

Correlated MIMO Rayleigh Fading Channels: Capacity, Optimal Signaling, and Asymptotics¹

Venugopal V. Veeravalli, Yingbin Liang² and Akbar Sayeed³

Abstract

The capacity of the MIMO channel is investigated under the assumption that the elements of the channel matrix are zero mean proper complex Gaussian random variables with a general correlation structure. 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. It is shown that in the virtual domain, the capacity achieving input vector consists of independent zero-mean proper complex entries, whose variances can be computed numerically. 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 are also derived. Finally, the capacity is investigated in the asymptotic regime where the number of receive and transmit antennas go to infinity, with their ratio 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.

1 Introduction

Multi-input multi-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

¹V.V. Veeravalli and Y. Liang were supported by NSF under CCR-9980616, through a subcontract with Cornell university, and by the NSF CAREER/PECASE award CCR-0049089. A. Sayeed was supported by NSF under CAREER grant CCR-9875805 and ITR grant CCR-0113385, and by the ONR under the YIP grant N00014-01-1-0825

²V. 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; e-mail: vvv@uiuc.edu, yliang1@uiuc.edu

³A. Sayeed is with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706; e-mail: akbar@engr.wisc.edu

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 maximizing) input is an i.i.d. zero-mean 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 [3, 4, 6, 7, 8, 10] have investigated the capacity and corresponding optimal input distributions for correlated 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 the corresponding transmit correlation and receive correlation. 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.

In this paper, we adopt a realistic model for MIMO channels with uniform linear arrays (ULAs), introduced in [11], where the elements of the channel matrix have a general correlation structure that encompasses the product form case. The channel matrix is related to the underlying scattering environment through a virtual channel matrix, 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 variance 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 correlation structure of the original channel is captured succinctly in the variances of the virtual channel coefficients. This obviously facilitates estimation of the correlation since all that is needed are estimates of the variances of the virtual coefficients. These estimates may be obtained by transmitting pilot signals sequentially beamformed to different virtual transmit angles and computing the average power along the different virtual receive angles [9]. Thus our assumption that the channel statistics are available at the transmitter is 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. If only one of these variances is non-zero, then the optimal transmit strategy is to beamform along the corresponding virtual angle. In the low SNR regime, we show that beamforming to the virtual transmit angle with the largest effective channel gain at the receiver is asymptotically optimal. Furthermore, for arbitrary SNR levels, we provide simple necessary and sufficient conditions for the optimality of beamforming in terms of the variances of the virtual channel coefficients.

We then move on to study the capacity in the asymptotic regime where the number of receive and transmit antennas go to infinity, with their ratio 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. 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 product-form correlation model discussed above was given [4]. Also, in recent work [13], the asymptotic capacity of wideband correlated channels was investigated using the virtual representation for the special case of D -diagonal channels, where the virtual channel coefficients are equal along D leading diagonals and zero elsewhere. While the D -diagonal channel model does not belong to the class of product-form correlation models, it is clearly a special case of the more general scenario where the virtual channel coefficients have arbitrary variances.

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-dimension 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 there is strong correlation among the elements.

1.1 Notation and organization

We use the following notation. For deterministic objects, we use uppercase letters for matrices, lowercase letters for scalars, and underlined lowercase letters for vectors. The only exception is the symbol C which is a scalar that is used to denote capacity. Random objects are identified by corresponding boldfaced 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 \mathbf{x} to denote a random scalar. To indicate the entries of matrices, we use subscripts. For example, the symbol $\mathbf{H}_{k,\ell}$ denotes the component at the k -th row and ℓ -th column of the random matrix \mathbf{H} .

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

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 2, we introduce the virtual representation of the channel. In Section 3, we characterize the optimal input distribution in the virtual domain, and discuss techniques for computing the resulting capacity. In Section 4, we exploit the virtual representation further in analyzing the asymptotic capacity of the channel for large numbers of transmit and receive antennas. In Section 5, we provide a set of numerical results that illustrate the theoretical results of the previous sections. Finally, in Section 6, we give some concluding remarks.

2 Channel Model and Virtual Representation

We consider the now standard channel model for MIMO systems 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 $\underline{\mathbf{x}}$ is the n_t -dim transmit vector, \mathbf{H} is the $n_r \times n_t$ channel matrix, and $\underline{\mathbf{w}}$ is complex additive white Gaussian noise. We normalize the noise so that $\underline{\mathbf{w}} \sim \mathcal{CN}(\mathbf{0}, I)$, and assume an average input power constraint of $\mathbb{E}[\underline{\mathbf{x}}^\dagger \underline{\mathbf{x}}] \leq n_t$. It is natural to assume the entries of the channel matrix \mathbf{H} are identically distributed since each pair of transmit and receive antennas sees the same scattering environment. If we further assume a purely diffuse rich scattering environment with no dominant (specular) paths, then the entries are well modeled as zero mean proper complex Gaussian random variables. Normalizing the channel gains to have unit variance, we have that the entries of \mathbf{H} are $\mathcal{CN}(0, 1)$ random variables, but not necessarily independent. With all the above normalization, Γ represents the effective signal-to-noise ratio (SNR) at each receive antenna. We also make the reasonable assumption that the channel changes in a stationary ergodic manner from symbol to symbol.

In general, the channel matrix \mathbf{H} will consist of correlated entries due to the physical characteristics of the scattering environment. We will exploit the recently introduced *virtual representation* of MIMO channels that provides a simple and insightful characterization of the physical scattering environment [11] for uniform linear arrays ULA's of antennas at the transmitter and receiver. To this end, let d_t and d_r denote the antenna spacing at the transmitter and receiver, respectively. The channel can be written in terms of the gains of all the paths joining the transmitter and

