Canonical statistical modeling and capacity analysis of correlated MIMO fading channels

Jayesh H. Kotecha* and Akbar M. Sayeed

Department of Electrical and Computer Engineering,

University of Wisconsin at Madison, Madison, 53706.

Email: jkotecha@ece.wisc.edu, akbar@engr.wisc.edu

This research is supported in part by NSF Grant Nos. CCR-9875805 and CCR-0113385 and ONR Grant No. N00014-01-1-0825.

Abstract

We introduce canonical statistical models for spatially correlated, multi-input multi-output (MIMO) channels with Rayleigh fading, based on canonical decompositions of the channel matrix. The proposed models encompass a large family of channels having structured covariances and generalize existing models like the Kronecker model and the virtual representation. An important property of these models is their characterization of MIMO channels as interactions among transmit and receive eigen spaces, which provides a structured framework and new insights to understand the complexities of correlated MIMO channels. The analysis points to the fundamental differences of correlated and i.i.d. MIMO channels, that unlike i.i.d channels, correlated channels have unequal degrees of freedom and parallel eigen-channels, and invite optimal design of the transmitted signal. We derive the optimal input covariance matrix which achieves coherent capacity assuming knowledge of only the channel covariance matrix at the transmitter and show that its eigenvectors are identical to the transmit eigenvectors. The optimal power assignment can be obtained numerically and it is shown that at high signal-to-noise ratio (SNR) all transmit eigen subspaces get equal power while at low SNR beamforming is optimal. Numerical results show a non-trivial gain in capacity due to efficient use of covariance feedback. Since the proposed models unify and extend many existing channel models, our results generalize many recent results.

Index Terms

Correlated MIMO fading channels, coherent capacity, optimal signal design, beamforming, canonical statistical models.

I. INTRODUCTION

Under the assumption of independent and identically distributed (i.i.d.) Rayleigh fading multi-antenna links, it was shown in [1],[2] that MIMO systems achieve a linear growth in multiplexing gain and coherent Shannon capacity with the number of antennas, which motivated extensive research in the field. However, the rich scattering, i.i.d assumption is idealistic and most realistic channels exhibit clustered scattering and spatially correlated links. This is motivated by the physics of the underlying scattering environment [3],[4],[5] and confirmed by measurement based analysis, see for example [6]. Correlated MIMO channels have been analyzed mainly in the contexts of the so-called Kronecker model [3],[5] and the virtual representation model for uniform linear arrays (ULAs) [4]. The Kronecker model assumes separability in channel statistics induced by the transmitter and receiver arrays, which limits the degrees of freedom in modeling the channel and though valid in certain scattering environments [6],[5] limits its generality. The virtual representation model does not assume such separability, but is applicable only for ULAs.

In this paper, we propose new canonical statistical models for Rayleigh fading MIMO channels based on canonical decompositions of the channel matrix which provide new insights in the structure of a large class of correlated MIMO channels. Motivated by the virtual representation model, these models do not assume separable statistics, but are applicable for general array geometries. A MIMO channel matrix can be decomposed into a canonical channel matrix sandwiched between transmit and receive unitary matrices. We obtain a rich family of channel models called canonical statistical models by making certain assumptions on the channel covariance structures which dictate the properties of the canonical channel. In particular, we assume that the auto and cross covariances of the rows and/or columns (the corresponding auto and cross covariances of the respective MISO and SIMO links) have identical matrices of eigenvectors (eigen-matrices). As a result, three canonical models arise in which the canonical channel matrix either has uncorrelatedness amongst the columns or rows or all its elements are uncorrelated, see Figure 1. The last case, denoted here as CM3, was first suggested in [7],[8] by the authors and is further generalized here into a richer class of models. Geometrically, this model characterizes MIMO channels as coupling gains (given by the canonical channel matrix) between the basis vectors of the transmit and receive eigen spaces, where the basis vectors are exactly the columns of the transmit and receive unitary matrices in the canonical decomposition. Furthermore, while the transmit and receive matrices provide directional information of the scattering in the channel, the canonical channel matrix embodies the true degrees of freedom and the uncorrelated parallel eigen-channels available in the channel. An important property of these models is that linear operations at the transmitter and receiver allow transmission in the eigen domain in which the parallel channels are decorrelated. Special cases of CM3 are the Kronecker model (stronger assumptions are made on the covariances) and the virtual representation for ULAs (the transmit and receive unitary matrices become discrete Fourier matrices). We note that CM3 has been independently developed in [9] with the motivation of fitting measured data.

The capacity achieving input distribution for MIMO channels with additive white Gaussian noise is zero mean Gaussian and for i.i.d. channels, its covariance matrix is a scalar multiple of the identity matrix [1]. A key result for the optimal covariance matrix for correlated multi-input single-output (MISO) channels was found in [10], and that for MIMO channels with only transmit covariance in [11], the Kronecker model in [12], [13], and the virtual representation for ULAs in [14]. Here, we characterize the capacity achieving optimal input covariance matrix for

the proposed channel models. It is assumed that the channel is perfectly known to the receiver but not to the transmitter and the transmitter has perfect knowledge of the channel covariance matrix. It is shown, that it is optimal to transmit along the transmit eigenvectors[1], which is consistent with all previously known results and generalizes them. The analysis also points at a fundamental difference between i.i.d. and correlated channels, mainly that, unlike i.i.d. channels, the degrees of freedom and parallel eigen-channels in correlated MIMO channels are non-identical. The chief outcome of this difference is that, unlike i.i.d. MIMO channels, optimal transmit signal design is a function of the SNR and in general non-isotropic[2]. This has been shown for capacity optimization in this paper and [10],[11],[12],[13],[15], for optimal channel estimation in [8] and for space-time coding in [16]. In particular, at high SNR the number of *dominant*[3] transmit subspaces d_s is equal to the number of transmit antennas[4] and all get equal power, d_s reduces with SNR and the eigen subspaces get unequal powers, and $\mathrm{finally}$ at low SNR beamforming becomes optimal, i.e. only the strongest transmit subspace gets all the power[5]. As a result, the spatial multiplexing gain (the number of independent parallel channels) decreases with SNR. Numerical results show a non-trivial gain in capacity due to $\mathrm{efficient}$ use of covariance feedback.

Section II describes the canonical statistical models for correlated MIMO channels. In section III, the capacity achieving optimal input covariance matrix is obtained. Section IV discusses related results on channel estimation and space-time coding. Some simulation results are presented in section V and section VI concludes with a summary and future directions.

Notation: If $\mathbf{X}$ is a $Q \times K$ matrix, $\mathbf{x} = \mathrm{vec}(\mathbf{X})$ denotes the $QK \times 1$ vector obtained by stacking columns of $\mathbf{X}$. $\otimes$ denotes the Kronecker product. $\mathbf{X}^*, \mathbf{X}^T, \mathbf{X}^H$ denote the complex conjugate, transpose, and Hermitian transpose of $\mathbf{X}$. The inverse, trace and determinant of $\mathbf{X}$ are denoted by $\mathbf{X}^{-1}$, $\mathrm{tr}(\mathbf{X})$ and $\det(\mathbf{X})$. $E(\cdot)$ denotes the expectation operator. $\mathbf{X}[m, n]$ denotes the entry in the m-th row and n-th column of $\mathbf{X}$.

[1]Transmission along a basis vector of the transmit eigen space is denoted herein as transmission along a transmit subspace.

[2]Isotropic transmission implies equal power is transmitted along the eigen-channels.

[3]A subspace is dominant if it gets a non-zero power assignment.

[4]In general, this is a maximum number, because if a particular transmit subspace does not have a non-zero coupling with atleast one receive subspace, it is dropped and the power is distributed equally among the rest of the transmit subspaces.

[5]The powers allocated are not only a function of SNR but also of the optimization criteria like capacity (discussed here), MMSE for channel estimation [8] and pairwise error probability for space-time coding[16].

II. Canonical modeling of correlated MIMO Channels

Consider a narrowband, frequency non-selective, Rayleigh fading MIMO channel with M_t transmit and M_r receive antennae. If $\mathbf{s}$ is the transmit vector of dimension M_t, then the M_r-dimensional received signal $\mathbf{x}$ can be written as

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n} \tag{1}$$

where $\mathbf{H}$ is the $M_r \times M_t$ channel matrix coupling the transmitter and receiver antennae. $\mathbf{n}$ is the M_t-dimensional noise vector, which is assumed to be zero mean, complex white Gaussian with covariance matrix $\sigma^2\mathbf{I}_Q$. The channel gain between the $n-th$ receive and $m-th$ transmit antenna is the corresponding entry in the matrix $\mathbf{H}$ denoted by $\mathbf{H}[n,m]$. Assume that the entries of $\mathbf{H}$ are zero mean complex Gaussian with channel covariance $\mathbf{R} = E(\mathbf{h}\mathbf{h}^H)$, where $\mathbf{h} = \mathrm{vec}(\mathbf{H})$.

We now describe some canonical decompositions of MIMO channels. Let $\mathbf{Q}_t := E\left(\mathbf{H}^H\mathbf{H}\right)$ and $\mathbf{Q}_r := E\left(\mathbf{H}\mathbf{H}^H\right)$ and their EVDs be given by $\mathbf{U}_t\mathbf{D}_t\mathbf{U}_t^H$ and $\mathbf{U}_r\mathbf{D}_r\mathbf{U}_r^H$ respectively, where $\mathbf{U}_t$ and $\mathbf{U}_r$ are the matrices of eigenvectors and $\mathbf{D}_t$ and $\mathbf{D}_r$ are diagonal matrices of eigenvalues. Then any channel matrix $\mathbf{H}$ can be written in the canonical form

$$CF1: \quad \mathbf{H} = \mathbf{H}_t\mathbf{U}_t^H \tag{2}$$

such that $E(\mathbf{H}_t^H\mathbf{H}_t) = \mathbf{D}_t$ and the elements of $\mathbf{H}_t$ are zero mean, Gaussian with covariance given by $\mathbf{R}_t = E(\mathbf{h}_t\mathbf{h}_t^H) = (\mathbf{U}_t^* \otimes \mathbf{I})^H\mathbf{R}(\mathbf{U}_t^* \otimes \mathbf{I})$, where $\mathbf{h}_t = \mathrm{vec}(\mathbf{H}_t)$.

Similarly, any channel matrix $\mathbf{H}$ can be written in the canonical form

$$CF2: \quad \mathbf{H} = \mathbf{U}_r\mathbf{H}_r \tag{3}$$

such that $E(\mathbf{H}_r\mathbf{H}_r^H) = \mathbf{D}_r$ and the elements of $\mathbf{H}_r$ are zero mean, Gaussian with covariance given by $\mathbf{R}_r = E(\mathbf{h}_r\mathbf{h}_r^H) = (\mathbf{I} \otimes \mathbf{U}_r)^H\mathbf{R}(\mathbf{I} \otimes \mathbf{U}_r)$, where $\mathbf{h}_r = \mathrm{vec}(\mathbf{H}_r)$.

