MIMO Systems with Restricted Pre/Post-coding – Capacity Analysis based on Coupled Doubly-correlated Wishart Matrices

Vishnu V. Ratnam, *Student Member, IEEE,* Andreas F. Molisch, *Fellow, IEEE,* and Haralabos C. Papadopoulos, *Member, IEEE*

***Abstract*—Many practical communication systems have some form of restricted precoding or postcoding such as Antenna Selection, Selection Combining, Beam Selection and Limited Feedback precoding, to name a few. The capacity analysis of such systems is, in general, difficult and previous works in the literature provide results only for certain simplified cases. The current paper derives a novel approach to analyze the capacity for such systems under a very generic setting. The results are based on asymptotic closed-form expressions for second-order statistics and joint distributions of eigenvalues for a set of coupled, doubly-correlated Wishart matrices. A tight approximation to the joint distribution of the eigenvalues in the non-asymptotic regime is also proposed. These results are then used to show that the system capacity can be approximated as the largest element of a correlated Gaussian vector. Showing that this is equivalent to the problem of finding the distribution of sum of lognormals, we propose a novel approach to characterize its distribution. As an application, the capacity for an antenna selection system and a limited feedback precoding system are compared to their respective approximations. The paper also demonstrates how the results can be used to design the precoding codebook in limited feedback systems.**

***Index Terms*—Restricted precoding, Antenna Selection, Limited Feedback precoding, Joint eigenvalue distribution, Wishart matrices, Capacity distribution.**

I. Introduction

With the rising amount of downlink data traffic but a limited available spectrum, there is an impending spectrum crunch and a need for higher spectral efficiency. Multiple-input-multiple-output (MIMO) transmission technologies promise large gains in spectral efficiency by offering spatial degrees of freedom for data transmission. In fact, with the progress in digital and radio-frequency (RF) hardware technology, the development of low complexity precoding algorithms and also the possibility of using the mm-wave frequency band for cellular transmission, cellular networks are moving towards the massive MIMO regime i.e., base stations with dozens or hundreds of antenna elements. First proposed in [1], data transmission in massive MIMO systems offers several advantages such as simplified precoding, higher beamforming gains etc. However, having a massive antenna array brings with it several problems. Firstly, the cost of channel state information (CSI) feedback from user terminal to the base-station, in the downlink scenario, increases significantly for frequency division duplexing systems[1]. This has led to the proposition of limited feedback precoding (LFP) [2], wherein the transmitter maintains a codebook of precoding vectors and the receiver feeds back only the codebood index based on the CSI. Secondly, since RF hardware such as analogue-to-digital converters, digital-to-analogue converters, mixers, RF filters etc. are power hungry and expensive, with massive antenna arrays it may be impractical to equip each antenna with a dedicated RF chain. This has led to the proposition of transmit antenna selection (TAS) [3], [4], wherein a smaller number of RF chains feed the transmit antennas via an array of RF switches. Another such example is a beam selection system [5]–[7]. In all of these precoding methods, the data precoding is restricted to take-on only a restricted set of values (either due to limited CSI or due to limitations in the RF hardware). We shall refer to any such system with restrictions on data precoding as a restricted pre-coded system.[2] Such systems form an important class of practically viable massive MIMO systems.

The performance analysis for restricted precoded systems is difficult in general and no closed form expressions (for the most general setting) are known to date. For the case of limited feedback precoding, lower bounds on the capacity with beamforming in an isotropic channel were proposed in [8]. System upper bounds under similar settings were considered in [9], [10] etc. The lower bound in [8] was extended to the case of spatial multiplexing in [11]. For correlated channels, heuristic designs for the codebook were suggested in [8], [12], [13]. However, bounds on the performance and the optimal design of the codebook for spatial multiplexing in a correlated channel are not available in literature to the best of the authors' knowledge. Even in the relatively simple case of random vector quantized beamforming, performance bounds and good codebook designs for a correlated MIMO channel were found only recently [14]. A more complete discussion of the results prior to 2008 are available in [2]. Similarly, for the case of

V. V. Ratnam and A. F. Molisch are with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA, 90089 USA (e-mail: {ratnam, molisch}@usc.edu)

H. C. Papadopoulos is with Docomo Innovations, Palo Alto, CA, 94304 USA (e-mail: hpapadopoulos@docomoinnovations.com)

[1] Unlike with time division duplexing, where CSI from uplink training can be used for downlink transmission, with frequency division duplexing, we need to rely on downlink training and uplink CSI feedback.

[2] A system with similar restrictions at the receiver, for example: Receive Antenna Selection, shall be referred to as a Restricted postcoded system

TAS[3], bounds on the distribution of the capacity for spatial multiplexing in an isotropic channel were found in [15]. The ergodic capacity in the high and low signal-to-noise ratio (SNR) regimes were discussed in [16]. A loose upper bound on the outage probability for a spatial multiplexing system with receive antenna selection, in a correlated fading channel was considered in [17]. However, the performance analysis of spatial multiplexing with TAS in a correlated channel is not available in the literature to the best of the authors' knowledge. A more complete review of literature on antenna selection is available in [2], [4], [18].

Evaluation of the system performance is important in designing good systems. For example, it can aid in the design of good codebooks for a LFP system. In this paper we develop a mathematical framework for the analysis of such systems. As shall be shown in Sec. II, some of the important performance measures like mean and outage capacity are a function of the eigenvalues of a set of coupled[4], doubly-correlated Wishart matrices. Therefore characterizing the joint eigenvalue distribution across these coupled Wishart matrices is an essential step towards characterizing the performance. The asymptotic eigenvalue distribution [19], [20], non-asymptotic diagonal distribution [21] and joint eigenvalue distributions [22] for a single Wishart matrix have been widely characterized both with and with-out correlated entries. The joint eigenvalue distribution for a pair of correlated Wishart matrices was characterized in [23], [24]. However, the joint eigenvalue distribution across a larger set of correlated, let alone coupled, Wishart matrices has not been studied in literature to the best of our knowledge.

The contributions of this paper are as follows: We derive the asymptotic second-order statistics and joint distribution of eigenvalues across a set of coupled, doubly-correlated Wishart matrices. We also propose a tight approximation for the joint distribution of eigenvalues in the non-asymptotic regime. These results are then used to approximate the distribution of capacity for a restricted precoded system. In the process, we propose a new technique for finding the distribution of the largest element of a correlated Gaussian vector. As an application of the proposed techniques, we also design an efficient codebook for a limited feedback system. Though we focus here on restricted precoding, the presented analysis can also be extended to the case of restricted postcoding.

The rest of the paper is organized as follows: In Sec. II, the channel model is introduced and the problem of finding the capacity of a restricted precoded system is formulated. The joint distribution of the channel eigenvalues are derived in Sec. III. Using these results, the approximate capacity distribution for a restricted precoded system is derived in Sec. IV. To study the effectiveness of the approximation, simulations under some practical channel parameters are performed in Sec. V. As an application of the results, we also demonstrate how it can be used to design a good codebook for a limited feedback system in Sec. VI. Finally, the conclusions are presented in Sec. VII.

Notation used in this work is as follows: scalars are represented by light-case letters; vectors by bold-case letters; matrices are represented by capitalized bold-case letters and sets; and subspaces are represented by calligraphic letters. Additionally, a_i represents the i-th element of a vector $\mathbf{a}$, $\|\mathbf{a}\|_P$ represents the L_P norm of a vector $\mathbf{a}$, $[\mathbf{A}]_{i,j}$ represents the (i, j)-th element of a matrix $\mathbf{A}$, $[\mathbf{A}]_{\mathbf{c}\{i\}}$ and $[\mathbf{A}]_{\mathbf{r}\{i\}}$ represent the i-th column and row vectors of matrix $\mathbf{A}$, respectively, $\|\mathbf{A}\|_\mathbf{F}$ represents the Frobenius norm of a matrix $\mathbf{A}$, $\mathbf{A}^\dagger$ is the conjugate transpose of a matrix $\mathbf{A}$ and $|\mathcal{A}|$ represents the cardinality of a set $\mathcal{A}$ or dimension of a space $\mathcal{A}$. Also, $\mathbb{E}\{\}$ represents the expectation operator, $\mathbb{P}$ is the probability operator, $\mathbb{I}_i$ and $\mathbb{O}_{i,j}$ are the $i \times i$ and $i \times j$ identity and zero matrices respectively, and $\mathbb{R}$ and $\mathbb{C}$ represent the field of real and complex numbers.

II. General Assumptions and Channel model

We consider a point-to-point MIMO link where the transmitter has an array with $N \gg 1$ antenna elements and the receiver has $M \leq N$ antenna elements, respectively. We assume a narrow-band system with a frequency flat and temporally block fading channel. The channel fading statistics are assumed to be Rayleigh in amplitude, doubly spatially correlated (both at transmitter and receiver end) and to follow the widely used Kronecker correlation model [25]. The transmitter is assumed to have restricted precoding, wherein, the transmit data vector is precoded by a precoding matrix of dimension $N \times K$, where $K \leq N$. For LFP, K corresponds to the number of transmit data streams and in the case of TAS, K corresponds to the number of RF chains, respectively. Note that in LFP, we typically have the number of data streams $K \leq M$. However, for analytical tractability, in this paper we assume $K \geq M$ and the case of $K < M$ is deferred to future work. Under these assumptions, the baseband downlink received signal vector can be expressed as:

$$\begin{aligned} \mathbf{y} &= \sqrt{\rho}\mathbf{H}\mathbf{T}\mathbf{x} + \mathbf{n} \\ &= \sqrt{\rho}\mathbf{R}_{\mathbf{rx}}^{1/2}\mathbf{G}\mathbf{R}_{\mathbf{tx}}^{1/2}\mathbf{T}\mathbf{x} + \mathbf{n} \end{aligned} \tag{1}$$

where $\mathbf{y}$ is the $M \times 1$ received signal vector, ρ is the SNR, $\mathbf{H}$ is the $M \times N$ small-scale fading channel matrix, $\mathbf{R}_{\mathbf{tx}}$ is the $N \times N$ transmit spatial-correlation matrix, $\mathbf{R}_{\mathbf{rx}}$ is the $M \times M$ receive spatial-correlation matrix, $\mathbf{G}$ is an $M \times N$ matrix with independent and identically distributed (i.i.d.) $\mathcal{CN}(0, 1)$ components (circularly symmetric zero-mean complex Gaussian entries with unit variance), $\mathbf{n} \sim \mathcal{CN}(\mathbb{O}_{M\times1}, \mathbb{I}_M)$ is the $M \times 1$ normalized AWGN noise vector, $\mathbf{T}$ is a $N \times K$ transmit precoding matrix and $\mathbf{x}$ is the $K \times 1$ transmit data vector.

In a restricted precoded system, the precoding matrix $\mathbf{T}$ can only attain a restricted set of values, i.e., $\mathbf{T} \in \mathcal{T}$. Here, we shall refer to this set $\mathcal{T}$ as a codebook. For example, in LFP, $\mathcal{T}$ is the set of all $N \times K$ precoding matrices in the codebook and in the case of TAS, it is the set of all possible $N \times K$ submatrices of $\mathbb{I}_N$ formed by picking K out of N (distinct) columns. Let $\mathcal{T} = \{\mathbf{T}_1, ..., \mathbf{T}_{|\mathcal{T}|}\}$. We assume that the precoding matrices are

[3]It is well known that transmit antenna selection without CSI at the transmitter (CSIT) can be interpreted as a type of LFP [2]. However with the presence of CSIT, this is not true.

[4]By coupling, we mean that the Gaussian matrices generating the set of Wishart matrices have some common elements.

semi-unitary i.e. $\mathbf{T}_i^\dagger \mathbf{T}_i = \mathbb{I}_K$ for all $i \in \{1,..,|\mathcal{T}|\}$[5]. We further assume that $\mathbf{H}$ is quasi-static and therefore the system capacity is computed for each channel realization as:

$$C(\mathbf{H}) = \max_{\mathbf{P}, 1\le i\le|\mathcal{T}|} \log\left|\mathbb{I}_M + \rho \mathbf{H}\mathbf{T}_i\mathbf{P}\mathbf{T}_i^\dagger\mathbf{H}^\dagger\right| \quad \text{s.t} \quad \mathrm{Tr}\{\mathbf{P}\} \le 1 \tag{2}$$

where $\mathbf{P} = \mathbb{E}\{\mathbf{x}\mathbf{x}^\dagger\}$ is the transmit power allocation matrix. In unitary LFP, since the transmitter does not have CSI, typically equal power allocation is used i.e. $\mathbf{P} = \frac{1}{K}\mathbb{I}_K$.[6] In the case of transmit antenna selection with CSI however, the capacity optimal water filling power allocation can be used. This is the major difference between LFP and TAS with CSIT. Though the results can also be extended to the case of water-filling with a little effort, for the case of TAS we assume that equal power is allocated to all the non-zero channel eigenvalues. This scheme is capacity optimal in the high SNR regime [26]. Furthermore, since with antenna selection the best precoder is likely to yield less skewed channel eigenvalues, the capacity loss due to equal power allocation is small even at low SNR. Under these assumptions, the capacity expressions[7] in either case can be expressed as:

$$C_{\mathrm{LFP}}(\mathbf{H}) = \max_{1\le i\le|\mathcal{T}|}\left\{\sum_{m=1}^{M}\log(1+\frac{\rho}{K}\tilde{\lambda}_{im})\right\} \tag{3}$$

$$C_{\mathrm{TAS}}(\mathbf{H}) = \max_{1\le i\le|\mathcal{T}|}\left\{\sum_{m=1}^{M}\log\left[1+\frac{\rho\tilde{\lambda}_{im}}{\mathrm{rank}\{\mathbf{H}\mathbf{T}_i\}}\right]\right\}$$
$$\approx \max_{1\le i\le|\mathcal{T}|}\left\{\sum_{m=1}^{M}\log\left[1+\frac{\rho\tilde{\lambda}_{im}}{M}\right]\right\} \tag{4}$$

where $\tilde{\lambda}_{im}$ is the m-th largest eigenvalue of $\mathbf{H}\mathbf{T}_i\mathbf{T}_i^\dagger\mathbf{H}^\dagger$ and the last step follows from the fact that $M \le K$ and the best channel is rank deficient with very low probability. Since the effective channel for a given precoder matrix $\mathbf{T}$ is $\mathbf{H}\mathbf{T}$, we shall henceforth refer to $\tilde{\lambda}_{im}$ as the m-th "channel" eigenvalue for precoder $\mathbf{T}_i$.

