Statistical analysis of the Capacity of MIMO Frequency selective Rayleigh Fading Channels with arbitrary number of inputs and outputs

Anna Scaglione
School of Electrical & Computer
Engineering, Cornell University
Ithaca, NY 14853, USA

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. Simple expressions are derived for the case of uncorrelated Rayleigh fading and an arbitrary finite number of transmit and receive antennas.

Keywords — Information Theory and Statistics

I. Introduction

This paper is concerned with the derivation of the statistics of the channel Capacity of a MIMO frequency selective system. This task requires the non trivial step of deriving the joint statistics of the eigenvalues of the random MIMO frequency response. 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. 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, which have provided asymptotic results for similar problems [see e.g. [11]] the analysis developed in this paper applies for an arbitrary finite number of inputs and outputs and paves the road for the derivation of other performance measures which depend on the channel eigenvalues. These results are useful especially in the context of Space-Time coding [12]-[9], where the number of inputs and outputs is naturally limited to a few elements. Our MIMO channel is frequency selective, thus our setting is analogous to the one in [10].

Notation: Boldface letters are vectors (lower case) or matrices (upper case). The $tr(\boldsymbol{A})$, $|\boldsymbol{A}|, \lambda(\boldsymbol{A})$ are the trace, determinant and eigenvalues of $\boldsymbol{A}$, $\mathbf{a} = vec(\boldsymbol{A})$ is formed stacking vertically the columns of $\boldsymbol{A}$. 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 indexes $\mathbf{k} = (k_1, \ldots, k_m)$ and the notation $\boldsymbol{A}[\mathbf{k}] \triangleq (\boldsymbol{A}[k_1]^T, \ldots, \boldsymbol{A}[k_m]^T)^T$.

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 $\boldsymbol{x}(t)$ is related to the (coded) symbol vector $\boldsymbol{x}[n]$ 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 $\boldsymbol{y}(t)$ 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

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

where $g_R(t)$ is the impulse response of the lowpass 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

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

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 $\boldsymbol{x}_i \triangleq vec([\boldsymbol{x}[iP], \ldots, \boldsymbol{x}[iP+P-1]])$ and K received snapshots in a $MN_R \times 1$ vector $\boldsymbol{y}_i \triangleq vec([\boldsymbol{y}[iP+L], \ldots, \boldsymbol{y}[iP+P-1]])$, where $\boldsymbol{y}_i$ starts from the Lth array snapshot so that the inter-block interference (IBI) is not considered, we have

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

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

$$\boldsymbol{H} = \begin{pmatrix} \mathbf{H}[L] & \cdots & \mathbf{H}[0] & 0 & \ldots & 0 \\ 0 & \mathbf{H}[L] & \cdots & \mathbf{H}[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{H}[L] & \cdots & \mathbf{H}[0] \end{pmatrix}_{RM \times KP}. \tag{6}$$

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

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

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

$$\tilde{\boldsymbol{H}} \quad \triangleq \quad \boldsymbol{E}_R\boldsymbol{H}\boldsymbol{E}_T = diag(\tilde{\mathbf{H}}[\boldsymbol{d}])\ , \quad \boldsymbol{d} \triangleq (0,\dots,L),$$

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

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

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 it the channel model is often expressed only in terms of the statistics of the fading envelope $\alpha_{r,t}[l] \triangleq |\{\mathbf{H}[l]\}_{r,t}|$ of each path coefficient for the (r,t) link. The interesting and challenging aspect of the MIMO case is that the performance are expressed in terms of the eigenvalues of the matrix $\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}}$ 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 $\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}}$ are linked to the joint statistics of $\mathbf{H}(\boldsymbol{d}) = (\mathbf{H}^T[0],\dots,\mathbf{H}^T[l])^T$ 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 [13], 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 [2].

