Theory of Large Dimensional Random Matrices for Engineers

Jack W. Silverstein* Department of Mathematics, Box 8205, North Carolina State University, Raleigh, NC 27695-8205, USA, Email: jack@math.ncsu.edu Antonia M. Tulino

Dip. di Ing. Elettronica e delle Telecomunicazioni, Universita' degli Studi di Napoli, "Fedrico I Via Claudio 21, Napoli Italy, Email: atulino@princeton.edu

Abstract— In the last few years, the asymptotic distribution of the singular values of certain random matrices has emerged as a key tool in the analysis and design of wireless communication channels. These channels are characterized by random matrices that admit various statistical descriptions depending on the actual application. The goal of this paper is the investigation and application of random matrix theory with particular emphasis on the asymptotic theorems on the distribution of the squared singular values under various assumption on the joint distribution of the random matrix entries.

I. INTRODUCTION

Random matrices have fascinated mathematicians and physicists since they were first introduced in mathematical statistics by Wishart in 1928. After a slow start, the subject gained prominence when Wigner introduced the concept of statistical distribution of nuclear energy levels in 1950. Since then, random matrix theory has matured into a field with applications in many branches of physics and mathematics, and nowadays random matrices find applications in fields as diverse as the Riemann hypothesis, stochastic differential equations, statistical physics, chaotic systems, numerical linear algebra, neural networks, etc. Recently random matrices are also finding an increasing number of applications in the context of information theory and signal processing, which include among others: wireless communications channels, learning and neural networks, capacity of ad hoc networks, direction of arrival estimation in sensor arrays, etc. The earliest applications to wireless communication were the pioneering works of Foschini and Telatar in the mid-90s on characterizing the capacity of multiantenna channels. With works like [1], [2], [3] which, initially, called attention to the effectiveness of asymptotic random matrix theory in wireless communication theory, interest in the study of random matrices began and the singular value densities of random matrices and their asymptotics, as the matrix size tends to infinity, became an active research area in information/communication theory. In the last few years a considerable body of results on the fundamental information-theoretic limits of various wireless communication channels that makes substantial use of asymptotic random matrix theory, has emerged in the communications and information theory literature. An extended survey of the these results and works can be found in [4].

The aim of this paper is dual: we first review some of the most interesting existing mathematical results that are relevant to the analysis of the statistics of random matrices arising in wireless communications. The emphasis will be on asymptotic distribution of the squared singular-values under various assumptions on the joint distribution of the random matrix coefficients. Those results are then exploited in order to assess the fundamental limits of wireless communication channels in the asymptotic regime where the number of columns and rows of the channel matrix H goes to infinity while the aspect ratio of the matrix is kept constant. Specifically we focus on two performance measures of engineering interest: Shannon capacity and linear minimum meansquare error, which are determined by the distribution of the squared singular values of the channel matrix.

II. WIRELESS COMMUNICATION CHANNELS

A typical wireless communication channel is described by the usual linear vector memoryless channel:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1}$$

where **x** is the *K*-dimensional vector of the signal input, **y** is the *N*-dimensional vector of the signal output, and the *N*-dimensional vector **n** is the additive Gaussian noise, whose components are independent complex Gaussian random variables with zero mean and independent real and imaginary parts with the same variance $\sigma^2/2$ (i.e., circularly distributed). **H**, in turn, is the $N \times K$ complex random matrix describing the channel whose entries admit various statistical descriptions depending on the actual applications. Fading, wideband, multiuser and multi-antenna are some of the key features

^{*}Research supported by the U.S. Army Research Office under Grant W911NF-05-1-0244 $\,$

that characterize wireless channels of contemporary interest. In each of these cases, N, K and **H** take different meanings. We will focus particular attention on a few models that capture various features of interest, each of them corresponding to a particular **H**-model:

- *A.* Randomly spread Code Division Multiple Access (CDMA) channels.
- *B.* Single-user multiantenna systems subject to frequency-fiat fading.
- C. CDMA channels with multiple receiving antennas.

Naturally, random matrices also arise in models that incorporate more than one of the above features (multiuser, multiantenna, fading, wideband). Although realistic models do include several (if not all) of the above features it is conceptually advantageous to start by deconstructing them into their essential ingredients. In the next subsections we describe the foregoing three scenarios and show how the distribution of the squared singular values of certain matrices determine communication limits in both the coded regime (Shannon capacity) and the uncoded regime (probability of error).

A. CDMA

An application that is very suitable is code-division multiple access channel or CDMA channel, ere each user is assigned a signature vector known at the receiver which n be seen as an element of an N dimensional signal space. Based on the nature of this signal space we can distinguish between

- Direct sequence CDMA used in many current cellular systems (IS-95, cdma2000, UMTS)
- Multi-Carrier CDMA being considered for Fourth Generation of cellular systems.

