

# Correlated MIMO Wireless Channels: Capacity, Optimal Signaling, and Asymptotics

Venugopal V. Veeravalli, *Senior Member, IEEE*, Yingbin Liang, *Student Member, IEEE*, and Akbar M. Sayeed, *Senior Member, IEEE*

**Abstract**—The capacity of the multiple-input multiple-output (MIMO) wireless channel with uniform linear arrays (ULAs) of antennas at the transmitter and receiver is investigated. It is assumed that the receiver knows the channel perfectly but that the transmitter knows only the channel statistics. The analysis is carried out using an equivalent virtual representation of the channel that is obtained via a spatial discrete Fourier transform. A key property of the virtual representation that is exploited is that the components of virtual channel matrix are approximately independent. With this approximation, the virtual representation allows for a general capacity analysis without the common simplifying assumptions of Gaussian statistics and product-form correlation (Kronecker model) for the channel matrix elements. A deterministic line-of-sight (LOS) component in the channel is also easily incorporated in much of the analysis. It is shown that in the virtual domain, the capacity-achieving input vector consists of independent zero-mean proper-complex Gaussian entries, whose variances can be computed numerically using standard convex programming algorithms based on the channel statistics. Furthermore, in the asymptotic regime of low signal-to-noise ratio (SNR), it is shown that *beamforming* along one virtual transmit angle is asymptotically optimal. Necessary and sufficient conditions for the optimality of beamforming, and the value of the corresponding optimal virtual angle, are also derived based on only the second moments of the virtual channel coefficients. Numerical results indicate that beamforming may be close to optimum even at moderate values of SNR for sparse scattering environments. Finally, the capacity is investigated in the asymptotic regime where the numbers of receive and transmit antennas go to infinity, with their ratio being kept constant. Using a result of Girko, an expression for the asymptotic capacity scaling with the number of antennas is obtained in terms of the two-dimensional spatial scattering function of the channel. This asymptotic formula for the capacity is shown to be accurate even for small numbers of transmit and receive antennas in numerical examples.

**Index Terms**—Beamforming, large random matrices, multiple-antenna channels, optimal input distribution, virtual representation.

## I. INTRODUCTION

**M**ULTIPLE-input multiple-output (MIMO) wireless systems, which use antenna arrays at the transmitter and receiver, have generated considerable interest in recent years due to their potential to provide dramatic increases in the information rates and reliability of wireless links. The information-theoretic capacity of the MIMO wireless channel has been characterized under various assumptions since the seminal works of Foschini [1] and Telatar [2]. The goal of this paper is to characterize the capacity under the most general and realistic assumptions on the channel.

In studying wireless channels, an important aspect is the availability of channel state information (CSI) at the transmitter and receiver. It is reasonable to assume that CSI is available at the receiver via training. While having CSI at the transmitter allows for better performance, this may not be possible in practice, especially in MIMO channels, due to rapid variations and limited feedback bandwidth. Nevertheless, it is reasonable to assume that the channel statistics are known at the transmitter since these statistics change over much larger time scales than the channel gains. In this paper, we assume that CSI is available at the receiver and that channel statistics are known at the transmitter. Such a channel is commonly referred to as a *coherent* channel.

The capacity of coherent MIMO channels was first analyzed in the work of Telatar [2]. The model used by Telatar was one where the channel matrix has independent and identically distributed (i.i.d.) zero-mean proper-complex Gaussian entries. Under this i.i.d. model, the optimal (capacity-achieving) input is an i.i.d. zero-mean proper-complex Gaussian vector. While the i.i.d. model facilitates analysis, it is an idealized model representing rich uniform scattering that seldom occurs in practice. It is hence of interest to study more general, realistic models where the elements of the channel matrix are correlated. To this end, some recent papers (see, e.g., [3], [4], [6], [7], [9]) have investigated the capacity and corresponding optimal input distributions for correlated proper-complex Gaussian MIMO channel models. Common to much of this work is the *product-form correlation* assumption, where the correlation between the fading of two distinct antenna pairs is the product of

Manuscript received September 15, 2003; revised November 24, 2004. This work was supported at the University of Illinois by the National Science Foundation under Grant CCR-9980616, through a subcontract with Cornell University, and by the National Science Foundation under CAREER/PECASE Award CCR-0049089; and at the University of Wisconsin by the National Science Foundation under CAREER Grant CCR-9875805 and ITR Grant CCR-0113385, and by the ONR under the YIP Grant N00014-01-1-0825. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Yokohama, Japan, June/July 2003, and at the IEEE Asilomar Conference on Signals, Systems, and Computers, Monterey, CA, November 2003.

V. V. Veeravalli and Y. Liang are with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: vvv@uiuc.edu; yliang1@uiuc.edu).

A. M. Sayeed is with the Department of Electrical and Computer Engineering, University of Wisconsin–Madison, Madison, WI 53706 USA (e-mail: akbar@engr.wisc.edu).

Communicated by D. N. C. Tse, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2005.847724

the corresponding transmit correlation and receive correlation. This correlation model is referred to as the *Kronecker* model in the literature. Unfortunately, such a correlation structure is still quite restrictive, and can only be justified in scenarios where the scattering is locally rich at either the transmitter or the receiver [10]. Furthermore, the elements of the channel matrix tend to be correlated in sparse scattering environments, precisely the scenario where the Gaussian model for the joint statistics may be not be reasonable.

Our goal in this paper is to provide a general analysis of the capacity of coherent MIMO channels without the simplifying assumptions of Gaussian statistics and Kronecker correlation for the channel matrix elements. To this end, we exploit an equivalent representation of MIMO channels with uniform linear arrays (ULAs), introduced in [11], which is obtained via a two-dimensional discrete spatial Fourier transform. This virtual representation directly relates the channel matrix to the physical scattering environment; each element of the virtual channel matrix corresponds to the effective channel gain obtained when the transmit and receive arrays are set to beamform in fixed (virtual) directions. The second moment of each entry in the virtual matrix reflects the channel power gain at the corresponding transmit–receive angle pair.

A key property of the virtual representation that we exploit is that the elements of the virtual channel matrix can be assumed to have independent entries without much loss of accuracy. There are two major consequences of this independence. First, the complex joint statistical structure of the original channel is captured succinctly in the marginals of the virtual channel coefficients. This obviously facilitates estimation of the channel at the receiver, making our assumption that the channel statistics are available at the transmitter reasonable. Secondly, and more importantly, the independence of the virtual channel coefficient greatly facilitates the analysis of the channel capacity as we will demonstrate in this paper.

We first show that the capacity-achieving input vector in the virtual domain is zero-mean proper-complex Gaussian with independent entries. The variances of these optimal inputs represent the amount of power assigned to the corresponding virtual transmit angles, and they can be obtained numerically using standard convex programming algorithms. If only one of these variances is nonzero, then the optimal transmit strategy is to beamform along the corresponding virtual angle. In the low signal-to-noise ratio (SNR) regime, we show that beamforming to the virtual transmit angle with the largest effective channel gain is asymptotically optimal. Furthermore, for arbitrary SNR levels, we provide necessary and sufficient conditions for the optimality of beamforming based on only the second moments of the virtual channel coefficients.

We then move on to study the capacity in the asymptotic regime where the numbers of transmit and receive antennas go to infinity, with their ratio being kept constant. In this regime, the quantity of interest is the asymptotic capacity normalized by the number of antennas, which characterizes the scaling of the capacity with the number of antennas. A closed-form formula for the asymptotic normalized capacity for the i.i.d. Gaussian channel model was obtained by Telatar in [2]; and in recent work by Kamath and Hughes [12], it was shown that this asymptotic

formula is extremely accurate even for small numbers of transmit and receive antennas. The asymptotic capacity analysis for the Kronecker correlation model discussed above was given in [4]. Also, in recent work [13], the asymptotic capacity of wide-band correlated channels was investigated using the virtual representation for the special case of Gaussian  $D$ -diagonal channels, where the virtual channel coefficients are equal along  $D$  leading diagonals and zero elsewhere. While the  $D$ -diagonal Gaussian channel model does not belong to the class of Kronecker models, it is clearly a special case of the model considered in this paper where the virtual channel coefficients are allowed to have fairly general statistics. In this paper, we use the virtual channel representation to obtain a general formula for the asymptotic normalized capacity, which is expressed directly in terms of the two-dimensional spatial scattering function of the channel. We also show through numerical results that this asymptotic formula is quite accurate for small numbers of antennas even when the scattering is sparse.

### A. Notation and Organization

We use the following notation. For deterministic objects, we use upper case letters for matrices, lower case letters for scalars, and underlined lower case letters for vectors. An exception is the symbol  $C$  which is a scalar that is used to denote capacity. Random objects are identified by corresponding bold-faced letters. For example, we use  $\mathbf{X}$  to denote a random matrix,  $X$  to denote the realization of  $\mathbf{X}$ ,  $\underline{\mathbf{x}}$  to denote a random vector, and  $\underline{x}$  to denote a random scalar. To indicate the entries of matrices, we use subscripts. For example,  $\mathbf{H}_{k,\ell}$  denotes the component at the  $k$ th row and  $\ell$ th column of the random matrix  $\mathbf{H}$ .

We use  $\mathcal{CN}(\mu, \sigma^2)$  to denote the proper-complex Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ , and  $\mathcal{CN}(\underline{\mu}, \underline{\Sigma})$  to denote the proper-complex Gaussian vector distribution with mean  $\underline{\mu}$  and covariance matrix  $\underline{\Sigma}$ .

We use  $\text{Tr}(\cdot)$  to denote the trace of a matrix, and  $(\cdot)^\dagger$  and  $(\cdot)^\top$  to denote the Hermitian transpose and the transpose of a matrix, respectively. The symbol  $\|\cdot\|$  denotes the Euclidean norm of a vector. The symbol  $I$  denotes the identity matrix.

The remainder of this paper is organized as follows. In Section II, we introduce the virtual representation of the channel. In Section III, we characterize the optimal input distribution in the virtual domain, and discuss techniques for computing the resulting capacity. In Section IV, we exploit the virtual representation further in analyzing the asymptotic capacity of the channel for large numbers of transmit and receive antennas. In Section V, we provide a set of numerical results that illustrate the theoretical results of the previous sections. Finally, in Section VI, we give some concluding remarks.

## II. CHANNEL MODEL AND VIRTUAL REPRESENTATION

We consider the now standard model for a frequency flat, slow-fading MIMO channel with  $n_t$  transmit and  $n_r$  receive antennas. In complex baseband, the received signal vector  $\underline{\mathbf{y}}$  corresponding to one symbol interval is given by

$$\underline{\mathbf{y}} = \sqrt{\frac{\Gamma}{n_t}} \mathbf{H} \underline{\mathbf{x}} + \underline{\mathbf{w}} \quad (1)$$

where  $\mathbf{x}$  is the  $n_t$ -dim transmit vector,  $\mathbf{H}$  is the  $n_r \times n_t$  channel matrix, and  $\mathbf{w}$  is complex additive white Gaussian noise. We normalize the noise so that  $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, I)$ , and assume an average input power constraint of  $\mathbb{E}[\mathbf{x}^\dagger \mathbf{x}] \leq n_t$ . It is natural to assume that the entries of the channel matrix  $\mathbf{H}$  are *identically* distributed since each pair of transmit and receive antennas sees the same scattering environment. We may further normalize the channel gains to unity so that  $\mathbb{E}[|\mathbf{H}_{k,\ell}|^2] = 1$ , for all  $k, \ell$ . With all the above normalization,  $\Gamma$  represents the effective SNR at each receive antenna. We also make the reasonable assumption that the channel changes in a stationary and ergodic manner from symbol to symbol.

For a purely diffuse rich scattering environment with no dominant (specular) paths, the entries of  $\mathbf{H}$  are well modeled as i.i.d. zero-mean proper-complex Gaussian random variables. In general, the entries of  $\mathbf{H}$  may not be modeled well as Gaussian, may have nonzero mean and may be dependent. In the following, we explore the statistical structure of  $\mathbf{H}$  in more detail based on a physical model for the channel.

