

Multiple-Antennas and Isotropically Random Unitary Inputs: The Received Signal Density in Closed Form

Babak Hassibi and Thomas L. Marzetta, *Senior Member, IEEE*

Invited Paper

Abstract—An important open problem in multiple-antenna communications theory is to compute the capacity of a wireless link subject to flat Rayleigh block-fading, with no channel-state information (CSI) available either to the transmitter or to the receiver. The isotropically random (i.r.) unitary matrix—having orthonormal columns, and a probability density that is invariant to premultiplication by an independent unitary matrix—plays a central role in the calculation of capacity and in some special cases happens to be capacity-achieving. In this paper, we take an important step toward computing this capacity by obtaining, in *closed form*, the probability density of the received signal when transmitting i.r. unitary matrices. The technique is based on analytically computing the expectation of an exponential quadratic function of an i.r. unitary matrix and makes use of a Fourier integral representation of the constituent Dirac delta functions in the underlying density. Our formula for the received signal density enables us to evaluate the mutual information for any case of interest, something that could previously only be done for single transmit and receive antennas. Numerical results show that at high signal-to-noise ratio (SNR), the mutual information is maximized for $M = \min(N, T/2)$ transmit antennas, where N is the number of receive antennas and T is the length of the coherence interval, whereas at low SNR, the mutual information is maximized by allocating all transmit power to a single antenna.

Index Terms—Isotropically random (i.r.) unitary matrix, multiple antennas, unitary space–time modulation (USTM), wireless communications.

I. INTRODUCTION

WE consider a single-user multiple-antenna wireless link, subject to flat Rayleigh block-fading, with no channel-state information (CSI) available either to the transmitter or to the receiver. The assumption of no CSI—as opposed to CSI available at the receiver, for example—converts a straightforward problem in Shannon theory [1]–[3] into a rather difficult, albeit more realistic, one [4]–[8]. The capacity-attaining signals

are non-Gaussian, the maximum-likelihood receiver is noncoherent, and there is a hard upper limit on the number of usable transmit antennas.

The isotropically random unitary matrix—having orthonormal columns, and a probability density that is invariant to premultiplication by an independent unitary matrix—plays a central role in the theory of our multiple-antenna link [6]. The capacity-attaining input signal is the product of an isotropically random unitary matrix, and an independent nonnegative real diagonal matrix. In certain limiting regimes [6]–[8], the diagonal matrix is constant, and the message is carried entirely by the unitary matrix: a type of modulation called unitary space–time (USTM) [7]. A number of practical considerations make USTM attractive for general usage.

It follows that calculations of capacity or mutual information invariably require taking expectations with respect to an isotropically random unitary matrix. This was done for single transmit and receive antennas [6], but has not been done for more complicated problems. The contribution of this paper is a technique for taking this expectation analytically for certain cases. In particular, when the input signal is isotropically random unitary we obtain a closed-form expression, in the form of a determinant, for the probability density of the received signal. When combined with a simple Monte Carlo integration, it is possible to compute mutual information for any case of interest.

Our computation of the mutual information resulting from an isotropically random unitary input yields a *lower* bound on capacity that is increasingly tight as the signal-to-noise ratio (SNR) grows [8]. At low SNR, however, it is well known that all the information is carried by the diagonal matrix that determines the magnitude of the transmitted matrix signal, and not by the isotropic unitary matrix that determines its directionality. At moderate values of SNR, some of the information is carried by the diagonal matrix and some by the unitary matrix (although a simple count of the degrees of freedom— M , as opposed to $(2TM - M^2)$ —suggests that most of the information will be carried by the unitary matrix). Finally, our results have ramifications elsewhere: for obtaining *upper* bounds on capacity by appealing to divergence-based methodologies (see, e.g., [9]), and for upper bounds on the mutual information of certain unitary space–time constellations [10].

Section II reviews the signal model, and known results. Section III contains a derivation of the probability density of the received signal. Our technique uses a Fourier representation for

Manuscript received March 7, 2001; revised November 2, 2001. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Washington DC, June 2001.

B. Hassibi was with the Mathematical Sciences Research Center, Bell Laboratories, Lucent Technologies, Murray Hills, NJ 07974 USA. He is now with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: hassibi@caltech.edu).

T. L. Marzetta is with the Mathematical Sciences Research Center, Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA (e-mail: tlm@research.bell-labs.com).

Communicated by S. Shamai, Guest Editor.

Publisher Item Identifier S 0018-9448(02)04011-7.

the Dirac delta functions that comprise the isotropic unitary density. An integration with respect to the received signal, which we cannot perform analytically, yields mutual information (Section IV). We perform this integration via Monte Carlo in Section V to compute mutual information. Appendix A reviews the isotropically random unitary density, and derives its Fourier-based version.¹ Appendix B reviews a technique for converting a particular M -dimensional integral into a determinant of an $M \times M$ Hankel matrix, whose entries are one-dimensional integrals. Finally, Appendix C discusses a certain complex exponential quadratic integral.

A. Notations and Definitions

This paper makes significant use of matrix notation. $\text{tr}(A)$ and $\det(A)$ denote the trace and determinant of a square matrix A . The superscript \dagger denotes conjugate transposition, so that A^\dagger is the conjugate transpose of A . For scalars, the complex conjugate is denoted by the superscript $*$. I_T is the $T \times T$ identity matrix. For a complex-valued argument the Dirac delta function is defined as $\delta(x) = \delta(\text{Real}(x))\delta(\text{Imag}(x))$. For a complex-valued $M \times N$ matrix A , we have

$$\begin{aligned}\delta(A) &= \prod_{m=1}^M \prod_{n=1}^N \delta(a_{mn}) \\ &= \prod_{m=1}^M \prod_{n=1}^N \delta(\text{Real}(a_{mn}))\delta(\text{Imag}(a_{mn}))\end{aligned}$$

and for a $M \times M$ Hermitian matrix A , we have

$$\begin{aligned}\delta(A) &= \prod_{m=1}^M \delta(a_{mm}) \prod_{n=m+1}^M \delta(a_{mn}) \\ &= \prod_{m=1}^M \delta(a_{mm}) \prod_{n=m+1}^M \delta(\text{Real}(a_{mn}))\delta(\text{Imag}(a_{mn})).\end{aligned}$$

Finally, $\Gamma(x) = \int_0^\infty \lambda^{x-1} e^{-\lambda} d\lambda$ is the gamma function, and $\text{CN}(0, 1)$ denotes a scalar, zero-mean, circularly symmetric, complex Gaussian with unit variance.

II. SIGNAL MODEL; MUTUAL INFORMATION

A. Signal Model

Consider a single-user block-fading multiple-antenna link with M transmit and N receive antennas described by a propagation matrix that is constant during coherence intervals of length T symbols, after which it jumps to a new independent value for T more symbols, and so on [6]. This is a reasonable model for systems employing some form of time-division multiplexing, or frequency hopping. During any coherence interval, a $T \times M$ complex matrix S is transmitted, and a $T \times N$ complex matrix X is received

$$X = \sqrt{\frac{\rho}{M}} SH + W \quad (1)$$

¹Throughout the paper, all probability densities are with respect to Lebesgue measure, rather than Haar measure.

where H is an $M \times N$ propagation matrix, and W is a $T \times N$ additive noise matrix. The elements of H and W are independent $\text{CN}(0, 1)$ -distributed. The values of H and W are unknown to both the transmitter and the receiver. At each time t , the transmitted signal has, on average, unit-variance entries

$$\mathbb{E} \left\{ \sum_{m=1}^M |s_{tm}|^2 \right\} = M. \quad (2)$$

Since S , H , and W are independent, the normalization $\sqrt{\frac{\rho}{M}}$ in (1) implies that ρ is equal to the SNR at each receive antenna, independently of M .

B. Mutual Information

The independence of H and W over different coherence intervals implies that Shannon coding, applied over many coherence intervals, yields reliable transmission at any rate less than the mutual information

$$I(X; S) = \mathbb{E} \left\{ \log \frac{p(X|S)}{p(X)} \right\}. \quad (3)$$

Since X given S is zero-mean and Gaussian (it is the sum of two zero-mean random Gaussian matrices, SH and W), a covariance computation shows that the conditional density $p(X|S)$ is

$$p(X|S) = \frac{1}{\pi^{TN}} \cdot \frac{\exp \left[-\text{tr} X^\dagger \left(I_T + \frac{\rho}{M} SS^\dagger \right)^{-1} X \right]}{\det \left(I_T + \frac{\rho}{M} SS^\dagger \right)^N}. \quad (4)$$

The maximization of the mutual information over the distribution of S , subject to the power constraint (2), yields the capacity. There is no point in making the number of transmit antennas greater than the symbol duration of the coherence interval [6], so we assume that $M \leq T$. The capacity-attaining signal is the product of two independent random matrices [6]

$$S = \sqrt{T} \Phi D \quad (5)$$

where Φ is a $T \times M$ isotropically distributed unitary matrix, and D is an $M \times M$ nonnegative real diagonal matrix. The matrix Φ has orthonormal columns $\Phi^\dagger \Phi = I_M$, and the joint probability density of its elements is unchanged when Φ is multiplied by any $T \times T$ unitary matrix Ψ that is independent of Φ , i.e., $p(\Phi) = p(\Psi\Phi)$. See [6], or Appendix A, for a derivation of this density. The power constraint (2), combined with (5), implies that

$$\mathbb{E} \{ \text{tr} D^2 \} = M. \quad (6)$$

Except for the case $T = M = N = 1$ [4], [5], and for certain limiting cases, little is known concerning the capacity-attaining joint density of the diagonal elements of D .