To keep the presentation self-cointained, next we informally introduce some of the concepts used in the statistical analysis on random matrices (see e.g. [4]). One of such tools is based on the seminal work of Élie Cartan on *exterior differential calculus* [3]. 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 two vectors $\alpha,\ \beta$ the basic axioms of Grassman algebra are:

- $\alpha\wedge\alpha = 0$
- $\alpha\wedge\ \beta = -\ \beta\wedge\alpha$
- $(a\alpha)\wedge\ \beta = a(\alpha\wedge\ \beta)$.

The axioms are sufficient to establish that:[1]

$$(\boldsymbol{A}\alpha)\wedge\beta = |\boldsymbol{A}|(\alpha\wedge\beta). \tag{9}$$

[1] If $\boldsymbol{A}$ is $m\times n$ and $m>n$ or if it is rank deficient $|\boldsymbol{A}|$ has to be replaced by 0. If $m\le n$ $|\boldsymbol{A}|$ and has to be replaced by the matrix compound $\wedge^m\boldsymbol{A}$, i.e. the matrix of all cofactors of order m, if $m\le n$ [3].

Cartan's exterior differential calculus [3] 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 applying Grassman algebra to 1-forms such as dx. There is an axiomatic definition of the d operator and, in particular, $d(r\text{-form}) = (r+1)\text{-form}$ and $d(dx)=0$ (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 [2] and indicate by $d\boldsymbol{A}$ the matrix of differentials and by $(d\boldsymbol{A})$ the exterior product of the independent entries in $d\boldsymbol{A}$, for example:

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

$$(x,\ y)\ \begin{matrix} a & -b \\ b & a \end{matrix}\ , \tag{10}$$

from (9) it follows that $(d\alpha z) = |\alpha|^2dxdy$. In general [4]:

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

$$det\ \frac{\partial(\boldsymbol{u},\boldsymbol{v})}{\partial(\boldsymbol{x},\boldsymbol{y})} = det\ \frac{\partial\boldsymbol{w}}{\partial\boldsymbol{z}}^{\,2} \tag{11}$$

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

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 $(d\boldsymbol{Q})$. A unitary $\boldsymbol{Q}$ 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 $\boldsymbol{Q}$. 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). It is to our advantage to remove this ambiguity by having the diagonal elements of $\boldsymbol{Q}$ set to be real. Note that, because of $\boldsymbol{Q}\boldsymbol{Q}^H = \boldsymbol{I} \ \rightarrow\ \boldsymbol{Q}d\boldsymbol{Q}^H = -d\boldsymbol{Q}\boldsymbol{Q}^H$: thus, when the diagonal elements of $\boldsymbol{Q}$ are real the diagonal elements of $\boldsymbol{Q}d\boldsymbol{Q}^H$ are zero and $\boldsymbol{Q}d\boldsymbol{Q}^H$ is antisymmetric. So, in most cases $(d\boldsymbol{Q})$ is replaced by

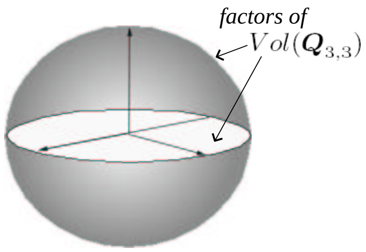

Figure 1: The factors in the $Vol(\boldsymbol{Q}_{3\times 3})$ for $\boldsymbol{Q}$ orthogonal.

$(\boldsymbol{Q}^H d\boldsymbol{Q}) = \wedge_{i>j} \boldsymbol{q}_i^H d\boldsymbol{q}_j$. Note also that, when $\boldsymbol{Q}$ is $m \times n$ and semi-unitary with $m \leq n$, 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 $\bar{\boldsymbol{Q}} = (\boldsymbol{Q}, \boldsymbol{Q}^{\perp})$ such that $\bar{\boldsymbol{Q}}^H \boldsymbol{Q} = \boldsymbol{I}_{m,n}$, so that $(d\boldsymbol{Q}) = (\bar{\boldsymbol{Q}}^H d\boldsymbol{Q})$.

