TI-Ph	FFT	Ant. Sel.
$\theta_r$	$\sigma_r$	$L$							
45°	6°	1	5.99	4.96	4.87	4.86	4.75	4.26	3.50
45°	6°	2	5.99	5.99	5.96	5.96	5.88	5.77	4.96
60°	15°	2	7.25	7.25	6.82	6.70	6.62	6.58	6.05

#### IV. SIMULATION RESULTS

For the purpose of subsequent illustrations and simulations, we assume that the transmitter and receiver are uniform linear arrays (ULA) with antenna element spacing  $d = 0.5\lambda$ , where  $\lambda$  is the wavelength. The AoA and AoD at the receiver and transmitter have Gaussian distributions  $\mathcal{N}(\theta_r, \sigma_r^2)$  and  $\mathcal{N}(\theta_t, \sigma_t^2)$ , respectively. Here,  $\theta_r$  ( $\theta_t$ ) is the mean AoA (AoD) and  $\sigma_r$  ( $\sigma_t$ ) is the root mean square (RMS) angle spread at the receiver (transmitter). For such a configuration,  $\mathbf{R}$  and  $\mathbf{T}$  are determined using [10].

We study and compare the performance of channel statistics-based pre-processing –  $L \times N_r$  time-invariant pre-processing (TI),  $L \times N_r$  time-invariant phase-approximated pre-processing (TI-Ph), and  $N_r \times N_r$  time-invariant pre-processing followed by selection (TI-S) – with several receiver structures studied in the literature such as full complexity (FC), conventional antenna selection, FFT pre-processing with selection [7], and time-variant pre-processing [8]. The number of antenna elements is chosen to be  $N_t = 2$  and  $N_r = 4$  and the SNR is  $\rho = 10$  dB. We assume that  $\mathbf{T} = \mathbf{I}$  for the sake of simplicity. Each receiver architecture has  $L$  demodulator chains available ( $L \leq N_r$ ), except for the FC receiver, which requires  $N_r$  demodulator chains.

Figure 2 plots the beam-patterns for a ULA (as a function of the azimuth angle) formed by the FFT and  $\mathbf{M}_{\text{TI-S}}$  pre-processing matrices. We can clearly see from Fig. 2(a) that FFT pre-processing gives high gains for rays with AoAs  $0^\circ$ ,  $60^\circ$ ,  $90^\circ$ ,  $270^\circ$ , and  $300^\circ$ ; however, the gains are lower for other directions. Figures 2(b) and 2(c) plot the beam-patterns of  $\mathbf{M}_{\text{TI-S}}$  for AoAs  $45^\circ$  and  $60^\circ$ , respectively. We see that, unlike FFT, the beam-pattern of  $\mathbf{M}_{\text{TI-S}}$  adapts to the mean AoA at the receiver and ensures good gains in both cases.

Table 1 compares the ergodic capacity of different receiver architectures. We see that the receivers employing pre-processing outperform conventional antenna selection. Also note that the phase approximation to TI is very accurate – TI-Ph comes within 0.1 bits/s/Hz of TI. We also plot the cumulative distribution function (CDF) of the capacity. The CDF provides a complete characterization of the capacity, as opposed to the expected value, which provides information on the first moment only.

Figure 3 plots the CDF of capacity for  $\sigma_r = 6^\circ$  and  $\theta_r = 45^\circ$ . We observe that for  $L = 1$ , TI, TI-S and TV achieve a 1.0 bits/s/Hz gain over pure antenna selection. Since the mean AoA falls on the minima of FFT beam-pattern, it performs 0.5 bits/s/Hz worse than TI. With  $L = 2$ , while con-

ventional antenna selection capacity is 1.0 bits/s/Hz less than FC, the receivers with pre-processing achieve performance parity with FC. Note that while FFT performs worse than TI for  $L = 1$ , it achieves the same performance as TI for  $L = 2$ .<sup>3</sup>

The effect of different angle spreads is studied in Fig. 4. As  $\sigma_r$  increases, the performance of pure antenna selection improves, while that of TI, TI-S and FFT degrades. In the extreme case of an uncorrelated channel, antenna selection, FFT and TI-S will have the same performance, while TI will be slightly worse. This is a consequence of the fact that efficacy of the statistics-based solutions reduces as the channel correlation reduces. However, the time-variant receiver (not shown in the figure) will perform better than pure antenna selection because it can adapt to the instantaneous channel state.