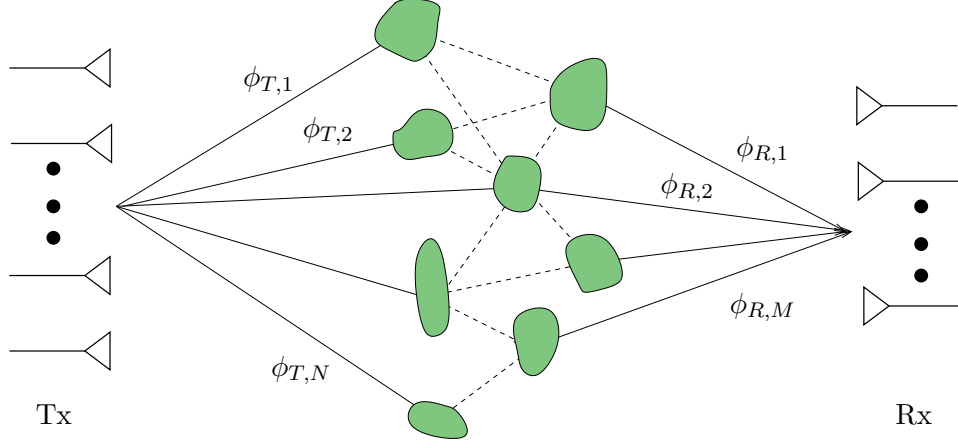


Figure 1: Physical model for MIMO channel

receiver via the following equation (also see Fig 1):

$$\mathbf{H} = \sqrt{n_t, n_r} \int_0^1 \int_0^1 \boldsymbol{\alpha}(\theta_r, \theta_t) \underline{\mathbf{a}}_r(\theta_r) \underline{\mathbf{a}}_t^\dagger(\theta_t) d\theta_r d\theta_t \quad (2)$$

where $\underline{\mathbf{a}}_t(\theta_t)$ and $\underline{\mathbf{a}}_r(\theta_r)$ are normalized array steering and response vectors, which are given by

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

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

The variable θ is related to the physical propagation angle ϕ as

$$\theta = \frac{d(\sin(\phi) + 1)}{\lambda_c} = \alpha(\sin(\phi) + 1)$$

with λ_c being the wavelength of propagation, and $\alpha = d/\lambda_c$. Note that θ can be thought of as a *scaled* angle in that the array and steering response vectors are unaffected if we replace θ by $[\theta]_{\text{mod}[0,1]}$. For simplicity, we assume critical ($\lambda_c/2$) antenna spacing⁴. In this case, there is a one-to-one mapping between the physical angle⁵ $\phi \in [-\pi/2, \pi/2]$ and $\theta \in [0, 1]$, and we refer to θ as a *virtual angle*.

Referring back to (2), the vector $\underline{\mathbf{a}}_r(\theta_r)$ represents the signal response at the receiver array due to a point source in the virtual direction θ_r . Similarly, $\underline{\mathbf{a}}_t(\theta_t)$ represents the array weights needed to transmit a beam focussed in the virtual direction θ_t . The function $\boldsymbol{\alpha}(\theta_r, \theta_t)$ represents

⁴The analysis of this paper can be generalized to arbitrary antenna spacing using the more general models described in [11]

⁵The 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 $[-\pi/2, \pi/2]$.

the random *angular spreading function*, i.e., the random complex channel gain density at virtual transmit angle θ_t and receive angle θ_r .

The model of (2) implicitly assumes an infinite spatial resolution by antenna arrays. The channel matrix \mathbf{H} is written as a double integral assuming that a continuum of paths joining the transmitter and receiver can be resolved. However, a ULA with n elements with $\lambda/2$ spacing is limited to a spatial resolution (in the θ variable) of $1/n$. Hence the double integral of (2) can be replaced by a double summation which can be written compactly as:

$$\mathbf{H} = A_r \tilde{\mathbf{H}} A_t^\dagger \quad (3)$$

where the matrices A_r ($n_r \times n_r$) and A_t ($n_t \times n_t$) are given by

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

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

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 $\theta_{t,\ell}$.

We refer to $\tilde{\mathbf{H}}$ as a *virtual representation* 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. Figure 2 shows the virtual representation of a physical MIMO channel. As

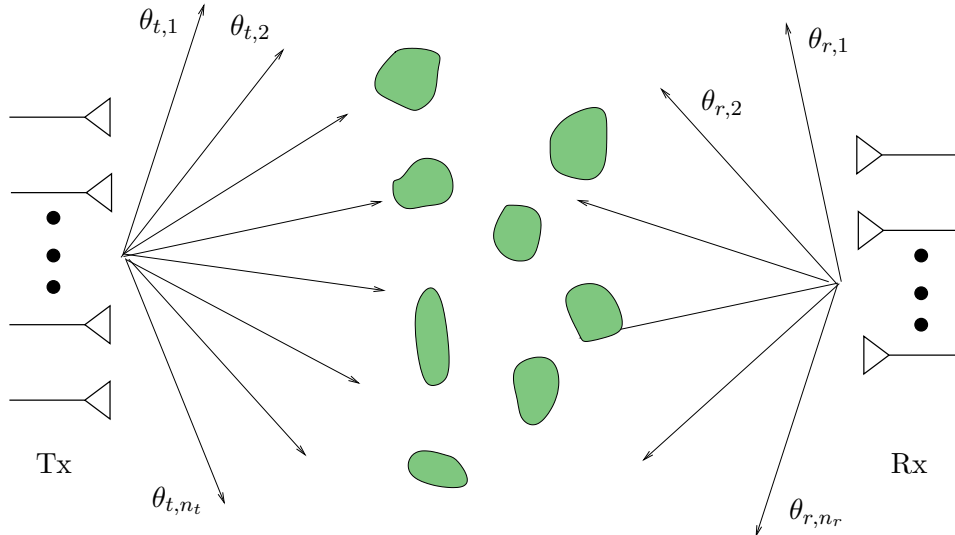


Figure 2: Virtual representation of MIMO Channel

noted in [11], the matrices A_r and A_t are *unitary* discrete Fourier transform matrices. Thus, the

virtual representation can be considered to be a two-dimensional spatial Fourier representation of the channel. The coefficients of the virtual channel matrix are related to the spreading function $\boldsymbol{\alpha}$ as

$$\tilde{\mathbf{H}}_{k,\ell} = \hat{\boldsymbol{\alpha}}(\theta_{r,k}, \theta_{t,\ell}), \quad k = 1, \dots, n_r, \quad \ell = 1, \dots, n_t \quad (4)$$

where

$$\hat{\boldsymbol{\alpha}}(\theta_r, \theta_t) = \int_0^1 \int_0^1 \boldsymbol{\alpha}(\theta'_r, \theta'_t) f_{n_r}(\theta'_r - \theta_r) f_{n_t}^*(\theta'_t - \theta_t) d\theta'_r d\theta'_t$$

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, (4) states that the virtual channel coefficients are samples (at the virtual angles) of a smoothed version of the angular spreading function. Furthermore, the smoothing kernel gets narrower⁶ 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. Furthermore, as we briefly discuss next, the non-vanishing virtual coefficients are approximately uncorrelated.