Several different approaches can be taken to parametrize $\boldsymbol{Q}$ in its independent parameters, for example:

- $\boldsymbol{Q}$ is product of Givens rotations [Ch.5 [5]], i.e. for $\boldsymbol{Q}$ $n \times n$:
$$\boldsymbol{Q} = \prod_{k=1}^{n} \prod_{i=k+1}^{n} \boldsymbol{G}(k,i) \tag{12}$$
each $\boldsymbol{G}(k,i)$ has one parameter (the Euler angle) when $\boldsymbol{Q}$ is orthogonal and two when it is unitary;
- $\boldsymbol{Q}$ is product of n Housenholder rotations $\boldsymbol{H}_i = \boldsymbol{I} - 2\boldsymbol{v}_i\boldsymbol{v}_i^H/(\boldsymbol{v}_i^H\boldsymbol{v}_i)$ [Ch.5 [5]], where for $i = 0, \ldots, n-1$ each $\boldsymbol{v}_i$ is described by $n - i$ complex parameters.
- Decomposing $\boldsymbol{Q} = \boldsymbol{\Omega}_1 \boldsymbol{D} \boldsymbol{\Omega}_2$, with $\boldsymbol{\Omega}_i$, $i = 1, 2$ orthogonal matrices and $\boldsymbol{D} = diag(e^{j\phi_1}, \ldots, e^{j\phi_n})$.
- Using $\boldsymbol{Q} = e^{j\boldsymbol{\Theta}}$ where $\boldsymbol{\Theta}$ is Hermitian (the description is unique $\forall\, \boldsymbol{\Theta} :\ \boldsymbol{0} \leq \boldsymbol{\Theta} \leq \pi \boldsymbol{I}$).
- For $\boldsymbol{Q}$ not having eigenvalues equal to -1 (a probability zero event for continuous random $\boldsymbol{Q}$), the Cayley transform $\boldsymbol{Q} = (\boldsymbol{I} + j\boldsymbol{S})^{-1}(\boldsymbol{I} + j\boldsymbol{S})$ where $\boldsymbol{S}$ is skew Hermitian, i.e. $\boldsymbol{S}^H = -\boldsymbol{S}$. Note that $\boldsymbol{S} = (\boldsymbol{I} + \boldsymbol{Q})^{-1}(\boldsymbol{I} - \boldsymbol{Q})$.

The 3×3 orthogonal matrix case is illustrated in Fig. IV. The uniform p.d.f. in the Stiefel group of orthogonal or unitary matrices is called *Haar distribution* [4, Ch.1]. The volume of $(\bar{\boldsymbol{Q}}^H d\boldsymbol{Q})$ integrated over $\boldsymbol{Q}^H \boldsymbol{Q} = \boldsymbol{I}$, for $\boldsymbol{Q}$ unitary with real diagonal elements, is:

$$Vol(\boldsymbol{Q}_{m,n}) = \frac{2^n (\pi)^{mn - n(n-1)/2}}{\prod_{i=0}^{n-1} \Gamma(m-i)}. \tag{13}$$

when the diagonal elements of $\boldsymbol{Q}_{m,n}$ are constrained to be real:

$$\overline{Vol}(\boldsymbol{Q}_{m,n}) = \frac{(\pi)^{(m-1)n - n(n-1)/2}}{\prod_{i=0}^{n-1} \Gamma(m-i)}. \tag{14}$$

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.[2]

[2]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.