#### V. CONCLUSIONS

In this paper, a novel RF pre-processing solution was proposed to recover the beamforming gain lost by conventional antenna selection. This architecture introduces a RF pre-processing matrix, followed by selection, if required. The optimal time-invariant  $L \times N_r$  RF pre-processing matrix,  $\mathbf{M}_{\text{TI}}$ , projects the received vector along the  $L$  largest eigenvectors of the receiver correlation matrix. By its very design, the  $L \times N_r$  matrix eliminates the need for subsequent antenna selection and its associated computational complexity. For a correlated channel, such pre-processing significantly outperformed antenna selection, despite using the same number of demodulator chains as antenna selection. A more sophisticated receiver architecture that uses an  $N_r \times N_r$  RF matrix followed by antenna selection was also investigated. At the expense of additional computations, this architecture outperformed the  $L \times N_r$  matrix discussed above and performed significantly better than pure antenna selection. Simulation results showed that in some cases, RF pre-processing achieved performance gains close to that of an FC receiver. We showed that implementing the pre-processor using low complexity RF phase-shifters incurs a negligible loss in performance. A beam-pattern-based intuition was also developed. We saw that the statistics-based solutions can adapt the beam-pattern to mean AoA direction.

While only receiver design was considered in this paper, antenna selection at the transmitter can be implemented in an analogous manner.

#### APPENDIX A-1: PROOF OF LEMMA 1

Let  $\mathbf{\Lambda}_1 = \text{diag}(\lambda_{11}, \dots, \lambda_{1N_r})$  and  $\mathbf{\Lambda}_\alpha = \text{diag}(\lambda_{11} + \alpha, \lambda_{12}, \dots, \lambda_{1N_r})$ , where  $\alpha$  is a positive constant. We first show that  $|\mathbf{I}_{N_t} + \mathbf{A}\mathbf{\Lambda}_\alpha\mathbf{A}^\dagger|$  increases monotonically with  $\alpha$ . Define  $\mathbf{\Sigma}_1 = \text{diag}(1, 0, \dots, 0)$ , and  $\mathbf{B}_1(\alpha) = \mathbf{I}_{N_t} + \mathbf{A}\mathbf{\Lambda}_\alpha\mathbf{A}^\dagger$ . Then,

$$|\mathbf{B}_1(\alpha)| = |\mathbf{I}_{N_t} + \mathbf{A}\mathbf{\Lambda}_1\mathbf{A}^\dagger + \alpha\mathbf{A}\mathbf{\Sigma}_1\mathbf{A}^\dagger|. \quad (13)$$

<sup>3</sup>This is because the gain from the two FFT beams add up and yield performance close to the full gains.

