The Statistics of the MIMO Frequency selective Fading AWGN Channel Capacity

Anna Scaglione (Contact Author), *Member*, *IEEE*

Atul Salhotra

A. Scaglione and A. Salhotra are with the Department of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA This work was supported by the NSF grant CCR-0133635.

Abstract

The classic problem of maximizing the information rate over parallel Gaussian independent sub-channels with a limit on the total power leads to the elegant closed form water-filling solution. In the case of multi-input multi-output MIMO frequency selective channel the solution requires the derivation of the eigenvalue decomposition of the MIMO frequency response which, for every frequency bin, have generalized Wishart distribution. This paper shows the methodology used to derive the statistics of eigenvalues and eigenvectors and applies this methodology to the derivation of the average channel Capacity and of its characteristic function. The bound on the outage Capacity is then obtained using the characteristic function. Simple expressions are derived for the case of uncorrelated Rayleigh fading and for an arbitrary finite number of transmit and receive antennas.

Keywords

Capacity, Information Theory, Statistics and Frequency Selective Channel

I. Introduction

Extensive efforts are being made to improve the spectral efficiency of communication channels. To this end, the study of MIMO channels has gained prominent attention. Rather than increasing the bandwidth of the channel, which is quite an expensive treatment, the MIMO architecture is employed to exploit the propagation diversity. This calls for the study of the channel's limitations. Hence, the analysis of the Capacity of the multiple-antenna channels is worth a task and has already attracted a number of researchers.

Telatar [1], Foschini in [2] and Marzetta and Hochwald in [3] studied the Capacity MIMO flat fading channels by using a block fading model. Specifically, Telatar [1] investigated the use of multiple antennas at both the transmitting and the receiving ends of a single user channel and found that the Capacity gain was considerable under the assumption of independently faded channels, compared to single-antenna environment. When the channel is known at both the transmitter and the receiver side, it was shown that the Capacity in Rayleigh flat-fading increases linearly with the minimum between the number of transmitters and receivers. The Rayleigh case allows to use several closed form results (see also e.g. [4]).

Zheng and Tse in [5] employed the same model to study the Capacity of the channel as a function of the number of transmitter and receiver antennas at high SNR. They

used the geometric interpretation of the Capacity expression as the sphere packing in the Grassmann manifold to obtain the ergodic (mean) Capacity expression for arbitrary number of transmitter and receiver antennas for the case of flat fading channels.

This paper is concerned mainly with the derivation of the statistics of the MIMO frequency selective channel Capacity. Deriving the channel Capacity for MIMO systems requires the non trivial step of deriving the joint statistics of the eigenvalues of the random MIMO frequency response. We use the insight similar to the one used by Zheng and Tse [5] and view the matrix factorization as mere change of variables. The contribution of this paper is twofold: 1) the condensed review on the key tools and results in the study of complex random matrices, 2) the derivation of the channel Capacity and of its characteristic function, that allows us to calculate the Chernoff bound on the outage Capacity. As part of our overview on the random matrix analysis, in Section III we will present the rules of *exterior differential calculus* which is used to compute the Jacobian of matrix decompositions and perform integration over matrix groups.

Contrary to other authors, who have provided asymptotic results for similar problems [see e.g. [3], [6]] the analysis developed in this paper applies for an arbitrary finite number of inputs and outputs and our review paves the road for the derivation of other performance measures which depend on the factors of MIMO channel decompositions. Unlike [1] and [5], we analyze the channel which is frequency selective and look at the outage Capacity rather than ergodic Capacity. Our results are particularly useful in the context of the broad-band multi-carrier Space-Time communications for wireless local area networks (WLAN), where the number of carriers will be relatively high but number of input and output antennas is naturally going to be limited to a few elements [7]. For the case of Rayleigh fading, we provide simplified expressions for the characteristic function, which are useful to provide the Chernoff bound for the outage Capacity. Finally, we present numerical examples that support the theoretical results obtained.

Notation: Boldface letters are vectors (lower case) or matrices (upper case). The $tr(\mathbf{A})$, $|\mathbf{A}|$, $\lambda(\mathbf{A})$ are the trace, determinant and eigenvalues of $\mathbf{A}$, $\mathbf{a} = vec(\mathbf{A})$ is formed stacking vertically the columns of $\mathbf{A}$. Continuous time signals vectors are like $\mathbf{a}(t)$ discrete time vector sequences like $\mathbf{a}[n]$. Sequences of vectors obtained by stacking consecutive

blocks, such as $\mathbf{a}_i = [\mathbf{a}[iM], \ldots, \mathbf{a}[iM + M - 1]]$, are characterized by a suffix i. To manipulate blocked matrices we introduce vectors of indices $\mathbf{k} = (k_1, \ldots, k_m)$ and the notation $\mathbf{A}[\mathbf{k}] \equiv (\mathbf{A}[k_1]^H, \ldots, \mathbf{A}[k_m]^H)^H$.

II. System Model

The system considered has N_T transmit and N_R receive antennas. The baseband equivalent transmitted signal is the vector $\boldsymbol{x}(t) := (x_1(t), \ldots, x_{N_T}(t))^T$ of complex envelopes emitted by the transmit antennas. We assume a digital link with linear modulation so that the vector $\mathbf{x}(t)$ is related to the (coded) symbol vector $\mathbf{x}[n]$ by

$$\boldsymbol{x}(t) = \sum_{n=-\infty}^{+\infty} \boldsymbol{x}[n] g_T(t - nT), \tag{1}$$

where $g_T(t)$ is the transmit pulse. Correspondingly, $\boldsymbol{z}(t) = \boldsymbol{y}(t) + \boldsymbol{n}(t)$ is the received $N_R \times 1$ vector which contains the channel output $\mathbf{y}(t)$ and additive noise $\boldsymbol{n}(t)$. For a linear (generally time-varying) channel, the input-output (I/O) relationship can be cast in the form of an integral equation

$$\boldsymbol{y}(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_R(t - \theta) \mathbf{H}(\theta, \tau) \mathbf{x}(\theta - \tau) d\tau d\theta. \tag{2}$$

where $g_R(t)$ is the impulse response of the receive filter (usually a square-root raised cosine filter) matched to the transmit filter $g_T(t)$, and the (k, l)th entry of matrix $\mathbf{H}(\theta, \tau)$ is the impulse response of the channel between the l-th transmit and the k-th receive antennas. Introducing the discrete-time time-varying impulse response

$$\mathbf{H}[k, n] \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{H}(\theta, \tau) g_T(\theta - \tau - (k - n)T) g_R(kT - \theta) d\tau d\theta, \tag{3}$$

we can write the vector of received samples $\boldsymbol{y}[k] := \boldsymbol{y}(kT)$ as

$$\boldsymbol{y}[k] = \sum_{n=-\infty}^{\infty} \mathbf{H}[k, k - n] \boldsymbol{x}[n]. \tag{4}$$

If the channel $\mathbf{H}[k, n]$ is causal and has finite memory L we can write the I/O relationship (4) as a finite linear system of equations. Specifically, stacking $P = K + L$ transmit snapshots in a $PN_T \times 1$ vector $\mathbf{x}_i \equiv vec([\boldsymbol{x}[iP], \ldots, \boldsymbol{x}[iP + P - 1]])$ and K received snapshots in a $KN_R \times 1$ vector $\mathbf{y}_i \equiv vec([\boldsymbol{y}[iP + L], \ldots, \boldsymbol{y}[iP + P - 1]])$, where $\mathbf{y}_i$ starts

from the Lth array snapshot so that the inter-block interference (IBI) is not considered, we have

$$\mathbf{y}_i = \mathbf{H}\ \mathbf{x}_i, \tag{5}$$

where $\mathbf{H}$ is an $N_R K \times N_T P$ block-Toeplitz matrix:

$$\mathbf{H} = \begin{pmatrix} \boldsymbol{H}[L] & \cdots & \boldsymbol{H}[0] & 0 & \dots & 0 \\ 0 & \boldsymbol{H}[L] & \cdots & \boldsymbol{H}[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \boldsymbol{H}[L] & \cdots & \boldsymbol{H}[0] \end{pmatrix}_{N_R K \times N_T P} . \tag{6}$$

Assuming that the Gaussian additive noise is spatially and temporally white, space-time OFDM will convert our frequency selective MIMO system into a set of K parallel independent MIMO systems. In fact, if the channel matrix $\boldsymbol{H}$ is sandwiched between the two matrices

$$\mathbf{E}_T \equiv (\bar{\mathbf{W}}_K \otimes \mathbf{I}_{N_T \times N_T}) \quad , \ \mathbf{E}_R \equiv (\mathbf{W}_K^H \otimes \mathbf{I}_{N_R \times N_R}) \tag{7}$$

where $\bar{\mathbf{W}}_{K+L,K}$ is an extended $(K+L) \times K$ IFFT matrix, i.e., $\{\mathbf{W}_K\}_{k,n} := e^{j2\pi \frac{k(n-L)}{K}}$, $k = 0, \dots, K-1$ and $n = 0, \dots, P-1$ with a proper phase shift that creates the so called cyclic prefix, and $\mathbf{W}_K$ is the $K \times K$ IFFT matrix, i.e., $\{\mathbf{W}_K\}_{k,n} := e^{j2\pi \frac{kn}{K}}$ $k = 0, \dots, K-1$ and $n = 0, \dots, K-1$, then similar to what happen in the scalar case, the equivalent channel is:

$$\tilde{\mathbf{H}} \ \equiv \ \mathbf{E}_R \mathbf{H} \mathbf{E}_T = diag(\tilde{\mathbf{H}}[\mathbf{k}]) \ , \quad \mathbf{k} \equiv (0, \dots, K-1),$$

where $\tilde{\boldsymbol{H}}[k]$ is the MIMO transfer function at the kth frequency bin:

$$\tilde{\boldsymbol{H}}[k] = \sum_{l=0}^{L} \mathbf{H}[l] e^{-j2\pi \frac{kl}{K}}. \tag{8}$$

Channel modelling and performance analysis over fading wireless channels have been studied extensively and in numerous cases the receiver performance can be expressed in closed form (see e.g. [8]). Most of the results apply to narrow-band SISO/SIMO transmission. In this context the channel model is often expressed only in terms of the statistics of the fading envelope $\alpha_{r,t}[l] \equiv |\{\boldsymbol{H}[l]\}_{r,t}|$ of each path coefficient for the (r,t) link. The interesting and challenging aspect of the MIMO case is that the performance is expressed in

terms of the eigenvalues of the matrix $\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}$ and thus the results for the scalar case are not generalized in a straightforward way to MIMO systems. The goal of this paper is twofold: we will first describe how the statistics of the eigenvalues of $\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}$ are linked to the joint statistics of $\mathbf{H}(\mathbf{d}) = (\boldsymbol{H}^H[0], \ldots, \boldsymbol{H}^H[L])^H$ and then we will specialize the analysis to the case of wide-sense stationary Rayleigh fading, deriving the statistics of the channel Capacity (C) for an arbitrary number of transmit and receive antennas. Prior to this we will provide an overview on *exterior differential forms* which explain the derivations done in the following.

III. Elements of exterior differential forms

The study of random eigenvalues, initiated by the pioneering work of Wigner [9], provides a wide range of tools to analyze the statistics of several matrix factorizations beside the eigenvalues decomposition (EVD). The first step in deriving the statistics of the resulting matrices consists in deriving the Jacobian of the change of variables from the original matrix to its factors. When the decomposition is unique (at least up to a sign and permutation), the number of independent variables in the matrix and in the corresponding factors is the same and the Jacobian matrix is square. This can be verified to be true in the case of EVD (eigenvalue), QR or LU (lower-upper) decompositions and Cholesky decomposition for example [10].

To keep the presentation self-cointained, next we informally introduce some of the concepts used in the statistical analysis of random matrices (see e.g. [11]). One of such tools is based on the seminal work of Élie Cartan on *exterior differential calculus* [12]. The concept of *exterior product*, denoted by the symbol $\wedge$, was introduced by Hermann Günter Grassmann in 1844 and was utilized by Cartan in the study of differential forms. Ordinary vectors are 1-vectors, wedge products of p independent vectors generates the space of p-vectors. Given vectors α, β, γ the basic axioms of Grassman algebra are:

- Associativity: $(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma)$
- Anti-Commutativity: $\alpha \wedge \beta = -\beta \wedge \alpha$
- Distributivity: $(a\alpha + b\beta) \wedge \gamma = a(\alpha \wedge \gamma) + b(\beta \wedge \gamma)$.

