On the Capacity of Large-MIMO Block-Fading Channels

Wei Yang, *Student Member, IEEE*, Giuseppe Durisi, *Senior Member, IEEE*, Erwin Riegler, *Member, IEEE*

Abstract—We characterize the capacity of Rayleigh block-fading multiple-input multiple-output (MIMO) channels in the noncoherent setting where transmitter and receiver have no *a priori* knowledge of the realizations of the fading channel. We prove that *unitary space-time modulation* (USTM) is not capacity-achieving in the high signal-to-noise ratio (SNR) regime when the total number of antennas exceeds the coherence time of the fading channel (expressed in multiples of the symbol duration), a situation that is relevant for MIMO systems with large antenna arrays (large-MIMO systems). This 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 [1], [2]. Specifically, when a genie provides the receiver with perfect channel state information (the so called *coherent setting*), the capacity of a multiple-input multiple-output (MIMO) fading channel grows linearly in the minimum between the number of transmit and receive antennas [2]. In practice, however, the fading channel is not known *a priori* at the receiver and must be estimated, for example through the transmission of pilot symbols. 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 [3]–[6], 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. We emphasize that in the noncoherent setting the receiver is allowed to try and gain channel knowledge. Channel estimation is simply viewed as a specific form of coding [7].

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 $T > 1$ 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 MIMO channel has been studied extensively in the literature [3], [8], [4], [9], no closed-form capacity expression is available to date. Zheng and Tse [4] showed that capacity behaves in the high signal-to-noise ratio (SNR) regime as[1]

$$C(\rho) = M^* \left(1 - \frac{M^*}{T}\right) \log(\rho) + \mathcal{O}(1), \quad \rho \to \infty. \quad (1)$$

Here, ρ denotes the SNR, $M^* \triangleq \min\{M, N, \lfloor T/2 \rfloor\}$ with M and N standing for the number of transmit and receive antennas, respectively, and $\mathcal{O}(1)$ indicates a bounded function of ρ (for sufficiently large ρ). The high-SNR capacity expression given in (1) is insightful as it allows one to determine the capacity loss (at high SNR) due to lack of *a priori* channel knowledge. Recalling that in the coherent case

$$C_{\mathrm{coh}}(\rho) = \min\{M, N\} \log(\rho) + \mathcal{O}(1), \quad \rho \to \infty$$

one sees that this loss is pronounced when the channel's coherence time T is small. The capacity expression (1) also implies that, for a given coherence time T and number of receive antennas N, the capacity *pre-log* (i.e., the asymptotic ratio between the capacity in (1) and $\log(\rho)$ as $\rho \to \infty$) is maximized by using $M = \min\{N, \lfloor T/2 \rfloor\}$ transmit antennas.[2]

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 [4, Sec. IV.B]:

$$C(\rho) = M^* \left(1 - \frac{M^*}{T}\right) \log(\rho) + c + o(1), \quad \rho \to \infty. \quad (2)$$

Here, c, which is given in [4, Eq. (24)], depends on T, M, and N but not on ρ, and $o(1) \to 0$ as $\rho \to \infty$. Differently from (1), the high-SNR expression (2) describes capacity accurately already at moderate SNR values [11], because it captures the first two terms in the asymptotic expansion of $C(\rho)$ for $\rho \to \infty$. The key element exploited in [4] to establish (2) is the optimality of *isotropically distributed unitary* input signals [3, Sec. A.2] at high SNR. The isotropic unitary input distribution is often referred to as *unitary space-time modulation* (USTM) [12], [9], [13]. Capacity-approaching coding schemes that are based on

---

[1]When $T = 1$, capacity grows double-logarithmically in ρ [10, Thm. 4.2].

[2]More generally, for fixed T and N, and for arbitrary SNR, the capacity for $M > T$ is equal to the capacity for $M = T$ [3, Thm. 1].

USTM and do not require the explicit estimation of the fading channel have been recently proposed in [13].

In this paper, we focus on the case $T < M + N$ (channel's coherence time smaller than the total number of antennas), which is of interest for point-to-point 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 co-operation among base stations and low complexity channel estimation algorithms [14]–[16].

Contributions: We prove that in the large-MIMO setting where $T < M + N$, USTM is not capacity-achieving at high SNR. The capacity-achieving input signal turns out to consist of 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 $T < M + N$. We show that using BSTM instead of USTM yields a rate gain of about 13% when SNR is 30 dB and $N \gg T$. Note that our result holds for all T, M, and N values satisfying $1 < T < M + N$. In other words, differently from most of the literature on large-MIMO systems, our analysis is not asymptotic in the number of antennas.

Our proof technique exploits the geometric structure in the MIMO block-fading channel input-output relation first observed in [4]. The set of tools used to establish our main result is, however, different from the one used in [4]. In particular, differently from [4], our proof is based on the duality approach [10], and on a novel closed-form characterization of the probability density function (pdf) of the MIMO block-fading channel output, which generalizes a previous result obtained in [9]. These two tools allow us to simplify the derivation of (2) for the case $T \geq M + N$ compared to the derivation provided in [4], and to generalize (2) to the large-MIMO setting $T < M + N$.

Notation: Uppercase boldface letters denote matrices and lowercase boldface letters designate vectors. Uppercase sans-serif letters (e.g., Q) denote probability distributions, while lowercase sans-serif letters (e.g., r) are reserved for pdfs. The superscripts $^\mathsf{T}$ and $^\mathsf{H}$ stand for transposition and Hermitian transposition, respectively. We denote the identity matrix of dimension $M \times M$ by $\mathbf{I}_M$; $\mathrm{diag}\{\mathbf{a}\}$ is the diagonal square matrix whose main diagonal contains the entries of the vector $\mathbf{a}$, and $\lambda_q\{\mathbf{A}\}$ stands for the qth largest eigenvalue of the Hermitian positive-semidefinite matrix $\mathbf{A}$. For a random matrix $\mathbf{X}$ with probability distribution $\mathsf{Q}_\mathbf{X}$, we write $\mathbf{X} \sim \mathsf{Q}_\mathbf{X}$. We denote expectation by $\mathbb{E}[\cdot]$, and use the notation $\mathbb{E}_\mathbf{X}[\cdot]$ or $\mathbb{E}_{\mathsf{Q}_\mathbf{X}}[\cdot]$ to stress that expectation is taken with respect to $\mathbf{X} \sim \mathsf{Q}_\mathbf{X}$. We write $D(\mathsf{Q}_\mathbf{Y}(\cdot) \,\|\, \mathsf{R}_\mathbf{Y}(\cdot))$ for the relative entropy between the probability distributions $\mathsf{Q}_\mathbf{Y}$ and $\mathsf{R}_\mathbf{Y}$. Furthermore, $\mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$ stands for the distribution of a circularly-symmetric complex Gaussian random vector with covariance matrix $\mathbf{\Sigma}$. For two functions $f(x)$ and $g(x)$, the notation $f(x) = \mathcal{O}(g(x))$, $x \to \infty$, means that $\limsup_{x\to\infty} |f(x)/g(x)| < \infty$, and $f(x) = o(g(x))$, $x \to \infty$, means that $\lim_{x\to\infty} |f(x)/g(x)| = 0$. Finally, $\log(\cdot)$ indicates the natural logarithm, $\Gamma(\cdot)$ denotes the Gamma

Parameter	Definition
$\overline{L}$	$\max\{N, T - M\}$
$\underline{L}$	$\min\{N, T - M\}$
$\overline{P}$	$\max\{N, T\}$
$\underline{P}$	$\min\{N, T\}$

TABLE I
FOUR PARAMETERS RELATED TO THE CHANNEL'S COHERENCE TIME T, THE NUMBER OF TRANSMIT ANTENNAS M, AND THE NUMBER OF RECEIVE ANTENNAS N.

function [17, Eq. (6.1.1)], and $\Gamma_m(a)$ designates the *complex multivariate Gamma function* [18, Eq. (44)]

$$\Gamma_m(a) = \pi^{m(m-1)/2} \prod_{k=1}^{m} \Gamma(a - k + 1). \tag{3}$$

II. SYSTEM MODEL AND KNOWN RESULTS

A. System Model

We consider a point-to-point Rayleigh block-fading MIMO channel with M transmit antennas, N receive antennas, and channel's coherence time $T > 1$. The channel input-output relation within a coherence interval can be compactly written in matrix notation as follows [4], [8], [9]:

$$\mathbf{Y} = \sqrt{\rho/M} \cdot \mathbf{XH} + \mathbf{W}. \tag{4}$$

Here, $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_M] \in \mathbb{C}^{T \times M}$ 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 (4) is given by

$$C(\rho) = \frac{1}{T} \sup_{\mathsf{Q}_\mathbf{X}} I(\mathbf{X}; \mathbf{Y}). \tag{5}$$

Here, $I(\mathbf{X}; \mathbf{Y})$ denotes the mutual information [19, Sec. 8.5] between the input matrix $\mathbf{X}$ and the output matrix $\mathbf{Y}$, and the supremum is over all probability distributions $\mathsf{Q}_\mathbf{X}$ on $\mathbf{X}$ that satisfy the average-power constraint

$$\mathbb{E}\big[\mathrm{tr}\{\mathbf{XX}^\mathsf{H}\}\big] \leq TM. \tag{6}$$

Since the variance of the entries of $\mathbf{H}$ and $\mathbf{W}$ is normalized to one, ρ in (4) can be interpreted as the SNR at each receive antenna.

Throughout the paper, we will often make use of four parameters $(\underline{L}, \overline{L}, \underline{P}, \overline{P})$ related to the coherence time T, the number of transmit antennas M, and the number of receiver antennas N. These parameters are listed in Table I for future reference.

B. Properties of the Capacity-Achieving Input Distribution

Even if no closed-form expression is available to date for $C(\rho)$, the structure of the capacity-achieving input distribution is partially known. We next review two properties of the capacity-achieving input distribution that will reveal useful for our analysis.

Lemma 1 ([3, Thm. 2]): The capacity-achieving input matrix $\mathbf{X}$ is the product of a $T \times M$ isotropically distributed unitary matrix $\mathbf{\Phi}$ and an independent $M \times M$ nonnegative diagonal matrix $\mathbf{D} = \mathrm{diag}\{[d_1 \cdots d_M]^{\mathsf{T}}\}$.

For the case $T \geq M + N$, taking $\mathbf{D}$ deterministic with diagonal entries equal to $\sqrt{T}$ turns out to be optimal at high SNR. In this case, the resulting input matrix $\mathbf{X}$ is a scaled isotropically distributed unitary matrix. This input distribution, which is known as USTM [12], [9], [13], is the one used in [4] to establish (2).

When $T < M + N$, USTM is not optimal at high SNR, as we shall illustrate in Section III. Nevertheless, the optimal distribution of $\mathbf{X} = \mathbf{\Phi}\mathbf{D}$ shares the following property with USTM: the probability distribution induced on $\sqrt{\rho}d_m = \sqrt{\rho}\|\mathbf{x}_m\|$, $m = 1, \ldots, M$, by the capacity-achieving input distribution *escapes to infinity* [10, Def. 4.11] as $\rho \to \infty$. Namely, it allocates vanishing probability to every interval of the form $[0, \sqrt{\rho_0}]$ with $\rho_0 > 0$. This property is formalized in the following lemma:

Lemma 2: Fix an arbitrary $\rho_0 > 0$ and let

$$\mathcal{K}(\rho_0) \triangleq \Big\{ \mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_M] \in \mathbb{C}^{T \times M} :$$
$$\min_{m=1,\ldots,M}\{\rho\|\mathbf{a}_m\|^2\} < \rho_0 \Big\}. \quad (7)$$

Let $\{\mathsf{Q}_{\mathbf{X}}^{(\rho)}, \rho > 0\}$ be a family of input distributions (parametrized with respect to the SNR ρ) satisfying (6) and the following additional property

$$\lim_{\rho \to \infty} \frac{I(\mathbf{X}; \mathbf{Y})}{C(\rho)} = 1, \quad \mathbf{X} \sim \mathsf{Q}_{\mathbf{X}}^{(\rho)}.$$

Then, $\lim_{\rho \to \infty} \mathbb{P}\Big(\mathbf{X} \in \mathcal{K}(\rho_0), \mathbf{X} \sim \mathsf{Q}_{\mathbf{X}}^{(\rho)}\Big) = 0$.

Proof: The proof follows along the same lines of the proofs of [5, Thm. 8] and [4, Lem 8]. ∎

An important consequence of the escape-to-infinity property of the capacity-achieving input distribution is that the asymptotic behavior of $C(\rho)$ as $\rho \to \infty$ does not change if we constrain the probability distribution of $\sqrt{\rho}d_m$ ($m = 1, 2, \ldots, M$) to be supported outside the interval $[0, \sqrt{\rho_0}]$, $\rho_0 > 0$. More precisely, we have the following result.

Lemma 3: Fix an arbitrary $\rho_0 > 0$ and let $\mathcal{K}(\rho_0)$ as in (7). Denote by $C_{\mathcal{K}}(\rho)$ the capacity of the channel (4) when the input $\mathbf{X}$ is subject to the average-power constraint (6) and to the additional constraint that $\mathbf{X} \notin \mathcal{K}(\rho_0)$ with probability 1 (w.p.1). Then, $C(\rho) = C_{\mathcal{K}}(\rho) + o(1)$, $\rho \to \infty$.

Proof: The proof follows from [10, Thm. 4.12]. ∎

III. CAPACITY IN THE HIGH-SNR REGIME

A. Asymptotic Characterization of Capacity

The main result of this paper is Theorem 4 below, which provides a high-SNR characterization of $C(\rho)$ that generalizes (2),

in that it holds also in the large-MIMO setting $T < M + N$.[3]

Theorem 4: The capacity $C(\rho)$ of the Rayleigh block-fading MIMO channel (4) with N receive antennas, coherence time T, and $M \leq \min\{N, \lfloor T/2 \rfloor\}$ transmit antennas is given by

$$C(\rho) = M\left(1 - \frac{M}{T}\right)\log(\rho) + c + o(1), \quad \rho \to \infty \quad (8)$$

where

$$c \triangleq \frac{1}{T}\log\left(\frac{\Gamma_M(M)\Gamma_M(\underline{L})}{\Gamma_M(N)\Gamma_M(T)}\right) + M\left(1 - \frac{M}{T}\right)\log\left(\frac{T}{M}\right)$$
$$+ \frac{M\underline{L}}{T}\log\left(\frac{N}{\underline{L}}\right) + \frac{\overline{L}}{T}\Big(\mathbb{E}\big[\log\det(\mathbf{H}\mathbf{H}^{\mathsf{H}})\big] - M\Big). \quad (9)$$

Here, $\underline{L}$ and $\overline{L}$ are defined in Table I, and

$$\mathbb{E}\big[\log\det(\mathbf{H}\mathbf{H}^{\mathsf{H}})\big] = \sum_{i=1}^{M}\psi(N - i + 1)$$
$$= -M\gamma + \sum_{i=1}^{M}\sum_{k=1}^{N-i}\frac{1}{k} \quad (10)$$

where $\psi(\cdot)$ denotes Euler's digamma function [17, Eq. (6.3.1)] and $\gamma \approx 0.577$ is Euler's constant.

Proof: See Section IV. A sketch of the proof for the single-input multiple-output case, which is simpler to analyze than the MIMO case, is given in [20]. ∎

In Section III-B below we compare $C(\rho)$ in (8) with the capacity lower bound obtained using USTM. The input distribution that achieves (8) is described in Section III-C. Numerical results illustrating the lack of tightness of the USTM-based capacity lower bound in the large-MIMO setting are provided in Section III-D.

B. Rates Achievable with USTM

For the case $T \geq M + N$, the high-SNR capacity expression (8) coincides with the one reported in [4, Sec. IV.B].[4] In this case, USTM, i.e., $\mathbf{X} = \sqrt{T}\mathbf{\Phi}$, with $\mathbf{\Phi}$ unitary and isotropically distributed, achieves (8). When $T < M + N$, the novel capacity characterization provided in Theorem 4 implies that USTM is not capacity-achieving at high SNR, as formalized in the following corollary.

Corollary 5: The rate achievable using USTM over the Rayleigh block-fading MIMO channel (4) with N receive antennas, coherence time T, and $M \leq \min\{N, \lfloor T/2 \rfloor\}$ transmit antennas is

$$C_{\mathrm{USTM}}(\rho) = M\left(1 - \frac{M}{T}\right)\log(\rho) + c_{\mathrm{USTM}} + o(1), \quad \rho \to \infty$$
$$(11)$$

---

[3]Because of the constraint $M \leq \min\{N, \lfloor T/2 \rfloor\}$, *large-MIMO setting* in this paper indicates a point-to-point MIMO uplink with a large antenna array at the receiver.