III. Joint distribution of channel eigenvalues

From (3) and (4) it is clear that the capacity distribution depends only on the eigenvalues of $\mathbf{H}\mathbf{T}_i\mathbf{T}_i^\dagger\mathbf{H}^\dagger$. It can be easily verified from (1) that $\{\mathbf{H}\mathbf{T}_i\mathbf{T}_i^\dagger\mathbf{H}^\dagger | 1 \le i \le |\mathcal{T}|\}$ forms a set of coupled, doubly-correlated Wishart matrices. The coupling comes from the fact that all these matrices are generated from the same i.i.d. random matrix $\mathbf{G}$. In this section, we characterize the joint distribution of the eigenvalues of these coupled Wishart matrices. We first derive the asymptotic second-order statistics and the joint distribution of the eigenvalues in the large antenna limit i.e., for $N, K \to \infty$ (with a fixed ratio) while M is kept fixed (finite). Note that this is counter intuitive in the case of LFP since there we typically have $K \le M$. However, this scaling is required for analytical tractability and where necessary, we shall also consider approximations for the, more practical, finite antenna regime (including the case of $K = M$).

For the large antenna limit, we define the scaled parameters $N = sN_o$ and $K = sK_o$, where N_o, K_o are constants and s is the scaling factor. We define a family of $N \times N$ transmit correlation matrices $\mathbf{R_{tx}}$ and a family of codebooks $\mathcal{T} = \{\mathbf{T}_1,..,\mathbf{T}_{|\mathcal{T}|}\}$ as a function of s. For the family of codebooks, the codebook size $|\mathcal{T}|$ is fixed but the precoding matrices $\mathbf{T}_i$ are $N \times K$ semi-unitary matrices as a function of s. The eigenvalues of $\mathbf{H}\mathbf{T}_i\mathbf{T}_i^\dagger\mathbf{H}^\dagger$ typically diverge as s increases. Therefore we shall instead characterize the eigenvalues of its normalized counterpart $\mathbf{Q}_i \triangleq \frac{\mathbf{H}\mathbf{T}_i\mathbf{T}_i^\dagger\mathbf{H}^\dagger}{\mathrm{Tr}\{\mathbf{T}_i^\dagger\mathbf{R_{tx}}\mathbf{T}_i\}}$.[8] We define the eigen decomposition $\mathbf{Q}_i = \mathbf{E}_i\Lambda_i\mathbf{E}_i^\dagger$ where $\mathbf{E}_i$ and Λ_i are the *unordered* eigenvector and eigenvalue matrices, respectively. We shall refer to these eigenvalues $\lambda_{im} = [\Lambda_i]_{m,m}$ as the normalized channel eigenvalues. We also define the eigen decompositions $\mathbf{R_{rx}} = \mathbf{E^{rx}}\Lambda^{\mathbf{rx}}[\mathbf{E^{rx}}]^\dagger$ and $\mathbf{R_{tx}} = \mathbf{E^{tx}}\Lambda^{\mathbf{tx}}[\mathbf{E^{tx}}]^\dagger$ where, $\lambda_k^{\mathbf{tx}} = [\Lambda^{\mathbf{tx}}]_{kk}$, $\lambda_k^{\mathbf{rx}} = [\Lambda^{\mathbf{tx}}]_{kk}$ are the k-th largest eigenvalues of $\mathbf{R_{tx}}$, $\mathbf{R_{rx}}$ respectively.

A. First-order approximation and second-order statistics

The expression for the normalized channel eigenvalues and their second-order statistics, in the large antenna limit, are given by the following theorem, which extends the results in [20] to the joint statistics case:

Theorem III.1. *Consider a family of transmit correlation matrices $\mathbf{R_{tx}}$ and a family of precoding matrices $\mathcal{T} = \{\mathbf{T}_1,..,\mathbf{T}_{|\mathcal{T}|}\}$ as a function of s. If the eigenvalues of $\mathbf{R_{rx}}$ are all distinct and $\lim_{s\to\infty}\frac{\|\mathbf{T}_i^\dagger\mathbf{R_{tx}}\mathbf{T}_i\|_{\mathbf{F}}}{\mathrm{Tr}\{\mathbf{T}_i^\dagger\mathbf{R_{tx}}\mathbf{T}_i\}} = 0$ for all $i \in \{1,..,|\mathcal{T}|\}$, then as $s \to \infty$:*

$$\lambda_{im} \simeq \mathbf{e}_m^{\mathbf{rx}\dagger}\mathbf{Q}_i\mathbf{e}_m^{\mathbf{rx}} \triangleq \dot{\lambda}_{im} \tag{5}$$

$$\mu_{im} = \mathbb{E}\{\lambda_{im}\} \simeq \lambda_m^{\mathbf{rx}} \triangleq \dot{\mu}_{im} \tag{6}$$

$$K_{mn}^{ij} = \mathbb{E}\left\{\lambda_{im}\lambda_{jn}\right\} - \mu_{im}\mu_{jn}$$
$$\simeq \frac{\delta_{mn}\lambda_m^{\mathbf{rx}}\lambda_n^{\mathbf{rx}}\left\|\mathbf{T}_j^\dagger\mathbf{R_{tx}}\mathbf{T}_i\right\|_{\mathbf{F}}^2}{\mathrm{Tr}\{\mathbf{T}_i^\dagger\mathbf{R_{tx}}\mathbf{T}_i\}\mathrm{Tr}\{\mathbf{T}_j^\dagger\mathbf{R_{tx}}\mathbf{T}_j\}} \triangleq \dot{K}_{mn}^{ij} \tag{7}$$

for all $1 \le m,n \le M$, $i,j \in \{1,..,|\mathcal{T}|\}$ and where we use "$\simeq$" to denote a first-order approximation (i.e., an equality in which the higher order terms that do not influence the asymptotic statistics of λ_{im} as $s \to \infty$ are neglected), $\lambda_{im} = [\Lambda_i]_{m,m}$ and $\mathbf{e}_m^{\mathbf{rx}} = [\mathbf{E^{rx}}]_{c\{m\}}$.

Proof. See Appendix A. □

These normalized channel eigenvalues $[\lambda_{i1},..,\lambda_{iM}]$ are unordered and are picked in the permutation such that the Hoffman-Weilandt inequality holds (see proof of Theorem III.1). Note that the conditions required for the above theorem

[5]This assumption is valid for TAS and also holds for the most common case of LFP called limited feedback unitary precoding.

[6]Note that with availability of second-order CSI at transmitter, statistical power allocation can also be used to improve performance slightly. Also, in the most general non-unitary LFP setting, the codebook can be augmented with additional feedback bits to convey power allocation information. Though such non-unitary precoding is beyond the scope of this work, the analysis presented here can also be extended to these scenarios with little effort.

[7]Due to the use of sub-optimal power allocation, strictly speaking, these expressions correspond to "achievable data-rate". However in this work, with a slight abuse of notation, we shall refer to them as capacity.

[8]Note that if $\mathrm{Tr}\{\mathbf{T}_i^\dagger\mathbf{R_{tx}}\mathbf{T}_i\} = 0$, the corresponding channel eigenvalues are trivially zero. Here we only consider the non-trivial case of $\mathrm{Tr}\{\mathbf{T}_i^\dagger\mathbf{R_{tx}}\mathbf{T}_i\} > 0$.

are somewhat difficult to verify since they depend on th codebook. A simpler sufficient condition, independent of th codebook, is given by the following proposition.

Proposition III.1.1 (Simpler sufficient condition)**.** *Theorer III.1 is satisfied if eigenvalues of* $\mathbf{R}_{\mathbf{rx}}$ *are all distinct and eithe*

$$\lim_{s\to\infty} \frac{\sum_{k=1}^{K} (\lambda_k^{\mathbf{tx}})^2}{\left[\sum_{\ell=1}^{K} \lambda_{N+1-\ell}^{\mathbf{tx}}\right]^2} = 0 \ \textit{or} \ \lim_{s\to\infty} \frac{(\lambda_1^{\mathbf{tx}})^2}{\sum_{\ell=1}^{K} (\lambda_{N+1-\ell}^{\mathbf{tx}})^2} = 0$$

Proof. See Appendix B.

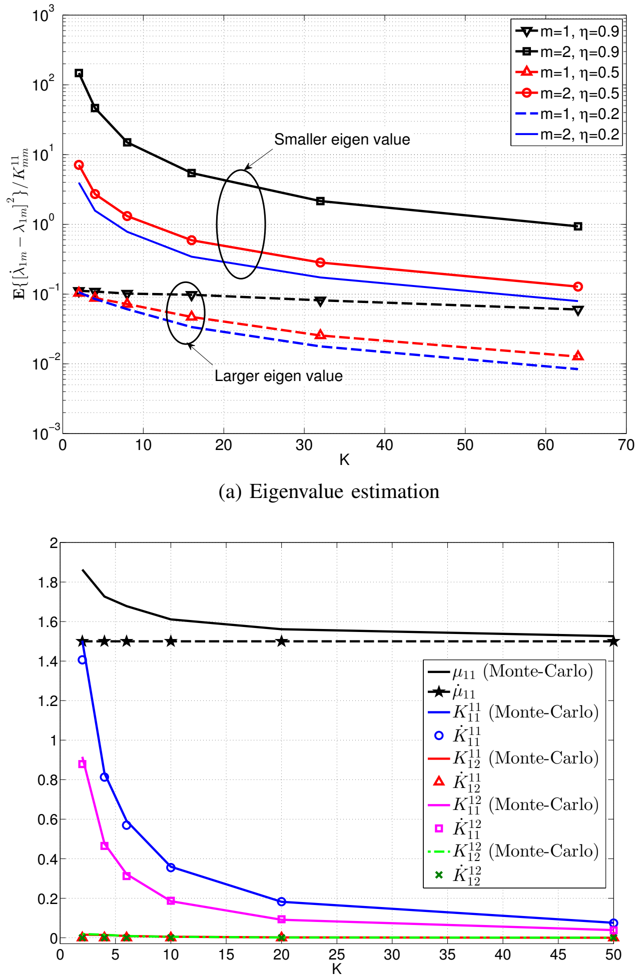
Intuitively, the theorem states that as long as the eigenvalue of the transmit correlation matrix are not too skewed (so tha the law of large numbers is applicable) the normalized channe eigenvalues asymptotically converge. Therefore, the first-orde approximations to the normalized channel eigenvalues ar valid for large s, and these are used to derive the second-orde statistics. Some examples of families of transmit correlation matrices which satisfy the skewness constraints in Proposition III.1.1 are discussed in Appendix D.

Though the presented results are asymptotic, we are interested in how quickly the terms in (5)-(7) converge to their first-order approximations as a function of s. It is worth mentioning that for the special case of a single receive antenna ($M = 1$), (5)-(7) are *exact* for *all* values of s. For larger values of M, a comparison of the convergence speeds of the first-order approximations to results from Monte-Carlo simulations are studied in Fig. 1 for a sample restricted precoded system. Here, the unordered eigenvalues of Λ_i are being compared with the ordered eigenvalues obtained from Monte-Carlo simulations. Such a comparison is reasonable if the overlap between the marginal distributions of the unordered eigenvalues is low i.e., if the eigenvalues of $R_{\mathbf{rx}}$ are sufficiently well separated and if $s \gg 1$. A comparison of the convergence of the asymptotic first-order expression (5) to Monte-Carlo simulation results is studied in Fig. 1a. It shows that while the convergence of (5) is very quick with K for large eigenvalues it is slower for the smaller eigenvalues. A comparison of the approximate second-order eigen statistics to Monte-Carlo simulations as a function of s is presented in Fig. 1b. It shows that the second-order statistics match even for $K = 2$, validating quick convergence. Similar results have been observed for a wide variety of system parameters. The seemingly slow convergence of $\dot{\lambda}_{im}$ for small eigenvalues in Fig. 1a and μ_{im} in Fig. 1b is a result of the comparison of ordered with unordered eigenvalues.

Approximation III.1. *Due to the accuracy and quick convergence of the first-order approximations, we will use the* $\dot{X}$*'s in place of the* X*'s in* (5)-(7)*, even in the non-asymptotic regime, i.e., for finite values of* s*.*

B. Joint Distribution of eigenvalues

In this section we find the joint distribution of the normalized channel eigenvalues. Since the actual distribution is hard to characterize, we first derive the asymptotic joint distribution and later consider approximations for finite values of K. The following theorem gives us partial results on the asymptotic joint distribution of eigenvalues:



(a) Eigenvalue estimation

(b) Eigenvalue statistics

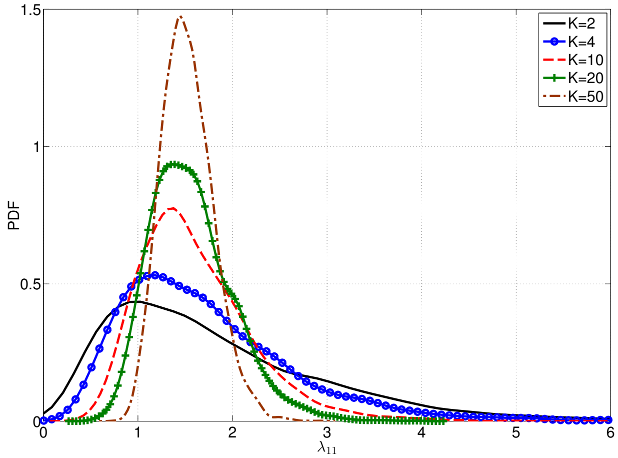
Fig. 1. Convergence of normalized channel eigenvalues to their first-order approximations for a restricted precoded system, as a function of K: (a) Compares the mean square error $\mathbb{E}\{[\lambda_{im} - \dot{\lambda}_{im}]^2\}$ normalized by K_{mm}^{11}, as a function of K for different values of transmit correlation η (b) Compares $\mu_{im}, K_{mn}^{ij}, \dot{\mu}_{im}, \dot{K}_{mn}^{ij}$ as a function of K for $\eta = 0.5$ (system parameters: $N = 2K$, $M = 2$, $[\mathbf{R}_{\mathbf{tx}}]_{ab} = \eta^{|a-b|}$, $[\mathbf{R}_{\mathbf{rx}}]_{ab} = (0.5)^{|a-b|}$, $\mathbf{T}_1 = [\mathbb{I}_N]_{\mathbf{c}\{1:K\}}$ and $\mathbf{T}_2 = [\mathbb{I}_N]_{\mathbf{c}\left\{1+\left\lfloor \frac{K}{2} \right\rfloor : \left\lfloor \frac{3K}{2} \right\rfloor\right\}}$)