For the two limiting cases: 1) $T \gg M$ [6], or 2) $M \leq \min(T/2, N)$ and $\rho \gg 1$ [8], capacity is attained when $D = I_M$, and the optimal transmitted signal is proportional to an isotropically random unitary matrix. The transmission of orthonormal signals is called unitary space-time modulation (USTM) [7], and its general use is motivated by several practical considerations [6], [11]. For either of the limiting cases above, mutual information can be evaluated indirectly through

asymptotic formulas for capacity [6], [2], [8]. In general, however, the optimizing joint distribution on D is not known. In fact, even the mutual information obtained by $D = I_M$ in the general case was not known and is the main focus of this paper.

C. Difficulties in Computing Mutual Information

The combination of (3) and (5) yields a formidable multiple integration

$$I(X; S) = \int dX \int d\Phi \int dD p(X | \Phi, D) \cdot p(\Phi) \cdot p(D) \cdot \log \left\{ \frac{p(X | \Phi, D)}{\int d\hat{\Phi} \int d\hat{D} p(X | \hat{\Phi}, \hat{D}) \cdot p(\hat{\Phi}) \cdot p(\hat{D})} \right\}. \quad (7)$$

The term that arises from the log-numerator reduces to an expectation with respect to D . Most of the trouble results from the log-denominator.

The outer integration of (7) with respect to Φ may be eliminated as follows. For any $T \times T$ unitary matrix Ψ , a straightforward computation shows (4) that $p(X | S) = p(\Psi^\dagger X | \Psi^\dagger S)$ (see also [6]). Now use this property for $\Psi = [\Phi \ \Phi_\perp]$, where Φ_\perp is the $T \times (T - M)$ orthogonal complement to Φ , and perform the following change of variable in (7): $\Psi^\dagger X \rightarrow X$. This yields

$$\begin{aligned} I(X; S) &= \int dX \int d\Phi \int dD p(\Psi^\dagger X | \Phi_0, D) \cdot p(\Phi) \cdot p(D) \\ &\quad \cdot \log \left\{ \frac{p(X | \Phi, D)}{\int d\hat{\Phi} \int d\hat{D} p(X | \hat{\Phi}, \hat{D}) \cdot p(\hat{\Phi}) \cdot p(\hat{D})} \right\} \\ &= \int dX \int d\Phi \int dD p(X | \Phi_0, D) \cdot p(\Phi) \cdot p(D) \\ &\quad \cdot \log \left\{ \frac{p(\Psi X | \Phi, D)}{\int d\hat{\Phi} \int d\hat{D} p(\Psi X | \hat{\Phi}, \hat{D}) \cdot p(\hat{\Phi}) \cdot p(\hat{D})} \right\} \\ &= \int dX \int d\Phi \int dD p(X | \Phi_0, D) \cdot p(\Phi) \cdot p(D) \\ &\quad \cdot \log \left\{ \frac{p(X | \Phi_0, D)}{\int d\hat{\Phi} \int d\hat{D} p(X | \Psi^\dagger \hat{\Phi}, \hat{D}) \cdot p(\hat{\Phi}) \cdot p(\hat{D})} \right\} \end{aligned}$$

where

$$\Phi_0 \triangleq \Psi^\dagger \Phi = \begin{bmatrix} I_M \\ 0 \end{bmatrix}. \quad (8)$$

Now make the change of variable $\Psi^\dagger \hat{\Phi} \rightarrow \hat{\Phi}$, and use the fact that $\hat{\Phi}$ is isotropically distributed to obtain

$$I(X; S) = \int dX \int dD p(X | \Phi_0, D) \cdot p(D) \cdot \log \left\{ \frac{p(X | \Phi_0, D)}{\int d\hat{\Phi} \int d\hat{D} p(X | \hat{\Phi}, \hat{D}) \cdot p(\hat{\Phi}) \cdot p(\hat{D})} \right\}. \quad (9)$$

Note that in the above expression we have easily integrated out Φ since the only dependence of the integrand on Φ was through $p(\Phi)$.

The log-numerator term in (9) is

$$\begin{aligned} E\{\log p(X | \Phi_0, D)\} &= -TN \log(\pi e) - NE \left\{ \log \det(I_T + \rho\beta\Phi_0 D^2 \Phi_0^\dagger) \right\} \\ &= -TN \log(\pi e) - NE \left\{ \log \det(I_M + \rho\beta D^2 \Phi_0^\dagger \Phi_0) \right\} \\ &= -TN \log(\pi e) - NE \left\{ \log \det(I_M + \rho\beta D^2) \right\} \quad (10) \end{aligned}$$

where we define $\beta \triangleq T/M$.

One further simplification of (9) occurs, because the logarithm depends only on the eigenvalues of XX^\dagger , as we discuss in the next section.

All of the above simplifications, combined with numerical integration, were used to evaluate capacity and mutual information for cases where $M = N = 1$ and $T > 1$ [6]. Classical numerical integration of the quadrature type is helpless when dealing with more than a few integration variables, so the same technique cannot be applied to bigger problems. Another possibility—not investigated in this paper—is Monte Carlo integration. However, this could only be used in conjunction with some clever importance sampling, for the reason that $p(X | \Phi, D)$ is sharply peaked with respect to Φ . In the next section, we show how to calculate analytically the inner integral with respect to the isotropically distributed unitary matrix for the case where $D = I_M$.

III. COMPUTING $p(X)$

Using (4) and (5), we may write

$$\begin{aligned} p(X|S) &= \frac{1}{\pi^{NT}} \cdot \frac{\exp[-\text{tr } X^\dagger (I_T + \rho\beta\Phi D^2 \Phi^\dagger)^{-1} X]}{\det(I_T + \rho\beta\Phi D^2 \Phi^\dagger)^N} \\ &= \frac{1}{\pi^{NT}} \cdot \frac{\exp\left\{-\text{tr} \left[X^\dagger X - X^\dagger \Phi \left(\Phi^\dagger \Phi + \frac{1}{\rho\beta} D^{-2} \right)^{-1} \Phi^\dagger X \right] \right\}}{\det(I_T + \rho\beta D^2 \Phi^\dagger \Phi)^N} \\ &= \frac{1}{\pi^{NT}} \cdot \frac{\exp(-\text{tr } X^\dagger X)}{\det(I_M + \rho\beta D^2)^N} \\ &\quad \cdot \exp\left[\text{tr } X^\dagger \Phi \left(I_M + \frac{1}{\rho\beta} D^{-2} \right)^{-1} \Phi^\dagger X \right]. \quad (11) \end{aligned}$$

This implies that

$$p(X|D) = \frac{1}{\pi^{NT}} \cdot \frac{\exp(-\text{tr } X^\dagger X)}{\det(I_M + \rho\beta D^2)^N} \cdot E_{|D} \exp\left[\text{tr } X^\dagger \Phi \left(I_M + \frac{1}{\rho\beta} D^{-2} \right)^{-1} \Phi^\dagger X \right] \quad (12)$$

where the expectation is taken over the isotropically distributed unitary matrix Φ . Consider now the eigenvalue decomposition of XX^\dagger

$$XX^\dagger = U \underbrace{\begin{bmatrix} \Sigma & 0_{K \times (T-K)} \\ 0_{(T-K) \times K} & 0_{(T-K) \times (T-K)} \end{bmatrix}}_{\triangleq A} U^\dagger$$

where U is $T \times T$ and unitary, Σ is $K \times K$ real positive diagonal, and $K \triangleq \min(T, N)$. Then

$$\begin{aligned} E_{|D} \exp \left[\text{tr} X^\dagger \Phi \left(I_M + \frac{1}{\rho\beta} D^{-2} \right)^{-1} \Phi^\dagger X \right] \\ &= E_{|D} \exp \left[\text{tr} XX^\dagger \Phi \left(I_M + \frac{1}{\rho\beta} D^{-2} \right)^{-1} \Phi^\dagger \right] \\ &= E_{|D} \exp \left[\text{tr} UAU^\dagger \Phi \left(I_M + \frac{1}{\rho\beta} D^{-2} \right)^{-1} \Phi^\dagger \right] \\ &= E_{|D} \exp \left[\text{tr} AU^\dagger \Phi \left(I_M + \frac{1}{\rho\beta} D^{-2} \right)^{-1} \Phi^\dagger U \right] \\ &= E_{|D} \exp \left[\text{tr} A\Phi \left(I_M + \frac{1}{\rho\beta} D^{-2} \right)^{-1} \Phi^\dagger \right] \end{aligned}$$

where to obtain the last equality, we have used the fact that, since Φ is isotropically distributed, $U^\dagger \Phi$ is also unitary and isotropically distributed. Defining $B \triangleq (I_M + \frac{1}{\rho\beta} D^{-2})^{-1}$, we note that to compute $p(X|D)$ we need to compute $E\{\exp(\text{tr} A\Phi B\Phi^\dagger)\}$, where the expectation is taken over the $T \times M$ isotropically distributed unitary matrix Φ . For reasons to be made clear shortly, it will be convenient to transform the diagonal matrix $A \geq 0$ into a negative-definite matrix via the following transformation. Choose the scalar $a > 0$ such that $aI_T > A$, then

$$\begin{aligned} E e^{\text{tr} A\Phi B\Phi^\dagger} &= e^{\text{tr} aB} E e^{\text{tr} A\Phi B\Phi^\dagger} e^{-\text{tr} aB} \\ &= e^{\text{tr} aB} E e^{\text{tr} A\Phi B\Phi^\dagger} e^{-\text{tr} aB\Phi^\dagger \Phi} \\ &= e^{\text{tr} aB} E e^{\text{tr} A\Phi B\Phi^\dagger} e^{-\text{tr} a\Phi B\Phi^\dagger} \\ &= e^{\text{tr} aB} E e^{-\text{tr}(aI_T - A)\Phi B\Phi^\dagger} \end{aligned}$$

which transforms $A \geq 0$ to $A - aI_T < 0$.