We now make the reasonable assumption that the physical scattering is uncorrelated across paths. This imposes the following second-order statistical structure on $\boldsymbol{\alpha}(\theta_r, \theta_t)$:

$$\mathbb{E}[\boldsymbol{\alpha}(\theta_r, \theta_t) \boldsymbol{\alpha}^*(\theta'_r, \theta'_t)] = \Psi(\theta_r, \theta_t) \delta(\theta_r - \theta'_r) \delta(\theta_t - \theta'_t) \quad (5)$$

where $\delta(\cdot)$ denotes the Dirac delta function and $\Psi(\theta_r, \theta_t) \geq 0$ is called the *angular scattering function* that reflects the channel power density for various transmit-receive angle pairs. As we mentioned earlier, we assume a purely diffuse scattering environment (without specular paths) for which $\Psi(\theta_r, \theta_t)$ is well modeled as a *bounded* piece-wise continuous function with

$$\int_0^1 \int_0^1 \Psi(\theta_r, \theta_t) d\theta_r d\theta_t = 1.$$

Using the uncorrelated scattering assumption, it can be shown that [11]

$$\mathbb{E}[\tilde{\mathbf{H}}_{k,\ell} \tilde{\mathbf{H}}_{k',\ell'}^*] \approx V_{k,\ell} \delta_{k-k'} \delta_{\ell-\ell'} \quad (6)$$

where δ_n denotes the Kronecker delta function and V is a $n_r \times n_t$ matrix that contains the variances of the components of the virtual channel matrix $\tilde{\mathbf{H}}$.

$$V_{k,\ell} = \mathbb{E}[|\tilde{\mathbf{H}}_{k,\ell}|^2] = \int_0^1 \int_0^1 \Psi(\theta_r, \theta_t) |f_{n_r}(\theta_r - \theta_{r,k})|^2 |f_{n_t}(\theta_t - \theta_{t,\ell})|^2 d\theta_r d\theta_t \quad (7)$$

where the approximation in (6) gets better with increasing n_r and n_t . The above equations state that the virtual channel coefficients are approximately uncorrelated, and for sufficiently large n_r and n_t we have

$$V_{k,\ell} \approx \Psi(\theta_{r,k}, \theta_{t,\ell}) \quad (8)$$

⁶The null-to-null mainlobe width of $f_{n_r}(\theta)$ is $2/n_r$.

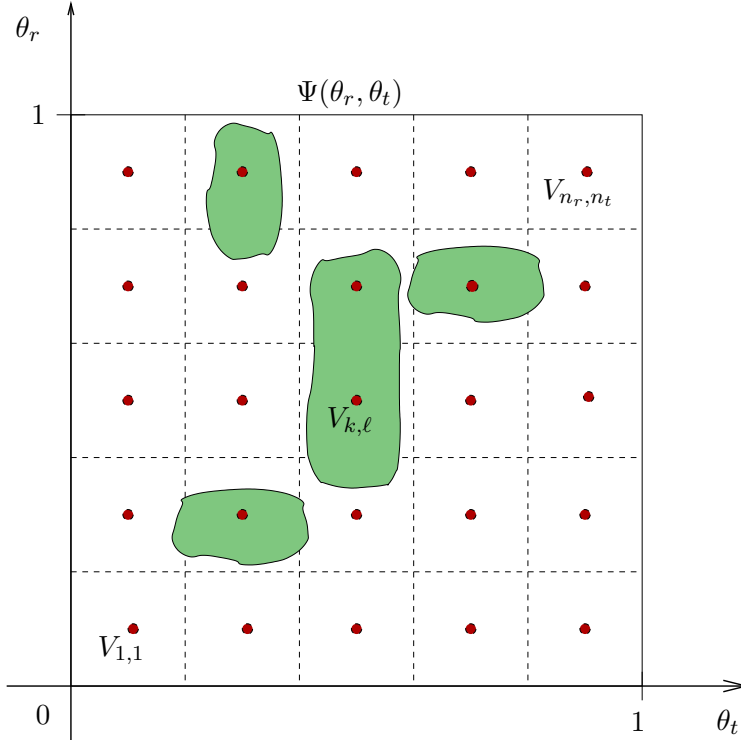


Figure 3: Example scattering function and corresponding variance matrix. The shaded regions in $\Psi(\theta_r, \theta_t)$ correspond to scattering clusters in the physical environment.

where the approximation improves with increasing n_t and n_r . Figure 3 illustrates the relationship between $\Psi(\theta_r, \theta_t)$ and $V_{k,\ell}$.

Remark 1. *As argued in [11], under uncorrelated physical scattering (5), the elements of \mathbf{H} are samples of a 2-D stationary (homogeneous) Gaussian random field and the virtual channel coefficients are samples of a smoothed version of the corresponding spectral representation and are hence approximately uncorrelated. In this interpretation, $\Psi(\theta_r, \theta_t)$ is the 2-D power spectral density associated with the random field and the angular spreads can be construed as bandwidths (support of $\Psi(\theta_r, \theta_t)$) associated with the random field.*

The above discussion is summarized in the following assumption that will be used in the remainder of this paper.

Assumption 1. *The entries of $\tilde{\mathbf{H}}$ are independent zero mean proper complex Gaussian random variables with variances given by the variance matrix V of (7).*

The statistical model for \mathbf{H} that results from this assumption is clearly only an approximation to the real channel, particularly when the number of antennas is small. Nevertheless, this approximation can be expected to be considerably more accurate than the special case of the i.i.d. model

in most realistic scenarios[11]. 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 entries. On the other hand if V has roughly uniform entries, then \mathbf{H} will have roughly i.i.d. entries.

Remark 2. We note that the product form correlation structure used in the analysis of [4, 6, 7, 8, 10] 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 $\Psi(\theta_r, \theta_t)$.

We now present some basic properties of the variance matrix V . First, since the original channel matrix \mathbf{H} is normalized to have unit variance entries, it follows from (3) that the variance matrix V satisfies

$$\sum_{k,\ell} V_{k,\ell} = n_r n_t. \quad (9)$$

Also, based on (7), 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

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

for $k = 1, \dots, n_r$ and $\ell = 1, \dots, n_t$, where $V_{k,\ell}$ is as given in (7). Then

$$\lim_{n_r, n_t \rightarrow \infty} g_{n_r, n_t}(\theta_r, \theta_t) = \Psi(\theta_r, \theta_t) \quad \text{uniformly for } (\theta_r, \theta_t) \in [0, 1]^2. \quad (10)$$

We will exploit this convergence when we study the asymptotic capacity in Section 4.