[4]The expression for c given in [4, 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 [4, Eq. (24)] with the result given in [4, Thm. 9] for the case $M = N$.

where

$$c_{\text{USTM}} \triangleq \frac{1}{T} \log\left(\frac{\Gamma_M(M)}{\Gamma_M(T)}\right) + M\left(1 - \frac{M}{T}\right) \log\left(\frac{T}{eM}\right)$$
$$+ \left(1 - \frac{M}{T}\right) \mathbb{E}\left[\log\det(\mathbf{HH}^{\mathsf{H}})\right].$$

Note that $c_{\text{USTM}} = c$ when $T \geq M + N$; however, $c_{\text{USTM}} < c$ when $T < M + N$.

Proof: The proof follows by repeating the same steps as in Section IV-B after having replaced the capacity-achieving input distribution (to be described in Section III-C) with USTM. ∎

C. The Capacity-Achieving Input Distribution at High SNR

1) Matrix-variate distributions: To describe the input probability distribution that achieves (8), we shall need the following preliminary results from multivariate statistics.

Definition 6: An $m \times m$ random matrix $\mathbf{A}$ is said to have a *complex Wishart distribution* with $n > 0$ degrees of freedom and covariance matrix $\boldsymbol{\Sigma}$ if $\mathbf{A} = \mathbf{BB}^{\mathsf{H}}$, where the columns of the $m \times n$ matrix $\mathbf{B}$ are independent and $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$-distributed. In this case, we shall write $\mathbf{A} \sim \mathcal{W}_m(n, \boldsymbol{\Sigma})$.

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

Definition 7: 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 $\mathbf{C} = \left(\mathbf{T}^{\mathsf{H}}\right)^{-1} \mathbf{AT}^{-1}$, where $\mathbf{A} \sim \mathcal{W}_m(p, \boldsymbol{\Sigma})$ and $\mathbf{B} \sim \mathcal{W}_m(n, \boldsymbol{\Sigma})$ are independent, and $\mathbf{A} + \mathbf{B} = \mathbf{T}^{\mathsf{H}}\mathbf{T}$, with $\mathbf{T}$ upper-triangular with positive diagonal elements [21, p. 406]. In this case, we shall write $\mathbf{C} \sim \text{Beta}_m(p, n)$.

For the case when $n < m$ or $p < m$, the probability distribution of $\mathbf{C}$ is usually referred to as singular complex matrix-variate Beta distribution because it involves singular Wishart distributions. In the next lemma, we state two properties of the complex matrix-variate Beta distribution that will be used in the proof of Theorem 4.

Lemma 8: Let $\mathbf{C} \sim \text{Beta}_m(p, n)$ with $p \geq m > 0$ and $n > 0$. The following properties hold:

1) $\mathbf{C}$ is unitarily invariant [22, Def. 2.6], i.e., $\mathbf{C} \sim \mathbf{UCU}^{\mathsf{H}}$ for every $m \times m$ unitary matrix $\mathbf{U}$ independent of $\mathbf{C}$.

2) The joint pdf of the ordered eigenvalues $\lambda_1 \geq \cdots \geq \lambda_m$ of $\mathbf{C}$ takes on two different forms according to the value of n. If $n \geq m$, then $1 > \lambda_1 > \cdots > \lambda_m > 0$ w.p.1, and the joint pdf of $\lambda_1, \ldots, \lambda_m$ is given by

$$f_{\lambda_1, \ldots, \lambda_m}(a_1, \ldots, a_m)$$
$$= \frac{\pi^{m(m-1)}}{\Gamma_m(m)} \cdot \frac{\Gamma_m(p+n)}{\Gamma_m(p)\Gamma_m(n)}$$
$$\cdot \prod_{i=1}^{m} a_i^{p-m}(1-a_i)^{n-m} \cdot \prod_{i<j}(a_i - a_j)^2. \quad (12)$$

If $0 < n < m$, then $\lambda_1 = \cdots = \lambda_{m-n} = 1$ w.p.1, and $1 > \lambda_{m-n+1} > \cdots > \lambda_m > 0$ w.p.1. Moreover, the joint pdf of $\lambda_{m-n+1}, \ldots, \lambda_m$ is given by

$$f_{\lambda_{m-n+1}, \ldots, \lambda_m}(a_{m-n+1}, \ldots, a_m)$$
$$= \frac{\pi^{n(n-1)}}{\Gamma_n(n)} \cdot \frac{\Gamma_n(p+n)}{\Gamma_n(m)\Gamma_n(p+n-m)}$$
$$\cdot \prod_{i=m-n+1}^{m} a_i^{p-m}(1-a_i)^{m-n} \cdot \prod_{m-n<i<j}^{m}(a_i - a_j)^2. \quad (13)$$

Proof: Part 1 and (12) in part 2 follow by extending to the complex case [23, Lem. 3.11] and [24, Thm. 3.3.4], respectively; to prove (13) it is sufficient to note that $\widetilde{\mathbf{C}} = (\mathbf{I}_m - \mathbf{C}) \sim \text{Beta}_m(n, p)$ (see [24, Def. 3.3.2]), that $\widetilde{\mathbf{C}}$ has rank $n < m$, and that its n nonzero eigenvalues are distributed as the eigenvalues of a $\text{Beta}_n(m, p + n - m)$-distributed random matrix. ∎

We shall also need the following result relating Wishart-distributed and Beta-distributed matrices.

Lemma 9: Let $\mathbf{S} \sim \mathcal{W}_m(p + n, \boldsymbol{\Sigma})$ 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 $\mathbf{S} = \mathbf{T}^{\mathsf{H}}\mathbf{T}$, where $\mathbf{T}$ is upper-triangular with positive diagonal elements. Then, $\mathbf{A} = \mathbf{T}^{\mathsf{H}}\mathbf{CT} \sim \mathcal{W}_m(p, \boldsymbol{\Sigma})$.

Proof: The lemma follows from a generalization to the complex case of [24, Thm. 3.3.1] for the nonsingular case $n \geq m$, and of [25, Thm. 1] for the singular case $0 < n < m$. ∎

Note that Lemma 8 (part 1) implies that the eigenvalues of $\mathbf{A}$ and of $\mathbf{CS}$ in Lemma 9 have the same distribution.

2) The Optimal Input Distribution: We are now ready to describe the input distribution that achieves (8). This distribution takes on two different forms according to the relation between T, M and N. Specifically, one should take $\mathbf{X} = \boldsymbol{\Phi}\mathbf{D}$ where $\boldsymbol{\Phi}$ is unitary and isotropically distributed, and $\mathbf{D} = \sqrt{TN/\underline{L}} \cdot \widetilde{\mathbf{D}}$ with $\underline{L}$ defined in Table I, and with $\widetilde{\mathbf{D}}$ being a diagonal matrix whose *ordered* positive entries $\{\tilde{d}_1, \ldots, \tilde{d}_M\}$ are distributed as follows:

a) Case $T < M + N$: The squared nonzero entries $\{\tilde{d}_1^2, \ldots, \tilde{d}_M^2\}$ of $\widetilde{\mathbf{D}}$ have the same joint pdf as the ordered eigenvalues of a positive-definite $M \times M$ random matrix $\mathbf{Z} \sim \text{Beta}_M(T - M, M + N - T)$. The resulting pdf of $\{\tilde{d}_1^2, \ldots, \tilde{d}_M^2\}$ is obtained by setting $p = T - M$ and $n = M + N - T$ in (12) if $T \leq N$, and in (13) if $N < T < M + N$.

b) Case $T \geq M + N$: The nonzero entries $\{\tilde{d}_1, \ldots, \tilde{d}_M\}$ of $\widetilde{\mathbf{D}}$ should be taken so that $\tilde{d}_1 = \cdots = \tilde{d}_M = 1$ w.p.1. This results in the USTM distribution used in [4].

In the remainder of the paper, we shall denote by $\mathsf{Q}_{\mathbf{D}}^{\text{opt}}$ the probability distribution of $\mathbf{D} = \sqrt{TN/\underline{L}} \cdot \widetilde{\mathbf{D}}$ we have just introduced. Furthermore, we shall refer to the probability distribution of $\mathbf{X} = \boldsymbol{\Phi}\mathbf{D}$ resulting by choosing $\boldsymbol{\Phi}$ unitary and isotropically distributed and $\mathbf{D} \sim \mathsf{Q}_{\mathbf{D}}^{\text{opt}}$ as BSTM. Note that BSTM reduces to USTM when $T \geq M + N$.

As shown in [4, p. 369], USTM is optimal for the case $T \geq M + N$ because it maximizes

$$h(\mathbf{UDH}) + (T - M - N)\,\mathbb{E}\left[\log\det\left(\mathbf{D}^2\right)\right] \quad (14)$$

where $\mathbf{U} \in \mathbb{C}^{M \times M}$ is an isotropically distributed unitary matrix independent of both $\mathbf{D}$ and $\mathbf{H}$, and $h(\cdot)$ denotes the differential entropy. In fact, the average-power constraint (6) implies that

$$h(\mathbf{UDH}) \leq MN \log(\pi e T); \quad \mathbb{E}\big[\log \det(\mathbf{D}^2)\big] \leq M \log(T)$$

and under USTM, which yields $\mathbf{D} = \sqrt{T} \cdot \mathbf{I}_M$, both inequalities hold with equality.

In the large-MIMO setting $T < M + N$, however, the second term in (14) turns negative and USTM does no longer maximize (14). As we shall now illustrate, the maximizing distribution of $\mathbf{D}$ turns out to be $\mathrm{Q}_\mathbf{D}^{\mathrm{opt}}$, which results in BSTM. Through algebraic manipulations similar to the ones leading to (46) and (47) in Section IV, it is possible to show that

$$h(\mathbf{UDH}) + (T - M - N)\,\mathbb{E}\big[\log\det(\mathbf{D}^2)\big] = h(\mathbf{G}) + k.$$

Here, k is a constant that does not dependent on $\mathbf{D}$, and $\mathbf{G} \in \mathbb{C}^{M \times (T-M)}$ is a random matrix with singular values jointly distributed as the singular values of $\mathbf{DH}$, and with isotropically distributed singular vectors. Lemma 10 below implies that the choice $\mathbf{D} \sim \mathrm{Q}_\mathbf{D}^{\mathrm{opt}}$ induces a matrix $\mathbf{G}$ that is Gaussian with i.i.d. $\mathcal{CN}(0, TN/(T-M))$ entries. But a Gaussian $\mathbf{G}$ with i.i.d. entries maximizes $h(\mathbf{G})$, and, hence, (14).

Lemma 10: Let $\mathbf{D} \sim \mathrm{Q}_\mathbf{D}^{\mathrm{opt}}$ and let $\mathbf{H}$ be an independent $M \times N$ random matrix with i.i.d. $\mathcal{CN}(0,1)$ entries. The singular values of $\mathbf{DH}$ are distributed as the singular values of an $M \times \underline{L}$ matrix $\mathbf{G}$ with i.i.d. $\mathcal{CN}(0, TN/\underline{L})$ entries.

Proof: For the case $T \geq M + N$, we have that $\underline{L} = N$ and, hence, $\mathbf{D} = \sqrt{T} \cdot \mathbf{I}_M$. Consequently, $\mathbf{DH} = (\sqrt{T} \cdot \mathbf{H}) \sim \mathbf{G}$, from which the statement in the lemma follows.

For the case $T < M + N$ (and, hence, $\underline{L} = T - M$) we shall proceed as follows. Let $\mathbf{D} = \sqrt{TN/(T-M)} \cdot \widetilde{\mathbf{D}}$, and let $\mathbf{U}$ be an $M \times M$ unitary and isotropically distributed random matrix independent of $\widetilde{\mathbf{D}}$ and $\mathbf{H}$. Since $\mathbf{HH}^\mathsf{H}$ is unitary invariant, we have that $\mathbf{HH}^\mathsf{H} \sim \mathbf{U}^\mathsf{H}\mathbf{HH}^\mathsf{H}\mathbf{U}$, and hence $\widetilde{\mathbf{D}}\mathbf{HH}^\mathsf{H}\widetilde{\mathbf{D}} \sim \widetilde{\mathbf{D}}\mathbf{U}^\mathsf{H}\mathbf{HH}^\mathsf{H}\mathbf{U}\widetilde{\mathbf{D}}$. Now note that $\widetilde{\mathbf{D}}\mathbf{U}^\mathsf{H}\mathbf{HH}^\mathsf{H}\mathbf{U}\widetilde{\mathbf{D}}$ and $\mathbf{U}\widetilde{\mathbf{D}}^2\mathbf{U}^\mathsf{H}\mathbf{HH}^\mathsf{H}$ have the same eigenvalues; furthermore, $\mathbf{U}\widetilde{\mathbf{D}}^2\mathbf{U}^\mathsf{H} \sim \mathrm{Beta}_M(T - M, M + N - T)$, which follows from Lemma 8 (part 1), and from [22, Lem. 2.6]; finally, $\mathbf{HH}^\mathsf{H} \sim \mathcal{W}_M(N, \mathbf{I}_M)$. Hence, by Lemmas 8 and 9 the eigenvalues of $\mathbf{U}\widetilde{\mathbf{D}}^2\mathbf{U}^\mathsf{H}\mathbf{HH}^\mathsf{H}$—and consequently also the eigenvalues of $\widetilde{\mathbf{D}}\mathbf{HH}^\mathsf{H}\widetilde{\mathbf{D}}$—have the same distribution as the eigenvalues of a $\mathcal{W}_M(T - M, \mathbf{I}_M)$-distributed random matrix. ∎

D. Gain of BSTM over USTM

The use of USTM is motivated by several practical considerations [8], [9], [13]. 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.

Asymptotic Analysis: In Corollary 11 below we show that the rate gain resulting from the use of BSTM instead of USTM grows logarithmically in the number of receive antennas.

Corollary 11: Let T and $M \leq \lfloor T/2 \rfloor$ be fixed. Then

$$\lim_{N \to \infty} \lim_{\rho \to \infty} \left(C(\rho) - C_{\mathrm{USTM}}(\rho) - \frac{M^2}{2T} \log(N) \right) = c_{M,T} \quad (15)$$

where $C(\rho)$ and $C_{\mathrm{USTM}}(\rho)$ are given in (8) and (11), respectively, and

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

Proof: As we are interested in the limit $N \to \infty$, we shall assume without loss of generality that $\underline{L} = T - M$ and $\overline{L} = N$. Since the first term in the high-SNR expansion of $C(\rho)$ and $C_{\mathrm{USTM}}(\rho)$ is the same,

$$\lim_{\rho \to \infty} \big(C(\rho) - C_{\mathrm{USTM}}(\rho) \big) = c - c_{\mathrm{USTM}} = c_0 + c_N$$

where c_0 and c_N are defined as follows:

$$T \cdot c_0 = \log\big(\Gamma_M(T - M)\big) + M(T - M)\log\left(\frac{e}{T - M}\right)$$
$$T \cdot c_N = (N - T + M)\,\mathbb{E}\big[\log\det(\mathbf{HH}^\mathsf{H})\big] - \log\big(\Gamma_M(N)\big)$$
$$- MN + M(T - M)\log(N). \quad (16)$$

Note that c_N is a function of N, while c_0 is not. Consequently, to establish (15) it is sufficient to study the limit $N \to \infty$ of the first two terms on the right-hand side (RHS) of (16). For the first term, we use (10) and the following asymptotic expansion of the Euler's digamma function [17, Eq. (6.3.18)]: $\psi(m) = \log(m) - 1/(2m) + o(1/m)$, $m \to \infty$, which yields

$$(N - T + M)\,\mathbb{E}\big[\log\det(\mathbf{HH}^\mathsf{H})\big]$$
$$= (N - T + M)\sum_{i=1}^{M} \psi(N - i + 1)$$
$$= -M(T - M)\log(N) + N \sum_{i=1}^{M} \log(N - i + 1)$$
$$- \frac{M}{2} + o(1), \quad N \to \infty. \quad (17)$$

For the second term on the RHS of (16) we proceed as follows:

$$\log\big(\Gamma_M(N)\big)$$
$$= \frac{M(M - 1)}{2} \log(\pi) + \sum_{i=1}^{M} \log\big((N - i)!\big)$$
$$\overset{(a)}{=} \sum_{i=1}^{M} \left((N - i)\log(N - i) + \frac{\log(N - i)}{2} + i \right) - MN$$
$$+ \frac{M}{2}\log(2) + \frac{M^2}{2}\log(\pi) + o(1), \quad N \to \infty$$
$$= N\left[\sum_{i=1}^{M} \log(N - i)\right] + \frac{M^2}{2}\log\left(\frac{\pi e}{N}\right) + \frac{M}{2}\log(2e)$$
$$- MN + o(1), \quad N \to \infty. \quad (18)$$

Here, (a) follows from Stirling's formula $n! = n^n e^{-n}\sqrt{2\pi n}\,(1 + o(1))$, $n \to \infty$. We complete the proof by substituting (17) and (18) into (16), and using that

$$\lim_{N \to \infty} N \log\left(\frac{N - i + 1}{N - i}\right) = 1. \qquad ∎$$
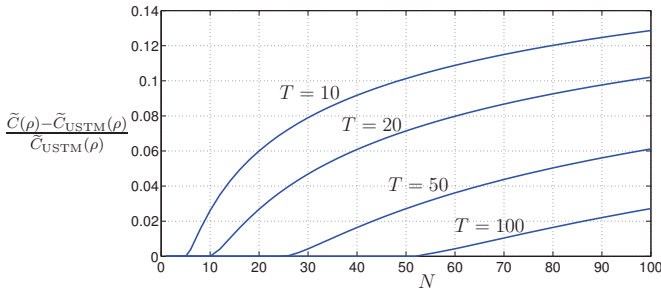
Fig. 1. Rate gain resulting from the use of BSTM instead of USTM as a function of the number of receive antennas N and the channel's coherence time T; in the figure, $\rho = 30$ dB, and $M = \min\{\lfloor T/2 \rfloor, N\}$.