Theorem III.2. *Consider a family of transmit correlation matrices* $\mathbf{R}_{\mathbf{tx}}$ *and a family of precoding matrices* $\mathcal{T} = \{\mathbf{T}_1, .., \mathbf{T}_{|\mathcal{T}|}\}$ *as a function of* s*. Then the vector of eigenvalues* $\mathbf{v} = [\lambda_{i_1 m_1}, \lambda_{i_2 m_2}, ..., \lambda_{i_L m_L}]$ *for any finite* L*,* $1 \leq i_1, ..., i_L \leq |\mathcal{T}|$ *and* $1 \leq m_1, ..., m_L \leq M$ *are jointly Gaussian distributed as* $s \to \infty$*, with second-order statistics as given in Theorem III.1, if:*

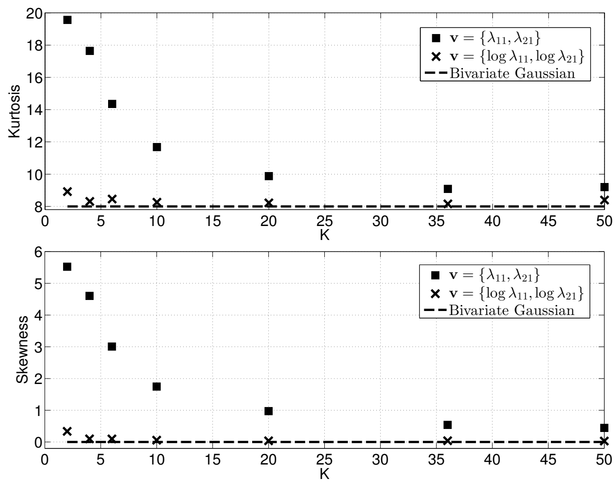
1) *The eigenvalues of* $\mathbf{R}_{\mathbf{rx}}$ *are all distinct.*
2) $\mathbf{T}_{i_\ell} = \mathbf{U}_{i_\ell} \otimes \mathbf{V}$ *for all* $1 \leq \ell \leq L$*, where* $\otimes$ *defines the Kronecker product,* $\mathbf{V}$ *is a* $s \times s$ *unitary matrix and* $\mathbf{U}_{i_\ell}$ *is any fixed* $N_o \times K_o$ *semi-unitary matrix.*
3) *The transmit-correlation matrix eigenvalues satisfy* $\lim_{s\to\infty} \frac{(\lambda_1^{\mathbf{tx}})^2}{\sum_{k=1}^{s} (\lambda_{N+1-k}^{\mathbf{tx}})^2} = 0.$

Proof. See Appendix C. □

Some examples of families of transmit correlation matrices



(a) Marginal PDF



(b) Skewness-Kurtosis

Fig. 2. Asymptotic convergence of the distribution of normalized channel eigen-values for a restricted precoded system, as a function K: (a) Plots the empirical probability distribution of the eigenvalue λ_{11} (b) Compares the kurtosis and skewness of the bi-variate random vectors $\mathbf{v} = \{\lambda_{11}, \lambda_{21}\}$ and $\mathbf{v} = \{\log \lambda_{11}, \log \lambda_{21}\}$ to a bi-variate Gaussian with same second-order statistics (system parameters: $N = 2K, M = 2, [\mathbf{R_{tx}}]_{ab} = (0.5)^{|a-b|}, [\mathbf{R_{rx}}]_{ab} = (0.5)^{|a-b|}, \mathcal{T} = \{\mathbf{T}_1, \mathbf{T}_2\}$ where $\mathbf{T}_1 = [\mathbb{I}_N]_{\mathbf{c}\{1:K\}}$, $\mathbf{T}_2 = [\mathbb{I}_N]_{\mathbf{c}\left\{1+\lfloor\frac{K}{2}\rfloor:\lfloor\frac{3K}{2}\rfloor\right\}}$ and 5000 samples)

which satisfy the skewness constraint (condition 3) are discussed in Appendix D. Intuitively, the theorem states that if the eigenvalues of the transmit correlation matrix are not too skewed (so that Lyapunov's central limit theorem is applicable) then the normalized channel eigenvalues corresponding to the precoding matrices that are sufficiently well separated in their column space are asymptotically jointly Gaussian. To characterize this convergence, the empirical distribution of the normalized channel eigenvalues for a sample restricted precoded system is plotted in Fig. 2a for different values of K. Following the approach in [27], to test the joint normality, the Kurtosis and Skewness of a vector of eigenvalues $\mathbf{v}$ (corresponding to a well-separated codebook) are plotted in Fig. 2b as a function of K. From [27], for a large sample set from a p-variate Gaussian distribution, the skewness and kurtosis converge to the values of 0 and $p(p+2)$, respectively. Both figures suggest that though the joint distribution is asymptotically Gaussian, the convergence is very slow. Such large values of K may be impractical and therefore, other approximations to the joint distribution are required in the finite antenna regime.

In this paper, we propose a joint lognormal distribution as an approximation for the normalized channel eigenvalue distribution in the finite antenna regime. We observe from Fig. 2a that in the non-asymptotic regime, a lognormal distribution may indeed be a better fit for the marginal distribution. In Fig. 2b, the kurtosis and skewness for the logarithm of eigenvalues are also depicted. The quick convergence of these parameters with K provides further credence to this hypothesis. Apart from ensuring that the eigenvalues are always non-negative, a joint-lognormal approximation for eigenvalues also ensures that the capacity is Gaussian distributed in the high SNR regime. This is an intuitively pleasing result and is consistent with prior literature [28], [29].

Approximation III.2. *In the rest of the paper, we shall approximate the normalized channel eigenvalues for any set of precoding matrices to be jointly lognormally distributed in the non-asymptotic regime.*

Unlike in Theorem III.2, which considers only precoding matrices that are sufficiently well separated, here we approximate the eigenvalues corresponding to any set of precoding matrices to be jointly lognormal distributed. The validity of this approximation is studied in Fig. 3 wherein the Skewness and Kurtosis for the eigenvalues corresponding to all precoding matrices of a sample antenna selection system are plotted. These results also show that a jointly lognormal distribution is a better fit than a Gaussian fit for the normalized channel eigenvalues. However, even for the logarithm of eigenvalues, the Skewness and Kurtosis values deviate partially from those of a Gaussian distribution, thereby suggesting that Approx. III.2 is not very accurate. However, this approximation is needed for analytical tractability. In Sec. IV, it is demonstrated that the resulting approximation error in estimating the channel capacity is relatively small.

IV. Capacity Analysis

The individual channel capacities for $i = 1, .., |\mathcal{T}|$ can be expressed in the form $C_i(\alpha_i, \mathbf{H}) \triangleq \sum_{m=1}^{M} \log(1 + \alpha_i \lambda_{im})$ where the α_is are suitably chosen constants. In particular, inspection of (3)-(4) and the definition of normalized channel eigenvalues (see Section III) reveals that for LFP $\alpha_i = \frac{\rho \mathrm{Tr}\{\mathbf{T}_i^\dagger \mathbf{R_{tx}} \mathbf{T}_i\}}{K}$ while for TAS, $\alpha_i = \frac{\rho \mathrm{Tr}\{\mathbf{T}_i^\dagger \mathbf{R_{tx}} \mathbf{T}_i\}}{M}$. For sufficiently large values of α_i, we can approximate:

$$C_i(\alpha_i, \mathbf{H}) \approx \sum_{m=1}^{M} \log(\alpha_i \lambda_{im}) \tag{8}$$

Now, from Approx. III.2, for moderately large values of α_i we have:

$$\{C_1(\alpha_1, \mathbf{H}), \ldots, C_{|\mathcal{T}|}(\alpha_{|\mathcal{T}|}, \mathbf{H}))\} \sim \text{Jointly Gaussian} \tag{9}$$

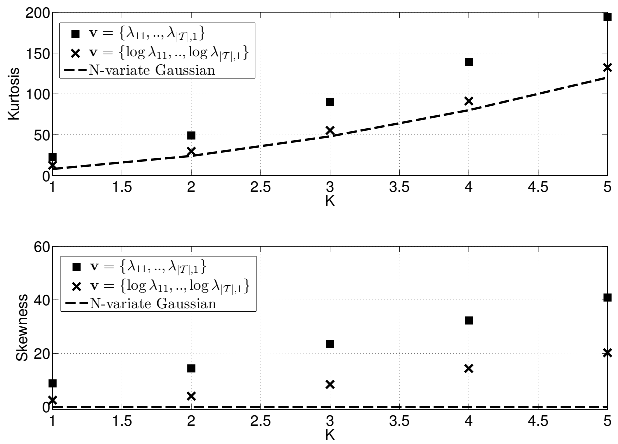


Fig. 3. Comparison of Skewness and Kurtosis of the set of normalized channel eigenvalues $\mathbf{v} = \{\lambda_{11}, .., \lambda_{|\mathcal{T}|,1}\}$ and their logarithms $\mathbf{v} = \{\log \lambda_{11}, .., \log \lambda_{|\mathcal{T}|,1}\}$ to a Gaussian distribution with same second-order statistics, in an antenna selection system (system parameters: $N = 2K, M = 1, [\mathbf{R_{tx}}]_{ab} = (0.5)^{|a-b|}, [\mathbf{R_{rx}}]_{ab} = (0.5)^{|a-b|}$, $\mathcal{T} = \{\mathbf{T}|\mathbf{T}$ is a $N \times K$ submatrx of $\mathbb{I}_N\}$ and 10000 samples) [9]

Additionally, using Approx. III.1, (8) and results on the second-order statistics of a lognormal random vector [30], we can easily show that:

$$\bar{C}_i \triangleq \mathbb{E}\{C_i(\alpha_i, \mathbf{H})\} \approx \sum_{m=1}^{M} \log\left[\frac{\alpha_i \dot{\mu}_{im}^2}{\sqrt{\dot{\mu}_{im}^2 + \dot{K}_{mm}^{ii}}}\right] \quad (10)$$

$$\begin{aligned} \kappa_{ij} &\triangleq \mathbb{E}\{C_i(\alpha_i, \mathbf{H})C_j(\alpha_j, \mathbf{H})\} \\ &\quad -\mathbb{E}\{C_i(\alpha_i, \mathbf{H})\}\mathbb{E}\{C_j(\alpha_j, \mathbf{H})\} \\ &\approx \sum_{m,n} \log\left[\frac{\dot{\mu}_{im}\dot{\mu}_{jn} + \dot{K}_{mn}^{ij}}{\dot{\mu}_{im}\dot{\mu}_{jn}}\right] \\ &= \sum_{m=1}^{M} \log\left[\frac{\dot{\mu}_{im}\dot{\mu}_{jm} + \dot{K}_{mm}^{ij}}{\dot{\mu}_{im}\dot{\mu}_{jm}}\right] \end{aligned} \quad (11)$$

A comparison of the approximate joint statistics and marginal distribution of $C_i(\alpha_i, \mathbf{H})$ to Monte-Carlo simulations for a sample antenna selection system is given in Fig. 4, as a function of K. The results show that the approximations are tight for $K \geq 4$. In Fig. 4c, the skewness and kurtosis of the vector of individual capacities corresponding to all precoding matrices for the antenna selection system are studied. The close fit to a Gaussian distribution suggests that the impact of Approx. III.2 on capacity is small. Similar results have been observed for a wide variety of system parameters.

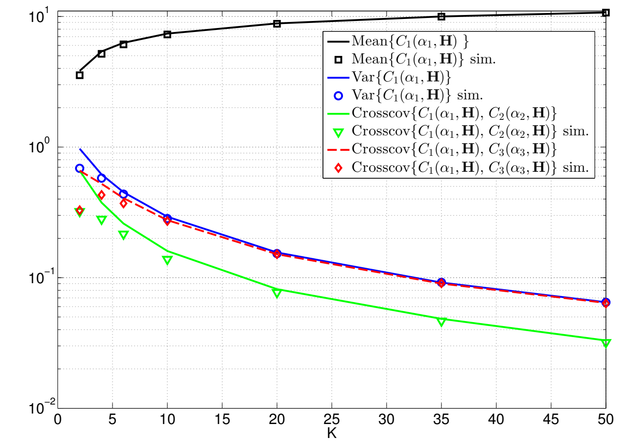
A. System capacity

A general abstraction of the system capacity expressions (3)–(4) can be given by:

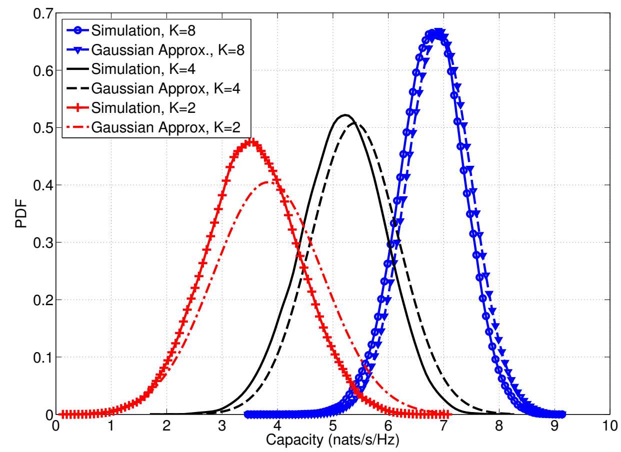
$$C_{\max}(\mathbf{H}) = \max_{1 \leq i \leq |\mathcal{T}|} \{C_i(\alpha_i, \mathbf{H})\} \quad (12)$$

Note that (9)–(11) provide a model that fully characterizes the joint distribution of the individual capacities. From (9),

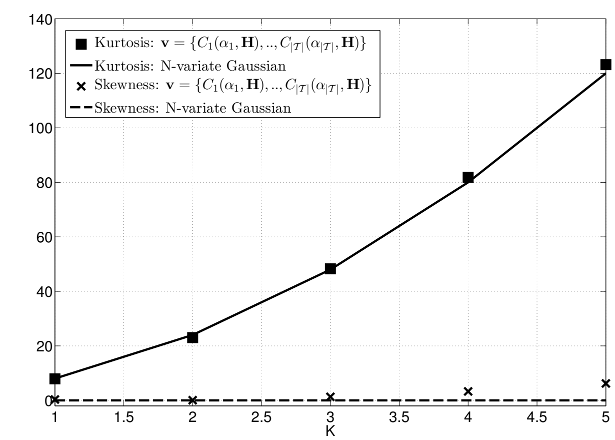
[9] Though $\mathbf{v}$ is a $\binom{N}{K}$ random vector, the Kurtosis and Skewness are computed only for the dominant subspace, which has N principal components.