Remark 3. The convergence result of (10) obviously also holds if we set $V_{k,\ell} = \Psi(\theta_{r,k}, \theta_{t,\ell})$ as in the approximation of (8).

3 Optimal Input Distribution and Capacity

Based on (3), we can rewrite the input-output relationship (1) in the virtual domain as

$$\tilde{\mathbf{y}} = \sqrt{\frac{\Gamma}{n_t}} \tilde{\mathbf{H}} \tilde{\mathbf{x}} + \tilde{\mathbf{w}} \quad (11)$$

where $\tilde{\mathbf{x}} = A_t^\dagger \mathbf{x}$, $\tilde{\mathbf{y}} = A_r^\dagger \mathbf{y}$, and $\tilde{\mathbf{w}} = A_r^\dagger \mathbf{w}$. 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 (11). As mentioned in Section 1 we make the coherent channel assumption throughout the paper.

Referring to (1), the results of [2, Section 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 $\underline{\mathbf{x}}$ with a covariance matrix Q that satisfies $\text{Tr}(Q) \leq n_t$. The capacity is hence given by

$$C = \max_{Q: \text{Tr}(Q) \leq n_t} \mathbb{E} \left[\log \det \left(I + \frac{\Gamma}{n_t} \mathbf{H} Q \mathbf{H}^\dagger \right) \right], \quad (12)$$

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.

3.1 Characterization of optimal input distribution

From (11), it is clear that the optimal input in the virtual domain is also a zero-mean proper complex Gaussian input vector $\tilde{\underline{\mathbf{x}}}$ 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

$$\mathcal{I}(\tilde{Q}) = \mathbb{E} \left[\log \det \left(I + \frac{\Gamma}{n_t} \tilde{\mathbf{H}} \tilde{Q} \tilde{\mathbf{H}}^\dagger \right) \right]. \quad (13)$$

The capacity of (12) can be rewritten as

$$C = \max_{\tilde{Q}: \text{Tr}(\tilde{Q}) \leq n_t} \mathcal{I}(\tilde{Q}) \quad (14)$$

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 (14), 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. *The optimal \tilde{Q} that maximizes the mutual information (13) is unique and it is diagonal.*

Proof. The proof follows the proof of Theorem 3.1 in [5]. 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 (13) with \tilde{Q} restricted to the set Ω' . Since the set Ω' is convex and compact, and the function $\mathcal{I}(\tilde{Q})$ is differentiable and strictly concave over that set, there exists a unique Λ° that maximizes $\mathcal{I}(\tilde{Q})$ over Ω' . By Theorem 2 in Chapter 7.4 of [14], Λ° satisfies the following necessary condition

$$\delta \mathcal{I}(\Lambda^\circ; \Lambda - \Lambda^\circ) \leq 0, \quad \forall \Lambda \in \Omega' \quad (15)$$

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 (15) can be computed as follows.

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

where we used the general formula

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

with A and B being Hermitian matrices and x being a real scalar. Then the condition (15) becomes

$$\mathbb{E} \operatorname{Tr} \left\{ \left(I + \frac{\Gamma}{n_t} \tilde{\mathbf{H}} \Lambda^\circ \tilde{\mathbf{H}}^\dagger \right)^{-1} \frac{\Gamma}{n_t} \tilde{\mathbf{H}}(\Lambda - \Lambda^\circ) \tilde{\mathbf{H}}^\dagger \right\} \leq 0, \quad \forall \Lambda \in \Omega' \quad (16)$$

Now we want to show that Λ° remains optimal even when the optimization is performed over the set Ω . Since $\mathcal{I}(\tilde{Q})$ is strictly concave over the convex set Ω , it is sufficient to show that

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

To that 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

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

Since $\Lambda_{\tilde{Q}} \in \Omega'$, the first term in the above equation is less than zero by (16). 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

$$\begin{aligned}
&\mathbb{E} \operatorname{Tr} \left\{ \left(I + \frac{\Gamma}{n_t} \tilde{\mathbf{H}} \Lambda^\circ \tilde{\mathbf{H}}^\dagger \right)^{-1} \frac{\Gamma}{n_t} \tilde{\mathbf{H}} A \tilde{\mathbf{H}}^\dagger \right\} \\
&= \sum_{\substack{k, \ell=1 \\ k \neq \ell}}^{n_t} \mathbb{E} \operatorname{Tr} \left\{ \left(I + \frac{\Gamma}{n_t} \sum_{i=1}^{n_t} \lambda_i^\circ \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\dagger \right)^{-1} \frac{\Gamma}{n_t} A_{k, \ell} \tilde{\mathbf{h}}_k \tilde{\mathbf{h}}_\ell^\dagger \right\} \quad (17)
\end{aligned}$$

where λ_i° is the i -th diagonal entry of Λ° .

In the above sum, consider a particular term

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

Since $\tilde{\mathbf{H}}$ has independent entries, its columns are independent. Thus, conditioned on $\tilde{\mathbf{h}}_2, \tilde{\mathbf{h}}_3, \dots, \tilde{\mathbf{h}}_{n_t}$, $\tilde{\mathbf{h}}_1$ is still zero mean proper complex Gaussian. 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, the inner expectation

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

Hence the particular term we considered in the sum on the right hand side of equation (17) 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. \square

We have shown that the optimal covariance matrix \tilde{Q} is a diagonal matrix Λ° . This means that the capacity achieving input vector in the virtual domain has independent entries, i.e., the optimal input signals transmitted at different virtual directions are independent. For diagonal input covariance Λ , diagonal element λ_i represents the power assigned to the i -th virtual transmit angle. We are still left with the problem of finding the λ_i 's that achieve the capacity. But this is a problem of optimizing a concave function over a convex set and can easily be solved numerically as we show in Section 5.

