

# MIMO Systems with Arbitrary Antenna Array Architectures: Channel Modeling, Capacity and Low-Complexity Signaling

Vasanthan Raghavan, Ada S. Y. Poon, Venugopal V. Veeravalli  
 University of Illinois, Urbana-Champaign, Urbana IL 61801, USA  
 E-mail: {vasanth, poon, vvv}@uiuc.edu

**Abstract**—The focus of this work is on multi-antenna systems with arbitrary antenna array architectures. First, we propose a multi-antenna channel model that encompasses many of the existing models and study the capacity of such systems. We show that capacity is achieved by pre-nulling the input signals along the transmit array manifold with a transmit-SNR dependent rank and power control. The rank of the optimal signaling scheme is monotonically non-decreasing as SNR increases. Thus our results generalize many known results in this area. Further, we compute two explicit SNR values,  $\rho_{low}$  below which rank-1 signaling and  $\rho_{high}$  above which full-rank signaling are near-optimal, respectively. Finally, with a focus on low-complexity scalar signaling schemes, we propose a sub-optimal beamforming approach that minimizes the statistical feedback necessary to implement such schemes. While almost all works in this area assume a genie-aided statistics acquisition process, we show that the proposed scheme trades-off performance with statistical feedback overhead attractively.

## I. INTRODUCTION

It is well-known that multiple antennas at the transmitter and the receiver can lead to substantial capacity gains over single antenna communication systems. While initial results in this area focussed on rich multipath modeled by channels with independent and identically distributed (i.i.d.) entries, realizing such gains in practice is critically dependent on realistic channel modeling followed by a performance analysis with low-complexity signaling schemes. We refer the reader to [1] for a summary of work in this area.

## II. SYSTEM MODEL

Our focus in this paper is on a multi-antenna communication system with  $N_t$  transmitters and  $N_r$  receivers and the following system equation:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (1)$$

where  $\mathbf{y}$  is the  $N_r \times 1$  received vector,  $\mathbf{x}$  is the  $N_t \times 1$  transmitted vector, and  $\mathbf{n}$  is the  $N_r \times 1$  additive white Gaussian noise vector. The channel matrix  $\mathbf{H}$  can be written as

$$\mathbf{H} \stackrel{(a)}{=} \int \int \int h(t, k_r, k_t) e^{-j2\pi t W} \mathbf{a}_r(k_r) \mathbf{a}_t(k_t)^H dt dk_r dk_t$$

$$\stackrel{(b)}{=} \sum_i \rho_i e^{-j2\pi \tau_i W} \mathbf{a}_r(k_{r,\ell_i}) \mathbf{a}_t(k_{t,\ell_i})^H$$

where  $W$  is the signaling bandwidth,  $\mathbf{a}_r(k_r)$  and  $\mathbf{a}_t(k_t)$  denote the  $N_r \times 1$  receive and  $N_t \times 1$  transmit array manifold, and  $h(t, k_r, k_t)$  corresponds to the complex gain between transmit direction  $k_t$  and receive direction  $k_r$  (both directions measured with respect to the horizontal axes without any loss in generality) at time  $t$  due to an impulse applied at time 0. In (a), we have used the continuous channel representation

and (b) follows from the following discrete representation to  $h(t, k_r, k_t)$ :

$$h(t, k_r, k_t) = \sum_{i=1}^L \rho_i \delta(t - \tau_i) \delta(k_r - k_{r,\ell_i}) \delta(k_t - k_{t,\ell_i})$$

where  $\rho_i, \tau_i, k_{r,\ell_i}, k_{t,\ell_i}$  correspond to the complex gain, delay, direction of arrival, and direction of departure of the  $i$ -th resolvable physical path. The number of such resolvable paths is assumed to be  $L$ . In wireless applications under the non line-of-sight setting, the channel gains rapidly change in time and frequency due to fading and with the widely used *wide-sense stationary uncorrelated scattering (WSSUS)* assumption, it is reasonable to model the complex gains of the resolvable physical paths as independent *complex normal* random variables with zero mean.

It can be shown that [2], irrespective of the antenna array geometry and spacing, there exist  $N_t$  transmitter directions,  $k_{t_i}, i = 1, \dots, N_t$  and  $N_r$  receiver directions,  $k_{r_i}, i = 1, \dots, N_r$  such that  $\mathbf{H}$  can be efficiently approximated as

$$\mathbf{H} \approx \sum_{i=1}^{N_r} \sum_{j=1}^{N_t} \mathbf{H}_{ang}[i, j] \mathbf{a}_r(k_{r_i}) \mathbf{a}_t(k_{t_j})^H$$

$$= \mathbf{A}_r \mathbf{H}_{ang} \mathbf{A}_t^H \quad (2)$$

where  $\mathbf{H}_{ang} = \{\mathbf{H}_{ang}[i, j]\}$  denotes the channel in the *angular* domain with  $\mathbf{H}_{ang}[i, j]$  approximately independent and complex normal with zero mean (because the underlying  $h(t, k_r, k_t)$  are complex normal with zero mean). Thus, the statistics of  $\mathbf{H}_{ang}$  are determined by the second moment of its entries.  $\mathbf{A}_t$  and  $\mathbf{A}_r$  are spatial array matrices defined as

$$\mathbf{A}_t = [\mathbf{a}_t(k_{t_1}), \dots, \mathbf{a}_t(k_{t_{N_t}})],$$

$$\mathbf{A}_r = [\mathbf{a}_r(k_{r_1}), \dots, \mathbf{a}_r(k_{r_{N_r}})].$$

While it may be possible to perform a *Karhunen-Loeve type decomposition of the channel  $\mathbf{H}$  as in (2) with a scattering environment-dependent choice of  $\mathbf{A}_t$  and  $\mathbf{A}_r$* , it is crucial to note that the choices of  $\mathbf{A}_t$  and  $\mathbf{A}_r$  proposed in (2) are in fact independent of the scattering environment and are dependent only on the antenna array geometry and spacing. Moreover, the choices of  $k_{t_i}$  and  $k_{r_i}$  are such that the matrices  $\mathbf{A}_t$  and  $\mathbf{A}_r$  are non-singular.