Numerical Results: Let $\widetilde{C}(\rho)$ be the high-SNR approximation of $C(\rho)$ obtained by neglecting the $o(1)$ term in (8). Similarly, let $\widetilde{C}_{\text{USTM}}(\rho)$ be the high-SNR approximation of $C_{\text{USTM}}(\rho)$ obtained by neglecting the $o(1)$ term in (11). As can be inferred from the results reported in [4], [9], [11], $\widetilde{C}_{\text{USTM}}(\rho)$ is a good approximation for $C_{\text{USTM}}(\rho)$ at $\rho \gtrsim 30$ dB. Numerical evidence suggests that the same holds for the pair $\widetilde{C}(\rho)$ 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 $[\widetilde{C}(\rho) - \widetilde{C}_{\text{USTM}}(\rho)]/\widetilde{C}_{\text{USTM}}(\rho)$ 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%.

IV. PROOF OF THEOREM 4

The proof is effected by exhibiting capacity upper and lower bounds that agree up to a $o(1)$ term.

A. Upper Bound

Fix $\rho_0 > 0$ and let $\mathcal{K}(\rho_0)$ as in (7); as a consequence of Lemma 3, we can restrict—without loss of generality—the supremum in (5) to input distributions $\mathsf{Q_X}$ satisfying the constraint $\mathbf{X} \notin \mathcal{K}(\rho_0)$ w.p.1. Our capacity upper bound is based on *duality* [10], [26], which is a technique that allows one to obtain tight upper bounds on $I(\mathbf{X}; \mathbf{Y})$ by carefully choosing a probability distribution of $\mathbf{Y}$. Specifically, let $\mathsf{P_{Y|X}}$ denote the conditional probability distribution of $\mathbf{Y}$ given $\mathbf{X}$, and $\mathsf{Q_Y}$ denote the distribution induced on $\mathbf{Y}$ by $\mathsf{Q_X}$ through (4). Finally, let $\mathsf{R_Y}$ be an arbitrary distribution of $\mathbf{Y}$ with pdf $\mathsf{r_Y}$. We use duality to upper-bound $I(\mathbf{X}; \mathbf{Y})$ in (5) as follows [10, Thm. 5.1]:

$$
\begin{aligned}
I(\mathbf{X}; \mathbf{Y}) &= \mathbb{E}_{\mathbf{X}}\big[D(\mathsf{P_{Y|X}} \,\|\, \mathsf{Q_Y})\big] \\
&\overset{(a)}{=} \mathbb{E}_{\mathbf{X}}\big[D(\mathsf{P_{Y|X}} \,\|\, \mathsf{R_Y})\big] - D(\mathsf{Q_Y} \,\|\, \mathsf{R_Y}) \\
&\overset{(b)}{\leq} \mathbb{E}_{\mathbf{X}}\big[D(\mathsf{P_{Y|X}} \,\|\, \mathsf{R_Y})\big] \\
&= -\mathbb{E}_{\mathsf{Q_Y}}\big[\log(\mathsf{r_Y}(\mathbf{Y}))\big] - h(\mathbf{Y} \,|\, \mathbf{X}). \quad (19)
\end{aligned}
$$

Here, (a) follows from Topsøe's identity [27], and (b) is a consequence of the nonnegativity of relative entropy [19, Thm. 2.6.3]. The conditional differential entropy $h(\mathbf{Y} \mid \mathbf{X})$ in (19) is given by

$$
h(\mathbf{Y} \mid \mathbf{X}) = N \sum_{i=1}^{M} \mathbb{E}\left[\log\left(1 + \frac{\rho \|\mathbf{x}_i\|^2}{M}\right)\right] + NT \log(\pi e). \quad (20)
$$

To evaluate the first term on the RHS of (19), we need to choose a specific output pdf $\mathsf{r_Y}$. Let us express $\mathbf{Y}$ in terms of its singular value decomposition (SVD)

$$
\mathbf{Y} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}} \quad (21)
$$

where $\mathbf{U} \in \mathbb{C}^{T \times \underline{P}}$ and $\mathbf{V} \in \mathbb{C}^{N \times \underline{P}}$ ($\underline{P}$ is defined in Table I) are (truncated) unitary matrices, and $\boldsymbol{\Sigma} = \text{diag}\{[\sigma_1(\mathbf{Y}) \cdots \sigma_{\underline{P}}(\mathbf{Y})]^{\mathsf{T}}\}$ contains the singular values of $\mathbf{Y}$ arranged in decreasing order. To make the SVD unique, we shall assume that the diagonal entries of $\mathbf{U}$ are real and nonnegative. Hence, $\mathbf{V}$ is an element of the complex *Stiefel manifold* $\mathcal{S}(N, \underline{P})$ [18], [4], while $\mathbf{U}$ belongs to a *submanifold* $\widetilde{\mathcal{S}}(T, \underline{P})$ of $\mathcal{S}(T, \underline{P})$. We put forward the following result about the volume of $\mathcal{S}(n, m)$ and $\widetilde{\mathcal{S}}(n, m)$ for the case $n \geq m$ (see [18, Sec. V])

$$
|\mathcal{S}(n, m)| = \frac{2^m \pi^{mn}}{\Gamma_m(n)}; \quad \left|\widetilde{\mathcal{S}}(n, m)\right| = \frac{|\mathcal{S}(n, m)|}{(2\pi)^m} = \frac{\pi^{m(n-1)}}{\Gamma_m(n)}.
$$

When $\mathsf{Q_X}$ is capacity-achieving, Lemma 1 and the Gaussianity of $\mathbf{H}$ and $\mathbf{W}$, imply that $\mathbf{U}$ and $\mathbf{V}$ are uniformly distributed on $\widetilde{\mathcal{S}}(T, \underline{P})$ and $\mathcal{S}(N, \underline{P})$, respectively, and independent of each other and of $\boldsymbol{\Sigma}$. We shall take an output pdf for which this property holds. Furthermore, we take the first M singular values of $\mathbf{Y}$ distributed as the ordered singular values of the noiseless channel output matrix $\sqrt{\rho/M} \cdot \mathbf{XH} = \sqrt{\rho/M} \cdot \boldsymbol{\Phi}\mathbf{DH}$, with $\boldsymbol{\Phi}$ unitary and isotropically distributed, and $\mathbf{D} \sim \mathsf{Q_D^{opt}}$. By Lemma 10, this implies that the first M singular values of $\mathbf{Y}$ are distributed as the singular values of an $M \times \underline{L}$ matrix with i.i.d. $\mathcal{CN}(0, \lambda)$ entries, where $\lambda \triangleq NT\rho/(M\underline{L})$. We take the remaining $\underline{P} - M$ singular values distributed as the singular values of an independent $(N - M) \times (T - M)$ matrix with i.i.d. $\mathcal{CN}(0, 1)$ entries. The intuition behind this choice is the following: in the absence of the additive noise $\mathbf{W}$ in (4), the output matrix $\mathbf{Y}$ has rank M; this suggests that, in the high-SNR regime, the smallest $\underline{P} - M$ singular values of $\mathbf{Y}$ carry information about $\mathbf{W}$ only. Summarizing, we take the pdf $\mathsf{r}_{\sigma_1, \ldots, \sigma_{\underline{P}}}$ of the ordered singular values of $\mathbf{Y}$ as follows[5]

$$
\begin{aligned}
\mathsf{r}_{\sigma_1, \ldots, \sigma_{\underline{P}}}(a_1, \cdots, a_{\underline{P}}) &= \mathsf{r}_{\sigma_1, \cdots, \sigma_M}(a_1, \cdots, a_M) \\
&\quad \cdot \mathsf{r}_{\sigma_{M+1}, \cdots, \sigma_{\underline{P}}}(a_{M+1}, \cdots, a_{\underline{P}})
\end{aligned}
$$

where

$$
\begin{aligned}
&\mathsf{r}_{\sigma_1, \cdots, \sigma_M}(a_1, \cdots, a_M) \\
&= \frac{k_1 e^{-\sum_{i=1}^{M} a_i^2/\lambda}}{\lambda^{M\underline{L}}} \prod_{i=1}^{M} a_i^{2(\underline{L}-M)+1} \cdot \prod_{i<j}^{M} \left(a_i^2 - a_j^2\right)^2, \\
&\hspace{6cm} a_1 > \cdots > a_M \quad (22)
\end{aligned}
$$

[5] We shall indicate $\sigma_i(\mathbf{Y})$ simply as σ_i whenever no ambiguity occurs.

with $k_1 \triangleq 2^M \pi^{M(M-1)} / \left(\Gamma_M(\underline{L}) \Gamma_M(M) \right)$ and

$$
\begin{aligned}
& \mathsf{r}_{\sigma_{M+1}, \cdots, \sigma_{\underline{P}}}(a_{M+1}, \cdots, a_{\underline{P}}) \\
& = k_2 e^{-\sum_{i=M+1}^{\underline{P}} a_i^2} \prod_{i=M+1}^{\underline{P}} a_i^{2(\overline{P}-\underline{P})+1} \cdot \prod_{M<i<j}^{\underline{P}} \left(a_i^2 - a_j^2 \right)^2, \\
& \hspace{6cm} a_{M+1} > \cdots > a_{\underline{P}} \quad (23)
\end{aligned}
$$

with $\overline{P}$ defined in Table I, and

$$
k_2 \triangleq \frac{2^{\underline{P}-M} \pi^{(\underline{P}-M)(\underline{P}-M-1)}}{\Gamma_{\underline{P}-M}(\overline{P}-M) \Gamma_{\underline{P}-M}(\underline{P}-M)}.
$$

Here, both (22) and (23) follow from [22, Thm. 2.17] and the change of variable theorem. We are now ready to evaluate the first term on the RHS of (19). Let

$$
J_{\overline{P},\underline{P}}(\sigma_1, \cdots, \sigma_{\underline{P}}) = \prod_{i=1}^{\underline{P}} \sigma_i^{2(\overline{P}-\underline{P})+1} \cdot \prod_{i<j}^{\underline{P}} \left(\sigma_i^2 - \sigma_j^2 \right)^2 \quad (24)
$$

be the Jacobian of the SVD transformation [4, App. A]. The change of variables theorem yields

$$
\begin{aligned}
- \mathbb{E}[\log(\mathsf{r}_{\mathbf{Y}}(\mathbf{Y}))] &= - \mathbb{E}[\log(\mathsf{r}_{\mathbf{U},\mathbf{\Sigma},\mathbf{V}}(\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}))] \\
& \quad + \mathbb{E}\left[\log\left(J_{\overline{P},\underline{P}}(\sigma_1, \cdots, \sigma_{\underline{P}}) \right)\right] \\
&= - \mathbb{E}[\log(\mathsf{r}_{\mathbf{U}}(\mathbf{U}))] - \mathbb{E}[\log(\mathsf{r}_{\mathbf{V}}(\mathbf{V}))] \\
& \quad - \mathbb{E}\left[\log\left(\mathsf{r}_{\sigma_{M+1},...,\sigma_{\underline{P}}}(\sigma_1, \ldots, \sigma_M)\right)\right] \\
& \quad - \mathbb{E}\left[\log\left(\mathsf{r}_{\sigma_{M+1},...,\sigma_{\underline{P}}}(\sigma_{M+1}, \ldots, \sigma_{\underline{P}})\right)\right] \\
& \quad + \mathbb{E}\left[\log\left(J_{\overline{P},\underline{P}}(\sigma_1, \cdots, \sigma_{\underline{P}}) \right)\right] \quad (25)
\end{aligned}
$$

where the second equality follows from the independence between $\mathbf{U}, \mathbf{V}$, and $\mathbf{\Sigma}$. Because $\mathbf{U}$ and $\mathbf{V}$ are uniformly distributed on the corresponding manifolds,

$$
- \mathbb{E}[\log(\mathsf{r}_{\mathbf{U}}(\mathbf{U}))] = \log\left| \widetilde{\mathcal{S}}(T, \underline{P}) \right| \quad (26)
$$

$$
- \mathbb{E}[\log(\mathsf{r}_{\mathbf{V}}(\mathbf{V}))] = \log\left| \mathcal{S}(N, \underline{P}) \right|. \quad (27)
$$

Substituting (22), (23), (24), (26) and (27) into (25) we obtain after simple algebraic manipulations

$$
\begin{aligned}
- \mathbb{E}[\log(\mathsf{r}_{\mathbf{Y}}(\mathbf{Y}))] &= M\underline{L}\log(\lambda) + \log\left(\frac{\Gamma_M(M)\Gamma_M(\underline{L})}{\Gamma_M(N)\Gamma_M(T)}\right) \\
& \quad + NT\log(\pi) + k_3 \mathbb{E}\left[\sum_{i=1}^M \log(\sigma_i^2)\right] \\
& \quad + \sum_{i=1}^M \sum_{j=M+1}^{\underline{P}} \mathbb{E}\left[\log\left((\sigma_i^2 - \sigma_j^2)^2\right)\right] \\
& \quad + \frac{1}{\lambda}\sum_{i=1}^M \mathbb{E}\left[\sigma_i^2\right] + \mathbb{E}\left[\sum_{i=M+1}^{\underline{P}} \sigma_i^2\right]. \quad (28)
\end{aligned}
$$

Here, $k_3 \triangleq \overline{P} - \underline{P} + M - \underline{L}$. We next upper-bound the last three terms on the RHS of (28). Using that the singular values are arranged in decreasing order we obtain

$$
\sum_{i=1}^M \sum_{j=M+1}^{\underline{P}} \mathbb{E}\left[\log\left((\sigma_i^2 - \sigma_j^2)^2\right)\right]
$$

$$
\leq 2(\underline{P}-M)\sum_{i=1}^M \mathbb{E}\left[\log(\sigma_i^2)\right]. \quad (29)
$$

For the second-last term, the power constraint (6) and the noise-variance normalization imply that

$$
\frac{1}{\lambda}\sum_{i=1}^M \mathbb{E}\left[\sigma_i^2\right] \leq \frac{NT(\rho+1)}{\lambda} = M\underline{L} + o(1), \quad \rho \to \infty \quad (30)
$$

where we used that $\lambda = NT\rho/(M\underline{L})$. Finally, to upper-bound the last term in (28) we proceed as in [4, p. 377] and obtain

$$
\mathbb{E}\left[\sum_{i=M+1}^{\underline{P}} \sigma_i^2\right] \leq (T-M)(N-M). \quad (31)
$$

Substituting (29), (30), and (31) into (28), and then (28) and (20) into (19), we get

$$
\begin{aligned}
& I(\mathbf{X}; \mathbf{Y}) \\
& \leq M\underline{L}\log(\rho) - M \underbrace{(N+T-M-\underline{L})}_{=\overline{L}} \\
& \quad + \log\left(\frac{\Gamma_M(M)\Gamma_M(\underline{L})}{\Gamma_M(M)\Gamma_M(T)}\right) + M\underline{L}\log\left(\frac{NT}{M\underline{L}}\right) \\
& \quad + (T-M-\underline{L}) \underbrace{\mathbb{E}\left[\sum_{i=1}^M \log(\sigma_i^2)\right]}_{\triangleq c_1(\rho)} \\
& \quad + N \underbrace{\left(\mathbb{E}\left[\sum_{i=1}^M \log(\sigma_i^2)\right] - \sum_{i=1}^M \mathbb{E}\left[\log\left(1 + \frac{\rho\|\mathbf{x}_i\|^2}{M}\right)\right] \right)}_{\triangleq c_2(\rho)} \\
& \quad + o(1), \quad \rho \to \infty. \quad (32)
\end{aligned}
$$

To conclude the proof, we bound $c_1(\rho)$ and $c_2(\rho)$ by exploiting that $\mathbf{X} \notin \mathcal{K}(\rho_0)$ w.p.1. Let $\mathbf{Z}$ be a $(T-M) \times N$ random matrix, independent of the channel matrix $\mathbf{H}$, and with i.i.d. $\mathcal{CN}(0,1)$ entries. Given $\mathbf{X} = \mathbf{\Phi}\mathbf{D}$, the matrix $\mathbf{Y}^\mathsf{H}\mathbf{Y}$ has the same conditional distribution as [4, p. 377]

$$
\underbrace{\mathbf{H}^\mathsf{H}\left(\mathbf{I}_M + \frac{\rho}{M}\mathbf{D}^2\right)\mathbf{H}}_{\triangleq \mathbf{A}} + \underbrace{\mathbf{Z}^\mathsf{H}\mathbf{Z}}_{\triangleq \mathbf{B}}.
$$