(a) Joint statistics of $C_i(\alpha_i, \mathbf{H})$



(b) PDF of $C_1(\alpha_1, \mathbf{H})$



(c) Skewness-Kurtosis

Fig. 4. Individual capacity distribution of an antenna selection system: (a) Compares the joint second-order statistics of capacities across different precoding matrices as a function of K (b) Compares the empirical PDF of $C_1(\alpha_1, \mathbf{H})$ to a Gaussian distribution with mean and variance as given by (10)-(11) (c) Compares skewness and kurtosis of the set of channel capacities $\mathbf{v} = \{C_1(\alpha_1, \mathbf{H}), .., C_{\mathcal{T}}(\alpha_{|\mathcal{T}|}, \mathbf{H})\}$ to a Gaussian distribution with same second-order statistics (system parameters: $N = 2K, M = 2, \rho = 10, [\mathbf{R_{tx}}]_{ab} = (0.5)^{|a-b|}, [\mathbf{R_{rx}}]_{ab} = (0.5)^{|a-b|}$, $\alpha_i = \frac{\rho \mathrm{Tr}\{\mathbf{T}_i^\dagger \mathbf{R_{tx}} \mathbf{T}_i\}}{M}$, $\mathbf{T}_1 = [\mathbb{I}_N]_{\mathbf{c}\{1:K\}}$, $\mathbf{T}_2 = [\mathbb{I}_N]_{\mathbf{c}\left\{1+\left\lfloor \frac{K}{2} \right\rfloor : \left\lfloor \frac{3K}{2} \right\rfloor\right\}}$, $\mathbf{T}_3 = [\mathbb{I}_N]_{\mathbf{c}\{2:(K+1)\}}$, $\mathcal{T} = \{\mathbf{T}|\mathbf{T}$ is a $N \times K$ submatrx of $\mathbb{I}_N\}$ and 10000 samples) [9]

$C_{\max}(\mathbf{H})$ can be expressed as the largest element of a correlated Gaussian vector.

For the maximum of a set of correlated Gaussian random variables, neither the exact distribution nor even the mean is not known in closed form. Existing methods [31] to solve for them are too cumbersome, especially when $|\mathcal{T}|$ is large. Though several bounds exist on the mean [32]–[36], they are not uniformly tight across all correlation structures. On the other hand, numerical approaches like in [37], [38] are recursive and therefore are likely to accumulate significant amount of error when the number of variables $|\mathcal{T}|$ are large. This is specifically relevant to our scenario since the codebook size $|\mathcal{T}|$ can be very large. We therefore formulate a new approach to compute the distribution and mean of the largest element of a correlated Gaussian vector. Note that:

$$\begin{aligned} C_{\max}(\mathbf{H}) &= \left\| [C_1(\alpha_1,\mathbf{H}),...,C_{|\mathcal{T}|}(\alpha_{|\mathcal{T}|},\mathbf{H})] \right\|_\infty \\ &= \log \left\| [e^{C_1(\alpha_1,\mathbf{H})},...,e^{C_{|\mathcal{T}|}(\alpha_{|\mathcal{T}|},\mathbf{H})}] \right\|_\infty \\ &\approx \log \left\| [e^{C_1(\alpha_1,\mathbf{H})},...,e^{C_{|\mathcal{T}|}(\alpha_{|\mathcal{T}|},\mathbf{H})}] \right\|_p \\ &\quad \text{for } p \geq \log|\mathcal{T}| \\ &= \frac{1}{p}\log\left[\sum_{i=1}^{|\mathcal{T}|} e^{pC_i(\alpha_i,\mathbf{H})}\right] \end{aligned} \quad (13)$$

where the second last step follows from the norm inequalities $L^{-1/p}\|\mathbf{a}\|_p \leq \|\mathbf{a}\|_\infty \leq \|\mathbf{a}\|_p$ for any vector $\mathbf{a}$ of length L. Since $e^{pC_i(\alpha_i,\mathbf{H})} \approx \left(\prod_{m=1}^{M}\alpha_i\lambda_{im}\right)^p$ is lognormal distributed, equation (13) above shows that the largest element of a correlated Gaussian vector can be approximately represented as the logarithm of a sum of correlated lognormals.

B. Sum of correlated lognormals

It is well known that the sum of correlated lognormal random variables is approximately lognormal (see [39] and references therein). Of the many approximations for characterizing this sum, the moment and cumulant matching approaches, such as [40], [41], yield a poor fit in the lower tail regions of the distribution. On the other hand, the moment generating function matching approaches, like [42], are too cumbersome when the number of variables $|\mathcal{T}|$ is large. Here, we propose the use of the approach in [43] (reproduced here as algorithm 1), which extends the work in [44] to the correlated case. Similar to [37], this algorithm is recursive and therefore also shares the same drawback of accumulating error. However, the drawback is a by-product of the algorithm in [43] and not of our approach in (13). Any new results on sum of correlated lognormals can readily be used to resolve this drawback. To check the goodness of fit, the empirical distribution of the largest element of a sample Gaussian vector, and its p-norm approximation are compared to the distributions obtained using Algorithm 1, Clark [38] and second-order moment-matching [40] in Fig. 5. The results show that both Clark as well as Algorithm 1 give good approximations to the distribution of the largest element.

In summary, the system capacity $C_{\max}(\mathbf{H})$ can be approximated as a Gaussian random variable and its mean and variance can be computed via Algorithm 1.

Algorithm 1: Statistics of the largest element of a correlated Gaussian vector

Inputs: p, $\bar{C}_i, \kappa_{ij}$ forall $1 \leq i,j \leq |\mathcal{T}|$ // Defined as in (10)-(13)
$\mu_w(1) = \mu_s(1) = p\bar{C}_1$
$\sigma_w^2(1) = \sigma_s^2(1) = p^2\kappa_{11}$
$Q(1,*) = p^2\kappa_{1*}$
for $i = 2$ to $|\mathcal{T}|$ **do**
  $\mu_w(i) = p\bar{C}_i - \mu_s(i-1)$
  $\sigma_w^2(i) = p^2\kappa_{ii} + \sigma_s^2(i-1) - 2Q(i-1,i)$
  $\mu_s(i) = \mu_s(i-1) + G_1(\sigma_w(i), \mu_w(i))$
  $\sigma_s^2(i) = \sigma_s^2(i-1) - G_1(\sigma_w(i), \mu_w(i))$
    $+G_2(\sigma_w(i), \mu_w(i))$
    $+2\left[Q(i-1,i) - \frac{\sigma_s^2(i-1)G_3(\sigma_w(i),\mu_w(i))}{\sigma_w^2(i)}\right]$
  for $j = 1$ to $|\mathcal{T}|$ **do**
    $Q(i,j) = Q(i-1,j)\left[1 - \frac{G_3(\sigma_w(i),\mu_w(i))}{\sigma_w^2(i)}\right]$
      $+\frac{p^2\kappa_{ij}G_3(\sigma_w(i),\mu_w(i))}{\sigma_w^2(i)}$
  end for
end for
// $G_1(\sigma,\mu), G_2(\sigma,\mu), G_3(\sigma,\mu)$ are as defined in Appendix of [43]
return $\mu_s(|\mathcal{T}|)/p$ // Mean of $C_{\max}$
return $\sigma_s^2(|\mathcal{T}|)/p^2$ // Variance of $C_{\max}$

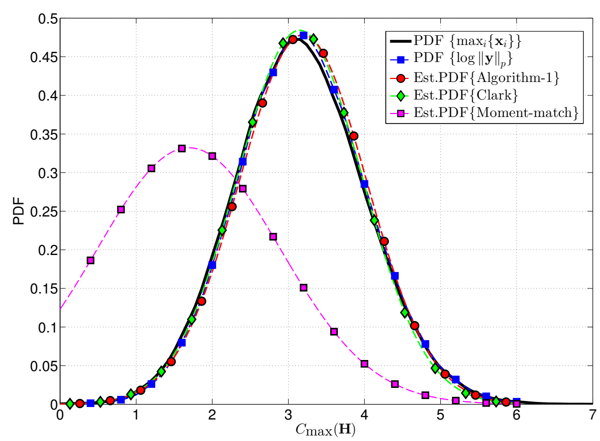
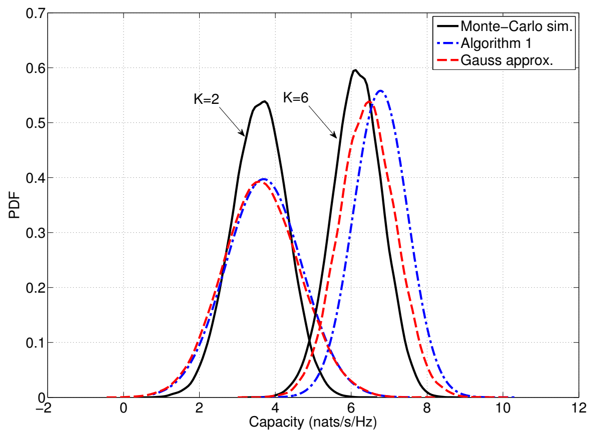


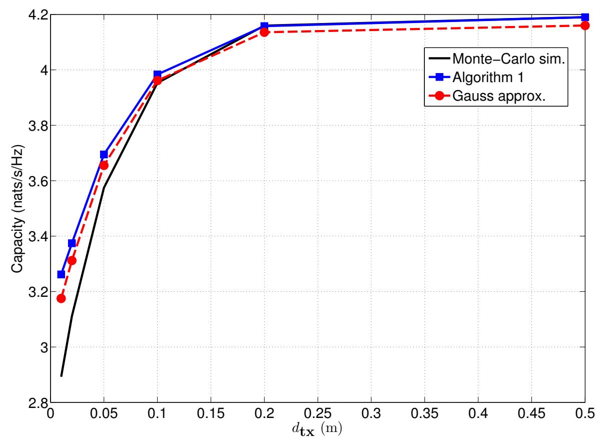
Fig. 5. Distribution of largest element of a Gaussian vector: $\mathbf{x}$-jointly Gaussian vector, $\mathbf{y} \triangleq \exp[\mathbf{x}]$ (element-wise); Clark, moment-match refer to solutions from [38] and [40], respectively (simulation parameters: $|\mathbf{X}| = 40$, $\mathbb{E}\{\mathbf{X}_i\} = 1$, $\mathbb{E}\{\mathbf{X}_i\mathbf{X}_j\} = 1.5 + \delta_{ij}$, $p = 8$)

V. Simulation Results

Using the results derived in the previous sections, we shall now analyze the system capacity of several practical restricted precoded systems. The simulation layout considers a single user, cellular downlink channel operating at 2.4 GHz. Both the transmitter and receiver have a uniform, linear antenna array with antenna spacings of $d_{\mathbf{tx}} = 5$cm and $d_{\mathbf{rx}} = 2$cm, respectively (unless otherwise stated). The transmitter experiences a Laplacian power angle spectrum (PAS) with mean angle of arrival (AoA) $= \pi/6$ rads and an angle spread (AS) of $\pi/10$ rads. The receiver on the other hand experiences a uniform power angle spectrum.[10] As a first simple example,



(a) PDF$\{C_{\max}(\mathbf{H})\}$

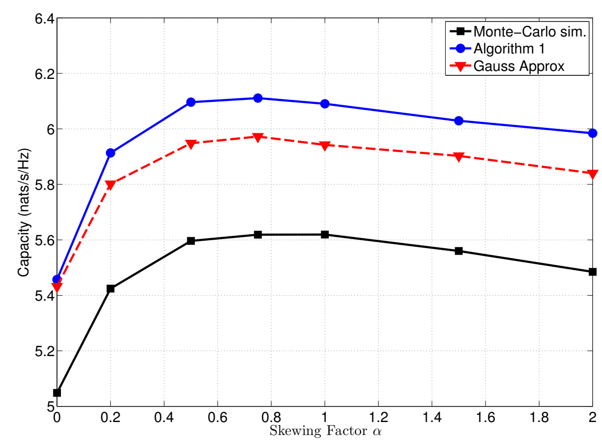


(b) Tx antenna spacing

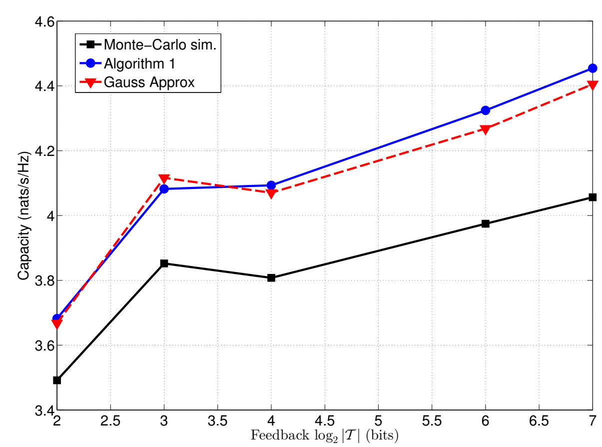
Fig. 6. Comparison of capacity of an antenna selection system as predicted by Algorithm1 to Monte-Carlo simulations and Gauss-approx (a) Plots PDF of system capacity for $K = 2, 6$ (b) Plots the mean capacity as a function of transmit antenna spacing ($K = 2$) (system parameters: $N = 2K$, $M = 2$, SNR $\rho = 10$)

we consider an antenna selection system in Fig. 6. In Fig. 6a, the PDF of capacity as computed by Algorithm 1 is compared to Monte-Carlo simulations. To quantify the origin of the mismatch in distributions, the PDF of the largest element of a Gaussian vector with the second-order statistics given by (10)-(11) is also plotted, labeled as Gauss-approx. The gap between Monte-Carlo and Gauss-approx quantifies the error due to inaccuracy of Approx. III.1 and III.2. We observe that this gap does not increase much with K. On the other hand, the gap between Gauss-approx and Algorithm 1 quantifies the error due to inaccuracy of the approach in [43]. This gap increases with K, owing to the error accumulation in the recursive steps of [43] for large codebooks $\left(|\mathcal{T}| = \binom{N}{K}\right)$. In Fig. 6b, the impact of transmit antenna spacing on ergodic capacity is compared. As seen from the results, though Algorithm 1 overestimates the capacity, it accurately reflects the impact of system parameters like antenna spacing on capacity.