We now illustrate the connections between the model in (2) and the many stochastic MIMO channel models that have been proposed in the literature. First, note that in the case of uniform linear arrays (ULAs) where with a choice of  $k_{t_i}$  and  $k_{r_i}$  that leads to uniform sampling in the  $\sin(k_t)$  and  $\sin(k_r)$  domains, respectively,  $\mathbf{A}_t$  and  $\mathbf{A}_r$  are the Fourier unitary matrices and

we arrive at what is known as the *virtual representation* [3]. More generally, if  $\mathbf{A}_t$  and  $\mathbf{A}_r$  are only assumed to be unitary, we arrive at what is called as the *canonical modeling framework* [4] (also called the *beam-space approach* in [5], [6]). When the discrete model in (2) is generalized, we arrive at the *continuous modeling framework* of [7]. On the other hand, with  $\mathbf{A}_t$  and  $\mathbf{A}_r$  unitary and restrictions on the variances of  $\mathbf{H}_{\text{ang}}[i, j]$ , we arrive at two well-studied channel models: 1) A separability property leads to the *Kronecker model*, that is,  $\sigma_{ij}^2 = \mathbf{E}[|\mathbf{H}_{\text{ang}}[i, j]|^2] = \Lambda_r[i]\Lambda_t[j]$  for some  $\Lambda_t$  and  $\Lambda_r$  diagonal, positive semi-definite, and 2) The *i.i.d. model* corresponds to assuming  $\sigma_{ij}^2 = 1$  for all  $i, j$ . Thus, we see that the various channel models that have been proposed in the literature can be incorporated as special cases of (2).

While  $\mathbf{A}_t$  and  $\mathbf{A}_r$  turn out to be unitary in the case of ULAs, for a general array geometry and spacing, it is *not necessary* that the columns of  $\mathbf{A}_t$  and  $\mathbf{A}_r$  are either orthogonal or even that they are normalized to 1. In this general case, the only condition that has to be met by  $\mathbf{A}_t$  and  $\mathbf{A}_r$  are that they are non-singular. Perfect knowledge of  $\mathbf{A}_t$  and  $\mathbf{A}_r$  are possible since these are scattering environment independent and can hence be learnt reliably at both the ends. On the other hand, assumption about the knowledge of  $\{\sigma_{ij}^2 = \mathbf{E}[|\mathbf{H}_{\text{ang}}[i, j]|^2]\}$  is reasonable if the channel statistics remain fixed for a sufficiently long period of time so that this acquisition cost can be neglected. In the first part of this work, we will assume knowledge of  $\sigma_{ij}^2$  at the transmitter while in the latter part, we will explicitly incorporate the statistical feedback cost.

### III. OPTIMAL SIGNALING

The capacity  $C$  of the MIMO system with an array architecture characterized by the spatial array matrices  $\{\mathbf{A}_r, \mathbf{A}_t\}$  is

$$\begin{aligned} C &= \sup_{\mathbf{Q} \in \mathcal{Q}} \mathbf{E} \left[ \log_2 \det \left( \mathbf{I}_{N_r} + \mathbf{A}_r \mathbf{H}_{\text{ang}} \mathbf{A}_t^H \mathbf{Q} \mathbf{A}_t \mathbf{H}_{\text{ang}}^H \mathbf{A}_r^H \right) \right] \\ &= \sup_{\mathbf{Q} \in \mathcal{Q}} \mathbf{E} \left[ \log_2 \det \left( (\mathbf{A}_r^H \mathbf{A}_r)^{-1} + \mathbf{H}_{\text{ang}} \mathbf{A}_t^H \mathbf{Q} \mathbf{A}_t \mathbf{H}_{\text{ang}}^H \right) \right] \\ &\quad + \log_2 \det(\mathbf{A}_r^H \mathbf{A}_r). \end{aligned} \quad (3)$$

In (3), the optimization is over the set  $\mathcal{Q}$  defined as  $\mathcal{Q} = \{\mathbf{Q} : \mathbf{Q} \succ \mathbf{0}, \text{Tr}(\mathbf{Q}) \leq \rho\}$ , the set of positive semi-definite input covariance matrices  $\mathbf{Q}$  that satisfy a power constraint of  $\rho$ . The following theorem addresses the question of the optimal input covariance matrix in (3).

*Theorem 1:* Given that the transmitter knows  $\mathbf{A}_t$ ,  $\mathbf{A}_r$  and the statistics of  $\mathbf{H}_{\text{ang}}$ , the optimal input covariance matrix is given by

$$\mathbf{Q}_{\text{opt}} = \mathbf{A}_t^{-H} \mathbf{D}_{\text{opt}} \mathbf{A}_t^{-1} \quad (4)$$

where  $\mathbf{D}_{\text{opt}}$  is a positive semi-definite, diagonal matrix that solves the following convex optimization problem:

$$\mathbf{D}_{\text{opt}} = \arg \max_{\mathbf{D}} \mathbf{E} \left[ \log_2 \det \left( (\mathbf{A}_r^H \mathbf{A}_r)^{-1} + \mathbf{H}_{\text{ang}} \mathbf{D} \mathbf{H}_{\text{ang}}^H \right) \right] \quad (5)$$

such that  $\text{Tr}(\mathbf{D}(\mathbf{A}_t^H \mathbf{A}_t)^{-1}) \leq \rho$ .