The axioms are sufficient to establish that:[1]

$$\alpha \wedge \alpha = 0 \quad \text{and} \quad (\mathbf{A}\alpha) \wedge \beta = |\mathbf{A}|(\alpha \wedge \beta). \tag{9}$$

Cartan's exterior differential calculus (a very clear book is [12]) is built around the observation that, if we do consider the sign in the Jacobian, products of differentials $dxdy$ behave as $dx \wedge dy$: this can be easily observed introducing a dummy transformation $x(u,v), y(u,v)$ and realizing that $dxdy = |\partial(x,y)/\partial(u,v)|dudv$ equals 0 for $u = v = x$ and equals $-dydx$ for $u = y$ and $v = x$. The rules of exterior differential calculus are derived by applying Grassman algebra to 1-forms such as dx or the gradient of a differentiable function ∇f. An $r-$form is:

$$\alpha = \sum_{k_1<k_2\ldots<k_r} A(x_1,\ldots,x_n)dx_{k_1} \wedge \ldots \wedge dx_{k_r} \tag{10}$$

There is an axiomatic definition of the d operator and, in particular:

- $d(r\text{-form}) = (r+1)\text{-form}$
- $d(\alpha + \beta) = d\alpha + d\beta$
- if α is an $r-$form and β is an $s-$form $d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^r \alpha \wedge d\beta$
- $d(dx) = 0$ (Poincarè Lemma)

These rules are systematic and the results are simpler to grasp than the theory of manifolds. In addition, they provide a way of deriving the Jacobian of an arbitrary matrix factorization, by applying the d operator first and then evaluating the $\wedge$ product of all the independent differentials. This last task entails some additional complexity, because it requires the description of the group of matrices by mean of their independent parameters (see e.g. Section IV). The evaluation of this Jacobian is essential to derive the probability density function (pdf) of the factors from the pdf of the original matrix. We will borrow the notation from [10] and indicate by $d\mathbf{A}$ the matrix of differentials and by $(d\mathbf{A})$ the exterior product of the independent entries in $d\mathbf{A}$, for example:

- for an arbitrary $\mathbf{A}$, $(d\mathbf{A}) = \wedge_i \wedge_j da_{ij}$
- if $\mathbf{A}$ is diagonal $(d\mathbf{A}) = \wedge_i da_{ii}$
- if $\mathbf{A} = \mathbf{A}^T$ or $\mathbf{A}$ is lower triangular $(d\mathbf{A}) = \wedge_{1\le i\le j\le n} da_{ij}$

[1]If $\mathbf{A}$ is $m \times n$ and $m > n$ or if it is rank deficient $|\mathbf{A}|$ has to be replaced by 0. If $m \le n$, $|\mathbf{A}|$ has to be replaced by the matrix compound $\wedge^m \mathbf{A}$, i.e. the matrix of all cofactors of order m, if $m \le n$ [12].

• see Section IV for $\mathbf{Q}$ unitary.

When dealing with complex matrices we can apply the same rules remembering that any complex dz has associated a $(dz) = d\Re[z]d\Im[z]$ or, more precisely, $(dz) = d\Re[z] \wedge d\Im[z]$. Therefore dz can be treated as a bidimensional vector. Since the multiplication of $z = x + jy$ by a complex number $\alpha = a + j\ b$ can be viewed as:

$$(x,\ y) \begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \tag{11}$$

from (9) it follows that $(d\alpha z) = |\alpha|^2 dxdy$. In general [11]:

Lemma 1: If $\mathbf{w}=\mathbf{u}+j\mathbf{v}$ are analytical functions of $\mathbf{z}=\boldsymbol{x}+j\boldsymbol{y}$ then

$$det\left(\frac{\partial(\mathbf{u},\mathbf{v})}{\partial(\boldsymbol{x},\boldsymbol{y})}\right) = \left|det\left(\frac{\partial \mathbf{w}}{\partial \mathbf{z}}\right)\right|^2 \tag{12}$$

Other properties of the complex case are easily derived, for example: i) $(dz) = -(dz^*)$; ii) $dz \wedge dz^* = 0$. Note that for $\mathbf{B} = \mathbf{XA}$ $(d\mathbf{B}) = |\mathbf{X}|^n (d\mathbf{A})$ in $\mathbb{R}^{\mathrm{n}}$ (the absolute value square of $|\mathbf{X}|$ in $\mathbb{C}^{\mathrm{n}}$). Because of (9) and Lemma 1, orthogonal or unitary linear mappings of $\mathbf{A}$ do no not change $(d\mathbf{A})$, i.e. if $\mathbf{Q}^H\mathbf{Q} = \mathbf{I}$ $(\mathbf{Q}^H d\mathbf{A}) \equiv (d\mathbf{A})$.

IV. The Stiefel Manifold

In the description of the joint distribution of matrix decompositions such as the QR the EVD etc., there is the clear need of identifying what is $(d\mathbf{Q})$. A unitary $\mathbf{Q}$ can be described by n^2 smooth functions that can be integrated over nice enough intervals which describe the so called Stiefel Manifold: clearly, the independent parameters of the Stiefel Manifold are not the real and imaginary parts of the elements of $\mathbf{Q}$, which would be $2n^2$ parameters. For the purpose of studying the statistics of matrix decompositions, such as the QR or the EVD, n out of the n^2 parameters are redundant (in the sense that the decomposition is unique up to n parameters). This ambiguity could be removed by having the diagonal elements of $\mathbf{Q}$ set to be real for example. Note that, because of $\mathbf{QQ}^H = \mathbf{I} \ \rightarrow \ \mathbf{Q}d\mathbf{Q}^H = -d\mathbf{QQ}^H$: thus, $\mathbf{Q}d\mathbf{Q}^H$ is antisymmetric and the diagonal elements of $\mathbf{Q}d\mathbf{Q}^H$ are purely imaginary. Note also that, when $\mathbf{Q}$ is $m \times n$ and semi-unitary with $n \leq m$, we have $2mn - n(n+1)$ real parameters (the roles are reversed if $n > m$) and we can always define an $m \times m$ matrix $\bar{\mathbf{Q}} = (\mathbf{Q}, \mathbf{Q}^{\perp})$ such that $\bar{\mathbf{Q}}^H\mathbf{Q} = \mathbf{I}_{m,n}$, so that $(d\mathbf{Q}) = (\bar{\mathbf{Q}}^H d\mathbf{Q})$.

Several different approaches can be taken to represent $\mathbf{Q}$ using n^2 independent parameters, for example:

- $\mathbf{Q}$ is product of Givens rotations [Ch.5 [13]], i.e. for $\mathbf{Q}$ $n \times n$:

$$\mathbf{Q} = \prod_{k=1}^{n} \prod_{i=1}^{k} \mathbf{G}(k,i) \tag{13}$$

each $\mathbf{G}(k,i)$ is all zero except for the sub-matrix formed with the $(k,k),(i,i),(k,i),(i,k)$ elements[2].

- $\mathbf{Q}$ is product of n Householder rotations $\mathbf{H}_i = \mathbf{I} - 2\mathbf{v}_i\mathbf{v}_i^H/(\mathbf{v}_i^H\mathbf{v}_i)$ [Ch.5 [13]], where for $i = 0, \ldots, n-1$ each $\mathbf{v}_i$ is described by $n-i$ complex parameters.
- Decomposing $\mathbf{Q} = \mathbf{\Omega}_1\mathbf{D}\mathbf{\Omega}_2$, with $\mathbf{\Omega}_i$, $i = 1, 2$ orthogonal matrices and $\mathbf{D} = diag(e^{j\phi_1}, \ldots, e^{j\phi_n})$.
- Using $\mathbf{Q} = e^{j\mathbf{\Theta}}$ where $\mathbf{\Theta}$ is Hermitian (the description is unique $\forall$ $\mathbf{\Theta}$: $\mathbf{0} \leq \mathbf{\Theta} \leq \pi\mathbf{I}$).
- For $\mathbf{Q}$ not having eigenvalues equal to -1 (a probability zero event for continuous random $\mathbf{Q}$), the Cayley transform $\mathbf{Q} = (\mathbf{I}+j\mathbf{S})^{-1}(\mathbf{I}+j\mathbf{S})$ where $\mathbf{S}$ is skew Hermitian, i.e. $\mathbf{S}^H = -\mathbf{S}$. Note that $\mathbf{S} = (\mathbf{I}+\mathbf{Q})^{-1}(\mathbf{I}-\mathbf{Q})$.

The 3×3 orthogonal matrix case is illustrated in Fig. 1. The uniform p.d.f. in the Stiefel group of orthogonal or unitary matrices is called *Haar distribution* [11, Ch.1]. If $\mathbf{Q}$ is $n \times n$ and is decomposed as in (13) with Euler angles θ_{ik} that define $\mathbf{G}(k,i)$ uniquely (the number of independent parameters of $\mathbf{Q}$ in this case is $n(n-1)/2$):

$$p_{\mathbf{Q}}(\mathbf{Q})(\boldsymbol{Q}^T d\boldsymbol{Q}) = \prod_{k=1}^{n} \frac{\Gamma(k/2)}{2(\pi)^{k/2}} \prod_{k=1}^{n-1} \prod_{i=1}^{k} \sin^{i-1}(\theta_{ik}) \bigwedge_{k=1}^{n-1} \bigwedge_{i=1}^{k} d\theta_{ik} \tag{14}$$

where $\Gamma(x)$ is the Gamma function and $2(\pi)^{k/2}/\Gamma(k/2)$ is the volume of a unit sphere in $\mathbb{R}^{\mathrm{k}}$ and $\theta_{1k} \in [0, 2\pi)$, $\theta_{ik} \in [0, \pi)$ $i = 2, \ldots, k;$ $k = 1, \ldots, n-1$.

To determine the volume of $(\mathbf{Q}^H d\mathbf{Q})$ in $\mathbb{C}^{\mathrm{m}}$ integrated over $\mathbf{Q}^H\mathbf{Q} = \mathbf{I}$, we can observe, as in [10], that the QR decomposition of an m vector $\mathbf{z} = r\mathbf{q}$ is trivially given by $r = \sqrt{\mathbf{z}^H\mathbf{z}}$

[2]

$$\begin{pmatrix} \{\mathbf{G}(k,i)\}_{i,i} & \{\mathbf{G}(i,k)\}_{i,k} \\ \{\mathbf{G}(k,i)\}_{k,i} & \{\mathbf{G}(i,k)\}_{k,k} \end{pmatrix} = \begin{pmatrix} c & -s \\ s^* & c^* \end{pmatrix}.$$

which is described by one parameter (the Euler angle) when $\mathbf{Q}$ is orthogonal ($c = \cos\theta_{ik}, s = \sin\theta_{ik}$). When it is unitary the extra constraint is that $|c|^2 + |s|^2 = e^{j\phi_i k}$, hence the parameters are 4, 3 if unimodular (i.e. $|c|^2 + |s|^2 = 1$).

and $\mathbf{q} = \mathbf{z}/r$. Note that $(d\mathbf{q}) = \wedge_{i=2}^{n} dq_i$ because only $n-1$ parameters are independent since $\|\mathbf{q}\| = 1$. Since $(d\mathbf{z}) = (dr\mathbf{q} + rd\mathbf{q})$, we can write $(d\mathbf{z}) = r^{2n-1}dr(d\mathbf{q})$, where $(d\mathbf{q}) = (1 - \sum_{i=2}^{n} |q_i|^2)u_1(1 - \sum_{i=2}^{n} |q_i|^2) \wedge_{i=2}^{n} dq_i^2$ and therefore

$$\int (d\mathbf{q}) = \frac{\int_{\mathbb{C}^m} e^{-|z|^2/2}(dz)}{\int_0^\infty r^{2m-1}e^{-r^2/2}dr} = \frac{(2\pi)^m}{2^{m-1}\Gamma(m)}. \tag{15}$$

This approach of finding the volume element of a unit sphere can be extended to the case of matrices and hence can be employed to find the element of volume of a unitary group. As mentioned before the element of volume of Stiefel manifold is given by $(d\mathbf{Q}) = (\bar{\mathbf{Q}}^H d\mathbf{Q})$. Extending the ideas given by Edelman[3] [10] to the case of unitary group, the element of volume can be found to be

$$(\bar{\mathbf{Q}}^H d\mathbf{Q}) = \bigwedge_{i \geq j} \boldsymbol{q}_i^H d\boldsymbol{q}_j \tag{16}$$

where $\bar{\mathbf{Q}} = [\boldsymbol{q}_1 \cdots \boldsymbol{q}_m]$ is the same as before and $\boldsymbol{q}_i$ is a complex unit vector. The details are given in the Appendix -A.