1) DS-CDMA Frequency-flat fading: Concerning the DS-CDMA, we first focus on channels whose response is flat over the signal bandwidth which implies that the received signature of each user is just a scaled version of the transmitted one where the scaling factors are the independent fading coefficients for each user. Considering the basic synchronous DS-CDMA [1, Sec. 2.9.2] with K users and spreading factor N in a frequency-flat fading environment, in this case, the vector \mathbf{x} contains the symbols transmitted by the K users while the role of \mathbf{H} is played by the product of two matrices, \mathbf{S} and \mathbf{A} , where \mathbf{S} is a $N \times K$ matrix whose columns are the spreading sequences

$$\mathbf{S} = [\mathbf{s}_1 \mid \dots \mid \mathbf{s}_K]. \tag{2}$$

and **A** is a $K \times K$ diagonal matrix of complex fading coefficients. The model thus specializes to

$$\mathbf{y} = \mathbf{S}\mathbf{A}\mathbf{x} + \mathbf{n}.$$
 (3)

The standard random signature model [1, Sec. 2.3.5] assumes that the entries of **S**, are chosen independently and equiprobably on $\{-\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{N}}\}^{1}$.

The unfaded equal power case is obtained by the the above model assuming $\mathbf{A} = A\mathbf{I}$, where A is the transmitted amplitude equal for all users.

Let us consider a synchronous DS-CDMA downlink with K active users employing random spreading codes and operating over a frequency-selective fading channel. Then **H** in (1) particularizes to

$$\mathbf{H} = \mathbf{CSA} \tag{4}$$

where **A** is a $K \times K$ deterministic diagonal matrix containing the amplitudes of the users and **C** is an $N \times N$ Toeplitz matrix defined as

$$(\mathbf{C})_{i,j} = \frac{1}{W_c} c\left(\frac{i-j}{W_c}\right)$$
(5)

with $c(\cdot)$ the impulse response of the channel.²

2) Multi-Carrier CDMA: If the channel is not flat over the signal bandwidth, then the received signature of each user is not simply a scaled version of the transmitted one. In this case, we can insert suitable transmit and receive interfaces and choose the signature space in such a way that the equivalent channel that encompasses the actual channel plus the interfaces can be modeled as a random matrix **H** given by:

$$\mathbf{H} = \mathbf{G} \circ \mathbf{S}$$

where \circ denotes the Hadamard (element-wise) product [5], *Sm* is the random signature matrix in the frequency domain, while **G** is an *N* × *K* matrix whose columns are independent *N*-dimensional random vectors whose (ℓ, k) -th element denotes the fading coefficients for the ℓ -th subcarrier of the *k*-th user, independent across the users. Thus the linear model (1) specializes to

$$\mathbf{y} = (\mathbf{G} \circ \mathbf{S})\mathbf{x} + \mathbf{n}.$$
 (6)

In the downlink case where **G** is given by the outer product of two vectors $\mathbf{G} = \mathbf{c}\mathbf{a}^T$ then $\mathbf{H} = \mathbf{CSA}$ with $\mathbf{C} = \text{diag}(\mathbf{c})$ and $\mathbf{A} = \text{diag}(\mathbf{a})$.

B. Multi-antenna Channels

Let us now consider a single-user channel where the transmitter has $n_{\rm T}$ antennas and the receiver has $n_{\rm R}$ antennas. In this case, **x** contains the symbols transmitted from the $n_{\rm T}$ transmit antennas and **y** the symbols received by the $n_{\rm R}$ receive antennas. With frequency-flat

²For contributions on the asymptotic analysis of uplink DS-CDMA systems in frequency selective fading channels see [4] and references therein.

¹One motivation for this is the use of "long sequences" in some commercial CDMA systems, where the period of the pseudo-random sequence spans many symbols. Another motivation is to provide a baseline of comparison for systems that use signature waveform families with low cross-correlations.

fading, the entries of \mathbf{H} represent the fading coefficients between each transmit and each receive antenna, typically modelled as zero-mean complex Gaussian. If all antennas are co-polarized, \mathbf{H} is identically distributed and thus we can factor out their variance g.

Antenna correlation at the transmitter and at the receiver, that is, between the columns and between the rows of **H**, respectively, can be accounted for through corresponding correlation matrices $\Theta_{\rm T}$ and $\Theta_{\rm R}$ [6], [7], [8]. A Ricean term can be incorporated through an additional deterministic matrix **H**₀ containing unitmagnitude entries [9], [10], [11]. With proper weighting of the random and deterministic matrices so that the entries of **H** retain their second-order moment of g, the model particularizes to

$$\mathbf{y} = \sqrt{\mathsf{g}} \left(\sqrt{\frac{1}{\mathsf{K}+1}} \mathbf{\Theta}_{\mathsf{R}}^{1/2} \mathbf{H}_{w} \mathbf{\Theta}_{\mathsf{T}}^{1/2} + \sqrt{\frac{\mathsf{K}}{\mathsf{K}+1}} \mathbf{H}_{0} \right) \mathbf{x} + \mathbf{n}$$
(7)

with \mathbf{H}_w an iid $\mathcal{N}(0,1)$ matrix and with the Ricean K-factor quantifying the ratio between the deterministic (unfaded) and the random (faded) energies [12].