Now denoting the distribution on Φ by $p(\Phi)$, we need to compute

$$E e^{\text{tr} A\Phi B\Phi^\dagger} = e^{\text{tr} aB} \int d\Phi p(\Phi) e^{-\text{tr}(aI_T - A)\Phi B\Phi^\dagger}.$$

Although it is possible to give an explicit formula for the density as $p(\Phi) = c\delta(\Phi^\dagger \Phi - I_M)$, where c is an appropriate normalization constant, direct integration of the above expression using the Dirac delta representation of the density appears to be formidable. To compute this integral, it is useful to introduce the following integral formula for $p(\Phi)$:

$$p(\Phi) = \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{2^M \pi^{\frac{M}{2}(2T+M+1)}} \int d\Omega \exp[i \text{tr} \Omega (\Phi^\dagger \Phi - I_M)] \quad (13)$$

where the integration is over the $M \times M$ Hermitian matrix variable $\Omega = \Omega^\dagger$. (For a proof see Appendix A.)

We may, therefore, write

$$\begin{aligned} E e^{\text{tr} A\Phi B\Phi^\dagger} &= e^{\text{tr} aB} \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{2^M \pi^{\frac{M}{2}(2T+M+1)}} \\ &\quad \cdot \int d\Phi e^{-\text{tr}(aI_T - A)\Phi B\Phi^\dagger} \int d\Omega e^{i \text{tr} \Omega (\Phi^\dagger \Phi - I_M)} \\ &= e^{\text{tr} aB} \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{2^M \pi^{\frac{M}{2}(2T+M+1)}} \\ &\quad \cdot \int d\Omega e^{-i \text{tr} \Omega} \int d\Phi e^{-\text{tr}[(aI_T - A)\Phi B\Phi^\dagger - \Phi i\Omega \Phi^\dagger]}. \end{aligned} \quad (14)$$

Note that the inner integration over Φ is feasible, since Φ is now an unconstrained matrix variable. To facilitate this integration, the expression $\text{tr}[(aI_T - A)\Phi B\Phi^\dagger - \Phi i\Omega \Phi^\dagger]$ can be reorganized as a quadratic form in the entries of Φ

$$\begin{aligned} \text{tr}[(aI_T - A)\Phi B\Phi^\dagger - \Phi i\Omega \Phi^\dagger] \\ &= \sum_{t=1}^T \sum_{m_1=1}^M \sum_{m_2=1}^M \phi_{tm_1} \phi_{tm_2}^* [(a - a_t)b_{m_1} \delta_{m_1 m_2} - i\omega_{m_1 m_2}] \\ &= \sum_{t=1}^T [\phi_{t1} \cdots \phi_{tM}] [(a - a_t)B - i\Omega] \begin{bmatrix} \phi_{t1}^* \\ \vdots \\ \phi_{tM}^* \end{bmatrix} \end{aligned} \quad (15)$$

where $\{a_t\}$ and $\{b_m\}$ denote the diagonal elements of A and B , respectively. Denoting the t th row of Φ by ϕ_t , then the integral over Φ in (14) is the product of T elementary integrals

$$\begin{aligned} \int d\Phi e^{-\text{tr}[(aI_T - A)\Phi B\Phi^\dagger - \Phi i\Omega \Phi^\dagger]} \\ &= \prod_{t=1}^T \int d\phi_t e^{-\phi_t [(a - a_t)B - i\Omega] \phi_t^*}. \end{aligned}$$

Note that the parameter a has been chosen so that $a - a_t > 0$ and hence $(a - a_t)B > 0$. This implies that each of the integrals converges absolutely for all Hermitian Ω and so the result is given by

$$\int d\Phi e^{-\text{tr}[(aI_T - A)\Phi B\Phi^\dagger - \Phi i\Omega \Phi^\dagger]} = \frac{\pi^{TM}}{\prod_{t=1}^T \det[(a - a_t)B - i\Omega]}. \quad (16)$$

(See Appendix C for an explanation.)

We substitute (16) into (14) to obtain

$$\begin{aligned} E e^{\text{tr} A\Phi B\Phi^\dagger} &= e^{\text{tr} aB} \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{2^M \pi^{\frac{M}{2}(M+1)}} \\ &\quad \cdot \int \frac{d\Omega e^{-i \text{tr} \Omega}}{\det[B(a - a_1) - i\Omega] \cdots \det[B(a - a_T) - i\Omega]}. \end{aligned} \quad (17)$$

Recall in our case, that

$$A = \text{diag}\{\sigma_1, \dots, \sigma_K, 0, \dots, 0\}$$

where the σ_k are the nonzero eigenvalues of XX^\dagger and $K = \min(T, N)$, and that $B = (I_M + \frac{1}{\rho\beta}D^{-2})^{-1}$. Therefore, we have (18), shown at the bottom of the page.

For the remainder of the paper, rather than consider an arbitrary diagonal B , we confine ourselves to the case of USTM.

A. The Case of Isotropically Distributed USTM

When $D = I_M$, the input signal takes on the form of the (scaled) isotropically distributed unitary matrix $S = \sqrt{T}\Phi$. In this case, we have

$$B = \frac{1}{1 + \frac{1}{\rho\beta}} I_M = \alpha I_M, \quad \alpha \triangleq \frac{\rho\beta}{1 + \rho\beta} \quad (19)$$

and therefore we have (20), also shown at the bottom of the page.

To carry out the integration in the above expression, it is useful to introduce the eigenvalue decomposition $\Omega = U\Lambda U^\dagger$ of the Hermitian matrix Ω . This eigenvalue decomposition can be regarded as a change of variables from Ω to $\{U, \Lambda\}$. Note that the integrand depends only on the eigenvalues. It is a well-known result in random matrix theory (see, e.g., [12], [13]) that, for any function $g(\Omega)$ whose value depends only on the eigenvalues of Ω (and not on the eigenvectors)

$$\begin{aligned} & \int d\Omega g(\Omega) \\ &= \int d\Lambda \frac{\pi^{M(M-1)/2}}{\Gamma(M+1) \cdots \Gamma(1)} \prod_{k>l} \prod_l (\lambda_k - \lambda_l)^2 \cdot g(\Lambda). \end{aligned} \quad (21)$$

(In fact, the term $\prod_{k>l} \prod_l (\lambda_k - \lambda_l)^2$ can be regarded as the Jacobian in the change of variables from Ω to $\{U, \Lambda\}$.) The combination of (21) with (20) gives

$$\begin{aligned} & E e^{\text{tr} \alpha X^\dagger \Phi \Phi^\dagger X} \\ &= \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{\Gamma(M+1) \cdots \Gamma(1) (2\pi)^M} \int d\lambda_1 \cdots \int d\lambda_M \prod_{m=1}^M \\ & \cdot \left(\frac{e^{\alpha a - i\lambda_m}}{[\alpha(a - \sigma_1) - i\lambda_m] \cdots [\alpha(a - \sigma_K) - i\lambda_m] [\alpha a - i\lambda_m]^{T-K}} \right) \\ & \cdot \prod_{l < m} (\lambda_m - \lambda_l)^2. \end{aligned}$$

Note that, since Ω is Hermitian, the λ_m , $m = 1, \dots, M$ are real and so the domain of integration over the λ_m is $[-\infty, \infty]$. We now perform the change of variables $\lambda_m \rightarrow \lambda_m - i\alpha a$, and introduce the factor $-i$ in the terms $(\lambda_m - \lambda_l)$. This leads to

$$\begin{aligned} & E e^{\text{tr} \alpha X^\dagger \Phi \Phi^\dagger X} \\ &= \frac{\Gamma(T) \cdots \Gamma(T+1-M) (-1)^{M(M-1)/2}}{\Gamma(M+1) \cdots \Gamma(1) (2\pi)^M} \int d\lambda_1 \cdots \int d\lambda_M \\ & \cdot \prod_{m=1}^M \left[\frac{e^{-i\lambda_m}}{(-\alpha\sigma_1 - i\lambda_m) \cdots (-\alpha\sigma_K - i\lambda_m) (-i\lambda_m)^{T-K}} \right] \\ & \cdot \prod_{l < m} (-i\lambda_m + i\lambda_l)^2 \end{aligned}$$

where we should now make a mental note of the fact that the domain of integration of the λ_m is no longer $[-\infty, \infty]$, but rather $[-\infty + i\alpha a, \infty + i\alpha a]$.

