

# Unitary Isotropically Distributed Inputs Are Not Capacity-Achieving for Large-MIMO Fading Channels

Wei Yang and Giuseppe Durisi  
Department of Signals and Systems  
Chalmers University of Technology  
41296 Gothenburg, Sweden  
E-mail: {ywei, durisi}@chalmers.se

Erwin Riegler  
Institute for Telecommunications  
Vienna University of Technology  
1040 Vienna, Austria  
E-mail: {erwin.riegler@nt.tuwien.ac.at}

**Abstract**—We analyze the capacity of Rayleigh block-fading multiple-input multiple-output (MIMO) channels in the noncoherent setting and prove that *unitary space-time modulation* (USTM) is not capacity-achieving when the total number of antennas exceeds the coherence time of the fading channel. This situation is relevant for MIMO systems with large antenna arrays (large-MIMO systems). Our result settles a conjecture by Zheng & Tse (2002) in the affirmative. The capacity-achieving input signal, which we refer to as *Beta-variate space-time modulation* (BSTM), turns out to be the product of a unitary isotropically distributed random matrix, and a diagonal matrix whose nonzero entries are distributed as the square-root of the eigenvalues of a Beta-distributed random matrix of appropriate size. Numerical results illustrate that using BSTM instead of USTM in large-MIMO systems yields a rate gain as large as 13% for SNR values of practical interest.

## I. INTRODUCTION

The use of multiple antennas increases tremendously the throughput of wireless systems operating over fading channels. Specifically, when a genie provides the receiver with perfect channel state information (the so called *coherent setting*), the capacity of a multiple-antenna fading channel grows linearly in the minimum between the number of transmit and receive antennas [1]. In practice, however, the fading channel is not known *a priori* at the receiver and must be estimated. Lack of *a priori* channel knowledge at the receiver determines a capacity loss compared to the coherent case. This loss, which depends on the rate at which the fading channel varies in time, frequency, and space [2]–[5], can be characterized in a fundamental way by studying capacity in the *noncoherent setting* where neither the transmitter nor the receiver are assumed to have *a priori* knowledge of the realizations of the fading channel (but both are assumed to know its statistics perfectly). In the remainder of the paper, we will refer to capacity in the noncoherent setting simply as capacity.

For frequency-flat fading channels, a simple model to capture channel variations in time is the Rayleigh block-fading model according to which the channel remains constant over a block of  $T \geq 1$  symbols and changes independently from block to

block. The parameter  $T$  can be thought of as the channel's coherence time. Even if the capacity of the Rayleigh block-fading multiple-input multiple-output (MIMO) channel has been studied extensively in the literature [2], [6], [3], [7], no closed-form capacity expression is available to date. Zheng and Tse [3] showed that the capacity  $C$  grows as

$$C(\rho) = M^*(1 - M^*/T) \log(\rho) + \mathcal{O}(1), \quad \rho \rightarrow \infty. \quad (1)$$

Here,  $\rho$  denotes the SNR,  $M^* = \min\{M, N, \lfloor T/2 \rfloor\}$ , with  $M$  and  $N$  standing for the number of transmit and receive antennas, respectively, and  $\mathcal{O}(1)$  indicates a bounded function of  $\rho$  (for sufficiently large  $\rho$ ). The capacity expression (1) implies that, at high SNR, the capacity loss due to lack of *a priori* channel knowledge is large when the channel's coherence time  $T$  is small. It also implies that at high SNR, the capacity-maximizing number of transmit antennas  $M$  (for fixed  $T$  and a fixed number of receive antennas  $N$ ) is  $\min\{N, \lfloor T/2 \rfloor\}$ .

When  $T \geq M + N$  (channel's coherence time larger or equal to the total number of antennas), the high-SNR expression (1) can be tightened as follows [3, Sec. IV.B]

$$C(\rho) = M^*(1 - M^*/T) \log(\rho) + c + o(1), \quad \rho \rightarrow \infty. \quad (2)$$

Here,  $o(1) \rightarrow 0$  as  $\rho \rightarrow \infty$ , and  $c$ , which is given in [3, Eq. (24)], depends on  $T$ ,  $M$ , and  $N$  but not on  $\rho$ . Differently from (1), the high-SNR expression (2) describes capacity accurately already at moderate SNR values [7], [8] because it captures the first two terms in the asymptotic expansion of capacity for  $\rho \rightarrow \infty$ . The key element exploited in [3] to establish (2) is the optimality of *isotropically distributed unitary* input signals [2, Sec. A.2] at high SNR; the isotropic unitary input distribution is often referred to as *unitary space-time modulation* (USTM) [9], [7].

In this paper, we shall focus on the case  $T < M + N$  (channel's coherence time smaller than the total number of antennas), which is of interest for communication systems using large antenna arrays. The use of large antenna arrays in MIMO systems (*large-MIMO* systems) has been recently advocated to reduce energy consumption in wireless networks, to combat the effect of small-scale fading, and to release multi-user MIMO gains with limited

The work of Erwin Riegler was supported by the WWTF under grant ICT10-066 (NOWIRE).

co-operation among base stations and low complexity channel estimation algorithms [10]–[12].