*Proof:* With a transformation  $\widehat{\mathbf{Q}} = \mathbf{A}_t^H \mathbf{Q} \mathbf{A}_t$ ,  $C$  can be rewritten as

$$C = \sup_{\widehat{\mathbf{Q}}} \mathbf{E} \left[ \log_2 \det \left( \mathbf{I}_{N_r} + \mathbf{H}_{\text{ang}} \widehat{\mathbf{Q}} \mathbf{H}_{\text{ang}}^H \mathbf{A}_r^H \mathbf{A}_r \right) \right] \quad (6)$$

where the set  $\widehat{\mathcal{Q}}$  is defined as  $\widehat{\mathcal{Q}} = \{\widehat{\mathbf{Q}} : \widehat{\mathbf{Q}} \succ \mathbf{0}, \text{Tr}(\widehat{\mathbf{Q}}(\mathbf{A}_t^H \mathbf{A}_t)^{-1}) \leq \rho\}$ . Note that for any pair of positive definite matrices  $\mathbf{X}$  and  $\mathbf{Z}$ , the function  $f(\widehat{\mathbf{Q}}) = \log_2 \det(\mathbf{X} + \mathbf{Y} \widehat{\mathbf{Q}} \mathbf{Y}^H)$  is concave over the convex set  $A$  defined as  $A = \{\widehat{\mathbf{Q}} : \text{Tr}(\widehat{\mathbf{Q}} \mathbf{Z}) \leq \rho\}$ .

Construct two sets  $\Omega$  and  $\Lambda$  of positive semi-definite matrices and positive semi-definite, diagonal matrices:

$$\begin{aligned} \Omega &= \{\widehat{\mathbf{Q}} : \widehat{\mathbf{Q}} \succ \mathbf{0}, \text{Tr}(\widehat{\mathbf{Q}}(\mathbf{A}_t^H \mathbf{A}_t)^{-1}) \leq \rho\}, \\ \Lambda &= \{\mathbf{D} : \mathbf{D} \succ \mathbf{0}, \text{Tr}(\mathbf{D}(\mathbf{A}_t^H \mathbf{A}_t)^{-1}) \leq \rho\}, \end{aligned}$$

respectively. Following identical lines to the proof in [8, Theorem 1], it can be shown that the global optimum of  $\mathbf{E} \left[ \log_2 \det \left( \mathbf{I}_{N_r} + \mathbf{H}_{\text{ang}} \widehat{\mathbf{Q}} \mathbf{H}_{\text{ang}}^H \mathbf{A}_r^H \mathbf{A}_r \right) \right]$  over  $\Omega$  is met with a local optimum over  $\Lambda$ . Since this result follows in a straightforward way from [8], we do not provide the details. ■

*Eigen-Decomposition of  $\mathbf{Q}_{\text{opt}}$ :* In the special case where  $\mathbf{A}_t$  is unitary, from (4)  $\mathbf{Q}_{\text{opt}}$  reduces to  $\mathbf{Q}_{\text{opt}} = \mathbf{A}_t \mathbf{D}_{\text{opt}} \mathbf{A}_t^H$  as shown in [8]. In the more general case, the following interpretation of  $\mathbf{Q}_{\text{opt}}$  is useful. Let the singular value decomposition (SVD) of  $\mathbf{A}_t$  be  $\mathbf{A}_t = \mathbf{U}_t \mathbf{\Lambda}_t \mathbf{V}_t^H$ . Then,  $\mathbf{Q}_{\text{opt}}$  can be written as

$$\begin{aligned} \mathbf{Q}_{\text{opt}} &= \mathbf{U}_t \left( \mathbf{\Lambda}_t^{-1} \mathbf{V}_t^H \mathbf{D}_{\text{opt}} \mathbf{V}_t \mathbf{\Lambda}_t^{-1} \right) \mathbf{U}_t^H \quad (7) \\ &= \mathbf{U}_t \widetilde{\mathbf{U}} \widetilde{\mathbf{\Lambda}} \widetilde{\mathbf{U}}^H \mathbf{U}_t^H \end{aligned}$$

where we have assumed an SVD of the form  $\widetilde{\mathbf{U}} \widetilde{\mathbf{\Lambda}} \widetilde{\mathbf{U}}^H$  for the matrix in parenthesis in (7). Thus, the eigenvector matrix of  $\mathbf{Q}_{\text{opt}}$  is a composition of two matrices: the left singular vector matrix of  $\mathbf{A}_t$  followed by a skewing with  $\widetilde{\mathbf{U}}$  to correct for the non-ULA array geometry. In the ULA case, check that  $\mathbf{U}_t \widetilde{\mathbf{U}}$  reduces to  $\mathbf{A}_t$  and  $\widetilde{\mathbf{\Lambda}}$  reduces to  $\mathbf{D}_{\text{opt}}$ .

*Low- and High-SNR Extremes:* The optimization problem in (5) is a very standard problem: maximizing a concave function over a convex set and hence many of the standard convex optimization procedures are applicable. In the special case of ULAs, it is known that in the low-SNR extreme, beamforming to the statistically dominant eigen-direction is capacity optimal while in the high-SNR extreme, uniform power signaling across the rank of  $\mathbf{H}_{\text{ang}}$  is capacity optimal. More generally, in the intermediate-SNR regime, the rank of  $\mathbf{Q}_{\text{opt}}$  is non-decreasing as  $\rho$  increases. We now consider these problems for the more general channel model.