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

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

3.2 Asymptotically optimal power allocation at low SNR

While it is relatively easy to compute Λ° 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 Λ is optimized by Λ° with all the elements equal to zero except that $\lambda_i^\circ = n_t$, 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 Λ° is such that

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

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

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

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

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

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

Hence

$$\mathbb{E}[\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}}] = \text{diag} \left(\sum_{k=1}^{n_r} V_{k,1}, \sum_{k=1}^{n_r} V_{k,2}, \dots, \sum_{k=1}^{n_r} V_{k,n_t} \right). \quad (18)$$

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

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

It is straightforward to see that the optimizing Λ° 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. \square

The above theorem suggests that as SNR approaches to zero, the optimal input strategy tends to transmit a strong signal along one virtual transmit angle with the largest channel gains rather than spread the power among all directions. 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, since the MIMO channel can be treated as an effective scalar channel for which one-dimensional codes can be designed to achieve capacity.

3.3 A 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*

$$\Gamma \sum_{k=1}^{n_r} (1 - \mu_{k,i}) V_{k,\ell^\circ} - \sum_{k=1}^{n_r} \mu_{k,i} \leq 0, \quad (19)$$

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

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

Among the n_t conditions in (19) corresponding to $1 \leq i \leq n_t$, at most one can be satisfied.

Proof. We note that the steps of the proof are similar to those given in [6] for the product-form correlation function.

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 Λ can be parameterized in the following way

$$\Lambda = \text{diag}\{n_t - p, p\beta_2, \dots, p\beta_{n_t}\} \quad (20)$$

where $0 \leq p \leq n_t$ and

$$\beta_i \geq 0, \quad \text{for } 2 \leq i \leq n_t; \quad \text{and} \quad \sum_{i=2}^{n_t} \beta_i \leq 1. \quad (21)$$

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

$$\mathcal{I}(p) = \mathbb{E} \log \det \left(I + \frac{\Gamma}{n_t} (n_t - p) \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^\dagger + \frac{\Gamma}{n_t} \sum_{i=2}^{n_t} p \beta_i \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^\dagger \right). \quad (22)$$

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

Lemma 1. *The function $\mathcal{I}(p)$ in (22) is a strict concave function over $0 \leq p \leq n_t$ for all $\{\beta_i\}$ that satisfy the condition (21).*

Based on Lemma 1, a necessary and sufficient condition for beamforming to the first transmit virtual angle is optimal is given by

$$\left. \frac{\partial \mathcal{I}(p)}{\partial p} \right|_{p=0} \leq 0 \quad (23)$$

for all $\{\beta_i\}$ such that (21) is satisfied. The following lemma, whose proof is given in the Appendix, provides an expression for the derivative on the left hand side of (23).

Lemma 2.

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

Thus, to satisfy condition (23), 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 $\beta_{\ell^o} = 1$ where ℓ^o 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:

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

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 for the case when beamforming to the i -th virtual angle is optimal as shown in (19).

We are now left to show that at most one of the n_t conditions in (19) 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 (20). 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 is not optimal. \square

Remark 4. *It is interesting to compare the results of Theorems 2 and 3 for the special case where two columns (say i and ℓ) 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 ℓ . In particular, it is asymptotically optimal to beamform to virtual angle i (or ℓ). However, in Theorem 3, if the beamforming condition (19) for angle i is satisfied, it will also be satisfied for angle ℓ . 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.*

4 Asymptotic Capacity for Large Number of Antennas

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

$$C = \mathbb{E} \left[\log \det \left(I + \frac{\Gamma}{n_t} \tilde{\mathbf{H}} \Lambda^\circ \tilde{\mathbf{H}}^\dagger \right) \right] \quad (24)$$

where Λ° 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

$$C = \mathbb{E} \left[\log \left(1 + \Gamma \|\tilde{\mathbf{h}}_i\|^2 \right) \right] \quad (25)$$

The expectation in (24) (or (25)) 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 τ (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

$$\bar{C} = \frac{1}{n} \mathbb{E} \left[\log \det \left(I + \frac{\Gamma}{\lfloor \tau n \rfloor} \tilde{\mathbf{H}} \Lambda^\circ \tilde{\mathbf{H}}^\dagger \right) \right]. \quad (26)$$

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

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 (26) it is clear that the evaluation of \bar{C}^∞ requires the characterization of the limiting value of the optimal diagonal input covariance matrix Λ° . However, this is a rather difficult task considering that in general it is not possible to obtain an explicit equation for Λ° 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 (26) with Λ° being replaced by some reasonable diagonal input covariance Λ that can be characterized explicitly⁷. In this case we need to interpret \bar{C} as simply the information rate that is achievable by using the input covariance Λ . To proceed with the analysis, we need to impose some regularity conditions on $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{n\tau})$.

Assumption 2. For each n , define the function $s_n : [0, \tau] \rightarrow \Re$ 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 5. A natural question to ask is whether Λ° satisfies Assumption 2. It is clear that the assumption will not hold for Λ° 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 piece-wise continuity and boundedness of the scattering function Ψ ; 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 a $n \times n$ Hermitian matrix A is defined as

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

⁷For example, we can consider $\Lambda = I$ corresponding to i.i.d. inputs in both the virtual and actual domains.

We now have the following lemma that follows quite easily from a similar result on the sum capacity of CDMA systems with random spreading [15]. A sketch of the proof is given in the Appendix.

Lemma 3. *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:*

$$\bar{C}^\infty = \int_0^1 \frac{1}{t} \left[1 - \frac{\tau}{t\Gamma} m\left(-\frac{\tau}{t\Gamma}\right) \right] dt \quad (27)$$

Lemma 3 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 [16, Corollary 10.1.2] which is restated below.

Theorem 4 (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*

$$f_n(u, v) = n \text{Var}(X_{i,j}), \quad \text{for } u \in \left[\frac{i-1}{n}, \frac{i}{n} \right], \quad v \in \left[\frac{j-1}{n}, \frac{j}{n} \right]$$

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

$$\frac{1}{n} \sum_{i=[an]}^{[bn]} (\mathbf{X}\mathbf{X}^\dagger - zI)_{ii}^{-1} \xrightarrow{n \rightarrow \infty} \int_a^b e(u, z) du \quad (28)$$

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 (28), 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 4, we give the following lemma whose proof follows easily from (10) and Assumption 2.

Lemma 4. *For each n , define the function $g_n : [0, 1] \times [0, \tau] \rightarrow \Re$ 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

$$g(u, v) = \Psi(u, v/\tau) s(v). \quad (29)$$

Using Lemma 4 in Theorem 4, we have the following result that characterizes the asymptotic capacity.