This property allows us to use Weyl's theorem [21, Thm. 4.3.1] to bound $c_1(\rho)$ as follows:

$$
\begin{aligned}
c_1(\rho) &= \mathbb{E}_{\mathbf{X}}\left[\mathbb{E}_{\mathbf{Y}\,|\,\mathbf{X}}\left[\sum_{i=1}^M \log\left(\lambda_i\{\mathbf{Y}^\mathsf{H}\mathbf{Y}\}\right)\Bigg| \mathbf{X}\right]\right] \\
& \leq \mathbb{E}_{\mathbf{X}}\left[\mathbb{E}_{\mathbf{H},\mathbf{Z}}\left[\sum_{i=1}^M \log\left(\lambda_i\{\mathbf{A}\} + \lambda_1\{\mathbf{B}\}\right)\Bigg| \mathbf{X}\right]\right] \\
& \leq \mathbb{E}_{\mathbf{X}}\left[\mathbb{E}_{\mathbf{H}}\left[\sum_{i=1}^M \log\left(\lambda_i\{\mathbf{A}\} + \underbrace{\mathbb{E}_{\mathbf{Z}}[\lambda_1\{\mathbf{B}\}]}_{\triangleq \eta}\right)\Bigg| \mathbf{X}\right]\right]. \quad (33)
\end{aligned}
$$

Here, in the last step we used Jensen's inequality. We next rewrite the argument in the expectation on the RHS of (33) in a

more convenient form:

$$\sum_{i=1}^{M} \log\big(\lambda_i\{\mathbf{A}\} + \eta\big)$$

$$\overset{(a)}{=} \log\det\left(\left(\mathbf{I}_M + \frac{\rho}{M}\mathbf{D}^2\right)\mathbf{H}\mathbf{H}^{\mathsf{H}} + \eta\mathbf{I}_M\right)$$

$$= \log\det\left(\mathbf{I}_M + \frac{\rho}{M}\mathbf{D}^2\right)$$

$$+ \log\det\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \mathrm{diag}\bigg\{\left[\eta\left(1 + \frac{\rho}{M}\|\mathbf{x}_1\|^2\right)^{-1}\right.\right.$$

$$\left.\left.\cdots\ \eta\left(1 + \frac{\rho}{M}\|\mathbf{x}_M\|^2\right)^{-1}\right]^{\mathsf{T}}\bigg\}\right)$$

$$\overset{(b)}{\leq} \log\det\left(\mathbf{I}_M + \frac{\rho}{M}\mathbf{D}^2\right)$$

$$+ \underbrace{\log\det\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \eta\left(1 + \frac{\rho_0}{M}\right)^{-1}\mathbf{I}_M\right)}_{\triangleq \kappa(\mathbf{H},\rho_0)}. \tag{34}$$

Here, (a) follows because $\mathbf{H}^{\mathsf{H}}\big(\mathbf{I}_M + (\rho/M)\mathbf{D}^2\big)\mathbf{H}$ and $\big(\mathbf{I}_M + (\rho/M)\mathbf{D}^2\big)\mathbf{H}\mathbf{H}^{\mathsf{H}}$ have the same M nonzero eigenvalues [21, Thm. 1.3.20], and (b) follows because $\mathbf{X} \notin \mathcal{K}(\rho_0)$ w.p.1 and because for two matrices $\mathbf{A}$ and $\mathbf{B}$, if $\mathbf{A} - \mathbf{B}$ is positive semidefinite then $\det(\mathbf{A}) \geq \det(\mathbf{B})$ [21, Cor. 7.7.4]. Substituting (34) into (33) we obtain

$$c_1(\rho) \leq \mathbb{E}\left[\log\det\left(\mathbf{I}_M + \frac{\rho}{M}\mathbf{D}^2\right)\right] + \mathbb{E}[\kappa(\mathbf{H},\rho_0)] \tag{35}$$

$$\leq M\log\left(1 + \frac{T\rho}{M}\right) + \mathbb{E}[\kappa(\mathbf{H},\rho_0)]$$

$$= M\log\left(\frac{T\rho}{M}\right) + \mathbb{E}[\kappa(\mathbf{H},\rho_0)] + o(1),\ \rho \to \infty. \tag{36}$$

To bound $c_2(\rho)$ we use (35) and obtain

$$c_2(\rho) \leq \mathbb{E}[\kappa(\mathbf{H},\rho_0)]. \tag{37}$$

Finally, substituting (36) and (37) into (32) we get

$$I(\mathbf{X};\mathbf{Y}) \leq M(T-M)\log(\rho) + T\cdot c_{\rho_0} + o(1),\ \rho \to \infty \tag{38}$$

where

$$c_{\rho_0} \triangleq \frac{1}{T}\log\left(\frac{\Gamma_M(M)\Gamma_M(\underline{L})}{\Gamma_M(N)\Gamma_M(T)}\right) + M\left(1 - \frac{M}{T}\right)\log\left(\frac{T}{M}\right)$$

$$+ \frac{M\underline{L}}{T}\log\left(\frac{N}{\underline{L}}\right) + \frac{\overline{L}}{T}\big(\mathbb{E}[\kappa(\mathbf{H},\rho_0)] - M\big). \tag{39}$$

Note that the RHS of (38) does not depend on the choice of the input distribution. Hence, (38) is an upper bound on capacity as well. Because $\mathbf{H}$ has i.i.d. Gaussian entries, and, hence, its singular values have finite differential entropy, we can apply [10, Lem. 6.7(b)] combined with the dominated convergence theorem [28, p. 180] and obtain

$$\lim_{\rho_0 \to \infty} \mathbb{E}[\kappa(\mathbf{H},\rho_0)] = \mathbb{E}\left[\lim_{\rho_0 \to \infty} \kappa(\mathbf{H},\rho_0)\right]$$

$$= \mathbb{E}\left[\log\det(\mathbf{H}\mathbf{H}^{\mathsf{H}})\right].$$

Hence, c_{ρ_0} in (39) can be made arbitrarily close to c in (9) by choosing ρ_0 sufficiently large.

B. Lower Bound

To obtain a capacity lower bound that matches the upper bound derived in Section IV-A, we evaluate $I(\mathbf{X};\mathbf{Y})$ for the BSTM input distribution introduced in Section III-C. More specifically, we proceed as follows. Fix $\rho_0 > 0$ and let

$$\widetilde{\mathcal{K}}(\rho_0) \triangleq \Big\{\mathbf{\Lambda} = \mathrm{diag}\{[\lambda_1\ \cdots\ \lambda_M]^{\mathsf{T}}\}:$$

$$0 < \min_{m=1,\ldots,M}\{\lambda_m^2\} < \rho_0/\rho\Big\}.$$

Starting from $\mathrm{Q}_{\mathbf{D}}^{\mathrm{opt}}$ (see Section III-C), we define the following family of probability distributions parameterized with respect to[6] ρ

$$\mathrm{Q}_{\mathbf{D}}^{\mathrm{opt},\rho}(\mathbf{\Lambda}) = \begin{cases} \frac{\mathrm{Q}_{\mathbf{D}}^{\mathrm{opt}}(\mathbf{\Lambda})}{1 - \mathbb{P}\big(\mathbf{D} \in \widetilde{\mathcal{K}}(\rho_0),\, \mathbf{D} \sim \mathrm{Q}_{\mathbf{D}}^{\mathrm{opt}}\big)}, & \text{if } \mathbf{\Lambda} \notin \widetilde{\mathcal{K}}(\rho_0) \\ 0, & \text{if } \mathbf{\Lambda} \in \widetilde{\mathcal{K}}(\rho_0). \end{cases} \tag{40}$$

Note that $\mathrm{Q}_{\mathbf{D}}^{\mathrm{opt},\rho}(\mathbf{\Lambda})$ is supported outside $\widetilde{\mathcal{K}}(\rho_0)$ and that $\lim_{\rho \to \infty} \mathrm{Q}_{\mathbf{D}}^{\mathrm{opt},\rho}(\mathbf{\Lambda}) = \mathrm{Q}_{\mathbf{D}}^{\mathrm{opt}}(\mathbf{\Lambda})$ for all $\mathbf{\Lambda}$.

1) Preliminary Results: In Lemma 12 below, we establish that when $\mathbf{X} = \mathbf{\Phi}\mathbf{D}$ with $\mathbf{D} \sim \mathrm{Q}_{\mathbf{D}}^{\mathrm{opt},\rho}$ and $\mathbf{\Phi}$ unitary and isotropically distributed, the joint pdf of the largest M singular values of the output matrix $\mathbf{Y}$ in (4) converges pointwise to the pdf of the nonzero singular values of $\sqrt{\rho/M}\cdot\mathbf{\Phi}\mathbf{D}\mathbf{H}$. Furthermore, the pdf of the remaining $\underline{P} - M$ singular values converge pointwise to the pdf of the singular values of an independent Gaussian matrix. We remark that we implicitly used this property to construct the output distribution in Section IV-A.

Lemma 12: Let $\mathbf{X} = \mathbf{\Phi}\mathbf{D}$ where $\mathbf{\Phi}$ is unitary and isotropically distributed and $\mathbf{D} \sim \mathrm{Q}_{\mathbf{D}}^{\mathrm{opt},\rho}$; let $\mathbf{Y}$ as in (4). Denote by $\sigma_1 > \cdots > \sigma_P$ the singular values of $\mathbf{Y}$ and let

$$\tilde{\boldsymbol{\sigma}} = \left[\big(\sqrt{M/\rho}\big)\sigma_1\ \cdots\ \big(\sqrt{M/\rho}\big)\sigma_M\ \sigma_{M+1}\ \cdots\ \sigma_{\underline{P}}\right]^{\mathsf{T}}. \tag{41}$$

The pdf of $\tilde{\boldsymbol{\sigma}}$ converges pointwise as $\rho \to \infty$ to the pdf of a vector $\mathbf{u} \in \mathbb{C}^{\underline{P}}$ whose first M entries are distributed as the ordered nonzero singular values of $\mathbf{D}\mathbf{H}$, with $\mathbf{D} \sim \mathrm{Q}_{\mathbf{D}}^{\mathrm{opt}}$ and $\mathbf{H}$ as in (4), and whose remaining $\underline{P} - M$ entries are distributed as the nonzero singular values of an independent $(T-M)\times(N-M)$ random matrix with i.i.d. $\mathcal{CN}(0,1)$ entries.

Proof: See Appendix A. ∎

Note that by Scheffé's Theorem [29], pointwise convergence of pdfs implies convergence in distribution of $\tilde{\boldsymbol{\sigma}}$ to $\mathbf{u}$. This weaker convergence result (which is not sufficient to establish our capacity lower bound) has been already pointed out (without proof) in [4, Lem. 16].

In Lemma 13 below we collect four asymptotic results regarding the differential entropy and the expected logarithm of the entries of $\tilde{\boldsymbol{\sigma}}$ in (41) that we shall need in the proof of the lower bound.

Lemma 13: Let $\tilde{\boldsymbol{\sigma}} = [\tilde{\sigma}_1\ \cdots\ \tilde{\sigma}_{\underline{P}}]^{\mathsf{T}}$ and $\mathbf{u} = [u_1\ \cdots\ u_{\underline{P}}]^{\mathsf{T}}$ as in Lemma 12. Then

1) $h(\tilde{\boldsymbol{\sigma}}) = h(\mathbf{u}) + o(1),\ \rho \to \infty$
2) $\mathbb{E}[\log(\tilde{\sigma}_i)] = \mathbb{E}[\log(u_i)] + o(1),\ \rho \to \infty,\ 1 \leq i \leq \underline{P}$
3) $\mathbb{E}\big[\log(\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2)\big] = \mathbb{E}\big[\log(u_i^2 - u_j^2)\big] + o(1),\ \rho \to \infty,$
   $1 \leq i < j \leq \underline{P}$

---

[6]Although $\mathrm{Q}_{\mathbf{D}}^{\mathrm{opt},\rho}$ depends on both ρ and ρ_0, the choice of ρ_0 in the proof of the lower bound will turn out to be immaterial.

4) $\mathbb{E}\left[\log\left(\tilde{\sigma}_i^2 - M\tilde{\sigma}_j^2/\rho\right)\right] = \mathbb{E}\left[\log(u_i^2)\right] + o(1),\ \rho \to \infty,$ $1 \leq i < j \leq \underline{P}.$

Proof: See Appendix B. ∎

2) The Actual Bound: We evaluate the mutual information

$$I(\mathbf{X};\mathbf{Y}) = h(\mathbf{Y}) - h(\mathbf{Y}\mid\mathbf{X}) \tag{42}$$

in (5) for $\mathbf{X} = \boldsymbol{\Phi}\mathbf{D}$ with $\boldsymbol{\Phi}$ unitary and isotropically distributed and $\mathbf{D} \sim Q_{\mathbf{D}}^{\mathrm{opt},\rho}$. The second term on the RHS of (42) is given by

$$
\begin{aligned}
h(\mathbf{Y}\mid\mathbf{X}) &= NT\log(\pi e) + N\sum_{i=1}^{M}\mathbb{E}\left[\log\left(1 + \rho\|\mathbf{x}_i\|^2/M\right)\right] \\
&= NT\log(\pi e) + N\sum_{i=1}^{M}\mathbb{E}\left[\log(\rho d_i^2/M)\right] \\
&\quad + N\sum_{i=1}^{M}\mathbb{E}\left[\log\left(1 + M/(\rho d_i^2)\right)\right] \\
&\stackrel{(a)}{=} NT\log(\pi e) + MN\log(\rho/M) \\
&\quad + N\,\mathbb{E}\left[\log\det\left(\mathbf{D}^2\right)\right] + o(1),\quad \rho \to \infty. \tag{43}
\end{aligned}
$$

Here, (a) follows because $\rho d_i^2 \geq \rho_0$ w.p.1, and hence, $0 \leq \log\left(1 + M/(\rho d_i^2)\right) \leq \log(1 + M/\rho_0)$ w.p.1, which implies that

$$\lim_{\rho\to\infty}\mathbb{E}\left[\log\left(1 + \frac{M}{\rho d_i^2}\right)\right] = \mathbb{E}\left[\lim_{\rho\to\infty}\log\left(1 + \frac{M}{\rho d_i^2}\right)\right] = 0$$

as a consequence of the dominated convergence theorem [28, p. 180]. We shall compute $h(\mathbf{Y})$ in SVD coordinates [cf., (21)] as follows:

$$
\begin{aligned}
h(\mathbf{Y}) &\stackrel{(a)}{=} \underbrace{h(\mathbf{U})}_{=\log|\widetilde{\mathcal{S}}(T,\underline{P})|} + \underbrace{h(\mathbf{V})}_{=\log|\mathcal{S}(N,\underline{P})|} + h(\sigma_1,\ldots,\sigma_{\underline{P}}) \\
&\quad + \mathbb{E}\left[\log\left(J_{\overline{P},\underline{P}}(\sigma_1,\ldots,\sigma_{\underline{P}})\right)\right] \\
&\stackrel{(b)}{=} \log|\widetilde{\mathcal{S}}(T,\underline{P})| + \log|\mathcal{S}(N,\underline{P})| + \frac{M}{2}\log\left(\frac{\rho}{M}\right) \\
&\quad + h(\tilde{\boldsymbol{\sigma}}) + \mathbb{E}\left[\log\left(J_{\overline{P},\underline{P}}(\sigma_1,\ldots,\sigma_{\underline{P}})\right)\right]. \tag{44}
\end{aligned}
$$

