Foundations and Trends™ in Communications and Information Theory

Volume 1 Issue 1, 2004

Editorial Scope

Foundations and Trends™ in Communications and Information Theory will publish survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and image compression
- Wireless communications

Information for Librarians

Foundations and Trends™ in Communications and Information Theory, 2004, Volume 1, 4 issues. ISSN paper version 1567-2190 (USD 200 N. America; EUR 200 Outside N. America). ISSN online version 1567-2328 (EUR 250 N. America; EUR 250 Outside N. America). Also available as a combined paper and online subscription (USD N. America; EUR 300 Outside N. America).

Random Matrix Theory and Wireless Communications

Antonia M. Tulino

Dept. Ingegneria Elettronica e delle Telecomunicazioni
Universitá degli Studi di Napoli "Federico II"
Naples 80125, Italy

atulino@ee.princeton.edu

Sergio Verdú

Dept. Electrical Engineering
Princeton University
Princeton, New Jersey 08544, USA

verdu@princeton.edu

now

the essence of knowledge

Foundations and Trends™ in Communications and Information Theory

Published, sold and distributed by:
PO Box 179
2600 AD Delft
The Netherlands
Tel: +31-6-51115274
www.nowpublishers.com
sales@nowpublishers.com

in North America:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510

Printed on acid-free paper

ISSNs: Paper version 1567-2190; Electronic version 1567-2328

Printed in Great Britain by Antony Rowe Limited.

Foundations and Trends™ in
Communications and Information Theory
Vol 1, No 1 (2004) 1-182

Random Matrix Theory and Wireless Communications

Antonia M. Tulino[1], Sergio Verdú[2]

[1] Dept. Ingegneria Elettronica e delle Telecomunicazion, i Universita degli Studi di Napoli "Federico II", Naples 80125, Italy
[2] Dept. Electrical Engineering, Princeton University, Princeton, New Jersey 08544, USA

Abstract
Random matrix theory has found many applications in physics, statistics and engineering since its inception. Although early developments were motivated by practical experimental problems, random matrices are now used in fields as diverse as Riemann hypothesis, stochastic differential equations, condensed matter physics, statistical physics, chaotic systems, numerical linear algebra, neural networks, multivariate statistics, information theory, signal processing and small-world networks. This article provides a tutorial on random matrices which provides an overview of the theory and brings together in one source the most significant results recently obtained. Furthermore, the application of random matrix theory to the fundamental limits of wireless communication channels is described in depth.

Table of Contents

1

Introduction

From its inception, random matrix theory has been heavily influenced by its applications in physics, statistics and engineering. The landmark contributions to the theory of random matrices of Wishart (1928) [311], Wigner (1955) [303], and Marčenko and Pastur (1967) [170] were motivated to a large extent by practical experimental problems. Nowadays, random matrices find applications in fields as diverse as the Riemann hypothesis, stochastic differential equations, condensed matter physics, statistical physics, chaotic systems, numerical linear algebra, neural networks, multivariate statistics, information theory, signal processing, and small-world networks. Despite the widespread applicability of the tools and results in random matrix theory, there is no tutorial reference that gives an accessible overview of the classical theory as well as the recent results, many of which have been obtained under the umbrella of free probability theory.

In the last few years, a considerable body of work has emerged in the communications and information theory literature on the fundamental limits of communication channels that makes substantial use of results in random matrix theory.

The purpose of this monograph is to give a tutorial overview of random matrix theory with particular emphasis on asymptotic theorems on the distribution of eigenvalues and singular values under various assumptions on the joint distribution of the random matrix entries. While results for matrices with fixed dimensions are often cumbersome and offer limited insight, as the matrices grow large with a given aspect ratio (number of columns to number of rows), a number of powerful and appealing theorems ensure convergence of the empirical eigenvalue distributions to deterministic functions.

The organization of this monograph is the following. Section 1.1 introduces the general class of vector channels of interest in wireless communications. These channels are characterized by random matrices that admit various statistical descriptions depending on the actual application. Section 1.2 motivates interest in large random matrix theory by focusing on two performance measures of engineering interest: Shannon capacity and linear minimum mean-square error, which are determined by the distribution of the singular values of the channel matrix. The power of random matrix results in the derivation of asymptotic closed-form expressions is illustrated for channels whose matrices have the simplest statistical structure: independent identically distributed (i.i.d.) entries. Section 1.3 gives a brief historical tour of the main results in random matrix theory, from the work of Wishart on Gaussian matrices with fixed dimension, to the recent results on asymptotic spectra. Section 2 gives a tutorial account of random matrix theory. Section 2.1 focuses on the major types of random matrices considered in the literature, as well on the main fixed-dimension theorems. Section 2.2 gives an account of the Stieltjes, η, Shannon, Mellin, R- and S-transforms. These transforms play key roles in describing the spectra of random matrices. Motivated by the intuition drawn from various applications in communications, the η and Shannon transforms turn out to be quite helpful at clarifying the exposition as well as the statement of many results. Considerable emphasis is placed on examples and closed-form expressions. Section 2.3 uses the transforms defined in Section 2.2 to state the main asymptotic distribution theorems. Section 2.4 presents an overview of the application of Voiculescu's free probability theory to random matrices. Recent results on the speed of convergence to the asymptotic limits are reviewed in Section 2.5. Section 3 applies the results in Section 2 to the fundamental limits of wireless communication channels described by random matrices. Section 3.1 deals with direct-sequence code-division multiple-access (DS-CDMA), with and without fading (both frequency-flat and frequency-selective) and with single and multiple receive antennas. Section 3.2 deals with multi-carrier code-division multiple access (MC-CDMA), which is the time-frequency dual of the model considered in Section 3.1. Channels with multiple receive and transmit antennas are reviewed in Section 3.3 using models that incorporate nonideal effects such as antenna correlation, polarization, and line-of-sight components.

1.1 Wireless Channels

The last decade has witnessed a renaissance in the information theory of wireless communication channels. Two prime reasons for the strong level of activity in this field can be identified. The first is the growing importance of the efficient use of bandwidth and power in view of the ever-increasing demand for wireless services. The second is the fact that some of the main challenges in the study of the capacity of wireless channels have only been successfully tackled recently. Fading, wideband, multiuser and multi-antenna are some of the key features that characterize wireless channels of contemporary interest. Most of the information theoretic literature that studies the effect of those features on channel capacity deals with linear vector memoryless channels of the form

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1.1}$$

where $\mathbf{x}$ is the K-dimensional input vector, $\mathbf{y}$ is the N-dimensional output vector, and the N-dimensional vector $\mathbf{n}$ models the additive circularly symmetric Gaussian noise. All these quantities are, in general, complex-valued. In addition to input constraints, and the degree of knowledge of the channel at receiver and transmitter, (1.1) is characterized by the distribution of the $N \times K$ random *channel matrix* $\mathbf{H}$ whose entries are also complex-valued.

The nature of the K and N dimensions depends on the actual application. For example, in the single-user narrowband channel with n_{T}

and n_{R} antennas at transmitter and receiver, respectively, we identify $K = n_{\mathrm{T}}$ and $N = n_{\mathrm{R}}$; in the DS-CDMA channel, K is the number of users and N is the spreading gain.

In the multi-antenna case, $\mathbf{H}$ models the propagation coefficients between each pair of transmit-receive antennas. In the synchronous DS-CDMA channel, in contrast, the entries of $\mathbf{H}$ depend on the received signature vectors (usually pseudo-noise sequences) and the fading coefficients seen by each user. For a channel with J users each transmitting with n_{T} antennas using spread-spectrum with spreading gain G and a receiver with n_{R} antennas, $K = n_{\mathrm{T}}J$ and $N = n_{\mathrm{R}}G$.

Naturally, the simplest example is the one where $\mathbf{H}$ has i.i.d. entries, which constitutes the canonical model for the single-user narrowband multi-antenna channel. The same model applies to the randomly spread DS-CDMA channel not subject to fading. However, as we will see, in many cases of interest in wireless communications the entries of $\mathbf{H}$ are not i.i.d.

1.2 The Role of the Singular Values

Assuming that the channel matrix $\mathbf{H}$ is completely known at the receiver, the capacity of (1.1) under input power constraints depends on the distribution of the singular values of $\mathbf{H}$. We focus in the simplest setting to illustrate this point as crisply as possible: suppose that the entries of the input vector $\mathbf{x}$ are i.i.d. For example, this is the case in a synchronous DS-CDMA multiaccess channel or for a single-user multi-antenna channel where the transmitter cannot track the channel.

The empirical cumulative distribution function of the eigenvalues (also referred to as the *spectrum* or empirical distribution) of an $n \times n$ Hermitian matrix $\mathbf{A}$ is denoted by $\mathsf{F}^n_{\mathbf{A}}$ defined as[1]

$$\mathsf{F}^n_{\mathbf{A}}(x) = \frac{1}{n}\sum_{i=1}^{n} 1\{\lambda_i(\mathbf{A}) \le x\}, \tag{1.2}$$

where $\lambda_1(\mathbf{A}), \ldots, \lambda_n(\mathbf{A})$ are the eigenvalues of $\mathbf{A}$ and $1\{\cdot\}$ is the indicator function.

[1] If $\mathsf{F}^n_{\mathbf{A}}$ converges as $n \to \infty$, then the corresponding limit (asymptotic empirical distribution or asymptotic spectrum) is simply denoted by $\mathsf{F}_{\mathbf{A}}(x)$.

Now, consider an arbitrary $N \times K$ matrix $\mathbf{H}$. Since the nonzero singular values of $\mathbf{H}$ and $\mathbf{H}^\dagger$ are identical, we can write

$$N\mathsf{F}^N_{\mathbf{HH}^\dagger}(x) - Nu(x) = K\mathsf{F}^K_{\mathbf{H}^\dagger\mathbf{H}}(x) - Ku(x) \tag{1.3}$$

where $u(x)$ is the unit-step function ($u(x) = 0$, $x \le 0$; $u(x) = 1$, $x > 0$).

With an i.i.d. Gaussian input, the normalized input-output mutual information of (1.1) conditioned on $\mathbf{H}$ is[2]

$$\begin{aligned} \frac{1}{N} I(\mathbf{x};\mathbf{y}|\mathbf{H}) &= \frac{1}{N} \log\det\left(\mathbf{I} + \mathsf{SNR}\,\mathbf{HH}^\dagger\right) && (1.4) \\ &= \frac{1}{N}\sum_{i=1}^{N} \log\left(1 + \mathsf{SNR}\,\lambda_i(\mathbf{HH}^\dagger)\right) \\ &= \int_0^\infty \log\left(1 + \mathsf{SNR}\,x\right) d\mathsf{F}^N_{\mathbf{HH}^\dagger}(x) && (1.5) \end{aligned}$$

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

$$\mathsf{SNR} = \frac{N\mathbb{E}[||\mathbf{x}||^2]}{K\mathbb{E}[||\mathbf{n}||^2]}, \tag{1.6}$$

and with $\lambda_i(\mathbf{HH}^\dagger)$ equal to the ith squared singular value of $\mathbf{H}$.

If the channel is known at the receiver and its variation over time is stationary and ergodic, then the expectation of (1.4) over the distribution of $\mathbf{H}$ is the channel capacity (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 (1.4) determines the outage capacity (e.g. [22]).

Another important performance measure for (1.1) is the minimum mean-square-error (MMSE) achieved by a linear receiver, which determines the maximum achievable output signal-to-interference-and-noise

[2] The celebrated log-det formula has a long history: In 1964, Pinsker [204] gave a general log-det formula for the mutual information between jointly Gaussian random vectors but did not particularize it to the linear model (1.1). Verdú [270] in 1986 gave the explicit form (1.4) as the capacity of the synchronous DS-CDMA channel as a function of the signature vectors. The 1991 textbook by Cover and Thomas [47] gives the log-det formula for the capacity of the power constrained vector Gaussian channel with arbitrary noise covariance matrix. In the mid 1990s, Foschini [77] and Telatar [250] gave (1.4) for the multi-antenna channel with i.i.d. Gaussian entries. Even prior to those works, the conventional analyses of Gaussian channels with memory via vector channels (e.g. [260, 31]) used the fact that the capacity can be expressed as the sum of the capacities of independent channels whose signal-to-noise ratios are governed by the singular values of the channel matrix.

ratio (SINR). For an i.i.d. input, the arithmetic mean over the users (or transmit antennas) of the MMSE is given, as function of $\mathbf{H}$, by [271]

$$\frac{1}{K}\min_{\mathbf{M}\in C^{K\times N}} \mathbb{E}\left[||\mathbf{x}-\mathbf{M}\mathbf{y}||^2\right] = \frac{1}{K}\mathrm{tr}\left\{\left(\mathbf{I}+\mathsf{SNR}\,\mathbf{H}^\dagger\mathbf{H}\right)^{-1}\right\} \tag{1.7}$$

$$= \frac{1}{K}\sum_{i=1}^{K}\frac{1}{1+\mathsf{SNR}\,\lambda_i(\mathbf{H}^\dagger\mathbf{H})} \tag{1.8}$$

$$= \int_0^\infty \frac{1}{1+\mathsf{SNR}\,x}d\mathsf{F}^K_{\mathbf{H}^\dagger\mathbf{H}}(x)$$

$$= \frac{N}{K}\int_0^\infty \frac{1}{1+\mathsf{SNR}\,x}d\mathsf{F}^N_{\mathbf{H}\mathbf{H}^\dagger}(x) - \frac{N-K}{K} \tag{1.9}$$

where the expectation in (1.7) is over $\mathbf{x}$ and $\mathbf{n}$ while (1.9) follows from (1.3). Note, incidentally, that both performance measures as a function of SNR are coupled through

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

As we see in (1.5) and (1.9), both fundamental performance measures (capacity and MMSE) are dictated by the distribution of the empirical (squared) singular value distribution of the random channel matrix. In the simplest case of $\mathbf{H}$ having i.i.d. Gaussian entries, the density function corresponding to the expected value of $\mathsf{F}^N_{\mathbf{H}\mathbf{H}^\dagger}$ can be expressed explicitly in terms of the Laguerre polynomials. Although the integrals in (1.5) and (1.9) with respect to such a probability density function (p.d.f.) lead to explicit solutions, limited insight can be drawn from either the solutions or their numerical evaluation. Fortunately, much deeper insights can be obtained using the tools provided by *asymptotic* random matrix theory. Indeed, a rich body of results exists analyzing the asymptotic spectrum of $\mathbf{H}$ as the number of columns and rows goes to infinity while the aspect ratio of the matrix is kept constant.

Before introducing the asymptotic spectrum results, some justification for their relevance to wireless communication problems is in order. In CDMA, channels with K and N between 32 and 64 would be fairly typical. In multi-antenna systems, arrays of 8 to 16 antennas would be at the forefront of what is envisioned to be feasible in the foreseeable future. Surprisingly, even quite smaller system sizes are large enough for the asymptotic limit to be an excellent approximation. Furthermore, not only do the averages of (1.4) and (1.9) converge to their limits surprisingly fast, but the randomness in those functionals due to the random outcome of $\mathbf{H}$ disappears extremely quickly. Naturally, such robustness has welcome consequences for the operational significance of the resulting formulas.

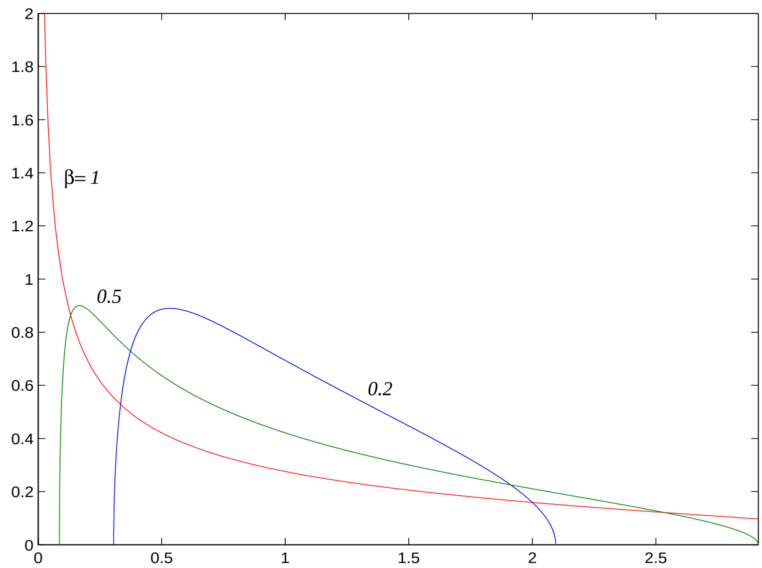

Fig. 1.1 The Marčenko-Pastur density function (1.10) for $\beta = 1, 0.5, 0.2$.

As we will see in Section 2, a central result in random matrix theory states that when the entries of $\mathbf{H}$ are zero-mean i.i.d. with variance $\frac{1}{N}$, the empirical distribution of the eigenvalues of $\mathbf{H}^\dagger\mathbf{H}$ converges almost surely, as $K, N \to \infty$ with $\frac{K}{N} \to \beta$, to the so-called Marčenko-Pastur law whose density function is

$$\mathsf{f}_\beta(x) = \left(1 - \frac{1}{\beta}\right)^+ \delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi\beta x} \tag{1.10}$$

where $(z)^+ = \max(0, z)$ and

$$a = (1 - \sqrt{\beta})^2 \qquad\qquad b = (1 + \sqrt{\beta})^2. \tag{1.11}$$

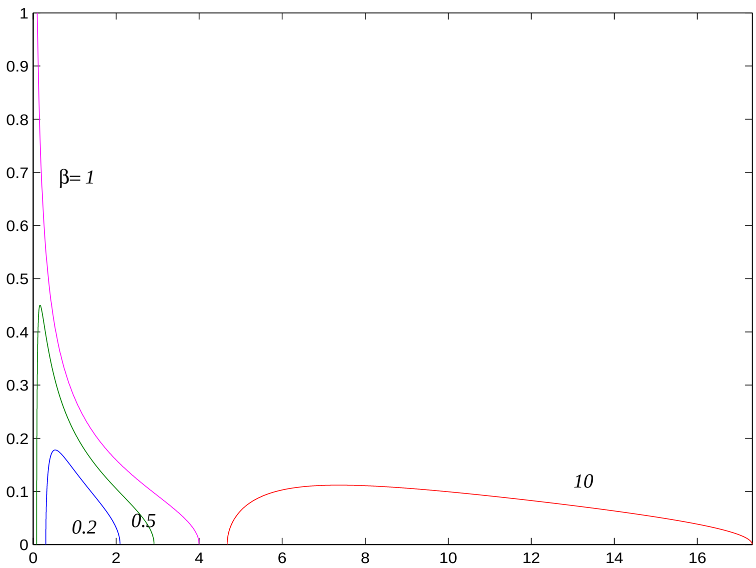

Fig. 1.2 The Marčenko-Pastur density function (1.12) for $\beta = 10, 1, 0.5, 0.2$. Note that the mass points at 0, present in some of them, are not shown.

Analogously, the empirical distribution of the eigenvalues of $\mathbf{HH}^\dagger$ converges almost surely to a nonrandom limit whose density function is (cf. Fig. 1.2)

$$\begin{aligned} \tilde{\mathsf{f}}_\beta(x) &= (1-\beta)\,\delta(x) + \beta\,\mathsf{f}_\beta(x) \\ &= (1-\beta)^+\,\delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi x}. \end{aligned} \tag{1.12}$$

Using the asymptotic spectrum, the following closed-form expressions for the limits of (1.4) [275] and (1.7) [271] can be obtained:

$$\tag{1.13}$$

$$\begin{aligned} \frac{1}{N}\log\det\left(\mathbf{I} + \mathsf{SNR}\,\mathbf{HH}^\dagger\right) &\rightarrow \beta\int_a^b \log(1+\mathsf{SNR}\;x)\mathsf{f}_\beta(x)\;dx \\ &= \beta\log\left(1+\mathsf{SNR} - \frac{1}{4}\mathcal{F}\left(\mathsf{SNR}\,,\beta\right)\right) \\ &+ \log\left(1+\mathsf{SNR}\,\beta - \frac{1}{4}\mathcal{F}\left(\mathsf{SNR}\,,\beta\right)\right) \\ &- \frac{\log e}{4\,\mathsf{SNR}}\mathcal{F}\left(\mathsf{SNR}\,,\beta\right) \end{aligned} \tag{1.14}$$

$$\frac{1}{K}\mathrm{tr}\left\{\left(\mathbf{I}+\mathsf{SNR}\,\mathbf{H}^{\dagger}\mathbf{H}\right)^{-1}\right\} \rightarrow \int_a^b \frac{1}{1+\mathsf{SNR}\,x}\mathsf{f}_{\beta}(x)\,dx \quad (1.15)$$

$$= 1-\frac{\mathcal{F}(\mathsf{SNR},\beta)}{4\,\beta\,\mathsf{SNR}} \quad (1.16)$$

with

$$\mathcal{F}(x,z)=\left(\sqrt{x(1+\sqrt{z})^2+1}-\sqrt{x(1-\sqrt{z})^2+1}\right)^2. \quad (1.17)$$

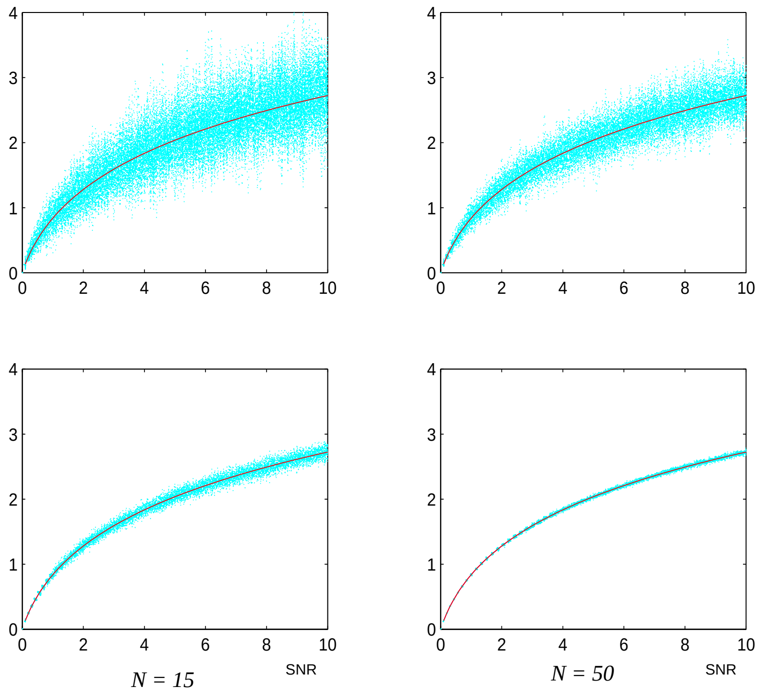

Fig. 1.3 Several realizations of the left-hand side of (1.13) are compared to the asymptotic limit in the right-hand side of (1.13) in the case of $\beta = 1$ for sizes: $N = 3, 5, 15, 50$.

The convergence of the singular values of $\mathbf{H}$ exhibits several key features with engineering significance:

- *Insensitivity* of the asymptotic eigenvalue distribution to the shape of the p.d.f. of the random matrix entries. This property implies, for example, that in the case of a single-user multi-antenna link, the results obtained asymptotically hold for any type of fading statistics. It also implies that restricting the CDMA waveforms to be binary-valued incurs no loss in capacity asymptotically.[3]

- *Ergodic* behavior: it suffices to observe a single matrix realization in order to obtain convergence to a deterministic limit. In other words, the eigenvalue histogram of any matrix realization converges almost surely to the average asymptotic eigenvalue distribution. This "hardening" of the singular values lends operational significance to the capacity formulas even in cases where the random channel parameters do not vary ergodically within the span of a codeword.
- *Fast convergence* of the empirical singular-value distribution to its asymptotic limit. Asymptotic analysis is especially useful when the convergence is so fast that, even for small values of the parameters, the asymptotic results come close to the finite-size results (cf. Fig. 1.3). Recent works have shown that the convergence rate is of the order of the reciprocal of the number of entries in the random matrix [8, 110].

It is crucial for the explicit expressions of asymptotic capacity and MMSE shown in (1.14) and (1.16), respectively, that the channel matrix entries be i.i.d. Outside that model, explicit expressions for the asymptotic singular value distribution such as (1.10) are exceedingly rare. Fortunately, in other random matrix models, the asymptotic singular value distribution can indeed be characterized, albeit not in explicit form, in ways that enable the analysis of capacity and MMSE through the numerical solution of nonlinear equations.

The first applications of random matrix theory to wireless communications were the works of Foschini [77] and Telatar [250] on narrowband multi-antenna capacity; Verdú [271] and Tse-Hanly [256] on the optimum SINR achievable by linear multiuser detectors for CDMA;

[3] The spacing between consecutive eigenvalues, when properly normalized, was conjectured in [65, 66] to converge in distribution to a limit that does not depend on the shape of the p.d.f. of the entries. The universality of the level spacing distribution and other *microscopic (local)* spectral characteristics has been extensively discussed in recent theoretical physics and mathematical literature [174, 106, 200, 52, 54].

Verdú [271] on optimum near-far resistance; Grant-Alexander [100], Verdú-Shamai [275, 217], Rapajic-Popescu [206], and Müller [185] on the capacity of CDMA. Subsequently, a number of works, surveyed in Section 3, have successfully applied random matrix theory to a variety of problems in the design and analysis of wireless communication systems.

Not every result of interest in the asymptotic analysis of channels of the form (1.1) has made use of the asymptotic eigenvalue tools that are of central interest in this paper. For example, the analysis of single-user matched filter receivers [275] and the analysis of the optimum asymptotic multiuser efficiency [258] have used various versions of the central-limit theorem; the analysis of the asymptotic uncoded error probability as well as the rates achievable with suboptimal constellations have used tools from statistical physics such as the replica method [249, 103].

1.3 Random Matrices: A Brief Historical Account

In this subsection, we provide a brief introduction to the main developments in the theory of random matrices. A more detailed account of the theory itself, with particular emphasis on the results that are relevant for wireless communications, is given in Section 2.

Random matrices have been a part of advanced multivariate statistical analysis since the end of the 1920s with the work of Wishart [311] on fixed-size matrices with Gaussian entries. The first asymptotic results on the limiting spectrum of large random matrices were obtained by Wigner in the 1950s in a series of papers [303, 305, 306] motivated by nuclear physics. Replacing the self-adjoint Hamiltonian operator in an infinite-dimensional Hilbert space by an ensemble of very large Hermitian matrices, Wigner was able to bypass the Schrödinger equation and explain the statistics of experimentally measured atomic energy levels in terms of the limiting spectrum of those random matrices. Since then, research on the limiting spectral analysis of large-dimensional random matrices has continued to attract interest in probability, statistics and physics.

Wigner [303] initially dealt with an $n \times n$ symmetric matrix $\mathbf{A}$ whose diagonal entries are 0 and whose upper-triangle entries are independent

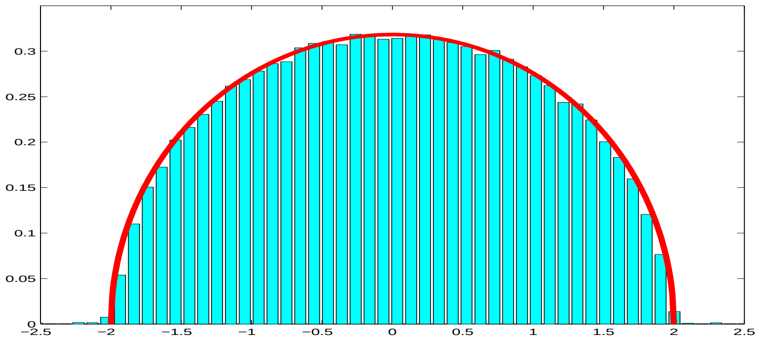

Fig. 1.4 The semicircle law density function (1.18) compared with the histogram of the average of 100 empirical density functions for a Wigner matrix of size $n = 100$.

and take the values ± 1 with equal probability. Through a combinatorial derivation of the asymptotic eigenvalue moments involving the Catalan numbers, Wigner showed that, as $n \to \infty$, the averaged empirical distribution of the eigenvalues of $\frac{1}{\sqrt{n}}\mathbf{A}$ converges to the *semicircle law* whose density is

$$w(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4 - x^2} & \text{if} \quad |x| \leq 2 \\ 0 & \text{if} \quad |x| > 2 \end{cases} \tag{1.18}$$

Later, Wigner [305] realized that the same result would be obtained if the random selection was sampled from a zero-mean (real or complex) Gaussian distribution. In that case, it is even possible to find an exact formula for the joint distribution of the eigenvalues as a function of n [176]. The matrices treated in [303] and [305] are special cases of Wigner matrices, defined as Hermitian matrices whose upper-triangle entries are zero-mean and independent. In [306], Wigner showed that the asymptotic distribution of any Wigner matrix is the semicircle law (1.18) even if only a unit second-moment condition is placed on its entries.

Figure 1.4 compares the semicircle law density function (1.18) with the average of 100 empirical density functions of the eigenvalues of a 10×10 Wigner matrix whose diagonal entries are 0 and whose upper-triangle entries are independent and take the values ± 1 with equal probability.

If no attempt is made to symmetrize the square matrix $\mathbf{A}$ and all its entries are chosen to be i.i.d., then the eigenvalues of $\frac{1}{\sqrt{n}}\mathbf{A}$ are asymptotically uniformly distributed on the unit circle of the complex plane. This is commonly referred to as Girko's *full-circle law*, which is exemplified in Figure 1.5. It has been proved in various degrees of rigor and generality in [173, 197, 85, 68, 9]. If the off-diagonal entries $A_{i,j}$ and $A_{j,i}$ are Gaussian and pairwise correlated with correlation coefficient ρ, then [238] shows that the eigenvalues of $\frac{1}{\sqrt{n}}\mathbf{A}$ are asymptotically uniformly distributed on an ellipse in the complex plane whose axes coincide with the real and imaginary axes and have radius $1+\rho$ and $1-\rho$, respectively. When $\rho = 1$, the projection on the real axis of such *elliptic law* is equal to the semicircle law.

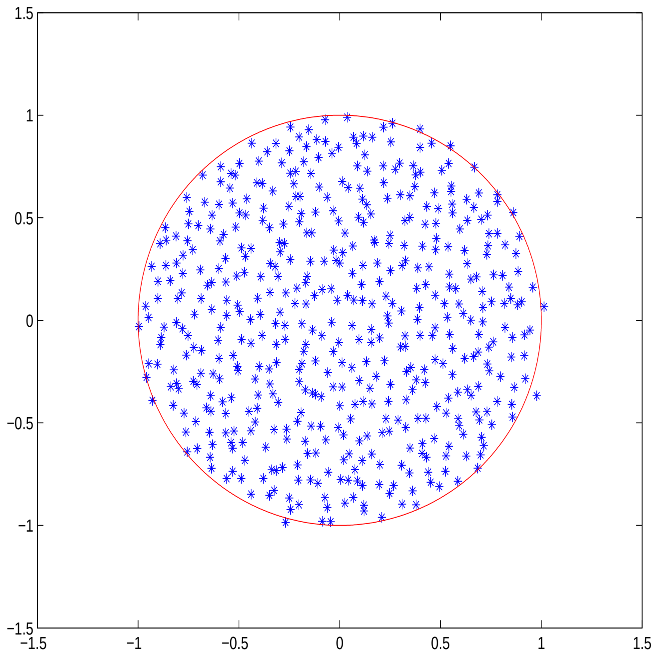

Fig. 1.5 The full-circle law and the eigenvalues of a realization of a matrix of size $n = 500$.

Most of the results surveyed above pertain to the eigenvalues of square matrices with independent entries. However, as we saw in Section 1.2, key problems in wireless communications involve the singular values of rectangular matrices $\mathbf{H}$; even if those matrices have independent entries, the matrices $\mathbf{H}\mathbf{H}^\dagger$ whose eigenvalues are of interest do not have independent entries.

When the entries of $\mathbf{H}$ are zero-mean i.i.d. Gaussian, $\mathbf{H}\mathbf{H}^\dagger$ is commonly referred to as a Wishart matrix. The analysis of the joint distribution of the entries of Wishart matrices is as old as random matrix theory itself [311]. The joint distribution of the eigenvalues of such matrices is known as the Fisher-Hsu-Roy distribution and was discovered simultaneously and independently by Fisher [75], Hsu [120], Girshick [89] and Roy [210]. The corresponding marginal distributions can be expressed in terms of the Laguerre polynomials [125].

The asymptotic theory of singular values of rectangular matrices has concentrated on the case where the matrix aspect ratio converges to a constant

$$\frac{K}{N} \to \beta \tag{1.19}$$

as the size of the matrix grows.

The first success in the quest for the limiting empirical singular value distribution of rectangular random matrices is due to Marčenko and Pastur [170] in 1967. This landmark paper considers matrices of the form

$$\mathbf{W} = \mathbf{W}_0 + \mathbf{H}\mathbf{T}\mathbf{H}^\dagger \tag{1.20}$$

where $\mathbf{T}$ is a real diagonal matrix independent of $\mathbf{H}$, $\mathbf{W}_0$ is a deterministic Hermitian matrix, and the columns of the $N \times K$ matrix $\mathbf{H}$ are i.i.d. random vectors whose distribution satisfies a certain symmetry condition (encompassing the cases of independent entries and uniform distribution on the unit sphere). In the special case where $\mathbf{W}_0 = \mathbf{0}$, $\mathbf{T} = \mathbf{I}$, and $\mathbf{H}$ has i.i.d. entries with variance $\frac{1}{N}$, the limiting spectrum of $\mathbf{W}$ found in [170] is the density in (1.10). In the special case of square $\mathbf{H}$, the asymptotic density function of the singular values, corresponding to the square root of the random variable whose p.d.f. is (1.10) with $\beta = 1$, is equal to the *quarter circle* law:

$$q(x) = \frac{1}{\pi}\sqrt{4 - x^2}, \quad 0 \le x \le 2. \tag{1.21}$$

As we will see in Section 2, in general ($\mathbf{W}_0 \neq \mathbf{0}$ or $\mathbf{T} \neq \mathbf{I}$) no closed-form expression is known for the limiting spectrum. Rather, [170] characterized it indirectly through its Stieltjes transform,[4] which uniquely determines the distribution function. Since [170], this transform, which can be viewed as an iterated Laplace transform, has played a fundamental role in the theory of random matrices.

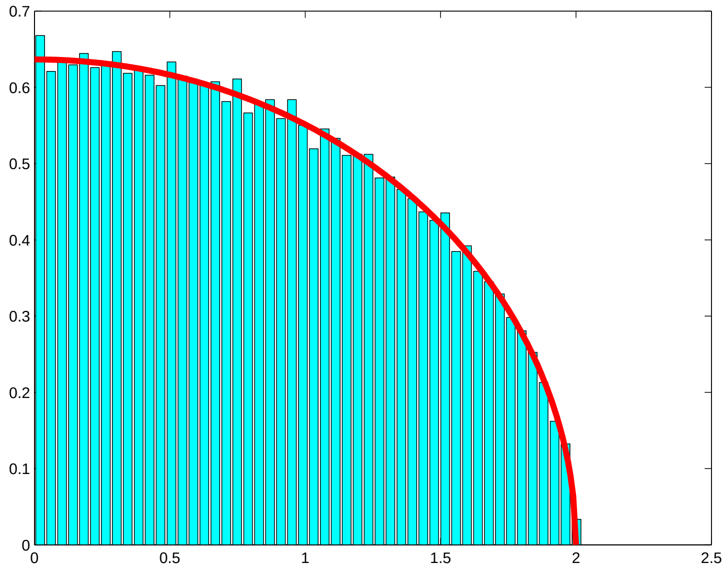

Fig. 1.6 The quarter circle law compared a histogram of the average of 100 empirical singular value density functions of a matrix of size 100×100.

Figure 1.6 compares the quarter circle law density function (1.21) with the average of 100 empirical singular value density functions of a 100×100 square matrix $\mathbf{H}$ with independent zero-mean complex Gaussian entries with variance $\frac{1}{100}$.

Despite the ground-breaking nature of Marčenko and Pastur's contribution, it remained in obscurity for quite some time. For example, in 1977 Grenander and Silverstein [101] rediscovered (1.10) motivated by a neural network problem where the entries of $\mathbf{H}$ take only two values. Also unaware of the in-probability convergence result of [170], in 1978 Wachter [296] arrived at the same solution but in the stronger sense of almost sure convergence under the condition that the entries of $\mathbf{H}$ have

[4] The Stieltjes transform is defined in Section 2.2.1. The Dutch mathematician T. J. Stieltjes (1856-1894) provided the first inversion formula for this transform in [246].

uniformly bounded central moments of order higher than 2 as well as the same means and variances within a row. The almost sure convergence for the model (1.20) considered in [170] was shown in [227]. Even as late as 1991, rediscoveries of the Marčenko-Pastur law can be found in the Physics literature [50].

The case where $\mathbf{W} = \mathbf{0}$ in (1.20), $\mathbf{T}$ is not necessarily diagonal but Hermitian and $\mathbf{H}$ has i.i.d. entries was solved by Silverstein [226] also in terms of the Stieltjes transform.

The special case of (1.20) where $\mathbf{W_0} = \mathbf{0}$, $\mathbf{H}$ has zero-mean i.i.d. Gaussian entries and

$$\mathbf{T} = (\mathbf{Y}\mathbf{Y}^{\dagger})^{-1}$$

where the $K \times m$ matrix $\mathbf{Y}$ has also zero-mean i.i.d. Gaussian entries with variance $\frac{1}{m}$, independent of $\mathbf{H}$, is called a (central) multivariate F-matrix. Because of the statistical applications of such matrix, its asymptotic spectrum has received considerable attention culminating in the explicit expression found by Silverstein [223] in 1985.

The speed of convergence to the limiting spectrum is studied in [8]. For our applications it is more important, however, to assess the speed of convergence of the performance measures (e.g. capacity and MMSE) to their asymptotic limits. Note that the sums in the right side of (1.4) involve dependent terms. Thanks to that dependence, the convergence in (1.13) and (1.15) is quite remarkable: the deviations from the respective limits *multiplied by* N converge to Gaussian random variables with fixed mean[5] and variance. This has been established for general continuous functions, not just the logarithmic and rational functions of (1.13) and (1.15), in [15] (see also [131]).

The matrix of eigenvectors of Wishart matrices is known to be uniformly distributed on the manifold of unitary matrices (the so-called *Haar measure*) (e.g. [125, 67]). In the case of $\mathbf{H}\mathbf{H}^{\dagger}$ where $\mathbf{H}$ has i.i.d. non-Gaussian entries, much less success has been reported in the asymptotic characterization of the eigenvectors [153, 224, 225].

For matrices whose entries are Gaussian and correlated according to a Toeplitz structure, an integral equation is known for the Stielt-

[5] The mean is zero in the interesting special case where $\mathbf{H}$ has i.i.d. complex Gaussian entries [15].

jes transform of the asymptotic spectrum as a function of the Fourier transform of the correlation function [147, 198, 55]. Other results on random matrices with correlated and weakly dependent entries can be found in [170, 196, 146, 53, 199, 145]. Reference [191], in turn, considers a special class of random matrices with dependent entries that falls outside the Marčenko-Pastur framework and that arises in the context of the statistical physics of disordered systems.

Incidentally, another application of the Stieltjes transform approach is the generalization of Wigner's semicircle law to the sum of a Wigner matrix and a deterministic Hermitian matrix. Provided Lindeberg-type conditions are satisfied by the entries of the random component, [147] obtained the *deformed semicircle law*, which is only known in closed-form in the Stieltjes transform domain.

Sometimes, an alternative to the characterization of asymptotic spectra through the Stieltjes transform is used, based on the proof of convergence and evaluation of moments such as $\frac{1}{N}\mathrm{tr}\{(\mathbf{HH}^\dagger)^k\}$. For most cases of practical interest, the limiting spectrum has bounded support. Thus, the moment convergence theorem can be applied to obtain results on the limiting spectrum through its moments [297, 314, 315, 313].

An important recent development in asymptotic random matrix analysis has been the realization that the non-commutative *free probability theory* introduced by Voiculescu [283, 285] in the mid-1980s is applicable to random matrices. In free probability, the classical notion of independence of random variables is replaced by that of "freeness" or "free independence".

The power of the concept of free random matrices is best illustrated by the following setting. In general, we cannot find the eigenvalues of the sums of random matrices from the eigenvalues of the individual matrices (unless they have the same eigenvectors), and therefore the asymptotic spectrum of the sum cannot be obtained from the individual asymptotic spectra. An obvious exception is the case of independent diagonal matrices in which case the spectrum of the sum is simply the convolution of the spectra. When the random matrices are asymptotically free [287], the asymptotic spectrum of the sum is also obtainable from the individual asymptotic spectra. Instead of convolution (or equivalently, summing the logarithms of the individual Fourier transforms), the "free convolution" is obtained through the sum of the so-called R-transforms introduced by Voiculescu [285]. Examples of asymptotically free random matrices include independent Gaussian random matrices, and **A** and $\mathbf{UBU}^*$ where **A** and **B** are Hermitian and **U** is uniformly distributed on the manifold of unitary matrices and independent of **A** and **B**.

In free probability, the role of the Gaussian distribution in classical probability is taken by the semicircle law (1.18) in the sense of the free analog of the central limit theorem [284]: the spectrum of the normalized sum of free random matrices (with given spectrum) converges to the semicircle law (1.18). Analogously, the spectrum of the normalized sum of free random matrices with unit rank converges to the Marčenko-Pastur law (1.10), which then emerges as the free counterpart of the Poisson distribution [239, 295]. In the general context of free random variables, Voiculescu has found an elegant definition of free-entropy [288, 289, 291, 292, 293]. A number of structural properties have been shown for free-entropy in the context of non-commutative probability theory (including the counterpart of the entropy-power inequality [248]). The free counterpart to Fisher's information has been investigated in [289]. However, a free counterpart to the divergence between two distributions is yet to be discovered.

A connection between random matrices and information theory was made by Balian [17] in 1968 considering the inverse problem in which the distribution of the entries of the matrix must be determined while being consistent with certain constraints. Taking a maximum entropy method, the ensemble of Gaussian matrices is the solution to the problem where only a constraint on the energy of the singular values is placed.

2

Random Matrix Theory

In this section, we review a wide range of existing mathematical results that are relevant to the analysis of the statistics of random matrices arising in wireless communications. We also include some new results on random matrices that were inspired by problems of engineering interest.

Throughout the monograph, complex Gaussian random variables are always circularly symmetric, i.e., with uncorrelated real and imaginary parts, and complex Gaussian vectors are always proper complex.[1]

2.1 Types of Matrices and Non-Asymptotic Results

We start by providing definitions for the most important classes of random matrices: Gaussian, Wigner, Wishart and Haar matrices. We also collect a number of results that hold for arbitrary (non-asymptotic) matrix sizes.

[1] In the terminology introduced in [188], a random vector with real and imaginary components $\mathbf{x}$ and $\mathbf{y}$, respectively, is *proper complex* if $\mathbb{E}\left[(\mathbf{x}-\mathbb{E}[\mathbf{x}])(\mathbf{y}-\mathbb{E}[\mathbf{y}])^T\right]=\mathbf{0}$.

2.1.1 Gaussian Matrices

Definition 2.1. A *standard real/complex Gaussian* $m \times n$ matrix $\mathbf{H}$ has i.i.d. real/complex zero-mean Gaussian entries with identical variance $\sigma^2 = \frac{1}{m}$. The p.d.f. of a complex Gaussian matrix with i.i.d. zero-mean Gaussian entries with variance σ^2 is

$$(\pi\sigma^2)^{-mn} \exp\left[-\frac{\mathrm{tr}\{\mathbf{H}\mathbf{H}^\dagger\}}{\sigma^2}\right]. \tag{2.1}$$

The following result is the complex counterpart of those given in [18, 78, 27, 245] and [182, Thm. 3.2.14]:

Lemma 2.1. [104] Let $\mathbf{H}$ be an $m \times n$ standard complex Gaussian matrix with $n \geq m$. Denote its QR-decomposition by $\mathbf{H} = \mathbf{Q}\mathbf{R}$. The upper triangular matrix $\mathbf{R}$ is independent of $\mathbf{Q}$, which is uniformly distributed over the manifold[2] of complex $m \times n$ matrices such that $\mathbf{Q}\mathbf{Q}^\dagger = \mathbf{I}$. The entries of $\mathbf{R}$ are independent and its diagonal entries, $\mathsf{R}_{i,i}$ for $i \in \{1, \ldots, m\}$, are such that $2m\mathsf{R}_{i,i}^2$ are χ^2 random variables with $2(n - i + 1)$ degrees of freedom while the off-diagonal entries, $\mathsf{R}_{i,j}$ for $i < j$, are independent zero-mean complex Gaussian with variance $\frac{1}{m}$.

The proof of Lemma 2.1 uses the expression of the p.d.f. of $\mathbf{H}$ given in (2.1) and [67, Theorem 3.1].

The p.d.f. of the eigenvalues of standard Gaussian matrices is studied in [32, 68]. If the $n \times n$ matrix coefficients are real, [69] gives an exact expression for the expected number of real eigenvalues which grows as $\sqrt{2n/\pi}$.

2.1.2 Wigner Matrices

Definition 2.2. An $n \times n$ Hermitian matrix $\mathbf{W}$ is a *Wigner* matrix if its upper-triangular entries are independent zero-mean random variables with identical variance. If the variance is $\frac{1}{n}$, then $\mathbf{W}$ is a standard Wigner matrix.

[2] This is called the Stiefel manifold and it is a subspace of dimension $2mn - m^2$ with total volume $2^m \pi^{mn - \frac{1}{2}m(m-1)} \prod_{i=1}^{m} \frac{1}{(n-i)!}$

Theorem 2.2. Let $\mathbf{W}$ be an $n \times n$ complex Wigner matrix whose (diagonal and upper-triangle) entries are i.i.d. zero-mean Gaussian with unit variance.[3] Then, its p.d.f. is

$$2^{-n/2}\pi^{-n^2/2}\exp\left[-\frac{\mathrm{tr}\{\mathbf{W}^2\}}{2}\right] \tag{2.2}$$

while the joint p.d.f. of its ordered eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n$ is

$$\frac{1}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}\sum_{i=1}^{n}\lambda_i^2} \prod_{i=1}^{n-1}\frac{1}{i!}\prod_{i<j}^{n}(\lambda_i-\lambda_j)^2. \tag{2.3}$$

Theorem 2.3. [307] Let $\mathbf{W}$ be an $n \times n$ complex Gaussian Wigner matrix defined as in Theorem 2.2. The marginal p.d.f. of the unordered eigenvalues is

$$\frac{1}{n}\sum_{i=0}^{n-1}\frac{1}{2^i\, i!\,\sqrt{2\pi}}\left(e^{-\frac{x^2}{4}}H_i(x)\right)^2 \tag{2.4}$$

with $H_i(\cdot)$ the ith Hermite polynomial [1].

As shown in [304, 172, 81, 175], the spacing between adjacent eigenvalues of a Wigner matrix exhibits an interesting behavior. With the eigenvalues of a Gaussian Wigner matrix sorted in ascending order, denote by L the spacing between adjacent eigenvalues relative to the mean eigenvalue spacing. The density of L in the large-dimensional limit is accurately approximated by[4]

$$f_{\mathsf{L}}(s) \approx \frac{\pi}{2}\, s\, e^{-\frac{\pi}{4}s^2} \tag{2.5}$$

For small values of s, (2.5) approaches zero implying that very small spacings are unlikely and that the eigenvalues somehow repel each other.

[3] Such matrices are often referred to as simply Gaussian Wigner matrices.

[4] Wigner postulated (2.5) in [304] by assuming that the energy levels of a nucleus behave like a modified Poisson process. Starting from the joint p.d.f. of the eigenvalues of a Gaussian Wigner matrix, (2.5) has been proved in [81, 175] where its exact expression has been derived. Later, Dyson conjectured that (2.5) may also hold for more general random matrices [65, 66]. This conjecture has been proved by [129] for a certain subclass of not necessarily Gaussian Wigner matrices.

2.1.3 Wishart Matrices

Definition 2.3. The $m \times m$ random matrix $\mathbf{A} = \mathbf{H}\mathbf{H}^\dagger$ is a (central) *real/complex Wishart* matrix with n degrees of freedom and covariance matrix $\mathbf{\Sigma}$, ($\mathbf{A} \sim \mathcal{W}_m(n, \mathbf{\Sigma})$), if the columns of the $m \times n$ matrix $\mathbf{H}$ are zero-mean independent real/complex Gaussian vectors with covariance matrix $\mathbf{\Sigma}$.[5] The p.d.f. of a complex Wishart matrix $\mathbf{A} \sim \mathcal{W}_m(n, \mathbf{\Sigma})$ for $n \geq m$ is [244, p. 84], [182, 125][6]

$$f_{\mathbf{A}}(\mathbf{B}) = \frac{\pi^{-m(m-1)/2}}{\det \mathbf{\Sigma}^n \prod_{i=1}^m (n-i)!} \exp\left[-\text{tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{B}\right\}\right] \det \mathbf{B}^{n-m}. \tag{2.6}$$

2.1.4 Haar Matrices

Definition 2.4. A square matrix $\mathbf{U}$ is *unitary* if

$$\mathbf{U}\mathbf{U}^\dagger = \mathbf{U}^\dagger\mathbf{U} = \mathbf{I}.$$

Definition 2.5. [107] An $n \times n$ random matrix $\mathbf{U}$ is a *Haar* matrix[7] if it is uniformly distributed on the set, $\mathcal{U}(n)$, of $n \times n$ unitary matrices.[8] Its density function on $\mathcal{U}(n)$ is given by [107, 67]

$$2^{-n}\pi^{-\frac{1}{2}n(n+1)} \prod_{i=1}^{n} (n-i)! \tag{2.7}$$

Lemma 2.4. [107] The eigenvalues, ζ_i for $i \in \{1, \ldots, n\}$, of an $n \times n$ Haar matrix lie on the unit circle, i.e., $\zeta_i = e^{j\theta_i}$, and their joint p.d.f. is

$$\frac{1}{n!} \prod_{i<\ell} |\zeta_i - \zeta_\ell|^2. \tag{2.8}$$

Lemma 2.5. (e.g. [110]) If $1 \leq i, j, k, \ell \leq n$, $i \neq k$, $j \neq \ell$, and $\mathbf{U}$ is an

---

[5] If the entries of $\mathbf{H}$ have nonzero mean, $\mathbf{H}\mathbf{H}^\dagger$ is a non-central Wishart matrix.

[6] The case $n < m$ is studied in [267].

[7] Also called *isotropic* in the multi-antenna literature [171].

[8] A real Haar matrix is uniformly distributed on the set of real orthogonal matrices.

$n \times n$ (complex) Haar matrix, then

$$\begin{aligned}
\mathbb{E}[|\mathsf{U}_{ij}|^2] &= \frac{1}{n} \\
\mathbb{E}[|\mathsf{U}_{ij}|^4] &= \frac{2}{n(n+1)} \\
\mathbb{E}[|\mathsf{U}_{ij}|^2|\mathsf{U}_{kj}|^2] = \mathbb{E}[|\mathsf{U}_{ij}|^2|\mathsf{U}_{i\ell}|^2] &= \frac{1}{n(n+1)} \\
\mathbb{E}[|\mathsf{U}_{ij}|^2|\mathsf{U}_{k\ell}|^2] &= \frac{1}{n^2-1} \\
\mathbb{E}[\mathsf{U}_{ij}\mathsf{U}_{k\ell}\mathsf{U}^*_{i\ell}\mathsf{U}^*_{kj}] &= -\frac{1}{n(n^2-1)}.
\end{aligned}$$

A way to generate a Haar matrix is the following: let $\mathbf{H}$ be an $n \times n$ standard complex Gaussian matrix and let $\mathbf{R}$ be the upper triangular matrix obtained from the QR decomposition of $\mathbf{H}$ chosen such that all its diagonal entries are nonnegative. Then, as a consequence of Lemma 2.1, $\mathbf{HR}^{-1}$ is a Haar matrix [245].

2.1.5 Unitarily Invariant Matrices

Definition 2.6. A Hermitian random matrix $\mathbf{W}$ is called *unitarily invariant* if the joint distribution of its entries equals that of $\mathbf{VWV}^\dagger$ for any unitary matrix $\mathbf{V}$ independent of $\mathbf{W}$.

Example 2.1. A Haar matrix is unitarily invariant.

Example 2.2. A Gaussian Wigner matrix is unitarily invariant.

Example 2.3. A central Wishart matrix $\mathbf{W} \sim \mathcal{W}_m(n, \mathbf{I})$ is unitarily invariant.

Lemma 2.6. (e.g [111]) If $\mathbf{W}$ is unitarily invariant, then it can be decomposed as

$$\mathbf{W} = \mathbf{U\Lambda U}^\dagger.$$

with $\mathbf{U}$ a Haar matrix independent of the diagonal matrix $\mathbf{\Lambda}$.

Lemma 2.7. [110, 111] If $\mathbf{W}$ is unitarily invariant and $f(\cdot)$ is a real continuous function defined on the real line, then $f(\mathbf{W})$, given via the functional calculus, is also unitarily invariant.

Definition 2.7. A rectangular random matrix $\mathbf{H}$ is called *bi-unitarily invariant* if the joint distribution of its entries equals that of $\mathbf{UHV}^\dagger$ for any unitary matrices $\mathbf{U}$ and $\mathbf{V}$ independent of $\mathbf{H}$.

Example 2.4. A standard Gaussian random matrix is bi-unitarily invariant.

Lemma 2.8. [111] If $\mathbf{H}$ is a bi-unitarily invariant square random matrix, then it admits a polar decomposition $\mathbf{H} = \mathbf{UC}$ where $\mathbf{U}$ is a Haar matrix independent of the unitarily-invariant nonnegative definite random matrix $\mathbf{C}$.

In the case of a rectangular $m \times n$ matrix $\mathbf{H}$, with $m \leq n$, Lemma 2.8 also applies with $\mathbf{C}$ an $n \times n$ unitarily-invariant nonnegative definite random matrix and with $\mathbf{U}$ uniformly distributed over the manifold of complex $m \times n$ matrices such that $\mathbf{UU}^\dagger = \mathbf{I}$.

2.1.6 Properties of Wishart Matrices

In this subsection we collect a number of properties of central and non-central Wishart matrices and, in some cases, their inverses. We begin by considering the first and second order moments of a central Wishart matrix and its inverse.

Lemma 2.9. [164, 96] For a central Wishart matrix $\mathbf{W} \sim \mathcal{W}_m(n, \mathbf{I})$,

$$\begin{aligned} \mathbb{E}[\mathrm{tr}\{\mathbf{W}\}] &= mn \\ \mathbb{E}[\mathrm{tr}\{\mathbf{W}^2\}] &= mn\,(m+n) \\ \mathbb{E}[\mathrm{tr}^2\{\mathbf{W}\}] &= mn\,(mn+1). \end{aligned}$$

Lemma 2.10. [164, 96](see also [133]) For a central Wishart matrix $\mathbf{W} \sim \mathcal{W}_m(n, \mathbf{I})$ with $n > m$,

$$\mathbb{E}\left[\mathrm{tr}\left\{\mathbf{W}^{-1}\right\}\right] = \frac{m}{n-m} \tag{2.9}$$

while, for $n > m+1$,

$$\begin{aligned} \mathbb{E}\left[\mathrm{tr}\left\{\mathbf{W}^{-2}\right\}\right] &= \frac{m\,n}{(n-m)^3 - (n-m)} \\ \mathbb{E}\left[\mathrm{tr}^2\left\{\mathbf{W}^{-1}\right\}\right] &= \frac{m}{n-m}\left(\frac{n}{(n-m)^2-1} + \frac{m-1}{n-m+1}\right). \end{aligned}$$

For higher order moments of Wishart and generalized inverse Wishart matrices, see [96].

From Lemma 2.1, we can derive several formulas on the determinant and log-determinant of a Wishart matrix.

Theorem 2.11. [182, 131][9] A central complex Wishart matrix $\mathbf{W} \sim \mathcal{W}_m(n, \mathbf{I})$, with $n \geq m$, satisfies

$$\mathbb{E}\left[\det \mathbf{W}^k\right] = \prod_{\ell=0}^{m-1} \frac{\Gamma(n-\ell+k)}{\Gamma(n-\ell)} \tag{2.10}$$

and hence the moment-generating function of $\log_e \det \mathbf{W}$ for $\zeta \geq 0$ is

$$\mathbb{E}\left[e^{\zeta \log_e \det \mathbf{W}}\right] = \prod_{\ell=0}^{m-1} \frac{\Gamma(n-\ell+\zeta)}{\Gamma(n-\ell)} \tag{2.11}$$

with $\Gamma(\cdot)$ denoting the Gamma function [97]

$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$$

which, for integer arguments, satisfies $\Gamma(n+1) = n!$ From (2.11),

$$\mathbb{E}[\log_e \det \mathbf{W}] = \sum_{\ell=0}^{m-1} \psi(n-\ell) \tag{2.12}$$

$$\mathrm{Var}[\log_e \det \mathbf{W}] = \sum_{\ell=0}^{m-1} \dot{\psi}(n-\ell) \tag{2.13}$$

where $\psi(\cdot)$ is Euler's digamma function [97], which for natural arguments can be expressed as

$$\psi(m) = \psi(1) + \sum_{\ell=1}^{m-1} \frac{1}{\ell} = \psi(m-1) + \frac{1}{m-1} \tag{2.14}$$

with $-\psi(1) = 0.577215...$ the Euler-Mascheroni constant. The derivative of $\psi(\cdot)$, in turn, can be expressed as

$$\dot{\psi}(m+1) = \dot{\psi}(m) - \frac{1}{m^2} \tag{2.15}$$

[9] Note that [182, 131] derive the real counterpart of Theorem 2.11, from which the complex case follows immediately.

with $\dot{\psi}(1) = \frac{\pi^2}{6}$.

If $\mathbf{\Sigma}$ and $\mathbf{\Phi}$ are positive definite deterministic matrices and $\mathbf{H}$ is an $n \times n$ complex Gaussian matrix with independent zero-mean unit-variance entries, then $\mathbf{W} = \mathbf{\Sigma}\mathbf{H}\mathbf{\Phi}\mathbf{H}^\dagger$ satisfies (using (2.10))

$$\mathbb{E}\left[\det \mathbf{W}^k\right] = \det(\mathbf{\Sigma}\mathbf{\Phi})^k \prod_{\ell=0}^{n-1} \frac{(n-\ell+k-1)!}{(n-\ell-1)!} \tag{2.16}$$

The generalization of (2.16) for rectangular $\mathbf{H}$ is derived in [165, 219]. Analogous relationships for the non-central Wishart matrix are derived in [5].