Now using the integral formula developed in Appendix B we obtain

$$\begin{aligned} & \int \frac{d\lambda_1}{2\pi} \cdots \int \frac{d\lambda_M}{2\pi} \\ & \cdot \prod_{m=1}^M \left[\frac{e^{-i\lambda_m}}{(-\alpha\sigma_1 - i\lambda_m) \cdots (-\alpha\sigma_K - i\lambda_m) (-i\lambda_m)^{T-K}} \right] \\ & \cdot \prod_{l < m} (-i\lambda_m + i\lambda_l)^2 = M! \det F, \end{aligned}$$

where the $M \times M$ Hankel matrix F is given by

$$\begin{aligned} F &\triangleq \int_{-\infty + i\alpha a}^{\infty + i\alpha a} \frac{e^{-i\lambda} d\lambda / (2\pi)}{(-\alpha\sigma_1 - i\lambda) \cdots (-\alpha\sigma_K - i\lambda) (-i\lambda)^{T-K}} \\ & \cdot \begin{bmatrix} (-i\lambda)^0 \\ \vdots \\ (-i\lambda)^{M-1} \end{bmatrix} [(-i\lambda)^0 \quad \cdots \quad (-i\lambda)^{M-1}]. \end{aligned} \quad (22)$$

We thus have

$$\begin{aligned} & E e^{\text{tr} \alpha X^\dagger \Phi \Phi^\dagger X} \\ &= (-1)^{M(M-1)/2} \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{\Gamma(M) \cdots \Gamma(1)} \det F \end{aligned}$$

$$\begin{aligned} E_{|D} e^{\text{tr} X^\dagger \Phi (I_M + \frac{1}{\rho\beta} D^{-2})^{-1} \Phi^\dagger X} &= \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{2^M \pi^{\frac{M}{2}(M+1)}} \\ & \times \int \frac{d\Omega e^{\text{tr}[Ba - i\Omega]}}{\det[B(a - \sigma_1) - i\Omega] \cdots \det[B(a - \sigma_K) - i\Omega] \det[Ba - i\Omega]^{T-K}}. \end{aligned} \quad (18)$$

$$\begin{aligned} E_{|D} e^{\text{tr} X^\dagger \Phi (I_M + \frac{1}{\rho\beta} D^{-2})^{-1} \Phi^\dagger X} &= E e^{\text{tr} \alpha X^\dagger \Phi \Phi^\dagger X} \\ &= \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{2^M \pi^{\frac{M}{2}(M+1)}} \times \int \frac{d\Omega e^{\text{tr}(\alpha a I_M - i\Omega)}}{\det[\alpha(a - \sigma_1) - i\Omega] \cdots \det[\alpha(a - \sigma_K) - i\Omega] \det[\alpha a - i\Omega]^{T-K}}. \end{aligned} \quad (20)$$

and, therefore, using (12)

$$p(X) = \frac{(-1)^{M(M-1)/2}}{\pi^{TN}} \cdot \frac{\exp(-\text{tr } X^\dagger X)}{(1 + \rho\beta)^{MN}} \cdot \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{\Gamma(M) \cdots \Gamma(1)} \det F \quad (23)$$

which is the desired expression for $p(X)$.

B. Computing the Entries of F

All that remains to give an explicit formula for $p(X)$ is to determine the entries of the Hankel matrix F . Note from (22) that the (m, n) th entry of F is given by

$$F_{mn} = \frac{1}{2\pi} \int_{-\infty+i\alpha}^{\infty+i\alpha} d\lambda \frac{(-i\lambda)^{-T+K+m+n-2} e^{-i\lambda}}{(-\alpha\sigma_1 - i\lambda) \cdots (-\alpha\sigma_K - i\lambda)}. \quad (24)$$

This expression may be evaluated by inspection. We introduce a function of a real variable τ

$$f(\tau) = \frac{1}{2\pi} \int_{-\infty+i\alpha}^{\infty+i\alpha} d\lambda \frac{e^{-i\lambda\tau}}{(-\alpha\sigma_1 - i\lambda) \cdots (-\alpha\sigma_K - i\lambda)} \quad (25)$$

which constitutes an inverse Fourier transform. Since the region of convergence includes $\text{Im}\{\lambda\} \geq \alpha$, $f(\tau)$ is causal. We recognize that the elements of F are merely derivatives or integrals of various orders of the function $f(\tau)$, evaluated at $\tau = 1$

$$F_{mn} = f^{(-T+K+m+n-2)}(\tau)|_{\tau=1}. \quad (26)$$

The formula for $f(\tau)$ is obtained by taking $-i2\pi$ times the sum of the residues of the K distinct poles in (25), $\lambda = i\alpha\sigma_k$, $k = 1, \dots, K$; the factor -1 results from the clockwise encirclements of the poles

$$f(\tau) = \begin{cases} \sum_{k=1}^K \frac{e^{\alpha\sigma_k\tau}}{\prod_{l \neq k} (\alpha\sigma_k - \alpha\sigma_l)}, & \tau \geq 0 \\ 0, & \tau < 0. \end{cases} \quad (27)$$

The r th derivative of $f(\tau)$ at $\tau = 1$ is

$$\left. \frac{d^r f(\tau)}{d\tau^r} \right|_{\tau=1} = \sum_{k=1}^K \frac{(\alpha\sigma_k)^r e^{\alpha\sigma_k}}{\prod_{l \neq k} (\alpha\sigma_k - \alpha\sigma_l)}. \quad (28)$$

The r th integral of $f(\tau)$ at $\tau = 1$ is

$$\begin{aligned} & \left(\int_0^\tau du \right)^r f(u) \\ &= \sum_{k=1}^K \frac{1}{(\alpha\sigma_k)^r \prod_{l \neq k} (\alpha\sigma_k - \alpha\sigma_l)} \left(e^{\alpha\sigma_k} - \sum_{q=0}^{r-1} \frac{(\alpha\sigma_k)^q}{q!} \right) \\ &= \sum_{k=1}^K \frac{e^{\alpha\sigma_k}}{(\alpha\sigma_k)^r \prod_{l \neq k} (\alpha\sigma_k - \alpha\sigma_l)} \frac{\gamma(r, \alpha\sigma_k)}{\Gamma(r)} \end{aligned} \quad (29)$$

where $\gamma(r, s)$ is the incomplete Gamma function

$$\gamma(r, s) = \int_0^s du u^{r-1} e^{-u}.$$

Finally, we combine (26) with (28) and (29) to obtain the desired formula for the entries of F

$$F_{mn} = \sum_{k=1}^K \frac{e^{\alpha\sigma_k}}{(\alpha\sigma_k)^Q \prod_{l \neq k} (\alpha\sigma_k - \alpha\sigma_l)} \cdot \begin{cases} \frac{\gamma(Q, \alpha\sigma_k)}{\Gamma(Q)}, & Q \geq 1 \\ 1, & Q \leq 0 \end{cases} \quad (30)$$

where $m, n = 1, \dots, M$ and

$$Q = T - K - m - n + 2.$$

We can also give an alternative expression for $p(X)$. To this end, note that

$$\begin{aligned} \left. \frac{d^r f(\tau)}{d\tau^r} \right|_{\tau=1} &= \sum_{k=1}^K \frac{(\alpha\sigma_k)^r e^{\alpha\sigma_k}}{\prod_{l \neq k} (\alpha\sigma_k - \alpha\sigma_l)} \\ &= \sum_{k=1}^K \frac{1}{\prod_{l \neq k} (\alpha\sigma_k - \alpha\sigma_l)} \sum_{q=0}^{\infty} \frac{(\alpha\sigma_k)^{q+r}}{q!}. \end{aligned}$$

If we adopt the convention that $q! = \infty$ for all $q < 0$, this last equation can be written as

$$\left. \frac{d^r f(\tau)}{d\tau^r} \right|_{\tau=1} = \sum_{k=1}^K \frac{1}{\prod_{l \neq k} (\alpha\sigma_k - \alpha\sigma_l)} \sum_{q=0}^{\infty} \frac{(\alpha\sigma_k)^q}{(q+r)!}. \quad (31)$$

On the other hand

$$\begin{aligned} & \left(\int_0^\tau du \right)^r f(u) \\ &= \sum_{k=1}^K \frac{1}{(\alpha\sigma_k)^r \prod_{l \neq k} (\alpha\sigma_k - \alpha\sigma_l)} \left(e^{\alpha\sigma_k} - \sum_{q=0}^{r-1} \frac{(\alpha\sigma_k)^q}{q!} \right) \\ &= \sum_{k=1}^K \frac{1}{(\alpha\sigma_k)^r \prod_{l \neq k} (\alpha\sigma_k - \alpha\sigma_l)} \sum_{q=0}^{\infty} \frac{(\alpha\sigma_k)^{q+r}}{(q+r)!} \\ &= \sum_{k=1}^K \frac{1}{\prod_{l \neq k} (\alpha\sigma_k - \alpha\sigma_l)} \sum_{q=0}^{\infty} \frac{(\alpha\sigma_k)^q}{(q+r)!}. \end{aligned} \quad (32)$$

Combining (31) and (32) yields

$$F_{mn} = \sum_{k=1}^K \frac{1}{\prod_{l \neq k} (\alpha\sigma_k - \alpha\sigma_l)} \sum_{q=0}^{\infty} \frac{(\alpha\sigma_k)^q}{(q+T-(m+n+K)+2)!}, \quad m, n = 1, \dots, M \quad (33)$$

where we have adopted the convention $q! = \infty$, for all $q < 0$.