Let us denote by $p_{\boldsymbol{A}}(A)$ and $p_{\boldsymbol{B}}(B)$ 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 [6] which is based on the QR and Cholesky decompositions of $\boldsymbol{B}$ and $\boldsymbol{A}$ respectively:

$$\boldsymbol{B} = \boldsymbol{Q}\boldsymbol{R}\,, \quad \boldsymbol{A} = \boldsymbol{R}^H\boldsymbol{R}. \tag{15}$$

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

$$\begin{aligned}(d\boldsymbol{A}) &= \bigwedge_{1\leq i<j\leq n} \left(\sum_{1\leq k\leq i} dr_{k,i}^* r_{k,j} + dr_{k,j}^* r_{k,i} \right) \\ &= 2^n \prod_{i=1}^{n} (|r_{ii}|^2)^{n+1-i} (d\boldsymbol{R}) \end{aligned} \tag{16}$$

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

$$p_{\boldsymbol{A}}(\boldsymbol{A})(d\boldsymbol{A}) = p_{\boldsymbol{A}}(\boldsymbol{R}^H\boldsymbol{R}) \prod_{i=1}^{n} 2^n \ \ |r_{ii}|^2 \ \ ^{n+1-i} (d\boldsymbol{R}). \tag{17}$$

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:

$$(d\boldsymbol{B}) = \prod_{i=1}^{n} (|r_{ii}|^2)^{m+1-i} (d\boldsymbol{R})(d\boldsymbol{Q}), \tag{18}$$

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

$$p_{\boldsymbol{B}}(\boldsymbol{B})(d\boldsymbol{B}) = p_{\boldsymbol{B}}(\boldsymbol{Q}\boldsymbol{R}) \prod_{i=1}^{n} \ |r_{ii}|^2 \ \ ^{m+1-i} (d\boldsymbol{R})(d\boldsymbol{Q}), \tag{19}$$

and, with $\sqrt{\boldsymbol{A}} \triangleq \boldsymbol{R}$, from (17) and (19) and $|\boldsymbol{A}| = \prod_{i=1}^{n} |r_{ii}|^2$ it follows:

$$p_{\boldsymbol{A}}(\boldsymbol{A}) = 2^{-n} |\boldsymbol{A}|^{m-n} \int p_{\boldsymbol{B}}(\boldsymbol{Q}\sqrt{\boldsymbol{A}})(\boldsymbol{Q}^H d\boldsymbol{Q}), \tag{20}$$

which is the form of the so called *generalized Wishart density* [4]. Generalizing the results in [4] to the complex case (17) 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*

$$p_{\boldsymbol{R}}(\boldsymbol{R}) = \prod_{i=1}^{n} \ |r_{ii}|^2 \ \ ^{m-n} p(\boldsymbol{R}^H\boldsymbol{R}) Vol(\boldsymbol{Q}_{m,n}); \tag{21}$$

2) The p.d.f. of $\boldsymbol{A}$ *is [c.f. (13)]:*

$$p_{\boldsymbol{A}}(\boldsymbol{A}) = 2^{-n} |\boldsymbol{A}|^{m-n} p(\boldsymbol{A}) Vol(\boldsymbol{Q}_{m,n}), \tag{22}$$

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

$$(d\boldsymbol{A}) = (d\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H + \boldsymbol{U}d\boldsymbol{\Lambda}\boldsymbol{U}^H + \boldsymbol{U}^H\boldsymbol{\Lambda}d\boldsymbol{U}) \tag{23}$$

$$\begin{aligned}(d\boldsymbol{A}) &\equiv (\boldsymbol{U}^H d\boldsymbol{A}\boldsymbol{U}) = (\boldsymbol{U}^H d\boldsymbol{U}\boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{U}^H d\boldsymbol{U} + d\boldsymbol{\Lambda}) \\ &= \prod_{1\leq i<k\leq n}^{n} (\lambda_k - \lambda_i)^2 (d\boldsymbol{\Lambda})(\boldsymbol{U}^H d\boldsymbol{U}). \end{aligned} \tag{24}$$

Equations (17) and (24) are the equations that can be used to address the general case of $\boldsymbol{A} = \boldsymbol{B}^H \boldsymbol{B}$:

$$p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}) = 2^{-n} \prod_{1 \le i < k \le n} (\lambda_k - \lambda_i)^2 \left. \prod_{i=1}^{n} \lambda_i \right)^{m-n} \Psi(\lambda) \quad (25)$$

$$\Psi(\lambda) \triangleq \int p_{\boldsymbol{B}}(\boldsymbol{Q} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{U}^H)(\boldsymbol{Q}^H d\boldsymbol{Q})(\boldsymbol{U}^H d\boldsymbol{U}). \quad (26)$$