Here, (a) follows because the isotropic distribution of $\boldsymbol{\Phi}$ and the Gaussianity of $\mathbf{H}$ and $\mathbf{W}$ imply that $\mathbf{U}$ and $\mathbf{V}$ are uniformly distributed on $\widetilde{\mathcal{S}}(T,\underline{P})$ and $\mathcal{S}(N,\underline{P})$, respectively, and independent of $\boldsymbol{\Sigma}$; In (b), we used (41) and that $h(\mathbf{A}\mathbf{x}) = h(\mathbf{x}) + \log\det(\mathbf{A})$ for a random vector $\mathbf{x}$ and a deterministic matrix $\mathbf{A}$ [19, Eq. (8.71)]. It is convenient to express also the Jacobian $J_{\overline{P},\underline{P}}$ in (44) in terms of $\tilde{\boldsymbol{\sigma}}$. Using (24) and (41) we obtain

$$
\begin{aligned}
&\mathbb{E}\left[\log\left(J_{\overline{P},\underline{P}}(\sigma_1,\ldots,\sigma_{\underline{P}})\right)\right] \\
&= k_4\log\left(\frac{\rho}{M}\right) + \sum_{i=1}^{M}\mathbb{E}\left[\log\left(\tilde{\sigma}_i^{2(\overline{P}-\underline{P})+1}\right)\right] \\
&\quad + \sum_{i<j}^{M}\mathbb{E}\left[\log\left((\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2)^2\right)\right] \\
&\quad + \sum_{i=M+1}^{\underline{P}}\mathbb{E}\left[\log\left(\tilde{\sigma}_i^{2(\overline{P}-\underline{P})+1}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \sum_{M<i<j}^{\underline{P}}\mathbb{E}\left[\log\left((\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2)^2\right)\right] \\
&\quad + \sum_{i=1}^{M}\sum_{j=M+1}^{\underline{P}}\mathbb{E}\left[\log\left((\tilde{\sigma}_i^2 - M\tilde{\sigma}_j^2/\rho)^2\right)\right] \tag{45}
\end{aligned}
$$

where $k_4 \triangleq M(\overline{P} + \underline{P} - M - 1/2)$. Substituting (45) into (44), and using Lemma 13, we obtain

$$
\begin{aligned}
h(\mathbf{Y}) &= \log|\widetilde{\mathcal{S}}(T,\underline{P})| + \log|\mathcal{S}(N,\underline{P})| \\
&\quad + M(\overline{P} + \underline{P} - M)\log(\rho/M) + h(u_1,\ldots,u_M) \\
&\quad + \sum_{i=1}^{M}\mathbb{E}\left[\log\left(u_i^{2(\overline{P}+\underline{P}-2M)+1}\right)\right] \\
&\quad + \sum_{i<j}^{M}\mathbb{E}\left[\log\left((u_i^2 - u_j^2)^2\right)\right] + h(u_{M+1},\ldots,u_{\underline{P}}) \\
&\quad + \sum_{i=M+1}^{\underline{P}}\mathbb{E}\left[\log\left(u_i^{2(\overline{P}-\underline{P})+1}\right)\right] \\
&\quad + \sum_{M<i<j}^{\underline{P}}\mathbb{E}\left[\log\left((u_i^2 - u_j^2)^2\right)\right] + o(1),\ \rho \to \infty. \tag{46}
\end{aligned}
$$

We next evaluate the terms on the RHS of (46) by proceeding as follows. By Lemmas 12 and 10, $\{u_1,\ldots,u_M\}$ are jointly distributed as the singular values of an $M \times \underline{L}$ Gaussian random matrix $\mathbf{G}$ with i.i.d. $\mathcal{CN}(0,\sqrt{NT/\underline{L}})$ entries. Evaluating $h(\mathbf{G})$ in the SVD coordinate system, we get

$$
\begin{aligned}
h(\mathbf{G}) &= \log|\widetilde{\mathcal{S}}(M,M)| + \log|\mathcal{S}(\underline{L},M)| + h(u_1,\ldots,u_M) \\
&\quad + \sum_{i=1}^{M}\mathbb{E}\left[\log\left(u_i^{2(\underline{L}-M)+1}\right)\right] \\
&\quad + \sum_{i<j}^{M}\mathbb{E}\left[\log\left((u_i^2 - u_j^2)^2\right)\right]. \tag{47}
\end{aligned}
$$

Similarly, by Lemma 12, $\{u_{M+1},\ldots,u_{\underline{L}}\}$ are jointly distributed as the singular values of a $(T-M) \times (N-M)$ random Gaussian matrix $\widetilde{\mathbf{W}}$ with i.i.d. $\mathcal{CN}(0,1)$ entries. Thus,

$$
\begin{aligned}
h(\widetilde{\mathbf{W}}) &= \log\left|\widetilde{\mathcal{S}}(T-M,\underline{P}-M)\right| + \log|\mathcal{S}(N-M,\underline{P}-M)| \\
&\quad + h(u_{M+1},\ldots,u_{\underline{P}}) + \sum_{i=M+1}^{\underline{P}}\mathbb{E}\left[\log\left(u_i^{2(\overline{P}-\underline{P})+1}\right)\right] \\
&\quad + \sum_{M<i<j}^{\underline{P}}\mathbb{E}\left[\log\left((u_i^2 - u_j^2)^2\right)\right]. \tag{48}
\end{aligned}
$$

Substituting (47) and (48) into (46), and then (43) and (46) into (42), we obtain

$$
\begin{aligned}
&I(\mathbf{X};\mathbf{Y}) \\
&= M(T-M)\log\left(\frac{\rho}{M}\right) + h(\mathbf{G}) + h(\widetilde{\mathbf{W}}) \\
&\quad + \underbrace{\sum_{i=1}^{M}\mathbb{E}\left[\log\left(u_i^{2(\overline{P}+\underline{P}-M-\underline{L})}\right)\right] - N\,\mathbb{E}\left[\log\det\left(\mathbf{D}^2\right)\right]}_{\triangleq\alpha} \\
&\quad + \log(k_5) - NT\log(\pi e) + o(1),\quad \rho \to \infty \tag{49}
\end{aligned}
$$

where $k_5 \triangleq |\widetilde{\mathcal{S}}(T,\underline{P})| \cdot |\mathcal{S}(N,\underline{P})| / \Big[|\widetilde{\mathcal{S}}(T-M,\underline{P}-M)| \cdot |\mathcal{S}(N-M,\underline{P}-M)| \cdot |\widetilde{\mathcal{S}}(M,M)| \cdot |\mathcal{S}(\underline{L},M)| \Big]$. The term denoted by α in (49) can be simplified as follows:

$$\alpha \overset{(a)}{=} \overline{L}\,\mathbb{E}\left[\sum_{i=1}^{M} \log\big(u_i^2\big)\right] - N\,\mathbb{E}\big[\log\det(\mathbf{D}^2)\big]$$

$$\overset{(b)}{=} \overline{L}\,\mathbb{E}\big[\log\det(\mathbf{D}^2\mathbf{H}\mathbf{H}^{\mathsf{H}})\big] - N\,\mathbb{E}\big[\log\det(\mathbf{D}^2)\big]$$

$$= (\overline{L}-N)\,\mathbb{E}\big[\log\det(\mathbf{D}^2)\big] + \overline{L}\,\mathbb{E}\big[\log\det(\mathbf{H}\mathbf{H}^{\mathsf{H}})\big]$$

$$\overset{(c)}{=} (\overline{L}-N)M\log(T) + \overline{L}\,\mathbb{E}\big[\log\det(\mathbf{H}\mathbf{H}^{\mathsf{H}})\big]. \quad (50)$$

Here, in (a) we used that $\overline{L} = \overline{P} + \underline{P} - M - \underline{L}$, (b) follows from Lemma 12, and (c) holds because when $T \geq M+N$ we have that $\mathbf{D} = \sqrt{T} \cdot \mathbf{I}_M$, and when $T < M+N$ we have that $\overline{L} - N = 0$. Finally, substituting (50) into (49), we get after straightforward algebraic manipulations

$$I(\mathbf{X};\mathbf{Y}) = M(T-M)\log(\rho) + T \cdot c + o(1), \quad \rho \to \infty$$

where c is given in (9). This concludes the proof.

V. Conclusions

It was shown in [4] that USTM achieves the high-SNR capacity of a Rayleigh block-fading MIMO channel in the regime where the channel's coherence time T is larger or equal to the sum of the number of transmit antennas M and receive antennas N. In the same paper, it was also conjectured that when $T < M + N$, a situation relevant for large-MIMO systems, USTM is no longer optimal. In this paper, we prove this conjecture. Specifically, we establish that USTM is not capacity-achieving when $T < M+N$ by determining the input distribution (which we refer to as BSTM) that achieves capacity at high SNR. The corresponding capacity-achieving input signal is the product of a unitary isotropically distributed matrix and a diagonal matrix whose nonzero entries are distributed as the square-root of the eigenvalues of a Beta-distributed matrix of appropriate size. The analytical and numerical results reported in Section III-D illustrate that the rate gain determined by using BSTM instead USTM grows logarithmically in the number of receive antennas N, and can be as large as 13% for practically relevant SNR values, when $N \gg T$ and $M = \lfloor T/2 \rfloor$.

Appendix A
Proof of Lemma 12

Throughout this appendix, we shall focus for simplicity on the case $T \leq N$. We shall, however, outline the additional steps needed to generalize the proof to the case $T > N$. Let $\mathsf{q}_{\mathbf{D}}^{\mathrm{opt},\rho}$ and $\mathsf{q}_{\mathbf{D}}^{\mathrm{opt}}$ be the pdfs corresponding to the probability distributions $\mathsf{Q}_{\mathbf{D}}^{\mathrm{opt},\rho}$ and $\mathsf{Q}_{\mathbf{D}}^{\mathrm{opt}}$, respectively (such pdfs exist when $T \leq N$). Let $\mathsf{f}_{\mathbf{Y}|\mathbf{D}}$ denote the conditional pdf of $\mathbf{Y}$ given $\mathbf{D}$. Denote by $\mathsf{f}_{\widetilde{\boldsymbol{\sigma}}}^{(\rho)}$ and $\mathsf{f}_{\mathbf{u}}$ the pdf of $\widetilde{\boldsymbol{\sigma}}$ and $\mathbf{u}$, respectively. Finally, denote by $\mathsf{f}_{\widetilde{\boldsymbol{\sigma}}|\mathbf{D}}^{(\rho)}$ and $\mathsf{f}_{\mathbf{u}|\mathbf{D}}$ the conditional pdf of $\widetilde{\boldsymbol{\sigma}}$ and $\mathbf{u}$ given $\mathbf{D}$. The proof consists of the following three steps:

1) We first obtain a closed-form expression for $\mathsf{f}_{\mathbf{Y}|\mathbf{D}}$, thus generalizing the result obtained in [9, Sec. III.A] (for the special case of $\mathbf{D}$ being a scaled identity matrix) to arbitrary diagonal matrices. This result is of independent interest.

2) We then calculate $\mathsf{f}_{\widetilde{\boldsymbol{\sigma}}|\mathbf{D}}^{(\rho)}$ from $\mathsf{f}_{\mathbf{Y}|\mathbf{D}}$ and show that $\mathsf{f}_{\widetilde{\boldsymbol{\sigma}}|\mathbf{D}}^{(\rho)}$ converges pointwise to $\mathsf{f}_{\mathbf{u}|\mathbf{D}}$ as $\rho \to \infty$.

3) Finally, we show that

$$\left| \mathsf{f}_{\widetilde{\boldsymbol{\sigma}}|\mathbf{D}}^{(\rho)}(\mathbf{a}\,|\,\boldsymbol{\Lambda}) \cdot \mathsf{q}_{\mathbf{D}}^{\mathrm{opt},\rho}(\boldsymbol{\Lambda}) \right| \leq k \quad (51)$$

where k is a finite constant that does not depend on $\mathbf{a}$ and $\boldsymbol{\Lambda}$, i.e., the bound is uniform in both $\mathbf{a}$ and $\boldsymbol{\Lambda}$. As $\mathbf{D} \sim \mathsf{Q}_{\mathbf{D}}^{\mathrm{opt},\rho}$ implies that $\mathbf{D}$ has compact support, we can invoke the dominated convergence theorem [30, Thm. 1.34] and conclude that

$$\lim_{\rho \to \infty} \mathsf{f}_{\widetilde{\boldsymbol{\sigma}}}^{(\rho)}(\mathbf{a}) = \lim_{\rho \to \infty} \int \mathsf{f}_{\widetilde{\boldsymbol{\sigma}}|\mathbf{D}}^{(\rho)}(\mathbf{a}\,|\,\boldsymbol{\Lambda})\mathsf{q}_{\mathbf{D}}^{\mathrm{opt},\rho}(\boldsymbol{\Lambda})d\boldsymbol{\Lambda}$$

$$= \int \lim_{\rho \to \infty} \left[\mathsf{f}_{\widetilde{\boldsymbol{\sigma}}|\mathbf{D}}^{(\rho)}(\mathbf{a}\,|\,\boldsymbol{\Lambda})\mathsf{q}_{\mathbf{D}}^{\mathrm{opt},\rho}(\boldsymbol{\Lambda}) \right] d\boldsymbol{\Lambda}$$

$$= \int \mathsf{f}_{\mathbf{u}|\mathbf{D}}(\mathbf{a}\,|\,\boldsymbol{\Lambda})\mathsf{q}_{\mathbf{D}}^{\mathrm{opt}}(\boldsymbol{\Lambda})d\boldsymbol{\Lambda} = \mathsf{f}_{\mathbf{u}}(\mathbf{a}).$$

A. Step 1

Set $\tilde{\rho} \triangleq \rho/M$. Since $\mathbf{Y}$ is conditionally Gaussian given $\mathbf{X}$,

$$\mathsf{f}_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}\,|\,\mathbf{X}) = \frac{1}{\pi^{NT}} \cdot \frac{\exp\Big[-\operatorname{tr}\Big(\mathbf{Y}^{\mathsf{H}}\big(\tilde{\rho}\mathbf{X}\mathbf{X}^{\mathsf{H}} + \mathbf{I}_T\big)^{-1}\mathbf{Y}\Big)\Big]}{\det(\tilde{\rho}\mathbf{X}\mathbf{X}^{\mathsf{H}} + \mathbf{I}_T)^N}.$$

To obtain $\mathsf{f}_{\mathbf{Y}|\mathbf{D}}$ from $\mathsf{f}_{\mathbf{Y}|\mathbf{X}}$, it is convenient to consider the eigenvalue decomposition of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$:

$$\mathbf{Y}\mathbf{Y}^{\mathsf{H}} = \widetilde{\mathbf{U}} \underbrace{\begin{pmatrix} \boldsymbol{\Sigma}^2 & \mathbf{0}_{\underline{P}\times(T-\underline{P})} \\ \mathbf{0}_{(T-\underline{P})\times\underline{P}} & \mathbf{0}_{T-\underline{P}} \end{pmatrix}}_{\triangleq\,\boldsymbol{\Delta}} \widetilde{\mathbf{U}}^{\mathsf{H}}. \quad (52)$$

Here, $\widetilde{\mathbf{U}}$ is a $T \times T$ unitary matrix, and $\boldsymbol{\Sigma}$, defined in (21), contains the singular values $\sigma_1 > \cdots > \sigma_{\underline{P}}$ of $\mathbf{Y}$. Set now $\boldsymbol{\Lambda} \triangleq (\tilde{\rho}^{-1}\mathbf{D}^{-2} + \mathbf{I}_M)^{-1}$ and recall that $\mathbf{X} = \boldsymbol{\Phi}\mathbf{D}$, where $\boldsymbol{\Phi}$ is unitary and isotropically distributed, and, hence, uniformly distributed on $\mathcal{S}(T,M)$. Proceeding as in [9, Sec. III],

$$\mathsf{f}_{\mathbf{Y}|\mathbf{D}}(\mathbf{Y}\,|\,\mathbf{D}) = \frac{1}{|\mathcal{S}(T,M)|} \int \mathsf{f}_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}\,|\,\boldsymbol{\Phi}\mathbf{D})d\boldsymbol{\Phi}$$

$$= \frac{1}{\pi^{NT}} \cdot \frac{\exp\big[-\operatorname{tr}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y})\big]}{\det(\tilde{\rho}\mathbf{D}^2 + \mathbf{I}_M)^N} \cdot \frac{1}{|\mathcal{S}(T,M)|}$$

$$\cdot \int_{\mathcal{S}(T,M)} \exp\big[\operatorname{tr}(\boldsymbol{\Delta}\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathsf{H}})\big]d\boldsymbol{\Phi}. \quad (53)$$

The integral on the RHS of (53) is computed in closed-form in [9, Sec. III.A] for the special case $\mathbf{D} = \sqrt{T} \cdot \mathbf{I}_M$, which corresponds to USTM. We shall next evaluate this integral (and, hence, $\mathsf{f}_{\mathbf{Y}|\mathbf{D}}$) in closed-form for an arbitrary diagonal matrix $\mathbf{D}$. We start by observing that the integral under examination resembles the well-known Itzykson-Zuber integral [31, Eq. (3.2)], with the crucial differences that, in our case, the integration is performed over the Stiefel manifold $\mathcal{S}(T,M)$ instead of the *unitary group* $\mathcal{U}(T) \triangleq \mathcal{S}(T,T)$. Let $\widetilde{\boldsymbol{\Phi}} = [\boldsymbol{\Phi}\ \boldsymbol{\Phi}_\perp]$ where $\boldsymbol{\Phi}_\perp$ is

a $T \times (T-M)$ matrix chosen so that $\widetilde{\boldsymbol{\Phi}}$ is unitary, i.e., $\widetilde{\boldsymbol{\Phi}} \in \mathcal{U}(T)$. Then [32, Eq. (5)]

$$\int_{\mathcal{S}(T,M)} e^{\text{tr}(\boldsymbol{\Delta}\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathsf{H}})} d\boldsymbol{\Phi} = \frac{1}{|\mathcal{U}(T-M)|} \int_{\mathcal{U}(T)} e^{\text{tr}(\boldsymbol{\Delta}\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathsf{H}})} d\widetilde{\boldsymbol{\Phi}}. \quad (54)$$