Furthermore, $\mathbf{H}$ can be written in the canonical form

$$CF3: \quad \mathbf{H} = \mathbf{U}_r\mathbf{H}_c\mathbf{U}_t^H \tag{4}$$

such that $E(\mathbf{H}_c^H\mathbf{H}_c) = \mathbf{D}_t$ and $E(\mathbf{H}_c\mathbf{H}_c^H) = \mathbf{D}_r$ and the elements of $\mathbf{H}_c$ are zero mean, Gaussian with covariance given by $\mathbf{R}_c = E(\mathbf{h}_c\mathbf{h}_c^H) = (\mathbf{U}_t^* \otimes \mathbf{U}_r)^H\mathbf{R}(\mathbf{U}_t^* \otimes \mathbf{U}_r)$, where $\mathbf{h}_c = \mathrm{vec}(\mathbf{H}_c)$.

It is possible to obtain some interesting models and tractable analysis if we make some simplifying assumptions. The following notation is used: For the channel matrix $\mathbf{H}$, the column vectors $\underline{\mathbf{g}}_i$ and $\underline{\mathbf{h}}_j$ denote the i-th column of $\mathbf{H}^H$ (row of of $\mathbf{H}$) and j-th column of $\mathbf{H}$ respectively, i.e. $\mathbf{H} = [\underline{\mathbf{h}}_1 \ldots \underline{\mathbf{h}}_{M_t}] = [\underline{\mathbf{g}}_1 \ldots \underline{\mathbf{g}}_{M_r}]^H$.

A. Canonical model 1 (CM1)

We shall denote as CM1, a channel which follows the following two assumptions:

Assumption 1: The covariance matrices of all rows of $\mathbf{H}$, given by $E(\underline{\mathbf{g}}_i\ \underline{\mathbf{g}}_i^H)\ \ \forall i$, have the same eigenvectors. Let these eigenvectors be denoted by the columns of $\mathbf{U}_t$.

Assumption 2: The cross covariance matrices of the rows of $\mathbf{H}$ also have the same eigenvectors given by the columns of $\mathbf{U}_t$, i.e. $E(\underline{\mathbf{g}}_i\ \underline{\mathbf{g}}_j^H) = \mathbf{U}_t\mathbf{\Lambda}_{t,ij}\mathbf{U}_t^H\ \ \forall i,j;\ i \neq j$.

From these two assumptions, $\mathbf{H}$ can be written in CF1 (2) where $\mathbf{H}_t$ has the following properties:

- The covariance matrix of each row of $\mathbf{H}_t$, given by $E(\underline{\mathbf{g}}_{ti}\ \underline{\mathbf{g}}_{ti}^H)$, is diagonal - This follows directly from Assumption 1 and $\underline{\mathbf{g}}_{ti} = \mathbf{U}_t^H\underline{\mathbf{g}}_i$.
- The columns of $\mathbf{H}_t$ are uncorrelated to each other i.e. $E(\underline{\mathbf{h}}_{ti}\ \underline{\mathbf{h}}_{tj}^H) = \mathbf{0},\ \ \forall i,j;\ i \neq j$ (the columns may have arbitrary covariances) - This follows from Assumptions 1 and 2, which state that $E(\underline{\mathbf{g}}_{ti}\ \underline{\mathbf{g}}_{tj}^H) = \mathbf{U}_t^H E(\underline{\mathbf{g}}_i\ \underline{\mathbf{g}}_j^H)\mathbf{U}_t = \mathbf{\Lambda}_{ij}$ is diagonal $\forall i,j$, i.e. any element in a given column of $\mathbf{H}_t$ is uncorrelated to elements in the rest of the columns.

We shall summarize CM1 in the following lemma:

Lemma 1: CM1: Under Assumptions 1 and 2, any channel can be written in CF1 $\mathbf{H} = \mathbf{H}_t\mathbf{U}_t^H$, where the columns of $\mathbf{H}_t$ are uncorrelated to each other.

B. Canonical model 2 (CM2)

We shall denote as CM2, a channel which follows the following two assumptions:

Assumption 3: The covariance matrix of the columns of $\mathbf{H}$, given by $E(\underline{\mathbf{h}}_i\ \underline{\mathbf{h}}_i^H)\ \ \forall i$, have identical eigenvectors. Let these eigenvectors be denoted by the columns of $\mathbf{U}_r$.

Assumption 4: The cross covariance matrices of the columns of $\mathbf{H}$ also have the same eigenvectors given by the columns of $\mathbf{U}_r$, i.e. $E(\underline{\mathbf{h}}_i\ \underline{\mathbf{h}}_j^H) = \mathbf{U}_r\mathbf{\Lambda}_{r,ij}\mathbf{U}_r^H\ \ \forall i,j;\ i \neq j$.

Lemma 2: CM2: Under Assumptions 3 and 4, any channel can be written in CF2 $\mathbf{H} = \mathbf{U}_r\mathbf{H}_r$, where the rows of $\mathbf{H}_r$ are uncorrelated to each other.

C. Canonical model 3 (CM3)

We shall denote as CM3, a channel which follows Assumptions 1-4. We have the following:

Lemma 3: CM3: Under Assumptions 1-4, any channel can be written in CF3 $\mathbf{H} = \mathbf{U}_r\mathbf{H}_c\mathbf{U}_t^H$, where the elements of $\mathbf{H}_c$ are uncorrelated (but not necessarily identically distributed).

Proof: From (2) and (4), we can write $\mathbf{H}_c = \mathbf{U}_r^H\mathbf{H}_t$. Then, the cross covariance of the columns of $\mathbf{H}_c$ is $E(\underline{\mathbf{h}}_{ci}\ \underline{\mathbf{h}}_{cj}^H) = \mathbf{U}_r^H E(\underline{\mathbf{h}}_{ti}\ \underline{\mathbf{h}}_{tj}^H)\mathbf{U}_r = \mathbf{0}\quad \forall\ i,j;\ i \neq j$, which follows from the

uncorrelatedness of the columns of $\mathbf{H}_t$ from Lemma 1. Thus, the columns of $\mathbf{H}_c$ are uncorrelated to each other. Similarly, from (3) and (4), we can write $\mathbf{H}_c = \mathbf{H}_r\mathbf{U}_t$. Then, the cross covariance of the rows of $\mathbf{H}_c$ is $E(\underline{\mathbf{g}}_{ci}\ \underline{\mathbf{g}}_{cj}^H) = \mathbf{U}_t^H E(\underline{\mathbf{g}}_{ri}\ \underline{\mathbf{g}}_{rj}^H)\mathbf{U}_t = \mathbf{0} \quad \forall\ i,j;\ i \neq j$, which follows from the uncorrelatedness of the rows of $\mathbf{H}_r$ from Lemma 2. Thus, the rows of $\mathbf{H}_c$ are uncorrelated to each other. Since the rows of $\mathbf{H}_c$ are uncorrelated to each other and the columns of $\mathbf{H}_c$ are uncorrelated to each other, all the elements of $\mathbf{H}_c$ are uncorrelated. ■

Thus in CM3, the channel covariance matrix can be written as

$$\mathbf{R} = (\mathbf{U}_t^* \otimes \mathbf{U}_r)\mathbf{R}_c(\mathbf{U}_t^* \otimes \mathbf{U}_r)^H, \tag{5}$$

where $\mathbf{R}_c$ is diagonal. Note that the RHS is an EVD of $\mathbf{R}$. The matrices $\mathbf{U}_t$ and $\mathbf{U}_r$, which are the eigen matrices of $\mathbf{Q}_t = E\left(\mathbf{H}^H\mathbf{H}\right)$ and $\mathbf{Q}_r = E\left(\mathbf{H}\mathbf{H}^H\right)$ respectively, can be interpreted as transmit and receive eigen-matrices. This model was first introduced by the authors in [7],[8]. Note that, clearly, CM3 is a special case of CM1 (CM2), where the covariance matrices of the columns (rows) of $\mathbf{H}_t$ ($\mathbf{H}_r$) have the same eigen matrix $\mathbf{U}_r$ ($\mathbf{U}_t$). In fact, CM3 is an intersection of CM1 and CM2, which points that CM3 captures the interaction between the transmit and receive spaces $\mathrm{identified}$ in CM1 and CM2. We elaborate on this interpretation in the following subsection. Then, we show that CM3 represents a large class of channels by discussing its correspondence to two well known channel models.

D. Transmit and receive eigen spaces and their interaction

The canonical decomposition in CM3 gives an accurate mathematical representation of the interaction of the transmit and receive eigen spaces whose basis vectors are precisely the columns of $\mathbf{U}_t$ and $\mathbf{U}_r$. Consider a transmitted signal vector $\mathbf{s}$ ($M_t \times 1$) and the ($M_r \times 1$) received vector $\mathbf{x} = \mathbf{H}\mathbf{s} = \mathbf{U}_r\mathbf{H}_c\mathbf{U}_t^H\mathbf{s}$ (ignoring noise). Thus, $\mathbf{s}$ is first projected onto the transmit basis, transmitted along the canonical channel $\mathbf{H}_c$ and then projected along the received basis. Since the elements of $\mathbf{H}_c$ are uncorrelated, the interaction between the transmit and receive spaces is governed completely by $\mathbf{R}_c$ on an average. Since $\mathbf{H}$ and $\mathbf{H}_c$ are unitarily equivalent, the MIMO channel can be completely characterized by $\mathbf{H}_c$ (given $\mathbf{U}_t$ and $\mathbf{U}_r$ which provide only directional information). This gives an equivalent transmission scheme in the eigen domain (see $\mathrm{Figure}\ 2$)

$$\mathbf{x}_c = \mathbf{H}_c\mathbf{s}_c + \mathbf{n}_c \tag{6}$$

where $\mathbf{x}_c := \mathbf{U}_r^H\mathbf{x}, \mathbf{s}_c := \mathbf{U}_t^H\mathbf{s}$ and $\mathbf{n}_c$ is complex white Gaussian noise. Denote transmission along the i-th transmit eigenvector as transmission along the i-th transmit subspace (similarly for

receive). Then, $\mathbf{H}_c[i,j]$ denotes the gain of a signal transmitted along the i-th transmit subspace and received along the j-th receive subspace. An important consequence of CM3 is that, since the elements of $\mathbf{H}_c$ are uncorrelated, it implies that a *linear transformation* at the transmitter and receiver results in $\min(M_t, M_r)$ uncorrelated parallel channels. Define the *subspace connectivity* matrix $\mathbf{P}_s$ as a $M_r \times M_t$ matrix of the variances of the elements of the canonical channel $\mathbf{H}_c$ obtained by stacking the diagonal elements of $\mathbf{R}_c$, i.e.

$$\mathbf{P}_s[i,j] = E\left(|\mathbf{H}_c[i,j]|^2\right). \tag{7}$$

We note the following:

- *Joint statistics* - $\mathbf{H}_c$ captures the *joint* transmitter and receiver statistics indicated by $\mathbf{P}_s$ which are in general non-separable, in comparison to the Kronecker model which assumes separable statistics at the transmitter and receiver (see next subsection). $\mathbf{U}_t$, $\mathbf{U}_r$ and $\mathbf{R}_c$ can be completely obtained from $\mathbf{R}$.
- *Degrees of freedom* - Define the degrees of freedom available in the channel as the elements of $\mathbf{H}_c$ having non-zero variances. Thus, the number of degrees of freedom $d = \mathrm{rank}(\mathbf{R}_c) = \mathrm{rank}(\mathbf{R}) \leq M_t M_r$ and their weights (or powers) are clearly exposed by the corresponding elements of $\mathbf{P}_s$. On the other hand, i.i.d. channels have $d = M_t M_r$ and all the degrees of freedom have equal weights. Transmitting along the (say) i-th transmit subspace excites a maximum of M_r degrees of freedom and their corresponding elements in $\mathbf{P}_s$ denote the powers with which they are excited. This implies that, in general, unlike i.i.d. channels, the degrees of freedom cannot be excited equally since they have unequal weights.
- *Parallel channels* - The number of parallel channels in i.i.d channels is $\min(M_t, M_r)$ and have identical statistics. However, for correlated channels, the non-zero columns of $\mathbf{P}_s$ expose the exact number of available parallel channels which may be less than or equal to $\min(M_t, M_r)$ and show that they are not all created equal.