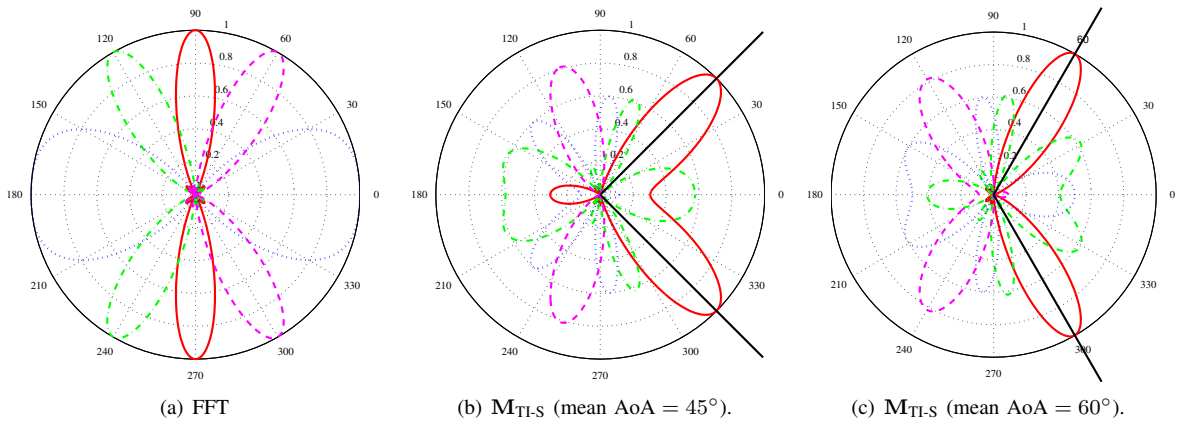


Fig. 2. Beam-patterns as a function of azimuth angles

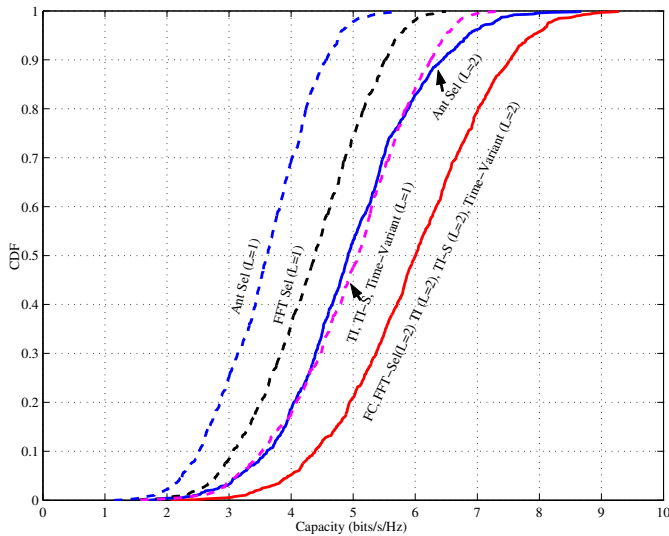


Fig. 3. CDF of capacity for spatial multiplexing as a function of  $L$  ( $N_r = 4$ ,  $\theta_r = 45^\circ$ ,  $\sigma_r = 6^\circ$ ).

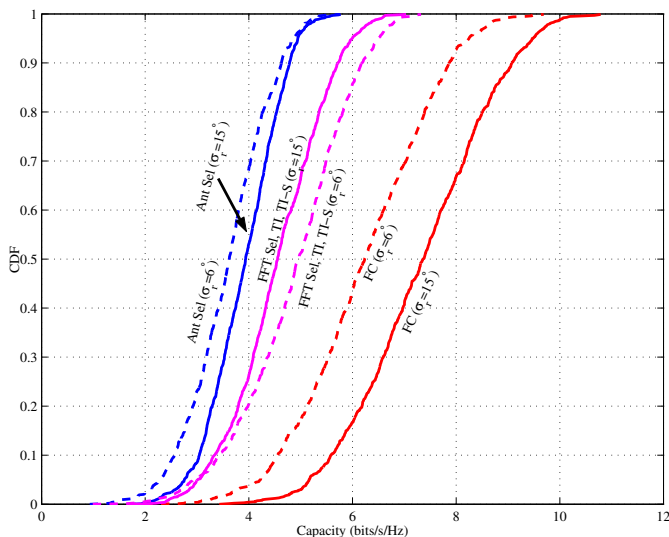


Fig. 4. Effect of spatial correlation  $\sigma_r$  on the CDF of capacity for a spatial multiplexing system ( $N_r = 4$ ,  $L = 1$ ,  $\theta_r = 60^\circ$ ).

