MIMO Capacity Through Correlated Channels in the Presence of Correlated Interferers and Noise: A (Not So) Large N Analysis

Aris L. Moustakas, *Member, IEEE*, Steven H. Simon, and Anirvan Mayukh Sengupta

***Abstract*—The use of multiple-antenna arrays in both transmission and reception promises huge increases in the throughput of wireless communication systems. It is therefore important to analyze the capacities of such systems in realistic situations, which may include spatially correlated channels and correlated noise, as well as correlated interferers with known channel at the receiver. Here, we present an approach that provides analytic expressions for the statistics, i.e., the moments of the distribution, of the mutual information of multiple-antenna systems with arbitrary correlations, interferers, and noise. We assume that the channels of the signal and the interference are Gaussian with arbitrary covariance. Although this method is valid formally for large antenna numbers, it produces extremely accurate results even for arrays with as few as two or three antennas. We also develop a method to analytically optimize over the input signal covariance, which enables us to calculate analytic capacities when the transmitter has knowledge of the statistics of the channel (i.e., the channel covariance). In many cases of interest, this capacity is very close to the full closed-loop capacity, in which the transmitter has instantaneous channel knowledge. We apply this analytic approach to a number of examples and we compare our results with simulations to establish the validity of this approach. This method provides a simple tool to analyze the statistics of throughput for arrays of any size. The emphasis of this paper is on elucidating the novel mathematical methods used.**

***Index Terms*—Beamforming, capacity, multiple antennas, random matrix theory (RMT), replicas, side information.**

I. Introduction

OVER the past few years there has been increased interest in using multiple-antenna arrays simultaneously in transmission and reception. Recent theoretical work in [1] and [2] has shown that for sufficiently rich scattering environments, the Shannon capacity of an n_t-element transmitting and an n_r-element receiving array is roughly proportional to $\min(n_r, n_t)$ for large numbers of antennas. In addition, both indoors [3] and outdoors [4] measurements have shown promising throughput gains for such multiple-input–multiple-output (MIMO) technologies.

Manuscript received October 30, 2002; revised June 25, 2003. The material in this paper was presented in part at the DIMACS Workshop on Signal Processing for Wireless Communications, Rutgers University, Piscataway, NJ, October 2002.

A. L. Moustakas is with AT&T Laboratories, Lucent Technologies, Whippany, NJ 07981 USA (e-mail: arislm@lucent.com).

S. H. Simon is with AT&T Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA (e-mail: shsimon@lucent.com).

A. M. Sengupta is with the Department of Physics, Rutgers–The State University of New Jersey, Piscataway, NJ 08854 USA (e-mail: anirvan@physics.rutgers.edu).

Communicated by G. Caire, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2003.817427

Of course, the use of multiple-antenna arrays comes with higher complexity requirements in transmission and reception. Therefore, it is important to be able to assess the capacity gains of MIMO technologies in realistic situations. Several factors have to be taken into account. First, the degree of spatial correlations between antennas at both transmitter and receiver determines the effective number of independent channels. This is especially important in the case of outdoor propagation where the base-station antennas are typically placed on high towers with a few nearby scatterers. Second, the spatial structure of interference may be significant. For example, in downlink transmission, the interference from other base stations may be coming from very different directions than the desired signal.

For large antenna numbers and Gaussian channel distributions, the analysis of MIMO ergodic capacities (expectation value of mutual information over channel realizations) is greatly facilitated by the use of asymptotic techniques of random matrix theory (RMT). These methods were introduced in this context by various authors, starting with Foschini [1] and Telatar [2]. Verdú and Shamai [5] derived the expression of the capacity for infinite antennas with uncorrelated channels in the context of code-division multiple-access (CDMA) codes, and more recently Rapajic and Popescu [6] derived it in the context of multiple-antenna systems. Lozano and Tulino [7] using methods developed by Tse and Hanley [8] calculated the infinite antenna number capacity with spatially uncorrelated channels and uncorrelated interferers. Very recently, Chuah *et al.* [9] extended the results of [5] to calculate the mutual information of spatially correlated channels in the infinite antenna limit. In all previous studies, the capacity has been assumed to be asymptotically proportional to $\min(n_t, n_r)$ and only the corresponding proportionality factor was studied as the number of antennas grew indefinitely. However, for finite antenna numbers or for slowly decaying spatial correlations, the capacity of arrays with finite antenna numbers cannot be simply described as linear in the antenna number. For example, as shown in [10, Fig. 2], for square arrays with $\lambda/2$ spacing the capacity per antenna does not reach the limiting value even at 10 000 antennas per array! In [10], using techniques developed by Sengupta and Mitra [11] in a different context, a method was developed to analytically calculate the capacity of spatially correlated channels for large but finite antenna numbers.

In this paper, we extend work done in [10], [12] to provide analytic expressions for the statistics of the mutual information in the presence of spatially correlated channels, interference, and noise. We apply a method from physics, known as the replica ap-

IEEE Proof

0018-9448/03$17.00 © 2003 IEEE

proach, to analyze the resulting random matrix problems. The replica approach, first introduced in [13], has been used heavily in physics for understanding random systems [14], [15]. One of the first applications of this method in communication theory was by Sourlas [16] in the field of error-correcting codes. More recently, this method has been seeing increased application in the field of information theory [17]–[19]. One of the main purposes of this paper is to elucidate this technique in a way that will allow it to become a common tool in the information theory field.

In this paper, we will use the replica method to average over the channel realizations and obtain moments of the distribution of mutual information. We find that for large antenna numbers n, the second and third moments for the correlated channels is of $\mathcal{O}(1)$ and $\mathcal{O}(1/n)$, respectively, which strongly indicates that the mutual information distribution is Gaussian. Very recently, similar results were verified analytically [20] and numerically for uncorrelated channels and no interference. Also in [22], a method was introduced to calculate the moment generating functional for uncorrelated channels and no interference. To illustrate the usefulness of our results we calculate the mutual information for a number of representative cases. Surprisingly, these methods give accurate results when applied to arrays with even few (three or four) antennas. Thus, this analytic approach provides a powerful tool for analyzing antenna systems with even a few antennas.

In the remainder of this section, we provide some notational definitions used in this paper (Section I-A) and define several quantities of interest (Section I-B). In the next section, we describe the mathematical framework of the methods used to calculate the statistics of the mutual information. In Section II-A, we start with the derivation of our results with white noise and no interferers. Then, in Section II-B, we generalize to the case of colored noise and interference. Section IV deals with the application of these methods to a set of representative cases, with aim to demonstrate the power of our methods, but also to compare them with numerical simulations and previously established results. Finally, in Appendix I, we provide proofs of several useful identities regarding complex integrals, while in Appendix II, we introduce the so-called Grassman variables and prove some of their properties that we use in this paper.

A. Notation

1) Vectors/Matrices: Throughout this paper we will use bold-faced upper case letters to denote matrices, e.g., $\boldsymbol{X}$, with elements given by X_{ab}, bold-faced lower case letters for column vectors, e.g., $\boldsymbol{x}$, with elements x_a, and light-face lower case letters for scalar quantities. Also the superscripts T and $\dagger$ will indicate transpose and Hermitian conjugate operations and $\boldsymbol{I}_n$ will represent the n-dimensional identity matrix.

For $\boldsymbol{A}$, an $m \times n$-dimensional matrix, it is useful to define $vec(\boldsymbol{A})$ as a mn-dimensional vector with entries the columns of $\boldsymbol{A}$ stacked on top of each other. In addition, $\boldsymbol{K} = \boldsymbol{M} \otimes \boldsymbol{N}$ will depict the outer product between the $m \times m$-dimensional matrix $\boldsymbol{M}$ and the $n \times n$-dimensional $\boldsymbol{N}$. $\boldsymbol{K}$ is $mn \times mn$-dimensional and its elements can be written as

$$K_{n(\mu-1)+\nu,\, n(\mu'-1)+\nu'} = M_{\mu,\,\mu'} N_{\nu,\,\nu'}$$

for $\mu,\, \mu' = 1 \cdots m$ and $\nu,\, \nu' = 1 \cdots n$.

Finally, the superscripts/subscripts t, r, i will be used for quantities referring to the transmitter, receiver, and interferers, respectively.

2) Order of Number of Antennas $\mathcal{O}(n^k)$: We will be examining quantities in the limit when all n_t, n_r, n_i are large but their ratios are fixed and finite. We will denote collectively the order in an expansion over the antenna numbers as $\mathcal{O}(n)$, $\mathcal{O}(1)$, $\mathcal{O}(1/n)$, etc., irrespective of whether the particular term involves n_t, n_r, or n_i.

3) Integral Measures: In this paper, we will be dealing with three general types of integrals over matrix elements. We will, therefore, adopt the following notation for their corresponding integration measures. In the first type we will be integrating over the real and imaginary part of the elements of a complex $m_{\text{rows}} \times m_{\text{cols}}$ matrix $\boldsymbol{X}$. The integral measure will be denoted by

$$D_c \boldsymbol{X} = \prod_{a=1}^{m_{\text{rows}}} \prod_{\alpha=1}^{m_{\text{cols}}} \frac{d\text{Re}X_{a\alpha} d\text{Im}X_{a\alpha}}{2\pi} \tag{1}$$

where the subscript c denotes complex matrices. The second type of integration is over pairs of complex square matrices $\boldsymbol{T}$ and $\boldsymbol{R}$. Each element of $\boldsymbol{T}$ and $\boldsymbol{R}$ will be integrated over a contour in the complex plane (to be specified). The corresponding measure will be described as

$$d\mu(\boldsymbol{T}, \boldsymbol{R}) = \prod_{a=1}^{m_{\text{rows}}} \prod_{\alpha=1}^{m_{\text{cols}}} \frac{dT_{a\alpha} dR_{\alpha a}}{2\pi i}. \tag{2}$$

Finally, the third integral type is over so-called Grassman variables, which are introduced and defined in Appendix II. The corresponding integral measure for the $m_{\text{rows}} \times m_{\text{cols}}$-dimensional matrix $\boldsymbol{A}$ and its companion $m_{\text{rows}} \times m_{\text{cols}}$-dimensional matrix $\overline{\boldsymbol{A}}$ is denoted as

$$D_g \boldsymbol{A} = \prod_{a=1}^{m_{\text{rows}}} \prod_{\alpha=1}^{m_{\text{cols}}} dA_{a\alpha} d\overline{A}_{a\alpha} \tag{3}$$

where the subscript g denotes "Grassman."

B. Definitions

We consider the case of single-user transmission from n_t transmit antennas to n_r receive antennas over a narrow-band fading channel. We also include a number n_i of interfering transmitters in addition to the noise at each receiver. The received n_r-dimensional complex signal vector $\boldsymbol{y}$ can be written as

$$\boldsymbol{y} = \sqrt{\frac{\rho_s}{n_t}}\, \boldsymbol{G^s x}^s + \sqrt{\frac{\rho_i}{n_i}}\, \boldsymbol{G^i x}^i + \boldsymbol{z}. \tag{4}$$

$\boldsymbol{G^s}$ is an $n_r \times n_t$ complex matrix with the channel coefficients from the transmitting to the receiving arrays. Similarly, $\boldsymbol{G^i}$ is an $n_r \times n_i$ matrix describing the channel from the interfering antennas to the receiver array. $\boldsymbol{x}^s$ and $\boldsymbol{x}^i$ are the n_t- and n_r-dimensional vectors of transmitted and interfering signals, respectively, both assumed to be zero mean Gaussian. The signal covariance $\boldsymbol{Q} = E[\boldsymbol{x^s x^{s\dagger}}]$ is normalized so that $\text{Tr}\{\boldsymbol{Q}\} = n_t$. Without loss of generality, the interfering signal elements $\boldsymbol{x}^i$ are assumed to be independent and identically distributed (i.i.d.) with identity variance (since they cannot be controlled by the

IEEE Proof

transmitter or receiver, nontrivial correlations in $\boldsymbol{x}^i$ can be absorbed in the channel matrix $\boldsymbol{G^i}$). ρ_s and ρ_i are the signal-to-noise (SNR) and interference-to-noise ratios (INR). The signal to interference and noise ratio can also be written as SINR $= \rho_s/(1+\rho_i)$. Finally, the Gaussian noise vector $\boldsymbol{z}$ has covariance matrix $\boldsymbol{N}$ given by $\boldsymbol{N} = E[\boldsymbol{z}\boldsymbol{z}^\dagger]$, normalized, so that $\mathrm{Tr}\{\boldsymbol{N}\} = n_r$. In general, $\boldsymbol{N}$ is spatially colored, i.e., $\boldsymbol{N} \neq \boldsymbol{I}_{n_r}$.

It is assumed that the receiver knows the channel matrices $\boldsymbol{G^s}$, $\boldsymbol{G^i}$, and the noise covariance $\boldsymbol{N}$. The difference between the terms $\boldsymbol{G^i}\boldsymbol{G^i}^\dagger$ and $\boldsymbol{N}$ is that in the former term the receiver has detailed knowledge of the interfering channel, while in the latter only average properties are available (i.e., the covariance). The transmitter, on the other hand, knows only the statistics of the "noise," $p(\boldsymbol{y}|\{\boldsymbol{x}^s, \boldsymbol{G^s}, \boldsymbol{G^i}\})$, as well as the statistics of the channel $p(\boldsymbol{G^s}, \boldsymbol{G^i})$ (both transmitter and receiver also know ρ_s, ρ_i).

The associated mutual information can be expressed as [7] (the derivation follows directly from [1, Appendix A])

$$I\left(\boldsymbol{y}; \boldsymbol{x}^s \middle| \boldsymbol{G^s}, \boldsymbol{G^i}\right) = h_s - h_i$$
$$h_s = \log\det\left(\boldsymbol{N} + \frac{\rho_i}{n_i}\boldsymbol{G^i}\boldsymbol{G^i}^\dagger + \frac{\rho_s}{n_t}\boldsymbol{G^s}\boldsymbol{Q}\boldsymbol{G^s}^\dagger\right)$$
$$h_i = \log\det\left(\boldsymbol{N} + \frac{\rho_i}{n_i}\boldsymbol{G^i}\boldsymbol{G^i}^\dagger\right). \quad (5)$$

The log above (and throughout the whole paper) represents the natural logarithm and thus I, h_s, and h_i are all expressed in nats. It should be noted that this equation for the mutual information is also applicable to multiuser detection. This case corresponds, e.g., to an uplink scenario with the interferer(s) representing other uplink users.

Due to the underlying randomness of $\boldsymbol{G^s}$ and $\boldsymbol{G^i}$, I is also a random quantity. To analyze its statistics we assume that $\boldsymbol{G^s}$, $\boldsymbol{G^i}$ are zero-mean Gaussian matrices with covariance

$$\langle G^s_{a\alpha} {G^s_{b\beta}}^* \rangle = T^s_{\alpha\beta} R^s_{ab}$$
$$\langle G^i_{a\alpha} {G^i_{b\beta}}^* \rangle = T^i_{\alpha\beta} R^i_{ab} \quad (6)$$

and $\langle G^s_{a\alpha} {G^i_{b\beta}}^* \rangle = 0$. From the above it follows that the correlations of $vec(\boldsymbol{G^s})$ and $vec(\boldsymbol{G^i})$ are $\boldsymbol{R}^s \otimes \boldsymbol{T}^s$ and $\boldsymbol{R}^i \otimes \boldsymbol{T}^i$, respectively. $\boldsymbol{R}^s$ and $\boldsymbol{R}^i$ are $n_r \times n_r$ correlation matrices of the incoming signal and interference at the receiver, respectively, and are normalized so that $\mathrm{Tr}\{\boldsymbol{R}^i\} = \mathrm{Tr}\{\boldsymbol{R}^s\} = n_r$. $\boldsymbol{T}^s$ is the $n_t \times n_t$ correlation matrix describing the antenna correlations at the transmitter array with $\mathrm{Tr}\{\boldsymbol{T}^s\} = n_t$. Similarly, $\boldsymbol{T}^i$ is the $n_i \times n_i$ correlation matrix of the transmitted interference with $\mathrm{Tr}\{\boldsymbol{T}^i\} = n_i$. The simple form of correlations of (6) has been shown [10] to be the leading term in a controlled approximation for diffusive environments. The dependence of these matrices on antenna array properties, polarization, and channel properties such as angle spread have been described in various references [9], [10], [23].

We define the notation $\langle\cdot\rangle$ to denote the ensemble average over channel realizations. Thus, for $O(\boldsymbol{G^s}, \boldsymbol{G^i})$ an arbitrary function of $\boldsymbol{G^s}$, $\boldsymbol{G^i}$, we have

$$\langle O\rangle = \frac{1}{\det(\boldsymbol{T}^s \otimes \boldsymbol{R}^s)\det(\boldsymbol{T}^i \otimes \boldsymbol{R}^i)} \int D_c\boldsymbol{G^s} D_c\boldsymbol{G^i}$$
$$\times e^{-\frac{1}{2}\mathrm{Tr}\{(\boldsymbol{T}^s)^{-1}\boldsymbol{G^s}(\boldsymbol{R}^s)^{-1}\boldsymbol{G^s}^\dagger + (\boldsymbol{T}^i)^{-1}\boldsymbol{G^i}(\boldsymbol{R}^i)^{-1}\boldsymbol{G^i}^\dagger\}} O. \quad (7)$$

Equation (7) is simply an integration over the appropriate Gaussian probability distribution of the $\boldsymbol{G}$'s. To see this more explicitly we can use (6) to define $\boldsymbol{G^s} = [\boldsymbol{R}^s]^{1/2}\boldsymbol{Z}^s[\boldsymbol{T}^s]^{1/2}$ and $\boldsymbol{G^i} = [\boldsymbol{R}^i]^{1/2}\boldsymbol{Z}^i[\boldsymbol{T}^i]^{1/2}$ where $\boldsymbol{Z}^s$ and $\boldsymbol{Z}^i$ are i.i.d. matrices of unit variance. The probability distribution for an i.i.d. matrix is, of course, proportional to $\exp(-\frac{1}{2}\mathrm{Tr}\{\boldsymbol{Z}^\dagger\boldsymbol{Z}\})$. Thus, taking an expectation over the $\boldsymbol{Z}$'s is simply given by

$$\langle O\rangle = \int D_c\boldsymbol{Z}^s D_c\boldsymbol{Z}^i e^{-\frac{1}{2}\mathrm{Tr}\{\boldsymbol{Z}^{s\dagger}\boldsymbol{Z}^s + \boldsymbol{Z}^{i\dagger}\boldsymbol{Z}^i\}}$$
$$\times O\Big(\boldsymbol{G^s} = [\boldsymbol{R}^s]^{1/2}\boldsymbol{Z}^s[\boldsymbol{T}^s]^{1/2}; \boldsymbol{G^i} = [\boldsymbol{R}^i]^{1/2}\boldsymbol{Z}^i[\boldsymbol{T}^i]^{1/2}\Big). \quad (8)$$

This equation can immediately be converted to (7) by a variable transformation, with the determinants out front of (7) being Jacobian of the transformation (see also Appendix I).

