Capacity of Differential versus Non-Differential Unitary Space-Time Modulation for MIMO channels

Aris L. Moustakas, *Member, IEEE*, Steven H. Simon and Thomas L. Marzetta, *Fellow, IEEE*

A. L. Moustakas (email: arislm@science.bell-labs.com), S. H. Simon (email: shsimon@bell-labs.com) and T. L. Marzetta (email: tlm@research.bell-labs.com) are with Lucent Technologies.

Abstract

Differential Unitary Space-Time Modulation (DUSTM) and its earlier nondifferential counterpart, USTM, permit high-throughput MIMO communication entirely without the possession of channel state information (CSI) by either the transmitter or the receiver. For an isotropically random unitary input we obtain the exact closed-form expression for the probability density of the DUSTM received signal, which permits the straightforward Monte Carlo evaluation of its mutual information. We compare the performance of DUSTM and USTM through both numerical computations of mutual information and through the analysis of low- and high-SNR asymptotic expressions. In our comparisons the symbol durations of the equivalent unitary space-time signals are both equal to T, as are the number of receive antennas N. For DUSTM the number of transmit antennas is constrained by the scheme to be $M = T/2$, while USTM has no such constraint. If DUSTM and USTM utilize the *same number of transmit antennas* at high SNR's the normalized mutual information of the differential and the nondifferential schemes expressed in bits/sec/Hz are asymptotically equal, with the differential scheme performing somewhat better, while at low SNR's the normalized mutual information of DUSTM is asymptotically twice the normalized mutual information of USTM. If, instead, USTM utilizes the optimum number of transmit antennas then USTM can outperform DUSTM at sufficiently low SNR's.

Index Terms

Non-coherent Communication, Capacity, Space-Time Coding, Multiple Antennas, Differential Encoding, Multiplicative Channels.

I. INTRODUCTION

Considerable volume of work has followed the prediction [1], [2] that the use of multiple antennas in transmitting and receiving signals can result to substantial increases in information throughput. The underlying assumptions of this effort have been that the receiver knows the channel through some training scheme and that the channel coefficients are statistically independent. In this case and for large signal to noise ratio ρ, the capacity is roughly

$$C_{coh} \approx \min(M, N) \log_2 \rho \text{ bits/sec/Hz} \tag{1}$$

where M, N are the numbers of transmitting and receiving antennas.

However, in a typical mobile wireless communication system the channel coefficients can be assumed to be constant over only limited periods of time. Hence, especially for large transmitting antenna numbers, training will require a substantial fraction of the coherence time of the channel and thus hamper the data throughput rates. To address this problem, Marzetta and Hochwald [3], [4] investigated the scenario where the receiver has no a-priori channel knowledge. In addition to the conventional additive Gaussian noise, this channel has also multiplicative noise, corresponding to the channel matrix, which is also assumed to be Gaussian. This is a "non-coherent" channel, as opposed to the additive white Gaussian noise channel with known (and static) "coherent" channel coefficients at the receiver. In an elegant group-theoretic approach, Zheng and Tse [5] found the capacity of this channel to scale as

$$C_{incoh} \approx M^*(1 - M^*/T) \log_2 \rho \text{ bits/sec/Hz} \tag{2}$$

for large ρ, where $M^* = \min(M, N, T/2)$ and T is the number of time intervals over which the channel is static.[1] This implies that for fixed T, there is no need to use more than $M = T/2$ transmitters.

To take advantage of the constancy of the channel over T time intervals, [4] proposed to encode the signal using $T \times M$ isotropic unitary matrices. In this encoding, called isotropic unitary space-time modulation (USTM), a symbol can be spread not only over M antennas, but also over T time intervals. Some analytic results on the mutual information of USTM already exist. In particular, it has been shown that for $T \gg M$ [3] and for $M < \min(N, T/2)$ and large ρ [5] the optimal input distribution is isotropic random unitary, i.e. that of USTM. Thus the asymptotic capacity is equal to the mutual information, as in [2], [3], [5]. More recently, Hassibi and Marzetta [6] analytically calculated the received signal distribution and thus were able to numerically evaluate the mutual information of USTM for a variety of M, T, N and ρ, confirming some of the above asymptotic results.

In the case of USTM it is implicitly assumed that, after T symbols the channel completely changes. However, in practice the channel changes continuously by small amounts. In contrast, differential phase-shift keying (DPSK) [7] has been used extensively to take advantage of the continuous slow-varying nature of the channel, without needing to perform any training. In this scheme, each transmitted symbol is encoded into a phase-difference from the previous symbol.

In [8], [9], the concept of differential modulation was extended to multi-antenna systems. In this method, called differential unitary space-time modulation (DUSTM), the signal is encoded over M transmitting antennas and M time intervals using an $M \times M$ unitary matrix. In each successive M time intervals, the transmitter encodes the input signal by multiplying a $M \times M$ unitary matrix to the unitary matrix transmitted during the previous M time intervals over the M antennas and then transmits the matrix product. In turn, the receiver decodes the signal by comparing the received signal from the M antennas and M time intervals to that received over the previous M time intervals. Thus this scheme requires no training and assumes that the channel is fixed over $T = 2M$ time intervals. The technique of DUSTM can be applied to the mathematically identical space-frequency channel that appears during a single OFDM symbol interval, resulting in a variation called differential unitary space-frequency modulation (DUSFM) [10].

Despite its importance in practical applications [8], no analytic results are available regarding the mutual information of DUSTM and its comparison with USTM for $T = 2M$. The main obstacle has been the difficulty in integrating over exponentials of unitary matrices. This is a problem that was tackled in the 80's by high-energy physicists in analyzing the nuclear strong interactions (quantum chromodynamics). Due to the $SU(3)$ symmetry of these interactions their fluctuations can in certain cases be represented by unitary matrices. Thus to integrate them out, one needs to make use of such integrals of exponentials of unitary matrices. In this paper we apply these results [11] to the context of DUSTM. In particular, we get the following results:

1) We analytically calculate the received signal distribution for the case of DUSTM (see section III).

[1] A similar approach was developed independently by A. M. Sengupta and P. P. Mitra, presented at the first BLAST workshop, Bell Labs, Murray Hill, NJ, February 2000.

2) Using this received signal distribution, we evaluate numerically I_{DUSTM}, the mutual information of DUSTM for a variety of M, N and ρ, and compare it to I_{USTM}, the mutual information of USTM setting $T = 2M$. At low ρ we find that the two mutual informations for the same M, N, $T = 2M$ are nearly identical. This implies that the number of bits *per symbol* i.e. I_{DUSTM}/M is twice $I_{USTM}/T = I_{USTM}/2M$. In contrast, at large ρ the number of bits/symbol of the two schemes approach each other, but with $I_{DUSTM}/M > I_{USTM}/T$.
3) We compare the maximum with respect to M of the two mutual informations per symbol. For fixed M, N, $T = 2M$, we find that while at large ρ we have $\max_{M^* \leq M} I_{DUSTM}(M^*, N, \rho)/M^* > \max_{M^* \leq M} I_{USTM}(M^*, N, \rho, T)/T$, at small ρ the opposite inequality holds.
4) We back the above numerical results by providing perturbative expansions of the mutual information for both small and large ρ.

II. Definitions

A. Notation

Throughout this paper we will denote the number of time-intervals, transmitting antennas and receiving antennas with T, M, N, respectively. R, K and Q will represent $R = \min(M, N)$, $K = \min(T, N)$ and $Q = \max(M, N) - \min(M, N)$.

In addition, we will use bold-faced upper-case letters to represent matrices, e.g. $\mathbf{X}$, with elements given by X_{ij}, bold-faced lower-case letters for column vectors, e.g. $\mathbf{x}$, with elements x_i, and non-bold lower-case letters for scalar quantities. $\mathrm{Tr}\{\mathbf{X}\}$ will represent the trace of $\mathbf{X}$, while the superscripts T and $\dagger$ will indicate transpose and Hermitian conjugate operations. The determinant of a matrix will be represented by $\det(\mathbf{X})$ or by $\det(X_{ij})$. Also, $\mathbf{I}_n$ will denote the n-dimensional identity matrix, while $\mathbf{J}_n$ will represent a $T \times T$ matrix with zeros in all elements other than the first n diagonals, which have unit value.

The complex, circularly symmetric Gaussian distribution with zero-mean and unit-variance will be denoted by $\mathcal{CN}(0, 1)$.

The per-symbol normalized mutual information will be given by $\hat{I}$, measured in bits/sec/Hz. Thus for the case of USTM, $\hat{I}_{USTM} = I_{USTM}/T$, while for DUSTM, $\hat{I}_{DUSTM} = I_{DUSTM}/M$.

B. System Model

We consider the case of single-user transmission from M transmit antennas to N receive antennas over a narrow-band block-fading channel. The channel coefficients are assumed to be constant over time intervals of length T, after which they acquire independent values, which in turn remain constant for the same time interval. The received $T \times N$-dimensional complex signal $\mathbf{X}$ can be written in terms of the $T \times M$-dimensional transmitted complex signal $\mathbf{\Phi}$ as

$$\mathbf{X} = \sqrt{\frac{\rho T}{M}}\mathbf{\Phi}\mathbf{H} + \mathbf{W} \tag{3}$$

where $\mathbf{H}$ is a $M \times N$ matrix with the channel coefficients from the transmitting to the receiving arrays and $\mathbf{W}$ is the $T \times N$ additive noise matrix. Both $\mathbf{H}$ and $\mathbf{W}$ are assumed to have elements that are independent and $\mathcal{CN}(0,1)$-distributed. Their instantaneous values are assumed to be unknown to both the transmitter and the receiver. The first term in (3) is normalized, so that ρ is the total average signal-to-noise ratio (SNR) transmitted from all antennas.

C. Unitary Matrices for Isotropic and Differential USTM

In this paper we will be dealing with unitary input distributions $\mathbf{\Phi}$. For the case of USTM $\mathbf{\Phi}$ is a member of the $S(M,T)$ Stiefel manifold (see [12]) i.e. the set of all complex $T \times M$ matrices, such that

$$\mathbf{\Phi}^\dagger \mathbf{\Phi} = \mathbf{I}_M \tag{4}$$

Note that it is implicitly assumed here that $T \geq M$, since only thus can M T-dimensional vectors be mutually orthogonal.

For the case of DUSTM, we restrict ourselves to the $U(M)$ subgroup of the $S(M,2M)$ Stiefel manifold, such that [8]

$$\mathbf{\Phi} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I}_M \\ \mathbf{U} \end{bmatrix} \tag{5}$$

where $\mathbf{U}$ is an $M \times M$ unitary matrix.