Consider signaling with an input covariance matrix of the form  $\mathbf{Q} = \mathbf{A}_t^{-H} \mathbf{D} \mathbf{A}_t^{-1}$  with  $\mathbf{D}$  diagonal subject to the power constraint  $\sum_k \mathbf{D}[j](\mathbf{A}_t^H \mathbf{A}_t)^{-1}[j] \leq \rho$ . We begin with the low-SNR analog.

*Proposition 1:* Let  $\sigma_{ij}^2$  denote the variance of  $\mathbf{H}_{\text{ang}}[i, j]$  and define  $j^*$  as

$$j^* = \arg \max_j \frac{\sum_i \sigma_{ij}^2 (\sum_k |\mathbf{A}_r[k, i]|^2)}{(\mathbf{A}_t^H \mathbf{A}_t)^{-1}[j]}. \quad (8)$$

Further, define

$$\rho_{\text{low}} \triangleq \frac{(\mathbf{A}_t^H \mathbf{A}_t)^{-1} [j^*]}{\sum_i \sigma_{i j^*}^2 (\sum_k |\mathbf{A}_\tau[k, i]|^2)}. \quad (9)$$

Also, let  $I_{\text{br}}(\rho)$  and  $C(\rho)$  denote the average mutual information achieved by beamforming (with  $\mathbf{Q} = \mathbf{A}_t^{-H} \mathbf{D} \mathbf{A}_t^{-1}$  where  $\mathbf{D}$  has only the diagonal entry in the  $j^*$ -position non-zero) and the capacity at an SNR of  $\rho$ . For any  $\rho < \rho_{\text{low}}$  and  $x = \frac{\rho}{\rho_{\text{low}}}$ , we have

$$C(\rho) \leq \log_2(e) \cdot x, \quad I_{\text{br}}(\rho) \geq \log_2(e) \cdot (x - 2x^2). \quad (10)$$

Thus, in the low-SNR extreme (that is, as  $\rho \rightarrow 0$  and hence,  $x \rightarrow 0$ ),

$$\frac{C(\rho) - I_{\text{br}}(\rho)}{I_{\text{br}}(\rho)} \rightarrow 0. \quad (11)$$

In other words, beamforming along  $k^*$  is optimal from an ergodic capacity view-point.

*Proof:* No proof is provided here due to lack of space. The readers are referred to [2] for details. ■

Note that for the ULA case where  $\mathbf{A}_t$  and  $\mathbf{A}_\tau$  are unitary,  $j^*$  reduces to  $\arg \max_j \sum_i \sigma_{i j}^2$  and  $\mathbf{Q}$  to  $\mathbf{A}_t \mathbf{D} \mathbf{A}_t^H$ , the optimal transmit eigen-direction. This result is well-known in MIMO literature; see [1], [8] etc. for more details. Also, under the assumption that  $\sigma_{i k^*} = 1$  for all  $i$ ,  $\rho_{\text{low}}$  turns out to be  $\frac{1}{N_r}$ . This characterization of  $\rho_{\text{low}}$  in the unitary  $\{\mathbf{A}_\tau, \mathbf{A}_t\}$  case is also known from prior results on MIMO capacity; see [9] which shows this up to an  $\mathcal{O}(1)$  term and [4] which shows this using more precise results on tail probabilities of sums of independent random variables.

We now consider the high-SNR extreme and show that uniform power signaling is optimal.

*Proposition 2:* Let  $\mathcal{R}$  denote the set of indices  $\mathcal{R} = \{j : \sum_i \sigma_{i j}^2 (\sum_k |\mathbf{A}_\tau[k, i]|^2) > 0\}$ . Given any  $\epsilon > 0$  and for any  $\rho > \rho_{\text{high}}$  where  $\log(\rho_{\text{high}}) \triangleq \frac{\log(|\mathcal{R}|)}{\epsilon}$ , the difference in average mutual information between the optimal scheme (denoted by  $C(\rho)$ ) and the uniform power signaling scheme across  $\mathcal{R}$  (denoted by  $I_{\text{eq}}(\rho)$ ) can be bounded as

$$\frac{|C(\rho) - I_{\text{eq}}(\rho)|}{I_{\text{eq}}(\rho)} \leq \epsilon.$$

Thus, uniform power signaling is optimal from an ergodic capacity viewpoint in the high-SNR extreme. ■

With  $\mathbf{A}_\tau$  unitary,  $\mathcal{R}$  corresponds to  $\mathcal{R} = \{j : \sum_i \sigma_{i j}^2 > 0\}$ , that is, the essential rank of  $\mathbf{H}_{\text{ang}}$  as seen from the transmitter. This result is well-known from the work in [8]. Thus, Propositions 1 and 2 provide the natural extension of these results to arbitrary array geometries. Further, in the unitary case,  $\rho_{\text{high}}$  has also been characterized in [9] up to an  $\mathcal{O}(1)$  factor.

#### IV. SUB-OPTIMAL BEAMFORMING STRATEGIES

In this section, we assume that we operate in the low-SNR regime,  $\rho < \rho_{\text{low}}$ . Thus, the focus here is on the optimal beamforming scheme. Further, for any SNR,  $\rho \in (\rho_{\text{low}}, \rho_{\text{high}})$ , while the rank of the optimal signaling scheme satisfies  $1 <$

$\text{rank}(\mathbf{D}_{\text{opt}}) < N_t$ , from an (implementation) complexity point-of-view, a scalar scheme like beamforming is preferred. In fact, this is a principal reason why there is a significant interest in both theory as well as practice on beamforming. We revisit this scheme and establish connections between beamforming in the ULAs case and the non-ULAs case. The following convention is followed: If  $\frac{I(\rho)}{\rho} = K + \mathcal{O}(\rho)$  for some  $K$  finite, we will restrict to writing  $I(\rho) = K\rho$  with an understanding that the focus is on the dominant term in the low-SNR expansion of average mutual information.