For the same simulation layout, the impact of the codebook on ergodic capacity for a limited feedback precoding system is studied in Fig. 7. The influence of the codebook shape on capacity is studied in Fig. 7a, where we use skewed codebooks $\mathcal{T}_\alpha$ generated from a Grassmannian packed codebook $\hat{\mathcal{T}}$ as: $\mathcal{T}_\alpha = \left\{(\mathbf{R_{tx}})^\alpha \hat{\mathbf{T}}_i \left[\hat{\mathbf{T}}_i^\dagger (\mathbf{R_{tx}})^{2\alpha} \hat{\mathbf{T}}_i\right]^{-1/2} \middle| \hat{\mathbf{T}}_i \in \hat{\mathcal{T}}\right\}$. The skewing factor α controls the spacing between the precoding matrices of the codebook. The results suggest that Algorithm 1 gives a good estimate of dependence of capacity on the codebook shape. The impact of codebook size on ergodic capacity is studied in Fig. 7b. Here, we use Grasmannian codebooks of different sizes. We observe that Algorithm 1 gives accurate results for small codebooks but the error increases with codebook size $|\mathcal{T}|$. This is again due to the error accumulation in the recursive steps of Algorithm 1. The sudden dip at 4 bits of feedback is because the codebook is arbitrary and not customized to the chosen PAS.



(a) Skewing Factor



(b) Codebook Size

Fig. 7. Impact of the codebook on capacity of a limited feedback precoding system, as predicted by Algorithm1, Monte-Carlo simulations and Gauss-approx (a) Studies impact of the codebook skewing-factor $\left(N = 8, K = M = 2\right.$, SNR $\rho = 10$, $|\hat{\mathcal{T}}| = 256$, $\hat{\mathcal{T}}$ is a Grassmannian codebook from [45]$\left.\right)$ (b) Studies impact of codebook size $\left(N = 4, K = M = 2\right.$, SNR $\rho = 10$, $\mathcal{T}$ for each codebook size is from [45] $\left.\right)$

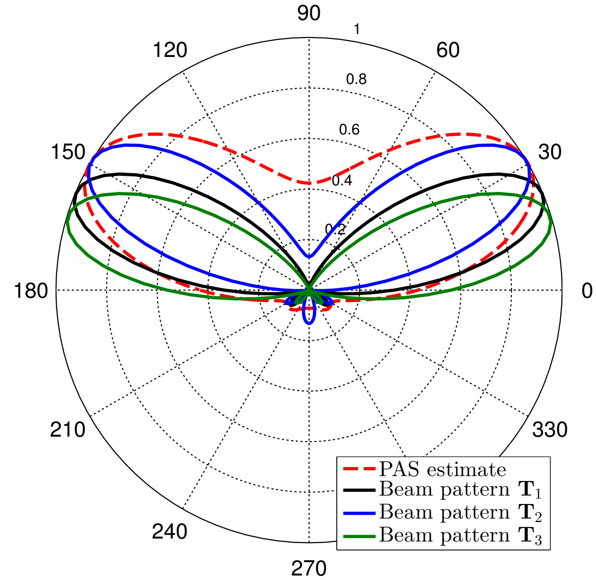
[10] Tx/Rx correlation matrix is calculated as: $[\mathbf{R_x}]_{ab} = \int_{-\pi}^{\pi} \mathrm{PAS}(\theta) e^{\frac{2\pi j(a-b)d_{\mathbf{x}} \sin\theta}{\lambda}} d\theta \Big/ \int_{-\pi}^{\pi} \mathrm{PAS}(\phi) d\phi$, where $j = \sqrt{-1}$, λ is the wavelength at 2.4GHz and **x=tx/rx**. All arrays and multipath components are in the horizontal plane.

VI. Application to codebook design for limited feedback systems

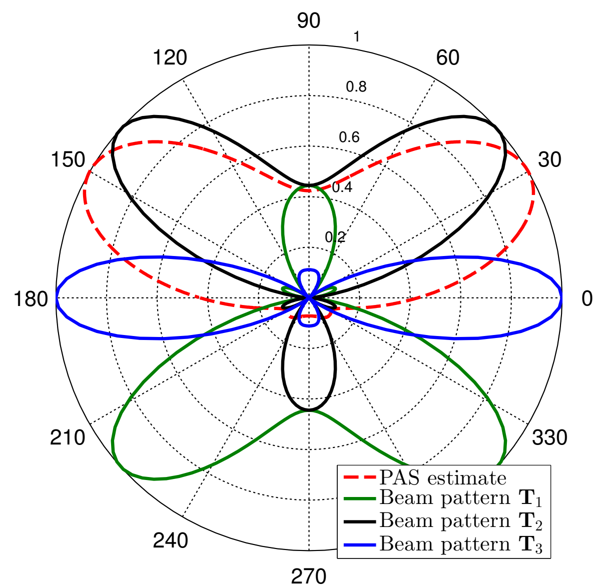
It is widely accepted that for an uncorrelated i.i.d. channel, a Grassmannian-packed codebook is near optimal for a limited feedback system [9], [11], [46]. However, for a correlated channel, no consensus exists on the best codebook design. This is partly because no uniformly tight (over the space of all codebooks) bounds or approximations to the system capacity are available for a correlated channel. In this section, using Algorithm 1 as the objective function, we shall use a numerical-gradient ascent algorithm to search for the codebook that maximizes the mean capacity. Here, the gradient of mean capacity (as predicted by Algorithm 1) with respect to a matrix $\mathbf{T}_{\text{CB}}$ (formed by appending all the precoders in the codebook $\mathcal{T}$) is computed numerically. For ease of pictorial representation, we consider a limited feedback beamforming setting with $N = 3$, $K = M = 1$ and other parameters same as in Sec. V. For a codebook size of $|\mathcal{T}| = 3$, we compare the beam-patterns formed by the precoding vectors of the optimized codebook to a Discrete Fourier Transform (DFT) codebook[12] in Fig. 8. The estimated PAS[11] is also plotted for comparison. As the results show, unlike the DFT codebook that is designed for generic correlated channels, the optimized codebook here adapts to the user PAS leading to an increase in capacity from 2.74 to 2.9 nats/s/Hz. Though several other families of codebooks have been proposed in literature [8], [12]–[14] for a correlated channel, they involve some parameters which need to be chosen. Algorithm 1 is also useful in such scenarios since it enables picking the mean/outage capacity maximizing values for these parameters.

VII. Conclusions

This paper analyzes a special class of MIMO systems called restricted precoded systems and discusses how many practically relevant systems like antenna selection, beam selection and limited feedback precoding fall under this class. It is shown that the system capacity of restricted precoded systems can be expressed as a function of the eigenvalues of a set of coupled doubly-correlated Wishart matrices. The eigenvalues are shown to be jointly Gaussian in the large antenna limit, if a set of conditions on the transmit correlation and codebook are satisfied. The asymptotic second-order statistics of the eigenvalues are also derived and the results suggest that their convergence is very quick. We propose, and verify, that in the finite antenna regime, a joint-lognormal distribution is a better fit to the eigenvalue distribution. Using these results, and a few simplifying approximations, we show that the system capacity for a restricted precoded system can be approximated as the largest element of a correlated Gaussian vector. We propose a new approach for characterizing its distribution and, in the process, show that the problem of finding the distribution of the sum of lognormals and the problem of finding the distribution of the largest element of a Gaussian vector are equivalent. Simulations results, for both an antenna selection system and a limited feedback precoded system, suggest that the proposed algorithm slightly overestimates the capacity but predicts the dependence of capacity on system parameters like number of antennas, antenna spacing and codebook shape & size accurately. We also observe that a significant portion of the mismatch comes from error accumulation in the recursive steps of the proposed algorithm. Any non-recursive approach to characterize the sum of lognormals can be useful in tackling this problem. We also demonstrate, via an example, that the



(a) Optimized codebook
$\mathbb{E}\{C_{\max}(\mathbf{H})\} = 2.91$nats/s/Hz



(b) DFT codebook
$\mathbb{E}\{C_{\max}(\mathbf{H})\} = 2.74$nats/s/Hz

Fig. 8. Comparison of beam-patterns for optimized codebook and the Grassmannian codebook (system parameters: $N = 3$, $K = M = 1$, SNR $\rho = 10$)[11]

[11] $\text{Est.PAS}(\theta) = \sum_{ab} [\mathbf{R}_{\text{tx}}]_{ab} e^{-\frac{2\pi j(a-b)d_{\text{tx}}\sin\theta}{\lambda}}$, where $j = \sqrt{-1}$ and λ is the wavelength at 2.4GHz.

[12] For $K = 1$, $|\mathcal{T}| = N$, the columns of a DFT matrix form a Grassmannian codebook that is well suited for correlated channels [6].

proposed algorithm can be used in the design of near-optimal codebooks for a limited feedback system.

APPENDIX A

(Proof of Theorem III.1). For any value of N, K and $\forall i \in \{1,..,|\mathcal{T}|\}$, from (1) we have:

$$\begin{aligned}\mathbf{Q}_i &= \frac{\mathbf{H}\mathbf{T}_i\mathbf{T}_i^\dagger\mathbf{H}^\dagger}{\mathrm{Tr}\{\mathbf{T}_i^\dagger\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}} = \frac{\mathbf{R}_{\mathbf{rx}}^{1/2}\mathbf{G}\mathbf{R}_{\mathbf{tx}}^{1/2}\mathbf{T}_i\mathbf{T}_i^\dagger\mathbf{R}_{\mathbf{tx}}^{1/2}\mathbf{G}^\dagger\mathbf{R}_{\mathbf{rx}}^{1/2}}{\mathrm{Tr}\{\mathbf{T}_i^\dagger\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}} \\ &= \mathbf{R}_{\mathbf{rx}}^{1/2}\left[\sum_{k=1}^{K}\hat{\mathbf{h}}_k^i[\hat{\mathbf{h}}_k^i]^\dagger\right]\mathbf{R}_{\mathbf{rx}}^{1/2} \qquad (14)\end{aligned}$$

where $\hat{\mathbf{h}}_k^i = \mathbf{G}\mathbf{R}_{\mathbf{tx}}^{1/2}[\mathbf{T}_i]_{\mathbf{c}\{k\}}/\sqrt{\mathrm{Tr}\{\mathbf{T}_i^\dagger\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}}$. Defining $\hat{\mathbf{Q}}_i = \sum_{k=1}^{K}\hat{\mathbf{h}}_k^i[\hat{\mathbf{h}}_k^i]^\dagger$ and taking expectations we get:

$$\begin{aligned}&\mathbb{E}\{[\hat{\mathbf{Q}}_i]_{ab}\} \\ &= \sum_k \frac{\mathbb{E}\{[\mathbf{G}]_{\mathbf{r}\{a\}}\mathbf{R}_{\mathbf{tx}}^{1/2}[\mathbf{T}_i]_{\mathbf{c}\{k\}}[\mathbf{T}_i]_{\mathbf{c}\{k\}}^\dagger\mathbf{R}_{\mathbf{tx}}^{1/2}\big([\mathbf{G}]_{\mathbf{r}\{b\}}\big)^\dagger\}}{\mathrm{Tr}\{\mathbf{T}_i^\dagger\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}} \\ &= \sum_k \frac{[\mathbf{T}_i]_{\mathbf{c}\{k\}}^\dagger\mathbf{R}_{\mathbf{tx}}^{1/2}\mathbb{E}\{\big([\mathbf{G}]_{\mathbf{r}\{b\}}\big)^\dagger[\mathbf{G}]_{\mathbf{r}\{a\}}\}\mathbf{R}_{\mathbf{tx}}^{1/2}[\mathbf{T}_i]_{\mathbf{c}\{k\}}}{\mathrm{Tr}\{\mathbf{T}_i^\dagger\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}} \\ &= \delta_{ab} \qquad (15)\end{aligned}$$

where $\delta_{ab} = 1$ if $a = b$ and $\delta_{ab} = 0$ if $a \neq b$ (the Kronecker delta function) and the last step follows from the fact that $\mathbf{G}$ has i.i.d. $\mathcal{CN}(0,1)$ entries.

$$\begin{aligned}\mathbb{E}\left\{\left|[\hat{\mathbf{Q}}_i]_{ab}\right|^2\right\} &= \sum_{k=1}^{K}\sum_{\ell=1}^{K}\mathbb{E}\left\{[\hat{\mathbf{h}}_k^i]_a[\hat{\mathbf{h}}_k^i]_b^\dagger[\hat{\mathbf{h}}_\ell^i]_b[\hat{\mathbf{h}}_\ell^i]_a^\dagger\right\} \\ &= \sum_{k,\ell}\Bigg[\mathbb{E}\left\{[\hat{\mathbf{h}}_k^i]_a[\hat{\mathbf{h}}_k^i]_b^\dagger\right\}\mathbb{E}\left\{[\hat{\mathbf{h}}_\ell^i]_b[\hat{\mathbf{h}}_\ell^i]_a^\dagger\right\} \\ &\quad + \mathbb{E}\left\{[\hat{\mathbf{h}}_k^i]_a[\hat{\mathbf{h}}_\ell^i]_a^\dagger\right\}\mathbb{E}\left\{[\hat{\mathbf{h}}_k^i]_b^\dagger[\hat{\mathbf{h}}_\ell^i]_b\right\}\Bigg] \qquad (16a) \\ &= \left|\mathbb{E}\{[\hat{\mathbf{Q}}_i]_{ab}\}\right|^2 + \sum_{k,\ell}\left|\mathbb{E}\left\{[\hat{\mathbf{h}}_k^i]_b^\dagger[\hat{\mathbf{h}}_\ell^i]_b\right\}\right|^2 \qquad (16b) \\ &= \left|\mathbb{E}\{[\hat{\mathbf{Q}}_i]_{ab}\}\right|^2 + \sum_{k,\ell}\frac{[\mathbf{T}_i{}^\dagger\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i]_{lk}^2}{\mathrm{Tr}\{\mathbf{T}_i^\dagger\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}^2} \\ &= \left|\mathbb{E}\{[\hat{\mathbf{Q}}_i]_{ab}\}\right|^2 + \frac{\|\mathbf{T}_i^\dagger\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\|_{\mathbf{F}}^2}{\mathrm{Tr}\{\mathbf{T}_i^\dagger\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}^2} \qquad (16c)\end{aligned}$$

where (16a) follows from the result on the expectation of the product of four circularly symmetric jointly Gaussian random variables [47] and (16b) follows from the fact that the vector $\hat{\mathbf{h}}_\beta^i$ has i.i.d entries $\forall\beta \in \{1,..,K\}$. Therefore, if $\lim_{s\to\infty}\frac{\|\mathbf{T}_i^\dagger\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\|_{\mathbf{F}}}{\mathrm{Tr}\{\mathbf{T}_i^\dagger\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}} = 0$, from (14), (16c) we have:

$$\lim_{s\to\infty}\hat{\mathbf{Q}}_i \overset{\mathrm{ms}}{=} \mathbb{I}_M \Rightarrow \lim_{s\to\infty}\mathbf{Q}_i \overset{\mathrm{ms}}{=} \mathbf{R}_{\mathbf{rx}} \qquad (17)$$

where $\overset{\mathrm{ms}}{=}$ denotes element-wise mean square convergence. From the Hoffman-Weilandt inequality [48], there exists a permutation matrix $\mathbf{P}$ such that $\|\mathbf{P}\Lambda_i\mathbf{P}^\dagger - \Lambda^{\mathbf{rx}}\|_{\mathbf{F}} \leq \|\mathbf{Q}_i - \mathbf{R}_{\mathbf{rx}}\|_{\mathbf{F}}$, where $\Lambda_i, \Lambda^{\mathbf{rx}}$ are the eigenvalue matrices of $\mathbf{Q}_i, \mathbf{R}_{\mathbf{rx}}$, respectively. Without loss of generality, assuming Λ_i is always picked in this permutation, we have:

$$\lim_{s\to\infty}\Lambda_i \overset{\mathrm{ms}}{=} \Lambda^{\mathbf{rx}} \qquad (18)$$