The last two important observations signify the key differences in correlated versus i.i.d channels. Since the degrees of freedom and parallel channels are unequal they should be excited appropriately to obtain optimal transmission. This manifests itself in a number of problems including capacity (see Section III), channel estimation [7], [8] and space-time coding [16].

E. Special cases

In the following, we discuss the correspondence of CM3 with two known channel models, namely the so-called Kronecker model and the virtual channel representation.

1) Uniform Linear Arrays and the Virtual Channel Representation: In [4], the *virtual* channel representation is proposed for ULAs at both the transmitter and receiver, according to which the MIMO channel can be written as $\mathbf{A}_r\mathbf{H}_v\mathbf{A}_t$ where $\mathbf{A}_r$ and $\mathbf{A}_t$ are discrete Fourier transform (DFT) matrices. It is argued in [4], that the elements of $\mathbf{H}_v$ are *approximately uncorrelated* for finite number of antennas and the approximation become increasingly better with the number of antennas. We shall now inspect the correspondence between CM3 and the virtual representation. It can be easily $\mathrm{verified}$ for ULAs that under the uncorrelated scattering assumption, the transmit and receive covariance matrices $\mathbf{R}_t = E(\mathbf{H}^H\mathbf{H})$ and $\mathbf{R}_r = E(\mathbf{H}\mathbf{H}^H)$ are Toeplitz matrices. It is well known that, asymptotically (in the number of antennas), the eigen matrices of Toeplitz matrices and hence $\mathbf{R}_t$ and $\mathbf{R}_r$ converge to DFT matrices. In addition, Assumptions 1-4 made for CM3 are $\mathrm{satisfied}$ by the virtual representation. This shows the asymptotic equivalence of the canonical decomposition and the virtual representation for ULAs. However, while the transmit and receive basis in CM3 are a function of the channel statistics and the elements of the canonical decomposition are *exactly* uncorrelated, the virtual representation assumes the transmit and receive eigen matrices to be fixed DFT matrices independent of the channel statistics and the elements of $\mathbf{H}_V$ to be *approximately* uncorrelated. Thus, the virtual representation can be very convenient to use especially for large number antennas, not only for the physical insights it provides but also because it makes transmit signal design easier since the transmit and receive basis are fixed and need not be obtained from the channel statistics. In this work, we extend the virtual representation framework to arbitrary arrays via CM3.

2) Kronecker Model: This channel model assumes separable statistics at the transmitter and receiver. The channel matrix and covariance matrix can be written as

$$\mathbf{H} = \mathbf{\Sigma}_r^{1/2}\mathbf{H}_w\mathbf{\Sigma}_t^{1/2}, \qquad \mathbf{R} = \mathbf{\Sigma}_t \otimes \mathbf{\Sigma}_r \tag{8}$$

where the elements of $\mathbf{H}_w$ are i.i.d. The matrices $\mathbf{\Sigma}_T$ and $\mathbf{\Sigma}_R$ are the transmit and receive covariance matrices. This model was assumed in [3], derived in [5] using the 'one ring' ray tracing model [17] assuming that the transmitter is relatively unobstructed while the receiver is surrounded by local scatterers (see [5] for other assumptions made), and $\mathrm{verified}$ by measurements under certain environments in [18],[6]. Denote the eigen value decompositions (EVD) of $\mathbf{\Sigma}_t$ and $\mathbf{\Sigma}_r$ as $\mathbf{U}_t\Lambda_t\mathbf{U}_t^H$ and $\mathbf{U}_r\Lambda_r\mathbf{U}_r^H$, where $\mathbf{U}_t$ and $\mathbf{U}_r$ are the matrices of eigenvectors and Λ_t and Λ_r are diagonal matrices of eigenvalues respectively. We can write (8) in the form CM3

$$\mathbf{H} = \mathbf{U}_r\Lambda_r^{1/2}\mathbf{U}_r^H\mathbf{H}_w\mathbf{U}_t\Lambda_t^{1/2}\mathbf{U}_t^H = \mathbf{U}_r\mathbf{H}_c\mathbf{U}_t^H \tag{9}$$

where the second equality arises from the following two observations. First, the elements of $\mathbf{U}_r^H \mathbf{H}_w \mathbf{U}_t$ are still i.i.d. and second, the pre- and post-multiplication of diagonal matrices $\Lambda_r^{1/2}$ and $\Lambda_t^{1/2}$ respectively makes the elements of $\mathbf{H}_c = \Lambda_r^{1/2} \mathbf{U}_r^H \mathbf{H}_w \mathbf{U}_t \Lambda_t^{1/2}$ uncorrelated with diagonal covariance matrix given by

$$\mathbf{R}_c = \Lambda_t \otimes \Lambda_r \tag{10}$$

which has a Kronceker form. The Kronecker form of $\mathbf{R}_c$ implies that the coupling vector between each transmit subspace and the receive subspaces is identical (up to a constant determined by the transmit covariance matrix), i.e

$$\mathbf{P}_{s,i} = \Lambda_t[i,i] \cdot [\ \Lambda_r[1,1], \Lambda_r[2,2], \ldots, \Lambda_r[M_r, M_r]\ ]^T, \tag{11}$$

where $\mathbf{P}_{s,i}$ is the i-th column of $\mathbf{P}_s$. However this is a direct result of assuming separable statistics at the transmitter and receiver and does not hold in general. Thus the Kronceker model is not rich enough in general (see [15] which points the same using scattering models) and an arbitrarily diagonal $\mathbf{R}_c$ may be needed as in CM3 which does not assume separable statistics.

III. Coherent Capacity

In this section, it is assumed that perfect channel estimates are available at the receiver, but the transmitter has perfect knowledge of only the channel covariance matrix $\mathbf{R}$. It has been shown in [1], that the Shannon capacity of a Rayleigh fading MIMO channel with additive Gaussian noise is achieved by a zero mean, complex, Gaussian vector with covariance matrix $\mathbf{Q}$ such that,

$$C = \max_{\mathbf{Q}\,:\,\mathrm{tr}(\mathbf{Q}) \le \beta} \; E_{\mathbf{H}} \left[\log\det\left(\mathbf{I} + \frac{\mathbf{H}\mathbf{Q}\mathbf{H}^H}{\sigma^2} \right) \right], \tag{12}$$

where β is the transmitted power and the signal-to-noise ratio (SNR) is $\mathrm{defined}$ as $\frac{\beta}{\sigma^2}$.

A. Capacity achieving input covariance matrix

In the following, we characterize $\mathbf{Q}$ for correlated MIMO channels with CM1, where the columns can have arbitrary covariance matrices. Since CM3 is a special case of CM1, the results hold for CM3 as well.

Theorem 1: For channels following CM1, the optimal $\mathbf{Q}$ which achieves capacity in (12) has an EVD given by $\mathbf{Q}_{opt} = \mathbf{U}_t \mathbf{\Lambda}_{opt} \mathbf{U}_t^H$ where $\mathbf{U}_t$ is the eigen matrix of $\mathbf{R}_t = E(\mathbf{H}^H \mathbf{H})$ and

$$\mathbf{\Lambda}_{opt} = \arg \max_{\mathbf{\Lambda}:\mathrm{tr}(\mathbf{\Lambda}) \le \beta} E_{\mathbf{H}_t} \left[\log\det\left(\mathbf{I} + \frac{\mathbf{H}_t \mathbf{\Lambda} \mathbf{H}_t^H}{\sigma^2} \right) \right]. \tag{13}$$

Proof: See Appendix. ■

For channels following CM3 also $\mathbf{Q}_{opt} = \mathbf{U}_t \mathbf{\Lambda}_{opt} \mathbf{U}_t^H$ where

$$\mathbf{\Lambda}_{opt} = \arg \max_{\mathbf{\Lambda}:\mathrm{tr}(\mathbf{\Lambda}) \leq \beta} E_{\mathbf{H}_c} \left[\log \det \left(\mathbf{I} + \frac{\mathbf{H}_c \mathbf{\Lambda} \mathbf{H}_c^H}{\sigma^2} \right) \right]. \tag{14}$$

and the capacity can be written as

$$C = E_{\mathbf{H}_c} \left[\log \det \left(\mathbf{I} + \frac{\mathbf{H}_c \Lambda_{opt} \mathbf{H}_c^H}{\sigma^2} \right) \right]. \tag{15}$$

The above theorem implies that it is optimal to transmit along the transmit subspaces with powers given by the elements of $\mathbf{\Lambda}_{opt}$. In other words, the transmitted signal should be aligned exactly with the basis vectors of the transmit eigen space and the powers with which the transmit subspaces are excited are governed by the statistical properties of the canonical channel.

B. Optimal power assignments

There is no known closed form solution for the optimization in (13) and (14), however approximate solutions can be obtained for the high and low SNR regimes. For the subsequent discussion, $\mathrm{define}\ \bar{M}_t$ as the number of columns of $\mathbf{P}_s$ which have at least one non-zero element.

1) Optimal powers at high SNR: The optimization in (13) can be written as

$$\mathbf{\Lambda}_{opt} = \arg \max_{\mathbf{\Lambda}:\mathrm{tr}(\mathbf{\Lambda}) \leq \beta} E_{\mathbf{H}_t} \left[\log \det \left(\mathbf{I} + \frac{\mathbf{\Lambda} \mathbf{H}_t^H \mathbf{H}_t}{\sigma^2} \right) \right].$$

It is clear that if the elements of the (say) i-th column of $\mathbf{H}_t$ have zero variance i.e. $E(\underline{\mathbf{h}}_{ti}^H \underline{\mathbf{h}}_{ti}) = \mathbf{0}$, then the power transmitted along the i-th transmit subspace $\mathbf{\Lambda}_{opt}[i,i] = 0$, irrespective of the SNR. Thus, β will be distributed only among those transmit subspaces whose corresponding columns in $\mathbf{H}_t$ have at least one element having non-zero power (i.e. the scalar $E(\underline{\mathbf{h}}_{ti}^H \underline{\mathbf{h}}_{ti})$ is non-zero). Hence, without loss of generality, it is assumed for the derivation below, that all columns of $\mathbf{H}_t$ have at least one element with non-zero power (i.e. $E(\underline{\mathbf{h}}_{ti}^H \underline{\mathbf{h}}_{ti}) \neq 0 \ \ \forall i$). Using (23) and the Taylor series expansion of $\log(1+x)$, it can be shown that for high SNR

$$E_{\mathbf{H}_t} \left[\log \det \left(\mathbf{I} + \frac{\mathbf{\Lambda} \mathbf{H}_t^H \mathbf{H}_t}{\sigma^2} \right) \right] \approx E_{\mathbf{H}_t} \left[\log \det \left(\frac{\mathbf{\Lambda} \mathbf{H}_t^H \mathbf{H}_t}{\sigma^2} \right) \right] = E_{\mathbf{H}_t} \left[\log \det \left(\frac{\mathbf{H}_t^H \mathbf{H}_t}{\sigma^2} \right) \right] + \log \mathbf{\Lambda}.$$

Then the optimal powers are assigned as

$$\mathbf{\Lambda}_{opt} = \arg \max_{\mathbf{\Lambda}:\mathrm{tr}(\mathbf{\Lambda}) \leq \beta} \log(\mathbf{\Lambda}) = \frac{\beta}{\bar{M}_t} \mathbf{I} \tag{16}$$

i.e. all the transmit subspaces get equal power. In general, if $\bar{M}_t$ columns have non-zero powers then the total transmitted power will be distributed equally amongst the corresponding subspaces.