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. (22)] and, for $\lambda_i > 0$:

$$p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}) = \chi_1 \prod_{1 \le i < k \le n} (\lambda_k - \lambda_i)^2 e^{-\frac{\sum_i \lambda_i}{\sigma^2}} \left. \prod_{i=1}^{n} \lambda_i \right)^{m-n} \quad (27)$$

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 [7]:

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

$$E\{f(\boldsymbol{A})\} = \int_0^\infty f(x) \mu_n^{m-n}(x) dx \quad (28)$$

$$\mu_n^{m-n}(x) \triangleq \frac{1}{n!} \int_0^\infty \cdots \int_0^\infty p_{\boldsymbol{\Lambda}}(x, \lambda_2, \ldots, \lambda_n) d\lambda_2 \ldots d\lambda_n. \quad (29)$$

Note that, for $f(\boldsymbol{A}) = \sum_{i=1}^{n} \delta(x - \lambda_i(\boldsymbol{A}))$, $E\{f(\boldsymbol{A})\}$ in (28) 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 (27) [1], with $\alpha = m - n$:

$$\mu_n^\alpha(x) = \frac{1}{n} \sum_{k=0}^{n-1} \phi_k^\alpha(x)^2 \quad (30)$$

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

$$\phi_k^\alpha(x) = \left(\frac{k!}{\Gamma(k+\alpha+1)} x^\alpha e^{-x} \right)^{1/2} L_k^\alpha(x). \quad (31)$$

VI. Statistics of the the MIMO frequency selective channel

We will assume that:

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

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

Let us also denote by:

$$n \triangleq \min(N_T, N_R)\,, \quad m \triangleq \max(N_T, N_R). \quad (32)$$

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

$$C = \log |\boldsymbol{I} + \gamma \tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}}| \quad (33)$$

therefore the average Capacity is:

$$E\{C\} = \sum_{k=0}^{K-1} \sum_{l=1}^{N_T} E\left\{\log(1 + \gamma \lambda_l[k])\right\}. \quad (34)$$

and the characteristic function of C is:

$$\Phi_C(s) = E\{e^{sC}\} = E\left\{ \prod_{k=0}^{K-1} |\boldsymbol{I} + \gamma \tilde{\mathbf{H}}[k]^H \tilde{\mathbf{H}}[k]|^s \right\} \quad (35)$$

both functions of the eigenvalues of $\tilde{\mathbf{H}}[k]^H \tilde{\mathbf{H}}[k]$, $k = 0, \ldots, K-1$. The average Capacity can be easily derived explicitly. In fact, $\tilde{\mathbf{H}}[k]$ is given by (8) thus, under **a3**, $\tilde{\mathbf{H}}[k]$, $k, 0, \ldots, K-1$ are also zero mean complex Gaussian with variance:

$$\sigma_H^2[k] = \sum_{(l_1,l_2)=0}^{L} R_H(l_1, l_2) e^{-j2\pi \frac{(l_1 - l_2)k}{K}} \quad (36)$$

as a direct consequence of Lemma 3 we can write:

Corollary 1 *Under* **a1, a2**, *the average Capacity for any (n, m) in (32) is:*

$$E\{C\} = \sum_{k=0}^{K-1} \int_0^\infty \log\left(1 + \gamma \sigma_H^2[k] x\right) \mu_n^{m-n}(x) dx \quad (37)$$

where $\mu_n^{m-n}(x)$ is given in (30).