If antennas with different polarizations are used, the entries of **H** are no longer identically distributed because of the different power transfer between co-polarized and differently polarized antennas. In that case,

$$\mathbf{H} = \mathbf{P} \circ \left(\sqrt{\frac{1}{\mathsf{K}+1}} \mathbf{\Theta}_{\mathsf{R}}^{1/2} \mathbf{H}_{w} \mathbf{\Theta}_{\mathsf{T}}^{1/2} + \sqrt{\frac{\mathsf{K}}{\mathsf{K}+1}} \mathbf{H}_{0} \right)$$
(8)

and with \mathbf{P} containing the square-root of the secondorder moment of each entry of \mathbf{H} , which is given by the relative polarization of the corresponding antenna pair. If all antennas are co-polar, then every entry of \mathbf{P} equals g. For an extended survey on contributions on the asymptotic analysis of multi-antenna channels see [4], [13] and references therein.

C. CDMA channels with multiple receiving antennas.

Suppose that we incorporate the features of *A*. and *B*. but with a single transmitter antenna (or equivalently *K* represents the number of users times the transmitting antennas). If we have $n_{\rm R}$ receiving antennas, then we have a model with $Nn_{\rm R}$ observables:

$$\mathbf{H} = \begin{bmatrix} \mathbf{SA}_1 \\ \cdots \\ \mathbf{SA}_{n_{\mathrm{R}}} \end{bmatrix}$$
(9)

where

$$\mathbf{A}_{\ell} = \operatorname{diag}\{\mathsf{A}_{1,\ell}, \dots, \mathsf{A}_{K,\ell}\}, \qquad \ell = 1, \dots n_{\mathrm{R}} \quad (10)$$

and $\{A_{k,\ell}\}$ indicates the i.i.d. fading coefficients of the *k*th user at the ℓ th antenna.

D. Why Asymptotic Random Matrix Theory?

Let us now talk about the role of random matrices and their singular values in wireless communication through the derivation of some key performance measures of wireless channels, which are determined by the distribution of the singular values of the channel matrix.

The empirical cumulative distribution function (c.d.f) of the eigenvalues (also referred to as the empirical spectral distribution (ESD)) of an $N \times N$ Hermitian matrix **A** is defined as

$$\mathsf{F}_{\mathbf{A}}^{N}(x) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{\lambda_{i}(\mathbf{A}) \le x\}$$
(11)

where $\lambda_1(\mathbf{A}), \ldots, \lambda_N(\mathbf{A})$ are the eigenvalues of \mathbf{A} and $1\{\cdot\}$ is the indicator function. If $\mathsf{F}^N_{\mathbf{A}}(\cdot)$ converges a.s as $N \to \infty$, then the corresponding limit (asymptotic ESD) is denoted by $\mathsf{F}_{\mathbf{A}}(\cdot)$.

The first performance measure that we are going to consider is the mutal information. If the channel is known by the receiver, and the input x is Gaussian the normalized mutual information in (1) conditioned on **H** is given by

$$\mathcal{J}(\mathsf{SNR}) = \frac{1}{N} I(\mathbf{x}; \mathbf{y} | \mathbf{H})$$
(12)

$$= \frac{1}{N} \log \det \left(\mathbf{I} + \operatorname{SNR} \mathbf{H} \boldsymbol{\Phi} \mathbf{H}^{\dagger} \right)$$
(13)

$$= \frac{1}{N} \sum_{i=1}^{N} \log \left(1 + \operatorname{SNR} \lambda_i (\mathbf{H} \boldsymbol{\Phi} \mathbf{H}^{\dagger}) \right)$$
$$= \int_0^{\infty} \log \left(1 + \operatorname{SNR} x \right) d\mathsf{F}_{\mathbf{H} \boldsymbol{\Phi} \mathbf{H}^{\dagger}}^N(x) \quad (14)$$

with the transmitted signal-to-noise ratio (SNR)

$$\mathsf{SNR} = \frac{N\mathbb{E}[||\mathbf{x}||^2]}{K\mathbb{E}[||\mathbf{n}||^2]},\tag{15}$$

 $\Phi = \frac{\mathbb{E}[\mathbf{x}\mathbf{x}^{\dagger}]}{\mathbb{E}[|\mathbf{x}||^2]}$, and finally $\lambda_i(\mathbf{H}\Phi\mathbf{H}^{\dagger})$ equal to the *i*th squared singular value of $\mathbf{H}\Phi^{1/2}$.

If the channel is known at the receiver and its variation over time is stationary and ergodic, then the expectation of (12) over the distribution of **H** is the ergodic mutual information (normalized to the number of receive antennas or the number of degrees of freedom per symbol in the CDMA channel). More generally, the distribution of the random variable (12) determines the outage capacity (e.g. [14]).

For $\text{SNR}\to\infty,$ a regime of interest in short-range applications, the mutual information admits the following affine expansion

$$\mathcal{I}(\mathsf{SNR}) = S_{\infty} \left(\log \mathsf{SNR} + \mathcal{L}_{\infty} \right) + o(1) \tag{16}$$

where the key measures are the high-SNR slope

$$S_{\infty} = \lim_{\mathsf{SNR} \to \infty} \frac{\mathfrak{I}(\mathsf{SNR}\,)}{\log \mathsf{SNR}} \tag{17}$$

which for most channels gives $S_{\infty} = \min \{\frac{K}{N}, 1\}$, and the *power offset*

$$\mathcal{L}_{\infty} = \lim_{\mathsf{SNR} \to \infty} \log \mathsf{SNR} - \frac{\mathsf{J}(\mathsf{SNR})}{S_{\infty}} \tag{18}$$

which essentially boils down to $\log \det(\mathbf{H}\mathbf{H}^{\dagger})$ or $\log \det(\mathbf{H}^{\dagger}\mathbf{H})$ depending on whether K > N or K < N.