*Contributions:* We prove that in the large-MIMO setting where  $T < M + N$ , USTM is not capacity-achieving at high SNR. We show that the capacity-achieving input signal is the product of a unitary isotropically distributed random matrix, and a diagonal matrix whose nonzero entries are distributed as the square-root of the eigenvalues of a Beta-distributed random matrix of appropriate size. Utilizing this input distribution, which we refer to as *Beta-variate space-time modulation (BSTM)*, we extend (2) to the case  $T < M + N$ . The use of BSTM instead of USTM when  $N \gg T$  turns out to yield a rate gain larger than 10% at moderate SNR values.

*Notation:* Uppercase boldface letters denote matrices and lowercase boldface letters designate vectors. The superscripts  $\top$  and  $\text{H}$  stand for transposition and Hermitian transposition, respectively. We denote the identity matrix of dimension  $M \times M$  by  $\mathbf{I}_M$ , and  $\text{diag}\{\mathbf{a}\}$  is the diagonal square matrix whose main diagonal contains the entries of the vector  $\mathbf{a}$ . The distribution of a circularly-symmetric complex Gaussian random vector with covariance matrix  $\Sigma$  is denoted by  $\mathcal{CN}(\mathbf{0}, \Sigma)$ ;  $\text{Gamma}(\cdot, \cdot)$  is the Gamma distribution, and  $\text{Beta}(\cdot, \cdot)$  denotes the Beta distribution [13]. Finally,  $\log(\cdot)$  indicates the natural logarithm, and  $\Gamma(\cdot)$  denotes the Gamma function.

## II. SYSTEM MODEL

We consider a Rayleigh block-fading MIMO channel with  $M$  transmit antennas,  $N$  receive antennas, and channel's coherence time  $T$ . The channel input-output relation within a coherence interval can be compactly written in matrix notation as follows [6], [7], [3]:

$$\mathbf{Y} = \sqrt{\rho/M} \mathbf{X} \mathbf{H} + \mathbf{W}. \quad (3)$$

Here,  $\mathbf{X} \in \mathbb{C}^{T \times M}$  contains the signal transmitted from the  $M$  antennas within the coherence interval,  $\mathbf{H} \in \mathbb{C}^{M \times N}$  is the channel's propagation matrix,  $\mathbf{W} \in \mathbb{C}^{T \times N}$  is the additive noise, and  $\mathbf{Y} \in \mathbb{C}^{T \times N}$  contains the signal received at the  $N$  antennas within the coherence interval. We will assume throughout the paper that  $M \leq \min\{N, \lfloor T/2 \rfloor\}$ . The random matrices  $\mathbf{H}$  and  $\mathbf{W}$  are independent of each other and have independent and identically distributed (i.i.d.)  $\mathcal{CN}(0, 1)$  entries. We consider the noncoherent setting where neither the transmitter nor the receiver have *a priori* knowledge of the realizations of  $\mathbf{H}$  and  $\mathbf{W}$ , but both know their statistics perfectly.

We assume that  $\mathbf{H}$  and  $\mathbf{W}$  take on independent realizations over successive coherence intervals. Under this *block-memoryless* assumption, the ergodic capacity of the channel in (3) is given by

$$C(\rho) = \frac{1}{T} \sup_{\mathbf{Q}_{\mathbf{X}}} I(\mathbf{X}; \mathbf{Y}). \quad (4)$$

The supremum is over all probability distributions  $\mathbf{Q}_{\mathbf{X}}$  on  $\mathbf{X}$  that satisfy the average-power constraint

$$\mathbb{E} [\text{tr}\{\mathbf{X} \mathbf{X}^{\text{H}}\}] \leq TM. \quad (5)$$

Since the variance of the entries of  $\mathbf{H}$  and  $\mathbf{W}$  is normalized to one, (5) implies that  $\rho$  in (3) can be interpreted as the SNR at each receive antenna. Throughout the paper, we will often use the following additional quantities:  $Q = \min\{N, T - M\}$ ,  $R = \max\{N, T - M\}$ ,  $P = \min\{N, T\}$ , and  $L = \max\{N, T\}$ .

## III. CAPACITY IN THE HIGH-SNR REGIME

### A. A Complete Asymptotic Characterization of Capacity

The main result of this paper is Theorem 1 below, which provides a high-SNR characterization of  $C(\rho)$  that generalizes (2), in that it holds also in the large-MIMO setting  $T < M + N$ .