Theorem 2.12. [166] Let $\mathbf{H}$ be an $n \times m$ complex Gaussian matrix with zero-mean unit-variance entries and let $\mathbf{W}$ be a complex Wishart matrix $\mathbf{W} \sim \mathcal{W}_n(p, \mathbf{I})$, with $m \le n \le p$. Then, for $\zeta \in (-1, 1)$,

$$\begin{aligned}
\mathbb{E}[\det(\mathbf{H}^\dagger \mathbf{W}^{-1}\mathbf{H})^\zeta] &= \prod_{\ell=0}^{m-1} \frac{\Gamma(m+p-n-\zeta-\ell)\,\Gamma(n+\zeta-\ell)}{\Gamma(n-\ell)\,\Gamma(m+p-n-\ell)} \\
\mathbb{E}[\log\det(\mathbf{H}^\dagger \mathbf{W}^{-1}\mathbf{H})] &= \sum_{\ell=0}^{m-1} \left(\psi(n-\ell) - \psi(m+p-n-\ell)\right).
\end{aligned}$$

Additional results on quadratic functions of central and non-central Wishart matrices can be found in [141, 142, 144] and the references therein.

Some results on the p.d.f. of complex pseudo-Wishart matrices[10] and their corresponding eigenvalues can be found in [58, 59, 168].

Next, we turn our attention to the determinant and log-determinant of matrices that can be expressed as a multiple of the identity plus a Wishart matrix, a familiar form in the expressions of the channel capacity.

[10] $\mathbf{W} = \mathbf{H}\mathbf{H}^\dagger$ is a pseudo-Wishart matrix if $\mathbf{H}$ is a $m \times n$ Gaussian matrix and the correlation matrix of the columns of $\mathbf{H}$ has a rank strictly larger than n [244, 267, 94, 58, 59].

Theorem 2.13. A complex Wishart matrix $\mathbf{W} \sim \mathcal{W}_m(n, \mathbf{I})$, with $n \geq m$, satisfies

$$\mathbb{E}[\det(\mathbf{I} + \gamma \mathbf{W})] = \sum_{i=0}^{m} \binom{m}{i} \frac{n!}{(n-i)!} \gamma^i. \tag{2.17}$$

Theorem 2.14. [38, 299] Let $\mathbf{W}$ be a central Wishart matrix $\mathbf{W} \sim \mathcal{W}_m(n, \mathbf{I})$ and let $t = \min\{n, m\}$ and $r = \max\{n, m\}$. The moment-generating function of $\log_e \det(\mathbf{I} + \gamma \mathbf{W})$ is

$$\mathbb{E}\left[e^{\zeta \log_e \det(\mathbf{I}+\gamma\mathbf{W})}\right] = \frac{\det \mathbf{G}(\zeta)}{\prod_{i=1}^{t}(r-i)!} \tag{2.18}$$

with $\mathbf{G}(\zeta)$ a $t \times t$ Hankel matrix whose (i,k)th entry is

$$\begin{aligned} \mathsf{G}_{i,k} &= \int_0^\infty (1+\gamma\lambda)^\zeta \, \lambda^{d-1} e^{-\lambda} d\lambda \\ &= \frac{\pi}{\Gamma(-\zeta)\sin(\pi(d-1+\zeta))} \left(\frac{\gamma^{-d}\ (d-1)!}{\Gamma(1+d+\zeta)} \, {}_1F_1\left(d, 1+d+\zeta, \tfrac{1}{\gamma}\right) \right. \\ &\qquad \left. - \frac{\gamma^\zeta\ \Gamma(-\zeta)}{\Gamma(1-d-\zeta)} \, {}_1F_1\left(-\zeta, 1-d-\zeta, \tfrac{1}{\gamma}\right) \right) \end{aligned} \tag{2.19}$$

with ${}_1F_1(\cdot)$ the confluent hypergeometric function [97] and with $d = r - t + i + k + 1$.

For a non-central Wishart matrix with covariance matrix equal to the identity, a series expression for $\mathbb{E}[\log\det(\mathbf{I} + \gamma\mathbf{W})]$ has been computed in [3] while the moment-generating function (2.18) has been computed in [134] in terms of the integral of hypergeometric functions.

For a central Wishart matrix $\mathbf{W} \sim \mathcal{W}_m(n, \mathbf{\Sigma})$ where $\mathbf{\Sigma}$ is positive definite with distinct eigenvalues, the moment-generating function (2.18) has been computed in [234] and [135].[11]

Theorem 2.15. [192] If $\mathbf{H}$ is an $m \times m$ zero-mean unit-variance complex Gaussian matrix and $\mathbf{\Sigma}$ and $\mathbf{\Upsilon}$ are positive definite matrices having

[11] Reference [234] evaluates (2.18) in terms of Gamma functions for $m > n$ while reference [135] evaluates it for arbitrary m and n, in terms of confluent hypergeometric functions of the second kind [97].

distinct eigenvalues a_i and ϕ_i, respectively, then for $\zeta \leq 0$

$$\mathbb{E}\left[\det\left(\mathbf{I}+\mathbf{\Sigma H \Upsilon H}^\dagger\right)^{\zeta}\right] \quad = \quad {}_2F_0(-\zeta, m \,|\, -\mathbf{\Sigma}, \mathbf{\Upsilon}) \qquad (2.20)$$

where the hypergeometric function with matrix arguments [192] is

$${}_2F_0(-\zeta, m \,|\, -\mathbf{\Sigma}, \mathbf{\Upsilon}) = \frac{\det\left(\{{}_2F_0(-\zeta - m + 1, 1| - a_i\phi_j)\}\right)}{\prod_{k=1}^{m-1}(-\zeta - k)^k \prod_{i<j}^{m}(\phi_i - \phi_j)\prod_{i<j}^{m}(a_j - a_i)}$$

with ${}_2F_0(\cdot, \cdot|\cdot)$ denoting the scalar hypergeometric function [1].[12]

For $\mathbf{\Upsilon} = \mathbf{I}$ (resp. $\mathbf{\Sigma} = \mathbf{I}$), (2.20) still holds but with ${}_2F_0(s, m \,|\, -\mathbf{\Sigma}, \mathbf{I})$ (resp. ${}_2F_0(-\zeta, m \,|\, \mathbf{I}, -\mathbf{\Upsilon})$) replaced by [192]

$${}_2F_0(-\zeta, m \,|\, \mathbf{\Theta}) = \frac{\det\left(\left\{\theta_j^{m-i}\, {}_2F_0(-\zeta - i + 1, m - i + 1 \,|\, \theta_j)\right\}\right)}{\prod_{i<j}^{n}(\theta_i - \theta_j)} \qquad (2.21)$$

with $\mathbf{\Theta} = -\mathbf{\Sigma}$ (resp. $\mathbf{\Theta} = -\mathbf{\Upsilon}$).

The counterpart of Theorem 2.15 for a rectangular matrix $\mathbf{H}$ is as follows.

Theorem 2.16. [148, 150] Let $\mathbf{H}$ be an $m \times n$ complex Gaussian matrix with zero-mean unit-variance entries with $m \leq n$ and define

$$M(\zeta) = \mathbb{E}\left[e^{\zeta \log\det(\mathbf{I}+\gamma\mathbf{\Sigma H \Upsilon H}^\dagger)}\right]$$

with $\mathbf{\Sigma}$ and $\mathbf{\Upsilon}$ positive definite matrices having distinct eigenvalues a_i and ϕ_i, respectively. Then for $\zeta \leq 0$

$$M(\zeta) = \frac{\det\mathbf{G}(\zeta)\det\mathbf{\Sigma}^{-d}}{(-1)^{\frac{d(d-1)}{2}}(-\gamma)^{\frac{n(n-1)}{2}}} \prod_{i=0}^{n-1}\frac{1}{(\zeta\log\frac{1}{e} - i)^i}\prod_{i<j}^{n}\frac{1}{\phi_i - \phi_j}\prod_{i<j}^{m}\frac{1}{a_i - a_j}$$

with $d = n - m$ and with $\mathbf{G}(\zeta)$ an $n \times n$ matrix whose (i,j)th entry is

$$\mathsf{G}_{i,j}(\zeta) = \begin{cases} {}_2F_0\left(\zeta\log\frac{1}{e} - n + 1,\, 1 \,|\, -\gamma\phi_j a_i\right) & \begin{array}{l} i \in \{1, \ldots, m\} \\ j \in \{1, \ldots, n\} \end{array} \\ (-\gamma\phi_j)^{i-1-m}\left[\zeta\log\frac{1}{e} - n + 1\right]_{i-1-m} & \begin{array}{l} i \in \{m+1, \ldots, n\} \\ j \in \{1 \ldots, n\} \end{array} \end{cases}$$

where $[b]_k = \frac{\Gamma(b+k)}{\Gamma(b)}$ indicates the Pochhammer symbol.[13]

[12] In the remainder, $\det(\{f(i,j)\})$ denotes the determinant of a matrix whose (i,j)th entry is $f(i,j)$.

[13] If b is an integer, $[b]_k = b(b+1)\ldots(b-1+k)$.

An alternative expression for the moment-generating function in Theorem 2.16 can be found in [231].

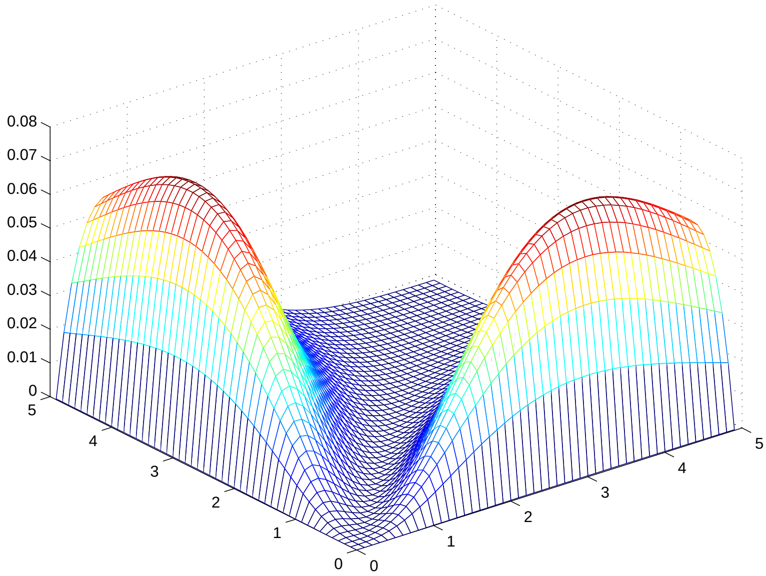

Fig. 2.1 Joint p.d.f. of the unordered positive eigenvalues of the Wishart matrix $\mathbf{H}\mathbf{H}^\dagger$ with $r = 3$ and $t = 2$. (Scaled version of (2.22).)

To conclude the exposition on properties of Wishart matrices, we summarize several results on the non-asymptotic distribution of their eigenvalues.

Theorem 2.17. [75, 120, 89, 210] Let the entries of $\mathbf{H}$ be i.i.d. complex Gaussian with zero mean and unit variance. The joint p.d.f. of the ordered strictly positive eigenvalues of the Wishart matrix $\mathbf{H}\mathbf{H}^\dagger$, $\lambda_1 \geq \ldots \geq \lambda_t$, equals

$$e^{-\sum_{i=1}^{t}\lambda_i} \prod_{i=1}^{t} \frac{\lambda_i^{r-t}}{(t-i)!\,(r-i)!} \prod_{i<j}^{t} (\lambda_i - \lambda_j)^2 \tag{2.22}$$

where t and r are the minimum and maximum of the dimensions of $\mathbf{H}$.

The marginal p.d.f. of the unordered eigenvalues is[14] (e.g. [32])

$$\mathsf{g}_{r,t}(\lambda) = \frac{1}{t} \sum_{k=0}^{t-1} \frac{k!}{(k+r-t)!} \left[L_k^{r-t}(\lambda)\right]^2 \lambda^{r-t} e^{-\lambda} \tag{2.23}$$

where the Laguerre polynomials are

$$L_k^n(\lambda) = \frac{e^\lambda}{k!\lambda^n} \frac{d^k}{d\lambda^k} \left(e^{-\lambda}\lambda^{n+k}\right). \tag{2.24}$$

Figure 2.1 depicts the joint p.d.f. of the unordered positive eigenvalues of the Wishart matrix $\mathbf{H}\mathbf{H}^\dagger$, $\lambda_1 > 0, \ldots, \lambda_t > 0$, which is obtained by dividing the joint p.d.f. of the ordered positive eigenvalues by $t!$

Theorem 2.18. Let $\mathbf{W}$ be a central complex Wishart matrix $\mathbf{W} \sim W_m(n, \mathbf{\Sigma})$ with $n \geq m$, where the eigenvalues of $\mathbf{\Sigma}$ are distinct and their ordered values are $a_1 > \ldots > a_m > 0$. The joint p.d.f. of the ordered positive eigenvalues of $\mathbf{W}$, $\lambda_1 \geq \ldots \geq \lambda_m$, equals [125]

$$\frac{\det(\{e^{-\lambda_j/a_i}\})}{\det \mathbf{\Sigma}^n} \prod_{\ell=1}^{m} \frac{\lambda_\ell^{n-m}}{(n-\ell)!} \prod_{k<\ell}^{m} \frac{\lambda_k - \lambda_\ell}{a_k - a_\ell} a_\ell \, a_k. \tag{2.25}$$

The marginal p.d.f. of the unordered eigenvalues is [2]

$$\mathsf{q}_{m,n}(\lambda) = \frac{\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{m} \mathcal{D}(i,j)\, \lambda^{n-m+j-1}\, e^{-\lambda/a_i}}{\displaystyle m \det \mathbf{\Sigma}^n \prod_{\ell=1}^{m} (n-\ell)! \prod_{k<\ell}^{m} (\frac{1}{a_\ell} - \frac{1}{a_k})} \tag{2.26}$$

where $\mathcal{D}(i,j)$ is the (i,j)th cofactor of the matrix $\mathbf{D}$ with entries

$$\mathsf{D}_{\ell,k} = \frac{(n-m+k-1)!}{a_\ell^{-n+m-k}}. \tag{2.27}$$

Figure 2.2 contrasts a histogram obtained via Monte Carlo simulation with the marginal p.d.f. of the unordered eigenvalues of $\mathbf{W} \sim W_m(n, \mathbf{\Sigma})$ with $n = 3$ and $m = 2$ and with the correlation matrix $\mathbf{\Sigma}$ chosen such that[15]

$$\mathsf{\Sigma}_{i,j} = e^{-0.2(i-j)^2}. \tag{2.28}$$

[14] An alternative expression for (2.23) can be found in [183, B.7].
[15] The correlation in (2.28) is typical of a base station in a wireless cellular system.

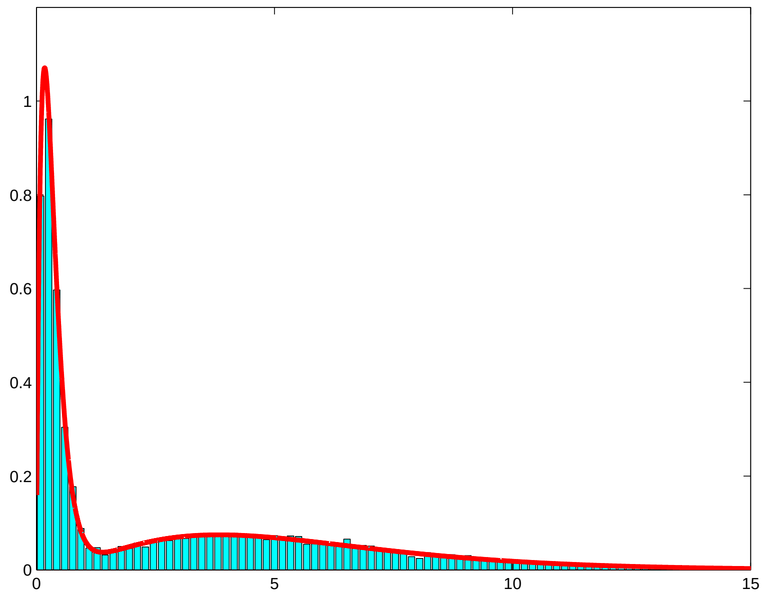

Fig. 2.2 Marginal p.d.f. of the unordered eigenvalues of $\mathbf{W} \sim W_m(n, \boldsymbol{\Sigma})$ with $n = 3$, $m = 2$ and $\boldsymbol{\Sigma}_{i,j} = e^{-0.2(i-j)^2}$, compared to an histogram obtained via Monte Carlo simulation.

Theorem 2.19. Let $\mathbf{W}$ be a central complex Wishart matrix $\mathbf{W} \sim W_m(n, \boldsymbol{\Sigma})$ with $m > n$, where the eigenvalues of $\boldsymbol{\Sigma}$ are distinct and their ordered values are $a_1 > \ldots > a_m > 0$. The joint p.d.f. of the unordered strictly positive eigenvalues of $\mathbf{W}$, $\lambda_1, \ldots, \lambda_n$, equals [80]

$$\det(\boldsymbol{\Xi}) \prod_{\ell=1}^{n} \frac{1}{\ell!} \prod_{k<\ell}^{m} \frac{1}{(a_\ell - a_k)} \prod_{k<\ell}^{n} (\lambda_\ell - \lambda_k) \tag{2.29}$$

with

$$\boldsymbol{\Xi} = \begin{bmatrix} 1 & a_1 & \ldots & a_1^{m-n-1} & a_1^{m-n-1} e^{-\frac{\lambda_1}{a_1}} & \ldots & a_1^{m-n-1} e^{-\frac{\lambda_n}{a_1}} \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ 1 & a_m & \ldots & a_m^{m-n-1} & a_m^{m-n-1} e^{-\frac{\lambda_1}{a_m}} & \ldots & a_m^{m-n-1} e^{-\frac{\lambda_n}{a_m}} \end{bmatrix}.$$

The marginal p.d.f. of the unordered eigenvalues is given in [2].

Let $\mathbf{H}$ be an $m \times m$ zero-mean unit-variance complex Gaussian matrix and $\boldsymbol{\Sigma}$ and $\boldsymbol{\Upsilon}$ be nonnegative definite matrices. Then the joint p.d.f. of the eigenvalues of $\mathbf{\Sigma H \Upsilon H}^\dagger$ is computed in [209] while the marginal p.d.f. has been computed in [230].

The distributions of the largest and smallest eigenvalues of a central and non-central Wishart matrix $\mathbf{W} \sim W_m(n, \mathbf{I})$ are given in [67] and [140, 143, 136]. The counterpart for a central Wishart matrix $\mathbf{W} \sim W_m(n, \mathbf{\Sigma})$ with $n \geq m$ can be found in [208].

2.1.7 Rank Results

Lemma 2.20. For any $N \times K$ matrices $\mathbf{A}$, $\mathbf{B}$,

$$\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}).$$

Moreover, the rank of $\mathbf{A}$ is less than or equal to the number of nonzero entries of $\mathbf{A}$.

Lemma 2.21. For any Hermitian $N \times N$ matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\sum_{i=1}^{N} (\lambda_i(\mathbf{A}) - \lambda_i(\mathbf{B}))^2 \leq \text{tr}\,(\mathbf{A} - \mathbf{B})^2\,.$$

Lemma 2.22. [313, 10] For any $N \times K$ matrices $\mathbf{A}$ and $\mathbf{B}$,

$$N \sup_{x \geq 0} |\mathsf{F}^N_{\mathbf{AA}^\dagger}(x) - \mathsf{F}^N_{\mathbf{BB}^\dagger}(x)| \leq \text{rank}(\mathbf{A} - \mathbf{B}). \tag{2.30}$$

Lemma 2.23. [313, 10] For any $N \times N$ Hermitian matrices $\mathbf{A}$ and $\mathbf{B}$,

$$N \sup_{x \geq 0} |\mathsf{F}^N_{\mathbf{A}}(x) - \mathsf{F}^N_{\mathbf{B}}(x)| \leq \text{rank}(\mathbf{A} - \mathbf{B}). \tag{2.31}$$

2.1.8 Karhunen-Loève Expansion

As will be illustrated in Section 3, this transformation, widely used in image processing, is a very convenient tool that facilitates the application of certain random matrix results to channels of practical interest.

Definition 2.8. Let $\mathbf{A}$ be an $N \times K$ random matrix. Denote the correlation between the (i, j)th and (i', j')th entries of $\mathbf{A}$ by

$$\mathsf{r}_{\mathbf{A}}(i, j; i', j') = \mathbb{E}\left[\mathsf{A}_{i,j}\mathsf{A}^*_{i',j'}\right]. \tag{2.32}$$

The Karhunen-Loève expansion of $\mathbf{A}$ yields an $N \times K$ *image* random matrix $\tilde{\mathbf{A}}$ whose entries are

$$\tilde{\mathsf{A}}_{k,\ell} = \sum_{i=1}^{N} \sum_{j=1}^{K} \mathsf{A}_{i,j} \, \psi^*_{k,\ell}(i,j)$$

where the so-called expansion kernel $\{\psi_{k,\ell}(i,j)\}$ is a set of complete orthonormal discrete basis functions formed by the eigenfunctions of the correlation function of $\mathbf{A}$, i. e., this kernel must satisfy for all $k \in \{1,\ldots,N\}$ and $\ell \in \{1,\ldots,K\}$

$$\sum_{i'=1}^{N} \sum_{j'=1}^{K} \mathsf{r}_{\mathbf{A}}(i,j;i',j') \, \psi_{k,\ell}(i',j') = \lambda_{k,\ell}(\mathsf{r}_{\mathbf{A}}) \, \psi_{k,\ell}(i,j) \tag{2.33}$$

where we indicate the eigenvalues of $\mathsf{r}_{\mathbf{A}}$ by $\lambda_{k,\ell}(\mathsf{r}_{\mathbf{A}})$.

Lemma 2.24. The entries of a Karhunen-Loève image are, by construction, uncorrelated and with variances given by the eigenvalues of the correlation of the original matrix, i.e.,

$$\mathbb{E}\left[\tilde{\mathsf{A}}_{k,\ell}\tilde{\mathsf{A}}^*_{j,i}\right] = \begin{cases} \lambda_{k,\ell}(\mathsf{r}_{\mathbf{A}}) & \text{if } k=j \text{ and } \ell=i, \\ 0 & \text{otherwise.} \end{cases} \tag{2.34}$$

Lemma 2.25. If the expansion kernel can be factored as

$$\psi_{k,\ell}(i,j) = u_k(i)\, v_\ell(j), \tag{2.35}$$

then

$$\mathbf{A} = \mathbf{U}\tilde{\mathbf{A}}\mathbf{V}^\dagger$$

with $\mathsf{U}_{k,i}{=}u_k(i)$ and $\mathsf{V}_{j,\ell}{=}v^*_\ell(j)$, which renders the matrices $\mathbf{U}$ and $\mathbf{V}$ unitary. As a consequence, $\mathbf{A}$ and its Karhunen-Loève image, $\tilde{\mathbf{A}}$, have the same singular values.

Thus, with the Karhunen-Loève expansion we can map the singular values of a matrix with correlated Gaussian entries and factorable kernel to those of another Gaussian matrix whose entries are independent.

Definition 2.9. The correlation of a random matrix $\mathbf{A}$ is said to be *separable* if $\mathsf{r}_{\mathbf{A}}(i,j;i',j')$ can be expressed as the product of two marginal correlations[16] that are functions, respectively, of (i,j) and (i',j').

If the correlation of $\mathbf{A}$ is separable, then the kernel is automatically factorable[17] and, furthermore, $\lambda_{k,\ell}(\mathsf{r}_{\mathbf{A}})=\lambda_k\lambda_\ell$ where λ_k and λ_ℓ are, respectively, the kth and ℓth eigenvalues of the two marginal correlations whose product equals $\mathsf{r}_{\mathbf{A}}$.

2.1.9 Regular Matrices

Definition 2.10. An $N \times K$ matrix $\mathbf{P}$ is *asymptotically row-regular* if

$$\lim_{K\to\infty} \frac{1}{K} \sum_{j=1}^{K} 1\{\mathsf{P}_{i,j} \leq \alpha\}$$

is independent of i for all $\alpha \in \mathbb{R}$, as the aspect ratio $\frac{K}{N}$ converges to a constant. A matrix whose transpose is asymptotically row-regular is called *asymptotically column-regular*. A matrix that is both asymptotically row-regular and asymptotically column-regular is called *asymptotically doubly-regular* and satisfies

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathsf{P}_{i,j} = \lim_{K\to\infty} \frac{1}{K} \sum_{j=1}^{K} \mathsf{P}_{i,j}. \tag{2.36}$$

If (2.36) is equal to 1, then $\mathbf{P}$ is *standard asymptotically doubly-regular*.

Example 2.5. An $N \times K$ rectangular Toeplitz matrix

$$\mathsf{P}_{i,j} = \varphi(i-j)$$

with $K \geq N$ is an asymptotically row-regular matrix. If either the function φ is periodic or $N = K$, then the Toeplitz matrix is asymptotically doubly-regular.

---

[16] Equivalently, the correlation matrix of the vector obtained by stacking up the columns of $\mathbf{A}$ can be expressed as the Kronecker product of two separate matrices that describe, respectively, the correlation between the rows and between the columns of $\mathbf{A}$.

[17] Another relevant example of a factorable kernel occurs with shift-invariant correlation functions such as $\mathsf{r}_{\mathbf{A}}(i,j;i',j') = \mathsf{r}_{\mathbf{A}}(i-i',j-j')$, for which the Karhunen-Loève image is equivalent to a two-dimensional Fourier transform.

2.1.10 Cauchy-Binet Theorem

The result reported below, which is the continuous analog of the Cauchy-Binet formula [121], has been applied in several contributions [39, 166, 2, 231, 219] in order to compute the capacity of multi-antenna channels and the marginal distributions of the singular values of matrices with correlated Gaussian entries.

Theorem 2.26. [144](see also [6]) Let $\mathbf{F}$ and $\mathbf{G}$ be $n \times n$ matrices parametrized by a real n-vector $(w_1, \ldots, w_n)$:

$$\mathbf{F}_{i,j} = f_j(w_i) \tag{2.37}$$

$$\mathbf{G}_{i,j} = g_j(w_i) \tag{2.38}$$

where f_j and g_j, $j = 1, \ldots, n$, are real-valued functions defined on the real line. Then, for $0 < a < b$,

$$\int_a^b \cdots \int_a^b \det\mathbf{F} \, \det\mathbf{G} \; dw_1, \ldots, dw_n = n! \det\mathbf{A}$$

where $\mathbf{A}$ is another $n \times n$ matrix whose (i,j)-th entry is

$$\mathbf{A} = \int_a^b f_i(w) g_j(w) \, dw.$$

Note that, in [144], the factor $n!$ does not appear because the variables $w_1, \ldots, w_n$ are ordered.

2.1.11 Lyapunov Exponent

The celebrated result in this subsection, although outside the main focus of this monograph, has been used in several engineering applications [114, 122, 83].

As $n \to \infty$, the growth of the maximum singular value of the product of n random matrices is exponential with a rate of increase given by the following result.

Theorem 2.27. [79, 193, 29, 44] Denote the maximum singular value of $\mathbf{A}$ (spectral norm of $\mathbf{A}$) by $\rho(\mathbf{A})$. Let $\mathbf{A}_1, \ldots, \mathbf{A}_n, \ldots$ be a stationary ergodic sequence of random matrices for which

$$\mathbb{E}[\log(\max\{\rho(\mathbf{A}_n), 1\}) < \infty.$$

Then, there exists a deterministic constant λ (the so-called Lyapunov exponent) such that almost surely[18]

$$\lim_{n\to\infty} \frac{1}{n} \log \rho \left(\prod_{i=1}^{n} \mathbf{A}_i \right) = \lambda. \tag{2.39}$$

2.2 Transforms

As mentioned in Section 1.3, it is often the case that the solution for the limiting spectrum is obtained in terms of a transform of its distribution. In this section, we review the most useful transforms including the Shannon transform and the η-transform which, suggested by problems of interest in communications, are introduced in this monograph.

For notational convenience, we refer to the transform of a random variable and the transform of its cumulative distribution or density function interchangeably. If the distribution of such variable equals the asymptotic spectrum of a random matrix, then we refer to the transform of the matrix and the transform of its asymptotic spectrum interchangeably.

2.2.1 Stieltjes Transform

Let X be a real-valued random variable with distribution $F_X(\cdot)$. Its Stieltjes transform is defined for complex arguments as[19]

$$\mathcal{S}_X(z) = \mathbb{E}\left[\frac{1}{X - z}\right] = \int_{-\infty}^{\infty} \frac{1}{\lambda - z}\, dF_X(\lambda). \tag{2.40}$$

Although (2.40) is an analytic function on the complement of the support of $F_X(\cdot)$ on the complex plane, it is customary to further restrict the domain of $\mathcal{S}_X(z)$ to arguments having positive imaginary parts. According to the definition, the signs of the imaginary parts of z and $\mathcal{S}_X(z)$ coincide. In the following examples, the sign of the square root should be chosen so that this property is satisfied.

[18] This property is satisfied by any conventional norm.

[19] The Stieltjes transform is also known as the Cauchy transform and it is equal to $-\pi$ times the Hilbert transform when defined on the real line. As with the Fourier transform there is no universal agreement on its definition, as sometimes the Stieltjes transform is defined as $\mathcal{S}_X(-z)$ or $-\mathcal{S}_X(z)$.

Example 2.6. The Stieltjes transform of the semi-circular law $w(\cdot)$ in (1.18) is

$$\mathcal{S}_w(z) = \frac{1}{2\pi}\int_{-2}^{2}\frac{\sqrt{4-\lambda^2}}{\lambda - z}d\lambda = \frac{1}{2}\left[-z \pm \sqrt{z^2-4}\right]. \tag{2.41}$$

Example 2.7. The Stieltjes transform of the Marčenko-Pastur law $\mathsf{f}_\beta(\cdot)$ in (1.10) is

$$\begin{aligned}\mathcal{S}_{\mathsf{f}_\beta}(z) &= \int_a^b \frac{1}{\lambda - z}\,\mathsf{f}_\beta(\lambda)\,d\lambda \\ &= \frac{1-\beta-z\pm\sqrt{z^2-2(\beta+1)z+(\beta-1)^2}}{2\beta z}.\end{aligned} \tag{2.42}$$

Example 2.8. The Stieltjes transform of $\tilde{\mathsf{f}}_\beta(\cdot)$ in (1.12) is

$$\begin{aligned}\mathcal{S}_{\tilde{\mathsf{f}}_\beta}(z) &= \int_a^b \frac{1}{\lambda - z}\,\tilde{\mathsf{f}}_\beta(\lambda)\,d\lambda \\ &= \frac{-1+\beta-z\pm\sqrt{z^2-2(\beta+1)z+(\beta-1)^2}}{2z}.\end{aligned} \tag{2.43}$$

Example 2.9. The Stieltjes transform of the averaged empirical eigenvalue distribution of the unit-rank matrix $\mathbf{s}\mathbf{s}^\dagger$ is equal to

$$\mathcal{S}(z) = \frac{1}{N}\,\mathcal{S}_P(z) - \left(1-\frac{1}{N}\right)\frac{1}{z} \tag{2.44}$$

where N is the dimension of $\mathbf{s}$ and $\mathcal{S}_P$ is the Stieltjes transform of the random variable $\|\mathbf{s}\|^2$.

Given $\mathcal{S}_X(\cdot)$, the inversion formula that yields the p.d.f. of X is [246, 222]

$$\mathsf{f}_X(\lambda) = \lim_{\omega\to 0^+}\frac{1}{\pi}\mathrm{Im}\Big[\mathcal{S}_X(\lambda + \mathrm{j}\,\omega)\Big]. \tag{2.45}$$

Assuming $F_X(\cdot)$ has compact support, we can expand $\mathcal{S}_X(\cdot)$ in a Laurent series involving the moments of X. Expanding $\frac{1}{\lambda - z}$ with respect to z, exchanging summation and integration and using analytical extension, (2.40) can be written as

$$\mathcal{S}_X(z) = -\frac{1}{z}\sum_{k=0}^{\infty}\frac{\mathbb{E}[X^k]}{z^k}. \tag{2.46}$$

If the distribution of X is the averaged empirical eigenvalue distribution of an $N \times N$ random matrix $\mathbf{A}$, then $\mathbb{E}[X^k]$ can be regarded as the kth moment $\mathbb{E}\left[\frac{1}{N}\mathrm{tr}\{\mathbf{A}^k\}\right]$. As a consequence, $\mathcal{S}_X(\cdot)$ can be regarded as a generating function for the moments of the random matrix whose averaged empirical eigenvalue distribution is F_X.

As indicated at the onset of Section 2.2, we often denote the Stieltjes transform of the asymptotic empirical distribution of a matrix $\mathbf{A}$ by $\mathcal{S}_{\mathbf{A}}(\cdot)$. However, as in Examples 2.6, 2.7 and 2.8, it is sometimes convenient to subscript $\mathcal{S}(\cdot)$ by its corresponding asymptotic empirical distribution or density function. Similar notational conventions will be applied to the transforms to be defined in the sequel.

2.2.2 η-transform

In the applications of interest, it is advantageous to consider a transform that carries some engineering intuition, while at the same time is closely related to the Stieltjes transform.

Interestingly, this transform, which has not been used so far in the random matrix literature, simplifies many derivations and statements of results.[20]

Definition 2.11. The η-transform of a nonnegative random variable X is

$$\eta_X(\gamma) = \mathbb{E}\left[\frac{1}{1+\gamma X}\right] \tag{2.47}$$

where γ is a nonnegative real number and thus $0 < \eta_X(\gamma) \leq 1$.

The rationale for introducing this quantity can be succinctly explained by considering a hypothetical situation where the sum of three components is observed (for example, at the output of a linear receiver): "desired signal" with strength γ, "background noise" with unit strength, and "multiuser interference" with strength γX. The reason the multiuser interference strength is scaled by γ is reminiscent of the fact that, in many systems, the power of the users either is equal (perfect power control) or scales linearly. The expected SINR divided by

[20] The η-transform was first used in [273].

the single-user (i.e. $X = 0$) signal-to-noise ratio is given by (2.47). Since this notion is reminiscent of the *multiuser efficiency* [271], we have chosen the notation η standard in multiuser detection.

Either with analytic continuation or including the negative real line in the domain of definition of the Stieltjes transform, we obtain the simple relationship with the η-transform:

$$\eta_X(\gamma) = \frac{\mathcal{S}_X(-\frac{1}{\gamma})}{\gamma}. \tag{2.48}$$

Given the η-transform, (2.48) gives the Stieltjes transform by analytic continuation in the whole positive upper complex half-plane, and then the distribution of X through the inversion formula (2.45).

From (2.46) and (2.48), the η-transform can be written in terms of the moments of X:

$$\eta_X(\gamma) = \sum_{k=0}^{\infty} (-\gamma)^k \mathbb{E}[X^k], \tag{2.49}$$

whenever the moments of X exist and the series in (2.49) converges.

From (1.8) it follows that the MMSE considered in Section 1.2 is equal to the η-transform of the empirical distribution of the eigenvalues of $\mathbf{H}^\dagger\mathbf{H}$.

Simple properties of the η-transform that prove useful are:

- $\eta_X(\gamma)$ is strictly monotonically decreasing with $\gamma \geq 0$ from 1 to $\mathbb{P}[X = 0]$. [21]
- $\gamma\eta_X(\gamma)$ is strictly monotonically increasing with $\gamma \geq 0$ from 0 to $\mathbb{E}[\frac{1}{X}]$.

Thus, the asymptotic fraction of zero eigenvalues of $\mathbf{A}$ is

$$\lim_{\gamma\to\infty} \eta_{\mathbf{A}}(\gamma) \tag{2.50}$$

while

$$\lim_{n\to\infty} \frac{1}{n}\mathrm{tr}\{\mathbf{A}^{-1}\} = \lim_{\gamma\to\infty} \gamma\eta_{\mathbf{A}}(\gamma). \tag{2.51}$$

[21] Note from (2.47) that it is easy (and, it will turn out, sometimes useful) to extend the definition of the η-transform to (generalized or defective) distributions that put some nonzero mass at $+\infty$. In this case, $\eta_X(0) = \mathbb{P}[X < \infty]$

Example 2.10. [271, p. 303] The η-transform of the Marčenko-Pastur law given in (1.10) is

$$\eta(\gamma) = 1 - \frac{\mathcal{F}(\gamma, \beta)}{4\,\beta\,\gamma}. \tag{2.52}$$

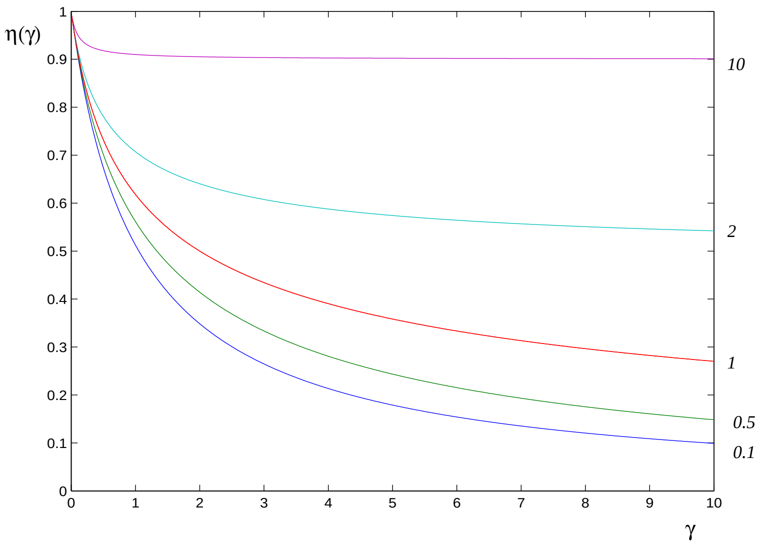

Fig. 2.3 η-transform of the Marčenko-Pastur law (1.10) evaluated for $\beta = 0.1, 0.5, 1, 2, 10$.

Example 2.11. The η-transform of the averaged empirical eigenvalue distribution of the unit-rank matrix $\mathbf{s}\mathbf{s}^\dagger$ is equal to

$$\eta(\gamma) = 1 - \frac{1}{N}\,(1 - \eta_P(\gamma)) \tag{2.53}$$

where N is the dimension of $\mathbf{s}$, and η_P is the η-transform of the random variable $\|\mathbf{s}\|^2$.

Example 2.12. The η-transform of the exponential distribution with unit mean is

$$\eta(\gamma) = -\frac{e^{\frac{1}{\gamma}}}{\gamma} E_i(-\tfrac{1}{\gamma}) \tag{2.54}$$

with $E_i(\cdot)$ denoting the exponential integral

$$E_i(z) = -\int_{-z}^{\infty} \frac{e^{-t}}{t} dt.$$

Example 2.13. Let $\mathbf{Q}$ be a $N \times K$ matrix uniformly distributed over the manifold of $N \times K$ complex matrices such that $\mathbf{Q}^\dagger \mathbf{Q} = \mathbf{I}$. Then

$$\eta_{\mathbf{Q}\mathbf{Q}^\dagger}(\gamma) = 1 - \beta + \frac{\beta}{1+\gamma}.$$

Lemma 2.28. For any $N \times K$ matrix $\mathbf{A}$ and $K \times N$ matrix $\mathbf{B}$ such that $\mathbf{AB}$ is nonnegative definite,

$$N\left(1 - \eta_{\mathsf{F}^N_{\mathbf{AB}}}(\gamma)\right) = K\left(1 - \eta_{\mathsf{F}^K_{\mathbf{BA}}}(\gamma)\right). \tag{2.55}$$

Consequently, for $K, N \to \infty$ with $\frac{K}{N} \to \beta$, if the spectra converge,

$$\eta_{\mathbf{AB}}(\gamma) = 1 - \beta + \beta \eta_{\mathbf{BA}}(\gamma). \tag{2.56}$$

Lemma 2.29.

(a) Let the components of the N-dimensional vector $\mathbf{x}$ be zero-mean and uncorrelated with second-order moment $\frac{1}{N}$. Then, for any $N \times N$ deterministic nonnegative definite matrix $\mathbf{A}$,

$$\mathbb{E}\left[\mathbf{x}^\dagger \left(\mathbf{I} + \gamma \mathbf{A}\right)^{-1} \mathbf{x}\right] = \eta_{\mathsf{F}^N_{\mathbf{A}}}(\gamma).$$

(b) [13] Let the components of the N-dimensional vector $\mathbf{x}$ be zero-mean and independent with variance $\frac{1}{N}$. For any $N \times N$ nonnegative definite random matrix $\mathbf{B}$ independent of $\mathbf{x}$ whose spectrum converges almost surely,

$$\lim_{N\to\infty} \mathbf{x}^\dagger \left(\mathbf{I} + \gamma \mathbf{B}\right)^{-1} \mathbf{x} = \eta_{\mathbf{B}}(\gamma) \quad \text{a.s.} \tag{2.57}$$

$$\lim_{N\to\infty} \mathbf{x}^\dagger \left(\mathbf{B} - z\mathbf{I}\right)^{-1} \mathbf{x} = \mathcal{S}_{\mathbf{B}}(z) \quad \text{a.s.} \tag{2.58}$$

2.2.3 Shannon Transform

Another transform motivated by applications is the following.[22]

Definition 2.12. The Shannon transform of a nonnegative random variable X is defined as

$$\mathcal{V}_X(\gamma) = \mathbb{E}[\log(1 + \gamma X)] \tag{2.59}$$

where γ is a nonnegative real number.

The Shannon transform is intimately related to the Stieltjes and η-transforms:

$$\begin{aligned}\frac{\gamma}{\log e}\frac{d}{d\gamma}\mathcal{V}_X(\gamma) &= 1 - \frac{1}{\gamma}\mathcal{S}_X\left(-\frac{1}{\gamma}\right) &(2.60)\\ &= 1 - \eta_X(\gamma). &(2.61)\end{aligned}$$

Since $\mathcal{V}_X(0) = 0$, $\mathcal{V}_X(\gamma)$ can be obtained for all $\gamma > 0$ by integrating the derivative obtained in (2.60). The Shannon transform contains the same information as the distribution of X, either through the inversion of the Stieltjes transform or from the fact that all the moments of X are obtainable from $\mathcal{V}_X(\gamma)$.

As we saw in Section 1.2, the Shannon transform of the empirical distribution of the eigenvalues of $\mathbf{HH}^\dagger$ gives the capacity of various communication channels of interest.

Example 2.14. [275] The Shannon transform of the Marčenko-Pastur law $\mathsf{f}_\beta(\cdot)$ in (1.10) is

$$\begin{aligned}\mathcal{V}(\gamma) &= \log\left(1 + \gamma - \frac{1}{4}\mathcal{F}(\gamma,\beta)\right) + \frac{1}{\beta}\log\left(1 + \gamma\beta - \frac{1}{4}\mathcal{F}(\gamma,\beta)\right)\\ &\quad -\frac{\log e}{4\,\beta\,\gamma}\mathcal{F}(\gamma,\beta). &(2.62)\end{aligned}$$

Example 2.15. [131] Denoting by $\mathcal{V}(\gamma)$ the Shannon transform of the Marčenko-Pastur law $\mathsf{f}_\beta(\cdot)$ in (1.10) with $\beta \leq 1$,

$$\lim_{\gamma\to\infty}(\log\gamma - \mathcal{V}(\gamma)) = \frac{1-\beta}{\beta}\log(1-\beta) + \log e. \tag{2.63}$$

[22] The Shannon transform was first introduced in [272, 273].

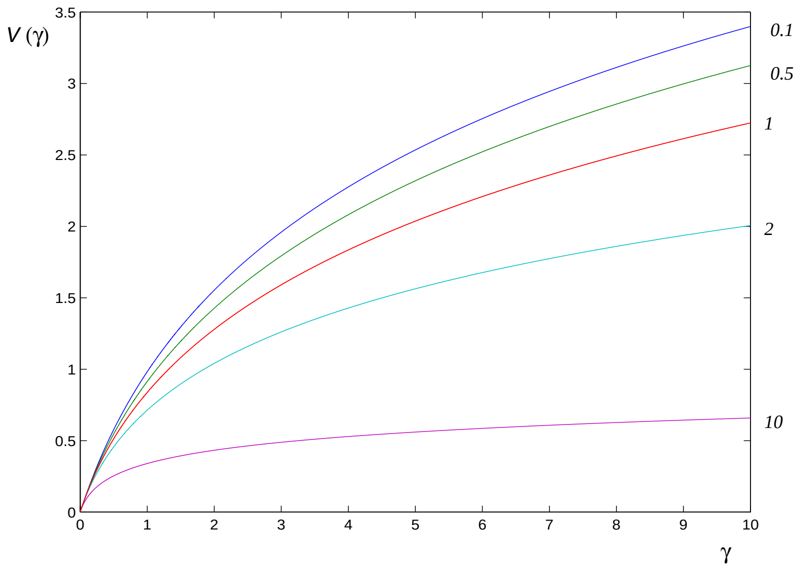

Fig. 2.4 Shannon transform of the Marčenko-Pastur law (1.10) for $\beta = 0.1, 0.5, 1, 2, 10$.

Example 2.16. The Shannon transform of the averaged empirical eigenvalue distribution of the unit-rank matrix $\mathbf{s}\mathbf{s}^\dagger$ equals

$$\mathcal{V}(\gamma) = \frac{1}{N}\mathcal{V}_P(\gamma) \tag{2.64}$$

where N is the dimension of $\mathbf{s}$ and $\mathcal{V}_P$ is the Shannon transform of the random variable $\|\mathbf{s}\|^2$.

Example 2.17. [61] The Shannon transform of $\mathsf{g}_{rt}(\cdot)$ in (2.23) is[23]

$$\mathcal{V}(\gamma) = \sum_{k=0}^{t-1}\sum_{\ell_1=0}^{k}\sum_{\ell_2=0}^{k}\binom{k}{\ell_1}\frac{(k+r-t)!(-1)^{\ell_1+\ell_2}I_{\ell_1+\ell_2+r-t}(\gamma)}{(k-\ell_2)!(r-t+\ell_1)!(r-t+\ell_2)!\ell_2!}$$

with $I_0(\gamma) = -e^{\frac{1}{\gamma}}E_i(-\frac{1}{\gamma})$ while

$$I_n(\gamma) = nI_{n-1}(\gamma) + (-\gamma)^{-n}\left(I_0(\gamma) + \sum_{k=1}^{n}(k-1)!\,(-\gamma)^k\right). \tag{2.65}$$

[23] Related expressions in terms of the exponential integral function [97] and the Gamma function can be found in [219] and [126], respectively.

An analytical expression for the Shannon transform of the marginal distribution, $\mathsf{q}_{m,n}(\cdot)$ in (2.26), of the eigenvalues of a central complex Wishart matrix $\mathbf{W} \sim W_m(n, \mathbf{\Sigma})$ with $n \geq m$ can be found in [2, 135]. For the converse case, $n \leq m$, defined in Theorem 2.19, the corresponding Shannon transform can be found in [2, 234, 135].

Example 2.18. [148] The Shannon transform of the asymptotic eigenvalue distribution of $\mathbf{\Sigma H \Phi H}^\dagger$ as defined in Theorem 2.16 is

$$\mathcal{V}(\gamma) = \frac{\log e \prod_{i=1}^{n-1} \frac{i!}{i^n}}{(-1)^{\frac{d(d-1)}{2}} \gamma^{\frac{n(n-1)}{2}}} \prod_{i<j}^{n} \frac{1}{\phi_i - \phi_j} \prod_{i<j}^{m} \frac{1}{a_i - a_j} \sum_{\ell=1}^{m} \det\left(\begin{bmatrix} \mathbf{X}_\ell \\ \mathbf{Y} \end{bmatrix} \right)$$

where $\mathbf{X}_\ell$ is a $m \times n$ matrix whose (i,j)th entry, for $i \in \{1 \ldots, m\}$ and $j \in \{1 \ldots, n\}$, is

$$(\mathbf{X}_\ell)_{i,j} = \begin{cases} -(n-1)! \frac{(\gamma\phi_j)^{n-1}}{a_i^{1-m}} e^{\frac{1}{\gamma \phi_j a_i}} E_i \left(-\frac{1}{\gamma \phi_j a_i} \right) & i = \ell \\ \sum_{k=n-m}^{n-1} \frac{(-\gamma \phi_j a_i)^k}{a_i^{n-m}} [1-n]_k & i \neq \ell \end{cases}$$

and $\mathbf{Y}$ is an $(n-m) \times n$ matrix whose (i,j)th entry, for $j \in \{1 \ldots, n\}$ and $i \in \{1 \ldots, n-m\}$, is

$$(\mathbf{Y})_{i,j} = [1-n]_{i-1} (-\gamma \phi_j)^{i-1}.$$

Example 2.19. The Shannon transform of the exponential distribution plays an important role in the capacity of fading channels and can be written in terms of its η-transform given in (2.54):

$$\mathcal{V}(\gamma) = \gamma \eta(\gamma). \tag{2.66}$$

2.2.4 Mellin Transform

The Mellin transform has been used in the non-asymptotic theory of random matrices. As we will see, it is related to the Shannon transform and can be used to find the capacity of multi-antenna channels with finite number of antennas in closed form.

Definition 2.13. The Mellin transform of a positive random variable X is given by

$$\mathcal{M}_X(z) = \mathbb{E}[X^{z-1}] \tag{2.67}$$

where z belongs to a strip of the complex plane where the expectation is finite.

The inverse Mellin transform of $\Omega(z)$ is given by

$$\mathcal{M}_{\Omega}^{-1}(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} t^{-z}\Omega(z)dz. \tag{2.68}$$

Notice that

$$\mathcal{M}_{\mathcal{M}_X}^{-1}(x) = \mathsf{f}_X(x)$$

with $\mathsf{f}_X(\cdot)$ denoting the p.d.f. of X.

Another interesting property of the Mellin transform is that the Mellin transform of the product of two independent random variables is equal to the product of the Mellin transforms:

$$\mathcal{M}_{XY} = \mathcal{M}_X\mathcal{M}_Y. \tag{2.69}$$

Example 2.20. If X is exponentially distributed with mean $\frac{1}{\mu}$, then

$$\mathcal{M}_X(z) = \mu^{1-z}\Gamma(z).$$

Example 2.21. If X is Nakagami distributed with parameter ν, $f_\nu(r) = \frac{2\nu^\nu}{\Gamma(\nu)}r^{2\nu-1}e^{-\nu r^2}$, then for $1-z<\nu$

$$\mathcal{M}_{X^2}(z) = \frac{\nu^{1-z}}{\Gamma(\nu)}\Gamma(\nu+z-1).$$

Example 2.22. [126] The Mellin transform of $\mathsf{g}_{r,r}(\cdot)$ in (2.23) is

$$\mathcal{M}_{\mathsf{g}_{r,r}}(1-z) = \frac{r^{-1}}{\Gamma(z)\,\Gamma(1-z)}\sum_{n=0}^{r-1}\frac{\Gamma^2(1-z+n)}{(n!)^2}\sum_{\ell=0}^{r-1-n}\frac{\Gamma(z+\ell)}{\ell!}.$$

Example 2.23. The Mellin transform of $\mathsf{q}_{m,n}(\cdot)$ in (2.26) is

$$\mathcal{M}_{\mathsf{q}_{m,n}}(z) = \frac{m^{-1}}{\det\mathbf{\Sigma}^n}\prod_{k<\ell}^{m}\frac{a_k a_\ell}{a_k-a_\ell}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\mathcal{D}(i,j)\,\Gamma(z+n-m+j-1)}{a_i^{m-z-n-j+1}\prod_{\ell=1}^{m}(n-\ell)!}$$

with $\mathcal{D}(\cdot,\cdot)$ given in (2.27).

Theorem 2.30. [126]

$$\mathcal{V}_X(\gamma) = \mathcal{M}_\Upsilon^{-1}(\gamma) \tag{2.70}$$

where $\mathcal{M}_\Upsilon^{-1}$ is the inverse Mellin transform of

$$\Upsilon(z) = z^{-1}\Gamma(z)\Gamma(1-z)\mathcal{M}_X(1-z). \tag{2.71}$$

Using Theorem 2.30, an explicit expression for the Shannon transform of $\mathsf{g}_{r,r}(\cdot)$ in (2.23) has been derived in [126].

2.2.5 R-transform

Another handy transform, on which we elaborate next, is the R-transform. In particular, as we shall see in detail in Section 2.4 once the concept of asymptotic freeness has been introduced, the R-transform enables the characterization of the asymptotic spectrum of a sum of suitable matrices (such as independent unitarily invariant matrices) from their individual asymptotic spectra.

Definition 2.14. [285] Let $\mathcal{S}_X^{-1}(z)$ denote the inverse (with respect to the composition of functions) of the Stieltjes transform of X, i. e., $z = \mathcal{S}_X^{-1}(\mathcal{S}_X(z))$. The R-transform of X is defined as the complex-valued function of complex argument

$$\mathsf{R}_X(z) = \mathcal{S}_X^{-1}(-z) - \frac{1}{z}. \tag{2.72}$$

As a consequence of (2.72), a direct relationship between the R-transform and the Stieltjes transform exists, namely

$$s = \frac{1}{\mathsf{R}_X(-s) - z} \tag{2.73}$$

where for notational simplicity we used $s = \mathcal{S}_X(z)$. For positive random variables, letting $z = -\frac{1}{\gamma}$ in (2.73), we obtain from (2.48) the following relationship between the R-transform and the η-transform:

$$\eta_X(\gamma) = \frac{1}{1 + \gamma\,\mathsf{R}_X(-\gamma\,\eta_X(\gamma))}. \tag{2.74}$$

A consequence of (2.74) is that the R-transform (restricted to the negative real axis) can be equivalently defined as

$$\mathsf{R}_X(\varphi) = \frac{\eta_X(\gamma) - 1}{\varphi} \tag{2.75}$$

with γ and φ satisfying

$$\varphi = -\gamma\,\eta_X(\gamma). \tag{2.76}$$

Example 2.24. The R-transform of a unit mass at a is

$$\mathsf{R}(z) = a. \tag{2.77}$$

Example 2.25. The R-transform of the semicircle law is

$$\mathsf{R}(z) = z. \tag{2.78}$$

Example 2.26. The R-transform of the Marčenko-Pastur law $\mathsf{f}_\beta(\cdot)$ in (1.10) is

$$\mathsf{R}(z) = \frac{1}{1 - \beta z}. \tag{2.79}$$

Example 2.27. The R-transform of $\tilde{\mathsf{f}}_\beta(\cdot)$ in (1.12) is

$$\mathsf{R}(z) = \frac{\beta}{1 - z}. \tag{2.80}$$

Example 2.28. The R-transform of the averaged empirical eigenvalue distribution of the N-dimensional unit-rank matrix $\mathbf{s}\mathbf{s}^\dagger$ such that $\|\mathbf{s}\|^2$ has η-transform η_P, satisfies the implicit equation

$$\mathsf{R}\left(\frac{\gamma}{N} - \frac{\gamma}{N}\eta_P(\gamma) - \gamma\right) = \frac{1}{\gamma}\frac{1 - \eta_P(\gamma)}{N - 1 + \eta_P(\gamma)}. \tag{2.81}$$

In the special case where the norm is deterministic, $\|\mathbf{s}\|^2 = c$,

$$\eta_P(\gamma) = \frac{1}{1 + \gamma c},$$

and an explicit expression for the R-transform can be obtained from (2.81) as

$$\begin{aligned}\mathsf{R}(z) &= \frac{-1 + cz + \sqrt{\frac{4cz}{N} + (1 - cz)^2}}{2z} \\ &= \frac{c}{(1 - cz)N} + O(N^{-2}).\end{aligned} \tag{2.82}$$

Theorem 2.31. For any $a > 0$,

$$\mathsf{R}_{aX}(z) = a\mathsf{R}_X(az). \tag{2.83}$$

We now outline how to obtain the moments of X from $\mathsf{R}_X(z)$. When the random variable X is compactly supported, the R-transform can be represented as a series (for those values in the region of convergence):

$$\mathsf{R}_X(z) = \sum_{k=1}^{\infty} c_k z^{k-1} \tag{2.84}$$

where the coefficients c_k, called the free cumulants of X, play a role akin to that of the classical cumulants. As in the classical case, the coefficients c_k are polynomial functions of the moments $\mathbb{E}[X^p]$ with $0 \le p \le k$. Given the free cumulants c_k, the moments of X can be obtained by the so-called free cumulant formula [241]

$$\mathbb{E}[X^m] = \sum_{k=1}^{m} c_k \sum_{m_1+\cdots+m_k=m} \mathbb{E}\left[X^{m_1-1}\right] \cdots \mathbb{E}\left[X^{m_k-1}\right]. \tag{2.85}$$

Note that $c_1 = \mathbb{E}[X]$, $c_2 = \mathrm{Var}(X)$, and $\mathsf{R}_X(0) = \mathbb{E}[X]$.

As hinted at the beginning of this section, the main usefulness of the R-transform stems from Theorem 2.192 stating that, for an important class of random matrices, the R-transform of the asymptotic spectrum of the sum is the sum of R-transforms of the individual spectra.

2.2.6 S-transform

Definition 2.15. The S-transform of a nonnegative random variable X is[24]

$$\Sigma_X(x) = -\frac{x+1}{x}\, \eta_X^{-1}(1+x), \tag{2.86}$$

which maps $(-1, 0)$ onto the positive real line.

[24] A less compact definition of the S-transform on the complex plane is given in the literature (since the η-transform had not been used before) for arbitrary random variables with nonzero mean. Note that the restriction to nonnegative random variables stems from the definition of the η-transform.

Example 2.29. The S-transform of the Marčenko-Pastur law $\mathsf{f}_\beta(\cdot)$ in (1.10) is

$$\Sigma(x) = \frac{1}{1+\beta x}. \tag{2.87}$$

Example 2.30. The S-transform of $\tilde{\mathsf{f}}_\beta(\cdot)$ in (1.12) is

$$\Sigma(x) = \frac{1}{\beta + x}. \tag{2.88}$$

Example 2.31. The S-transform of the averaged empirical eigenvalue distribution of the N-dimensional unit-rank matrix $\mathbf{s}\mathbf{s}^\dagger$ such that $\|\mathbf{s}\|^2 = c$ is equal to

$$\Sigma(x) = \frac{1+x}{c\,(x+1/N)}. \tag{2.89}$$

The S-transform was introduced by Voiculescu [286] in 1987. As we will see, its main usefulness lies in the fact that the S-transform of the product of certain random matrices is the product of the corresponding S-transforms in the limit.