**Optimality of Beamforming Along Eigenvectors of  $\mathbf{R}_t$ :** Consider the case where  $\mathbf{A}_t$  and  $\mathbf{A}_\tau$  are unitary. As noted before, in this case  $\mathbf{Q}_{\text{opt}}$  reduces to  $\mathbf{A}_t \mathbf{D}_{\text{opt}} \mathbf{A}_t^H$  (with the beamforming direction  $j^* = \arg \max_j \sum_i \sigma_{i j}^2$ ). The transmit covariance matrix  $\mathbf{R}_t$  is defined as  $\mathbf{R}_t \triangleq \mathbf{E}[\mathbf{H}^H \mathbf{H}]$  and in this setting  $\mathbf{R}_t$  turns out to be  $\mathbf{A}_t \mathbf{E}[\mathbf{H}_{\text{ang}}^H \mathbf{H}_{\text{ang}}] \mathbf{A}_t^H = \mathbf{A}_t \mathbf{D}_t \mathbf{A}_t^H$  where  $\mathbf{D}_t[j] = \sum_i \sigma_{i j}^2$ . Note that the eigenvectors of the optimal input covariance matrix are also the eigenvectors of  $\mathbf{R}_t$ . It is natural to question if this correspondence also holds in the non-ULAs case. Towards answering this question, note that a similar relationship holds in the general case:

$$\begin{aligned} \mathbf{R}_t &= \mathbf{A}_t \mathbf{D}_t \mathbf{A}_t^H \\ &= \sum_{j=1}^{N_t} \mathbf{D}_t[j] \mathbf{a}_t(k_{t_j}) \mathbf{a}_t^H(k_{t_j}) \\ \mathbf{D}_t[j] &= \sum_i \sigma_{i j}^2 \sum_k |\mathbf{A}_\tau[k, i]|^2. \end{aligned} \quad (12)$$

Using the SVD of  $\mathbf{A}_t (= \mathbf{U}_t \mathbf{\Lambda}_t \mathbf{V}_t^H)$ , we have

$$\begin{aligned} \mathbf{R}_t &= \mathbf{U}_t (\mathbf{\Lambda}_t \mathbf{V}_t^H \mathbf{D}_t \mathbf{V}_t \mathbf{\Lambda}_t) \mathbf{U}_t^H \\ &= \mathbf{U}_t \tilde{\mathbf{U}} \tilde{\mathbf{\Lambda}} \tilde{\mathbf{U}}^H \mathbf{U}_t^H \end{aligned} \quad (13)$$

where we assume an SVD of the form  $\tilde{\mathbf{U}} \tilde{\mathbf{\Lambda}} \tilde{\mathbf{U}}^H$  for the matrix in parenthesis. From (4), we see that

$$\begin{aligned} \mathbf{Q}_{\text{opt}} &= \mathbf{A}_t^{-H} \mathbf{D}_{\text{opt}} \mathbf{A}_t^{-1} \\ &= \mathbf{U}_t (\mathbf{\Lambda}_t^{-1} \mathbf{V}_t^H \mathbf{D}_{\text{opt}} \mathbf{V}_t \mathbf{\Lambda}_t^{-1}) \mathbf{U}_t^H \\ &= \mathbf{U}_t \tilde{\tilde{\mathbf{U}}} \tilde{\tilde{\mathbf{\Lambda}}} \tilde{\tilde{\mathbf{U}}}^H \mathbf{U}_t^H \end{aligned} \quad (14)$$

where an SVD of the form  $\tilde{\tilde{\mathbf{U}}} \tilde{\tilde{\mathbf{\Lambda}}} \tilde{\tilde{\mathbf{U}}}^H$  is assumed for the matrix in the parenthesis. From (13) and (14), we thus see that, in general the eigenvectors of  $\mathbf{Q}_{\text{opt}}$  and  $\mathbf{R}_t$  are not the same. In other words, beamforming along the eigenvectors of  $\mathbf{R}_t$  is strictly sub-optimal from a capacity viewpoint in the non-ULA case. The following proposition captures the loss in performance by pursuing this sub-optimal scheme. For this, we need the following definition.

*Definition 1:* Let  $\mathbf{B}$  be an  $n \times n$  Hermitian matrix with

$$\delta_{\mathbf{B}, j} = \frac{\left( \sum_{k \neq j} |\mathbf{B}[j, k]| \right)^{1-\alpha} \left( \sum_{k \neq j} |\mathbf{B}[k, j]| \right)^\alpha}{|\mathbf{B}[j]|}.$$

Then,  $\mathbf{B}$  is said to satisfy the *Ostrowskii condition* for some  $\alpha \in (0, 1)$  if

$$\Delta_{\mathbf{B}} \triangleq \max_{j=1, \dots, n} \delta_{\mathbf{B}, j} < 1. \quad (15)$$

The Ostrowskii condition is a measure of the diagonal dominance of  $\mathbf{B}$ . Note that the extreme cases of  $\alpha = 0$  and  $\alpha = 1$  correspond to whether the diagonal entries dominate the off-diagonal entries in the row or the column corresponding to that diagonal entry. The general case of  $\alpha \in (0, 1)$  captures diagonal dominance over the weighted off-diagonal entries.