C. An Alternative Expression for $p(X)$

The expressions (23), (30), and (33) for $p(X)$ can, in principle, be used to compute the mutual information $E \log \frac{p(X|S)}{p(X)}$ via Monte Carlo integration. However, practice has shown that Monte Carlo evaluation of the mutual information using these expressions leads to numerical problems at high SNR and at large numbers of receive and transmit antennas. The main reason for this is that the entries of the Hankel matrix F (in (30) and (33)) contain exponential terms with positive exponent (the $e^{\alpha\sigma_k}$ terms) which can lead to numerical instability, overflow, etc. Therefore, it would be useful to avoid such expressions.

This, indeed, can be done: one can derive an equivalent formulation that involves only negative exponents. To this end, let Φ_\perp denote the $T \times (T - M)$ orthogonal complement of Φ , i.e.,

$$I_T = \Phi\Phi^\dagger + \Phi_\perp\Phi_\perp^\dagger \quad \text{and} \quad \Phi_\perp^\dagger\Phi_\perp = I_{T-M}.$$

Then one can write

$$\begin{aligned} p(X) &= \frac{e^{-\text{tr} X^\dagger X}}{\pi^{TN}(1 + \rho\beta)^{MN}} E e^{\text{tr} \alpha X X^\dagger \Phi \Phi^\dagger} \\ &= \frac{e^{-\text{tr}(1-\alpha)X^\dagger X}}{\pi^{TN}(1 + \rho\beta)^{MN}} E e^{-\text{tr} \alpha X X^\dagger \Phi_\perp \Phi_\perp^\dagger}. \end{aligned}$$

Note that in the above expression Φ_\perp is also isotropically random. We may, therefore, repeat the above steps to compute the expectation on the right-hand side. We need only note two differences: 1) α has changed to $-\alpha$ in the exponent of the isotropic term, and 2) the isotropically random (i.r.) unitary matrix has $T - M$, rather than M , columns, which means that we only need to replace M by $T - M$ in all our expressions. This leads to

$$p(X) = \frac{(-1)^{(T-M)(T-M-1)/2}}{\pi^{TN}} \cdot \frac{\exp(-\text{tr}(1-\alpha)X^\dagger X)}{(1 + \rho\beta)^{MN}} \cdot \frac{\Gamma(T) \cdots \Gamma(M+1)}{\Gamma(T-M) \cdots \Gamma(1)} \det G \quad (34)$$

where G is a $(T - M) \times (T - M)$ Hankel matrix whose entries are given by

$$G_{mn} = \sum_{k=1}^K \frac{e^{-\alpha\sigma_k}}{(-\alpha\sigma_k)^Q \prod_{l \neq k} (-\alpha\sigma_k + \alpha\sigma_l)} \cdot \begin{cases} \frac{\gamma(Q, -\alpha\sigma_k)}{\Gamma(Q)}, & Q \geq 1 \\ 1, & Q \leq 0 \end{cases} \quad (35)$$

where we define the incomplete Gamma function for a negative-valued argument as

$$\frac{\gamma(Q, -\alpha\sigma_k)}{\Gamma(Q)} = 1 - e^{\alpha\sigma_k} \sum_{q=0}^{Q-1} \frac{(-\alpha\sigma_k)^q}{q!}$$

or equivalently

$$G_{mn} = \sum_{k=1}^K \frac{1}{\prod_{l \neq k} (-\alpha\sigma_k + \alpha\sigma_l)} \cdot \sum_{q=0}^{\infty} \frac{(-\alpha\sigma_k)^q}{(q + T - (m + n + K) + 2)!} \quad (36)$$

where $m, n = 1, \dots, T - M$.

IV. THE MUTUAL INFORMATION FOR ISOTROPICALLY DISTRIBUTED USTM

Having obtained various formulas for $p(X)$, we can now readily write the mutual information $I(X; \Phi)$, using (10) with $D = I_M$, and (23) as

$$\begin{aligned} I(X; \Phi) &= -TN \log(\pi e) - MN \log(1 + \rho\beta) \\ &\quad - \left[-TN \log \pi - MN \log(1 + \rho\beta) \right. \\ &\quad \left. + \log \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{\Gamma(M) \cdots \Gamma(1)} \right. \\ &\quad \left. - E \{ \text{tr} X^\dagger X \} \log e \right. \\ &\quad \left. + E \left\{ \log \left[(-1)^{M(M-1)/2} \det F \right] \right\} \right] \\ &= MN \rho\beta \log e - \log \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{\Gamma(M) \cdots \Gamma(1)} \\ &\quad - E \left\{ \log \left[(-1)^{M(M-1)/2} \det F \right] \right\} \quad (37) \end{aligned}$$

where the expectations are taken with respect to X , conditioned on Φ_0 being transmitted. A similar expression follows from (34).

Theorem 1 (Mutual Information for Isotropically Distributed Unitary Input): Consider the channel model

$$X = \sqrt{\frac{\rho}{M}} SH + W$$

where the matrix $X \in \mathcal{C}^{T \times N}$ is the received signal, $S \in \mathcal{C}^{T \times M}$ is the transmitted signal, and $W \in \mathcal{C}^{T \times N}$ is the additive noise. The entries of H and W are assumed to be independent $\text{CN}(0, 1)$ distributed. Then if $S = \sqrt{T}\Phi$, where Φ is an isotropically distributed unitary matrix, we have

$$\begin{aligned} I(X; \Phi) &= \rho TN \log e - \log \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{\Gamma(M) \cdots \Gamma(1)} \\ &\quad - E \left\{ \log \left((-1)^{M(M-1)/2} \det F \right) \right\} \quad (38) \end{aligned}$$

or equivalently

$$\begin{aligned} I(X; \Phi) &= \frac{1 - \beta}{1 + \rho\beta} \rho TN \log e - \log \frac{\Gamma(T) \cdots \Gamma(M+1)}{\Gamma(T-M) \cdots \Gamma(1)} \\ &\quad - E \left\{ \log \left((-1)^{(T-M)(T-M-1)/2} \det G \right) \right\} \quad (39) \end{aligned}$$

where F and G are $M \times M$ and $(T - M) \times (T - M)$ Hankel matrices whose entries are given by (30) and (35), respectively.

V. NUMERICAL RESULTS

The formulas (38) and (39) are expressions for mutual information for an isotropically random unitary space-time input. At present, however, we are unable to evaluate analytically the expectation with respect to X . This may be done by Monte Carlo integration. We generate L independent realizations $\{X_1, \dots, X_L\}$, where the X_l are distributed according to (11), with $D = I_M$ and $S = \sqrt{T}\Phi_0$, $\Phi_0^\dagger = [I_M \ 0]$. As mentioned earlier, because the entries of G have exponential terms with negative exponent, whereas those of F have exponential terms

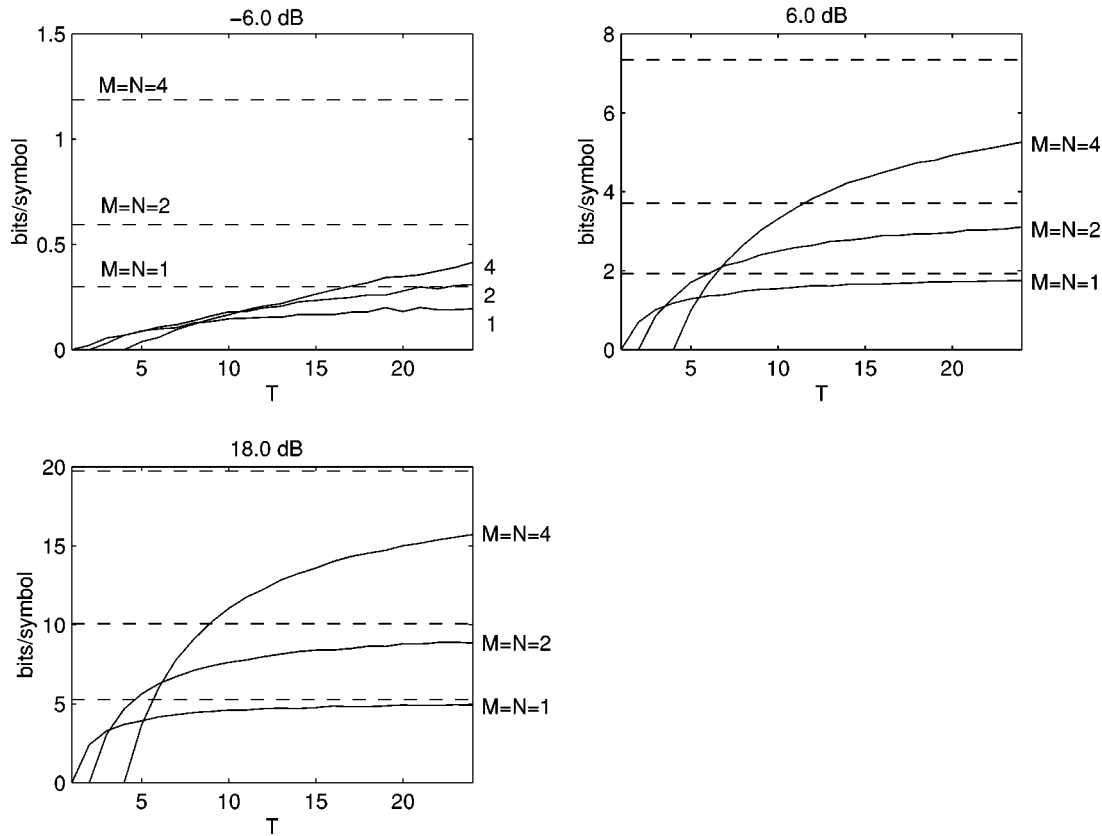


Fig. 1. Mutual information per symbol versus duration of coherence interval T , for $M = N = 1, 2, 4$ transmit and receive antennas, and for $\rho = -6.0, 6.0, 18.0$ dB.