From (2.56), we obtain

$$\eta_{\mathbf{AB}}^{-1}(\gamma) = \eta_{\mathbf{BA}}^{-1}\left(\frac{\gamma-1}{\beta} + 1\right) \tag{2.90}$$

and hence the S-transform counterpart to (2.56):

Theorem 2.32. For any $N \times K$ matrix $\mathbf{A}$ and $K \times N$ matrix $\mathbf{B}$ such that, as $K, N \to \infty$ with $\frac{K}{N} \to \beta$, the spectra converge while $\mathbf{AB}$ is nonnegative definite,

$$\Sigma_{\mathbf{AB}}(x) = \frac{x+1}{x+\beta}\Sigma_{\mathbf{BA}}\left(\frac{x}{\beta}\right). \tag{2.91}$$

Example 2.32. Let $\mathbf{Q}$ be a $N \times K$ matrix uniformly distributed over the manifold of $N \times K$ complex matrices such that $\mathbf{Q}^\dagger\mathbf{Q} = \mathbf{I}$. Then

$$\Sigma_{\mathbf{Q}\mathbf{Q}^\dagger}(x) = \frac{1+x}{\beta+x}. \tag{2.92}$$

2.3 Asymptotic Spectrum Theorems

In this section, we give the main results on the limit of the empirical distributions of the eigenvalues of various random matrices of interest. For pedagogical purposes we will give results in increasing level of generality.

2.3.1 The Semicircle Law

Theorem 2.33. [308, 305] Consider an $N \times N$ standard Wigner matrix $\mathbf{W}$ such that, for some constant κ, and sufficiently large N

$$\max_{1 \le i \le j \le N} \mathbb{E}\left[|\mathsf{W}_{i,j}|^4\right] \le \frac{\kappa}{N^2}. \tag{2.93}$$

Then, the empirical distribution of $\mathbf{W}$ converges almost surely to the semicircle law whose density is

$$w(x) = \frac{1}{2\pi}\sqrt{4 - x^2} \tag{2.94}$$

with $|x| \le 2$.

Wigner's original proof [305] of the convergence to the semicircle law consisted of showing convergence of the empirical moments $\frac{1}{N}\mathrm{tr}\left\{\mathbf{W}^{2k}\right\}$ to the even moments of the semicircle law, namely, the Catalan numbers:

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N}\mathrm{tr}\left\{\mathbf{W}^{2k}\right\} &= \int_{-2}^{2} x^{2k} w(x)\, dx \\ &= \frac{1}{k+1}\binom{2k}{k}. \end{aligned} \tag{2.95}$$

The zero-mean assumption in the definition of a Wigner matrix can be relaxed to an identical-mean condition using Lemma 2.23. In fact, it suffices that the rank of the mean matrix does not grow linearly with N for Theorem 2.33 to hold.

Assuming for simplicity that the diagonal elements of the Wigner matrix are zero, we can give a simple sketch of the proof of Theorem 2.33 based on the matrix inversion lemma:

$$(\mathbf{A}^{-1})_{i,i} = \frac{1}{\mathsf{A}_{i,i} - \mathbf{a}_i^\dagger \mathbf{A}_i^{-1}\mathbf{a}_i} \tag{2.96}$$

with $\mathbf{a}_i$ representing the ith column of $\mathbf{A}$ excluding the i-element and $\mathbf{A}_i$ indicating the $(n-1)\times(n-1)$ submatrix obtained by eliminating from $\mathbf{A}$ the ith column and the ith row. Thus

$$\frac{1}{N}\,\mathrm{tr}\left\{(-z\mathbf{I}+\mathbf{W})^{-1}\right\} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{-z-\mathbf{w}_i^{\dagger}\left(-z\mathbf{I}+\mathbf{W}_i\right)^{-1}\mathbf{w}_i}. \tag{2.97}$$

Moreover, $\mathbf{W}_i$ is independent of $\mathbf{w}_i$, whose entries are independent with identical variance $\frac{1}{n}$. Then, taking the limit of (2.97) and applying (2.58) to the right-hand side, we obtain the quadratic equation

$$\mathcal{S}_{\mathbf{W}}(z) = \frac{1}{-z-\mathcal{S}_{\mathbf{W}}(z)}$$

which admits the closed-form solution given in (2.41).

Condition (2.93) on the entries of $\sqrt{N}\,\mathbf{W}$ can be replaced by the Lindeberg-type condition on the whole matrix [10, Thm. 2.4]:

$$\frac{1}{N}\sum_{i,j}\mathbb{E}\left[|\mathsf{W}_{i,j}|^2\, 1\left\{|\mathsf{W}_{i,j}|\geq\delta\right\}\right]\to 0 \tag{2.98}$$

for any $\delta>0$.

2.3.2 The Full-Circle Law

Theorem 2.34. [173, 197, 85, 68, 9] Let $\mathbf{H}$ be an $N\times N$ complex random matrix whose entries are independent random variables with identical mean, variance $\frac{1}{N}$ and finite kth moments for $k\geq 4$. Assume that the joint distributions of the real and imaginary parts of the entries have uniformly bounded densities. Then, the asymptotic spectrum of $\mathbf{H}$ converges almost surely to the circular law, namely the uniform distribution over the unit disk on the complex plane $\{\zeta\in\mathbb{C}:|\zeta|\leq 1\}$ whose density is given by

$$\mathsf{f}_c(\zeta)=\frac{1}{\pi}\qquad |\zeta|\leq 1. \tag{2.99}$$

Theorem 2.34 also holds for real matrices replacing the assumption on the joint distribution of real and imaginary parts with the one-dimensional distribution of the real-valued entries.

2.3.3 The Marčenko-Pastur Law and its Generalizations

Theorem 2.35. [170, 296, 131, 10] Consider an $N \times K$ matrix $\mathbf{H}$ whose entries are independent zero-mean complex (or real) random variables with variance $\frac{1}{N}$ and fourth moments of order $O(\frac{1}{N^2})$. As $K, N \to \infty$ with $\frac{K}{N} \to \beta$, the empirical distribution of $\mathbf{H}^\dagger \mathbf{H}$ converges almost surely to a nonrandom limiting distribution with density

$$\mathsf{f}_\beta(x) = \left(1 - \frac{1}{\beta}\right)^+ \delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi\beta x} \tag{2.100}$$

where

$$a = (1-\sqrt{\beta})^2 \qquad\qquad b = (1+\sqrt{\beta})^2.$$

The above limiting distribution is the Marčenko-Pastur law with ratio index β. Using Lemma 2.22, the zero-mean condition can be relaxed to having identical mean. The condition on the fourth moments can be relaxed [10, Thm. 2.8] to a Lindeberg-type condition:

$$\frac{1}{K}\sum_{i,j} \mathbb{E}\left[|\mathsf{H}_{i,j}|^2 \, 1\{|\mathsf{H}_{i,j}| \geq \delta\}\right] \to 0 \tag{2.101}$$

for any $\delta > 0$.

Using (1.3) and (2.100), the empirical distribution of $\mathbf{H}\mathbf{H}^\dagger$, with $\mathbf{H}$ as in Theorem 2.35, converges almost surely to a nonrandom limiting distribution with density (1.12) whose moments are given by

$$\int_a^b x^k \tilde{\mathsf{f}}_\beta(x)\, dx = \sum_{i=1}^{k} \frac{1}{k}\binom{k}{i}\binom{k}{i-1}\beta^i \tag{2.102}$$

$$= \lim_{N\to\infty} \frac{1}{N}\mathrm{tr}\left\{(\mathbf{H}\mathbf{H}^\dagger)^k\right\}. \tag{2.103}$$

Furthermore, from Lemma 2.10, it follows straightforwardly that the first and second order asymptotic moments of $(\mathbf{H}\mathbf{H}^\dagger)^{-1}$ with $\beta > 1$ converge to

$$\lim_{N\to\infty} \frac{1}{N}\mathrm{tr}\left\{(\mathbf{H}\mathbf{H}^\dagger)^{-1}\right\} = \frac{1}{\beta - 1} \tag{2.104}$$

$$\lim_{N\to\infty} \frac{1}{N}\mathrm{tr}\left\{(\mathbf{H}\mathbf{H}^\dagger)^{-2}\right\} = \frac{\beta}{(\beta-1)^3}. \tag{2.105}$$

The convergence in (2.103)–(2.105) is almost surely. If $\mathbf{H}$ is square, then the empirical distribution of its singular values converges almost surely to the quarter circle law with density $q(\cdot)$ given in (1.21). The even moments of the quarter circle law coincide with the corresponding moments of the semicircle law. Unlike those of the semicircle law, the odd moments of the quarter circle law do not vanish. For all positive integers k the moments of the quarter circle law are given by

$$\int_0^2 x^k q(x) dx = \frac{2^k}{\sqrt{\pi}} \frac{\Gamma(\frac{1+k}{2})}{\Gamma(2+\frac{k}{2})}. \tag{2.106}$$

In the important special case of square $\mathbf{H}$ with independent Gaussian entries, the speed at which the minimum singular value vanishes (and consequently the growth of the condition number) is characterized by the following result.

Theorem 2.36. [67, Thm. 5.1],[218] Consider an $N \times N$ standard complex Gaussian matrix $\mathbf{H}$. The minimum singular value of $\mathbf{H}$, $\sigma_{\min}$, satisfies

$$\lim_{N\to\infty} P[N\sigma_{\min} \geq x] = e^{-x-x^2/2}. \tag{2.107}$$

A summary of related results for both the minimum and maximum singular values of $\mathbf{H}$ can be found in [67, 10].

The following theorem establishes a link between asymptotic random matrix theory and recent results on the asymptotic distribution of the zeros of classical orthogonal polynomials.

Theorem 2.37. [57] Let $\lambda_1 \leq \ldots \leq \lambda_K$ denote the ordered eigenvalues of $\mathbf{H}^\dagger\mathbf{H}$ with $\mathbf{H}$ an $N \times K$ standard complex Gaussian matrix and let $x_1 \leq \ldots \leq x_K$ denote the zeros of the Laguerre polynomial $L_K^{N-K+1}(Nx)$. If $K, N \to \infty$ with $\frac{K}{N} \to \beta \in (0,\infty)$, then almost surely

$$\frac{1}{K}\sum_{\ell=1}^{K} |\lambda_\ell - x_\ell|^2 \stackrel{a.s}{\to} 0. \tag{2.108}$$

Moreover, if $d_1 \leq d_2 \leq \ldots \leq d_K$ denote the ordered differences $|\lambda_i - x_i|$, then

$$d_{\lfloor yK \rfloor} \stackrel{a.s}{\to} 0 \tag{2.109}$$

for all $y \in (0,1)$. For the smallest and largest eigenvalues of $\mathbf{H}^\dagger\mathbf{H}$, and for the smallest and largest zero of the polynomial $L_K^{N-K+1}(Nx)$, we have that almost surely

$$\lim_{K\to\infty} x_1 = \lim_{K\to\infty} \lambda_1 = (1-\sqrt{\beta})^2 \tag{2.110}$$

$$\lim_{K\to\infty} x_K = \lim_{K\to\infty} \lambda_K = (1+\sqrt{\beta})^2 \tag{2.111}$$

for $\beta \leq 1$ while, for $\beta > 1$,

$$\lim_{K\to\infty} x_{K-N+1} = \lim_{K\to\infty} \lambda_{K-N+1} = (1-\sqrt{\beta})^2. \tag{2.112}$$

Theorem 2.37 in conjunction with recent results on the asymptotic distribution of the zeros of scaled generalized Laguerre polynomials, $L_K^{N-K+1}(Nx)$, also provides an alternative proof of the semicircle and Marčenko-Pastur laws.

In [57], using results on the asymptotics of classical orthogonal polynomials, results analogous to Theorem 2.37 are also derived for centered sample covariance matrices

$$\sqrt{\frac{N}{K}}\left(\mathbf{H}^\dagger\mathbf{H} - \kappa\mathbf{I}\right) \tag{2.113}$$

with $\kappa = \max\left\{1, \frac{K}{N}\right\}$. For such matrices, it is proved that if $K, N \to \infty$ with $\frac{K}{N} \to \infty$ or with $\frac{K}{N} \to 0$, the extremal eigenvalues converge almost surely to 2 and -2, while the corresponding eigenvalue distribution converges to the semicircle law (cf. Example 2.50).

Theorem 2.38. [170, 227] Let $\mathbf{H}$ be an $N \times K$ matrix whose entries are i.i.d. complex random variables with zero-mean and variance $\frac{1}{N}$. Let $\mathbf{T}$ 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 $\mathcal{S}_0$. If $\mathbf{H}$, $\mathbf{T}$, and $\mathbf{W}_0$ are independent, the empirical eigenvalue distribution of

$$\mathbf{W} = \mathbf{W}_0 + \mathbf{H}\mathbf{T}\mathbf{H}^\dagger \tag{2.114}$$

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

$$\mathcal{S}(z) = \mathcal{S}_0 \left(z - \beta \, \mathbb{E} \left[\frac{\mathsf{T}}{1 + \mathsf{T}\mathcal{S}(z)} \right] \right). \tag{2.115}$$

The case $\mathbf{W}_0 = \mathbf{0}$ merits particular attention. Using the more convenient η-transform and Shannon transform, we derive the following result from [226]. (The proof is given in Appendix 4.1 under stronger assumptions on $\mathbf{T}$.)

Theorem 2.39. Let $\mathbf{H}$ be an $N \times K$ matrix whose entries are i.i.d. complex random variables with variance $\frac{1}{N}$. Let $\mathbf{T}$ be a $K \times K$ Hermitian nonnegative random matrix, independent of $\mathbf{H}$, whose empirical eigenvalue distribution converges almost surely to a nonrandom limit. The empirical eigenvalue distribution of $\mathbf{HTH}^\dagger$ converges almost surely, as $K, N \to \infty$ with $\frac{K}{N} \to \beta$, to a distribution whose η-transform satisfies

$$\beta = \frac{1 - \eta}{1 - \eta_{\mathbf{T}}(\gamma\eta)} \tag{2.116}$$

where for notational simplicity we have abbreviated $\eta_{\mathbf{HTH}^\dagger}(\gamma) = \eta$. The corresponding Shannon transform satisfies[25]

$$\mathcal{V}_{\mathbf{HTH}^\dagger}(\gamma) = \beta \mathcal{V}_{\mathbf{T}}(\eta\gamma) + \log \frac{1}{\eta} + (\eta - 1) \log e. \tag{2.117}$$

The condition of i.i.d. entries can be relaxed to independent entries with common mean and variance $\frac{1}{N}$ satisfying the Lindeberg-type condition (2.101). The mth moment of the empirical distribution of $\mathbf{HTH}^\dagger$ converges almost surely to [313, 116, 158]:

$$\sum_{i=1}^{m} \beta^i \sum_{\substack{m_1 + \cdots + m_i = m \\ m_1 \le \cdots \le m_i}} \frac{m!}{(m - i + 1)! f(m_1, \dots m_i)} \mathbb{E}[\mathsf{T}^{m_1}] \dots \mathbb{E}[\mathsf{T}^{m_i}] \tag{2.118}$$

where T is a random variable with distribution equal to the asymptotic spectrum of $\mathbf{T}$ and, $\forall \; 1 \le \ell \le m$,

$$f(i_1, \dots, i_\ell) = f_1! \cdots f_m! \tag{2.119}$$

[25] The derivation of (2.117) from (2.116) is given in Section 3.1.2.

with f_i the number of entries of the vector $[i_1, \ldots, i_\ell]$ equal to i.[26]

Figure 2.5 depicts the Shannon transform of $\mathbf{HTH}^\dagger$ given in (2.117) for $\beta = \frac{2}{3}$ and T exponentially distributed.

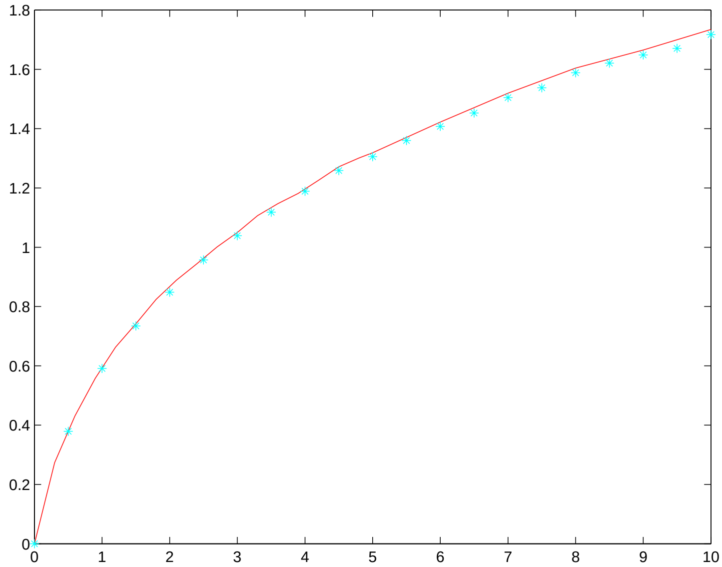

Fig. 2.5 Shannon transform of the asymptotic spectrum of $\mathbf{HTH}^\dagger$ for $\beta = \frac{2}{3}$ and T exponentially distributed. The stars indicate the Shannon transform, obtained via Monte Carlo simulation, of the averaged empirical distribution of the eigenvalues of $\mathbf{HTH}^\dagger$ where $\mathbf{H}$ is 3×2.

If $\mathbf{T} = \mathbf{I}$, then $\eta_{\mathbf{T}}(\gamma) = \frac{1}{1+\gamma}$, and (2.116) becomes

$$\eta = 1 - \beta + \frac{\beta}{1 + \gamma\eta} \tag{2.120}$$

whose explicit solution is the η-transform of the Marčenko-Pastur distribution, $\tilde{\mathsf{f}}_\beta(\cdot)$, in (1.12):

$$\eta(\gamma) = 1 - \frac{\mathcal{F}(\gamma, \beta)}{4\,\gamma}. \tag{2.121}$$

Equation (2.116) admits an explicit solution in a few other cases, one of which is illustrated by the result that follows.

[26] For example, $f(1, 1, 4, 2, 1, 2) = 3! \cdot 2! \cdot 1!$.

Theorem 2.40. [223] If, in Theorem 2.39, $\mathbf{T} = (\mathbf{Y}\mathbf{Y}^\dagger)^{-1}$ with $\mathbf{Y}$ a $K \times m$ ($K \le m$) Gaussian random matrix whose entries have zero-mean and variance $\frac{1}{m}$, then, using (2.121),[27]

$$\eta_{\mathbf{T}}(\gamma) = \frac{\gamma}{4\tilde{\beta}} \mathcal{F}\left(\frac{1}{\gamma}, \tilde{\beta}\right) \tag{2.122}$$

where $\frac{K}{m} \to \tilde{\beta}$. Thus, solving (2.116) we find that the asymptotic spectrum of $\mathbf{W} = \mathbf{H}(\mathbf{Y}\mathbf{Y}^\dagger)^{-1}\mathbf{H}^\dagger$ is given by

$$\mathsf{f}_{\mathbf{W}}(x) = \left(1 - \frac{1}{\beta}\right)^+ \delta(x) \; + \; \frac{(1-\tilde{\beta})\sqrt{(x-a^2)^+ \, (b^2 - x)^+}}{2\pi x (x\tilde{\beta} + \beta)} \tag{2.123}$$

with

$$a = \frac{1 - \sqrt{1 - (1-\beta)(1-\tilde{\beta})}}{1 - \tilde{\beta}} \qquad b = \frac{1 + \sqrt{1 - (1-\beta)(1-\tilde{\beta})}}{1 - \tilde{\beta}}.$$

Using (2.56) and (2.116), we can give an equivalent expression for the η-transform of the asymptotic spectrum of $\mathbf{T}^{1/2}\mathbf{H}^\dagger\mathbf{H}\mathbf{T}^{1/2}$:

$$\eta_{\mathbf{T}}(\gamma(1 - \beta + \beta\eta)) = \eta \tag{2.124}$$

where $\eta = \eta_{\mathbf{T}^{1/2}\mathbf{H}^\dagger\mathbf{H}\mathbf{T}^{1/2}}(\gamma)$. Note that, as $\beta \to 0$,

$$\eta_{\mathbf{T}^{1/2}\mathbf{H}^\dagger\mathbf{H}\mathbf{T}^{1/2}}(\gamma) \to \eta_{\mathbf{T}}(\gamma) \tag{2.125}$$

and thus the spectrum of $\mathbf{T}^{1/2}\mathbf{H}^\dagger\mathbf{H}\mathbf{T}^{1/2}$ converges to that of $\mathbf{T}$.

Theorem 2.41. [178] Let $\mathbf{\Sigma}$ be a positive definite matrix whose asymptotic spectrum has the p.d.f.

$$\mathsf{f}_{\mathbf{\Sigma}}(\lambda) = \frac{1}{2\pi\mu\lambda^2}\sqrt{\left(\frac{\lambda}{\sigma_1} - 1\right)\left(1 - \frac{\lambda}{\sigma_2}\right)} \tag{2.126}$$

with $\sigma_1 \le \lambda \le \sigma_2$ and

$$\mu = \frac{(\sqrt{\sigma_2} - \sqrt{\sigma_1})^2}{4\sigma_1\sigma_2}. \tag{2.127}$$

[27] Although [223] obtained (2.123) with the condition that $\mathbf{Y}$ be Gaussian, it follows from (2.121) and Theorem 2.39 that this condition is not required for (2.122) and (2.123) to hold.

If $\mathbf{H}$ is an $N \times K$ standard complex Gaussian matrix, then, as $K, N \to \infty$ with $\frac{K}{N} \to \beta$, the asymptotic spectrum of $\mathbf{W} = \mathbf{\Sigma}^{1/2}\mathbf{H}\mathbf{H}^\dagger\mathbf{\Sigma}^{1/2}$ has the p.d.f.[28]

$$\mathsf{f}_{\mathbf{W}}(\lambda) = (1-\beta)^+\delta(\lambda) + \frac{\sqrt{(\lambda-a)^+(b-\lambda)^+}}{2\pi\lambda(1+\lambda\mu)} \tag{2.128}$$

with

$$a = 1+\beta+2\mu\beta - 2\sqrt{\beta}\sqrt{(1+\mu)(1+\mu\beta)} \tag{2.129}$$

$$b = 1+\beta+2\mu\beta + 2\sqrt{\beta}\sqrt{(1+\mu)(1+\mu\beta)}. \tag{2.130}$$

The Shannon transform of (2.128) is

$$\begin{aligned}\mathcal{V}_{\mathbf{W}}(\gamma) &= \log(\gamma\omega_1(\gamma,\beta,\mu)) + \frac{1}{\mu}\log|1-\mu\omega_2(\gamma,\beta,\mu)| \\ &\quad -(\beta-1)\log|\omega_3(\gamma,\beta,\mu)| \end{aligned} \tag{2.131}$$

with

$$\omega_1(\gamma,\beta,\mu) = \frac{(1+(1+\beta)\mu)[1+\gamma(1+\beta)+\sqrt{\omega_4}]-2\mu\beta(\gamma-\mu)}{2\beta\gamma[1+(1+\beta)\mu+\beta\mu^2]}$$

$$\omega_2(\gamma,\beta,\mu) = \frac{\beta+\gamma(1+\beta)-\sqrt{\omega_4}+2\gamma\beta\mu}{2\gamma[1+(1+\beta)\mu+\beta\mu^2]}$$

$$\omega_3(\gamma,\beta,\mu) = \begin{cases} \dfrac{1+(1-\beta)\gamma+2\mu\beta-\sqrt{\omega_4}}{2\beta(\gamma-\mu)} & \text{if } \gamma \neq \mu, \\ -\dfrac{(1+\gamma\beta)}{1+(1+\beta)\gamma} & \text{if } \gamma = \mu \end{cases}$$

$$\omega_4 = (1+(1+\beta)\gamma)^2 - 4\beta\gamma(\gamma-\mu).$$

Returning to the setting of Theorem 2.38 but interchanging the assumptions on $\mathbf{W}_0$ and $\mathbf{T}$, i.e., with $\mathbf{W}_0$ diagonal and $\mathbf{T}$ Hermitian, the result that follows (proved in Appendix 4.2) states that the asymptotic spectrum in Theorem 2.38 still holds under the condition that $\mathbf{W}_0$ and $\mathbf{T}$ be nonnegative definite. Consistent with our emphasis, this result is formulated in terms of the η-transform rather than the Stieltjes transform used in Theorem 2.38.

[28] Theorem 2.39 indicates that (2.128) holds even without the Gaussian condition on $\mathbf{H}$.

Theorem 2.42. Let $\mathbf{H}$ be an $N \times K$ matrix whose entries are i.i.d. complex random variables with zero-mean and variance $\frac{1}{N}$. Let $\mathbf{T}$ be a $K \times K$ positive definite random matrix whose empirical eigenvalue distribution converges almost surely to a nonrandom limit. Let $\mathbf{W}_0$ be an $N \times N$ nonnegative definite diagonal random matrix with empirical eigenvalue distribution converging almost surely to a nonrandom limit. Assuming that $\mathbf{H}$, $\mathbf{T}$, and $\mathbf{W}_0$ are independent, the empirical eigenvalue distribution of

$$\mathbf{W} = \mathbf{W}_0 + \mathbf{H}\mathbf{T}\mathbf{H}^\dagger \tag{2.132}$$

converges almost surely, as $K, N \to \infty$ with $\frac{K}{N} \to \beta$, to a nonrandom limiting distribution whose η-transform is the solution of the following pair of equations:

$$\gamma\,\eta = \varphi\,\eta_0\,(\varphi) \tag{2.133}$$

$$\eta = \eta_0\,(\varphi) - \beta\,(1 - \eta_{\mathbf{T}}(\gamma\,\eta)) \tag{2.134}$$

with η_0 and $\eta_{\mathbf{T}}$ the η-transforms of $\mathbf{W}_0$ and $\mathbf{T}$ respectively.

Notice that the function $\eta(\gamma)$ can be immediately evaluated from (2.133) and (2.134) since every $\varphi \in (0, \infty)$ determines a pair of values $(\gamma, \eta(\gamma)) \in (0, \infty) \times [0, 1]$: the product $(\gamma\,\eta)$ is obtained from (2.133) (which is strictly monotonically increasing in φ), then η is obtained from (2.134) and, finally, $\gamma = \frac{(\gamma\,\eta)}{\eta}$.

Figure 2.6 shows the η-transform of $\mathbf{W} = \mathbf{H}\mathbf{T}\mathbf{H}^\dagger$ where the asymptotic spectrum of $\mathbf{T}$ converges almost surely to an exponential distribution.

Theorem 2.43. [86, 55, 159] Define $\mathbf{H} = \mathbf{CSA}$ where $\mathbf{S}$ is an $N \times K$ matrix whose entries are independent complex random variables (arbitrarily distributed) satisfying the Lindeberg condition (2.101) with identical means and 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 $\mathbf{D} = \mathbf{C}\mathbf{C}^\dagger$ and $\mathbf{T} = \mathbf{A}\mathbf{A}^\dagger$ converge almost surely to compactly supported measures.[29] If $\mathbf{C}$, $\mathbf{A}$ and $\mathbf{S}$ are independent, as $K, N \to \infty$ with $\frac{K}{N} \to \beta$,

[29] In the case that $\mathbf{C}$ and $\mathbf{A}$ are diagonal deterministic matrices, Theorem 2.43 is a special case of Theorem 2.50.

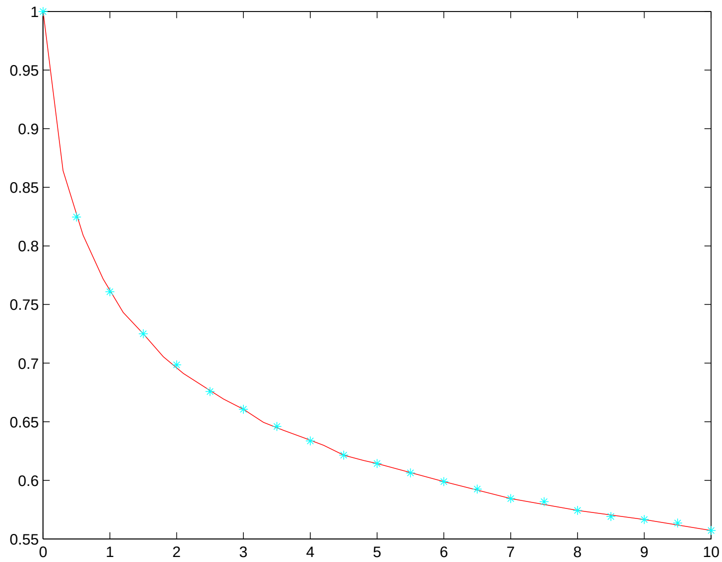

Fig. 2.6 η-transform of $\mathbf{HTH}^\dagger$ with $\beta = \frac{2}{3}$ and $\eta_{\mathbf{T}}$ given by (2.54). The stars indicate the η-transform of the averaged empirical spectrum of $\mathbf{HTH}^\dagger$ for a 3×2 matrix $\mathbf{H}$.

the η-transform of $\mathbf{HH}^\dagger$ is

$$\eta_{\mathbf{HH}^\dagger}(\gamma) = \mathbb{E}\,[\,\Gamma_{\mathbf{HH}^\dagger}(\mathsf{D},\gamma)\,] \tag{2.135}$$

where $\Gamma_{\mathbf{HH}^\dagger}(d,\gamma)$ satisfies

$$\Gamma_{\mathbf{HH}^\dagger}(d,\gamma) = \frac{1}{1+\gamma\,\beta\,d\,\mathbb{E}\left[\frac{\mathsf{T}}{1+\gamma\,\mathsf{T}\,\mathbb{E}[\mathsf{D}\,\Gamma_{\mathbf{HH}^\dagger}(\mathsf{D},\gamma)]}\right]} \tag{2.136}$$

with D and T independent random variables whose distributions are the asymptotic spectra of $\mathbf{D}$ and $\mathbf{T}$ respectively. The asymptotic fraction of zero eigenvalues of $\mathbf{HH}^\dagger$ equals

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

The following result, proved in Appendix 4.3, finds the Shannon transform of $\mathbf{HH}^\dagger$ in terms of the Shannon transforms of $\mathbf{D}$ and $\mathbf{T}$.

Theorem 2.44. Let $\mathbf{H}$ be an $N \times K$ matrix as defined in Theorem 2.43. The Shannon transform of $\mathbf{HH}^\dagger$ is given by:

$$\mathcal{V}_{\mathbf{HH}^\dagger}(\gamma) = \mathcal{V}_{\mathbf{D}}(\beta\gamma_d) + \beta\mathcal{V}_{\mathbf{T}}(\gamma_{\mathsf{t}}) - \beta\frac{\gamma_d\gamma_{\mathsf{t}}}{\gamma}\log e \tag{2.137}$$

where

$$\frac{\gamma_d\gamma_{\mathsf{t}}}{\gamma} = 1 - \eta_{\mathbf{T}}(\gamma_{\mathsf{t}}) \qquad \beta\frac{\gamma_d\gamma_{\mathsf{t}}}{\gamma} = 1 - \eta_{\mathbf{D}}(\beta\gamma_d). \tag{2.138}$$

From (2.138), an alternative expression for $\eta_{\mathbf{HH}^\dagger}(\gamma)$ with $\mathbf{H}$ as in Theorem 2.43, can be obtained as

$$\eta_{\mathbf{HH}^\dagger}(\gamma) = \eta_{\mathbf{D}}(\beta\,\gamma_d(\gamma)) \tag{2.139}$$

where $\gamma_d(\gamma)$ is the solution to (2.138).

Theorem 2.45. [262, 165] Let $\mathbf{H}$ be an $N \times K$ matrix defined as in Theorem 2.43. Defining

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

$$\lim_{\gamma\to\infty}\left(\log(\gamma\,\beta) - \frac{\mathcal{V}_{\mathbf{HH}^\dagger}(\gamma)}{\min\{\beta\,\mathbb{P}[\mathsf{T} \neq 0], \mathbb{P}[\mathsf{D} \neq 0]\}}\right) = \mathcal{L}_\infty \tag{2.140}$$

with

$$\mathcal{L}_\infty = \begin{cases} -\mathbb{E}\left[\log\frac{\mathsf{D}'}{\alpha\beta' e}\right] - \beta'\mathcal{V}_{\mathsf{T}'}(\alpha) & \beta' > 1 \\ -\mathbb{E}\left[\log\frac{\mathsf{T}'\mathsf{D}'}{e}\right] & \beta' = 1 \\ -\mathbb{E}\left[\log\frac{\Gamma_\infty\mathsf{T}'}{e}\right] - \frac{1}{\beta'}\mathcal{V}_{\mathsf{D}'}\left(\frac{1}{\Gamma_\infty}\right) & \beta' < 1 \end{cases} \tag{2.141}$$

with α and Γ_∞, respectively, solutions to

$$\eta_{\mathsf{T}'}(\alpha) = 1 - \frac{1}{\beta'}, \qquad \eta_{\mathsf{D}'}\left(\frac{1}{\Gamma_\infty}\right) = 1 - \beta'. \tag{2.142}$$

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

Corollary 2.1. As $\gamma \to \infty$, we have that

$$\lim_{\gamma\to\infty} \gamma\, \eta_{\mathbf{HH}^\dagger}(\gamma) = \begin{cases} \mathbb{E}\left[\frac{1}{\mathsf{D}}\right]\alpha, & \beta' > 1 \text{ and } \mathbb{P}[\mathsf{D} > 0] = 1 \\ \infty, & \text{otherwise} \end{cases}$$

with α solution to (2.142).

Theorem 2.46. [262] Let $\mathbf{H}$ be an $N \times K$ matrix defined as in Theorem 2.43. Further define

$$\mathcal{F}^{(N)}(y,\gamma) = \frac{1}{\|\mathbf{h}_j\|^2}\, \mathbf{h}_j^\dagger \left(\mathbf{I} + \gamma \sum_{\ell \neq j} \mathbf{h}_\ell \mathbf{h}_\ell^\dagger\right)^{-1} \mathbf{h}_j \quad \text{with} \quad \tfrac{j-1}{K} \leq y < \tfrac{j}{K}.$$

As $K, N \to \infty$, $\mathcal{F}^{(N)}(y,\gamma)$ converges almost surely to

$$\mathcal{F}^{(N)}(y,\gamma) \stackrel{a.s.}{\to} \frac{\gamma_{\mathsf{t}}(\gamma)}{\gamma\mathbb{E}[\mathsf{D}]} \quad y \in [0,1]$$

with $\gamma_{\mathsf{t}}(\gamma)$ satisfying (2.138).

Corollary 2.2. As $\gamma \to \infty$, we have that

$$\lim_{\gamma\to\infty} \frac{\gamma_{\mathsf{t}}(\gamma)}{\gamma} = \beta\, \mathbb{P}[\mathsf{T} > 0]\, \Gamma_\infty \tag{2.143}$$

where $\gamma_{\mathsf{t}}(\gamma)$ is the solution to (2.138) while Γ_∞ is the solution to (2.142) for $\beta' < 1$ and 0 otherwise.

Theorem 2.47. [159] Let $\mathbf{H}$ be an $N \times K$ matrix defined as in Theorem 2.43. The mth moment of the empirical eigenvalue distribution of $\mathbf{HH}^\dagger$ converges almost surely to

$$\sum_{k=1}^{m} \beta^k \sum_{\substack{m_1+\cdots+m_k=m \\ m_1\leq\cdots\leq m_k}} \sum_{\substack{n_1+\cdots+n_{m+1-k}=m \\ n_1\leq\cdots\leq n_{m+1-k}}} B(m_1,\ldots,m_k,n_1,\ldots,n_{m+1-k}) \cdot \\ \mathbb{E}[\mathsf{T}^{m_1}]\cdots\mathbb{E}[\mathsf{T}^{m_k}]\mathbb{E}[\mathsf{D}^{n_1}]\cdots\mathbb{E}[\mathsf{D}^{n_{m+1-k}}]. \tag{2.144}$$

with D and T defined as in Theorem 2.43, $f(i_1,\ldots,i_\ell)$ defined as in

(2.119), while[30]

$$B(m_1,\ldots,m_k,n_1,\ldots,n_{m+1-k}) = \frac{m(m-k)!(k-1)!}{f(m_1,\ldots,m_k)\cdot f(n_1,\ldots,n_{m+1-k})}.$$

Equation (2.144) is obtained in [159] using combinatorial tools. An alternative derivation can be obtained using Theorem 2.55, from which the nth moment of $\mathbf{HH}^\dagger$ given by (2.144) is also seen to equal $\mathbb{E}[\tilde{m}_n(\mathsf{D})]$ with $\tilde{m}_n$ admitting the following recursive re-formulation:

$$\tilde{m}_n(d) = \sum_{\ell=1}^{n} c_\ell(d) \sum_{n_1+n_2+\cdots+n_\ell=n} \tilde{m}_{n_1-1}(d)\ldots\tilde{m}_{n_\ell-1}(d) \tag{2.145}$$

with

$$c_{\ell+1}(d) = \beta\, d\,\mathbb{E}[\mathsf{T}^{\ell+1}]\mathbb{E}[\mathsf{D}]^\ell.$$

Theorem 2.48. [159] Let $\mathbf{H}$ be an $N\times K$ matrix defined as in Theorem 2.43 whose jth column is $\mathbf{h}_j$. Further define

$$\delta_n^{(N)}(y) = \frac{1}{\|\mathbf{h}_j\|^2}\mathbf{h}_j^\dagger\left(\sum_{\ell\neq j}\mathbf{h}_\ell\mathbf{h}_\ell^\dagger\right)^n\mathbf{h}_j \quad \text{with} \quad \tfrac{j-1}{K}\leq y<\tfrac{j}{K} \tag{2.146}$$

then, as $K,N\to\infty$ with $\frac{K}{N}\to\beta$, almost surely

$$\delta_n^{(N)}(y) \overset{a.s.}{\to} \frac{\mathbb{E}[\mathsf{D}m_n(\mathsf{D})]}{\mathbb{E}[\mathsf{D}]} = \frac{\xi_n}{\mathbb{E}[\mathsf{D}]} \tag{2.147}$$

where ξ_n can be computed through the following recursive equation

$$\xi_n = \beta\sum_{\ell=1}^{n}\mathbb{E}\left[\mathsf{D}^2\, m_{\ell-1}(\mathsf{D})\right]\sum_{\substack{n_1+\cdots+n_i=n-\ell\\ 1\leq i\leq n-\ell}}\mathbb{E}\left[\mathsf{T}^{i+1}\right]\xi_{n_1-1}\ldots\xi_{n_i-1}$$

with

$$m_n(d) = \beta d\sum_{\ell=1}^{n} m_{\ell-1}(d)\sum_{\substack{n_1+\cdots+n_i=n-\ell\\ 1\leq i\leq n-\ell}}\mathbb{E}\left[\mathsf{T}^{i+1}\right]\xi_{n_1-1}\ldots\xi_{n_i-1}. \tag{2.148}$$

[30] Note that $B(m_1,\ldots,m_k,n_1,\ldots,n_{m+1-k})$ can be interpreted as the number of non-crossing partitions (cf. Section 2.4.4) ϖ on $\{1,\ldots,m\}$ satisfying the conditions:
(i) the cardinalities of the subsets in ϖ, in increasing order, are $m_1,\ldots,m_k$,
(ii) the cardinalities of the subsets in the complementation map (cf. Section 2.4.4) of ϖ are, in increasing order, $n_1,\ldots,n_{m+1-k}$.

Moreover, $\mathbb{E}[m_n(\mathsf{D})]$ yields yet another way to compute the nth moment of the asymptotic spectrum of $\mathbf{HH}^\dagger$.

Under mild assumptions on the distribution of the independent entries of $\mathbf{H}$, the following convergence result is shown in Appendix 4.4.

Theorem 2.49. Define an $N \times K$ complex random matrix $\mathbf{H}$ whose entries are independent complex random variables (arbitrarily distributed) satisfying the Lindeberg condition (2.101) and with identical means. Let their variances be

$$\mathrm{Var}\left[\mathsf{H}_{i,j}\right] = \frac{\mathsf{P}_{i,j}}{N} \tag{2.149}$$

with $\mathbf{P}$ an $N \times K$ deterministic standard asymptotically doubly-regular matrix whose entries are uniformly bounded for any N. The asymptotic empirical eigenvalue distribution of $\mathbf{H}^\dagger\mathbf{H}$ converges almost surely to the Marčenko-Pastur distribution whose density is given by (2.100).

Using Lemma 2.22, Theorem 2.49 can be extended to matrices whose mean has rank r where $r > 1$ but such that

$$\lim_{N\to\infty} \frac{r}{N} = 0.$$

Definition 2.16. Consider an $N \times K$ random matrix $\mathbf{H}$ whose entries have variances

$$\mathrm{Var}[\mathsf{H}_{i,j}] = \frac{\mathsf{P}_{i,j}}{N} \tag{2.150}$$

with $\mathbf{P}$ an $N \times K$ deterministic matrix whose entries are uniformly bounded. For each N, let

$$v^N : [0,1) \times [0,1) \to \mathbb{R}$$

be the *variance profile* function given by

$$v^N(x,y) = \mathsf{P}_{i,j} \qquad \tfrac{i-1}{N} \le x < \tfrac{i}{N}, \qquad \tfrac{j-1}{K} \le y < \tfrac{j}{K}. \tag{2.151}$$

Whenever $v^N(x,y)$ converges uniformly to a limiting bounded measurable function, $v(x,y)$, we define this limit as the *asymptotic variance profile* of $\mathbf{H}$.

Theorem 2.50. [86, 102, 221] Let $\mathbf{H}$ be an $N \times K$ random matrix whose entries are independent zero-mean complex random variables (arbitrarily distributed) satisfying the Lindeberg condition (2.101) and with variances

$$\mathbb{E}\left[|\mathsf{H}_{i,j}|^2\right] = \frac{\mathsf{P}_{i,j}}{N} \tag{2.152}$$

where $\mathbf{P}$ is an $N \times K$ deterministic matrix whose entries are uniformly bounded and from which the asymptotic variance profile of $\mathbf{H}$, denoted $v(x, y)$, can be obtained as per Definition 2.16. As $K, N \to \infty$ with $\frac{K}{N} \to \beta$, the empirical eigenvalue distribution of $\mathbf{HH}^\dagger$ converges almost surely to a limiting distribution whose η-transform is

$$\eta_{\mathbf{HH}^\dagger}(\gamma) = \mathbb{E}\left[\Gamma_{\mathbf{HH}^\dagger}(\mathsf{X}, \gamma)\right] \tag{2.153}$$

with $\Gamma_{\mathbf{HH}^\dagger}(x, \gamma)$ satisfying the equations,

$$\Gamma_{\mathbf{HH}^\dagger}(x, \gamma) = \frac{1}{1 + \beta\gamma\mathbb{E}[v(x, \mathsf{Y})\Upsilon_{\mathbf{HH}^\dagger}(\mathsf{Y}, \gamma)]} \tag{2.154}$$

$$\Upsilon_{\mathbf{HH}^\dagger}(y, \gamma) = \frac{1}{1 + \gamma\,\mathbb{E}[v(\mathsf{X}, y)\Gamma_{\mathbf{HH}^\dagger}(\mathsf{X}, \gamma)]} \tag{2.155}$$

where X and Y are independent random variables uniform on $[0, 1]$.

The zero-mean hypothesis in Theorem 2.50 can be relaxed using Lemma 2.22. Specifically, if the rank of $\mathbb{E}[\mathbf{H}]$ is $o(N)$, then Theorem 2.50 still holds.

The asymptotic fraction of zero eigenvalues of $\mathbf{HH}^\dagger$ is equal to

$$\lim_{\gamma\to\infty} \eta_{\mathbf{HH}^\dagger}(\gamma) = 1 - \min\left\{\beta\,\mathbb{P}[\mathbb{E}[v(\mathsf{X}, \mathsf{Y})|\mathsf{Y}] \neq 0], \mathbb{P}[\mathbb{E}[v(\mathsf{X}, \mathsf{Y})|\mathsf{X}] \neq 0]\right\}.$$

Lemma 2.51. [86] Let $\mathbf{H}$ be an $N \times K$ complex random matrix defined as in Theorem 2.50. For each $a, b \in [0, 1]$, $a < b$

$$\frac{1}{N}\sum_{i=\lfloor aN\rfloor}^{\lfloor bN\rfloor} (\gamma\mathbf{HH}^\dagger + \mathbf{I})^{-1}_{i,i} \to \int_a^b \Gamma_{\mathbf{HH}^\dagger}(x, \gamma)dx. \tag{2.156}$$

Theorem 2.52. [262] Let $\mathbf{H}$ be an $N \times K$ matrix defined as in Theorem 2.50. Further define

$$\mathcal{F}^{(N)}(y,\gamma) = \frac{1}{\|\mathbf{h}_j\|^2}\mathbf{h}_j^\dagger \left(\mathbf{I} + \gamma \sum_{\ell \neq j} \mathbf{h}_\ell \mathbf{h}_\ell^\dagger\right)^{-1} \mathbf{h}_j, \qquad \tfrac{j-1}{K} \leq y < \tfrac{j}{K}.$$

As $K, N \to \infty$, $\mathcal{F}^{(N)}$ converges almost surely to $\frac{\mathcal{F}(y,\gamma)}{\mathbb{E}[v(\mathsf{X},y)]}$, with $\mathcal{F}(y,\gamma)$ solution to the fixed-point equation

$$\mathcal{F}(y,\gamma) = \mathbb{E}\left[\frac{v(\mathsf{X},y)}{1+\gamma\,\beta\,\mathbb{E}\left[\frac{v(\mathsf{X},\mathsf{Y})}{1+\gamma\mathcal{F}(\mathsf{Y},\gamma)}|\mathsf{X}\right]}\right] \quad y \in [0,1]. \tag{2.157}$$

The transform of the asymptotic spectrum of $\mathbf{HH}^\dagger$ is given by the following result proved in Appendix 4.5.

Theorem 2.53. Let $\mathbf{H}$ be an $N\times K$ complex random matrix defined as in Theorem 2.50. The Shannon transform of the asymptotic spectrum of $\mathbf{HH}^\dagger$ is

$$\begin{aligned}\mathcal{V}_{\mathbf{HH}^\dagger}(\gamma) &= \beta\,\mathbb{E}\left[\log(1+\gamma\,\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Gamma_{\mathbf{HH}^\dagger}(\mathsf{X},\gamma)|\mathsf{Y}])\right] \\ &\quad +\mathbb{E}\left[\log(1+\gamma\,\beta\,\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Upsilon_{\mathbf{HH}^\dagger}(\mathsf{Y},\gamma)|\mathsf{X}])\right] \\ &\quad -\gamma\,\beta\,\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\Gamma_{\mathbf{HH}^\dagger}(\mathsf{X},\gamma)\Upsilon_{\mathbf{HH}^\dagger}(\mathsf{Y},\gamma)\right]\log e\end{aligned} \tag{2.158}$$

with $\Gamma_{\mathbf{HH}^\dagger}(\cdot,\cdot)$ and $\Upsilon_{\mathbf{HH}^\dagger}(\cdot,\cdot)$ satisfying (2.154) and (2.155).

Theorem 2.54. [262] Let $\mathbf{H}$ be an $N \times K$ complex random matrix defined as in Theorem 2.50. Then, denoting

$$\beta' = \beta\frac{\mathbb{P}[\mathbb{E}[v(\mathsf{X},\mathsf{Y})|\mathsf{Y}] \neq 0]}{\mathbb{P}[\mathbb{E}[v(\mathsf{X},\mathsf{Y})|\mathsf{X}] \neq 0]},$$

we have that

$$\lim_{\gamma\to\infty}\left(\log(\gamma\beta) - \frac{\mathcal{V}_{\mathbf{HH}^\dagger}(\gamma)}{\min\{\beta\mathbb{P}[\mathbb{E}[v(\mathsf{X},\mathsf{Y})|\mathsf{Y}]\neq 0], \mathbb{P}[\mathbb{E}[v(\mathsf{X},\mathsf{Y})|\mathsf{X}]\neq 0]\}}\right) = \mathcal{L}_\infty$$

with

$$\mathcal{L}_\infty \stackrel{a.s.}{\to} \begin{cases} -\mathbb{E}\left[\log\left(\frac{1}{e}\mathbb{E}\left[\frac{v(\mathsf{X}',\mathsf{Y}')}{1+\alpha(\mathsf{Y}')}|\mathsf{X}'\right]\right)\right] - \beta'\mathbb{E}\left[\log\left(1+\alpha(\mathsf{Y}')\right)\right] & \beta' > 1 \\ -\mathbb{E}\left[\log\frac{v(\mathsf{X}',\mathsf{Y}')}{e}\right] & \beta' = 1 \\ -\mathbb{E}\left[\log\frac{\Gamma_\infty(\mathsf{Y}')}{e}\right] - \frac{1}{\beta'}\mathbb{E}\left[\log\left(1+\mathbb{E}\left[\frac{v(\mathsf{X}',\mathsf{Y}')}{\Gamma_\infty(\mathsf{Y}')}|\mathsf{X}'\right]\right)\right] & \beta' < 1 \end{cases}$$

with X' and Y' the restrictions of X and Y to the events $\mathbb{E}[v(\mathsf{X},\mathsf{Y})|\mathsf{X}]\neq 0$ and $\mathbb{E}[v(\mathsf{X},\mathsf{Y})|\mathsf{Y}]\neq 0$, respectively. The function $\alpha(\cdot)$ is the solution, for $\beta'>1$, of

$$\alpha(y) = \frac{1}{\beta'}\mathbb{E}\left[\frac{v(\mathsf{X}',y)}{\mathbb{E}\left[\frac{v(\mathsf{R}',\mathsf{Y}')}{1+\alpha(\mathsf{Y}')}|\mathsf{X}'\right]}\right] \tag{2.159}$$

whereas $\Gamma_\infty(\cdot)$ is the solution, for $\beta'<1$, of

$$\mathbb{E}\left[\frac{1}{1+\mathbb{E}\left[\frac{v(\mathsf{X}',\mathsf{Y}')}{\Gamma_\infty(\mathsf{Y}')}|\mathsf{X}'\right]}\right] = 1-\beta'. \tag{2.160}$$

Corollary 2.3. As $\gamma\to\infty$, if $\beta' > 1$ and $\mathbb{P}[\mathbb{E}[v(\mathsf{X},\mathsf{Y})|\mathsf{x}] > 0]] = 1$, then

$$\lim_{\gamma\to\infty} \gamma\,\eta_{\mathbf{HH}^\dagger}(\gamma) = \frac{1}{\beta\,\mathbb{P}[\mathbb{E}[v(\mathsf{X},\mathsf{Y})|\mathsf{Y}]\neq 0]}\mathbb{E}\left[\frac{1}{\mathbb{E}\left[\frac{v(\mathsf{X}',\mathsf{Y}')}{1+\alpha(\mathsf{Y}')}|\mathsf{X}'\right]}\right] \tag{2.161}$$

with $\alpha(\cdot)$ solution to (2.159). Otherwise the limit in (2.161) diverges.

Corollary 2.4. As $\gamma\to\infty$, we have that

$$\lim_{\gamma\to\infty} \mathcal{F}(y,\gamma) = \beta\,\mathbb{P}[\mathbb{E}[v(\mathsf{X},\mathsf{Y})|\mathsf{Y}]\neq 0]\Gamma_\infty(y) \tag{2.162}$$

where $\Gamma_\infty(y)$ is the solution to (2.160) for $\beta' < 1$ and 0 otherwise while $\mathcal{F}(y,\gamma)$ is the solution to (2.157).

Theorem 2.55. [159] Let $\mathbf{H}$ be an $N\times K$ matrix defined as in Theorem 2.50. The nth moment of the empirical eigenvalue distribution of $\mathbf{HH}^\dagger$ converges almost surely to

$$\lim_{N\to\infty}\frac{1}{N}\mathrm{tr}\left\{(\mathbf{HH}^\dagger)^n\right\} = \mathbb{E}[m_n(\mathsf{X})] \tag{2.163}$$

with $m_n(x)$ satisfying the recursive equation

$$m_n(x) = \beta \sum_{\ell=1}^{n} m_{\ell-1}(x)\, \mathbb{E}[\, v(x,\mathsf{Y}) \sum_{\substack{n_1+\cdots+n_i=n-\ell \\ 1\le i\le n-\ell}} \mathbb{E}\,[v(\mathsf{X},\mathsf{Y}) m_{n_1-1}(\mathsf{X})|\mathsf{Y}] \ldots \mathbb{E}\,[v(\mathsf{X},\mathsf{Y}) m_{n_i-1}(\mathsf{X})|\mathsf{Y}]\,] \tag{2.164}$$

where $m_0(x) = 1$ and where, in the second summation, the n_k's with $k \in \{1,\ldots,i\}$ are strictly positive integers. In turn, X and Y are independent random variables uniform on [0,1].

Theorem 2.56. [159] Consider an $N \times K$ matrix $\mathbf{H}$ defined as in Theorem 2.50 whose jth column is $\mathbf{h}_j$. As $K, N \to \infty$, the quadratic form

$$\delta_n^{(N)}(y) = \frac{1}{\|\mathbf{h}_j\|^2} \mathbf{h}_j^\dagger \left(\sum_{\ell \neq j} \mathbf{h}_\ell \mathbf{h}_\ell^\dagger \right)^n \mathbf{h}_j \qquad \tfrac{j-1}{K} \le y < \tfrac{j}{K} \tag{2.165}$$

converges almost surely to a function $\delta_n(y)$ given by

$$\delta_n(y) = \frac{\mathbb{E}[m_n(\mathsf{X}) v(\mathsf{X},y)]}{\mathbb{E}[v(\mathsf{X},y)]} = \frac{\xi_n(y)}{\mathbb{E}[v(\mathsf{X},y)]} \tag{2.166}$$

where X is a random variable uniform on [0,1] and $m_n(x)$ is given by (2.164) in Theorem 2.55.

From Theorems 2.55 and 2.56 it follows that:

Corollary 2.5. The relationships between the moments, $\mathbb{E}[m_n(\mathsf{X})]$, and $\xi_n(y)$ are:

$$\mathbb{E}[m_n(\mathsf{X})] = \beta \sum_{\ell=1}^{n} \mathbb{E}\left[\xi_{\ell-1}(\mathsf{Y}) \sum_{\substack{n_1+\cdots+n_i=n-\ell \\ 1\le i\le n-\ell}} \xi_{n_1-1}(\mathsf{Y}) \ldots \xi_{n_i-1}(\mathsf{Y}) \right] \tag{2.167}$$

with $\xi_n(y) = \mathbb{E}[m_n(\mathsf{X}) v(\mathsf{X},y)]$.

In the case that $v(x,y)$ factors as $v(x,y) = v_{\mathrm{X}}(x) v_{\mathrm{Y}}(y)$, then (2.164) becomes

$$m_n(r) = \beta\, r \sum_{\ell=1}^{n} m_{\ell-1}(r) \sum_{\substack{n_1+\cdots+n_i=n-\ell \\ 1\le i\le n-\ell}} \mathbb{E}[\mathsf{D}^{i+1}]\, \mathbb{E}[\mathsf{C} m_{n_1-1}(\mathsf{C})] \cdot \cdot \mathbb{E}[\mathsf{C} m_{n_i-1}(\mathsf{C})]$$

where C and D are independent random variables whose distribution equals the distributions of $v_{\mathrm{X}}(\mathsf{X})$ and $v_{\mathrm{Y}}(\mathsf{Y})$, respectively, with X and Y uniform on $[0,1]$. From the above recursive formula, the closed-form expression given in (2.144) can be found by resorting to techniques of non-crossing partitions and the complementation map.

Remark 2.3.1. If $v(x,y)$ factors, Theorems 2.50-2.56 admit simpler formulations. The Shannon transform, η-transform, $\digamma(y,\gamma)$ and moments of the asymptotic spectrum of $\mathbf{HH}^\dagger$, with $\mathbf{H}$ defined as in Theorem 2.50, coincide with those of Theorems 2.43-2.48: in this case D and T represent independent random variables whose distributions are given by the distributions of $v_{\mathrm{X}}(\mathsf{X})$ and $v_{\mathrm{Y}}(\mathsf{Y})$, respectively.

An example of $v(x,y)$ that factors is when the $N \times K$ matrix of variances, $\mathbf{P}$, introduced in (2.152), is the outer product of two vectors

$$\mathbf{P} = \mathsf{d}\mathsf{t}^T. \tag{2.168}$$

where the N-vector d and the K-vector t have nonnegative deterministic entries.

Definition 2.17. Let $\mathbf{B}$ be an $N \times K$ random matrix with independent columns. Denoting by $\lfloor \cdot \rfloor$ the closest smaller integer, $\mathbf{B}$ behaves ergodically if, for a given $x \in [0,1)$, the empirical distribution of

$$|(\mathbf{B})_{\lfloor xN \rfloor,1}|^2, \ldots, |(\mathbf{B})_{\lfloor xN \rfloor,K}|^2$$

converges almost surely to a nonrandom limit $F_x(\cdot)$ and, for a given $y \in [0,1)$, the empirical distribution of

$$|(\mathbf{B})_{1,\lfloor yK \rfloor}|^2, \ldots, |(\mathbf{B})_{N,\lfloor yK \rfloor}|^2$$

converges almost surely to a nonrandom limit $F_y(\cdot)$.

Definition 2.18. Let $\mathbf{B}$ be a random matrix that behaves ergodically in the sense of Definition 2.17. Assuming that $F_x(\cdot)$ and $F_y(\cdot)$ have all their moments bounded, the two-dimensional channel profile of $\mathbf{B}$ is defined as the function $\rho(x,y) : [0,1]^2 \to \mathbb{R}$ such that, if X is uniform on $[0,1]$, the distribution of $\rho(\mathsf{X},y)$ equals $F_y(\cdot)$ whereas, if Y is uniform on $[0,1]$, then the distribution of $\rho(x,\mathsf{Y})$ equals $F_x(\cdot)$.

Analogously, the one-dimensional channel profile of $\mathbf{B}$ for a given k is the function $\rho_k(x) : [0,1] \to \mathbb{R}$ such that, if X is uniform on $[0,1]$, the distribution of $\rho_k(\mathsf{X})$ equals the nonrandom asymptotic empirical distribution of $|(\mathbf{B})_{1,k}|^2, \ldots, |(\mathbf{B})_{N,k}|^2$.