*Theorem 1:* The capacity  $C(\rho)$  of the MIMO Rayleigh block-fading channel (3) with  $N$  receive antennas, coherence time  $T$ , and  $M \leq \min\{N, \lfloor T/2 \rfloor\}$  transmit antennas is given by

$$C(\rho) = M(1 - M/T) \log(\rho) + c + o(1), \quad \rho \rightarrow \infty \quad (6)$$

where

$$c = \frac{1}{T} \log \left( \frac{\Gamma_M(M) \Gamma_M(Q)}{\Gamma_M(N) \Gamma_M(T)} \right) + M \left( 1 - \frac{M}{T} \right) \log \left( \frac{T}{M} \right) + \frac{MQ}{T} \log \left( \frac{N}{Q} \right) + \frac{R}{T} \left( \mathbb{E} [\log \det(\mathbf{H} \mathbf{H}^{\text{H}})] - M \right). \quad (7)$$

Here,  $\Gamma_m(a) = \pi^{m(m-1)/2} \prod_{k=1}^m \Gamma(a - k + 1)$  is the *complex multivariate Gamma function*.

*Proof:* The proof, which is omitted for space limitations and can be found in [14, Sec. IV], exploits the geometric structure in the input-output relation (3) first observed in [3]. The tools used to establish (6) are, however, different from the ones used in [3]. In particular, differently from [3], our proof is based on the duality approach [4], and a novel closed-form characterization of the probability density function (pdf) of the channel output  $\mathbf{Y}$  in (3), which generalizes a previous result obtained in [7]. These two tools allow us not only to generalize (2) to the large-MIMO setting  $T < M + N$ , but also to simplify the corresponding derivation, compared to the one provided in [3] for the case  $T \geq M + N$ . An outline of the proof for the single-input multiple-output (SIMO) case, which sheds light on the structure of the capacity-achieving input distribution, is provided in Section III-C.

### B. Rate Achievable with USTM

For the case  $T \geq M + N$ , the high-SNR capacity expression (6) coincides with the one reported in [3, Sec. IV.B].<sup>1</sup> In this case, USTM, i.e., setting  $\mathbf{X} = \sqrt{T} \Phi$ , with  $\Phi$  unitary and isotropically distributed, achieves (6). When  $T < M + N$ , the novel high-SNR capacity characterization provided in Theorem 1 implies that USTM is not capacity-achieving at high SNR, as formalized in the following corollary.

*Corollary 2:* The rate achievable using USTM over the MIMO Rayleigh block-fading channel (3) with  $N$  receive antennas, coherence time  $T$ , and  $M \leq \min\{N, \lfloor T/2 \rfloor\}$  transmit antennas is given by

$$L(\rho) = M(1 - M/T) \log(\rho) + l + o(1), \quad \rho \rightarrow \infty \quad (8)$$

<sup>1</sup>The expression for  $c$  given in [3, Eq. (24)] contains a typo: the argument of the logarithm in the second addend should be divided by  $M$  as one can verify by comparing [3, Eq. (24)] with the result given in [3, Thm. 9] for the case  $M = N$ .

where

$$l = \frac{1}{T} \log \left( \frac{\Gamma_M(M)}{\Gamma_M(T)} \right) + M \left( 1 - \frac{M}{T} \right) \log \left( \frac{T}{eM} \right) + \left( 1 - \frac{M}{T} \right) \mathbb{E} [\log \det(\mathbf{H}\mathbf{H}^H)].$$

Note that  $l = c$  when  $T \geq M + N$ ; however,  $l < c$  when  $T < M + N$ .

### C. Why Is USTM Not Capacity Achieving?

We next present a sketch of the proof of Theorem 1. Our aim is to provide an intuitive explanation on why USTM is not capacity-achieving at high SNR when  $T < M + N$ , and to explain why the matrix-variate Beta distribution arises in this case. We recall that a complete proof of Theorem 1 can be found in [14, Sec. IV]. For simplicity, we shall focus in this section on the SIMO case ( $M = 1$ ), for which the input-output relation (3) reduces to  $\mathbf{Y} = \sqrt{\rho} \mathbf{x}\mathbf{h}^T + \mathbf{W}$ . Here,  $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ , and  $\mathbf{x} \in \mathbb{C}^T$  is subject to the average-power constraint  $\mathbb{E}[\|\mathbf{x}\|^2] \leq T$ . We need to show that (see Theorem 1)

$$C(\rho) = \left( 1 - \frac{1}{T} \right) \log(\rho) + \frac{1}{T} \log \left( \frac{\Gamma(Q)}{\Gamma(N)\Gamma(T)} \right) + \left( 1 - \frac{1}{T} \right) \log(T) + \frac{Q}{T} \log \left( \frac{N}{Q} \right) + \frac{R}{T} \left( \mathbb{E}[\log(\|\mathbf{h}\|^2)] - 1 \right) + o(1), \quad \rho \rightarrow \infty. \quad (9)$$

As the capacity-achieving distribution is isotropic [2, Thm. 2], we shall assume, without loss of optimality, that  $\mathbf{x}$  is isotropically distributed. To establish (9), we analyze separately the two differential entropy terms in the definition of mutual information

$$I(\mathbf{x}; \mathbf{Y}) = h(\mathbf{Y}) - h(\mathbf{Y} | \mathbf{x}). \quad (10)$$

