RANDOM MATRICES, FREE PROBABILITY AND THE REPLICA METHOD

Ralf R. Müller

Forschungszentrum Telekommunikation Wien (ftw.)
Tech-Gate Vienna, Donau-City Str. 1/3, 1220 Wien, Austria, e-mail: mueller@ftw.at.

ABSTRACT

This survey gives an overview of analytic tools to the design, analysis, and modelling of communication systems which can be described by linear vector channels such as $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$ where the number of components in each vector is large. Tools from probability theory, operator algebra, and statistical physics are reviewed.

Asymptotic eigenvalue distributions of some classes of random matrices are given in terms of densities, moments and/or Stieltjes transforms. Free probability theory which evolved from non-commutative operator algebras is explained from a probabilistic point of view in order to better fit the engineering community. For that purpose freeness is defined without reference to non-commutative algebras. The treatment includes additive and multiplicative free convolution, the R-transform, and the S-transform. The replica method developed in statistical physics for the purpose of analyzing spin glasses is reviewed from the view point of its applications in communications engineering. Correspondences between free energy and mutual information as well as energy functions and detector metrics are established.

1. INTRODUCTION

In a multi-dimensional communication system many data streams are transmitted from various sources to various sinks via a common medium called a channel. Technical systems handling this task are telephone networks, both fixed and wireless, the internet, local area networks, computers' data buses, etc.

The complexity of such communication systems increases with the number of people or data streams to be handled simultaneously. This rise in complexity is not limited to the hardware to be deployed, but also affects the design, the modeling, and the analysis of the system. From an engineering point of view, it is particularly important to be able to predict the behavior of a technical system, before it is actually built. With respect to the steadily increasing number of people using various kinds of communications technology, this seems to become, sooner or later, a hopeless task at first sight.

In a combustion engine, many molecules of fuel and air interact with each other. However, though we cannot control the individual behavior of each molecule, and even do not intend to do so, we can trust that the mixture of gas and air will explode when it is lighted, heat up, expand, and drive the engine. Physicists have successfully build the theory of thermodynamics to explain the evolution of macroscopic values like temperature and pressure though the microscopic behavior of the molecules is only described statistically. Simply the fact, that there are enough objects which interact randomly, makes the unity of these objects to obey certain rules. These rules depend on the kind of interaction and some other things and can be understood as generalizations of the law of large numbers.

Communication systems for multiple data streams can be modeled as well by statistical interactions between the signals belonging to different data streams. Provided that the number of data streams transmitted simultaneously through the system is large enough, similar effects as in thermodynamics occur.

The use of microscopic statistical models to predict macroscopic quantities in physics is not limited to thermodynamics. It was already used by Wigner in the 1950s in order to predict the spacings of nuclear energy levels. Random matrices and their applications in communications engineering are discussed in Section 3.

Advances in *free probability theory* are driven by mathematicians interested in operator algebras and the Riemann hypothesis. Since numerical evidence showed a striking connection between the zeros of Riemann's zeta function and the spacings of adjacent eigenvalues of large random matrices, physicist started to hope for new insights into their problems. However, the first concrete meaning for the *R-transform*, one of the most fundamental concepts in free probability theory, was found in the theory of large CDMA systems. An overview on free probability is given in Section 4.

Random matrix and free probability theory are concepts well suited to analyze the interaction of many Gaussian random processes. However, they do not respect the binary nature of most signals in modern communications. Driven by the laws of quantum mechanics which allow only two values for the spin of an electron, physicists have developed powerful tools to analyze the thermodynamics of magnetic materials, one of which is called the *replica method*. Though from a mathematical point of view not rigorously established yet, it is able to predict the macroscopic behavior of spin glasses just as well as bit error rates of maximum a-posteriori detectors for CDMA signals. An overview on the replica method is given in Section 5.

2. APPLICATIONS

The principles to be surveyed here apply to a broad class of communication channels. To just pick one of them, consider the vector-valued additive white Gaussian noise channel

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

with the $K \times 1$ vector of transmitted symbols $\mathbf{x}[\nu]$, the $N \times 1$ vector of received symbols $\mathbf{y}[\nu]$, the $N \times K$ channel matrix $\mathbf{H}[\nu]$, the $N \times 1$ vector of additive white Gaussian noise $\mathbf{n}[\nu]$, and discrete time ν. The time index will be dropped whenever it is not needed to express the dependency on discrete time explicitly.