### A. Physical Channel Model

As we mentioned earlier, we will restrict our attention to MIMO channels with ULAs of antennas at the transmitter and receiver. Let  $d_t$  and  $d_r$  denote the antenna spacing at the transmitter and receiver, respectively. The channel can be described in terms of all the paths joining the transmitter and receiver as shown in Fig. 1. If the receiver is in the line of sight (LOS) of the transmitter, there is a single path corresponding to the LOS. In addition, there would be reflections from other scatterers in the environment. Physics dictates that each such scatterer (since it cannot be perfectly smooth) produces a continuum of paths joining the transmitter and receiver. Nevertheless, it is convenient to first consider the following discrete path model to describe the channel before we generalize it to the continuous path case:

$$\mathbf{H} = \sqrt{n_t n_r} \left[ \beta_{\text{LOS}} \underline{\mathbf{a}}_r(\theta_{r,\text{LOS}}) \underline{\mathbf{a}}_t^\dagger(\theta_{t,\text{LOS}}) + \sum_{i=1}^N \beta_i e^{j\varphi_i} \underline{\mathbf{a}}_r(\theta_{r,i}) \underline{\mathbf{a}}_t^\dagger(\theta_{t,i}) \right] \quad (2)$$

where  $\beta_{\text{LOS}}$  is the path gain of the LOS path (if one exists), and  $\{\beta_i\}_{i=1}^N$  are the path gains associated with the non-LOS paths. The path gains are normalized so that  $\beta_{\text{LOS}}^2 + \sum_i \beta_i^2 = 1$ . The LOS path typically has a gain that is larger than the sum of all the non-LOS path gains, and the ratio of  $\beta_{\text{LOS}}^2$  to  $\sum_i \beta_i^2$  is called the *Rice factor*. The phase of the LOS path can be tracked accurately at the receiver and is hence well modeled as deterministic, and without loss of generality, we may set the LOS path phase to 0. The randomness of the channel is due to the randomness of the non-LOS path phases  $\{\varphi_i\}$ . The vectors  $\underline{\mathbf{a}}_t(\theta_t)$  and  $\underline{\mathbf{a}}_r(\theta_r)$  are the normalized array steering and response vectors, which are given by

$$\underline{\mathbf{a}}_t(\theta_t) = \frac{1}{\sqrt{n_t}} \left[ 1, e^{-j2\pi(\theta_t-0.5)}, \dots, e^{-j2\pi(n_t-1)(\theta_t-0.5)} \right]^\top$$

$$\underline{\mathbf{a}}_r(\theta_r) = \frac{1}{\sqrt{n_r}} \left[ 1, e^{-j2\pi(\theta_r-0.5)}, \dots, e^{-j2\pi(n_r-1)(\theta_r-0.5)} \right]^\top.$$

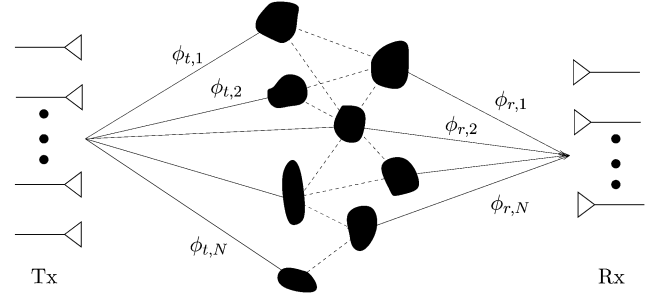


Fig. 1. Physical model for MIMO channel.

The variable  $\theta$  is related to the physical propagation angle  $\phi$  as

$$\theta = \left[ \frac{d(\sin(\phi) + 1)}{\lambda_c} \right]_{\text{mod}[0,1]} = [\alpha(\sin(\phi) + 1)]_{\text{mod}[0,1]} \quad (3)$$

with  $\lambda_c$  being the wavelength of propagation, and  $\alpha = d/\lambda_c$ . Note that  $\theta$  can be thought of as a *scaled* angle. In particular, if we assume that  $\alpha \leq 1/2$ , there is a one-to-one mapping between the physical angle<sup>1</sup>  $\phi \in [-\pi/2, \pi/2]$  and  $\theta \in [0, 2\alpha]$ . For  $\alpha > 1/2$ , there may be more than one (but less than  $\lceil 2\alpha \rceil$ ) physical angle  $\phi \in [-\pi/2, \pi/2]$  that maps to each  $\theta \in [0, 1]$ . We refer to  $\theta$  as a *virtual angle* and use it to describe the spatial scattering environment in place of  $\phi$  in the remainder of this paper.

We may now generalize the model of (2) to a continuum of paths as follows:

$$\mathbf{H} = \sqrt{n_t n_r} \left[ \beta_{\text{LOS}} \underline{\mathbf{a}}_r(\theta_{r,\text{LOS}}) \underline{\mathbf{a}}_t^\dagger(\theta_{t,\text{LOS}}) + \int_0^1 \int_0^1 \alpha(\theta_r, \theta_t) \underline{\mathbf{a}}_r(\theta_r) \underline{\mathbf{a}}_t^\dagger(\theta_t) d\theta_r d\theta_t \right] \quad (4)$$

where the function  $\alpha(\theta_r, \theta_t)$  represents the random *angular spreading function*, i.e., the random complex channel gain density at virtual transmit angle  $\theta_t$  and receive angle  $\theta_r$ . We note that it is often useful to approximate the continuous path model of (4) by a discrete path model for statistical characterization (see Section II-C).

### B. Virtual Representation

Consider the matrices  $A_r$  ( $n_r \times n_r$ ) and  $A_t$  ( $n_t \times n_t$ ) that are given in terms of the array steering and response vectors as

$$A_r = [\underline{\mathbf{a}}_r(\theta_{r,1}), \underline{\mathbf{a}}_r(\theta_{r,2}), \dots, \underline{\mathbf{a}}_r(\theta_{r,n_r})]$$

$$A_t = [\underline{\mathbf{a}}_t(\theta_{t,1}), \underline{\mathbf{a}}_t(\theta_{t,2}), \dots, \underline{\mathbf{a}}_t(\theta_{t,n_t})]$$

and where  $\theta_{r,1}, \theta_{r,2}, \dots, \theta_{r,n_r}$  (similarly,  $\theta_{t,1}, \theta_{t,2}, \dots, \theta_{t,n_t}$ ) are equally spaced angles in the range  $[0, 1]$ . That is,

$$\theta_{r,k} = \frac{k - 0.5}{n_r}, \quad k = 1, 2, \dots, n_r$$

and a similar equation holds for the angles  $\theta_{t,\ell}$ . As noted in [11], the matrices  $A_r$  and  $A_t$  are *unitary* discrete Fourier transform matrices.

<sup>1</sup>The physical angle can be restricted to the range  $[-\pi/2, \pi/2]$  without loss of generality since paths corresponding to angles in the ranges  $[-\pi, -\pi/2]$  and  $[\pi/2, \pi]$  can be mapped into equivalent paths in the range  $[-\pi/2, \pi/2]$ .

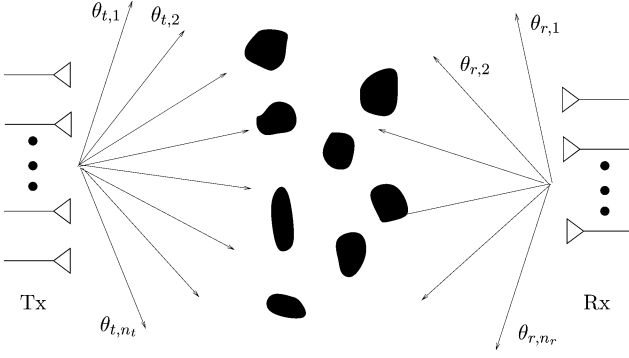


Fig. 2. Virtual representation of MIMO channel.

A two-dimensional spatial Fourier representation of the channel is obtained via the matrices  $A_r$  and  $A_t$  as follows:

$$\mathbf{H} = A_r \tilde{\mathbf{H}} A_t^\dagger. \quad (5)$$

The matrix  $\tilde{\mathbf{H}}$  is referred to as a *virtual representation* [11] of the channel matrix  $\mathbf{H}$ , with the understanding that it corresponds to transmitting and receiving in fixed, virtual directions determined by the spatial resolution of the array. Fig. 2 shows the virtual representation of a physical MIMO channel.

We note that there are alternative ways to obtain a representation for  $\mathbf{H}$  by pre- and post-multiplication by other unitary matrices. What we show in the following is that using the specific unitary matrices  $A_r$  and  $A_t$  leads to some very interesting and useful properties for the representation  $\tilde{\mathbf{H}}$ .

The elements of the virtual channel matrix are related to the spreading function  $\alpha$  as

$$\tilde{\mathbf{H}}_{k,\ell} = \hat{\alpha}(\theta_{r,k}, \theta_{t,\ell}), \quad k = 1, \dots, n_r, \quad \ell = 1, \dots, n_t \quad (6)$$

where

$$\hat{\alpha}(\theta_r, \theta_t) = \beta_{\text{LOS}} f_{n_r}(\theta_r - \theta_{r,\text{LOS}}) f_{n_t}^*(\theta_t - \theta_{t,\text{LOS}}) + \int_0^1 \int_0^1 \alpha(\theta'_r, \theta'_t) f_{n_r}(\theta'_r - \theta_r) f_{n_t}^*(\theta'_t - \theta_t) d\theta'_r d\theta'_t$$

with

$$f_{n_r}(\theta) = \sqrt{n_r} \frac{\text{sinc}(n_r \theta)}{\text{sinc}(\theta)} e^{-j\pi\theta(n_r-1)}$$

and where  $\text{sinc}(x) = \sin(\pi x)/(\pi x)$ . A similar equation holds for  $f_{n_t}(\theta)$ . Thus, (6) states that the virtual channel matrix elements are samples (at the fixed virtual angles) of a smoothed version of the angular spreading function and the LOS component. Furthermore, the smoothing kernel gets narrower<sup>2</sup> with increasing  $n_t$  and  $n_r$ . Due to this sampling property, the virtual matrix can be considered as an “image” of the physical scattering environment. We exploit this property of the virtual representation to approximate the statistics of the channel matrix.

### C. Virtual Path Partitioning and Channel Statistics

For sufficiently large  $n_r$  and  $n_t$ , each virtual channel matrix element  $\tilde{\mathbf{H}}_{k,\ell}$  can be considered to be an aggregation of disjoint sets of paths with transmit and receive angles in the neighborhood of the fixed angles  $\theta_{t,\ell}$  and  $\theta_{r,k}$ , respectively, for

<sup>2</sup>The null-to-null main-lobe width of  $f_{n_r}(\theta)$  is  $2/n_r$ .

$\ell = 1, \dots, n_t, k = 1, \dots, n_r$  [11]. In particular, the LOS path will contribute to only one of the virtual channel components, and we use  $\check{k}$  and  $\check{\ell}$  to denote the indices of this coefficient. Thus, for sufficiently large  $n_r$  and  $n_t$ , we can approximate

$$\tilde{\mathbf{H}}_{k,\ell} \approx \sqrt{n_r n_t} \int_{\theta_{r,k}-0.5/n_r}^{\theta_{r,k}+0.5/n_r} \int_{\theta_{t,\ell}-0.5/n_t}^{\theta_{t,\ell}+0.5/n_t} \alpha(\theta_r, \theta_t) d\theta_r d\theta_t \quad (7)$$

for  $k \neq \check{k}, \ell \neq \check{\ell}$ , with  $\tilde{\mathbf{H}}_{\check{k},\check{\ell}}$  having the term

$$m_{\text{LOS}} = \sqrt{n_r n_t} \beta_{\text{LOS}}$$

added to the double integral. We reiterate that for  $\alpha > 1/2$ , there may be more than one physical path at each virtual angle  $\theta$ , and for  $\alpha < 1/2$ , there can be no paths at angles  $\theta \in (2\alpha, 1]$ .

For the purposes of understanding the statistics of  $\tilde{\mathbf{H}}$ , we may further simplify the right-hand side of (7) using a discrete path approximation. Let  $S_{k,\ell}$  denote the set of non-LOS paths that contribute to  $\tilde{\mathbf{H}}_{k,\ell}$ . Then we can write

$$\tilde{\mathbf{H}}_{k,\ell} \approx \sum_{i \in S_{k,\ell}} \beta_i e^{j\varphi_i}, \quad \text{for } k \neq \check{k}, \ell \neq \check{\ell} \quad (8)$$

with  $\tilde{\mathbf{H}}_{\check{k},\check{\ell}}$  having the term  $m_{\text{LOS}}$  added to the summation.