The volume of $(\bar{\mathbf{Q}}^H d\mathbf{Q})$ integrated over $\mathbf{Q}^H\mathbf{Q} = \mathbf{I}$, for $\mathbf{Q}$ unitary, is:

$$Vol(\mathbf{Q}_{m,n}) = \frac{2^n(\pi)^{mn-n(n-1)/2}}{\prod_{i=0}^{n-1}\Gamma(m-i)}. \tag{17}$$

when the n constraints are added to $\mathbf{Q}_{m,n}$ (for example the diagonal elements are constrained to be real):

$$\overline{Vol}(\mathbf{Q}_{m,n}) = \frac{(\pi)^{(m-1)n-n(n-1)/2}}{\prod_{i=0}^{n-1}\Gamma(m-i)}. \tag{18}$$

V. The statistics of $\boldsymbol{A} = \boldsymbol{B}^H\boldsymbol{B}$ and its EVD

The matrix we are interested in has the form $\boldsymbol{A} = \boldsymbol{B}^H\boldsymbol{B}$, where $\boldsymbol{B}$ is a random $m \times n$ matrix with continuous p.d.f and we will assume that $m \geq n$ in which case $\boldsymbol{A}$ is full rank with probability one.[4] Let us denote by $p_{\boldsymbol{A}}(\boldsymbol{A})$ and $p_{\boldsymbol{B}}(\boldsymbol{B})$ the pdfs of the random matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively: the pdf of $\boldsymbol{A}$ is called generalized Wishart distribution. To derive $p_{\boldsymbol{A}}(A)$ one can follow the approach in [14] which is based on the QR and Cholesky

[3]In [10] the derivation of the volume element of the group of orthogonal matrix is found to be $(\bar{\mathbf{Q}}^T d\mathbf{Q}) = \bigwedge_{i>j} \boldsymbol{q}_i^T d\boldsymbol{q}_j$. Note that the elements $\boldsymbol{q}_i^T d\boldsymbol{q}_i = 0$ for the real case.

[4]In case $m < n$ $\boldsymbol{A}$ has $n-m$ zero eigenvalues. Because the non null eigenvalues of $\boldsymbol{B}^H\boldsymbol{B}$ and $\boldsymbol{B}\boldsymbol{B}^H$ coincide, the case $m \geq n$ is general enough to provide the distribution of the non zero eigenvalues for any choice of n, m.

decompositions of $\boldsymbol{B}$ and $\boldsymbol{A}$ respectively:

$$\boldsymbol{B} = \boldsymbol{Q}\boldsymbol{R}\ , \qquad \boldsymbol{A} = \boldsymbol{R}^H\boldsymbol{R}. \tag{19}$$

Considering that $(d\boldsymbol{A}) = (d\boldsymbol{R}^H\boldsymbol{R} + \boldsymbol{R}^H d\boldsymbol{R})$:

$$\begin{aligned}(d\boldsymbol{A}) &= \bigwedge_{1\le i<j\le n}\left(\sum_{1\le k\le i} dr^*_{k,i}r_{k,j} + dr^*_{k,j}r_{k,i}\right)\\ &= 2^n\prod_{i=1}^{n}(|r_{ii}|)^{2(n-i)+1}(d\boldsymbol{R})\end{aligned} \tag{20}$$

with $(d\boldsymbol{R}) = \wedge_{i<j}(dr_{ij})$. Therefore:

$$p_{\boldsymbol{A}}(\boldsymbol{A})(d\boldsymbol{A}) = p_{\boldsymbol{A}}(\boldsymbol{R}^H\boldsymbol{R})\prod_{i=1}^{n} 2^n\left(|r_{ii}|\right)^{2(n-i)+1}(d\boldsymbol{R}). \tag{21}$$

For QR factorization to be unique, we constrain the diagonal elements of $\boldsymbol{R}$ to be real (the number of independent parameters of $\boldsymbol{B}$ now equals those of $\boldsymbol{Q}$ and $\boldsymbol{R}$ put together). Denoting by $\bar{\boldsymbol{Q}} = (\boldsymbol{Q}, \boldsymbol{Q}^{\perp})$ the $m \times m$ matrix such that $\bar{\boldsymbol{Q}}^H\boldsymbol{Q} = \boldsymbol{I}_{m,n}$ has the top $n \times n$ portion equal to an identity matrix and the bottom $m - n$ rows equal to zero, $(d\boldsymbol{B})$ $= (\bar{\boldsymbol{Q}}^H d\boldsymbol{B}) = (\bar{\boldsymbol{Q}}^H d\boldsymbol{Q}\boldsymbol{R} + \boldsymbol{I}_{m,n}d\boldsymbol{R})$, taking the wedge product we have (the details can be found in the Appendix -A):

$$(d\boldsymbol{B}) = \prod_{i=1}^{n}(|r_{ii}|)^{2(m-i)+1}(d\boldsymbol{R})(d\boldsymbol{Q}), \tag{22}$$

where $(d\boldsymbol{Q}) = (\bar{\boldsymbol{Q}}^H d\boldsymbol{Q})$ is the element of volume of the Stiefel manifold. Hence:

$$p_{\boldsymbol{B}}(\boldsymbol{B})(d\boldsymbol{B}) = p_{\boldsymbol{B}}(\boldsymbol{Q}\boldsymbol{R})\prod_{i=1}^{n}\left(|r_{ii}|\right)^{2(m-i)+1}(d\boldsymbol{R})(d\boldsymbol{Q}), \tag{23}$$

and, with $\sqrt{\boldsymbol{A}} \triangleq \boldsymbol{R}$, from (21) and (23) and $|\boldsymbol{A}| = \prod_{i=1}^{n}|r_{ii}|^2$ it follows:

$$p_{\boldsymbol{A}}(\boldsymbol{A}) = 2^{-n}|\boldsymbol{A}|^{m-n}\int p_{\boldsymbol{B}}(\boldsymbol{Q}\sqrt{\boldsymbol{A}})(\bar{\boldsymbol{Q}}^H d\boldsymbol{Q}), \tag{24}$$

which is the form of the so called *generalized Wishart density* [11]. Generalizing the results in [11] to the complex case (21) implies:

Lemma 2: When the p.d.f. $p_{\boldsymbol{B}}(\boldsymbol{B}) = p(\boldsymbol{B}^H\boldsymbol{B})$ then:

1) $\boldsymbol{Q}$ and $\boldsymbol{R}$ in the QR decomposition $\boldsymbol{B} = \boldsymbol{Q}\boldsymbol{R}$, are independent. The p.d.f. of $\boldsymbol{Q}$ is

uniform over the unit $\boldsymbol{Q}\boldsymbol{Q}^H = \boldsymbol{I}$ (*Haar distribution*) and $\boldsymbol{R}$ is

$$p_{\boldsymbol{R}}(\boldsymbol{R}) = \prod_{i=1}^{n} \left(|r_{ii}|^2\right)^{m-n} p(\boldsymbol{R}^H\boldsymbol{R}) Vol(\boldsymbol{Q}_{m,n}); \tag{25}$$

2) The p.d.f. of $\boldsymbol{A}$ is [c.f. (17)]:

$$p_{\boldsymbol{A}}(\boldsymbol{A}) = 2^{-n}|\boldsymbol{A}|^{m-n} p(\boldsymbol{A}) Vol(\boldsymbol{Q}_{m,n}), \tag{26}$$

The Jacobian of the EVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathbf{H}}$ can also be obtained by fixing the diagonal element of $\boldsymbol{U}$ to be real so that the EVD is unique:

$$\begin{aligned} (d\boldsymbol{A}) &= (d\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H + \boldsymbol{U}d\boldsymbol{\Lambda}\boldsymbol{U}^H + \boldsymbol{U}\boldsymbol{\Lambda}d\boldsymbol{U}^H) & (27)\\ (d\boldsymbol{A}) &\equiv (\boldsymbol{U}^H d\boldsymbol{A}\boldsymbol{U}) = (\boldsymbol{U}^H d\boldsymbol{U}\boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{U}^H d\boldsymbol{U} + d\boldsymbol{\Lambda}) \\ &= \prod_{1\le i<k\le n} (\lambda_k - \lambda_i)^2 (d\boldsymbol{\Lambda})(\boldsymbol{U}^H d\boldsymbol{U}). & (28) \end{aligned}$$

Equations (21) and (28) are the equations that can be used to address the general case of $\boldsymbol{A} = \boldsymbol{B}^H\boldsymbol{B}$:

$$\begin{aligned} p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}) &= 2^{-n}\prod_{1\le i<k\le n} (\lambda_k - \lambda_i)^2 \left(\prod_{i=1}^{n}\lambda_i\right)^{m-n} \Psi(\lambda_1,\ldots,\lambda_n) & (29)\\ \Psi(\lambda_1,\ldots,\lambda_n) &\triangleq \int p_{\boldsymbol{B}}(\boldsymbol{Q}\sqrt{\boldsymbol{\Lambda}}\boldsymbol{U}^H)(\bar{\boldsymbol{Q}}^H d\boldsymbol{Q})(\boldsymbol{U}^H d\boldsymbol{U}). & (30) \end{aligned}$$

When in Lemma 2 $p(\boldsymbol{A}) \equiv p(\boldsymbol{\Lambda})$, the density of the eigenvalues is simple to derive: for example, in the multivariate Gaussian case $\{\boldsymbol{B}\}_{i,j} \sim \mathcal{N}(0,\sigma^2)$, $p(\boldsymbol{A}) = (\pi\sigma^2)^{-mn}\exp(-\frac{tr(\boldsymbol{A})}{\sigma^2})$ [c.f. (26)] and, for $\lambda_i > 0$:

$$p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}) = \chi_1 \prod_{1\le i<k\le n} (\lambda_k - \lambda_i)^2 e^{-\frac{\sum_i \lambda_i}{\sigma^2}} \left(\prod_{i=1}^{n}\lambda_i\right)^{m-n} \tag{31}$$

where $\chi_1 = 2^{-n}(\pi\sigma^2)^{-mn} Vol(\boldsymbol{Q}_{m,n})\overline{Vol}(\boldsymbol{U}_{n,n})$.

Using Wigner's approach, the density function obtained by averaging over all permutations $p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda})$ is $\frac{1}{n!}p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda})$, thus [15]:

Lemma 3: For $m \ge n$ and any continuous real$f(\boldsymbol{A}) = \sum_{i=1}^{n} f(\lambda_i(\boldsymbol{A}))$

$$\begin{aligned} E\{f(\boldsymbol{A})\} &= \int_0^\infty f(x)\mu_n^{m-n}(x)dx & (32)\\ \mu_n^{m-n}(x) &\triangleq \frac{1}{n!}\int_0^\infty \cdots \int_0^\infty p_{\boldsymbol{\Lambda}}(x,\lambda_2,\ldots,\lambda_n)d\lambda_2\ldots d\lambda_n. & (33) \end{aligned}$$

Note that, for $f(\boldsymbol{A}) = \sum_{i=1}^{n} \delta(x - \lambda_i(\boldsymbol{A}))$, $E\{f(\boldsymbol{A})\}$ in (32) is the *empirical distribution* of the eigenvalues or, in other words, the average histogram of the eigenvalues of random matrix samples.

When $p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda})$ is as in (31) [16], with $\alpha = m - n$:

$$\mu_n^{\alpha}(x) = \frac{1}{n} \sum_{k=0}^{n-1} \phi_k^{\alpha}(x)^2 \tag{34}$$

where, denoting by $L_k^{\alpha}(x)$ the Laguerre polynomials of order α

$$\phi_k^{\alpha}(x) = \left[\frac{k!}{\Gamma(k+\alpha+1)} x^{\alpha} e^{-x} \right]^{1/2} L_k^{\alpha}(x). \tag{35}$$

VI. Statistics of the the MIMO frequency selective channel

We will assume that:

a1. The noise is AWGN with variance $\sigma_n^2 = 1$

and some of the results will be given for the case where:

a2. $\{\boldsymbol{H}[l]\}_{r,t}^{*}$ are spatially uncorrelated circularly symmetric zero mean complex Gaussian random variables (Rayleigh fading) with $\boldsymbol{R}_H[l_1, l_2, r_1, r_2, t_1, t_2] \triangleq E\{\{\boldsymbol{H}[l_1]\}_{r_1,t_1}^{*} \{\boldsymbol{H}[l_2]\}_{r_2,t_2}\} = \delta(t_1 - t_2)\ \delta(r_1 - r_2) \boldsymbol{R}_H(l_2, l_1)$.