It is well known in literature [1] that the signal

$$\mathbf{r}[\nu] = \mathbf{H}^{\mathrm{H}}[\nu]\mathbf{y}[\nu] = \mathbf{H}^{\mathrm{H}}[\nu]\mathbf{H}[\nu]\mathbf{x}[\nu] + \mathbf{H}^{\mathrm{H}}[\nu]\mathbf{n}[\nu] \tag{2}$$

provides sufficient statistics for the estimation of the signal $\mathbf{x}[\nu]$. All information about $\mathbf{x}[\nu]$ that could be extracted from the received signal $\mathbf{y}[\nu]$ can also be extracted from the signal $\mathbf{r}[\nu]$.

The two equivalent channels (1) and (2) appear in several areas of wireless and wireline communications:

- In CDMA, the components of the vector $\mathbf{x}$ are regarded as the signals of K individual users while the matrix $\mathbf{H}$ contains their spreading sequences as columns.
- In antenna array communications, the components of the vectors $\mathbf{x}$ and $\mathbf{y}$ represent the signals sent and received by the K transmit and N receive antenna elements, respectively.
- In cable transmission, the components of the vector $\mathbf{x}$ contains the signals sent on the bundled twisted pairs within a cable. The coefficients in the matrix $\mathbf{H}^{\mathrm{H}}\mathbf{H}$ describe the electromagnetic crosstalk between the respective twisted pairs.
- For block transmission over a dispersive channel, the components of the vectors $\mathbf{x}$ and $\mathbf{y}$ contain the symbols sent and received consecutively in time. Discrete time ν counts blocks,

and the matrix $\mathbf{H}$ is a circulant matrix of the channel's discrete-time impulse response.

- In orthogonal frequency-division multiple access (OFDM), the components of the vectors $\mathbf{x}$ and $\mathbf{r}$ represent the K sub-carriers at transmitter and receiver site, respectively, and the matrix $\mathbf{H}^{\mathrm{H}}\mathbf{H}$ accounts for inter-carrier interference.

Regardless of the application one has in mind, performances can be analyzed for a variety of receiver algorithms and assumptions on the properties of the channel matrix $\mathbf{H}$. Numerous results are reported in literature [1, 2] and we do not try to make any effort to be comprehensive here.

3. RANDOM MATRIX THEORY

The channel matrix $\mathbf{H}$ in (1) is composed of NK random elements. Though it can be simply considered as an NK dimensional random object, it has also some more interesting interpretations.

Consider a scalar zero-mean random process $H_\eta[\mu]$ over discrete time μ. Stack the time samples into the row dimensions of the matrix $\mathbf{H}$ and the ensembles η into the column dimensions of $\mathbf{H}$ such that

$$\mathbf{H} = \begin{bmatrix} H_1[0] & H_1[1] & H_1[2] & \cdots \\ H_2[0] & H_2[1] & H_2[2] & \cdots \\ H_3[0] & H_3[1] & H_3[2] & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{3}$$

If we let the dimension $N, K \to \infty$, the matrix $\mathbf{H}$ describes a whole random process. Nevertheless, we can still think of it as a single realization of a many-dimensional random variable. This double interpretation results in the *self-averaging property* of many functions of a large dimensional random matrices.

Consider the function rowsum : $\mathbf{X} \in \mathbb{C}^{N\times K} \mapsto K^{-\frac{1}{2}}\mathbf{X}\mathbf{1} \in \mathbb{C}^N$ with $\mathbf{1}$ denoting the all one vector. It simply sums up the rows of its argument and normalizes the result. As $K \to \infty$, the value of the function $\mathbf{h} = \mathrm{rowsum}(\mathbf{H})$ is an N-dimensional Gaussian random vector due to the central limit theorem. As $N \to \infty$, the empirical distribution function of its components h_i

$$\mathrm{P}_{\mathbf{h}}(x) = \frac{1}{N}\left|\left\{h_i : h_i < x\right\}\right| \tag{4}$$

converges to a Gaussian distribution. Communications engineering builds upon this result, whenever something is modeled as a Gaussian random process.