The assumption $T \leq N$ entails that the nonzero entries of the diagonal matrix $\mathbf{D}$ are distinct (see Section III-C); hence, the nonzero entries of the diagonal matrix $\boldsymbol{\Lambda}$ are distinct as well. Furthermore, when $T \leq N$ we have that $\underline{P} = T$ (and $\overline{P} = N$) and, hence, $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^2$ [see (52)]. Starting from $\boldsymbol{\Lambda} = \text{diag}\{[\lambda_1 \cdots \lambda_M]^{\mathsf{T}}\}$, we next define the following full-rank $T \times T$ diagonal matrix:

$$\boldsymbol{\Lambda}_\epsilon \triangleq \text{diag}\{[\lambda_1 \cdots \lambda_M \; \epsilon'_{M+1} \cdots \epsilon'_T]^{\mathsf{T}}\}.$$

Here, $\epsilon'_{M+1}, \ldots, \epsilon'_T$ are nonnegative real numbers chosen so that the nonzero entries of $\boldsymbol{\Lambda}_\epsilon$ are distinct. As the unitary group $\mathcal{U}(T)$ is compact,

$$\int_{\mathcal{U}(T)} e^{\text{tr}(\boldsymbol{\Sigma}^2\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathsf{H}})} d\widetilde{\boldsymbol{\Phi}} = \lim_{\epsilon'_{M+1} \to 0, \ldots, \epsilon'_T \to 0} \int_{\mathcal{U}(T)} e^{\text{tr}(\boldsymbol{\Sigma}^2\widetilde{\boldsymbol{\Phi}}\boldsymbol{\Lambda}_\epsilon\widetilde{\boldsymbol{\Phi}}^{\mathsf{H}})} d\widetilde{\boldsymbol{\Phi}}. \quad (55)$$

The argument of the $\lim$ operator on the RHS of (55) is the Itzykson-Zuber integral. Hence, by [31, Eq. (3.4)] we get[7]

$$\int_{\mathcal{U}(T)} e^{\text{tr}(\boldsymbol{\Sigma}^2\widetilde{\boldsymbol{\Phi}}\boldsymbol{\Lambda}_\epsilon\widetilde{\boldsymbol{\Phi}}^{\mathsf{H}})} d\widetilde{\boldsymbol{\Phi}} = \frac{|\mathcal{U}(T)| \cdot \prod_{i=1}^{T} \Gamma(i) \cdot \det(\mathbf{A})}{\prod_{i<j}^{T} (\sigma_i^2 - \sigma_j^2) \cdot \prod_{i<j}^{T} (\tilde{\lambda}_i - \tilde{\lambda}_j)}. \quad (56)$$

Here, $\{\tilde{\lambda}_j\}_{j=1}^{T}$ are the diagonal entries of $\boldsymbol{\Lambda}_\epsilon$, and $\mathbf{A}$ is a $T \times T$ matrix defined as follows: $[\mathbf{A}]_{i,j} = \exp(\sigma_j^2 \tilde{\lambda}_i)$. We next compute the limit $\epsilon'_{M+1} \to 0, \ldots, \epsilon'_T \to 0$ of the RHS of (56) using l'Hôpital's Theorem, substitute the final result into (54), and obtain [33, Lem. 5]

$$\int_{\mathcal{S}(T,M)} e^{\text{tr}(\boldsymbol{\Delta}\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathsf{H}})} d\boldsymbol{\Phi}$$
$$= \underbrace{\frac{|\mathcal{U}(T)|}{|\mathcal{U}(T-M)|}}_{=|\mathcal{S}(T,M)|} \cdot \frac{\det(\mathbf{M}) \det(\boldsymbol{\Lambda}^{M-T}) \cdot \prod_{i=T-M+1}^{T} \Gamma(i)}{\prod_{i<j}^{T} (\sigma_i^2 - \sigma_j^2) \cdot \prod_{i<j}^{M} (\lambda_i - \lambda_j)} \quad (57)$$

with $\mathbf{M}$ being a $T \times T$ matrix defined as follows:

$$[\mathbf{M}]_{i,j} = \begin{cases} e^{\lambda_i \sigma_j^2}, & 1 \leq i \leq M, \; 1 \leq j \leq T \\ \sigma_j^{2(T-i)}, & M < i \leq T, \; 1 \leq j \leq T. \end{cases}$$

Substituting (57) into (53) and using that $\boldsymbol{\Lambda} = (\tilde{\rho}^{-1}\mathbf{D}^{-2} + \mathbf{I}_M)^{-1}$, we obtain the following closed-form

---

[7]Note that—differently from our setup—in [31, Eq. (3.4)] the Haar measure on the unitary group is normalized.

---

expression for the conditional pdf $\mathsf{f}_{\mathbf{Y}\,|\,\mathbf{D}}$:

$$\mathsf{f}_{\mathbf{Y}\,|\,\mathbf{D}}(\mathbf{Y}\,|\,\mathbf{D})$$
$$= \frac{1}{\pi^{NT}} \cdot \prod_{i=T-M+1}^{T} \Gamma(i) \cdot \frac{\exp[-\text{tr}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y})]}{\det(\tilde{\rho}\mathbf{D}^2 + \mathbf{I}_M)^N}$$
$$\cdot \frac{\det(\mathbf{M}) \det(\boldsymbol{\Lambda}^{M-T})}{\prod_{i<j}^{T} (\sigma_i^2 - \sigma_j^2) \cdot \prod_{i<j}^{M} (\lambda_i - \lambda_j)}$$
$$= \frac{\tilde{\rho}^{-M(M-1)/2}}{\pi^{NT}} \cdot \prod_{i=T-M+1}^{T} \Gamma(i) \cdot \frac{\det(\tilde{\rho}^{-1}\mathbf{D}^{-2} + \mathbf{I}_M)^{T-M}}{\det(\tilde{\rho}\mathbf{D}^2 + \mathbf{I}_M)^{N-M+1}}$$
$$\cdot \frac{\exp(-\sum_{i=1}^{P} \sigma_i^2)}{\prod_{i<j}^{T} (\sigma_i^2 - \sigma_j^2)} \cdot \frac{\det(\mathbf{M})}{\prod_{i<j}^{M} (d_i^2 - d_j^2)}. \quad (58)$$

We remark that (58) holds under the assumption that $T \leq N$, which ensures that the $\{d_i\}_{i=1}^{M}$ are all distinct.

When $T > N$, we have that $d_1 = \cdots = d_l = \sqrt{TN/\underline{L}}$, where $l = T - \overline{L} = \min\{M, T-N\}$ (see Section III-C). Hence, $\lambda_1 = \cdots = \lambda_l = \lambda \triangleq [\underline{L}/(TN\tilde{\rho}) + 1]^{-1}$. Let in this case

$$\boldsymbol{\Lambda}'_\epsilon \triangleq \text{diag}\{[\lambda + \epsilon'_1 \cdots \lambda + \epsilon'_l \; \lambda_{l+1} \cdots \lambda_M \; \epsilon'_{M+1} \cdots \epsilon'_T]^{\mathsf{T}}\}$$

where $\epsilon'_1, \ldots, \epsilon'_l$ and $\epsilon'_{M+1}, \ldots, \epsilon'_T$ are positive real numbers chosen so that the diagonal elements of $\boldsymbol{\Lambda}'_\epsilon$ are distinct. Let also $\boldsymbol{\Delta}_\epsilon \triangleq \text{diag}\{[\sigma_1^2 \cdots \sigma_N^2 \; \epsilon_{N+1} \cdots \epsilon_T]^{\mathsf{T}}\}$, where $\epsilon_{N+1}, \ldots, \epsilon_T$ are positive real numbers chosen so that the diagonal elements of $\boldsymbol{\Delta}_\epsilon$ are distinct. To obtain $\mathsf{f}_{\mathbf{Y}\,|\,\mathbf{D}}$, we need to replace (55) with (59) on the top of next page, and then follow the same steps leading to (58). The corresponding steps are omitted. For simplicity, in the remainder of the proof we shall focus exclusively on the case $T \leq N$.

B. Step 2

1) Computing $\mathsf{f}_{\tilde{\boldsymbol{\sigma}}\,|\,\mathbf{D}}^{(\rho)}$: To obtain $\mathsf{f}_{\tilde{\boldsymbol{\sigma}}\,|\,\mathbf{D}}^{(\rho)}$ from $\mathsf{f}_{\mathbf{Y}\,|\,\mathbf{D}}$, we express $\mathbf{Y}$ in terms of its SVD [see (21)], which yields

$$\mathsf{f}_{\mathbf{U},\boldsymbol{\Sigma},\mathbf{V}\,|\,\mathbf{D}}(\mathbf{U},\boldsymbol{\Sigma},\mathbf{V}\,|\,\mathbf{D})$$
$$= \mathsf{f}_{\mathbf{Y}\,|\,\mathbf{D}}(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}\,|\,\mathbf{D}) \cdot J_{N,T}(\sigma_1, \cdots, \sigma_T) \quad (60)$$

where $J_{N,T}$ is the Jacobian of the SVD transformation given in (24) (recall that we assumed $T \leq N$, and, hence, $\underline{P} = T$ and $\overline{P} = N$).

Next, we integrate the RHS of (60) over $\mathbf{U}$ and $\mathbf{V}$ and then operate the change of variable $\boldsymbol{\sigma} \mapsto \tilde{\boldsymbol{\sigma}}$ defined in (41). These two steps yield

$$\mathsf{f}_{\tilde{\boldsymbol{\sigma}}\,|\,\mathbf{D}}^{(\rho)}(\tilde{\boldsymbol{\sigma}}\,|\,\mathbf{D})$$
$$= \frac{2^T e^{-\sum_{i=M+1}^{T} \tilde{\sigma}_i^2}}{\prod_{i=N-T+1}^{N} \Gamma(i) \cdot \prod_{i=1}^{T-M} \Gamma(i)} \cdot \frac{\det(\tilde{\rho}^{-1}\mathbf{D}^{-2} + \mathbf{I}_M)^{T-M}}{\det(\mathbf{D}^2 + \tilde{\rho}^{-1}\mathbf{I}_M)^{N-M+1}}$$
$$\cdot \prod_{i<j}^{M} (\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2) \cdot \prod_{M<i<j}^{T} (\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2) \cdot \prod_{i=1}^{M} \prod_{j=M+1}^{T} \left(\tilde{\sigma}_i^2 - \frac{\tilde{\sigma}_j^2}{\tilde{\rho}}\right)$$

$$\int_{\mathcal{U}(T)} e^{\mathrm{tr}(\mathbf{\Delta\Phi\Lambda\Phi}^{\mathsf{H}})}d\widetilde{\mathbf{\Phi}} = \lim_{\epsilon_{N+1}\to 0,\dots,\epsilon_T\to 0}\ \lim_{\epsilon_1'\to 0,\dots,\epsilon_l'\to 0}\ \lim_{\epsilon_{M+1}'\to 0,\dots,\epsilon_T'\to 0}\int_{\mathcal{U}(T)} e^{\mathrm{tr}(\mathbf{\Delta}_\epsilon\widetilde{\mathbf{\Phi}}\mathbf{\Lambda}_\epsilon'\widetilde{\mathbf{\Phi}}^{\mathsf{H}})}d\widetilde{\mathbf{\Phi}} \tag{59}$$

$$\cdot \underbrace{e^{-\tilde\rho\sum_{i=1}^M\tilde\sigma_i^2}\det(\mathbf{M})}_{\triangleq c_\rho(\tilde{\boldsymbol{\sigma}})} \cdot \frac{\prod_{i=1}^T \tilde\sigma_i^{2(N-T)+1}}{\prod_{i<j}^M (d_i^2 - d_j^2)}. \tag{61}$$

2) Convergence of $\mathsf{f}_{\tilde{\boldsymbol{\sigma}}\,|\,\mathbf{D}}^{(\rho)}$ to $\mathsf{f}_{\mathbf{u}\,|\,\mathbf{D}}$ as $\rho\to\infty$: We start by characterizing the limit $\rho\to\infty$ of $c_\rho(\tilde{\boldsymbol{\sigma}})$ in (61). Let $\mathbf{L}$ be a $T\times T$ matrix defined as in (62) on the top of next page. Observe now that $c_\rho(\tilde{\boldsymbol{\sigma}}) = \det(\mathbf{L})$ and that $\mathbf{L}_{21}$ vanishes as $\rho\to\infty$. These two facts imply that

$$\lim_{\rho\to\infty} c_\rho(\tilde{\boldsymbol{\sigma}}) = \lim_{\rho\to\infty}\det(\mathbf{L})$$
$$= \lim_{\rho\to\infty}\Big(\det(\mathbf{L}_{11})\det(\mathbf{L}_{22})\Big)$$
$$= \det(\widetilde{\mathbf{L}})\cdot\prod_{M<i<j}^T(\tilde\sigma_i^2 - \tilde\sigma_j^2) \tag{63}$$

with $\widetilde{\mathbf{L}}$ being a $M\times M$ matrix defined by $\big[\widetilde{\mathbf{L}}\big]_{i,j} = e^{-\tilde\sigma_j^2/d_i^2}$. Substituting (63) into (61), we get after some algebraic manipulations

$$\lim_{\rho\to\infty}\mathsf{f}_{\tilde{\boldsymbol{\sigma}}\,|\,\mathbf{D}}^{(\rho)}(\tilde{\boldsymbol{\sigma}}\,|\,\mathbf{D})$$
$$= \underbrace{\frac{2^M\det(\widetilde{\mathbf{L}})\cdot\prod_{i=1}^M\tilde\sigma_i^{2(N-M)+1}}{\det(\mathbf{D})^{2N}\cdot\prod_{i=N-M+1}^N\Gamma(i)}\cdot\prod_{i<j}^M\frac{\tilde\sigma_i^2-\tilde\sigma_j^2}{d_i^2-d_j^2}d_i^2 d_j^2}_{\triangleq \mathsf{f}_1(\tilde\sigma_1,\dots,\tilde\sigma_M)}$$
$$\cdot\underbrace{2^{T-M}e^{-\sum_{i=M+1}^T\tilde\sigma_i^2}\cdot\frac{\prod_{i=M+1}^T\tilde\sigma_i^{2(N-T)+1}\cdot\prod_{M<i<j}^T(\tilde\sigma_i^2-\tilde\sigma_j^2)^2}{\prod_{i=1}^{T-M}\Gamma(i)\cdot\prod_{i=N-T+1}^{N-M}\Gamma(i)}}_{\triangleq \mathsf{f}_2(\tilde\sigma_{M+1},\dots,\tilde\sigma_T)} \tag{64}$$

The proof of part 2 is concluded by noting that

$$\lim_{\rho\to\infty}\mathsf{f}_{\tilde{\boldsymbol{\sigma}}\,|\,\mathbf{D}}^{(\rho)}(\tilde{\boldsymbol{\sigma}}\,|\,\mathbf{D}) = \mathsf{f}_1(\tilde\sigma_1,\dots,\tilde\sigma_M)\cdot\mathsf{f}_2(\tilde\sigma_{M+1},\dots,\tilde\sigma_T)$$
$$= \mathsf{f}_{\mathbf{u}\,|\,\mathbf{D}}(\tilde{\boldsymbol{\sigma}}\,|\,\mathbf{D})$$

where the last equality follows from [22, Thms. 2.17 and 2.18].