D. Mutual Information

For fixed $\mathbf{\Phi}$, $\mathbf{X}$ in (3) is a sum of two Gaussian matrices, therefore its probability density conditional on $\mathbf{\Phi}$ can be written as

$$\begin{aligned} p(\mathbf{X}|\mathbf{\Phi}) &= \frac{\exp\left(-\mathrm{Tr}\left\{\mathbf{X}^\dagger \left[\mathbf{I}_T + \frac{\rho T}{M}\mathbf{\Phi}\mathbf{\Phi}^\dagger\right]^{-1}\mathbf{X}\right\}\right)}{\pi^{TN}\det(\mathbf{I}_T + \frac{\rho T}{M}\mathbf{\Phi}\mathbf{\Phi}^\dagger)^{\mathbf{N}}} \\ &= \frac{\exp\left(-\mathrm{Tr}\left\{\mathbf{X}^\dagger \left[\mathbf{I}_T - \frac{\rho T}{M+\rho T}\mathbf{\Phi}\mathbf{\Phi}^\dagger\right]\mathbf{X}\right\}\right)}{\pi^{TN}(1+\frac{\rho T}{M})^{MN}} \end{aligned} \tag{6}$$

where the second equation results from $\mathbf{\Phi}$ being unitary, (4).

The mutual information between $\mathbf{X}$ and $\mathbf{\Phi}$ is given by

$$I(\mathbf{X};\mathbf{\Phi}) = \int d\mathbf{\Phi}\, p(\mathbf{\Phi}) \int d\mathbf{X}\, p(\mathbf{X}|\mathbf{\Phi}) \log_2\left(\frac{p(\mathbf{X}|\mathbf{\Phi})}{p(\mathbf{X})}\right) \tag{7}$$

$p(\mathbf{X})$ is the received signal probability density given by

$$p(\mathbf{X}) = \int d\mathbf{\Phi}\, p(\mathbf{X}|\mathbf{\Phi}) \equiv \langle p(\mathbf{X}|\mathbf{\Phi})\rangle \tag{8}$$

where we introduced the notation $\langle\cdot\rangle$ as the integration over $\mathbf{\Phi}$.

The integration over $\mathbf{\Phi}$ in (7) can be eliminated by noting [6] first that

$$p(\mathbf{X}|\mathbf{\Phi}) = p(\bar{\mathbf{\Phi}}^\dagger \mathbf{X}|\mathbf{\Phi}_0) \tag{9}$$

where $\bar{\mathbf{\Phi}} = [\mathbf{\Phi}\, \mathbf{\Phi}_\perp]$, with $\mathbf{\Phi}_\perp$ the $T \times (T-M)$ orthogonal complement of $\mathbf{\Phi}$, i.e. with

$$\mathbf{\Phi}\mathbf{\Phi}^\dagger + \mathbf{\Phi}_\perp \mathbf{\Phi}_\perp^\dagger = \mathbf{I}_T \quad \text{and} \quad \mathbf{\Phi}_\perp^\dagger \mathbf{\Phi}_\perp = \mathbf{I}_{T-M} \tag{10}$$

The choice of $\mathbf{\Phi}_0$ depends on the particular application. Thus, for the case of USTM the following expression can be used

$$\mathbf{\Phi}_0 = \begin{bmatrix} \mathbf{I}_M \\ \mathbf{0}_{T-M} \end{bmatrix} \tag{11}$$

while for DUSTM it is convenient to use

$$\mathbf{\Phi}_0 = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I}_M \\ \mathbf{I}_M \end{bmatrix} \tag{12}$$

which is the identity matrix of matrices of the form of (5). Using (9) and through the change of variables $\mathbf{X} \rightarrow \bar{\mathbf{\Phi}}^\dagger \mathbf{X}$, which leaves the $\mathbf{X}$-integration measure unaffected, we completely eliminate any non-trivial $\mathbf{\Phi}$-dependence of the integrand of (7). The remaining $\int d\mathbf{\Phi} p(\mathbf{\Phi})$ can be easily integrated to give unity and thus is disregarded. This results to

$$I(\mathbf{X};\mathbf{\Phi}) = \int d\mathbf{X}\, p(\mathbf{X}|\mathbf{\Phi}_0) \log_2 \left(\frac{p(\mathbf{X}|\mathbf{\Phi}_0)}{p(\mathbf{X})} \right) \tag{13}$$

III. Closed-form solution of $p(\mathbf{X})$ for DUSTM

When dealing with DUSTM, it is convenient to express the conditional probability in terms of $\mathbf{U}$, defined in (5). Thus, if we express $\mathbf{X}$ as

$$\mathbf{X} = [\mathbf{X}_1\, \mathbf{X}_2]^T \tag{14}$$

where both $\mathbf{X}_1$ and $\mathbf{X}_2$ have dimensions $M \times N$, then (6) can be rewritten in terms of $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{U}$ as

$$\begin{aligned} p(\mathbf{X}|\mathbf{U}) &= \frac{\exp\left(-\frac{1+\rho}{1+2\rho}\mathrm{Tr}\left\{\mathbf{X}_1^\dagger \mathbf{X}_1 + \mathbf{X}_2^\dagger \mathbf{X}_2\right\}\right)}{\pi^{2MN}(1+2\rho)^{MN}} \\ &\times \exp\left(\frac{\rho}{1+2\rho}\mathrm{Tr}\left\{\mathbf{X}_2\mathbf{X}_1^\dagger \mathbf{U} + \mathbf{X}_1\mathbf{X}_2^\dagger \mathbf{U}^\dagger\right\}\right) \end{aligned} \tag{15}$$

Combining this with (8) we get

$$\begin{aligned} p(\mathbf{X}) &= \frac{\exp\left(-\frac{1+\rho}{1+2\rho}\mathrm{Tr}\left\{\mathbf{X}_1^\dagger \mathbf{X}_1 + \mathbf{X}_2^\dagger \mathbf{X}_2\right\}\right)}{\pi^{2MN}(1+2\rho)^{MN}} \\ &\times \left\langle \exp\left(\beta \mathrm{Tr}\left\{\mathbf{X}_2\mathbf{X}_1^\dagger \mathbf{U} + \mathbf{X}_1\mathbf{X}_2^\dagger \mathbf{U}^\dagger\right\}\right)\right\rangle \end{aligned} \tag{16}$$

where

$$\beta = \frac{\rho}{1+2\rho} \tag{17}$$

We can now use the result of [11] to get

$$\begin{aligned} &\left\langle \exp\left(\beta \mathrm{Tr}\left\{\mathbf{X}_2\mathbf{X}_1^\dagger \mathbf{U} + \mathbf{X}_1\mathbf{X}_2^\dagger \mathbf{U}^\dagger\right\}\right)\right\rangle = \\ &\qquad \prod_{k=0}^{M-1} k! \frac{\det\left(y_j^{(i-1)/2} I_{i-1}(2y_j^{1/2})\right)}{\det\left(y_j^{i-1}\right)} \end{aligned} \tag{18}$$

where y_j for $j = 1 \ldots M$ are the eigenvalues of $\beta^2 \mathbf{X}_1 \mathbf{X}_2^\dagger \mathbf{X}_2 \mathbf{X}_1^\dagger$ (or the squares of the svd's of $\beta \mathbf{X}_2 \mathbf{X}_1^\dagger$). This equation is essentially the generating functional of $\mathbf{U}$: Any moment of $\mathbf{U}$ can be evaluated by taking arbitrary derivatives with respect of elements of the matrix $\mathbf{X}_2 \mathbf{X}_1^\dagger$ on both sides of (18) and subsequently setting this matrix to zero.

The determinant in the denominator is the Vandermonde determinant

$$\begin{aligned} \Delta(\{y_j\}) &= \det(y_j^{i-1}) \\ &= \begin{vmatrix} 1 & 1 & \cdots & 1 \\ y_1 & y_2 & \cdots & y_M \\ y_1^2 & y_2^2 & \cdots & y_M^2 \\ \vdots & \vdots & \cdots & \vdots \\ y_1^{M-1} & y_2^{M-1} & \cdots & y_M^{M-1} \end{vmatrix} \end{aligned} \tag{19}$$

while the determinant in the numerator can be written explicitly as

$$\det\left(y_j^{(i-1)/2} I_{i-1}(2y_j^{1/2})\right) = \tag{20}$$

$$\begin{vmatrix} I_0(2y_1^{1/2})) & .. & I_0(2y_M^{1/2}) \\ y_1^{1/2} I_1(2y_1^{1/2}) & .. & y_M^{1/2} I_1(2y_M^{1/2}) \\ y_1 I_2(2y_1^{1/2}) & .. & y_M I_2(2y_M^{1/2}) \\ \vdots & .. & \vdots \\ y_1^{(M-1)/2} I_{M-1}(2y_1^{1/2}) & .. & y_M^{(M-1)/2} I_{M-1}(2y_M^{1/2}) \end{vmatrix}$$

where $I_n(x)$ is the modified Bessel function of order n.

One has to exercise caution in evaluating (18) in the case $M < N$. The reason is that only R singular values of $\mathbf{X}_1 \mathbf{X}_2^\dagger$ are non-zero. Therefore, both the determinants in the numerator and the denominator vanish. However, the ratio remains finite. Using Lemma I we can show that

$$\langle \exp\left(\beta \mathrm{Tr}\left\{\mathbf{X}_2 \mathbf{X}_1^\dagger \mathbf{U} + \mathbf{X}_1 \mathbf{X}_2^\dagger \mathbf{U}^\dagger\right\}\right) \rangle = \tag{21}$$

$$\prod_{k=M-R}^{M-1} k! \, \frac{\det\left(y_j^{(M-R+i-1)/2} I_{M-R+i-1}(2y_j^{1/2})\right)}{\det\left(y_j^{M-R+i-1}\right)}$$

where the range of the indices in the determinants are $i, j = 1, \ldots R$.

IV. Mutual Information of DUSTM

Using (15), (16) and (21) we can now express the ratio $p(\mathbf{X}|\mathbf{\Phi}_0)/p(\mathbf{X})$ as

$$\log_2\left(\frac{p(\mathbf{X}|\mathbf{\Phi}_0)}{p(\mathbf{X})}\right) = \tag{22}$$
$$\log_2\left[\frac{\prod_{k=M-R}^{M-1}\frac{1}{k!}\ \det\left(y_j^{M-R+i-1}\right)}{\det\left(y_j^{(M-R+i-1)/2}\tilde{I}_{M-R+i-1}(2y_j^{1/2})\right)}\right]$$
$$+\left(\beta\mathrm{Tr}\left\{\mathbf{X}_1^\dagger\mathbf{X}_2+\mathbf{X}_2^\dagger\mathbf{X}_1\right\}-2\sum_{i=1}^{R}y_i^{1/2}\right)\log_2 e$$

In the above equation we have defined $\tilde{I}_n(x)=I_n(x)e^{-x}$ and we have multiplied both numerator and denominator of the expression inside the log with $\exp\left(-2\sum_{i=1}^{R}y_i^{1/2}\right)$, so that neither will have exponentially increasing terms for large y_i.