Another important performance measure for (1) is the minimum mean-square-error (MMSE) achieved by a linear receiver, which determines the maximum achievable output signal-to-interference-and-noise ratio (SINR). For an i.i.d. input, the arithmetic mean over the users (or transmit antennas) of the MMSE is given, as a function of \mathbf{H} , by [1]

$$\mathsf{MMSE}(\mathsf{SNR}) = \frac{1}{K} \min_{\mathbf{M} \in \boldsymbol{C}^{K \times N}} \mathbb{E}\left[||\mathbf{x} - \mathbf{My}||^2 \right]$$
(19)

$$= \frac{1}{K} \operatorname{tr} \left\{ \left(\mathbf{I} + \operatorname{snr} \mathbf{H}^{\dagger} \mathbf{H} \right)^{-1} \right\}$$
(20)

$$= \frac{1}{K} \sum_{i=1}^{K} \frac{1}{1 + \operatorname{SNR} \lambda_i(\mathbf{H}^{\dagger} \mathbf{H})}$$
(21)
$$= \int_{-\infty}^{\infty} \frac{1}{-1} d\mathbf{F}^K (x)$$

$$= \int_{0}^{\infty} \frac{1}{1 + \operatorname{SNR} x} d\mathbf{F}_{\mathbf{H}^{\dagger}\mathbf{H}}(x)$$
$$= \frac{N}{K} \int_{0}^{\infty} \frac{1}{1 + \operatorname{SNR} x} d\mathbf{F}_{\mathbf{H}\mathbf{H}^{\dagger}}^{N}(x) - \frac{N - K}{K}$$
(22)

where the expectation in (19) is over \mathbf{x} and \mathbf{n} while (22) follows from

$$N\mathsf{F}_{\mathbf{H}\mathbf{H}^{\dagger}}^{N}(x) - Nu(x) = K\mathsf{F}_{\mathbf{H}^{\dagger}\mathbf{H}}^{K}(x) - Ku(x)$$
(23)

where u(x) is the unit-step function $(u(x) = 0, x \le 0; u(x) = 1, x > 0)$. Note, incidentally, that both performance measures as a function of SNR are coupled through

$$\frac{d}{d \mathrm{SNR}} \log_e \det \left(\mathbf{I} + \mathrm{SNR} \, \mathbf{H} \mathbf{H}^\dagger \right) = \frac{K - \mathrm{tr} \left\{ \left(\mathbf{I} + \mathrm{SNR} \, \mathbf{H}^\dagger \mathbf{H} \right)^{-1} \right\}}{\mathrm{SNR}}.$$

As we see in (14) and (22), both fundamental performance measures (mutual information and MMSE) are dictated by the distribution of the empirical (squared) singular value distribution of the random channel matrix. It is thus of paramount importance, in order to evaluate these—and other—performance measures, to be able to express this empirical distribution. Since $F_{HH^{\dagger}}^{N}$ clearly depends on the specific realization of H, so do (12) and (19) above. In terms of engineering insight, however, it is crucial to obtain expressions for the performance measures that do not depend on the single matrix realization, to which end two approaches are possible:

- To study the average behavior³ by taking an expectation of the performance measures over **H**, which requires assigning a probabilistic structure to it.
- The second approach is to consider an operative regime where the performance measures (12) and

(19) do not depend on the specific choice of signatures.

Asymptotic analysis (in the sense of large dimensional systems, i.e $K, N \to \infty$ with $\frac{K}{N} \to \beta$) is where both these approaches meet. First, the computation of the average performance measures simplifies as the dimensions grow to infinity. Second, the asymptotic regime turn out to be the operative regime where the dependencies of (12) and (19) on the realization of **H** disappear. Specifically, in most of the cases, asymptotic random matrix theory guarantees that as the dimensions of **H** go to infinity but their ratio is kept constant, its empirical singular-value distribution displays the following properties, which are key to the applicability to wireless communication problems:

- Insensitivity of the asymptotic eigenvalue distribution to the probability density function of the random matrix entries.
- An "ergodic" nature in the sense that—with probability one—the eigenvalue histogram of any matrix realization converges almost surely to the asymptotic eigenvalue distribution.
- Fast convergence rate of the empirical singularvalue distribution to its asymptotic limit [15], [16], which implies that that even for small values of the parameters, the asymptotic results come close to the finite-parameter results.

All these properties are very attractive in terms of analysis but are also of paramount importance at the design level (see[4, Sec. 3.1.6] and references therein).