*Proposition 3:* Let the average mutual information of the above sub-optimal scheme be  $I_{\mathbf{R}_t, \text{bf}}(\rho)$  and consider the low-SNR regime,  $\rho < \rho_{\text{low}}$  where  $\rho_{\text{low}}$  is defined in (9). If  $\mathbf{A}_t^H \mathbf{A}_t$  satisfies the Ostrowskii condition for some  $\alpha \in (0, 1)$ , we have

$$\frac{C(\rho) - I_{\mathbf{R}_t, \text{bf}}(\rho)}{C(\rho)} \leq \frac{(\Delta_{\mathbf{A}_t^H \mathbf{A}_t})^2}{1 - (\Delta_{\mathbf{A}_t^H \mathbf{A}_t})^2}.$$

The Ostrowskii condition on  $\mathbf{A}_t^H \mathbf{A}_t$  implies the following:

$$\begin{aligned} \lambda_{\max}(\mathbf{A}_t^H \mathbf{A}_t) &\leq \max_j \left( \mathbf{A}_t^H \mathbf{A}_t[j] + \delta_{\mathbf{A}_t^H \mathbf{A}_t, j} \right) \\ &\leq 2 \max_j \mathbf{A}_t^H \mathbf{A}_t[j], \\ \lambda_{\min}(\mathbf{A}_t^H \mathbf{A}_t) &\geq \min_j \left( \mathbf{A}_t^H \mathbf{A}_t[j] - \delta_{\mathbf{A}_t^H \mathbf{A}_t, j} \right) > 0. \end{aligned}$$

The above conditions imply that  $\mathbf{A}_t$  is well-conditioned and hence, invertible. However, it is not very clear if well-conditioning of  $\mathbf{A}_t$  is equivalent to (or sufficient to imply) the Ostrowskii condition. But, the following general principle holds: *The more diagonally dominant  $\mathbf{A}_t^H \mathbf{A}_t$  is, smaller  $\Delta_{\mathbf{A}_t^H \mathbf{A}_t}$  is and the closer the performance of the sub-optimal scheme when compared with the optimal.* For example, in the special case where  $\mathbf{A}_t$  is unitary, as pointed out earlier, both the schemes coincide and similar conclusions can be inferred from Prop. 3.

**Feedback of Statistical Information:** Most works in the MIMO literature assume that the channel statistics change at a sufficiently slow rate so that they can be acquired at the transmitter with negligible cost. While this may be reasonable in certain low mobility situations, in general, the receiver has to estimate statistical information and feed back the same to the transmitter. We now account for the feedback cost of statistical channel information that is necessary to perform the low-SNR optimal beamforming scheme.

The transmit/receive array manifold information can be obtained at the transmitter/receiver, respectively via deterministic techniques upon knowledge of antenna geometry and spacing or long-term averaging techniques that extract the array information by averaging out the imprint of the scattering environment in the short-term channel statistics. Knowledge of  $\mathbf{A}_t$  and  $\mathbf{A}_r$  at the transmitter and the receiver, respectively implies that determining the optimal beamforming direction requires knowledge of only  $\{\mathbf{D}_t[j] = \sum_i \sigma_{ij}^2 \sum_k |\mathbf{A}_r[k, i]|^2\}$  at the transmitter. Using its knowledge of  $\mathbf{A}_t$ , the transmitter then computes  $j^*$  as in (8). Thus, optimal beamforming requires feedback of  $N_t$  positive numbers,  $\{\mathbf{D}_t[j]\}$  that are a function of the short-term channel statistics.

In general, accurate information of  $\{\mathbf{D}_t[j]\}$  is not possible at the transmitter. We will assume that error-free estimates

$\widehat{\mathbf{D}}_t[j]$  from a  $B$ -bit quantization scheme are available at the transmitter. The transmitter can use this quantized information to compute the optimal beamforming direction akin to (8). Alternatively, as described before, the transmitter could estimate  $\mathbf{R}_t$  as

$$\widehat{\mathbf{R}}_t = \sum_{j=1}^{N_t} \widehat{\mathbf{D}}_t[j] \mathbf{a}_t(k_{t_j}) \mathbf{a}_t^H(k_{t_j}). \quad (16)$$

Then, the transmitter can signal sub-optimally along the dominant eigenvector of  $\widehat{\mathbf{R}}_t$ . In the discussion that follows, we study both these schemes more carefully.

We now describe the  $B$ -bit quantization scheme for feedback of  $\{\mathbf{D}_t[j]\}$ . We assume that the transmitter and the receiver have an *a priori* knowledge of the quantization range  $[0, \sigma_{\max}^2]$ . For example, a suitable choice of  $\sigma_{\max}^2$  is

$$\sigma_{\max}^2 = c \cdot \max_{\mathcal{E}} \max_j \sum_i \sigma_{ij}^2 \sum_k |\mathbf{A}_r[k, i]|^2 \quad (17)$$

for some  $c > 1$ . The class  $\mathcal{E}$  corresponds to a representative family of scattering environments that the transmitter and the receiver encounter in a given communication scenario. The choice of  $c$  is such that for any scattering environment encountered and the associated statistics  $\{\sigma_{ij}^2\}$ , we have

$$\frac{\sigma_{\max}^2}{\sum_i \sigma_{ij}^2 \sum_k |\mathbf{A}_r[k, i]|^2} = \mathcal{O}(1). \quad (18)$$

The range  $[0, \sigma_{\max}^2]$  is quantized with  $B$ -bits and hence the length of each quantization interval is  $\frac{\sigma_{\max}^2}{2^B}$ . We neglect the error and delay associated with the feedback link. Thus, the error in estimating  $\mathbf{D}_t[j]$  for any  $j$  is at most  $\delta_{\text{err}} = \frac{\sigma_{\max}^2}{2^{B+1}}$ .