To proceed we make the following reasonable and now standard assumption about the physical scattering environment [14].

*Assumption 1:* The path phases  $\{\varphi_i\}$  are independent random variables that are uniformly distributed on  $[0, 2\pi]$ .

This assumption along with the path partitioning of (8) immediately imply the following result that will be crucial in characterizing the channel capacity.

*Lemma 1:* The statistics of the virtual coefficients satisfy the following properties.

- (P1) The elements of  $\tilde{\mathbf{H}}$  are *independent* random variables.
- (P2) For  $k \neq \check{k}, \ell \neq \check{\ell}$ ,  $\tilde{\mathbf{H}}_{k,\ell}$  is a zero-mean proper-complex random variable with a symmetric distribution around the origin. In particular,  $\tilde{\mathbf{H}}_{k,\ell}$  has the same distribution as  $-\tilde{\mathbf{H}}_{k,\ell}$ .
- (P3) The LOS term  $\tilde{\mathbf{H}}_{\check{k},\check{\ell}}$  is a proper-complex random variable, with mean  $m_{\text{LOS}} = \sqrt{n_r n_t} \beta_{\text{LOS}}$  and a symmetric distribution around the mean.

Note that the elements of  $\tilde{\mathbf{H}}$  are not necessarily well modeled as jointly Gaussian, particularly for large  $n_r$  and  $n_t$  where only a small fraction of the paths contribute to each coefficient. However, the fact that the elements of  $\tilde{\mathbf{H}}$  are well modeled as being independent implies that the joint distribution of  $\tilde{\mathbf{H}}$  (and hence the joint distribution of  $\mathbf{H}$ ) is completely characterized by the marginals of the elements  $\tilde{\mathbf{H}}_{k,\ell}$ . Furthermore, the independence of the elements of  $\tilde{\mathbf{H}}$  greatly facilitates the capacity analysis as we will demonstrate in Sections III and IV.

### D. Second-Order Statistics of Virtual Channel Matrix

Assumption 1 imposes the following second-order statistical structure on  $\alpha(\theta_r, \theta_t)$ :

$$E[\alpha(\theta_r, \theta_t) \alpha^*(\theta'_r, \theta'_t)] = \Psi(\theta_r, \theta_t) \delta(\theta_r - \theta'_r) \delta(\theta_t - \theta'_t) \quad (9)$$

where  $\delta(\cdot)$  denotes the Dirac delta function and  $\Psi(\theta_r, \theta_t) \geq 0$  is called the *angular scattering function* for the non-LOS

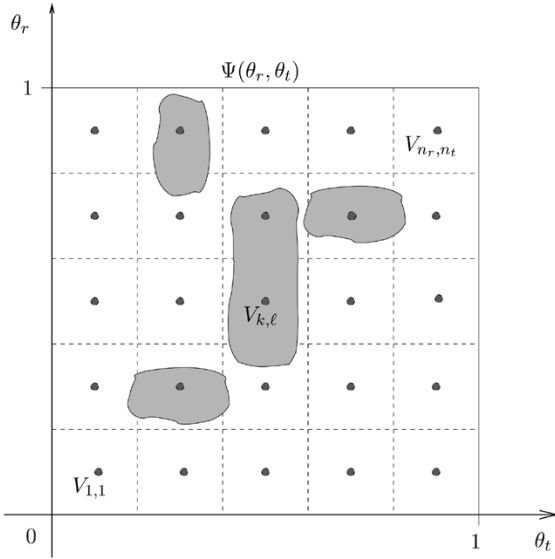


Fig. 3. Example scattering function and corresponding variance matrix. The shaded regions in  $\Psi(\theta_r, \theta_t)$  correspond to scattering clusters in the physical environment.

paths, which reflects the channel power density for various transmit–receive angle pairs. The function  $\Psi(\theta_r, \theta_t)$  is well modeled as *bounded* and piecewise continuous with

$$\int_0^1 \int_0^1 \Psi(\theta_r, \theta_t) d\theta_r d\theta_t = 1 - \beta_{\text{LOS}}^2.$$

We now explore the second-order statistics of the elements of  $\tilde{\mathbf{H}}$  and their relationship to the scattering function. From Lemma 1 we know that elements of  $\tilde{\mathbf{H}}$  are independent and have zero mean (except for the LOS component). It is easily shown that the variances of the elements are given by

$$\begin{aligned} V_{k,\ell} &= \text{Var}(\tilde{\mathbf{H}}_{k,\ell}) \\ &= \int_0^1 \int_0^1 \Psi(\theta_r, \theta_t) |f_{n_r}(\theta_r - \theta_{r,k})|^2 \\ &\quad \cdot |f_{n_t}(\theta_t - \theta_{t,\ell})|^2 d\theta_r d\theta_t \\ &\approx n_t n_r \int_{\theta_{r,k}-0.5/n_r}^{\theta_{r,k}+0.5/n_r} \int_{\theta_{t,\ell}-0.5/n_t}^{\theta_{t,\ell}+0.5/n_t} \Psi(\theta_r, \theta_t) d\theta_r d\theta_t \end{aligned} \quad (10)$$

where the approximation in (10) gets better with increasing  $n_r$  and  $n_t$ , and for sufficiently large  $n_r$  and  $n_t$  we have

$$V_{k,\ell} \approx \Psi(\theta_{r,k}, \theta_{t,\ell}). \quad (11)$$

Fig. 3 illustrates the relationship between  $\Psi(\theta_r, \theta_t)$  and  $V_{k,\ell}$ .

The variation in the elements of  $V$  can be used as a measure of the correlation in the original channel matrix  $\mathbf{H}$ . If  $V$  has only a small fraction of dominant entries,  $\mathbf{H}$  will have highly correlated elements. On the other hand, if  $V$  has roughly uniform entries, then  $\mathbf{H}$  will have roughly uncorrelated elements with equal variances.

*Remark 1:* We note that the Kronecker correlation structure used in the analysis of [4], [6], [7], [9] is obtained in the rare special case where the scattering function  $\Psi(\theta_r, \theta_t)$  is in product form, i.e.,  $\Psi(\theta_r, \theta_t) = \Psi_r(\theta_r)\Psi_t(\theta_t)$ . The analysis given in this

paper is valid for general  $\Psi(\theta_r, \theta_t)$  and without the Gaussian assumption for the statistics.

We now present some basic properties of the variance matrix  $V$ . First, since the original channel matrix  $\mathbf{H}$  is normalized to have unit power entries, it follows from (5) that the variance matrix  $V$  satisfies

$$\sum_{k,\ell} V_{k,\ell} = n_r n_t (1 - \beta_{\text{LOS}}^2). \quad (12)$$

Also, based on (10), it easily follows that a “tile” approximation to  $V$  converges to  $\Psi(\theta_r, \theta_t)$  as  $n_r, n_t \rightarrow \infty$ . In particular, define the piecewise constant function

$$g_{n_r, n_t}(\theta_r, \theta_t) = V_{k,\ell},$$

$$\text{for } \theta_r \in \left[ \frac{k-1}{n_r}, \frac{k}{n_r} \right], \quad \theta_t \in \left[ \frac{\ell-1}{n_t}, \frac{\ell}{n_t} \right]$$

for  $k = 1, \dots, n_r$  and  $\ell = 1, \dots, n_t$ , where  $V_{k,\ell}$  is as given in (10). Then

$$\lim_{n_r, n_t \rightarrow \infty} g_{n_r, n_t}(\theta_r, \theta_t) = \Psi(\theta_r, \theta_t) \quad (13)$$

uniformly for  $(\theta_r, \theta_t) \in [0, 1]^2$ . We will exploit this convergence when we study the asymptotic capacity in Section IV.

*Remark 2:* The convergence result of (13) obviously also holds if we set  $V_{k,\ell} = \Psi(\theta_{r,k}, \theta_{t,\ell})$  as in the approximation of (11).

### III. OPTIMAL INPUT DISTRIBUTION AND CAPACITY

Based on (5), we can rewrite the input–output relationship (1) in the virtual domain as

$$\tilde{\mathbf{y}} = \sqrt{\frac{\Gamma}{n_t}} \tilde{\mathbf{H}} \tilde{\mathbf{x}} + \tilde{\mathbf{w}} \quad (14)$$

where  $\tilde{\mathbf{x}} = A_t^\dagger \mathbf{x}$ ,  $\tilde{\mathbf{y}} = A_t^\dagger \mathbf{y}$ , and  $\tilde{\mathbf{w}} = A_t^\dagger \mathbf{w}$ . Due to the unitarity of  $A_t$ , the input power constraint in the virtual domain is unchanged, i.e.,  $\mathbb{E}[\tilde{\mathbf{x}}^\dagger \tilde{\mathbf{x}}] \leq n_t$ .

Our main goal in this paper is to analyze the channel capacity using the model given in (14). As mentioned in Section I, we make the coherent channel assumption throughout the paper.

Referring to (1), the results of [2, Sec. 4] show that the (ergodic) capacity of the MIMO channel described in the previous section is achieved by a zero-mean proper-complex Gaussian input vector  $\mathbf{x}$  with a covariance matrix  $Q$  that satisfies  $\text{Tr}(Q) \leq n_t$ . The capacity is hence given by

$$C = \max_{Q: \text{Tr}(Q) \leq n_t} \mathbb{E} \left[ \log \det \left( I + \frac{\Gamma}{n_t} \mathbf{H} Q \mathbf{H}^\dagger \right) \right]. \quad (15)$$

If the entries of  $\mathbf{H}$  are i.i.d. (as in [2]), then the optimal  $Q$  is the identity matrix. In the general correlated case that we consider here, the optimal  $Q$  is difficult to characterize, and we hence turn to the virtual domain to facilitate the analysis.

In the remainder of this section, we present several results characterizing the optimal input distribution. In order to keep the notation from getting cumbersome, the main results are presented for the case where there is no LOS path in the channel. Extensions to the case with an LOS path are discussed in remarks following the main results.

### A. Characterization of Optimal Input Distribution

From (14), it is clear that the optimal input in the virtual domain is also a zero-mean proper complex Gaussian input vector  $\tilde{\mathbf{x}}$  with a covariance matrix  $\tilde{Q}$  that satisfies  $\text{Tr}(\tilde{Q}) \leq n_t$ . For such a choice of input, the mutual information between channel input and output in the virtual domain is given by

$$\mathcal{I}(\tilde{Q}) = \mathbb{E} \left[ \log \det \left( I + \frac{\Gamma}{n_t} \tilde{\mathbf{H}} \tilde{Q} \tilde{\mathbf{H}}^\dagger \right) \right]. \quad (16)$$

The capacity of (15) can be rewritten as

$$C = \max_{\tilde{Q}: \text{Tr}(\tilde{Q}) \leq n_t} \mathcal{I}(\tilde{Q}). \quad (17)$$

Note that for a given  $\tilde{Q}$  in the virtual domain, the actual input covariance is given by  $Q = A_t \tilde{Q} A_t^\dagger$ .

By (17), the problem of finding the capacity  $C$  reduces to the problem of finding the optimal  $\tilde{Q}$ . We characterize the optimal  $\tilde{Q}$  in the following result.

*Theorem 1:* Suppose that the matrix  $\tilde{\mathbf{H}}$  has independent entries with each entry  $\tilde{\mathbf{H}}_{k,\ell}$  being a proper-complex random variable (not necessarily Gaussian), with  $\tilde{\mathbf{H}}_{k,\ell}$  having the same distribution as  $-\tilde{\mathbf{H}}_{k,\ell}$  (as in Lemma 1). Then the optimal  $\tilde{Q}$  that maximizes the mutual information (16) is unique and it is *diagonal*.

*Proof:* The proof follows a technique that was introduced in [5] for proving a similar result for the multiple-input single-output (MISO) proper-complex Gaussian channel. The details are given in Appendix I.  $\square$

*Remark 3:* If there is a LOS path that contributes to element  $\tilde{\mathbf{H}}_{k,\tilde{\ell}}$ , then  $\tilde{\mathbf{H}}_{k,\tilde{\ell}}$ , and hence the  $\tilde{\ell}$ th column of  $\tilde{\mathbf{H}}$ , does not have a symmetric distribution around zero. However, as can be seen in (32) in Appendix I, the crucial step in the proof of Theorem 1 requires only one column in any pair of columns of  $\tilde{\mathbf{H}}$  to have a symmetric distribution around zero. Thus, the proof given in Appendix I easily extends to the LOS case.