II. Mathematical Framework

The purpose of this paper is to analyze the statistics of the mutual information I in (5) for Gaussian channels having correlations given by (6). In this section, we introduce the mathematical framework necessary for deriving analytic expressions of the cumulant moments of I. This method was introduced in this context in [11], [12]. We first introduce the generating function $g(\nu)$ of I

$$g(\nu) = \left\langle \frac{\left[\det\left(\boldsymbol{N} + \frac{\rho_i}{n_i}\boldsymbol{G^i}\boldsymbol{G^i}^\dagger + \frac{\rho_s}{n_t}\boldsymbol{G^s}\boldsymbol{Q}\boldsymbol{G^s}^\dagger\right)\right]^{-\nu}}{\left[\det\left(\boldsymbol{N} + \frac{\rho_i}{n_i}\boldsymbol{G^i}\boldsymbol{G^i}^\dagger\right)\right]^{-\nu}} \right\rangle$$
$$= \left\langle e^{-\nu I}\right\rangle$$
$$= 1 - \nu\langle I\rangle + \frac{\nu^2}{2}\langle I^2\rangle + \cdots. \quad (9)$$

Assuming that $g(\nu)$ is analytic at least in the vicinity of $\nu = 0$, we can express $\log g(\nu)$ as follows:

$$\log g(\nu) = -\nu\langle I\rangle + \sum_{p=2}^{\infty}\frac{(-\nu)^p}{p!}\mathcal{C}_p \quad (10)$$

where $\mathcal{C}_p$ is the pth cumulant moment of I. For example

$$\mathcal{C}_2 = Var(I) = \langle (I - \langle I\rangle)^2\rangle$$

is the variance and

$$\mathcal{C}_3 = Sk(I) = \langle (I - \langle I\rangle)^3\rangle$$

is the skewness of the distribution. It should be noted that $\langle I\rangle$ and $\mathcal{C}_p$ for $p > 2$ are implicit functions of $\boldsymbol{Q}$, ρ_s, ρ_i, etc. For notational simplicity, we will be suppressing this dependence. Thus, to obtain the moments of the mutual information distribution we need to calculate $g(\nu)$ for ν in the vicinity of $\nu = 0$. This is not necessarily any easier than evaluating the moments $\mathcal{C}_p$ directly, which is a notoriously difficult task, since one has to average products of logarithms of random quantities. In contrast, averaging $g(\nu)$ for integer values of ν involves averages over integer powers of determinants of random quantities, in which case some analytic progress can be made. We will therefore make the following assumption.

IEEE Proof

Assumption 1 (Replica Method): $g(\nu)$ evaluated for positive integer values of ν can be analytically continued for real ν, specifically in the vicinity of $\nu = 0^+$.

This assumption, used also in [14], [15], [19], alleviates the problem of dealing with averages of logarithms of random quantities, since the logarithm is obtained after calculating $g(\nu)$. It has seen widespread use in the field of physics for more than 25 years [13], and, in many cases [11] can be been shown to produce exactly the same results as systematic series expansions. Thus, the replica method can be seen essentially a bookkeeping tool. Here, we will be using it without any direct proof, although we will be comparing some of our final results to Monte Carlo simulations to demonstrate their validity.

In the next sections we will be making use of Identities 1–6, which are stated and proved in the Appendixes.

A. Case of White Noise and No Interference

In order to expose the basic steps of the calculation with more clarity, we will start with the case of i.i.d. noise $\boldsymbol{N} = \boldsymbol{I}_{n_r}$ and no interference $\rho_i = 0$. In this case, $g(\nu)$ of (9) can be simplified to

$$g(\nu) = \left\langle \left[\det\left(\boldsymbol{I}_{n_r} + \frac{\rho_s}{n_t} \boldsymbol{G^s Q G^{s\dagger}} \right) \right]^{-\nu} \right\rangle. \tag{11}$$

In the following analysis, we will assume ν to be an arbitrary positive integer. Using Identity 1 in Appendix I we can write $g(\nu)$ as

$$g(\nu) = \int D_c \boldsymbol{X} e^{-\frac{1}{2}\mathrm{Tr}\{\boldsymbol{X}^\dagger \boldsymbol{X}\}} \left\langle e^{-\frac{\rho_s}{2n_t}\mathrm{Tr}\{\boldsymbol{X}^\dagger \boldsymbol{G^s Q G^{s\dagger}} \boldsymbol{X}\}} \right\rangle \tag{12}$$

where $\boldsymbol{X}$ is a complex $n_r \times \nu$ matrix. The word "replica" stems from thinking of the matrix $\boldsymbol{X}$ as being ν versions (or "replicas") of an n_r-dimensional vector, thus producing ν-replicas of the determinant.

The bracketed quantity in (12) can be rewritten, using Identity 2 (with $\boldsymbol{M} = \boldsymbol{I}_{n_t}$, $\boldsymbol{N} = \boldsymbol{I}_{n_r}$, $\boldsymbol{A} = \boldsymbol{B} = \boldsymbol{Q}^{1/2} \boldsymbol{G^s}^\dagger \boldsymbol{X} \sqrt{\rho_s/n_t}$, where $\boldsymbol{Q}^{1/2}$ is the matrix square root of $\boldsymbol{Q}$, which is well defined since $\boldsymbol{Q}$ is nonnegative definite) by introducing an integral over a complex $n_t \times \nu$ matrix $\boldsymbol{Y}$

$$\left\langle e^{-\frac{\rho_s}{2n_t}\mathrm{Tr}\{\boldsymbol{X}^\dagger \boldsymbol{G^s Q G^{s\dagger}} \boldsymbol{X}\}} \right\rangle = \int D_c \boldsymbol{Y} e^{-\frac{1}{2}\mathrm{Tr}\boldsymbol{Y}^\dagger \boldsymbol{Y}} \times \left\langle e^{-\sqrt{\frac{\rho_s}{4n_t}}\mathrm{Tr}\{\boldsymbol{X}^\dagger \boldsymbol{G^s} \boldsymbol{Q}^{1/2} \boldsymbol{Y} - \boldsymbol{Y}^\dagger \boldsymbol{Q}^{1/2} \boldsymbol{G^{s\dagger}} \boldsymbol{X}\}} \right\rangle. \tag{13}$$

At this point, we can use Identity 2 to integrate over $\boldsymbol{G^s}$. Combining the result of (7), (12), and (13), $g(\nu)$ can be expressed as

$$g(\nu) = \int D_c \boldsymbol{X} D_c \boldsymbol{Y} \cdot e^{-\frac{1}{2}\mathrm{Tr}\{\boldsymbol{X}^\dagger \boldsymbol{X} + \boldsymbol{Y}^\dagger \boldsymbol{Y} + \frac{\rho_s}{2n_t} \boldsymbol{Y}^\dagger \boldsymbol{Q}^{1/2} \boldsymbol{T^s} \boldsymbol{Q}^{1/2} \boldsymbol{Y} \boldsymbol{X}^\dagger \boldsymbol{R^s} \boldsymbol{X}\}}. \tag{14}$$

To make progress we now need to express the last term in the exponent in a quadratic form in terms of $\boldsymbol{X}, \boldsymbol{Y}$. This can be done by using Identity 3 and introducing $\nu \times \nu$ matrices $\boldsymbol{R}_1, \boldsymbol{T}_1$. Thus, the last term in the exponent of (14) becomes

$$\exp\left[\frac{\rho_s}{2n_t} \mathrm{Tr}\left\{ \boldsymbol{Y}^\dagger \boldsymbol{Q}^{1/2} \boldsymbol{T^s} \boldsymbol{Q}^{1/2} \boldsymbol{Y} \boldsymbol{X}^\dagger \boldsymbol{R^s} \boldsymbol{X} \right\} \right] = \int d\mu(\boldsymbol{T}_1, \boldsymbol{R}_1) \exp\left[\mathrm{Tr}\{\boldsymbol{T}_1 \boldsymbol{R}_1\} - \sqrt{\frac{\rho_s}{4n_t}} \times \mathrm{Tr}\left\{ \boldsymbol{T}_1 \boldsymbol{Y}^\dagger \boldsymbol{Q}^{1/2} \boldsymbol{T^s} \boldsymbol{Q}^{1/2} \boldsymbol{Y} + \boldsymbol{X}^\dagger \boldsymbol{R^s} \boldsymbol{X} \boldsymbol{R}_1 \right\} \right]. \tag{15}$$

The application of Identity 3 and the introduction of matrices $\boldsymbol{R}_1$ and $\boldsymbol{T}_1$ is a particular form of a Hubbard–Stratonovich transformation, also used in [11]. The Hubbard–Stratonovich transformation is generally a convenient tool to decouple terms in exponents, such as in (14), at the expense of introducing auxiliary variables, and has been used extensively in physics [14], [15]. The usefulness of this method is in that it allows the integration of certain quantities $(\boldsymbol{X}, \boldsymbol{Y})$ of limited relevance, and introduces auxiliary quantities (such as $\boldsymbol{R}_1$ and $\boldsymbol{T}_1$), which will prove to have particular importance in the final answer.

Combining (14), (15), and using Identity 1, we can now integrate out $\boldsymbol{X}, \boldsymbol{Y}$, resulting in

$$g(\nu) = \int d\mu\,(\boldsymbol{T}_1, \boldsymbol{R}_1)\, e^{-\mathcal{S}} \tag{16}$$

where

$$\mathcal{S} = \log\det\left(\boldsymbol{I}_{n_r} \otimes \boldsymbol{I}_\nu + \sqrt{\frac{\rho_s}{n_t}} \boldsymbol{R^s} \otimes \boldsymbol{R}_1 \right) + \log\det\left(\boldsymbol{I}_{n_t} \otimes \boldsymbol{I}_\nu + \sqrt{\frac{\rho_s}{n_t}} \boldsymbol{Q}^{1/2} \boldsymbol{T^s} \boldsymbol{Q}^{1/2} \otimes \boldsymbol{T}_1 \right) - \mathrm{Tr}\{\boldsymbol{T}_1\, \boldsymbol{R}_1\}. \tag{17}$$

At this point, ν is still a positive integer, which should be taken to zero following Assumption 1. However, before doing this, we will first take the limit of large antenna numbers n_t, $n_r \gg 1$. In this limit, the saddle point method of evaluating the integral (described below) becomes accurate. Subsequently, we will take the $\nu \to 0^+$ limit.

Assumption 2 (Interchanging Limits): The limits $n \to \infty$ and $\nu \to 0^+$ in evaluating $g(\nu)$ in (16) can be interchanged by first taking the former and then the latter without changing the final answer.

This assumption will be applied without proof. As we shall see later, the two limits of large antenna numbers and small ν are related: Higher terms in the expansion in ν involve successively higher terms in a $1/n$ expansion.

The saddle point approximation is a well known method to obtain a systematic expansion of an integral of a rapidly varying function in complex space. This methods involves deforming the contour of integration to pass through one point (or more), where the exponent of the integrand is stationary, i.e., the point where the derivatives of the exponent with respect to all integration variables vanish [24]. Thus integration around the point(s) along the steepest descent paths carries asymptotically the highest weight of the integral. Usually, this approximation is valid when the exponent of the integrand is large, in which

IEEE Proof

case the method has been also called Laplace's method (see [24]), or Varadhan's theorem [25]. In the case of (16) and (17), when $n_t, n_r \gg 1$, the exponent $\mathcal{S}$ is nominally of order $\mathcal{O}(\nu n)$. Following Assumption 2 and, thus, keeping ν *fixed* we may apply the saddle point approximation to calculate the integral of (16) and *then* take the $\nu \to 0^+$ limit. It should be stressed that for a fixed positive integer ν, the saddle point analysis of $\mathcal{S}$ and of $g(\nu)$ is a straightforward exercise in asymptotic analysis. The only additional complexity to the standard textbook treatment of this topic [24] is that $\mathcal{S}$ involves integrals over multiple variables (the elements of $\boldsymbol{T}_1$, $\boldsymbol{R}_1$).

Rather than search for saddle point solutions over all possible complex $\boldsymbol{T}_1$ and $\boldsymbol{R}_1$ matrices, we are going to invoke the following hypothesis without proof.

Assumption 3 (Replica Invariance): The relevant saddle-point solution for (16) involves matrices $\boldsymbol{T}_1$, $\boldsymbol{R}_1$, which are invariant in replica space and thus are proportional to the identity matrix $\boldsymbol{I}_\nu$.

The above hypothesis, used heavily in physics [14], [15], basically states that there is no preferred direction in the space of replicas, and thus if any saddle point solution is valid, so is any unitary transformation thereof in replica space. Although we are not going to provide a proof here, it should be noted that in [11] it is shown that the results obtained by this method are identical to those using a systematic expansion.

1) Saddle Point Analysis: The assumed form of $\boldsymbol{T}_1$ and $\boldsymbol{R}_1$ at the saddle point is $t_1\sqrt{n_t}\,\boldsymbol{I}_\nu$ and $r_1\sqrt{n_t}\,\boldsymbol{I}_\nu$, respectively. The extra factor of $\sqrt{n_t}$ has been included for convenience, as will become evident later. To consider the vicinity around the saddle point, we thus rewrite $\boldsymbol{T}_1$, $\boldsymbol{R}_1$ as

$$\begin{aligned}\boldsymbol{T}_1 &= t_1\sqrt{n_t}\,\boldsymbol{I}_\nu + \boldsymbol{\delta T}_1\\ \boldsymbol{R}_1 &= r_1\sqrt{n_t}\,\boldsymbol{I}_\nu + \boldsymbol{\delta R}_1\end{aligned} \tag{18}$$

where $\boldsymbol{\delta T}_1$, $\boldsymbol{\delta R}_1$ are $\nu \times \nu$ matrices representing deviations around the saddle point. One can then expand $\mathcal{S}$ of (17) in a Taylor series of increasing powers of $\boldsymbol{\delta T}_1$, $\boldsymbol{\delta R}_1$ as follows:

$$\mathcal{S} = \mathcal{S}_0 + \mathcal{S}_1 + \mathcal{S}_2 + \mathcal{S}_3 + \mathcal{S}_4 \cdots \tag{19}$$

with $\mathcal{S}_p$ containing pth-order terms in $\boldsymbol{\delta T}_1$, $\boldsymbol{\delta R}_1$. These terms can be obtained explicitly by differentiating (17)

$$\begin{aligned}\mathcal{S}_0 &= \nu[\log\det(\boldsymbol{I}_{n_t} + \sqrt{\rho_s}\,\boldsymbol{Q}\boldsymbol{T}^s t_1)\\ &\quad + \log\det(\boldsymbol{I}_{n_r} + \sqrt{\rho_s}\,\boldsymbol{R}^s r_1) - n_t r_1 t_1] \qquad (20)\\ &= \nu\Gamma_s \qquad (21)\end{aligned}$$

$$\mathcal{S}_1 = (M_{r,1} - \sqrt{n_t}\,t_1)\,\mathrm{Tr}\{\boldsymbol{\delta R}_1\} + (M_{t,1} - \sqrt{n_t}\,r_1)\,\mathrm{Tr}\{\boldsymbol{\delta T}_1\} \tag{22}$$

$$\begin{aligned}\mathcal{S}_2 &= -\frac{1}{2}\mathrm{Tr}\{M_{r,2}\,(\boldsymbol{\delta R}_1)^2 + M_{t,2}\,(\boldsymbol{\delta T}_1)^2 + 2\boldsymbol{\delta R}_1\,\boldsymbol{\delta T}_1\}\\ &= \frac{1}{2}\mathrm{Tr}\left\{\begin{bmatrix}\boldsymbol{\delta R}_1\\ \boldsymbol{\delta T}_1\end{bmatrix}^T \boldsymbol{V}\begin{bmatrix}\boldsymbol{\delta R}_1\\ \boldsymbol{\delta T}_1\end{bmatrix}\right\} \qquad (23)\end{aligned}$$

where the Hessian $\boldsymbol{V}$ is given by

$$\boldsymbol{V} = \begin{bmatrix}-M_{r,2} & -1\\ -1 & -M_{t,2}\end{bmatrix}. \tag{24}$$

For $p > 2$, the expanded terms simplify to

$$\mathcal{S}_p = \frac{(-1)^p}{p}\,\mathrm{Tr}\{M_{r,p}(\boldsymbol{\delta R}_1)^p + M_{t,p}(\boldsymbol{\delta T}_1)^p\}. \tag{25}$$

The coefficients in this Taylor expansion have the particularly simple form

$$\begin{aligned}M_{t,p} &= \left(\frac{\rho_s}{n_t}\right)^{p/2}\mathrm{Tr}\Big\{\Big[\boldsymbol{Q}\boldsymbol{T}^s(\boldsymbol{I}_{n_t} + \sqrt{\rho_s}\,t_1\boldsymbol{Q}\boldsymbol{T}^s)^{-1}\Big]^p\Big\}\\ M_{r,p} &= \left(\frac{\rho_s}{n_t}\right)^{p/2}\mathrm{Tr}\Big\{\Big[\boldsymbol{R}^s(\boldsymbol{I}_{n_r} + \sqrt{\rho_s}\,r_1\boldsymbol{R}^s)^{-1}\Big]^p\Big\}. \qquad (26)\end{aligned}$$

Note that in (20), (26) it is possible to combine the two $\boldsymbol{Q}^{1/2}$ into $\boldsymbol{Q}$. The saddle point solution of (16) and, hence, the corresponding values of t_1, r_1 is found by demanding that $\mathcal{S}$ is stationary with respect to variations in $\boldsymbol{T}_1$, $\boldsymbol{R}_1$ [24]. This means that $\mathcal{S}_1 = 0$, which is analogous to setting the first derivative of a function to zero, in order to find its maximum or minimum. This produces the following saddle point equations:

$$r_1 = \frac{1}{n_t}\,M_{t,1} = \frac{\sqrt{\rho_s}}{n_t}\,\mathrm{Tr}\Big\{\boldsymbol{Q}\boldsymbol{T}^s[\boldsymbol{I}_{n_t} + \sqrt{\rho_s}\,t_1\,\boldsymbol{Q}\boldsymbol{T}^s]^{-1}\Big\} \tag{27}$$

$$t_1 = \frac{1}{n_t}\,M_{r,1} = \frac{\sqrt{\rho_s}}{n_t}\,\mathrm{Tr}\Big\{\boldsymbol{R}^s[\boldsymbol{I}_{n_r} + \sqrt{\rho_s}\,\boldsymbol{R}^s r_1]^{-1}\Big\}. \tag{28}$$

It is interesting to note that the solutions to (27) and (28) maximize Γ_s for real and positive t_1, r_1. It is important to note that for generic full-rank matrices $\boldsymbol{Q}\boldsymbol{T}^s$, $\boldsymbol{R}^s$ both r_1 and t_1 are generally of order unity, $r_1, t_1 = \mathcal{O}(1)$. Thus, the expansion coefficients $M_{t,p}, M_{r,p}$ are generally of order $\mathcal{O}(n^{1-p/2})$, successively decreasing in size for increasing p.