Note that  $\mathbf{Y}$  is conditionally Gaussian given  $\mathbf{x}$ . Hence, the second term on the right-hand side (RHS) of (10) is given by

$$h(\mathbf{Y} | \mathbf{x}) = N \mathbb{E}[\log(\rho\|\mathbf{x}\|^2 + 1)] + NT \log(\pi e) = N \log(\rho) + N \mathbb{E}[\log(\|\mathbf{x}\|^2)] + NT \log(\pi e) + o(1), \quad \rho \rightarrow \infty. \quad (11)$$

To compute  $h(\mathbf{Y})$ , we observe that, in the absence of additive noise  $\mathbf{W}$ , the columns of  $\mathbf{Y}$  are collinear with  $\mathbf{x}$  and, hence,  $\mathbf{Y}$  has rank 1. Once Gaussian noise is added,  $\mathbf{Y}$  becomes full rank. However, because a rank-1 matrix of dimension  $T \times N$  is characterized by  $T + N - 1$  parameters, the remaining  $TN - (T + N - 1) = (T - 1)(N - 1)$  parameters describing  $\mathbf{Y}$  must contain information about the additive noise only in the  $\rho \rightarrow \infty$  limit. Hence, we expect that—for an appropriate choice of the input distribution— $h(\mathbf{Y})$  should grow as  $(T + N - 1) \log(\rho)$ .

To establish this result, it is convenient to express  $\mathbf{Y}$  in terms of its singular value decomposition (SVD). Specifically, we write  $\mathbf{Y}$  as

$$\mathbf{Y} = \tilde{\mathbf{U}} \tilde{\Sigma} \tilde{\mathbf{V}}^H = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^H + \mathbf{U} \Sigma \mathbf{V}^H.$$

Here,  $\tilde{\mathbf{U}} \in \mathbb{C}^{T \times P}$  and  $\tilde{\mathbf{V}} \in \mathbb{C}^{N \times P}$  (recall that  $P = \min\{N, T\}$ ) are (truncated) unitary matrices and  $\tilde{\Sigma} = \text{diag}\{\sigma_1 \cdots \sigma_P\}^T$

contains the singular values of  $\mathbf{Y}$  arranged in decreasing order;  $\mathbf{u}_1$  and  $\mathbf{v}_1$  stand for the first column of  $\tilde{\mathbf{U}}$  and  $\tilde{\mathbf{V}}$ , respectively;  $\mathbf{U} \in \mathbb{C}^{T \times (P-1)}$  and  $\mathbf{V} \in \mathbb{C}^{N \times (P-1)}$  contain the remaining columns of  $\tilde{\mathbf{U}}$  and  $\tilde{\mathbf{V}}$ , respectively; finally,  $\Sigma = \text{diag}\{\sigma_2 \cdots \sigma_P\}^T$ . To make the SVD unique, we shall assume that the first entry of  $\mathbf{u}_1$  and the diagonal entries of  $\tilde{\mathbf{U}}$  are real and nonnegative [15, Sec. IV.5]. Let now  $\mathbf{P}_u = [\mathbf{u}_1 \tilde{\mathbf{P}}_u]$  where  $\tilde{\mathbf{P}}_u \in \mathbb{C}^{T \times (T-1)}$  is a deterministic function of  $\mathbf{u}_1$  chosen so that  $\mathbf{P}_u$  is a  $T \times T$  unitary matrix. Similarly, let  $\mathbf{P}_v = [\mathbf{v}_1 \tilde{\mathbf{P}}_v]$  where  $\tilde{\mathbf{P}}_v \in \mathbb{C}^{N \times (N-1)}$  is a deterministic function of  $\mathbf{v}_1$  chosen so that  $\mathbf{P}_v$  is a  $N \times N$  unitary matrix. By construction, we have that

$$\mathbf{Y} = \mathbf{P}_u \mathbf{P}_u^H (\sigma_1 \mathbf{u}_1 \mathbf{v}_1^H + \mathbf{U} \Sigma \mathbf{V}^H) \mathbf{P}_v \mathbf{P}_v^H = \mathbf{P}_u \begin{pmatrix} \sigma_1 & \mathbf{0}_{1 \times (N-1)} \\ \mathbf{0}_{(T-1) \times 1} & \tilde{\mathbf{Y}} \end{pmatrix} \mathbf{P}_v^H$$

where  $\tilde{\mathbf{Y}} = \tilde{\mathbf{P}}_u^H \mathbf{U} \Sigma \mathbf{V}^H \tilde{\mathbf{P}}_v \in \mathbb{C}^{(T-1) \times (N-1)}$ . The transformation  $\mathbf{Y} \mapsto (\sigma_1, \mathbf{u}_1, \mathbf{v}_1, \tilde{\mathbf{Y}})$  is one-to-one by construction, and its Jacobian  $J(\cdot)$  can be easily obtained from the Jacobian of the SVD given in [3, App. A]:

$$J(\sigma_1, \dots, \sigma_P) = \sigma_1^{2(L-P)+1} \cdot \prod_{i=2}^P (\sigma_1^2 - \sigma_i^2)^2$$

with  $L = \max\{T, N\}$ . We next compute  $h(\mathbf{Y})$  in the coordinate system induced by the transformation<sup>2</sup>  $\mathbf{Y} \mapsto (\sigma_1, \mathbf{u}_1, \mathbf{v}_1, \tilde{\mathbf{Y}})$

$$h(\mathbf{Y}) = h(\sigma_1) + h(\mathbf{u}_1) + h(\mathbf{v}_1) + h(\tilde{\mathbf{Y}} | \sigma_1) + \mathbb{E}[\log(J(\sigma_1, \dots, \sigma_P))]. \quad (12)$$