Theorem 5. *As $n \rightarrow \infty$, the Stieltjes transform $\tilde{\mathbf{H}}\Lambda\tilde{\mathbf{H}}^\dagger/n$ converges in probability to a deterministic limit*

$$m(z) = \int_0^1 e(u, z) du \quad (30)$$

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

$$e(u, z) = \left[-z + \int_0^\tau \frac{g(u, v) dv}{1 + \int_0^1 e(w, z) g(w, v) dw} \right]^{-1} \quad (31)$$

with $g(u, v)$ defined in (29).

Using (27), (29), (30) and (31), the asymptotic normalized capacity \bar{C}^∞ can be computed numerically for any input covariance satisfying Assumption 2 and any bounded spatial scattering function Ψ . Sample numerical results are presented in the following section.

The asymptotic normalized capacity \bar{C}^∞ 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:

$$C = n_r \bar{C}^\infty(\tau). \quad (32)$$

We will investigate the accuracy of this approximation in the following section.

5 Numerical Results and Discussion

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

5.1 Optimal input distribution and beamforming

We begin with an example illustrating the results of Section 3. Consider a system with 5 transmit and 5 receive antennas, where the channel is such that the variance matrix is given by:

$$V = \frac{25}{5.7} \begin{pmatrix} 0.1 & 0 & 1 & 0 & 0 \\ 0 & 0.1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.25 & 0 \\ 0 & 0 & 1 & 0 & 0.25 \end{pmatrix}. \quad (33)$$

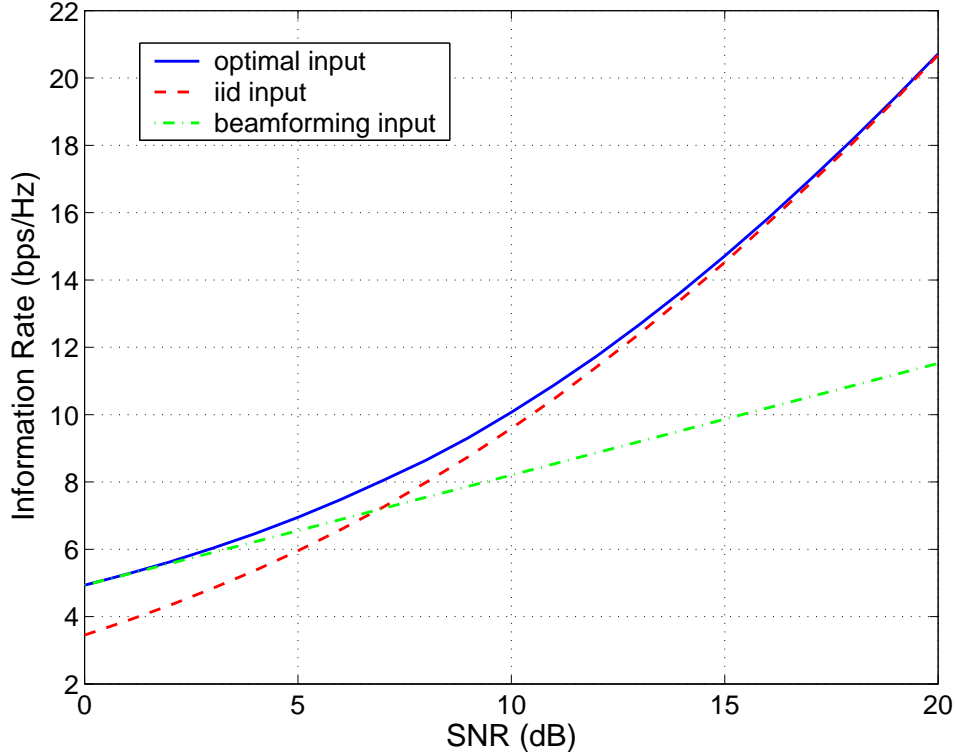


Figure 4: Comparison of information rates obtained by using the optimal inputs, i.i.d. inputs and beamforming to the 5-th antenna for the variance matrix shown in (33)

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 $\{\lambda_i^o\}$ can easily be obtained numerically for this example. Figure 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 increases⁸. 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 5-th virtual transmit direction is optimal for SNR's below 0.29 dB. Figure 4 also plots the the information rate obtained by beamforming along the 5-th virtual transmit direction for SNR's ranging from 0 to 20 dB. As

⁸It is easy to show for this example that i.i.d. inputs are indeed asymptotically optimal as the SNR goes to infinity. This would not be the case if the variance matrix V has one or more all-zero columns.

can be seen in the figure, beamforming is indeed optimal for SNR's below 0.29 dB, and it remains close to optimal for SNR's below 5 dB.

5.2 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 (33). It is interesting to see in Figure 5, that for SNR's below 2 dB,

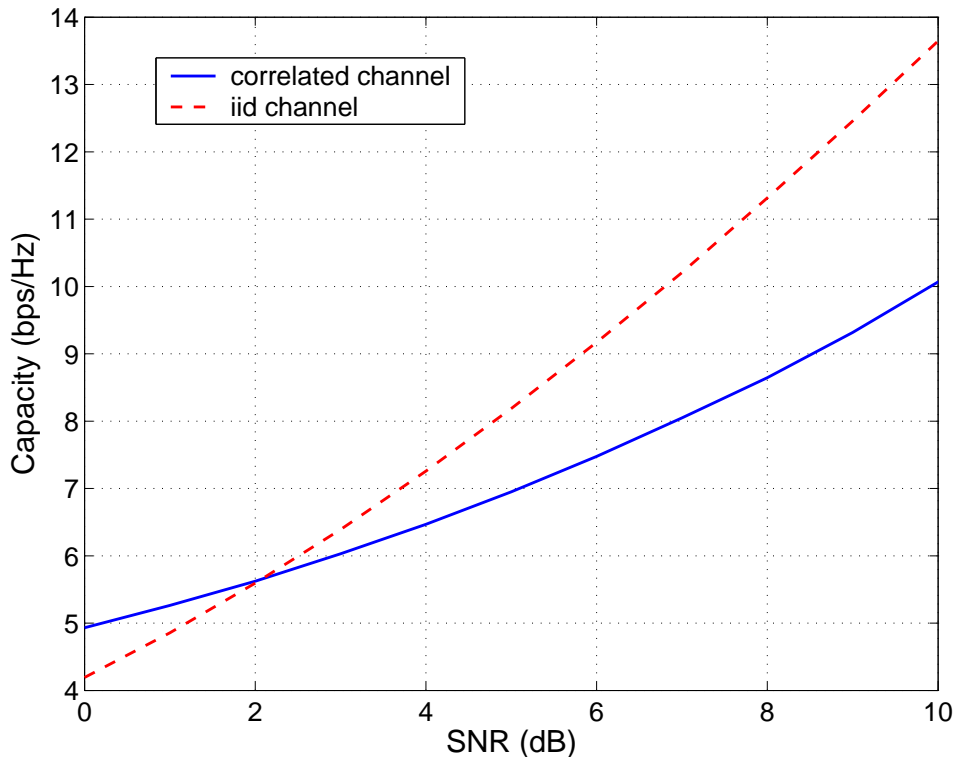


Figure 5: Comparison of capacities of the i.i.d. and correlated channel with V given in (33)

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 SNR's. We note further that the capacity of the correlated channel can be approached by using beamforming inputs at low SNR's as we saw previously in Figure 4.