Applying the formula for the derivatives of determinants on (13), we get

$$\begin{aligned} \frac{d|\mathbf{B}_1(\alpha)|}{d\alpha} &= |\mathbf{B}_1(\alpha)| \text{Trace} \left( \mathbf{B}_1^{-1}(\alpha) \frac{d\mathbf{B}_1(\alpha)}{d\alpha} \right), \\ &= |\mathbf{B}_1(\alpha)| \text{Trace} \left( \mathbf{B}_1^{-1}(\alpha) \mathbf{A} \Sigma_1 \mathbf{A}^\dagger \right), \\ &= |\mathbf{B}_1(\alpha)| \left( \mathbf{a}_1^\dagger \mathbf{B}_1^{-1}(\alpha) \mathbf{a}_1 \right), \end{aligned} \quad (14)$$

where  $\mathbf{a}_1$  is the first column of  $\mathbf{A}$ . Since  $\mathbf{B}_1(\alpha)$  is positive-definite,  $\frac{d|\mathbf{B}_1(\alpha)|}{d\alpha} > 0$ . Thus  $|\mathbf{B}(\alpha)|$  increases with  $\alpha$ . Similarly, we can show that  $|\mathbf{I}_{N_t} + \mathbf{A} \mathbf{\Lambda} \mathbf{A}^\dagger|$  increases as each non-zero element of the diagonal matrix  $\mathbf{\Lambda}$  increases.

#### REFERENCES

- [1] G. J. Foschini, M. J. Gans, "On the Limits of Wireless Communications in a Fading Environment When Using Multiple Antennas," *Wireless Pers. Commun.*, Vol. 6, No. 3, pp. 311–335, Mar. 1998.
- [2] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen "A stochastic MIMO radio channel model with experimental validation," *IEEE JSAC*, pp. 1211–1226, Aug. 2002.
- [3] A. F. Molisch, M. Z. Win "MIMO systems with Antenna Selection," *IEEE Microwave Mag.*, Vol. 5, No. 1, pp. 46–56, Mar. 2004.
- [4] D. A. Gore, A. Paulraj, "MIMO Antenna Subset Selection With Space-Time Coding," *IEEE Trans. Signal Processing*, Vol. 50, No. 10, pp. 2580–2588, Oct. 2002.
- [5] A. Gorokhov, D. Gore, A. Paulraj, "Receive Antenna Selection for MIMO Flat-Fading Channels: Theory and Algorithms," *IEEE Trans. Inform. Theory*, Vol. 49, No. 10, pp. 2687–2696, Oct. 2003.
- [6] M. G. Alkhanari, A. B. Gershman, "Fast Antenna Subset Selection in MIMO Systems," *IEEE Trans. Signal Processing*, Vol. 52, No. 2, pp. 339–347, Feb. 2004.
- [7] A. F. Molisch, X. Zhang, S. Y. Kung, J. Zhang, "FFT-Based Hybrid Antenna Selection Schemes for Spatially Correlated MIMO Channels," *IEEE Communication Letters* Vol. 8, Issue 1, pp. 36–38, Jan. 2004.
- [8] X. Zhang, A. F. Molisch, S. Y. Kung, "Phase-shift-based antenna selection for MIMO channels" in *Proc. Globecom*, pp. 1089–1093, 2003.
- [9] P. Sudarshan, N. B. Mehta, A. F. Molisch, J. Zhang "Channel Statistics-Based RF Pre-Processing with Antenna Selection," *submitted to IEEE Trans. Wireless Comm.* 2004.
- [10] D. Asztely, "On Antenna Arrays in Mobile Communication Systems: Fast Fading and GSM Base Station Receiver Algorithms," *Technical Report IR-S3-SB-9611*, Royal Institute of Technology, Stockholm, Sweden, Mar. 1996.
- [11] D. Hand, H. Mannila, P. Smyth, "Principles of Data Mining," *The MIT Press*, Cambridge, MA, 2001.
- [12] T. Ohira, "Analog Smart Antennas: An Overview," *Int. Symposium on Personal, Indoor and Mobile Radio Commun.*, vol. 4, pp. 1502–1506, Lisboa, Portugal, Sep. 2002.
- [13] R. A. Horn, C. R. Johnson, "Topics in Matrix Analysis," *Cambridge Univ. Press*, Cambridge U. K., 1991.