Theorem 2.57. [159, 160] Consider an $N \times K$ matrix $\mathbf{H} = \mathbf{S} \circ \mathbf{B}$ with $\circ$ denoting the Hadamard (element-wise) product and with $\mathbf{S}$ and $\mathbf{B}$ independent $N \times K$ random matrices. The entries of $\mathbf{S}$ are zero-mean i.i.d. complex random variables arbitrarily distributed with variance $\frac{1}{N}$ while $\mathbf{B}$ is as in Definition 2.18 with $F_x(\cdot)$ and $F_y(\cdot)$ having all their moments bounded. Denoting by $\rho_{\mathbf{B}}(x,y)$ the channel profile of $\mathbf{B}$, then, as $K, N \to \infty$ with $\frac{K}{N} \to \beta$, the empirical eigenvalue distribution of $\mathbf{H}\mathbf{H}^\dagger$ converges almost surely to a nonrandom limit whose η-transform, Shannon transform and moments are given by (2.153), (2.158) and (2.163-2.164) respectively with $v(x,y)$ replaced by $\rho_{\mathbf{B}}(x,y)$. Analogous considerations hold for the functions $\digamma(y,\gamma)$ and $\delta_n(y)$.

Theorem 2.58. [262] Consider an $N \times K$ matrix $\mathbf{H}$ whose entries are zero-mean correlated Gaussian random variables with correlation function $\mathsf{r}_{\mathbf{H}}(i,j;i',j')$ whose eigenvalues are $\lambda_{i,j}(\mathsf{r}_{\mathbf{H}})$, for $1 \le i \le N$ and $1 \le j \le K$ (cf. Definition 2.8) and whose kernel factors as in (2.35). Assume that $N\lambda_{i,j}(\mathsf{r}_{\mathbf{A}})$ are uniformly bounded for any N. Theorems 2.49-2.56 hold by redefining $v(x,y)$ as the asymptotic variance profile of the Karhunen-Loève image of $\mathbf{H}$, which corresponds to the limit for $N \to \infty$ of

$$v^N(x,y) = N\lambda_{i,j}(\mathsf{r}_{\mathbf{H}}) \qquad \tfrac{i-1}{N} \le x < \tfrac{i}{N}, \quad \tfrac{j-1}{K} \le y < \tfrac{j}{K}.$$

Therefore, the asymptotic spectrum of $\mathbf{H}$ is fully characterized by the variances of the entries of its Karhunen-Loève image.

A special case of Theorem 2.58 is illustrated in [55] for $\mathsf{r}_{\mathbf{H}}(i,j;i',j') = f(i-i', j-j')$, in which case $\mathbf{H}$ is termed a *band matrix*.

Theorem 2.59. [159] Consider the $N \times K$ random matrix

$$\mathbf{H} = [\mathbf{A}_1\mathbf{s}_1, \ldots, \mathbf{A}_K\mathbf{s}_K]\bar{\mathbf{A}} \tag{2.169}$$

where $\mathbf{S} = [\mathbf{s}_1 \dots \mathbf{s}_K]$ is an $N \times K$ matrix with zero-mean i.i.d. entries with variance $\frac{1}{N}$, $\bar{\mathbf{A}}$ is a deterministic diagonal matrix and $\mathbf{A}_k$ $k \in \{1, \dots, K\}$ are either *finite order* or *infinite order absolutely summable* $N \times N$ Toeplitz independent matrices, independent of $\mathbf{S}$. Let $\rho(x, y)$ be the two-dimensional channel profile of the $N \times K$ matrix $\mathbf{\Lambda}$ whose (i, j)th entry is[31]

$$\Lambda_{i,j} = |\bar{\mathsf{A}}_j|^2 \, \lambda_i(\mathbf{A}_j) \tag{2.170}$$

with $\lambda_i(\mathbf{A}_j)$ the ith eigenvalue of $\mathbf{A}_j \mathbf{A}_j^\dagger$. As $K, N \to \infty$ with $\frac{K}{N} \to \beta$, the empirical eigenvalue distribution of $\mathbf{H}\mathbf{H}^\dagger$ converges almost surely to a nonrandom limiting distribution whose η-transform is [159]

$$\eta_{\mathbf{H}\mathbf{H}^\dagger}(\gamma) = \mathbb{E}\left[\, \Gamma_{\mathbf{H}\mathbf{H}^\dagger}(\mathsf{X}, \gamma) \,\right] \tag{2.171}$$

where $\Gamma_{\mathbf{H}\mathbf{H}^\dagger}(\cdot, \cdot)$ satisfies the equations

$$\Gamma_{\mathbf{H}\mathbf{H}^\dagger}(x, \gamma) = \frac{1}{1 + \beta\, \gamma \mathbb{E}[\rho(x, \mathsf{Y}) \Upsilon_{\mathbf{H}\mathbf{H}^\dagger}(\mathsf{Y}, \gamma)]} \tag{2.172}$$

$$\Upsilon_{\mathbf{H}\mathbf{H}^\dagger}(y, \gamma) = \frac{1}{1 + \gamma\, \mathbb{E}[\rho(\mathsf{X}, y) \Gamma_{\mathbf{H}\mathbf{H}^\dagger}(\mathsf{X}, \gamma)]} \tag{2.173}$$

with X and Y independent random variables uniform on $[0, 1]$.

Consequently, Theorems 2.49-2.56 still hold with the function $v(x, y)$ replaced by $\rho(x, y)$.

Define

$$\mathcal{R}(N, m) = \frac{1}{N} \sum \mathsf{H}^*_{i_1, j_m} \mathsf{H}_{i_1, j_1} \cdots \mathsf{H}^*_{i_m, j_{m-1}} \mathsf{H}_{i_m, j_m}, \tag{2.174}$$

where the summation ranges over all $2m$-tuples $i_1, \dots, i_m, j_1, \dots, j_m$ satisfying $1 \leq i_\ell \leq N$ and $1 \leq j_\ell \leq K$, such that the cardinality of the set of distinct values of i_ℓ plus the cardinality of the set of distinct values of j_ℓ equals $k + 1$, and such that there is one-to-one pairing of the unconjugate and the conjugate terms in the products.

[31] The existence of $\rho(x, y)$ implies that $\mathbf{\Lambda}$ is a matrix that behaves ergodically in the sense of Definition 2.17.

Lemma 2.60. [296] Let $\mathbf{H}$ be an $N \times K$ real or complex random matrix whose entries are independent with

$$\mathbb{E}[\mathsf{H}_{i,j}] = \frac{\mu_i}{\sqrt{N}}$$

regardless of j and with

$$\mathbb{E}\left[|\mathsf{H}_{i,j} - \tfrac{\mu_i}{\sqrt{N}}|^{2+\delta}\right] < \frac{\kappa^2}{N^{1+\delta/2}}$$

for some $\delta > 0$ and $\kappa > 0$. The empirical eigenvalue distribution of $\mathbf{H}\mathbf{H}^\dagger$ converges almost surely to a nonrandom limit $F_{\mathbf{H}\mathbf{H}^\dagger}(\cdot)$ if and only if, for each m, $\mathbb{E}[\mathcal{R}(N,m)]$ in (2.174) converges as $N \to \infty$. Furthermore,

$$\int \lambda^m dF_{\mathbf{H}\mathbf{H}^\dagger}(\lambda) = \lim_{N\to\infty} \int \lambda^m dF^N_{\mathbf{H}\mathbf{H}^\dagger}(\lambda) \quad (2.175)$$

$$= \lim_{N\to\infty} \mathbb{E}[\mathcal{R}(N,m)]. \quad (2.176)$$

2.4 Free Probability

In the last few years, a large fraction of the new results on the asymptotic convergence of the eigenvalues of random matrices has been obtained using the tools of free probability. This is a discipline founded by Voiculescu [283] in the 1980s that spawned from his work on operator algebras. Unlike classical scalar random variables, random matrices are noncommutative objects whose large-dimension asymptotics have provided the major applications of the theory of free probability.

Knowing the eigenvalues of two matrices is, in general, not enough to find the eigenvalues of the sum of the two matrices (unless they commute). However, it turns out that free probability identifies a certain sufficient condition (called asymptotic freeness) under which the asymptotic spectrum of the sum can be obtained from the individual asymptotic spectra without involving the structure of the eigenvectors of the matrices.

When two matrices are asymptotically free, there exists a rule to compute any asymptotic moment of the sum of the matrices (and thus the asymptotic spectrum) as a function of the individual moments. The combinatorics of the rule are succinctly described by recourse to the R-transform. Indeed, the central result in the application of free probability to random matrices is that the R-transform of the asymptotic spectrum of the sum of asymptotically free matrices is equal to the sum of the individual R-transforms. Analogously, the S-transform of the product of asymptotically free random matrices is equal to the product of the individual S-transforms. Computation of the R-transform, S-transform and the mixed moments of random matrices is often aided by a certain combinatorial construct based on noncrossing partitions due to Speicher [240, 241, 242].

The power of free probability is evident, not only in the new results on random matrices it unveils, but on the fresh view it provides on established results. For example, it shows that the semicircle law and the Marčenko-Pastur laws are the free counterparts of the Gaussian and Poisson distributions, respectively, in classical probability. Furthermore, using the central R-transform result it is possible to provide different proof techniques for the major results reviewed in Section 2.3.

2.4.1 Asymptotic Freeness

For notational convenience, we define the following functional for sequences of Hermitian matrices:

$$\phi(\mathbf{A}) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}[\mathrm{tr}\mathbf{A}]. \tag{2.177}$$

Note that the expected asymptotic pth moment of $\mathbf{A}$ is $\phi(\mathbf{A}^p)$ and $\phi(\mathbf{I}) = 1$.

Definition 2.19. [287] The Hermitian random matrices $\mathbf{A}$ and $\mathbf{B}$ are asymptotically free if for all ℓ and for all polynomials $p_i(\cdot)$ and $q_i(\cdot)$ with $1 \leq i \leq \ell$ such that[32]

$$\phi(p_i(\mathbf{A})) = \phi(q_i(\mathbf{B})) = 0, \tag{2.178}$$

we have

$$\phi(p_1(\mathbf{A})\, q_1(\mathbf{B}) \ldots p_\ell(\mathbf{A})\, q_\ell(\mathbf{B})) = 0. \tag{2.179}$$

Definition 2.19 generalizes to several random matrices as follows.

[32] This includes polynomials with constant (zero-order) terms.

Definition 2.20. The Hermitian random matrices $\mathbf{A}_1, \ldots, \mathbf{A}_m$ are asymptotically free if, for all $\ell \in \mathbb{N}$ and all polynomials $p_1(\,\cdot\,), \ldots, p_\ell(\,\cdot\,)$,

$$\phi\left(p_1(\mathbf{A}_{j(1)}) \cdot p_2(\mathbf{A}_{j(2)}) \cdots p_\ell(\mathbf{A}_{j(\ell)})\right) = 0 \tag{2.180}$$

whenever

$$\phi\left(p_i(\mathbf{A}_{j(i)})\right) = 0 \qquad \forall i = 1, \ldots, \ell \tag{2.181}$$

where $j(i) \neq j(i+1)$ (i.e., consecutive indices are distinct, but non-neighboring indices are allowed to be equal).

It is also of interest to define asymptotic freeness between pairs of Hermitian random matrices.

Definition 2.21. [287] The pairs of Hermitian matrices $\{\mathbf{A}_1, \mathbf{A}_2\}$ and $\{\mathbf{B}_1, \mathbf{B}_2\}$ are asymptotically free if, for all ℓ and for all polynomials $p_i(\cdot)$ and $q_i(\cdot)$ in two noncommuting indeterminates with $1 \leq i \leq \ell$ such that

$$\phi(p_i(\mathbf{A}_1, \mathbf{A}_2)) = \phi(q_i(\mathbf{B}_1, \mathbf{B}_2)) = 0, \tag{2.182}$$

we have

$$\phi(p_1(\mathbf{A}_1, \mathbf{A}_2)\, q_1(\mathbf{B}_1, \mathbf{B}_2) \ldots p_\ell(\mathbf{A}_1, \mathbf{A}_2)\, q_\ell(\mathbf{B}_1, \mathbf{B}_2)) = 0. \tag{2.183}$$

As a shorthand, when $\{\mathbf{A}_1, \mathbf{A}_2\}$ and $\{\mathbf{B}_1, \mathbf{B}_2\}$ are asymptotically free, we will say that $(\{\mathbf{A}_1, \mathbf{A}_2\}, \{\mathbf{B}_1, \mathbf{B}_2\})$ are asymptotically free.

Let us now incorporate, in the definition of asymptotic freeness, the class of non-Hermitian matrices. If $\mathbf{H}_1$ and $\mathbf{H}_2$ are rectangular non-Hermitian matrices, we say that $\{\mathbf{H}_1, \mathbf{H}_1^\dagger\}$ and $\{\mathbf{H}_2, \mathbf{H}_2^\dagger\}$ are asymptotically free, or equivalently that $\mathbf{H}_1$ and $\mathbf{H}_2$ are asymptotically *-free, if the relations given in Definition 2.21 apply with $p_i(\mathbf{H}_1, \mathbf{H}_1^\dagger)$ and $q_i(\mathbf{H}_2, \mathbf{H}_2^\dagger)$ polynomials of two noncommuting variables.

The definition of asymptotic freeness is somewhat reminiscent of the concept of independent random variables. However, as the following example shows, statistical independence does not imply asymptotic freeness.

Example 2.33. Suppose that X_1 and X_2 are independent zero-mean random variables with nonzero variance. Then, $X_1\mathbf{I}$ and $X_2\mathbf{I}$ are not asymptotically free. More generally, if two matrices are asymptotically free and they commute, then one of them is necessarily deterministic.

An alternative to the foregoing definitions is obtained by dropping the expectation from the definition of the operator ϕ in (2.177) and assuming that the spectra of the matrices converge almost surely to a nonrandom limit. This notion is known as *almost surely asymptotic freeness* [110, 111]. As will be pointed out, some of properties and examples discussed in the sequel for asymptotic freeness also hold for almost surely asymptotic freeness.

To illustrate the usefulness of the definition of asymptotic freeness, we will start by computing various mixed moments of random matrices. If $\mathbf{A}_1, \ldots, \mathbf{A}_\ell$ are asymptotically free random matrices, a number of useful relationships can be obtained by particularizing the following identity:

$$\phi\left((\mathbf{A}_1^{k_1} - \phi(\mathbf{A}_1^{k_1})\mathbf{I}) \cdot (\mathbf{A}_2^{k_2} - \phi(\mathbf{A}_2^{k_2})\mathbf{I}) \cdots (\mathbf{A}_\ell^{k_\ell} - \phi(\mathbf{A}_\ell^{k_\ell})\mathbf{I})\right) = 0 \tag{2.184}$$

which is obtained from (2.180) by considering the ℓ polynomials

$$p_i(\mathbf{A}_i) = \mathbf{A}_i^{k_i} - \phi(\mathbf{A}_i^{k_i})\mathbf{I}$$

which obviously satisfy $\phi(p_i(\mathbf{A}_i)) = 0$.

Applying (2.184), we can easily obtain the following relationships for asymptotically free $\mathbf{A}$ and $\mathbf{B}$:

$$\phi(\mathbf{A}^k\mathbf{B}^\ell) = \phi(\mathbf{A}^k)\phi(\mathbf{B}^\ell) \tag{2.185}$$

$$\phi(\mathbf{ABAB}) = \phi^2(\mathbf{B})\phi(\mathbf{A}^2) + \phi^2(\mathbf{A})\phi(\mathbf{B}^2) - \phi^2(\mathbf{A})\phi^2(\mathbf{B}). \tag{2.186}$$

As mentioned, one approach to characterize the asymptotic spectrum of a random matrix is to obtain its moments of all orders. Frequent applications of the concept of asymptotic freeness stem from the fact that the moments of a noncommutative polynomial $p(\mathbf{A}, \mathbf{B})$ of two asymptotically free random matrices can be computed from the individual moments of $\mathbf{A}$ and $\mathbf{B}$. Thus, if $p(\mathbf{A}, \mathbf{B})$, $\mathbf{A}$, $\mathbf{B}$ are Hermitian, the asymptotic spectrum of $p(\mathbf{A}, \mathbf{B})$ depends only on those of $\mathbf{A}$ and $\mathbf{B}$ even if they do not have the same eigenvectors. To illustrate this point, when $p(\mathbf{A}, \mathbf{B}) = \mathbf{A} + \mathbf{B}$ we can use (2.184) to obtain the first few moments:

$$\phi(\mathbf{A}+\mathbf{B}) = \phi(\mathbf{A}) + \phi(\mathbf{B}) \tag{2.187}$$

$$\phi((\mathbf{A}+\mathbf{B})^2) = \phi(\mathbf{A}^2) + \phi(\mathbf{B}^2) + 2\phi(\mathbf{A})\phi(\mathbf{B}) \tag{2.188}$$

$$\begin{aligned}\phi((\mathbf{A}+\mathbf{B})^3) &= \phi(\mathbf{A})^3 + \phi(\mathbf{B}^3) + 3\phi(\mathbf{A})\phi(\mathbf{B}^2) \\ &+ 3\phi(\mathbf{B})\phi(\mathbf{A}^2)\end{aligned} \tag{2.189}$$

$$\begin{aligned}\phi((\mathbf{A}+\mathbf{B})^4) &= \phi(\mathbf{A}^4) + \phi(\mathbf{B}^4) + 4\phi(\mathbf{A})\phi(\mathbf{B}^3) \\ &+ 4\phi(\mathbf{B})\phi(\mathbf{A}^3) + 2\phi^2(\mathbf{B})\phi(\mathbf{A}^2) \\ &+ 2\phi^2(\mathbf{A})\phi(\mathbf{B}^2) + 2\phi(\mathbf{B}^2)\phi(\mathbf{A}^2).\end{aligned} \tag{2.190}$$

All other higher moments can be computed analogously. As we will see below, the R-transform defined in Section 2.2.5 circumvents the increasingly cumbersome derivations required to derive other moments.[33]

Next, we compile a list of some of the most useful instances of asymptotic freeness that have been shown so far. In order to ease the exposition, we state them without including all the technical sufficient conditions (usually on the higher order moments of the matrix entries) under which they have been proved so far. For the exact technical conditions, the reader can refer to the pertinent citations.

Example 2.34. Any random matrix and the identity are asymptotically free.

Example 2.35. [287] Independent Gaussian standard Wigner matrices are asymptotically free.

Example 2.36. [287] Let $\mathbf{X}$ and $\mathbf{Y}$ be independent standard Gaussian matrices. Then $\{\mathbf{X}, \mathbf{X}^\dagger\}$ and $\{\mathbf{Y}, \mathbf{Y}^\dagger\}$ are asymptotically free.

[33] Notice that the first three moments of $\mathbf{A}+\mathbf{B}$ can be obtained from formulas identical to those pertaining to classical independent random variables. A difference appears from the fourth moment (2.190) on.

Historically, Examples 2.35 and 2.36 are the first results on the freeness of random matrices.

Example 2.37. [63] Independent standard Wigner matrices are asymptotically free.

Example 2.38. [63] A standard Wigner matrix and a diagonal deterministic matrix (or a block diagonal deterministic matrix with bounded block size) are asymptotically free.

Example 2.39. [211] Let $\mathbf{X}$ and $\mathbf{Y}$ be independent square matrices whose entries are zero-mean independent random variables (arbitrarily distributed), with variance vanishing inversely proportionally to the size. Then, $(\{\mathbf{X}, \mathbf{X}^\dagger\}, \{\mathbf{Y}, \mathbf{Y}^\dagger\})$ are asymptotically free. Furthermore, these matrices and block diagonal deterministic matrices with bounded block size are also asymptotically free.

Example 2.40. Suppose that the N-vectors $\mathbf{h}_i$, $i \in \{1, \dots, \ell\}$, are independent and have independent entries with variances equal to $\frac{1}{N}$ and identical means. Furthermore, let $X_1, \dots X_\ell$ be independent random variables with finite moments of all order and also independent of the random vectors. Then,

$$X_1 \mathbf{h}_1 \mathbf{h}_1^\dagger, X_2 \mathbf{h}_2 \mathbf{h}_2^\dagger, \dots, X_\ell \mathbf{h}_\ell \mathbf{h}_\ell^\dagger$$

are asymptotically free.

Example 2.41. [287] If $\mathbf{U}$ and $\mathbf{V}$ are independent Haar matrices, then $(\{\mathbf{U}, \mathbf{U}^\dagger\}, \{\mathbf{V}, \mathbf{V}^\dagger\})$ are asymptotically free.

Example 2.42. [287] If $\mathbf{U}$ is a Haar matrix and $\mathbf{D}$ is a deterministic matrix with bounded eigenvalues, then $(\{\mathbf{U}, \mathbf{U}^\dagger\}, \{\mathbf{D}, \mathbf{D}^\dagger\})$ are asymptotically free.

Example 2.43. [294] Let $\mathbf{X}$ be a standard Gaussian matrix and let $\mathbf{D}$ be a deterministic matrix with bounded eigenvalues. Then $(\{\mathbf{X}, \mathbf{X}^\dagger\}, \{\mathbf{D}, \mathbf{D}^\dagger\})$ are asymptotically free.

Example 2.44. [240] $\mathbf{UAU}^\dagger$ and $\mathbf{B}$ are asymptotically free if $\mathbf{A}$ and $\mathbf{B}$ are Hermitian matrices whose asymptotic averaged empirical eigenvalue distributions are compactly supported and $\mathbf{U}$ is a Haar matrix independent of $\mathbf{A}$ and $\mathbf{B}$.

Example 2.45. [240] A unitarily invariant matrix with compactly supported asymptotic spectrum and a deterministic matrix with bounded eigenvalues are asymptotically free.

Example 2.46. [295] Independent unitarily invariant matrices with compactly supported asymptotic spectra are asymptotically free.

Example 2.47. [295] Let $\mathbf{A}$ and $\mathbf{B}$ be $N \times K$ independent biunitarily invariant random matrices whose asymptotic averaged empirical singular value distributions are compactly supported. Then, $(\{\mathbf{A}, \mathbf{A}^\dagger\}, \{\mathbf{B}, \mathbf{B}^\dagger\}, \{\mathbf{D}, \mathbf{D}^\dagger\})$ are asymptotically free for any deterministic $N \times K$ matrix $\mathbf{D}$ with bounded eigenvalues.

Example 2.48. Let $\mathbf{H}_1$ and $\mathbf{H}_2$ be independent standard Gaussian matrices and let $\mathbf{T}$ be a random Hermitian matrix independent of $\mathbf{H}_1$ and $\mathbf{H}_2$ with compactly supported asymptotic averaged empirical eigenvalue distribution. Then it follows from Lemma 2.7 and Examples 2.45–2.46 that $(f_1(\mathbf{H}_1\mathbf{T}\mathbf{H}_1^\dagger), f_2(\mathbf{H}_2\mathbf{T}\mathbf{H}_2^\dagger), \{\mathbf{D}, \mathbf{D}^\dagger\})$ are asymptotically free for any real continuous functions $f_1(\cdot)$ and $f_2(\cdot)$, defined on the real line, and any deterministic square matrix $\mathbf{D}$ with bounded asymptotic spectrum.

Examples 2.41–2.48 are not only instances of asymptotic freeness, but also of almost surely asymptotic freeness [111]. In particular, for Example 2.48 the almost surely convergence holds if the asymptotic empirical eigenvalue distribution of $\mathbf{T}$ converges almost surely to a compactly supported probability measure. Note also that Examples 2.35 and 2.36 are special cases of Example 2.46 and 2.47, respectively.

Theorem 2.61. [64] Let $(\mathbf{A}, \{\mathbf{P}_1, \mathbf{V}_1, \ldots, \mathbf{P}_\ell, \mathbf{V}_\ell\})$ be asymptotically free. If

$$\mathbf{P}_i\mathbf{V}_i = \mathbf{V}_i\mathbf{P}_i = \mathbf{I} \qquad \text{and} \qquad \phi(\mathbf{P}_i\mathbf{V}_j) = 0$$

for all $i \in \{1, \ldots, \ell\}$ and $i \neq j$, then $\mathbf{P}_1\mathbf{A}\mathbf{V}_1, \ldots, \mathbf{P}_\ell\mathbf{A}\mathbf{V}_\ell$ are asymptotically free.

Example 2.49. [73] Let $\mathbf{P}_\ell$ be the permutation matrix corresponding to a cyclic shift by $\ell - 1$ entries, and $\mathbf{S}$ be a complex standard Gaussian matrix. Notice that $\mathbf{P}_\ell\mathbf{P}_\ell^\dagger = \mathbf{I}$ and that, for $\ell \neq 1(\text{mod}N)$, $\text{tr}\{\mathbf{P}_\ell\} = 0$. Consequently, for $N \to \infty$

$$\text{tr}\{\mathbf{P}_i\mathbf{P}_j^\dagger\} - \text{tr}\{\mathbf{P}_{i-j}\} = \delta_{i,j}. \tag{2.191}$$

Since $\mathbf{S}\mathbf{S}^\dagger$ and $\{\mathbf{P}_1, \mathbf{P}_1^\dagger, \ldots, \mathbf{P}_L, \mathbf{P}_L^\dagger\}$ are asymptotically free (e.g. Example 2.45), it follows from Theorem 2.61 that

$$\mathbf{P}_1\mathbf{S}\mathbf{S}^\dagger\mathbf{P}_1^\dagger, \ldots, \mathbf{P}_L\mathbf{S}\mathbf{S}^\dagger\mathbf{P}_L^\dagger$$

are asymptotically free. Let $\mathbf{S}_1, \ldots, \mathbf{S}_L$ be independent complex standard Gaussian matrices. The foregoing asymptotic freeness together with the fact that the asymptotic distribution of the asymptotically free matrices $\mathbf{P}_\ell\mathbf{S}_\ell\mathbf{S}_\ell^\dagger\mathbf{P}_\ell^\dagger$ does not depend on ℓ, implies that the asymptotic averaged empirical distributions of $\sum_{\ell=1}^{L}\mathbf{P}_\ell\mathbf{S}\mathbf{S}^\dagger\mathbf{P}_\ell^\dagger$ and of $\sum_{\ell=1}^{L}\mathbf{P}_\ell\mathbf{S}_\ell\mathbf{S}_\ell^\dagger\mathbf{P}_\ell$ are the same.

Theorem 2.62. [290, 190] Let $(\mathbf{P}, \{\mathbf{W}_1, \ldots, \mathbf{W}_\ell\})$ be asymptotically free Hermitian random matrices. $\mathbf{P}\mathbf{W}_1\mathbf{P}, \ldots, \mathbf{P}\mathbf{W}_\ell\mathbf{P}$ are asymptotically free if $\mathbf{P}$ is idempotent.

We note that, under the condition that $\mathbf{P}_\ell$ and $\mathbf{V}_\ell$ are unitary Haar matrices, Theorems 2.61 and 2.62 hold not only in terms of asymptotic freeness but also in terms of almost surely asymptotic freeness.

Theorem 2.63. [290, 190] Let $\mathbf{W}$ be a random matrix whose averaged spectrum converges to the circular law (2.99). Let $\mathbf{P}_1, \ldots, \mathbf{P}_\ell$ be a family of Hermitian random matrices asymptotically free of $\mathbf{W}$ such that $\mathbf{P}_i\mathbf{P}_j = \mathbf{P}_j\mathbf{P}_i = \delta_{i,j}\mathbf{P}_i$, then $\mathbf{W}\mathbf{P}_1\mathbf{W}^\dagger, \ldots, \mathbf{W}\mathbf{P}_\ell\mathbf{W}^\dagger$ are asymptotically free. This result also holds if the spectrum of $\mathbf{W}$ converges to the quarter circle law (1.21) or to the semicircle law (2.94), in which case the spectrum of $\mathbf{W}\mathbf{P}_j\mathbf{W}^\dagger$ converges to the Marčenko-Pastur law.

2.4.2 Sums of Asymptotically Free Random Matrices

Much of the practical usefulness of free probability stems from the following result.

Theorem 2.64. [285] If **A** and **B** are asymptotically free random matrices, then the R-transform of their sum satisfies

$$\mathsf{R}_{\mathbf{A}+\mathbf{B}}(z) = \mathsf{R}_{\mathbf{A}}(z) + \mathsf{R}_{\mathbf{B}}(z). \tag{2.192}$$

As a simple application of this important result, and in view of Example 2.24, we can verify the translation property

$$\mathsf{R}_{\mathbf{A}+\gamma\mathbf{I}}(z) = \mathsf{R}_{\mathbf{A}}(z) + \mathsf{R}_{\gamma\mathbf{I}}(z) = \mathsf{R}_{\mathbf{A}}(z) + \gamma. \tag{2.193}$$

Using Theorem 2.64 and the relationship between the R-transform and the η-transform (2.75)–(2.76) we can obtain:

Theorem 2.65. The η-transform of the sum of asymptotically free random matrices is

$$\eta_{\mathbf{A}+\mathbf{B}}(\gamma) = \eta_{\mathbf{A}}(\gamma_a) + \eta_{\mathbf{B}}(\gamma_b) - 1 \tag{2.194}$$

with γ_a, γ_b and γ satisfying the following pair of equations:

$$\gamma_a\, \eta_{\mathbf{A}}(\gamma_a) = \gamma\, \eta_{\mathbf{A}+\mathbf{B}}(\gamma) = \gamma_b\, \eta_{\mathbf{B}}(\gamma_b). \tag{2.195}$$

As a simple application of Theorem 2.64, let us sketch a heuristic argument for the key characterization (2.116) of the η-transform of the asymptotic spectrum of $\mathbf{HTH}^\dagger$. Let us assume that $\mathbf{H}$ is an $N \times K$ matrix whose entries are independent random variables with common variance $\frac{1}{N}$, while $\mathbf{T}$ is a deterministic positive real diagonal matrix. According to Example 2.40, we can write $\mathbf{HTH}^\dagger$ as the sum of asymptotically free matrices

$$\mathbf{HTH}^\dagger = \sum_{k=1}^{K} T_k \mathbf{h}_k \mathbf{h}_k^\dagger. \tag{2.196}$$

Thus, with $\zeta \geq 0$

$$\mathsf{R}_{\mathbf{HTH}^\dagger}(-\zeta) = \lim_{K\to\infty} \sum_{k=1}^{K} \mathsf{R}_{T_k \mathbf{h}_k \mathbf{h}_k^\dagger}(-\zeta) \tag{2.197}$$

$$= \lim_{K\to\infty} \frac{\beta}{K} \sum_{k=1}^{K} \frac{T_k}{1+T_k\zeta} \tag{2.198}$$

$$= \beta\, \frac{1-\eta_{\mathbf{T}}(\zeta)}{\zeta} \tag{2.199}$$

where (2.198) follows from (2.82) whereas (2.199) follows from the law of large numbers. Finally, using the relationship between the η-transform and the R-transform in (2.74) we obtain (2.116) letting $\zeta = \gamma \eta_{\mathbf{HTH}^\dagger}(\gamma)$, i.e.

$$\eta_{\mathbf{HTH}^\dagger}(\gamma) = 1 - \beta\,(1 - \eta_{\mathbf{T}}(\gamma\, \eta_{\mathbf{HTH}^\dagger}(\gamma)))\,. \tag{2.200}$$

Note that (2.197) has not been rigorously justified above, since it involves both the limit in the size of the matrices which is the basis for the claim of asymptotic freeness and a limit in the number of matrices.

The more general result (2.133)–(2.134) can be readily obtained from (2.194), (2.195) and (2.200).

For $\mathbf{T} = \mathbf{I}$, we recover the η-transform in (2.121) of the Marčenko-Pastur law. It is interesting to note that, in this special case, we are summing unit-rank matrices whose spectra consist of a $1 - \frac{1}{N}$ mass at 0 and a $\frac{1}{N}$ mass at a location that converges to 1. If we were to take the Nth classical convolution (inverting the sum of log-moment generating functions) of those distributions we would obtain asymptotically the Poisson distribution; however, the distribution we obtain by taking the Nth free convolution (inverting the sum of R-transforms) is the Marčenko-Pastur law. Thus, we can justifiably claim that the Marčenko-Pastur law is the free analog of the classical Poisson law.

The free analog of the Gaussian law is the semicircle law according to the celebrated free probability central limit theorem:

Theorem 2.66. [284] Let $\mathbf{A}_1, \mathbf{A}_2, \ldots$ be a sequence of $N \times N$ asymptotically free random matrices. Assume that $\phi(\mathbf{A}_i) = 0$ and $\phi(\mathbf{A}_i^2) = 1$.

Further assume that $\sup_i |\phi(\mathbf{A}_i^k)| < \infty$ for all k. Then, as $m, N \to \infty$, the asymptotic spectrum of

$$\frac{1}{\sqrt{m}}(\mathbf{A}_1 + \mathbf{A}_2 + \cdots \mathbf{A}_m) \tag{2.201}$$

converges in distribution to the semicircle law, that is, for every k,

$$\phi\left(\frac{(\mathbf{A}_1 + \mathbf{A}_2 + \cdots + \mathbf{A}_m)^k}{m^{\frac{k}{2}}}\right) \to \begin{cases} 0 & k \text{ odd} \\ \dfrac{1}{1+\frac{k}{2}}\dbinom{k}{\frac{k}{2}} & k \text{ even.} \end{cases}$$

A simple sketch of the main idea behind the proof of this result can be given in the case of asymptotically free matrices identically distributed. In this case, Theorem 2.64 implies that the R-transform of (2.201) equals

$$\sqrt{m}\,\mathsf{R}_{\mathbf{A}_1}\left(\tfrac{z}{\sqrt{m}}\right) = \sqrt{m}\phi(\mathbf{A}_1) + z\phi(\mathbf{A}_1^2) + \sqrt{m}\sum_{k=3}^{\infty} c_k \left(\tfrac{z}{\sqrt{m}}\right)^{k-1} \tag{2.202}$$

$$\to \quad z \tag{2.203}$$

which is the R-transform of the semicircle law (Example 2.25). Note that (2.202) follows from (2.84) while (2.203) follows from the fact that the free cumulants are bounded because of the assumption in Theorem 2.66. A similar approach can be followed to prove that the spectra of Gaussian Wigner matrices converges to the semicircle law. The key idea is that a Gaussian standard Wigner matrix can be written as the sum of two independent rescaled Gaussian standard Wigner matrices

$$\mathbf{W} = \frac{1}{\sqrt{2}}(\mathbf{X}_1 + \mathbf{X}_2). \tag{2.204}$$

Since the two matrices in the right side of (2.204) are asymptotically free, the R-transforms satisfy

$$\begin{aligned} \mathsf{R}_{\mathbf{W}}(z) &= \mathsf{R}_{\frac{\mathbf{X}_1}{\sqrt{2}}}(z) + \mathsf{R}_{\frac{\mathbf{X}_2}{\sqrt{2}}}(z) \\ &= \sqrt{2}\,\mathsf{R}_{\mathbf{W}}\left(\frac{z}{\sqrt{2}}\right) \end{aligned} \tag{2.205}$$

which admits the solution (cf. Example 2.25)

$$\mathsf{R}_{\mathbf{W}}(z) = z. \tag{2.206}$$

Example 2.50. Let $\mathbf{H}$ be an $N \times m$ random matrix whose entries are zero-mean i.i.d. Gaussian random variables with variance $\frac{1}{\sqrt{m\,N}}$ and denote $\frac{1}{N}\sqrt{m} = \varsigma$. Using Example 2.46, Theorem 2.66, and the fact that we can represent

$$\mathbf{H}\mathbf{H}^{\dagger} = \frac{1}{\sqrt{m}} \sum_{i}^{m} \mathbf{s}_i \mathbf{s}_i^{\dagger}$$

with $\mathbf{s}_i$ an N-dimensional vector whose entries are zero-mean i.i.d. with variance $\frac{1}{\sqrt{N}}$, it can be shown that as $N, m \to \infty$ with $\frac{N}{m} \to 0$, the asymptotic spectrum of the matrix

$$\mathbf{H}\mathbf{H}^{\dagger} - \varsigma\sqrt{N}\mathbf{I}$$

is the semicircle law. This result was also found using the moment approach, based on combinatorial tools, in [16] (without invoking Gaussianity) and in [57] using results on the asymptotic distribution of the zeros of Laguerre polynomials $L_N^m(\sqrt{Nm}x + m + N)$.

2.4.3 Products of Asymptotically Free Matrices

The S-transform plays an analogous role to the R-transform for products (instead of sums) of asymptotically free matrices, as the following theorem shows:[34]

Theorem 2.67. Let $\mathbf{A}$ and $\mathbf{B}$ be nonnegative asymptotically free random matrices. The S-transform of their product satisfies

$$\Sigma_{\mathbf{AB}}(x) = \Sigma_{\mathbf{A}}(x)\Sigma_{\mathbf{B}}(x). \tag{2.207}$$

Because of (2.69), it follows straightforwardly that the S-transform is the free analog of the Mellin transform in classical probability theory, whereas recall that the R-transform is the free analog of the log-moment generating function in classical probability theory.

Theorem 2.67 together with (2.86) yields

[34] Given the definition of the S-transform, we shall consider only nonnegative random matrices whose trace does not vanish asymptotically.

Theorem 2.68. Let **A** and **B** be nonnegative asymptotically free random matrices, then for $0 < \gamma < 1$,

$$\eta_{\mathbf{AB}}^{-1}(\gamma) = \frac{\gamma}{1-\gamma}\,\eta_{\mathbf{A}}^{-1}(\gamma)\,\eta_{\mathbf{B}}^{-1}(\gamma). \tag{2.208}$$

In addition, the following implicit relation is also useful:

$$\eta_{\mathbf{AB}}(\gamma) = \eta_{\mathbf{A}}\left(\frac{\gamma}{\Sigma_{\mathbf{B}}(\eta_{\mathbf{AB}}(\gamma)-1)}\right). \tag{2.209}$$

As an application of (2.209), we can obtain the key relation (2.116) from the S-transform of the Marčenko-Pastur law in (2.87)

$$\Sigma_{\mathbf{H}^\dagger\mathbf{H}}(x) = \frac{1}{1+\beta\, x}$$

provided that **T** and $\mathbf{H}^\dagger\mathbf{H}$ are asymptotically free. According to (2.209)

$$\begin{aligned}\eta_{\mathbf{TH}^\dagger\mathbf{H}}(\gamma) &= \eta_{\mathbf{T}}\left(\gamma\,(1-\beta+\beta\eta_{\mathbf{TH}^\dagger\mathbf{H}}(\gamma))\right)\\ &= \eta_{\mathbf{T}}\left(\gamma\,\eta_{\mathbf{HTH}^\dagger}(\gamma)\right) \qquad (2.210)\end{aligned}$$

where (2.210) follows from (2.56). Applying (2.56) again,

$$\eta_{\mathbf{HTH}^\dagger}(\gamma) = 1-\beta+\beta\,\eta_{\mathbf{T}}\left(\gamma\,\eta_{\mathbf{HTH}^\dagger}(\gamma)\right). \tag{2.211}$$

From (2.209), Examples 2.13, 2.32 and 2.45 we obtain the following result.

Example 2.51. Let **Q** be a $N \times K$ matrix uniformly distributed over the manifold of $N \times K$ complex matrices such that $\mathbf{Q}^\dagger\mathbf{Q} = \mathbf{I}$ and let **A** be an $N \times N$ nonnegative Hermitian random matrix independent of **Q** whose empirical eigenvalue distribution converges almost surely to a compactly supported measure. Then

$$\eta_{\mathbf{QQ}^\dagger\mathbf{A}}(\gamma) = \eta_{\mathbf{A}}\left(\gamma + \gamma\,\frac{\beta-1}{\eta_{\mathbf{QQ}^\dagger\mathbf{A}}(\gamma)}\right) \tag{2.212}$$

with $\frac{K}{N} \to \beta$.

Example 2.52. Define two $N \times N$ independent random matrices, $\mathbf{H}_1$ and $\mathbf{H}_2$, each having zero-mean i.i.d. entries with variance $\frac{1}{N}$ and higher order moments of order $o(1/N)$. From Example 2.39, $(\{\mathbf{H}_1, \mathbf{H}_1^\dagger\}, \{\mathbf{H}_2, \mathbf{H}_2^\dagger\})$ are asymptotically free and, consequently, we can compute the S-transform of $\mathbf{A}_2 = \mathbf{H}_1\mathbf{H}_2\mathbf{H}_2^\dagger\mathbf{H}_1^\dagger$ by simply applying Example 2.29 and Theorems 2.67 and 2.32:

$$\Sigma_{\mathbf{A}_2}(x) = \frac{1}{(x+1)^2} \tag{2.213}$$

from which it follows that the η-transform of $\mathbf{A}_2$, $\eta_{\mathbf{A}_2}(\gamma)$, is the solution of the fixed-point equation

$$\eta(1 + \gamma\eta^2) = 1. \tag{2.214}$$

Example 2.52 can be extended as follows.

Example 2.53. [184] Let $\mathbf{H}$ and $\mathbf{T}$ be as in Theorem 2.39. Then,

$$\Sigma_{\mathbf{HTH}^\dagger}(x) = \frac{x+1}{x+\beta}\,\Sigma_{\mathbf{H}^\dagger\mathbf{H}}\left(\frac{x}{\beta}\right)\,\Sigma_{\mathbf{T}}\left(\frac{x}{\beta}\right) \tag{2.215}$$

$$= \frac{1}{x+\beta}\,\Sigma_{\mathbf{T}}\left(\frac{x}{\beta}\right) \tag{2.216}$$

where (2.216) follows from Example 2.29.

Example 2.53 follows from the fact that, if $\mathbf{H}$ in Theorem 2.39 is a standard complex Gaussian matrix, then $(\{\mathbf{H}, \mathbf{H}^\dagger\}, \mathbf{T})$ are asymptotically free (cf. Example 2.43) and thus it follows from Theorem 2.67 that the S-transform of $\mathbf{HTH}^\dagger$ is given by (2.216). On the other hand, since the validity of Theorem 2.39 depends on the distribution of $\mathbf{H}$ only through the first and second order moments, every matrix $\mathbf{HTH}^\dagger$ defined as in Theorem 2.39 with $\mathbf{H}$ arbitrarily distributed admits the same asymptotic spectrum and the same R- and S-transforms and hence Example 2.53 follows straightforwardly. Analogous considerations hold for Theorems 2.38, 2.42 and 2.43. More precisely, the hypotheses in those theorems are sufficient to guarantee the additivity of the R-transforms and factorability of the S-transforms therein. Note, however, that the factorability of the S-transforms in (2.216) and the additivity of the R-transforms in Theorem 2.38 do not imply, in general, that $(\{\mathbf{H}, \mathbf{H}^\dagger\}, \mathbf{T})$ are asymptotically free.

2.4.4 Freeness and Non-Crossing Partitions

The combinatorial description of the freeness developed by Speicher in [241, 243] and in some of his joint works with A. Nica [189] has succeeded in obtaining a number of new results in free probability theory. It is well known that there exists a combinatorial description of the classical cumulants that is related to the partition theory of sets. In the same way, a noncommutative analogue to the classical cumulants, the so-called free cumulants, can be also described combinatorially. The key difference with the classical case is that one has to replace the partitions by so-called non-crossing partitions [241, 243].

Definition 2.22. Consider the set $\{1, \ldots, n\}$ and let ϖ be a partition of this set,

$$\varpi = \{V_1, \ldots, V_k\},$$

where each V_i is called a block of ϖ. A partition ϖ is called non-crossing if the following does not occur: there exist $1 \leq p_1 \leq q_1 \leq p_2 \leq q_2$ such that p_1 and p_2 belong to the same block, q_1 and q_2 belong to the same block, but q_1 and p_2 do not belong to the same block.

Example 2.54. Consider the set $\{1, 2, 3, 4\}$ and the non-crossing partition $\varpi = \{\{1, 3\}, \{2\}, \{4\}\}$. Definition 2.22 is interpreted graphically in Figure 2.7(a) by connecting elements in the same block with a line. The fact that these lines do not cross evidences the non-crossing nature of the partition. In contrast, the crossing partition $\varpi = \{\{1, 3\}, \{2, 4\}\}$ of the same set is also shown in Figure 2.7(b).

Example 2.55. Consider the set $\{1, 2, \ldots, 7\}$. Let V_1, V_2 and V_3 be a partition of $\{1, 2, \ldots, 7\}$ with $V_1 = \{1, 5, 7\}$, $V_2 = \{2, 3, 4\}$, and $V_3 = \{6\}$. Then $\{V_1, V_2, V_3\}$ is a non-crossing partition.

Every non-crossing partition ϖ, can be associated to a complementation map [154], denoted by $K(\varpi)$. Figure 2.8 depicts the non-crossing partition $\varpi = \{\{1, 5, 7\}, \{2, 3, 4\}, \{6\}\}$ and the corresponding complementation map $K(\varpi) = \{\{1, 4\}, \{2\}, \{3\}, \{5, 6\}, \{7\}\}$. The complementation map $K(\varpi)$ can be found graphically as follows: duplicate the

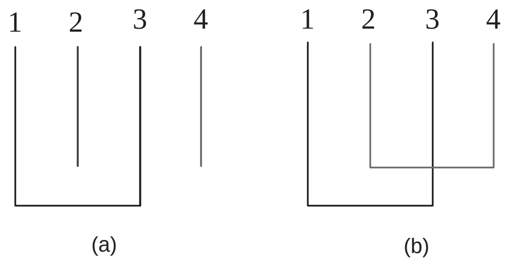

Fig. 2.7 Figures (a) and (b) depict a non-crossing and a crossing partition respectively.

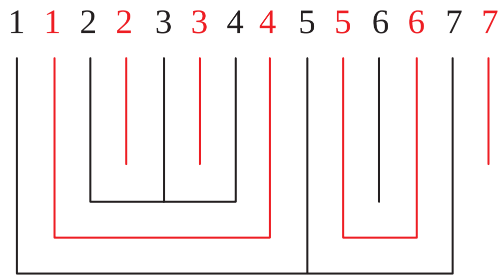

Fig. 2.8 The non-crossing partition $\varpi = \{\{1,5,7\},\{2,3,4\},\{6\}\}$ and the complementation map $K(\varpi) = \{\{1,4\},\{2\},\{3\},\{5,6\},\{7\}\}$ obtained with the repeated integers.

elements of the set placing them between the elements of the old set; then connect with a line as many elements of the new set as possible without crossing the lines of the original partition.

The number of non-crossing partitions of the set $\{1, 2, \ldots, n\}$ into i blocks equals[35]

$$Q_i = \frac{1}{n}\binom{n}{i}\binom{n}{i-1}.$$

Moreover, the number of non-crossing partitions of $\{1, 2, \ldots, n\}$ equals the nth Catalan number. This follows straightforwardly from the fact that

$$\sum_{i=1}^{n} Q_i = \frac{1}{n+1}\binom{2n}{n}.$$

[35] Note that $\sum_{i=1}^{n} Q_i \beta^i$ equals the n-th moment of $\tilde{\mathsf{f}}_\beta(\cdot)$ given in (1.12).

The following result gives a general expression of the joint moments of asymptotically free random matrices.

Theorem 2.69. [20, 21] Consider matrices $\mathbf{A}_1, \ldots, \mathbf{A}_\ell$ whose size is such that the product $\mathbf{A}_1 \ldots \mathbf{A}_\ell$ is defined. Some of these matrices are allowed to be identical. Omitting repetitions, assume that the matrices are asymptotically free.[36] Let ϱ be the partition of $\{1, \ldots, \ell\}$ determined by the equivalence relation[37] $j \equiv k$ if $i_j = i_k$. For each partition ϖ of $\{1, \ldots, \ell\}$, let

$$\phi_\varpi = \prod_{\substack{\{j_1, \ldots, j_r\} \in \varpi \\ j_1 < \ldots < j_r}} \phi(\mathbf{A}_{j_1} \ldots \mathbf{A}_{j_r}).$$

There exist universal coefficients $c(\varpi, \varrho)$ such that

$$\phi(\mathbf{A}_1 \ldots \mathbf{A}_\ell) = \sum_{\varpi \leq \varrho} c(\varpi, \varrho) \phi_\varpi$$

where $\varpi \leq \varrho$ indicates that ϖ is finer [38] than ϱ.

Finding an explicit formula for the coefficients $c(\varpi, \varrho)$ is a nontrivial combinatorial problem which has been solved by Speicher [241, 243]. From Theorem 2.69 it follows that $\phi(\mathbf{A}_1 \ldots \mathbf{A}_\ell)$ is completely determined by the moments of the individual matrices.

It is useful to highlight a special case of Theorem 2.69.

Theorem 2.70. [111] Assume that $\mathbf{A}$ and $\mathbf{B}$ are asymptotically free random matrices. Then, the moments of $\mathbf{A} + \mathbf{B}$ are expressed by the free cumulants of $\mathbf{A}$ and $\mathbf{B}$ as

$$\phi((\mathbf{A} + \mathbf{B})^n) = \sum_{\varpi} \prod_{V \in \varpi} (c_{|V|}(\mathbf{A}) + c_{|V|}(\mathbf{B})) \tag{2.217}$$

[36] For example, $(\mathbf{A}_1, \ldots, \mathbf{A}_4) = (\mathbf{B}, \mathbf{C}, \mathbf{C}, \mathbf{B})$ with $\mathbf{B}$ and $\mathbf{C}$ asymptotically free.

[37] If an equivalence relation is given on the set Ω, then the set of all equivalence classes forms a partition of Ω. Conversely, if a partition ϖ_1 is given on Ω, we can define an equivalence relation on Ω by writing $x \equiv y$ if and only if there exists a member of ϖ_1 which contains both x and y. The notions of "equivalence relation" and "partition" are thus essentially equivalent.

[38] Given two partitions ϖ_1 and ϖ_2 of a given set Ω, we say that ϖ_1 is finer than ϖ_2 if it splits the set Ω into smaller blocks, i.e., if every element of ϖ_1 is a subset of an element of ϖ_2. In that case, one writes $\varpi_1 \leq \varpi_2$.

where the summation is over all non-crossing partitions of $\{1, \ldots, n\}$, $c_\ell(\mathbf{A})$ denotes the ℓth free cumulant of $\mathbf{A}$ (cf. Section 2.2.5) and $|V|$ denotes the cardinality of V.

Theorem 2.70 is based on the fact that, if $\mathbf{A}$ and $\mathbf{B}$ are asymptotically free random matrices, the free cumulants of the sum satisfy $c_\ell(\mathbf{A}+\mathbf{B}) = c_\ell(\mathbf{A}) + c_\ell(\mathbf{B})$.

The counterpart of Theorem 2.70 for the product of two asymptotically free random matrices $\mathbf{A}$ and $\mathbf{B}$ is given by the following theorem.

Theorem 2.71. [111] Assume that $\mathbf{A}$ and $\mathbf{B}$ are asymptotically free random matrices. Then the moments of $\mathbf{AB}$ are expressed by the free cumulants of $\mathbf{A}$ and $\mathbf{B}$ as follows:

$$\phi((\mathbf{AB})^n) = \sum_{\varpi_1, \varpi_2} \prod_{V_1 \in \varpi_1} c_{|V_1|}(\mathbf{A}) \prod_{V_2 \in \varpi_2} c_{|V_2|}(\mathbf{B}) \tag{2.218}$$

where the summation is over all non-crossing partitions of $\{1, \ldots, n\}$.

2.5 Convergence Rates and Asymptotic Normality

Most of the literature on large random matrices has focused on the existence of the limiting spectral distributions employing the moment convergence theorem, i.e., verifying the convergence of the kth moments of the $N \times N$ random matrix to the moments of the target distribution either almost surely or in probability. While this method guarantees convergence, it gives no information on the speed of convergence. Loose bounds on the convergence rate to the semicircle law were put forth in 1998 by Girko [88]. A sharper result, but probably not the final word on the matter, was obtained recently:

Theorem 2.72. [95] Let $\mathbf{W}$ be an $N \times N$ Gaussian standard Wigner matrix. The maximal absolute difference between the expected empirical eigenvalue distribution of $\mathbf{W}$ and the semicircle law, F_w, whose density is given in (2.94), vanishes as

$$\|\mathbb{E}[\mathsf{F}^N_{\mathbf{W}}] - \mathsf{F}_w\| \leq \kappa N^{-2/3} \tag{2.219}$$

with κ a positive constant and with $\|f - g\| = \sup_x |f(x) - g(x)|$.

For an arbitrary deterministic sequence a_N, the notation

$$\xi_N = O_p(a_N) \tag{2.220}$$

means[39] that, for any ϵ, there exists an $\varsigma > 0$ such that

$$\sup_N P[|\xi_N| \geq \varsigma a_N] < \epsilon. \tag{2.221}$$

Similarly, the notation

$$\xi_N = o(a_N) \quad \text{a. s.} \tag{2.222}$$

means that $a_N^{-1}\xi_N \to 0$ almost surely.

Theorem 2.73. [11] Let $\mathbf{W}$ be an $N \times N$ standard Wigner matrix such that $\sup_{i,j,N} \mathbb{E}[|\sqrt{N}\,\mathsf{W}_{i,j}|^8] < \infty$ and that, for any positive constant δ,

$$\sum_{i,j} \mathbb{E}\left[\left|\sqrt{N}\,\mathsf{W}_{i,j}\right|^8 1\{|\mathsf{W}_{i,j}| \geq \delta\}\right] = o(N^2). \tag{2.223}$$

Then,

$$\|\mathsf{F}_{\mathbf{W}}^N - \mathsf{F}_w\| = O_p(N^{-2/5}). \tag{2.224}$$

If we further assume that all entries of $\sqrt{N}\mathbf{W}$ have finite moments of all orders, then for any $\eta > 0$, the empirical distribution of the Wigner matrix tends to the semicircle law as

$$\|\mathsf{F}_{\mathbf{W}}^N - \mathsf{F}_w\| = o(N^{-2/5+\eta}) \qquad \text{a. s.} \tag{2.225}$$

If we relax the assumption on the entries of $\sqrt{N}\mathbf{W}$ to simply finite fourth-order moments, then the convergence rates for $\mathsf{F}_{\mathbf{W}}^N$ and $\mathbb{E}[\mathsf{F}_{\mathbf{W}}^N]$ have been proved in [8] to reduce to

$$\|\mathbb{E}[\mathsf{F}_{\mathbf{W}}^N] - \mathsf{F}_w\| = O(N^{-1/4}) \tag{2.226}$$

$$\|\mathsf{F}_{\mathbf{W}}^N - \mathsf{F}_w\| = O_p(N^{-1/4}). \tag{2.227}$$

In the context of random matrices of the form $\mathbf{H}\mathbf{H}^\dagger$ the following results have been obtained.

[39] It is common in the literature to say that a sequence of random variables is *tight* if it is $O_p(1)$.

Theorem 2.74. [12] Let $\mathbf{H}$ be an $N \times K$ matrix whose entries are mutually independent with zero mean and variance $\frac{1}{N}$. Assume that

$$\sup_{i,j,N} \mathbb{E}\left[|\sqrt{N}\,\mathsf{H}_{i,j}|^8\right] < \infty \tag{2.228}$$

and for any positive constant δ

$$\sum_{i,j} \mathbb{E}\left[\left|\sqrt{N}\,\mathsf{H}_{i,j}\right|^8 1\{|\mathsf{H}_{i,j}| \geq \delta\}\right] = o(N^2). \tag{2.229}$$

Then, the maximal absolute difference between the expected empirical eigenvalue distribution of $\mathbf{H}^\dagger\mathbf{H}$ and the Marčenko-Pastur law, F_β, whose density is given in (1.10), vanishes as

$$\|\mathbb{E}[\mathsf{F}^N_{\mathbf{H}^\dagger\mathbf{H}}] - \mathsf{F}_\beta\| = O\left(\frac{N^{-\frac{1}{4\theta+2}}}{1-\sqrt{\beta}+N^{-\frac{1}{8\theta+4}}}\right) \tag{2.230}$$

and

$$\|\mathsf{F}^N_{\mathbf{H}^\dagger\mathbf{H}} - \mathsf{F}_\beta\| = O_p\left(\max\left\{\frac{N^{-\frac{2}{5+\theta}}}{1-\sqrt{\beta}+N^{-\frac{1}{5+\theta}}}, \frac{N^{-\frac{1}{4\theta+2}}}{1-\sqrt{\beta}+N^{-\frac{1}{8\theta+4}}}\right\}\right) \tag{2.231}$$

with

$$\theta = \begin{cases} \frac{-2\log(1-\sqrt{\frac{K}{N}})}{\log N + 4\log(1-\sqrt{\frac{K}{N}})} & \text{if } \frac{K}{N} \leq (1-N^{\frac{1}{8}})^2, \\ \frac{1}{2} & \text{otherwise.} \end{cases} \tag{2.232}$$

Summarizing, if $\beta < 1$ then $\theta \sim c/\log N$ and hence the convergence rates in (2.230) and (2.231) are $O(N^{-1/2})$ and $O_p(N^{-2/5})$, respectively. When $\beta > 1$, $\theta = \frac{1}{2}$ and the rates are $O(N^{-1/8})$ and $O_p(N^{-1/8})$, respectively. For $\beta = 1$, the exact speed at which $\frac{K}{N} \to 1$ matters as far as Theorem 2.74 is concerned.

Theorem 2.75. [87, 15] Let $\mathbf{H}$ be an $N \times K$ complex matrix whose entries are i.i.d. zero-mean random variables with variance $\frac{1}{N}$ such that $\mathbb{E}[|\sqrt{N}\mathsf{H}_{i,j}|^4] = 2$. Define the random variable

$$\begin{aligned} \Delta_N &= \log\det(\mathbf{H}^\dagger\mathbf{H}) - K\int_a^b \log(x)\,\mathsf{f}_\beta(x)dx \qquad (2.233) \\ &= \log\det(\mathbf{H}^\dagger\mathbf{H}) + K\left(\frac{1-\beta}{\beta}\log(1-\beta)+\log e\right) \end{aligned}$$

with $\mathsf{f}_\beta(\cdot)$ the density of the Marčenko-Pastur law in (1.10). As $K, N \to \infty$ with $\frac{K}{N} \to \beta \leq 1$, Δ_N converges to a Gaussian random variable with zero mean and variance

$$\mathbb{E}\left[|\Delta|^2\right] = \log \frac{1}{1-\beta}. \tag{2.234}$$

The counterpart of Theorem 2.75 for real $\mathbf{H}$ was first derived by Jonsson in [131] for a real zero-mean matrix with Gaussian i.i.d. entries and an analogous result has been found by Girko in [87] for real (possible nonzero-mean) matrix with i.i.d. entries and variance $\frac{1}{N}$. In the special case of Gaussian entries, Theorem 2.75 can be easily obtained following [131] using the expression of the moment-generating function of $\log\det(\mathbf{H}^\dagger\mathbf{H})$ in (2.11). In the general case, Theorem 2.75 can be easily verified using the result given in [15].