The small parameter controlling this approximation is, therefore, $n^{-1/2}$, making this saddle point solution increasingly accurate for large n. Therefore, the aim of this analysis is to calculate successively higher order terms in $n^{-1/2}$ and classify each resulting quantity in terms of their powers of ν to the appropriate cumulant moment $\mathcal{C}_p$ in (10). This matching of powers of ν implicitly assumes that the expansion in (10) is valid for integer ν, as described in Assumption 1. Thus, for example, terms of order $\mathcal{O}(\nu n)$ and $\mathcal{O}(\nu/n)$ will both contribute to $\langle I\rangle$, as we shall see.

2) Mean Mutual Information: We start with the leading term of $g(\nu)$ in the saddle point approximation:

$$g(\nu) \approx \exp(-\mathcal{S}_0) = \exp(-\nu\Gamma_s)$$

where Γ_s is evaluated in (21) using (27) and (28). We thus see from (10) and (11) that the leading term in the expansion of $\langle I\rangle$ is Γ_s and note that $\langle I\rangle = \mathcal{O}(n)$.

3) Variance of the Mutual Information: To obtain the $\mathcal{O}(\nu^2)$ term in the expansion of $\log g(\nu)$ in (10), we need to include the next nonvanishing term $\mathcal{S}_2$. Thus, for the moment we neglect higher order terms $\mathcal{S}_p$ for $p > 2$.

Noting the measure-preserving transformation $\boldsymbol{T}_1$, $\boldsymbol{R}_1 \to \boldsymbol{\delta T}_1$, $\boldsymbol{\delta R}_1$ of (18) we have

$$\begin{aligned}g(\nu) &= e^{-\mathcal{S}_0}\int d\mu(\boldsymbol{\delta T}_1, \boldsymbol{\delta R}_1)e^{-\mathcal{S}_2}\\ &= e^{-\mathcal{S}_0}\int d\mu(\boldsymbol{\delta T}_1, \boldsymbol{\delta R}_1)\\ &\quad\times\exp\left(-\frac{1}{2}\sum_{a,b=1}^{\nu}[\delta T_{1,ab}\,\delta R_{1,ab}]\boldsymbol{V}[\delta T_{1,ba}\,\delta R_{1,ba}]^T\right).\end{aligned} \tag{29}$$

IEEE Proof

To diagonalize the exponent of the preceding equation, we rotate each pair $\boldsymbol{\delta T}_1, \boldsymbol{\delta R}_1$ to a new basis of $\nu \times \nu$-dimensional matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$. In particular,

$$[W_{1,ab}\, W_{2,ab}]^T = \boldsymbol{U}[\delta T_{1,ab}\, \delta R_{1,ab}]^T$$

where $\boldsymbol{U}$ is an orthogonal matrix such that $\boldsymbol{UVU^T} = \mathrm{diag}(\boldsymbol{v})$ and $\boldsymbol{v} = [v_1\ v_2]^T$ is the vector of eigenvalues of $\boldsymbol{V}$. We may now rewrite (29) as

$$g(\nu) = e^{-\mathcal{S}_0} \int d\mu(\boldsymbol{W}_1, \boldsymbol{W}_2) \times \exp\left(-\frac{1}{2}\sum_{a,b=1}^{\nu}\left[v_1 W_{1,ab} W_{1,ba} + v_2 W_{2,ab} W_{2,ba}\right]\right). \quad (30)$$

We note that since $M_{t,2}$, $M_{r,2} < 1$, the eigenvalues v_1 and v_2 will have opposite signs, namely, $v_1 > 0$ and $v_2 < 0$. Thus, to specify the steepest descent paths for $W_{1,ab}$ we take $\boldsymbol{W}_1 = \boldsymbol{W}_1^\dagger$, while for $\boldsymbol{W}_2$ we take $\boldsymbol{W}_2 = i\boldsymbol{W}_2^\dagger$. The extra factor of i cancels out with the one appearing in the integration measure (see (2)). This results in

$$\begin{aligned} g(\nu) &= e^{-\mathcal{S}_0} \prod_{k=1}^{2}\left\{\prod_{a=1}^{\nu}\int\frac{dW_{k,aa}}{\sqrt{2\pi}}\prod_{b>a}^{\nu}\iint\frac{d\mathrm{Re}W_{k,ab}\, d\mathrm{Im}W_{k,ab}}{2\pi} \times \exp\left[-\tfrac{1}{2}|v_k|\,\mathrm{Tr}\left\{\boldsymbol{W}_k^\dagger \boldsymbol{W}_k\right\}\right]\right\} \\ &= e^{-\mathcal{S}_0}|v_1 v_2|^{-\frac{\nu^2}{2}} = e^{-\mathcal{S}_0}|\det \boldsymbol{V}|^{-\frac{\nu^2}{2}}. \end{aligned} \quad (31)$$

Comparing (10) to (31) and matching order by order the terms of the ν-Taylor expansion of the exponent of $g(\nu)$, we can identify the leading term in the variance of the mutual information to be

$$\langle I^2\rangle - \langle I\rangle^2 = \mathcal{C}_2 = -\log|\det \boldsymbol{V}| + \cdots \quad (32)$$

$$= -\log(1 - M_{r,2}M_{t,2}) + \cdots. \quad (33)$$

We note that since $M_{r,2}$, $M_{t,2} = \mathcal{O}(1)$, the variance is also $\mathcal{O}(1)$ in the expansion of $n^{-1/2}$ when both n_t and n_r are of the same order. (However, if n_r is fixed while n_t increases, we find that $\mathcal{C}_2 = \mathcal{O}(n^{-1})$, in agreement with [20].) Also, we see that no term proportional to ν is produced from $\mathcal{S}_2$. Thus, no term of $\mathcal{O}(1)$ in the antenna number appears in $\langle I\rangle$, resulting in $\langle I\rangle = \Gamma_s + o(1)$.

4) Higher Order Terms: To obtain higher order corrections in the small parameter $n^{-1/2}$, we need to take into account the terms $\mathcal{S}_p$ for $p > 2$ in (25). Here we will set the framework for the perturbative analysis and will use it to calculate the leading additional term in the $n^{-1/2}$ expansion.

We define an expectation bracket of an arbitrary operator $O(\boldsymbol{\delta T}_1, \boldsymbol{\delta R}_1)$ as

$$\langle\langle O\rangle\rangle = |\det \boldsymbol{V}|^{\nu^2/2}\int d\mu(\boldsymbol{\delta T}_1,\ \boldsymbol{\delta R}_1) e^{-\mathcal{S}_2} O(\boldsymbol{\delta T}_1,\ \boldsymbol{\delta R}_1). \quad (34)$$

The integration over $\boldsymbol{\delta T}_1, \boldsymbol{\delta R}_1$ is performed as described in the previous section. Using Wick's theorem discussed in Appendix I and (129) we can obtain the following expectations:

$$\langle\langle \delta R_{1,ab}\delta R_{1,cd}\rangle\rangle = -\delta_{ad}\delta_{bc} M_{t,2}/\det \boldsymbol{V} \quad (35)$$

$$\langle\langle \delta T_{1,ab}\delta T_{1,cd}\rangle\rangle = -\delta_{ad}\delta_{bc} M_{r,2}/\det \boldsymbol{V} \quad (36)$$

$$\langle\langle \delta T_{1,ab}\delta R_{1,cd}\rangle\rangle = \delta_{ad}\delta_{bc}/\det \boldsymbol{V}. \quad (37)$$

In addition, the expectation of any odd power of $\boldsymbol{\delta T}_1,\ \boldsymbol{\delta R}_1$ must vanish by symmetry (see Appendix I). As a result, the perturbative expansion discussed here involves only integer powers of $1/n$.

With this bracket notation we can rewrite

$$g(\nu) = e^{-\mathcal{S}_0}\, e^{-\mathcal{S}_2}\, e^{-(\mathcal{S}_3+\mathcal{S}_4+\cdots)} \quad (38)$$

$$= e^{-\nu\Gamma_s}|\det \boldsymbol{V}|^{-\nu^2/2}\left\langle\left\langle 1 - (\mathcal{S}_3+\mathcal{S}_4+\cdots) + \tfrac{1}{2}(\mathcal{S}_3+\mathcal{S}_4+\cdots)^2 + \cdots\right\rangle\right\rangle. \quad (39)$$

At this point, it is interesting to count powers of n in the various terms of the expansion. Since $M_{t,2}$ and $M_{r,2}$ are of $\mathcal{O}(1)$, any quadratic correlation, such as $\langle\langle \delta R_{1,ab}\delta T_{1,cd}\rangle\rangle$, is $\mathcal{O}(1)$. Thus, we can count powers of $1/n$ by inspecting the various terms. For example, a term $\langle\langle \mathcal{S}_p\rangle\rangle$ is of order $n^{-p/2+1}$, but is zero for p odd. A term $\langle\langle \mathcal{S}_p\mathcal{S}_q\rangle\rangle$ is of order $n^{-p/2-q/2+2}$ but is zero for $p+q$ odd; and so forth. We now regroup the terms in the above expansion by their order in $1/n$

$$g(\nu) = e^{-\mathcal{S}_0}\left[|\det \boldsymbol{V}|^{-\nu^2/2}\right][1 + D_1 + D_2 + D_3 + \cdots] \quad (40)$$

where

$$D_1 = \left\langle\left\langle \mathcal{S}_4 + \tfrac{1}{2}\mathcal{S}_3^2\right\rangle\right\rangle \quad (41)$$

$$D_2 = \left\langle\left\langle \mathcal{S}_6 + \tfrac{1}{2}\mathcal{S}_4^2 + \mathcal{S}_3\mathcal{S}_5 + \tfrac{1}{24}\mathcal{S}_3^4\right\rangle\right\rangle \quad (42)$$

$$\vdots\ . \quad (43)$$

Here, we have regrouped all terms which are order $1/n^p$ into the term D_p. Thus, for example, D_1 contains all terms of order $1/n$.

To evaluate D_1 we need to calculate $\langle\langle \mathcal{S}_4\rangle\rangle$ and $\langle\langle \mathcal{S}_3^2\rangle\rangle$. Referring back to the definition of $\mathcal{S}_4$ in (25), we see that we will need to evaluate $\langle\langle \mathrm{Tr}\{(\boldsymbol{\delta R}_1)^4\}\rangle\rangle$. We do this as follows:

$$\left\langle\left\langle \mathrm{Tr}\left\{(\boldsymbol{\delta R}_1)^4\right\}\right\rangle\right\rangle = \sum_{a,b,c,d=1}^{\nu}\langle\langle \delta R_{1,ab}\delta R_{1,bc}\delta R_{1,cd}\delta R_{1,da}\rangle\rangle \quad (44)$$

$$= \sum_{a,b,c,d=1}^{\nu}\left[\langle\langle \delta R_{1,ab}\delta R_{1,bc}\rangle\rangle\langle\langle \delta R_{1,cd}\delta R_{1,da}\rangle\rangle + \langle\langle \delta R_{1,ab}\delta R_{1,cd}\rangle\rangle\langle\langle \delta R_{1,bc}\delta R_{1,da}\rangle\rangle + \langle\langle \delta R_{1,ab}\delta R_{1,da}\rangle\rangle\langle\langle \delta R_{1,cd}\delta R_{1,bc}\rangle\rangle\right] \quad (45)$$

$$= (2\nu^3 + \nu)(M_{t,2}/\det \boldsymbol{V})^2. \quad (46)$$

We have used Wick's theorem (127) to go from (44) to (45) and in the last step we substituted the value of the quadratic expectations given by (35). Similarly, we obtain

$$\left\langle\left\langle \mathrm{Tr}\left\{(\boldsymbol{\delta T}_1)^4\right\}\right\rangle\right\rangle = (2\nu^3 + \nu)(M_{r,2}/\det \boldsymbol{V})^2. \quad (47)$$

Combining (44)–(47) results in

$$\langle\langle \mathcal{S}_4\rangle\rangle = (2\nu^3 + \nu)a_1 \quad (48)$$

where

$$a_1 = \frac{1}{4}\frac{1}{(\det \boldsymbol{V})^2}\left(M_{t,2}^2 M_{r,4} + M_{r,2}^2 M_{t,4}\right) \quad (49)$$

which we see is indeed of order $1/n$. The evaluation of $\langle\langle \mathcal{S}_3^2\rangle\rangle$ can be performed similarly (with a bit more tedious algebra) to obtain

$$\langle\langle \mathcal{S}_3^2\rangle\rangle = (3\nu^3 + 3\nu)a_2 + 9(\nu^3)a_3 \quad (50)$$

IEEE Proof

where

$$a_2 = \frac{1}{9(\det \boldsymbol{V})^3}\left(-M_{t,3}^2 M_{r,2}^3 - M_{r,3}^2 M_{t,2}^3 + 2M_{t,3}M_{r,3}\right) \tag{51}$$

$$a_3 = \frac{1}{9(\det \boldsymbol{V})^3}\left(-M_{t,3}^2 M_{r,2}^3 - M_{r,3}^2 M_{t,2}^3 + 2M_{t,3}M_{r,3}M_{r,2}M_{t,2}\right). \tag{52}$$

We note that all three a_1, a_2, $a_3 = \mathcal{O}(n^{-1})$. Thus, we have

$$D_1 = \left(2a_1 + \frac{3}{2}a_2 + \frac{9}{2}a_3\right)\nu^3 + \left(a_1 + \frac{3}{2}a_2\right)\nu. \tag{53}$$

Calculation of higher terms, such as D_2 is mainly a matter of careful bookkeeping. Although we will not perform such a calculation here, it is useful to note that D_2 generates only terms proportional to ν^4 and ν^2.

Using the cumulant expansion notation of (10) we can write

$$\log g(\nu) = -\nu\Gamma_s + \frac{\nu^2}{2}\mathcal{C}_2 - \frac{\nu^3}{6}\mathcal{C}_3 + \cdots \tag{54}$$

$$= -\nu\Gamma_s - \frac{\nu^2}{2}\log|\det \boldsymbol{V}| + \log(1 + D_1 + D_2 + \cdots) \tag{55}$$

$$= -\nu\Gamma_s - \frac{\nu^2}{2}\log|\det \boldsymbol{V}| + (D_1 + D_2 + \cdots) - \frac{1}{2}(D_1 + \cdots)^2 + \cdots \tag{56}$$

$$= -\nu\left[\Gamma_s - \left(a_1 + \frac{3}{2}a_2\right)\right] - \frac{\nu^2}{2}\log|\det \boldsymbol{V}| + \nu^3\left(2a_1 + \frac{3}{2}a_2 + \frac{9}{2}a_3\right) + \mathcal{O}(1/n^2) \tag{57}$$

such that we can simply extract the desired cumulants.

5) Summary of Results for $\boldsymbol{N} = \boldsymbol{I}_{n_r}$, $\rho_i = 0$: To summarize the case of white noise and no interferers, we have derived the following results, with Γ_s given by (20); the a_k's given in (49), (51), and (52); the M parameter's given in (26); and $\boldsymbol{V}$ given by (24); with the parameters r_1 and t_1 given by (27) and (28)

$$\langle I\rangle = \Gamma_s - \left(a_1 + \frac{3}{2}a_2\right) + \mathcal{O}(1/n^3) = \Gamma_s + \left[\frac{1}{6\det \boldsymbol{V}^3}\left(M_{t,3}^2 M_{r,2}^3 + M_{r,3}^2 M_{t,2}^3 - 2M_{t,3}M_{r,3}\right) - \frac{1}{4\det \boldsymbol{V}^2}\left(M_{t,2}^2 M_{r,4} + M_{r,2}^2 M_{t,4}\right)\right] + \mathcal{O}(1/n^3) \tag{58}$$

with Γ_s being of order n and the correction in square brackets being of order $1/n$. The next correction to $\langle I\rangle$ will be in order $1/n^3$, because we know that all terms of order $1/n^2$ will generate only even powers of ν, thus, not providing a contribution to $\langle I\rangle$. From (32) the variance is given by

$$\mathcal{C}_2 = Var(I) = \langle I^2\rangle - \langle I\rangle^2 = -\log|\det \boldsymbol{V}| + \mathcal{O}(1/n^2). \tag{59}$$

Here the leading term is order 1, and with no correction at order $1/n$. This term, however, may have a correction at order $1/n^2$. Finally, by combining all terms proportional to ν^3 in (53) we obtain the skewness

$$\mathcal{C}_3 = Sk(I) = -6\left(2a_1 + \tfrac{3}{2}a_2 + \tfrac{9}{2}a_3\right) + \mathcal{O}(1/n^3) = \frac{4}{\det \boldsymbol{V}^3}\left(M_{t,3}^2 M_{r,2}^3 + M_{r,3}^2 M_{t,2}^3 - 2M_{t,3}M_{r,3}\right) - \frac{3}{\det \boldsymbol{V}^2}\left(M_{t,2}^2 M_{r,4} + M_{r,2}^2 M_{t,4} + 2M_{t,3}M_{r,3}\right) + \mathcal{O}(1/n^3). \tag{60}$$

The expansion can be continued to higher order in $1/n$ straightforwardly, requiring only additional algebraic effort.