The derivation of $\Phi_C(s)$ is more complicated, since it requires averaging over the joint density of the eigenvalues of all $\tilde{\mathbf{H}}[k]^H \tilde{\mathbf{H}}[k]$, $k = 0, \ldots, K-1$. In general $K \ge L$ therefore from (8) the joint density of the MIMO channel response at all frequency bins is:

$$p_{\tilde{H}}(\tilde{\mathbf{H}}[\mathbf{k}]) = p(\tilde{\mathbf{H}}[\bar{\mathbf{p}}] \mid \tilde{\mathbf{H}}[\mathbf{p}]) p(\tilde{\mathbf{H}}[\mathbf{p}]) \quad (38)$$

where $\mathbf{k} = (0, \ldots, K)$, $\mathbf{p} = (k_0, \ldots, k_L)$ is a vector with as elements $L+1$ distinct, but otherwise arbitrary, indexes extracted from $\mathbf{k}$ and $\bar{\mathbf{p}}$ is the vector of the complementary indexes. The blocks of $\tilde{\mathbf{H}}[\mathbf{p}] = (\tilde{\mathbf{H}}^T[k_0], .., \tilde{\mathbf{H}}^T[k_L])^T$ are in a one to one mapping with the blocks of $\mathbf{H}(\boldsymbol{d}) = (\mathbf{H}^T[0], .., \mathbf{H}^T[L])^T$: 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 $\boldsymbol{W}_{L+1}^H$:

$$\{\boldsymbol{W}_{L+1}\}_{il} = e^{-j\frac{2\pi}{K} k_i l} \quad l \in [0, L],\ k_i \in \mathbf{p}, \quad (39)$$

Thus, $\exists\, c_{k_i l} = \{\boldsymbol{W}_{L+1}^{-1}\}_{k_i l}$ such that $\sum_{l=0}^{L} c_{k_i l} e^{-j\frac{2\pi h_j l}{K}} = \delta_{k_i h_j}$ (Kroneker δ). The $c_{k_i l}$ are computable as the coefficients of the Lth order Lagrange polynomials

$$C_{k_i}(z) \triangleq \prod_{j \ne i, 0 \le j \le L} \frac{z - e^{-j2\pi k_j / K}}{e^{-j2\pi k_i / K} - e^{-j2\pi k_j / K}} \quad (40)$$

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

$$\tilde{\mathbf{H}}[h_j] = \sum_{l=0}^{L} \mathbf{H}[l] e^{-j2\pi k_i l / K} = \sum_{l=0}^{L} \sum_{k_i \in \mathbf{p}} \tilde{\mathbf{H}}[k_i] c_{k_i l} e^{-j2\pi k_i l / K}. \quad (41)$$

From (40) and (41), it follows that $p(\tilde{\mathbf{H}}[\bar{\mathbf{p}}] \mid \tilde{\mathbf{H}}[\mathbf{p}])$ is product of Dirac deltas. With $C_{k_i, h_j} \triangleq C_{k_i}(e^{-j2\pi h_j / K})$ we have

$$p(\tilde{\mathbf{H}}[\bar{\mathbf{p}}] \mid \tilde{\mathbf{H}}[\mathbf{p}]) = \prod_{h_j \in \bar{\mathbf{p}}} \delta \left(\tilde{\mathbf{H}}[h_j] - \sum_{k_i \in \mathbf{p}} \tilde{\mathbf{H}}[k_i] C_{k_i h_j} \right) \quad (42)$$

$$p_{\tilde{\mathbf{H}}}(\tilde{\mathbf{H}}[\mathbf{p}]) = |\boldsymbol{W}_{L+1}|^{-N_R N_T} p_{\mathbf{H}}((\boldsymbol{W}_{L+1} \otimes \boldsymbol{I})^{-1} \tilde{\mathbf{H}}[\mathbf{p}]). \quad (43)$$

Gathering these results we can state the following:

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 $p_{\mathbf{H}}(\mathbf{H}(\boldsymbol{d}))$, $\boldsymbol{d} = (0,\ldots,L)$, $\mathbf{H}(\boldsymbol{d}) = (\mathbf{H}^T(0),\ldots,\mathbf{H}^T(L))^T$, the characteristic function of the mutual information is equal to:*

$$\Phi_c(s) = \chi_2 \int \prod_{h=0}^{K-1} \Upsilon^s(\boldsymbol{x}_h) p_{\mathbf{H}}((\boldsymbol{W}_{L+1}^{-1} \otimes \boldsymbol{I})\tilde{\mathbf{H}}[\mathbf{p}])(d\tilde{\mathbf{H}}[\mathbf{p}]) \quad (44)$$

where $\chi_2 = |\boldsymbol{W}_{L+1}|^{-N_R N_T}$, $\boldsymbol{W}_{L+1}$ is defined in (39), $\boldsymbol{W}_{L+1}^{-1}$ can be expressed in terms of the coefficients of the Lagrange polynomials in (40) and, with $\boldsymbol{x}_h \triangleq (C_{k_0 h},\ldots,C_{k_L h})$

$$\Upsilon(\boldsymbol{x}_h) \triangleq \boldsymbol{I} + \gamma \sum_{(i,\,l)=0}^{L} \tilde{\mathbf{H}}^H[k_i]\tilde{\mathbf{H}}[k_l] C^*_{k_i h} C_{k_l h} \quad (45)$$

where $C_{k_i h} \triangleq C_{k_i}(e^{-j2\pi h/K})$ and $C_{k_i}(z)$ is defined in (40).

To reach a simple expression for $\Phi_c(s)$ when $K \gg L$, 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 $e^{-j\frac{2\pi}{Q(L+1)} lQd} = e^{-j\frac{2\pi}{(L+1)} ld}$, $\boldsymbol{W}_{L+1}$ is unitary and

$$C_{nQ}(e^{-j2\pi(lQ+q)/K}) = \frac{1}{L+1}\,\frac{e^{-j2\pi\frac{q}{Q}} - 1}{e^{-j2\pi[\frac{(l-n)}{L+1} + \frac{q}{K}]} - 1} \quad (46)$$

In addition, let us assume that:
a4. In assumption **a3.** $R_H(l_1,l_2) = R_H(l_2 - l_1)$.
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 indexes, in force of Szëgo theorem for $L \gg 1$ the elements of $\tilde{\mathbf{H}}[\mathbf{p}]$ will be approximately uncorrelated not only in space, but also across the frequency bins. Under (**a3**) the p.d.f. of $\mathbf{H}(\boldsymbol{d})$ with $\boldsymbol{d} = (0,\ldots,L)$, is $\mathcal{N}(0,(\boldsymbol{I}\otimes \boldsymbol{R}_h))$ and therefore $\tilde{\mathbf{H}}[\mathbf{p}] \backsim \mathcal{N}(\mathbf{0},(\boldsymbol{W}_{L+1}^H \boldsymbol{R}_H \boldsymbol{W}_{L+1} \otimes \boldsymbol{I}))$ where $(\boldsymbol{W}_{L+1}^H \boldsymbol{R}_H \boldsymbol{W}_{L+1}) \approx diag(\Sigma_h^2[\boldsymbol{d}])$ where $\Sigma_h^2[\boldsymbol{d}] = (\sigma_H^2[0],..,\sigma_H^2[LQ])$ and $\sigma_h^2[lQ] = \sum_n R_H[n]e^{-j2\pi nl/(L+1)}$. Therefore, writing h as $h = lQ + q$, for $(l,q) = 0,\ldots,L$:

$$\Phi_C(s) = E\left\{\prod_{l=0}^{L}\prod_{q=0}^{Q-1} \Upsilon^s(\boldsymbol{x}_{(lQ+q)})\right\}, \quad (47)$$