Theorem 2.76. [15] Let $\mathbf{H}$ be an $N \times K$ complex matrix whose entries are i.i.d. zero-mean random variables with variance $\frac{1}{N}$ such that $\mathbb{E}[|\mathsf{H}_{i,j}|^4] = \frac{2}{N^2}$. Denote by $\mathcal{V}_\beta(\gamma)$ the Shannon transform of $\tilde{\mathsf{f}}_\beta(\cdot)$ (Example 2.117). As $K, N \to \infty$ with $\frac{K}{N} \to \beta$, the random variable

$$\begin{aligned} \Delta_N &= \log\det(\mathbf{I} + \gamma\mathbf{H}\mathbf{H}^\dagger) - N\int_a^b \log(1+\gamma x)\,\tilde{\mathsf{f}}_\beta(x)\,dx \\ &= \log\det(\mathbf{I} + \gamma\mathbf{H}\mathbf{H}^\dagger) - N\mathcal{V}_\beta(\gamma) \end{aligned} \tag{2.235}$$

is asymptotically Gaussian with zero mean and variance

$$\begin{aligned} \mathbb{E}\left[\Delta^2\right] &= -\log\left(1 - \frac{(1-\eta_{\mathbf{H}\mathbf{H}^\dagger}(\gamma))^2}{\beta}\right) \\ &= -\log\left(1 - \frac{1}{\beta}\left(\frac{\mathcal{F}(\gamma,\beta)}{4\gamma}\right)^2\right). \end{aligned} \tag{2.236}$$

Notice that

$$\lim_{\gamma\to\infty} \frac{\mathcal{F}(\gamma,\beta)}{4\gamma} = \min\{1,\beta\} \tag{2.237}$$

and Theorem 2.75 can be obtained as special case.

Theorem 2.77. [15] Let $\mathbf{H}$ be an $N \times K$ complex matrix defined as in Theorem 2.76. Let $\mathbf{T}$ be an Hermitian random matrix independent of $\mathbf{H}$ with bounded spectral norm and whose asymptotic spectrum converges almost surely to a nonrandom limit. Denote by $\mathcal{V}_{\mathbf{HTH}^\dagger}(\gamma)$ the Shannon transform of $\mathbf{HTH}^\dagger$. As $K, N \to \infty$ with $\frac{K}{N} \to \beta$, the random variable

$$\Delta_N \quad = \quad \log\det(\mathbf{I} + \gamma\mathbf{HTH}^\dagger) - N\mathcal{V}_{\mathbf{HTH}^\dagger}(\gamma) \tag{2.238}$$

is asymptotically zero-mean Gaussian with variance

$$\mathbb{E}[\Delta^2] = -\log\left(1 - \frac{(1-\eta_{\mathbf{HTH}^\dagger}(\gamma))^2}{\beta}\right). \tag{2.239}$$

More general results (for functions other than $\log(1+\gamma x)$) are given in [15].

Theorem 2.78. [15] Let $\mathbf{H}$ be an $N \times K$ complex matrix defined as in Theorem 2.76. Let $\mathbf{T}$ be a $K \times K$ nonnegative definite deterministic matrix defined as in Theorem 2.77. Let $g(\cdot)$ be a continuous function on the real line with bounded and continuous derivatives, analytic on a open set containing the interval[40]

$$\left[\liminf_N \phi_N \max{}^2\{0, 1-\sqrt{\beta}\},\ \limsup_N \phi_1(1+\sqrt{\beta})^2\right]. \tag{2.240}$$

where $\phi_1 \geq \ldots \geq \phi_N$ are the eigenvalues of $\mathbf{T}$. Denoting by λ_i the ith eigenvalue of $\mathbf{HTH}^\dagger$, the random variable

$$\boldsymbol{\Delta}_N = \sum_{i=1}^{N} g(\lambda_i) - N\int g(x)\, d\mathsf{F}_{\mathbf{HTH}^\dagger} \tag{2.241}$$

converges, as $K, N \to \infty$ with $\frac{K}{N} \to \beta$, to a zero-mean Gaussian random variable.[41]

[40] In [14, 13, 170, 222] this interval contains the spectral support of $\mathbf{H}^\dagger\mathbf{HT}$.

[41] See [15] for an expression of the variance of the limit.

3

Applications to Wireless Communications

In this section, we detail some of the more representative problems described by (1.1) that capture various features of interest in wireless communications and we show how random matrix results have been used to characterize the fundamental limits of the various channels that arise in wireless communications.

Unless otherwise stated, the analysis applies to coherent reception and thus it is presumed that the state of the channel is perfectly tracked by the receiver. The degree of channel knowledge at the transmitter, on the other hand, is specified for each individual setting.

3.1 Direct-Sequence CDMA

The analysis of randomly-spread DS-CDMA in the asymptotic regime of number of users, K, and spreading gain, N, going to infinity with $\frac{K}{N} \to \beta$ provides valuable insight into the behavior of multiuser receivers for large DS-CDMA systems employing pseudo-noise spreading sequences (e.g. [167, 275, 256, 100, 217, 30]).

The standard random signature model [271, Sec. 2.3.5] assumes that the entries of the matrix $\mathbf{S}$, whose columns are the spreading sequences, are chosen independently and equiprobably on $\{\frac{-1}{\sqrt{N}}, \frac{1}{\sqrt{N}}\}$. A motivation for this is the use of "long sequences" in 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. Sometimes (particularly when the random sequence setting is used to model to some extent nonideal effects such as asynchronism and the frequency selectivity of the channel) the signatures are assumed to be uniformly distributed on the unit Euclidean N-dimensional sphere (a case for which the Marcenko-Pastur law also applies). In the analysis that follows, the only condition on the signature sequences is that their entries be i.i.d. zero-mean with variance $\frac{1}{N}$.

Specializing the general model in (1.1) to DS-CDMA, the vector $\mathbf{x}$ contains the symbols transmitted by the K users, which have zero-mean and equal variance. The entries of $\mathbf{x}$ correspond to different users and are therefore independent. (Unequal-power users will be accommodated by pre-multiplying $\mathbf{x}$ by an additional diagonal matrix of amplitudes.)

3.1.1 Unfaded Equal-Power DS-CDMA

With equal-power transmission at every user and no fading, the multi-access channel model becomes [271, Sec. 2.9.2]

$$\mathbf{y} = \mathbf{S}\mathbf{x} + \mathbf{n}, \tag{3.1}$$

where the energy per symbol transmitted from each user divided by the noise variance per chip is denoted by SNR, i.e.,

$$\mathsf{SNR} = \frac{\mathbb{E}[\|\mathbf{x}\|^2]}{\frac{1}{N}\mathbb{E}[\|\mathbf{n}\|^2]}.$$

Asymptotic analyses have been reported in the literature for various receivers, including:

- Single-user matched filter
- Decorrelator
- MMSE
- Optimum
- Iterative nonlinear.

The asymptotic analysis of the single-user matched filter (both uncoded error probability and capacity) has relied on the central limit theorem rather than on random matrix techniques [275]. The asymptotic analysis of the uncoded error probability has not used random matrix techniques either: [258] used large-deviation techniques to obtain the asymptotic efficiency and [249] used the replica method of statistical physics to find an expression for the uncoded bit error rate (see also [103]). The optimum near-far resistance and the MMSE were obtained in [271] using the Marčenko-Pastur law (Theorem 2.35). Recall, from (1.12), that the asymptotic fraction of zero eigenvalues of $\mathbf{HH}^\dagger$ is given by $(1-\beta)^+$. Then, for $\beta \leq 1$, using (2.57), the decorrelator achieves an output SINR that converges asymptotically to [271, (4.111)]

$$(1-\beta)\,\mathsf{SNR}\,. \tag{3.2}$$

When $\beta > 1$, the Moore-Penrose generalized-inverse decorrelator [271, Sec. 5.1] is shown in [70] (also using the Marčenko-Pastur law) to attain an asymptotic SINR ratio equal to

$$\frac{\beta-1}{(\beta-1)^2+\beta/\mathsf{SNR}}. \tag{3.3}$$

Using (2.57) and (2.121), the maximum SINR (achieved by the MMSE linear receiver) converges to [271, (6.59)]

$$\mathsf{SNR} - \frac{\mathcal{F}\left(\mathsf{SNR}\,,\beta\right)}{4} \tag{3.4}$$

with $\mathcal{F}(\cdot,\cdot)$ defined in (1.17) while the MMSE converges to

$$1-\frac{\mathcal{F}\left(\mathsf{SNR}\,,\beta\right)}{4\,\mathsf{SNR}\,\beta}. \tag{3.5}$$

Incidentally, note that, as $\mathsf{SNR} \to \infty$, (3.3) and (3.4) converge to the same quantity if $\beta > 1$.

The total capacity (sum-rate) of the multiaccess channel (3.1) was obtained in [275] for the linear receivers listed above and the optimum receiver also using the Marčenko-Pastur law. These expressions for the decorrelator and MMSE receiver are

$$\mathcal{C}^{\mathsf{dec}}(\beta,\mathsf{SNR}\,) = \beta\log\left(1+\mathsf{SNR}\,(1-\beta)\right), \quad 0\leq\beta\leq 1 \tag{3.6}$$

and

$$\mathcal{C}^{\mathsf{mmse}}(\beta, \mathsf{SNR}) = \beta \log\left(1 + \mathsf{SNR} - \frac{\mathcal{F}(\mathsf{SNR}, \beta)}{4}\right) \tag{3.7}$$

while the capacity achieved with the optimum receiver is (1.14)

$$\begin{aligned}\mathcal{C}^{\mathsf{opt}}(\beta, \mathsf{SNR}) &= \beta \log\left(1 + \mathsf{SNR} - \frac{\mathcal{F}(\mathsf{SNR}, \beta)}{4}\right) \\ &\quad + \log\left(1 + \mathsf{SNR}\,\beta - \frac{\mathcal{F}(\mathsf{SNR}, \beta)}{4}\right) - \frac{\mathcal{F}(\mathsf{SNR}, \beta)}{4\mathsf{SNR}} \log e.\end{aligned} \tag{3.8}$$

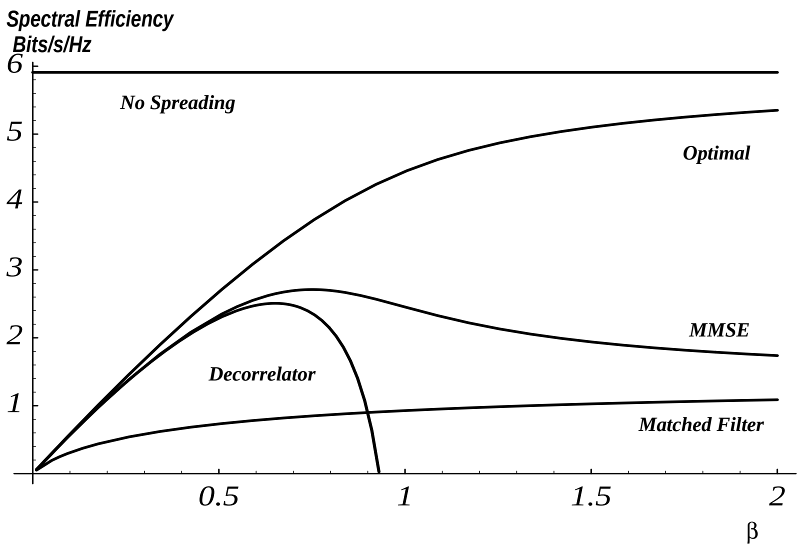

Fig. 3.1 Capacity of CDMA without fading for $\frac{E_b}{N_0} = 10$dB.

Figure 3.1 (from [275]) compares (3.6), (3.7) and (3.8) as a function of the number of users to spreading gain β, choosing SNR so that $\beta \mathsf{SNR} / \mathcal{C}(\beta, \mathsf{SNR}) = \frac{E_b}{N_0} = 10$.

3.1.2 DS-CDMA with Frequency-Flat Fading

When the users are affected by different attenuations which may vary from symbol to symbol, it is convenient to model the channel gains seen by each user as random quantities $\{|\mathsf{A}_1|^2, \ldots, |\mathsf{A}_K|^2\}$ whose empirical distribution converges almost surely to a nonrandom limit as the number of users goes to infinity. In this case, the channel matrix $\mathbf{H}$ can be written as the product of the $N \times K$ matrix $\mathbf{S}$ containing the spreading sequences with a $K \times K$ diagonal matrix $\mathbf{A}$ of complex fading coefficients such that the linear model in (1.1) becomes

$$\mathbf{y} = \mathbf{SAx} + \mathbf{n}. \tag{3.9}$$

Here, the role of the received signal-to-noise ratio of the kth user is taken by $|\mathsf{A}_k|^2 \mathsf{SNR}$.

The η-transform is intimately related to the performance of MMSE multiuser detection of (3.1). The arithmetic mean of the MMSEs for the K users satisfies [271, (6.27)]

$$\frac{1}{K}\sum_{k=1}^{K} \mathsf{MMSE}_k = \frac{1}{K}\mathrm{tr}\left\{\left(\mathbf{I} + \mathsf{SNR}\,\mathbf{A}^\dagger\mathbf{S}^\dagger\mathbf{SA}\right)^{-1}\right\} \tag{3.10}$$

$$\to \eta_{\mathbf{A}^\dagger\mathbf{S}^\dagger\mathbf{SA}}(\mathsf{SNR}) \tag{3.11}$$

whereas the multiuser efficiency of the kth user (output SINR relative to the single-user signal-to-noise ratio) achieved by the MMSE receiver, $\eta_k^{\mathsf{mmse}}(\mathsf{SNR})$, is[1]

$$\eta_k^{\mathsf{mmse}}(\mathsf{SNR}) = \mathbf{s}_k^T\left(\mathbf{I} + \sum_{i\neq k}\mathsf{SNR}\,|\mathsf{A}_i|^2\mathbf{s}_i\mathbf{s}_i^T\right)^{-1}\mathbf{s}_k \tag{3.12}$$

$$\to \eta_{\mathbf{SAA}^\dagger\mathbf{S}^\dagger}(\mathsf{SNR}) \tag{3.13}$$

where the limit follows from (2.57). According to Theorem 2.39, the MMSE multiuser efficiency, abbreviated as

$$\eta = \eta_{\mathbf{SAA}^\dagger\mathbf{S}^\dagger}(\mathsf{SNR}), \tag{3.14}$$

[1] The conventional notation for multiuser efficiency is η [271]; the relationship in (3.13) is the motivation for the choice of the η-transform terminology introduced in Section 2.2.2.

is the solution to the fixed-point equation

$$1 - \eta = \beta \left(1 - \eta_{|\mathsf{A}|^2}(\mathsf{SNR}\,\eta)\right), \tag{3.15}$$

where $\eta_{|\mathsf{A}|^2}$ is the η-transform of the asymptotic empirical distribution of $\{|\mathsf{A}_1|^2, \ldots, |\mathsf{A}_K|^2\}$. A fixed-point equation equivalent to (3.15) was given in [256] and its generalization to systems with symbol-level asynchronism (but still chip-synchronous) is studied in [152].

The distribution of the output SINR is asymptotically Gaussian [257], in the sense of Theorem 2.78, and its variance decreases as $\frac{1}{N}$. The same holds for the decorrelator. Closed-form expressions for the asymptotic mean are $\mathsf{SNR}\;\eta_{\mathbf{SAA}^\dagger\mathbf{S}^\dagger}$ for the MMSE receiver and $\mathsf{SNR}\,(1 - \beta\mathbb{P}[|\mathsf{A}| > 0])$ for the decorrelator with $\beta < 1$ while the variance, for both receivers, is obtained in [257].[2]

In [217], the spectral efficiencies achieved by the MMSE receiver and the decorrelator are given respectively by

$$\mathcal{C}^{\mathsf{mmse}}(\beta, \mathsf{SNR}) = \beta\,\mathbb{E}\left[\log\left(1 + |\mathsf{A}|^2 \mathsf{SNR}\,\eta_{\mathbf{SAA}^\dagger\mathbf{S}^\dagger}(\mathsf{SNR})\right)\right] \tag{3.16}$$

and, for $\beta \leq 1$,

$$\mathcal{C}^{\mathsf{dec}}(\beta, \mathsf{SNR}) = \beta\,\mathbb{E}\left[\log\left(1 + |\mathsf{A}|^2 \mathsf{SNR}\,(1 - \beta\mathbb{P}[|\mathsf{A}| > 0])\right)\right] \tag{3.17}$$

where the distribution of $|\mathsf{A}|^2$ is given by the asymptotic empirical distribution of $\mathbf{AA}^\dagger$ and (3.17) follows from Corollary 2.2 using the fact that the multiuser efficiency of the kth user achieved by the decorrelator, η_k^{dec}, equals that of the MMSE as the noise vanishes [271].

Also in [217], the capacity of the optimum receiver is characterized in terms of the MMSE spectral efficiency: [3]

$$\begin{aligned}\mathcal{C}^{\mathsf{opt}}(\beta, \mathsf{SNR}) \;&=\; \mathcal{C}^{\mathsf{mmse}}(\beta, \mathsf{SNR}) + \log \frac{1}{\eta_{\mathbf{SAA}^\dagger\mathbf{S}^\dagger}(\mathsf{SNR})} \\ &\quad + (\eta_{\mathbf{SAA}^\dagger\mathbf{S}^\dagger}(\mathsf{SNR}) - 1)\log e. \end{aligned} \tag{3.18}$$

[2] Although most fading distributions of practical interest do not have any point masses at zero, we express various results without making such an assumption on the fading distribution. For example, the inactivity of certain users or groups of users can be modelled by nonzero point masses in the fading distribution.

[3] Equation (3.18) also holds for the capacity with non-Gaussian inputs, as shown in [186] and [103] using statistical-physics methods.

This result can be immediately obtained by specializing Theorem 2.44 to the case where $\mathbf{T} = \mathbf{A}\mathbf{A}^\dagger$ and $\mathbf{D} = \mathbf{I}$. Here we give the derivation in [217], which illustrates the usefulness of the interplay between the η and Shannon transforms. From the definition of Shannon transform, the capacity of the optimum receiver coincides with the Shannon transform of the matrix evaluated at SNR, i.e.,

$$\mathcal{C}^{\mathsf{opt}}(\beta, \mathsf{SNR}) = \mathcal{V}_{\mathbf{S}\mathbf{A}\mathbf{A}^\dagger\mathbf{S}^\dagger}(\mathsf{SNR}). \tag{3.19}$$

Furthermore, also from the definition of Shannon transform and (3.16), it follows that

$$\mathcal{C}^{\mathsf{mmse}}(\beta, \mathsf{SNR}) = \beta\, \mathcal{V}_{\mathbf{A}\mathbf{A}^\dagger}\left(\mathsf{SNR}\, \eta_{\mathbf{S}\mathbf{A}\mathbf{A}^\dagger\mathbf{S}^\dagger}(\mathsf{SNR})\right) \tag{3.20}$$

and we know from (2.61) that

$$\frac{\gamma}{\log e}\frac{d}{d\gamma}\mathcal{V}_X(\gamma) = 1 - \eta_X(\gamma). \tag{3.21}$$

Thus, using the shorthand in (3.14),

$$\begin{aligned} \frac{d}{d\mathsf{SNR}}\mathcal{C}^{\mathsf{mmse}}(\mathsf{SNR}, \beta) &= \beta\, \frac{1 - \eta_{\mathbf{A}\mathbf{A}^\dagger}(\mathsf{SNR}\,\eta)}{\mathsf{SNR}} \left(1 + \frac{\mathsf{SNR}\,\dot{\eta}}{\eta}\right) \log e \\ &= \frac{1-\eta}{\mathsf{SNR}} \left(1 + \frac{\mathsf{SNR}\,\dot{\eta}}{\eta}\right) \log e \end{aligned} \tag{3.22}$$

where we used (3.15) to write (3.22). The derivative of (3.19) yields

$$\frac{d}{d\mathsf{SNR}}\mathcal{C}^{\mathsf{opt}}(\beta, \mathsf{SNR}) = \frac{1-\eta}{\mathsf{SNR}} \log e. \tag{3.23}$$

Subtracting the right-hand sides of (3.22) and (3.23),

$$\frac{d}{d\mathsf{SNR}}\mathcal{C}^{\mathsf{opt}}(\beta, \mathsf{SNR}) - \frac{d}{d\mathsf{SNR}}\mathcal{C}^{\mathsf{mmse}}(\mathsf{SNR}, \beta) = \dot{\eta}\left(1 - \frac{1}{\eta}\right)\log e, \tag{3.24}$$

which is equivalent to (3.18) since, at $\mathsf{SNR} = 0$, both functions equal 0.

Random matrix methods have also been used to optimize power control laws in DS-CDMA, as the number of users goes to infinity, for various receivers: matched filter, decorrelator, MMSE and optimum receiver [217, 281].

Departing from the usual setup where the channel and spreading sequences are known by the receiver, the performance of blind and group-blind linear multiuser receivers that have access only to the received spreading sequence of the user of interest is carried out via random matrix techniques in [318]. The asymptotic SINR at the output of direct matrix inversion blind MMSE, subspace blind MMSE and group-blind MMSE receivers with binary random spreading is investigated and an interesting saturation phenomenon is observed. This indicates that the performance of blind linear multiuser receivers is not only limited by interference, but by estimation errors as well. The output residual interference is shown to be zero-mean and Gaussian with variance depending on the type of receiver.

3.1.3 DS-CDMA with Flat Fading and Antenna Diversity

Let us now study the impact of having, in addition to frequency-flat fading, L receive antennas at the base station. The channel matrix is now the $NL \times K$ array

$$\mathbf{H} = \begin{bmatrix} \mathbf{SA}_1 \\ \cdots \\ \mathbf{SA}_L \end{bmatrix} \tag{3.25}$$

where

$$\mathbf{A}_\ell = \mathrm{diag}\{\mathsf{A}_{1,\ell}, \ldots, \mathsf{A}_{K,\ell}\}, \qquad \ell = 1, \ldots L \tag{3.26}$$

and $\{\mathsf{A}_{k,\ell}\}$ indicates the i.i.d. fading coefficients of the kth user at the ℓth antenna.

Assuming that the fading coefficients are bounded,[4] using Lemma 2.60, [108] shows that the asymptotic averaged empirical singular value distribution of (3.25) is the same as that of

$$\begin{bmatrix} \mathbf{S}_1\mathbf{A}_1 \\ \cdots \\ \mathbf{S}_L\mathbf{A}_L \end{bmatrix}$$

where $\mathbf{S}_k$ for $k \in \{1, \ldots, L\}$ are i.i.d. matrices. Consequently, Theorem 2.50 leads to the conclusion that

$$1 - \eta_{\mathbf{HH}^\dagger} = \frac{\beta}{L}\left(1 - \eta_{\mathsf{P}}\left(\mathrm{SNR}\ \eta_{\mathbf{HH}^\dagger}\right)\right), \tag{3.27}$$

[4] This assumption is dropped in [160].

where η_{P} is the η-transform of the asymptotic empirical distribution of $P_1,\ldots,P_K$ with $P_k = \sum_{\ell=1}^{L} |\mathsf{A}_{k,\ell}|^2$. This result admits the pleasing engineering interpretation that the effective spreading gain is equal to the CDMA spreading gain times the number of receive antennas (but, of course, the bandwidth only grows with the CDMA spreading gain).

From the above result it follows that the expected arithmetic mean of the MMSE's for the K users converges to

$$\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[\mathsf{MMSE}_k] = \frac{1}{K}\mathbb{E}\left[\mathrm{tr}\left\{\left(\mathbf{I}+\mathsf{SNR}\,\mathbf{H}^\dagger\mathbf{H}\right)^{-1}\right\}\right] \tag{3.28}$$

$$\to \eta_{\mathbf{H}^\dagger\mathbf{H}}(\mathsf{SNR}). \tag{3.29}$$

Moreover, the MMSE multiuser efficiency, $\eta_k^{\mathsf{mmse}}(\mathsf{SNR})$, converges in probability as $K, N \to \infty$ to [108]

$$\eta_k^{\mathsf{mmse}}(\mathsf{SNR}) \to \eta_{\mathbf{H}\mathbf{H}^\dagger} \tag{3.30}$$

while the asymptotic multiuser efficiency is given by

$$\lim_{\mathsf{SNR}\to\infty}\eta_k^{\mathsf{mmse}}(\mathsf{SNR}) = 1-\min\left\{\frac{\beta}{L}\,\mathbb{P}[\mathsf{P}\neq 0],1\right\} \tag{3.31}$$

where P is a random variable distributed according to the asymptotic empirical distribution of $P_1,\ldots,P_K$. The spectral efficiency for MMSE and decorrelator and the capacity of the optimum receiver are

$$\begin{aligned}\mathcal{C}^{\mathsf{mmse}}(\beta,\mathsf{SNR}) &= \beta\,\mathcal{V}_{\mathsf{P}}\left(\mathsf{SNR}\,\eta_{\mathbf{H}\mathbf{H}^\dagger}(\mathsf{SNR})\right)\\ &= \beta\,\mathbb{E}\left[\log\left(1+\mathsf{SNR}\,\mathsf{P}\,\eta_{\mathbf{H}\mathbf{H}^\dagger}(\mathsf{SNR})\right)\right]\end{aligned} \tag{3.32}$$

and, using Corollary 2.2, for $\beta \le 1$

$$\mathcal{C}^{\mathsf{dec}}(\beta,\mathsf{SNR}) = \beta\,\mathbb{E}\left[\log\left(1+\mathsf{SNR}\,\mathsf{P}\left(1-\frac{\beta}{L}\,\mathbb{P}[\mathsf{P}>0]\right)\right)\right] \tag{3.33}$$

while

$$\begin{aligned}\mathcal{C}^{\mathsf{opt}}(\beta,\mathsf{SNR}) &= \mathcal{C}^{\mathsf{mmse}}(\beta,\mathsf{SNR}) + \log\frac{1}{\eta_{\mathbf{H}\mathbf{H}^\dagger}(\mathsf{SNR})}\\ &\quad +(\eta_{\mathbf{H}\mathbf{H}^\dagger}(\mathsf{SNR})-1)\log e.\end{aligned} \tag{3.34}$$

Note the parallel between (3.32–3.34) and (3.16–3.18).

3.1.4 DS-CDMA with Frequency-Selective Fading

Let us consider a synchronous DS-CDMA uplink with K active users employing random spreading codes and operating over a frequency-selective fading channel. The base station is equipped with a single receive antenna.

Assuming that the symbol duration ($T_s \approx \frac{N}{W_c}$ with W_c the chip-bandwidth) is much larger than the delay spread, we can disregard the intersymbol interference. In this case, the channel matrix in (1.1) particularizes to

$$\mathbf{H} = [\mathbf{C}_1\,\mathbf{s}_1\,, \ldots\,, \mathbf{C}_K\,\mathbf{s}_K]\mathbf{A} \tag{3.35}$$

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

$$(\mathbf{C}_k)_{i,j} = \frac{1}{W_c} c_k\left(\frac{i-j}{W_c}\right) \tag{3.36}$$

with $c_k(\cdot)$ the impulse response of the channel for the kth user independent across users.

Let $\mathbf{\Lambda}$ be an $N \times K$ matrix whose (i,j)th entry is

$$\mathsf{\Lambda}_{i,j} = \lambda_i(\mathbf{C}_j)|\mathsf{A}_j|^2$$

with $\lambda_i(\mathbf{C}_j)$ the ith eigenvalue of $\mathbf{C}_j\mathbf{C}_j^\dagger$. Assuming that $\mathbf{\Lambda}$ behaves ergodically (cf. Definition 2.17), from Theorem 2.59 it follows that the arithmetic mean of the MMSE's satisfies

$$\begin{aligned}
\frac{1}{K}\sum_{k=1}^{K}\mathsf{MMSE}_k &= \frac{1}{K}\mathrm{tr}\left\{\left(\mathbf{I} + \mathsf{SNR}\,\mathbf{H}^\dagger\mathbf{H}\right)^{-1}\right\} && (3.37)\\
&\to \eta_{\mathbf{H}^\dagger\mathbf{H}}(\mathsf{SNR}) && (3.38)\\
&= 1 - \frac{1}{\beta} + \frac{1}{\beta}\mathbb{E}\left[\,\Gamma_{\mathbf{H}\mathbf{H}^\dagger}(\mathsf{X}, \mathsf{SNR})\,\right] && (3.39)
\end{aligned}$$

where in (3.39) we have used (2.56). The function $\Gamma_{\mathbf{H}\mathbf{H}^\dagger}(\cdot,\cdot)$, in turn, satisfies the fixed-point equation

$$\Gamma_{\mathbf{H}\mathbf{H}^\dagger}(x, \mathsf{SNR}) + \beta\,\mathsf{SNR}\,\mathbb{E}\left[\frac{\rho(x,\mathsf{Y})\Gamma_{\mathbf{H}\mathbf{H}^\dagger}(x,\mathsf{SNR})}{1 + \mathsf{SNR}\;\mathbb{E}[\rho(\mathsf{X},\mathsf{Y})\Gamma_{\mathbf{H}\mathbf{H}^\dagger}(\mathsf{X},\mathsf{SNR})|\mathsf{Y}]}\right] = 1 \tag{3.40}$$

where X and Y are independent random variables uniform on $[0,1]$ and $\rho(\cdot,\cdot)$ is the channel profile of $\mathbf{\Lambda}$ (cf. Definition 2.18). Note that the received signal-to-noise ratio of the kth user is $\textsf{SNR}\,\|\mathbf{h}_k\|^2$ with

$$\begin{aligned}\|\mathbf{h}_k\|^2 &\to |\mathsf{A}_k|^2 \lim_{N\to\infty} \frac{1}{N}\mathrm{tr}\{\mathbf{C}_k^\dagger \mathbf{C}_k\} \\ &= \mathbb{E}[\rho_k(\mathsf{X})]. \end{aligned} \tag{3.41}$$

with $\mathbb{E}[\rho_k(X)]$ representing the one-dimensional channel profile (cf. Definition 2.18) of $\mathbf{\Lambda}$. The multiuser efficiency of the kth user achieved by the MMSE receiver is [159]

$$\eta_k^{\mathsf{mmse}}(\textsf{SNR}) = \frac{\mathsf{SINR}_k}{\textsf{SNR}\,\|\mathbf{h}_k\|^2} \tag{3.42}$$

$$= \frac{\mathbf{h}_k^\dagger\left(\mathbf{I}+\textsf{SNR}\sum_{i\neq k}\mathbf{h}_i\mathbf{h}_i^\dagger\right)^{-1}\mathbf{h}_k}{\|\mathbf{h}_k\|^2} \tag{3.43}$$

$$\to \frac{\mathcal{F}(y,\textsf{SNR})}{\mathbb{E}[\rho_k(\mathsf{X})]} \tag{3.44}$$

with $\frac{k-1}{K} \leq y < \frac{k}{K}$ and $\mathcal{F}(\cdot,\cdot)$ defined as the solution to the fixed-point equation (cf. (2.157))

$$\mathcal{F}(y,\textsf{SNR}) = \mathbb{E}\left[\frac{\rho(\mathsf{X},y)}{1+\textsf{SNR}\,\beta\,\mathbb{E}\left[\frac{\rho(\mathsf{X},\mathsf{Y})}{1+\mathsf{SNR}\,\mathcal{F}(\mathsf{Y},\mathsf{SNR})}\,|\mathsf{X}\right]}\right]. \tag{3.45}$$

Let the ratio between the effective number of users and the effective processing gain be defined as

$$\beta' = \beta\frac{\mathbb{P}[\mathbb{E}\left[\rho(\mathsf{X},\mathsf{Y})|\mathsf{Y}\right]>0]}{\mathbb{P}[\mathbb{E}\left[\rho(\mathsf{X},\mathsf{Y})|\mathsf{X}\right]>0]}. \tag{3.46}$$

Using Corollary 2.4, we obtain that the asymptotic MMSE multiuser efficiency admits the following expression for $\beta' < 1$:

$$\begin{aligned}\eta_k^{\mathsf{dec}} &= \lim_{\mathsf{SNR}\to\infty}\eta_k^{\mathsf{mmse}}(\textsf{SNR}) \\ &= \frac{\beta\,\mathbb{P}[\mathbb{E}[\rho(\mathsf{X},\mathsf{Y})|\mathsf{Y}]\neq 0]\Gamma_\infty(y)}{\mathbb{E}[\rho_k(\mathsf{X})]}\end{aligned} \tag{3.47}$$

where $\Gamma_\infty(\cdot)$ satisfies (2.142) with the role of $v(x,y)$ played by $\rho(x,y)$.

Specializing (3.39) to the case that the signal transmitted by each user propagates through L discrete i.i.d. chip-spaced paths (where L does not grow with N), the η-transform of the asymptotic averaged eigenvalue distribution of $\mathbf{HH}^\dagger$, $\eta_{\mathbf{HH}^\dagger}$, satisfies the fixed-point equation [159]

$$1 - \eta_{\mathbf{HH}^\dagger} = \beta \left(1 - \eta_{\mathsf{P}} \left(\mathsf{SNR}\ \eta_{\mathbf{HH}^\dagger} \right) \right) \tag{3.48}$$

where η_{P} is the η-transform of the almost sure asymptotic empirical distribution of[5]

$$\left\{ \frac{|A_1|^2}{W_c^2} \sum_{\ell=1}^{L} \left| c_1 \left(\frac{\ell}{2W_c} \right) \right|^2, \ldots, \frac{|A_K|^2}{W_c^2} \sum_{\ell=1}^{L} \left| c_K \left(\frac{\ell}{2W_c} \right) \right|^2 \right\}.$$

Using this result, [159] concludes that, asymptotically as $N \to \infty$, each multipath interferer with a fixed number of resolvable paths acts like a single path interferer with received power equal to the total received power from all the paths of that user. From this it follows that, in the special case of a fixed number of i.i.d. resolvable paths, the expressions obtained for the SINR at the output of the decorrelator and MMSE receiver in a frequency-selective channel are equivalent to those for a flat fading channel. This result has been found also in [73] under the assumption that the spreading sequences are either independent across users and paths or independent across users and cyclically shifted across the paths (cf. Section 3.1.5).

In the downlink, every user experiences the same frequency-selective fading, i.e., $\mathbf{C}_k = \mathbf{C}\ \forall k$, where the empirical distribution of $\mathbf{CC}^\dagger$ converges almost surely to a nonrandom limit $F_{|\mathsf{C}|^2}$. Consequently, (3.35) particularizes to

$$\mathbf{H} = \mathbf{CSA}. \tag{3.49}$$

Using Theorem 2.46 and with the aid of an auxiliary function $\chi(\mathsf{SNR})$, abbreviated as χ, we obtain that the MMSE multiuser efficiency of the

[5] Whenever we refer to an almost sure asymptotic empirical distribution, we are implicitly assuming that the corresponding empirical distribution converges almost surely to a nonrandom limit.

kth user, abbreviated as $\eta = \eta^{\mathsf{mmse}}(\mathsf{SNR})$, is the solution to

$$\beta\,\eta\,\chi \;=\; \frac{1-\eta_{|\mathsf{C}|^2}(\beta\,\chi)}{\mathbb{E}[|\mathsf{C}|^2]} \qquad (3.50)$$

$$\eta\,\chi \;=\; \frac{1-\eta_{|\mathsf{A}|^2}(\mathsf{SNR}\,\mathbb{E}[|\mathsf{C}|^2]\eta)}{\mathbb{E}[|\mathsf{C}|^2]} \qquad (3.51)$$

where $|\mathsf{C}|^2$ and $|\mathsf{A}|^2$ are independent random variables with distributions given by the asymptotic spectra of $\mathbf{CC}^\dagger$ and $\mathbf{AA}^\dagger$, respectively, while $\eta_{|\mathsf{C}|^2}(\cdot)$ and $\eta_{|\mathsf{A}|^2}(\cdot)$ represent their respective η-transforms. Note that, instead of (3.51) and (3.50), we may write [37, 159]

$$\eta = \frac{1}{\mathbb{E}[|\mathsf{C}|^2]}\mathbb{E}\left[\frac{|\mathsf{C}|^2}{1+\beta\,\mathsf{SNR}\,|\mathsf{C}|^2\mathbb{E}\left[\frac{|\mathsf{A}|^2}{1+\mathsf{SNR}\,\mathbb{E}[|\mathsf{C}|^2]\,|\mathsf{A}|^2\eta}\right]}\right]. \qquad (3.52)$$

From Corollary 2.2 we have that, for

$$\beta\frac{\mathbb{P}[|\mathsf{A}|>0]}{\mathbb{P}[|\mathsf{C}|>0}\leq 1,$$

η_k^{dec} converges almost surely to the solution to

$$1 = \mathbb{E}\left[\frac{|\mathsf{C}|^2}{\eta^{\mathsf{dec}}\,\mathbb{E}[|\mathsf{C}|^2]+\beta\,\mathbb{P}[|\mathsf{A}|>0]\,|\mathsf{C}|^2}\right]. \qquad (3.53)$$

Note that both the MMSE and the decorrelator multiuser efficiencies are asymptotically the same for every user.

From Theorem 2.43, the downlink counterpart of (3.39) is [159]

$$\frac{1}{K}\sum_{k=1}^{K}\mathsf{MMSE}_k \;=\; \frac{1}{K}\mathrm{tr}\left\{\left(\mathbf{I}+\mathsf{SNR}\,\mathbf{H}^\dagger\mathbf{H}\right)^{-1}\right\}$$
$$=\; 1-\frac{1}{\beta}+\frac{1}{\beta}\,\eta_{|\mathsf{C}|^2}\left(\beta\chi(\mathsf{SNR})\right) \qquad (3.54)$$

with $\chi(\cdot)$ solution to (3.51) and (3.50). The special case of (3.52) for equal-power users was given in [56].

For the sake of brevity, we will not explicitly extend the analysis to the case in which both frequency selectivity and multiple receive antennas are present. This can be done by blending the results obtained in

Sections 3.1.3 and 3.1.4. Moreover, multiple transmit antennas can be further incorporated as done explicitly in [169], where analytical tools already leveraged in Sections 3.1.2-3.1.3 are applied to the asymptotic characterization of the single-user matched filter and MMSE detectors. It is found that DS-CDMA, even with single-user decoding, can outperform orthogonal multiaccess with multiple antennas provided the number of receive antennas is sufficiently large.

In most of the literature, the DS-CDMA channel spans only the users within a particular system cell with the users in other cells regarded as a collective source of additive white Gaussian noise. While it is reasonable to preclude certain forms of multiuser detection of users in other cells, on the basis that their codebooks may be unknown, the structure in the signals of those other-cell users can be exploited even without access to their codebooks. This, however, requires more refined models that incorporate this structure explicitly within the noise. For some simple such models, the performance of various receivers has been evaluated asymptotically in [317, 237]. Since the expression for the capacity of a DS-CDMA channel with colored noise parallels that of the corresponding multi-antenna channel, we defer the details to Section 3.3.8.

3.1.5 Channel Estimation for DS-CDMA

Reference [73] applies the concept of asymptotic freeness to the same setup of Section 3.1.4 (linear DS-CDMA receivers and a fading channel with L discrete chip-spaced paths), but departing from the usual assumption that the receiver has perfect side information about the state of the channel. Incorporating channel estimation, the receiver consists of two distinct parts:

- The channel estimator, which provides linear MMSE joint estimates of the channel gains for every path of every user.
- The data estimator, which uses those channel estimates to detect the transmitted data using a one-shot linear receiver.

In order to render the problem analytically tractable, the delay spread is considered small relative to the symbol time and, more importantly, the time delays of the resolvable paths of all users are assumed known. Thus, the channel estimation encompasses only the path gains and it is further conditioned on the data (hypothesis that is valid during training or with error-free data detection). The joint estimation of the channel path gains for all the users is performed over an estimation window of Q symbols, presumed small relative to the channel coherence time. For the ith symbol within this window, the output of the chip matched filter is

$$\mathbf{y}(i) = \sum_{k=1}^{K} \sum_{\ell=1}^{L} \mathsf{C}_{k,\ell}\, \mathbf{s}_{k,\ell}(i)(\mathbf{x}(i))_k + \mathbf{n}(i) \tag{3.55}$$

where $\mathsf{C}_{k,\ell}$ represents the channel fading coefficient for path ℓ of user k such that $\mathbb{E}[|\mathsf{C}_{k,\ell}|^2] = \frac{1}{L}$, $\mathbf{s}_{k,\ell}(i)$ is the spreading sequence for the ℓth path of the kth user for the ith symbol interval, $\mathbf{n}(i)$ is the additive Gaussian noise in the ith symbol interval, and $(\mathbf{x}(i))_k$ represents the ith symbol of the kth user.

With long (i.e., changing from symbol to symbol) random spreading sequences independent across users and paths, [73] shows using Theorem 2.38 that, as $K, N \to \infty$ with $\frac{K}{N} \to \beta$, the mean-square error of the estimation of every path gain coefficient converges to

$$\xi^2 = \left(\mathsf{SNR}\, \frac{Q - \beta L}{2} + \frac{L}{2} + \sqrt{\mathsf{SNR}^{\,2} \frac{(Q-\beta L)^2}{4} + \mathsf{SNR}\; L \frac{Q+\beta}{2} + \frac{L^2}{4}} \right)^{-1}.$$

This result, in fact, holds under alternative conditions as well:

- If the spreading sequences are independent across users and paths but they repeat from symbol to symbol, i.e., $\mathbf{s}_{k,\ell}(i) = \mathbf{s}_{k,\ell}\ \forall i$ (this can be proved using Theorem 2.61).
- The sequences received over the L paths are cyclically shifted versions of each other but independent across users, i.e., $\mathbf{s}_{k,\ell}(i)$ is a cyclically shifted replica of $\mathbf{s}_{k,1}(i)$ by $\ell - 1$ chips (this can be proved using Example 2.49).

The linear receiver performing data estimation operates under the belief that the estimate of the ℓth path gain of the kth user has mean $\bar{\mathsf{C}}_{k,\ell}$ and variance ξ_k^2. These estimates are further assumed uncorrelated and with equal variance for all paths of each user. (When the channel is perfectly known, $\mathsf{C}_{k,\ell} = \bar{\mathsf{C}}_{k,\ell}$ and $\xi^2_{k,\ell} = 0$ and the results reduce to their counterparts in Section 3.1.4.) The linear receiver is designed with all expectations being conditional on the spreading sequences and the mean and variance supplied by the channel estimator.

From Theorem 2.46, the output SINR for user k converges asymptotically in probability to

$$\frac{1}{1+\xi_k^2 \mathsf{SINR}_d} \sum_{\ell=1}^{L} |\bar{\mathsf{C}}_{k,\ell}|^2 \mathsf{SINR}_d \tag{3.56}$$

where SINR_d is the corresponding output SINR *without* the effect of other-user channel estimation errors. Implicit expressions for SINR_d, depending on the type of linear receiver, are

$$\frac{1}{\mathsf{SINR}_d} = \begin{cases} \dfrac{1}{\mathsf{SNR}} + \beta L \mathbb{E}\left[\dfrac{\mathsf{P}}{1+\mathsf{P}\,\mathsf{SINR}_d}\right] & \text{MMSE} \\ \dfrac{1}{\mathsf{SNR}} + \dfrac{1}{\mathsf{SNR}} \dfrac{\beta L}{1-\beta L} & \text{decorrelator} \\ \dfrac{1}{\mathsf{SNR}} + \beta L \mathbb{E}[\mathsf{P}] & \text{single-user matched filter} \end{cases} \tag{3.57}$$

with expectation over P, whose distribution equals the asymptotic empirical eigenvalue distribution of the matrix $\mathbb{E}[\mathrm{diag}(\mathbf{c}_2\mathbf{c}_2^\dagger, \ldots, \mathbf{c}_K\mathbf{c}_K^\dagger)]$ (assumed to converge to a nonrandom limit) with $\mathbf{c}_k = [\mathsf{C}_{k,1} \ldots \mathsf{C}_{k,L}]^T$. The main finding of the analysis in [73] is that, provided the channel estimation window (in symbols) exceeds the number of resolvable paths, the resulting estimates enable near-optimal performance of the linear data estimator.

In [46], the impact of channel estimator errors on the performance of the linear MMSE multistage receiver (cf. Section 3.1.6) for large multiuser systems with random spreading sequences is analyzed.

3.1.6 Reduced-Rank Receivers for DS-CDMA

Both the MMSE and the decorrelator receivers need to invert a matrix whose dimensionality is equal to either the number of users or the spreading gain. In large-dimensional systems, this is a computationally intensive operation. It is therefore of interest to pursue receiver structures that approach the performance of these linear receivers at a lower computational cost.

Invoking the Cayley-Hamilton Theorem,[6] the MMSE receiver can be synthesized as a polynomial expansion that yields the soft estimate of the kth user symbol in (1.1) as

$$\hat{x}_k = \mathbf{h}_k^\dagger \sum_{m=0}^{D-1} w_m \mathbf{R}^m \mathbf{y} \tag{3.58}$$

where $\mathbf{R} = \mathbf{H}\mathbf{H}^\dagger$ and $D = N$ (the rank or the number of stages of the receiver). Since the coefficients w_m, $m \in \{0, \dots, D-1\}$ must be obtained from the characteristic polynomial of the matrix whose inverse is being expanded, this expansion by itself does not reduce the computational complexity. It does, however, enable the possibility of a flexible tradeoff between performance and complexity controlled through D. The first proposal for a reduced-complexity receiver built around this idea came in [179], where it was suggested approximating (3.58) with $D < N$ and with the coefficients w_m computed using as cost function the mean-square error between $\hat{x}_k$ obtained with the chosen D and the actual $\hat{x}_k$ obtained with a true MMSE receiver. Then, the w_m's become a function of the first D moments of the empirical distribution of $\mathbf{R}$. With $D < N$, the linear receiver in (3.58) projects the received vector on the subspace (of the signal space) spanned by the vectors $\{\mathbf{h}_k, \mathbf{R}\mathbf{h}_k, \dots, \mathbf{R}^{D-1}\mathbf{h}_k\}$.[7] Reduced-rank receivers have been put forth for numerous signal processing applications such as array processing, radar, model order reduction (e.g. [214, 215, 130]), where the signal is effectively projected onto a lower-dimensional subspace and the filter optimization then occurs within that subspace. This subspace can be chosen using a variety of criteria:

Principal components. The projection occurs onto an estimate of the lower-dimensional signal subspace with the largest energy

[6] The Cayley-Hamilton Theorem ensures that the inverse of a $K \times K$ nonsingular matrix can always expressed as a $(K-1)$th order polynomial [117].

[7] These vectors are also known as a Krylov sequence [117]. For a given matrix $\mathbf{A}$ and vector $\mathbf{x}$, the sequence of vectors $\mathbf{x}, \mathbf{A}\mathbf{x}, \mathbf{A}^2\mathbf{x}, \dots$ or a truncated portion of this sequence is known as the Krylov sequence of $\mathbf{A}$. The subspace spanned by a Krylov sequence is called Krylov space of $\mathbf{A}$.

[298, 115, 247].

Cross-spectral method. The eigenvector basis which minimizes the mean-square error is chosen [34, 92] based on an eigenvalue decomposition of the correlation matrix.

Partial despreading. The lower dimensional subspace of the reduced rank receiver is spanned by non-overlapping segments of the matched filter [232].

Reduced-rank multistage Wiener filter. The multi-stage Wiener (MSW) filter and its reduced-rank version were proposed in [91, 93].

These various techniques have been analyzed asymptotically, in terms of SINR, in [116]. In particular, it is shown in [116] for the MSW filter with equal-power users that, as $K, N \to \infty$, the output SINR converges in probability to a nonrandom limit

$$\mathsf{SINR}_{D+1} = \frac{\mathsf{SNR}}{1 + \beta \frac{\mathsf{SNR}}{1+\mathsf{SINR}_D}} \tag{3.59}$$

for $D \geq 0$, where $\mathsf{SINR}_0 = 0$ and $\mathsf{SINR}_1 = \frac{P}{1+\beta\,\mathsf{SNR}}$ is the SINR at the output of the matched filter. The analysis for unequal-power users can be found in [253, 255]. A generalization of the analysis in [116] and [253] can be found in [162] where a connection between the asymptotic behavior of the SINR at the output of the reduced rank Wiener filter and the theory of orthogonal polynomials for the so-called power moments is established. It is further demonstrated in [116] and [162], numerically and analytically respectively, that the number of stages D needed in the reduced-rank MSW filter to achieve a desired output SINR does not scale with the dimensionality; in fact, a few stages are usually sufficient to achieve near-full-rank output SINR regardless of the dimension of the signal space. However, the weights of the reduced-rank receiver do depend on the spreading sequences. Therefore, in long-sequence CDMA they have to be reevaluated from symbol to symbol, which hampers real-time implementation.

To lift the burden of computing the weights from the spreading sequences for every symbol interval, [187, 265, 159] proposed the asymptotic reduced-rank MMSE receiver, which replaces the weights in (3.58) with their limiting values in the asymptotic regime. Following this approach, various scenarios described by (1.1) have been evaluated in [45, 105, 158, 159, 187, 265].[8] For all these different scenarios it has been proved that, in contrast with the exact weights, the asymptotic weights do not depend on the realization of $\mathbf{H}$ and hence they do not need to be updated from symbol to symbol. The asymptotic weights are determined only by the number of users per chip and by the asymptotic moments of $\mathbf{HH}^\dagger$ and thus, in order to compute these weights explicitly, it is only necessary to obtain explicit expressions for the asymptotic eigenvalue moments of the interference autocorrelation matrix. Numerical results show that the asymptotic weights work well for even modest dimensionalities.

Alternative low-complexity implementations of both the decorrelator and the MMSE receiver can be realized using the concepts of iterative linear interference cancellation [84, 124, 33, 207, 71, 72], which rely on well-known iterative methods for the solution of systems of linear equations (and consequently for matrix inversion) [7]. This connection has been recently established in [99, 251, 72]. In particular, parallel interference cancellation receivers are an example of application of the Jacobi method, first- and second-order stationary methods and Chebyshev methods, while serial interference cancellation receivers are an example of application of Gauss-Seidel and successive relaxation methods. For all these linear (parallel and serial) interference cancellation receivers, the convergence properties to the true decorrelator or MMSE solution have been studied in [99] for large systems. For equal-power users, the asymptotic convergence of the output SINR of the linear multistage parallel interference cancellation receiver (based on the first

[8] In [187], DS-CDMA with equal-power users and no fading is studied. In turn, [158] considers the more general scenario of DS-CDMA with unequal-power users and flat-fading. Related results in the context of the reduced-rank MSW and of the receiver originally proposed by [179] were reported in [45]. In [158, 159], the analysis is extended to multi-antenna receivers and further extended to include frequency selectivity in [105, 159]. Specifically, the frequency-selective CDMA downlink is studied in [105] with the restriction that the signature matrix be unitarily invariant with i.i.d. entries. In [159], in contrast, the analysis with frequency-selectivity is general enough to encompass uplink and downlink as well as signature matrices whose entries are independent with common mean and variance but otherwise arbitrarily distributed. The case of frequency-selective CDMA downlink with orthogonal signatures has been treated in [105].

and second-order stationary linear iterative method) to a nonrandom limit has been analyzed in [252, 254].

We now summarize some of the results on linear polynomial MMSE receivers for DS-CDMA. The linear expansion of the MMSE receiver is built using a finite-order Krylov sequence of the matrix $\mathbf{H}_k\mathbf{H}_k^\dagger + \sigma^2\mathbf{I}$ and the coefficients of the expansion are chosen to minimize MSE. The soft estimate of the kth user symbol is given by (3.58) with $\mathbf{R}$ replaced by

$$\sum_{i \neq k} \mathbf{h}_i\mathbf{h}_i^\dagger + \sigma^2\mathbf{I} = \mathbf{H}_k\mathbf{H}_k^\dagger + \sigma^2\mathbf{I} \tag{3.60}$$

where $\mathbf{H}_k$ indicates the matrix $\mathbf{H}$ with the kth column removed. The weights that minimize the mean-squared error are

$$\mathbf{w} = \begin{bmatrix} \mathcal{H}_1 + \mathcal{H}_0\mathcal{H}_0 & \cdots & \mathcal{H}_D + \mathcal{H}_{D-1}\mathcal{H}_0 \\ \vdots & \ddots & \vdots \\ \mathcal{H}_D + \mathcal{H}_{D-1}\mathcal{H}_0 & \cdots & \mathcal{H}_{2D-1} + \mathcal{H}_{D-1}\mathcal{H}_{D-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{H}_0 \\ \vdots \\ \mathcal{H}_{D-1} \end{bmatrix} \tag{3.61}$$

where the (i,j)th entry of the above matrix is $\mathcal{H}_{i+j-1} + \mathcal{H}_{i-1}\mathcal{H}_{j-1}$ with

$$\mathcal{H}_m = \mathbf{h}_k^\dagger \left(\mathbf{H}_k\mathbf{H}_k^\dagger + \sigma^2\mathbf{I}\right)^m \mathbf{h}_k. \tag{3.62}$$

Denoting the asymptotic value of $\mathcal{H}_m$ as

$$\mathcal{H}_m^\infty = \lim_{K\to\infty} \mathcal{H}_m, \tag{3.63}$$

the asymptotic weights are given by (3.61) where each $\mathcal{H}_m$ is replaced by its asymptotic counterpart, $\mathcal{H}_m^\infty$. The calculation of these asymptotic weights is closely related to the evaluation of the asymptotic eigenvalue moments of $\mathbf{H}\mathbf{H}^\dagger$, which can be done using the results laid down in Section 2.3. In the following, all the hypotheses made in the previous sections dealing with DS-CDMA are upheld.

In the case of unfaded equal power DS-CDMA, with $\mathbf{H} = \mathbf{S}$ as in Section 3.1.1, using (2.102) we have that [187, 158]

$$\mathcal{H}_m^\infty = \sum_{n=0}^{m} \binom{m}{n} \sigma^{2m-2n} \sum_{i=1}^{n} \binom{n}{i} \binom{n}{i-1} \frac{\beta^i}{n}. \tag{3.64}$$

In the case of faded DS-CDMA with a single receive antenna, where $\mathbf{H} = \mathbf{SA}$ as in Section 3.1.2,

$$\mathcal{H}_m^\infty = \sum_{n=0}^{m} \binom{m}{n} \sigma^{2m-2n} |\mathsf{A}_k|^2 \mu_n \tag{3.65}$$

with μ_n, from (2.118), given by

$$\mu_n = \sum_{i=1}^{n} \beta^{n-i} \sum \frac{n!}{m_1! \dots m_i! \; i!} \mathbb{E}\left[|\mathsf{A}|^{2m_1}\right] \dots \mathbb{E}\left[|\mathsf{A}|^{2m_i}\right]. \tag{3.66}$$

where $|\mathsf{A}|$ is a random variable whose distribution equals the asymptotic empirical singular value distribution of $\mathbf{A}$ and the inner sum is over all i-tuples of nonnegative integers $(m_1, \dots, m_i)$ such that [158, 45]

$$\sum_{\ell=1}^{i} m_\ell = n - i + 1 \tag{3.67}$$

$$\sum_{\ell=1}^{i} \ell m_\ell = n, \tag{3.68}$$

A similar result holds for the faded DS-CDMA with antenna diversity described in Section 3.1.3 with $|\mathsf{A}|$ now equal to the square root of the random variable whose distribution is given by the asymptotic empirical distribution of $P_1, \dots, P_K$ as defined in Section 3.1.3.

For the frequency-selective faded downlink, applying Theorem 2.48 to the model in Section 3.1.4 we have [159]

$$\mathcal{H}_m^\infty = \sum_{n=0}^{m} \binom{m}{n} \sigma^{2m-2n} |\mathsf{A}_k|^2 \mathbb{E}\left[|\mathsf{C}|^2 m_n(|\mathsf{C}|^2)\right] \tag{3.69}$$

where

$$m_n(r) = \beta\, r \sum_{\ell=1}^{n} m_{\ell-1}(r) \sum_{\substack{n_1+\dots+n_i=n-\ell \\ 1\le i\le n-\ell}} \mathbb{E}\left[|\mathsf{A}|^{2i+2}\right] \mathbb{E}\left[|\mathsf{C}|^2 m_{n_1-1}(|\mathsf{C}|^2)\right] \dots \mathbb{E}\left[|\mathsf{C}|^2 m_{n_i-1}(|\mathsf{C}|^2)\right] \tag{3.70}$$

with $|\mathsf{C}|^2$ as in Section 3.1.4 and with $|\mathsf{A}|$ representing a random variable, independent of $|\mathsf{C}|^2$, whose distribution equals the asymptotic empirical singular value distribution of $\mathbf{A}$. The counterpart of (3.69) for orthogonal Haar distributed spreading signatures and for unitarily invariant i.i.d. spreading sequences has been analyzed in [105], where the asymptotic weights are calculated using free probability.

In the frequency-selective faded uplink, in turn, $\mathbf{H}$ is given by (3.35) and straight application of Theorem 2.59 yields

$$\begin{aligned} \mathcal{H}_m^\infty &= \sum_{n=0}^{m} \binom{m}{n} \sigma^{2m-2n} \delta_{n,k} \\ &= \sum_{n=0}^{m} \binom{m}{n} \sigma^{2m-2n} \mathbb{E}[\rho(\mathsf{X},k)]\mathbb{E}[m_n(\mathsf{X})\rho_k(\mathsf{X})] \end{aligned} \tag{3.71}$$

with $\rho(\cdot,\cdot)$ and $\rho_k(\cdot)$ as in Section 3.1.4 and with $m_n(\cdot)$ obtained through the recursive equation given by (2.164) in Theorem 2.55.

3.2 Multi-Carrier CDMA

Multi-Carrier CDMA (MC-CDMA) is the frequency dual of DS-CDMA. Hence, a MC-CDMA transmitter uses a given spreading sequence to spread the original signal in the frequency domain. In other words, each fraction of the symbol corresponding to a chip of the spreading code is transmitted through a different subcarrier. It is essential that the sub-band corresponding to each subcarrier be narrow enough for its fading to be frequency non-selective. The basic transmitter structure of MC-CDMA is similar to that of OFDM [109], with the main difference being that the MC-CDMA scheme transmits the same symbol in parallel through the various subcarrier whereas an OFDM scheme transmits different symbols. The spreading gain N is equal to the number of frequency subcarriers. Each symbol of the data stream generated by user k is replicated into N parallel copies. Each copy is then multiplied by a chip from the corresponding spreading sequence. Finally, an inverse discrete Fourier transform (IDFT) is used to convert those N parallel copies back into serial form for transmission. A cyclic or empty prefix is appended to facilitate demodulation, at the expense of some loss in efficiency. A possible receiver front-end consists of N matched filters, one for each subcarrier.