Since the optimal covariance matrix  $\tilde{Q}$  is a diagonal matrix  $\Lambda^\circ$ , the capacity-achieving input vector in the virtual domain has independent entries, i.e., the optimal input signals transmitted at the different virtual directions are independent. For diagonal input covariance  $\Lambda$ , diagonal element  $\lambda_i$  represents the power assigned to the  $i$ th virtual transmit angle. We are still left with the problem of finding the  $\lambda_i$ 's that achieve the capacity. But this is a problem of optimizing a concave function over a convex set and can be solved numerically using the statistics of  $\tilde{\mathbf{H}}$ . An example is given in Section V.

Note also that the actual optimal input vector has correlated entries in general with covariance matrix  $Q^\circ$  that is given by

$$Q^\circ = A_t \Lambda^\circ A_t^\dagger.$$

### B. Asymptotically Optimal Power Allocation at Low SNR

While it is possible to compute  $\Lambda^\circ$  numerically, further simplification in the optimization occurs in the asymptotic regime where the SNR is small. The following result shows that in this regime, beamforming along one of the virtual transmit angles is optimal.

*Theorem 2:* The first-order low-SNR expansion of the mutual information as a function of the diagonal covariance matrix  $\Lambda$  is optimized by  $\Lambda^\circ$  with all the elements equal to zero except  $\lambda_i^\circ$ , where  $i$  is the index identified by

$$i = \arg \max_{1 \leq \ell \leq n_t} \sum_{k=1}^{n_r} V_{k,\ell}.$$

If the maximizing index is not unique, define the index set

$$\mathcal{T} = \left\{ i : i = \arg \max_{1 \leq \ell \leq n_t} \sum_{k=1}^{n_r} V_{k,\ell} \right\}.$$

Then  $\Lambda^\circ$  is such that

$$\sum_{i \in \mathcal{T}} \lambda_i^\circ = n_t, \quad \lambda_i^\circ \geq 0, \text{ for } i \in \mathcal{T}$$

and  $\lambda_i^\circ = 0$ , for  $i \notin \mathcal{T}$ .

i.e., the power is arbitrarily assigned to the diagonal elements corresponding to those maximizing indexes without changing the capacity as long as the total power is  $n_t$ .

*Proof:* See Appendix II.  $\square$

*Remark 4:* Theorem 2 continues to hold if there is a LOS path that contributes to element  $\tilde{\mathbf{H}}_{k,\tilde{\ell}}$ , with the modification that the beamforming virtual angle is identified by

$$i = \arg \max_{1 \leq \ell \leq n_t} \left\{ \underbrace{\sum_{k=1}^{n_r} V_{k,1}}_{\ell=1} \underbrace{\sum_{k=1}^{n_r} V_{k,2}}_{\ell=2}, \dots, \underbrace{\sum_{k=1}^{n_r} V_{k,\tilde{\ell}} + m_{\text{LOS}}^2}_{\ell=\tilde{\ell}}, \dots, \underbrace{\sum_{k=1}^{n_r} V_{k,n_t}}_{\ell=n_t} \right\}$$

where  $m_{\text{LOS}} = \sqrt{n_t n_r} \beta_{\text{LOS}}$  is the mean of  $\tilde{\mathbf{H}}_{k,\tilde{\ell}}$ . Since the LOS path typically dominates all the non-LOS in gain, beamforming to angle  $\theta_{t,\tilde{\ell}}$  will become optimal at low SNR.

The preceding theorem suggests that as the SNR approaches zero, the optimal input strategy favors allocating all the transmit power to the virtual transmit angle with the largest channel gains over spreading the power among all directions. Referring back to (3), we note that beamforming to a given virtual angle  $\theta$  corresponds to beamforming to a single physical angle  $\phi$  as long as the antenna spacing  $d$  is less than  $\lambda_c/2$ . In general, beamforming to a given  $\theta$  could correspond to transmitting power in more than one direction  $\phi$ , with  $Q^\circ$  being a rank-1 matrix.

If beamforming to virtual angle  $i$  is optimal, then the capacity simplifies to

$$C = \mathbb{E} \left[ \log \left( 1 + \Gamma \|\tilde{\mathbf{h}}_i\|^2 \right) \right].$$

This beamforming strategy is also considerably easier to implement than a general input strategy, because the MIMO channel can be treated as an effective scalar channel for which one-dimensional codes can be designed to achieve capacity.

### C. Necessary and Sufficient Condition for Beamforming to be Optimal

We now sharpen the result of Theorem 2 to precisely characterize the threshold on the SNR below which beamforming is optimal.

*Theorem 3:* A necessary and sufficient condition for beamforming to the  $i$ th virtual angle to be optimal is given by

$$\Gamma \sum_{k=1}^{n_r} (1 - \mu_{k,i}) V_{k,\ell^\circ} - \sum_{k=1}^{n_r} \mu_{k,i} \leq 0 \quad (18)$$

where  $\ell^\circ = \arg \max_{\substack{1 \leq \ell \leq n_t \\ \ell \neq i}} \sum_{k=1}^{n_r} (1 - \mu_{k,i}) V_{k,\ell}$ . The functions  $\mu_{k,i}$  are defined as

$$\mu_{k,i} := \mu_{k,i}(V_{1,i}, V_{2,i}, \dots, V_{n_r,i}) = \mathbb{E} \left[ \frac{\Gamma |\tilde{\mathbf{H}}_{k,i}|^2}{1 + \Gamma \|\tilde{\mathbf{h}}_i\|^2} \right]$$

for  $1 \leq i \leq n_t$ ,  $1 \leq k \leq n_r$ . Among the  $n_t$  conditions in (18) corresponding to  $1 \leq i \leq n_t$ , at most one can be satisfied.

*Proof:* The proof of this result follows steps similar to those used in [6, Theorem 2] for the Gaussian product-form correlation model. The details are given in Appendix III.  $\square$

It is interesting to compare the results of Theorems 2 and 3 for the special case where two columns (say  $i$  and  $\ell$ ) of the variance matrix  $V$  are identical, and these columns have the maximum sum. According to Theorem 2, for asymptotically small SNR, it is optimal to assign power in an arbitrary manner among the virtual transmit angles  $i$  and  $\ell$ . In particular, it is asymptotically optimal to beamform to virtual angle  $i$  (or  $\ell$ ). However, in Theorem 3, if the beamforming condition (18) for angle  $i$  is satisfied, it will also be satisfied for angle  $\ell$ . This is not possible by the uniqueness clause in Theorem 3. Hence, the asymptotic behavior of Theorem 2 does not necessarily hold for any nonzero SNR.

*Remark 5:* Theorem 3 can easily be modified to incorporate the case where there is an LOS path by simply adding  $m_{\text{LOS}}^2$  to  $V_{k,\tilde{\ell}}$ . As we commented earlier in Remark 4, beamforming to angle  $\theta_{t,\tilde{\ell}}$  will be optimal below a threshold SNR.

We end this section by noting that the optimal virtual angle for beamforming can be determined using only the second moments of the virtual coefficients. In contrast, determining the optimal input distribution in situations where beamforming is not optimal requires the entire distribution of the virtual coefficients.

## IV. ASYMPTOTIC CAPACITY FOR A LARGE NUMBER OF ANTENNAS

In this section, we study the capacity in the asymptotic regime where the numbers of receive and transmit antennas go to infinity, with their ratio being kept constant. The results given in this section are restricted to the case where there is no LOS path connecting the transmitter and receiver.

Based on the results of the previous section, we can write the ergodic capacity of the MIMO channel as

$$C = \mathbb{E} \left[ \log \det \left( I + \frac{\Gamma}{n_t} \tilde{\mathbf{H}} \Lambda^\circ \tilde{\mathbf{H}}^\dagger \right) \right] \quad (19)$$

where  $\Lambda^\circ$  is the optimal diagonal input covariance matrix. In scenarios where beamforming in the  $i$ th virtual transmit angle is optimal, the capacity further simplifies to

$$C = \mathbb{E} \left[ \log \left( 1 + \Gamma \|\tilde{\mathbf{h}}_i\|^2 \right) \right]. \quad (20)$$

The expectation in (19) (or (20)) is easily evaluated numerically in general, and in some special cases, closed-form expressions may also be obtained. But to gain further analytical insight into the capacity of the channel, we turn to the asymptotic scenario where the number of antennas is large.

Let  $n_t$  and  $n_r$  go to infinity with their ratio  $n_t/n_r$  being kept constant at  $\tau$  (say). To simplify notation, let  $n = n_r$ . Then  $n_t = \lfloor \tau n \rfloor$ , and we are interested in the limit as  $n \rightarrow \infty$ . Since  $C$  can possibly grow without bound under this limit, we normalize  $C$  by  $n$ . The normalized capacity is given by

$$\bar{C} = \frac{1}{n} \mathbb{E} \left[ \log \det \left( I + \frac{\Gamma}{\lfloor \tau n \rfloor} \tilde{\mathbf{H}} \Lambda^\circ \tilde{\mathbf{H}}^\dagger \right) \right]. \quad (21)$$

Our goal is to evaluate  $\lim_{n \rightarrow \infty} \bar{C}$ , which we denote by  $\bar{C}^\infty$ .

Note that  $\bar{C}$  captures how the capacity scales with the number of antennas. In particular, if  $0 < \bar{C} < \infty$ , then the capacity scales linearly with  $n$  as it does in the case of the i.i.d. channel.

From (21) it is clear that the evaluation of  $\bar{C}^\infty$  requires the characterization of the limiting value of the optimal diagonal input covariance matrix  $\Lambda^\circ$ . However, this is a rather difficult task considering that, in general, it is not possible to obtain an explicit equation for  $\Lambda^\circ$  for finite  $n$ . For this reason, we relax our requirement that input distribution be optimal, and instead consider computing the limit of the right-hand side of (21) with  $\Lambda^\circ$  being replaced by some reasonable diagonal input covariance  $\Lambda$  that can be characterized explicitly.<sup>3</sup> In this case, we need to interpret  $\bar{C}$  as simply the information rate that is achievable by using the input covariance  $\Lambda$ . To proceed with the analysis, we need to impose some regularity conditions on  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{n_t})$ .

*Assumption 2:* For each  $n$ , define the function  $s_n : [0, \tau] \rightarrow \mathfrak{R}$  by

$$s_n(v) = \lambda_\ell, \quad \text{for } v \in \left[ \frac{\ell-1}{n}, \frac{\ell}{n} \right]$$

where  $\ell = 1, \dots, \lfloor n\tau \rfloor$ . Then  $s_n(v)$  is bounded for each  $n$ , and converges uniformly to a limiting bounded function  $s(v)$  as  $n \rightarrow \infty$ .

Clearly,  $s(v)$  has to satisfy the power constraint

$$\int_0^\tau s(v) dv = \tau.$$

*Remark 6:* A natural question to ask is whether  $\Lambda^\circ$  satisfies Assumption 2. It is clear that the assumption will not hold for  $\Lambda^\circ$  if the optimal transmit power along any of the virtual angles grows without bound as  $n \rightarrow \infty$ . It seems reasonable that such unbounded input power allocations should be ruled out by the piecewise continuity and boundedness of the scattering

<sup>3</sup>For example, we can consider  $\Lambda = I$  corresponding to i.i.d. inputs in both the virtual and actual domains.

function  $\Psi$ ; however, we have not been able to establish a rigorous result along these lines.

The asymptotic evaluation of  $\bar{C}$  is facilitated by relating the limiting value of  $\bar{C}$  to the limiting eigenvalue distribution of  $\tilde{\mathbf{H}}\Lambda\tilde{\mathbf{H}}^\dagger/n$ . To this end, we first give the following definition.

*Definition 1:* The Stieltjes transform  $m_A$  of an  $n \times n$  Hermitian matrix  $A$  is defined as

$$m_A(z) = \frac{1}{n} \text{Tr}\{(A - zI)^{-1}\}.$$

We now have the following theorem that follows quite easily from a similar result on the sum capacity of code-division multiple-access (CDMA) systems with random spreading [16]. A sketch of the proof is given in Appendix V.