In CM3, the canonical decomposition clearly shows the powers of the degrees of freedom through the subspace connectivity matrix $\mathbf{P}_s$. Similar to the above argument, from (14) it is clear that if the i-th column of $\mathbf{P}_s$ has no non-zero elements, then the i-th subspace will get zero power irrespective of the SNR. Thus, if $\bar{M}_t$ columns of $\mathbf{P}_s$ have at least one non-zero element, then the total power β will be distributed among the corresponding subspaces.

2) Optimal powers at low SNR: From (24), the optimal $\mathbf{\Lambda}_{opt}$ satisfies the condition

$$E_{\mathbf{H}_t}\left[\mathrm{tr}\left(\left(\mathbf{I}+\frac{\mathbf{H_t}\mathbf{\Lambda}_{opt}\mathbf{H}_t^H}{\sigma^2}\right)^{-1}\frac{\mathbf{H_t}(\mathbf{\Lambda}-\mathbf{\Lambda}_{opt})\mathbf{H}_t^H}{\sigma^2}\right)\right]\leq 0,$$

where $\mathbf{\Lambda}, \mathbf{\Lambda}_{opt} \in \mathbf{\Omega}_d$, and $\mathbf{\Omega}_d := \{\mathbf{\Lambda} : \mathbf{\Lambda} \text{ is diagonal}, \ \mathrm{tr}(\mathbf{\Lambda}) = \beta\}$. Using the matrix inversion identity $(\mathbf{I}+\mathbf{C})^{-1} = \mathbf{I}-\mathbf{C}(\mathbf{I}+\mathbf{C})^{-1}$, the LHS of the above inequality can be approximated for low SNR as

$$E_{\mathbf{H}_t}\left[\mathrm{tr}\left(\left(\mathbf{I}+\frac{\mathbf{H_t}\mathbf{\Lambda}_{opt}\mathbf{H}_t^H}{\sigma^2}\right)^{-1}\frac{\mathbf{H_t}(\mathbf{\Lambda}-\mathbf{\Lambda}_{opt})\mathbf{H}_t^H}{\sigma^2}\right)\right]\approx\frac{1}{\sigma^2}E_{\mathbf{H}_t}\left[\mathrm{tr}\left(\mathbf{H_t}(\mathbf{\Lambda}-\mathbf{\Lambda}_{opt})\mathbf{H}_t^H\right)\right].$$

This implies that at low SNR, $\mathbf{\Lambda}_{opt}$ satisfies

$$\mathrm{tr}\left(\mathbf{\Lambda}E_{\mathbf{H}_t}(\mathbf{H}_t^H\mathbf{H_t})\right)\leq\mathrm{tr}\left(\mathbf{\Lambda}_{opt}E_{\mathbf{H}_t}(\mathbf{H}_t^H\mathbf{H_t})\right).$$

Then the optimization problem becomes

$$\mathbf{\Lambda}_{opt}=\arg\max_{\mathbf{\Lambda}}\mathrm{tr}\left(\mathbf{\Lambda}\mathbf{D}_t\right)=\arg\max_{\mathbf{\Lambda}}\sum_{i=1}^{M_t}\lambda[i]\mathbf{D}_t[i,i],$$

where $\mathbf{D}_t = E_{\mathbf{H}_t}(\mathbf{H}_t^H\mathbf{H_t})$ is diagonal and $\lambda[i]$ is the i-th diagonal element of $\mathbf{\Lambda}$. It is straightforward to show that,

$$\lambda_{opt}[i]=1,\quad \lambda_{opt}[j]=0,\ \ \forall\, j\neq i\quad s.t.\quad i=\arg\max_i\mathbf{D}_t[i,i]$$

i.e. all the power is assigned to the transmit subspace whose corresponding $E(\underline{\mathbf{h}}_{ti}^H\underline{\mathbf{h}}_{ti})$ is maximum. In the event of multiple maxima, which happens for instance when the two strongest columns of $\mathbf{H}_t$ have identical statistics, the optimal transmit subspace can be chosen arbitrarily.

For CM3, it can be shown that this optimization becomes

$$\lambda_{opt}[i]=1,\quad \lambda_{opt}[j]=0,\ \ \forall\, j\neq i\quad s.t.\quad i=\arg\max_i\sum_{j=1}^{M_r}\mathbf{P}_s[j,i]$$

i.e. all the power is assigned to the transmit subspace whose sum of coupling powers is maximum.

3) Dominant subspaces and Spatial multiplexing gain: Define the transmit subspaces getting non-zero power as *dominant* and d_s as the number of dominant transmit subspaces. Thus, at high SNR's, $d_s = \bar{M}_t$ and the total power is equally transmitted along all $\bar{M}_t$ transmit subspaces. In general, as SNR is reduced, $d_s \leq \bar{M}_t$, the power assignments become unequal and the weaker transmit subspaces get dropped till finally, at low SNR, $d_s = 1$ and beamforming becomes optimal. The spatial multiplexing gain is defined as the number of dominant parallel channels and is equal to $\min(d_s, M_r) \leq \min(M_t, M_r)$. The above analysis shows that, unlike i.i.d. channels, the spatial multiplexing gain for correlated MIMO channels is a function of SNR and decreases with decreasing SNR [19],[15].

IV. Related results for optimal channel estimation and code design

In [7], [8], the optimal transmit signal is designed which minimizes the minimum mean square estimation error at the receiver for CM3 under the assumptions that the channel is block fading and the channel statistics are known at the transmitter. It is shown that the optimal signal is a $M_t \times N$ block signal of the form $\mathbf{S}_{opt} = \mathbf{U}_t \mathbf{\Lambda}^*$, where $\mathbf{\Lambda}^*$ is a diagonal matrix and N is determined by the SNR and $\mathbf{R}_c$. Thus, similar to capacity optimality, it is optimal to transmit along the transmit subspaces. In particular, the optimal signal *scans* the scattering environment by transmitting sequentially along the transmit subspaces with powers (the elements of $\mathbf{\Lambda}^*$) determined by the SNR and $\mathbf{R}_c$. Similar to the capacity analysis, beamforming is optimal at low SNRs $(N = 1)$ i.e. only the strongest subspace is estimated. As SNR increases the number of subspaces estimated increase $(1 \leq N \leq M_r)$ and the power transmitted along the different subspaces are possibly unequal depending on $\mathbf{R}_c$ and SNR. Finally, at high SNRs all subspaces are estimated and the optimal signal transmits equal power along all the subspaces with non-zero variances (N takes a maximum value of $\bar{M}_t$). Thus the optimal block length increases with increasing SNR. In essence, those canonical channel coefficients whose power is small compared to the background noise are not important from a communication viewpoint and the transmit power is better utilized in estimating channel coefficients that exhibit a higher SNR.

In [16], space-time coding is investigated for correlated channels modeled as CM3. Assuming a block fading channel, the block transmission scheme is given by $\mathbf{X} = \sqrt{\frac{\mathcal{E}_s}{M_r}}\mathbf{HS} + \mathbf{N}$, where $\mathbf{S}$ and $\mathbf{X}$ are the $M_t \times T$ transmitted and $M_r \times T$ received block signals, T is the code length, and $\mathcal{E}_s$ is the transmitted power. The performance of space-time block codes (STBC) is measured

by the pair-wise error probability (PEP), which is the probability of erroneously decoding $\mathbf{S}^m$ when $\mathbf{S}^n$ was transmitted. The PEP can be bounded as

$$P(\mathbf{S}^m \rightarrow \mathbf{S}^n) \leq |\mathbf{I}_{PQ} + \frac{\mathcal{E}_s}{4\sigma^2 M_t}\mathbf{R}_c(\mathbf{R}_e \otimes \mathbf{I}_Q)|^{-1} \tag{17}$$

where $\mathbf{R}_e := (\mathbf{S}^m - \mathbf{S}^n)(\mathbf{S}^m - \mathbf{S}^n)^H$ is the code error covariance matrix. In [16],[20], it is argued that since the PEP depends on the interaction of $\mathbf{R}_e$ and $\mathbf{R}_c$, the performance of STBCs designed for the i.i.d channel degrades over correlated channels [21],[22],[23] because the codes are not *matched* to channel statistics, and can be improved by designing STBCs that match $\mathbf{R}_c$ which effectively excite the different weighted degrees of freedom optimally. One approach to achieve this is to modify existing STBCs, like the orthogonal STBC, using *precoding*. Precoding involves applying a non-unitary linear transformation to the codeword before transmitting it. Define $\mathbf{W}$ as the $M_t \times M_t$ precoding matrix such that $\mathrm{tr}(\mathbf{W}\mathbf{W}^H) = M_t$ so that the total transmitted power per block is preserved. With precoded symbols $\mathbf{WS}$, the PEP expressions can be obtained by substituting $\mathbf{R}_e$ with $\mathbf{W}\mathbf{R}_e\mathbf{W}^H$ in (17). For orthogonal STBCs [24], assuming knowledge of $\mathbf{R}$ at the transmitter, it is shown in [16] that the optimal precoding matrix which minimizes PEP is of the form $\mathbf{W} = \mathbf{U}_t\mathbf{\Lambda}^*$ where $\mathbf{\Lambda}^*$ is a diagonal matrix. Thus, similar to the capacity analysis, it is optimal to transmit the STBC streams along the transmit subspaces. The powers are given by the non-zero values of $\mathbf{\Lambda}^*$ and d_s is the number of dominant transmit subspaces as a function of SNR. Again, beamforming is optimal at low SNR ($d_s = 1$) i.e. only one data stream is transmitted along the strongest transmit subspace. As SNR increases, more streams are transmitted with possibly nonequal powers along the dominant transmit subspaces ($1 \leq d_s \leq \bar{M}_t$) and $\mathrm{finally}$ at high SNR $d_s = \bar{M}_t$ streams are transmitted with equal powers. The result is *reduced* dimensional codes, since $d_s \times T$ blocks are transmitted for a given SNR. Reduced dimensional codes have smaller diversity than OSTBCs without precoding, since only the strongest subset of the transmit subspaces are used and a subset of the available degrees of freedom are excited. However, the coding gain is higher since the total power is transmitted along the strongest subset. This diversity-coding gain tradeoff results in improved performance compared to orthogonal STBCs without precoding (see [16] for details).