To evaluate the mutual information, (22) needs to averaged over realizations of $\mathbf{X}_1$, $\mathbf{X}_2$, which are generated with probability distribution $p(\mathbf{X}|\mathbf{\Phi}_0)$. This corresponds to $\mathbf{X}_1$, $\mathbf{X}_2$ having Gaussian correlations given by

$$E\left[X_{1i\alpha}^* X_{1j\beta}\right] = (1+\rho)\delta_{ij}\delta_{\alpha\beta} \tag{23}$$

$$E\left[X_{2i\alpha}^* X_{2j\beta}\right] = (1+\rho)\delta_{ij}\delta_{\alpha\beta} \tag{24}$$

$$E\left[X_{1i\alpha}^* X_{2j\beta}\right] = \rho\delta_{ij}\delta_{\alpha\beta} \tag{25}$$

V. Mutual Information of USTM

In the next section we will compare the mutual information of DUSTM to that of USTM. Thus, for completeness, we review here the results obtained in [6] regarding USTM. We start with the conditional probability $p(\mathbf{X}|\mathbf{\Phi})$

$$p(\mathbf{X}|\mathbf{\Phi}) = \frac{\exp\left(-\mathrm{Tr}\left\{\mathbf{X}^\dagger\mathbf{X}\right\}\right)\exp\left(\alpha\mathrm{Tr}\left\{\mathbf{X}^\dagger\mathbf{\Phi}\mathbf{\Phi}^\dagger\mathbf{X}\right\}\right)}{\pi^{TN}(1+\rho T/M)^{MN}} \tag{26}$$

where $\mathbf{X}$ is a $T\times N$ complex matrix, $\mathbf{\Phi}$ is a $T\times N$ unitary matrix and

$$\alpha = \frac{\rho T}{M+\rho T} \tag{27}$$

In [6] the received signal probability density was found to be

$$p(\mathbf{X}) = \int d\mathbf{\Phi}\, p(\mathbf{X}|\mathbf{\Phi}) \tag{28}$$
$$= \frac{\exp\left(-\mathrm{Tr}\left\{\mathbf{X}^\dagger\mathbf{X}\right\}\right)}{\pi^{TN}\left(1+\rho T/M\right)^{MN}}$$
$$\times \left\langle\exp\left(\alpha\mathrm{Tr}\left\{\mathbf{X}^\dagger\mathbf{\Phi}\mathbf{\Phi}^\dagger\mathbf{X}\right\}\right)\right\rangle$$

where the average over $\mathbf{\Phi}$, expressed as $\langle \cdots \rangle$ was performed as follows:

$$\begin{aligned} & \langle \exp\left(\alpha \mathrm{Tr}\left\{\mathbf{X}^\dagger \mathbf{\Phi}\mathbf{\Phi}^\dagger \mathbf{X}\right\}\right)\rangle \\ = \quad & \frac{C_{TM}}{M!}\int \frac{dt_1}{2\pi}\cdots\int \frac{dt_M}{2\pi} \\ \times \quad & \prod_{m=1}^{M}\left[\frac{e^{-it_m}}{(-\alpha y_1 - it_m)\cdots(-\alpha y_K - it_m)(-it_m)^{T-K}}\right] \\ \times \quad & \prod_{l<m}(-it_m - it_l)^2 \\ = \quad & C_{TM}\,|\det \mathbf{F}| \end{aligned} \tag{29}$$

where the constant C_{TM} is equal to

$$C_{TM} = \frac{(T-1)!\cdots(T-M)!}{(M-1)!\cdots(0)!} \tag{30}$$

and $\mathbf{F}$ is a $M \times M$ Hankel matrix with entries given by

$$\begin{aligned} F_{mn} \quad = \quad & \sum_{k=1}^{K}\frac{e^{\alpha y_k}}{(\alpha y_k)^q \prod_{l\neq k}(\alpha y_k - \alpha y_l)} \\ \times \quad & \begin{cases} \frac{\gamma(q,\alpha y_k)}{\Gamma(q)}, & q \geq 1 \\ 1, & q \leq 0 \end{cases} \end{aligned} \tag{31}$$

In the above expression, $q = T-K-m-n+2$, $\gamma(n,x)$ is the incomplete Γ function and y_n, for $n = 1,\ldots,K$ are the non-zero eigenvalues of the $N \times N$ matrix $\mathbf{X}^\dagger\mathbf{X}$. As in the case of DUSTM, to numerically calculate the mutual information one needs to average the log-ratio $\log_2(p(\mathbf{X}|\mathbf{\Phi}_0)/p(\mathbf{X}))$, where $\mathbf{\Phi}_0$ is given by (11), with respect to $\mathbf{X}$, which has probability density $p(\mathbf{X}|\mathbf{\Phi}_0)$. It is convenient to write $\mathbf{X}^\dagger\mathbf{X}$ as

$$\mathbf{X}^\dagger\mathbf{X} = \left(1 + \frac{\rho T}{M}\right)\mathbf{X_1^\dagger X_1} + \mathbf{X_2^\dagger X_2} \tag{32}$$

where $\mathbf{X}_1$, $\mathbf{X}_2$ are $M \times N$, $(T-M) \times N$ complex Gaussian, unit-variance matrices.

VI. Analysis and Comparison to Isotropic USTM

In section VI-C below, we present numerical results on the mutual information of DUSTM and compare them to corresponding USTM results. However, before that, it is instructive to analyze the asymptotic behavior of the mutual information in both small and large SNR regimes. As we shall see, this exact asymptotic analysis of both USTM and DUSTM will provide insight and quantitative agreement with numerical simulations.

A. Low ρ region

To obtain the small ρ behavior we expand the exponent in the log-ratio $\log_2(p(\mathbf{X}|\mathbf{\Phi}_0)/p(\mathbf{X}))$ and integrate over the fields. For the DUSTM case in Appendix II-A we obtain

$$\hat{I}_{DUSTM} \approx \rho^2 N\left[1 - 2\rho + \frac{\rho^2}{2}\left(5 - \frac{N}{M}\right)\right]\log_2 e \tag{33}$$

For small small ρ, we see that $\hat{I}_{DUSTM}$ is an increasing function of M. As a result, under the constraint of the channel being constant over T time-intervals, the optimal number of transmitting antennas is $M_{opt} = T/2$.

For comparison, in Appendix II-B we calculate the mutual information for USTM for small ρ. The final result up to $\mathcal{O}(\rho^3)$ is

$$\begin{aligned} \hat{I}_{USTM} &\approx \frac{N\rho^2}{2M}(T-M)\left[1-\frac{2\rho T}{3M}\left(1+\frac{M}{T}\right)\right]\log_2 e \\ &= \frac{\rho^2 N}{2}(1-2\rho)\log_2 e \end{aligned} \tag{34}$$

where the last equality holds for $T = 2M$. We see that for $T = 2M$, $\hat{I}_{DUSTM} \approx 2\hat{I}_{USTM}$ up to order $\mathcal{O}(\rho^3)$! Also, for fixed T and N, $\hat{I}_{USTM}$ is actually a decreasing function of the number of transmitting antennas M, with optimal $M = 1$. This can be seen in Fig. 3, where the optimal M at low ρ is 1.

It is important to note that for $\rho \ll 1$, the mutual information for both schemes scales as ρ^2, rather than ρ as in the coherent case. Thus, at small SNR, the lack of knowledge of the channel becomes increasingly problematic. This is generally the case for unitary space-time modulated schemes.

B. High ρ region

In Appendix III-A we obtain the large ρ behavior of the mutual information of DUSTM, which to $\mathcal{O}(\log_2 \rho/\rho)$ is

$$\begin{aligned} \hat{I}_{DUSTM} &= \frac{1}{M}\left[R\left(M-\frac{R}{2}\right)\log_2\rho + A_{MN}\right] \\ &+ \mathcal{O}\left(\frac{\log_2\rho}{\rho}\right) \end{aligned} \tag{35}$$

where

$$\begin{aligned} A_{MN} &= \frac{R}{2}\log_2(4\pi) - R\left(M-\frac{R}{2}\right)\log_2(2e) \\ &- \sum_{k=M-R}^{M-1}\log_2 k! + R\left(M-R+\frac{1}{2}\right)\mathcal{L}_1(M,N) \\ &+ \frac{1}{2}R(R-1)\mathcal{L}_2(M,N) \end{aligned} \tag{36}$$

is a constant, independent of ρ. In (36) we have defined the quantities $\mathcal{L}_1(M,N) = E\left[\log_2\lambda_1\right]$ and $\mathcal{L}_2(M,N) = E\left[\log_2(\lambda_1+\lambda_2)\right]$, where $\lambda_{1,2}$ are distinct non-zero singular values of an $M \times N$ matrix with independent $\mathcal{CN}(0,1)$ entries. Their explicit expressions are given in (84), (85).

Similarly, in Appendix III-B we derive the asymptotic large-ρ form of the mutual information for USTM (for $T \geq M$)

$$\begin{aligned} \hat{I}_{USTM} &= \frac{1}{T}\left[R(T-M)\log_2\rho + B_{TMN}\right] \\ &+ \mathcal{O}\left(\frac{\log_2\rho}{\rho}\right) \end{aligned} \tag{37}$$

with

$$\begin{aligned} B_{TMN} &= R(T-M)\left(\log_2 \frac{T}{Me} + \mathcal{L}_1(M,N)\right) \\ &- \log C_{TM} - \log_2 |\det \mathbf{G}| \end{aligned} \tag{38}$$

with $\mathcal{L}_1(M,N)$ given in (84). The last term appears only for $M < N$ and the elements of $\mathbf{G}$ are given in (92). It is important to note that for $T = M$ the mutual information vanishes to the order calculated above, since in that case the mutual information is identically zero.

The leading terms, proportional to $\log_2 \rho$ in (35) and (37) provide insight on the large ρ behavior of DUSTM and USTM. Starting with (35), we find that for fixed N, the mutual information $\hat{I}_{DUSTM}$ is an increasing function of M. Thus, as we found in the small ρ case in the previous section, to maximize the mutual information, one should use the maximum number of transmitting antennas consistent with the constraint that the channel is constant over $2M$ time-intervals.

In the case of USTM we find that, for $T > 2N$ the optimal transmitting antenna number is $M_{opt} = N$, while in the opposite case $T \leq 2N$, the leading term is optimized for $M_{opt} = T/2$.