B. Analysis of General Case

Now we will generalize the methods discussed in the preceding section to include the presence of interferers $\boldsymbol{G^i}$ with $\rho_i \neq 0$ and a general noise variance $\boldsymbol{N}$ in the calculation of the statistical properties of the mutual information I.

We thus start with the general form of the generating function $g(\nu)$ in (9). One difference from the simpler case discussed above is that the interference channel matrix $\boldsymbol{G^i}$ appears in two distinct determinants and, most importantly, that these determinants appear with oppositely signed powers. This difference will require us to introduce a somewhat more complicated formalism, as we shall see later.

We first apply Identities 1 and 2 to the determinant in the numerator of (9), as we did in (12), (13). The only difference is that we need to introduce an additional $n_i \times \nu$ matrix $\boldsymbol{Z}$ to disentangle the product $\boldsymbol{G^i}\boldsymbol{G^i}^\dagger$. We thus have

$$\left[\det\left(\boldsymbol{N} + \frac{\rho_i}{n_i}\boldsymbol{G^i}\boldsymbol{G^i}^\dagger + \frac{\rho_s}{n_t}\boldsymbol{G^s}\boldsymbol{G^s}^\dagger\right)\right]^{-\nu} = \int D_c\boldsymbol{X}D_c\boldsymbol{Y}D_c\boldsymbol{Z} \times \exp\left[-\frac{1}{2}\mathrm{Tr}\{\boldsymbol{X}^\dagger\boldsymbol{N}\boldsymbol{X} + \boldsymbol{Y}^\dagger\boldsymbol{Y} + \boldsymbol{Z}^\dagger\boldsymbol{Z} + \boldsymbol{Y}^\dagger\boldsymbol{G^s}^\dagger\boldsymbol{X} - \boldsymbol{X}^\dagger\boldsymbol{G^s}\boldsymbol{Y} + \boldsymbol{Z}^\dagger\boldsymbol{G^i}^\dagger\boldsymbol{X} - \boldsymbol{X}^\dagger\boldsymbol{G^i}\boldsymbol{Z}\}\right] \tag{61}$$

where $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ are $n_r \times \nu$-, $n_t \times \nu$-, and $n_i \times \nu$-dimensional complex matrices, respectively. Note that we have suppressed $\boldsymbol{Q}$ here for simplicity and we will reintroduce it at the end of the calculation.

Next we would like to express the other determinant in (9) in terms of complex integrals, in order to bring the matrix $\boldsymbol{G^i}$ to the exponent, and thus be able to average over $\boldsymbol{G^i}$ simultaneously from both determinants. However, Identity 1 applies only to determinants with *negative* (integer) powers. We thus need to introduce integrals over Grassman, instead of complex, variables, which yield determinants with *positive* (integer) powers. A summary of the properties of these variables as well as the proof of the necessary identities can be found in Appendix II-B.

Thus, applying Identity 4 we have the desired effect, namely,

$$\left[\det\left(\boldsymbol{N} + \frac{\rho_i}{n_i}\boldsymbol{G^i}\boldsymbol{G^i}^\dagger\right)\right]^{\nu} = \int D_g\boldsymbol{A}\exp\left[\mathrm{Tr}\left\{\overline{\boldsymbol{A}}\boldsymbol{N}\boldsymbol{A} + \frac{\rho_i}{n_i}\overline{\boldsymbol{A}}\boldsymbol{G^i}\boldsymbol{G^i}^\dagger\boldsymbol{A}\right\}\right] \tag{62}$$

IEEE Proof

where $\overline{\boldsymbol{A}}$, $\boldsymbol{A}$ are independent $\nu \times n_i$ and $n_i \times \nu$ matrices of Grassman variables. The integration measure $D_g\boldsymbol{A}$ is given in (3). We note that due to the "anticommutation" rules between Grassman variables (see (131) and the discussion after (136)) the order of integration and thus the ordering of $dA_{\alpha a}$ is important. Applying now Identity 5, we disentangle the quadratic terms of $\boldsymbol{G^i G^{i\dagger}}$ by introducing integrals over the $\nu \times n_i$- and $n_i \times \nu$-dimensional Grassman matrices variables $\overline{\boldsymbol{B}}$, $\boldsymbol{B}$ so that

$$\begin{aligned}&\left[\det\left(\boldsymbol{N}+\boldsymbol{G^i G^{i\dagger}}\right)\right]^{\nu}\\&=\int D_g\boldsymbol{A}D_g\boldsymbol{B}\\&\quad\times\exp\left[\mathrm{Tr}\left\{\overline{\boldsymbol{A}}\boldsymbol{N}\boldsymbol{A}+\sqrt{\frac{\rho_i}{n_i}}\left(\overline{\boldsymbol{B}}\boldsymbol{B}+\overline{\boldsymbol{B}}\boldsymbol{G^{i\dagger}}\boldsymbol{A}-\overline{\boldsymbol{A}}\boldsymbol{G^i}\overline{\boldsymbol{B}}\right)\right\}\right].\end{aligned}\tag{63}$$

Thus, the quantity to be averaged over $\boldsymbol{G^s}$ and $\boldsymbol{G^i}$ to produce $g(\nu)$ in (9) is the product of (61) and (63). The integration over the channel realizations may now be performed. As in the case of (14), we apply Identities 2 and 5, resulting in the following expression:

$$\begin{aligned}g(\nu)=&\int D_g\boldsymbol{A}D_g\boldsymbol{B}\\&\times\exp\left[\mathrm{Tr}\left\{\overline{\boldsymbol{A}}\boldsymbol{N}\boldsymbol{A}+\overline{\boldsymbol{B}}\boldsymbol{B}+\frac{\rho_i}{n_i}\overline{\boldsymbol{B}}\boldsymbol{T}^i\boldsymbol{B}\overline{\boldsymbol{A}}\boldsymbol{R}^i\boldsymbol{A}\right\}\right]\\&\times\int D_c\boldsymbol{X}D_c\boldsymbol{Y}D_c\boldsymbol{Z}\\&\times\exp\left[-\tfrac{1}{2}\mathrm{Tr}\{\boldsymbol{X}^\dagger\boldsymbol{N}\boldsymbol{X}+\boldsymbol{Y}^\dagger\boldsymbol{Y}+\boldsymbol{Z}^\dagger\boldsymbol{Z}\}\right.\\&\left.-\mathrm{Tr}\left\{\frac{\rho_s}{4n_t}\boldsymbol{Y}^\dagger\boldsymbol{T}^s\boldsymbol{Y}\boldsymbol{X}^\dagger\boldsymbol{R}^s\boldsymbol{X}+\frac{\rho_i}{4n_i}\boldsymbol{Z}^\dagger\boldsymbol{T}^i\boldsymbol{Z}\boldsymbol{X}^\dagger\boldsymbol{R}^i\boldsymbol{X}\right\}\right]\\&\times\exp\left[\frac{\rho_i}{2n_i}\mathrm{Tr}\left\{\overline{\boldsymbol{A}}\boldsymbol{R}^i\boldsymbol{X}\boldsymbol{Z}^\dagger\boldsymbol{T}^i\boldsymbol{B}+\overline{\boldsymbol{B}}\boldsymbol{T}^i\boldsymbol{Z}\boldsymbol{X}^\dagger\boldsymbol{R}^i\boldsymbol{A}\right\}\right].\end{aligned}\tag{64}$$

We see that the first exponential involves only $\boldsymbol{A}$, $\boldsymbol{B}$ (and $\overline{\boldsymbol{A}}$, $\overline{\boldsymbol{B}}$), the second only $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$, while the last involves all variables. As in the previous section, we need to introduce auxiliary $\nu\times\nu$ matrices to separate the terms that are quartic in the integration matrices. As in the case of (14) in the previous section, we introduce the $\nu \times \nu$ complex matrices $(\boldsymbol{T}_1, \boldsymbol{R}_1)$, $(\boldsymbol{T}_2, \boldsymbol{R}_2)$ and $(\boldsymbol{T}_3, \boldsymbol{R}_3)$ to separate the $\mathcal{O}(\boldsymbol{Y}^2\boldsymbol{X}^2)$, $\mathcal{O}(\boldsymbol{Z}^2\boldsymbol{X}^2)$, and $\mathcal{O}(\overline{\boldsymbol{A}}\boldsymbol{A}\overline{\boldsymbol{B}}\boldsymbol{B})$ terms in (64), respectively. The substitutions have the form of (15) so we will not repeat them. We note that although the last term involves Grassman variables, they need to be separated in pairs ($\overline{\boldsymbol{A}}\boldsymbol{R}^i\boldsymbol{A}$ and $\overline{\boldsymbol{B}}\boldsymbol{T}^i\boldsymbol{B}$) and, therefore, each term can be treated as a complex number, see (133). However, the quartic terms in the last line of (64) need to be separated into $\overline{\boldsymbol{A}}\boldsymbol{R}^i\boldsymbol{X}$ and $\boldsymbol{Z}^\dagger\boldsymbol{T}^i\boldsymbol{B}$. These $\nu\times\nu$-dimensional terms, being linear in Grassman terms, are anticommuting, see (131). Thus, we need to introduce the $\nu\times\nu$-dimensional matrices $\boldsymbol{T}_4$, $\overline{\boldsymbol{T}}_4$, $\boldsymbol{R}_4$, $\overline{\boldsymbol{R}}_4$ of Grassman variables, in order to separate these terms, using the Grassman version of the Hubbard–Stratonovich transformation of Identity 6. This transformation can be expressed as

$$\begin{aligned}&\exp\left[\frac{\rho_i}{2n_i}\mathrm{Tr}\left\{\overline{\boldsymbol{A}}\boldsymbol{R}^i\boldsymbol{X}\boldsymbol{Z}^\dagger\boldsymbol{T}^i\boldsymbol{B}+\overline{\boldsymbol{B}}\boldsymbol{T}^i\boldsymbol{Z}\boldsymbol{X}^\dagger\boldsymbol{R}^i\boldsymbol{A}\right\}\right]\\&\quad=\int D_g\boldsymbol{T}_4D_g\boldsymbol{R}_4\exp\left[-\mathrm{Tr}\left\{\overline{\boldsymbol{T}}_4\boldsymbol{T}_4+\overline{\boldsymbol{R}}_4\boldsymbol{R}_4\right\}\right.\\&\qquad+\sqrt{\frac{\rho_i}{2n_i}}\mathrm{Tr}\left\{\overline{\boldsymbol{A}}\boldsymbol{R}^i\boldsymbol{X}\boldsymbol{T}_4+\overline{\boldsymbol{R}}_4\boldsymbol{X}^\dagger\boldsymbol{R}^i\boldsymbol{A}\right.\\&\qquad\left.\left.+\overline{\boldsymbol{B}}\boldsymbol{T}^i\boldsymbol{Z}\boldsymbol{R}_4+\overline{\boldsymbol{T}}_4\boldsymbol{Z}^\dagger\boldsymbol{T}^i\boldsymbol{B}\right\}\right].\end{aligned}\tag{65}$$

As in the previous section, the introduction of these $\nu\times\nu$ auxiliary matrices allows us to integrate over the variables $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$, $\boldsymbol{A}$, and $\boldsymbol{B}$. The resulting form of $g(\nu)$ is

$$g(\nu)=\int d\mu(\boldsymbol{T}_1,\boldsymbol{R}_1;\boldsymbol{T}_2,\boldsymbol{R}_2;\boldsymbol{T}_3,\boldsymbol{R}_3)D_g\boldsymbol{T}_4D_g\boldsymbol{R}_4e^{-\mathcal{S}}\tag{66}$$

where now $\mathcal{S}$ is given by

$$\begin{aligned}\mathcal{S}=&\log\det\left\{(\boldsymbol{N}\otimes\boldsymbol{I}_\nu)+\sqrt{\frac{\rho_s}{n_t}}(\boldsymbol{R}^s\otimes\boldsymbol{R}_1)\right.\\&+\sqrt{\frac{\rho_i}{n_i}}(\boldsymbol{R}^i\otimes\boldsymbol{R}_2)+\frac{\rho_i}{n_i}(\boldsymbol{R}^i\otimes\overline{\boldsymbol{R}}_4)\\&\left.\times\left[(\boldsymbol{N}\otimes\boldsymbol{I}_\nu)+\sqrt{\frac{\rho_i}{n_i}}(\boldsymbol{R}^i\otimes\boldsymbol{R}_3)\right]^{-1}(\boldsymbol{R}^i\otimes\boldsymbol{T}_4)\right\}\\&+\log\det\left\{(\boldsymbol{I}_{n_i}\otimes\boldsymbol{I}_\nu)+\sqrt{\frac{\rho_i}{n_i}}(\boldsymbol{T}^i\otimes\boldsymbol{T}_2)+\frac{\rho_i}{n_i}(\boldsymbol{T}^i\otimes\overline{\boldsymbol{T}}_4)\right.\\&\left.\times\left[(\boldsymbol{I}_{n_t}\otimes\boldsymbol{I}_\nu)+\sqrt{\frac{\rho_i}{n_i}}(\boldsymbol{T}^i\otimes\boldsymbol{T}_3)\right]^{-1}(\boldsymbol{T}^i\otimes\boldsymbol{R}_4)\right\}\\&+\log\det\left\{(\boldsymbol{I}_{n_t}\otimes\boldsymbol{I}_\nu)+\sqrt{\frac{\rho_s}{n_t}}(\boldsymbol{Q}^{1/2}\boldsymbol{T}^s\boldsymbol{Q}^{1/2}\otimes\overline{\boldsymbol{T}}_1)\right\}\\&-\mathrm{Tr}\{\boldsymbol{T}_1\boldsymbol{R}_1+\boldsymbol{T}_2\boldsymbol{R}_2\}\\&-\log\det\left\{(\boldsymbol{I}_{n_i}\otimes\boldsymbol{I}_\nu)+\sqrt{\frac{\rho_i}{n_i}}(\boldsymbol{T}^i\otimes\boldsymbol{T}_3)\right\}\\&-\log\det\left\{(\boldsymbol{N}\otimes\boldsymbol{I}_\nu)+\sqrt{\frac{\rho_i}{n_i}}(\boldsymbol{R}^i\otimes\boldsymbol{R}_3)\right\}\\&+\mathrm{Tr}\{\boldsymbol{T}_3\boldsymbol{R}_3+\overline{\boldsymbol{T}}_4\boldsymbol{T}_4+\overline{\boldsymbol{R}}_4\boldsymbol{R}_4\}.\end{aligned}\tag{67}$$

Note that the first matrix of each outer product pair has indexes representing transmitters and receivers, whereas the second matrix is always of dimension $\nu\times\nu$. Also note that in the above equation we have reintroduced the signal covariance matrix $\boldsymbol{Q}$.

As in the preceding section, we now look for replica-invariant saddle point solutions of $\mathcal{S}$, in accordance with Assumption 3. As in (18), we take into account the assumed form of the saddle point values of $\boldsymbol{T}_k$, $\boldsymbol{R}_k$ for $k=1-4$ by expressing them as

$$\begin{aligned}\boldsymbol{T}_1&=t_1\sqrt{n_t}\,\boldsymbol{I}_\nu+\boldsymbol{\delta T}_1 & \boldsymbol{R}_1&=r_1\sqrt{n_t}\,\boldsymbol{I}_\nu+\boldsymbol{\delta R}_1\\\boldsymbol{T}_2&=t_2\sqrt{n_i}\,\boldsymbol{I}_\nu+\boldsymbol{\delta T}_2 & \boldsymbol{R}_2&=r_2\sqrt{n_i}\,\boldsymbol{I}_\nu+\boldsymbol{\delta R}_2\\\boldsymbol{T}_3&=t_3\sqrt{n_i}\,\boldsymbol{I}_\nu+\boldsymbol{\delta T}_3 & \boldsymbol{R}_3&=r_3\sqrt{n_i}\,\boldsymbol{I}_\nu+\boldsymbol{\delta R}_3\end{aligned}\tag{68}$$

IEEE Proof

$$\begin{aligned} \boldsymbol{T}_4 &= \boldsymbol{\delta T}_4 & \boldsymbol{R}_4 &= \boldsymbol{\delta R}_4 \\ \overline{\boldsymbol{T}}_4 &= \overline{\boldsymbol{\delta T}}_4 & \overline{\boldsymbol{R}}_4 &= \boldsymbol{\delta}\overline{\boldsymbol{R}}_4 \end{aligned} \tag{69}$$

where the $\boldsymbol{\delta T}_k$, $\boldsymbol{\delta R}_k$, for $k = 1-3$ are $\nu \times \nu$ complex matrices representing, as in (18), deviations around the saddle point. It is important to note that, while $\boldsymbol{T}_k$, $\boldsymbol{R}_k$ for $k = 1-3$ are assumed to be nonzero and proportional to $\boldsymbol{I}_\nu$ at the saddle point, the Grassman matrices $\boldsymbol{R}_4$, $\boldsymbol{T}_4$ have been chosen to vanish at the saddle point, with $\boldsymbol{\delta T}_4$, $\boldsymbol{\delta R}_4$, etc., representing also deviations around the saddle point in the Grassman space. The reason is that at the saddle point we expect $\mathcal{S}$ to be real, while for nonzero Grassman $\overline{\boldsymbol{T}}_4$, etc., this will not be true (Grassman variables and their products are not real numbers).