Here, the equality follows from the isotropic nature of the distribution of  $\mathbf{x}$ , which implies that  $\mathbf{u}_1$  is independent of  $\mathbf{v}_1$ , and  $(\mathbf{u}_1, \mathbf{v}_1)$  is independent of  $(\sigma_1, \tilde{\mathbf{Y}})$ . Furthermore [14, Sec. IV.B]

$$h(\mathbf{u}_1) = \log(\pi^{T-1}/\Gamma(T)); \quad h(\mathbf{v}_1) = \log(2\pi^N/\Gamma(N)). \quad (13)$$

We now exploit the *escape-to-infinity* property of the capacity-achieving distribution [14, Lem. 2], which implies that (see [14, Lem. 12 and Lem. 13])

$$h(\sigma_1) = \log(\sqrt{\rho}) + h(\|\mathbf{h}\| \cdot \|\mathbf{x}\|) + o(1), \quad \rho \rightarrow \infty$$

$$h(\tilde{\mathbf{Y}} | \sigma_1) = (N-1)(T-1) \log(\pi e) + o(1), \quad \rho \rightarrow \infty$$

$$\mathbb{E}[\log(J(\sigma_1, \dots, \sigma_P))] = (2N + 2T - 3) \mathbb{E}[\log(\|\mathbf{h}\| \cdot \|\mathbf{x}\|)] + (N + T - 3/2) \log(\rho) + o(1), \quad \rho \rightarrow \infty. \quad (14)$$

Substituting (13) and (14) into (12), and then (11) and (12) into (10), we get

$$I(\mathbf{x}; \mathbf{Y}) = (T-1) \log(\rho) + h(\|\mathbf{h}\| \cdot \|\mathbf{x}\|) + (2T-3) \mathbb{E}[\log(\|\mathbf{x}\|)] + k_1 + o(1), \quad \rho \rightarrow \infty \quad (15)$$

where

$$k_1 = \log(2) - \log(\Gamma(T) \cdot \Gamma(N)) - (T + N - 1) + (2T + 2N - 3) \mathbb{E}[\log(\|\mathbf{h}\|)].$$

<sup>2</sup>The differential entropy terms on the RHS of (12) are computed with respect to the appropriate area measure.

To conclude the proof, we need to determine the distribution on  $\|\mathbf{x}\|$  that maximizes (15). To solve this problem, it is convenient to operate one more transformation. Let  $\mathbf{g} = \|\mathbf{h}\| \cdot \|\mathbf{x}\| \cdot \hat{\mathbf{g}}$ , where  $\hat{\mathbf{g}}$  is taken uniformly distributed on the unit sphere in  $\mathbb{C}^Q$  (recall that  $Q = \min\{N, T - M\}$ ) and independent of  $\mathbf{x}$  and  $\mathbf{h}$ . By using polar coordinates, we can relate  $h(\|\mathbf{h}\| \cdot \|\mathbf{x}\|)$  and  $h(\mathbf{g})$  as follows [4, Lem. 6.17]

$$h(\|\mathbf{h}\| \cdot \|\mathbf{x}\|) = h(\mathbf{g}) - \log\left(\frac{2\pi^Q}{\Gamma(Q)}\right) - (2Q - 1) \mathbb{E}[\log(\|\mathbf{h}\| \cdot \|\mathbf{x}\|)]. \quad (16)$$

Substituting (16) into (15) yields

$$\begin{aligned} I(\mathbf{x}; \mathbf{Y}) &= (T - 1) \log(\rho) + h(\mathbf{g}) \\ &+ (T - Q - 1) \mathbb{E}[\log(\|\mathbf{x}\|^2)] + k_2 + o(1), \quad \rho \rightarrow \infty \end{aligned} \quad (17)$$

where

$$\begin{aligned} k_2 &= \log\left(\frac{\Gamma(Q)}{\Gamma(T)\Gamma(N)}\right) - (T + N - 1) \\ &+ \underbrace{(T + N - Q - 1)}_{=\max\{N, T-1\}=R} \mathbb{E}[\log(\|\mathbf{h}\|^2)] - Q \log(\pi). \end{aligned}$$

Note that maximizing the RHS of (15) amounts to maximizing the second and the third term on the RHS of (17). We next analyze these two terms separately.