Once optimized over M, the leading terms of both (35) and (37) are identical to 2). Thus, to leading order in ρ, both DUSTM and USTM are capacity achieving schemes. Comparing the next-to-leading ρ-independent terms in (35), (37) we find that, after optimizing over M, the mutual information of DUSTM is larger than that of USTM. This can be seen in Fig. 3, where the optimized-over-M $\hat{I}_{DUSTM}$ and $\hat{I}_{USTM}$ of (35) and (37) are plotted (dashed lines). This may come as a surprise if one takes into account that for $T = 2M$, the manifold of constellations used for DUSTM (5) is a subgroup of those used in USTM. However, one should take into account that in DUSTM, although information is sent over M time-intervals, the receiver exploits the side information that the channel has not changed over the previous M time-intervals.

C. Numerical Simulations

We now discuss the numerical simulations performed to evaluate the mutual information for USTM and DUSTM. The simulation procedure consists of the following steps: First we generate L instances of Gaussian complex random matrices with covariance given by (23) and (32) for the DUSTM and USTM cases. For each matrix instantiation we calculate the singular values and then we apply them to evaluate the log-ratio $\log_2(p(\mathbf{X}|\mathbf{\Phi})/p(\mathbf{X}))$, which we then average over its L values. For intermediate and large ρ we have found that $L \approx 4 - 5 \cdot 10^4$ are sufficient. However, for smaller ρ, at least $L = 5 \cdot 10^5$ are required. The reason is that the mutual information, being $\mathcal{O}(\rho^2)$, is quite small and therefore fluctuations have a more pronounced effect.

In Fig. 1 we compare the numerically evaluated mutual information of USTM and DUSTM for low, intermediate and relatively large SNR values. We find that for small $\rho = -6dB$ the normalized mutual information $\hat{I}_{DUSTM}$ is nearly exactly twice $\hat{I}_{USTM}$. This is in agreement with (33) and (34). Even for intermediate SNR, $\rho = 6dB$ we find the approximate relation $\hat{I}_{USTM}(T = 2M, M, 2N) \approx \hat{I}_{DUSTM}(T = 2M, M, N)$. This approximation breaks down for larger ρ.

Motivated by these ratio dependencies and scaling relations, in Fig. 2 we analyze the dependence of ratios of $\hat{I}_{DUSTM}$ and $\hat{I}_{USTM}$ on SNR. In Fig. 2(a) we plot the ratio $\hat{I}_{DUSTM}(T = 2M, M, N = rM)/\hat{I}_{USTM}(T = 2M, M, N = rM)$ as a function of ρ for various values of M and for $r = 1/2$, $r = 1$ and $r = 2$. We find that for fixed r, the ratios fall close (but not on top) to each other. Their value starts from very close to 2, for small ρ and in accordance with (33), (34), and approaches $2(1 - 0.5\min(1, r))$, in agreement with (35), (37). We note however the slow convergence to their asymptotic values for large ρ, which can be explained by the fact that both mutual informations increase only logarithmically with ρ. The closeness of the curves for fixed r indicates that the ratio has weak dependence on the actual values of T, M, N. Thus a large-T, M, N analysis is expected to give good results even for small antenna numbers.

In Fig. 2(b) we plot the ratios $\hat{I}_{DUSTM}(T = 2M, M, N = rM)/(M\hat{I}_{DUSTM}(2, 1, N = 1)$ as a function of ρ for various values of M and r.

In Fig. 3 we analyze the mutual information of DUSTM and USTM optimized over the number of transmitting antennas M with T fixed to $T = 8$ and for various values of N. In Figs 3(a),(b) we plot the capacity of DUSTM and USTM defined as

$$\begin{aligned} C_{DUSTM} &= \max_{M^* \le T/2} \hat{I}_{DUSTM}(T^* = 2M^*, M^*, N) \qquad (39) \\ &= \hat{I}_{DUSTM}(T, T/2, N) \\ C_{USTM} &= \max_{M^*} \hat{I}_{USTM}(T, M^*, N) \qquad (40) \end{aligned}$$

as a function of ρ (solid curves). In Fig. 3(c) the solid curves depict the optimal number of M that maximizes $\hat{I}_{USTM}(T, M, N)$

$$M_{opt} = \arg\max_{M^*} \hat{I}_{USTM}(T, M^*, N) \qquad (41)$$

as a function of ρ. As seen in (39), the optimal M for DUSTM is always equal to $M = T/2$, consistent with both low and large ρ analysis. In Figs. 3(b),(c) the dashed curves represent the capacity and optimal M values as evaluated using the large-ρ asymptotic expressions of (35), (37). Very good agreement with the exact values (solid curves) can be seen down to moderate SNR. However, one should note, that even though (37) describes the capacity accurately down to moderate SNR, the large-ρ optimal value of M as predicted by simply maximizing the $\log\rho$ term in (2) [5] and in (37), actually becomes optimal at very large $\rho \sim 50dB$.

Turning now to Fig. 3(a) we see that at relatively small SNR, C_{USTM} and C_{DUSTM} actually cross each other. At high SNR, DUSTM consistently performs better than USTM. At low SNR, USTM, when optimized over M performs better than DUSTM. This can be explained by looking at the leading term of (34): the optimal M is $M_{opt} = 1$ and $\hat{I}_{USTM}(T, 1, N)$ can be higher than $\hat{I}_{DUSTM}(T, T/2, N)$. Interestingly, the analytic estimates at low SNR do not match very accurately to the behavior at $\rho \approx -6dB$.

VII. Conclusions

In conclusion, we have found a closed-form expression for the probability density of the received signal for differential unitary space-time modulated (DUSTM) signals. This allowed us to evaluate numerically the corre-

sponding mutual information. In addition, we calculated analytically the asymptotic form of the mutual information for DUSTM and USTM for small and large SNR's. At low SNR's the nondifferential form of USTM can outperform the differential form if the number of transmit antennas is optimized. However, at high enough SNR's the differential USTM outperforms its nondifferential counterpart with respect to mutual information. An additional advantage of DUSTM over USTM is its simplicity of decoding, though recent progress has been reported for decoding of nondifferential USTM [13]. This suggests that DUSTM is a promising type of transmission for non-coherent MIMO channels.

APPENDIX I

Lemma 1 (Lemma): Let $f_j(x_i)$ represent the i,j-th element of a $T \times T$ dimensional matrix. Here $f_j(x)$ for $j = 1, \ldots, T$ is a family of analytic functions and $\{x_i\}$ is a T-plet of real numbers. For simplicity we represent this matrix in terms of its columns denoted by $\mathbf{f}(x_i) = [f_1(x_i)\, f_2(x_i)\, \ldots\, f_T(x_i)]^T$. Also we denote by $\Delta(\{x_j\})$ the Vandermonde determinant of the x_j's

$$\Delta(\{x_i\}) = \det(x_i^{(j-1)}) = \prod_{j>i}(x_j - x_i) \tag{42}$$

Thus, in the limit that a subset of k members of the T-plet are equal with each other (i.e. $x_1 = \ldots = x_k$, for $k \leq T$), then the ratio of $\det f_i(x_j)/\Delta(\{x_i\})$ exists and is equal to

$$\lim_{x_i \to x_1\, i=2,\ldots,k} \frac{\det\left[\mathbf{f}(x_1)\, \mathbf{f}(x_2)\, \ldots\, \mathbf{f}(x_T)\right]}{\prod_{j=1}^{T} \prod_{i>j}(x_i - x_j)} = \tag{43}$$
$$\frac{\det\left[\mathbf{f}(x_1)\, \mathbf{f}^{(1)}(x_1)\, \ldots\, \mathbf{f}^{(k-1)}(x_1)\, \mathbf{f}(x_{k+1})\, \ldots\, \mathbf{f}(x_T)\right]}{\prod_{p=0}^{k-1} p! \left[\prod_{i>j,\, j=k+1}^{T}(x_i - x_j)\right] \prod_{m=k+1}^{T}(x_m - x_1)^k}$$

where $\mathbf{f}^{(n)}(x)$ denotes the n-th derivative of each of the elements of the vector $\mathbf{f}(x)$ evaluated at x.

Proof: Omitted ■

APPENDIX II

SMALL ρ ANALYSIS

In this section we will calculate the first four terms in the Taylor expansion in ρ of the mutual information for both the differential and the isotropic USTM cases. The mutual information $I(\mathbf{X}; \mathbf{\Phi})$, (13) can be rewritten as

$$I(\mathbf{X}; \mathbf{\Phi}) = \int d\mathbf{X}\, p(\mathbf{X}|\mathbf{\Phi}_0) \log_2 \frac{p(\mathbf{X}|\mathbf{\Phi}_0)}{< p(\mathbf{X}|\mathbf{\Phi}) >} \tag{44}$$

where $\langle \cdot \rangle$ denotes average over $\mathbf{\Phi}$. The simplest way to proceed is to expand both logarithms in powers of ρ and, where convenient, interchange the integration over $\mathbf{X}$ and $\mathbf{\Phi}$. The expectation over $\mathbf{X}$ will be denoted by $E[\cdot]$.

A. Differential USTM

In the case of DUSTM, we see that by taking the ratio of $p(\mathbf{X}|\mathbf{U}=\mathbf{I}_M)$ in (15) and $p(\mathbf{X})$ in (16) the mutual information can be written as

$$\begin{aligned} I &= \beta E\left[\mathrm{Tr}\left\{\mathbf{X}_2\mathbf{X}_1^\dagger+\mathbf{X}_1\mathbf{X}_2^\dagger\right\}\right]\log_2 e \\ &- E\left[\log_2\langle\exp\left(\beta\mathrm{Tr}\left\{\mathbf{X}_2\mathbf{X}_1^\dagger\mathbf{U}+\mathbf{X}_1\mathbf{X}_2^\dagger\mathbf{U}^\dagger\right\}\right)\rangle\right] \end{aligned} \tag{45}$$

Since $\mathbf{X}_1$, $\mathbf{X}_2$ are zero-mean Gaussian quantities, we only need to specify their variances given by (23). As a result, the first term in (45) can be easily evaluated to give

$$\beta E\left[\mathrm{Tr}\left\{\mathbf{X}_2\mathbf{X}_1^\dagger+\mathbf{X}_1\mathbf{X}_2^\dagger\right\}\right]=2\beta\rho MN \tag{46}$$

To deal with the second term in (45), we also need the following identities for the averages over $\mathbf{U}$.

$$\langle U_{ij}U_{lk}^*\rangle = \frac{1}{M}\delta_{il}\delta_{jk} \tag{47}$$

$$\langle\prod_{q=1}^{2d+1}U_{i_qj_q}\rangle = 0 \tag{48}$$

where $r=0,1,\cdots$. Note that, since $\mathbf{U}$ is an element of $U(M)$, all odd moments vanish. However, even moments other than the second one are not easy to evaluate. In fact, even using the simple-looking form of (18) does not simplify matters too much.