Let us also denote by:

$$n \triangleq \min(N_T, N_R)\ , \quad m \triangleq \max(N_T, N_R). \tag{36}$$

In the MIMO case described in Section II, denoting by γ the signal to noise ratio dictated by the large-scale fading and receiver noise power, the conditional channel Capacity is:

$$C = \log |\boldsymbol{I} + \gamma \tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}}| \tag{37}$$

therefore the average Capacity is:

$$E\{C\} = \sum_{k=0}^{K-1} \sum_{l=1}^{n} E\left\{\log(1 + \gamma \lambda_l[k])\right\}. \tag{38}$$

and the characteristic function of C is:

$$\Phi_C(s) = E\{e^{sC}\} = E\left\{ \prod_{k=0}^{K-1} |\boldsymbol{I} + \gamma \tilde{\boldsymbol{H}}^H[k] \tilde{\boldsymbol{H}}[k]|^s \right\} \tag{39}$$

both functions of the eigenvalues of $\tilde{\boldsymbol{H}}[k]^H\tilde{\boldsymbol{H}}[k]$, $k=0,\ldots,K-1$. The average Capacity can be expressed in integral form and solved in the case when **a2** applies. In fact, $\tilde{\boldsymbol{H}}[k]$ is given by (8) thus, under **a2**, $\tilde{\boldsymbol{H}}[k]$, $k,0,\ldots,K-1$ are also zero mean complex Gaussian with variance:

$$\sigma_H^2[k]=\sum_{(l_1,l_2)=0}^{L}\boldsymbol{R}_H(l_1,l_2)e^{-j2\pi\frac{(l_1-l_2)k}{K}} \tag{40}$$

as a direct consequence of Lemma 3 we can write:

Corollary 1: The average Capacity for any (n,m) in (36) is (Telatar in [1] derived the expression of mean Capacity for the MIMO channel with Rayleigh faded coefficients):

$$E\{C\}=\sum_{k=0}^{K-1}\int_0^\infty \log\left(1+\gamma\sigma_H^2[k]x\right)\mu_n^{m-n}(x)dx \tag{41}$$

where $\mu_n^{m-n}(x)$ is given by (29), (32) and under **a2**, $\mu_n^{m-n}(x)$ is given by (34).

The derivation of $\Phi_C(s)$ is in general more complicated, since it requires averaging over the joint density of the eigenvalues of all $\tilde{\boldsymbol{H}}[k]^H\tilde{\boldsymbol{H}}[k]$, $k=0,\ldots,K-1$ and the matrices are dependent. However, it is worth noticing that an approximate result for $\gamma\ll 1$ can be obtained quite easily:

Proposition 1: For $\gamma\ll 1$

$$\Phi_C(s)\approx E\left\{|\boldsymbol{I}+\gamma K\boldsymbol{H}^H[\mathbf{d}]\boldsymbol{H}[\mathbf{d}]|^s\right\} \tag{42}$$

The eigenvalues of the product $\boldsymbol{H}^H[\mathbf{d}]\boldsymbol{H}[\mathbf{d}]$ can be calculated as described in Section V. The interesting consequence of (42) is that at low SNR (in the so called *low power regime* [17]), the statistics of the Capacity of the frequency selective channel are approximately equivalent to the ones of a MIMO flat fading channel with $N_R(L+1)$ antennas rather than N_R antennas.

Proof: When $\gamma\ll 1$, $\prod_{k=0}^{K-1}\left|\boldsymbol{I}+\gamma\tilde{\boldsymbol{H}}[k]^H\tilde{\boldsymbol{H}}[k]\right|\approx\left|\boldsymbol{I}+\gamma\sum_{k=0}^{K-1}\tilde{\boldsymbol{H}}[k]^H\tilde{\boldsymbol{H}}[k]\right|$, therefore (39) is approximately:

$$\Phi_C(s)\approx E\left\{\left|\boldsymbol{I}+\gamma\sum_{k=0}^{K-1}\tilde{\boldsymbol{H}}^H[k]\tilde{\boldsymbol{H}}[k]\right|^s\right\}. \tag{43}$$

Recalling that $\tilde{\boldsymbol{H}}[k]$ is the Fourier transform of $\boldsymbol{H}[l]$, $\tilde{\boldsymbol{H}}[\boldsymbol{k}]=(\boldsymbol{W}_K\otimes\boldsymbol{I})\boldsymbol{H}[\mathbf{d}]$, where $\boldsymbol{W}_K$ is the $(K\times L+1)$ DFT matrix with elements $\{\boldsymbol{W}_K\}_{k,l}=\exp(-j2\pi kl/K)$, $k\in$

$[0, K-1]$, $l \in [0, L]$. Therefore, $\sum_{k=0}^{K-1} \tilde{\boldsymbol{H}}^H[k]\tilde{\boldsymbol{H}}[k] = \tilde{\boldsymbol{H}}^H[\boldsymbol{k}]\tilde{\boldsymbol{H}}[\boldsymbol{k}] = K\boldsymbol{H}^H[\mathbf{d}]\boldsymbol{H}[\mathbf{d}]$, which proves our statement. ■

Calculating the dimensions m and n in (36) with $N_R(L+1)$ in place of N_R, under **a.2** the approximate characteristic function in (42) can be expressed in a rather complex closed form [18], [19] which is the exact solution for the Rayleigh flat fading case:

$$\Phi_C(s) \approx \chi_3 |\boldsymbol{G}|, \tag{44}$$

where $\boldsymbol{G}$ is $n \times n$ and

$$\{\boldsymbol{G}\}_{i,j} = G(i+j-2) \qquad i,j = 1, \ldots, m, \tag{45}$$

with

$$\begin{aligned} G(k) &= \frac{1}{\Gamma(-s/\ln 2)} \gamma^{m-n-k-1} \Gamma(1+k+m-n)\Gamma(-1-k-m+n-s/\ln 2) \\ &\cdot \quad {}_1F_1(1+k+m-n, 2+k+m-n+s/\ln 2, \gamma^{-1}) \\ &+ \quad \gamma^{s/\ln 2}\Gamma(1+k+m-n+s/\ln 2)\ {}_1F_1(-s/\ln 2, -k-m+n-s/\ln 2, \gamma^{-1}) \end{aligned} \tag{46}$$

The above expression looks rather cumbersome to deal with. However, when γ is really small, further approximations allow us to obtain a simpler closed form expression for the characteristic function which can be handled analytically quite easily. The expression in (42) can be written as

$$\Phi_C(s) \approx E\left\{|1+\gamma K tr(\boldsymbol{H}^H[\mathbf{d}]\boldsymbol{H}[\mathbf{d}])|^s\right\} \tag{47}$$

$$= E\left\{e^{s\ln(1+\gamma K vec(\boldsymbol{H}[\mathbf{d}])^H vec(\boldsymbol{H}[\mathbf{d}]))}\right\} \tag{48}$$

where $\boldsymbol{H}[\mathbf{d}] \sim \mathcal{N}(0, \boldsymbol{R}_r \otimes \boldsymbol{R}_t \otimes \boldsymbol{R}_H)$. Approximating $\ln(1+x)$ as x, for small x, and using the multivariate Gaussian density for $\boldsymbol{H}[\mathbf{d}]$, we obtain the following

$$\Phi_C(s) \approx \frac{1}{|\boldsymbol{I} - \gamma s K \overline{\boldsymbol{R}}|}, \qquad \overline{\boldsymbol{R}} = \boldsymbol{R}_r \otimes \boldsymbol{R}_t \otimes \boldsymbol{R}_H \tag{49}$$

The Capacity in this case looks like a standard $\chi - square$ which, if the number of degrees of freedom is large enough, can be further approximated by a Gaussian distribution. The corresponding mean and variance for the Gaussian approximation are derived in the section VI-A.

To address the opposite case of high γ, we have to consider that $K \geq L$ and therefore (8) implies that the joint density of the MIMO channel response at all frequency bins is dependent. We can decompose $p_{\tilde{\boldsymbol{H}}}(\tilde{\boldsymbol{H}}[\mathbf{k}])$ as follows

$$p_{\tilde{\boldsymbol{H}}}(\tilde{\boldsymbol{H}}[\mathbf{k}]) = p(\tilde{\boldsymbol{H}}[\bar{\mathbf{p}}] \mid \tilde{\boldsymbol{H}}[\mathbf{p}])p(\tilde{\boldsymbol{H}}[\mathbf{p}]) \tag{50}$$

where $\mathbf{k} = (0, \ldots, K-1)$, $\mathbf{p} = (k_0, \ldots, k_L)$ is a vector that has, as elements, $L+1$ distinct, but otherwise arbitrary indices extracted from $\mathbf{k}$ and $\bar{\mathbf{p}}$ is the vector of the complementary indices. For every choice of the vector of frequency indices $\mathbf{p}$, the blocks of $\tilde{\boldsymbol{H}}[\mathbf{p}] = (\tilde{\boldsymbol{H}}^H[k_0], .., \tilde{\boldsymbol{H}}^H[k_L])^H$ are in a one to one mapping with the blocks of $\boldsymbol{H}[\mathbf{d}] = (\boldsymbol{H}^H[0], .., \boldsymbol{H}^H[L])^H$; in fact, (8) for each antenna pair represents a system of linear equations, each corresponding to a different index $k_i \in \mathbf{p}$, with coefficients forming a full rank Vandermonde matrix $\boldsymbol{W}_{L+1}(\mathbf{p})$:

$$\{\boldsymbol{W}_{L+1}(\mathbf{p})\}_{li} = e^{-j\frac{2\pi}{K}lk_i} \quad l \in [0, L],\ k_i \in \mathbf{p}, \quad i \in [0, L], \tag{51}$$

and we can write:

$$\tilde{\boldsymbol{H}}[\mathbf{p}] = (\boldsymbol{W}_{L+1}(\mathbf{p}) \otimes \boldsymbol{I})\boldsymbol{H}[\mathbf{d}] \quad \mathbf{p} = (k_0, \ldots, k_L)^T\ , \quad \mathbf{d} = (0, \ldots, L)^T. \tag{52}$$

Because the k_i are distinct $\boldsymbol{W}_{L+1}^{-1}$ must exist. Note that $\{\boldsymbol{W}_{L+1}^{-1}\boldsymbol{W}_{L+1}\}_{iq} = \sum_{l=0}^{L}\{\boldsymbol{W}_{L+1}^{-1}\}_{il}$ $e^{-j\frac{2\pi lk_q}{K}} = \delta_{iq}$ (Kroneker δ). Hence, the ith row of $\boldsymbol{W}_{L+1}^{-1}$ is computable as the coefficients of the Lth order Lagrange polynomials

$$P_{k_i}(z) \triangleq \prod_{j \neq i, 0 \leq j \leq L} \frac{z - e^{-j2\pi k_j/K}}{e^{-j2\pi k_i/K} - e^{-j2\pi k_j/K}} \tag{53}$$

with $(k_i, k_j) \in \mathbf{p}$. Thus, for any h_j we can write:

$$\tilde{\boldsymbol{H}}[h_j] = \sum_{l=0}^{L} \boldsymbol{H}[l]e^{-j2\pi h_j l/K} = \sum_{k_i \in \mathbf{p}} \tilde{\boldsymbol{H}}[k_i]P_{k_i}(e^{-j2\pi h_j/K}), \tag{54}$$

and, for $h_j = k_j$ obviously $P_{k_i}(e^{-j2\pi k_j/K}) = \delta_{k_j k_i}$. From (53) and (54), it follows that $p(\tilde{\boldsymbol{H}}[\bar{\mathbf{p}}] \mid \tilde{\boldsymbol{H}}[\mathbf{p}])$ is product of Dirac deltas. With

$$\boldsymbol{P}[\mathbf{p}, h_j] \triangleq ([P_{k_0}(e^{-j2\pi h_j/K}), \ldots, P_{k_L}(e^{-j2\pi h_j/K})] \otimes \boldsymbol{I}), \tag{55}$$

we have

$$p(\tilde{\boldsymbol{H}}[\bar{\mathbf{p}}] \mid \tilde{\boldsymbol{H}}[\mathbf{p}]) = \prod_{h_j \in \bar{\mathbf{p}}} \delta\left(\tilde{\boldsymbol{H}}[h_j] - \boldsymbol{P}[\mathbf{p}, h_j]\tilde{\boldsymbol{H}}[\mathbf{p}]\right) \tag{56}$$

$$p_{\tilde{H}}(\tilde{\boldsymbol{H}}[\mathbf{p}]) = |\boldsymbol{W}_{L+1}(\mathbf{p})|^{-N_R N_T} p_{\boldsymbol{H}}((\boldsymbol{W}_{L+1}(\mathbf{p}) \otimes \boldsymbol{I})^{-1}\tilde{\boldsymbol{H}}[\mathbf{p}]). \tag{57}$$

Gathering these results we can state the following lemma (valid for any γ):

Lemma 4: Under **a1**, for an FIR N_T input N_R output MIMO frequency selective channel having probability density function of the MIMO impulse response $p_{\mathbf{H}}(\boldsymbol{H}(\mathbf{d}))$, $\mathbf{d} = (0, \ldots, L)$, $\boldsymbol{H}(\mathbf{d}) = (\boldsymbol{H}^H(0), \ldots, \boldsymbol{H}^H(L))^H$, the characteristic function of the mutual information is equal to:

$$\Phi_c(s) = \chi_2 \int \prod_{h=0}^{K-1} \left|\boldsymbol{I} + \gamma \tilde{\boldsymbol{H}}^H[\mathbf{p}]\boldsymbol{P}^H[\mathbf{p}, h]\boldsymbol{P}[\mathbf{p}, h]\tilde{\boldsymbol{H}}[\mathbf{p}]\right|^s p_{\boldsymbol{H}}((\boldsymbol{W}_{L+1}^{-1}(\mathbf{p}) \otimes \boldsymbol{I})\tilde{\boldsymbol{H}}[\mathbf{p}])(d\tilde{\boldsymbol{H}}[\mathbf{p}]) \tag{58}$$

where $\boldsymbol{W}_{L+1}(\mathbf{p})$ is defined in (51), $\tilde{\boldsymbol{H}}[\mathbf{p}]$ is defined in (52), $\boldsymbol{W}_{L+1}^{-1}(\mathbf{p})$ can be expressed in terms of the coefficients of the Lagrange polynomials in (53) and $\chi_2 = |\boldsymbol{W}_{L+1}(\mathbf{p})|^{-N_R N_T}$. Lemma 4, unfortunately, is preserving all the challenge of the calculation of the Capacity characteristic function, since $\left|\boldsymbol{I} + \gamma \tilde{\boldsymbol{H}}^H[\mathbf{p}]\boldsymbol{P}^H[\mathbf{p}, h]\boldsymbol{P}[\mathbf{p}, h]\tilde{\boldsymbol{H}}[\mathbf{p}]\right|$ is the determinant of a linear combination of matrices (c.f. (54)) and thus, it is not explicitly related to the eigenvalues of the blocks $\tilde{\boldsymbol{H}}[k_i]$, $k_i \in \mathbf{p}$.