Closely related to the MMSE is the SINR achieved at the output of a MMSE receiver. Denote by \hat{x}_k the estimate of the *k*th component of **x** and by MMSE_{*k*} the corresponding MMSE, we have

$$\mathsf{SINR}_k = \frac{E[|\hat{x}_k|^2] - \mathsf{MMSE}_k}{\mathsf{MMSE}_k}$$

Typically, the estimator sets $E[|\hat{x}_k|^2] = 1$ and thus

$$ext{SINR}_k = rac{1 - ext{MMSE}_k}{ ext{MMSE}_k} = ext{SNR}\, \mathbf{h}_k^\dagger \left(\mathbf{I} + ext{SNR}\, \sum_{j
eq k} \mathbf{h}_j \mathbf{h}_j^\dagger
ight)^{-1} \mathbf{h}_k$$

with the aid of the matrix inversion lemma. Often it is convenient to work with the normalized version

$$\frac{\mathsf{SINR}_k}{\mathsf{SNR}\,\|\mathbf{h}_k\|^2}$$

For $K, N \to \infty$ with $\frac{K}{N} \to \beta$, both SINR and normalized SINR can be written as a function of the asymptotic ESD of **HH**[†].

III. MATHEMATICAL BACKGROUND

In this section, we review a wide range of existing mathematical results that are very relevant to the analysis of the statistics of random matrices (and of their

³It is worth emphasizing that, in many cases, resorting to the expected value of the mutual information is motivated by the stronger consideration that: in problems such as aperiodic DS-CDMA or multiantenna with an ergodic channel, it is precisely the expected capacity that has real operational meaning.

matrix factorizations, such as the singular-value decomposition) arising in single- and multiuser communication theory.

For our purposes, it is advantageous to make use of the η -transform and the Shannon transform, which were motivated by the application of random matrix theory to various problems in the information theory of noisy communication channels [4]. These transforms, intimately related with each other and with the Stieltjes transform traditionally used in random matrix theory [17], characterize the spectrum of a random matrix while carrying certain engineering intuition.

Definition 1 Given an $N \times N$ Hermitian matrix **A** whose ESD converges almost surely, its Stieltjes transform is

$$S_{\mathbf{A}}(z) = \mathbb{E}\left[\frac{1}{X-z}\right] = \int \frac{1}{\lambda-z} dF_{\mathbf{A}}(\lambda)$$

where X denotes a random variable whose distribution is the asymptotic ESD of **A**.

Definition 2 *Given an* $N \times N$ *nonnegative definite random matrix* **A** *whose ESD converges almost surely, its* η *-transform is*

$$\eta_{\mathbf{A}}(\gamma) = \mathbb{E}\left[\frac{1}{1+\gamma X}\right]$$
(24)

while its Shannon transform is defined as

$$\mathcal{V}_{\mathbf{A}}(\gamma) = \mathbb{E}[\log(1 + \gamma X)] \tag{25}$$

where X is a nonnegative random variable whose distribution is the asymptotic ESD of **A** while γ is a nonnegative real number.

From the definition $0 < \eta_X(\gamma) \le 1$. $\eta_{\mathbf{A}}(\gamma)$ can be regarded as a generating function for the asymptotic moments of **A** [4]. As it turns out, the Shannon and η -transforms are intimately related with each other and with the Stieltjes transform though the following relationships:

$$\frac{\gamma}{\log e} \frac{d}{d\gamma} \mathcal{V}_{\mathbf{A}}(\gamma) = 1 - \frac{1}{\gamma} \mathcal{S}_{\mathbf{A}}\left(-\frac{1}{\gamma}\right) = 1 - \eta_{\mathbf{A}}(\gamma)$$

Assume that, as $K, N \to \infty$ with $\frac{K}{N} \to \beta$, the ESD of \mathbf{HH}^{\dagger} in (1) converges almost surely to a nonrandom limit. Based on the above definitions we immediately recognize from (14) and (22) that for an i.i.d. Gaussian input \mathbf{x} , as $K, N \to \infty$ with $\frac{K}{N} \to \beta$ the normalized mutual information and the MMSE of (1) are related to η -and Shannon transform of \mathbf{HH}^{\dagger} by:

$$\mathfrak{I}(\mathsf{SNR}\,) \to \mathcal{V}_{\mathbf{HH}^{\dagger}}(\mathsf{SNR}\,) \tag{26}$$

$$\mathsf{MMSE}(\mathsf{SNR}) \to \eta_{\mathbf{H}^{\dagger}\mathbf{H}}(\mathsf{SNR}) = 1 - \frac{1 - \eta_{\mathbf{H}\mathbf{H}^{\dagger}}(\mathsf{SNR})}{\beta} \qquad (27)$$

where (27) follows from (23).

In the following, we give some of the more representative results on the η - and Shannon transform (and thus on the fundamental limits: normalized mutual information, MMSE, etc.) of the various random (channel) matrices that arise in the analysis of the wireless communications models described in Section II.