with positive exponent, Monte Carlo simulation of (39) is much more numerically stable. The Monte Carlo estimate of mutual information is

$$\begin{aligned} \hat{I}(X; S) &= \frac{1}{L} \sum_{l=1}^L \log \frac{p(X_l | \Phi_0)}{p(X_l)} \\ &= \frac{1}{L} \sum_{l=1}^L \log \left(\frac{\Gamma(T-M) \cdots \Gamma(1)}{\Gamma(T) \cdots \Gamma(M+1)} \right. \\ &\quad \left. \cdot \frac{\exp(\alpha \text{tr } X_l X_l^\dagger (-I_T + \Phi_0 \Phi_0^\dagger))}{(-1)^{(T-M)(T-M-1)/2} \det G_l} \right) \end{aligned} \quad (40)$$

where we have used (11) for $p(X_l | \Phi_0)$, and (34) for $p(X_l)$. We avoid the form (39), where the expectation of $\log p(X | \Phi_0)$ is taken analytically, and instead, for greater stability, we apply Monte Carlo integration to a log-likelihood ratio.

A. Mutual Information as a Function of T

Fig. 1 shows the mutual information per symbol (i.e., (40) is normalized through a division by T) as a function of $T = 1, \dots, 25$, for $M = N = 1, 2, 4$, and $\rho = -6.0, 6.0, 18.0$ dB, with $L = 4 \times 10^4$ Monte Carlo trials. The $M = N = 1$ curves match similar curves obtained in [6], where the expectation with respect to X was performed by numerical quadrature. The dashed lines represent the perfect knowledge capacity. The figure shows the rate of convergence of the mutual information

to the perfect knowledge capacity as $T \rightarrow \infty$. As can be seen, convergence is faster at higher SNRs.

B. Mutual Information as a Function of M and N

Fig. 2 shows the mutual information per symbol as a function of $M = 1, \dots, 8$, and $N = 1, \dots, 8$, for $T = 8$, and $\rho = -6.0, 6.0, 18.0$ dB, and $L = 2500$. As can be seen, at very low SNR (here -6.0 dB) USTM is unable to utilize multiple transmit antennas efficiently, and it is best to reserve all of the power for a single antenna. This contrasts sharply with the high SNR regime (here 18.0 dB), where the mutual information peaks at about $M = \min(T/2, N)$ transmit antennas, thereby confirming the high SNR results of [8]. Intermediate SNR values (here 6.0 dB) present an intermediate stage. We conclude that, as the SNR increases, more and more transmit antennas should be powered successively.

VI. DISCUSSION AND CONCLUSION

We have obtained a closed-form expression in the form of a determinant, for the probability density of the received signal, when the input is isotropically random unitary. This expression, when combined with a simple Monte Carlo integration enables us to compute mutual information for all cases of interest. At sufficiently low SNR, we have found that mutual information is maximized by allocating all power to a single transmit antenna. At sufficiently high SNR, we have confirmed the result of [8] that $M = \min(T/2, N)$ transmit antennas maximizes mutual information.

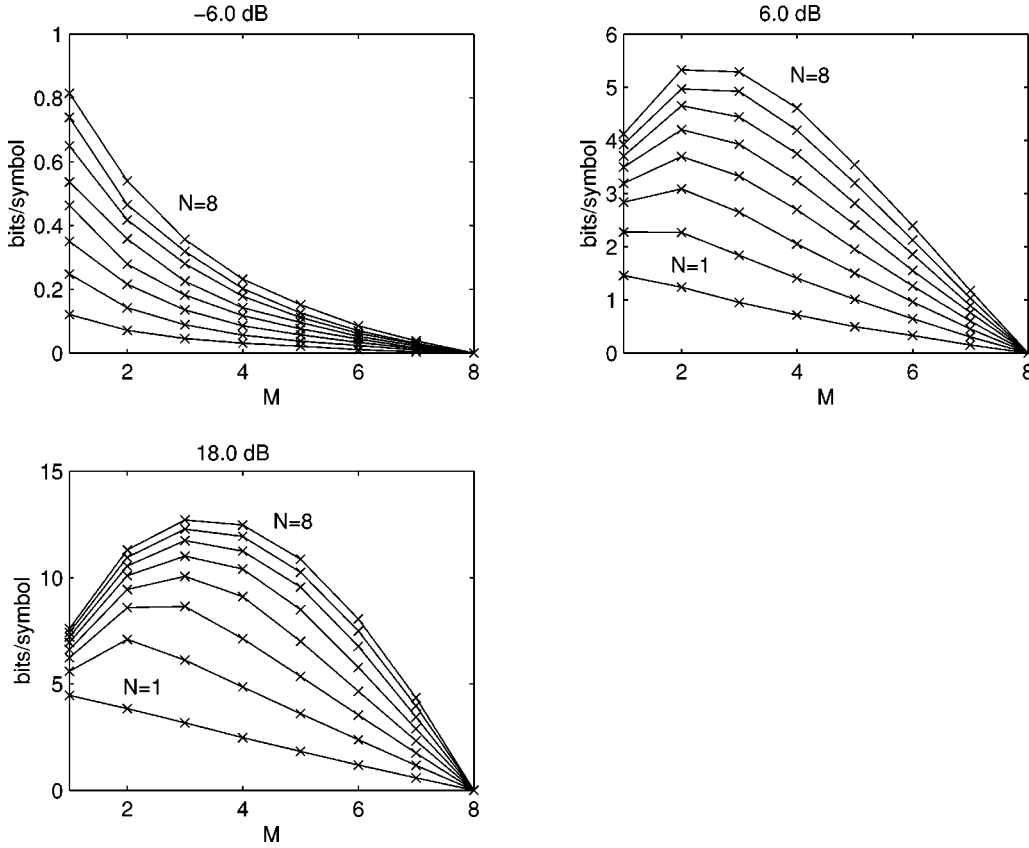


Fig. 2. Mutual information per symbol versus number of transmit antennas M , for $N = 1, \dots, 8$ receive antennas, for $T = 8$, and $\rho = -6.0, 6.0, 18.0$ dB. At this low SNR it is best to feed all power to a single antenna. At high SNR it is best to use $M = \min(T/2, N)$ transmit antennas.

The mutual information we have computed yields a lower bound on the actual channel capacity. In future work we will attempt to compute the mutual information for arbitrary D , rather than $D = I_M$. This should lead to further insight into the problem at moderate values of SNR and to tighter lower bounds on capacity. The fact that we have been able to find the output density in closed form should also be useful for obtaining upper bounds on capacity via divergence-based techniques [9].

Our method of taking expectations with respect to the isotropically random unitary density has other interesting ramifications. In particular, it enables us to compute any desired moments for an isotropically random unitary matrix. Also, we can obtain a variety of new bounds on the random coding exponent for space-time autocoding,² which should provide improved estimates, compared with the simple union bound, of the autocoding performance of codebooks of isotropically random unitary signals [14], [15].

APPENDIX A INTEGRAL FORMULA FOR $p(\Phi)$

Let us begin by obtaining the formula for the density of a $T \times M$ identically distributed (i.d.) unitary matrix $\Phi = [\phi_1 \dots \phi_M]$, where the ϕ_m , $m = 1, \dots, M$ denote the $T \times 1$ column vectors of Φ . This density was derived in [6], but we include it here to make our treatment self-contained and

because our derivation will be useful for obtaining the desired Fourier integral expression.

We first note that, due to the unitary constraint $\Phi^\dagger \Phi = I_M$, the density must have the form $p(\Phi) = f(\Phi) \delta(\Phi^\dagger \Phi - I_M)$, where $f(\Phi)$ is some smooth function and where

$$\delta(\Phi^\dagger \Phi - I_M) = \prod_{m=1}^M \delta(\phi_m^\dagger \phi_m - 1) \prod_{m>n} \delta(\text{Re}(\phi_m^\dagger \phi_n)) \cdot \delta(\text{Im}(\phi_m^\dagger \phi_n)). \quad (\text{A1})$$

Now, for Φ to be isotropically distributed, the density $p(\Phi)$ must be invariant to pre- and postmultiplication by any $T \times T$ and $M \times M$ unitary matrices, say U and V . Clearly, $\delta(\Phi^\dagger \Phi - I_M)$ is invariant under such a transformation. For $f(\cdot)$ to be invariant, we require that $f(U\Phi V) = f(\Phi)$. But since U and V are arbitrary this implies that $f(\Phi)$ cannot be a function of the singular vectors of Φ and must, therefore, be only a function of the singular values. But since Φ is unitary, its singular values are all unity, which implies that $f(\Phi)$ must be a constant. In other words, we have $p(\Phi) = c \delta(\Phi^\dagger \Phi - I_M)$, for some normalizing constant c .