For the second term on the RHS of (17), we note that, as  $\mathbb{E}[\|\mathbf{g}\|^2] = \mathbb{E}[\|\mathbf{h}\|^2 \cdot \|\mathbf{x}\|^2] \leq TN$ , then

$$h(\mathbf{g}) \leq Q \log(\pi e TN / Q) \quad (18)$$

with equality achieved if  $\mathbf{g} = (\|\mathbf{h}\| \cdot \|\mathbf{x}\| \cdot \hat{\mathbf{g}}) \sim \mathcal{CN}\left(\mathbf{0}, \frac{TN}{Q} \mathbf{I}_Q\right)$ , or, equivalently, if

$$\|\mathbf{h}\|^2 \cdot \|\mathbf{x}\|^2 \sim \text{Gamma}(Q, TN/Q). \quad (19)$$

Now note that  $\|\mathbf{h}\|^2 \sim \text{Gamma}(N, 1)$ . Hence, for the case  $Q = N$ , we can attain (19) by setting  $\|\mathbf{x}\|^2 = T$  with probability one (w.p.1). When  $Q = T - 1$ , however, we need to choose  $\|\mathbf{x}\|^2 = TN\bar{d}^2/(T - 1)$  with  $\bar{d}^2 \sim \text{Beta}(T - 1, N + 1 - T)$  to fulfill (19). This follows from Lemma 3 below, which is a special case of Lemma 6 in Section III-D.

*Lemma 3:* Let  $u \sim \text{Beta}(p, n)$  with  $p, n \geq 0$ ; let also  $r \sim \text{Gamma}(p + n, 1)$  independent of  $u$ . Then  $(u \cdot r) \sim \text{Gamma}(p, 1)$ .

For the third term on the RHS of (17), we note that, as  $T - Q - 1 \geq 0$ , Jensen's inequality yields

$$(T - Q - 1) \mathbb{E}[\log(\|\mathbf{x}\|^2)] \leq (T - Q - 1) \log(T). \quad (20)$$

Equality in (20) is achieved if  $\|\mathbf{x}\|^2 = T$  w.p.1, or if  $Q = T - 1$ , in which case both sides of (20) vanish.

Summarizing, when  $T \geq N + 1$ , it is sufficient to take  $\|\mathbf{x}\|^2 = T$  w.p.1 to achieve equality in (18) and (20). As  $\mathbf{x}$  was taken isotropically distributed, the resulting input distribution is USTM. However, when  $T < N + 1$ , USTM is no longer optimal: achieving equality in (18) and (20) requires taking  $[(T - 1)\|\mathbf{x}\|^2/(TN)] \sim \text{Beta}(T - 1, N + 1 - T)$ . Substituting (18) and (20) in (17) and dividing by  $T$  yields (9).

#### D. The Capacity-Achieving Input Distribution

*Matrix-variate distributions:* We are now ready to describe the input distribution that achieves (6) for the general MIMO case. The following preliminary results from multivariate statistics will be needed.

*Definition 4:* An  $m \times m$  random matrix  $\mathbf{A}$  is said to have the complex Wishart distribution with  $n > 0$  degrees of freedom and covariance matrix  $\Sigma$  if  $\mathbf{A} = \mathbf{B}\mathbf{B}^H$ , where the columns of the  $m \times n$  matrix  $\mathbf{B}$  are independent and  $\mathcal{CN}(\mathbf{0}, \Sigma)$ -distributed. In this case, we shall write  $\mathbf{A} \sim \mathcal{W}_m(n, \Sigma)$ .

Note that when  $m > n$ , the matrix  $\mathbf{A}$  is singular and, hence, does not admit a pdf. In this case, the probability distribution on  $\mathbf{A}$  is sometimes referred to as *pseudo-Wishart* or *singular Wishart*. When  $m = 1$ , the Wishart distribution reduces to the Gamma distribution.

*Definition 5:* An  $m \times m$  random matrix  $\mathbf{C}$  is said to have a complex matrix-variate Beta distribution of parameters  $p > 0$  and  $n > 0$  if  $\mathbf{C}$  can be written as  $\mathbf{C} = (\mathbf{T}^H)^{-1} \mathbf{A} \mathbf{T}^{-1}$ , where  $\mathbf{A} \sim \mathcal{W}_m(p, \Sigma)$  and  $\mathbf{B} \sim \mathcal{W}_m(n, \Sigma)$  are independent, and  $\mathbf{A} + \mathbf{B} = \mathbf{T}^H \mathbf{T}$ , with  $\mathbf{T}$  upper-triangular with positive diagonal elements (Cholesky factorization). In this case, we shall write  $\mathbf{C} \sim \text{Beta}_m(p, n)$ .