C. Step 3

We next establish that the function $\mathsf{f}_{\tilde{\boldsymbol{\sigma}}\,|\,\mathbf{D}}^{(\rho)}(\tilde{\boldsymbol{\sigma}}\,|\,\mathbf{D})\cdot\mathsf{q}_{\mathbf{D}}^{\mathrm{opt},\rho}(\mathbf{D})$ is bounded. By (40), this is obviously true for the case when $\mathbf{D}\in\widetilde{\mathcal{K}}(\rho_0)$. We analyze next the case $\mathbf{D}\notin\widetilde{\mathcal{K}}(\rho_0)$. Throughout this appendix, we shall use k to indicate a constant term that does not depend on $\tilde{\boldsymbol{\sigma}}$, $\mathbf{D}$, and ρ. Note that k can take on different values at each appearance. We start by observing that for an arbitrary

$\zeta\in(0,1)$, there exists a $\rho_{\mathrm{th}} > 0$ such that $\mathbb{P}\{\mathbf{D}\in\widetilde{\mathcal{K}}(\rho_0),\mathbf{D}\sim\mathsf{Q}_{\mathbf{D}}^{\mathrm{opt}}\}\le\zeta$ for all $\rho > \rho_{\mathrm{th}}$. Hence, for $\rho > \rho_{\mathrm{th}}$,

$$\mathsf{q}_{\mathbf{D}}^{\mathrm{opt},\rho}(\mathbf{D}) = \frac{\mathsf{q}_{\mathbf{D}}^{\mathrm{opt}}(\mathbf{D})}{1 - \mathbb{P}\{\mathbf{D}\in\widetilde{\mathcal{K}}(\rho_0),\mathbf{D}\sim\mathsf{Q}_{\mathbf{D}}^{\mathrm{opt}}\}} \le \frac{\mathsf{q}_{\mathbf{D}}^{\mathrm{opt}}(\mathbf{D})}{1-\zeta}. \tag{65}$$

Since we are interested in the limit $\rho\to\infty$, we will assume throughout that $\rho > \rho_{\mathrm{th}}$, so that (65) holds. Let $d^2\triangleq TN/\underline{L}$. It follows from (12) and from the change of variable theorem that

$$\mathsf{q}_{\mathbf{D}}^{\mathrm{opt}}(\mathbf{D}) = \mathsf{f}_{\lambda_1,\dots,\lambda_M}\left(\frac{d_1^2}{d^2},\dots,\frac{d_M^2}{d^2}\right)\cdot\prod_{i=1}^M\frac{2d_i}{d}$$
$$= k\cdot\prod_{i=1}^M\left(\frac{d_i}{d}\right)^{2(T-2M)+1}\left(1-\frac{d_i^2}{d^2}\right)^{N-T}$$
$$\cdot\prod_{i<j}^M\left(\frac{d_i^2-d_j^2}{d^2}\right)^2. \tag{66}$$

Here, the second equality follows by setting $m = M$, $p = T - M$, and $n = M+N-T$ in (12). Substituting (66) into (65) we obtain

$$\mathsf{f}_{\tilde{\boldsymbol{\sigma}}\,|\,\mathbf{D}}^{(\rho)}(\tilde{\boldsymbol{\sigma}}\,|\,\mathbf{D})\cdot\mathsf{q}_{\mathbf{D}}^{\mathrm{opt},\rho}(\mathbf{D})$$
$$\le k\cdot\frac{\det\big(\tilde\rho^{-1}\mathbf{D}^{-2}+\mathbf{I}_M\big)^{T-M}}{\det(\mathbf{D}^2+\tilde\rho^{-1}\mathbf{I}_M)^{N-M+1}}\cdot\det(\mathbf{M})e^{-\tilde\rho\sum_{i=1}^M\tilde\sigma_i^2}$$
$$\cdot\underbrace{\prod_{i=1}^M\tilde\sigma_i^{2(N-T)+1}\cdot\prod_{i<j}^M(\tilde\sigma_i^2-\tilde\sigma_j^2)\cdot\prod_{i=1}^M\prod_{j=M+1}^T\left(\tilde\sigma_i^2-\frac{\tilde\sigma_j^2}{\tilde\rho}\right)}_{\le\prod_{i=1}^M\tilde\sigma_i^{2(N-i)+1}}$$
$$\cdot e^{-\sum_{i=M+1}^T\tilde\sigma_i^2}\cdot\underbrace{\prod_{i=M+1}^T\tilde\sigma_i^{2(N-T)+1}\cdot\prod_{M<i<j}^T({\tilde\sigma_i}^2-{\tilde\sigma_j}^2)}_{\le\prod_{i=M+1}^T\tilde\sigma_i^{2(N-i)+1}}$$
$$\cdot\underbrace{\prod_{i<j}^M(d_i^2-d_j^2)\cdot\prod_{i=1}^M d_i^{2(T-2M)+1}(d^2-d_i^2)^{N-T}}_{\le k\cdot\prod_{i=1}^M d_i}$$
$$\le k\cdot\underbrace{\prod_{i=1}^M\Big[1+1/(\tilde\rho d_i^2)\Big]^{T-M}}_{\le(1+M/\rho_0)^{M(T-M)}=k}\cdot\frac{\det(\mathbf{M})\prod_{i=1}^T\tilde\sigma_i^{2(N-i)+1}}{\prod_{i=1}^M(\tilde\rho^{-1}+d_i^2)^{N-M+1/2}}$$
$$\cdot\underbrace{\prod_{i=1}^M\frac{d_i}{(\tilde\rho^{-1}+d_i^2)^{1/2}}}_{\le 1}\cdot\exp\left(-\tilde\rho\sum_{i=1}^M\tilde\sigma_i^2-\sum_{i=M+1}^T\tilde\sigma_i^2\right)$$
$$\le k\cdot\det(\mathbf{N}) \tag{67}$$

$$\mathbf{L} \triangleq \mathbf{M} \cdot \mathrm{diag}\Big\{[e^{-\tilde{\rho}\tilde{\sigma}_1^2} \cdots e^{-\tilde{\rho}\tilde{\sigma}_M^2} \underbrace{1 \cdots 1}_{T-M}]^\mathsf{T}\Big\}$$

$$= \left(\begin{array}{ccc|ccc}
e^{-\frac{\tilde{\rho}}{1+\tilde{\rho}d_1^2}\tilde{\sigma}_1^2} & \cdots & e^{-\frac{\tilde{\rho}}{1+\tilde{\rho}d_1^2}\tilde{\sigma}_M^2} & e^{\lambda_1\tilde{\sigma}_{M+1}^2} & \cdots & e^{\lambda_1\tilde{\sigma}_T^2} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
e^{-\frac{\tilde{\rho}}{1+\tilde{\rho}d_M^2}\tilde{\sigma}_1^2} & \cdots & e^{-\frac{\tilde{\rho}}{1+\tilde{\rho}d_M^2}\tilde{\sigma}_M^2} & e^{\lambda_M\tilde{\sigma}_{M+1}^2} & \cdots & e^{\lambda_M\tilde{\sigma}_T^2} \\
(\tilde{\rho}\tilde{\sigma}_1^2)^{T-M-1}\cdot e^{-\tilde{\rho}\tilde{\sigma}_1^2} & \cdots & (\tilde{\rho}\tilde{\sigma}_M^2)^{T-M-1}\cdot e^{-\tilde{\rho}\tilde{\sigma}_M^2} & \tilde{\sigma}_{M+1}^{2(T-M-1)} & \cdots & \tilde{\sigma}_T^{2(T-M-1)} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
e^{-\tilde{\rho}\tilde{\sigma}_1^2} & \cdots & e^{-\tilde{\rho}\tilde{\sigma}_M^2} & 1 & \cdots & 1
\end{array}\right)$$

$$= \begin{pmatrix} \mathbf{L}_{11} & \mathbf{L}_{12} \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{pmatrix}. \tag{62}$$

where the $T \times T$ matrix $\mathbf{N}$ is defined as follows:

$$\mathbf{N} \triangleq \mathrm{diag}\Big\{\Big[(\tilde{\rho}^{-1} + d_1^2)^{-(N-M+1/2)}$$
$$\cdots (\tilde{\rho}^{-1} + d_M^2)^{-(N-M+1/2)}, \underbrace{1 \cdots 1}_{T-M}\Big]^\mathsf{T}\Big\}$$
$$\cdot \mathbf{M} \cdot \mathrm{diag}\Big\{\Big[e^{-\tilde{\rho}\tilde{\sigma}_1^2} \cdots e^{-\tilde{\rho}\tilde{\sigma}_M^2}\, e^{-\tilde{\sigma}_{M+1}^2} \cdots e^{-\tilde{\sigma}_T^2}\Big]^\mathsf{T}\Big\}$$
$$\cdot \mathrm{diag}\Big\{\Big[\tilde{\sigma}_1^{2(N-1)+1} \cdots \tilde{\sigma}_T^{2(N-T)+1}\Big]^\mathsf{T}\Big\}.$$

Next, we upper-bound $\det(\mathbf{N})$ by bounding the entries $n_{i,j}$ of $\mathbf{N}$, which are given by

$$n_{i,j} = \begin{cases} \dfrac{\exp\left[-\left(d_i^2 + \tilde{\rho}^{-1}\right)^{-1}\tilde{\sigma}_j^2\right]}{\left(d_i^2 + \tilde{\rho}^{-1}\right)^{N-M+1/2}} \cdot \tilde{\sigma}_j^{2(N-j)+1}, \\ \qquad\qquad 1 \le i \le M,\ 1 \le j \le M \\[2mm] \dfrac{\exp\left[-(1 + \tilde{\rho}d_i^2)^{-1}\tilde{\sigma}_j^2\right]}{\left(d_i^2 + \tilde{\rho}^{-1}\right)^{N-M+1/2}} \cdot \tilde{\sigma}_j^{2(N-j)+1}, \\ \qquad\qquad 1 \le i \le M,\ M < j \le T \\[2mm] \left(\tilde{\rho}\tilde{\sigma}_j^2\right)^{T-i} \cdot e^{-\tilde{\rho}\tilde{\sigma}_j^2} \cdot \tilde{\sigma}_j^{2(N-j)+1}, \\ \qquad\qquad M < i \le T,\ 1 \le j \le M \\[2mm] \tilde{\sigma}_j^{2(T+N-i-j)+1} \cdot e^{-\tilde{\sigma}_j^2},\ M < i \le T,\ M < j \le T. \end{cases}$$

In the following, we shall repeatedly make use of the fact that the function $f(x) = e^{-\beta x^2} x^\alpha$ with $\alpha, \beta, x > 0$ is maximized for $x = x^* = \sqrt{\alpha/(2\beta)}$, and that the corresponding maximum values is $f(x^*) = \left[\alpha/(2\beta e)\right]^{\alpha/2}$. This implies that

$$f(x) = e^{-\beta x^2} x^\alpha \le \left(\frac{\alpha}{2\beta e}\right)^{\alpha/2}, \quad x > 0. \tag{68}$$

1) *Case* $1 \le i \le M,\ 1 \le j \le M$:

$$n_{i,j} \overset{(a)}{\le} k \cdot \left(d_i^2 + \tilde{\rho}^{-1}\right)^{N-j+1/2} \left(d_i^2 + \tilde{\rho}^{-1}\right)^{-(N-M+1/2)}$$
$$= k \cdot \left(d_i^2 + \tilde{\rho}^{-1}\right)^{M-j}$$
$$\overset{(b)}{\le} k \cdot \left(d^2 + \frac{M}{\rho_{\mathrm{th}}}\right)^{M-j} = k.$$

Here, (a) follows from (68) by setting $\alpha = 2(N-j)+1$, $\beta = (d_i^2 + \tilde{\rho}^{-1})^{-1}$, and $x = \tilde{\sigma}_j$; (b) follows because $d_i \le d$ and $j \le M$.

2) *Case* $1 \le i \le M,\ M < j \le T$:

$$n_{i,j} = \underbrace{\exp\left[-(1 + \tilde{\rho}d_i^2)^{-1}\tilde{\sigma}_j^2\right]}_{\le 1} \cdot \tilde{\sigma}_j^{2(N-j)+1}$$
$$\cdot \underbrace{\left(d_i^2 + \tilde{\rho}^{-1}\right)^{-(N-M+1/2)}}_{\overset{(a)}{\le} \left(\frac{\tilde{\rho}}{1+\rho_0/M}\right)^{N-M+1/2}}$$
$$\le k \cdot \tilde{\sigma}_j^{2(N-j)+1}\tilde{\rho}^{N-M+1/2}. \tag{69}$$

Here, (a) follows because $d_i^2\rho \ge \rho_0$.

3) *Case* $M < i \le T,\ 1 \le j \le M$:

$$n_{i,j} = \tilde{\rho}^{T-i}e^{-\tilde{\rho}\tilde{\sigma}_j^2}\tilde{\sigma}_j^{2(T+N-i-j)+1}$$
$$\overset{(a)}{\le} k \cdot \tilde{\rho}^{T-i} \cdot \left(\frac{1}{\tilde{\rho}}\right)^{T+N-i-j+1/2}$$
$$= k \cdot \tilde{\rho}^{-(N-j+1/2)}$$
$$\le k \cdot \rho_{\mathrm{th}}^{-(N-j+1/2)} = k.$$

Here, (a) follows from (68) by setting $\alpha = 2(T+N-i-j)+1$, $\beta = \tilde{\rho}$, and $x = \tilde{\sigma}_j$.

4) *Case* $M < i \le T,\ M < j \le T$: We have $n_{i,j} \le k$, which follows directly from (68) by setting $\alpha = 2(T+N-i-j)+1$, $\beta = 1$, and $x = \tilde{\sigma}_j$.

To show that $\det(\mathbf{N})$ is bounded, it remains to further analyze case 2, where $n_{i,j}$ is not bounded. Let $\{i_1, \ldots, i_T\}$ be an arbitrary permutation of $\{1, \ldots, T\}$. Then [21, Sec. 0.3]

$$\det(\mathbf{N}) = \sum_{(i_1, \ldots, i_T)} \mathrm{sgn}(i_1, \ldots, i_T) \cdot n_{i_1,1} \cdot \ldots \cdot n_{i_T,T}. \tag{70}$$

Here, the sum is over all the $(T!)$ permutations of $\{1, \ldots, T\}$ and $\mathrm{sgn}(\cdot)$ denotes the sign of the permutation [21, p. 8]. We observe that for each $n_{i,j}$ $(1 \le i \le M, M < j \le T)$ that appears in the product on the RHS of (70), there exists a factor

$n_{i',j'}$ in the same product with $M < i' \le T$, $1 \le j' \le M$ (case 3). Note now that

$$n_{i,j} \cdot n_{i',j'}$$
$$\overset{(a)}{\le} k \cdot \underbrace{\tilde{\sigma}_j^{2(N-j)+1}}_{\overset{(b)}{\le} \left(\tilde{\rho}\tilde{\sigma}_{j'}^2\right)^{N-j+1/2}} \cdot \tilde{\rho}^{N-M+1/2}$$
$$\cdot \left(\tilde{\rho}\tilde{\sigma}_{j'}^2\right)^{T-i'} \cdot e^{-\tilde{\rho}\tilde{\sigma}_{j'}^2} \cdot \tilde{\sigma}_{j'}^{2(N-j')+1}$$
$$\le k \cdot \tilde{\rho}^{2N+T-M+1-j-i'} \cdot \underbrace{\tilde{\sigma}_{j'}^{2(2N+T+1-j-i'-j')} e^{-\tilde{\rho}\tilde{\sigma}_{j'}^2}}_{\overset{(c)}{\le} k \cdot \tilde{\rho}^{-(2N+T+1-j-i'-j')}}$$
$$\le k \cdot \tilde{\rho}^{-(M-j')}$$
$$\le k \cdot \rho_{\text{th}}^{-(M-j')} = k.$$

Here, (a) follows from (69), in (b) we used that $\tilde{\rho}\tilde{\sigma}_{j'}^2 = \sigma_{j'}^2 \ge \sigma_j^2 = \tilde{\sigma}_j^2$ for $j' \le M < j$, and (c) follows from (68) by setting $\alpha = 2(2N+T+1-j-i'-j')$, $\beta = \tilde{\rho}$, and $x = \tilde{\sigma}_{j'}$.

Summarizing, we showed that

$$\left| f_{\tilde{\boldsymbol{\sigma}}|\mathbf{D}}^{(\rho)}(\tilde{\boldsymbol{\sigma}}\,|\,\mathbf{D}) \cdot q_{\mathbf{D}}^{\text{opt},\rho}(\mathbf{D}) \right| \le k \cdot \det(\mathbf{N}) \le k$$

which concludes the proof of the lemma.

APPENDIX B
PROOF OF LEMMA 13

Throughout this appendix, we shall set $d \triangleq \sqrt{NT/\underline{L}}$, and $\tilde{\rho} \triangleq \rho/M$, and denote by $q_{\mathbf{D}}^{\text{opt},\rho}$ and $q_{\mathbf{D}}^{\text{opt}}$ the pdfs corresponding to the probability distributions $Q_{\mathbf{D}}^{\text{opt},\rho}$ and $Q_{\mathbf{D}}^{\text{opt}}$, respectively, by $f_{\tilde{\boldsymbol{\sigma}}}^{(\rho)}$ and $f_{\mathbf{u}}$ the pdf of $\tilde{\boldsymbol{\sigma}}$ and $\mathbf{u}$, respectively, and by $f_{\tilde{\boldsymbol{\sigma}}|\mathbf{D}}^{(\rho)}$ and $f_{\mathbf{u}|\mathbf{D}}$ the conditional pdf of $\tilde{\boldsymbol{\sigma}}$ and $\mathbf{u}$ given $\mathbf{D}$, respectively. We shall use k to denote a finite constant; its value might change at every appearance. Since the lemma only addresses limiting behaviors as $\rho \to \infty$, we shall assume throughout that $\rho > \rho_{\text{th}} > 0$. Finally, for simplicity we shall focus exclusively on the case $T \le N$; the proof for the case $T > N$ follows from analogous steps.

A. Proof of Part 1

The proof is based on the following theorem.