To reach a simpler expression for $\Phi_c(s)$ and provide also approximate formulas that facilitate the interpretation of the structure of the Capacity p.d.f., we can restrict our attention to the cases where the following assumption is valid, interpolating $\Phi_c(s)$ for the intermediate values of K:

a3. The number of frequency bins is an integer multiple of the channel duration, i.e. $K = Q(L+1)$.

Choosing $\mathbf{p} = (0, Q, \ldots, QL)$, since $e^{-j\frac{2\pi}{Q(L+1)}lQd} = e^{-j\frac{2\pi}{(L+1)}ld}$, $\boldsymbol{W}_{L+1}(\mathbf{p})$ is unitary and

$$P_{nQ}(e^{-j2\pi(lQ+q)/K}) = \frac{1}{L+1}\,\frac{e^{-j2\pi\frac{q}{Q}} - 1}{e^{-j2\pi[\frac{(l-n)}{L+1}+\frac{q}{K}]} - 1} = \frac{e^{j2\pi\left(\frac{qL}{Q(L+1)}-\frac{l-n}{L+1}\right)}}{L+1}\,\frac{\sin\left(\frac{q}{Q}\right)}{\sin\left(\frac{\pi(l-n+q/Q)}{L+1}\right)}. \tag{59}$$

As can be noted from Fig.2 under (**a3.**), with the choice of $\mathbf{p} = (0, Q, \ldots, QL)$ the coefficients of the linear combination in (54) that corresponds to $\tilde{\boldsymbol{H}}[lQ+q]$ are for the most

part highly concentrated around $l-n=0$, which suggest the approximation:

$$\tilde{\boldsymbol{H}}[lQ+q] \approx P_{lQ}(e^{-j2\pi(lQ+q)/K})\tilde{\boldsymbol{H}}[lQ] = \alpha(q)\tilde{\boldsymbol{H}}[lQ]. \tag{60}$$

where

$$\alpha(q) \triangleq \frac{e^{j2\pi\left(\frac{qL}{Q(L+1)}\right)}}{L+1}\frac{\sin\left(\frac{q}{Q}\right)}{\sin\left(\frac{\pi q}{Q(L+1)}\right)} \tag{61}$$

In addition, let us assume that:

a4. In assumption **a2.** $\boldsymbol{R}_H(l_1,l_2) = \boldsymbol{R}_H(l_2-l_1)$.

In general, this condition rarely applies because, for example, the paths are likely not to have the same average power. However, this assumption describes a worse case scenario in terms of the frequency selectivity of the channel and helps simplifying the derivations considerably. In fact, if $\mathbf{p}$ is selected to have uniformly spaced frequency indices, in force of Szëgo theorem for $L \gg 1$ the elements of $\tilde{\boldsymbol{H}}[\mathbf{p}]$ will be approximately uncorrelated not only in space, but also across the frequency bins. Since the correlation matrix of $\boldsymbol{H}(\mathbf{d})$ is $(\boldsymbol{I}\otimes\mathbf{R}_H)$, using the central limit theorem the p.d.f. of $\tilde{\boldsymbol{H}}[\mathbf{p}]$ is approximately[5] $\backsim$ $\mathcal{N}(\mathbf{0}\ ,(\boldsymbol{W}_{L+1}^H(\mathbf{p})\mathbf{R}_H\boldsymbol{W}_{L+1}(\mathbf{p})\otimes\boldsymbol{I}))$ where $(\boldsymbol{W}_{L+1}^H\boldsymbol{R}_H\boldsymbol{W}_{L+1}) \approx diag(\sigma_h^2[\mathbf{p}])$ where $\sigma_h^2[\mathbf{p}] = (\sigma_h^2[0],..,\sigma_h^2[LQ])$ and $\sigma_h^2[lQ] = \sum_n \boldsymbol{R}_H[n]e^{-j2\pi nl/(L+1)}$. This leads to the following:

Proposition 2: Under **a1, a3, a4** for $L \gg 1$ and assuming $E\{\tilde{\boldsymbol{H}}[\mathbf{p}]\} = \mathbf{0}$ the $\tilde{\boldsymbol{H}}^H[lQ]$ are approximately Gaussian and independent and

$$\Phi_c(s) \approx \chi_3 \prod_{l=0}^{L}\int\prod_{q=0}^{Q-1}\left|\boldsymbol{I}+\gamma\alpha(q)\tilde{\boldsymbol{H}}^H[lQ]\tilde{\boldsymbol{H}}[lQ]\right|^s e^{-\frac{tr(\tilde{\boldsymbol{H}}^H[lQ]\tilde{\boldsymbol{H}}[lQ])}{\sigma_h^2[l]}}(d\tilde{\boldsymbol{H}}[lQ]), \tag{62}$$

where $\chi_3 = \prod_{l=0}^{L}(\pi\sigma_h^2[l])^{N_TN_R} = \pi^{N_TN_R(L+1)}|\boldsymbol{R}_H|^{N_TN_R}$. The closed form expression of the integral on the right side of (62) is analogous to (44) [18], [19].

Because (44) is not intelligible anyways, we prefer to proceed in our approximation and exploit the fact that $\gamma \gg 1$. In this case is also easier to consider spatial correlation when it is reasonable to assume that the correlation does not change with the path and the following separable model applies:

$$E\{\{\boldsymbol{H}[l]\}_{r_1,t_1}^*\{\boldsymbol{H}[l]\}_{r_2,t_2}\} \ = \ \{\boldsymbol{R}_r\}_{r_1,r_2}\{\boldsymbol{R}_t\}_{t_1,t_2}. \tag{63}$$

[5]Under (**a2**) this is obviously exactly true.

Overall this means that the covariance matrix of $\boldsymbol{H}[\mathbf{d}]$ is $\boldsymbol{R}_r \otimes \boldsymbol{R}_t \otimes \boldsymbol{R}_H$. Since the DFT operates in time, the spatial correlation of the MIMO frequency response $\tilde{\boldsymbol{H}}[\mathbf{p}]$ for $L \gg 1$ tend to be $\boldsymbol{R}_r \otimes \boldsymbol{R}_t \otimes diag(\sigma_h^2[\mathbf{p}])$, or in other words

$$E[vec(\tilde{\boldsymbol{H}}[lQ])vec(\tilde{\boldsymbol{H}}[lQ])^H] = \sigma^2[lQ](\boldsymbol{R}_t \otimes \boldsymbol{R}_r). \tag{64}$$

With this in mind we can state the following

Proposition 3: Under the same assumptions and using the same approximations that lead to Proposition 2, if $\gamma \gg 1$:

$$\Phi_C(s) \approx |\boldsymbol{R}_t|^{sK}|\boldsymbol{R}_r|^{sK}\left(\gamma|\boldsymbol{R}_H|^{\frac{1}{L+1}}\right)^{sKn}\left(\prod_{q=0}^{Q-1}\alpha(q)\right)^{s(L+1)n}\left(\prod_{i=1}^{n}\frac{\Gamma(Qs+m-n+i)}{\Gamma(m-n+i)}\right)^{L+1} \tag{65}$$

Proof: When $\gamma \gg 1$ the contribution of the identity matrix to the determinant in (62) can be neglected leading to:

$$\begin{aligned}
\Phi_C(s) &\approx \prod_{l=0}^{L} E\left\{\prod_{q=0}^{Q-1}|\gamma\alpha(q)\tilde{\boldsymbol{H}}^H[lQ]\tilde{\boldsymbol{H}}[lQ]|^s\right\} \\
&= \prod_{l=0}^{L} E\left\{\prod_{q=0}^{Q-1}|\gamma\alpha(q)\boldsymbol{R}_t^{1/2}\boldsymbol{R}_t^{-1/2}\tilde{\boldsymbol{H}}^H[lQ]\boldsymbol{R}_r^{-1/2}\boldsymbol{R}_r\boldsymbol{R}_r^{-1/2}\tilde{\boldsymbol{H}}[lQ]\boldsymbol{R}_t^{-1/2}\boldsymbol{R}_t^{1/2}|^s\right\} \\
&= \gamma^{Ksn}|\boldsymbol{R}_t|^{Ks}|\boldsymbol{R}_r|^{Ks}\left(\prod_{l=0}^{L}\sigma^2[lQ]\right)^{Qsn}\left(\prod_{q=0}^{Q-1}\alpha(q)\right)^{sn(L+1)} \\
&\cdot \quad E\left\{|\boldsymbol{R}_r^{-1/2}\tilde{\boldsymbol{H}}[lQ]\boldsymbol{R}_t^{-1/2}/\sigma[lQ]|^{2Qs}\right\} \\
&= \gamma^{Ksn}|\boldsymbol{R}_t|^{Ks}|\boldsymbol{R}_r|^{Ks}|\boldsymbol{R}_H|^{Qsn}\left(\prod_{q=0}^{Q-1}\alpha^{sn(L+1)}(q)\right)E\left\{|\boldsymbol{W}|^{2Qs}\right\}
\end{aligned} \tag{66}$$

where $E\left\{|\boldsymbol{W}|^{2Qs}\right\}$ is the expectation of the random determinant of a complex Gaussian zero mean matrix with independent entries having unit variance. The expression of the expectation is:

$$\begin{aligned}
E\left\{|\boldsymbol{W}|^{2Qs}\right\} &= E\left\{\prod_{i=1}^{n}\lambda_i^{Qs}\right\} = \int_{\lambda\geq 0}\left(\prod_{i=1}^{n}\lambda_i^{Qs}\right)p_\Lambda(\lambda_1\cdots\lambda_n)(d\boldsymbol{\Lambda}) \\
&= \int_{\lambda\geq 0}\left(\prod_{i=1}^{n}\lambda_i^{Qs}\right)\chi\prod_{1\leq i<k\leq n}(\lambda_k-\lambda_i)^2e^{-\sum_i\lambda_i}\left(\prod_{i=1}^{n}\lambda_i\right)^{m-n}(d\boldsymbol{\Lambda})
\end{aligned}$$

where χ is given by (c.f. (17), (18) for $Vol(\boldsymbol{Q}_{m,n})\ \overline{Vol(\boldsymbol{U}_{n,n})}$)

$$\chi \ = \ 2^{-n}\pi^{-mn}\ Vol(\boldsymbol{Q}_{m,n})\ \overline{Vol(\boldsymbol{U}_{n,n})} = \frac{1}{\prod_{i=1}^{n}\Gamma(i)\Gamma(m-n+i)}. \tag{67}$$

Because:

$$\prod_{i=1}^{n}\Gamma(i)\Gamma(\alpha+i) = \int_{\lambda_i\geq 0}\left(\prod_{i=1}^{n}\lambda_i\right)^{\alpha} e^{-\sum_i\lambda_i}\prod_{1\leq i<k\leq n}^{n}(\lambda_k-\lambda_i)^2(d\boldsymbol{\Lambda}), \tag{68}$$

from (68) we have

$$E\left\{\prod_{i=1}^{n}\lambda_i^{Qs}\right\} = \chi\prod_{i=1}^{n}\Gamma(i)\Gamma(Qs+m-n+i), \tag{69}$$

and this leads to (65). ■

A. The Gaussian approximation

As pointed out earlier in this section, the Capacity for the case of $\gamma \ll 1$ takes the following form

$$C \approx \gamma K tr(\boldsymbol{H}^H[\mathbf{d}]\boldsymbol{H}[\mathbf{d}]) \tag{70}$$

This is a $\chi - square$ distribution which, under the limit of large number of antennas and paths, will closely approximate the Gaussian density. However, even for the reasonable values of N_R and N_T, we show that the Gaussian fit is accurate even when γ is really small. To do so, we first find parameters of the corresponding Gaussian random variable using (49). First and second order derivatives of $\Phi_C(s)$ evaluated at $s=0$ yield the mean and variance [6] :

$$\mu = \gamma K tr(\boldsymbol{R}_r)tr(\boldsymbol{R}_t)tr(\boldsymbol{R}_H), \qquad \sigma^2 = (\gamma K)^2 tr(\boldsymbol{R}_r^2)tr(\boldsymbol{R}_t^2)tr(\boldsymbol{R}_H^2) \tag{71}$$

Hence, we have, in this case, the following

$$p_C(C) \approx \frac{e^{-\frac{(C-\gamma K tr(\overline{\boldsymbol{R}}))^2}{2(\gamma K)^2 tr(\overline{\boldsymbol{R}}^2)}}}{\sqrt{2\pi(\gamma K)^2 tr(\overline{\boldsymbol{R}}^2)}}. \tag{72}$$

It is interesting to note that mean and variance are proportional to the signal to noise ratio (γ). Also, while there is no explicit dependence on m and n, the dependence on the

[6] We use the identity $\frac{\partial}{\partial\alpha}\{\ln|\boldsymbol{A}|\} = tr\left\{\boldsymbol{A}^{-1}\frac{\partial\boldsymbol{A}}{\partial\alpha}\right\}$

correlation matrices of paths and the array elements is through their trace. The Gaussian approximation is compared with the Capacity histogram in Fig.6 (See section VIII for details). It is interesting to observe that the mean normalized by the standard deviation is 1. As we shall see later, the dependence of γ, $\boldsymbol{R}_H$, $\boldsymbol{R}_t$, $\boldsymbol{R}_r$, m and n on the mean and variance is entirely different when we consider γ to be high.