Theorem 1 [2] If the entries of **H** are zero-mean i.i.d. with variance $\frac{1}{N}$, as $K, N \to \infty$ with $\frac{K}{N} \to \beta$, the ESD of **HH**[†] converges a.s. to the Marčenko-Pastur law whose density function is

$$\tilde{\mathsf{f}}_{\beta}(x) = (1-\beta)^+ \,\delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi x}$$
 (28)

while the η - and Shannon transforms are

$$\eta_{\mathbf{H}\mathbf{H}^{\dagger}}(\gamma) = 1 - \frac{\mathcal{F}(\gamma,\beta)}{4\gamma}$$
(29)

and

$$\mathcal{V}_{\mathbf{H}\mathbf{H}^{\dagger}}(\gamma) = \beta \log \left(1 + \gamma - \frac{\mathcal{F}(\gamma, \beta)}{4} \right) \\ + N \log \left(1 + \gamma \beta - \frac{\mathcal{F}(\gamma, \beta)}{4} \right) \\ - \frac{N \log e}{4\gamma} \mathcal{F}(\gamma, \beta)$$
(30)

with

$$\mathcal{F}(x,z) = \left(\sqrt{x(1+\sqrt{z})^2 + 1} - \sqrt{x(1-\sqrt{z})^2 + 1}\right)^2$$

Theorem 2 [18], [19] Let **S** be an $N \times K$ matrix whose entries are i.i.d. complex random variables with zero-mean and variance $\frac{1}{N}$. Let **T** be a $K \times K$ real diagonal random matrix whose empirical eigenvalue distribution converges almost surely to the distribution of a random variable **T**. Let \mathbf{W}_0 be an $N \times N$ Hermitian complex random matrix with empirical eigenvalue distribution converging almost surely to a nonrandom distribution whose Stieltjes transform is S_0 . If \mathbf{H} , \mathbf{T} , and \mathbf{W}_0 are independent, the empirical eigenvalue distribution of

$$\mathbf{W} = \mathbf{W}_0 + \mathbf{STS}^\dagger \tag{31}$$

converges, as $K, N \to \infty$ with $\frac{K}{N} \to \beta$, almost surely to a nonrandom limiting distribution whose Stieltjes transform, $S(\cdot)$, satisfies $z \in \mathbb{C}^+$

$$S(z) = S_0 \left(z - \beta \mathbb{E} \left[\frac{\mathsf{T}}{1 + \mathsf{T}S(z)} \right] \right)$$
(32)

Using the η -transform, we reformulate the following results from [20] in terms of the η -transform.

Theorem 3 [4] Let **S** be an $N \times K$ complex random matrix whose entries are i.i.d. with variance $\frac{1}{N}$. Let **T** be a $K \times K$ nonnegative definite random matrix, whose ESD converges almost surely to a nonrandom distribution. The ESD of **STS**[†] converges a.s., as $K, N \to \infty$ with $\frac{K}{N} \to \beta$, to a distribution whose η -transform satisfies

$$\beta = \frac{1 - \eta}{1 - \eta_{\mathbf{T}}(\gamma \eta)} \tag{33}$$

where we have compactly abbreviated $\eta_{\mathbf{STS}^{\dagger}}(\gamma) = \eta$. The corresponding Shannon transform is

$$\mathcal{V}_{\mathbf{STS}^{\dagger}}(\gamma) = \beta \mathcal{V}_{\mathbf{T}}(\eta \gamma) + \log \frac{1}{\eta} + (\eta - 1) \log e$$
 (34)

Theorem 4 [21] Define $\mathbf{H} = \mathbf{CSA}$ where \mathbf{S} is an $N \times K$ matrix whose entries are i.i.d. complex random variables with variance $\frac{1}{N}$. Let \mathbf{C} and \mathbf{A} be, respectively, $N \times N$ and $K \times K$ random matrices such that the asymptotic spectra of $\mathbf{D} = \mathbf{CC}^{\dagger}$ and $\mathbf{T} = \mathbf{AA}^{\dagger}$ converge almost surely to a nonrandom limit. If \mathbf{C} , \mathbf{A} and \mathbf{S} are independent, as $K, N \to \infty$ with $\frac{K}{N} \to \beta$, the Stieltjes transform of \mathbf{HH}^{\dagger} for each $z \in \mathbb{C}^+$ satisfies

$$S_{\mathbf{H}\mathbf{H}^{\dagger}}(z) = \mathbb{E}\left[\frac{1}{\beta \, \mathsf{D} \, \mathbb{E}\left[\frac{\mathsf{T}}{1+F(z)\mathsf{T}}\right] - z}\right]$$
(35)

where F(z) satisfies

$$F(z) = \mathbb{E}\left[\frac{\mathsf{D}}{\beta \,\mathsf{D}\,\mathbb{E}\left[\frac{\mathsf{T}}{1+F(z)\,\mathsf{T}}\right] - z}\right] \tag{36}$$

with D and T independent random variables whose distributions are the asymptotic spectra of D and T respectively.

The following result finds the η - and the Shannon transform of \mathbf{HH}^{\dagger} in terms of the η - and Shannon transforms of **D** and **T**.