To compute the normalizing constant c , we can rewrite the density as

$$\begin{aligned} p(\Phi) &= p(\phi_1, \dots, \phi_M) \\ &= p(\phi_1) p(\phi_2, \dots, \phi_M | \phi_1) \\ &= p(\phi_1) \prod_{m=2}^M p(\phi_m | \phi_1, \dots, \phi_{m-1}). \end{aligned} \quad (\text{A2})$$

²This was pointed out independently to us by Prof. S. Shamai.

Now since ϕ_1 is $T \times 1$ isotropically-distributed unitary, we have and

$$p(\phi_1) = c\delta(\phi_1^\dagger \phi_1 - 1)$$

for some constant c . To determine c it is useful to represent the delta function via its Fourier transform $\delta(x) = \frac{1}{2\pi} \int d\omega e^{i\omega x}$, and to proceed as follows:

$$\begin{aligned} 1 &= \int d\phi_1 p(\phi_1) \\ &= \frac{c}{2\pi} \int d\phi_1 \int d\omega e^{i\omega(\phi_1^\dagger \phi_1 - 1)} \\ &= \frac{c}{2\pi} \int d\phi_1 e^{\alpha\phi_1^\dagger \phi_1} e^{-\alpha\phi_1^\dagger \phi_1} \int d\omega e^{i\omega(\phi_1^\dagger \phi_1 - 1)}, \quad \alpha > 0 \\ &= \frac{c e^\alpha}{2\pi} \int d\omega e^{-i\omega} \int d\phi_1 e^{-(\alpha - i\omega)\phi_1^\dagger \phi_1} \\ &= \frac{c e^\alpha}{2\pi} \int d\omega e^{-i\omega} \cdot \frac{\pi^T}{(\alpha - i\omega)^T} = \frac{c\pi^T}{\Gamma(T)} \end{aligned}$$

which implies that

$$p(\phi_1) = \frac{\Gamma(T)}{\pi^T} \delta(\phi_1^\dagger \phi_1 - 1).$$

Now we also have

$$p(\phi_m|\phi_1, \dots, \phi_{m-1}) = c\delta(\phi_m^\dagger \phi_m - 1) \prod_{n=1}^{m-1} \delta(\phi_m^\dagger \phi_n)$$

for some other normalizing constant c , and where

$$\delta(\phi_m^\dagger \phi_n) = \delta(\text{Re}(\phi_m^\dagger \phi_n))\delta(\text{Im}(\phi_m^\dagger \phi_n)).$$

Proceeding with a similar argument to the one presented above to determine c , or, alternatively, by simply noting that, in the subspace that is orthogonal to $\{\phi_1, \dots, \phi_{M-1}\}$, $\phi_m|\phi_1, \dots, \phi_{m-1}$ has the same distribution as a $T+1-M$ -dimensional isotropically distributed unitary vector, we may write

$$\begin{aligned} p(\phi_m|\phi_1, \dots, \phi_{m-1}) &= \frac{\Gamma(T+1-m)}{\pi^{T+1-m}} \delta(\phi_m^\dagger \phi_m - 1) \prod_{n=1}^{m-1} \delta(\phi_m^\dagger \phi_n). \end{aligned}$$

Using the expressions for $p(\phi_1)$ and $p(\phi_m|\phi_1, \dots, \phi_{m-1})$, and using (A2), we may write

$$\begin{aligned} p(\Phi) &= \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{\pi^{\frac{M(2T-M+1)}{2}}} \prod_{m=1}^M \delta(\phi_m^\dagger \phi_m - 1) \\ &\quad \cdot \prod_{m>n} \delta(\phi_m^\dagger \phi_n) \end{aligned} \quad (\text{A3})$$

$$= \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{\pi^{\frac{M(2T-M+1)}{2}}} \delta(\Phi^\dagger \Phi - I_M) \quad (\text{A4})$$

which is the desired density function for $p(\Phi)$.

A. The Integral Formula

To obtain an integral formula for $p(\Phi)$, we shall replace the delta functions in (A3) by their Fourier integrals. Thus,

$$\delta(\phi_m^\dagger \phi_m - 1) = \frac{1}{2\pi} \int d\omega_{mm} e^{i\omega_{mm}(\phi_m^\dagger \phi_m - 1)} \quad (\text{A5})$$

$$\begin{aligned} \delta(\phi_m^\dagger \phi_n) &= \delta(\text{Re}(\phi_m^\dagger \phi_n)) \cdot \delta(\text{Im}(\phi_m^\dagger \phi_n)) \\ &= \frac{1}{(2\pi)^2} \int d\alpha_{mn} d\beta_{mn} \\ &\quad \cdot e^{i(\alpha_{mn} \text{Re}(\phi_m^\dagger \phi_n) + \beta_{mn} \text{Im}(\phi_m^\dagger \phi_n))} \\ &= \frac{1}{(2\pi)^2} \int d\alpha_{mn} d\beta_{mn} \\ &\quad \cdot e^{i(\frac{\alpha_{mn} - i\beta_{mn}}{2} \phi_m^\dagger \phi_n + \frac{\alpha_{mn} + i\beta_{mn}}{2} \phi_n^\dagger \phi_m)}. \end{aligned} \quad (\text{A6})$$

Now, if we define the off-diagonal parts of the Hermitian matrix Ω via

$$\omega_{mn} = \frac{\alpha_{mn} - i\beta_{mn}}{2} \quad \text{and} \quad \omega_{nm} = \omega_{mn}^* = \frac{\alpha_{mn} + i\beta_{mn}}{2}$$

we clearly have

$$d\omega_{mn} = \frac{d\alpha_{mn} d\beta_{mn}}{4}$$

so that

$$\delta(\phi_m^\dagger \phi_n) = \frac{1}{\pi^2} \int d\omega_{mn} e^{i(\omega_{mn} \phi_m^\dagger \phi_n + \omega_{nm} \phi_n^\dagger \phi_m)}. \quad (\text{A7})$$

Combining (A5) and (A7) with (A3), we arrive at

$$\begin{aligned} p(\Phi) &= \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{\pi^{\frac{M(2T-M+1)}{2}}} \\ &\quad \cdot \frac{1}{(2\pi)^M} \int \prod_{m=1}^M d\omega_{mm} e^{i\omega_{mm}[(\Phi^\dagger \Phi)_{mm} - \delta_{mm}]} \\ &\quad \cdot \frac{1}{\pi^{M(M-1)}} \int \prod_{m>n} d\omega_{mn} \\ &\quad \cdot e^{i[\omega_{mn}((\Phi^\dagger \Phi)_{mn} - \delta_{nm}) + \omega_{nm}((\Phi^\dagger \Phi)_{nm} - \delta_{mn})]} \\ &= \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{2^M \pi^{\frac{M}{2}(2T+M+1)}} \int d\Omega e^{i \text{tr} \Omega^T (\Phi^\dagger \Phi - I_M)} \\ &= \frac{\Gamma(T) \cdots \Gamma(T+1-M)}{2^M \pi^{\frac{M}{2}(2T+M+1)}} \int d\Omega e^{i \text{tr} \Omega (\Phi^\dagger \Phi - I_M)} \end{aligned} \quad (\text{A8})$$

where the integration is over the Hermitian matrix Ω . But this is simply the desired result (13).

APPENDIX B

A USEFUL INTEGRAL FORMULA

In the next result we generalize a formula of Wigner [16] for positive functions $f(\cdot)$ to arbitrary ones.

Lemma 1 (An Integral Formula): Consider the single-variable function $f(\cdot)$. Then we have the following identity for the M -fold integral:

$$\int \left(\prod_{m=1}^M d\lambda_m f(\lambda_m) \right) \prod_{l<m} (-i\lambda_m + i\lambda_l)^2 = M! \det F \quad (\text{B1})$$

where F is an $M \times M$ Hankel matrix given by

$$F \triangleq \int d\lambda f(\lambda) \begin{bmatrix} (-i\lambda)^0 \\ \vdots \\ (-i\lambda)^{M-1} \end{bmatrix} [(-i\lambda)^0 \quad \dots \quad (-i\lambda)^{M-1}]. \quad (\text{B2})$$

Proof: Note that the product $\prod_{l < m} (-i\lambda_m + i\lambda_l)$ can be written as the determinant of a Vandermonde matrix

$$\prod_{l < m} (-i\lambda_m + i\lambda_l) = \det \begin{bmatrix} (-i\lambda_1)^0 & \cdots & (-i\lambda_M)^0 \\ \vdots & \ddots & \vdots \\ (-i\lambda_1)^{M-1} & \cdots & (-i\lambda_M)^{M-1} \end{bmatrix}.$$

Assume now that F is nonsingular,³ and define the polynomials $g_1(\cdot), \dots, g_M(\cdot)$ via

$$\begin{bmatrix} g_1(\lambda) \\ \vdots \\ g_M(\lambda) \end{bmatrix} \triangleq F^{-1} \begin{bmatrix} (-i\lambda)^0 \\ \vdots \\ (-i\lambda)^{M-1} \end{bmatrix}. \quad (\text{B3})$$