we can simplify (47) as in the following

Corollary 2 *Under* **a1, a2, a3, a4** *for $L \gg 1$*

$$\Phi_c(s) \approx \chi_3 \prod_{l=0}^{L} \int \prod_{q=0}^{Q-1} \Upsilon^s(\boldsymbol{x}_{(lQ+q)}) e^{-\frac{tr(\tilde{\mathbf{H}}^H[lQ]\tilde{\mathbf{H}}[lQ])}{\sigma_h^2[l]}} (d\tilde{\mathbf{H}}[lQ]) \quad (48)$$

where $\chi_3 = \prod_{l=0}^{L} (\pi\sigma_h^2[l])^{N_T N_R}$

Knowing that $C_{nQ\,lQ} = \delta_{nl} \rightarrow \boldsymbol{x}_{lQ} = \ _l$ where $\ _l$ is the lth canonical vector, if Q small compared to L, we shall consider $\boldsymbol{x}_{(lQ+q)} \approx \boldsymbol{x}_{lQ} = \ _l$ and then replace $\Upsilon^s(\boldsymbol{x}_{(lQ+q)})$ for $|q| < Q/2$ and $l = 1,\ldots,L-1$[3], with a zero order approximation $\Upsilon^s(\boldsymbol{x}_{(lQ+q)}) \approx \Upsilon^s(\boldsymbol{x}_{(lQ)}) = \Upsilon^s(\ _l)$. This coarse approximation applied to (47) in conjunction with assumption **a3** for $L \gg 1$ leads to

$$\Phi_C(s) \approx \prod_{l=0}^{L} E\left\{|\boldsymbol{I} + \gamma\tilde{\mathbf{H}}^H[lQ]\tilde{\mathbf{H}}[lQ]|^{Qs}\right\}, \quad (49)$$

which has the advantage of being expressed directly in terms of the eigenvalues of $\tilde{\mathbf{H}}^H[lQ]\tilde{\mathbf{H}}[lQ]$. Therefore, using $p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda})$ in (27) for $\sigma^2 = 1$ we can write:

$$\Phi_C(s) \approx \prod_{l=0}^{L} \int_{\lambda_i \geq 0} \prod_{i=1}^{n} (1 + \gamma\sigma^2[lQ]\lambda_i)^{Qs} p_{\boldsymbol{\Lambda}}(\lambda_1,\ldots,\lambda_n)(d\boldsymbol{\Lambda}). \quad (50)$$

In case $\gamma\sigma^2[lQ] \gg 1$ to reduce the multivariate we can note that $(1+\gamma\sigma^2[lQ]\lambda_i)^{Qs} \approx (\gamma\sigma^2[lQ]\lambda_i)^{Qs}$ which gives us:

$$\begin{aligned}\Phi_C(s) \quad &\approx \quad \gamma^{Qsn} \quad \prod_{l=0}^{L} (\sigma^2[lQ])^{Qs+m-\frac{n(n+1)}{2}} \chi_1(l)\Bigg) \quad (51)\\ &\cdot \quad \prod_{i=1}^{n} (\Gamma(i)\Gamma(m-n+Qs+i))^{L+1},\end{aligned}$$

where $\chi_1(l)$ is given in (27) and

$$\prod_{i=1}^{n}\Gamma(i)\Gamma(\alpha+i) = \int_{\lambda_i \geq 0} \quad \prod_{i=1}^{n}\lambda_i\Bigg)^{\alpha} e^{-\sum_i \lambda_i} \prod_{1\leq i<k\leq n} (\lambda_k - \lambda_i)^2 (d\boldsymbol{\Lambda}). \quad (52)$$

References

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

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

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

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

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

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

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

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

[9] 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.

[10] G. G. Raleigh and J. M. Cioffi, "Spatiotemporal coding for wireless communications", *IEEE Trans. on Commun.*, Vol. 46, No. 3, pp. 357–366, March 1998.

[11] 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.

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

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

[3]For $l = 0$, we take $0 \leq q \leq \frac{Q}{2}$ and for $l = L$ we take $-\frac{Q}{2} \leq q \leq 0$.