The following results characterize the average mutual information of the optimal beamforming scheme and the sub-optimal scheme, both with statistical feedback (as described above). Given any fixed scattering environment, if  $B$  is large enough to quantize the  $\{\mathbf{D}_t[j]\}$  finely, the first proposition claims no loss due to statistical feedback.

*Proposition 4:* For any given scattering environment, if  $B$  and  $\{\mathbf{D}_t[j]\}$  satisfy

$$\begin{aligned} B > \log_2 \left( \frac{1}{\min_{j \neq j^*} (\mathbf{A}_t^H \mathbf{A}_t)^{-1}[j]} + \frac{1}{(\mathbf{A}_t^H \mathbf{A}_t)^{-1}[j^*]} \right) \\ + \log_2 \left( \frac{\sigma_{\max}^2}{\frac{\mathbf{D}_t[j^*]}{(\mathbf{A}_t^H \mathbf{A}_t)^{-1}[j^*]} - \max_{j \neq j^*} \frac{\mathbf{D}_t[j]}{(\mathbf{A}_t^H \mathbf{A}_t)^{-1}[j]}} \right) - 1 \triangleq B_{\min} \end{aligned}$$

then, the optimal scheme does not suffer any loss due to statistical feedback. ■

Following along the same logic of Prop. 4, we have the following corollary.

*Corollary 1:* Let  $I_{\text{bf}, B}(\rho)$  denote the average mutual information of the optimal scheme with  $B$ -bit statistical feedback. There exist scattering environments and realizations  $\{\sigma_{ij}^2\}$  associated with them such that

$$\frac{C(\rho) - I_{\text{bf}, B}(\rho)}{C(\rho)} \approx \mathcal{O}(1) \cdot \frac{1}{2^B}. \quad (19)$$

*Proposition 5:* Let  $I_{\mathbf{R}_t, \text{bf}, B}(\rho)$  denote the average mutual information of the sub-optimal scheme with  $B$ -bit statistical feedback. Let  $\mathbf{U}_{\mathbf{R}_t}$  denote an eigenvector matrix of  $\mathbf{R}_t$  and denote by  $\mathbf{V}$  the matrix  $\mathbf{U}_{\mathbf{R}_t}^H \mathbf{A}_t$ . Then, we have

$$\begin{aligned} & \frac{I_{\mathbf{R}_t, \text{bf}}(\rho) - I_{\mathbf{R}_t, \text{bf}, B}(\rho)}{C(\rho)} \\ & \leq \frac{\lambda_{\max}(\mathbf{A}_t^H \mathbf{A}_t)}{\lambda_{\min}(\mathbf{A}_t^H \mathbf{A}_t)} \cdot \sum_{i>1} \frac{|\sum_k \mathbf{V}[i, k] (\widehat{\mathbf{D}}_t[k] - \mathbf{D}_t[k]) \mathbf{V}^*[1, k]|^2}{(\lambda_{\max}(\mathbf{R}_t) - \lambda_i(\mathbf{R}_t))^2}. \quad \blacksquare \end{aligned}$$

Combining Prop. 3-5 and Cor. 1, we have the main statement of this section.

*Theorem 2:* For any realization of the channel statistics, the relative loss in average mutual information of the sub-optimal scheme satisfies

$$\begin{aligned} & \frac{C(\rho) - I_{\mathbf{R}_t, \text{bf}, B}(\rho)}{C(\rho)} \leq \frac{(\Delta_{\mathbf{A}_t^H \mathbf{A}_t})^2}{1 - (\Delta_{\mathbf{A}_t^H \mathbf{A}_t})^2} \\ & + \frac{\lambda_{\max}(\mathbf{A}_t^H \mathbf{A}_t)}{\lambda_{\min}(\mathbf{A}_t^H \mathbf{A}_t)} \cdot \sum_{i>1} \frac{|\sum_k \mathbf{V}[i, k] (\widehat{\mathbf{D}}_t[k] - \mathbf{D}_t[k]) \mathbf{V}^*[1, k]|^2}{(\lambda_{\max}(\mathbf{R}_t) - \lambda_i(\mathbf{R}_t))^2}. \end{aligned}$$

If  $B > B_{\min}$  (which is as above), then the optimal scheme incurs no loss due to statistical feedback. However, for any fixed  $B$ , there exists channel statistics such that the relative loss in average mutual information of the sub-optimal scheme satisfies

$$\begin{aligned} \frac{C(\rho) - I_{\text{bf}, B}(\rho)}{C(\rho)} & \approx \frac{\mathcal{O}(1)}{2^B} \cdot \left( \frac{1}{\min_{j \neq j^*} (\mathbf{A}_t^H \mathbf{A}_t)^{-1}[j]} \right. \\ & \left. + \frac{1}{(\mathbf{A}_t^H \mathbf{A}_t)^{-1}[j^*]} \right). \quad \blacksquare \end{aligned}$$

Apart from the low-SNR regime, it is well-known that beamforming is capacity optimal for a considerably large SNR range if the channel  $\mathbf{H}$  has one strong eigen-mode and many other weak eigen-modes. In this context, it is worth noting that  $B_{\min}$  measures how ill-conditioned  $\mathbf{H}_{\text{ang}}$  is. The more ill-conditioned  $\mathbf{H}_{\text{ang}}$  is, the smaller  $B_{\min}$  has to be to incur negligible loss with the optimal scheme and *vice versa*. If  $\mathbf{H}_{\text{ang}}$  is reasonably well-conditioned, then  $B_{\min}$  can be rather large. However, in many practical situations, the amount of feedback  $B$  that can be afforded is usually small. Thus, the small- $B$  regime is an important (and practical) regime and will be the focus of our attention in the ensuing discussion.