Let $\mathbf{e}_{im}$ be an eigenvector of $\mathbf{Q}_i$ corresponding to an eigenvalue $\lambda_{im} = [\Lambda_i]_{m,m}$. Now as $s \to \infty$:

$$\begin{aligned}\mathbf{Q}_i\mathbf{e}_{im} &= \lambda_{im}\mathbf{e}_{im} \qquad (19a) \\ \Rightarrow \mathbf{R}_{\mathbf{rx}}\mathbf{e}_{im} &\overset{\mathrm{ms}}{=} \lambda_m^{\mathbf{rx}}\mathbf{e}_{im} \quad \text{[From (17) and (18)]} \\ \Rightarrow \mathbf{e}_{im} &\overset{\mathrm{ms}}{=} e^{j\phi}\mathbf{e}_m^{\mathbf{rx}} \qquad (19b)\end{aligned}$$

for some angle ϕ (which may be a function of $\mathbf{G}$), where (19b) follows from the fact that all the eigenvalues of $\mathbf{R}_{\mathbf{rx}}$ are distinct. Now since both the eigenvalues and eigenvectors converge in mean square sense, following a similar procedure to [20], from (19a) we have:

$$\begin{aligned}&[\mathbf{Q}_i - \mathbf{R}_{\mathbf{rx}} + \mathbf{R}_{\mathbf{rx}}][\mathbf{e}_{im} - e^{j\phi}\mathbf{e}_m^{\mathbf{rx}} + e^{j\phi}\mathbf{e}_m^{\mathbf{rx}}] \\ &\qquad = [\lambda_{im} - \lambda_m^{\mathbf{rx}} + \lambda_m^{\mathbf{rx}}][\mathbf{e}_{im} - e^{j\phi}\mathbf{e}_m^{\mathbf{rx}} + e^{j\phi}\mathbf{e}_m^{\mathbf{rx}}] \\ &\Rightarrow [\mathbf{Q}_i - \mathbf{R}_{\mathbf{rx}}]e^{j\phi}\mathbf{e}_m^{\mathbf{rx}} + \mathbf{R}_{\mathbf{rx}}[\mathbf{e}_{im} - e^{j\phi}\mathbf{e}_m^{\mathbf{rx}}] \\ &\qquad \simeq [\lambda_{im} - \lambda_m^{\mathbf{rx}}]e^{j\phi}\mathbf{e}_m^{\mathbf{rx}} + \lambda_m^{\mathbf{rx}}[\mathbf{e}_{im} - e^{j\phi}\mathbf{e}_m^{\mathbf{rx}}] \qquad (20a) \\ &\Rightarrow \mathbf{e}_m^{\mathbf{rx}\dagger}[\mathbf{Q}_i - \mathbf{R}_{\mathbf{rx}}]\mathbf{e}_m^{\mathbf{rx}} \simeq \lambda_{im} - \lambda_m^{\mathbf{rx}} \qquad (20b) \\ &\Rightarrow \lambda_{im} \simeq \mathbf{e}_m^{\mathbf{rx}\dagger}\mathbf{Q}_i\mathbf{e}_m^{\mathbf{rx}} \qquad (20c)\end{aligned}$$

where, as in [20], we use "$\simeq$" to denote a first-order approximation (i.e., an equality in which the higher order terms that do not influence the asymptotic statistics of λ_{im} as $s \to \infty$ are neglected). Note that (20a) follows by neglecting the higher order terms and (20b) follows by premultiplying both sides by $e^{-j\phi}[\mathbf{e}_m^{\mathbf{rx}}]^\dagger$. This proves the asymptotic first-order expression for the eigenvalues (5).

By taking expectations on both sides of (20c), the asymptotic mean can be expressed as:

$$\begin{aligned}\mu_{im} &= \mathbb{E}\{\lambda_{im}\} \simeq \mathbf{e}_m^{\mathbf{rx}\dagger}\mathbb{E}\{\mathbf{Q}_i\}\mathbf{e}_m^{\mathbf{rx}} \\ &\simeq \lambda_m^{\mathbf{rx}} \qquad \text{[From (17)]} \qquad (21)\end{aligned}$$

Similarly, for cross-correlation we have:

$$\begin{aligned}&\mathbb{E}\left\{\lambda_{im}\lambda_{jn}\right\} \simeq \mathbb{E}\left\{\mathbf{e}_m^{\mathbf{rx}\dagger}\mathbf{Q}_i\mathbf{e}_m^{\mathbf{rx}}\mathbf{e}_n^{\mathbf{rx}\dagger}\mathbf{Q}_j\mathbf{e}_n^{\mathbf{rx}}\right\} \\ &= \lambda_m^{\mathbf{rx}}\lambda_n^{\mathbf{rx}}\sum_{k=1}^{K}\sum_{\ell=1}^{K}\mathbb{E}\left\{\mathbf{e}_m^{\mathbf{rx}\dagger}\hat{\mathbf{h}}_k^i[\hat{\mathbf{h}}_k^i]^\dagger\mathbf{e}_m^{\mathbf{rx}}\mathbf{e}_n^{\mathbf{rx}\dagger}\hat{\mathbf{h}}_\ell^j[\hat{\mathbf{h}}_\ell^j]^\dagger\mathbf{e}_n^{\mathbf{rx}}\right\} \qquad (22)\end{aligned}$$

where the last step follows from (14). Notice that $\mathbf{e}_\eta^{\mathbf{rx}\dagger}\hat{\mathbf{h}}_\beta^\alpha$ are all circularly symmetric jointly Gaussian random variables for all $1 \leq \eta \leq M$, $1 \leq \alpha \leq |\mathcal{T}|$, $1 \leq \beta \leq K$. From the result on the expectation of the product of four complex, circularly symmetric jointly Gaussian random variables [47], we have:

$$\begin{aligned}&\mathbb{E}\left\{\lambda_{im}\lambda_{jn}\right\} \\ &\simeq \lambda_m^{\mathbf{rx}}\lambda_n^{\mathbf{rx}}\sum_{k=1}^{K}\sum_{\ell=1}^{K}\mathbf{e}_m^{\mathbf{rx}\dagger}\mathbb{E}\left\{\hat{\mathbf{h}}_k^i[\hat{\mathbf{h}}_k^i]^\dagger\right\}\mathbf{e}_m^{\mathbf{rx}}\mathbf{e}_n^{\mathbf{rx}\dagger}\mathbb{E}\left\{\hat{\mathbf{h}}_\ell^j[\hat{\mathbf{h}}_\ell^j]^\dagger\right\}\mathbf{e}_n^{\mathbf{rx}} \\ &+ \lambda_m^{\mathbf{rx}}\lambda_n^{\mathbf{rx}}\sum_{k=1}^{K}\sum_{\ell=1}^{K}\mathbf{e}_m^{\mathbf{rx}\dagger}\mathbb{E}\left\{\hat{\mathbf{h}}_k^i[\hat{\mathbf{h}}_\ell^j]^\dagger\right\}\mathbf{e}_n^{\mathbf{rx}}\mathbf{e}_n^{\mathbf{rx}\dagger}\mathbb{E}\left\{\hat{\mathbf{h}}_\ell^j[\hat{\mathbf{h}}_k^i]^\dagger\right\}\mathbf{e}_m^{\mathbf{rx}} \\ &\Rightarrow \mathbb{E}\left\{\lambda_{im}\lambda_{jn}\right\} - \mu_{im}\mu_{jn}\end{aligned}$$

$$\simeq \sum_{k,\ell} \frac{\lambda_m^{\mathbf{rx}}\lambda_n^{\mathbf{rx}}\left|\mathbf{e}_m^{\mathbf{rx}\dagger}\mathbb{E}\{\mathbf{G}\mathbf{R}_{\mathbf{tx}}^{1/2}[\mathbf{T}_i]_{\mathbf{c}\{k\}}[\mathbf{T}_j]_{\mathbf{c}\{\ell\}}^{\dagger}\mathbf{R}_{\mathbf{tx}}^{1/2}\mathbf{G}^{\dagger}\}\mathbf{e}_n^{\mathbf{rx}}\right|^2}{[\mathrm{Tr}\{\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}\mathrm{Tr}\{\mathbf{T}_j^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_j\}]}$$
$$= \sum_{k,\ell} \frac{\lambda_m^{\mathbf{rx}}\lambda_n^{\mathbf{rx}}\left|[\mathbf{T}_j]_{\mathbf{c}\{\ell\}}^{\dagger}\mathbf{R}_{\mathbf{tx}}^{1/2}\mathbb{E}\{\mathbf{G}^{\dagger}\mathbf{e}_n^{\mathbf{rx}}\mathbf{e}_m^{\mathbf{rx}\dagger}\mathbf{G}\}\mathbf{R}_{\mathbf{tx}}^{1/2}[\mathbf{T}_i]_{\mathbf{c}\{k\}}\right|^2}{[\mathrm{Tr}\{\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}\mathrm{Tr}\{\mathbf{T}_j^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_j\}]}$$
$$= \sum_{k,\ell} \frac{\delta_{mn}\lambda_m^{\mathbf{rx}}\lambda_n^{\mathbf{rx}}\left|\left([\mathbf{T}_j]_{\mathbf{c}\{\ell\}}\right)^{\dagger}\mathbf{R}_{\mathbf{tx}}[\mathbf{T}_i]_{\mathbf{c}\{k\}}\right|^2}{[\mathrm{Tr}\{\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}\mathrm{Tr}\{\mathbf{T}_j^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_j\}]}$$
$$= \frac{\delta_{mn}\lambda_m^{\mathbf{rx}}\lambda_n^{\mathbf{rx}}\left\|\mathbf{T}_j^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\right\|_{\mathbf{F}}^2}{\mathrm{Tr}\{\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}\mathrm{Tr}\{\mathbf{T}_j^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_j\}} \tag{23}$$

where, the penultimate step follows from the fact that $\mathbf{G}$ has i.i.d. entries and $\mathbf{e}_m^{\mathbf{rx}\dagger}\mathbf{e}_n^{\mathbf{rx}} = \delta_{mn}$. This concludes the proof. □

APPENDIX B

(Proof of proposition III.1.1). From Theorem III.1, the required condition is that

$$\lim_{K\to\infty} \frac{\|\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\|_{\mathbf{F}}}{\mathrm{Tr}\{\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}} = 0 \quad \forall i \in \{1,..,|\mathcal{T}|\}$$

Note that:

$$\frac{\|\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\|_{\mathbf{F}}}{\mathrm{Tr}\{\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}} = \frac{\|\boldsymbol{\lambda}\{\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}\|_2}{\|\boldsymbol{\lambda}\{\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}\|_1} = \frac{\|\boldsymbol{\lambda}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\}\|_2}{\|\boldsymbol{\lambda}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\}\|_1} \tag{24}$$

where $\boldsymbol{\lambda}\{\mathbf{A}\}$ is the vector of eigenvalues for a square matrix $\mathbf{A}$. We define $\boldsymbol{\lambda}^{\uparrow}\{\mathbf{A}\}$ and $\boldsymbol{\lambda}^{\downarrow}\{\mathbf{A}\}$ as sortings of $\boldsymbol{\lambda}\{\mathbf{A}\}$ in ascending and descending orders, respectively. From results on eigenvalue majorization [49, Eqn 3.20], we have for all $1 \le L \le N$:

$$\prod_{\ell=1}^{L} \lambda_\ell^{\downarrow}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\}\lambda_\ell^{\uparrow}\{\mathbf{R}_{\mathbf{tx}}\} \le \prod_{\ell=1}^{L} \lambda_\ell^{\downarrow}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\}$$
$$\le \prod_{\ell=1}^{L} \lambda_\ell^{\downarrow}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\}\lambda_\ell^{\downarrow}\{\mathbf{R}_{\mathbf{tx}}\} \tag{25}$$

where $\lambda_\ell^{\cdot}\{\mathbf{A}\}$ is the ℓ-th element of $\boldsymbol{\lambda}^{\cdot}\{\mathbf{A}\}$. By taking the logarithm on both sides we get:

$$\log\left[\boldsymbol{\lambda}^{\downarrow}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\}\circ\boldsymbol{\lambda}^{\uparrow}\{\mathbf{R}_{\mathbf{tx}}\}\right] \prec \log\left[\boldsymbol{\lambda}^{\downarrow}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\}\right]$$
$$\prec \log\left[\boldsymbol{\lambda}^{\downarrow}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\}\circ\boldsymbol{\lambda}^{\downarrow}\{\mathbf{R}_{\mathbf{tx}}\}\right] \tag{26}$$

where $\circ$ denotes the Hadamard product, the logarithm is taken element-wise and $\mathbf{a} \prec \mathbf{b}$ implies $\mathbf{b}$ majorizes $\mathbf{a}$. Now consider the function $f_\alpha(\mathbf{x}) = \|e^{\mathbf{x}}\|_\alpha$ where $\alpha \in \{1, 2, \ldots\}$ and the exponent $e^{\mathbf{x}}$ is taken element wise. The function is clearly permutation invariant in vector $\mathbf{x}$ and for any elements x_a, x_b of vector $\mathbf{x}$, we have:

$$\left[\frac{\partial f_\alpha(\mathbf{x})}{\partial x_a} - \frac{\partial f_\alpha(\mathbf{x})}{\partial x_b}\right](x_a - x_b) = (x_a - x_b)\frac{e^{\alpha x_a} - e^{\alpha x_b}}{\|e^{\mathbf{x}}\|_\alpha^{\alpha-1}}$$
$$\ge 0 \tag{27}$$