The analysis in [7], [8] and [16] is presented for CM3, but can be extended to CM1. To summarize, the optimal transmit signal design in all the three cases of achieving capacity, minimizing MMSE in channel estimation and minimizing PEP in STBCs, requires transmission along the

transmit subspaces. The transmission powers however are dependent on the optimization problem involved and are dependent on the SNR and $\mathbf{R}_c$. It is clear that since i.i.d. channels have equally weighted degrees of freedom, the optimal transmit signal in all these cases is a scalar multiple of the identity matrix. These observations signify the key insights (and differences with i.i.d channels) provided by the canonical statistical models for correlated channels.

V. Covariance structures and simulation results

A. Covariance structures

The spatial covariance matrix of a narrowband frequency non-selective MIMO fading channel depends on many physical parameters which include antenna geometry (arrangement and spacing), angular spreading function and path distribution, the last two of which are dependent on the propagation environment. For a given array geometry, the angular spreads and path distribution determine the distribution of power in $\mathbf{P}_s$, i.e. the number of degrees of freedom d and their powers. Thus, $\mathbf{P}_s$ captures the *richness* of the scattering environment. In general, a larger angular spread implies a larger d while a larger number of paths and path gains imply larger powers of the canonical $\mathrm{coefficients}$. Thus, the idealistic rich scattering model has i.i.d statistics, i.e. $d = M_t M_r$ and all canonical elements have equal powers. On the other hand, clustered environments are correlated with $d \leq M_t M_r$ and unequal powers, for example, a highly correlated channel may have $\mathrm{significant}$ non-zero elements only in one column of $\mathbf{P}_s$.

It is useful to compare the capacities of different scattering environments which depend on the distribution of $\mathbf{P}_s$ as seen in the previous section. The standard normalization used is $\mathrm{tr}(\mathbf{R}) = M_t M_r$ [19],[25],[4],[14], so that the total received power is identical irrespective of the scattering environment. Denote this as normalization-I. Under normalization-I, in a highly correlated channel (with one non-zero column in $\mathbf{P}_s$) where beamforming is optimal, the received power is M_t times higher than that of an i.i.d channel where isotropic transmission is optimal and correlated channels have higher capacity than i.i.d channels below some low SNR value. However, normalization-I compares the idealistic rich scattering environment with smaller fading gains per degree of freedom with a highly clustered environment with relatively larger fading gains per degree of freedom. Clearly, this normalization is applicable in this scenario but in general fails to make a clear comparison between environments where the total received power may not be equal for the same transmitted power. Hence, the observation that correlated channels have higher capacity than i.i.d channels at low SNRs is not a general one and should be made

with care. Channels with different total received power can be compared by keeping the average power per degree of freedom constant, i.e. $\mathrm{tr}(\mathbf{R}_c)/d$. Denote this as normalization-II. This ensures that the average power in the non-zero canonical $\mathrm{coefficients}$ is the same. If $\mathbf{R}_{c1}$ and $\mathbf{R}_{c2}$ are two channels with degrees of freedom d_1 and d_2, then normalization-II implies $\frac{\mathrm{tr}(\mathbf{R}_{c1})}{d_1} = \frac{\mathrm{tr}(\mathbf{R}_{c2})}{d_2}$. Thus, channels with richer scattering having higher d will have a higher total received power than highly correlated channels which have smaller d.

In order to illustrate the effect of the degrees of freedom and available parallel channels, we compare different scattering environments by assuming equal *average* propagation loss per path in all cases. We assume a normalization method based only on the canonical channel $\mathbf{H}_c$ assuming CM3 given by

$$\mathbf{P}_s[i,j] \leq 1, \quad \forall \ i,j \tag{18}$$

which $\mathrm{reflects}$ the powers of the degrees of freedom available in the channel. Denote this as normalization-III. Under normalization-III, for the rich scattering i.i.d channel, the covariance matrix is normalized such that $\mathbf{P}_s[i,j] = 1, \ \forall\, i,j$. As the scattering environment gets sparser, lesser columns of $\mathbf{P}_s$ have non-zero values indicating the smaller degrees of freedom due to sparser scattering. A highly correlated channels may have non-zero powers only in one column of $\mathbf{P}_s$ subject to (18). This is illustrated with the following example which compares four channels with increasingly richer scattering, starting for a highly correlated channel to the i.i.d channel:

$$\mathbf{P}_s^{(1)} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \mathbf{P}_s^{(2)} = \begin{bmatrix} 1 & 0 \\ 1 & 0.5 \end{bmatrix}, \mathbf{P}_s^{(3)} = \begin{bmatrix} 1 & 0.8 \\ 1 & 0.5 \end{bmatrix}, \mathbf{P}_s^{(4)} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \tag{19}$$

Under this normalization, a richer scattering environment projects more power to the receiver than a sparser scattering environment and for a given transmit SNR, the receive SNR (ratio of total received power to noise) is lower for clustered environments. Under this normalization, an i.i.d channel will always have higher capacity than correlated channels irrespective of the SNR, since more disparate and independent paths should lead to higher received power and higher degrees of freedom [15].

B. Simulation results

The analytical results obtained in the previous sections on capacity are $\mathrm{verified}$ with simulation examples, which are illustrated only for CM3 without loss of generality. First, simulations are

presented using normalization-III for channels with scattering subspace matrices

$$\mathbf{P}_s^{(5)} = \begin{bmatrix} 1 & 0.5 & 0.3 & 0.1 \\ 1 & 0.5 & 0.2 & 0.1 \\ 1 & 0.4 & 0.2 & 0.1 \\ 1 & 0.4 & 0.3 & 0.1 \end{bmatrix}, \quad \mathbf{P}_s^{(6)} = \begin{bmatrix} 1 & 0.1 & 0 & 0 \\ 1 & 0 & 0.1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0.1 \end{bmatrix}. \tag{20}$$

Figure 3 shows the optimal power allocations to the transmit subspaces as a function of SNR for a channel with subspace connectivity matrix $\mathbf{P}_s^{(5)}$. At high SNR, $d_s = 4$ and all subspaces get equal power. As SNR decreases, subspace 4 is dropped at 24 dB, then subpace 3 is dropped at 10 dB, until finally below 0 dB beamforming becomes optimal with subspace 1 getting all the power. Figure 4 shows the capacity of this channel and compares it with the mutual information or achievable data rate with equal power allocation which corresponds to the optimal transmission scheme with no covariance knowledge. For the given scattering environment, optimal power allocation achieves a gain of approximately 1 to 0.7 bits/sec/Hz at SNRs 5 to 20 dB over isotropic power allocation. This shows that considerable rate gains can be achieved by efficient use of covariance information at the transmitter. Figure 5 shows the optimal power allocation for a highly correlated channel with subspace connectivity matrix $\mathbf{P}_s^{(6)}$ where the first subspace is much stronger than the others and as a result beamforming is optimal at SNRs even as high as 20 dB. Figure 6 shows that the achievable rate difference between the two transmission schemes is even larger, about 2 bits/sec/Hz at SNRs 5 to 20 dB.

Figure 7 shows the effect of degrees of freedom on channel capacity for a 2×2 antenna system. For the four curves shown, the respective subspace connectivity matrices are given in (19). The dotted curve shows the capacity of a highly correlated channel corresponding to $\mathbf{P}_s^{(1)}$. The curves corresponding to $\mathbf{P}_s^{(2)}$ and $\mathbf{P}_s^{(3)}$ indicate stronger scattering than the channel corresponding to $\mathbf{P}_s^{(1)}$, but weaker than the channel corresponding to $\mathbf{P}_s^{(4)}$. The solid curve is the capacity of a rich scattering channel model or i.i.d channel corresponding to $\mathbf{P}_s^{(4)}$. Note that the slope of the curves corresponding to $\mathbf{P}_s^{(2)}$, $\mathbf{P}_s^{(3)}$ and $\mathbf{P}_s^{(4)}$ have the same slope at high SNR since the spatial multiplexing gain is the same for those channels, $d_s = 2$, while the slope of the dotted curve corresponding to $\mathbf{P}_s^{(1)}$ has smaller slope since $d_s = 1$. The figure shows that, under normalization (18) capacity increases as the scattering environment gets richer for any given SNR.

The next set of figures are based on measurement data obtained for different scattering environments. The subspace scattering matrices of two channels with 8 antennas at the transmitter

and receiver are shown in the contour plots of Figure 8. Both channels measured show clustered scattering. The second channel shows only one strong cluster of scatterers, while the first channel shows richer scattering with two strong clusters and a much wider spread of significant elements. The degrees of freedom, i.e. canonical elements having significant gains, in the channels are $d_1 = 45$ and $d_2 = 19$. Figure 9 shows the capacities of the two channels under normalization-I. As expected, the richer scattering environment has much higher capacity than the highly correlated channel for medium and high SNRs. Figure 10 shows the capacities of the two channels under normalization-II. Again, the richer scattering environment has much higher capacity than the highly correlated channel, but the gap in the capacities is larger since the total power at the receiver is larger for the richer scattering channel than the sparser scattering channel. As discussed in the previous subsection, in both figures it is observed that at low SNRs the sparser scattering channel has higher capacity than the richer scattering channel.

VI. Conclusion

The canonical statistical models for correlated MIMO channels presented in this paper are based on canonical decompositions of the MIMO channel matrix and represent a large family of channel models. Geometrically, these models characterize MIMO channels as interactions among transmit and receive eigen spaces, which provides a structured framework for understanding the complexities of correlated MIMO channels. The analysis points to the fundamental differences of correlated and i.i.d. MIMO channels, that unlike i.i.d channels, correlated channels have unequal degrees of freedom and parallel eigen-channels, and invite optimal design of the transmitted signal. We have shown that the eigenvectors of the coherent capacity achieving optimal input covariance matrix are the same as those of the transmit covariance matrix, and the spatial multiplexing gain decreases with decreasing SNR. Since popular channel models like the Kronecker model and virtual representation models are special cases of our model, our results generalize and extend many previously known results. The geometrical interpretation of the interaction of the transmit and receive eigen spaces can be further used in other transmit signal design problems like minimizing channel estimation error [8], space-time block codes[16] and non-coherent communication.

VII. Acknowledgment

We thank Prof. Ernst Bonek of the Technical University of Vienna for graciously providing the measurement data and also Yan Zhou of the University of Wisconsin-Madison.

APPENDIX

Proof of Theorem 1: The proof proceeds by first 'guessing' the optimal eigenvectors and then it is shown that they are indeed optimal. Using (2), the capacity can be written as

$$C = \max_{\mathbf{Q}:\mathrm{tr}(\mathbf{Q})\leq\beta} E_{\mathbf{H}_t}\left[\log\det\left(\mathbf{I}+\frac{\mathbf{H_t}\mathbf{U}_t^H\mathbf{Q}\mathbf{U}_t\mathbf{H}_t^H}{\sigma^2}\right)\right] = \max_{\hat{\mathbf{Q}}:\mathrm{tr}(\hat{\mathbf{Q}})\leq\beta} E_{\mathbf{H}_t}\left[\log\det\left(\mathbf{I}+\frac{\mathbf{H_t}\hat{\mathbf{Q}}\mathbf{H}_t^H}{\sigma^2}\right)\right] \tag{21}$$

where $\hat{\mathbf{Q}} := \mathbf{U}_t^H\mathbf{Q}\mathbf{U}_t$. Define $I(\hat{\mathbf{Q}}) := E_{\mathbf{H}_t}\left[\log\det\left(\mathbf{I}+\frac{\mathbf{H_t}\hat{\mathbf{Q}}\mathbf{H}_t^H}{\sigma^2}\right)\right]$, then the maximization problem is to maximize $I(\hat{\mathbf{Q}})$ over $\hat{\mathbf{Q}}$ subject to the trace constraint.