Theorem 5 [13] Let **H** be an $N \times K$ matrix as defined in Theorem 4. The Shannon transform of **HH**^{\dagger} is given by:

$$\mathcal{V}_{\mathbf{H}\mathbf{H}^{\dagger}}(\gamma) = \mathcal{V}_{\mathbf{D}}(\beta\gamma_d) + \beta\mathcal{V}_{\mathbf{T}}(\gamma_t) - \beta\frac{\gamma_d\gamma_t}{\gamma}\log e \qquad (37)$$

where

$$\frac{\gamma_d \gamma_{\mathbf{t}}}{\gamma} = 1 - \eta_{\mathbf{T}}(\gamma_{\mathbf{t}}) \qquad \beta \frac{\gamma_d \gamma_{\mathbf{t}}}{\gamma} = 1 - \eta_{\mathbf{D}}(\beta \gamma_d) \tag{38}$$

while the η -transform of **HH**^{\dagger} can be obtained as

$$\eta_{\mathbf{H}\mathbf{H}^{\dagger}}(\gamma) = \eta_{\mathbf{D}}(\beta \, \gamma_d(\gamma)) \tag{39}$$

where $\gamma_d(\gamma)$ is the solution to (38).

The asymptotic fraction of zero eigenvalues of \mathbf{HH}^{\dagger} equals

$$\lim_{\gamma \to \infty} \eta_{\mathbf{H}\mathbf{H}^{\dagger}}(\gamma) = 1 - \min\left\{\beta \, \mathbb{P}[\mathsf{T} \neq 0], \mathbb{P}[\mathsf{D} \neq 0]\right\}$$

It is easy to verify that:

$$F\left(-\frac{1}{\gamma}\right) = \gamma_{t}(\gamma). \tag{40}$$

Moreover, given a linear memoryless vector channel as in (1) with the channel matrix **H** defined as in Theorem 4, it has been proved in [4] and [13] that:

$$\frac{\operatorname{SINR}_{k}}{\operatorname{SNR} \|\mathbf{h}_{k}\|^{2}} \to \frac{\gamma_{\mathsf{t}}(\operatorname{SNR})}{\operatorname{SNR} \mathbb{E}[\mathsf{D}]}.$$
(41)

 $\beta' > 1$

Theorem 6 [13], [22] Let **H** be an $N \times K$ matrix defined as in Theorem 4. Defining

$$\beta' = \beta \frac{\mathbb{P}[\mathsf{T} \neq 0]}{\mathbb{P}[\mathsf{D} \neq 0]},$$

$$\lim_{\gamma \to \infty} \left(\log(\gamma \beta) - \frac{\mathcal{V}_{\mathbf{H}\mathbf{H}^{\dagger}}(\gamma)}{\min\left\{\beta \mathbb{P}[\mathsf{T} \neq 0], \mathbb{P}[\mathsf{D} \neq 0]\right\}} \right) = \mathcal{L}_{\infty}$$
(42)
with

$$\int -\mathbb{E}\left[\log \frac{\mathbb{P}[\mathsf{T}\neq 0]\mathsf{D}'}{\mathbb{P}[\mathsf{T}\neq 0]\mathsf{D}'}\right] = \beta' \mathcal{V}_{\mathsf{T}'}(\alpha)$$

$$\mathcal{L}_{\infty} = \begin{cases} -\mathbb{E} \left[\log \frac{T'D'}{e} \right] & \beta & \forall \Gamma'(\Omega) \\ -\mathbb{E} \left[\log \frac{T'D'}{e} \right] & \beta' = 1 \\ -\mathbb{E} \left[\log \frac{\Gamma_{\infty}T'}{e} \right] - \frac{1}{\beta'} \mathcal{V}_{\mathsf{D}'} \left(\frac{\mathbb{P}[\mathsf{T}\neq 0]}{\Gamma_{\infty}} \right) & \beta' < 1 \end{cases}$$
(43)

with α and Γ_{∞} , respectively, solutions to

$$\eta_{\mathsf{T}'}(\alpha) = 1 - \frac{1}{\beta'}, \qquad \eta_{\mathsf{D}'}\left(\frac{\mathbb{P}[\mathsf{T}\neq 0]}{\Gamma_{\infty}}\right) = 1 - \beta'.$$
(44)

and with D' and T' the restrictions of D and T to the events $D \neq 0$ and T $\neq 0$.

Theorem 7 [23] Let \mathbf{H}_w be an $N \times K$ matrix with i.i.d. entries with variance $\frac{1}{N}$ and \mathbf{H}_0 and $N \times K$ deterministic random matrix such that the asymptotic spectra of $\mathbf{M} = \mathbf{H}_0 \mathbf{H}_0^{\dagger}$ converge almost surely to a nonrandom limit. Define \mathbf{H} as the $N \times K$ random matrix

$$\mathbf{H} = \mathbf{H}_w + \sqrt{\mathsf{K}}\mathbf{H}_0,\tag{45}$$

then the ESD of **HH**[†] converges, as $K, N \to \infty$ with $\frac{K}{N} \to \beta$, almost surely to a nonrandom limit whose Stieltjes transform satisfies for each $z \in \mathbb{C}^+$

$$\mathfrak{S}(z) = \mathbb{E}\left[\frac{1}{\frac{\mathsf{KM}}{1+\mathfrak{S}(z)} - z(1+\mathfrak{S}(z)) + (\beta-1)}\right]$$
(46)

with K > 0 a nonrandom positive value and M a random variable whose distribution is the asymptotic spectrum of **M**. S(z) is the only solution in the set $\{S(z) \in \mathbb{C}^+ : zS(z) \in \mathbb{C}^+\}$.