Since in the case of frequency-flat fading the analysis of MC-CDMA is mathematically equivalent to that of its DS-CDMA counterpart (see Section 3.1.2), we proceed directly to consider the more general case of frequency-selective fading.

3.2.1 MC-CDMA Uplink

In synchronous MC-CDMA with K active users and frequency-selective fading, the vector $\mathbf{x}$ contains the signals transmitted by each of the users and the kth column of $\mathbf{H}$ is

$$\mathbf{h}_k = [h_k^{(1)}, \ldots, h_k^{(N)}]^T \tag{3.72}$$

where

$$h_k^{(\ell)} = \mathsf{A}_k \mathsf{C}_{\ell,k} s_k^{(\ell)}, \tag{3.73}$$

with $\mathbf{s}_k = [s_k^{(1)}, \ldots, s_k^{(N)}]^T$ denoting the unit-energy transmitted spreading sequence of the kth user, A_k indicating the received amplitude of that kth user, which accounts for its average path loss, and with $\mathsf{C}_{\ell,k}$ denoting the fading for the ℓth subcarrier of the kth user, independent across the users. In this subsection we refer to $\mathbf{h}_k$ as the *received signature* of the kth user. Notice that $\mathbf{H}$ incorporates both the spreading and the frequency-selective fading. More precisely, denoting by $\mathbf{C}$ the $N \times K$ matrix whose (ℓ,k)th entry is $\mathsf{C}_{\ell,k}$, we can write the received signature matrix $\mathbf{H}$ as

$$\mathbf{H} = \mathbf{C} \circ \mathbf{SA} \tag{3.74}$$

with $\circ$ denoting element-wise (Hadamard) product and

$$\mathbf{A} = \mathrm{diag}(\mathsf{A}_1, \ldots, \mathsf{A}_K) \tag{3.75}$$

$$\mathbf{S} = [\ \mathbf{s}_1 \mid \ \ldots \ \mid \mathbf{s}_K\ ] \tag{3.76}$$

$$\mathbf{C} = [\ \mathbf{c}_1 \mid \ \ldots \ \mid \mathbf{c}_K\ ] \tag{3.77}$$

where the entries of $\mathbf{S}$ are i.i.d. zero-mean with variance $\frac{1}{N}$ and thus the general model becomes

$$\mathbf{y} = (\mathbf{C} \circ \mathbf{SA})\mathbf{x} + \mathbf{n}. \tag{3.78}$$

Each user experiences independent fading and hence the columns of $\mathbf{C}$ are independent. The relationship between the fading at different subcarriers of any given user, in turn, is dictated by the power-delay response of the channel. More precisely, we can define a frequency covariance matrix of the kth user as

$$\mathbf{M}_k = E[\mathbf{c}_k \mathbf{c}_k^\dagger]. \tag{3.79}$$

The (p,q)th entry of $\mathbf{M}_k$ is given by the correlation between the channel response at subcarriers p and q, separated by frequency $(p-q)\Delta_f$, i.e.,

$$(\mathbf{M}_k)_{p,q} = \int_{-\infty}^{\infty} \phi_k(\tau) e^{-j2\pi(p-q)\tau\Delta_f} d\tau = \Phi_k\left((p-q)\Delta_f\right) \tag{3.80}$$

with ϕ_k and Φ_k the power-delay response and the frequency correlation function of the kth user channel, respectively.

The received energy at the ℓth subcarrier, $\ell \in \{1,\ldots,N\}$, for the kth user, $k \in \{1,\ldots,K\}$, is $|\mathsf{C}_{\ell,k}\,\mathsf{A}_k|^2$.

Let $\mathbf{B}$ be the $N \times K$ matrix whose (i,j)th element is

$$\mathbf{B}_{i,j} = \mathsf{C}_{i,j}\,\mathsf{A}_j \tag{3.81}$$

and let $v(\cdot,\cdot)$ be the two-dimensional channel profile of $\mathbf{B}$ assumed to behave ergodically (cf. Definition 2.17). Then, the SINR at the output of the MMSE receiver is

$$\mathsf{SINR}_k^{\mathsf{mmse}} = \mathsf{SNR}\,|\mathsf{A}_k|^2 \,(\mathbf{c}_k \circ \mathbf{s}_k)^\dagger \left(\mathbf{I} + \mathsf{SNR}\,\mathbf{H}_k \mathbf{H}_k^\dagger\right)^{-1} (\mathbf{c}_k \circ \mathbf{s}_k)$$

where, recall from the DS-CDMA analysis, $\mathbf{H}_k$ indicates the matrix $\mathbf{H}$ with the kth column removed. Using Theorems 2.57 and 2.52, the multiuser efficiency is given by the following result.

Theorem 3.1. [160] For $0 \le y \le 1$, the multiuser efficiency of the MMSE receiver for the $\lfloor yK \rfloor$th user converges almost surely, as $K, N \to \infty$ with $\frac{K}{N} \to \beta$, to

$$\lim_{K\to\infty} \eta_{\lfloor yK \rfloor}^{\mathsf{mmse}}(\mathsf{SNR}) = \frac{\Psi(y,\mathsf{SNR})}{\mathbb{E}\left[v(\mathsf{X},y)\right]} \tag{3.82}$$

where $\Psi(\cdot,\cdot)$ is a positive function solution to

$$\Psi(y,\mathsf{SNR}) = \mathbb{E}\left[\frac{v(\mathsf{X},y)}{1 + \mathsf{SNR}\,\beta\mathbb{E}\left[\frac{v(\mathsf{X},\mathsf{Y})}{1+\mathsf{SNR}\,\Psi(\mathsf{Y},\mathsf{SNR})}\middle|\mathsf{X}\right]}\right] \tag{3.83}$$

where the expectations are with respect to independent random variables X and Y both uniform on [0,1].

Most quantities of interest such as the multiuser efficiency and the capacity approach their asymptotic behaviors very rapidly as K and N grow large. Hence, we can get an extremely accurate approximation of the multiuser efficiency and consequently of the capacity with an arbitrary number of users, K, and a finite processing gain, N, simply by resorting to their asymptotic approximation with $v(x,y)$ replaced in Theorem 3.1 by

$$v(x,y) \approx |\mathsf{A}_k|^2 \, |\mathsf{C}_{\ell,k}|^2 \qquad \frac{\ell-1}{N} \le x < \frac{\ell}{N} \quad \frac{k-1}{K} \le y < \frac{k}{K}.$$

Thus, we have that the multiuser efficiency of uplink MC-CDMA is closely approximated by

$$\eta_k^{\mathsf{mmse}}(\mathsf{SNR}) \approx \frac{\Phi_k^N(\mathsf{SNR})}{\frac{1}{N}\sum_{\ell=1}^{N} |\mathsf{C}_{\ell,k}|^2} \tag{3.84}$$

with

$$\Phi_k^N(\mathsf{SNR}) = \frac{1}{N}\sum_{\ell=1}^{N} \frac{|\mathsf{C}_{\ell,k}|^2}{1 + \mathsf{SNR}\,\frac{\beta}{K}\sum_{j=1}^{K} \frac{|\mathsf{A}_j|^2\,|\mathsf{C}_{\ell,j}|^2}{1+\mathsf{SNR}\,|\mathsf{A}_j|^2\,\Phi_j^N(\mathsf{SNR})}}. \tag{3.85}$$

From Theorem 3.1, the MMSE spectral efficiency converges, as $K, N \to \infty$, to

$$\mathcal{C}^{\mathsf{mmse}}(\beta,\mathsf{SNR}) = \beta\,\mathbb{E}\left[\log\left(1 + \mathsf{SNR}\,\Psi(\mathsf{Y},\mathsf{SNR})\right)\right] \tag{3.86}$$

where the function $\Psi(\cdot,\cdot)$ is the solution of (3.83).

Let the ratio between the effective number of users and the effective processing gain be defined as

$$\beta' = \beta\,\frac{\mathbb{P}[\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})|\mathsf{Y}\right] > 0]}{\mathbb{P}[\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})|\mathsf{X}\right] > 0]} \tag{3.87}$$

where only the contribution of users and subcarriers that are active and not completely faded is accounted for. For all y if $\beta' < 1$, as SNR goes to infinity, the solution to (3.83), $\Psi(y,\mathsf{SNR})$, converges to $\Psi_\infty(\cdot)$, which is the solution to the fixed-point equation

$$\Psi_\infty(y) = \mathbb{E}\left[\frac{v(\mathsf{X},y)}{1 + \beta\,\mathbb{E}\left[\frac{v(\mathsf{X},\mathsf{Y})}{\Psi_\infty(\mathsf{Y})}|\mathsf{X}\right]}\right]. \tag{3.88}$$

If $\beta' < 1$, the spectral efficiency of the decorrelator is

$$\mathcal{C}^{\mathsf{dec}}(\beta, \mathsf{SNR}) = \beta\,\mathbb{E}\left[\log\left(1 + \mathsf{SNR}\,\Psi_\infty(\mathsf{Y})\right)\right]. \tag{3.89}$$

As an application of Theorem 2.53, the following generalization of (3.18) to the multicarrier CDMA channel is obtained.

Theorem 3.2. [160] The capacity of the optimum receiver is

$$\begin{aligned}\mathcal{C}^{\mathsf{opt}}(\beta, \mathsf{SNR}) &= \mathcal{C}^{\mathsf{mmse}}(\beta, \mathsf{SNR}) \\ &\quad + \mathbb{E}\left[\log(1 + \mathsf{SNR}\,\beta\,\mathbb{E}\left[\upsilon(\mathsf{X},\mathsf{Y})\Upsilon(\mathsf{Y}, \mathsf{SNR})|\mathsf{X}\right]\right] \\ &\quad - \beta\,\mathsf{SNR}\,\mathbb{E}\left[\Psi(\mathsf{Y}, \mathsf{SNR})\Upsilon(\mathsf{Y}, \mathsf{SNR})\right]\log e \end{aligned} \tag{3.90}$$

with $\Psi(\cdot,\cdot)$ and $\Upsilon(\cdot,\cdot)$ satisfying the coupled fixed-point equations

$$\Psi(y, \mathsf{SNR}) = \mathbb{E}\left[\frac{\upsilon(\mathsf{X}, y)}{1 + \beta\,\mathsf{SNR}\,\mathbb{E}[\upsilon(\mathsf{X},\mathsf{Y})\Upsilon(\mathsf{Y}, \mathsf{SNR})|\mathsf{X}]}\right] \tag{3.91}$$

$$\Upsilon(y, \mathsf{SNR}) = \frac{1}{1 + \mathsf{SNR}\;\Psi(y, \mathsf{SNR})} \tag{3.92}$$

where X and Y are independent random variables uniform on $[0,1]$.

As an alternative to (3.90), the asymptotic capacity per dimension can also be expressed as

$$\begin{aligned}\mathcal{C}^{\mathsf{opt}}(\beta, \mathsf{SNR}) &= \mathcal{C}^{\mathsf{mmse}}(\beta, \mathsf{SNR}) + E\left[\log\left(\frac{1}{\mathcal{D}(\mathsf{X}, \mathsf{SNR})}\right)\right] \\ &\quad + (E\left[\mathcal{D}(\mathsf{X}, \mathsf{SNR})\right] - 1)\log e \end{aligned} \tag{3.93}$$

with $\mathcal{D}(\cdot,\cdot)$ the solution to

$$\mathcal{D}(x, \mathsf{SNR}) = \frac{1}{1 + \mathsf{SNR}\,\beta\,E\left[\frac{\upsilon(x,\mathsf{Y})}{1+\mathsf{SNR}\;\mathbb{E}[\mathcal{D}(\mathsf{X},\mathsf{SNR})\upsilon(\mathsf{X},\mathsf{Y})|\mathsf{Y}]}\right]}. \tag{3.94}$$

This alternative expression can be easily derived from (3.90) by virtue of the fact that $\Psi(\cdot,\cdot)$ and $\mathcal{D}(\cdot,\cdot)$ relate through

$$\Psi(y, \mathsf{SNR}) = \mathbb{E}[\upsilon(\mathsf{X}, y)\mathcal{D}(\mathsf{X}, \mathsf{SNR})].$$

Although (3.90) and (3.93) are equivalent, they admit different interpretations. The latter is a generalization of the capacity given in (3.18).

The former, on the other hand, appears as function of quantities with immediate engineering meaning. More precisely, $\mathsf{SNR}\,\Psi(y, \mathsf{SNR})$ is easily recognized from Theorem 3.1 as the SINR exhibited by the $\lfloor yK \rfloor$th user at the output of a linear MMSE receiver. In turn $\Upsilon(y, \mathsf{SNR})$ is the corresponding mean-square error.

An alternative characterization of the capacity (inspired by the optimality by successive cancellation with MMSE protection against uncancelled users) is given by

$$\mathcal{C}^{\mathsf{opt}}(\beta, \mathsf{SNR}) = \beta \mathbb{E}\left[\log(1 + \mathsf{SNR}\,\daleth(\mathsf{Y}, \mathsf{SNR}))\right] \tag{3.95}$$

where

$$\daleth(y, \mathsf{SNR}) = \mathbb{E}\left[\frac{\upsilon(\mathsf{X}, y)}{1 + \mathsf{SNR}\,\beta(1-y)\mathbb{E}\left[\frac{\upsilon(\mathsf{X},\mathsf{Z})}{1+\mathsf{SNR}\,\daleth(\mathsf{Z},\mathsf{SNR})}|\mathsf{X}\right]}\right] \tag{3.96}$$

where X, and Z are independent random variables uniform on $[0, 1]$ and $[y, 1]$, respectively.

A slight variation of the standard uplink MC-CDMA setup, namely a multicode version where users are allowed to signal using several simultaneous spreading signatures, is treated in [201]. The asymptotic output SINR of the linear MMSE receiver and the corresponding spectral efficiency with both i.i.d. and orthogonal signatures are computed accounting also for frequency selectivity in the channel. The derivations rely on approximating the user covariance matrices with suitable asymptotically free independent unitarily invariant matrices having compactly supported asymptotic spectra (cf. Example 2.46). The accuracy of this approximation is verified through simulation.

3.2.2 MC-CDMA Downlink

We now turn our attention to the MC-CDMA downlink, where the results take simpler forms.

For the downlink, the structure of the transmitted MC-CDMA signal is identical to that of the uplink, but the difference with (3.74) is that every user experiences the same channel and thus $\mathbf{c}_k = \mathbf{c}$ for all

$1 \leq k \leq K$. As a result, the use of easily detectable orthogonal spreading sequences becomes enticing. We shall thus consider, in addition to sequences with i.i.d. entries, a scenario where the transmitted spreading matrix $\mathbf{S}$ is an $N \times K$ isotropic unitary matrix $\mathbf{Q}$ and thus

$$\mathbf{H} = \mathbf{CQA}. \tag{3.97}$$

with $\mathbf{C} = \text{diag}(\mathbf{c})$.

The role of the received signal-to-noise ratio of the kth user is, in this scenario, taken by $|\mathsf{A}_k|^2\mathsf{SNR}\,\mathbb{E}[|\mathsf{C}|^2]$ where $|\mathsf{C}|$ is a random variable whose distribution equals the asymptotic empirical singular value distribution of $\mathbf{C}$.

In our asymptotic analysis, we assume that the empirical singular value distribution of $\mathbf{A}$ and $\mathbf{C}$ converge almost surely to respective nonrandom limiting distributions $F_{|\mathsf{A}|}$ and $F_{|\mathsf{C}|}$.

3.2.2.1 Sequences with i.i.d. Entries

It follows from Remark 2.3.1 that the results for the downlink can be obtained as special cases of those derived for the uplink in Section (3.2.1).

Application of Theorems 2.43 and 2.46 yields the following:

Theorem 3.3. The multiuser efficiency, η_k^{mmse}, of the MMSE receiver for the kth user converges almost surely to the solution, $\eta^{\mathsf{mmse}}(\mathsf{SNR})$, of the fixed-point equation

$$\eta^{\mathsf{mmse}} = \frac{1}{\mathbb{E}\left[|\mathsf{C}|^2\right]}\mathbb{E}\left[\frac{|\mathsf{C}|^2}{1 + \mathsf{SNR}\,\beta|\mathsf{C}|^2\mathbb{E}\left[\frac{|\mathsf{A}|^2}{1+|\mathsf{A}|^2\,\mathsf{SNR}\,\mathbb{E}[|\mathsf{C}|^2]\,\eta^{\mathsf{mmse}}}\right]}\right]. \tag{3.98}$$

In the equal-power case, [202] arrived at (3.98) for a specific choice of the distribution of $|\mathsf{C}|$.

Unlike in the uplink, in the downlink the asymptotic multiuser efficiency is the same for every user. This means that, asymptotically, all the users are equivalent. The asymptotic Gaussianity of the multiaccess interference at the output of the MMSE transformation [275] leads to the following asymptotic spectral efficiency for the MMSE receiver:

$$\mathcal{C}^{\mathsf{mmse}}(\beta, \mathsf{SNR}) = \beta\, \mathbb{E}\left[\log\left(1 + |\mathsf{A}|^2\, \mathsf{SNR}\; \mathbb{E}\left[|\mathsf{C}|^2\right]\; \eta^{\mathsf{mmse}}(\mathsf{SNR})\right)\right]. \quad (3.99)$$

Let β' be the ratio between the effective number of users and the effective processing gain:

$$\beta' = \beta\, \frac{\mathbb{P}[|\mathsf{A}| > 0]}{\mathbb{P}[|\mathsf{C}| > 0]}.$$

The asymptotic spectral efficiency of the decorrelator for $\beta' \leq 1$ is

$$\mathcal{C}^{\mathsf{dec}} = \beta\, \mathbb{E}\left[\log\left(1 + \mathsf{SNR}\; \eta_0 |\mathsf{A}|^2\right)\right] \quad (3.100)$$

where η_0 is the decorrelator multiuser efficiency, positive solution to (cf. Corollary 2.2)

$$\mathbb{E}\left[\frac{|\mathsf{C}|^2}{\mathbb{E}[|\mathsf{C}|^2]\eta_0 + \beta\, \mathbb{P}[|\mathsf{A}| > 0]|\mathsf{C}|^2}\right] = 1. \quad (3.101)$$

Applying Theorem 2.44, we obtain the central characterization of the capacity of downlink MC-CDMA.

Theorem 3.4. In the MC-CDMA downlink, the capacity of the optimum receiver admits the expression

$$\mathcal{C}^{\mathsf{opt}}(\beta, \mathsf{SNR}) = \mathcal{C}^{\mathsf{mmse}}(\beta, \mathsf{SNR}) + \mathbb{E}\left[\log(1 + \beta|\mathsf{C}|^2\rho)\right] - \beta\,\theta\,\rho \log e$$

where

$$\theta\,\rho = 1 - \eta_{|\mathsf{A}|^2}(\mathsf{SNR}\,\theta) \quad (3.102)$$

$$\beta\,\theta\,\rho = 1 - \eta_{|\mathsf{C}|^2}(\rho\beta). \quad (3.103)$$

Note that $\theta(\mathsf{SNR}) = \mathbb{E}\left[|\mathsf{C}|^2\right]\; \eta^{\mathsf{mmse}}(\mathsf{SNR})$.

3.2.2.2 Orthogonal Sequences

In this setting we assume that $K \leq N$ and the channel matrix $\mathbf{H}$ can be written as the product of the $N \times N$ diagonal matrix $\mathbf{C} = \mathrm{diag}(\mathbf{c})$, an $N \times K$ matrix $\mathbf{Q}$ containing the spreading sequences and the $K \times K$ diagonal matrix $\mathbf{A}$ of complex fading coefficients:

$$\mathbf{y} = \mathbf{CQAx} + \mathbf{n}. \quad (3.104)$$

Here, $\mathbf{Q}$ is independent of $\mathbf{C}$ and of $\mathbf{A}$ and uniformly distributed over the manifold[9] of complex $N \times K$ matrices such that $\mathbf{Q}^\dagger\mathbf{Q} = \mathbf{I}$.

The arithmetic mean of the MMSE's for the K users satisfies

$$\frac{1}{K}\sum_{k=1}^{K} \mathsf{MMSE}_k = \frac{1}{K}\mathrm{tr}\left\{\left(\mathbf{I} + \mathsf{SNR}\,\mathbf{A}^\dagger\mathbf{Q}^\dagger\mathbf{C}^\dagger\mathbf{C}\mathbf{Q}\mathbf{A}\right)^{-1}\right\} \quad (3.105)$$

$$\xrightarrow{a.s.} \eta_{\mathbf{A}^\dagger\mathbf{Q}^\dagger\mathbf{C}^\dagger\mathbf{C}\mathbf{Q}\mathbf{A}}(\mathsf{SNR}) \quad (3.106)$$

$$= 1 - \frac{1}{\beta}(1 - \eta_{\mathbf{C}\mathbf{Q}\mathbf{A}\mathbf{A}^\dagger\mathbf{Q}^\dagger\mathbf{C}^\dagger}(\mathsf{SNR})) \quad (3.107)$$

where (3.107) comes from (2.56). For equal-power users ($\mathbf{A} = \mathbf{I}$), from Example 2.51 we have that

$$\eta_{\mathbf{C}\mathbf{Q}\mathbf{Q}^\dagger\mathbf{C}^\dagger}(\mathsf{SNR}) = \eta_{\mathbf{C}\mathbf{C}^\dagger}\left(\mathsf{SNR}\,\frac{\beta - 1 + \eta_{\mathbf{C}\mathbf{Q}\mathbf{Q}^\dagger\mathbf{C}^\dagger}}{\eta_{\mathbf{C}\mathbf{Q}\mathbf{Q}^\dagger\mathbf{C}^\dagger}(\mathsf{SNR})}\right). \quad (3.108)$$

From

$$\frac{1}{K}\sum_{k=1}^{K} \mathsf{MMSE}_k = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{1+\mathsf{SINR}_k} \quad (3.109)$$

it follows that, as $K, N \to \infty$,

$$\frac{1}{K}\sum_{k=1}^{K}\frac{1}{1+\mathsf{SINR}_k} \xrightarrow{a.s.} 1 - \frac{1}{\beta}(1 - \eta_{\mathbf{C}\mathbf{Q}\mathbf{Q}^\dagger\mathbf{C}^\dagger}(\mathsf{SNR})). \quad (3.110)$$

For equal-power users, the unitary invariance of $\mathbf{Q}$ results in each user admitting the same limiting MMSE and, from $\mathsf{MMSE}_k = \frac{1}{1+\mathsf{SINR}_k}$, the same limiting SINR:

$$\frac{1}{1+\mathsf{SINR}_k} \xrightarrow{a.s.} \frac{1}{1+\mathsf{SINR}}. \quad (3.111)$$

Consequently, (3.110) implies that

$$\beta\frac{\mathsf{SINR}}{1+\mathsf{SINR}} = 1 - \eta_{\mathbf{C}\mathbf{Q}\mathbf{Q}^\dagger\mathbf{C}^\dagger}(\mathsf{SNR})$$

which, in conjunction with (3.108), means that SINR is the solution to

$$\beta\frac{\mathsf{SINR}}{1+\mathsf{SINR}} = 1 - \eta_{\mathbf{C}\mathbf{C}^\dagger}\left(\mathsf{SNR}\,\frac{\beta}{1+\mathsf{SINR}(1-\beta)}\right) \quad (3.112)$$

[9] This is called the Stiefel manifold (cf. Section 2, Footnote 2).

whereas the multiuser efficiency of the kth user achieved by the MMSE receiver, $\eta_k^{\mathsf{mmse}}(\mathsf{SNR})$, converges almost surely to

$$\eta_k^{\mathsf{mmse}}(\mathsf{SNR}) \to \eta^{\mathsf{mmse}}\left(\mathsf{SNR}\,\mathbb{E}\left[|\mathsf{C}|^2\right]\right)$$

where the right side is the solution to the following equation at the point $\tau = \mathsf{SNR}\,\mathbb{E}\left[|\mathsf{C}|^2\right]$

$$\frac{\eta^{\mathsf{mmse}}}{1+\tau\eta^{\mathsf{mmse}}} = \mathbb{E}\left[\frac{|\tilde{\mathsf{C}}|^2}{\beta\tau|\tilde{\mathsf{C}}|^2 + 1 + (1-\beta)\tau\eta^{\mathsf{mmse}}}\right] \tag{3.113}$$

with $|\tilde{\mathsf{C}}|^2 = \frac{|\mathsf{C}|^2}{\mathbb{E}[|\mathsf{C}|^2]}$. A fixed-point equation equivalent to (3.113) was derived in [56].

For equal-power users, the spectral efficiencies achieved by the MMSE receiver and the decorrelator are

$$\mathcal{C}^{\mathsf{mmse}}(\beta, \mathsf{SNR}) = \beta \log\left(1 + \mathsf{SNR}\,\mathbb{E}\left[|\mathsf{C}|^2\right]\eta^{\mathsf{mmse}}(\mathsf{SNR})\right) \tag{3.114}$$

and, for $0 \le \beta \le 1$,

$$\mathcal{C}^{\mathsf{dec}}(\beta, \mathsf{SNR}) = \beta \log\left(1 + \mathsf{SNR}\,\mathbb{E}\left[|\mathsf{C}|^2\right](1-\beta)\right). \tag{3.115}$$

In parallel with [217, Eqn. (141)], the capacity of the optimum receiver is characterized in terms of the η-transform of $\mathbf{HH}^\dagger = \mathbf{CQQ}^\dagger\mathbf{C}^\dagger$

$$\mathcal{C}^{\mathsf{opt}}(\beta, \mathsf{SNR}) \quad = \quad \int_0^{\mathsf{SNR}} \frac{1}{x}\left(1 - \eta_{\mathbf{CQQ}^\dagger\mathbf{C}^\dagger}(x)\right) dx \tag{3.116}$$

with $\eta_{\mathbf{CQQ}^\dagger\mathbf{C}^\dagger}(\cdot)$ satisfying (3.108). An alternative characterization of the capacity (inspired by the optimality by successive cancellation with MMSE protection against uncancelled users) is given by

$$\mathcal{C}^{\mathsf{opt}}(\beta, \mathsf{SNR}) \quad = \quad \beta\,\mathbb{E}\left[\log\left(1 + \daleth(\mathsf{Y}, \mathsf{SNR})\right)\right] \tag{3.117}$$

with

$$\frac{\daleth(y, \mathsf{SNR})}{1 + \daleth(y, \mathsf{SNR})} = \mathbb{E}\left[\frac{\mathsf{SNR}\,|\mathsf{C}|^2}{\beta\, y\,\mathsf{SNR}\,|\mathsf{C}|^2 + 1 + (1-\beta\, y)\daleth(y, \mathsf{SNR})}\right] \tag{3.118}$$

where Y is a random variable uniform on $[0, 1]$.

The case of unequal-power users has been analyzed in [37] with the restrictive setup of a finite number of user classes where the power is allowed to vary across classes but not over users within each class. Reference [37] shows that the SINR of the kth user at the output of the MMSE receiver, SINR_k, and consequently $\eta_k^{\mathsf{mmse}}(\mathsf{SNR})$, converge almost surely to nonrandom limits. Specifically, the multiuser efficiency converges to the solution η of

$$\mathbb{E}\left[\frac{|\tilde{\mathsf{C}}|^2}{\beta|\tilde{\mathsf{C}}|^2\left(1-\eta_{|\mathsf{A}|^2}(\tau\eta)\right)+\eta\left(1-\beta+\beta\eta_{|\mathsf{A}|^2}(\tau\eta)\right)}\right]=1 \tag{3.119}$$

with $\tau = \mathsf{SNR}\,\mathbb{E}[|\mathsf{C}|^2]$. From the multiuser efficiency, the capacity can be readily obtained using the optimality of successive interference cancellation as done in (3.117).

3.2.2.3 Orthogonal Sequences vs i.i.d. Sequences

The multiuser efficiency achieved by the MMSE receiver where i.i.d. spreading sequences are utilized, given in (3.98), can be rewritten as

$$\frac{\eta^{\mathsf{mmse}}}{1+\tau\eta^{\mathsf{mmse}}}=\mathbb{E}\left[\frac{|\tilde{\mathsf{C}}|^2}{\beta\tau|\tilde{\mathsf{C}}|^2+1+\tau\eta^{\mathsf{mmse}}}\right] \tag{3.120}$$

with $\tau = \mathsf{SNR}\,\mathbb{E}[|\mathsf{C}|^2]$. A comparison of (3.120) and (3.113) reveals that, for a fixed $\beta > 0$, the SINR in the i.i.d. case is always less than in the orthogonal case. Moreover, the performance gain induced by the use of orthogonal instead of i.i.d. spreading sequences grows when β approaches 1. If $\beta \sim 0$, then the output SINR in the two cases is basically equal. Moreover, from (3.98) and (3.113) it follows respectively that

$$\mathsf{SINR}_{\mathsf{i.i.d}}=\mathbb{E}\left[\frac{|\tilde{\mathsf{C}}|^2}{\frac{1}{\tau}+\beta\frac{|\tilde{\mathsf{C}}|^2}{1+\mathsf{SINR}_{\mathsf{i.i.d}}}}\right] \tag{3.121}$$

and

$$\mathsf{SINR}_{\mathsf{orth}}=\mathbb{E}\left[\frac{|\tilde{\mathsf{C}}|^2}{\frac{1}{\tau}\left(1-\beta\frac{\mathsf{SINR}_{\mathsf{orth}}}{1+\mathsf{SINR}_{\mathsf{orth}}}\right)+\beta\frac{|\tilde{\mathsf{C}}|^2}{1+\mathsf{SINR}_{\mathsf{orth}}}}\right]. \tag{3.122}$$

Notice, by comparing (3.121) and (3.122), that in the latter the term $\frac{1}{\tau}=\frac{1}{\mathsf{SNR}\,\mathbb{E}[|\mathsf{C}|^2]}$ is multiplied by $\left(1-\beta\frac{\mathsf{SINR}_{\mathsf{orth}}}{\mathsf{SINR}_{\mathsf{orth}}+1}\right)$, which is less than 1.

Accordingly, for a given SINR the required SNR is reduced with respect to the one required with i.i.d sequences.

3.2.3 Reduced Rank Receiver for MC-CDMA

In the downlink, the fading experienced by the N subcarriers is common to all users. The asymptotic weights of the rank-D MMSE receiver for the downlink can be easily derived from

$$\mathcal{H}_m^\infty = |A_k|^2 \sum_{n=0}^{m} \binom{m}{n} \sigma^{2m-2n} \xi_n \tag{3.123}$$

where, in the case of i.i.d. spreading sequences,

$$\xi_n = \beta \sum_{\ell=1}^{n} E\left[m_{\ell-1}(|\mathsf{C}|^2)\,|\mathsf{C}|^4\right] \sum_{\substack{n_1+\cdots+n_i=n-\ell \\ 1\le i\le n-\ell}} E\left[|\mathsf{A}|^{2i+2}\right] \xi_{n_1-1}\ldots\xi_{n_i-1} \tag{3.124}$$

and

$$m_n(r) = \beta r \sum_{\ell=1}^{n} m_{\ell-1}(r) \sum_{\substack{n_1+\cdots+n_i=n-\ell \\ 1\le i\le n-\ell}} E\left[|\mathsf{A}|^{2i+2}\right] \xi_{n_1-1}\ldots\xi_{n_i-1} \tag{3.125}$$

with $|\mathsf{C}|$ and $|\mathsf{A}|$ random variables whose distributions equal the asymptotic empirical distributions of the singular values of $\mathbf{C}$ and $\mathbf{A}$, respectively. In the case of orthogonal sequences, the counterparts of (3.124) and (3.125) can be found in [105].

For the uplink, the binomial expansion (3.62) becomes

$$\mathcal{H}_m^\infty = \sum_{n=0}^{m} \binom{m}{n} \sigma^{2m-2n} \xi_{n,k} \tag{3.126}$$

where

$$\xi_{n,k} = \mathbb{E}[m_n(\mathsf{X}) v_k(\mathsf{X})] \tag{3.127}$$

with $m_n(\cdot)$ solution to the recursive equation

$$\begin{aligned} m_n(x) \quad = \quad & \beta \sum_{\ell=1}^{n} m_{\ell-1}(x)\, \mathbb{E}[\, v(x,\mathsf{Y}) \sum_{\substack{n_1+\cdots+n_i=n-\ell \\ 1\le i\le n-\ell}} \mathbb{E}\left[v(\mathsf{X},\mathsf{Y}) m_{n_1-1}(\mathsf{X})|\mathsf{Y}\right] \\ & \ldots \mathbb{E}\left[v(\mathsf{X},\mathsf{Y}) m_{n_i-1}(\mathsf{X})|\mathsf{Y}\right]\,] \end{aligned} \tag{3.128}$$

where $v(\cdot,\cdot)$ is the two-dimensional channel profile of $\mathbf{B}$ as defined in Section 3.2.1.

3.3 Single-User Multi-Antenna Channels

Let us now consider the problem of a single-user channel where the transmitter has $n_{\rm T}$ antennas while the receiver has $n_{\rm R}$ antennas. (See [250, 76] for the initial contributions on this topic and [60, 90, 82, 24, 23] for recent articles of tutorial nature.)

3.3.1 Preliminaries

With reference to the general model in (1.1), $\mathbf{x}$ contains the symbols transmitted from the $n_{\rm T}$ transmit antennas and $\mathbf{y}$ the symbols received by the $n_{\rm R}$ receive antennas with $\frac{n_{\rm T}}{n_{\rm R}} \to \beta$ when $n_{\rm T}$ and $n_{\rm R}$ grow large. The entries of $\mathbf{H}$ represent the fading coefficients between each transmit and each receive antenna normalized such that[10]

$$\mathbb{E}\left[\mathrm{tr}\left\{\mathbf{H}\mathbf{H}^\dagger\right\}\right] = n_{\rm R} \tag{3.129}$$

while

$$\mathsf{SNR} = \frac{\mathbb{E}[\|\mathbf{x}\|^2]}{\frac{1}{n_{\rm R}}\mathbb{E}[\|\mathbf{n}\|^2]}. \tag{3.130}$$

In contrast with the multiaccess scenarios, in this case the signals transmitted by different antennas can be advantageously correlated and thus the covariance of $\mathbf{x}$ becomes relevant. Normalized by its energy per dimension, the input covariance is denoted by

$$\mathbf{\Phi} = \frac{\mathbb{E}[\mathbf{x}\mathbf{x}^\dagger]}{\frac{1}{n_{\rm T}}\mathbb{E}[\|\mathbf{x}\|^2]} \tag{3.131}$$

where the normalization ensures that $\mathbb{E}[\mathrm{tr}\{\mathbf{\Phi}\}] = n_{\rm T}$. It is useful to decompose this input covariance in its eigenvectors and eigenvalues,

[10] Although, in most of the multi-antenna literature, $\mathbb{E}\left[\mathrm{tr}\left\{\mathbf{H}\mathbf{H}^\dagger\right\}\right] = n_{\rm T} n_{\rm R}$, for consistency with the rest of the paper we use the normalization in (3.129). In the case that the entries of $\mathbf{H}$ are identically distributed, the resulting variance of each entry is $\frac{1}{n_{\rm T}}$.

$\mathbf{\Phi} = \mathbf{V}\mathbf{P}\mathbf{V}^\dagger$. Each eigenvalue represents the (normalized) power allocated to the corresponding signalling eigenvector. Associated with $\mathbf{P}$, we define an input *power profile*

$$\mathcal{P}^{(n_\mathrm{R})}(t, \mathsf{SNR}) = \mathsf{P}_{j,j} \qquad\qquad \frac{j}{n_\mathrm{R}} \leq t < \frac{j+1}{n_\mathrm{R}}$$

supported on $t \in (0, \beta]$. This profile specifies the power allocation at each SNR. As the number of antennas is driven to infinity, $\mathcal{P}^{(n_\mathrm{R})}(t, \mathsf{SNR})$ converges uniformly to a nonrandom function, $\mathcal{P}(t, \mathsf{SNR})$, which we term *asymptotic power profile.*

In order to achieve capacity, the input covariance $\mathbf{\Phi}$ must be properly determined depending on the channel-state information (CSI) available to the transmitter. In this respect, there are three main regimes of interest:

- The transmitter has full CSI, i.e., access to $\mathbf{H}$ instantaneously. In this case, $\mathbf{\Phi}$ can be made a function of $\mathbf{H}$. This operational regime applies, for example, to fixed wireless access systems where transmitter and receiver are stationary (backhaul, local loop, broadband residential) and to low-mobility systems (local-area networks, pedestrians). It is particularly appealing whenever uplink and downlink are reciprocal (time-duplexed systems) [48].
- The transmitter has only statistical CSI, i.e., access to the distribution of $\mathbf{H}$ but not to its realization. In this case, $\mathbf{\Phi}$ cannot depend on $\mathbf{H}$. This is the usual regime in high-mobility and wide-area systems, especially if link reciprocity does not hold.
- The transmitter has no CSI whatsoever.

For all these scenarios, the capacity per receive antenna is given by the maximum over $\mathbf{\Phi}$ of the Shannon transform of the averaged empirical distribution of $\mathbf{H}\mathbf{\Phi}\mathbf{H}^\dagger$, i.e.

$$\mathcal{C}(\mathsf{SNR}) \quad = \max_{\mathbf{\Phi}:\mathrm{tr}\mathbf{\Phi}=n_\mathrm{T}} \mathcal{V}_{\mathbf{H}\mathbf{\Phi}\mathbf{H}^\dagger}(\mathsf{SNR}). \tag{3.132}$$

If full CSI is available at the transmitter, then $\mathbf{V}$ should coincide with the eigenvector matrix of $\mathbf{H}^\dagger\mathbf{H}$ and $\mathbf{P}$ should be obtained through a waterfill process on the eigenvalues of $\mathbf{H}^\dagger\mathbf{H}$ [260, 47, 250, 205]. The resulting jth diagonal entry of $\mathbf{P}$ is

$$\mathsf{P}_{j,j} = \left(\nu - \frac{1}{\mathsf{SNR}\ \lambda_j(\mathbf{H}^\dagger\mathbf{H})}\right)^+ \tag{3.133}$$

where ν is such that $\mathrm{tr}\{\mathbf{P}\} = n_\mathrm{T}$. Then, substituting in (3.132),

$$\begin{aligned} \mathcal{C}(\mathsf{SNR}) &= \frac{1}{n_\mathrm{R}} \log\det(\mathbf{I} + \mathsf{SNR}\,\mathbf{P}\boldsymbol{\Lambda}) & (3.134)\\ &= \beta \int (\log(\mathsf{SNR}\,\nu\lambda))^+ d\mathsf{F}^{n_\mathrm{T}}_{\mathbf{H}^\dagger\mathbf{H}}(\lambda) & (3.135) \end{aligned}$$

with $\boldsymbol{\Lambda}$ equal to the diagonal eigenvalue matrix of $\mathbf{H}^\dagger\mathbf{H}$.

If, instead, only statistical CSI is available, then $\mathbf{V}$ should be set, for all the channels that we will consider, to coincide with the eigenvectors of $\mathbb{E}[\mathbf{H}^\dagger\mathbf{H}]$ while the capacity-achieving power allocation, $\mathbf{P}$, can be found iteratively [264].

With no CSI, the most reasonable strategy is to transmit an isotropic signal ($\boldsymbol{\Phi} = \mathbf{I}$) [195, 300]. In fact, because of its simplicity and because many space-time coding schemes conform to it, this strategy may be appealing even if some degree of CSI is available.

3.3.2 Canonical Model

The pioneering analyses that ignited research on this topic [250, 76] started with $\mathbf{H}$ having i.i.d. zero-mean complex Gaussian random entries (all antennas implicitly assumed identical and uncorrelated).

For this canonical channel, the capacity with full CSI converges asymptotically to [43, 98, 177, 212]

$$\mathcal{C}(\mathsf{SNR}) = \beta \int_{\max\{a,\nu^{-1}\}}^{b} \log\left(\frac{\nu\,\mathsf{SNR}}{\beta}\lambda\right) \mathsf{f}_\beta(\lambda)\, d\lambda \tag{3.136}$$

where ν satisfies

$$\int_{\max\{a,\nu^{-1}\}}^{b} \left(\nu - \frac{\beta}{\mathsf{SNR}\ \lambda}\right)^+ \mathsf{f}_\beta(\lambda)\, d\lambda = 1 \tag{3.137}$$

with a, b and $\mathsf{f}_\beta(\cdot)$ given in (1.10).

If $\nu \geq \frac{\beta}{\mathsf{SNR}\, a}$, then the integrals in (3.136) and (3.137) admit closed-form expressions. Since, with full CSI at the transmitter, the capacity is reciprocal in terms of the roles played by transmitter and receiver [250], we have that

$$\mathcal{C}(\beta, \mathsf{SNR}) = \beta\, \mathcal{C}\left(\frac{1}{\beta}, \mathsf{SNR}\right) \tag{3.138}$$

and thus we need only solve the integrals for $\beta < 1$. Applying Example 2.15 to (3.136) and Theorem 2.10 to (3.137) and exploiting (3.138), the following result is obtained.

Theorem 3.5. [263] For

$$\mathsf{SNR} \geq \frac{2\min\{1, \beta^{3/2}\}}{|1-\sqrt{\beta}||1-\beta|} \tag{3.139}$$

the capacity of the canonical channel with full CSI at the transmitter converges almost surely to

$$\mathcal{C}(\mathsf{SNR}) = \begin{cases} \beta \log\left(\frac{\mathsf{SNR}}{\beta} + \frac{1}{1-\beta}\right) + (1-\beta)\log\frac{1}{1-\beta} - \beta\log e & \beta < 1 \\ \log\left(\beta\,\mathsf{SNR} + \frac{\beta}{\beta-1}\right) + (\beta-1)\log\frac{\beta}{\beta-1} - \log e & \beta > 1. \end{cases}$$

Theorem 3.5 is illustrated in Figure 3.2 for various numbers of antennas. The solid lines indicate the asymptotic solutions, with the role of β played by $\frac{n_{\mathrm{T}}}{n_{\mathrm{R}}}$, while the circles show the outcome of corresponding Monte-Carlo simulations. Notice the power of the asymptotic analysis for SNR levels satisfying (3.139).

For $\beta = 1$, the asymptotic capacity with full CSI is known only for $\mathsf{SNR} \to \infty$, in which case it coincides with the mutual information achieved by an isotropic input, presented later in this section [43].

Non-asymptotically in the number of antennas, the capacity with full transmit CSI is studied in [4, 127]. In [127], in particular, an explicit expression is given although as function of a parameter that must be solved for numerically.

With statistical CSI, it was shown in [250] that capacity is achieved with $\mathbf{\Phi} = \mathbf{I}$. For fixed number of antennas, [250] gave an integral expression (integrating $\log(1 + \mathsf{SNR}\,\lambda)$ with respect to the p.d.f. in (2.23))

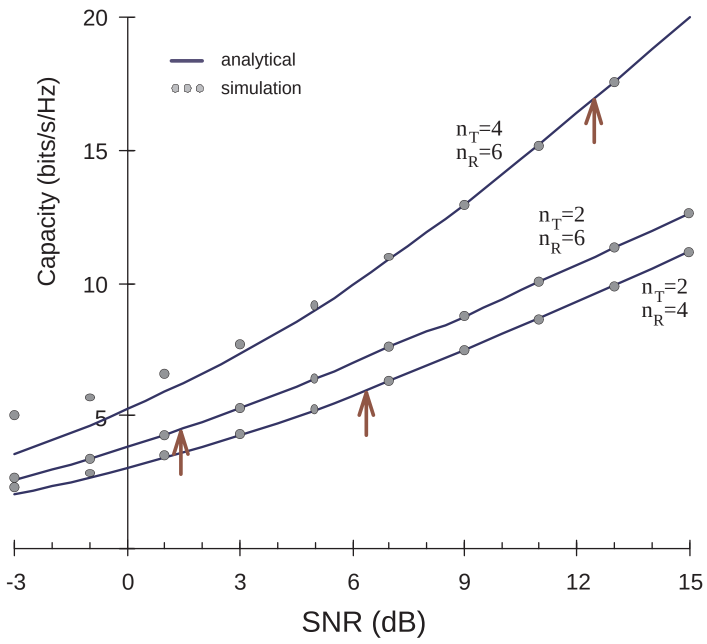

Fig. 3.2 Capacity of a canonical channel with various numbers of transmit and receive antennas. The arrows indicate the SNR above which (3.139) is satisfied.

for the expected capacity as a function of the signal-to-noise ratio and the number of transmit and receive antennas. This integral involving the Laguerre polynomials lends itself to an explicit expression. This has been accomplished in [219, 61, 126]. In particular, [126] uses the Mellin transform and Theorem 2.30 to arrive at a closed-form expression, and [61] gives the expression in Example 2.17.

Asymptotically, as the numbers of transmit and receive antennas grow with ratio β, the capacity per receive antenna converges almost surely to [275, 206]

$$\begin{aligned}\mathcal{C}(\beta, \mathsf{SNR}) \quad = \quad & \beta \log\left(1 + \frac{\mathsf{SNR}}{\beta} - \frac{1}{4}\mathcal{F}\left(\frac{\mathsf{SNR}}{\beta}, \beta\right)\right) \\ & + \log\left(1 + \mathsf{SNR} - \frac{1}{4}\mathcal{F}\left(\frac{\mathsf{SNR}}{\beta}, \beta\right)\right) - \beta \frac{\log e}{4\,\mathsf{SNR}} \mathcal{F}\left(\frac{\mathsf{SNR}}{\beta}, \beta\right)\end{aligned} \tag{3.140}$$

with $\mathcal{F}(\cdot,\cdot)$ given in (1.17). Notice that this capacity coincides, except for a signal-to-noise scaling, with that of an unfaded equal-power DS-CDMA channel.[11]

If $\beta = 1$, the asymptotic capacity per receive antenna with statistical CSI at the transmitter is equal to

$$\mathcal{C}(\beta, \mathsf{SNR}) = 2\log\left(\frac{1+\sqrt{1+4\mathsf{SNR}}}{2}\right) - \frac{\log e}{4\,\mathsf{SNR}}\left(\sqrt{1+4\,\mathsf{SNR}}-1\right)^2$$

evidencing the linear growth with the number of antennas originally observed in [250, 76]. Further insight can be drawn, for arbitrary β, from the high-SNR behavior of the capacity (cf. Example 2.15):

$$\mathcal{C}(\mathsf{SNR}) = \begin{cases} \log\frac{\mathsf{SNR}}{e} - (\beta-1)\log\frac{\beta-1}{\beta} + o(1) & \beta > 1 \\ \log\frac{\mathsf{SNR}}{e} + o(1) & \beta = 1 \\ \beta\log\frac{\mathsf{SNR}}{\beta e} - (1-\beta)\log(1-\beta) + o(1) & \beta < 1. \end{cases}$$

Besides asymptotically in the number of antennas, the high-SNR capacity can be characterized for fixed n_T and n_R via (2.12) in Theorem 2.11. Also in this case, the capacity is seen to scale linearly with the number of antennas, more precisely with $\min(n_\mathrm{T}, n_\mathrm{R})$. While this scaling makes multi-antenna communication highly appealing, it hinges on the validity of the idealized canonical channel model. Much of the research that has ensued, surveyed in the remainder of this section, is geared precisely at accounting for various nonidealities (correlation, deterministic channel components, etc) that have the potential of compromising this linear scaling.

3.3.3 Separable Correlation Model

The most immediate effect that results from locating various antennas in close proximity is that their signals tend to be, to some extent,

[11] In addition to its role in the analysis of multiaccess and single-user multi-antenna channels, (3.140) also plays a role in the analysis of the total capacity of the Gaussian broadcast channel with multiple antennas at the transmitter [112, 282]. As shown in [35, 316, 276, 280, 128, 259, 278, 36, 277, 302], in various degrees of generality, the multi-antenna broadcast channel capacity region is equal to the union of capacity regions of the dual multiaccess channel, where the union is taken over all individual power constraints that sum to the averaged power constraint.

correlated. In its full generality, the correlation between the (i, j) and (i', j') entries of $\mathbf{H}$ is given by

$$\mathsf{r}_{\mathbf{H}}(i, i', j, j') = \mathbb{E}\left[\mathsf{H}_{i,j}\mathsf{H}^*_{i',j'}\right]. \tag{3.141}$$

In a number of interesting cases, however, correlation turns out to be a strictly local phenomenon that can be modeled in a simplified manner. To that end, the so-called *separable* (also termed *Kronecker* or *product*) correlation model was proposed by several authors [220, 40, 203]. According to this model, an $n_{\mathrm{R}} \times n_{\mathrm{T}}$ matrix $\mathbf{H}_w$, whose entries are i.i.d. zero-mean with variance $\frac{1}{n_{\mathrm{T}}}$, is pre- and post-multiplied by the square root of deterministic matrices, $\mathbf{\Theta}_{\mathrm{T}}$ and $\mathbf{\Theta}_{\mathrm{R}}$, whose entries represent, respectively, the correlation between the transmit antennas and between the receive antennas:

$$\mathbf{H} = \mathbf{\Theta}_{\mathrm{R}}^{1/2}\mathbf{H}_w\mathbf{\Theta}_{\mathrm{T}}^{1/2}. \tag{3.142}$$

Implied by this model is that the correlation between two transmit antennas is the same regardless of the receive antenna at which the observation is made and viceversa. As confirmed experimentally in [41], this condition is often satisfied in outdoor environments if the arrays are composed by antennas with similar polarization and radiation patterns.

When (3.142) holds, the correlation in (3.141) can be expressed (cf. Definition 2.9) as

$$\mathsf{r}_{\mathbf{H}}(i, i', j, j') = \frac{(\mathbf{\Theta}_{\mathrm{R}})_{i,i'}(\mathbf{\Theta}_{\mathrm{T}})_{j,j'}}{n_{\mathrm{T}}}. \tag{3.143}$$

Results on the asymptotic capacity and mutual information, with various levels of transmitter information, of channels that obey this model can be found in [181, 262, 43, 263, 178]. Analytical non-asymptotic expressions have also been reported: in [208, 209, 2], the capacity of one-sided correlated channels is obtained starting from the joint distribution of the eigenvalues of a Wishart matrix $\sim \mathcal{W}_m(n, \mathbf{\Sigma})$ given in Theorem 2.18 and (2.19). References [135, 234, 149, 39] compute the moment generating function of the mutual information of a one-sided correlated MIMO channel, constraining the eigenvalues of the correlation matrix to be distinct. The two-sided correlated MIMO channel is analyzed in [148, 231, 149] also through the moment generating function of the mutual information (cf. (2.16)).

With full CSI at the transmitter, the asymptotic capacity is [43]

$$\mathcal{C}(\mathsf{SNR}) = \beta \int_0^\infty (\log(\mathsf{SNR}\,\nu\lambda))^+ dG(\lambda) \tag{3.144}$$

where ν satisfies

$$\int_0^\infty \left(\nu - \frac{1}{\mathsf{SNR}\,\lambda}\right)^+ dG(\lambda) = 1 \tag{3.145}$$

with $G(\cdot)$ the asymptotic spectrum of $\mathbf{H}^\dagger\mathbf{H}$ whose η-transform can be derived using Theorem 2.43 and Lemma 2.28. Invoking Theorem 2.45, the capacity in (3.144) can be evaluated as follows.

Theorem 3.6. [263] Let Λ_R and Λ_T be independent random variables whose distributions are the asymptotic spectra of the full-rank matrices $\mathbf{\Theta}_\mathrm{R}$ and $\mathbf{\Theta}_\mathrm{T}$ respectively. Further define

$$\Lambda_1 = \begin{cases} \Lambda_\mathsf{T} & \beta < 1 \\ \Lambda_\mathsf{R} & \beta > 1 \end{cases} \qquad \Lambda_2 = \begin{cases} \Lambda_\mathsf{R} & \beta < 1 \\ \Lambda_\mathsf{T} & \beta > 1 \end{cases} \tag{3.146}$$

and let κ be the infimum (excluding any mass point at zero) of the support of the asymptotic spectrum of $\mathbf{H}^\dagger\mathbf{H}$. For

$$\mathsf{SNR} \geq \frac{1}{\kappa} - \delta\mathbb{E}\left[\frac{1}{\Lambda_1}\right] \tag{3.147}$$

with δ satisfying

$$\eta_{\Lambda_2}(\delta) = 1 - \min\{\beta, \tfrac{1}{\beta}\},$$

the asymptotic capacity of a channel with separable correlations and full CSI at the transmitter is

$$\mathcal{C}(\mathsf{SNR}) = \begin{cases} \beta\,\mathbb{E}\left[\log\frac{\Lambda_\mathsf{T}}{e\vartheta}\right] + \mathcal{V}_{\Lambda_\mathsf{R}}(\vartheta) + \beta\log\left(\mathsf{SNR} + \vartheta\mathbb{E}[\frac{1}{\Lambda_\mathsf{T}}]\right) & \beta < 1 \\ \mathbb{E}\left[\log\frac{\Lambda_\mathsf{R}}{\alpha e}\right] + \beta\,\mathcal{V}_{\Lambda_\mathsf{T}}(\alpha) + \log\left(\mathsf{SNR} + \alpha\mathbb{E}[\frac{1}{\Lambda_\mathsf{R}}]\right) & \beta > 1 \end{cases}$$

with α and ϑ the solutions to

$$\eta_{\Lambda_\mathsf{T}}(\alpha) = 1 - \frac{1}{\beta} \qquad \eta_{\Lambda_\mathsf{R}}(\vartheta) = 1 - \beta.$$

As for the canonical channel, no asymptotic characterization of the capacity with full CSI at the transmitter is known for $\beta = 1$ and arbitrary SNR.

When the correlation is present only at either the transmit or receive ends of the link, the solutions in Theorem 3.6 sometimes become explicit:

Corollary 3.1. With correlation at the end of the link with the fewest antennas, the capacity per antenna with full CSI at the transmitter converges to

$$\mathcal{C} = \begin{cases} \beta\,\mathbb{E}\left[\log\frac{\Lambda_{\mathsf{T}}}{e}\right] + \log\frac{1}{1-\beta} + \beta\log\left(\mathsf{SNR}\,\frac{1-\beta}{\beta} + \mathbb{E}[\frac{1}{\Lambda_{\mathsf{T}}}]\right) & \begin{matrix}\beta < 1 \\ \Lambda_{\mathsf{R}} = 1\end{matrix} \\ E\left[\log\frac{\Lambda_{\mathsf{R}}}{e}\right] - \beta\log\frac{\beta-1}{\beta} + \log\left(\mathsf{SNR}\,(\beta-1) + \mathbb{E}[\frac{1}{\Lambda_{\mathsf{R}}}]\right) & \begin{matrix}\beta > 1 \\ \Lambda_{\mathsf{T}} = 1.\end{matrix} \end{cases}$$

With statistical CSI at the transmitter, achieving capacity requires that the eigenvectors of the input covariance, $\mathbf{\Phi}$, coincide with those of $\mathbf{\Theta}_{\mathrm{T}}$ [279, 123]. Consequently, denoting by $\mathbf{\Lambda}_{\mathrm{T}}$ and $\mathbf{\Lambda}_{\mathrm{R}}$ the diagonal eigenvalue matrices of $\mathbf{\Theta}_{\mathrm{T}}$ and $\mathbf{\Theta}_{\mathrm{R}}$, respectively, we have that

$$\mathcal{C}(\beta, \mathsf{SNR}) = \frac{1}{N}\log\det\left(\mathbf{I} + \mathsf{SNR}\,\mathbf{\Lambda}_{\mathrm{R}}^{1/2}\mathbf{H}_w\mathbf{\Lambda}_{\mathrm{T}}^{1/2}\mathbf{P}\mathbf{\Lambda}_{\mathrm{T}}^{1/2}\mathbf{H}_w^{\dagger}\mathbf{\Lambda}_{\mathrm{R}}^{1/2}\right)$$

where $\mathbf{P}$ is the capacity-achieving power allocation [264]. Applying Theorem 2.44, we obtain:

Theorem 3.7. [262] The capacity of a Rayleigh-faded channel with separable transmit and receive correlation matrices $\mathbf{\Theta}_{\mathrm{T}}$ and $\mathbf{\Theta}_{\mathrm{R}}$ and statistical CSI at the transmitter converges to

$$\begin{aligned} \mathcal{C}(\beta, \mathsf{SNR}) &= \beta E\left[\log(1 + \mathsf{SNR}\,\Lambda\Gamma(\mathsf{SNR}))\right] + E\left[\log(1 + \mathsf{SNR}\,\Lambda_{\mathsf{R}}\Upsilon(\mathsf{SNR})\right] \\ &\quad -\beta\,\mathsf{SNR}\,\Gamma(\mathsf{SNR})\Upsilon(\mathsf{SNR})\log e \end{aligned} \tag{3.148}$$

where

$$\Gamma(\mathsf{SNR}) = \frac{1}{\beta}E\left[\frac{\Lambda_{\mathsf{R}}}{1 + \mathsf{SNR}\,\Lambda_{\mathsf{R}}\Upsilon(\mathsf{SNR})}\right] \tag{3.149}$$

$$\Upsilon(\mathsf{SNR}) = E\left[\frac{\Lambda}{1 + \mathsf{SNR}\,\Lambda\Gamma(\mathsf{SNR})}\right] \tag{3.150}$$

with expectation over Λ and Λ_{R} whose distributions are given by the asymptotic empirical eigenvalue distributions of $\mathbf{\Lambda}_{\mathrm{T}}\mathbf{P}$ and $\mathbf{\Theta}_{\mathrm{R}}$, respectively.

If the input is isotropic, the achievable mutual information is easily found from the foregoing result.

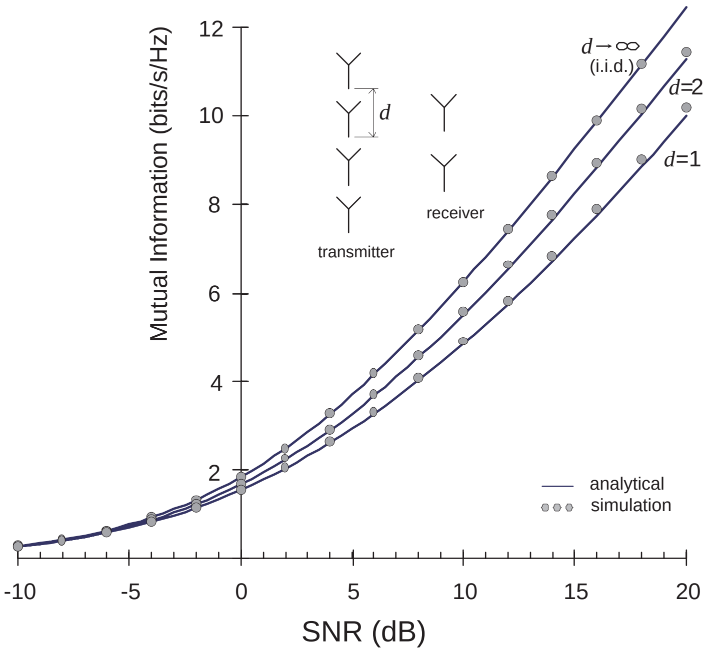

Fig. 3.3 Mutual information achieved by an isotropic input on a Rayleigh-faded channel with $n_{\mathrm{T}} = 4$ and $n_{\mathrm{R}} = 2$. The transmitter is a uniform linear array whose antenna correlation is given by (3.151) where d is the spacing (wavelengths) between adjacent antennas. The receive antennas are uncorrelated.