The type of distribution $\mathbf{h}$ follows does not depend on the distribution of $\mathbf{H}$ as $N \to \infty$. Instead, the distribution of $\mathbf{h}$ is determined by the mapping from $N \times K$-dimensional space into N-dimensional space. For many linear mappings, such as rowsum$(\cdot)$, the projection follows a Gaussian law. For non-linear mappings, a rich plurality of other limit distributions occurs.

3.1 Convergence of Eigenvalues

The eigenvalues of the channel matrix are important to characterize performance measures in communications engineering. Calculating the eigenvalues of (a function of) a random matrix, is a projection, though a non-linear one, similar to the rowsum function.

Calculation of the moments of the eigenvalue distribution is conveniently done by a normalized trace since

$$\frac{1}{N}\sum_{k=1}^{N}\lambda_k^m = \frac{1}{N}\mathrm{trace}\left(\mathbf{H}^m\right). \tag{5}$$

In the following, we also use

$$\mathrm{Tr}(\mathbf{H}) \triangleq \lim_{N\to\infty}\frac{1}{N}\mathrm{trace}(\mathbf{H}). \tag{6}$$

to denote the normalized trace in the large matrix limit. The eigenvalue distributions of several types of random matrices are examined in greater detail in the following.

3.1.1 Quarter Circle Law

Let the random matrix $\mathbf{H}$ be square, $N \times N$, with independent identically distributed entries with zero mean and variance $1/N$. Let

$$\mathbf{Q} = \sqrt{\mathbf{H}\mathbf{H}^{\mathrm{H}}}. \tag{7}$$

Let $\mathscr{L}$ denote the set containing the eigenvalues of $\mathbf{Q}$, i.e. the singular values of $\mathbf{H}$. Then, the empirical distribution of the eigenvalues

$$\mathrm{P}_{\mathbf{Q}}(x) = \frac{1}{N}\left|\left\{\lambda \in \mathscr{L} : \lambda < x\right\}\right| \tag{8}$$

converges to a non-random distribution function as $N \to \infty$ whose density is given by

$$\mathrm{p}_{\mathbf{Q}}(x) = \begin{cases} \frac{1}{\pi}\sqrt{4-x^2} & 0 \le x < 2 \\ 0 & \text{elsewhere} \end{cases}. \tag{9}$$

This distribution is called the quarter circle distribution.

With standard methods for the transformation of probability densities, see [3], the asymptotic eigenvalue distribution of $\mathbf{Q}^2 = \mathbf{H}\mathbf{H}^{\mathrm{H}}$ can be derived. It reads

$$\mathrm{p}_{\mathbf{Q}^2}(x) = \begin{cases} \frac{1}{2\pi}\sqrt{\frac{4-x}{x}} & 0 < x < 4 \\ 0 & \text{elsewhere} \end{cases}. \tag{10}$$

Its m^{th} moments are the Catalan numbers which play an important role in combinatorics [4].

3.1.2 Deformed Quarter Circle Law

The quarter circle law is part of a more general result for rectangular matrices: Let the entries of the $N \times K$ matrix $\mathbf{H}$ be independent identically distributed with zero mean and variance $1/N$. Then, the empirical distribution of the singular values of $\mathbf{H}$, i.e. the eigenvalues of

$$\mathbf{R} = \sqrt{\mathbf{H}\mathbf{H}^{\mathrm{H}}} \tag{11}$$

converges to a non-random distribution function as $N, K \to \infty$ with $\beta = K/N$ fixed and its density is given by

$$\mathrm{p}_{\mathbf{R}}(x) = \begin{cases} \frac{\sqrt{4\beta-(x^2-1-\beta)^2}}{\pi\sqrt{x}} & |1-\sqrt{\beta}| < x < 1+\sqrt{\beta} \\ [1-\beta]^+\,\delta(x) & \text{elsewhere} \end{cases}. \tag{12}$$

Again, we also consider the eigenvalue distribution of $\mathbf{R}^2 = \mathbf{H}\mathbf{H}^{\mathrm{H}}$ and find

$$\mathrm{p}_{\mathbf{R}^2}(x) = \begin{cases} \frac{\sqrt{4\beta-(x-1-\beta)^2}}{2\pi x} & (1-\sqrt{\beta})^2 < x < (1+\sqrt{\beta})^2 \\ [1-\beta]^+\,\delta(x) & \text{elsewhere} \end{cases}. \tag{13}$$

This distribution is known as the Marčenko-Pastur distribution. It has been used in [5] to calculate channel capacity for CDMA with random spreading.