To expand the exponent of the second term in (45) in powers of β we use the notation

$$A_n=\left(\mathrm{Tr}\left\{\mathbf{X}_2\mathbf{X}_1^\dagger\mathbf{U}+\mathbf{X}_1\mathbf{X}_2^\dagger\mathbf{U}^\dagger\right\}\right)^n \tag{49}$$

We see that due to (48), all odd terms vanish, $\langle A_{2r+1}\rangle=0$. Thus, to 4^{th} order in β, (45) can be written as

$$\begin{aligned} I &= \Bigg(2\beta\rho MN-\frac{\beta^2}{2}E\left[\langle A_2\rangle\right] \\ &- \frac{\beta^4}{24}\left(E\left[\langle A_4\rangle\right]-3E\left[\langle A_2\rangle^2\right]\right)\Bigg)\log_2 e \end{aligned} \tag{50}$$

From (47) we get

$$< A_2 >= \frac{2}{M}\mathrm{Tr}\left\{\mathbf{X}_2\mathbf{X}_1^\dagger\mathbf{X}_1\mathbf{X}_2^\dagger\right\} \tag{51}$$

which results to $E\left[\langle A_2\rangle\right]=\rho^2MN^2+(1+\rho)^2MN^2$. Since $\beta=\rho/(1+2\rho)$ is $\mathcal{O}(\rho)$ for small ρ, we only need to evaluate the averages involving A_4 and A_2^2 to leading order in ρ, i.e. to $\mathcal{O}(1)$. Thus, we may neglect the $\mathcal{O}(\rho)$ terms in the correlations between $\mathbf{X}_1$ and $\mathbf{X}_2$ (see (23)). As a result,

$$\begin{aligned} E\left[\langle A_2\rangle^2\right] &= 4N^2(1+M^2)+\mathcal{O}(\rho) \\ E\left[\langle A_4\rangle\right] &= \langle E\left[A_4\right]\rangle \\ &= 12N(1+MN)\mathrm{Tr}\left\{\langle\mathbf{U}\mathbf{U}^\dagger\mathbf{U}\mathbf{U}^\dagger\rangle\right\}+\mathcal{O}(\rho) \\ &= 12MN(1+MN)+\mathcal{O}(\rho) \end{aligned} \tag{52}$$

Collecting all terms from above and expanding them to $\mathcal{O}(\rho^4)$ we obtain the mutual information of (33).

B. Isotropic USTM

In the case of USTM, we will expand I_{USTM} to order ρ^3. Here, the analog of (45) is

$$\begin{aligned} I &= \alpha E\left[\mathrm{Tr}\left\{\mathbf{X}^\dagger \mathbf{J}_M \mathbf{X}\right\}\right]\log_2 e \\ &- E\left[\log_2\langle \exp\left(\alpha \mathrm{Tr}\left\{\mathbf{X}^\dagger \mathbf{U}\mathbf{J}_M\mathbf{U}^\dagger\mathbf{X}\right\}\right)\rangle\right] \end{aligned} \tag{53}$$

where $\mathbf{U}$ is a $T\times T$ unitary matrix and the $T\times N$ Gaussian random matrix $\mathbf{X}$ has the following correlations, which follow from (32)

$$E\left[X^*_{ij}X_{kl}\right] = \delta_{ik}\delta_{jl}(1+J_{ii}\frac{\alpha}{1-\alpha}) \tag{54}$$

As a result, the first term in (53) can be easily evaluated to

$$\alpha E\left[\mathrm{Tr}\left\{\mathbf{X}^\dagger \mathbf{J}_M \mathbf{X}\right\}\right] = MN\alpha/(1-\alpha) = TN\rho \tag{55}$$

Similarly to the previous section, we define B_n as

$$B_n = \left(\mathrm{Tr}\left\{\mathbf{X}^\dagger \mathbf{U}\mathbf{J}_M\mathbf{U}^\dagger\mathbf{X}\right\}\right)^n \tag{56}$$

Then, after expanding the second term in (53) to order α^3, I becomes

$$\begin{aligned} I &= \log_2 e\left(MN\frac{\alpha}{1-\alpha} - \alpha E\left[\langle B_1\rangle\right]\right. \\ &- \frac{\alpha^2}{2}\left(E\left[\langle B_2\rangle\right] - E\left[\langle B_1\rangle^2\right]\right) \\ &- \left.\frac{\alpha^3}{6}\left(E\left[\langle B_3\rangle\right] + 2E\left[\langle B_1\rangle^3\right] - 3E\left[\langle B_1\rangle\langle B_2\rangle\right]\right)\right) \end{aligned} \tag{57}$$

Using the orthogonality relation for $U(T)$ unitary matrices

$$\langle U_{ij}U^*_{lk}\rangle = \frac{1}{T}\delta_{il}\delta_{jk} \tag{58}$$

we can calculate $\langle B_1\rangle$ to be

$$< B_1 >= \frac{M}{T}\mathrm{Tr}\left\{\mathbf{X}^\dagger\mathbf{X}\right\} \tag{59}$$

We can now calculate the terms in (57) explicitly:

$$E\left[\langle B_1\rangle\right] = MN\left(1+\frac{M}{T}\frac{\alpha}{1-\alpha}\right) \tag{60}$$

$$\begin{aligned} E\left[\langle B_2\rangle\right] - E\left[\langle B_1\rangle^2\right] &= \langle E\left[B_2\right]\rangle - E\left[\langle B_1\rangle^2\right] \\ &= MN\left(1-\frac{M}{T}\right)\left(1+2\rho\frac{M}{T}\right)+\mathcal{O}(\rho^2) \end{aligned} \tag{61}$$

$$\begin{aligned} &E\left[\langle B_3\rangle\right] + 2E\left[\langle B_1\rangle^3\right] - 3E\left[\langle B_1\rangle\langle B_2\rangle\right] \\ &= 2MN\left(1-\frac{M}{T}\right)\left(1-2\rho\frac{M}{T}\right)+\mathcal{O}(\rho) \end{aligned} \tag{62}$$

Note that the last two equations were only calculated to $\mathcal{O}(\rho)$ and $\mathcal{O}(1)$, given that their proportionality constants in (57) are $\mathcal{O}(\rho^2)$ and $\mathcal{O}(\rho^3)$, respectively. Collecting all terms (60), (61) and (62) together in (57), we get the mutual information for USTM to $\mathcal{O}(\rho^3)$ expressed in (34).

APPENDIX III
LARGE ρ ANALYSIS

A. Differential USTM

We wish to calculate the asymptotic behavior of the DUSTM mutual information for large ρ. Using (46), we rewrite the log-ratio of (22) as

$$\begin{aligned}\log_2\left(\frac{p(\mathbf{X}|\mathbf{\Phi}_0)}{p(\mathbf{X})}\right) &= \left(2\beta\rho MN - 2\sum_{i=1}^{R}\sqrt{y_i}\right)\log_2 e \\ &- \sum_{k=M-R}^{M-1}\log_2 k! \quad + \quad \log_2\left[\frac{\det\left(y_j^{M-R+i-1}\right)}{\det\left(y_j^{\frac{M-R+i-1}{2}}\tilde{I}_{M-R+i-1}(2\sqrt{y_j})\right)}\right]\end{aligned} \tag{63}$$

where y_i, for $i = 1, \cdots, R$ are the R eigenvalues of the matrix $\beta^2\mathbf{X_2}\mathbf{X_1}^\dagger\mathbf{X_1}\mathbf{X_2}^\dagger$. The above equation is averaged over the $M \times N$ Gaussian matrices $\mathbf{X_1}$, $\mathbf{X_2}$ with correlations given by (23). To analyze the large ρ behavior, it is convenient to use the independent $M \times N$ matrices $\mathbf{Z}_\pm$ with $\mathcal{CN}(0,1)$ entries, defined as

$$\begin{aligned}\mathbf{Z}_+ &= \frac{\mathbf{X}_1 + \mathbf{X}_2}{\sqrt{2(1+2\rho)}} \\ \mathbf{Z}_- &= \frac{\mathbf{X}_1 - \mathbf{X}_2}{\sqrt{2}}\end{aligned} \tag{64}$$

Thus $\beta^2\mathbf{X_2}\mathbf{X_1}^\dagger\mathbf{X_1}\mathbf{X_2}^\dagger$ can be written as a sum of terms with decreasing powers of ρ:

$$\beta^2\mathbf{X_1}\mathbf{X_2}^\dagger\mathbf{X_2}\mathbf{X_1}^\dagger = \beta^2\rho^2\left(\boldsymbol{\mathcal{H}}_0{}^2 + \frac{\boldsymbol{\mathcal{H}}_1}{\sqrt{\rho}} + \frac{\boldsymbol{\mathcal{H}}_2}{\rho} + \mathcal{O}\left(\frac{1}{\rho^{3/2}}\right)\right) \tag{65}$$

where

$$\boldsymbol{\mathcal{H}}_0 = \boldsymbol{\mathcal{N}}_+ \tag{66}$$

$$\boldsymbol{\mathcal{H}}_1 = \frac{1}{\sqrt{2}}\left(\left(\mathbf{Z}_-\mathbf{Z}_+^\dagger - \mathbf{Z}_+\mathbf{Z}_-^\dagger\right)\boldsymbol{\mathcal{N}}_+ + h.c.\right) \tag{67}$$

$$\begin{aligned}\boldsymbol{\mathcal{H}}_2 &= \boldsymbol{\mathcal{N}}_+^2 - \frac{1}{2}\left(\boldsymbol{\mathcal{N}}_+\boldsymbol{\mathcal{N}}_- + \boldsymbol{\mathcal{N}}_-\boldsymbol{\mathcal{N}}_+\right) \\ &- \frac{1}{2}\left(\mathbf{Z}_-\mathbf{Z}_+^\dagger - \mathbf{Z}_+\mathbf{Z}_-^\dagger\right)^2\end{aligned} \tag{68}$$

and $\boldsymbol{\mathcal{N}}_\pm = \mathbf{Z}_\pm\mathbf{Z}_\pm^\dagger$.