Therefore from [49, Th 2.3.14], $f_\alpha(\mathbf{x})$ is a Schur-convex function. From the definition of a Schur convex function and from (26) we have:

$$\frac{f_2\left(\log[\boldsymbol{\lambda}^{\downarrow}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\}]\right)}{f_1\left(\log[\boldsymbol{\lambda}^{\downarrow}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\}]\right)} \le \frac{f_2\left(\log[\boldsymbol{\lambda}^{\downarrow}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\}\circ\boldsymbol{\lambda}^{\downarrow}\{\mathbf{R}_{\mathbf{tx}}\}]\right)}{f_1\left(\log[\boldsymbol{\lambda}^{\downarrow}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\}\circ\boldsymbol{\lambda}^{\uparrow}\{\mathbf{R}_{\mathbf{tx}}\}]\right)} \tag{28}$$

$$\Rightarrow \frac{\|\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\|_{\mathbf{F}}}{\mathrm{Tr}\{\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}} \le \frac{\sqrt{\sum_{k=1}^{K}(\lambda_k^{\mathbf{tx}})^2}}{\left[\sum_{\ell=1}^{K}\lambda_{N+1-\ell}^{\mathbf{tx}}\right]} \tag{29}$$

where in the last step we have used the fact that $\mathbf{T}_i$ is semi-unitary with dimension $N \times K$. Alternately, from Hölder's inequality:

$$\frac{\|\boldsymbol{\lambda}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\}\|_2}{\|\boldsymbol{\lambda}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\}\|_1} \le \frac{\|\boldsymbol{\lambda}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\}\|_\infty}{\|\boldsymbol{\lambda}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\}\|_2}$$

Now following similar steps to before, we have:

$$\frac{\|\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\|_{\mathbf{F}}}{\mathrm{Tr}\{\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}} \le \frac{f_\infty\left(\log[\boldsymbol{\lambda}^{\downarrow}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\}\circ\boldsymbol{\lambda}^{\downarrow}\{\mathbf{R}_{\mathbf{tx}}\}]\right)}{f_2\left(\log[\boldsymbol{\lambda}^{\downarrow}\{\mathbf{T}_i\mathbf{T}_i^{\dagger}\}\circ\boldsymbol{\lambda}^{\uparrow}\{\mathbf{R}_{\mathbf{tx}}\}]\right)}$$
$$\le \frac{\lambda_1^{\mathbf{tx}}}{\sqrt{\sum_{\ell=1}^{K}\left(\lambda_{N+1-\ell}^{\mathbf{tx}}\right)^2}} \tag{30}$$

Therefore a sufficient condition for Theorem III.1 is that the eigenvalues of $\mathbf{R}_{\mathbf{rx}}$ be distinct, and either the right hand side in (29) or (30) go to zero as $s \to \infty$. □

APPENDIX C

(Proof of Theorem III.2). Since $K = sK_o \ge s$, we have:

$$\frac{(\lambda_1^{\mathbf{tx}})^2}{\sum_{k=1}^{K}(\lambda_{N+1-k}^{\mathbf{tx}})^2} \le \frac{(\lambda_1^{\mathbf{tx}})^2}{\sum_{k=1}^{s}(\lambda_{N+1-k}^{\mathbf{tx}})^2} \tag{31}$$

Therefore using (31) and conditions 1,3 of the theorem statement, from proposition III.1.1, for any $1 \le i \le |\mathcal{T}|$, $1 \le m \le M$ as $s \to \infty$:

$$\lambda_{im} = \mathbf{e}_m^{\mathbf{rx}\dagger}\mathbf{Q}_i\mathbf{e}_m^{\mathbf{rx}} = \frac{\lambda_m^{\mathbf{rx}}\hat{\mathbf{g}}_m\mathbf{R}_{\mathbf{tx}}^{1/2}\mathbf{T}_i\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}^{1/2}\hat{\mathbf{g}}_m^{\dagger}}{\mathrm{Tr}\{\mathbf{T}_i^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_i\}} \tag{32}$$

where $\hat{\mathbf{g}}_m = \mathbf{e}_m^{\mathbf{rx}\dagger}\mathbf{G}$ is a $1 \times N$ vector with i.i.d. $\mathcal{CN}(0,1)$ entries. Additionally, the second-order statistics of λ_{im} are also as given by (6) and (7).

Now, for any $1 \le m \le M$, consider a set of eigenvalues $\mathbf{v} = [\lambda_{i_1 m}, \lambda_{i_2 m}, \ldots, \lambda_{i_L m}]$. From the Cramer-Wold theorem [50], these eigenvalues are asymptotically jointly Gaussian iff any weighted sum converges to a Gaussian distribution as $s \to \infty$. A weighted sum is given by:

$$\mathbf{w}^{\dagger}\mathbf{v} = \sum_{\ell=1}^{L} w_\ell \lambda_{i_\ell m}$$
$$= \lambda_m^{\mathbf{rx}}\hat{\mathbf{g}}_m\mathbf{R}_{\mathbf{tx}}^{1/2}\left[\sum_{\ell=1}^{L} w_\ell \frac{\mathbf{T}_{i_\ell}\mathbf{T}_{i_\ell}^{\dagger}}{\mathrm{Tr}\{\mathbf{T}_{i_\ell}^{\dagger}\mathbf{R}_{\mathbf{tx}}\mathbf{T}_{i_\ell}\}}\right]\mathbf{R}_{\mathbf{tx}}^{1/2}\hat{\mathbf{g}}_m^{\dagger}$$

Using condition 2 of the theorem statement, we have:

$\mathbf{w}^{\dagger}\mathbf{v}$

$$= \lambda_m^{\mathbf{rx}} \hat{\mathbf{g}}_m \mathbf{R}_{\mathbf{tx}}^{1/2} \left[\left(\underbrace{\sum_{\ell=1}^{L} \frac{w_\ell \mathbf{U}_{i_\ell} \mathbf{U}_{i_\ell}^\dagger}{\mathrm{Tr}\{\mathbf{T}_{i_\ell}^\dagger \mathbf{R}_{\mathbf{tx}} \mathbf{T}_{i_\ell}\}}}_{\mathbf{W}} \right) \otimes \mathbf{V}\mathbf{V}^\dagger \right] \mathbf{R}_{\mathbf{tx}}^{1/2} \hat{\mathbf{g}}_m^\dagger$$

$$\overset{\mathrm{d}}{=} \sum_{n=1}^{N} \lambda_m^{\mathbf{rx}} \left| [\hat{\mathbf{g}}_m]_n \right|^2 \lambda_n^{\downarrow} \{ \mathbf{R}_{\mathbf{tx}}^{1/2} [\mathbf{W} \otimes \mathbf{V}\mathbf{V}^\dagger] \mathbf{R}_{\mathbf{tx}}^{1/2} \} \tag{33}$$

where $\overset{\mathrm{d}}{=}$ represents equality in distribution and $\lambda_n^{\downarrow}\{\mathbf{A}\}$ is the n-th largest eigenvalue of a matrix $\mathbf{A}$. Using Lyapunov's central limit theorem, it is easy to show that (33) is asymptotically Gaussian if (see [20], [51] for details):

$$\lim_{s\to\infty} \frac{\|\boldsymbol{\lambda}\{\mathbf{R}_{\mathbf{tx}}^{1/2} [\mathbf{W} \otimes \mathbf{V}\mathbf{V}^\dagger] \mathbf{R}_{\mathbf{tx}}^{1/2}\}\|_\infty}{\|\boldsymbol{\lambda}\{\mathbf{R}_{\mathbf{tx}}^{1/2} [\mathbf{W} \otimes \mathbf{V}\mathbf{V}^\dagger] \mathbf{R}_{\mathbf{tx}}^{1/2}\}\|_2} = 0 \tag{34}$$

where, $\boldsymbol{\lambda}\{\mathbf{A}\}$ represents the vector of eigenvalues of a matrix $\mathbf{A}$. Now following similar steps to those in Appendix B (see (28)), we have:

$$\frac{\|\boldsymbol{\lambda}\{\mathbf{R}_{\mathbf{tx}}^{1/2} [\mathbf{W} \otimes \mathbf{V}\mathbf{V}^\dagger] \mathbf{R}_{\mathbf{tx}}^{1/2}\}\|_\infty}{\|\boldsymbol{\lambda}\{\mathbf{R}_{\mathbf{tx}}^{1/2} [\mathbf{W} \otimes \mathbf{V}\mathbf{V}^\dagger] \mathbf{R}_{\mathbf{tx}}^{1/2}\}\|_2} = \frac{\|\boldsymbol{\lambda}\{[\mathbf{W} \otimes \mathbb{I}_s] \mathbf{R}_{\mathbf{tx}}\}\|_\infty}{\|\boldsymbol{\lambda}\{[\mathbf{W} \otimes \mathbb{I}_s] \mathbf{R}_{\mathbf{tx}}\}\|_2} \tag{35}$$

$$\leq \frac{\|\boldsymbol{\lambda}^{\downarrow}\{\mathbf{W} \otimes \mathbb{I}_s\} \circ \boldsymbol{\lambda}^{\downarrow}\{\mathbf{R}_{\mathbf{tx}}\}\|_\infty}{\|\boldsymbol{\lambda}^{\downarrow}\{\mathbf{W} \otimes \mathbb{I}_s\} \circ \boldsymbol{\lambda}^{\uparrow}\{\mathbf{R}_{\mathbf{tx}}\}\|_2}$$

$$\leq \frac{\lambda_1^{\downarrow}\{\mathbf{R}_{\mathbf{tx}}\} \lambda_1^{\downarrow}\{\mathbf{W}\}}{\sqrt{\sum_{\ell=1}^{s} \left[\lambda_\ell^{\uparrow}\{\mathbf{R}_{\mathbf{tx}}\} \lambda_1^{\downarrow}\{\mathbf{W}\} \right]^2}}$$

$$= \frac{\lambda_1^{\mathbf{tx}}}{\sqrt{\sum_{\ell=1}^{s} (\lambda_{N+1-\ell}^{\mathbf{tx}})^2}} \tag{36}$$

where we use the fact that $\mathbf{V}\mathbf{V}^\dagger = \mathbb{I}_s$ and $\boldsymbol{\lambda}^{\downarrow}\{\mathbf{W} \otimes \mathbb{I}_s\} = \boldsymbol{\lambda}^{\downarrow}\{\mathbf{W}\} \otimes \boldsymbol{\lambda}^{\downarrow}\{\mathbb{I}_s\}$. From (34), (36) and condition 3 of the theorem, $\mathbf{w}^\dagger \mathbf{v}$ is Gaussian distributed $\forall \mathbf{w}$ as $s \to \infty$. Therefore the vector $\mathbf{v} = [\lambda_{i_1 m}, \lambda_{i_2 m}, ..., \lambda_{i_L m}]$ is asymptotically jointly Gaussian distributed. Note that the joint Gaussianity above trivially also ensures marginal Gaussianity. Now, the joint Gaussianity of λ_{im} and λ_{jn} for $n \neq m$ directly follows from the marginal Gaussianity and the independence of $\hat{\mathbf{g}}_m$ and $\hat{\mathbf{g}}_n$ in (32). □

Appendix D

In this section we shall enumerate some families of $N \times N$ transmit correlation matrices, where $N = N_o s$, which satisfy: $\lim_{s\to\infty} \frac{(\lambda_1^{\mathbf{tx}})^2}{\sum_{k=1}^{s} (\lambda_{N+1-k}^{\mathbf{tx}})^2} = 0$. As discussed in Appendix C, this condition also implies Proposition III.1.1.

1) Exponential Correlation: In this case, the elements of the correlation matrix are given by $[\mathbf{R}_{\mathbf{tx}}]_{ab} = \rho^{|a-b|}$ for $|\rho| < 1$. That this matrix satisfies the required constraint can be verified using the bounds on eigenvalues as derived in [52].

2) Arbitrary power angle spectrum (PAS): We consider a uniform linear antenna array at the transmitter with spacing $d_{\mathbf{tx}}$. Assuming the multipath components to be only in the horizontal plane, the elements of the transmit correlation matrix can be expressed as:

$$[\mathbf{R}_{\mathbf{tx}}]_{ab} = \int_0^1 e^{j2\pi f(a-b)} \underbrace{\left[\sum_{n=-\infty}^{\infty} \xi(f-n) \right]}_{\zeta(f)} df \tag{37}$$

where, $\xi(f)$

$$= \begin{cases} \frac{\lambda \overline{\mathrm{PAS}}\left(\arcsin(\frac{f\lambda}{d_{\mathbf{tx}}})\right) + \lambda \overline{\mathrm{PAS}}\left(\pi - \arcsin(\frac{f\lambda}{d_{\mathbf{tx}}})\right)}{\sqrt{d_{\mathbf{tx}}^2 - \lambda^2 f^2}} & \text{for } \frac{\lambda|f|}{d_{\mathbf{tx}}} \leq 1 \\ 0 & \text{for } \frac{\lambda|f|}{d_{\mathbf{tx}}} > 1 \end{cases} \tag{38}$$

$j = \sqrt{-1}$, λ is the wavelength and $\overline{\mathrm{PAS}}(\theta)$ is the normalized PAS computed as $\overline{\mathrm{PAS}}(\theta) = \mathrm{PAS}(\theta) \big/ \int_{-\pi}^{\pi} \mathrm{PAS}(\phi) d\phi$. We define $\mathcal{F}_a$ as the smallest, possibly non-contiguous, sub-interval of $[0, 1]$ such that:

$$\min_{f\in[0,1]-\mathcal{F}_a} \{\zeta(f)\} \geq \max_{f\in\mathcal{F}_a} \{\zeta(f)\} \text{ and } \int_{\mathcal{F}_a} f df = 1/a$$

where $\zeta(f)$ is as defined in (37). Since $\mathbf{R}_{\mathbf{tx}}$ is a Toeplitz matrix, from Szego's results on eigenvalues of Toeplitz matrices [53], as $s \to \infty$ we have:

$$\frac{(\lambda_1^{\mathbf{tx}})^2}{\sum_{k=1}^{s} (\lambda_{N+1-k}^{\mathbf{tx}})^2} = \frac{\max_{f\in\mathcal{F}_1} \{\zeta^2(f)\}}{N \int_{f'\in\mathcal{F}_{N_o}} \zeta^2(f') df'} \tag{39}$$

It can be easily verified that the right hand side of (39) goes to zero if:

- $\mathrm{PAS}(\theta)$ is continuous, $\max_{\theta\in(-\pi,\pi]} \{\mathrm{PAS}(\theta)\} < \infty$ and $\mathrm{PAS}(\frac{\pi}{2}) = \mathrm{PAS}(-\frac{\pi}{2}) = 0$
- The constant N_o is such that $\int_{f'\in\mathcal{F}_{N_o}} \zeta^2(f') df' > 0$.