The next step is to expand $\mathcal{S}$ into powers of $\boldsymbol{\delta T}$, $\boldsymbol{\delta R}$, resulting in an expansion in the small parameter $n^{-1/2}$ of the type of (19). Then, setting the coefficients of all $\boldsymbol{\delta T}$, $\boldsymbol{\delta R}$ in $\mathcal{S}_1$ to zero will produce the generalization of the saddle point equations of (27) and (28). The resulting equations can be expressed as

$$r_1 = \frac{\sqrt{\rho_s}}{n_t} \operatorname{Tr}\left\{ \boldsymbol{Q}\boldsymbol{T}^s [\boldsymbol{I}_{n_t} + \sqrt{\rho_s}\, t_1 \boldsymbol{Q}\boldsymbol{T}^s]^{-1} \right\} \tag{70}$$

$$t_1 = \frac{\sqrt{\rho_s}}{n_t} \operatorname{Tr}\left\{ \boldsymbol{R}^s \left[\boldsymbol{N} + \sqrt{\rho_s}\, \boldsymbol{R}^s r_1 + \sqrt{\rho_i}\, \boldsymbol{R}^i r_2 \right]^{-1} \right\} \tag{71}$$

$$r_2 = \frac{\sqrt{\rho_i}}{n_i} \operatorname{Tr}\left\{ \boldsymbol{T}^i \left[\boldsymbol{I}_{n_i} + \sqrt{\rho_i}\, t_2 \boldsymbol{T}^i \right]^{-1} \right\} \tag{72}$$

$$t_2 = \frac{\sqrt{\rho_i}}{n_i} \operatorname{Tr}\left\{ \boldsymbol{R}^i \left[\boldsymbol{N} + \sqrt{\rho_s}\, \boldsymbol{R}^s r_1 + \sqrt{\rho_i}\, \boldsymbol{R}^i r_2 \right]^{-1} \right\} \tag{73}$$

$$r_3 = \frac{\sqrt{\rho_i}}{n_i} \operatorname{Tr}\left\{ \boldsymbol{T}^i \left[\boldsymbol{I}_{n_i} + \sqrt{\rho_i}\, t_3 \boldsymbol{T}^i \right]^{-1} \right\} \tag{74}$$

$$t_3 = \frac{\sqrt{\rho_i}}{n_i} \operatorname{Tr}\left\{ \boldsymbol{R}^i \left[\boldsymbol{N} + \sqrt{\rho_i}\, R^i r_3 \right]^{-1} \right\}. \tag{75}$$

Note again that in (70) we have again combined the two $\boldsymbol{Q}^{1/2}$ factors into $\boldsymbol{Q}$. Despite the apparent complexity of the above six equations, we point out several simplifying properties. First, they are split into a set of four (70)–(73) and a set of two (74)–(75) coupled equations. Second, (72) and (74) can be written in terms of only the eigenvalues of $\boldsymbol{T}^i$, while (70) and (75) can be expressed in terms of only the eigenvalues of $\boldsymbol{Q}\boldsymbol{T}^s$ and $\boldsymbol{N}^{-1}\boldsymbol{R}^i$, respectively, thus simplifying their solution. Also, the convexity properties of these equations make solution by iteration simple and efficient. In addition, as we shall see, in many useful cases additional simplifications may be made.

1) Mean Mutual Information: At the saddle point, $\mathcal{S}$ is equal to

$$\mathcal{S}_0 = \nu(\Gamma_s - \Gamma_i) \tag{76}$$

where

$$\begin{aligned} \Gamma_s = {} & \log\det(\boldsymbol{I}_{n_t} + \sqrt{\rho_s}\, t_1 \boldsymbol{Q}\boldsymbol{T}^s) \\ & + \log\det\left(\boldsymbol{I}_{n_i} + \sqrt{\rho_i}\, t_2 \boldsymbol{T}^i \right) \\ & + \log\det\left(\boldsymbol{N} + \sqrt{\rho_s}\, r_1 \boldsymbol{R}^s + \sqrt{\rho_i}\, r_2 \boldsymbol{R}^i \right) \\ & - (n_t t_1 r_1 + n_i t_2 r_2) \end{aligned} \tag{77}$$

$$\begin{aligned} \Gamma_i = {} & \log\det\left(\boldsymbol{I}_{n_i} + \sqrt{\rho_i}\, t_3 \boldsymbol{T}^i \right) \\ & + \log\det\left(\boldsymbol{N} + \sqrt{\rho_i}\, r_3 \boldsymbol{R}^i \right) - n_i t_3 r_3. \end{aligned} \tag{78}$$

As in the previous section, the leading term in the expansion of $n^{-1/2}$ is proportional to ν. Thus, it contributes only to the first moment of the mutual information, namely, the mean. Thus, we see that to $\mathcal{O}(n)$ the mean mutual information is $\Gamma_s - \Gamma_i$. These terms have a simple interpretation: Γ_s is, to leading order, the average of h_s in (5), while Γ_i is the average of h_i. Since also the evaluation of the corresponding parameters t_1, r_1, etc., in (71)–(74), decouples to two independent sets, we see that to calculate the means of h_s and h_i to leading order, we can analyze them separately. This will not be the case for higher order terms, because of the presence of $\boldsymbol{G}^{\boldsymbol{i}}$ in both h_s and h_i, which will couple them.

2) Variance of Mutual Information: To calculate $\mathcal{S}$ to the next order, we need to integrate out the quadratic fluctuations around the saddle point appearing in $\mathcal{S}_2$, which can be expressed as (expanding $\mathcal{S}$ in (67) to second order)

$$\begin{aligned} \mathcal{S}_2 = \sum_{a,\,b=1}^{\nu} \big[& (\boldsymbol{x}^1_{ab})^T \boldsymbol{V}_1 \left(\boldsymbol{x}^1_{ba} \right) + (\boldsymbol{x}^2_{ab})^T \boldsymbol{V}_2 \left(\boldsymbol{x}^2_{ba} \right) \\ & + (\overline{\boldsymbol{x}}^3_{ab})^T \boldsymbol{V}_3 \left(\boldsymbol{x}^3_{ba} \right) \big] \end{aligned} \tag{79}$$

where, for every fixed pair of indexes a, b, $\boldsymbol{x}^1_{ab}$, $\boldsymbol{x}^2_{ab}$, $\overline{\boldsymbol{x}}^3_{ab}$, and $\boldsymbol{x}^3_{ab}$ are vectors given by

$$\begin{aligned} \boldsymbol{x}^1_{ab} &= [\delta T_{2,\,ab} \delta R_{2,\,ab} \delta T_{1,\,ab} \delta R_{1,\,ab}]^T \\ \boldsymbol{x}^2_{ab} &= [\delta T_{3,\,ab} \delta R_{3,\,ab}]^T \\ \overline{\boldsymbol{x}}^3_{ab} &= \left[\overline{\delta R}_{4,\,ab} \overline{\delta T}_{4,\,ab} \right]^T \\ \boldsymbol{x}^3_{ab} &= [\delta R_{4,\,ab} \delta T_{4,\,ab}]^T \end{aligned} \tag{80}$$

and where $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, and $\boldsymbol{V}_3$ are given by

$$\boldsymbol{V}_1 = \begin{pmatrix} -M^{i1}_{t,2} & -1 & 0 & 0 \\ -1 & -M^{i1}_{r,2} & 0 & -M^{is}_{r,2} \\ 0 & 0 & -M^{s1}_{t,2} & -1 \\ 0 & -M^{is}_{r,2} & -1 & -M^{s1}_{r,2} \end{pmatrix} \tag{81}$$

$$\boldsymbol{V}_2 = \begin{pmatrix} -M^{i2}_{t,2} & -1 \\ -1 & -M^{i2}_{r,2} \end{pmatrix} \tag{82}$$

$$\boldsymbol{V}_3 = \begin{pmatrix} 1 & M^{i3}_{r,2} \\ M^{i3}_{t,2} & 1 \end{pmatrix}. \tag{83}$$

The elements of the above matrices are given by

$$M^{i1}_{t,2} = \frac{\rho_i}{n_i} \operatorname{Tr}\left\{ \left(\boldsymbol{T}^i \left[\boldsymbol{I}_{n_i} + \sqrt{\rho_i}\, \boldsymbol{T}^i t_2 \right]^{-1} \right)^2 \right\} \tag{84}$$

$$M^{i1}_{r,2} = \frac{\rho_i}{n_i} \operatorname{Tr}\left\{ \left(\boldsymbol{R}^i \left[\boldsymbol{N} + \sqrt{\rho_i}\, \boldsymbol{R}^i r_2 + \sqrt{\rho_s}\, \boldsymbol{R}^s r_1 \right]^{-1} \right)^2 \right\} \tag{85}$$

$$M^{s1}_{t,2} = \frac{\rho_s}{n_t} \operatorname{Tr}\left\{ \left(\boldsymbol{Q}\boldsymbol{T}^s [\boldsymbol{I}_{n_t} + \sqrt{\rho_s}\, \boldsymbol{Q}\boldsymbol{T}^s t_1]^{-1} \right)^2 \right\} \tag{86}$$

$$M^{s1}_{r,2} = \frac{\rho_s}{n_t} \operatorname{Tr}\left\{ \left(\boldsymbol{R}^s \left[\boldsymbol{N} + \sqrt{\rho_i}\, \boldsymbol{R}^i r_2 + \sqrt{\rho_s}\, \boldsymbol{R}^s r_1 \right]^{-1} \right)^2 \right\} \tag{87}$$

$$\begin{aligned} M^{is}_{r,2} = \sqrt{\frac{\rho_i \rho_s}{n_t n_i}} \operatorname{Tr}\Big\{ & \boldsymbol{R}^s \left[\boldsymbol{N} + \sqrt{\rho_i}\, \boldsymbol{R}^i r_2 + \sqrt{\rho_s}\, \boldsymbol{R}^s r_1 \right]^{-1} \\ & \times \boldsymbol{R}^i \left[\boldsymbol{N} + \sqrt{\rho_i}\, \boldsymbol{R}^i r_2 + \sqrt{\rho_s}\, \boldsymbol{R}^s r_1 \right]^{-1} \Big\} \end{aligned} \tag{88}$$

$$M^{i2}_{t,2} = \frac{\rho_i}{n_i} \operatorname{Tr}\left\{ \left(\boldsymbol{T}^i \left[\boldsymbol{I}_{n_i} + \sqrt{\rho_i} \boldsymbol{T}^i t_3 \right]^{-1} \right)^2 \right\} \tag{89}$$

IEEE Proof

$$M_{r,2}^{i2}=\frac{\rho_i}{n_i}\,\mathrm{Tr}\left\{\left(\boldsymbol{R}^i\left[\boldsymbol{N}+\sqrt{\rho_i}\boldsymbol{R}^i r_3\right]^{-1}\right)^2\right\} \tag{90}$$

$$M_{t,2}^{i3}=\frac{\rho_i}{n_i}\,\mathrm{Tr}\left\{\left(\boldsymbol{T}^i\right)^2\left[\left(\boldsymbol{I}_{n_i}+\sqrt{\rho_i}\boldsymbol{T}^i t_3\right)\left(\boldsymbol{I}_{n_i}+\sqrt{\rho_i}\boldsymbol{T}^i t_3\right)\right]^{-1}\right\} \tag{91}$$

$$M_{r,2}^{i3}=\frac{\rho_i}{n_i}\,\mathrm{Tr}\left\{\boldsymbol{R}^i\left[\boldsymbol{N}+\sqrt{\rho_i}\boldsymbol{R}^i r_2+\sqrt{\rho_s}\boldsymbol{R}^s r_1\right]^{-1}\times\boldsymbol{R}^i\left[\boldsymbol{N}+\sqrt{\rho_i}\boldsymbol{R}^i r_3\right]^{-1}\right\}. \tag{92}$$

As in the absence of interferers, simple power counting shows that all the above elements are formally of order $\mathcal{O}(1)$.

We see that $\mathcal{S}_2$ in (79) is decoupled into three separate terms, with covariance matrices $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, and $\boldsymbol{V}_3$. The underlying reason for this decoupling is that these three terms correspond to $\langle h_s^2\rangle-\langle h_s\rangle^2$, $\langle h_s h_i\rangle-\langle h_s\rangle\langle h_i\rangle$, and $\langle h_i^2\rangle-\langle h_i\rangle^2$, respectively. This can be supported by analyzing the t's and r's that appear in their matrix elements. Indeed, $\boldsymbol{V}_1$ contains only t_k, r_k, for $k=1-2$, which only appear in Γ_s; $\boldsymbol{V}_3$ contains only t_3, r_3, which appear only in Γ_i; while $\boldsymbol{V}_2$ has t's and r's from both sets.

Following the methodology described in Section II-A3, for every a, b we rotate each of the vectors $\boldsymbol{x}_{ab}^1$, $\boldsymbol{x}_{ab}^2$ to the eigenvectors of the matrices $\bar{\boldsymbol{V}}_1$, $\bar{\boldsymbol{V}}_2$, and then integrate each of the eigenvectors along its corresponding steepest descent direction in the complex plane. We can similarly perform the Gaussian integral over the Grassman vectors $\boldsymbol{x}_{ab}^3$, $\bar{\boldsymbol{x}}_{ab}^3$ using Identity 4. The resulting form of $g(\nu)$ is

$$\begin{aligned}g(\nu)&=e^{-\mathcal{S}_0}\left[\left|\det\boldsymbol{V}_1\det\boldsymbol{V}_2(\det\boldsymbol{V}_3)^{-2}\right|^{-\nu^2/2}+\cdots\right]\\&=\exp\left[-\nu(\Gamma_s-\Gamma_i)-\frac{\nu^2}{2}\left(\log\left|\det\boldsymbol{V}_1\right|+\log\left|\det\bar{\boldsymbol{V}}_2\right|-2\log\left|\det\bar{\boldsymbol{V}}_3\right|\right)+\cdots\right]\end{aligned} \tag{93}$$

where the dots correspond to terms obtained by taking into account higher order terms, such as $\mathcal{S}_3$, $\mathcal{S}_4$, etc.

3) Summary of Results for the General Case: To summarize this section, we have derived the following results.

The mean mutual information is equal to

$$\langle I\rangle=\Gamma_s-\Gamma_i+\mathcal{O}(1/n^2) \tag{94}$$

where Γ_s, Γ_i are given in (77) and (78). As in the case of no interferers, we also find that the next term linear in ν appears to $\mathcal{O}(1/n^2)$.

In addition, the variance in the mutual information is given by

$$\begin{aligned}\mathcal{C}_2&=Var(I)=\langle I^2\rangle-\langle I\rangle^2\\&=-\log\left|\det\mathcal{V}_1\right|-\log\left|\det\mathcal{V}_2\right|+2\log\left|\det\mathcal{V}_3\right|+\mathcal{O}(1/n^2).\end{aligned} \tag{95}$$

By analyzing the next-order terms $\mathcal{S}_3$, $\mathcal{S}_4$, etc., we find that the next correction to $\mathcal{C}_2$ is of order $1/n^2$.

Similarly, higher moments can be calculated in this case as well with only additional algebraic effort.

III. Capacity-Achieving Signal Covariance $\boldsymbol{Q}$

As discussed in Section I-B, instead of instantaneous channel information, the transmitter has statistical information for the channel $\boldsymbol{G^s}$, $\boldsymbol{G^i}$, namely, only $\boldsymbol{T}^s$, $\boldsymbol{R}^s$, $\boldsymbol{T}^i$ and $\boldsymbol{R}^i$, and ρ_s, ρ_i, $\boldsymbol{N}$ are known. Based on this information, the signal covariance can be optimized to maximize a particular metric of the mutual information distribution. Here we describe the method to optimize $\boldsymbol{Q}$, in order to maximize the average mutual information, keeping only the $\mathcal{O}(n)$ term in (94), i.e., to find $\max_{\boldsymbol{Q}}\langle I\rangle$ in the large antenna limit. This method was introduced for MIMO systems without interference in [10] and is also discussed in [12].

We start by observing that $\langle I\rangle$ depends on $\boldsymbol{Q}$ only through the first term in (77). Expressing the determinant of this term in the eigenbasis of $\boldsymbol{T}^s$, it can be written as

$$\det\left(\boldsymbol{1}+\sqrt{\rho_s}\,t_1\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{T}}^s\right) \tag{96}$$

where $\tilde{\boldsymbol{T}}^s$ is a diagonal matrix with the eigenvalues of $\boldsymbol{T}^s$, given by T_k^s for $k=1,\ldots,n_t$, on the diagonals and $\tilde{\boldsymbol{Q}}$ the original matrix $\boldsymbol{Q}$ expressed in the eigenbasis of $\boldsymbol{T}^s$. We note that for any nonnegative definite matrix $\boldsymbol{A}$, we have $\det(\boldsymbol{A})\le\prod_k A_{kk}$, where A_{kk} are the diagonal elements [2]. Applying this inequality to (96) we get

$$\det\left(\boldsymbol{1}+\sqrt{\rho_s}\,t_1\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{T}}^s\right)\le\prod_{k=1}^{n_t}\left(1+\sqrt{\rho_s}\,t_1\tilde{Q}_{kk}T_k^s\right) \tag{97}$$

with equality when $\tilde{\boldsymbol{Q}}$ is diagonal. Thus, we conclude that the $\boldsymbol{Q}$ maximizing $\langle I\rangle$ is simultaneously diagonalizable with $\bar{\boldsymbol{T}}^s$. This result has been shown in [26] to hold in general in the absence of interference for arbitrary antenna numbers (but can be generalized to hold in the presence of interference). Here we have shown this to be true in the presence of interference, at least within the large antenna number approximation.

Once the optimal eigenbasis of $\boldsymbol{Q}$ has been determined to be the same as $\bar{\boldsymbol{T}}^s$, one needs to find its optimal eigenvalues q_k for $k=1,\ldots,n_t$. $\langle I\rangle$ has to be optimized subject to the power constraint $\mathrm{Tr}\boldsymbol{Q}=n_t$. This constraint is enforced by adding a Lagrange multiplier to $\langle I\rangle$, i.e.,

$$\langle I\rangle\to\langle I\rangle-\Lambda\left(\sum_{k=1}^{n_t}q_k-n_t\right). \tag{98}$$

Incorporating the Lagrange multiplier to (77) and maximizing, it is easy to see that the optimal eigenvalues of $\boldsymbol{Q}$ are then given by

$$q_k=\left[\frac{1}{\Lambda}-\frac{1}{\sqrt{\rho_s}\,T_k^s t_1}\right]_+ \tag{99}$$

where $[x]_+=\{x+\mathrm{sgn}(x)\}/2$. Here, $\Lambda>0$ is determined by imposing the power constraint

$$\mathrm{Tr}\{\boldsymbol{Q}\}=\sum_{k=1}^{n_t}q_k=n_t. \tag{100}$$

Essentially, the optimization of $\boldsymbol{Q}$ amounts to water-filling over the modes of $\boldsymbol{T}^s$ rather than those of the instantaneous channel $\boldsymbol{G^s}^\dagger\boldsymbol{G^s}$ itself. Thus, the "statistically water-filled" capacity of the link is found by solving for five unknowns

IEEE Proof

$(t_1, r_1, t_2, r_2, \Lambda)$ from (70)–(73) and (100) with (99) and using them to evaluate Γ_s in (77). In addition, to evaluate Γ_i in (78), we need to evaluate t_3, r_3 from (74) and (75).