To leading order in ρ, $y_i = (\beta\rho\lambda_i)^2$, where λ_i are the eigenvalues of $\boldsymbol{\mathcal{N}}_+$. We will need to calculate y_i to next to leading order, focusing on the R non-zero ones. This perturbative analysis of eigenvalues is treated in detail in standard textbooks, see for example [14]. Below we simply quote the answer:

$$\begin{aligned}y_i &= \beta^2\rho^2\left(\lambda_i^2 + \frac{\mathbf{n}_i^\dagger\boldsymbol{\mathcal{H}}_1\mathbf{n}_i}{\sqrt{\rho}} + \frac{\mathbf{n}_i^\dagger\boldsymbol{\mathcal{H}}_2\mathbf{n}_i}{\rho}\right. \\ &+ \left.\frac{1}{\rho}\sum_{j\neq i}\frac{\left|\mathbf{n}_i^\dagger\boldsymbol{\mathcal{H}}_1\mathbf{n}_j\right|^2}{\lambda_i^2 - \lambda_j^2} + \mathcal{O}(\rho^{-3/2})\right)\end{aligned} \tag{69}$$

where $\mathbf{n}_i$ are the eigenvectors corresponding to λ_i. The last term in the above equation is summed over all λ_j, including zeros, and is well behaved because the eigenvalues λ_i are unequal with probability 1. We next observe that, since $\mathbf{n}_i$ are eigenvectors of $\boldsymbol{\mathcal{H}}_0$, $\mathbf{n}_i^\dagger \boldsymbol{\mathcal{H}}_1 \mathbf{n}_i = 0$. We now can expand the second term in (63):

$$\begin{aligned} 2\sum_{i=1}^{R}\sqrt{y_i} &= 2\beta\rho\sum_{i}^{R}\left[\lambda_i + \frac{\mathbf{n}_i^\dagger \boldsymbol{\mathcal{H}}_2 \mathbf{n}_i}{2\rho\lambda_i}\right. \\ &+ \left.\frac{1}{2\rho\lambda_i}\sum_{j\neq i}\frac{\left|\mathbf{n}_i^\dagger \boldsymbol{\mathcal{H}}_1 \mathbf{n}_j\right|^2}{\lambda_i^2-\lambda_j^2} + \mathcal{O}(\rho^{-2})\right] \end{aligned} \tag{70}$$

To proceed further, we integrate out $\mathbf{Z}_-$ in the above equation (but not $\mathbf{Z}_+$). As a result we get

$$\begin{aligned} 2\sum_{i=1}^{R}\sqrt{y_i} &= 2\beta\rho\sum_{i}^{R}\lambda_i \\ &+ \beta\left(\sum_{i}^{R}\lambda_i + (M-N)R\right) + \mathcal{O}(\rho^{-1}) \end{aligned} \tag{71}$$

which, after integrating over $\mathbf{Z}_+$ gives

$$\begin{aligned} 2E\left[\sum_{i=1}^{R}\sqrt{y_i}\right] &= 2\beta\rho MN \\ &+ R\left(M-\frac{R}{2}\right) + \mathcal{O}(\rho^{-1}) \end{aligned} \tag{72}$$

Thus, the first term in the above equation, cancels the first $\mathcal{O}(\rho)$ term in (63), with the remainder being only of order unity.

We now turn to the asymptotic treatment of the determinants in (63). Since for large ρ the non-zero y_i's will be large, we may use the asymptotic form of the normalized modified Bessel function

$$\begin{aligned} \tilde{I}_n(x) &\equiv e^{-x} I_n(x) \\ &\approx \frac{1}{\sqrt{2\pi x}}(1+\mathcal{O}(x^{-1})) \end{aligned} \tag{73}$$

in the determinant of the denominator in (63) to obtain

$$\begin{aligned} &\det\left(\frac{y_j^{\frac{M-R+i-1}{2}}}{\sqrt{4\pi\sqrt{y_j}}}(1+\mathcal{O}(y_j^{-1/2}))\right) \\ =\ &\det\left(\frac{(\beta\rho\lambda_j)^{M-R+i-1}}{\sqrt{4\pi\beta\rho\lambda_j}}(1+\mathcal{O}(\rho^{-1}))\right) \\ =\ &\frac{(\beta\rho)^{(2M-R-2)R/2}}{(4\pi)^{R/2}}\prod_{i,j<i}\lambda_i^{M-R-1/2}(\lambda_j-\lambda_i) \\ \times\ &\left[1+\mathcal{O}(\rho^{-1})\right] \end{aligned} \tag{74}$$

The first equality follows from the fact that $y_i = \beta^2\rho^2(\lambda_i^2 + \mathcal{O}(\rho^{-1}))$. Similarly, the Vandermonde determinant can

be expressed as

$$\begin{aligned}\det(y_j^{M-R+i-1}) &= \det\left(\left[\beta^2\rho^2\lambda_j^2(1+\mathcal{O}(\rho^{-1}))\right]^{M-R+i-1}\right) \\ &= (\beta\rho)^{R(2M-R-1)}\prod_{i,j<i}\lambda_i^{2(M-R)}(\lambda_j^2-\lambda_i^2) \\ &\times \left[1+\mathcal{O}(\rho^{-1})\right] \end{aligned} \tag{75}$$

Taking the logarithm of the ratio of the two determinants (74), (75), we get

$$\begin{aligned}\log_2\frac{\det(\cdots)}{\det(\cdots)} &= R\left(M-\frac{R}{2}\right)\log\frac{\rho}{2}+\frac{R}{2}\log_2 4\pi \\ &+ \left(M-R+\frac{1}{2}\right)\sum_{i=1}^{R}\log_2\lambda_i \\ &+ \sum_{i,j<i}\log_2(\lambda_i+\lambda_j)+\log_2(1+\mathcal{O}(\rho^{-1}))\end{aligned} \tag{76}$$

Since the eigenvalues of $\boldsymbol{\mathcal{H}}_0$ are equivalent, we need only to evaluate the averages $E\left[\log_2\lambda_1\right]$ and $E\left[\log_2(\lambda_1+\lambda_2)\right]$ over the $M\times N$ Gaussian matrix $\mathbf{Z}_+$. Careful analysis of the correction term shows that it is $\mathcal{O}(\log_2\rho/\rho)$.

To calculate these quantities we need the single eigenvalue probability density $\rho(\lambda)$ as well as the joint two eigenvalue probability density $\rho(\lambda_1,\lambda_2)$ for the random matrix $\boldsymbol{\mathcal{H}}_0=\mathbf{Z}_+^\dagger\mathbf{Z}_+$. Using Telatar's analysis [2], it can be shown that

$$\rho(\lambda) = \frac{\lambda^Q e^{-\lambda}}{R}\mu_2(\lambda,\lambda) \tag{77}$$

$$\begin{aligned}\rho(\lambda_1,\lambda_2) &= \frac{\lambda_1^Q\lambda_2^Q e^{-(\lambda_1+\lambda_2)}}{R(R-1)} \\ &\times \left(\mu_2(\lambda_1,\lambda_1)\mu_2(\lambda_2,\lambda_2)-\mu_2(\lambda_1,\lambda_2)^2\right)\end{aligned} \tag{78}$$

where $\mu_2(\lambda_1,\lambda_2)$ is given by

$$\mu_2(\lambda_1,\lambda_2)=\sum_{k=0}^{R-1}\frac{k!}{(k+Q)!}L_k^Q(\lambda_1)L_k^Q(\lambda_2) \tag{79}$$

and $L_k^Q(x)$ is the associated Laguerre polynomial of order k. Since both $\rho(\lambda)$ and $\rho(\lambda_1,\lambda_2)$ are finite polynomials in λ_1, λ_2 times a exponential factor, they can be explicitly integrated using the following identities several times:

$$\begin{aligned}\int_0^\infty d\lambda\,\lambda^n\log_2\lambda e^{-\lambda} &= n!\Psi(n+1) \\ &\equiv n!\left(1+\frac{1}{2}+\cdots\frac{1}{n}-\mathcal{C}\right)\log_2 e\end{aligned} \tag{80}$$

$$\begin{aligned}\int_0^\infty d\lambda_1\int_0^\infty d\lambda_2\,\lambda_1^n\lambda_1^m e^{-(\lambda_1+\lambda_2)}\log_2(\lambda_1+\lambda_2) \\ = n!m!\Psi(n+m+2)\log_2 e\end{aligned} \tag{81}$$

where $\mathcal{C}$ is the Euler constant $\mathcal{C} = 0.57721\cdots$. To somewhat simplify the procedure, we apply the Christofel-Darboux identity (see [15])

$$\begin{aligned}\mu_2(\lambda_1,\lambda_2) &= \sum_{k=0}^{R-1}\frac{k!}{(k+Q)!}L_k^Q(\lambda_1)L_k^Q(\lambda_2) \\ &= \frac{R!}{(R+Q-1)!} \\ &\times \frac{L_{R-1}^Q(\lambda_1)L_R^Q(\lambda_2)-L_R^Q(\lambda_2)L_{R-1}^Q(\lambda_1)}{\lambda_1-\lambda_2}\end{aligned} \tag{82}$$

which, in the limit $\lambda_2 \to \lambda_1$ becomes

$$\begin{aligned}\mu_2(\lambda_1,\lambda_1) &= \frac{R!}{(R+Q-1)!} \\ &\times \left(L_{R-1}^Q(\lambda_1)L_{R-1}^{Q+1}(\lambda_1)-L_R^Q(\lambda_1)L_{R-2}^{Q+1}(\lambda_1)\right)\end{aligned} \tag{83}$$

Combining (77), (80) and (83), we get

$$\begin{aligned}\mathcal{L}_1(M,N) = E\left[\log\lambda\right] &= \frac{(R-1)!}{(R+Q-1)!} \\ \times\Bigg[\sum_{k,m=0}^{R-1}(-1)^{k+m}&\begin{pmatrix} Q+R-1 \\ R-1-k\end{pmatrix}\begin{pmatrix} Q+R \\ R-1-m\end{pmatrix} \\ -\sum_{k,m=0}^{R,R-2}(-1)^{k+m}&\begin{pmatrix} Q+R \\ R-k\end{pmatrix}\begin{pmatrix} Q+R-1 \\ R-2-m\end{pmatrix}\Bigg] \\ &\times\frac{(Q+k+m)!\Psi(Q+k+m+1)}{k!m!}\log_2 e\end{aligned} \tag{84}$$