The rest of the proof follows [14] and the proof of Theorem 3.1 in [10]. It is first assumed that $\hat{\mathbf{Q}}$ belongs to a set of diagonal matrices and then shown that an optimal $\hat{\mathbf{Q}}$ exists in the set. Then it is proved that this diagonal $\hat{\mathbf{Q}}$ is also optimal over the entire space of $\hat{\mathbf{Q}}$. Define

$$\mathbf{\Omega} := \{\hat{\mathbf{Q}} : \hat{\mathbf{Q}} \text{ is positive semidefinite, } \mathrm{tr}(\hat{\mathbf{Q}})\leq\beta\}, \quad \mathbf{\Omega}_d := \{\mathbf{\Lambda} : \mathbf{\Lambda} \text{ is diagonal, } \mathbf{\Lambda}\in\mathbf{\Omega}\}.$$

Consider the optimization in (21) over $\mathbf{\Omega}_d$. Since $\mathbf{\Omega}_d$ is a convex and compact set and $\hat{\mathbf{Q}}$ is differentiable and strictly concave over $\mathbf{\Omega}_d$, there exists a unique $\mathbf{\Lambda}_{opt}$ that maximizes $I(\hat{\mathbf{Q}})$ over $\mathbf{\Omega}_d$, and $\mathbf{\Lambda}_{opt}$ satisfies the necessary condition

$$\delta I(\mathbf{\Lambda}_{opt};\mathbf{\Lambda}-\mathbf{\Lambda}_{opt}) := \lim_{\alpha\to 0}\frac{1}{\alpha}\left[I(\mathbf{\Lambda}_{opt}+\alpha(\mathbf{\Lambda}-\mathbf{\Lambda}_{opt}))-I(\mathbf{\Lambda}_{opt})\right]\leq 0, \quad \forall\mathbf{\Lambda}\in\mathbf{\Omega}_d. \tag{22}$$

Simple algebraic manipulations show that,

$$\delta I(\mathbf{\Lambda}_{opt};\mathbf{\Lambda}-\mathbf{\Lambda}_{opt}) = \lim_{\alpha\to 0} E_{\mathbf{H}_t}\left[\log\det\left(\mathbf{I}+\alpha\left(\mathbf{I}+\frac{\mathbf{H_t}\mathbf{\Lambda}_{opt}\mathbf{H}_t^H}{\sigma^2}\right)^{-1}\frac{\mathbf{H_t}(\mathbf{\Lambda}-\mathbf{\Lambda}_{opt})\mathbf{H}_t^H}{\sigma^2}\right)\right].$$

Using the expansion

$$\det(\mathbf{I}+\alpha\mathbf{A}) = \prod_i(1+\alpha\lambda_i(\mathbf{A})) = 1+\alpha\mathrm{tr}(\mathbf{A})+\ldots+\alpha^M\det(\mathbf{A}) \tag{23}$$

where $\lambda_i(\mathbf{A})$ is the i-th eigenvalue of $M\times M$ matrix $\mathbf{A}$, and the Taylor series expansion of $\log(1+x)$, some algebraic manipulations give

$$\delta I(\mathbf{\Lambda}_{opt};\mathbf{\Lambda}-\mathbf{\Lambda}_{opt}) = E_{\mathbf{H}_t}\left[\mathrm{tr}\left(\left(\mathbf{I}+\frac{\mathbf{H_t}\mathbf{\Lambda}_{opt}\mathbf{H}_t^H}{\sigma^2}\right)^{-1}\frac{\mathbf{H_t}(\mathbf{\Lambda}-\mathbf{\Lambda}_{opt})\mathbf{H}_t^H}{\sigma^2}\right)\right]\leq 0. \tag{24}$$

The rest of the proof shows that $\mathbf{\Lambda}_{opt}$ is also optimal over $\mathbf{\Omega}$. Since $I(\hat{\mathbf{Q}})$ is strictly concave over the convex set $\mathbf{\Omega}$, it is sufficient to show that

$$\delta I(\mathbf{\Lambda}_{opt};\hat{\mathbf{Q}}-\mathbf{\Lambda}_{opt})\leq 0, \quad \forall\hat{\mathbf{Q}}\in\mathbf{\Omega}. \tag{25}$$

Split $\hat{\mathbf{Q}}$ into $\hat{\mathbf{Q}} = \mathbf{\Lambda} + \mathbf{A}$ where $\mathbf{\Lambda} \in \mathbf{\Omega}$ with diagonal entries equal to the diagonal entries of $\hat{\mathbf{Q}}$ and $\mathbf{A}$ is the matrix of off-diagonal elements of $\hat{\mathbf{Q}}$. Then,

$$\begin{aligned}\delta I(\mathbf{\Lambda}_{opt}; \hat{\mathbf{Q}} - \mathbf{\Lambda}_{opt}) &= E_{\mathbf{H}_t}\left[\mathrm{tr}\left(\left(\mathbf{I} + \frac{\mathbf{H_t}\mathbf{\Lambda}_{opt}\mathbf{H}_t^H}{\sigma^2}\right)^{-1}\frac{\mathbf{H_t}(\hat{\mathbf{Q}} - \mathbf{\Lambda}_{opt})\mathbf{H}_t^H}{\sigma^2}\right)\right] \\ &= E_{\mathbf{H}_t}\left[\mathrm{tr}\left(\left(\mathbf{I} + \frac{\mathbf{H_t}\mathbf{\Lambda}_{opt}\mathbf{H}_t^H}{\sigma^2}\right)^{-1}\frac{\mathbf{H_t}(\mathbf{\Lambda} - \mathbf{\Lambda}_{opt})\mathbf{H}_t^H}{\sigma^2}\right)\right] \\ &\quad + E_{\mathbf{H}_t}\left[\mathrm{tr}\left(\left(\mathbf{I} + \frac{\mathbf{H_t}\mathbf{\Lambda}_{opt}\mathbf{H}_t^H}{\sigma^2}\right)^{-1}\frac{\mathbf{H_t}\mathbf{A}\mathbf{H}_t^H}{\sigma^2}\right)\right]. \end{aligned} \tag{26}$$

The first term on the RHS is ≤ 0 from (24). The second term can be written as

$$\begin{aligned}&E_{\mathbf{H}_t}\left[\mathrm{tr}\left(\left(\mathbf{I} + \frac{\mathbf{H_t}\mathbf{\Lambda}_{opt}\mathbf{H}_t^H}{\sigma^2}\right)^{-1}\frac{\mathbf{H_t}\mathbf{A}\mathbf{H}_t^H}{\sigma^2}\right)\right] \\ &\quad = \sum_{k,l=1;k\neq l}^{M_t} E_{\mathbf{H}_t}\left[\mathrm{tr}\left(\left(\mathbf{I} + \frac{\mathbf{H_t}\mathbf{\Lambda}_{opt}\mathbf{H}_t^H}{\sigma^2}\right)^{-1}\frac{\mathbf{A}[k,l]\underline{\mathbf{h}}_{tk}\underline{\mathbf{h}}_{tl}^H}{\sigma^2}\right)\right]. \end{aligned} \tag{27}$$

Consider the first term of the summation on the RHS, which can be written as

$$\begin{aligned}&E_{\mathbf{H}_t}\left[\mathrm{tr}\left(\left(\mathbf{I} + \frac{\mathbf{H_t}\mathbf{\Lambda}_{opt}\mathbf{H}_t^H}{\sigma^2}\right)^{-1}\frac{\mathbf{A}[1,2]\underline{\mathbf{h}}_{t1}\underline{\mathbf{h}}_{t2}^H}{\sigma^2}\right)\right] \\ &= \mathrm{tr}\left(E\left[E\left[\left(\mathbf{I} + \frac{\mathbf{H_t}\mathbf{\Lambda}_{opt}\mathbf{H}_t^H}{\sigma^2}\right)^{-1}\frac{\mathbf{A}[1,2]\underline{\mathbf{h}}_{t1}\underline{\mathbf{h}}_{t2}^H}{\sigma^2}|\underline{\mathbf{h}}_{t2}, \underline{\mathbf{h}}_{t3}, \ldots, \underline{\mathbf{h}}_{t,M_t}\right]\right]\right). \end{aligned} \tag{28}$$

Since the columns of $\mathbf{H}_t$ are independent and zero mean complex Gaussian, the conditional distribution of $\mathbf{h}_{t1}$ given $\mathbf{h}_{t2}, \mathbf{h}_{t3}, \ldots, \mathbf{h}_{t3}$ is also zero mean complex Gaussian. Each element of the matrix $\left(\mathbf{I} + \frac{\mathbf{H_t}\mathbf{\Lambda}_{opt}\mathbf{H}_t^H}{\sigma^2}\right)^{-1}\frac{\mathbf{A}[1,2]\underline{\mathbf{h}}_{t1}\underline{\mathbf{h}}_{t2}^H}{\sigma^2}$ is an odd function of $\mathbf{h}_{t1}$ which implies that

$$E\left[\left(\mathbf{I} + \frac{\mathbf{H_t}\mathbf{\Lambda}_{opt}\mathbf{H}_t^H}{\sigma^2}\right)^{-1}\frac{\mathbf{A}[1,2]\underline{\mathbf{h}}_{t1}\underline{\mathbf{h}}_{t2}^H}{\sigma^2}|\underline{\mathbf{h}}_{t2}, \underline{\mathbf{h}}_{t3}, \ldots, \underline{\mathbf{h}}_{t,M_t}\right] = 0 \tag{29}$$

Hence the first term shown in (28) is zero. Following the same reasoning, all the terms on the RHS of (27) are zero. This shows that (25) is $\mathrm{satisfied}$ and $\mathbf{\Lambda}_{opt}$ is optimal over $\mathbf{\Omega}$.

REFERENCES

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *AT & T Bell Labs Internal Technical Memo*, June 1995.

[2] G. J. Foschini, "Layered space-time architechture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[3] C. N. Chua, J. M. Kahn, and D. Tse, "Capacity of multi-antenna array systems in indoor wireless environment," *Proceeding of the IEEE Global Telecommunications Conference*, pp. 1894–1899, 1998.

[4] A. M. Sayeed, "Deconstructing multi-antenna fading channels," *IEEE Transactions of Signal Processing*, Oct. 2002.

[5] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 502–512, Mar. 2000.

[6] J. R. Fonollosa, R. Gaspa, X. Mestre, A. Pages, M. Heikkila, J. P. Kermoal, L. Schumacher, A. Pollard, and J. Ylitalo, "The IST METRA project," *IEEE Communications Magazine*, vol. 40, no. 7, pp. 78–86, July 2002.