*Theorem 4:* Assume that the Stieltjes transform of  $\tilde{\mathbf{H}}\Lambda\tilde{\mathbf{H}}^\dagger/n$  converges in probability to a deterministic limit denoted by  $m$  as  $n \rightarrow \infty$ . Then the asymptotic normalized capacity is given by

$$\bar{C}^\infty = \int_0^1 \frac{1}{t} \left[ 1 - \frac{\tau}{t\Gamma} m\left(-\frac{\tau}{t\Gamma}\right) \right] dt. \quad (22)$$

Theorem 4 can be used to evaluate the asymptotic capacity if we can determine the limiting Stieltjes transform of  $\tilde{\mathbf{H}}\Lambda\tilde{\mathbf{H}}^\dagger/n$ . The latter limit requires the application of a result of Girko [17, Corollary 10.1.2] which is restated in Theorem 5. We note that this proof technique using Girko's result was also used in [4] in analyzing the asymptotic mutual information for i.i.d. zero-mean complex-Gaussian inputs for the special case of the Kronecker Gaussian channel model.

*Theorem 5 (Girko):* Let  $\mathbf{X}$  be a  $[cn] \times [dn]$  matrix with zero-mean independent entries. Define the variance function from  $[0, c] \times [0, d]$  to  $\Re$  by

$$f_n(u, v) = n \text{Var}(X_{i,j}),$$

$$\text{for } u \in \left[ \frac{i-1}{n}, \frac{i}{n} \right], \quad v \in \left[ \frac{j-1}{n}, \frac{j}{n} \right].$$

Assume that the variance function is uniformly bounded  $\forall n, i, j$  and converges uniformly to a bounded function  $f(u, v)$ . Then

$$\frac{1}{n} \sum_{i=[an]}^{[bn]} (\mathbf{X}\mathbf{X}^\dagger - zI)_{ii}^{-1} \xrightarrow{n \rightarrow \infty} \int_a^b e(u, z) du \quad (23)$$

where the convergence is in probability and where  $e(u, z)$  satisfies the equation

$$e(u, z) = \left[ -z + \int_0^d \frac{f(u, v) dv}{1 + \int_0^c e(w, z) f(w, v) dw} \right]^{-1}.$$

As a special case, setting  $a = 0$  and  $b = c$  in (23), the Stieltjes transform of  $\mathbf{X}\mathbf{X}^\dagger$  converges in probability to the deterministic limit

$$m(z) = \int_0^c e(u, z) du.$$

In order to apply Theorem 5, we give the following lemma whose proof follows easily from (13) and Assumption 2.

*Lemma 2:* For each  $n$ , define  $g_n : [0, 1] \times [0, \tau] \rightarrow \Re$  by

$$g_n(u, v) = \lambda_\ell V_{k,\ell}, \quad \text{for } u \in \left[ \frac{k-1}{n}, \frac{k}{n} \right], \quad v \in \left[ \frac{\ell-1}{n}, \frac{\ell}{n} \right]$$

where  $k = 1, \dots, n$  and  $\ell = 1, \dots, [n\tau]$ . Then, as  $n \rightarrow \infty$ ,  $g_n(u, v)$  converges uniformly to the limiting bounded function

$$g(u, v) = \Psi(u, v/\tau) s(v). \quad (24)$$

Using Lemma 2 in Theorem 5, we have the following result that characterizes the asymptotic capacity.

*Theorem 6:* As  $n \rightarrow \infty$ , the Stieltjes transform  $\tilde{\mathbf{H}}\Lambda\tilde{\mathbf{H}}^\dagger/n$  converges in probability to a deterministic limit

$$m(z) = \int_0^1 e(u, z) du \quad (25)$$

where  $e(u, t)$  satisfies the equation

$$e(u, z) = \left[ -z + \int_0^\tau \frac{g(u, v) dv}{1 + \int_0^1 e(w, z) g(w, v) dw} \right]^{-1} \quad (26)$$

with  $g(u, v)$  defined in (24).

Using (22), (24)–(26), the asymptotic normalized capacity  $\bar{C}^\infty$  can be computed numerically for any input covariance satisfying Assumption 2 and any bounded spatial scattering function  $\Psi$ . Sample numerical results are presented in Section V.

The asymptotic normalized capacity  $\bar{C}^\infty$  provides a convenient measure with which to compare various scattering environments. It also provides an analytical approximation to the capacity of finite antenna channels. In particular, if  $n_t = \tau n_r$ , then we can approximate the capacity of the  $(n_t, n_r)$  MIMO channel by

$$C = n_r \bar{C}^\infty(\tau). \quad (27)$$

We will investigate the accuracy of this approximation in Section V.

We end this section by noting that while the computation of the nonasymptotic capacity requires detailed information about the marginals of the virtual channel elements, the computation of the asymptotic capacity requires only second-order statistics (the scattering function).

## V. NUMERICAL RESULTS AND DISCUSSION

In this section, we provide a set of examples that illustrate the theoretical results of the previous sections.

### A. Optimal Input Distribution and Beamforming

We begin with an example illustrating the results of Section III. Consider a system with five transmit and five re-



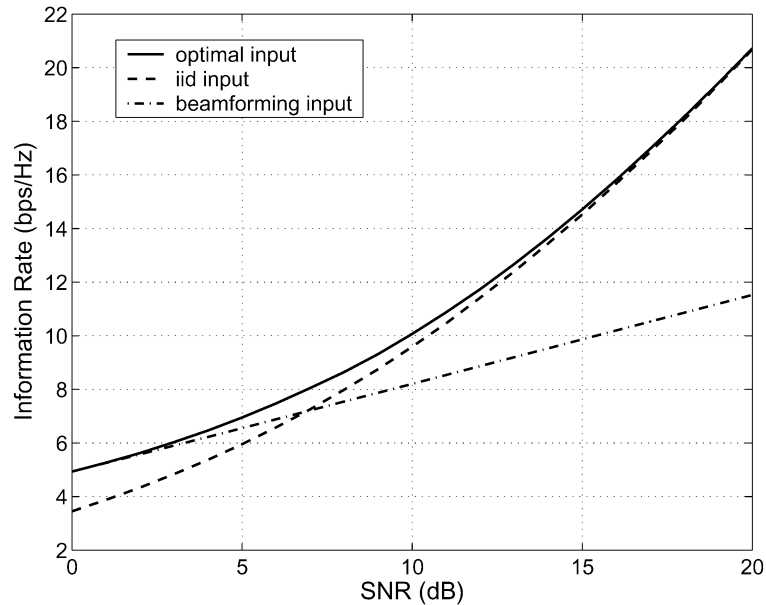


Fig. 4. Comparison of information rates obtained by using the optimal inputs, i.i.d. inputs, and beamforming along the best virtual transmit direction for the variance matrix shown in (28).

ceive antennas, where the  $\tilde{\mathbf{H}}$  is a zero-mean complex-Gaussian channel with variance matrix given by

$$V = \frac{25}{5.7} \begin{pmatrix} 0.1 & 0 & 1 & 0 & 0 \\ 0 & 0.1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.25 & 0 \\ 0 & 0 & 1 & 0 & 0.25 \end{pmatrix}. \quad (28)$$

Note that we have normalized the entries so that  $\sum_{k,\ell} V_{k,\ell} = 25$ . Such a variance matrix could represent a physical environment with two very small scatterers and two bigger scatterers and one large scattering cluster. The optimal input variances  $\{\lambda_i^\circ\}$  can easily be obtained numerically for this example. Fig. 4 plots the capacity achieved by using the optimal input distribution and compares it with the information rate obtained by using i.i.d. inputs, i.e.,  $\Lambda = I$ . It is clear from the figure that the information rate is improved by using the optimal inputs; however, the improvement in the capacity becomes less significant as SNR increases.<sup>4</sup> Note also that the improvement in information rate by optimizing the input distribution depends on the scattering function. If the scattering function is sufficiently rich, or if it is close to being symmetric along each of virtual transmit angles, then i.i.d. inputs may achieve very good performance.

Using Theorem 3 we can easily determine that beamforming along the third virtual transmit direction is optimal for SNRs below 0.29 dB. Fig. 4 also plots the information rate obtained by beamforming along the third virtual transmit direction for SNRs ranging from 0 to 20 dB. As can be seen in the figure, beamforming is indeed optimal for SNRs below 0.29 dB, and it remains close to optimal for SNRs below 5 dB.

<sup>4</sup>It is easy to show for this example that i.i.d. inputs are indeed asymptotically optimal as the SNR goes to infinity (see [19] for a proof). This would not be the case if the variance matrix  $V$  had one or more all-zero columns.

### B. Effect of Correlation on Capacity

We now provide an example comparing the capacity of the standard i.i.d. MIMO channel, for which  $V$  has all entries equal to 1, with the capacity of the correlated MIMO channel with variance matrix given in (28). It is interesting to see in Fig. 5, that for SNRs below 2 dB, the correlated channel has a larger capacity than the i.i.d. channel. This is somewhat surprising given that it is generally believed that rich scattering environments are needed for optimal use of multiple antennas. The reason for the crossover of course is that the multiplexing gain offered by the i.i.d. channel manifests itself only at sufficiently high SNRs. We note further that the capacity of the correlated channel can be approached by using beamforming inputs at low SNRs as we saw previously in Fig. 4.

### C. Accuracy of Asymptotics

In the following, we compute the asymptotic normalized capacity of Section IV for a specific example and investigate the accuracy of these asymptotics. Consider the scattering function shown in Fig. 6, which corresponds to a moderately correlated channel since the support of the scattering function is moderately limited in transmit and receive angles. For a MIMO system with  $n_t$  transmit and  $n_r$  receive antennas operating in this scattering environment, let the variance matrix be  $V$ . To simplify the calculations, we set  $V_{k,\ell} = \Psi(\theta_{r,k}, \theta_{t,\ell})$  (see Remark 2). Then, based on symmetry arguments, we immediately see that the optimal input variances are given by

$$\lambda_\ell^\circ = \begin{cases} \frac{5}{3}, & \text{if } V_{k,\ell} = \frac{25}{6} \text{ for some } k \\ 0, & \text{otherwise.} \end{cases}$$

Thus,  $\Lambda^\circ$  satisfies Assumption 2 and the corresponding limiting function  $s(v)$  is given by

$$s(v) = \begin{cases} \frac{5}{3}, & \text{if } v \in [0.2, 0.8] \\ 0, & \text{otherwise.} \end{cases}$$

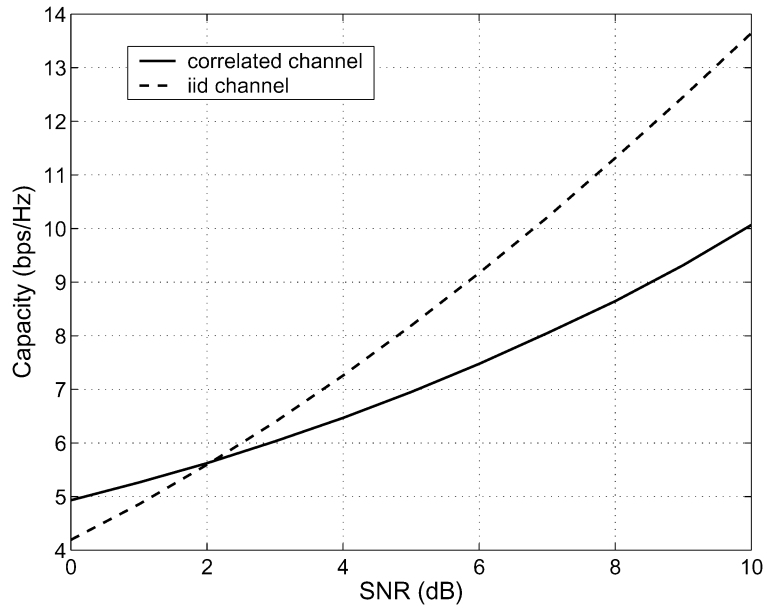


Fig. 5. Comparison of capacities of the i.i.d. and correlated channel with  $V$  given in (28).