5.3 Accuracy of Asymptotics

In the following we compute the asymptotic normalized capacity of Section 4 for a specific example and investigate the accuracy of these asymptotics. Consider the scattering function shown in

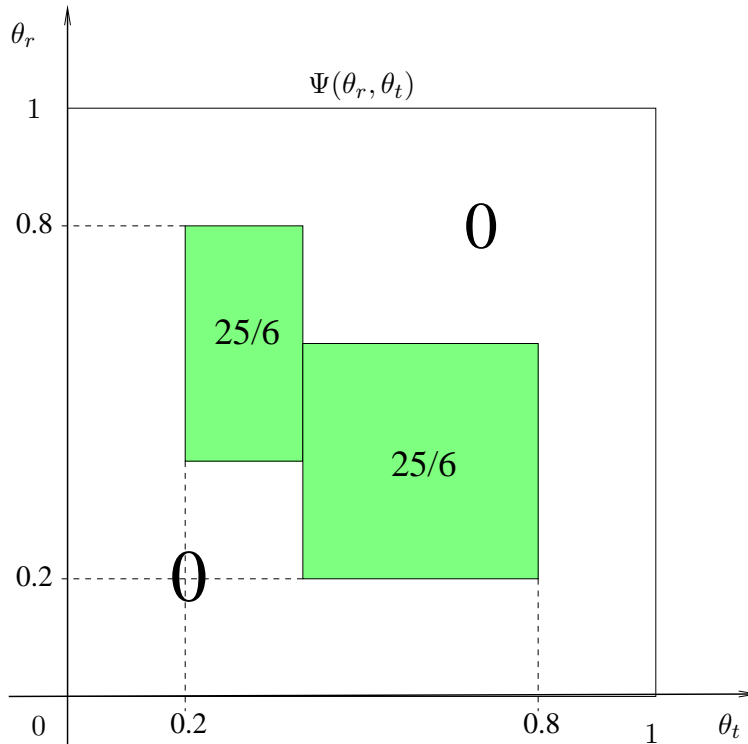


Figure 6: Spatial scattering function.

Figure 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, based on Assumption 1, all that is needed to characterize the channel is the variance matrix V . To simplify the calculations, we set $V_{k,\ell} = \Psi(\theta_{r,k}, \theta_{t,\ell})$ (see Remark 3). 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 Λ° satisfies Assumption 2 and the corresponding limiting function $s(v)$ is given by:

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

Plugging in $\Psi(\theta_r, \theta_t)$ and $s(v)$ in (27), (29), (30) and (31), we can numerically evaluate the

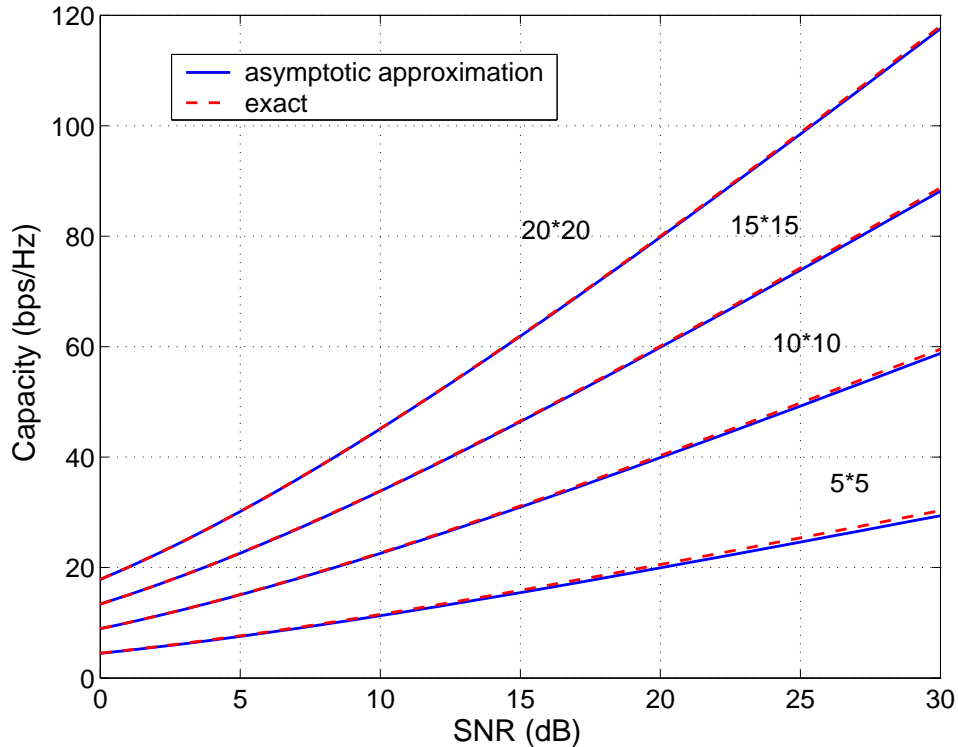


Figure 7: Comparison of the exact and asymptotic approximations for the capacity for the scattering function shown in Figure 6.

asymptotic normalized capacity \bar{C}^∞ for various values of $\tau = n_t/n_r$ and SNR Γ . We can then use (32) to approximate the capacity of any given (n_t, n_r) MIMO system operating in this scattering environment. Figure 7 plots the direct numerical evaluation of the capacity 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.