Corollary 3.2. [266] Consider a channel defined as in Theorem 3.7 and an isotropic input. Expression (3.148) yields the mutual information with the distribution of Λ given by the asymptotic empirical eigenvalue distribution of $\mathbf{\Theta}_{\mathrm{T}}$.

This corollary is illustrated in Fig. 3.3, which depicts the mutual information (bits/s/Hz) achieved by an isotropic input for a wide range of SNR. The channel is Rayleigh-faded with $n_{\mathrm{T}} = 4$ correlated antennas and $n_{\mathrm{R}} = 2$ uncorrelated antennas. The correlation between the ith and jth transmit antennas is

$$(\mathbf{\Theta}_{\mathrm{T}})_{i,j} = e^{-0.05d^2(i-j)^2} \tag{3.151}$$

which corresponds to a uniform linear array with antenna separation d (wavelengths) exposed to a broadside Gaussian azimuth angular spectrum with a 2° root-mean-square spread [42]. Such angular spread is typical of an elevated base station in rural or suburban areas. The solid lines depict the analytical solution obtained by applying Theorem 3.7 with $\mathbf{P} = \mathbf{I}$ and $\mathbf{\Theta}_{\mathrm{R}} = \mathbf{I}$ and with the expectations over Λ replaced with arithmetic averages over the eigenvalues of $\mathbf{\Theta}_{\mathrm{T}}$. The circles, in turn, show the result of Monte-Carlo simulations. Notice the excellent agreement even for such small numbers of antennas.

The high-SNR behaviors of the capacity with statistical CSI and of the mutual information achieved by an isotropic input can be characterized, asymptotically in the number of antennas, using Theorem 2.45. For arbitrary n_{T} and n_{R}, such characterizations can be found in [165].

3.3.4 Non-Separable Correlations

While the separable correlation model is relatively simple and analytically appealing, it also has clear limitations, particularly in terms of representing indoor propagation environments [194]. Also, it does not accommodate diversity mechanisms such as polarization[12] and radiation pattern diversity[13] that are becoming increasingly popular as they enable more compact arrays. The use of different polarizations and/or radiation patterns creates correlation structures that cannot be represented through the separable model.

In order to encompass a broader range of correlations, we model the channel as

$$\mathbf{H} = \mathbf{U}_{\mathrm{R}} \tilde{\mathbf{H}} \mathbf{U}_{\mathrm{T}}^{\dagger} \tag{3.152}$$

where $\mathbf{U}_{\mathrm{R}}$ and $\mathbf{U}_{\mathrm{T}}$ are unitary while the entries of $\tilde{\mathbf{H}}$ are independent zero-mean Gaussian. This model is advocated and experimentally supported in [301] and its capacity is characterized asymptotically in [262].

[12] Polarization diversity: Antennas with orthogonal polarizations are used to ensure low levels of correlation with minimum or no antenna spacing [156, 236] and to make the communication link robust to polarization rotations in the channel [19].

[13] Pattern diversity: Antennas with different radiation patterns or with rotated versions of the same pattern are used to discriminate different multipath components and reduce correlation.

For the more restrictive case where $\mathbf{U}_\mathrm{R}$ and $\mathbf{U}_\mathrm{T}$ are Fourier matrices, the model (3.152) was proposed earlier in [213].

The matrices $\mathbf{H}$ and $\tilde{\mathbf{H}}$ are directly related through the Karhunen-Loève expansion (cf. Lemma 2.25) with the variances of the entries of $\tilde{\mathbf{H}}$ given by the eigenvalues of $\mathsf{r}_{\mathbf{H}}$ obtained by solving the system of equations in (2.33). Furthermore, from Theorem 2.58, the asymptotic spectrum of $\mathbf{H}$ is fully characterized by the variances of the entries of $\tilde{\mathbf{H}}$, which we assemble in a matrix $\mathbf{G}$ such that $\mathsf{G}_{i,j} = n_\mathrm{T}\mathbb{E}[|\tilde{\mathsf{H}}_{i,j}|^2]$ with

$$\sum_{ij} \mathsf{G}_{i,j} = n_\mathrm{T} n_\mathrm{R}. \tag{3.153}$$

Invoking Definition 2.16, we introduce the *variance profile* of $\tilde{\mathbf{H}}$, which maps the entries of $\mathbf{G}$ onto a two-dimensional piece-wise constant function

$$\mathcal{G}^{(n_\mathrm{R})}(r,t) = \mathsf{G}_{i,j} \qquad \tfrac{i}{n_\mathrm{R}} \le r < \tfrac{i+1}{n_\mathrm{R}}, \quad \tfrac{j}{n_\mathrm{T}} \le t < \tfrac{j+1}{n_\mathrm{T}} \tag{3.154}$$

supported on $r, t \in [0,1]$. We can interpret r and t as normalized receive and transmit antenna indices. It is assumed that, as the number of antennas grows, $\mathcal{G}^{(n_\mathrm{R})}(r,t)$ converges uniformly to the *asymptotic variance profile*, $\mathcal{G}(r,t)$. The normalization condition in (3.153) implies that

$$\mathbb{E}[\mathcal{G}(\mathsf{R},\mathsf{T})] = 1 \tag{3.155}$$

with R and T independent random variables uniform on $[0,1]$.

With full CSI at the transmitter, the asymptotic capacity is given by (3.144) and (3.145) with $G(\cdot)$ representing the asymptotic spectrum of $\mathbf{H}^\dagger\mathbf{H}$. Using Theorems 2.58 and 2.54, an explicit expression for $\mathcal{C}(\mathsf{SNR})$ can be obtained for sufficiently high SNR.

With statistical CSI at the transmitter, the eigenvectors of the capacity-achieving input covariance coincide with the columns of $\mathbf{U}_\mathrm{T}$ in (3.152) [261, 268]. In order to characterize the capacity, we invoke Theorem 2.53 to obtain the following.

Theorem 3.8. [262] Consider the channel $\mathbf{H} = \mathbf{U}_\mathrm{R}\tilde{\mathbf{H}}\mathbf{U}_\mathrm{T}^\dagger$ where $\mathbf{U}_\mathrm{R}$ and $\mathbf{U}_\mathrm{T}$ are unitary while the entries of $\tilde{\mathbf{H}}$ are zero-mean Gaussian and independent. Denote by $\mathcal{G}(r,t)$ the asymptotic variance profile of $\tilde{\mathbf{H}}$. With statistical CSI at the transmitter, the asymptotic capacity is

$$\begin{aligned} \mathcal{C}(\beta, \mathsf{SNR}) &= \beta\,\mathbb{E}\left[\log(1+\mathsf{SNR}\,\mathbb{E}\left[\mathcal{G}(\mathsf{R},\mathsf{T})\mathcal{P}(\mathsf{T},\mathsf{SNR})\Gamma(\mathsf{R},\mathsf{SNR})|\,\mathsf{T}\right])\right] \\ &\quad +\mathbb{E}\left[\log(1+\mathbb{E}[\mathcal{G}(\mathsf{R},\mathsf{T})\mathcal{P}(\mathsf{T},\mathsf{SNR})\Upsilon(\mathsf{T},\mathsf{SNR})|\mathsf{R}])\right] \\ &\quad -\beta\,\mathbb{E}\left[\mathcal{G}(\mathsf{R},\mathsf{T})\mathcal{P}(\mathsf{T},\mathsf{SNR})\Gamma(\mathsf{R},\mathsf{SNR})\Upsilon(\mathsf{T},\mathsf{SNR})\right]\log e \end{aligned}$$

with expectation over the independent random variables R and T uniform on $[0,1]$ and with

$$\begin{aligned} \beta\,\Gamma(r,\mathsf{SNR}) &= \frac{1}{1+\mathbb{E}[\mathcal{G}(r,\mathsf{T})\mathcal{P}(\mathsf{T},\mathsf{SNR})\Upsilon(\mathsf{T},\mathsf{SNR})]} \\ \Upsilon(t,\mathsf{SNR}) &= \frac{\mathsf{SNR}}{1+\mathsf{SNR}\,\mathbb{E}\left[\mathcal{G}(\mathsf{R},t)\mathcal{P}(t,\mathsf{SNR})\Gamma(\mathsf{R},\mathsf{SNR})\right]} \end{aligned}$$

where $\mathcal{P}(t,\mathsf{SNR})$ is the asymptotic power profile of the capacity achieving power allocation at each SNR.

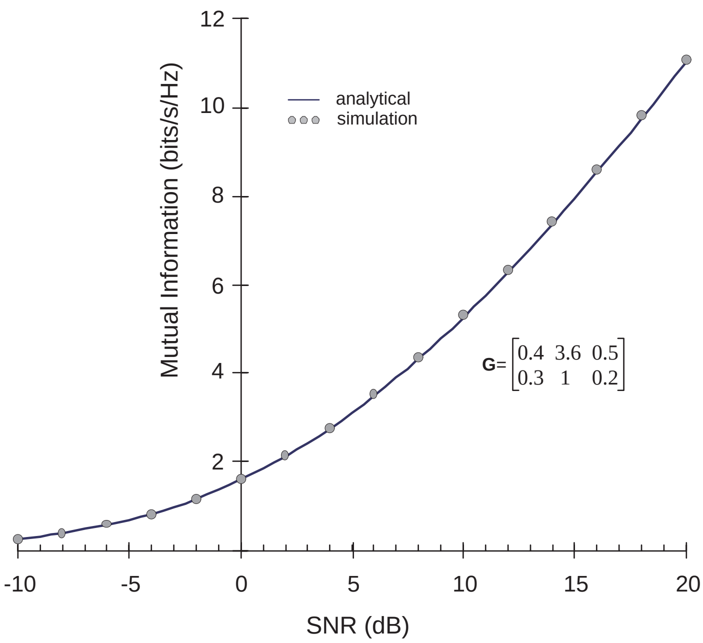

Fig. 3.4 Mutual information achieved by an isotropic input on a Rayleigh-faded channel with $n_{\mathrm{T}} = 3$ and $n_{\mathrm{R}} = 2$ for the variance matrix $\mathbf{G}$ in (3.156).

Corollary 3.3. [266] Consider a channel defined as in Theorem 3.8 but with an isotropic input. Theorem 3.8 yields the mutual information by setting $\mathcal{P}(t, \text{SNR}) = 1$.

This corollary is illustrated in Fig. 3.4 for a Rayleigh-faded channel with $n_{\mathrm{T}} = 3$ and $n_{\mathrm{R}} = 2$ where $\mathbf{H} = \mathbf{U}_{\mathrm{R}}\tilde{\mathbf{H}}\mathbf{U}_{\mathrm{T}}^{\dagger}$ with the entries of $\tilde{\mathbf{H}}$ being independent with zero-mean and variances given by

$$\mathbf{G} = \begin{bmatrix} 0.4 & 3.6 & 0.5 \\ 0.3 & 1 & 0.2 \end{bmatrix}. \tag{3.156}$$

Despite the very small numbers of antennas, there is full agreement between the analytical values (obtained by applying Theorem 3.8 with $\mathcal{P}(t, \text{SNR}) = 1$ and with the expectations replaced by arithmetic averages over the entries of $\mathbf{G}$) and the outcome of corresponding Monte-Carlo simulations.

Asymptotic characterizations of the high-SNR capacity with statistical CSI and of the mutual information achieved by an isotropic input can be obtained via Theorem 2.54.

Asymptotic spectrum results have also been used in [161] to characterize the wideband capacity of correlated multi-antenna channel using the tools of [274].

3.3.5 Polarization Diversity

A particularly interesting channel is generated if antennas with mixed polarizations are used and there is no correlation, in which case the entries of $\mathbf{H}$ are independent but not identically distributed because of the different power gain between co-polarized and differently polarized antennas. In this case, the eigenvalues of $\mathsf{r}_{\mathbf{H}}$ coincide with the variance of the entries of $\mathbf{H}$, which we can model as

$$\mathbf{H} = \mathbf{A} \circ \mathbf{H}_w \tag{3.157}$$

where $\circ$ indicates Hadamard (element-wise) multiplication, $\mathbf{H}_w$ is composed of zero-mean i.i.d. Gaussian entries with variance $\frac{1}{n_{\mathrm{T}}}$ and $\mathbf{A}$ is a deterministic matrix with nonnegative entries. Each $|\mathsf{A}_{i,j}|^2$ symbolizes the power gain between the jth transmit and ith receive antennas, determined by their relative polarizations.[14]

The asymptotic capacity with full CSI at the transmitter can be found, for sufficiently high SNR, by invoking Theorems 2.58 and 2.54.

Since the entries of $\mathbf{H}$ are independent, the input covariance that achieves capacity with statistical CSI is diagonal [261, 268]. The corresponding asymptotic capacity per antenna equals the one given in Theorem 3.8 with $\mathcal{G}(r,t)$ the asymptotic variance profile of $\mathbf{H}$. Corollary 3.3 holds similarly. Furthermore, these solutions do not require that the entries of $\mathbf{H}$ be Gaussian but only that their variances be uniformly bounded.

A common structure for $\mathbf{A}$, arising when the transmit and receive arrays have an equal number of antennas on each polarization, is that of a doubly-regular form (cf. Definition 2.10). For such channels, the capacity-achieving input is not only diagonal but isotropic and, applying Theorem 2.49, the capacity admits an explicit form.

Theorem 3.9. Consider a channel $\mathbf{H} = \mathbf{A} \circ \mathbf{H}_w$ where the entries of $\mathbf{A}$ are deterministic and nonnegative while those of $\mathbf{H}_w$ are zero-mean and independent, with variance $\frac{1}{n_{\mathrm{T}}}$ but not necessarily identically distributed. If $\mathbf{A}$ is doubly-regular (cf. Definition 2.10), the asymptotic capacity per antenna, with full CSI or with statistical CSI at the transmitter, coincides with that of the canonical channel, given in Theorem 3.5 and Eq. (3.140) respectively.

3.3.6 Progressive Scattering

Let us postulate the existence of $L-1$ clusters of scatterers each with n_ℓ scattering objects, $1 \le \ell \le L-1$, such that the signal propagates from the transmit array to the first cluster, from there to the second cluster and so on, until it is received from the $(L-1)$th cluster by the receiver

[14] If all antennas are co-polar, then every entry of $\mathbf{A}$ equals 1.

array. This practically motivated model provides a nice application of the S-transform.

The matrix $\mathbf{H}$ describing the communication link with progressive scattering be written as the product of L independent random matrices [184]

$$\mathbf{H} = \prod_{\ell=1}^{L} \mathbf{H}_\ell \tag{3.158}$$

where the $n_\ell \times n_{\ell-1}$ matrix $\mathbf{H}_\ell$ describes the subchannel between the $(\ell-1)$th and ℓth clusters. (We conventionally denote as 1st and Lth clusters the transmit and the receive arrays themselves.) If the matrices $\mathbf{H}_\ell$ are mutually independent with zero-mean i.i.d. entries having variance $\frac{1}{n_\ell}$, and defining $\beta_\ell = \frac{n_\ell}{n_L}$, the S-transform of the matrix

$$\mathbf{A}_L = \prod_{\ell=1}^{L} \mathbf{H}_\ell \left(\prod_{\ell=1}^{L} \mathbf{H}_\ell \right)^\dagger \tag{3.159}$$

$$= \mathbf{H}_L \mathbf{A}_{L-1} \mathbf{H}_L^\dagger \tag{3.160}$$

can be computed using Example 2.53 as

$$\Sigma_{\mathbf{A}_L}(x) = \frac{1}{x + \beta_{L-1}} \Sigma_{\mathbf{A}_{L-1}}(\tfrac{x}{\beta_{L-1}}) \tag{3.161}$$

which, applying Example 2.53 iteratively, yields

$$\Sigma_{\mathbf{A}_L}(x) = \prod_{\ell=1}^{L} \frac{\beta_\ell}{x + \beta_{\ell-1}} \tag{3.162}$$

from which it follows that the η-transform of $\mathbf{A}_L$ is the solution to

$$\mathsf{SNR} \, \frac{\eta_{\mathbf{A}_L}(\mathsf{SNR})}{1 - \eta_{\mathbf{A}_L}(\mathsf{SNR})} = \prod_{\ell=1}^{L} \frac{\beta_\ell}{\eta_{\mathbf{A}_L}(\mathsf{SNR}) + \beta_{\ell-1} - 1}. \tag{3.163}$$

3.3.7 Ricean Channel

Every zero-mean multi-antenna channel model analyzed thus far can be made Ricean by incorporating an additional deterministic component $\bar{\mathbf{H}}$ [62, 74, 236]. With proper weighting of the random and deterministic components so that condition (3.129) is preserved, the general model then becomes

$$\mathbf{y} = \left(\sqrt{\tfrac{1}{K+1}}\mathbf{H} + \sqrt{\tfrac{K}{K+1}}\bar{\mathbf{H}}\right)\mathbf{x} + \mathbf{n} \tag{3.164}$$

with the scalar Ricean factor K quantifying the ratio between the Frobenius norm of the deterministic (unfaded) component and the expected Frobenius norm of the random (faded) component. Considered individually, each (i, j)th channel entry has a Ricean factor given by

$$K\frac{|\bar{\mathsf{H}}_{i,j}|^2}{\mathbb{E}[|\mathsf{H}_{i,j}|^2]}.$$

Using Lemma 2.22 the next result follows straightforwardly.

Theorem 3.10. Consider a channel with a Ricean term whose rank is finite. The asymptotic capacity per antenna, $\mathcal{C}^{\mathsf{rice}}(\beta, \mathsf{SNR})$, equals the corresponding asymptotic capacity per antenna in the absence of the Ricean component, $\mathcal{C}(\beta, \mathsf{SNR})$, with a simple SNR penalty:

$$\mathcal{C}^{\mathsf{rice}}(\beta, \mathsf{SNR}) = \mathcal{C}(\beta, \frac{\mathsf{SNR}}{K+1}). \tag{3.165}$$

Note that, while the value of the capacity depends on the degree of CSI available at the transmitter, (3.165) holds regardless.

Further applications of random matrix methods to Ricean multi-antenna channels in the non-asymptotic regime, can be found in [134, 137, 3, 118, 151, 269].

3.3.8 Interference-limited Channel

Since efficient bandwidth utilization requires aggressive frequency reuse across adjacent cells and sectors, mature wireless systems tend to be, by design, limited by out-of-cell interference rather than by thermal noise. Unlike thermal noise, which is spatially and temporally white, interference is in general spatially colored. The impact of colored interference on the capacity has been studied asymptotically in [163, 181, 51], and non-asymptotically in [28, 138].

Out-of-cell interference can be incorporated into the model (1.1) by representing the noise as

$$\mathbf{n} = \sum_{\ell=1}^{L} \mathbf{H}_\ell \mathbf{x}_\ell + \mathbf{n}_{\mathsf{th}} \tag{3.166}$$

where L is the number of interferers, $\mathbf{x}_\ell$ the signal transmitted by the ℓ-th interferer, $\mathbf{H}_\ell$ the channel from such interferer and $\mathbf{n}_{\rm th}$ the underlying thermal noise. Thus, (1.1) becomes

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \sum_{\ell=1}^{L} \mathbf{H}_\ell \mathbf{x}_\ell + \mathbf{n}_{\mathsf{th}}. \tag{3.167}$$

In what follows, we consider a homogeneous system where the entries of $\mathbf{x}_\ell$, $\ell \in \{1, \ldots, L\}$, to be i.i.d. zero-mean Gaussian and the number of transmit antennas at each interferer to coincide with $n_{\rm T}$. Furthermore, the channels $\mathbf{H}$ and $\mathbf{H}_\ell$, $\ell \in \{1, \ldots, L\}$, are modeled as canonical. We define the signal-to-interference with respect to each interferer as

$$\mathsf{SIR}_{\,\ell} = \frac{\mathbb{E}[\|\mathbf{x}\|^2]}{\mathbb{E}[\|\mathbf{x}_\ell\|^2]} \tag{3.168}$$

and use SNR to specify the signal-to-thermal-noise ratio. With that, the overall SINR satisfies

$$\frac{1}{\mathsf{SINR}} = \frac{1}{\mathsf{SNR}} + \sum_{\ell=1}^{L} \frac{1}{\mathsf{SIR}_{\,\ell}} \tag{3.169}$$

and the capacity can be expressed as

$$\mathcal{C} = \frac{1}{n_{\rm R}} E\left[\log\det\left(\mathbf{I} + \mathbf{H}\mathbf{H}^\dagger \left(\sum_{\ell=1}^{L} \frac{1}{\mathsf{SIR}_{\,\ell}} \mathbf{H}_\ell \mathbf{H}_\ell^\dagger + \frac{n_{\rm T}}{\mathsf{SNR}}\mathbf{I}\right)^{-1}\right)\right] \tag{3.170}$$

with expectation over the distributions of $\mathbf{H}$ and $\mathbf{H}_\ell$, $\ell \in \{1, \ldots, L\}$. The impact of interference on the capacity essentially mirrors that of receive correlations except for the fact that the interference is subject to fading. Asymptotically, however, this becomes immaterial and hence Theorem 2.44 can be applied to obtain:

Theorem 3.11. [163, 317][15] Consider a Rayleigh-faded channel with i.i.d. zero-mean unit-variance entries exposed to L i.i.d. Gaussian interferers whose channels are similarly distributed. Let the user of interest and each interferer be equipped with n_T transmit antennas. As $n_\mathrm{T}, n_\mathrm{R} \to \infty$ with $\beta \to \frac{n_\mathrm{T}}{n_\mathrm{R}}$, the capacity converges to

$$\begin{aligned} \mathcal{C}\left(\beta, \mathsf{SNR}, \{\mathsf{SIR}_\ell\}\right) &= \beta \sum_{\ell=1}^{L} \log \left(\frac{\mathsf{SIR}_\ell + \mathsf{SNR}\,\frac{\eta_1}{\beta}}{\mathsf{SIR}_\ell + \mathsf{SNR}\,\frac{\eta_2}{\beta}} \right) + \beta \log(1 + \mathsf{SNR}\,\frac{\eta_1}{\beta}) \\ &\quad + \log \frac{\eta_2}{\eta_1} + (\eta_1 - \eta_2) \log e \end{aligned} \tag{3.171}$$

with η_1 and η_2 solutions to

$$\eta_1 + \frac{\mathsf{SNR}\,\eta_1}{\mathsf{SNR}\,\frac{\eta_1}{\beta} + 1} + \sum_{\ell=1}^{L} \frac{\mathsf{SNR}\,\eta_1}{\mathsf{SNR}\,\frac{\eta_1}{\beta} + \mathsf{SIR}_\ell} = 1 \tag{3.172}$$

$$\eta_2 + \sum_{\ell=1}^{L} \frac{\mathsf{SNR}\,\eta_2}{\mathsf{SNR}\,\frac{\eta_2}{\beta} + \mathsf{SIR}_\ell} = 1. \tag{3.173}$$

Obtaining explicit expressions requires solving for η_1 and η_2 in equations of order $L+2$ and $L+1$, respectively. Hence, the complexity of the solution is directly determined by the number of interferers. Nonetheless, solving for η_1 and η_2 becomes trivial in some limiting cases [163]:

- For growing β,

$$\lim_{\beta \to \infty} \eta_1 = \frac{1}{1 + \mathsf{SNR}\left(1 + \sum_{\ell=1}^{L} \frac{1}{\mathsf{SIR}_\ell}\right)} \tag{3.174}$$

$$\lim_{\beta \to \infty} \eta_2 = \frac{1}{1 + \mathsf{SNR} \sum_{\ell=1}^{L} \frac{1}{\mathsf{SIR}_\ell}} \tag{3.175}$$

  which are function of only the relative powers of the desired user, the interferers and the thermal noise. Plugging these into (3.171) yields an asymptotic capacity that is identical to that which would be attained if the interference was replaced with white noise. Hence, as the total number of interfering antennas grows much larger than the number of

[15] Although the analysis in [317] considers multicell DS-CDMA, the expression for the capacity maps exactly onto (3.170) except for a simple SNR scaling.

receive antennas, the progressively fine color of the interference cannot be discerned. The capacity depends only on the total interference-plus-thermal power, irrespective of how it breaks down.

- For diminishing β and finite L,

$$\lim_{\beta \to 0} \eta_1 = \lim_{\beta \to 0} \eta_2 = 1 \tag{3.176}$$

indicating that the capacity penalty due to a fixed number of interfering antennas vanishes as the number of receive antennas grows without bound. The performance becomes dictated only by the underlying thermal noise, irrespective of the existence of the interference [309, 310].

3.3.9 Outage Capacity

The ergodic capacity has operational meaning only if the distribution of $\mathbf{H}$ is revealed by the realizations encountered by each codeword. In some situations, however, $\mathbf{H}$ is held approximately constant during the transmission of a codeword, in which case a more suitable performance measure is the outage capacity, which coincides with the classical Shannon-theoretic notion of ϵ-capacity [49], namely the maximal rate for which block error probability ϵ is attainable. Under certain conditions, the outage capacity can be obtained through the probability that the transmission rate R exceeds the input-output mutual information (conditioned on the channel realization) [77, 250, 22]. Thus, given a rate R an outage occurs when the random variable

$$\mathcal{I} = \log\det(\mathbf{I} + \mathsf{SNR}\,\mathbf{H}\boldsymbol{\Phi}\mathbf{H}^\dagger)$$

whose distribution is induced by $\mathbf{H}$, falls below R. Establishing the input covariance that maximizes the rate supported at some chosen outage level is a problem not easily tackled analytically. (Some results on the eigenvectors of $\boldsymbol{\Phi}$ can be found in [229].) Hereafter $\boldsymbol{\Phi}$ is allowed to be an arbitrary deterministic matrix except where otherwise noted.

The distribution of $\mathcal{I}$ can be obtained via its moment-generating function

$$M(\zeta) = \mathbb{E}\left[e^{\zeta \mathcal{I}}\right] \tag{3.177}$$

which, for the canonical channel with $\mathbf{\Phi} = \mathbf{I}$, is given by (2.18) as derived in [38, 299]. The corresponding function for one-sided correlation, in the case of square channels, is for $\zeta \leq 0$

$$M(\zeta) = {}_2F_0(\zeta \log \tfrac{1}{e}\,,\ m \,|\, -\gamma\mathbf{\Theta}) \tag{3.178}$$

where ${}_2F_0(\cdot, \cdot\,|\,\cdot)$ is given by (2.21) with $\mathbf{\Theta} = \mathbf{\Theta}_{\mathrm{R}}$ if the correlation takes place at the receiver whereas $\mathbf{\Theta} = \mathbf{\Theta}_{\mathrm{T}}^{1/2}\mathbf{\Phi}\mathbf{\Theta}_{\mathrm{T}}^{1/2}$ if it takes place at the transmitter. With both transmit and receive correlations, $M(\cdot)$ is given by Theorem 2.16 with $\mathbf{\Sigma} = \mathbf{\Theta}_{\mathrm{R}}$ and $\mathbf{\Upsilon} = \mathbf{\Theta}_{\mathrm{T}}^{1/2}\mathbf{\Phi}\mathbf{\Theta}_{\mathrm{T}}^{1/2}$.

For uncorrelated Ricean channels with $\mathbf{\Phi} = \mathbf{I}$, $M(\cdot)$ is provided in [134] in terms of the integral of hypergeometric functions.

For $n_{\mathrm{R}} = 1$, the distribution of $\mathcal{I}$ is found directly, bypassing the moment-generating function, for correlated Rayleigh-faded channels in [180, 132] and for uncorrelated Ricean channels in [180, 233].[16]

An interesting property of the distribution of $\mathcal{I}$ is the fact that, for many of the multi-antenna channels of interest, it can be approximated as Gaussian as the number of antennas grows. A number of authors have explored this property using two distinct approaches in the engineering literature:

(1) The mean and variance of $\mathcal{I}$ are obtained through the moment generating function (for fixed number of antennas). A Gaussian distribution with such mean and variance is then compared, through Monte Carlo simulation, to the empirical distribution of $\mathcal{I}$. This approach is followed in [235, 299, 26] for the canonical channel, in [234] for channels with one-sided correlation, and in [235] for uncorrelated Ricean channels. Although, in every case, the match is excellent, no proof of asymptotic Gaussianity is provided. Only for SNR $\to \infty$ with $\mathbf{\Phi} = \mathbf{I}$ and with $\mathbf{H}$ being a real Gaussian matrix with i.i.d. entries has it been shown that $\mathcal{I} - \mathbb{E}[\mathcal{I}]$ converges to a Gaussian random variable [87].

[16] The input covariance is constrained to be $\mathbf{\Phi} = \mathbf{I}$ in [233], which also gives the corresponding distribution of $\mathcal{I}$ for $\min(n_{\mathrm{T}}, n_{\mathrm{R}}) = 2$ and arbitrary $\max(n_{\mathrm{T}}, n_{\mathrm{R}})$ although in the form of an involved integral expression.

(2) The random variable

$$\Delta_{n_{\mathrm{R}}} = \mathcal{I}(\mathsf{SNR}) - n_{\mathrm{R}}\mathcal{V}_{\mathbf{H\Phi H}^\dagger}(\mathsf{SNR}) \qquad (3.179)$$

is either shown or conjectured to converge to a zero-mean Gaussian random variable as $n_{\mathrm{R}} \to \infty$. For Rayleigh-faded channels with one-sided correlation (at either transmitter or receiver), the asymptotic Gaussianity of $\Delta_{n_{\mathrm{R}}}$ follows from Theorem 2.77.[17] The convergence rate to the Gaussian distribution is analyzed in [15]. With both transmit and receive correlations, the asymptotic Gaussianity of $\Delta_{n_{\mathrm{R}}}$ is conjectured in [216, 181] by observing the behavior of the second- and third-order moments obtained via the replica method.

The appeal of the Gaussian behavior, of course, is that its characterization entails finding only the mean and variance of $\mathcal{I}$. In how these are found, and in some others respects, the differences between both approaches are subtle but important:

- When approximating $\mathcal{I}$ as a Gaussian random variable for finite n_{T} and n_{R}, the first approach uses exact expressions for its mean and variance. These expressions, which can be obtained from the moment-generating function, tend to be rather involved and are often not in closed form. The second approach, on the other hand, relies on functionals of the asymptotic spectrum. Although exact only in the limit, these functionals are tight even for small values of n_{T} and n_{R} and tend to have simpler and more insightful forms.
- If the moment convergence theorem does not apply to the asymptotic spectrum, as in the case of Ricean channels where the rank of $\mathbb{E}[\mathbf{H}]$ is $o(n_{\mathrm{R}})$, then the second approach results in a bias that stems from the fact that $\mathbb{E}[\mathbf{H}]$ is not reflected in the asymptotic spectrum (cf. Lemma 2.22).

Denoting $\Delta = \lim_{n_{\mathrm{R}}\to\infty} \Delta_{n_{\mathrm{R}}}$, $E[\Delta^2]$ can be found by applying [15,

[17] The more restrictive case of a canonical channel at either low or high SNR is analyzed in [113].

(1.17)]. For the canonical channel, this yields (cf. Theorem 2.76)

$$\begin{aligned}\mathbb{E}[\Delta^2] &= -\log\left(1 - \frac{(1-\eta_{\mathbf{HH}^\dagger}(\gamma))^2}{\beta}\right) \\ &= -\log\left(1 - \frac{1}{\beta}\left(\frac{\mathcal{F}(\gamma,\beta)}{4\gamma}\right)^2\right). \qquad (3.180)\end{aligned}$$

With Rayleigh fading and correlation at the transmitter, in turn,

$$\mathbb{E}[\Delta^2] = -\log\left(1 - \frac{(1-\eta_{\mathbf{HTH}^\dagger}(\gamma))^2}{\beta}\right) \qquad (3.181)$$

where $\mathbf{T} = \mathbf{\Theta}_{\mathrm{T}}^{1/2}\mathbf{\Phi}\mathbf{\Theta}_{\mathrm{T}}^{1/2}$ with $\mathbf{\Phi}$ the capacity-achieving power allocation.

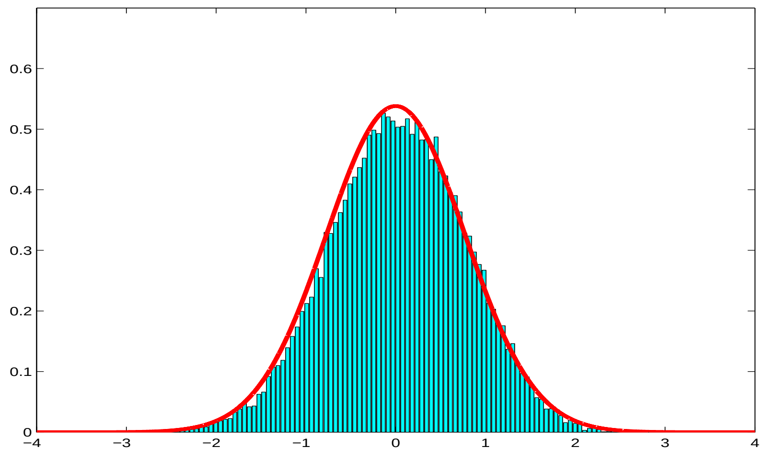

Fig. 3.5 Histogram of $\Delta_{n_{\mathrm{R}}}$ for a Rayleigh-faded channel with $n_{\mathrm{T}} = 5$ and $n_{\mathrm{R}} = 10$. The transmit antennas are correlated as per (3.182) while the receive antennas are uncorrelated. Solid line indicates the corresponding limiting Gaussian distribution.

Figure 3.5 compares the limiting Gaussian distribution of Δ with a histogram of $\Delta_{n_{\mathrm{R}}}$ for $n_{\mathrm{T}} = 5$ and $n_{\mathrm{R}} = 10$ with a transmit correlation matrix $\mathbf{\Theta}_{\mathrm{T}}$ such that

$$(\mathbf{\Theta}_{\mathrm{T}})_{i,j} = e^{-0.8(i-j)^2}. \qquad (3.182)$$

For channels with both transmit and receive correlation, the characteristic function found through the replica method yields to the expression of $\mathbb{E}[\Delta^2]$ given in [181].

3.3.10 Space-Time Coding

Besides the characterizations of the capacity for the various channels described throughout this section, random matrix theory (and specifically free probability) has also been used to obtain design criteria for space-time codes [25]. In [25], the behavior of space-time codes is characterized asymptotically in the number of antennas. Specifically, the behavior of pairwise error probabilities is determined with three types of receivers: maximum-likelihood (ML), decorrelator and linear MMSE, It is shown that with ML or linear receivers the asymptotic performance of space-time codes is determined by the Euclidean distances between codewords. This holds for intermediate signal-to-noise ratios even when the number of antennas is small. Simulations show how asymptotic results are quite accurate in the non-asymptotic regime. This has the interesting implication that even for few antennas, off-the-shelf codes may outperform special-purpose space-time codes.

3.4 Other Applications

In addition to the foregoing recent applications of random matrix theory in characterizing the fundamental limits of wireless communication channels, several other applications of the results in Section 2 can be found in the information theory, communications and signal processing literature:

- Speed of convergence of iterative algorithms for multiuser detection [312].
- Direction of arrival estimation in sensor arrays [228].
- Learning and neural networks [50].
- Capacity of ad hoc networks [157].
- Data mining and multivariate time series modelling and analysis [155, 139].
- Principal components analysis [119].
- Maximal entropy methods [17, 292].
- Information theory and free probability [288, 289, 248, 292, 293].

4

Appendices

4.1 Proof of Theorem 2.39

In this section we give a multiuser-detection argument for the proof of Theorem 2.39 in the special case where $\mathbf{T}$ is diagonal. To use the standard notation in multiuser detection [271], we replace $\mathbf{H}$ with $\mathbf{S}$ and let $\mathbf{T} = \mathbf{A}\mathbf{A}^\dagger$.

An important non-asymptotic relationship between the eigenvalues $\lambda_1, \ldots, \lambda_N$ of the matrix $\mathbf{STS}^\dagger$ and the signal-to-interference ratios achieved by the MMSE detector $\mathsf{SIR}_1, \ldots \mathsf{SIR}_K$ is [256]

$$\sum_{i=1}^{N} \frac{\lambda_i}{\lambda_i + \sigma^2} = \sum_{k=1}^{K} \frac{\mathsf{SIR}_k}{\mathsf{SIR}_k + 1} \tag{4.1}$$

where σ^2 is the variance of the noise components in (3.1). To show (4.1), we can write its right-hand side as

$$\begin{aligned} \sum_{i=1}^{N} \frac{\lambda_i}{\lambda_i + \sigma^2} &= \mathrm{tr}\left(\left(\sigma^2 \mathbf{I} + \mathbf{STS}^\dagger\right)^{-1} \mathbf{STS}^\dagger\right) \\ &= \mathrm{tr}\left(\left(\sigma^2 \mathbf{I} + \mathbf{STS}^\dagger\right)^{-1} \sum_{k=1}^{K} \mathsf{T}_k \mathbf{s}_k \mathbf{s}_k^\dagger\right) \end{aligned}$$

which can be further elaborated into

$$\begin{aligned}\sum_{i=1}^{N}\frac{\lambda_i}{\lambda_i+\sigma^2} &= \sum_{k=1}^{K}\mathsf{T}_k\mathbf{s}_k^\dagger\left(\sigma^2\mathbf{I}+\mathbf{STS}^\dagger\right)^{-1}\mathbf{s}_k\\ &= \sum_{k=1}^{K}\frac{\mathsf{SIR}_k}{\mathsf{SIR}_k+1} \qquad (4.2)\end{aligned}$$

where (4.2) follows from [271, (6.40)].

Denote for brevity

$$\eta = \eta_{\mathbf{STS}^\dagger}\left(\frac{1}{\sigma^2}\right). \qquad (4.3)$$

From the the fact that the η-transform of $\mathbf{STS}^\dagger$ evaluated at σ^{-2} is the multiuser efficiency achieved by each of the users asymptotically,

$$\mathsf{SIR}_k = \frac{\mathsf{T}_k}{\sigma^2}\eta \qquad (4.4)$$

we obtain

$$\begin{aligned}\lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\frac{\mathsf{SIR}_k}{\mathsf{SIR}_k+1} &= 1-\lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\frac{1}{\frac{\mathsf{T}_k}{\sigma^2}\eta+1}\\ &= 1-\eta_{\mathbf{T}}\left(\frac{\eta}{\sigma^2}\right) \qquad (4.5)\end{aligned}$$

almost surely, by the law of large numbers and the definition of η-transform. Also by definition of η-transform,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\frac{\lambda_i}{\lambda_i+\sigma^2} = 1-\eta \qquad (4.6)$$

Equations (4.1), (4.5) and (4.6) lead to the sought-after relationship

$$\beta\left(1-\eta_{\mathbf{T}}\left(\frac{\eta}{\sigma^2}\right)\right) = 1-\eta. \qquad (4.7)$$

4.2 Proof of Theorem 2.42

The first step in the proof is to convert the problem to one where $\mathbf{T}$ is replaced by a diagonal matrix $\mathbf{D}_{\mathbf{T}}$ of the same size and with the same limiting empirical eigenvalue distribution. To that end, denote the diagonal matrix

$$\mathbf{Q} = \mathbf{I} + \gamma \mathbf{W}_0 \tag{4.8}$$

and note that

$$\begin{aligned} \det\left(\mathbf{I} + \gamma(\mathbf{W}_0 + \mathbf{H}\mathbf{T}\mathbf{H}^\dagger)\right) &= \det(\mathbf{T}) \det(\mathbf{Q}) \\ &\quad \cdot \det\left(\mathbf{T}^{-1} + \gamma(\mathbf{H}\mathbf{Q}^{-1}\mathbf{H}^\dagger)\right). \end{aligned} \tag{4.9}$$

Using Theorem 2.38 with $\mathbf{W}_0$ and $\mathbf{T}$ therein equal to $\mathbf{T}^{-1}$ and $\mathbf{Q}^{-1}$ (this is a valid choice since $\mathbf{Q}^{-1}$ is diagonal), it follows that the asymptotic spectrum of $\mathbf{T}^{-1} + \gamma(\mathbf{H}\mathbf{Q}^{-1}\mathbf{H}^\dagger)$ depends on $\mathbf{T}^{-1}$ only through its asymptotic spectrum. Therefore, when we take $\frac{1}{N}\log$ of both sides of (4.9) we are free to replace $\mathbf{T}$ by $\mathbf{D}_\mathbf{T}$. Thus,

$$\mathcal{V}_\mathbf{W}(\gamma) = \lim_{N\to\infty} \frac{1}{N} \log\det\left(\mathbf{I} + \gamma(\mathbf{W}_0 + \mathbf{H}\mathbf{T}\mathbf{H}^\dagger)\right) \tag{4.10}$$

$$= \; = \lim_{N\to\infty} \frac{1}{N} \log\det\left(\mathbf{I} + \gamma(\mathbf{W}_0 + \mathbf{H}\mathbf{D}_\mathbf{T}\mathbf{H}^\dagger)\right) \tag{4.11}$$

$$= \mathcal{V}_{\mathbf{W}_0 + \mathbf{H}\mathbf{D}_\mathbf{T}\mathbf{H}^\dagger}(\gamma) \tag{4.12}$$

Since the Shannon transforms are identical, so are the η-transforms. Using Theorem 2.38 and (2.48), it follows that the η-transform of $\mathbf{W}_0 + \mathbf{H}\mathbf{D}_\mathbf{T}\mathbf{H}^\dagger$ and consequently of $\mathbf{W}$ is

$$\eta\gamma = \mathbb{E}\left[\frac{1}{\mathsf{W}_0 + \frac{1}{\gamma} + \beta\,\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\eta\gamma}\right]}\right] \tag{4.13}$$

where T and W_0 are independent random variables whose distributions equal the asymptotic empirical eigenvalue distributions of $\mathbf{T}$ and $\mathbf{W}_0$, respectively. From (4.13),

$$\eta\gamma = \varphi\eta_0(\varphi) \tag{4.14}$$

with

$$\varphi = \frac{\gamma}{1 + \beta\,\gamma\,\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\eta\gamma}\right]} \tag{4.15}$$

$$= \frac{\gamma}{1 + \frac{\beta}{\eta}\left(1 - \eta_\mathbf{T}(\eta\gamma)\right)} \tag{4.16}$$

from which

$$\begin{aligned} \eta\varphi + \varphi\beta\,(1 - \eta_{\mathbf{T}}(\eta\gamma)) &= \gamma\eta & (4.17) \\ &= \varphi\eta_0(\varphi). & (4.18) \end{aligned}$$

4.3 Proof of Theorem 2.44

From Theorem 2.43 and from Remark 2.3.1 it follows straightforwardly that the η-transform of $\mathbf{HH}^\dagger$ with $\mathbf{H} = \mathbf{CSA}$ equals the η-transform of a matrix $\tilde{\mathbf{H}}$ whose entries are independent zero-mean random variables with variance

$$\mathbb{E}[|\tilde{\mathsf{H}}_{i,j}|^2] = \frac{\mathsf{P}_{i,j}}{N}$$

and whose variance profile is

$$v(x, y) = v_{\mathrm{X}}(x)\, v_{\mathrm{Y}}(y)$$

with $v_{\mathrm{X}}(x)$ and $v_{\mathrm{Y}}(y)$ such that the distributions of $v_{\mathrm{X}}(\mathsf{X})$ and $v_{\mathrm{Y}}(\mathsf{Y})$ (with X and Y independent random variables uniform on $[0, 1]$) equal the asymptotic empirical distributions of $\mathbf{D}$ and $\mathbf{T}$ respectively. In turn, (2.137) can be proved as special case of (2.158) when the function $v(x, y)$ can be factored. In this case, the expressions of $\Gamma_{\mathbf{HH}^\dagger}(x, \gamma)$ and $\Upsilon_{\mathbf{HH}^\dagger}(y, \gamma)$ given by Equations (2.154) and (2.155) in Theorem 2.50 become

$$\begin{aligned} \Gamma_{\mathbf{HH}^\dagger}(x, \gamma) &= \frac{1}{1 + \beta\,\gamma\, v_{\mathrm{X}}(x)\mathbb{E}[v_{\mathrm{Y}}(\mathsf{Y})\, \Upsilon_{\mathbf{HH}^\dagger}(\mathsf{Y}, \gamma)]} \\ &= \frac{1}{1 + \beta\,\gamma\, v_{\mathrm{X}}(x)\, \tilde{\Upsilon}_{\mathbf{HH}^\dagger}(\gamma)} \end{aligned} \qquad (4.19)$$

where we have denoted

$$\tilde{\Upsilon}_{\mathbf{HH}^\dagger}(\gamma) = \mathbb{E}[v_{\mathrm{Y}}(\mathsf{Y})\Upsilon_{\mathbf{HH}^\dagger}(\mathsf{Y}, \gamma)].$$

For convenience, in the following we drop the subindices from $\Gamma_{\mathbf{HH}^\dagger}$, $\Upsilon_{\mathbf{HH}^\dagger}$, $\tilde{\Upsilon}_{\mathbf{HH}^\dagger}$. Let us further denote

$$\tilde{\Gamma}(\gamma) = \mathbb{E}[v_{\mathrm{X}}(\mathsf{X})\Gamma(\mathsf{X}, \gamma)].$$

Using (2.154) we obtain

$$
\begin{aligned}
\tilde{\Gamma}(\gamma) &= \mathbb{E}\left[\frac{v_{\mathrm{X}}(\mathsf{X})}{1+\beta\,\gamma\,v_{\mathrm{X}}(\mathsf{X})\,\mathbb{E}[v_{\mathrm{Y}}(\mathsf{Y})\,\Upsilon(\mathsf{Y},\gamma)]}\right] \\
&= \mathbb{E}\left[\frac{v_{\mathrm{X}}(\mathsf{X})}{1+\beta\,\gamma\,v_{\mathrm{X}}(\mathsf{X})\,\tilde{\Upsilon}(\gamma)}\right] \\
&= \mathbb{E}\left[\frac{\mathsf{\Lambda}_{\mathbf{D}}}{1+\beta\gamma\,\mathsf{\Lambda}_{\mathbf{D}}\tilde{\Upsilon}(\gamma)}\right] \\
&= \frac{1}{\beta\,\gamma\,\tilde{\Upsilon}(\gamma)}\Big(1-\eta_{\mathbf{D}}(\beta\,\gamma\,\tilde{\Upsilon}(\gamma))\Big) \qquad (4.20)
\end{aligned}
$$

where we have indicated by $\Lambda_{\mathbf{D}}$ a nonnegative random variable whose distribution is given by the asymptotic spectrum of $\mathbf{D}$. Likewise, using the definition of $\Upsilon(y,\gamma)$ in (2.155) we obtain

$$
\begin{aligned}
\tilde{\Upsilon}(\gamma) &= \mathbb{E}\left[\frac{v_{\mathrm{Y}}(\mathsf{Y})}{1+\gamma\,v_{\mathrm{Y}}(\mathsf{Y})\tilde{\Gamma}(\gamma)}\right] \\
&= \mathbb{E}\left[\frac{\mathsf{\Lambda}_{\mathbf{T}}}{1+\gamma\,\mathsf{\Lambda}_{\mathbf{T}}\,\tilde{\Gamma}(\gamma)}\right] \\
&= \frac{1}{\gamma\,\tilde{\Gamma}(\gamma)}\Big(1-\eta_{\mathbf{T}}(\gamma\,\tilde{\Gamma}(\gamma))\Big) \qquad (4.21)
\end{aligned}
$$

where we have denoted by $\Lambda_{\mathbf{T}}$ a nonnegative random variable whose distribution is given by the asymptotic spectrum of the matrix $\mathbf{T}$. Notice also that

$$
\begin{aligned}
\log\left(1+\gamma\mathbb{E}\left([v(\mathsf{X},\mathsf{Y})\Gamma(\mathsf{X},\gamma)|\mathsf{Y}]\right)\right. &= \log\left(1+\gamma\,v_{\mathrm{Y}}(\mathsf{Y})\mathbb{E}[v_{\mathrm{X}}(\mathsf{X})\Gamma(\mathsf{X},\gamma)]\right) \\
&= \log\Big(1+\gamma\,v_{\mathrm{Y}}(\mathsf{Y})\tilde{\Gamma}(\gamma)\Big) \qquad (4.22)
\end{aligned}
$$

and thus

$$
\begin{aligned}
\mathbb{E}\left[\log\left(1+\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Gamma(\mathsf{X},\gamma)|\mathsf{Y}]\right)\right] &= \mathbb{E}\Big[\log(1+\gamma\,\mathsf{\Lambda}_{\mathbf{T}}\,\tilde{\Gamma}(\gamma))\Big] \\
&= \mathcal{V}_{\mathbf{T}}(\gamma\tilde{\Gamma}(\gamma)). \qquad (4.23)
\end{aligned}
$$

Likewise,

$$
\begin{aligned}
\mathbb{E}\left[\log(1+\gamma\,\beta\,\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)|\mathsf{X}])\right] &= \mathbb{E}\Big[\log(1+\gamma\,\beta\,\mathsf{\Lambda}_{\mathbf{D}}\,\tilde{\Upsilon}(\gamma))\Big] \\
&= \mathcal{V}_{\mathbf{D}}(\gamma\,\beta\,\tilde{\Upsilon}(\gamma)). \qquad (4.24)
\end{aligned}
$$

Moreover,

$$\begin{aligned}\gamma\,\beta\,\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Upsilon(\mathsf{X},\gamma)\Upsilon(\mathsf{Y},\gamma)] &= \gamma\,\beta\,\mathbb{E}\left[v_{\mathrm{X}}(\mathsf{X})\,v_{\mathrm{Y}}(\mathsf{Y})\Gamma(\mathsf{X},\gamma)\Upsilon(\mathsf{Y},\gamma)\right] \\ &= \gamma\,\beta\,\tilde{\Gamma}(\gamma)\,\tilde{\Upsilon}(\gamma). \qquad (4.25)\end{aligned}$$

Defining

$$\gamma_{\mathrm{t}} = \gamma\,\tilde{\Gamma}(\gamma) \qquad\qquad \gamma_d = \gamma\,\tilde{\Upsilon}(\gamma), \qquad (4.26)$$

plugging (4.25), (4.24), (4.23) into (2.158), and using (4.26), (4.21) and (4.20), the expression for $\mathcal{V}_{\mathbf{HH}^\dagger}$ in Theorem 2.44 is found.

4.4 Proof of Theorem 2.49

From (2.153) it follows that

$$\eta_{\mathbf{HH}^\dagger}(\gamma) = \mathbb{E}[\Gamma_{\mathbf{HH}^\dagger}(\mathsf{X},\gamma)]$$

with $\Gamma_{\mathbf{HH}^\dagger}(\cdot,\cdot)$ satisfying the equation

$$\Gamma_{\mathbf{HH}^\dagger}(x,\gamma) = \frac{1}{1+\beta\gamma\,\mathbb{E}\left[\frac{v(x,\mathsf{Y})}{1+\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Gamma_{\mathbf{HH}^\dagger}(\mathsf{X},\gamma)|\mathsf{Y}]}\right]} \qquad (4.27)$$

where X and Y are independent random variables uniformly distributed on $[0,1]$. Again for convenience, in the following we abbreviate $\Gamma_{\mathbf{HH}^\dagger}(\cdot,\cdot)$ and $\Upsilon_{\mathbf{HH}^\dagger}(\cdot,\cdot)$ as $\Gamma(\cdot,\cdot)$ and $\Upsilon(\cdot,\cdot)$.

From the definition of doubly-regular matrix, we have that $\mathbb{E}\left[1\{v(\mathsf{X},t)\le x\}\right]$ does not depend on t and thus $\mathbb{E}[v(\mathsf{X},t)\Gamma(\mathsf{X},\gamma)]$ does not depend on t. At the same time, from the definition of doubly-regular $\mathbb{E}\left[1\{v(r,\mathsf{Y})\le x\}\right]$, does not depend on r and thus the expectation

$$E\left[\frac{v(r,\mathsf{Y})}{1+\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Gamma(\mathsf{X},\gamma)|\mathsf{Y}]}\right]$$

does not depend on r. Consequently, $\Gamma(r,\gamma)=\Gamma(\gamma)$ for all r. Thus, we can rewrite the fixed-point equation in (4.27) as

$$\begin{aligned}\Gamma(\gamma) &= \frac{1}{1+\beta\gamma\mathbb{E}\left[\frac{v(x,\mathsf{Y})}{1+\gamma\,\Gamma(\gamma)\mathbb{E}[v(\mathsf{X},\mathsf{Y})|\,\mathsf{Y}]}\right]}\\ &= \frac{1}{1+\beta\,\gamma\mathbb{E}\left[\frac{v(x,\mathsf{Y})}{1+\gamma\,\Gamma(\gamma)\mu}\right]}\\ &= \frac{1}{1+\beta\,\gamma\frac{\mathbb{E}\,[v(x,\mathsf{Y})]}{1+\gamma\,\Gamma(\gamma)\mu}}\end{aligned}$$

resulting in

$$\Gamma(\gamma) = \frac{1}{1+\beta\gamma\mu\frac{1}{1+\gamma\Gamma(\gamma)\mu}}$$

with $\mu = \mathbb{E}\,[v(\mathsf{X},y)] = \mathbb{E}\,[v(x,\mathsf{Y})] = 1$ since we have assumed $\mathbf{P}$ to be a standard double-regular matrix. The above equation can be solved to yield the η-transform of $\mathbf{HH}^\dagger$ as

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

Using (2.48) and the inverse Stieltjes formula, the claim is proved.

4.5 Proof of Theorem 2.53

From (2.59), the Shannon transform of $\mathbf{HH}^\dagger$ is given by

$$\mathcal{V}_{\mathbf{HH}^\dagger}(\gamma) = \int \log(1+\gamma\lambda)d\mathsf{F}_{\mathbf{HH}^\dagger}(\lambda)$$

where $\mathsf{F}_{\mathbf{HH}^\dagger}(\cdot)$ represents the limiting distribution to which the empirical eigenvalue distribution of $\mathbf{HH}^\dagger$ converges almost surely. The derivative with respect to γ is

$$\begin{aligned}
\dot{\mathcal{V}}_{\mathbf{HH}^\dagger}(\gamma) &= \log e \int \frac{\lambda}{1+\gamma\lambda} d\mathsf{F}_{\mathbf{HH}^\dagger}(\lambda) \\
&= \int \frac{\log e}{\gamma}\left(1 - \frac{1}{1+\gamma\lambda}\right) d\mathsf{F}_{\mathbf{HH}^\dagger}(\lambda) \\
&= \frac{\log e}{\gamma}\left(1 - \int \frac{1}{1+\gamma\lambda} d\mathsf{F}_{\mathbf{HH}^\dagger}(\lambda)\right) \\
&= \frac{\log e}{\gamma}\left(1 - \mathbb{E}\left[\Gamma_{\mathbf{HH}^\dagger}(\mathsf{X},\gamma)\right]\right) \qquad (4.28)
\end{aligned}$$

where, in the last equality, we have invoked Theorem 2.50 and where $\Gamma_{\mathbf{HH}^\dagger}(\cdot,\cdot)$ satisfies the equations given in (2.154) and (2.155), namely

$$\Gamma_{\mathbf{HH}^\dagger}(x,\gamma) = \frac{1}{1+\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\Upsilon_{\mathbf{HH}^\dagger}(\mathsf{Y},\gamma)]} \qquad (4.29)$$

$$\Upsilon_{\mathbf{HH}^\dagger}(y,\gamma) = \frac{1}{1+\gamma\mathbb{E}[v(\mathsf{X},y)\Gamma_{\mathbf{HH}^\dagger}(\mathsf{X},\gamma)]} \qquad (4.30)$$

with X and Y independent random variables uniform on $[0,1]$. For brevity, we drop the subindices from $\Gamma_{\mathbf{HH}^\dagger}$ and $\Upsilon_{\mathbf{HH}^\dagger}$. Using (4.29) we can write

$$\frac{1-\Gamma(x,\gamma)}{\gamma} = \frac{\beta\mathbb{E}[v(x,\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)]}{1+\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)]},$$

which, after adding and subtracting to the right-hand side

$$\frac{\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\dot{\Upsilon}(\mathsf{Y},\gamma)]}{1+\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)]},$$

becomes

$$\begin{aligned}
\frac{1-\Gamma(x,\gamma)}{\gamma} &= \frac{\beta\mathbb{E}[v(x,\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)] + \beta\gamma\mathbb{E}[v(x,\mathsf{Y})\dot{\Upsilon}(\mathsf{Y},\gamma)]}{1+\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)]} \\
&\quad - \frac{\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\dot{\Upsilon}(\mathsf{Y},\gamma)]}{1+\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)]} \\
&= \frac{d}{d\gamma}\ln(1+\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)]) \\
&\quad - \frac{\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\dot{\Upsilon}(\mathsf{Y},\gamma)]}{1+\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)]} \qquad (4.31)
\end{aligned}$$

where $\dot{\Upsilon}(\cdot,\gamma)=\frac{d}{d\gamma}\Upsilon(\cdot,\gamma)$. From (4.28) and (4.29) it follows that

$$\begin{aligned}\dot{\mathcal{V}}_{\mathbf{HH}^\dagger}(\gamma) &= \mathbb{E}\left[\frac{d}{d\gamma}\log(1+\beta\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)])\right] \\ &\quad -\beta\,\gamma\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\,\Gamma(\mathsf{X},\gamma)\,\dot{\Upsilon}(\mathsf{Y},\gamma)\right]\log e. \qquad (4.32)\end{aligned}$$

Notice that

$$\begin{aligned}-\gamma\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\,\Gamma(\mathsf{X},\gamma)\,\dot{\Upsilon}(\mathsf{Y},\gamma)\right] &= -\frac{d}{d\gamma}\left(\gamma\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\,\Gamma(\mathsf{X},\gamma)\,\Upsilon(\mathsf{Y},\gamma)\right]\right) \\ &\quad +\mathbb{E}\left[\gamma\,v(\mathsf{X},\mathsf{Y})\dot{\Gamma}(\mathsf{X},\gamma)\Upsilon(\mathsf{Y},\gamma)\right] \\ &\quad +\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\Gamma(\mathsf{X},\gamma)\Upsilon(\mathsf{Y},\gamma)\right] \qquad (4.33)\end{aligned}$$

with $\dot{\Gamma}(\cdot,\gamma)=\frac{d}{d\gamma}\Gamma(\cdot,\gamma)$. From (4.29),

$$\begin{aligned}\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\left(\gamma\,\dot{\Gamma}(\mathsf{X},\gamma)+\Gamma(\mathsf{X},\gamma)\right)\Upsilon(\mathsf{Y},\gamma)\right] &= \mathbb{E}\left[\frac{v(\mathsf{X},\mathsf{Y})\left(\gamma\,\dot{\Gamma}(\mathsf{X},\gamma)+\Gamma(\mathsf{X},\gamma)\right)}{1+\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Gamma(\mathsf{X},\gamma)|\mathsf{Y}]}\right] \\ &= \mathbb{E}\left[\frac{\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\left(\gamma\,\dot{\Gamma}(\mathsf{X},\gamma)+\Gamma(\mathsf{X},\gamma)\right)|\mathsf{Y}\right]}{1+\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Gamma(\mathsf{X},\gamma)|\mathsf{Y}]}\right]\end{aligned}$$

from which integrating (4.32) with respect to γ and using (4.33) we have that

$$\begin{aligned}\mathcal{V}_{\mathbf{HH}^\dagger}(\gamma) &= \mathbb{E}\left[\log(1+\beta\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)])\right] \\ &\quad -\beta\,\gamma\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\,\Gamma(\mathsf{X},\gamma)\,\Upsilon(\mathsf{Y},\gamma)\right]\log e \\ &\quad +\beta\,\mathbb{E}\left[\log(1+\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Gamma(\mathsf{X},\gamma)|\mathsf{Y}])\right]+\kappa \\ &\qquad (4.34)\end{aligned}$$

with κ the integration constant which must be set to $\kappa = 0$ so that $\mathcal{V}_{\mathbf{HH}^\dagger}(0) = 0$.