IV. Specific Cases

We now apply the equations of the previous section to a number of representative situations with aim to describe the basic characteristics and dependencies of the MIMO capacity for correlated channels and interferers. We will be keeping only the leading terms of calculated quantities in the $1/n$ expansion. Thus, for example, we will take $\langle I\rangle = \Gamma_s - \Gamma_i$ using (77), (78) and similarly we will neglect the $\mathcal{O}(1/n^2)$ terms in (74) and (75).

In the cases with spatially correlated antennas, we will use the following simple model for the dependence of the correlation coefficient T_{ab} on the antenna separation and angle spread. For concreteness, we assume that the antennas form a uniform linear ideal antenna array with $d_\lambda = d_{\min}/\lambda$ the nearest neighbor antenna spacing in wavelengths and we assume a Gaussian power azimuth spectrum (with two-dimensional propagation), i.e., the average incoming power at the antenna array is $P(\theta) \propto \exp[-(\theta/\delta)^2/2]$, [23], [27], where δ is the angle spread in degrees measured from the vertical to the array. This results in a $\boldsymbol{T}$ matrix with elements

$$T_{ab} = \int_{-180}^{180} \frac{d\phi}{\sqrt{2\pi\delta^2}}\, e^{2\pi i(a-b)d_\lambda \sin(\phi\pi/180) - \phi^2/(2\delta^2)} \tag{101}$$

with $a, b = 1\cdots n_t$ being the index of transmitting antennas. For simplicity, we assume no correlations at the receiver ($\boldsymbol{R} = \boldsymbol{I}_{n_r}$). This situation corresponds to a receiving mobile array deep inside the clutter and a transmitting base-station array with correlations due to finite angle spread.

A. Capacity for i.i.d. Matrices

It is instructive to start by verifying that $\langle I\rangle$ reduces to known results in the previously studied cases of i.i.d. channels. In the absence of interference (94) reduces to

$$\langle I\rangle = n_t \log\left(1+\sqrt{\rho_s}\, t_1\right) + n_r \log\left(1+\sqrt{\rho_s}\, r_1\right) - n_t t_1 r_1 \tag{102}$$

with the extremizing values of r_1 and t_1 from (70), (71) given by

$$r_1 = \frac{\sqrt{\rho_s}}{1+\sqrt{\rho_s} t_1} \tag{103}$$

$$t_1 = \frac{n_r\sqrt{\rho_s}}{n_t\left(1+\sqrt{\rho_s} r_1\right)}. \tag{104}$$

Solving for positive r_1, t_1 we get

$$\langle I\rangle = (n_t - n_r)\log u + n_r \log(\rho_s + u) - n_t\left(1-u^{-1}\right) \tag{105}$$

with

$$u = \frac{1}{2}\left[1+\rho_s\left(\frac{n_r}{n_t}-1\right) + \sqrt{\left(1+\rho_s\left(\frac{n_r}{n_t}-1\right)\right)^2 + 4\rho_s}\right] \tag{106}$$

which is identical to the result obtained by Verdú and Shamai [5].

B. Throughput Gain by Optimizing $\boldsymbol{Q}$ for Spatially Correlated Transmission $\boldsymbol{T}^s \neq \boldsymbol{I}$

Next, we analyze the benefits of statistical water-filling, i.e., knowing $\boldsymbol{T}^s$ at the transmitter. For simplicity, we neglect interferers and set $\rho_i = 0$. In addition, we assume that $\boldsymbol{R}^s = \boldsymbol{N} = \boldsymbol{I}_{n_r}$. This situation corresponds to a receiving mobile array deep inside the clutter and a transmitting base-station array with correlations due to finite angle spread. In this case, (94) becomes

$$\langle I\rangle = \sum_{i=1}^{n_t} \log\left(1+\sqrt{\rho_s}\, t_1 q_i T_i^s\right) + n_r \log\left(1+\sqrt{\rho_s}\, r_1\right) - n_t t_1 r_1. \tag{107}$$

To get the capacity we need to optimize over r_1 (70), t_1 (71), and q_i (99). Our optimization equations can be reduced to the following. We solve for r_1 numerically

$$\frac{n_r r_1\sqrt{\rho_s}}{1+\sqrt{\rho_s}} = \sum_{i=1}^{n_t}\left[1 - \frac{r_1}{\sqrt{\rho_s} T_i^s}\right]_+ \tag{108}$$

and then substitute r_1 into

$$t_1 = \frac{n_r\sqrt{\rho_s}}{n_t\left(1+\sqrt{\rho_s} r_1\right)} \tag{109}$$

$$q_i = \left[\frac{1}{t_1 r_1} - \frac{1}{t_1\sqrt{\rho_s} T_i^s}\right]_+. \tag{110}$$

Subsequently, r_1, t_1, and q_i are substituted into (107). Interestingly, applying these equations to find the optimality of beamforming (i.e., the stability of the solution with one nonzero q_i) produces extremely accurate results when compared with the exact criterion for the optimality of beamforming [28].

We compare this result to the $\boldsymbol{Q} = \boldsymbol{I}_{n_t}$ or $q_i \equiv 1$ case, where $\boldsymbol{Q}$ is not optimized and corresponds to independent transmission . Here one needs to use (71) and (70). The optimal values for t_1, r_1 are found by solving for t_1 in the following equation:

$$\frac{t_1}{\sqrt{\rho_s}} = \frac{n_r}{n_t} - 1 + \frac{1}{n_t}\sum_{i=1}^{n_t}\frac{1}{1+\sqrt{\rho_s} t_1 T_i^s} \tag{111}$$

and then solving for r_1 in (109). We note that (111) clearly has a single real solution, which can be found numerically very efficiently due to the monotonicity of its argument t_1. We now wish to quantify the throughput gain due to optimization of $\boldsymbol{Q}$. This gain depends not only on the amount of antenna correlation but also on the SNR ρ_s. For large ρ_s, the relative gain is small, since the channel of each transmitting antenna has enough power to successfully transmit without signal power reallocation. In contrast, for small ρ_s, it is advantageous to allocate more power to the eigenmodes of $\boldsymbol{T}^s$ with the larger eigenvalues and, therefore, the case with optimized $\boldsymbol{Q}$ will be markedly higher than the case of independent transmission ($\boldsymbol{Q} = \boldsymbol{I}$) for modest correlations. In fact, for low enough ρ_s, the optimized $\boldsymbol{Q}$ capacity is even higher than the capacity with uncorrelated (i.i.d.) channels. This effect is exemplified by comparing Figs. 1 and 2, where the open and closed loop capacities are plotted as a function of nearest neighbor antenna separation of a uniform linear transmitting array. The analytic solution is also compared to simulated results using 10^6 realizations of the channel matrices. The agreement is extremely good.

IEEE Proof

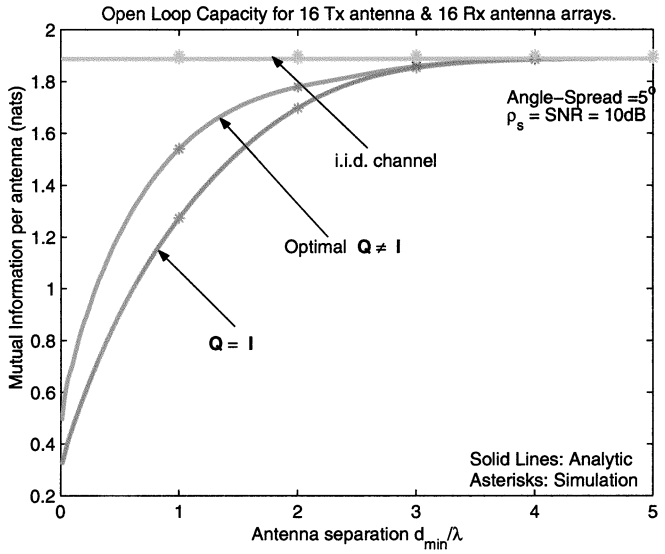


Fig. 1. Plot of mutual information per antenna for a 16-transmitting-antenna and a 16-receiving-antenna array pair. The receiving array has no antenna correlations ($\boldsymbol{R}^s = \boldsymbol{I}_{16}$). The transmitting array is a uniformly spaced array with neighboring antennas spaced by $d_{\min}$. The spatial correlations between antennas are calculated from (101) with angle-spread $\delta = 5°$. The lower curve depicts the mutual information per antenna with all antennas transmitting independently with equal power. The middle curve is the capacity with optimized transmission based on knowledge of the channel covariance: the transmitter here adjusts to the *typical* channel. The straight line is the capacity per antenna in the absence of any antenna correlations ($\boldsymbol{T}^s = \boldsymbol{R}^s = \boldsymbol{I}_{16}$). The asterisks represent the values of the simulated average capacities. We see that the agreement with analytic results are very good.

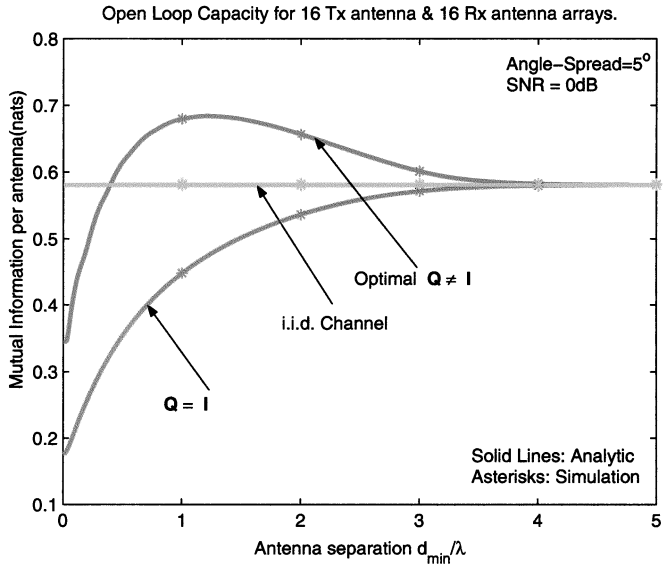


Fig. 2. Plot of mutual information per antenna for the same 16×16 antenna array pair as in Fig. 1, but with SNR $= 0$ dB. The relative gain in throughput from optimizing transmission from channel covariance knowledge is higher at lower SNRs. In fact, as can be seen in the figure, at low SNRs the transmitter can do better than the uncorrelated case, for a region of antenna separations. This is the effect of beam-steering. Again here the asterisks represent simulated capacity results.

C. Consequences of Interference Channel Information $\boldsymbol{G}^i$ at the Receiver

In this subsection, we pose the following question: What is the relative gain of knowing the interfering channel compared to just knowing its spatial covariance at the receiver? We treat the case of uncorrelated channels setting all $\boldsymbol{T}^i, \boldsymbol{R}^i, \boldsymbol{T}^s, \boldsymbol{R}^s, \boldsymbol{N}$ equal to the identity. In this case, $\beta = \rho_i/(1+\rho_i)$ determines the fraction of the noise that has channel known to the receiver. As we see in Fig. 3, when most of the noise is due to interference with known channel at the receiver, i.e., when $\beta \to 1$, the capacity depends strongly on the number of interfering antennas. If $n_i < \min(n_t, n_r)$, the capacity diverges. The reason is that the receiver can successfully null out all interferers, resulting in $\min(n_t, n_r) - n_i$ noiseless bit streams.

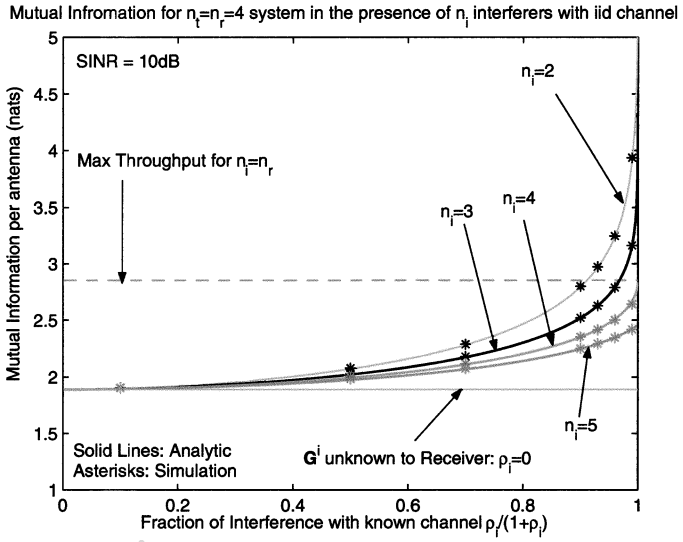


Fig. 3. Plot of mutual information per antenna for $\boldsymbol{G}^s$ and $\boldsymbol{G}^i$ matrices with $\boldsymbol{R}^i = \boldsymbol{R}^s = \boldsymbol{T}^s = \boldsymbol{I}_4$ and $\boldsymbol{T}^i = \boldsymbol{I}_{n_i}$, as a function of $\beta = \rho_i/(1+\rho_i)$, the fraction of interference power with channel known to the receiver to the total interference-plus-noise power. For $n_i < n_t$, the capacity diverges for $\beta \to 1$. For $n_i = n_t = n_r$, the capacity in this limit becomes $2\log(1+\sqrt{\text{SINR}})$. For $n_i \geq n_t = n_r$, it is finite. Again, we see good agreement with simulations (represented by asterisks).

D. Gaussian Approximation of the Distribution of Mutual Information: No Interference

We have seen that for large antenna numbers, the distribution of the mutual information is very peaked and all but its mean and variance tend to zero with varying powers of $1/n$. For few antennas this is generally not true. In this case, at least in principle, we cannot neglect higher moments as being small. However, we find that in most cases, the variance together with the mean describe the distribution very accurately. We demonstrate this feature in this and the next subsection, by analyzing two specific examples and by comparing numerically the Gaussian distribution $\mathcal{N}(\langle I\rangle, Var(I))$ calculated using (94) and (95) with the simulated distribution resulting from the generation of a large number of random matrix realizations.

We start by considering the case of a correlated transmitting array and an uncorrelated receiver array, both with three elements. The correlations of the transmitter are the same as in Section IV-B and Figs. 1 and 2 for antenna spacing $d_\lambda = 1$. Using (94) and (95), we calculate the average and variance of the mutual information I for $\boldsymbol{Q}$ set to either the identity matrix (the case of independent transmission) or the optimal $\boldsymbol{Q}$, which maximizes the average mutual information $\langle I\rangle$ in (94).

IEEE Proof

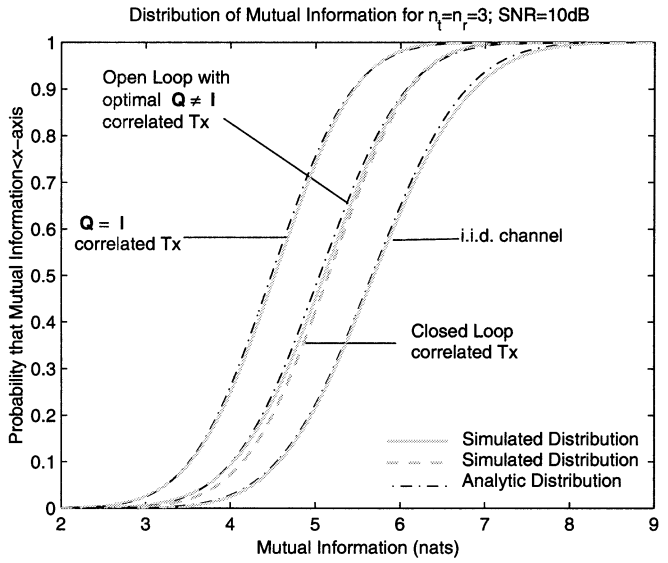


Fig. 4. Cumulative distribution function (CDF) of mutual information for three-antenna transmitting and receiving arrays in the absence of interferers ($\rho_i = 0$). Each closely spaced pair of curves corresponds to a simulated mutual information distribution (solid curves) and a normal distribution (dash-dotted curves) with mean and variance calculated from (94) and (95). The agreement between the analytic estimate and the simulated distributions is impressive. This is especially so given that $n_t = n_r = 3$ while the analytic method described in this paper is presumably valid for large antenna numbers. The three pairs of curves plotted correspond to the three curves plotted in Fig. 1, with SNR = 10 dB, $d_\lambda = 1$, and $\delta = 5°$. The $\boldsymbol{Q} = \boldsymbol{I}$ curve corresponds to independent transmission and the open loop curve has the optimal $\boldsymbol{Q}$, which maximizes the mean mutual information $\langle I \rangle$ based on channel covariance information. As an example, the analytically calculated skewness from (60) for the optimal $\boldsymbol{Q}$, $\boldsymbol{Q} = \boldsymbol{I}$, and i.i.d. channel cases for SNR = 10 dB are $Sk = -0.15$, $Sk = -0.12$, and $Sk = 0.05$, respectively, which are all small. For comparison, the simulated skewness for the i.i.d. case is $Sk_{sim} = 0.074 \pm 0.007$. The dashed curve represents the closed-loop capacity distribution (simulated but not calculated analytically), where the instantaneous correlated channel matrix is fed back to the transmitter. The mean throughput of the two closed-loop schemes are within 5% of each other.