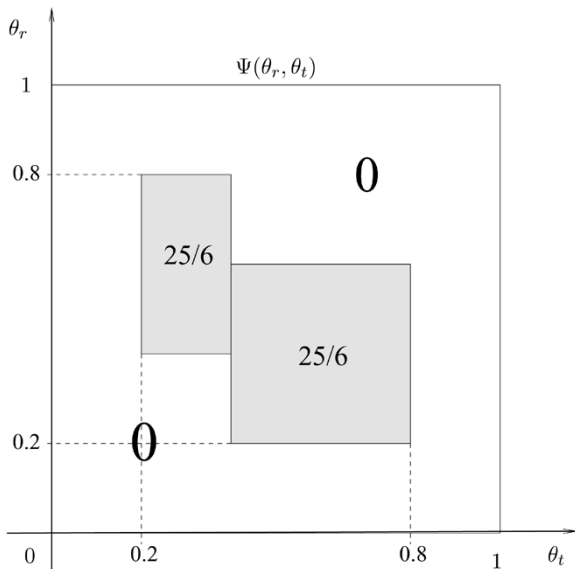


Fig. 6. Spatial scattering function.

Plugging  $\Psi(\theta_r, \theta_t)$  and  $s(v)$  in (22), (24)–(26), we can numerically evaluate the asymptotic normalized capacity  $\bar{C}^\infty$  for various values of  $\tau = n_t/n_r$  and SNR  $\Gamma$ . We can then use (27) to approximate the capacity of any given  $(n_t, n_r)$  MIMO system operating in this scattering environment. Fig. 7 plots the direct numerical evaluation of the capacity for a zero-mean complex Gaussian channel along with the asymptotic approximation for various values of  $n_t$  and  $n_r$ . As can be seen in this figure, the asymptotic approximation is quite accurate even for a  $(5, 5)$  MIMO system.

## VI. CONCLUSION

We have exploited the virtual representation of the MIMO wireless channel with ULAs to study its capacity without

the common simplifying assumptions of Gaussian statistics and Kronecker correlation for the channel matrix elements. In particular, the channel capacity is achieved by transmitting independent zero-mean proper-complex Gaussian input symbols along the  $n_t$  virtual transmit angles. In other words, the optimal input covariance matrix is always diagonal in the virtual (Fourier) domain.

Correlated MIMO channels possess fewer degrees of freedom compared to i.i.d. channels and these degrees of freedom are captured by the dominant virtual channel coefficients. Our results indicate that the effective number of dominant virtual coefficients depends on the SNR as well. In general, as the SNR decreases, fewer parallel channels can be used for reliable communication, thereby effectively reducing the multiplexing gain of the MIMO system. This is also consistent with recent results on minimum mean-square error (MMSE) estimation of correlated MIMO channels [8] which show that it is efficient to only estimate the channel coefficients corresponding to a decreasing set of dominant virtual transmit angles as the SNR decreases. In particular, we have shown in this paper that beamforming to one of the virtual transmit angles becomes optimal at low SNRs. Our results also suggest that for moderately correlated channels, beamforming may be nearly optimal for a large range of practical SNRs. In such scenarios, the MIMO channel can be effectively treated as a scalar channel and space–time coding is not required.

Some further comments regarding beamforming are in order in scenarios where it is not optimal but nearly optimal. First, the best virtual angle for beamforming can be determined using only the second moments of the virtual coefficients. In contrast, determining the optimal input distribution requires the entire distribution of the virtual coefficients. Second, the best angle for beamforming can be determined at the receiver and fed back to the transmitter using  $\log_2 n_t$  bits. This is considerably less than the amount of information that needs to be fed back to the

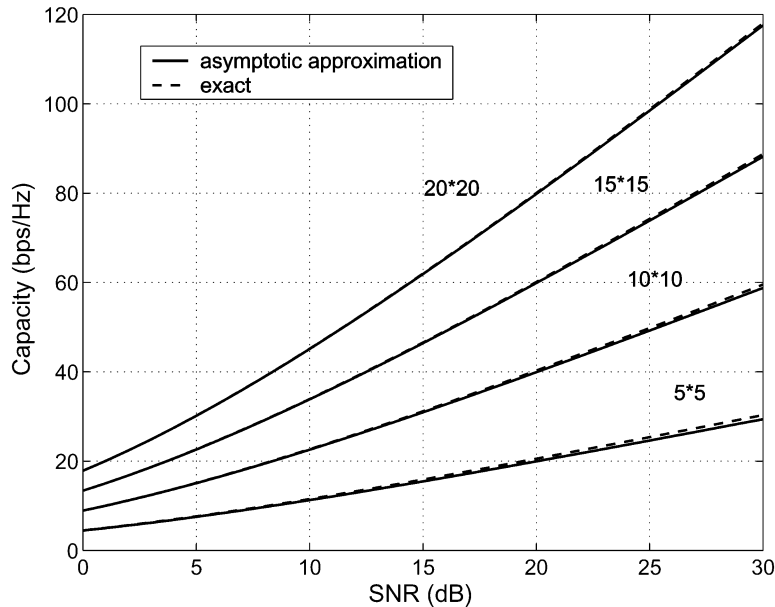


Fig. 7. Comparison of the exact and asymptotic approximations for the capacity for the scattering function shown in Fig. 6.

transmitter for optimal signaling. This indicates that beamforming would be more robust to errors in channel estimation at the receiver. It is of interest to further study this robustness and compare optimal signaling and beamforming in the presence of estimation errors.

The virtual representation also allows for the application of Girko's random matrix result to compute the asymptotic normalized capacity. We have expressed the asymptotic capacity directly in terms of the two-dimensional spatial scattering function of the channel. Thus, while the computation of the nonasymptotic capacity requires detailed information about the marginals of the virtual channel elements, the computation of the asymptotic capacity requires only second-order statistics. Our numerical results show that these asymptotics are accurate for moderate numbers of transmit and receive antennas. This further indicates that the variances of the virtual channel elements may be sufficient in accurately characterizing the capacity even for non-Gaussian channels. The asymptotic results were obtained under the restriction that there is no LOS path connecting the transmitter and receiver. It is clearly of interest to generalize these results to the LOS path case.

The virtual representation can be extended to the most general setting of time- and frequency-selective MIMO channels [18], and in future work we plan to investigate the wide-band channel capacity in both coherent and noncoherent settings. We have taken a first step in this direction in [13], where we investigated the coherent capacity scaling in wide-band correlated MIMO Rayleigh channels for the special case of the  $D$ -diagonal channel model. Furthermore, in [13], we related the capacity scaling to the number of physical propagation paths and identified situations in which capacity scaling can or cannot occur. It is clearly of interest to extend these results to general MIMO channels.

We end with some comments on the restriction of the virtual representation framework to ULAs and possible extensions. We first note that the Kronecker model, while being restrictive

in terms of its assumptions on the channel statistics, applies in principle to arbitrary array geometries. Recently, extensions of the virtual modeling approach have been proposed that apply to arbitrary array geometries without the Kronecker restriction [10], [19], [20]. The basic idea is similar to the virtual representation if we replace the discrete Fourier transform (DFT) matrices (eigenfunctions for ULAs) in the virtual representation with the transmit and receive eigenmatrices for arbitrary channels. The papers [10], [19] investigate the physical conditions under which the resulting transformed channel matrix has uncorrelated entries, whereas in [20] the uncorrelatedness is simply assumed. It is worth noting that such a transformation is environment dependent for arbitrary array geometries (the transmit and receive eigenfunctions change with the environment), whereas for ULAs, the DFT eigenfunctions are fixed regardless of the environment. Furthermore, while the general approach only guarantees uncorrelatedness of the transformed channel coefficients, the elements of the virtual representation for ULAs are also approximately independent due to the virtual path partitioning. Thus, in order to apply the capacity analysis of this paper to arbitrary array geometries, we would need to make the restriction that the channel matrix has proper-complex Gaussian entries.

#### APPENDIX I PROOF OF THEOREM 1

The proof follows the technique in [5, Theorem 3.1]. This technique was also used in [6, Theorem 3].

We define two sets of matrices

$$\begin{aligned}\Omega &:= \{\tilde{Q} : \tilde{Q} \text{ is positive semidefinite, and } \text{Tr}\{\tilde{Q}\} \leq n_t\} \\ \Omega' &:= \{\Lambda : \Lambda \text{ is diagonal, and } \Lambda \in \Omega\}.\end{aligned}$$

We first consider the optimization of the mutual information (16) with  $\tilde{Q}$  restricted to the set  $\Omega'$ . Since the set  $\Omega'$  is convex and compact, and the function  $\mathcal{I}(\tilde{Q})$  is differentiable and strictly

concave over that set, there exists a unique  $\Lambda^\circ$  that maximizes  $\mathcal{I}(\tilde{Q})$  over  $\Omega'$ . By [15, Ch. 7.4, Theorem 2],  $\Lambda^\circ$  satisfies the following necessary condition:

$$\delta\mathcal{I}(\Lambda^\circ; \Lambda - \Lambda^\circ) \leq 0, \quad \forall \Lambda \in \Omega' \quad (29)$$

where

$$\delta\mathcal{I}(\Lambda^\circ; \Lambda - \Lambda^\circ) := \lim_{\alpha \rightarrow 0} \frac{1}{\alpha} [\mathcal{I}(\Lambda^\circ + \alpha(\Lambda - \Lambda^\circ)) - \mathcal{I}(\Lambda^\circ)].$$

The left-hand side of (29) can be computed as follows:

$$\begin{aligned} & \delta\mathcal{I}(\Lambda^\circ; \Lambda - \Lambda^\circ) \\ &= \frac{d}{dx} \mathcal{I}(\Lambda^\circ + x(\Lambda - \Lambda^\circ)) \Big|_{x=0} \\ &= \frac{d}{dx} \mathbb{E} \left[ \log \det \left( I + \frac{\Gamma}{n_t} \tilde{\mathbf{H}}(\Lambda^\circ + x(\Lambda - \Lambda^\circ)) \tilde{\mathbf{H}}^\dagger \right) \right] \Big|_{x=0} \\ &= \mathbb{E} \operatorname{Tr} \left\{ \left( I + \frac{\Gamma}{n_t} \tilde{\mathbf{H}} \Lambda^\circ \tilde{\mathbf{H}}^\dagger \right)^{-1} \frac{\Gamma}{n_t} \tilde{\mathbf{H}} (\Lambda - \Lambda^\circ) \tilde{\mathbf{H}}^\dagger \right\} \end{aligned}$$

where we used the general formula

$$\frac{d}{dx} \log \det(A + xB) = \operatorname{Tr}\{(A + xB)^{-1}B\}$$

with real scalar  $x$  being such that matrix  $A + xB$  is positive definite [21, Example A.3, p. 643]. Then the condition (29) becomes

$$\mathbb{E} \operatorname{Tr} \left\{ \left( I + \frac{\Gamma}{n_t} \tilde{\mathbf{H}} \Lambda^\circ \tilde{\mathbf{H}}^\dagger \right)^{-1} \frac{\Gamma}{n_t} \tilde{\mathbf{H}} (\Lambda - \Lambda^\circ) \tilde{\mathbf{H}}^\dagger \right\} \leq 0 \quad (30)$$

$\forall \Lambda \in \Omega'$ . We now need to show that  $\Lambda^\circ$  remains optimal even when the optimization is performed over the set  $\Omega$ . Since  $\mathcal{I}(\tilde{Q})$  is strictly concave over the convex set  $\Omega$ , it is sufficient to show that

$$\delta\mathcal{I}(\Lambda^\circ; \tilde{Q} - \Lambda^\circ) \leq 0, \quad \forall \tilde{Q} \in \Omega.$$

To this end, consider a  $\tilde{Q} \in \Omega$ . We split  $\tilde{Q}$  into  $\tilde{Q} = \Lambda_{\tilde{Q}} + A$ , where  $\Lambda_{\tilde{Q}}$  is a diagonal matrix with components equal to diagonal entries of  $\tilde{Q}$ , and  $A$  contains the off-diagonal entries of  $\tilde{Q}$ . Then

$$\begin{aligned} & \delta\mathcal{I}(\Lambda^\circ; \tilde{Q} - \Lambda^\circ) \\ &= \mathbb{E} \operatorname{Tr} \left\{ \left( I + \frac{\Gamma}{n_t} \tilde{\mathbf{H}} \Lambda^\circ \tilde{\mathbf{H}}^\dagger \right)^{-1} \frac{\Gamma}{n_t} \tilde{\mathbf{H}} (\tilde{Q} - \Lambda^\circ) \tilde{\mathbf{H}}^\dagger \right\} \\ &= \mathbb{E} \operatorname{Tr} \left\{ \left( I + \frac{\Gamma}{n_t} \tilde{\mathbf{H}} \Lambda^\circ \tilde{\mathbf{H}}^\dagger \right)^{-1} \frac{\Gamma}{n_t} \tilde{\mathbf{H}} (\Lambda_{\tilde{Q}} - \Lambda^\circ) \tilde{\mathbf{H}}^\dagger \right\} \\ &+ \mathbb{E} \operatorname{Tr} \left\{ \left( I + \frac{\Gamma}{n_t} \tilde{\mathbf{H}} \Lambda^\circ \tilde{\mathbf{H}}^\dagger \right)^{-1} \frac{\Gamma}{n_t} \tilde{\mathbf{H}} A \tilde{\mathbf{H}}^\dagger \right\}. \end{aligned}$$