Theorem 14 ([34, Thm. 1]): Let $\{\mathbf{x}_i \in \mathbb{C}^m\}$ be a sequence of random vectors with pdfs f_i and let $\mathbf{x} \in \mathbb{C}^m$ be a random vector with pdf f. Assume that f_i converges to f pointwise. If there exist i) a finite constant $F > 0$ such that $\max\{\sup_{\mathbf{x}} f_i(\mathbf{x}), \sup_{\mathbf{x}} f(\mathbf{x})\} \le F$ for all i, and ii) a finite constant $L > 0$ such that $\max\{\int \|\mathbf{x}\|^\kappa f_i(\mathbf{x})d\mathbf{x}, \int \|\mathbf{x}\|^\kappa f(\mathbf{x})d\mathbf{x}\} \le L$ for some $\kappa > 1$ and all i, then $h(\mathbf{x}_i) \to h(\mathbf{x})$.

Since we established in Appendix A that $f_{\tilde{\boldsymbol{\sigma}}}^{(\rho)}$ converges to $f_{\mathbf{u}}$ pointwise as $\rho \to \infty$, we just need to verify that both $f_{\tilde{\boldsymbol{\sigma}}}^{(\rho)}$ and $f_{\mathbf{u}}$ satisfy the conditions i) and ii) in Theorem 14.

1) $f_{\tilde{\boldsymbol{\sigma}}}^{(\rho)}$ and $f_{\mathbf{u}}$ are bounded: Because of (51), and since $0 < d_i \le d$, we have that

$$f_{\tilde{\boldsymbol{\sigma}}}^{(\rho)}(\tilde{\boldsymbol{\sigma}}) = \int f_{\tilde{\boldsymbol{\sigma}}|\mathbf{D}}^{(\rho)}(\tilde{\boldsymbol{\sigma}}|\mathbf{D}) q_{\mathbf{D}}^{\text{opt},\rho}(\mathbf{D}) d\mathbf{D} \le k \cdot d^M = k.$$

To show that $f_{\mathbf{u}}$ is bounded, we first prove that $f_{\mathbf{u}|\mathbf{D}} \cdot q_{\mathbf{D}}^{\text{opt}}$ is bounded by using (64) and (66):

$$f_{\mathbf{u}|\mathbf{D}}(\mathbf{u}\,|\,\mathbf{D}) \cdot q_{\mathbf{D}}^{\text{opt}}(\mathbf{D})$$
$$= k \cdot \frac{\det(\widetilde{\mathbf{L}})}{\prod_{i=1}^{M} d_i^{2(N-M+1)}} \cdot \underbrace{\prod_{i=1}^{M} u_i^{2(N-M)+1} \cdot \prod_{i<j} \left(u_i^2 - u_j^2\right)}_{\le \prod_{i=1}^{M} u_i^{2(N-i)+1}}$$
$$\cdot e^{-\sum_{i=M+1}^{T} u_i^2} \cdot \underbrace{\prod_{i=M+1}^{T} u_i^{2(N-T)+1} \cdot \prod_{M<i<j} \left(u_i^2 - u_j^2\right)^2}_{\le \prod_{i=M+1}^{T} u_i^{2(N+T-2i)+1}}$$
$$\cdot \underbrace{\prod_{i<j}^{M} \left(d_i^2 - d_j^2\right) \cdot \left(\prod_{i=1}^{M} d_i^{2(T-2M)+1} \left(d^2 - d_i^2\right)^{N-T}\right)}_{\le k \cdot \prod_{i=1}^{M} d_i}$$
$$\le k \cdot \det(\widetilde{\mathbf{L}}) \cdot \underbrace{\prod_{i=1}^{M} d_i^{-(2(N-M)+1)} \cdot \prod_{j=1}^{M} u_j^{2(N-j)+1}}_{\overset{(a)}{\le} k}$$
$$\cdot \prod_{i=M+1}^{T} \underbrace{e^{-u_i^2} u_i^{2(N+T-2i)+1}}_{\overset{(b)}{\le} k}$$
$$\le k.$$

Here, (a) follows from (68) with the choice $\alpha = 2(N-j)+1$, $\beta = d_i^{-2}$, and $x = u_j$, as detailed below

$$e^{-u_j^2/d_i^2} \cdot d_i^{-(2(N-M)+1)} \cdot u_j^{2(N-j)+1}$$
$$\le k \cdot d_i^{2(M-j)} \le k \cdot d^{2(M-j)} = k$$

for all $1 \le i, j \le M$; (b) follows again from (68). Thus

$$f_{\mathbf{u}}(\mathbf{u}) = \int f_{\mathbf{u}|\mathbf{D}}(\mathbf{u}\,|\,\mathbf{D}) \cdot q_{\mathbf{D}}^{\text{opt}}(\mathbf{D}) d\mathbf{D} \le k \cdot d^M = k.$$

2) $f_{\tilde{\boldsymbol{\sigma}}}^{(\rho)}$ and $f_{\mathbf{u}}$ have finite second moment: We take $\kappa = 2$, and obtain

$$\int \|\tilde{\boldsymbol{\sigma}}\|^2 f_{\tilde{\boldsymbol{\sigma}}}^{(\rho)}(\tilde{\boldsymbol{\sigma}}) d\tilde{\boldsymbol{\sigma}} = \frac{1}{\tilde{\rho}} \sum_{i=1}^{M} \mathbb{E}\left[\sigma_i^2\right] + \sum_{i=M+1}^{P} \mathbb{E}\left[\sigma_i^2\right]$$
$$\overset{(a)}{\le} \frac{1}{\tilde{\rho}} NT(\rho+1) + (N-M)(T-M)$$
$$\le (1 + 1/\rho_{\text{th}}) MNT + (N-M)(T-M)$$
$$= k. \tag{71}$$

Here, (a) follows from (30) and (31). Furthermore,

$$\int \|\mathbf{u}\|^2 f_{\mathbf{u}}(\mathbf{u}) d\mathbf{u} = \mathbb{E}\left[\text{tr}\left(\mathbf{D}\mathbf{H}\mathbf{H}^{\mathsf{H}}\mathbf{D}\right)\right] + (N-M)(T-M)$$
$$= k.$$

This concludes the proof.

B. Proof of Part 2

Let $0 < \delta < 1$ and let r be a positive integer satisfying $r > 1$. Denote by $f_{\tilde{\sigma}_i}^{(\rho)}$ the pdf of $\tilde{\sigma}_i$ and by f_{u_i} the pdf of u_i. The expectation on Lemma 13–Part-2 can be rewritten as follows:

$$\lim_{\rho \to \infty} \mathbb{E}_{f_{\tilde{\sigma}_i}^{(\rho)}}[\log(x)] = \lim_{\rho \to \infty} \Big\{ \mathbb{E}_{f_{\tilde{\sigma}_i}^{(\rho)}}[\log(x) \cdot I\{x < \delta\}]$$
$$+ \mathbb{E}_{f_{\tilde{\sigma}_i}^{(\rho)}}[\log(x) \cdot I\{\delta \le x \le r\}]$$
$$+ \mathbb{E}_{f_{\tilde{\sigma}_i}^{(\rho)}}[\log(x) \cdot I\{x > r\}] \Big\} \quad (72)$$

where $I\{\cdot\}$ is the indicator function. We analyze the three terms on the RHS of (72) separately. For the first term, [10, Lemma 6.7(a)] and Lemma 13–Part 1 imply that

$$\lim_{\rho \to \infty} \mathbb{E}_{f_{\tilde{\sigma}_i}^{(\rho)}}[\log(x) \cdot I\{x < \delta\}] = \epsilon_1(\delta)$$

where $\epsilon_1(\delta) \to 0$ as $\delta \to 0$. For the second term, we have that

$$\lim_{\rho \to \infty} \mathbb{E}_{f_{\tilde{\sigma}_i}^{(\rho)}}[\log(x) \cdot I\{\delta \le x \le r\}]$$
$$= \mathbb{E}_{f_{u_i}}[\log(x) \cdot I\{\delta \le x \le r\}]$$

as a consequence of the dominated convergence theorem. Finally, for the third term we proceed as follows:

$$\mathbb{E}_{f_{\tilde{\sigma}_i}^{(\rho)}}[\log(x) \cdot I\{x > r\}] = \sum_{l=r}^{\infty} \int_{l}^{l+1} f_{\tilde{\sigma}_i}^{(\rho)}(x) \log(x) dx$$
$$\overset{(a)}{\le} \sum_{l=r}^{\infty} \int_{l}^{l+1} f_{\tilde{\sigma}_i}^{(\rho)}(x) \sqrt{x} \, dx$$
$$\le \sum_{l=r}^{\infty} \sqrt{l+1} \int_{l}^{l+1} f_{\tilde{\sigma}_i}^{(\rho)}(x) \, dx$$
$$\le \sum_{l=r}^{\infty} \sqrt{l+1} \int_{l}^{\infty} f_{\tilde{\sigma}_i}^{(\rho)}(x) \, dx$$
$$\overset{(b)}{\le} \sum_{l=r}^{\infty} \left[\sqrt{l+1} \cdot \frac{\mathbb{E}_{f_{\tilde{\sigma}_i}^{(\rho)}}[x^2]}{l^2} \right]$$
$$\overset{(c)}{\le} \sum_{l=r}^{\infty} \left[\sqrt{l+1} \cdot \frac{k}{l^2} \right]$$
$$\le \sqrt{2} k \sum_{l=r}^{\infty} l^{-3/2}. \quad (73)$$

Here, (a) follows because $\log(x) \le \sqrt{x}$, $x \ge 1$, (b) follows from Markov's inequality, and (c) is a consequence of (71). Note that (73) holds for all $\rho > \rho_{\text{th}}$. Hence, we have

$$0 \le \epsilon_2(r) \triangleq \lim_{\rho \to \infty} \mathbb{E}_{f_{\tilde{\sigma}_i}^{(\rho)}}[\log(x) \cdot I\{x > r\}] \le \sqrt{2} k \sum_{l=r}^{\infty} l^{-3/2}.$$

Since $\sum_{l=r}^{\infty} l^{-3/2}$ converges, we can make $\epsilon_2(r)$ arbitrarily close to 0 by choosing r sufficiently large. Summarizing, we showed that

$$\lim_{\rho \to \infty} \mathbb{E}_{f_{\tilde{\sigma}_i}^{(\rho)}}[\log(x)] = \mathbb{E}_{f_{u_i}}[\log(x) \cdot I\{\delta \le x \le r\}]$$
$$+ \epsilon_1(\delta) + \epsilon_2(r).$$

The RHS of this equality can be made arbitrarily close to $\mathbb{E}_{f_{u_i}}[\log(x)]$ by choosing δ sufficiently small and r sufficiently large. This concludes the proof.

C. Proof of Part 3

To establish the desired result, it is sufficient to show that

$$\mathbb{E}[\log(\tilde{\sigma}_i - \tilde{\sigma}_j)] = \mathbb{E}[\log(u_i - u_j)] + o(1), \; \rho \to \infty \quad (74)$$

and that

$$\mathbb{E}[\log(\tilde{\sigma}_i + \tilde{\sigma}_j)] = \mathbb{E}[\log(u_i + u_j)] + o(1), \; \rho \to \infty. \quad (75)$$

Lemma 13–Part 1 implies that, for sufficiently large ρ, $h(\tilde{\sigma}_i - \tilde{\sigma}_j) > -\infty$ and $h(\tilde{\sigma}_i + \tilde{\sigma}_j) > -\infty$; We can now establish (74) and (75) through steps similar to the ones in Part 2.
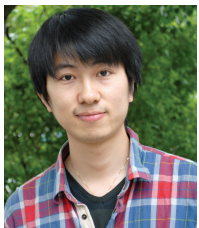
D. Proof of Part 4

The proof is analogous to the proof of part 2 and part 3.

REFERENCES

[1] G. J. Foschini and M. J. Ganz, "On limits of wireless communications in a fading environment when using multiple antennas," *Wirel. Personal Commun.*, vol. 6, pp. 311–335, 1998.

[2] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.

[3] 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.

[4] 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.

[5] A. Lapidoth and S. M. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 437–453, Feb. 2006.

[6] 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.

[7] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.

[8] 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.

[9] 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.

[10] 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.

[11] 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

[12] 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.

[13] 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.

[14] 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.

[15] 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

[16] 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.*, 2012, to appear.

[17] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*, 10th ed. Government Printing Office, 1972.

[18] 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.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, U.S.A.: Wiley, 2006.

[20] W. Yang, G. Durisi, and E. Riegler, "Unitary isotropically distributed inputs are not capacity-achieving for large-MIMO fading channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Boston, MA, U.S.A., Jul. 2012, pp. 1717–1721.

[21] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[22] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," in *Foundations and Trends in Communications and Information Theory*. Delft, The Netherlands: now Publishers, 2004, vol. 1, no. 1, pp. 1–182.

[23] S. K. Mitra, "A density-free approach to the matrix variate Beta distribution," *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, vol. 32, pp. 81–88, Mar. 1970.

[24] R. I. Muirhead, *Aspects of multivariate statistical theory*. Hoboken, NJ, U.S.A.: Wiley, 2005.

[25] H. Uhlig, "On singular Wishart and singular multivariate Beta distribution," *Ann. Statist.*, vol. 22, no. 1, pp. 395–405, 1994.

[26] G. Durisi and H. Bölcskei, "High-SNR capacity of wireless communication channels in the noncoherent setting: A primer," *Int. J. Electron. Commun. (AEÜ)*, vol. 65, no. 8, pp. 707–712, Aug. 2011.

[27] F. Topsøe, "An information theoretical identity and a problem involving capacity," *Studia Scientiarum Math. Hung.*, vol. 2, pp. 291–292, 1967.

[28] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, 3rd ed. Oxford, U.K.: Oxford Univ. Press, 2001.

[29] H. Scheffé, "A useful convergence theorem for probability distributions," *Ann. Math. Stat.*, vol. 18, no. 3, pp. 434–438, 1947.

[30] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY, U.S.A.: McGraw-Hill, 1987.

[31] C. Itzykson and J. B. Zuber, "The planar approximation. II," *J. Math. Phys.*, vol. 21, pp. 411–421, 1980.

[32] A. Onatski, "The Tracy–Widom limit for the largest eigenvalues of singular complex Wishart matrices," *Ann. Appl. Probab.*, vol. 18, no. 2, pp. 470–490, Apr. 2008.

[33] A. Ghaderipoor, C. Tellambura, and A. Paulraj, "On the application of character expansions for MIMO capacity analysis," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2950–2962, May 2012.

[34] M. Godavarti and A. Hero, "Convergence of differential entropies," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 171–176, Jan. 2004.

Giuseppe Durisi (S'02–M'06–SM'12) received the Laurea degree summa cum laude and the Doctor degree both from Politecnico di Torino, Italy, in 2001 and 2006, respectively. From 2002 to 2006, he was with Istituto Superiore Mario Boella, Torino, Italy. From 2006 to 2010 he was a postdoctoral researcher at ETH Zurich, Zurich, Switzerland. Since 2010 he has been an assistant professor at Chalmers University of Technology, Gothenburg, Sweden. He held visiting researcher positions at IMST (Germany), University of Pisa (Italy), and Vienna University of Technology (Austria).

Dr. Durisi is co-author of a paper that won a student paper award at the International Symposium on Information Theory (ISIT 2012). He served as TPC member in several IEEE conferences, and is currently publications editor of the IEEE Transactions on Information Theory. His research interests are in the areas of information theory, communication theory, and compressive sensing.
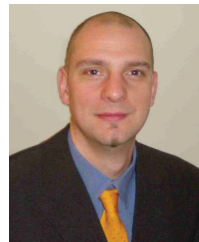
Erwin Riegler (M'07) received the Dipl-Ing. degree in Technical Physics (with distinction) in 2001 and the Dr. techn. degree in Technical Physics (with distinction) in 2004 from Vienna University of Technology.

He was a visiting researcher at the Max Planck Institute for Mathematics in the Sciences in Leipzig, Germany (Sep. 2004 – Feb. 2005), the Communication Theory Group at ETH Zürich, Switzerland (Sep. 2010 – Feb. 2011 and Jun. 2012 – Nov. 2012), and the Department of Electrical and Computer Engineering at The Ohio State University in Columbus, Ohio (Mar. 2012). From 2005 to 2006, he was a post-doctoral fellow at the Institute for Analysis and Scientific Computing, Vienna University of Technology. From 2007 to 2010, he was a senior researcher at the Telecommunications Research Center Vienna (FTW). Since 2010, he has been a post-doctoral fellow at the Institute of Telecommunications at Vienna University of Technology.

His research interests include noncoherent communications, machine learning, interference management, large system analysis, and transceiver design.

Wei Yang (S'09) received the B.E. degree in communication engineering and M.E. degree in communication and information systems from the Beijing University of Posts and Telecommunications, Beijing, China, in 2008 and 2011, respectively. He is currently pursuing a Ph.D. degree in electrical engineering at Chalmers University of Technology, Gothenburg, Sweden. From July to August 2012, he was a visiting student at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA.

Mr. Yang is the recipient of a Student Paper Award at the 2012 IEEE International Symposium on Information Theory (ISIT), Cambridge, MA. His research interests are in the areas of information and communication theory.