In fact, for high γ, (65) shows that the channel gain takes the form of the geometric mean of the eigenvalues of the channel covariance matrix. The form of the characteristic function in (65) motivates the idea of approximating the p.d.f. of Capacity with a Gaussian p.d.f. whose parameters can be easily computed. From (65),

$$\begin{aligned}\Phi_C(s) &= \prod_{q=0}^{Q-1}(\alpha(q)\gamma|\boldsymbol{R}_H|^{\frac{1}{L+1}}|\boldsymbol{R}_t|^{1/n}|\boldsymbol{R}_r|^{1/n})^{sn(L+1)}\left(\prod_{i=1}^{n}\frac{\Gamma(Qs+m-n+i)}{\Gamma(m-n+i)}\right)^{L+1} \\ &= e^{sn(L+1)\log\left(\prod_{q=0}^{Q-1}\left(\alpha(q)\gamma|\boldsymbol{R}_H|^{\frac{1}{L+1}}|\boldsymbol{R}_t|^{1/n}|\boldsymbol{R}_r|^{1/n}\right)\right)}\left(\prod_{i=1}^{n}\frac{\Gamma(Qs+m-n+i)}{\Gamma(m-n+i)}\right)^{L+1} \quad (73)\end{aligned}$$

The first factor in (73) implies that the Capacity p.d.f. $p_C(C)$ is a shifted version of the inverse Laplace transform of the term

$$\left(\prod_{i=1}^{n}\frac{\Gamma(Qs+m-n+i)}{\Gamma(m-n+i)}\right)^{L+1}. \quad (74)$$

The factor (74) is the Lth power of a product of functions. Since in our approximations $L \gg 1$, we can infer that the inverse Laplace transform of (74) will be very close to a Gaussian p.d.f., using the same arguments used in the proof of the central limit theorem. Even when L is moderately large, the product inside indicates that several convolutions take place in the inverse domain, so that the same conclusion holds approximately true.

From the first and second order derivatives of $\frac{\Gamma(Qs+m-n+i)}{\Gamma(m-n+i)}$ in $s=0$, one can easily obtain the first order moment $\mu_i^{(1)}$ and the variance σ_i^2 of its inverse Laplace transform, which are:

$$\mu_i = Q\psi^{(0)}(m-n+i)\ , \qquad \sigma_i^2 = Q^2\psi^{(2)}(m-n+i), \quad (75)$$

where $\psi^n(x)$ is the nth derivative of the Polygamma function also known as Digamma function [20]. Therefore, shifting by $\left(\prod_{q=0}^{Q-1}\alpha(q)\gamma|\boldsymbol{R}_H|^{\frac{1}{L+1}}|\boldsymbol{R}_t|^{\frac{1}{n}}|\boldsymbol{R}_r|^{\frac{1}{n}}\right)$ the inverse Laplace

transform of $\left[\left(\frac{\Gamma(Qs+m-n+i)}{\Gamma(m-n+i)}\right)^{L+1}\right]$, we have that approximately:

$$p_C(C) \approx \frac{e^{-\frac{\left(C-\left(\prod_{q=0}^{Q-1}\alpha(q)\gamma|\boldsymbol{R}_H|^{\frac{1}{L+1}}|\boldsymbol{R}_t|^{1/n}|\boldsymbol{R}_r|^{1/n}\right)-K\sum_{i=1}^{n}\psi^{(0)}(m-n+i)\right)^2}{2KQ\sum_{i=1}^{n}\psi^{(2)}(m-n+i)}}}{\sqrt{2\pi KQ\sum_{i=1}^{n}\psi^{(2)}(m-n+i)}}. \quad (76)$$

It should be noted that the variance of the Capacity (75) obtained in the high SNR regime is independent of γ and the same conclusion was reached in [21]. This is in direct contrast to the low γ scenario. Here we also notice that, even if under many approximations, the effect of correlation and SNR is only to shift the mean of the distribution but the same parameters have no impact on the Capacity variance, which is only a function of m, n. The peculiar dependence on the correlation of paths and array elements through the geometric mean of the eigenvalues is also quite interesting and provides further confirmation of the fact that antenna elements and multi-path have similar effects on the Capacity in broadband channels.

In the section (VIII), we present the plots of characteristic function given by (65) for a frequency selective channel for different values of Q . We also provide the corresponding plots, using Monte Carlo simulation, of (39) and show that both the expressions for characteristic function are same when number of frequency bins equals the channel duration, i.e. when $Q = 1$.

VII. Outage Capacity and Chernoff Bound

We consider the case of $\gamma \gg 1$ only. The case of low γ requires calculating the cumulative distribution function of a $\chi - square$ density and the comparison with its Gaussian approximation is available in literature [22] and will not be considered here for brevity. With the characteristic function and the Gaussian approximation (76) of the p.d.f for high γ case at hand, it is possible to derive curves that come close to the Outage Capacity. Using the Chernoff bound:

$$Pr(C > V) \leq min_{s\geq 0} e^{-sV+\log E\{e^{sC}\}} \quad (77)$$

To find the tightest bound we equate to zero the derivative of $-sV + \log E\{e^{sC}\}$ with respect to s and obtain the following equivalent equation:

$$\Rightarrow V = n(L+1)\log\left(\prod_{q=0}^{Q-1}\alpha(q)\gamma|\boldsymbol{R}_H|^{\frac{1}{L+1}}|\boldsymbol{R}_t|^{1/n}|\boldsymbol{R}_r|^{1/n}\right) + (L+1)\sum_{i=1}^{n}\frac{d}{ds}\left\{\log(\Gamma(Qs+m-n+i))\right\} \quad (78)$$

Let $\rho \triangleq V - n(L+1)\log\left(\prod_{q=0}^{Q-1}\alpha(q)\gamma|\boldsymbol{R}_H|^{\frac{1}{L+1}}|\boldsymbol{R}_t|^{1/n}|\boldsymbol{R}_r|^{1/n}\right)$, then from (78)

$$\rho = (L+1)\sum_{i=1}^{n}\frac{d}{ds}\left\{\log(\Gamma(Qs+m-n+i))\right\} = K\sum_{i=1}^{n}\psi^{(0)}(Qs+m-n+i). \quad (79)$$

The expression on the right hand side of (79), say $f(s)$, is a monotonic function of Qs (see Fig.3)

The intersection of this monotonic function $f(s)$ with horizontal family of curves for different values of V gives the optimum value of s that minimizes (77) [c.f. Section VIII].

With the Gaussian approximation derived in Section VI-A, the Outage Capacity is simply

$$Pr(C > V) = Q\left(\frac{V - \left(\prod_{q=0}^{Q-1}\alpha(q)\gamma|\boldsymbol{R}_H|^{\frac{1}{L+1}}|\boldsymbol{R}_t|^{1/n}|\boldsymbol{R}_r|^{1/n}\right) - K\sum_{i=1}^{n}\psi^{(0)}(m-n+i)}{\sqrt{KQ\sum_{i=1}^{n}\psi^{(2)}(m-n+i)}}\right) \quad (80)$$

where $Q(x)$ is the Q-function and $\psi^{(n)}(x)$ is, as mentioned before, the nth derivative of the Polygamma function also known as Digamma function.

VIII. Numerical Examples

In this section we conclude our work by presenting some numerical examples supporting the theoretical results obtained in the paper. We first present plots of the characteristic function for different channel parameters. The characteristic function given by (65) is plotted and compared with the the characteristic function obtained by Monte Carlo simulations (39). In all examples, the number of frequency bins is $K = 8$ and the signal to noise ratio (γ) of the channel is assumed to be $20dB$. In (Fig.4), the characteristic functions from (65) and (39) for various values of L are plotted together. We obtained these

curves for $N_R = 3,\ N_T = 4$ for the case where the paths and array elements correlations are homogeneous and decay exponentially, i.e. $\boldsymbol{R}_H(l_2, l_1) = \rho^{|l_2 - l_1|}$, $\boldsymbol{R}_t(t_2, t_1) = \rho^{|t_2 - t_1|}$ and $\boldsymbol{R}_r(r_2, r_1) = \rho^{|r_2 - r_1|}$ respectively, with $\rho = 0.5$. We can see that as L increases, the approximate expression (the solid curve) closes towards the exact one, and the approximation becomes exact when the number of frequency bins equals the channel duration, i.e. when $Q = 1$. This is shown by the pair of curves for $K = 8,\ L = 7$.

Fig.5 shows the effect of the signal to noise ratio (γ) on the accuracy of the approximate expression in (65). In this example $K = 8,\ L = 7,\ N_R = 3,\ N_T = 4$ and the correlations in time and space are identical to the ones used in the previous example. Since we arrived at (65) under high SNR assumption, we can expect intuitively the approximation to be better at high γ. This is evident from the (Fig.5), where we observe that the plots merge as γ is increased from $5dB$ to $20dB$.

The probability density function of the Capacity is plotted (see Fig.6) using Monte Carlo simulations and compared with the one obtained by Gaussian density (72) when γ is very small. The Gaussian approximation is really good and supports the claim made in the section VI-A. Here, we have used $K = 4,\ L = 3,\ \rho = 0.1,\ N_R = 6,\ N_T = 6$ and γ = -35 dB. The normalized histogram of Capacity for high γ is also shown in (Fig.7) and it is in a very good agreement with (73), which appears to be accurate even for low values of the Outage Capacity. $K = 8,\ L = 7,\ \rho = 0.3,\ N_R = 3,\ N_T = 3$ and $\gamma = 35\ dB$ are used. Hence, the Gaussian approximation works fine for low as well as high signal to noise ratio even for moderate values of N_R and N_T, as observed by other authors [21], [23].

In Fig.8 we show the Chernoff bound for the Capacity, the complementary cumulative distribution function (ccdf) of Capacity using (80) and compare them with the corresponding values obtained by Monte Carlo simulations. The parameters used are $K = 8,\ L = 7,\ \rho = 0.3,\ N_R = 2,\ N_T = 2$ and $\gamma = 20\ dB$. The effect of correlated fading is illustrated in (Fig.9) using the same exponential model for path and array correlation described above and changing the value of ρ. We consider $\rho = 0$ (uncorrelated case), $\rho = 0.2$ and $\rho = 0.4$. and, as before, the parameters are $K = 8,\ L = 7,\ N_R = 3,\ N_T = 3$ and $\gamma = 20dB$. It is interesting to observe that the outage Capacity increases as ρ increases. In fact, the correlation between the antenna elements reduces the diversity and

hence increases the outage.