Note that  $\mathbf{V}$  and  $\Delta_{\mathbf{A}_t^H \mathbf{A}_t}$  capture how close  $\mathbf{A}_t$  is to a unitary matrix. While  $\mathbf{V}$  is the projection of the eigenvectors of  $\mathbf{R}_t$  on to the transmit array manifold,  $\Delta_{\mathbf{A}_t^H \mathbf{A}_t}$  is a measure of the diagonal dominance of  $\mathbf{A}_t^H \mathbf{A}_t$ . In the special case when  $\mathbf{A}_t$  is unitary,  $\mathbf{V}$  reduces to  $\mathbf{I}$  and  $\mathbf{A}_t^H \mathbf{A}_t$  is diagonal, that is,  $\Delta_{\mathbf{A}_t^H \mathbf{A}_t}$  reduces to 0. In this setting, it is straightforward to check that, irrespective of the value of  $B$ , the upper bound to the loss term for the sub-optimal scheme reduces to zero. *That is, given that  $\mathbf{A}_t$  is known at the transmitter (and is unitary), statistical feedback has negligible impact on the sub-optimal scheme.* When  $\mathbf{A}_t$  is near-unitary,  $\mathbf{V}$  is close to the identity matrix and  $\mathbf{A}_t^H \mathbf{A}_t$  is near-diagonally dominant ( $\Delta_{\mathbf{A}_t^H \mathbf{A}_t}$  is

close to zero). Thus, in this case, the upper bound to the loss term is close to zero even for small values of  $B$ .

On the other hand, for any fixed scattering environment, depending on if  $B < B_{\min}$  (corresponding to that environment), the sub-optimal scheme *can* lead to a higher throughput than the optimal<sup>1</sup> scheme. However, more important is the behavior of the optimal scheme in response to change in channel statistics (possibly due to mobility in practical situations). Cor. 1 establishes that the low-SNR optimal beamforming scheme is very sensitive to feedback in this regime and suffers a worst-case relative distortion on the order of  $2^{-B}$ . In the small- $B$  regime, this relative loss could be significantly larger than that incurred by the sub-optimal scheme (with feedback). Thus, from a robustness perspective, the sub-optimal scheme can perform much better than the optimal scheme in practice.

The advantage of the sub-optimal scheme in the case of statistical feedback is also intuitive upon closer introspection of the structure of the two signaling schemes. Note that the optimal signaling scheme performs a transmit array (transposition and) inversion which reduces to signaling along the transmit array manifold when  $\mathbf{A}_t$  is near-unitary. In this case, even with low levels of feedback (small- $B$ ), the eigenvectors of  $\widehat{\mathbf{R}}_t$  are close to those of  $\mathbf{A}_t$ . On the other extreme, when  $\mathbf{A}_t$  is very far from being unitary, the eigenvectors of  $\widehat{\mathbf{R}}_t$  are critically dependent on the diagonal matrix  $\widehat{\mathbf{D}}_t$  and the proposed scheme has to be studied more carefully. Thus, our work illustrates the importance of revisiting the case of optimal signaling under realistic system models that are often encountered in practice.

#### ACKNOWLEDGMENT

The work of V. Raghavan and V. V. Veeravalli was supported in part by the NSF under grants CCF-0049089 and CCF 0431088 through the University of Illinois.

#### REFERENCES

- [1] A. Goldsmith, S. Jafar, N. Jindal, S. Vishwanath, "Capacity Limits of MIMO Channels," *IEEE Journ. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684-702, June 2003.
- [2] V. Raghavan, A. S. Y. Poon, V. V. Veeravalli, "Angular Domain Processing for Multi-Antenna Channels," Submitted, [Online]. Available: <http://www.ifp.uiuc.edu/~vasanth>.
- [3] A. M. Sayeed, "Deconstructing Multi-Antenna Fading Channels," *IEEE Trans. Sig. Proc.*, vol. 50, no. 10, pp. 2563-2579, Oct. 2002.
- [4] V. Raghavan, J. H. Kotecha, A. M. Sayeed, "Canonical Statistical Modeling and Capacity Analysis of Correlated MIMO Fading Channels," *To be submitted to the IEEE Trans. Inform. Theory*.
- [5] W. Weichselberger, M. Herdin, H. Ozelik, E. Bonek, "A Stochastic MIMO Channel Model With Joint Correlation of Both Link Ends," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 90-99, Jan. 2006.
- [6] H. Ozelik, M. Herdin, W. Weichselberger, J. Wallace, E. Bonek, "Deficiencies of 'Kronecker' MIMO Radio Channel Model," *Electronics Letters*, vol. 39, no. 16, pp. 1209-1210, Aug. 2003.
- [7] A. S. Y. Poon, R. W. Brodersen, D. N. C. Tse, "Degrees of Freedom in Multiple-Antenna Channels: A Signal Space Approach," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 523-536, Feb. 2005.
- [8] V. V. Veeravalli, Y. Liang, A. M. Sayeed, "Correlated MIMO Rayleigh fading channels: Capacity, Optimal Signaling, and Scaling Laws," *IEEE Trans. Inform. Th.*, June 2005.
- [9] V. Raghavan, V. V. Veeravalli, R. W. Heath, Jr., "Reduced Rank Signaling in Spatially Correlated MIMO Channels," Proc. IEEE ISIT 2007.

<sup>1</sup>Optimal and sub-optimal from a genie-aided statistics acquisition perspective.