Since  $\Lambda_{\tilde{Q}} \in \Omega'$ , the first term in the preceding equation is less than zero by (30). To evaluate the second term, we denote the

columns of matrix  $\tilde{\mathbf{H}}$  by  $\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \dots, \tilde{\mathbf{h}}_{n_t}$ . Then the second term can be written as

$$\begin{aligned} & \mathbb{E} \operatorname{Tr} \left\{ \left( I + \frac{\Gamma}{n_t} \tilde{\mathbf{H}} \Lambda^\circ \tilde{\mathbf{H}}^\dagger \right)^{-1} \frac{\Gamma}{n_t} \tilde{\mathbf{H}} A \tilde{\mathbf{H}}^\dagger \right\} \\ &= \sum_{\substack{k, \ell=1 \\ k \neq \ell}}^{n_t} \mathbb{E} \operatorname{Tr} \left\{ \left( I + \frac{\Gamma}{n_t} \sum_{i=1}^{n_t} \lambda_i^\circ \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\dagger \right)^{-1} \frac{\Gamma}{n_t} A_{k, \ell} \tilde{\mathbf{h}}_k \tilde{\mathbf{h}}_\ell^\dagger \right\} \end{aligned} \quad (31)$$

where  $\lambda_i^\circ$  is the  $i$ th diagonal entry of  $\Lambda^\circ$ .

In the above sum, consider a particular term

$$\begin{aligned} & \mathbb{E} \operatorname{Tr} \left\{ \left( I + \frac{\Gamma}{n_t} \sum_{i=1}^{n_t} \lambda_i^\circ \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\dagger \right)^{-1} \frac{\Gamma}{n_t} A_{1,2} \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_2^\dagger \right\} \\ &= \operatorname{Tr} \left\{ \mathbb{E} \left[ \mathbb{E} \left[ \left( I + \frac{\Gamma}{n_t} \sum_{i=1}^{n_t} \lambda_i^\circ \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\dagger \right)^{-1} \right. \right. \right. \\ &\quad \left. \left. \left. \cdot \frac{\Gamma}{n_t} A_{1,2} \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_2^\dagger \middle| \tilde{\mathbf{h}}_2, \tilde{\mathbf{h}}_3, \dots, \tilde{\mathbf{h}}_{n_t} \right] \right] \right\}. \end{aligned}$$

Since  $\tilde{\mathbf{H}}$  has independent entries, its columns are independent. Thus, conditioned on  $\tilde{\mathbf{h}}_2, \dots, \tilde{\mathbf{h}}_{n_t}$ , the distribution of  $\tilde{\mathbf{h}}_1$  is unchanged. Since each element of the matrix

$$\left( I + \frac{\Gamma}{n_t} \sum_{i=1}^{n_t} \lambda_i^\circ \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\dagger \right)^{-1} \frac{\Gamma}{n_t} A_{1,2} \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_2^\dagger$$

is an odd function of  $\tilde{\mathbf{h}}_1$ , i.e., if  $\tilde{\mathbf{h}}_1$  is replaced by  $-\tilde{\mathbf{h}}_1$ , each entry of the matrix changes to its antisymmetric value, and by Lemma 1, the inner expectation

$$\begin{aligned} & \mathbb{E} \left[ \left( I + \frac{\Gamma}{n_t} \sum_{i=1}^{n_t} \lambda_i^\circ \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\dagger \right)^{-1} \right. \\ &\quad \left. \cdot \frac{\Gamma}{n_t} A_{1,2} \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_2^\dagger \middle| \tilde{\mathbf{h}}_2, \tilde{\mathbf{h}}_3, \dots, \tilde{\mathbf{h}}_{n_t} \right] = 0. \end{aligned} \quad (32)$$

Hence, the particular term we considered in the sum on the right-hand side of (31) is zero. Following the same reason, all the terms in the sum are zero. Therefore,

$$\delta\mathcal{I}(\Lambda^\circ; \tilde{Q} - \Lambda^\circ) \leq 0, \quad \forall \tilde{Q} \in \Omega$$

which concludes our proof.

## APPENDIX II PROOF OF THEOREM 2

Let  $\{\gamma_k\}_{k=1}^{n_r}$  denote the eigenvalues of  $\tilde{\mathbf{H}} \Lambda \tilde{\mathbf{H}}^\dagger$ .

$$\begin{aligned} \mathcal{I}(\Lambda) &= \mathbb{E} \left[ \log \det \left( I + \frac{\Gamma}{n_t} \tilde{\mathbf{H}} \Lambda \tilde{\mathbf{H}}^\dagger \right) \right] \\ &= \mathbb{E} \left[ \log \prod_{k=1}^{n_r} \left( 1 + \frac{\Gamma}{n_t} \gamma_k \right) \right] \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[ \sum_{k=1}^{n_r} \log \left( 1 + \frac{\Gamma}{n_t} \boldsymbol{\gamma}_k \right) \right] \\
&= \mathbb{E} \left[ \sum_{k=1}^{n_r} \frac{\Gamma}{n_t} \boldsymbol{\gamma}_k \right] + O(\Gamma^2) \\
&= \frac{\Gamma}{n_t} \mathbb{E} \left[ \text{Tr}(\tilde{\mathbf{H}} \Lambda \tilde{\mathbf{H}}^\dagger) \right] + O(\Gamma^2) \\
&= \frac{\Gamma}{n_t} \text{Tr} \left( \Lambda \mathbb{E}[\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}}] \right) + O(\Gamma^2).
\end{aligned}$$

Using the assumption that  $\tilde{\mathbf{H}}$  has independent components, we can compute

$$\begin{aligned}
\mathbb{E} \left[ \tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}} \right]_{i,\ell} &= \mathbb{E} \left[ \sum_{k=1}^{n_r} \tilde{\mathbf{H}}_{i,k}^\dagger \tilde{\mathbf{H}}_{k,\ell} \right] \\
&= \mathbb{E} \left[ \sum_{k=1}^{n_r} \tilde{\mathbf{H}}_{k,i}^* \tilde{\mathbf{H}}_{k,\ell} \right] = \sum_{k=1}^{n_r} V_{k,i} \delta_{i,\ell}.
\end{aligned}$$

Hence,

$$\mathbb{E}[\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}}] = \text{diag} \left( \sum_{k=1}^{n_r} V_{k,1}, \sum_{k=1}^{n_r} V_{k,2}, \dots, \sum_{k=1}^{n_r} V_{k,n_t} \right). \quad (33)$$

Then the mutual information is given by

$$\mathcal{I}(\Lambda) = \frac{\Gamma}{n_t} \sum_{i=1}^{n_t} \lambda_i \left( \sum_{k=1}^{n_r} V_{k,i} \right) + O(\Gamma^2).$$

We want to maximize the first-order low-SNR expansion term of  $\mathcal{I}(\Lambda)$  subject to the constraint

$$\sum_{i=1}^{n_t} \lambda_i \leq n_t \text{ and } \lambda_i \geq 0, \quad \text{for } 1 \leq i \leq n_t.$$

It is straightforward to see that the optimizing  $\Lambda^\circ$  is the beamforming solution with all the power allocated to the transmit virtual angle  $i$  with the largest  $\sum_{k=1}^{n_r} V_{k,i}$  value. If the largest  $\sum_{k=1}^{n_r} V_{k,i}$  is not unique, the power can be spread over those virtual angles corresponding to the largest  $\sum_{k=1}^{n_r} V_{k,i}$  without affecting the capacity.

### APPENDIX III PROOF OF THEOREM 3

We note that the steps of the proof are similar to those given in [6, Theorem 2].

We first consider the condition for beamforming to the first transmit virtual angle to be optimal, and then generalize the condition to the other cases.

The matrix  $\Lambda$  can be parameterized in the following way:

$$\Lambda = \text{diag}\{n_t - p, p\beta_2, \dots, p\beta_{n_t}\} \quad (34)$$

where  $0 \leq p \leq n_t$  and

$$\beta_i \geq 0, \quad \text{for } 2 \leq i \leq n_t; \quad \text{and } \sum_{i=2}^{n_t} \beta_i \leq 1. \quad (35)$$

The mutual information can then be expressed in terms of  $p, \beta_2, \dots, \beta_{n_t}$  as

$$\mathcal{I}(p) = \mathbb{E} \log \det \left( I + \frac{\Gamma}{n_t} (n_t - p) \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^\dagger + \frac{\Gamma}{n_t} \sum_{i=2}^{n_t} p \beta_i \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\dagger \right). \quad (36)$$

The following lemma, which follows directly by the concavity of  $\log \det(\cdot)$  function, establishes a useful property of  $\mathcal{I}(p)$ .

*Lemma 3:* The function  $\mathcal{I}(p)$  in (36) is a strict concave function over  $0 \leq p \leq n_t$  for all  $\{\beta_i\}$  that satisfies the condition (35).

Based on Lemma 3, a necessary and sufficient condition for beamforming to the first transmit virtual angle being optimal is given by

$$\left. \frac{\partial \mathcal{I}(p)}{\partial p} \right|_{p=0} \leq 0 \quad (37)$$

for all  $\{\beta_i\}$  such that (35) is satisfied. The following lemma, whose proof is given in Appendix IV, provides an expression for the derivative on the left-hand side of (37).

*Lemma 4:*

$$\left. \frac{\partial \mathcal{I}(p)}{\partial p} \right|_{p=0} = \frac{\Gamma}{n_t} \sum_{i=2}^{n_t} \beta_i \sum_{k=1}^{n_r} (1 - \mu_{k,1}) V_{k,i} - \frac{1}{n_t} \sum_{k=1}^{n_r} \mu_{k,1}.$$

Thus, to satisfy condition (37), we need

$$\frac{\Gamma}{n_t} \sum_{i=2}^{n_t} \beta_i \sum_{k=1}^{n_r} (1 - \mu_{k,1}) V_{k,i} - \frac{1}{n_t} \sum_{k=1}^{n_r} \mu_{k,1} \leq 0$$

for all  $\{\beta_i\}$ . The first term on the left-hand side of the inequality is maximized when  $\beta_{\ell^\circ} = 1$  where  $\ell^\circ$  is the index with the largest value for  $\sum_{k=1}^{n_r} (1 - \mu_{k,1}) V_{k,\ell}$ . We can hence write the necessary and sufficient condition equivalently as follows:

$$\Gamma \sum_{k=1}^{n_r} (1 - \mu_{k,1}) V_{k,\ell^\circ} - \sum_{k=1}^{n_r} \mu_{k,1} \leq 0$$

where  $\ell^\circ = \arg \max_{1 \leq \ell \leq n_t, \ell \neq 1} \sum_{k=1}^{n_r} (1 - \mu_{k,1}) V_{k,\ell}$ .

Thus far, we have given the necessary and sufficient condition for beamforming to the first virtual angle to be optimal. It is straightforward to generalize the condition to the case where beamforming to the  $i$ th virtual angle is optimal as shown in (18).

We are now left to show that at most one of the  $n_t$  conditions in (18) can be satisfied. Without loss of generality, assume that the condition for beamforming to the first angle is satisfied. Then  $\Lambda^\circ = \text{diag}\{n_t, 0, \dots, 0\}$  which corresponds to  $p = 0$  in (34). Beamforming to the  $i$ th angle with  $i \neq 1$  corresponds to  $p = n_t$ ,  $\beta_i = 1$ , and  $\beta_j = 0$  for all  $j \neq i$ . For such  $\{\beta_i\}$ , since we know that the function  $\mathcal{I}(p)$  is strictly concave, there cannot be another optimal solution other than  $p = 0$ . Hence, beamforming to the other virtual angles cannot be optimal.