6 Conclusions

We have exploited the virtual representation of a MIMO Rayleigh fading channel to study its capacity under the coherent channel assumption. The virtual representation provides a simple characterization of arbitrary correlated MIMO channels corresponding to systems with uniform linear arrays (ULA's) of antennas. In particular, we have shown that regardless of the channel correlation, the channel capacity is achieved by transmitting independent 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 a MIMO system. This is also consistent with recent results on MMSE estimation of correlated MIMO channels [9] which show that it is only efficient to 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 SNR's. Our results also suggest that for moderately correlated channels, beamforming may be nearly optimal for a large range of practical SNR's. In such scenarios, the MIMO channel can be effectively treated as a scalar channel and space-time coding is not required.

The virtual representation also allowed 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. Our numerical results indicate that these asymptotics are accurate for moderate numbers of transmit and receive antennas even when the channel is correlated.

The results in this paper generalize all previous results on the capacity of coherent MIMO channels and illustrate the power of the virtual representation introduced in [11]. Unlike existing results, the virtual representation is not restricted to product-form correlation model but includes it as a special case.

Finally, we note that the virtual representation can be extended to the most general setting of time- and frequency-selective correlated MIMO channels [17], and in future work we plan to investigate wideband 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 wideband correlated MIMO 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 correlated MIMO channels.

Appendix

A Proof of Lemma 2

We compute $\left. \frac{\partial \mathcal{I}(p)}{\partial p} \right|_{p=0}$ as follows.

$$\begin{aligned}
\left. \frac{\partial \mathcal{I}(p)}{\partial p} \right|_{p=0} &= \mathbb{E} \operatorname{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} \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} \operatorname{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\} \\
&= \operatorname{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\} \\
&= \operatorname{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 \begin{pmatrix} V_{1,i} & & 0 \\ & \ddots & \\ 0 & & V_{n_r,i} \end{pmatrix} \right. \\
&\qquad \qquad \qquad \left. - \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\} \\
&= \operatorname{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 \begin{pmatrix} V_{1,i} & & 0 \\ & \ddots & \\ 0 & & V_{n_r,i} \end{pmatrix} \right\} \\
&\qquad \qquad \qquad - \frac{1}{n_t} \mathbb{E} \left[\frac{\Gamma \|\tilde{\mathbf{h}}_1\|^2}{1 + \Gamma \|\tilde{\mathbf{h}}_1\|^2} \right].
\end{aligned}$$

In the above equation, 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

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

B Proof of Lemma 3

Replacing Λ° with Λ in (26), 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 τn 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

$$\mathbf{d}_n = \frac{1}{n} \sum_{k=1}^n \left[\log \left(\gamma_n^{(k)} + \sigma \right) - \log \sigma \right] = \frac{1}{n} \sum_{k=1}^n \int_0^1 \frac{\gamma_n^{(k)}}{\sigma + t \gamma_n^{(k)}} dt. \quad (34)$$

Let

$$\mathbf{f}_n(t) = \frac{1}{n} \sum_{k=1}^n \frac{\gamma_n^{(k)}}{\sigma + t \gamma_n^{(k)}} = \frac{1}{t} - \frac{\sigma}{t^2} \mathbf{m}_n \left(-\frac{\sigma}{t} \right) \quad (35)$$

where $\mathbf{m}_n(z)$ is the Stieltjes transform of $\tilde{\mathbf{H}} \Lambda \tilde{\mathbf{H}}^\dagger / n$. Based on (34) and (35), 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 $n \rightarrow \infty$

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

The exchange of the limit and the integral follows from the Dominated Convergence Theorem

because of the following boundedness of $\mathbb{E}[\mathbf{f}_n(t)]$.

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

where the last equality follows from (18). Now using (9) and Assumption 2, it is clear that right hand side of the above inequality is uniformly bounded for all n and t . Thus

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

Finally, we have

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

by using the convergence assumed in the statement of the lemma 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 Pers. Commun.* 6: 311-335, 1998.
- [2] I.E. Telatar. Capacity of multiantenna Gaussian channels. *European Trans. Telecommun.* 10(6): 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.*, 20(3):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. Inform. Th.*, 48(3):637-650, March 2002.
- [5] E. Visotsky and U. Madhow. Space-time transmit precoding with imperfect feedback. *IEEE Trans. Inform. Th.*, 47(6):2632-2639, Sept. 2001.
- [6] S. Jafar and A. Goldsmith. Transmitter optimization and optimality of beamforming for multiple antenna systems. *submitted to IEEE Trans. wireless. Commun.*, 2003.

- [7] E. Jorswieck and H. Boche. Characterization of the optimality range of beamforming for multiple transmit and receive antennas with covariance feedback. *Proc. Conf. Inform. Sci. Sys. (CISS)*, Princeton, March 2002.
- [8] H. Boche and E. Jorswieck. Optimal transmit strategies for MIMO systems with partial channel state information. *Proc. SPIE International Symposium ITCOM*, Boston, July 2002.
- [9] J. Kotecha and A. M. Sayeed. Optimal signal design for estimation of correlated MIMO channels. *to appear in the IEEE Trans. Signal Processing*, 2003.
- [10] J. Kotecha and A. Sayeed. On the capacity of correlated MIMO channels with covariance feedback. In *Proc. of the 2003 International Symposium on Information Theory*, Yokohama, Japan, July 2003.
- [11] A.M. Sayeed. Deconstructing multi-antenna fading channels. *IEEE Trans. Signal Processing*, 50(10):2563–2579, Oct. 2002.
- [12] M.A. Kamath and B.L. Hughes. The asymptotic capacity of multiple-antenna Rayleigh fading channels. *Proc. 40th Annual Allerton Conference on Communication, Control, and Computing.*, Oct. 2002. Also submitted to the *IEEE Transactions on Information Theory*.
- [13] K. Liu, V. Raghavan, and A. M. Sayeed. Capacity scaling and spectral efficiency in wideband correlated MIMO channels. *to appear in the IEEE Trans. Inf. Th. (special issue on MIMO systems)*, 2003.
- [14] D.G. Luenberger. *Optimization by vector space methods*. John Wiley and Sons, 1997.
- [15] A. Mantravadi. *Analysis and Design of Wideband Multiantenna CDMA Systems*. PhD thesis, Cornell University, 2002.
- [16] V. L. Girko. *Theory of Random Determinants*. Kluwer Academic Publishers, Amsterdam, 1990.
- [17] A. M. Sayeed and V. Veeravalli. The Essential Degrees of Freedom in Space-Time Fading Channels. *Proc. 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'02)*, Lisbon, Portugal, September 2002.