[7] J. H. Kotecha and A. M. Sayeed, "Optimal signal design for estimation of correlated MIMO channels," *International Conference on Communications*, May 2003.

[8] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Transactions on Signal Processing*, Feb 2004.

[9] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "A stochastic mimo channel model with joint correlation of both link ends," *preprint*, 2003.

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

[11] S. A. Jafar, S. Vishwanath, and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," *International Conference on Communications*, pp. 2266–2270, 2001.

[12] J. H. Kotecha and A. M. Sayeed, " On the Capacity of Correlated MIMO channels with Covariance Feedback," *International Symposium on Information Theory*, 2003.

[13] H. Boche and E. Jorswieck, "Optimal transmission with imperfect channel state information at the transmit antenna array," *Wireless Personal Communications*, July 2003.

[14] V. Veeravalli, Y. Liang, and A. Sayeed, "Correlated MIMO Rayleigh fading channels: capacity, optimal signaling and asymptotics," *preprint*, 2003.

[15] A. L. Moustakas, H. U. Baranger, L. Balents, A. M. Sengupta, and S. H. Simon, "Communication through a diffusive medium: Coherence and capacity," *Science*, vol. 287, pp. 287–290, 2000.

[16] J. H. Kotecha, Z. Hong, and A. M. Sayeed, "Coding and Diversity Gain Tradeoff in Space-time Codes for Correlated MIMO Channels," *Globecomm 2003*.

[17] W. C. Jakes, *Microwave Mobile Communication*, IEEE Press, New York, 1993.

[18] Kai Yu, M. Bengtsson, B. Ottersten, Peter Karlsson, and Mark Beach, "Second order statistics of NLOS indoor MIMO channels based on 5.2 Ghz measurements," *Proceedings of the IEEE Global Telecommunications Conference*, 2001.

[19] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity Limits of MIMO Channels ," *IEEE Journal on Selected Areas in Communications*, June 2003.

[20] Z. Hong, K. Liu, R. Heath Jr, and A. M. Sayeed, "Spatial multiplexing in correlated fading via the virtual channel representation," *Journal on Selected Areas in Communications*, April 2003.

[21] M. O. Damen, A. Abdi, and M. Kaveh, "On the effect of fading on several space-time coding and detection schemes," *IEEE Vehicular Technology Conference, Oct. 2001*, vol. 1, pp. 13–16.

[22] M. Uysal and C. N. Georghiades, "Effect of spatial fading correlation on performance of space-time codes," *IEEE Electronics Letters*, vol. 37, no. 3, pp. 181–183, Feb 2001.

[23] H. El Gamal, "On the robustness of space-time coding," *IEEE Transactions on Signal Processing*, Oct. 2002.

[24] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, pp. 1456 –1467, Jul. 1999.

[25] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO Capacity Through Correlated Channels in the Presence of Correlated Interferers and Noise: A (not so) Large N Analysis," *IEEE Transactions on Information Theory*, Oct 2003.

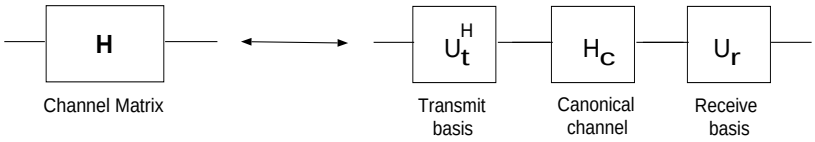

Fig. 1. Canonical channel modeling based on the above decomposition - depending on the assumptions three models arise where either the columns of $\mathbf{H}_c$ are uncorrelated or the rows of $\mathbf{H}_c$ are uncorrelated or all the elements of $\mathbf{H}_c$ are uncorrelated. $\mathbf{H}_c$ behaves as a coupling matrix between $\mathbf{U}_t$ and $\mathbf{U}_r$.

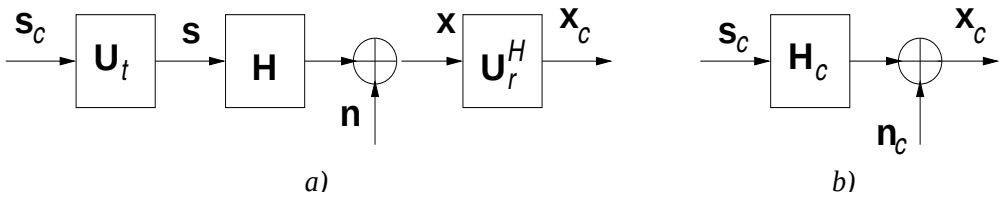

Fig. 2. Signaling and reception in the eigen domain. $\mathbf{U}_t$ and $\mathbf{U}_r$ are pre and post-processing matrices and expose the degrees of freedom of the correlated channel.

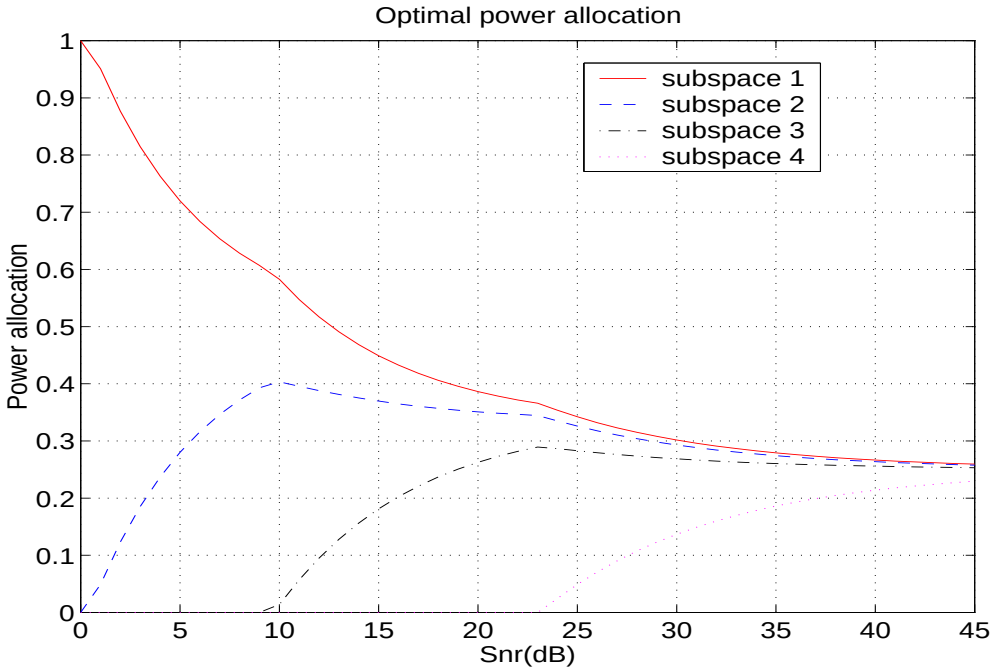

Fig. 3. Optimal power distribution for a 4×4 antenna system with $\mathbf{P}_s^{(5)}$ given in (20).

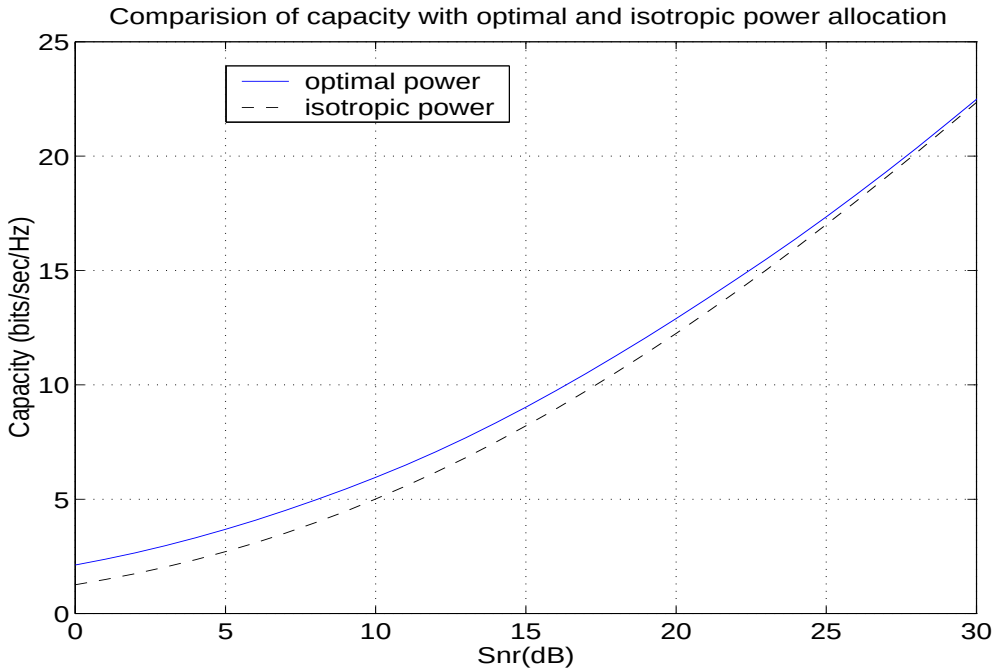

Fig. 4. Capacity of a 4×4 antenna system with $\mathbf{P}_s$ given in (20).

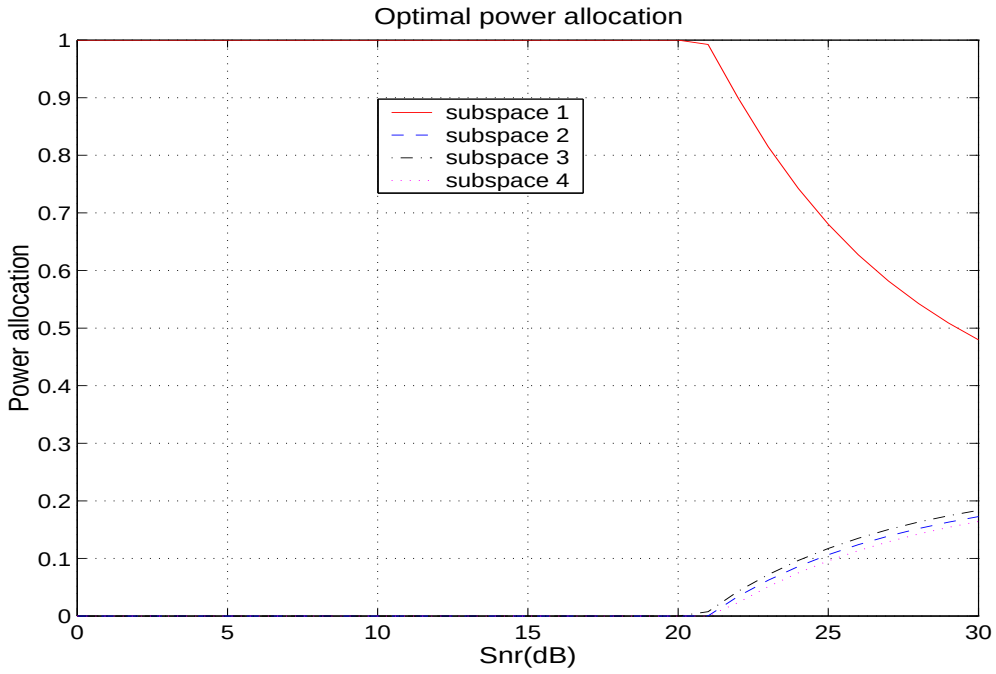

Fig. 5. Optimal power distribution for a 4×4 antenna system with $\mathbf{P}_s^{(6)}$ given in (20).