Analytic properties of the limiting ESD's in Theorems 3 and 7, can be found in [24] and [23], among them being the existence of continuous densities. In [25], comparisons between individual eigenvalues of the random matrix STS^{\dagger} as in Theorem 3 and the matrix T are explored, enabling the solution of the detection problem in array signal processing [26], that is, determining the number of sources emitting signals through a noise filled environment, impinging on a collection of sensors.

Acknowledgment

The authors gratefully acknowledge helpful suggestions by Dr. A. Lozano and Prof. S. Verdu.

References

- S. Verdú, Multiuser Detection, Cambridge University Press, Cambridge, UK, 1998.
- [2] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. on Information Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.
- [3] D. Tse and S. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans.* on Information Theory, vol. 45, no. 2, pp. 641–657, Mar. 1999.
- [4] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," Foundations and Trends in Communications and Information Theory, vol. 1, no. 1, 2004.
- [5] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [6] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effects on the capacity of multi-element antenna systems," *IEEE Trans. on Communications*, vol. 48, no. 3, pp. 502–511, Mar. 2000.
- [7] D. Chizhik, F. R. Farrokhi, J. Ling, and A. Lozano, "Effect of antenna separation on the capacity of BLAST in correlated channels," *IEEE Communications Letters*, vol. 4, no. 11, pp. 337–339, Nov. 2000.
- [8] K. I. Pedersen, J. B. Andersen, J. P. Kermoal, and P. E. Mogensen, "A stochastic multiple-input multiple-output radio channel model for evaluations of space-time coding algorithms," *Proc. IEEE Vehicular Technology Conf. (VTC'2000 Fall)*, pp. 893–897, Sep. 2000.
- [9] P. Driessen and G. J. Foschini, "On the capacity formula for multiple-input multiple-output channels: A geometric interpretation," *IEEE Trans. Communications*, vol. 47, no. 2, pp. 173–176, Feb. 1999.
- [10] F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-optimal space-time processing with multiple transmit and receive antennas," *IEEE Communications Letters*, vol. 5, no. 3, pp. 85–87, Mar. 2001.
- [11] P. Soma, D. S. Baum, V. Erceg, R. Krishnamoorthy, and A. Paulraj, "Analysis and modelling of multiple-input multipleoutput (MIMO) radio channel based on outdoor measurements conducted at 2.5 GHz for fixed BWA applications," *Proc. IEEE Int. Conf. on Communications (ICC'02), New York City, NY*, pp. 272– 276, 28 Apr.-2 May 2002.
- [12] S. Rice, "Mathematical analysis of random noise," Bell System Technical Journal, vol. 23, 1944.
- [13] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of correlation on the capacity of multi-antenna channels," *BIEEE Trans. on Information Theory*), vol. 51, no. 7, pp. 2491–2509, July 2005.
 [14] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels:
- [14] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans.* on *Information Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [15] Z. D. Bai, "Convergence rate of expected spectral distributions of large random matrices. Part I: Wigner matrices," Annals of Probability, vol. 21, no. 2, pp. 625–648, 1993.
- [16] F. Hiai and D. Petz, "Asymptotic freeness almost everywhere for random matrices," Acta Sci. Math. Szeged, vol. 66, pp. 801–826, 2000.
- [17] T. J. Stieltjes, "Recherches sur les fractions continues," Annales de la Faculte des Sciences de Toulouse, vol. 8 (9), no. A (J), pp. 1–47 (1–122), 1894 (1895).
- [18] V. A. Marčenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math. USSR-Sbornik*, vol. 1, pp. 457–483, 1967.
- [19] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," J. of Multivariate Analysis, vol. 54, pp. 175–192, 1995.
- [20] J. W. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *J. of Multivariate Analysis*, vol. 55, pp. 331–339, 1995.
 [21] J. W. Silverstein, "The limiting spectral distribution on a gener-
- [21] J. W. Silverstein, "The limiting spectral distribution on a generalized class of large dimensional sample covariance matrices," in preparation, 2006.

- [22] A. Lozano, A. M. Tulino, and S. Verdú, "High-SNR power offset in multi-antenna communication," in *Proc. IEEE Int. Symp. on Information Theory (ISIT'04)*, Chicago, IL, June 2004.
 [23] R.B. Dozier and J.W. Silverstein, "On the empirical distribution
- [23] R.B. Dozier and J.W. Silverstein, "On the empirical distribution of eigenvalues of large dimensional information-plus-noise type matrices," submitted 2005.
- [24] J. Silverstein and S. Choi, "Analysis of the limiting spectral distribution of large dimensional random matrices," J. of Multivariate Analysis, vol. 54(2), pp. 295–309, 1995.
- [25] Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *Annals of Probability*, vol. 26, pp. 316–345, 1998.
- [26] J. W. Silverstein and P. L. Combettes, "Signal detection via spectral theory of large dimensional random matrices," *IEEE Trans. on Signal Processing*, vol. 40, pp. 2100–2105, Aug. 1992.