Similarly, by combining (79), (81) and (82), we get

$$\begin{aligned}
\mathcal{L}_2(M,N) = E\left[\log_2\left(\lambda_1+\lambda_2\right)\right] = \qquad (85)\\
\left\{\left[\sum_{k_1,m_1=0}^{R-1}(-1)^{k_1+m_1}\left(\begin{array}{c}Q+R-1\\R-1-k_1\end{array}\right)\left(\begin{array}{c}Q+R\\R-1-m_1\end{array}\right)\right.\right.\\
\left.-\sum_{k_1,m_1=0}^{R,R-2}(-1)^{k_1+m_1}\left(\begin{array}{c}Q+R\\R-k_1\end{array}\right)\left(\begin{array}{c}Q+R-1\\R-2-m_1\end{array}\right)\right]\\
\times\left[\sum_{k_2,m_2=0}^{R-1}(-1)^{k_2+m_2}\left(\begin{array}{c}Q+R-1\\R-1-k_2\end{array}\right)\left(\begin{array}{c}Q+R\\R-1-m_2\end{array}\right)\right.\\
\left.\left.-\sum_{k_2,m_2=0}^{R,R-2}(-1)^{k_2+m_2}\left(\begin{array}{c}Q+R\\R-k_2\end{array}\right)\left(\begin{array}{c}Q+R-1\\R-2-m_2\end{array}\right)\right]\right\}\\
\times\frac{(Q+k_1+m_1)!(Q+k_2+m_2)!}{k_1!k_2!m_1!m_2!}\\
\times\Psi(2Q+k_1+m_1+k_2+m_2+2)\log_2 e\\
-\left\{\left[\sum_{k_1,m_1=0}^{R-1,R}(-1)^{k_1+m_1}\left(\begin{array}{c}Q+R-1\\R-1-k_1\end{array}\right)\left(\begin{array}{c}Q+R\\R-m_1\end{array}\right)\right.\right.\\
\left.\times\frac{sgn(k_1-m_1)}{k_1!m_1!}\sum_{p_1=0}^{|k_1-m_1|-1}\right]\\
\times\left[\sum_{k_2,m_2=0}^{R-1,R}(-1)^{k_2+m_2}\left(\begin{array}{c}Q+R-1\\R-1-k_2\end{array}\right)\left(\begin{array}{c}Q+R\\R-m_2\end{array}\right)\right.\\
\left.\left.\times\frac{sgn(k_2-m_2)}{k_2!m_2!}\sum_{p_2=0}^{|k_2-m_2|-1}\right]\right\}\\
\times(Q+\max(k_1,m_1)-1-p_1+\max(k_2,m_2)-1-p_2)!\\
\times(Q+\min(k_1,m_1)+\min(k_2,m_2))!\\
\times\Psi(2Q+k_1+m_1+k_2+m_2)\log_2 e
\end{aligned}$$

After collecting all terms we can now evaluate the DUSTM mutual information to order $\mathcal{O}(\log_2\rho/\rho)$.

B. Isotropic USTM

To analyze the large ρ behavior of mutual information of USTM, we start by writing the mutual information as

$$\begin{aligned}
I_{USTM} &= E\left[\log_2\frac{p(\mathbf{X}|\mathbf{\Phi}_0)}{p(\mathbf{X})}\right] \qquad (86)\\
&= \alpha E\left[\mathrm{Tr}\left\{\mathbf{X}^\dagger\mathbf{J}_M\mathbf{X}\right\}\right]\log_2 e\\
&- E\left[\log_2\langle\exp\left(\alpha\mathrm{Tr}\left\{\mathbf{X}^\dagger\mathbf{\Phi}\mathbf{\Phi}^\dagger\mathbf{X}\right\}\right)\rangle\right]\\
&= TN\rho\log_2 e - E\left[\log_2\langle\exp\left(\alpha\mathrm{Tr}\left\{\mathbf{X}^\dagger\mathbf{\Phi}\mathbf{\Phi}^\dagger\mathbf{X}\right\}\right)\rangle\right]
\end{aligned}$$

where the third equality is obtained by integrating over $\mathbf{X}$, see (55). To evaluate the second term we will perform an asymptotic analysis of the multiple integration in (29), which is performed by evaluating the residues of the poles of the t-integrals. We will assume that $T > M$, since otherwise the mutual information is identically zero. We also use the fact that at large ρ from (32), the eigenvalues of $\mathbf{X}^\dagger\mathbf{X}$ generally split into three groups: the first R being large $\mathcal{O}(\rho^{-1})$, $K-R$ eigenvalues being $\mathcal{O}(1)$, while the remaining $N-K$ being zero. For simplicity, we assume they are ordered in magnitude, i.e. $y_1 \geq y_2 \geq \ldots$. Note first that the last term in (29) guarantees that no two t_i's are evaluated at the residue of the same pole with $y_n \neq 0$. As a result the leading term will entail $\min(K,M)$ t's evaluated at the poles of the $\mathcal{O}(\rho)$ eigenvalues of $\mathbf{X}^\dagger\mathbf{X}$. All other terms will be exponentially smaller. Let us start with the simpler case of $M < K$. Here the M t-integrals are all performed by taking their residues at the M $\mathcal{O}(\rho)$ y_i's. Thus we get

$$\begin{aligned} & \langle \exp\left(\alpha \mathrm{Tr}\left\{\mathbf{X}^\dagger \mathbf{\Phi}\mathbf{\Phi}^\dagger \mathbf{X}\right\}\right)\rangle \\ \approx\ & C_{TM} \prod_{m=1}^{M} \frac{e^{\alpha y_m}}{\prod_{q=1,q\neq m}^{K}(\alpha y_m - \alpha y_q)(\alpha y_m)^{T-K}} \\ & \times \prod_{l<m} (\alpha y_l - \alpha y_m)^2 \\ =\ & C_{TM} \prod_{m=1}^{M} \left[\frac{e^{\alpha y_m}}{\prod_{q=m+1}^{K}(\alpha y_m - \alpha y_q)(\alpha y_m)^{T-K}}\right] \\ \approx\ & C_{TM} \prod_{m=1}^{M} \left[\frac{e^{\alpha y_m}}{(\alpha y_m)^{T-M}}\right] \left(1+\mathcal{O}(\rho^{-1})\right) \end{aligned} \tag{87}$$

where in the last step we used the fact that the eigenvalues $y_{M+1},\ldots,y_K$ are $\mathcal{O}(1)$, while y_m for $m=1,\ldots,M$ are $y_m = \mathcal{O}(\rho)$. Thus for $M \leq K$ the mutual information can be written as

$$\begin{aligned} I_{USTM} = & \left[TN\rho - \alpha \sum_{m=1}^{M} E\left[y_m\right]\right] \log_2 e - \log_2 C_{T,M} \\ & + (T-M) \sum_{m=1}^{M} E\left[\log_2(\alpha y_m)\right] + \mathcal{O}(\log_2 \rho/\rho) \end{aligned} \tag{88}$$

Using a similar analysis as in Appendix III-A, it can be shown that for $M \leq K$

$$\alpha \sum_{m=1}^{M} E\left[y_m\right] = TN\rho + (T-M)R + \mathcal{O}(\rho^{-1}) \tag{89}$$

To calculate the expectation of $\log_2 \alpha y_m$, we note that to leading order we have $y_m \approx \rho T \lambda_m/M + \mathcal{O}(1)$, where λ_m are the eigenvalues of $\mathbf{X}_1^\dagger\mathbf{X}_1$, with $\mathbf{X}_1$ a $M \times N$ Gaussian random, unit-variance matrix. Thus we can use (84).

When $M > K \equiv N$, we have the added complexity that only K y_n's are $\mathcal{O}(\rho)$. After performing the first K

t-integrals by evaluating them at the poles of these K $\mathcal{O}(\rho)$ y's, (29) becomes

$$\begin{aligned} & \langle \exp\left(\alpha \mathrm{Tr}\left\{\mathbf{X}^\dagger \mathbf{\Phi}\mathbf{\Phi}^\dagger \mathbf{X}\right\}\right)\rangle \\ \approx \quad & \frac{C_{TM}}{(M-K)!} \prod_{m=1}^{K} \frac{e^{\alpha y_m}}{(\alpha y_m)^{T-K}} \\ & \times \prod_{m=K+1}^{M} \int \frac{dt_m}{2\pi} \frac{e^{-i\lambda_m} \prod_{q=1}^{K}(-\alpha y_q - i t_m)}{(-i\lambda_m)^{T-K}} \\ & \times \prod_{l>m} (-i\lambda_m - i\lambda_l)^2 \end{aligned} \tag{90}$$

The $M-K$ remaining integrals have high-order poles at zero. It is straightforward to show that the above equation becomes

$$\begin{aligned} & \langle \exp\left(\alpha \mathrm{Tr}\left\{\mathbf{X}^\dagger \mathbf{\Phi}\mathbf{\Phi}^\dagger \mathbf{X}\right\}\right)\rangle \\ \approx \quad & C_{TM} \prod_{m=1}^{K} \frac{e^{\alpha y_m}}{(\alpha y_m)^{T-M}} \left|\det \mathbf{G}\right| \left(1 + \mathcal{O}(\rho^{-1})\right) \end{aligned} \tag{91}$$

where $\mathbf{G}$ is an $(M-K)$-dimensional square Hankel matrix with elements

$$G_{mn} = \begin{cases} \frac{1}{(T-K-m-n+1)!} & m+n \leq T-K+1 \\ 0 & \text{otherwise} \end{cases} \tag{92}$$

As a result, for $M > N$ and large ρ the mutual information is asymptotically equal to

$$\begin{aligned} I_{USTM} &= (T-M)R\left[\log_2 \frac{\rho T}{Me} + \mathcal{L}_1\right] - \log_2 C_{T,M} \\ &- \log_2 \left|\det \mathbf{G}\right| + \mathcal{O}(\log_2 \rho/\rho) \end{aligned} \tag{93}$$

REFERENCES

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

[2] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications and Related Technoilogies*, vol. 10, no. 6, pp. 585–596, November 1999.

[3] T. L. Marzetta and B. M. Hochwald, "Capacity of mobile multiple-antenna communication link in a Rayleigh flat-fading environment," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.

[4] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–565, Mar. 2000.

[5] L. Zheng and D. N. C. Tse, "Communication on the grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 359–383, Feb 2002.

[6] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: The received signal density in closed form," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1473–1484, June 2002.

[7] J. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill Press, 2001.

[8] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2041–2052, Dec. 2000.

[9] B. L. Hughes, "Differential space-time modulation," *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2567–2578, Nov 2000.

[10] T. L. Marzetta, "Differential Unitary Space-Frequency Modulation," in *Proceedings 41st Annual Allerton Conference on Communication, Control, and Computing*, Allerton Conference Center, Oct. 1-3 2003, pp. 344–345.