Acknowledgements

The authors gratefully acknowledge helpful suggestions by Prof. Z. D. Bai, National University of Singapore, Dr. A. Lozano, Bell Laboratories, Prof. R. Speicher, Queen's University at Kingston, Prof. D. Voiculescu, University of California at Berkeley, and the anonymous reviewers.

References

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions.* Dover Publications, Inc., New York, 1964.

[2] G. Alfano, A. Lozano, A. M. Tulino, and S. Verdú, "Capacity of MIMO channels with one-sided correlation," in *Proc. IEEE Int. Symp. on Spread Spectrum Tech. and Applications (ISSSTA'04)*, Aug. 2004.

[3] G. Alfano, A. Lozano, A. M. Tulino, and S. Verdú, "Mutual information and eigenvalue distribution of MIMO Ricean channels," in *Proc. IEEE Int. Symp. Information Theory & Applications (ISITA'04)*, (Parma, Italy), Oct. 2004.

[4] J. B. Andersen, "Array gain and capacity for known random channels with multiple element arrays at both ends," *IEEE J. on Selected Areas in Communications*, vol. 18, pp. 2172–2178, Nov. 2000.

[5] T. W. Anderson, "The non-central Wishart distribution and certain problems of multivariate statistics," *Annals of Math. Statistics*, vol. 17, no. 4, pp. 409–431, 1946.

[6] M. Andréief, "Note sur une relation entre les intégrals définies des produits des fonctions," *Mém. de la Soc. Sci. de Bordeaux*, vol. 2, pp. 1–14, 1883.

[7] O. Axelsson, *Iterative Solution Methods.* Cambridge University Press, 1994.

[8] 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.

[9] Z. D. Bai, "The circle law," *Annals of Probability*, vol. 25, pp. 494–529, 1997.

[10] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices," *Statistica Sinica*, vol. 9, no. 3, pp. 611–661, 1999.

[11] Z. D. Bai, B. Miao, and J. Tsay, "Convergence rates of spectral distributions of large Wigner matrices," *Int. Mathematical Journal*, vol. 1, no. 1, pp. 65–90, 2002.

[12] Z. D. Bai, B. Miao, and J. Yao, "Convergence rates of spectral distributions of large sample covariance matrices," *SIAM J. of Matrix Analysis and Applications*, vol. 25, no. 1, pp. 105–127, 2003.

[13] 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.

[14] Z. D. Bai and J. W. Silverstein, "Exact separation of eigenvalues of large dimensional sample covariance matrices," *Annals of Probability*, vol. 27, no. 3, pp. 1536–1555, 1999.

[15] Z. D. Bai and J. W. Silverstein, "CLT of linear spectral statistics of large dimensional sample covariance matrices," *Annals of Probability*, vol. 32, no. 1A, pp. 553–605, 2004.

[16] Z. D. Bai and Y. Q. Yin, "Convergence to the semicircle law," *Annals of Probability*, vol. 16, no. 2, pp. 863–875, 1988.

[17] R. Balian, "Random matrices in information theory," *Il Nuovo Cimento*, pp. 183–193, Sep. 1968.

[18] M. S. Barlett, "On the theory of statistical regression," in *Proc. R. Soc. Edinb. 53*, pp. 260–283, 1933.

[19] S. A. Bergmann and H. W. Arnold, "Polarization diversity in portable communications environment," *IEE Electronics Letters*, vol. 22, pp. 609–610, May 1986.

[20] P. Biane, "Minimal factorizations of cycle and central multiplicative functions on the infinite symmetric group," *J. Combinatorial Theory*, vol. A 76, no. 2, pp. 197–212, 1996.

[21] P. Biane, "Free probability for probabilists," *Report MSRI (http://arXiv.org/abs/math/9809193)*, Sep. 1998.

[22] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. on Information Theory*, vol. 44, pp. 2619–2692, Oct. 1998.

[23] E. Biglieri and G. Taricco, "Transmission and reception with multiple antennas: Theoretical foundations," *submitted to Foundations and Trends in Communications and Information Theory*, 2004.

[24] E. Biglieri and G. Taricco, "Large-system analyses of multiple-antenna system capacities," *Journal of Communications and Networks*, vol. 5, No. 2, p. 5764, June 2003.

[25] E. Biglieri, G. Taricco, and A. M. Tulino, "Performance of space-time codes for a large number of antennas," *IEEE Trans. on Information Theory*, vol. 48, pp. 1794–1803, July 2002.

[26] E. Biglieri, A. M. Tulino, and G. Taricco, "How far away is infinity? Using asymptotic analyses in multiple antenna systems," *Proc. IEEE Int. Symp. on Spread Spectrum Techn. and Applications (ISSSTA'02)*, vol. 1, pp. 1–6, Sep. 2002.

[27] G. Birkhoff and S. Gulati, "Isotropic distributions of test matrices," *J. Appl. Math. Physics*, vol. 30, pp. 148–157, 1979.

[28] R. S. Blum, "MIMO capacity with interference," *IEEE J. on Selected Areas in Communications*, vol. 21, pp. 793–801, June 2003.

[29] P. Bougerol and J. Lacroix, *Random Products of Matrices with Applications to Schrödinger Operators.* Basel, Switzerland: Birkhuser, 1985.

[30] J. Boutros and G. Caire, "Iterative multiuser joint decoding: unified framework and asymptotic analysis," *IEEE Trans. on Information Theory*, vol. 48, pp. 1772–1793, July 2002.

[31] L. H. Brandenburg and A. D. Wyner, "Capacity of the Gaussian channel with memory: The multivariate case," *Bell System Technical Journal*, vol. 53, pp. 745–778, May-June 1974.

[32] B. Bronk, "Exponential ensembles for random matrices," *J. of Math. Physics*, vol. 6, pp. 228–237, 1965.

[33] R. Buehrer and B. Woerner, "The asymptotic multiuser efficiency of m-stage interference cancellation receivers," in *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Communications (PIMRC'97)*, (Helsinki, Finland), pp. 570–574, Sep. 1997.

[34] K. A. Byerly and R. A. Roberts, "Output power based partial adaptive array design," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, (Pacific Grove, CA), pp. 576–580, Oct. 1989.

[35] G. Caire and S. Shamai, "On achievable rates in a multi-antenna broadcast downlink," in *Proc. Allerton Conf. on Communications, Control and Computing*, pp. 1188–1193, Oct. 2000.

[36] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. on Information Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

[37] J. M. Chaufray, W. Hachem, and P. Loubaton, "Asymptotic analysis of optimum and sub-optimum CDMA MMSE receivers," *Proc. IEEE Int. Symp. on Information Theory (ISIT'02)*, p. 189, July 2002.

[38] M. Chiani, "Evaluating the capacity distribution of MIMO Rayleigh fading channels," *Proc. IEEE Int. Symp. on Advances in Wireless Communications*, pp. 3–4, Sep. 23-24 2002, Victoria, Canada.

[39] M. Chiani, M. Z. Win, and A. Zanella, "On the capacity of spatially correlated MIMO Rayleigh-fading channels," *IEEE Trans. on Information Theory*, vol. 49, pp. 2363–2371, Oct. 2003.

[40] 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, pp. 337–339, Nov. 2000.

[41] D. Chizhik, J. Ling, P. Wolniansky, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple-input multiple-output measurements and modelling in Manhattan," *IEEE J. on Selected Areas in Communications*, vol. 21, pp. 321–331, Apr. 2003.

[42] T.-S. Chu and L. J. Greenstein, "A semiempirical representation of antenna diversity gain at cellular and PCS base stations," *IEEE Trans. on Communications*, vol. 45, pp. 644–656, June 1997.

[43] C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, “Capacity scaling in dual-antenna-array wireless systems,” *IEEE Trans. on Information Theory*, vol. 48, pp. 637–650, March 2002.

[44] J. Cohen, H. Kesten, and M. Newman, “Oseledec’s multiplicative ergodic theorem: a proof,” in *Random matrices and their applications*, (J. Cohen, H. Kesten, and M. Newman, eds.), Providence, RI: American Mathematical Society, 1986.

[45] L. Cottatellucci and R. R. Müller, “Asymptotic design and analysis of multistage detectors with unequal powers,” in *Proc. IEEE Information Theory Workshop (ITW’02)*, (Bangalore, India), Oct. 2002.

[46] L. Cottatellucci and R. R. Müller, “Asymptotic design and analysis of multistage detectors and multistage channel estimators for multipath fading channels,” in *Proc. IEEE Int. Symp. on Information Theory (ISIT’02)*, (Yokohama, Japan), June 2003.

[47] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* John Wiley and Sons, Inc., 1991.

[48] D. Cox, “Universal digital portable radio communications,” *Proc. IEEE*, vol. 75, No. 4, pp. 436–477, Apr. 1987.

[49] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic, 1981.

[50] Y. L. Cun, I. Kanter, and S. A. Solla, “Eigenvalues of covariance matrices: Application to neural-network learning,” *Physical Review Letters*, vol. 66, pp. 2396–2399, 1991.

[51] H. Dai, A. F. Molisch, and H. V. Poor, “Downlink capacity of interference-limited MIMO systems with joint detection,” *IEEE Trans. on Wireless Communications*, vol. 3, pp. 442– 453, Mar. 2004.

[52] A. B. de Monvel and A. Khorunzhy, “Asymptotic distribution of smoothed eigenvalue density. I. Gaussian random matrices & II. Wigner random matrices,” *Random Operators and Stochastic Equations*, vol. 7, pp. 1–22 & 149–168, 1999.

[53] A. B. de Monvel and A. Khorunzhy, “On the norm and eigenvalue distribution of large random matrices,” *Annals of Probability*, vol. 27, no. 2, pp. 913–944, 1999.

[54] A. B. de Monvel and A. Khorunzhy, “On universality of smoothed eigenvalue density of large random matrices,” *J. of Physics A: Mathematical and General*, vol. 32, pp. 413–417, 1999.

[55] A. B. de Monvel, A. Khorunzhy, and V. Vasilchuk, “Limiting eigenvalue distribution of random matrices with correlated entries,” *Markov Processes and Related Fields*, vol. 2, no. 2, pp. 607–636, 1996.

[56] M. Debbah, W. Hachem, P. Loubaton, and M. de Courville, “MMSE analysis of certain large isometric random precoded systems,” *IEEE Trans. on Information Theory*, vol. 49, pp. 1293–1311, May 2003.

[57] H. Dette, “Strong approximation of the eigenvalues of large dimensional Wishart matrices by roots of generalized Laguerre polynomials,” *J. of Approximation Theory*, vol. 118, no. 2, pp. 290–304, 2002.

[58] J. A. Díaz-García and J. R. Gutierrez, "Proof of the conjectures of H. Uhlig on the singular multivariate beta and the Jacobian of a certain matrix transformation," *Annals of Statistics*, vol. 25, pp. 2018–2023, 1997.

[59] J. A. Díaz-García, J. R. Gutierrez, and K. V. Mardia, "Wishart and pseudo-Wishart distributions and some applications to shape theory," *J. of Multivariate Analysis*, vol. 63, pp. 73–87, 1997.

[60] S. N. Diggavi, N. Al-Dhahir, A. Stamoulis, and A. R. Calderbank, "Great expectations: The value of spatial diversity in wireless networks," *Proc. IEEE*, vol. 92, pp. 219–270, Feb. 2004.

[61] M. Dohler and H. Aghvami, "A closed form expression of MIMO capacity over ergodic narrowband channels," *preprint*, 2003.

[62] P. Driessen and G. J. Foschini, "On the capacity formula for multiple-input multiple-output channels: A geometric interpretation," *IEEE Trans. Communications*, vol. 47, pp. 173–176, Feb. 1999.

[63] K. J. Dykema, "On certain free product factors via an extended matrix model," *J. of Funct. Analysis*, vol. 112, pp. 31–60, 1993.

[64] K. J. Dykema, "Interpolated free group factors," *Pacific J. Math.*, vol. 163, no. 1, pp. 123–135, 1994.

[65] F. Dyson, "Statistical theory of the energy levels of complex systems," *J. of Math. Physics*, vol. 3, pp. 140–175, 1962.

[66] F. Dyson, "A class of matrix ensembles," *J. of Math. Physics*, vol. 13, p. 90, 1972.

[67] A. Edelman, *Eigenvalues and condition number of random matrices.* PhD thesis, Dept. Mathematics, MIT, Cambridge, MA, 1989.

[68] A. Edelman, "The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law," *J. Multivariate Analysis*, vol. 60, pp. 203–232, 1997.

[69] A. Edelman, E. Kostlan, and M. Shub, "How many eigenvalues of a random matrix are real?," *J. Amer. Math. Soc.*, vol. 7, pp. 247–267, 1994.

[70] Y. C. Eldar and A. M. Chan, "On the asymptotic performance of the decorrelator," *IEEE Trans. on Information Theory*, vol. 49, pp. 2309–2313, Sep. 2003.

[71] H. Elders-Boll, A. Busboom, and H. Schotten, "Implementation of linear multiuser detectors for asynchronous CDMA systems by linear interference cancellation algorithms," in *Proc. IEEE Vehicular Technology. Conf. (VTC'98)*, (Ottawa, Canada), pp. 1849–1853, May 1998.

[72] H. Elders-Boll, H. Schotten, and A. Busboom, "Efficient implementation of linear multiuser detectors for asynchronous CDMA systems by linear interference cancellation," *Euro. Trans. Telecommunications*, vol. 9, pp. 427–438, Sep./Oct. 1998.

[73] J. Evans and D. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *IEEE Trans. on Information Theory*, vol. 46, pp. 2059–2078, Sep. 2000.

[74] 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, pp. 85–87, Mar. 2001.

[75] R. A. Fisher, "The sampling distribution of some statistics obtained from non-linear equations," *Annals of Eugenics*, vol. 9, pp. 238–249, 1939.
[76] G. Foschini and M. Gans, "On limits of wireless communications in fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, No. 6, pp. 315–335, Mar. 1998.
[77] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, pp. 41–59, 1996.
[78] D. A. S. Fraser, *The Structure of Inference.* John Wiley and Sons, New York, 1968.
[79] H. Furstenberg and H. Kesten, "Products of random matrices," *Annals of Math. Statistics*, pp. 457–469, 1960.
[80] H. Gao and P. J. Smith, "A determinant representation for the distribution of quadratic forms in complex normal vectors," *J. Multivariate Analysis*, vol. 73, pp. 155–165, May 2000.
[81] M. Gaudin, "Sur la loi limite de l'espacement des valeurs propres d'une matrice aleatoire," *Nuclear Physics*, vol. 25, pp. 447–455, 1961.
[82] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO Space-Time coded wireless systems," *J. on Selected Areas in Communications*, vol. 21, pp. 281–302, Apr. 2003.
[83] R. Gharavi and V. Anantharam, "An upper bound for the largest Lyapunov exponent of a Markovian product of nonnegative matrices," *submitted to Theoretical Computer Science*, 2004.
[84] V. Ghazi-Moghadam, L. Nelson, and M. Kaveh, "Parallel interference cancellation for CDMA systems," in *Proc. Allerton Conf. on Communication, Control and Computing*, (Monticello, IL), pp. 216–224, Oct. 1995.
[85] V. L. Girko, "Circular law," *Theory Prob. Appl.*, vol. 29, pp. 694–706, 1984.
[86] V. L. Girko, *Theory of Random Determinants.* Dordrecht: Kluwer Academic Publishers, 1990.
[87] V. L. Girko, "A refinement of the central limit theorem for random determinants," *Theory Prob. Appl.*, vol. 42, no. 1, pp. 121–129, 1997.
[88] V. L. Girko, "Convergence rate of the expected spectral functions of symmetric random matrices equal to $o(n^{-1/2})$," *Random Operator and Stochastic Equations*, vol. 6, pp. 359–408, 1998.
[89] M. A. Girshick, "On the sampling theory of roots of determinantal equations," *Annals of Math. Statistics*, vol. 10, pp. 203–204, 1939.
[90] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. on Selected Areas on Communications*, vol. 21, pp. 684–702, June 2003.
[91] J. S. Goldstein and I. S. Reed, "A new method of Wiener filtering and its application to interference mitigation for communications," in *Proc. IEEE MILCOM*, (Monterey, CA), pp. 1087–1091, Nov. 1997.
[92] J. S. Goldstein and I. S. Reed, "Reduced rank adaptive filtering," *IEEE Trans. Signal Processing*, vol. 45, pp. 492–496, Feb. 1997.

[93] J. S. Goldstein, I. S. Reed, and L. L. Scharf, “A multistage representation of the Wiener filter based on orthogonal projections,” *IEEE Trans. on Information Theory*, vol. 44, pp. 2943–2959, Nov. 1998.

[94] C. R. Goodall and K. V. Mardia, “Multivariate aspects of shape theory,” *Annals of Statistics*, vol. 21, pp. 848–866, 1993.

[95] F. Gotze and A. N. Tikhomirov, “Rate of convergence to the semicircular law for the Gaussian unitary ensemble,” *Theory Prob. Appl.*, vol. 47, no. 2, pp. 323–330, 2003.

[96] P. Graczyk, G. Letac, and H. Massam, “The complex Wishart distribution and the symmetric group,” *Annals of Statistics*, vol. 31, no. 1, pp. 287–309, 2003.

[97] S. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products.* New York: Academic, 1965.

[98] A. Grant, “Rayleigh fading multi-antenna channels,” *EURASIP J. on Applied Signal Processing*, vol. 3, pp. 316–329, 2002.

[99] A. Grant and C. Schlegel, “Convergence of linear interference cancellation multiuser receivers,” *IEEE Trans. on Communications*, vol. 10, pp. 1824–1834, Oct. 2001.

[100] A. J. Grant and P. D. Alexander, “Random sequences multisets for synchronous code-division multiple-access channels,” *IEEE Trans. on Information Theory*, vol. 44, pp. 2832–2836, Nov. 1998.

[101] U. Grenander and J. W. Silverstein, “Spectral analysis of networks with random topologies,” *SIAM J. of Applied Mathematics*, vol. 32, pp. 449–519, 1977.

[102] A. Guionnet and O. Zeitouni, “Concentration of the spectral measure for large matrices,” *Electronic Communications in Probability*, vol. 5, pp. 119–136, 2000.

[103] D. Guo and S. Verdú, “Multiuser detection and statistical mechanics,” in *Communications Information and Network Security*, (V. Bhargava, H. V. Poor, V. Tarokh, and S. Yoon, eds.), pp. 229–277, Ch. 13, Kluwer Academic Publishers, 2002.

[104] A. K. Gupta and D. K. Nagar, “Matrix variate distributions,” in *Monographs and Surveys in pure and applied mathematics*, Boca Raton, FL: Chapman and Hall/CRC, 2000.

[105] W. Hachem, “Low complexity polynomial receivers for downlink CDMA,” in *Proc. Asilomar Conf. on Systems, Signals and Computers*, (Pacific Grove, CA), Nov. 2002.

[106] G. Hackenbroich and H. A. Weidenmueller, “Universality of random-matrix results for non-Gaussian ensembles,” *Physics Review Letters*, vol. 74, pp. 4118–4121, 1995.

[107] P. R. Halmos, *Measure Theory.* Van Nostrand, Princeton, NJ, 1950.

[108] S. V. Hanly and D. N. C. Tse, “Resource pooling and effective bandwidths in CDMA networks with multiuser receivers and spatial diversity,” *IEEE Trans. on Information Theory*, vol. 47, pp. 1328–1351, May 2001.

[109] S. Hara and R. Prasad, “Overview of multicarrier CDMA,” *IEEE Communications Magazine*, vol. 35, pp. 126–133, Dec. 1997.

[110] F. Hiai and D. Petz, "Asymptotic freeness almost everywhere for random matrices," *Acta Sci. Math. Szeged*, vol. 66, pp. 801–826, 2000.

[111] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy.* American Mathematical Society, 2000.

[112] B. Hochwald and S. Vishwanath, "Space-time multiple access: Linear growth in the sum-rate," in *Proc. Allerton Conf. on Communication, Control and Computing*, (Monticello, IL), Oct. 2002.

[113] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multi-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. on Information Theory*, submitted May 2002.

[114] T. Holliday, A. J. Goldsmith, and P. Glynn, "On entropy and Lyapunov exponents for finite state channels," *submitted to IEEE Trans. on Information Theory*, 2004.

[115] M. L. Honig, "Adaptive linear interference suppression for packet DS-CDMA," *Euro. Trans. Telecommunications*, vol. 9, pp. 173–182, Mar./Apr. 1998.

[116] M. L. Honig and W. Xiao, "Performance of reduced-rank linear interference suppression for DS-CDMA," *IEEE Trans. on Information Theory*, vol. 47, pp. 1928–1946, July 2001.

[117] R. Horn and C. Johnson, *Matrix Analysis.* Cambridge University Press, 1985.

[118] D. Hösli and A. Lapidoth, "The capacity of a MIMO Ricean channel is monotonic in the singular values of the mean," *5th Int. ITG Conf. on Source and Channel Coding*, Jan. 2004.

[119] D. C. Hoyle and M. Rattray, "Principal component analysis eigenvalue spectra from data with symmetry breaking structure," *Physical Review E*, vol. 69, 026124, 2004.

[120] P. L. Hsu, "On the distribution of roots of certain determinantal equations," *Annals of Eugenics*, vol. 9, pp. 250–258, 1939.

[121] L. K. Hua, *Harmonic analysis of functions of several complex variables in the classical domains.* Providence, RI: American Mathematical Society, 1963.

[122] P. Jacquet, G. Seroussi, and W. Szpankowski, "On the entropy of a hidden Markov process," in *Proc. Data Compression Conference*, Mar. 23–25 2004.

[123] S. A. Jafar, S. Vishwanath, and A. J. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," *Proc. IEEE Int. Conf. on Communications (ICC'01)*, vol. 7, pp. 2266–2270, 2001.

[124] K. Jamal and E. Dahlman, "Multi-stage interference cancellation for DS-CDMA," in *Proc. IEEE Vehicular Technology Conf. (VTC'96)*, (Atlanta, GA), pp. 671–675, Apr. 1996.

[125] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *Annals of Math. Statistics*, vol. 35, pp. 475–501, 1964.

[126] R. Janaswamy, "Analytical expressions for the ergodic capacities of certain MIMO systems by the Mellin transform," *Proc. IEEE Global Telecomm. Conf. (GLOBECOM'03)*, vol. 1, pp. 287–291, Dec. 2003.

[127] S. K. Jayaweera and H. V. Poor, "Capacity of multiple-antenna systems with both receiver and transmitter channel state information," *IEEE Trans. on Information Theory*, vol. 49, pp. 2697–2709, Oct. 2003.

[128] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," *IEEE Trans. on Information Theory*, vol. 50, pp. 768–783, May 2004.
[129] K. Johansson, "Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices," *Comm. Math. Phys.*, vol. 215, pp. 683–705, 2001.
[130] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[131] D. Jonsson, "Some limit theorems for the eigenvalues of a sample covariance matrix," *J. Multivariate Analysis*, vol. 12, pp. 1–38, 1982.
[132] E. Jorswieck and H. Boche, "On transmit diversity with imperfect channel state information," *Proc. IEEE Conf. Acoustics, Speech, and Signal Processing (ICASSP'02)*, vol. 3, pp. 2181–2184, May 2002.
[133] V. Jungnickel, T. Haustein, E. Jorswieck, and C. von Helmolt, "On linear preprocessing in multi-antenna systems," in *Proc. IEEE Global Telecomm. Conf. (GLOBECOM'02)*, pp. 1012–1016, Dec. 2002.
[134] M. Kang and M. S. Alouini, "On the capacity of MIMO Rician channels," in *Proc. 40th Annual Allerton Conference on Communication, Control, and Computing*, (Monticello, IL), Oct. 2002.
[135] M. Kang and M. S. Alouini, "Impact of correlation on the capacity of MIMO channels," in *Proc. IEEE Int. Conf. in Communications (ICC'04)*, pp. 2623–2627, May 2003.
[136] M. Kang and M. S. Alouini, "Largest eigenvalue of complex Wishart matrices and performance analysis of MIMO MRC systems," *IEEE J. on Selected Areas in Communications*, vol. 21, pp. 418–426, Apr. 2003.
[137] M. Kang, L. Yang, and M. S. Alouini, "Capacity of MIMO Rician channels with multiple correlated Rayleigh co-channel interferers," in *Proc. 2003 IEEE Globecom*, (San Francisco, CA), Dec. 2003.
[138] M. Kang, L. Yang, and M. S. Alouini, "Performance analysis of MIMO systems in presence of co-channel interference and additive Gaussian noise," in *Proc. 37th Annual Conf. on Information Sciences and Systems (CISS'2003)*, (Baltimore, MD), Mar. 2003.
[139] H. Kargupta, K. Sivakumar, and S. Ghosh, "Dependency detection in MobiMine and random matrices," *Proc. 6th European Conference on Principles and Practice of Knowledge Discovery in Databases*, pp. 250–262, 2002.
[140] C. G. Khatri, "Distribution of the largest or the smallest characteristic root under null hyperthesis concerning complex multivariate normal populations," *Annals of Math. Statistics*, vol. 35, pp. 1807–1810, Dec. 1964.
[141] C. G. Khatri, "Some results on the non-central multivariate beta distribution and moments of trace of two matrices," *Annals of Math. Statistics*, vol. 36, no. 5, pp. 1511–1520, 1965.
[142] C. G. Khatri, "On certain distribution problems based on positive definite quadratic functions in normal vectors," *Annals of Math. Statistics*, vol. 37, pp. 467–479, Apr. 1966.

[143] C. G. Khatri, "Non-central distributions of i-th largest characteristic roots of three matrices concerning complex multivariate normal populations," *Annals of the Inst. Statist. Math.*, vol. 21, pp. 23–32, 1969.

[144] C. G. Khatri, "On the moments of traces of two matrices in three situations for complex multivariate normal population," *Sankhya: Indian J. of Statistics*, vol. 32, series A, pp. 65–80, 1970.

[145] A. Khorunzhy, "On spectral norm of large band random matrices," *preprint*, Apr. 2004.

[146] A. M. Khorunzhy, B. A. Khoruzhenko, L. A. Pastur, and M. V. Shcherbina, "The large-n limit in statistical mechanics and the spectral theory of disordered systems," *Phase Transitions*, vol. 15, pp. 73–239, 1992.

[147] A. M. Khorunzhy and L. A. Pastur, "Random eigenvalue distributions," *Adv. Soviet Math.*, vol. 19, pp. 97–127, 1994.

[148] M. Kiessling and J. Speidel, "Exact ergodic capacity of MIMO channels in correlated Rayleigh fading environments," in *Proc. Int. Zurich Seminar on Communications (IZS)*, (Zurich, Switzerland), Feb. 2004.

[149] M. Kiessling and J. Speidel, "Mutual information of MIMO channels in correlated Rayleigh fading environments - a general solution," in *Proc. of IEEE Int. Conf. in Communications. (ICC'04)*, (Paris, France), June 2004.

[150] M. Kiessling and J. Speidel, "Unifying analysis of ergodic MIMO capacity in correlated Rayleigh fading environments," in *Fifth European Wireless Conference Mobile and Wireless Systems beyond 3G*, (Barcelona, Catalonia, Spain), Feb. 24-27 2004.

[151] Y.-H. Kim and A. Lapidoth, "On the log-determinant of non-central Wishart matrices," *Proc. 2003 IEEE Int. Symp. on Information Theory*, pp. 54–54, Jul 2003.

[152] Kiran and D. Tse, "Effective interference and effective bandwidth of linear multiuser receivers in asynchronous systems," *IEEE Trans. on Information Theory*, vol. 46, pp. 1426–1447, July 2000.

[153] T. Kollo and H. Neudecker, "Asymptotics of eigenvalues and unit-length eigenvectors of sample variance and correlation matrices," *J. Multivariate Analysis*, vol. 47, pp. 283–300, 1993.

[154] G. Kreweras, "Sur les partitions non-croisees d'un cycle," *Discrete Math.*, vol. 1, pp. 333–350, 1972.

[155] L. Laloux, P. Cizeau, M. Potters, and J. P. Bouchaud, "Random matrix theory and financial correlations," *Int. J. of Theoretical and Applied Finance*, vol. 3, No. 3, pp. 391–397, 2000.

[156] W. C. Y. Lee and Y. S. Yeh, "Polarization diversity system for mobile radio," *IEEE Trans. on Communications*, vol. 20, pp. 912–923, Oct 1972.

[157] O. Lévêque, E. Telatar, and D. Tse, "Upper bounds on the capacity of ad-hoc wireless networks," in *Proc. 2003 Winter School on Coding and Information Theory*, (Monte Veritá, Switzerland), Feb. 24-27, 2003.

[158] L. Li, A. M. Tulino, and S. Verdú, "Asymptotic eigenvalue moments for linear multiuser detection," *Communications in Information and Systems*, vol. 1, pp. 273–304, Sep. 2001.

[159] L. Li, A. M. Tulino, and S. Verdú, "Design of reduced-rank MMSE multiuser detectors using random matrix methods," *IEEE Trans. on Information Theory*, vol. 50, June 2004.

[160] L. Li, A. M. Tulino, and S. Verdú, "Spectral efficiency of multicarrier CDMA," *submitted to IEEE Trans. on Information Theory*, 2004.

[161] K. Liu, V. Raghavan, and A. M. Sayeed, "Capacity and spectral efficiency in wideband correlated MIMO channels," *Proc. 2003 IEEE Int. Symp. on Information Theory (ISIT 2003)*, July, 2003.

[162] P. Loubaton and W. Hachem, "Asymptotic analysis of reduced rank Wiener filter," *Proc. IEEE Information Theory Workshop (ITW'03)*, pp. 328–332, Paris, France, 2003.

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

[164] A. Lozano, A. M. Tulino, and S. Verdú, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. on Information Theory*, vol. 49, pp. 2527–2544, Oct. 2003.

[165] 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.

[166] A. Lozano, A. M. Tulino, and S. Verdú, "High-SNR power offset in multi-antenna communication," *Bell Labs Technical Memorandum*, June 2004.

[167] U. Madhow and M. Honig, "On the average near-far resistance for MMSE detection of direct sequence CDMA signals with random spreading," *IEEE Trans. on Information Theory*, vol. 45, pp. 2039–2045, Sep. 1999.

[168] R. K. Mallik, "The pseudo-Wishart distribution and its application to MIMO systems," *IEEE Trans. on Information Theory*, vol. 49, pp. 2761–2769, Oct. 2003.

[169] A. Mantravadi, V. V. Veeravalli, and H. Viswanathan, "Spectral efficiency of MIMO multiaccess systems with single-user decoding," *IEEE J. on Selected Areas in Communications*, vol. 21, pp. 382–394, Apr. 2003.

[170] 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.

[171] T. Marzetta and B. Hochwald, "Capacity of mobile multiple-antenna communication link in a Rayleigh flat-fading environment," *IEEE Trans. on Information Theory*, vol. 45, pp. 139–157, Jan. 1999.

[172] M. L. Mehta, "On the statistical properties of the level-spacings in nuclear spectra," *Nuclear Physics*, vol. 18, p. 395, 1960.

[173] M. L. Mehta, *Random Matrices and the Statistical Theory of Energy Levels.* New York, Academic Press, 1967.

[174] M. L. Mehta, "Power series of level spacing functions of random matrix ensembles," *Z. Phys. B*, vol. 86, pp. 258–290, 1992.

[175] M. L. Mehta and F. Dyson, "Statistical theory of the energy levels of complex systems," *J. Math. Physics*, vol. 4, pp. 713–719, 1963.

[176] M. L. Mehta and M. Gaudin, "On the density of the eigenvalues of a random matrix," *Nuclear Physics*, vol. 18, pp. 420–427, 1960.

[177] X. Mestre, *Space processing and channel estimation: performance analysis and asymptotic results.* PhD thesis, Dept. de Teoria del Senyal i Comunicacions, Universitat Politècnica de Catalunya, Barcelona, Catalonia, Spain, 2002.

[178] X. Mestre, J. R. Fonollosa, and A. Pages-Zamora, "Capacity of MIMO channels: asymptotic evaluation under correlated fading," *IEEE J. on Selected Areas in Communications*, vol. 21, pp. 829– 838, June 2003.

[179] S. Moshavi, E. G. Kanterakis, and D. L. Schilling, "Multistage linear receivers for DS-CDMA systems," *Int. J. of Wireless Information Networks*, vol. 39, no. 1, pp. 1–17, 1996.

[180] A. L. Moustakas and S. H. Simon, "Optimizing multiple-input single-output (MISO) communication systems with general Gaussian channels: nontrivial covariance and nonzero mean," *IEEE Trans. on Information Theory*, vol. 49, pp. 2770–2780, Oct. 2003.

[181] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: a (not so) large N analysis," *IEEE Trans. on Information Theory*, vol. 49, pp. 2545–2561, Oct. 2003.

[182] R. J. Muirhead, *Aspects of multivariate statistical theory.* New York, Wiley, 1982.

[183] R. R. Müller, *Power and Bandwidth Efficiency of Multiuser Systems with Random Spreading.* PhD thesis, Universtät Erlangen-Nürnberg, Erlangen, Germany, Nov. 1998.

[184] R. R. Müller, "On the asymptotic eigenvalue distribution of concatenated vector-valued fading channels," *IEEE Trans. on Information Theory*, vol. 48, pp. 2086–2091, July 2002.

[185] R. R. Müller, "Multiuser receivers for randomly spread signals: Fundamental limits with and without decision-feedback," *IEEE Trans. on Information Theory*, vol. 47, no. 1, pp. 268–283, Jan. 2001.

[186] R. R. Müller and W. Gerstacker, "On the capacity loss due to separation of detection and decoding in large CDMA systems," in *IEEE Information Theory Workshop (ITW)*, p. 222, Oct. 2002.

[187] R. R. Müller and S. Verdú, "Design and analysis of low-complexity interference mitigation on vector channels," *IEEE J. on Selected Areas on Communications*, vol. 19, pp. 1429–1441, Aug. 2001.

[188] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. on Information Theory*, vol. 39, pp. 1293–1302, July 1993.

[189] A. Nica, *R-transforms in free probability.* Paris, France: Henri Poincare Institute, 1999.

[190] A. Nica and R. Speicher, "On the multiplication of free n-tuples of non-commutative random variables," *American J. Math.*, vol. 118, no. 4, pp. 799–837, 1996.

[191] B. Niederhauser, "Norms of certain random matrices with dependent entries," *Random Operators and Stochastic Equations*, vol. 11, no. 1, pp. 83–101, 2003.

[192] A. Y. Orlov, "New solvable matrix integrals," *Acta Sci. Math*, vol. 63, pp. 383–395, 1997.

[193] V. I. Oseledec, "A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems," *Trans. Moscow Math. Soc.*, vol. 19, pp. 197–231, 1968.

[194] H. Ozcelik, M. Herdin, W. Weichselberger, G. Wallace, and E. Bonek, "Deficiencies of the Kronecker MIMO channel model," *IEE Electronics Letters*, vol. 39, pp. 209–210, Aug. 2003.

[195] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: A game theoretic approach," *IEEE Trans. on Information Theory*, vol. 49, pp. 1707–1727, July 2003.

[196] L. A. Pastur, "On the spectrum of random matrices," *Teoret. Mat. Fiz. (English translation: Theoretical and Mathematical Physics)*, vol. 10, pp. 67–74, 1972.

[197] L. A. Pastur, "Spectra of random self-adjoint operators," *Uspekhi Mat. Nauk. (Russian Math. Surveys)*, vol. 28, pp. 1–67, 1973.

[198] L. A. Pastur, "Eigenvalue distribution of random matrices: some recent results," *Annals Inst. Henri Poincaré*, vol. 64, pp. 325–337, 1996.

[199] L. A. Pastur, "Random matrices as paradigm," *Mathematical Physics 2000*, pp. 216–265, 2000.

[200] L. A. Pastur and M. Shcherbina, "Universality of the local eigenvalue statistics for a class of unitary invariant ensembles," *J. Stat. Physics*, vol. 86, pp. 109–147, 1997.

[201] M. Peacock, I. Collings, and M. L. Honig, "Asymptotic spectral efficiency of multi-user multi-signature CDMA in frequency-selective channels," *preprint*, 2004.

[202] M. Peacock, I. Collings, and M. L. Honig, "Asymptotic SINR analysis of single-user MC-CDMA in Rayleigh fading," *Proc. IEEE Int. Symp. on Spread Spectrum Systems and Applications (ISSSTA'02)*, pp. 338–342, Sep. 2002.

[203] 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.

[204] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes.* San Francisco, CA: Holden-Day, 1964.

[205] G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communications," *IEEE Trans. on Communications*, vol. 46, pp. 357–366, Mar. 1998.

[206] P. Rapajic and D. Popescu, "Information capacity of a random signature multiple-input multiple-output channel," *IEEE Trans. on Communications*, vol. 48, pp. 1245–1248, Aug. 2000.

[207] L. Rasmussen, A. Johansson, and T. Lim, "One-shot filtering equivalence for linear successive interference cancellation in CDMA," in *Proc. IEEE Vehicular. Technology. Conf. (VTC'97)*, (Phoenix, AZ), pp. 2163–2167, May 1997.

[208] T. Ratnarajah, R. Vaillancourt, and M. Alvo, "Complex random matrices and applications," *Math. Rep. of the Acad. of Sci. of the Royal Soc. of Canada*, vol. 25, pp. 114–120, Dec. 2003.

[209] T. Ratnarajah, R. Vaillancourt, and M. Alvo, "Complex random matrices and Rayleigh channel capacity," *Communications in Information and Systems*, pp. 119–138, Oct. 2003.

[210] S. N. Roy, "p-statistics or some generalizations in the analysis of variance appropriate to multivariate problems," *Sankhya: Indian J. of Statistics*, vol. 4, pp. 381–396, 1939.

[211] O. Ryan, "On the limit distributions of random matrices with independent or free entries," *Communications in Mathematical Physics*, vol. 193, pp. 595–626, 1998.

[212] U. Sacoglu and A. Scaglione, "Asymptotic capacity of space-time coding for arbitrary fading: a closed form expression using Girko's law," in *Proc. Intl. Conf. on Acoust. Speech and Signal Proc., ICASSP(2001)*, (Salt Lake City, UT), May 7-12 2001.

[213] A. Sayeed, "Deconstructing multi-antenna channels," *IEEE Trans. on Signal Processing*, vol. 50, pp. 2563–2579, Oct. 2002.

[214] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. New York, NY: Addison-Wesley, 1990.

[215] L. L. Scharf, "The SVD and reduced rank signal processing," *Signal Processing*, vol. 24, pp. 113–133, Nov. 1991.

[216] A. M. Sengupta and P. P. Mitra, "Capacity of multivariate channels with multiplicative noise: I. Random matrix techniques and large-N expansions for full transfer matrices," *LANL arXiv:physics*, Oct. 2000.

[217] S. Shamai and S. Verdú, "The effect of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. on Information Theory*, vol. 47, pp. 1302–1327, May 2001.

[218] J. Shen, "On the singular values of Gaussian random matrices," *Linear Algebra and its Applications*, vol. 326, no. 1-3, pp. 1–14, 2001.

[219] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering and keyhole," *IEEE Trans. on Information Theory*, vol. 49, pp. 2636–2647, Oct. 2003.

[220] 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, pp. 502–511, Mar. 2000.

[221] D. Shlyankhtenko, "Random Gaussian band matrices and freeness with amalgamation," *Int. Math. Res. Note*, vol. 20, pp. 1013–1025, 1996.

[222] 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.

[223] J. W. Silverstein, "The limiting eigenvalue distribution of a multivariate F-matrix," *SIAM J. of Math. Analysis*, vol. 16, pp. 641–646, 1985.

[224] J. W. Silverstein, "On the eigenvectors of large dimensional sample covariance matrices," *J. of Multivariate Analysis*, vol. 30, pp. 1–16, 1989.

[225] J. W. Silverstein, "Weak-convergence of random functions defined by the eigenvectors of sample covariance matrices," *Annals of Probability*, vol. 18, pp. 1174–1194, July 1990.

[226] 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.

[227] 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.

[228] 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.

[229] S. H. Simon and A. L. Moustakas, "Optimizing MIMO systems with channel covariance feedback," *IEEE J. on Selected Areas in Communications*, vol. 21, pp. 406–417, Apr. 2003.

[230] S. H. Simon and A. L. Moustakas, "Eigenvalue density of correlated complex random Wishart matrices," *Physical Review E*, vol. 69, to appear, 2004.

[231] S. H. Simon, A. L. Moustakas, and L. Marinelli, "Capacity and character expansions: Moment generating function and other exact results for MIMO correlated channels," *Bell Labs Technical Memorandum ITD-04-45211T*, Mar. 2004.

[232] R. Singh and L. Milstein, "Interference suppression for DS-CDMA," *IEEE Trans. on Communications*, vol. 47, pp. 446–453, Mar. 1999.

[233] P. J. Smith and L. M. Garth, "Exact capacity distribution for dual MIMO systems in Ricean fading," *IEEE Communications Letters*, vol. 8, pp. 18–20, Jan. 2004.

[234] P. J. Smith, S. Roy, and M. Shafi, "Capacity of MIMO systems with semicorrelated flat-fading," *IEEE Trans. on Information Theory*, vol. 49, pp. 2781–2788, Oct. 2003.

[235] P. J. Smith and M. Shafi, "On a Gaussian approximation to the capacity of wireless MIMO systems," *Proc. IEEE Int. Conf. in Communications. (ICC'02)*, pp. 406–410, Apr. 2002.

[236] P. Soma, D. S. Baum, V. Erceg, R. Krishnamoorthy, and A. Paulraj, "Analysis and modelling of multiple-input multiple-output (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.

[237] O. Somekh, B. M. Zaidel, and S. Shamai, "Spectral efficiency of joint multiple cell-site processors for randomly spread DS-CDMA systems," in *Proc. IEEE Int. Symp. on Information Theory (ISIT'04)*, (Chicago, IL), June 2004.

[238] H. J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein, "Spectrum of large random asymmetric matrices," *Physical Review Letters*, vol. 60, pp. 1895–1899, May 1988.

[239] R. Speicher, "A new example of independence and white noise," *Probability Theory and Related Fields*, vol. 84, pp. 141–159, 1990.

[240] R. Speicher, "Free convolution and the random sum of matrices," *Publ. Res. Inst. Math. Sci.*, vol. 29, pp. 731–744, 1993.

[241] R. Speicher, *Free probability theory and non-crossing partitions.* 39^{e} Séminaire Lotharingien de Combinatoire, 1997.

[242] R. Speicher, “Free calculus,” in *Summer school on Quantum Probability*, (Grenoble, France), 1998.

[243] R. Speicher, *Freie Wahrscheinlichkeitstheorie (WS 97/98).* Heidelberg University, 1998.

[244] M. S. Srivastava and C. G. Khatri, *An Introduction to Multivariate Statistics.* North-Holland, Amsterdam, 1979.

[245] G. W. Stewart, “The efficient generation of random orthogonal matrices with an application to conditional estimation,” *SIAM J. Numer. Analysis*, vol. 17, pp. 403–409, June 1980.

[246] 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).

[247] E. G. Ström and S. L. Miller, “Properties of the single-bit single-user MMSE receiver for DS-CDMA system,” *IEEE Trans. on Communications*, vol. 47, pp. 416–425, Mar. 1999.

[248] S. Szarek and D. Voiculescu, “Volumes of restricted Minkowski sums and the free analogue of the entropy power inequality,” *Communications in Mathematical Physics*, vol. 178, pp. 563–570, July 1996.

[249] T. Tanaka, “A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors,” *IEEE Trans. on Information Theory*, vol. 48, pp. 2888–2910, Nov. 2002.

[250] E. Telatar, “Capacity of multi-antenna Gaussian channels,” *Euro. Trans. Telecommunications*, vol. 10, pp. 585–595, Nov.-Dec. 1999.

[251] L. Trichard, I. Collings, and J. Evans, “Parameter selection for multiuser receivers based on iterative methods,” in *Proc. IEEE Vehicular Techn. Conf. (VTC’00)*, (Tokyo, Japan), pp. 926–930, May 2000.

[252] L. Trichard, J. Evans, and I. Collings, “Large system analysis of linear multistage parallel interference cancellation,” *IEEE Trans. on Communications*, vol. 50, pp. 1778–1786, Nov. 2002.

[253] L. Trichard, J. Evans, and I. Collings, “Optimal linear multistage receivers and the recursive large system SIR,” in *Proc. IEEE Int. Symp. on Information Theory (ISIT’02)*, (Lausanne, Switzerland), p. 21, June 2002.

[254] L. Trichard, J. Evans, and I. Collings, “Large system analysis of second-order linear multistage CDMA receivers,” *IEEE Trans. on Wireless Communications*, vol. 2, pp. 591–600, May 2003.

[255] L. Trichard, J. Evans, and I. Collings, “Optimal linear multistage receivers with unequal power users,” in *Proc. IEEE Int. Symp. on Information Theory (ISIT’03)*, (Yokohama, Japan), p. 21, June 2003.

[256] D. Tse and S. Hanly, “Linear multiuser receivers: Effective interference, effective bandwidth and user capacity,” *IEEE Trans. on Information Theory*, vol. 45, pp. 641–657, Mar. 1999.

[257] D. Tse and O. Zeitouni, “Linear multiuser receivers in random environments,” *IEEE Trans. on Information Theory*, vol. 46, pp. 171–188, Jan. 2000.

[258] D. N. Tse and S. Verdú, “Optimum asymptotic multiuser efficiency of randomly spread CDMA,” *IEEE Trans. on Information Theory*, vol. 46, no. 6, pp. 2718–2723, Nov. 2000.

[259] D. N. Tse and P. Viswanath, "On the capacity of the multiple antenna broadcast channel," in *Multiantenna channels: Capacity, Coding and Signal Processing*, (G. Foschini and S. Verdú, eds.), pp. 87–106, American Mathematical Society Press, 2003.

[260] B. S. Tsybakov, "The capacity of a memoryless Gaussian vector channel," *Problems of Information Transmission*, vol. 1, pp. 18–29, 1965.

[261] A. M. Tulino, A. Lozano, and S. Verdú, "Capacity-achieving input covariance for single-user multi-antenna channels," *Bell Labs Tech. Memorandum ITD-04-45193Y (also submitted to IEEE Trans. on Wireless Communications.)*, Sep. 2003.

[262] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of correlation on the capacity of multi-antenna channels," *Bell Labs Technical Memorandum ITD-03-44786F (also submitted to IEEE Trans. on Information Theory)*, Sep. 2003.

[263] A. M. Tulino, A. Lozano, and S. Verdú, "MIMO capacity with channel state information at the transmitter," in *Proc. IEEE Int. Symp. on Spread Spectrum Tech. and Applications (ISSSTA'04)*, Aug. 2004.

[264] A. M. Tulino, A. Lozano, and S. Verdú, "Power allocation in multi-antenna communication with statistical channel information at the transmitter," in *Proc. IEEE Int. Conf. on Personal, Indoor and Mobile Radio Communications. (PIMRC'04)*, (Barcelona, Catalonia, Spain), Sep. 2004.

[265] A. M. Tulino and S. Verdú, "Asymptotic analysis of improved linear receivers for BPSK-CDMA subject to fading," *IEEE J. on Selected Areas in Communications*, vol. 19, pp. 1544–1555, Aug. 2001.

[266] A. M. Tulino, S. Verdú, and A. Lozano, "Capacity of antenna arrays with space, polarization and pattern diversity," *Proc. 2003 IEEE Information Theory Workshop (ITW'03)*, pp. 324–327, Apr. 2003.

[267] H. Uhlig, "On singular Wishart and singular multivariate beta distributions," *Annals of Statistics*, vol. 22, pp. 395–405, 1994.

[268] V. V. Veeravalli, Y. Liang, and A. Sayeed, "Correlated MIMO Rayleigh fading channels: Capacity, optimal signalling and asymptotics," *submitted to IEEE Trans. on Information Theory*, 2003.

[269] S. Venkatesan, S. H. Simon, and R. A. Valenzuela, "Capacity of a Gaussian MIMO channel with nonzero mean," *Proc. 2003 IEEE Vehicular Technology Conf. (VTC'03)*, Oct. 2003.

[270] S. Verdú, "Capacity region of Gaussian CDMA channels: The symbol synchronous case," in *Proc. Allerton Conf. on Communication, Control and Computing*, (Monticello, IL), pp. 1025–1034, Oct. 1986.

[271] S. Verdú, *Multiuser Detection.* Cambridge, UK: Cambridge University Press, 1998.

[272] S. Verdú, "Random matrices in wireless communication, proposal to the National Science Foundation," Feb. 1999.

[273] S. Verdú, "Large random matrices and wireless communications," *2002 MSRI Information Theory Workshop*, Feb 25–Mar 1, 2002.

[274] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. on Information Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[275] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. on Information Theory*, vol. 45, pp. 622–640, Mar. 1999.

[276] S. Vishwanath, N. Jindal, and A. Goldsmith, "On the capacity of multiple input multiple output broadcast channels," in *Proc. IEEE Int. Conf. in Communications (ICC'02)*, pp. 1444–1450, Apr. 2002.

[277] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. on Information Theory*, vol. 49, pp. 2658–2668, Oct. 2003.

[278] S. Vishwanath, G. Kramer, S. Shamai, S. Jafar, and A. Goldsmith, "Capacity bounds for Gaussian vector broadcast channels," in *Multiantenna channels: Capacity, Coding and Signal Processing*, (G. Foschini and S. Verdú, eds.), pp. 107–122, American Mathematical Society Press, 2003.

[279] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. on Information Theory*, vol. 47, pp. 2632–2639, Sep. 2001.

[280] P. Viswanath and D. N. Tse, "Sum capacity of the multiple antenna Gaussian broadcast channel," in *Proc. IEEE Int. Symp. Information Theory (ISIT'02)*, p. 497, June 2002.

[281] P. Viswanath, D. N. Tse, and V. Anantharam, "Asymptotically optimal water-filling in vector multiple-access channels," *IEEE Trans. on Information Theory*, vol. 47, pp. 241–267, Jan. 2001.

[282] H. Viswanathan and S. Venkatesan, "Asymptotics of sum rate for dirty paper coding and beamforming in multiple antenna broadcast channels," in *Proc. Allerton Conf. on Communication, Control and Computing*, (Monticello, IL), Oct. 2003.

[283] D. Voiculescu, "Asymptotically commuting finite rank unitary operators without commuting approximants," *Acta Sci. Math.*, vol. 45, pp. 429–431, 1983.

[284] D. Voiculescu, "Symmetries of some reduced free product c^*-algebra," in *Operator algebras and their connections with topology and ergodic theory, Lecture Notes in Mathematics, vol. 1132*, pp. 556–588, Berlin: Springer, 1985.

[285] D. Voiculescu, "Addition of certain non-commuting random variables," *J. Funct. Analysis*, vol. 66, pp. 323–346, 1986.

[286] D. Voiculescu, "Multiplication of certain non-commuting random variables," *J. Operator Theory*, vol. 18, pp. 223–235, 1987.

[287] D. Voiculescu, "Limit laws for random matrices and free products," *Inventiones Mathematicae*, vol. 104, pp. 201–220, 1991.

[288] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory, I," *Communications in Math. Physics*, vol. 155, pp. 71–92, July 1993.

[289] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory, II," *Inventiones Mathematicae*, vol. 118, pp. 411–440, Nov. 1994.

[290] D. Voiculescu, "Alternative proofs for the type II free Poisson variables and for the free compression results (appendix to a paper by A. Nica and R. Speicher)," *American J. Math.*, vol. 118, pp. 832–837, 1996.

[291] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory III: The absence of Cartan subalgebras," *Geometric and Functional Analysis*, vol. 6, pp. 172–199, 1996.

[292] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory, IV: Maximum entropy and freeness," in *Free Probability Theory*, (D. Voiculescu, ed.), pp. 293–302, Fields Inst. Communications, Amer. Math. Soc., 1997.

[293] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory V: noncommutative Hilbert transforms," *Inventiones Mathematicae*, vol. 132, pp. 189–227, Apr. 1998.

[294] D. Voiculescu, "A strengthened asymptotic freeness result for random matrices with applications to free entropy," *Int. Math. Res. Notices*, vol. 1, pp. 41–63, 1998.

[295] D. Voiculescu, "Lectures on free probability theory," in *Lectures on Probability theory and Statistics: Ecole d'Ete de Probabilites; Lecture Notes in Mathematics*, pp. 283–349, Springer, 2000.

[296] K. W. Wachter, "The strong limits of random matrix spectra for sample matrices of independent elements," *Annals of Probability*, vol. 6, no. 1, pp. 1–18, 1978.

[297] K. W. Wachter, "The limiting empirical measure of multiple discriminant ratios," *Annals of Statistics*, vol. 8, pp. 937–957, 1980.

[298] X. Wang and H. V. Poor, "Blind multiuser detection: A subspace approach," *IEEE Trans. on Information Theory*, vol. 44, pp. 677–690, Mar. 1998.

[299] Z. Wang and G. Giannakis, "Outage mutual information of space-time MIMO channels," *IEEE Trans. on Information Theory*, vol. 50, pp. 657–663, Apr. 2004.

[300] S. Wei and D. Goeckel, "On the minimax robustness of the uniform transmission power strategy in MIMO systems," *IEEE Communications Letters*, vol. 7, pp. 523–524, Nov. 2003.

[301] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "Stochastic MIMO channel model with joint correlation of both link ends," *to appear in IEEE Trans. on Wireless Communications*, 2004.

[302] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," *Proc. Conf. on Information Sciences and Systems (CISS'04)*, pp. 7–12, Mar. 2004.

[303] E. Wigner, "Characteristic vectors of bordered matrices with infinite dimensions," *Annals of Mathematics*, vol. 62, pp. 546–564, 1955.

[304] E. Wigner, "Results and theory of resonance absorption," in *Conference on Neutron Physics by Time-of-Flight*, Nov. 1-2 1956. Oak Ridge National Lab. Report ORNL-2309.

[305] E. Wigner, "On the distribution of roots of certain symmetric matrices," *Annals of Mathematics*, vol. 67, pp. 325–327, 1958.

[306] E. Wigner, "Statistical properties of real symmetric matrices with many dimensions," *Proc. 4th Canadian Math. Congress*, pp. 174–176, 1959.

[307] E. Wigner, "Distribution laws for the roots of a random Hermitian matrix," in *Statistical Theories of Spectra: Fluctuations*, (C. E. Porter, ed.), New York: Academic, 1965.

[308] E. Wigner, "Random matrices in physics," *SIAM Review*, vol. 9, pp. 1–123, 1967.

[309] J. H. Winters, "Optimum combining in digital mobile radio with cochannel interference," *IEEE J. on Selected Areas in Communications*, vol. 2, pp. 528–539, July 1984.

[310] J. H. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. on Communications*, vol. 42, pp. 1740–1751, Feb./Mar./Apr. 1994.

[311] J. Wishart, "The generalized product moment distribution in samples from a normal multivariate population," *Biometrika*, vol. 20 A, pp. 32–52, 1928.

[312] W. Xiao and M. L. Honig, "Large system convergence analysis of adaptive reduced- and full-rank least squares algorithms," *IEEE Trans. on Information Theory*, 2004, to appear.

[313] Y. Q. Yin, "Limiting spectral distribution for a class of random matrices," *J. of Multivariate Analysis*, vol. 20, pp. 50–68, 1986.

[314] Y. Q. Yin and P. R. Krishnaiah, "A limit theorem for the eigenvalues of product of two random matrices," *J. of Multivariate Analysis*, vol. 13, pp. 489–507, 1984.

[315] Y. Q. Yin and P. R. Krishnaiah, "Limit theorem for the eigenvalues of the sample covariance matrix when the underlying distribution is isotropic," *Theory Prob. Appl.*, vol. 30, pp. 861–867, 1985.

[316] W. Yu and J. Cioffi, "Trellis precoding for the broadcast channel," in *Proc. IEEE Global Telecomm. Conf. (GLOBECOM'01)*, pp. 1344–1348, Oct. 2001.

[317] B. M. Zaidel, S. Shamai, and S. Verdú, "Multicell uplink spectral efficiency of coded DS-CDMA with random signatures," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 1556–1569, Aug. 2001.

[318] J. Zhang and X. Wang, "Large-system performance analysis of blind and group-blind multiuser receivers," *IEEE Trans. on Information Theory*, vol. 48, pp. 2507–2523, Sep. 2002.