With this definition of the $g_m(\cdot)$, it is important to note that

$$\begin{aligned} I_M &= F^{-1} F \\ &= \int d\lambda f(\lambda) F^{-1} \begin{bmatrix} (-i\lambda)^0 \\ \vdots \\ (-i\lambda)^{M-1} \end{bmatrix} [(-i\lambda)^0 \cdots (-i\lambda)^{M-1}] \\ &= \int d\lambda f(\lambda) \begin{bmatrix} g_1(\lambda) \\ \vdots \\ g_M(\lambda) \end{bmatrix} [(-i\lambda)^0 \cdots (-i\lambda)^{M-1}] \end{aligned}$$

which implies that the $g_m(\cdot)$ and λ^{n-1} satisfy the following bi-orthogonality condition

$$\int d\lambda f(\lambda) g_m(\lambda) (-i\lambda)^{n-1} = \delta_{mn}, \quad m, n = 1, \dots, M. \quad (\text{B4})$$

Denoting the integral in the statement of the lemma by J , we can now write

$$\begin{aligned} J &= \int \left(\prod_{m=1}^M d\lambda_m f(\lambda_m) \right) \det F \det F^{-1} \\ &\quad \cdot \det \begin{bmatrix} (-i\lambda_1)^0 & \cdots & (-i\lambda_M)^0 \\ \vdots & \ddots & \vdots \\ (-i\lambda_1)^{M-1} & \cdots & (-i\lambda_M)^{M-1} \end{bmatrix}^2 \\ &= \det F \int \left(\prod_{m=1}^M d\lambda_m f(\lambda_m) \right) \\ &\quad \cdot \det \begin{bmatrix} g_1(\lambda_1) & \cdots & g_1(\lambda_M) \\ \vdots & \ddots & \vdots \\ g_M(\lambda_1) & \cdots & g_M(\lambda_M) \end{bmatrix} \\ &\quad \cdot \begin{bmatrix} (-i\lambda_1)^0 & \cdots & (-i\lambda_M)^0 \\ \vdots & \ddots & \vdots \\ (-i\lambda_1)^{M-1} & \cdots & (-i\lambda_M)^{M-1} \end{bmatrix}. \end{aligned}$$

If we now expand the above two determinants as sums of the products

$$g_{m_1}(\lambda_1) g_{m_2}(\lambda_2) \cdots g_{m_M}(\lambda_M)$$

³We shall momentarily show that if F is singular then the integral is zero.

$$(-i\lambda_1)^{n_1-1} (-i\lambda_2)^{n_2-1} \cdots (-i\lambda_M)^{n_M-1}$$

where the $\{m_i\}$ and $\{n_i\}$ denote some permutation of the integers $\{1, \dots, M\}$, we note that, due to the bi-orthogonality condition, whenever the $\{m_i\}$ and $\{n_i\}$ are different permutations their product integrates to zero. When they are the same permutation they integrate to unity. Since there are $M!$ such permutations, we conclude that

$$J = M! \det F$$

which is the desired result.

All that remains to be shown is that when F is singular the integral is zero. Suppose that F has $R > 0$ zero eigenvalues, so that its singular value decomposition takes on the form

$$F = U \begin{bmatrix} \Sigma & \\ & 0_{R \times R} \end{bmatrix} V^\dagger.$$

Now defining the polynomials $q_1(\cdot), \dots, q_M(\cdot)$ and $r_1(\cdot), \dots, r_M(\cdot)$

$$\begin{bmatrix} q_1(\lambda) \\ \vdots \\ q_M(\lambda) \end{bmatrix} \triangleq U^\dagger \begin{bmatrix} (-i\lambda)^0 \\ \vdots \\ (-i\lambda)^{M-1} \end{bmatrix}$$

and

$$\begin{bmatrix} r_1(\lambda) \\ \vdots \\ r_M(\lambda) \end{bmatrix} \triangleq V^T \begin{bmatrix} (-i\lambda)^0 \\ \vdots \\ (-i\lambda)^{M-1} \end{bmatrix}$$

it is straightforward to see that

$$\begin{aligned} &\int d\lambda f(\lambda) q_m(\lambda) r_n(\lambda) \\ &= \begin{cases} \sigma_n \delta_{mn}, & \text{if } m, n \leq M - R \\ 0, & \text{if } m > M - R \text{ or } n > M - R. \end{cases} \quad (\text{B5}) \end{aligned}$$

Using an argument similar to the one presented above, one can write

$$\begin{aligned} J &= \int \left(\prod_{m=1}^M d\lambda_m f(\lambda_m) \right) \det \begin{bmatrix} q_1(\lambda_1) & \cdots & q_1(\lambda_M) \\ \vdots & \ddots & \vdots \\ q_M(\lambda_1) & \cdots & q_M(\lambda_M) \end{bmatrix} \\ &\quad \cdot \begin{bmatrix} r_1(\lambda_1) & \cdots & r_1(\lambda_M) \\ \vdots & \ddots & \vdots \\ r_M(\lambda_1) & \cdots & r_M(\lambda_M) \end{bmatrix}. \end{aligned}$$

Now expanding the above two determinants and using (B5) it follows that all terms integrate to zero, so that $J = 0$. \square

APPENDIX C

AN EXPONENTIAL QUADRATIC INTEGRAL

Let P and Q be $M \times M$ Hermitian matrices such that $P > 0$. In this appendix, we verify the integral formula

$$\int dy e^{-y^*(P+iQ)y} = \frac{\pi^M}{\det(P+iQ)}. \quad (\text{C1})$$

Note that this result only requires a positivity assumption on P (and not on Q).

Since $P > 0$, for any square-root factorization $P = P^{1/2} P^{*/2}$, the matrix $P^{1/2}$ will be nonsingular. Consider the change of variables $z = P^{*/2} y$. Then clearly

$$dz = dy \left| \det P^{*/2} \right|^2 = dy \det P$$

and the integral becomes

$$\int dy e^{-y^*(P+iQ)y} = \frac{1}{\det P} \int dz e^{-z^*(I_M + iP^{-1/2}QP^{-*/2})z}.$$

Consider now the eigenvalue decomposition of the Hermitian matrix $P^{-1/2}QP^{-*/2} = U\Lambda U^*$, where U is unitary and Λ is diagonal with real diagonal entries $\lambda_1, \dots, \lambda_M$. The change of variable $w = U^*z$ yields $dw = dz$ (since the transformation matrix is unitary) and so

$$\begin{aligned} \int dy e^{-y^*(P+iQ)y} &= \frac{1}{\det P} \int dw e^{-w^*(I_M + i\Lambda)w} \\ &= \frac{1}{\det P} \prod_{m=1}^M \int dw_m e^{-(1+i\lambda_m)|w_m|^2}. \end{aligned}$$

It is straightforward to show that

$$\int dw_m e^{-(1+i\lambda_m)|w_m|^2} = \frac{\pi}{1+i\lambda_m}$$

so that

$$\begin{aligned} \int dy e^{-y^*(P+iQ)y} &= \frac{\pi^M}{\det P} \prod_{m=1}^M \frac{1}{1+i\lambda_m} \\ &= \frac{\pi^M}{\det P} \cdot \frac{1}{\det(I_M + i\Lambda)} \\ &= \frac{\pi^M}{\det P} \cdot \frac{1}{\det(I_M + iU\Lambda U^*)} \\ &= \frac{\pi^M}{\det(P + iQ)} \end{aligned}$$

which is the desired result.

ACKNOWLEDGMENT

The authors would like to thank the editor and reviewers for useful comments that led to the improvement of this paper.

REFERENCES

- [1] J. Salz and A. Wyner, "On data transmission over cross coupled linear channels," AT&T Bell Labs Tech. Memo., 1985.
- [2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.
- [3] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
- [4] G. Taricco and M. Elia, "Capacity of fading channels with no side information," *Electron. Lett.*, vol. 33, pp. 1368–1370, July 1997.
- [5] I. C. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz), "The capacity of discrete-time memoryless Rayleigh fading channels," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1290–1301, May 2001.
- [6] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 138–157, Jan. 1999.
- [7] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communication in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, Mar. 2000.
- [8] L. Zheng and D. N. C. Tse, "Communication on the Grassman manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359–384, Feb. 2002.
- [9] E. Csizsár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest, Hungary: Akademiai Kiado, 1997.
- [10] A. Steiner, M. Peleg, and S. Shamai (Shitz), "SVD iterative detection of turbo coded multi element unitary matrix differential modulation," *IEEE Trans. Commun.*, submitted for publication.
- [11] B. Hochwald, T. Marzetta, T. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1962–1994, Sept. 2000.
- [12] M. L. Mehta, *Random Matrices and Statistical Theory*. New York: Academic, 1967.
- [13] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.
- [14] B. Hochwald, T. Marzetta, and B. Hassibi, "Space-time autocoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2761–2681, Nov. 2001.
- [15] T. L. Marzetta, B. Hassibi, and B. M. Hochwald, "Structured unitary space-time autocoding constellations," *IEEE Trans. Inform. Theory*, vol. 48, pp. 942–950, Apr. 2002.
- [16] E. P. Wigner, "Random matrices in physics," *SIAM Rev.*, vol. 9, pp. 1–23, 1967.