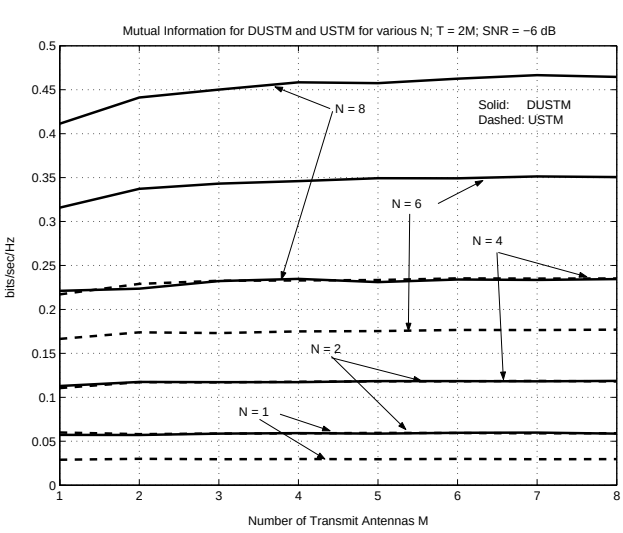
[11] R. C. Brower, P. Rossi, and C.-I. Tan, "Chiral chains for lattice quantum chromodynamics at $N_c = \infty$," *Physical Review D*, vol. 23, no. 4, pp. 942–952, February 1981.

[12] W. M. Boothby, *An Introduction to Differential Manifolds and Riemannian Geometry*, 2nd ed. San Diego, CA: Academic, 1986.

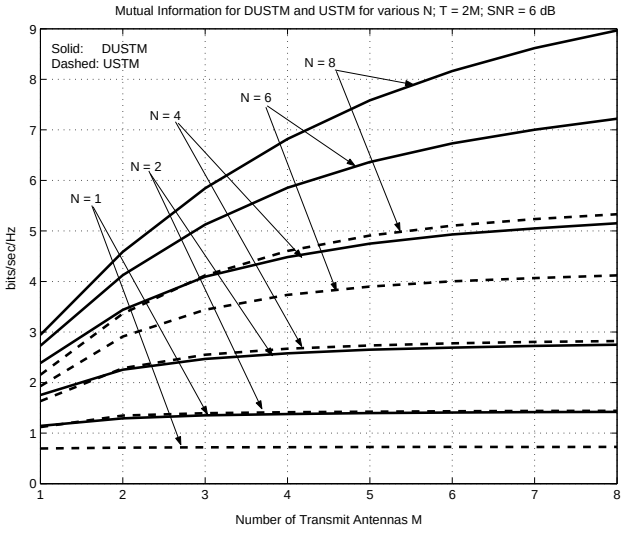
[13] B. Hassibi and Y. Jing, "Unitary space-time modulation via the cayley transform," in *Proceeding 2002 IEEE International Symposium on Information Theory*, Lausanne, Switzerland, June 2001, pp. 1663–1667.

[14] C. Cohen-Tannoudji, F. Laloe, and B. Diu, *Quantum Mechanics*. New York: John Wiley and Sons, 1977.

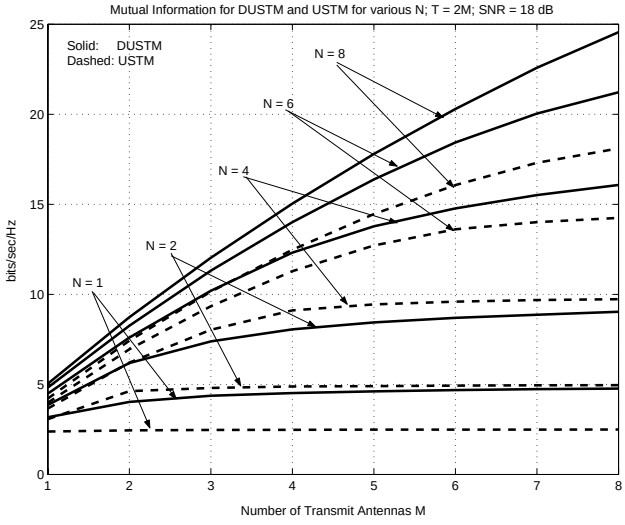
[15] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*. New York: Academic Press, 1995.



(a) $\rho = -6.0dB$

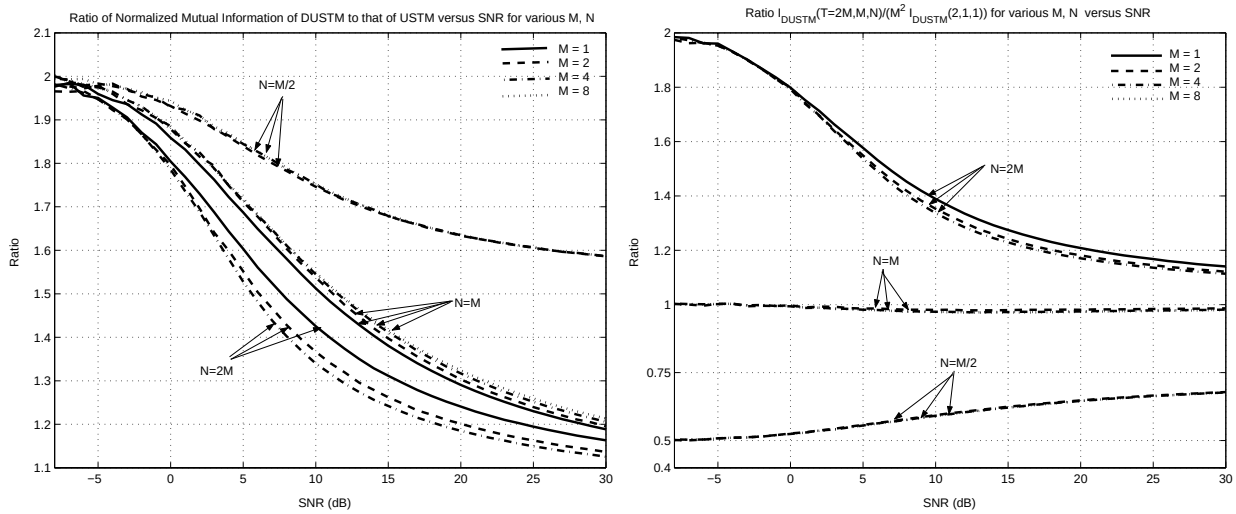


(b) $\rho = 6.0dB$



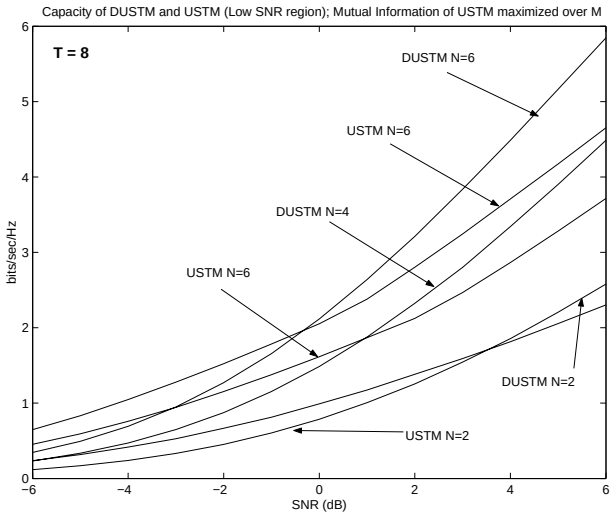
(c) $\rho = 18.0dB$

Fig. 1. Plot of normalized mutual information $\hat{I}_{DUSTM}$ (solid lines) and $\hat{I}_{USTM}$ (dashed lines) as a function of the number of transmit antennas M for different receive antenna numbers N and three SNR levels ρ. The coherence interval T is chosen to be $T = 2M$ for proper comparison.

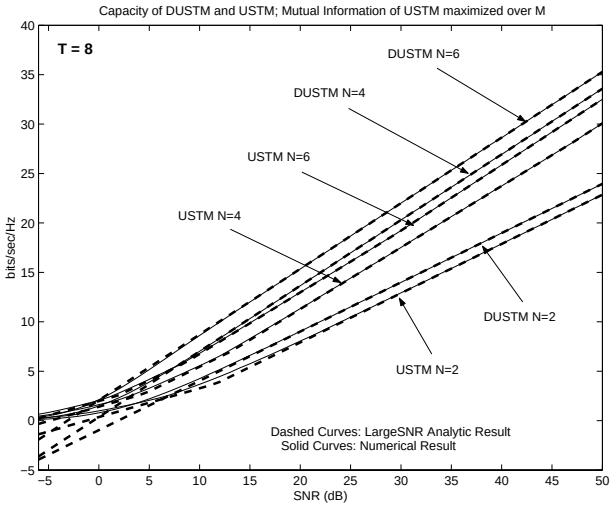


(a) $\hat{I}_{IUST}(T = 2M, M, N)/\hat{I}_{DUST}(T = 2M, M, N)$ (b) $\hat{I}_{DUST}(T = 2M, M, N)/(M\hat{I}_{DUST}(T = 2, 1, 1))$

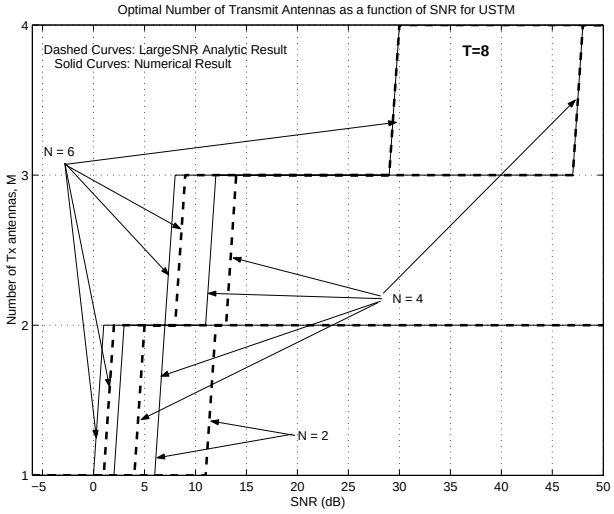
Fig. 2. (a) Plot of the ratio between the mutual information per symbol for the Differential USTM and the Isotropic USTM (with $T = 2M$) as a function of SNR. For low SNR, the ratio approaches 2, as seen in the previous figure and in agreement with (33), (34). For large SNR the ratio appears to approach 1 (for $M = N$ and $M = N/2$) and 1.5 (for $M = 2N$), as predicted from (35), (37). For $M = N$ and $M = 2N$, the ratio appears not to depend on the number of antennas for any intermediate ρ. (b) Plot of ratio between the mutual information per symbol for the Differential USTM for various antenna numbers to that for $M = N = 1$.



(a) Capacity–small ρ region



(b) Capacity–full region



(c) Optimal number of M for USTM

Fig. 3. Plot of capacity of DUSTM and USTM as a function of ρ. Here T is fixed to $T = 8$. Fig. 3(a) depicts the small-ρ region, while Fig. 3(b) shows the full region of ρ-values simulated. Fig. 3(c) represents the optimal number of transmitting antennas for USTM as a function of SNR and for various values of N. The dashed curves in Fig. 3(b),(c) correspond to the analytic expressions valid for large ρ (see (35), (37)).