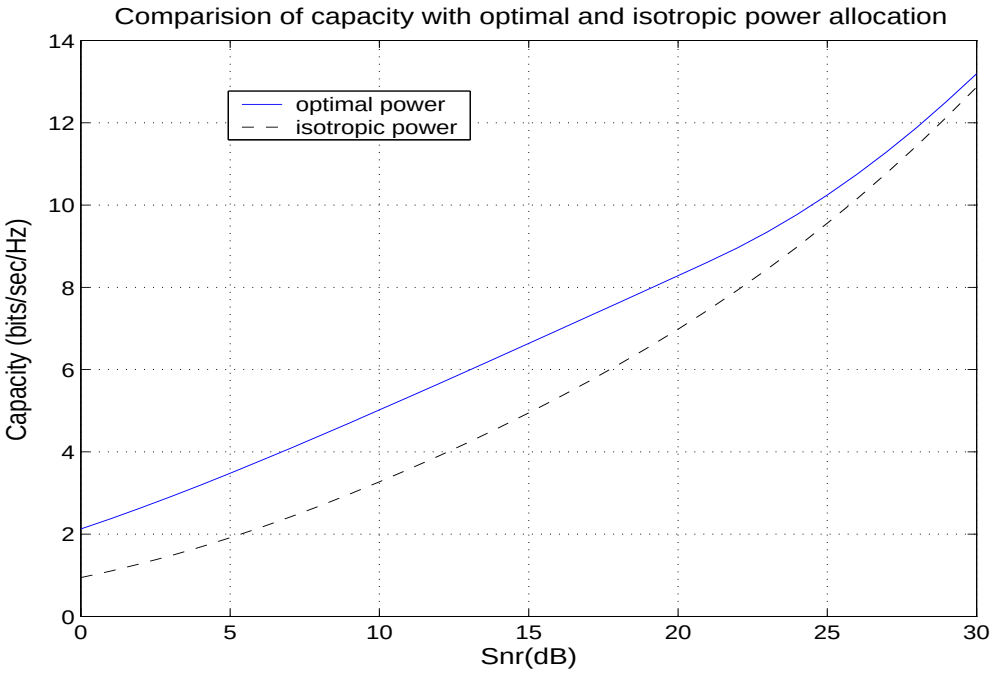

Fig. 6. Capacity of a highly correlated 4×4 antenna system with $\mathbf{P}_s$ given in (20).

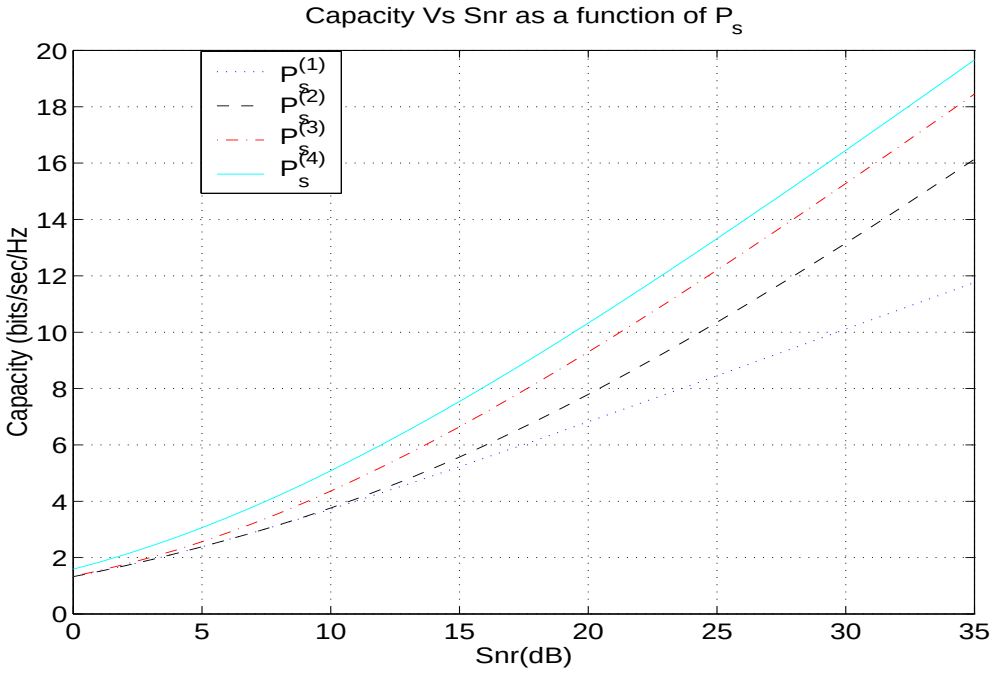

Fig. 7. Effect of scattering environment on the capacity of a 2×2 antenna system with $\mathbf{P}_s^{(i)}$ given in (19).

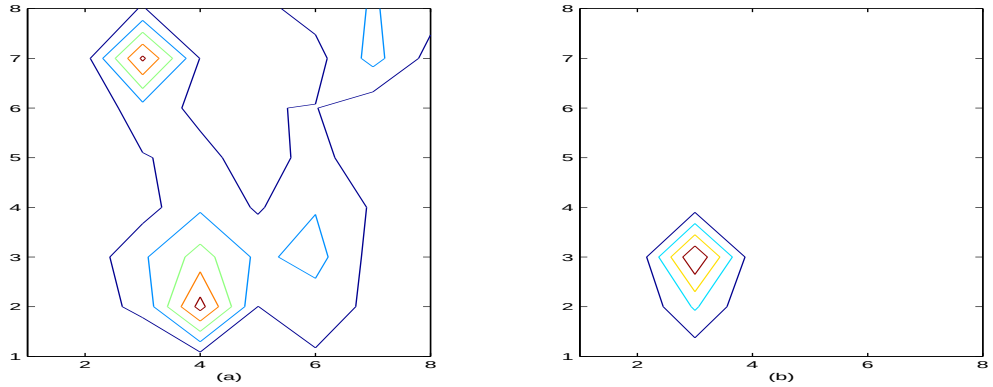

Fig. 8. Contour plots of the subspace scattering matrices of measured channels. (a) richer scattering with two scattering clusters - less correlated. (b) highly correlated channel with one strong cluster.

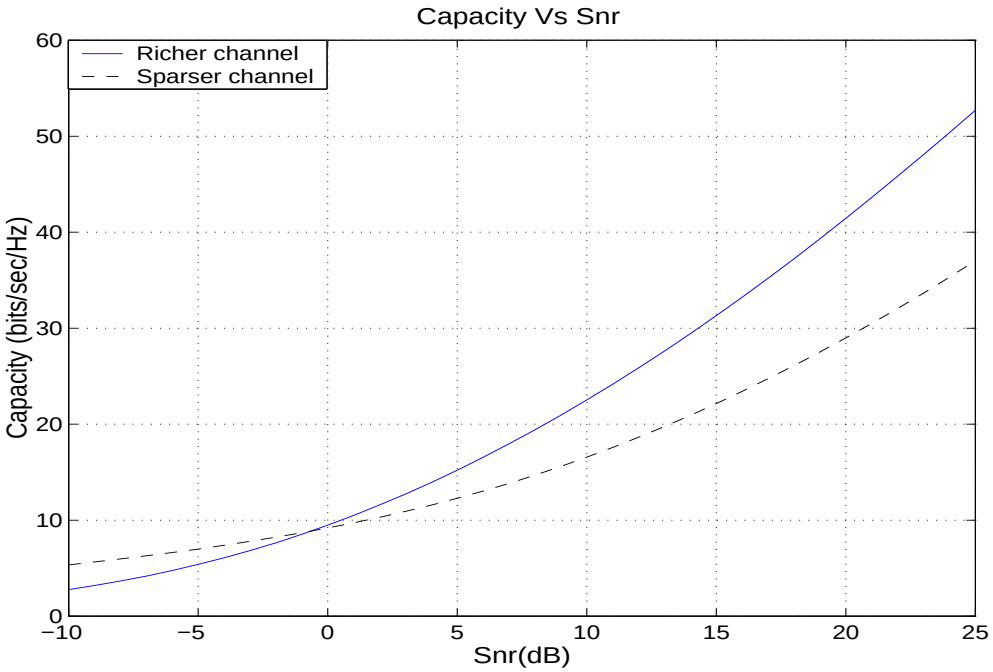

Fig. 9. Capacity comparison of scattering environments under normalization-I. Richer scattering shows higher capacity at medium to high SNRs than sparser scattering, reverse is true at low SNRs.

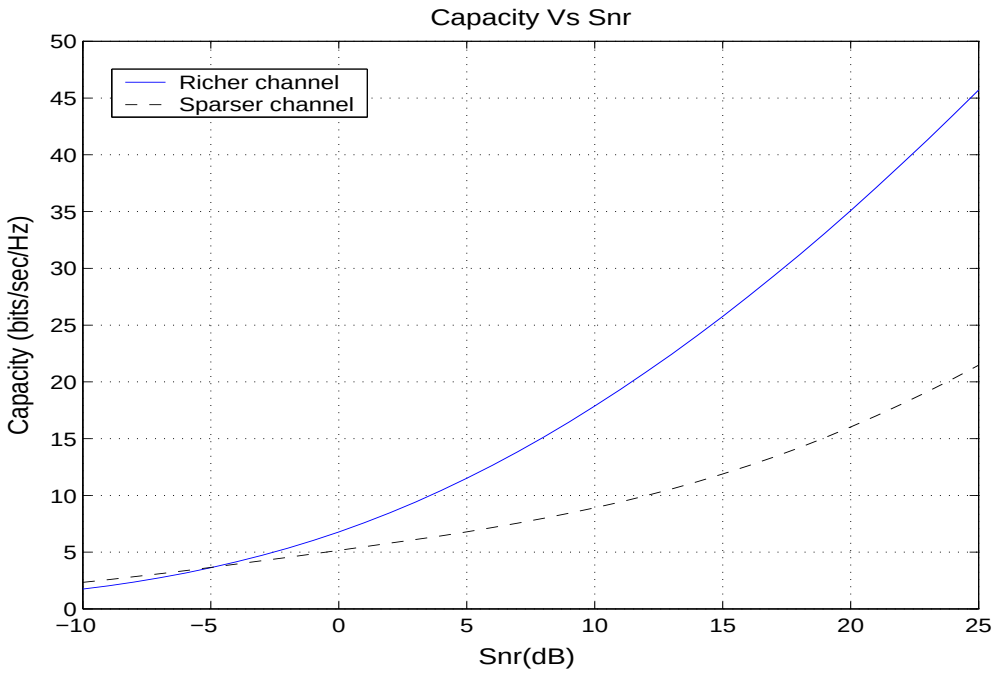

Fig. 10. Capacity comparison of scattering environments under normalization-II. Richer scattering shows higher capacity at medium to high SNRs than sparser scattering, gap in capacities is higher than under normalization-I.