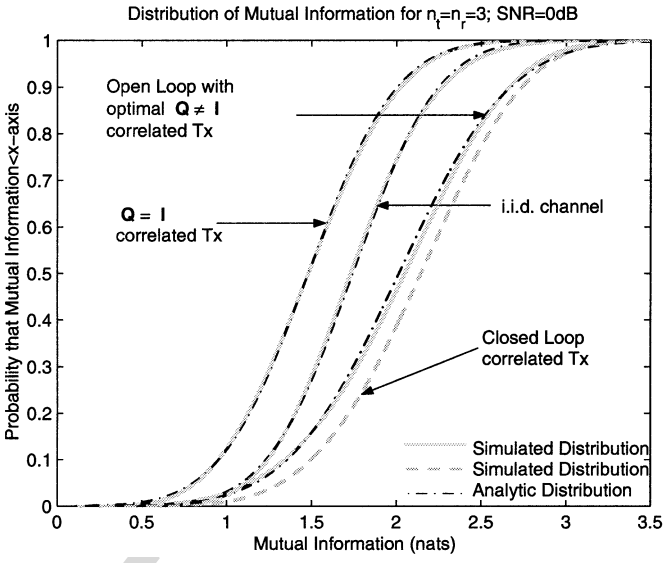


Fig. 5. CDF of mutual information for three-antenna transmitting and receiving arrays in the absence of interferers ($\rho_i = 0$) with SNR = 0 dB. The significance of these curves is identical to the previous figure. The main difference lies in that the open-loop capacity with optimal $\boldsymbol{Q}$ in the correlated channel is even higher than the open-loop capacity with uncorrelated channels (i.i.d. channel). This is also seen clearly in Fig. 2. As in the previous figure, the agreement between analytical results and simulations is impressive.

We compare the derived distribution with the distribution resulting from the generation of 10^6 random matrix realizations with the proper statistics. Figs. 4 and 5 show the results for two different ρ_s values, which indicate an impressive agreement. We also compare the optimized $\boldsymbol{Q}$ case where the transmitter only has knowledge of the statistics of the channel (the case we have been concerned with throughout this paper) with the full closed-loop capacity, where the instantaneous channel matrix is known at the transmitter. We see that the mean throughput of the two closed-loop schemes are within 5% of each other even for low SNR.

E. Gaussian Approximation of the Distribution of Mutual Information: i.i.d Interferers

We next analyze the case of uncorrelated transmitting and receiving antennas with interferers. For simplicity, we set all arrays equal in size $n_t = n_r = n_i = n$ and consider $n = 2, 3$. Using (94) and (95), we calculate the average and variance of the average mutual information I. Fig. 6 displays the cumulative distribution of the I based on these two parameters, for signal to interference-plus-noise ratio SINR $= \rho_s/(1+\rho_i) = 10$ and interference-to-noise ratio $\rho_i = 1$. Also displayed are the simulated distributions for the same cases. The agreement between analytic and simulated distributions is very good for two antennas and impressive for three (less than 1% difference).

The results of Sections IV-D and IV-E suggest that even for a low number of antennas, the mutual information distribution is very well approximated by a Gaussian.

V. Conclusion

In conclusion, we have presented an analytic approach to calculate the statistics of the mutual information of MIMO systems in the presence of spatially correlated channels, correlated interference with known channel at the receiver, and noise. In addition, we have used this method to find the optimal signal covariance $\boldsymbol{Q}$ and thus analytically calculate the capacity when the statistics of the (Gaussian) channel (with zero mean but nontrivial covariance) are known at the transmitter. The capacity calculated this way is shown in a few relevant cases to be very close to the closed-loop capacity obtained when the transmitter has full knowledge of the instantaneous channel. The ergodic capacity was calculated for a number of realistic cases to demonstrate the applicability of this approach. These methods, although formally valid for large antenna numbers, apply with very high accuracy to arrays with only a few antennas. To demonstrate this, we have simulated the distributions of small antenna arrays with spatial correlations and in the presence of interferers and have compared them with the corresponding distributions that we get by calculating the mean mutual information and its variance.

This analytic approach provides the framework and a simple tool to accurately analyze the statistics of throughput of even small arrays in the presence of arbitrary channel and noise correlations, as well as interferers with known channel at the receiver. It can be used, for example, in multiuser detection and dirty paper coding, where at each stage of detection the undecoded users are treated as interferers.

IEEE Proof

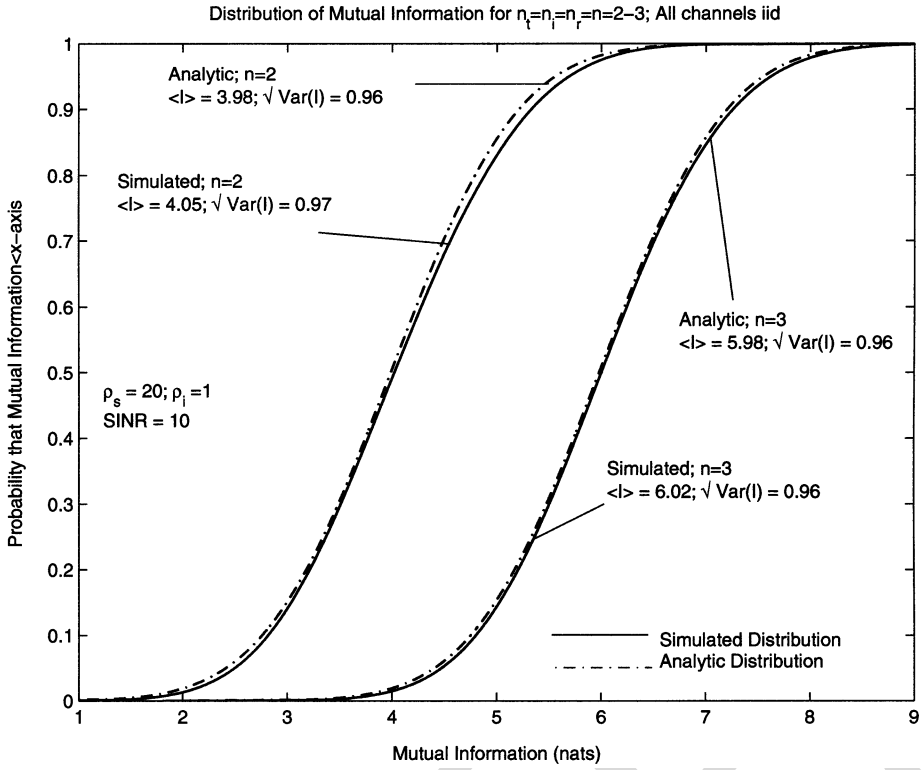


Fig. 6. CDF of the mutual information for two- and three-antenna arrays for all transmitter, receiver, and interferer with i.i.d. signal and interfering channels ($\boldsymbol{R}^i = \boldsymbol{R}^s = \boldsymbol{T}^s = \boldsymbol{N} = \boldsymbol{I}_n$). For both $n = 2$ and $n = 3$, we plot the CDF of the simulated mutual information (solid lines) and a Gaussian distribution with average and variance analytically calculated from (94) and (95). We also display the average $\langle I \rangle$ and the square root of the variance $\mathrm{Var}(I)$ of the simulated curves. The error bar in the simulated values of the variance is approximately 2%. Despite the relatively large variance, the analytical results acquired from a large antenna number approximation agree quite well with simulations.

APPENDIX I
COMPLEX INETRALS

Identity 1: Let $\boldsymbol{M}$ and $\boldsymbol{N}$ be a Hermitian positive-definite square $m \times m$ and $n \times n$ matrices, respectively, and let $\boldsymbol{X}$ be a complex $m \times n$ matrix. Then

$$(\det[\boldsymbol{N} \otimes \boldsymbol{M}])^{-1} = (\det \boldsymbol{N})^{-m} (\det \boldsymbol{M})^{-n} = \int D_c \boldsymbol{X} e^{-\frac{1}{2}\mathrm{Tr}\{\boldsymbol{N}\boldsymbol{X}^\dagger \boldsymbol{M}\boldsymbol{X}\}} \quad (112)$$

where the integration measure $D_c\boldsymbol{X}$ is given by (1). An obvious result of this identity is the case where $\boldsymbol{N} = \boldsymbol{I}_n$, which yields

$$(\det \boldsymbol{M})^{-n} = \int D_c \boldsymbol{X} e^{-\frac{1}{2}\mathrm{Tr}\{\boldsymbol{X}^\dagger \boldsymbol{M}\boldsymbol{X}\}}. \quad (113)$$

Proof: The equality

$$(\det[\boldsymbol{N} \otimes \boldsymbol{M}])^{-1} = (\det \boldsymbol{N})^{-m} (\det \boldsymbol{M})^{-n}$$

is proven by noting that the nm eigenvalues of $\boldsymbol{N} \otimes \boldsymbol{M}$ are all combinations $N_i M_j$ where N_i are the n eigenvalues of $\boldsymbol{N}$ and M_j are the m eigenvalues of $\boldsymbol{M}$.

To evaluate the integral, it is convenient to consider the $m \times n$ matrix $\boldsymbol{X}$ as a one-dimensional vector of dimension mn which we will call $\boldsymbol{x}$. We can then write the exponent $\mathrm{Tr}\{\boldsymbol{N}\boldsymbol{X}^\dagger \boldsymbol{M}\boldsymbol{X}\} = \boldsymbol{x}^\dagger \boldsymbol{P}\boldsymbol{x}$ where $\boldsymbol{P} = \boldsymbol{N} \otimes \boldsymbol{M}$ is an $mn \times mn$ matrix. We can then rewrite (112) as

$$(\det \boldsymbol{P})^{-1} = \int D_c \boldsymbol{x} e^{-\frac{1}{2}\boldsymbol{x}^\dagger \boldsymbol{P}\boldsymbol{x}} \quad (114)$$

which is proven trivially by defining $\boldsymbol{y} = \boldsymbol{P}^{1/2}\boldsymbol{x}$ such that

$$\int D_c \boldsymbol{x} e^{-\frac{1}{2}\boldsymbol{x}^\dagger \boldsymbol{P}\boldsymbol{x}} = (\det \boldsymbol{P})^{-1} \int D_c \boldsymbol{y} e^{-\frac{1}{2}\boldsymbol{y}^\dagger \boldsymbol{y}} = (\det \boldsymbol{P})^{-1}. \quad (115)$$

The factor of $\det \boldsymbol{P}$ here arises entirely as the Jacobian of the transformation. (Note that since $D_c\boldsymbol{y}$ involves integration over both real and imaginary parts, we obtain a Jacobian $\det \boldsymbol{P}$ rather than $\det \boldsymbol{P}^{1/2}$.) □

Identity 2: Let $\boldsymbol{X}$, $\boldsymbol{A}$, $\boldsymbol{B}$ be $m \times n$ complex matrices and $\boldsymbol{N}$, $\boldsymbol{M}$ positive-definite Hermitian $n \times n$ and $m \times m$, respectively. Then, the following equality holds:

$$\int D_c \boldsymbol{X} e^{-\frac{1}{2}\mathrm{Tr}\{\boldsymbol{N}\boldsymbol{X}^\dagger \boldsymbol{M}\boldsymbol{X} + \boldsymbol{A}^\dagger \boldsymbol{X} + \boldsymbol{X}^\dagger \boldsymbol{B}\}} = (\det[\boldsymbol{N} \otimes \boldsymbol{M}])^{-1} e^{-\frac{1}{2}\mathrm{Tr}\{\boldsymbol{N}^{-1}\boldsymbol{A}^\dagger \boldsymbol{M}^{-1}\boldsymbol{B}\}}. \quad (116)$$

Proof: As in Identity 1, it is convenient to rephrase this identity in terms of nm-dimensional vectors $\boldsymbol{x}$, $\boldsymbol{a}$, and $\boldsymbol{b}$, rather than $n \times m$-dimensional matrices $\boldsymbol{X}$, $\boldsymbol{A}$, and $\boldsymbol{B}$. Thus, (116) can be rewritten as

$$\int D_c \boldsymbol{x} e^{-\frac{1}{2}\left[\boldsymbol{x}^\dagger \boldsymbol{P}\boldsymbol{x} + \boldsymbol{a}^\dagger \boldsymbol{x} + \boldsymbol{x}^\dagger \boldsymbol{b}\right]} = (\det \boldsymbol{P})^{-1} e^{-\frac{1}{2}\boldsymbol{a}^\dagger \boldsymbol{P}^{-1}\boldsymbol{b}} \quad (117)$$

where $\boldsymbol{P} = \boldsymbol{N} \otimes \boldsymbol{M}$ (which implies $\boldsymbol{P}^{-1} = \boldsymbol{N}^{-1} \otimes \boldsymbol{M}^{-1}$). Analogous to Identity 1, to prove (117) we define $\boldsymbol{y} = \boldsymbol{P}^{1/2}\boldsymbol{x}$ as well as $\tilde{\boldsymbol{b}} = \boldsymbol{P}^{-1/2}\boldsymbol{b}$ and $\tilde{\boldsymbol{a}} = \boldsymbol{P}^{-1/2}\boldsymbol{a}$ so that

$$\int D_c \boldsymbol{x} e^{-\frac{1}{2}\left[\boldsymbol{x}^\dagger \boldsymbol{P}\boldsymbol{x} + \boldsymbol{a}^\dagger \boldsymbol{x} + \boldsymbol{x}^\dagger \boldsymbol{b}\right]} = (\det \boldsymbol{P})^{-1} \int D_c \boldsymbol{y} e^{-\frac{1}{2}\left[\boldsymbol{y}^\dagger \boldsymbol{y} + \tilde{\boldsymbol{a}}^\dagger \boldsymbol{y} + \boldsymbol{y}^\dagger \tilde{\boldsymbol{b}}\right]}. \quad (118)$$

IEEE Proof

It is then convenient to define real variables $z_j = \mathrm{Re} y_j$ and $w_j = \mathrm{Im} y_j$ so that the integration measure is

$$D_c \boldsymbol{y} = \prod_j \frac{dw_j\, dz_j}{2\pi}. \tag{119}$$

The exponent can then be rewritten as

$$\begin{aligned}&\boldsymbol{y}^\dagger \boldsymbol{y} + \tilde{\boldsymbol{a}}^\dagger \boldsymbol{y} + \boldsymbol{y}^\dagger \tilde{\boldsymbol{b}} \\ &= -\tilde{\boldsymbol{a}}^\dagger \tilde{\boldsymbol{b}} + \sum_j \left[\left(z_j + \frac{\tilde{a}_j^* + \tilde{b}_j}{2} \right)^2 + \left(w_j - \frac{\tilde{a}_j^* - \tilde{b}_j}{2i} \right)^2 \right].\end{aligned} \tag{120}$$

We thus have

$$\begin{aligned}&\int D_c \boldsymbol{x} e^{-\frac{1}{2}[\boldsymbol{x}^\dagger \boldsymbol{P} \boldsymbol{x} + \boldsymbol{a}^\dagger \boldsymbol{x} + \boldsymbol{x}^\dagger \boldsymbol{b}]} \\ &= \frac{e^{-\frac{1}{2}\tilde{\boldsymbol{a}}^\dagger \tilde{\boldsymbol{b}}}}{\det \boldsymbol{P}} \prod_j \left[\int \frac{dz_j dw_j}{2\pi} e^{-\frac{1}{2}(z_j + \frac{\tilde{a}_j^* + \tilde{b}_j}{2})^2 - \frac{1}{2}(w_j - \frac{\tilde{a}_j^* - \tilde{b}_j}{2i})^2} \right]\end{aligned} \tag{121}$$

$$= (\det \boldsymbol{P})^{-1} e^{-\frac{1}{2}\tilde{\boldsymbol{a}}^\dagger \tilde{\boldsymbol{b}}} = (\det \boldsymbol{P})^{-1} e^{-\frac{1}{2}\boldsymbol{a}^\dagger \boldsymbol{P}^{-1} \boldsymbol{b}}. \tag{122}$$

□

Identity 3 (Hubbard–Stratonovich Transformation): Let $\boldsymbol{U}$, $\boldsymbol{V}$ be arbitrary complex $\nu \times \nu$ matrices, where ν here is assumed to be an arbitrary positive integer. Then the following identity holds:

$$\exp[-\mathrm{Tr}\{\boldsymbol{UV}\}] = \int d\mu(\boldsymbol{T}, \boldsymbol{R}) \exp[\mathrm{Tr}\{\boldsymbol{RT} - \boldsymbol{UT} - \boldsymbol{RV}\}]. \tag{123}$$

In the above equation, the auxiliary matrices $\boldsymbol{T}$, $\boldsymbol{R}$ are general complex matrices $\nu \times \nu$ and their integration measure is given by (2). The integration of the elements of $\boldsymbol{R}$ and $\boldsymbol{T}$ is along contours in complex space parallel to the real and imaginary axis.

Proof: We start with the simple case, where $\boldsymbol{U}, \boldsymbol{V}$ are both scalars u, v. Then

$$\begin{aligned}&\frac{1}{2\pi i} \int_{\mathrm{Re\ axis}} dr \int_{\mathrm{Im\ axis}} dt\, e^{[rt - ut - rv]} \\ &= \frac{1}{2\pi i} \int_{\mathrm{Re\ axis}} dr \int_{\mathrm{Im\ axis}} dt\, e^{[(r-u)(t-v) - uv]} \\ &= \frac{1}{2\pi i} \int_{\mathrm{Re\ axis}} dr' \int_{\mathrm{Im\ axis}} dt'\, e^{[r't' - uv]} \\ &= e^{-uv} \int_{\mathrm{Re\ axis}} dr'\, \delta(r') = e^{-uv}\end{aligned} \tag{124}$$

where $r' = r - u$ and $t' = t - v$.

To prove (123) for general complex matrices $\boldsymbol{U}$, $\boldsymbol{V}$, we note that

$$\mathrm{Tr}\{\boldsymbol{UV}\} = \sum_{a,b=1}^{\nu} U_{ab} V_{ba}$$

and thus,

$$\exp[-\mathrm{Tr}\{\boldsymbol{UV}\}] = \prod_{a,b} \exp[-U_{ab} V_{ba}].$$

As a result, we can apply the result of (124) for every pair of elements U_{ab}, V_{ba}, resulting in the left-hand side product of (123) with measure of (2). □

A. Wick's Theorem

Wick's theorem is a general method for simplifying the calculation of moments of Gaussian variables. It was introduced and has been extensively used in physics [14], [15]. Here, we will only quote the results we use (as above in the discussions of Identities 1 and 2, it is useful to rewrite $m \times n$ matrices as vectors of length mn).