APPENDIX IV  
PROOF OF LEMMA 4

We compute  $\frac{\partial \mathcal{I}(p)}{\partial p} \Big|_{p=0}$  as shown in the equation at the bottom of the page, where  $\mathcal{V}_i = \text{diag}(V_{1,i}, \dots, V_{n_r,i})$ . In the equation below, consider  $\mathbb{E} \left[ \frac{\Gamma \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^\dagger}{1 + \Gamma \|\tilde{\mathbf{h}}_1\|^2} \right]$ . Since the column  $\tilde{\mathbf{h}}_1$  has independent entries

$$\mathbb{E} \left[ \frac{\Gamma \tilde{\mathbf{H}}_{k,1} \tilde{\mathbf{H}}_{m,1}^*}{1 + \Gamma \|\tilde{\mathbf{h}}_1\|^2} \right] = 0, \quad \text{for } k \neq m.$$

Now define

$$\mu_{k,1} := \mu_{k,1}(V_{1,1}, V_{2,1}, \dots, V_{n_r,1}) = \mathbb{E} \left[ \frac{\Gamma |\tilde{\mathbf{H}}_{k,1}|^2}{1 + \Gamma \|\tilde{\mathbf{h}}_1\|^2} \right]$$

and note that  $0 \leq \mu_{k,1} \leq 1$ . Then

$$\mathbb{E} \left[ \frac{\Gamma \|\tilde{\mathbf{h}}_1\|^2}{1 + \Gamma \|\tilde{\mathbf{h}}_1\|^2} \right] = \sum_{k=1}^{n_r} \mu_{k,1}$$

and we have

$$\begin{aligned} \frac{\partial \mathcal{I}(p)}{\partial p} \Big|_{p=0} &= \text{Tr} \left\{ \mathcal{M} \frac{\Gamma}{n_t} \sum_{i=2}^{n_t} \beta_i \mathcal{V}_i \right\} - \frac{1}{n_t} \sum_{k=1}^{n_r} \mu_{k,1} \\ &= \frac{\Gamma}{n_t} \sum_{i=2}^{n_t} \beta_i \sum_{k=1}^{n_r} (1 - \mu_{k,1}) V_{k,i} - \frac{1}{n_t} \sum_{k=1}^{n_r} \mu_{k,1} \end{aligned}$$

where  $\mathcal{M} = \text{diag}(1 - \mu_{1,1}, \dots, 1 - \mu_{n_r,1})$ .

APPENDIX V  
PROOF OF THEOREM 4

Replacing  $\Lambda^\circ$  with  $\Lambda$  in (21), we get

$$\bar{C} = \frac{1}{n} \mathbb{E} \left[ \log \det \left( I + \frac{\Gamma}{[\tau n]} \tilde{\mathbf{H}} \Lambda \tilde{\mathbf{H}}^\dagger \right) \right].$$

To streamline the proof, we assume that  $\tau n$  is an integer for all  $n$ , with the understanding that proof is straightforwardly modified when this assumption does not hold. Also, we denote the dependence of certain quantities on  $n$  explicitly using the subscript  $n$ .

Consider the sequence of random variables  $\{\mathbf{d}_n\}_{n=1}^\infty$  defined by

$$\mathbf{d}_n = \frac{1}{n} \log \det \left( I + \frac{\Gamma}{[\tau n]} \tilde{\mathbf{H}} \Lambda \tilde{\mathbf{H}}^\dagger \right)$$

i.e.,  $\bar{C}_n = \mathbb{E}[\mathbf{d}_n]$ . Now let  $\{\gamma_n^{(k)}\}_{k=1}^{k=n}$  denote the eigenvalues of  $\tilde{\mathbf{H}} \Lambda \tilde{\mathbf{H}}^\dagger / n$ , and let  $\sigma = \tau / \Gamma$ . Then

$$\mathbf{d}_n = \frac{1}{n} \sum_{k=1}^n \left[ \log \left( \gamma_n^{(k)} + \sigma \right) - \log \sigma \right] \quad (38)$$

$$= \frac{1}{n} \sum_{k=1}^n \int_0^1 \frac{\gamma_n^{(k)}}{\sigma + t \gamma_n^{(k)}} dt. \quad (39)$$

Let

$$\mathbf{f}_n(t) = \frac{1}{n} \sum_{k=1}^n \frac{\gamma_n^{(k)}}{\sigma + t \gamma_n^{(k)}} = \frac{1}{t} - \frac{\sigma}{t^2} \mathbf{m}_n \left( -\frac{\sigma}{t} \right) \quad (40)$$

where  $\mathbf{m}_n(z)$  is the Stieltjes transform of  $\tilde{\mathbf{H}} \Lambda \tilde{\mathbf{H}}^\dagger / n$ . Based on (39) and (40), we have

$$\bar{C}_n = \mathbb{E}[\mathbf{d}_n] = \mathbb{E} \int_0^1 \mathbf{f}_n(t) dt = \int_0^1 \mathbb{E}[\mathbf{f}_n(t)] dt$$

where the last equality follow from the Fubini theorem, since  $\mathbf{f}_n(t) > 0$ . Taking limits as  $n \rightarrow \infty$

$$\begin{aligned} \bar{C}^\infty &= \lim_{n \rightarrow \infty} \bar{C}_n = \lim_{n \rightarrow \infty} \int_0^1 \mathbb{E}[\mathbf{f}_n(t)] dt \\ &= \int_0^1 \lim_{n \rightarrow \infty} \mathbb{E}[\mathbf{f}_n(t)] dt. \end{aligned}$$

The exchange of the limit and the integral follows from the Dominated Convergence Theorem because of the following boundedness of  $\mathbb{E}[\mathbf{f}_n(t)]$ :

$$\begin{aligned} \mathbb{E}[\mathbf{f}_n(t)] &= \mathbb{E} \left[ \frac{1}{n} \sum_{k=1}^n \frac{\gamma_n^{(k)}}{\sigma + t \gamma_n^{(k)}} \right] \leq \mathbb{E} \left[ \frac{1}{n} \sum_{k=1}^n \frac{\gamma_n^{(k)}}{\sigma} \right] \\ &= \frac{1}{n^2 \sigma} \mathbb{E} \left[ \text{Tr}(\tilde{\mathbf{H}} \Lambda \tilde{\mathbf{H}}^\dagger) \right] \end{aligned}$$

---


$$\begin{aligned} \frac{\partial \mathcal{I}(p)}{\partial p} \Big|_{p=0} &= \mathbb{E} \text{Tr} \left\{ \left( I + \frac{\Gamma}{n_t} (n_t - p) \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^\dagger + \frac{\Gamma}{n_t} \sum_{i=2}^{n_t} p \beta_i \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\dagger \right)^{-1} \cdot \frac{\Gamma}{n_t} \left( \sum_{i=2}^{n_t} \beta_i \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\dagger - \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^\dagger \right) \right\} \Big|_{p=0} \\ &= \mathbb{E} \text{Tr} \left\{ \left( I + \Gamma \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^\dagger \right)^{-1} \frac{\Gamma}{n_t} \left( \sum_{i=2}^{n_t} \beta_i \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\dagger - \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^\dagger \right) \right\} \\ &= \text{Tr} \left\{ \mathbb{E} \left[ \left( I + \Gamma \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^\dagger \right)^{-1} \right] \frac{\Gamma}{n_t} \mathbb{E} \left[ \sum_{i=2}^{n_t} \beta_i \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\dagger \right] - \mathbb{E} \left[ \left( I + \Gamma \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^\dagger \right)^{-1} \frac{\Gamma}{n_t} \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^\dagger \right] \right\} \\ &= \text{Tr} \left\{ \mathbb{E} \left[ I - \frac{\Gamma \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^\dagger}{1 + \Gamma \|\tilde{\mathbf{h}}_1\|^2} \right] \frac{\Gamma}{n_t} \sum_{i=2}^{n_t} \beta_i \mathcal{V}_i - \mathbb{E} \left[ \left( I - \frac{\Gamma \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^\dagger}{1 + \Gamma \|\tilde{\mathbf{h}}_1\|^2} \right) \frac{\Gamma}{n_t} \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^\dagger \right] \right\} \\ &= \text{Tr} \left\{ \left( I - \mathbb{E} \left[ \frac{\Gamma \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^\dagger}{1 + \Gamma \|\tilde{\mathbf{h}}_1\|^2} \right] \right) \frac{\Gamma}{n_t} \sum_{i=2}^{n_t} \beta_i \mathcal{V}_i \right\} - \frac{1}{n_t} \mathbb{E} \left[ \frac{\Gamma \|\tilde{\mathbf{h}}_1\|^2}{1 + \Gamma \|\tilde{\mathbf{h}}_1\|^2} \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n^2\sigma} \text{Tr} \left( \Lambda \mathbf{E}[\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}}] \right) \\
&= \frac{1}{n^2\sigma} \sum_{i=1}^{n\tau} \lambda_i \left( \sum_{k=1}^n V_{k,i} \right)
\end{aligned}$$

where the last equality follows from (33). Now using (12) and Assumption 2, it is clear that the right-hand side of the preceding inequality is uniformly bounded for all  $n$  and  $t$ . Thus,

$$\bar{C}^\infty = \int_0^1 \left( \frac{1}{t} - \frac{\sigma}{t^2} \lim_{n \rightarrow \infty} \mathbf{E} \left[ \mathbf{m}_n \left( -\frac{\sigma}{t} \right) \right] \right) dt.$$

Finally, we have

$$\lim_{n \rightarrow \infty} \mathbf{E} \left[ \mathbf{m}_n \left( -\frac{\sigma}{t} \right) \right] = m \left( -\frac{\sigma}{t} \right)$$

by using the convergence assumed in the statement of the theorem and the fact that  $|\mathbf{m}_n(\sigma/t)| \leq t/\sigma$ .

#### REFERENCES

- [1] G. J. Foschini and M. Gans, "On limits of wireless communications in fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, 1998.
- [2] İ. E. Telatar, "Capacity of multiantenna gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.
- [3] A. F. Molisch, M. Steinbauer, M. Toeltsch, E. Bonek, and R. S. Thomä, "Capacity of MIMO systems based on measured wireless channels," *IEEE J. Sel. Areas. Commun.*, vol. 20, no. 3, pp. 561–569, Apr. 2002.
- [4] C.-N. Chua, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.
- [5] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.
- [6] S. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, Jul. 2004.
- [7] E. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless. Commun.*, vol. 3, no. 5, pp. 1543–1553, Sep. 2004.
- [8] J. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.
- [9] J. Kotecha and A. Sayeed, "On the capacity of correlated MIMO channels with covariance feedback," in *Proc. 2003 Int. Symp. Information Theory*, Yokohama, Japan, Jun./Jul. 2003, p. 355.
- [10] W. Weichselberger, M. Herdin, H. Ozelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. Wireless Commun.*, to be published.
- [11] A. M. Sayeed, "Deconstructing multi-antenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.
- [12] M. A. Kamath and B. L. Hughes, "On the capacity of large arrays over fading channels," in *Proc. 40th Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2002, pp. 807–816.
- [13] K. Liu, V. Raghavan, and A. M. Sayeed, "Capacity scaling and spectral efficiency in wide-band correlated MIMO channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2504–2526, Oct. 2003.
- [14] W. C. Jakes Jr, *Microwave Mobile Communications*. New York: Wiley, 1974.
- [15] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1997.
- [16] A. Mantravadi, "Analysis and Design of Wideband Multiantenna CDMA Systems," Ph.D. dissertation, Cornell University, Ithaca, NY, 2002.
- [17] V. L. Girko, *Theory of Random Determinants*. Amsterdam, The Netherlands: Kluwer Academic, 1990.
- [18] A. M. Sayeed and V. Veeravalli, "The essential degrees of freedom in space-time fading channels," in *Proc. 13th IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications (PIMRC'02)*, vol. 4, Lisbon, Portugal, Sep. 2002, pp. 1512–1516.
- [19] J. Kotecha and A. M. Sayeed, "Canonical statistical modeling and capacity analysis of correlated MIMO channels," *IEEE Trans. Wireless Commun.*, submitted for publication.
- [20] A. M. Tulino, A. Lozano, and S. Verdú, "Capacity-achieving input covariance matrix for correlated multi-antenna channels," in *Proc. 2003 Allerton Conf. Computing, Control and Communications*, Monticello, IL, Oct. 2003, pp. 242–251.
- [21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.