Let  $\mathbf{C} \sim \text{Beta}_m(p, n)$  with  $p \geq m > 0$  and  $n > 0$ . The pdf of the ordered eigenvalues  $\lambda_1 > \dots > \lambda_m$  of  $\mathbf{C}$  takes on two different forms according to the value of  $n$ . If  $n \geq m$ , then

$$\begin{aligned} f(\lambda_1, \dots, \lambda_m) &= \frac{\pi^{m(m-1)}}{\Gamma_m(m)} \cdot \frac{\Gamma_m(p+n)}{\Gamma_m(p)\Gamma_m(n)} \\ &\cdot \prod_{i=1}^m \lambda_i^{p-m} (1 - \lambda_i)^{n-m} \cdot \prod_{i < j}^m (\lambda_i - \lambda_j)^2, \\ &1 > \lambda_1 > \dots > \lambda_m > 0. \end{aligned} \quad (21)$$

If  $0 < n < m$ , the eigenvalues of  $\mathbf{C}$  are distributed as follows

$$\begin{aligned} \lambda_1 = \dots = \lambda_{m-n} &= 1 \text{ w.p.1, and} \\ f(\lambda_{m-n+1}, \dots, \lambda_m) &= \frac{\pi^{n(n-1)}}{\Gamma_n(n)} \cdot \frac{\Gamma_n(p+n)}{\Gamma_n(m)\Gamma_n(p+n-m)} \\ &\cdot \prod_{i=m-n+1}^m (\lambda_i)^{p-m} (1 - \lambda_i)^{m-n} \cdot \prod_{m-n < i < j}^m (\lambda_i - \lambda_j)^2, \\ &1 > \lambda_{m-n+1} > \dots > \lambda_m > 0. \end{aligned} \quad (22)$$

The following lemma generalizes Lemma 3 to matrix-variate distributions

*Lemma 6:* Let  $\mathbf{S} \sim \mathcal{W}_m(p+n, \Sigma)$  with  $m > 0$ ,  $n > 0$ , and  $p \geq m$ . Furthermore, let  $\mathbf{C} \sim \text{Beta}_m(p, n)$  be independent of  $\mathbf{S}$ . Finally, put  $\mathbf{S} = \mathbf{T}^H \mathbf{T}$ , where  $\mathbf{T}$  is upper-triangular with positive diagonal elements. Then,  $\mathbf{A} = \mathbf{T}^H \mathbf{C} \mathbf{T} \sim \mathcal{W}_m(p, \Sigma)$ .

*Proof:* The theorem follows from a generalization to the complex case of [16, Thm. 3.3.1] for the nonsingular case  $n \geq m$ , and of [17, Thm. 1] for the singular case  $0 < n < m$ . ■

*The Optimal Input Distribution:* Similarly to the SIMO case (see Section III-C), the capacity-achieving distribution for the general MIMO case takes on two different forms according to the relation between  $T$ ,  $M$ , and  $N$ . Specifically, one should take  $\mathbf{X} = \Phi \mathbf{D}$  where  $\Phi$  is unitary and isotropically distributed, and

$\mathbf{D} = \sqrt{TN/Q} \cdot \tilde{\mathbf{D}}$  with  $\tilde{\mathbf{D}}$  being a diagonal matrix whose ordered positive entries  $\{\tilde{d}_1, \dots, \tilde{d}_M\}$  are distributed as follows:

a) Case  $T < M + N$ : The squared nonzero entries  $\{\tilde{d}_1^2, \dots, \tilde{d}_M^2\}$  of  $\tilde{\mathbf{D}}$  have the same joint pdf as the ordered eigenvalues of a positive-definite  $M \times M$  random matrix  $\mathbf{G} \sim \text{Beta}_M(T - M, M + N - T)$ . The resulting pdf of  $\{\tilde{d}_1^2, \dots, \tilde{d}_M^2\}$  is obtained by setting  $p = T - M$  and  $n = M + N - T$  in (21) if  $T \leq N$ , and in (22) if  $N < T < M + N$ .

b) Case  $T \geq M + N$ : The nonzero entries  $\{\tilde{d}_1, \dots, \tilde{d}_M\}$  of  $\tilde{\mathbf{D}}$  should be taken so that  $\tilde{d}_1 = \dots = \tilde{d}_M = 1$  w.p.1. This results in the USTM distribution used in [3].

We shall denote by  $\mathbf{Q}_D^{\text{opt}}$  the probability distribution of  $\mathbf{D}$  we have just introduced, and refer to the probability distribution on  $\mathbf{X} = \Phi \mathbf{D}$  resulting by choosing  $\Phi$  unitary and isotropically distributed, and  $\mathbf{D} \sim \mathbf{Q}_D^{\text{opt}}$  as BSTM. Note that BSTM reduces to USTM when  $T \geq M + N$ .

### E. Gain of BSTM over USTM

The use of USTM is motivated by several practical considerations [6], [7], [18]. Is it then worth to replace USTM by the capacity-achieving BSTM in the large-MIMO setting? In this section, we shall investigate the rate gain that results from the use of BSTM instead of USTM when  $T < M + N$ .

*Asymptotic Analysis:* In Corollary 7 below we show that the rate gain resulting from using BSTM instead of USTM grows logarithmically in the number of receive antennas.