3.1.3 Haar Distribution

Let the random matrix $\mathbf{H}$ be square $N \times N$ with independent identically complex Gaussian distributed proper[1] entries with zero mean and finite positive variance. Then, the empirical distribution of the eigenvalues of the unitary random matrix

$$\mathbf{T} = \mathbf{H}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-\frac{1}{2}} \tag{14}$$

converges to a non-random distribution function as $N \to \infty$ that is uniform on the complex unit circle and vanishes elsewhere.

[1] A complex random variable is said to be *proper* if real and imaginary part are independent and identically distributed [6].

3.2 Stieltjes Transform

The task of finding an unknown probability distribution given its moments is known as the *problem of moments*. It was addressed by Stieltjes in 1894 using the integral transform

$$\mathrm{G}(s) \triangleq \int \frac{\mathrm{dP}(x)}{x-s} \tag{15}$$

with $\Im s > 0$. It is now commonly referred to as the *Stieltjes transform*.

A Taylor series expansion of its kernel

$$-\lim_{s\to 0} \frac{\mathrm{d}^m}{\mathrm{d}s^m} \frac{\mathrm{G}\left(s^{-1}\right)}{s} = m! \int x^m \mathrm{dP}(x) \tag{16}$$

shows how the moments can be found given the Stieltjes transform without the need for integration. The probability density function can be obtained from the Stieltjes transform taking the limit

$$\mathrm{p}(x) = \lim_{y\to 0+} \frac{1}{\pi} \Im \, \mathrm{G}(x+\mathrm{j}y) \tag{17}$$

which is often called the *Stieltjes inversion formula* [7].

3.2.1 Products of Random Matrices

Let the random matrix $\mathbf{H}$ fulfill the same conditions as needed for the deformed quarter circle law. Moreover, let $\mathbf{X} = \mathbf{X}^{\mathrm{H}}$ be an $N \times N$ Hermitian matrix, independent of $\mathbf{H}$, with an empirical eigenvalue distribution converging almost surely in distribution to a distribution function $\mathrm{P}_{\mathbf{X}}(x)$ as $N \to \infty$. Then, almost surely, the eigenvalue distribution of the matrix product $\mathbf{P} = \mathbf{H}\mathbf{H}^{\mathrm{H}}\mathbf{X}$ converges in distribution, as $K, N \to \infty$, but $\beta = K/N$ fixed, to a nonrandom distribution function whose Stieltjes transform satisfies for $\Im s > 0$ [8, 9]

$$\mathrm{G}_{\mathbf{P}}(s) = \int \frac{\mathrm{dP}_{\mathbf{X}}(x)}{x\big(1-\beta-\beta s \mathrm{G}_{\mathbf{P}}(s)\big) - s}. \tag{18}$$

3.2.2 Sums of Random Matrices

Let the random matrix $\mathbf{H}$ fulfill the same conditions as needed for the deformed quarter circle law. Let $\mathbf{X} = \mathbf{X}^{\mathrm{H}}$ be an $N \times N$ Hermitian matrix with an eigenvalue distribution function converging weakly to $\mathrm{P}_{\mathbf{X}}(x)$ almost surely. Let $\mathbf{Y} = \mathrm{diag}(y_1, \ldots, y_K)$ be a $K \times K$ diagonal matrix and the empirical distribution function of $\{y_1, \ldots, y_K\} \in \mathbb{R}^K$ converge almost surely in distribution to a probability distribution function $\mathrm{P}_{\mathbf{Y}}(x)$ as $K \to \infty$. Let $\mathbf{H}, \mathbf{X}, \mathbf{Y}$ be independent. Then, almost surely, the empirical eigenvalue distribution of the random matrix $\mathbf{S} = \mathbf{X} + \mathbf{H}\mathbf{Y}\mathbf{H}^{\mathrm{H}}$ converges weakly, as $K, N \to \infty$, but $\beta = K/N$ fixed, to a nonrandom distribution function whose Stieltjes transform satisfies for $\Im s > 0$ [10, 11]

$$\mathrm{G}_{\mathbf{S}}(s) = \mathrm{G}_{\mathbf{X}}\left(s - \beta \int \frac{y \, \mathrm{dP}_{\mathbf{Y}}(y)}{1 + y \mathrm{G}_{\mathbf{S}}(s)}\right). \tag{19}$$

This result was used in [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22] to derive results on the capacity, the SINR, and the output statistics of CDMA or antenna array channels.