Appendix

A. The computation of $(d\boldsymbol{B})$ where $\boldsymbol{B} = \boldsymbol{Q}\boldsymbol{R}$ and the volume element of a unitary group

Since we could not find the derivation of (16) anywhere, we thought of adding this appendix. The Householder reflections can be used to decompose any matrix, $\boldsymbol{A} \in \mathbb{C}^{\mathrm{m,n}}$ $(m \geq n)$ as a product of a unitary and an upper triangular matrix. The complex Householder reflection is given by

$$\boldsymbol{H} = \boldsymbol{I} - \frac{2\mathbf{v}\mathbf{v}^H}{\mathbf{v}^H\mathbf{v}}, \tag{81}$$

and it is easy to verify that $\boldsymbol{H}$ is Hermitian and unitary. Similar to the what is done for the real case in [13], we can write:

$$\boldsymbol{H}_n\boldsymbol{H}_{n-1}\cdots\boldsymbol{H}_1\boldsymbol{A} = \begin{pmatrix} \boldsymbol{R}_{n\times n} \\ \mathbf{0} \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{A} = \boldsymbol{H}_1\boldsymbol{H}_2\cdots\boldsymbol{H}_n \begin{pmatrix} \boldsymbol{R}_{n\times n} \\ \mathbf{0} \end{pmatrix}, \tag{82}$$

where $\bar{\boldsymbol{Q}} = \boldsymbol{H}_1\boldsymbol{H}_2\cdots\boldsymbol{H}_n$ is the product of the Householder reflections. From (82), we can write $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$, where $\boldsymbol{Q}$ includes only the first n columns of $\boldsymbol{H}_1\boldsymbol{H}_2\cdots\boldsymbol{H}_n$ (i.e. $\boldsymbol{Q} = [\boldsymbol{q}_1\cdots\boldsymbol{q}_n]$).

$\boldsymbol{A}$ has $2mn$ real parameters and, therefore, the total number of independent parameters of $\boldsymbol{R}$ and $\boldsymbol{Q}$ needs to be $2mn$ for $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$ to be a unique transformation. $\boldsymbol{Q}$ satisfies $\boldsymbol{Q}^H\boldsymbol{Q} = I$, which corresponds to $n + \frac{2n(n-1)}{2} = n^2$ equations. Hence, the Stiefel manifold has $2mn - n^2$ real independent parameters. $\boldsymbol{R}$ has $n(n+1)/2$ complex parameters and twice as many real, which is n too many. Hence in order to have unique factorization in (82), we can constrain the diagonal elements of $\boldsymbol{R}$ to be real so that $\boldsymbol{R}$ has n^2 independent parameters only. Now, we can proceed to calculate $(d\boldsymbol{A}) = (\bar{\boldsymbol{Q}}^H d\boldsymbol{A})$.

$$\begin{aligned} (\bar{\boldsymbol{Q}}^H d\boldsymbol{A}) &= (\bar{\boldsymbol{Q}}^H\boldsymbol{Q}d\boldsymbol{R} + \bar{\boldsymbol{Q}}^H d\boldsymbol{Q}\boldsymbol{R}) \\ &= (\boldsymbol{I}_{m,n}d\boldsymbol{R} + \bar{\boldsymbol{Q}}^H d\boldsymbol{Q}\boldsymbol{R}). \end{aligned} \tag{83}$$

The first term is an $m \times n$ upper triangular matrix while $\bar{\boldsymbol{Q}}^H d\boldsymbol{Q}$ is antisymmetric with purely imaginary diagonal elements[7]. In fact, $\boldsymbol{q}_i^H\boldsymbol{q}_i = 1$ implies $\boldsymbol{q}_i^H d\boldsymbol{q}_i = -d\boldsymbol{q}_i^H\boldsymbol{q}_i = -(\boldsymbol{q}_i^H d\boldsymbol{q}_i)^*$ and then $\Re(\boldsymbol{q}_i^H d\boldsymbol{q}_i) = 0$. Also, $\boldsymbol{q}_i^H d\boldsymbol{q}_j = -(\boldsymbol{q}_j^H d\boldsymbol{q}_i)^*$.

[7]The presence of these imaginary diagonal entries makes the case of unitary and orthogonal case different.

In matrix form,

$$\bar{\boldsymbol{Q}}^H d\boldsymbol{A} = \begin{pmatrix} dr_{11} & dr_{12} & \cdots & dr_{1n} \\ & dr_{22} & \cdots & dr_{2n} \\ & & \ddots & \vdots \\ & & & dr_{nn} \\ & & & \\ & & & \end{pmatrix} + \begin{pmatrix} \boldsymbol{q}_1^H d\boldsymbol{q}_1 & -(\boldsymbol{q}_2^H d\boldsymbol{q}_1)^* & \cdots & -(\boldsymbol{q}_n^H d\boldsymbol{q}_1)^* \\ \boldsymbol{q}_2^H d\boldsymbol{q}_1 & \boldsymbol{q}_2^H d\boldsymbol{q}_2 & \cdots & -(\boldsymbol{q}_n^H d\boldsymbol{q}_2)^* \\ & & \ddots & \\ \boldsymbol{q}_n^H d\boldsymbol{q}_1 & \boldsymbol{q}_n^H d\boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n^H d\boldsymbol{q}_n \\ \cdots & \cdots & \cdots & \cdots \\ \boldsymbol{q}_m^H d\boldsymbol{q}_1 & \boldsymbol{q}_m^H d\boldsymbol{q}_1 & \cdots & \boldsymbol{q}_m^H d\boldsymbol{q}_n \end{pmatrix} \boldsymbol{R}, \quad (84)$$

where $(d\boldsymbol{R})$ is $\bigwedge_{i\leq j} dr_{ij}$. Since $\bar{\boldsymbol{Q}}^H d\boldsymbol{Q}$ is antisymmetric, we can take the exterior product of all entries below the diagonal. The next step in calculating $(\bar{\boldsymbol{Q}}^H d\boldsymbol{A})$ is to compute the exterior product of the elements in the above matrix equation. Since the entries below the main diagonal in the first part of (84) are zero, only second part should be considered for the position below the main diagonal. We consider one column at a time starting from the left one. Because every entry in the first column of $\bar{\boldsymbol{Q}}^H d\boldsymbol{Q}$ get to be multiplied by r_{11} and because of the fact that $(\alpha dz) = |\alpha|^2 (dz)$, for all $\alpha \in \mathbb{C}$, the first column below the main diagonal becomes $(r_{11}^2)^{m-1} \bigwedge_{j=2}^m \boldsymbol{q}_j^H d\boldsymbol{q}_1$. The second column is multiplied by r_{22} and then r_{12} times the first column is added to it. However, the addition of the first column in this case does not contribute to the wedge product because wedge product of a vector with itself is zero and the first column was already in the first term of the wedge product. This implies that combining the independent entries of the first and the second column gives $(r_{11}^2)^{m-1}(r_{22}^2)^{m-2} \bigwedge_{j=2}^m \boldsymbol{q}_j^H d\boldsymbol{q}_1 \bigwedge_{j=3}^m \boldsymbol{q}_j^H d\boldsymbol{q}_2$. Continuing in a similar fashion for all the columns we have that only r_{ii} keep appearing in the product (note that the r_{ii}'s are all real), thus obtaining

$$\prod_{i=1}^{n} (r_{ii}^2)^{m-i} \bigwedge_{i=1}^{n} \bigwedge_{j=i+1}^{m} \boldsymbol{q}_j^H d\boldsymbol{q}_i. \quad (85)$$

Next, we incorporate the diagonal elements of both the parts of (84) in the wedge product and then finally consider the contribution of entries above the main diagonal to complete the process of calculating $(d\boldsymbol{Q})$. The contributing diagonal entries in $\bar{\boldsymbol{Q}}^H d\boldsymbol{A}$ are of the form $(dr_{ii} + r_{ii}\boldsymbol{q}_i^H d\boldsymbol{q}_i)$ and we can note that the first term is real while the second one is imaginary, so that the contributions from the diagonal entries to the wedge product

appear as

$$\prod_{i=1}^{n}(r_{ii})(dr_{11}\wedge dr_{22}\cdots\wedge dr_{nn})(\boldsymbol{q}_1^H d\boldsymbol{q}_1\wedge\cdots\wedge\boldsymbol{q}_n^H d\boldsymbol{q}_n). \tag{86}$$

For the entries above the main diagonal, only the first part in (84) contributes to the wedge product (because of the fact that $dz\wedge dz^*=0$) taking the form

$$\bigwedge_{i<j} dr_{ij}. \tag{87}$$

Equations (85), (86) and (87) can now be combined to produce $(\bar{\boldsymbol{Q}}^H d\boldsymbol{A})=(d\boldsymbol{A})$ as

$$\begin{aligned}(\bar{\boldsymbol{Q}}^H d\boldsymbol{A}) &= \prod_{i=1}^{n}(r_{ii})^{2(m-i)+1}\bigwedge_{i=1}^{n}\bigwedge_{j=i}^{m}\boldsymbol{q}_j^H d\boldsymbol{q}_i\bigwedge_{i\le j}dr_{ij}\\ &= \prod_{i=1}^{n}(r_{ii})^{2(m-i)+1}\bigwedge_{i=1}^{n}\bigwedge_{j=i}^{m}\boldsymbol{q}_j^H d\boldsymbol{q}_i(d\boldsymbol{R}).\end{aligned} \tag{88}$$

Writing $\bigwedge_{i=1}^{n}\bigwedge_{j=i}^{m}\boldsymbol{q}_j^H d\boldsymbol{q}_i$ in the above equation as $(d\boldsymbol{Q})$, $(d\boldsymbol{A})$ can finally be written as

$$(d\boldsymbol{A}) = \prod_{i=1}^{n}(r_{ii})^{2(m-i)+1}(d\boldsymbol{Q})(d\boldsymbol{R}) \tag{89}$$

The element of volume of Stiefel manifold is thus simply

$$(d\boldsymbol{Q})=\bigwedge_{i=1}^{n}\bigwedge_{j=i}^{m}\boldsymbol{q}_j^H d\boldsymbol{q}_i, \tag{90}$$

where, noticeably the elements for $j=i$ have to be included, contrary to the case of orthogonal matrices $\boldsymbol{Q}\in\mathbb{R}^{\mathrm{m,n}}$, $\boldsymbol{Q}^T\boldsymbol{Q}=\boldsymbol{I}$ for which:

$$(d\boldsymbol{Q})=\bigwedge_{i=1}^{n}\bigwedge_{j=i+1}^{m}\boldsymbol{q}_j^T d\boldsymbol{q}_i. \tag{91}$$

As an example, we give the volume element of a 2×2 unitary group for the following parametrization [24]

$$\boldsymbol{Q}(\xi_1,\xi_2,\xi_3,\xi_4)=\begin{pmatrix}\cos\xi_1 e^{-\jmath\,\xi_2} & \sin\xi_1 e^{-\jmath\,\xi_3}\\ -\sin\xi_1 e^{\jmath\,\xi_3} & \cos\xi_1 e^{\jmath\,\xi_2}\end{pmatrix}e^{\jmath\,\xi_4} \tag{92}$$

The element of volume for such a group can be calculated(using (90))to be

$$(d\boldsymbol{Q})=2|\sin\xi_1\cos\xi_1|\;d(\xi_1,\xi_2,\xi_3,\xi_4),\qquad 0\le\xi_1,\xi_2,\xi_3,\xi_4<2\pi \tag{93}$$

References

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels, Tech. Rep. Bell Laboratories, 1995.

[2] G.J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multielement antennas," *Bell Labs. Tech. Journ.*, Vol. 1, No. 2, pp. 41–59, 1996.

[3] T.L. Marzetta, B. H. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading", *IEEE Trans. on Info. Theory*, Vol. 45, No. 1, pp.139-157, Jan. 1999.

[4] Alex Grant, "Rayleigh Fading Multi-Antenna Channels," *EURASIP Journal on Applied Signal processing*, Vol.2002,Issue 3 pp. 316–329.

[5] Lizhong Zheng, David N.C. Tse, "Communication on the Grassmann Manifold: A Geometric Approach to the Noncoherent Multiple-Antenna Channel", *IEEE Trans. on Info. Theory*, Vol. 48, No. 2, pp.359-383, Feb. 2002.

[6] P. B. Rapajic and D. Popescu, "Information Capacity of a Random Signature Multiple-Input Multiple-Output Channel," *IEEE Trans. on Commun.*, Vol. 48, No. 8, pp. 1245–1248, August 2000.

[7] Y.Li, J. Chuang, N. R. Sollenberger, "Transmitter diversity for ofdm systems and its impacts on high-rate wireless networks," *Proc. of IEEE Int. Conf. on Comm.*, Vol. 1, pp. 534–538, Vancouver, Canada, June 1999.

[8] M.K Simon, M.S. Alouini "Digital Communication over Fading Channels, A Unified approach to Performance Analysis", *Wiley Interscience*, 2000.

[9] E. Wigner, "On the distribution of the roots of certain symmetric matrices," *Annales of Math.* Vol. 67, pp.325–327, 1958.

[10] A. Edelman, "Random Eigenvalues," Notes for Math 273 UC Berkeley, Feb. 27, 1999 *http://www.math.berkeley.edu/ edelman/math273.html*

[11] V. Girko, "Theory of Random Determinants" *Kluwer Publishers*(Netherlands) 1990.

[12] H. Flandes, "Diffrential Forms with Applications to teh Physical Sciences", *Dover Publications Inc.*, New York, II Edition, 1989.

[13] G. H. Golub, C. F. Van Loan, "Matrix Computations", *The Johns Hopkins University Press*, III Edition,1996.