Theorem 1: Given a $p \times p$ matrix $\boldsymbol{P}$ and an arbitrary operator $O(\boldsymbol{x})$, we define the bracket notation

$$\langle\langle O(\boldsymbol{x}) \rangle\rangle = (\det \boldsymbol{P})^{1/2} \int d\mu(\boldsymbol{x}) e^{-\frac{1}{2}\boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x}} O(\boldsymbol{x}) \tag{125}$$

with the measure defined

$$d\mu(\boldsymbol{x}) = \prod_{j=1}^{p} \frac{dx_j}{\sqrt{2\pi}} \tag{126}$$

and the integration is along an appropriately chosen contour. Given a set of indexes $t_k \in 1 \cdots p$ for $k = 1 \cdots q$, we have

$$\left\langle\left\langle \prod_{k=1}^{q} x_{t_k} \right\rangle\right\rangle = \sum_{pairs} \langle\langle x_{t_1} x_{t_2} \rangle\rangle \langle\langle x_{t_3} x_{t_4} \rangle\rangle \cdots \langle\langle x_{t_{q-1}} x_{t_q} \rangle\rangle \tag{127}$$

and equals zero if q is odd. The sum here is over all possible rearrangements of the orderings of the indexes such that different indexes are paired with each other (with each distinct pairing being counted once). For example (with $p \geq 8$)

$$\begin{aligned}\langle\langle x_1 x_3 x_4 x_8 \rangle\rangle &= \langle\langle x_1 x_3 \rangle\rangle \langle\langle x_4 x_8 \rangle\rangle + \langle\langle x_1 x_4 \rangle\rangle \langle\langle x_3 x_8 \rangle\rangle \\ &\quad + \langle\langle x_1 x_8 \rangle\rangle \langle\langle x_3 x_4 \rangle\rangle\end{aligned} \tag{128}$$

but $\langle\langle x_1 x_3 x_8 \rangle\rangle = 0$ (since q is odd). Further, it can be shown that

$$\langle\langle x_a x_b \rangle\rangle = [\boldsymbol{P}^{-1}]_{ab} \tag{129}$$

and $\langle\langle 1 \rangle\rangle = 1$. Thus, Wick's theorem allows us to simply write down any moment of a Gaussian variable.

Proof: Omitted. □

Appendix II
Grassman Variables

A. Properties

Grassman variables have seen widespread use in theoretical physics and can be used as an algebraic trick for doing certain integrals. (For more information see, for example, [15, Ch. 2].) Grassman variables have the following multiplication properties:

$$\lambda A = A \lambda \tag{130}$$

$$AB = -BA \tag{131}$$

$$A^2 = 0 \tag{132}$$

$$(AB)C = C(AB) \tag{133}$$

$$e^{AB} = 1 + AB + \frac{1}{2} ABAB + \cdots = 1 + AB \tag{134}$$

where A, B, and C are Grassman variables and λ is an arbitrary complex number. Equations (130) and (131) can be taken as definitions of how to multiply Grassman numbers with each other, or with complex numbers. Equation (132) is derived from (131) by taking $A = B$. Equation (133) is also derived from (131) and signifies that a product of two Grassman variables behaves as a complex number in the sense that it can be written either to the left or the right of a Grassman number (as in (130)). Finally, (134) results from expanding the exponential and using (132) to eliminate all higher terms in the expansion.

IEEE Proof

We also make the following definitions for integration over Grassman variables:

$$\int dA = 0 \tag{135}$$

$$\int dA A = 1 \tag{136}$$

where $\int dA$ is a linear operator. The differentials also anticommute such that $dB\,dA = -dA\,dB$ and $A\,dB = -dB\,A$. A useful relation, which we will need later can be derived as follows. For A and B Grassman variables

$$\begin{aligned}\int dA\,dB\,e^{BA} &= \int dA\,dB(1+BA)\\ &= \lambda \int dA \left[\int dB\,B\right] A\\ &= \int dA\,A = 1. \end{aligned}\tag{137}$$

Making variable transformations of Grassman variables is nontrivial. For example, to preserve (136), if $B = \lambda A$ (with λ a complex number, and A and B Grassman variables) then $dB = A/\lambda$ which looks backward from the usual Jacobian of a transformation. Making a change of variables with multiple Grassman variables is similar. Given a p-dimensional vector of Grassman variables $\boldsymbol{a}$ with Grassman elements a_i, we define a new p-dimensional vector $\boldsymbol{b} = \boldsymbol{M}^{-1}\boldsymbol{a}$ where $\boldsymbol{M}$ is a $p \times p$ complex matrix. To find the Jacobian of the transformation, we need only preserve the property

$$\int da_p \cdots da_1\, a_1 \cdots a_p = 1. \tag{138}$$

Now with $a_i = \sum_j M_{ij} b_j$, we have

$$a_1 \cdots a_p = \sum_{j_1} \cdots \sum_{j_p} M_{1j_1} \cdots M_{pj_p} b_{j_1} \cdots b_{j_n}. \tag{139}$$

However, the product of the b's is equal to a sign ($\pm$) times $b_1 \cdots b_p$ when all the j indexes are different, and is zero if any two j's are equal. Thus, we can write

$$a_1 \cdots a_p = b_1 \cdots b_p \sum M_{1j_1} \cdots M_{pj_p} P \tag{140}$$

where the sum is over all permutaions of the indexes $j_1, \ldots j_p = 1 \ldots p$ and P is the sign of the permutation ($+1$ for even permutations and -1 for odd permutations). We recognize this expression immediately as $\det \boldsymbol{M}$. To assure that the bs also obey (138) we must then have

$$db_p \cdots db_1 = \det \boldsymbol{M}\, da_p \cdots da_1 \tag{141}$$

for $\boldsymbol{a} = \boldsymbol{M}\boldsymbol{b}$, which again looks backward from the way normal complex vectors transform.

B. Integrals Over Grassman Variables

Identity 4: Let $\boldsymbol{A}$, $\overline{\boldsymbol{A}}$ be $n \times m$- and $m \times n$-dimensional matrices with elements $A_{a\alpha}$ and $\overline{A}_{\alpha a}$ that are Grassman variables ($a \in 1 \cdots n$ and $\alpha \in 1 \cdots m$). Let $\boldsymbol{M}$, $\boldsymbol{N}$ be $m \times m$- , $n \times n$-dimensional Hermitian matrices (with traditional complex entries)

$$\begin{aligned}\int D_g \boldsymbol{A} e^{\mathrm{Tr}\{\boldsymbol{M}\overline{\boldsymbol{A}}\boldsymbol{N}\boldsymbol{A}\}} &= \det(\boldsymbol{M}\otimes\boldsymbol{N})\\ &= (\det \boldsymbol{M})^n (\det \boldsymbol{N})^m \end{aligned}\tag{142}$$

where the integration measure is given by (3). This result can be compared with (112).

Proof: As in Identity 1, it is convenient to define nm-dimensional vectors $\boldsymbol{a}$ and $\overline{\boldsymbol{a}}$ which are simply rearrangements of the elements of $\boldsymbol{A}$ and $\overline{\boldsymbol{A}}$. We also define $\boldsymbol{P} = \boldsymbol{M} \otimes \boldsymbol{N}$. In this language, (142) is thus equivalent to

$$\int da_1 d\overline{a}_1 \cdots da_{(nm)} d\overline{a}_{(nm)} e^{\overline{\boldsymbol{a}}\boldsymbol{P}\boldsymbol{a}} = \det \boldsymbol{P}. \tag{143}$$

Changing variables as discussed above by writing $\boldsymbol{b} = \boldsymbol{P}\boldsymbol{a}$, we then have

$$\begin{aligned}&\int da_1 d\overline{a}_1 \cdots da_{(nm)} d\overline{a}_{(nm)} e^{\overline{\boldsymbol{a}}\boldsymbol{P}\boldsymbol{a}}\\ &= (\det \boldsymbol{P})(-1)^q \int d\overline{a}_1 \cdots d\overline{a}_{(nm)} db_1 \cdots db_{(nm)} e^{\overline{\boldsymbol{a}}\boldsymbol{b}} \end{aligned}\tag{144}$$

$$= (\det \boldsymbol{P}) \prod_{j=1}^{nm} \left[\int db_j d\overline{a}_j e^{\overline{a}_j b_j}\right] = \det \boldsymbol{P} \tag{145}$$

where the integral in brackets becomes unity as shown in (137) (and the factor of $(-1)^q$ comes from reordering where $q = (nm)(nm-1)/2$). □

Identity 5 (Completing the Square): Let $\boldsymbol{A}$, $\boldsymbol{C}$ and $\overline{\boldsymbol{A}}$, $\overline{\boldsymbol{B}}$ be $n \times m$- and $m \times n$-dimensional matrices, respectively, and $\boldsymbol{M}$, $\boldsymbol{N}$ be $m \times m$-, $n \times n$-dimensional positive-definite Hermitian matrices. Then

$$\begin{aligned}\det(\boldsymbol{M}\otimes\boldsymbol{N})^{-1} &\int D_g \boldsymbol{A} e^{\mathrm{Tr}\{\boldsymbol{M}\overline{\boldsymbol{A}}\boldsymbol{N}\boldsymbol{A} + \overline{\boldsymbol{B}}\boldsymbol{A} + \overline{\boldsymbol{A}}\boldsymbol{C}\}}\\ &= e^{-\mathrm{Tr}\{\boldsymbol{M}^{-1}\overline{\boldsymbol{B}}\boldsymbol{N}^{-1}\boldsymbol{C}\}} \end{aligned}\tag{146}$$

where $D_g \boldsymbol{A}$ is given by (3).

Proof: First, let us define the auxiliary matrices given by

$$\begin{aligned}\overline{\boldsymbol{A}}' &= \overline{\boldsymbol{A}} + \boldsymbol{M}^{-1}\overline{B}N^{-1}\\ \boldsymbol{A}' &= \boldsymbol{A} + \boldsymbol{N}^{-1}CM^{-1}. \end{aligned}\tag{147}$$

It is easy to check that the integration measure of the new variables is equal to the old, i.e., $D\boldsymbol{A}' = D\boldsymbol{A}$. The exponent can be expressed in terms of $\boldsymbol{A}'$, $\overline{\boldsymbol{A}}'$ as follows:

$$\begin{aligned}&\mathrm{Tr}\{\boldsymbol{M}\overline{\boldsymbol{A}}\boldsymbol{N}\boldsymbol{A} + \overline{\boldsymbol{B}}\boldsymbol{A} + \overline{\boldsymbol{A}}\boldsymbol{C}\}\\ &\qquad = \mathrm{Tr}\{\boldsymbol{M}\overline{\boldsymbol{A}}'\boldsymbol{N}\boldsymbol{A}' - \boldsymbol{M}^{-1}\overline{\boldsymbol{B}}\boldsymbol{N}^{-1}\boldsymbol{C}\}. \end{aligned}\tag{148}$$

Substituting (148) into the left-hand side of (146) and noting that the resulting integral over $\boldsymbol{A}'$, $\overline{\boldsymbol{A}}'$ is given by Identity 5, we prove (146). □

Identity 6 (Hubbard–Stratonovitch): Let $\overline{\boldsymbol{B}}$, $\boldsymbol{C}$ be arbitrary Grassman $\nu \times \nu$ matrices. Then the following identity holds:

$$\int D_g \boldsymbol{A} \exp\left[\mathrm{Tr}\{\overline{\boldsymbol{A}}\boldsymbol{A} - \overline{\boldsymbol{B}}\boldsymbol{A} - \overline{\boldsymbol{A}}\boldsymbol{C}\}\right] = \exp\left[-\mathrm{Tr}\{\overline{\boldsymbol{B}}\boldsymbol{C}\}\right]. \tag{149}$$

In the preceding equation, the auxiliary matrices $\overline{\boldsymbol{A}}$, $\boldsymbol{A}$ are Grassman $\nu \times \nu$ matrices and the integration measure $D_g \boldsymbol{A}$ is given by (3).

Proof: We use Identity 5 with both $\boldsymbol{M}$ and $\boldsymbol{N}$ equal the identity matrix. □

ACKNOWLEDGMENT

We wish to acknowledge enlightening discussions with Harold Baranger and Ridhar Gollamudi. We also wish to thank the Editor and the anonymous reviewers for their insightful comments.

IEEE Proof

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, 1988.

[2] İ. E. Telatar, "Capacity of multi-antenna gaussian channels," *Europ. Trans. Telecommun. and Related Technol.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[3] P. W. Wolniansky *et al.*, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI Int. Symp. Signals, Systems and Electronics*, 1998, pp. 295–300.

[4] J. Ling *et al.*, "Multiple transmit multiple receive (MTMR) capacity survey in Manhattan," *IEE Electron. Lett.*, vol. 37, no. 16, pp. 1041–1042, Aug. 2001.

[5] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 622–640, Mar. 1999.

[6] P. B. Rapajic and D. Popescu, "Information capacity of a random signature multiple-input multiple-output chanel," *IEEE Trans. Commun.*, vol. 48, pp. 1245–1248, Aug. 2000.

[7] A. Lozano and A. M. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Trans. Inform. Theory*, vol. 48, pp. 3117–3128, Dec. 2002.

[8] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inform. Theory*, vol. 45, pp. 641–657, Mar. 1999.

[9] C. N. Chuah *et al.*, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inform. Theory*, vol. 48, pp. 637–650, Mar. 2002.

[10] A. L. Moustakas *et al.*, "Communication through a diffusive medium: Coherence and capacity," *Science*, vol. 287, pp. 287–290, Jan. 2000.

[11] A. M. Sengupta and P. P. Mitra, "Distributions of singular values for some random matrices," *Phys. Rev. E*, vol. 60, no. 3, pp. 3389–3392, 1999.

[12] ——, "Capacity of multivariate channels with multiplicative noise: I. Random matrix techniques and large-n expansions for full transfer matrices," *LANL arXiv:physics/0010081*, Oct. 2000.

[13] S. F. Edwards and P. W. Anderson, *J. Phys., F 5*, 1975. **AU: PLEASE CITE TITLE OF PAPER, VOLUME, INCLUSIVE PAGE NUMBERS, AND MONTH OF PUBLICATION IF AVAILABLE. THANK YOU**.

[14] G. P. M. Mézard and M. A. Virasoro, *Spin Glass Theory and Beyond*. Singapore: World Scientific, 1987.

[15] C. Itzykson and J.-M. Drouffe, *Statistical Field Theory*. Cambridge, U.K.: Cambridge Univ. Press, 1989.

[16] N. Sourlas, "Spin glass models as error correcting codes," *Nature*, vol. 339, no. 6227, pp. 693–695, June 1989.

[17] A. Montanari and N. Sourlas, "Statistical mechanics of turbo codes," *Europ. Phys. J. B*, vol. 18, no. 1, pp. 107–119, 2000.

[18] A. Montanari, "Turbo codes: The phase transition," *Europ. Phys. J. B*, vol. 18, no. 1, pp. 121–136, 2000.

[19] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2888–2910, Nov. 2002.

[20] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multi-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inform. Theory*, submitted for publication.

[21] P. J. Smith and M. Shafi, "On the Gaussian approximation to the capacity of wireless MIMO systems," in *Proc. IEEE Int. Conf. Communications*, 2002, p. 406. **AU: PLEASE SUPPLY INCLUSIVE PAGE NUMBERS AND LOCATION OF CONFERENCE IF AVAILABLE. THANK YOU**.

[22] Z. Wang and G. B. Giannakis, "Outage mutual information of space-time MIMO channels," in *Proc. Asilomar Conf.*, Oct. 2002. **AU: PLEASE SUPPLY PAGE NUMBERS AND LOCATION OF CONFERENCE IF AVAILABLE. THANK YOU**.

[23] D. Chizhik *et al.*, "Effect of antenna separation on the capacity of BLAST in correlated channels," *IEEE Commun. Lett.*, vol. 4, pp. 337–339, Nov. 2000.

[24] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*. New York: McGraw-Hill, 1978.

[25] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York: Springer-Verlag, 1998.

[26] S. A. Jafar and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," preprint, 2001.

[27] R. M. Buehrer, "Generalized equations for spatial correlation for low to moderate angle-spread," in *Proc. 10th Annu. Virginia Tech Symp. Wireless Personal Communications*, July 2000, pp. 70–77.

[28] S. H. Simon and A. L. Moustakas, "Optimizing MIMO systems with channel covariance feedback," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 406–417, Apr. 2003.

Aris L. Moustakas (M'02) received the B.S. degree in physics from the California Institute of Technology, Pasadena, in 1990 and the M.S. and Ph.D. degrees in theoretical solid state physics from Harvard University, Cambridge, MA, in 1992 and 1996, respectively.

In 1998, after a two-year stint in the Greek Navy, he joined Bell Labs, Lucent Technologies, Whippany, NJ, first in the Physical Sciences Division and then in the Wireless Advanced Technology Laboratory. His current research interests lie in the areas of information theory of multiple-antenna systems, signal processing for smart antennas, and wide-band channel modeling. He has authored several papers and patents on these topics.

Steven H. Simon received the B.Sc. degree in physics and mathematics from Brown University, Providence, RI, in 1989 and the M.S. and Ph.D. degrees in theoretical condensed matter physics from Harvard University, Cambridge, MA, in 1992 and 1995, respectively.

From 1995 to 1997, he was a Postdoctoral Researcher in the Physics Department, Massachusettes Institute of Technology, Cambridge. In 1997, he joined Bell Labs, Lucent Technologies, Murray Hill, NJ, as a Member of Technical Staff in the Theoretical Physics Department. In January of 2000, he was promoted to Director of Theoretical Physics Research. His current research interests include multiple-antenna communications, information theory, materials physcs, quantum phenomena, and nanotechnology.

Anirvan Mayukh Sengupta received the B.Sc. degree from Calcutta University, Calcutta, India, and the Ph.D. degree from Bombay University, Bombay, India, in 1986 and 1994, respectively. His doctoral work at Tata Institute, Bombay, was on applying RMT to noncritical string theory.

He spent the years 1992–1994 at Ecole Normale Superieure, Paris, France, applying tools from statistical field theory to problems in condensed matter physics. From 1994 to 2002, with the exception of the year 1996-1997, he was at Bell Laboratories, Murray Hill, NJ, first as a Postdoctoral Fellow and then as a Member of Technical Staff. Although he joined Bell Labs as a theoretical condensed matter physicist, his work there gradually focused on issues of information transmission/processing in physical and biological systems. Currently, he is an Associate Professor in the Department of Physics and Astronomy and a member of the BioMaPS Instituet at Rutgers University, Piscataway, NJ, where he works on signaling and gene regulation in biomolecular networks.

IEEE Proof