*Corollary 7:* Let  $T$  and  $M \leq \lfloor T/2 \rfloor$  be fixed. Then

$$\lim_{N \rightarrow \infty} \lim_{\rho \rightarrow \infty} \left( C(\rho) - L(\rho) - \frac{M^2}{2T} \log(N) \right) = c_{M,T}$$

where  $C(\rho)$  and  $L(\rho)$  are given in (6) and (8), respectively, and

$$c_{M,T} = \frac{1}{T} \log(\Gamma_M(T - M)) + \frac{M(T - M)}{T} \log\left(\frac{e}{T - M}\right) - \frac{M}{2T} [M \log(\pi e) + \log(2)].$$

*Proof:* The proof is omitted for space limitations and can be found in [14, Sec. III.C].

*Numerical Results:* Let  $\tilde{C}(\rho)$  be the high-SNR approximation of  $C(\rho)$  obtained by neglecting the  $o(1)$  term in (6). Similarly, let  $\tilde{L}(\rho)$  be the high-SNR approximation of  $L(\rho)$  obtained by neglecting  $o(1)$  in (8). As can be inferred from the results reported in [3], [7], [8],  $\tilde{L}(\rho)$  is a good approximation for  $L(\rho)$  when  $\rho \gtrsim 20$  dB. Numerical evidence suggests that the same holds for  $\tilde{C}(\rho)$  and  $C(\rho)$ . To illustrate the gain resulting from the use of BSTM instead of USTM for a finite (but large) number of receive antennas, we plot in Fig. 1 the ratio  $[\tilde{C}(\rho) - \tilde{L}(\rho)]/\tilde{L}(\rho)$  for different values of  $T$  and  $N$ , when  $\rho = 30$  dB and  $M = \min\{\lfloor T/2 \rfloor, N\}$ . We observe from Fig. 1 that the rate gain resulting from the use of BSTM instead of USTM becomes significant when the number of receive antennas  $N$  is much larger than the channel's coherence time  $T$ . For example, when  $N = 100$  and  $T = 10$ , the rate gain amounts to 13%. However, when  $T = N = 100$  the rate gain is below 3%.

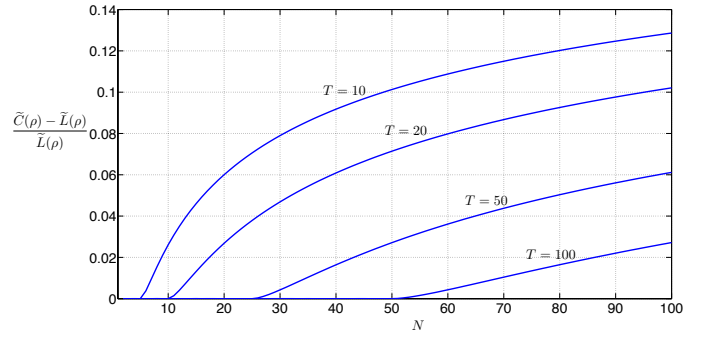


Fig. 1. Rate gain resulting from the use of BSTM instead of USTM as a function of the number of receive antennas  $N$  and the channel's coherence time  $T$ ; in the figure,  $\rho = 30$  dB, and  $M = \min\{\lfloor T/2 \rfloor, N\}$ .

### REFERENCES

- [1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.
- [2] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.
- [3] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.
- [4] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.
- [5] U. G. Schuster, G. Durisi, H. Bölcskei, and H. V. Poor, "Capacity bounds for peak-constrained multiantenna wideband channels," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2686–2696, Sep. 2009.
- [6] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.
- [7] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: the received signal density in closed form," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1473–1484, Jun. 2002.
- [8] K. Takeuchi, R. R. Müller, M. Vehkaperä, and T. Tanaka, "On an achievable rate of large Rayleigh block-fading MIMO channels with no CSI," Nov. 2011. [Online]. Available: <http://arxiv.org/abs/1111.4626v1>
- [9] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 1962–1973, Sep. 2000.
- [10] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [11] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," Dec. 2011. [Online]. Available: <http://arxiv.org/abs/1112.3810>
- [12] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, 2011, to appear.
- [13] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, 2nd ed. New York, NY, U.S.A.: Wiley, 1995, vol. 2.
- [14] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channels," *IEEE J. Sel. Areas Commun.*, Feb. 2012, submitted. [Online]. Available: <http://arxiv.org/abs/1202.0168>
- [15] P. Marques and S. Abrantes, "On the derivation of the exact, closed-form capacity formulas for receiver-sided correlated MIMO channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1139–1161, Mar. 2008.
- [16] R. I. Muirhead, *Aspects of multivariate statistical theory*. Hoboken, NJ, U.S.A.: Wiley, 2005.
- [17] H. Uhlig, "On singular Wishart and singular multivariate Beta distribution," *Ann. Statist.*, vol. 22, no. 1, pp. 395–405, 1994.
- [18] A. Ashikhmin and R. Calderbank, "Grassmannian packings from operator Reed–Muller codes," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5689–5714, Nov. 2010.