[14] A.T. James, "Distributions of matrix variates and laltent roots derived from normal samples," *Ann. Math. Statistics*, Vol. 35, pp.475-501, 1964.

[15] Haagerup and S. Thorbjrnsen, "Random matrices with complex Gaussian entries," preprint (49 pp.), Odense 1998. *http://www.imada.ou.dk/ haagerup/2000-.html.*

[16] B.V. Bronk, "Exponential ensembles for Random matrices", *Jour. of Math. Physics*, Vol. 6, pp. 228-237, 1965.

[17] A. Lozano, A. Tulino and S. Verdù, "Multi-Antenna Capacity in the Low-power Regime", *DIMACS Workshop on Signal Processing for Wireless Transmission, Rutgers*,October 2002.

[18] M. Kang and M. -S. Alouini, "Quadratic forms in complex Gaussian matrices and performance analysis of MIMO systems with co-channel interference",*Proceedings of IEEE International on Information Theory (ISIT'2002), p. , Lausanne, Switzerland*, June 2002.

[19] Z. Wang and G. B. Giannakis, "Outage mutual information of space-time MIMO channels", *In Proc. of 40th Allerton Conf*,October 2002.

[20] E. Artin, "The Gamma Function," *New York, Holt, Rinehart and Winston*, 1964.

[21] T.L. Marzetta, B. M. Hochwald and V. Tarokh"Multi-Antenna Channel-Hardening and its Implications for Rate Feedback and Scheduling" Submitted to IEEE Transactions on Information Theory, May 2002. preprint*http://mars.bell-labs.com/cm/ms/what/mars/papers/ratefeedback/*

[22] Steven M. Kay, "Fundamentals of Statistical Signal Processing- Detection Theory", *Prentice Hall*, 1998.

[23] P.J. Smith and M. Shafi, "On a Gaussian Approximation to the Capacity of Wireless MIMO Systems", *IEEE International Conference on Communications(ICC),New York city*,April-May 2002.

[24] F. D. Murnaghan, "The Theory of Group Representations", *The John Hopkins Press*, Baltimore, 1938.

List of Figures

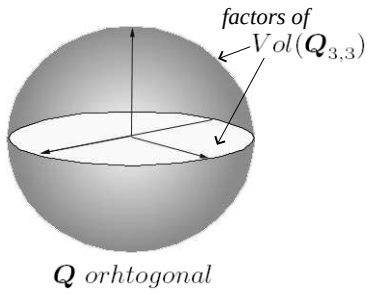

Fig. 1. The factors in the $Vol(\mathbf{Q}_{3\times 3})$ for $\boldsymbol{Q}$ orthogonal.

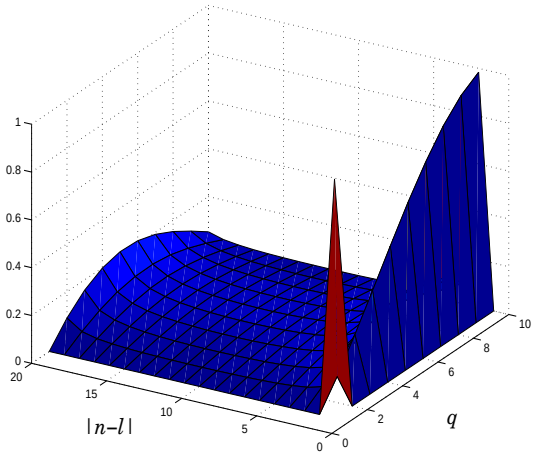

Fig. 2. The coefficients $P_{nQ}(e^{-j2\pi(lQ+q)/K})$ in (59).

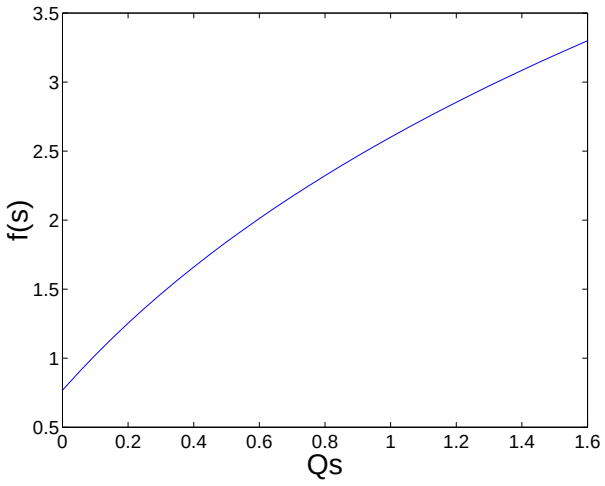

Fig. 3. The function $f(s)$ on the right hand side of (79)

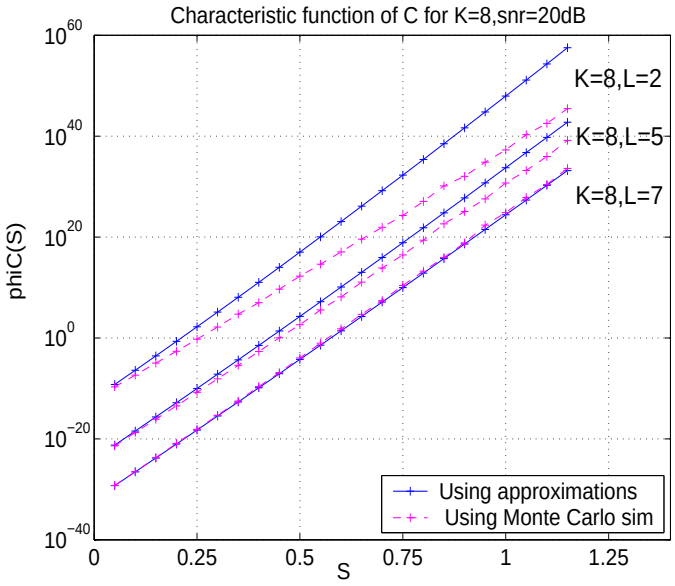

Fig. 4. Characteristic functions for $K = 8,\ \gamma = 20dB,\ \rho = 0.5,\ N_R = 3,\ N_T = 4$ and varying L.

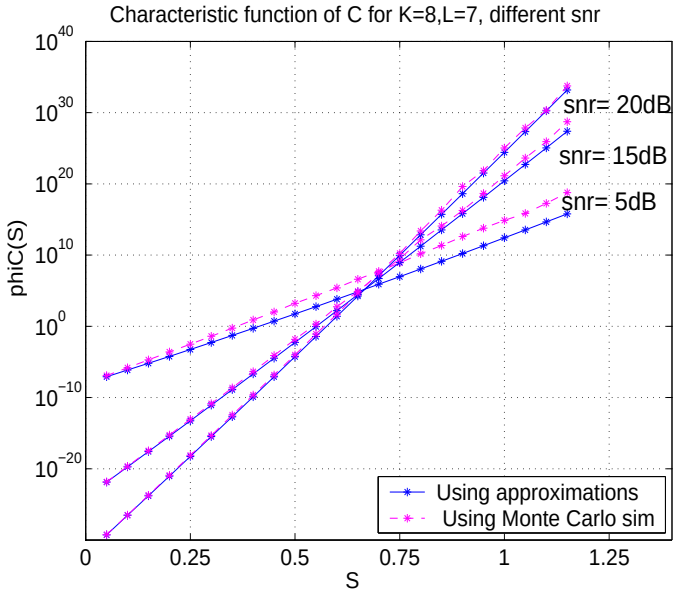

Fig. 5. Characteristic functions for $K = 8,\ L = 7,\ \rho = 0.5,\ N_R = 3,\ N_T = 4$ and varying γ.

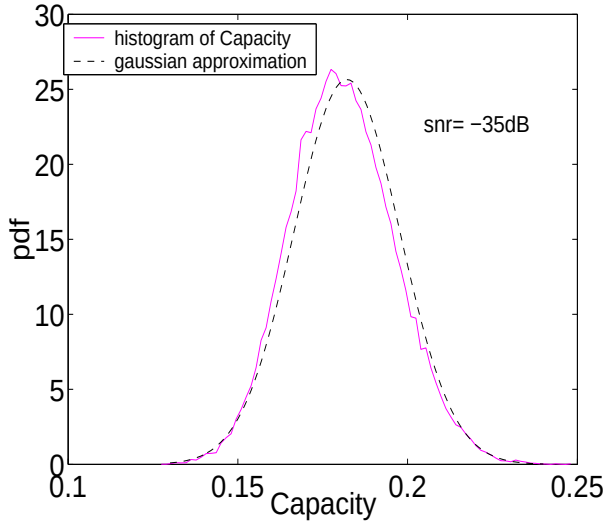

Fig. 6. The density of Capacity obtained by Monte Carlo simulations compared with the Gaussian approximation for $K = 4,\ L = 3,\ \rho = 0.1,\ N_R = 6,\ N_T = 6$ and $\gamma = -35dB$.

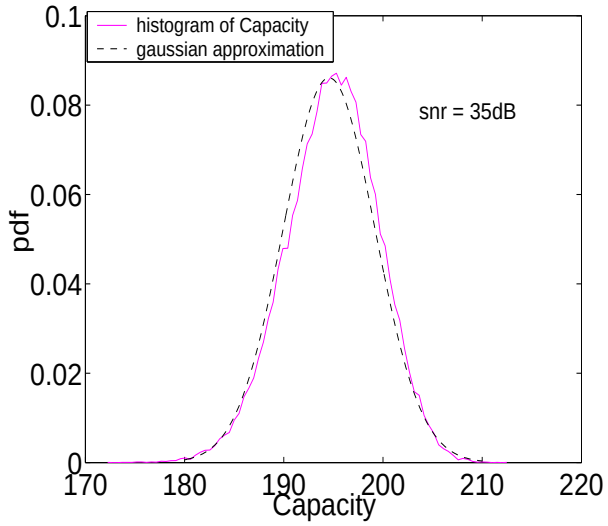

Fig. 7. The normalized histogram of Capacity superimposed by its density function approximated by Gaussian for $K = 8,\ L = 7,\ \rho = 0.3,\ N_R = 3,\ N_T = 3$ and $\gamma = 35dB$.

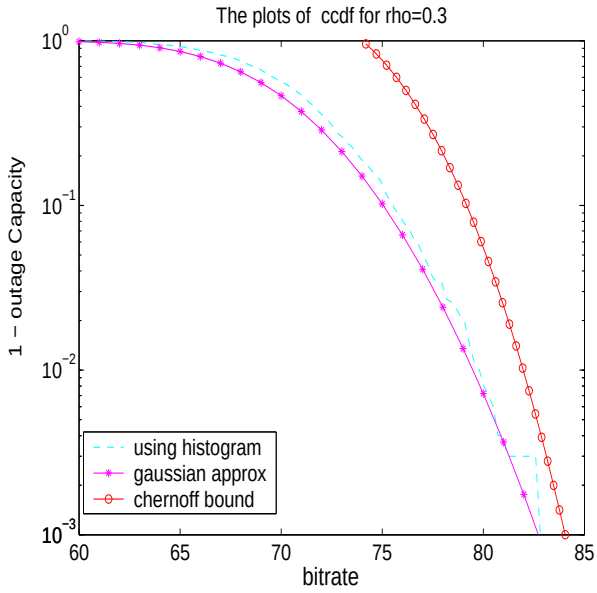

Fig. 8. The plots of ccdf for $K = 8,\ L = 7,\ \rho = 0.3,\ N_R = 2,\ N_T = 2$ and $\gamma = 20dB$.

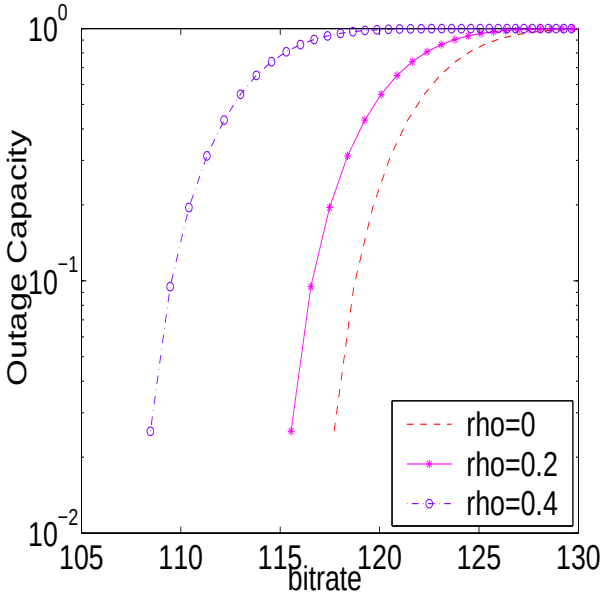

Fig. 9. The plots of Outage Capacity calculated theoretically for $K = 8,\ L = 7,\ N_R = 3,\ N_T = 3,\ \gamma = 20dB$ and varying ρ.