The former condition is almost always satisfied for a sectored base-station. The latter condition is more stringent. However, it can be relaxed if we restrict the codebook $\mathcal{T}$ such that $\forall i \ \lim_{s\to\infty} \mathrm{rank}\{\mathbf{T}_i^\dagger \mathbf{R}_{\mathbf{tx}} \mathbf{T}_i\}/s > K_o - 1$.

References

[1] T. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *Wireless Communications, IEEE Transactions on*, vol. 9, pp. 3590–3600, November 2010.

[2] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *Selected Areas in Communications, IEEE Journal on*, vol. 26, pp. 1341–1365, October 2008.

[3] D. GORE and A. Paulraj, "MIMO antenna subset selection with space-time coding," *Signal Processing, IEEE Transactions on*, vol. 50, pp. 2580–2588, Oct 2002.

[4] A. Molisch, "MIMO systems with antenna selection - an overview," in *Radio and Wireless Conference, 2003. RAWCON '03. Proceedings*, pp. 167–170, Aug 2003.

[5] X. Zhang, A. Molisch, and S.-Y. Kung, "Phase-shift-based antenna selection for MIMO channels," in *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*, vol. 2, pp. 1089–1093 Vol.2, Dec 2003.

[6] A. Molisch and X. Zhang, "Fft-based hybrid antenna selection schemes for spatially correlated MIMO channels," *Communications Letters, IEEE*, vol. 8, pp. 36–38, Jan 2004.

[7] P. Sudarshan, N. Mehta, A. Molisch, and J. Zhang, "Channel statistics-based RF pre-processing with antenna selection," *Wireless Communications, IEEE Transactions on*, vol. 5, pp. 3501–3511, December 2006.

[8] D. Love and R. Heath, "Grassmannian beamforming on correlated MIMO channels," in *Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE*, vol. 1, pp. 106–110 Vol.1, Nov 2004.

[9] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, “On beamforming with finite rate feedback in multiple-antenna systems,” *Information Theory, IEEE Transactions on*, vol. 49, pp. 2562–2579, Oct 2003.

[10] B. Mondal and R. Heath, “Performance analysis of quantized beamforming MIMO systems,” *Signal Processing, IEEE Transactions on*, vol. 54, pp. 4753–4766, Dec 2006.

[11] D. Love and R. Heath, “Limited feedback unitary precoding for spatial multiplexing systems,” *Information Theory, IEEE Transactions on*, vol. 51, pp. 2967–2976, Aug 2005.

[12] V. Raghavan, R. Heath, and A. Sayeed, “Systematic codebook designs for quantized beamforming in correlated MIMO channels,” *Selected Areas in Communications, IEEE Journal on*, vol. 25, pp. 1298–1310, September 2007.

[13] V. Raghavan, A. Sayeed, and V. Veeravalli, “Limited feedback precoder design for spatially correlated MIMO channels,” in *Information Sciences and Systems, 2007. CISS ’07. 41st Annual Conference on*, pp. 113–118, March 2007.

[14] V. Raghavan and V. Veeravalli, “Ensemble properties of RVQ-based limited-feedback beamforming codebooks,” *Information Theory, IEEE Transactions on*, vol. 59, pp. 8224–8249, Dec 2013.

[15] A. Molisch, M. Win, Y. seok Choi, and J. Winters, “Capacity of MIMO systems with antenna selection,” *Wireless Communications, IEEE Transactions on*, vol. 4, pp. 1759–1772, July 2005.

[16] S. Sanayei and A. Nosratinia, “Capacity of MIMO channels with antenna selection,” *Information Theory, IEEE Transactions on*, vol. 53, pp. 4356–4362, Nov 2007.

[17] H. Shen and A. Ghrayeb, “Analysis of the outage probability for spatially correlated MIMO channels with receive antenna selection,” in *Global Telecommunications Conference, 2005. GLOBECOM ’05. IEEE*, vol. 5, pp. 5 pp.–2464, Dec 2005.

[18] S. Sanayei and A. Nosratinia, “Antenna selection in MIMO systems,” *Communications Magazine, IEEE*, vol. 42, pp. 68–73, Oct 2004.

[19] A. M. Tulino and S. Verdu, “Random matrix theory and wireless communications,” *Foundations and TrendsÂő in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.

[20] C. Martin and B. Ottersten, “Asymptotic eigenvalue distributions and capacity for MIMO channels under correlated fading,” *Wireless Communications, IEEE Transactions on*, vol. 3, pp. 1350–1359, July 2004.

[21] D. Morales-Jimenez, J. Paris, J. Entrambasaguas, and K.-K. Wong, “On the diagonal distribution of a complex wishart matrix and its application to the analysis of MIMO systems,” *Communications, IEEE Transactions on*, vol. 59, pp. 3475–3484, December 2011.

[22] A. Maaref and S. Aissa, “Joint and marginal eigenvalue distributions of (non)central complex wishart matrices and pdf-based approach for characterizing the capacity statistics of MIMO ricean and rayleigh fading channels,” *Wireless Communications, IEEE Transactions on*, vol. 6, pp. 3607–3619, October 2007.

[23] P. J. Smith and L. M. Garth, “Distribution and characteristic functions for correlated complex wishart matrices,” *Journal of Multivariate Analysis*, vol. 98, no. 4, pp. 661 – 677, 2007.

[24] P.-H. Kuo, P. Smith, and L. Garth, “Joint density for eigenvalues of two correlated complex wishart matrices: Characterization of MIMO systems,” *Wireless Communications, IEEE Transactions on*, vol. 6, pp. 3902–3906, November 2007.

[25] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen, and F. Frederiksen, “A stochastic MIMO radio channel model with experimental validation,” *Selected Areas in Communications, IEEE Journal on*, vol. 20, pp. 1211–1226, Aug 2002.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[27] K. V. Mardia, “Measures of multivariate skewness and kurtosis with applications,” *Biometrika*, vol. 57, no. 3, pp. 519–530, 1970.

[28] P. Smith and M. Shafi, “On a gaussian approximation to the capacity of wireless MIMO systems,” in *Communications, 2002. ICC 2002. IEEE International Conference on*, vol. 1, pp. 406–410, 2002.

[29] P. Smith, S. Roy, and M. Shafi, “Capacity of MIMO systems with semicorrelated flat fading,” *Information Theory, IEEE Transactions on*, vol. 49, pp. 2781–2788, Oct 2003.

[30] “Fundamental multidimensional variables,” in *Probability Distributions Involving Gaussian Random Variables*, pp. 17–23, Springer US, 2002.

[31] R. B. Arellano-Valle and M. G. Genton, “On the exact distribution of linear combinations of order statistics from dependent random variables,” *Journal of Multivariate Analysis*, vol. 98, no. 10, pp. 1876 – 1894, 2007.

[32] H. Nagaraja, “Order statistics from independent exponential random variables and the sum of the top order statistics,” in *Advances in Distribution Theory, Order Statistics, and Inference* (N. Balakrishnan, J. Sarabia, and E. Castillo, eds.), Statistics for Industry and Technology, pp. 173–185, BirkhÃď’user Boston, 2006.

[33] A. Ross, “Computing bounds on the expected maximum of correlated normal variables,” *Methodology and Computing in Applied Probability*, vol. 12, no. 1, pp. 111–138, 2010.

[34] D. R. Hoover, “Bounds on expectations of order statistics for dependent samples,” *Statistics and Probability Letters*, vol. 8, no. 3, pp. 261 – 265, 1989.

[35] R. A. G. Barry C. Arnold, “Bounds on expectations of linear systematic statistics based on dependent samples,” *The Annals of Statistics*, vol. 7, no. 1, pp. 220–223, 1979.

[36] Y. Tong, “Order statistics of normal variables,” in *The Multivariate Normal Distribution*, Springer Series in Statistics, pp. 123–149, Springer New York, 1990.

[37] C. E. Clark, “The greatest of a finite set of random variables,” *Oper. Res.*, vol. 9, pp. 145–162, Apr. 1961.

[38] D. Sinha, H. Zhou, and N. Shenoy, “Advances in computation of the maximum of a set of random variables,” in *Quality Electronic Design, 2006. ISQED ’06. 7th International Symposium on*, pp. 6 pp.–311, March 2006.

[39] S. Asmussen, J. L. Jensen, and L. Rojas-Nandayapa, “A literature review on lognormal sums,”

[40] L. Fenton, “The sum of log-normal probability distributions in scatter transmission systems,” *Communications Systems, IRE Transactions on*, vol. 8, pp. 57–67, March 1960.

[41] A. Abu-Dayya and N. Beaulieu, “Outage probabilities in the presence of correlated lognormal interferers,” *Vehicular Technology, IEEE Transactions on*, vol. 43, pp. 164–173, Feb 1994.

[42] N. Mehta, J. Wu, A. Molisch, and J. Zhang, “Approximating a sum of random variables with a lognormal,” *Wireless Communications, IEEE Transactions on*, vol. 6, pp. 2690–2699, July 2007.

[43] A. Safak, “Statistical analysis of the power sum of multiple correlated log-normal components,” *Vehicular Technology, IEEE Transactions on*, vol. 42, pp. 58–61, Feb 1993.

[44] S. C. Schwartz and Y. S. Yeh, “On the distribution function and moments of power sums with log-normal components,” *Bell System Technical Journal*, vol. 61, no. 7, pp. 1441–1462, 1982.

[45] D. J. Love, “Tables of complex grassmannian packings [online],” 2004. Available: https://engineering.purdue.edu/ djlove/grass.html.

[46] D. Love, R. Heath, and T. Strohmer, “Grassmannian beamforming for multiple-input multiple-output wireless systems,” *Information Theory, IEEE Transactions on*, vol. 49, pp. 2735–2747, Oct 2003.

[47] F. Dittrich, “Useful formula for moment computation of normal random variables with nonzero means,” *Automatic Control, IEEE Transactions on*, vol. 16, pp. 263–265, Jun 1971.

[48] A. J. Hoffman and H. W. Wielandt, “The variation of the spectrum of a normal matrix,” *Duke Math. J.*, vol. 20, pp. 37–39, 03 1953.

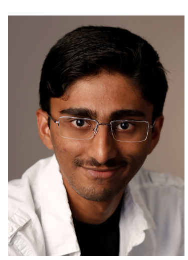
[49] R. Bhatia, *Matrix Analysis (Graduate Texts in Mathematics)*. Springer New York, 1997.

[50] H. Cramér and H. Wold, “Some theorems on distribution functions,” *Journal of the London Mathematical Society*, vol. s1-11, no. 4, pp. 290–294, 1936.

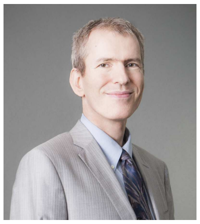
[51] G. Levin and S. Loyka, “Comments on "Asymptotic eigenvalue distributions and capacity for MIMO channels under correlated fading",” *Wireless Communications, IEEE Transactions on*, vol. 7, pp. 475–479, February 2008.

[52] J. Pierce and S. Stein, “Multiple diversity with nonindependent fading,” *Proceedings of the IRE*, vol. 48, pp. 89–104, Jan 1960.

[53] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. AMS Chelsea Publishing Series, University of California Press, 2001.



Vishnu V. Ratnam (S’10) received his Bachelor of Technology (Hons.) in electronics and electrical communication engineering from Indian Institute of Technology, Kharagpur in 2012. He graduated as the Salutatorian for the class of 2012. He is currently pursuing a Ph.D in Electrical Engineering at University of Southern California. His research interests are in: the design and analysis of low complexity transceivers for large antenna systems (Massive MIMO) and Ultra Wide-Band systems; and resource allocation problems in multi-antenna networks such as cooperative/ distributed antenna systems.



Andreas F. Molisch (S'89–M'95–SM'00–F'05) received the Dipl. Ing., Ph.D., and habilitation degrees from the Technical University of Vienna, Vienna, Austria, in 1990, 1994, and 1999, respectively. He subsequently was with AT&T (Bell) Laboratories Research (USA); Lund University, Lund, Sweden, and Mitsubishi Electric Research Labs (USA). He is now a Professor of Electrical Engineering at the University of Southern California, Los Angeles. His current research interests are the measurement and modeling of mobile radio channels, ultra-wideband communications and localization, cooperative communications, multiple-inputâĂŞmultiple-output systems, wireless systems for healthcare, and novel cellular architectures. He has authored, coauthored, or edited four books (among them the textbook Wireless Communications, Wiley-IEEE Press), 16 book chapters, some 200 journal papers, 270 conference papers, as well as more than 80 patents and 70 standards contributions.

Dr. Molisch has been an Editor of a number of journals and special issues, General Chair, Tecnical Program Committee Chair, or Symposium Chair of multiple international conferences, as well as Chairman of various international standardization groups. He is a Fellow of the National Academy of Inventors, Fellow of the AAAS, Fellow of the IET, an IEEE Distinguished Lecturer, and a member of the Austrian Academy of Sciences. He has received numerous awards, among them the Donald Fink Prize of the IEEE, and the Eric Sumner Award of the IEEE.

Haralabos C. Papadopoulos (S'92–M'98) received the S.B., S.M., and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, MA, all in electrical engineering and computer science, in 1990, 1993, and 1998, respectively.

Since December 2005, he has been with DOCOMO Innovations, Palo Alto, CA, working on physical-layer algorithms for wireless communication systems and architectures. From 1998 to 2005, he was on the faculty of the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, and held a joint appointment with the Institute of Systems Research. During his 1993–1995 summer visits to AT&T Bell Labs, Murray Hill, NJ, he worked on shared time-division duplexing systems and digital audio broadcasting. His research interests are in the areas of communications and signal processing, with emphasis on resource-efficient algorithms and architectures for wireless communication systems.

Dr. Papadopoulos is the recipient of an NSF CAREER Award (2000), the G. Corcoran Award (2000) given by the University of Maryland, College Park, and the 1994 F. C. Hennie Award (1994) given by the MIT EECS department. He is also a coauthor of the VTC Fall 2009 Best Student Paper Award. He is a member of Eta Kappa Nu and Tau Beta Pi. He is also active in the industry and an inventor on several issued and pending patents.