3.2.3 Girko's Law

Let the $N \times K$ random matrix $\mathbf{H}$ be composed of independent entries $(\mathbf{H})_{ij}$ with zero-mean and variances w_{ij}/N such that all w_{ij} are uniformly bounded from above. Assume that the empirical joint distribution of variances $w : [0,1] \times [0,\beta] \mapsto \mathbb{R}$ defined by $w(x,y) = w_{ij}$ for i, j satisfying $i/N \le x \le (i+1)/N$ and $j/N \le y \le (j+1)/N$ converges to a bounded joint limit distribution $w(x,y)$ as $K = \beta N \to \infty$. Then, for each $a, b \in [0,1], a < b$, and $\Im(s) > 0$

$$\frac{1}{N} \sum_{i=\lceil aN \rceil}^{\lceil bN \rceil} \left(\mathbf{H}\mathbf{H}^{\mathrm{H}} - s\mathbf{I}\right)^{-1}_{ii} \longrightarrow \int_a^b u(x,s) \, \mathrm{d}x \tag{20}$$

where convergence is in probability and $u(x,s)$ satisfies the fixed point equation

$$u(x,s) = \left[-s + \int_0^\beta \frac{w(x,y) \, \mathrm{d}y}{1 + \int_0^1 u(x',s) w(x',y) \, \mathrm{d}x'} \right]^{-1} \tag{21}$$

for every $x \in [0,1]$. The solution to (21) exists and is unique in the class of functions $u(x,s) \ge 0$, analytic for $\Im(s) > 0$ and continuous on $x \in [0,1]$.

Moreover, almost surely, the empirical eigenvalue distribution of $\mathbf{H}\mathbf{H}^{\mathrm{H}}$ converges weakly to a limiting distribution whose Stieltjes transform is given by [23]

$$G_{\mathbf{H}\mathbf{H}^{\mathrm{H}}}(s) = \int_0^1 u(x,s) \, \mathrm{d}x. \tag{22}$$

This result was used in [24] to analyze asynchronous CDMA in the large system limit, in [25] to proof resource pooling of chips and receive antennas in CDMA systems with antenna diversity, and in [26] to study capacity scaling in large dual antenna array systems.

3.3 Convergence Properties of Eigenvectors

While there are many results known in literature about the eigenvalues of large random matrices, few is known about the eigenvectors. However, there is one particular result which proves helpful for communications engineering applications:

Let $\mathbf{H}$ be an $N \times K$ random matrix with independent identically distributed real-valued random entries with zero mean and all positive moments bounded from above. Let the orthogonal matrix $\mathbf{U}$ be defined be the eigenvalue decomposition

$$\mathbf{U}^{\mathrm{T}} \Lambda \mathbf{U} = \mathbf{H}^{\mathrm{T}} \mathbf{H}. \tag{23}$$

Note that the rows of $\mathbf{U}$ are the eigenvectors of $\mathbf{H}^{\mathrm{T}}\mathbf{H}$. Let $\mathbf{x} \in \mathbb{R}^N$ with $||\mathbf{x}|| = 1$ be an arbitrary vector with unit Euclidean norm and the random vector $\mathbf{y} = [y_1, \ldots, y_N]^{\mathrm{T}}$ be defined as $\mathbf{y} = \mathbf{U}\mathbf{x}$. Then, as $N, K \to \infty$, but $\beta = K/N$ fixed,

$$\sum_{k=1}^{\lceil tN \rceil} y_k^2 \longrightarrow t \tag{24}$$

almost surely for every $t \in [0;1]$ with $\lceil x \rceil$ denoting the nearest integer to x which is not smaller than x [27].

This result is like a law of large numbers for the components of any linear combination of the components of the eigenvectors of $\mathbf{H}^{\mathrm{T}}\mathbf{H}$. The elements of the eigenvector matrix $\mathbf{U}^{\mathrm{T}}$ behave, for the purpose of summing its squared elements, as they were statistically independent in the large matrix limit. This property was used in [28] to show that the SINR of linear multiuser receivers is asymptotically Gaussian distributed for a random signature assignment.

4. FREE PROBABILITY THEORY

While random matrix theory considers a large random matrix as a whole ensemble and proves convergence results, free probability looks at a random matrix from a different point of view: A random matrix is primarily seen as a linear random operator. Free probability theory provides a framework for dealing with certain classes of linear random operators.

The essential feature that distinguishes random operators including random matrices from scalar random variables is the commutative law which, in general, does not hold for matrices and others operators. In order to see what causes problems for probability

theory, if a random matrix is seen as a single (non-commutative) random variable, consider the expectations

$$\mathcal{E}\{(xy)^m\} = \mathcal{E}\{x^m y^m\} \quad (25)$$

$$\underset{\text{free}}{\mathcal{E}}\{(\mathbf{XY})^m\} \neq \underset{\text{free}}{\mathcal{E}}\{\mathbf{X}^m\mathbf{Y}^m\} \quad (26)$$

where x, y are standard scalar random variables and $\mathbf{X}, \mathbf{Y}$ are random operators. For independent random variables, all joint moments must factorize. For statistically independent random matrices $\mathbf{X}$ and $\mathbf{Y}$, this is impossible, in general, due to the non-commutative nature of matrix multiplication. Thus, the fundamental concept of statistical independence does not make sense, if a random matrix is considered as a single random object. Random matrix theory circumvented this problem, considering a random matrix as being composed of standard scalar random variables. Thus, it defines statistical independence of two random matrices if all entries of the one matrix are jointly independent from all entries of the other matrix. In this section, we take the viewpoint of free probability. However, we restrict ourselves to asymptotically large random matrices as free random variables. Free probability theory also applies to other classes of random operators.

4.1 Free Expectation

An expectation operator should be linear and should assign 1 to the identity matrix (the unit element of the matrix algebra). It turns out that

$$\underset{\text{free}}{\mathcal{E}}\{\cdot\} \triangleq \text{Tr}(\cdot) \quad (27)$$

is the right definition for some random matrices to fit into the framework of free probability theory. For those random matrices, $\text{Tr}(\cdot)$ is indeed (almost surely) a deterministic quantity due to the asymptotic convergence of their eigenvalues.

4.2 Freeness

Freeness is the conceptual counterpart in free probability to independence in classical probability theory. Unfortunately, defining freeness is considerably more involved than defining independence.

Consider the following example of four random matrices and assume that they satisfy

$$\text{Tr}(\mathbf{ABCD}) = \text{Tr}(\mathbf{AB})\text{Tr}(\mathbf{CD}) \quad (28)$$

$$\text{Tr}(\mathbf{ACBD}) \neq \text{Tr}(\mathbf{AB})\text{Tr}(\mathbf{CD}). \quad (29)$$

For commutative random variables, (28) and (29) would contradict each other. For non-commutative multiplication, however, (28) and (29) can be true at the same time.

4.2.1 Non-commutative Polynomials

Due to the non-commutative nature of matrix-multiplication, there are more different matrix polynomials of two or more variables for a fixed degree than for commutative variables such as the real or complex numbers. Let $\mathbf{A}, \mathbf{B}$ be real matrices. The set of all second order non-commutative polynomials in two variables $\mathbf{A}$ and $\mathbf{B}$ is given by

$$\begin{aligned}&\alpha_1\mathbf{A}^2\mathbf{B}^2 + \alpha_2\mathbf{AB}^2\mathbf{A} + \alpha_3\mathbf{ABAB} + \alpha_4\mathbf{BABA} + \\ &+ \alpha_5\mathbf{BA}^2\mathbf{B} + \alpha_6\mathbf{B}^2\mathbf{A}^2 + \alpha_7\mathbf{A}^2\mathbf{B} + \alpha_8\mathbf{ABA} + \\ &+ \alpha_9\mathbf{BAB} + \alpha_{10}\mathbf{B}^2\mathbf{A} + \alpha_{11}\mathbf{A}^2 + \alpha_{12}\mathbf{AB} + \\ &+ \alpha_{13}\mathbf{BA} + \alpha_{14}\mathbf{B}^2 + \alpha_{15}\mathbf{A} + \alpha_{16}\mathbf{B} + \alpha_{17}\mathbf{I}.\end{aligned} \quad (30)$$

A non-commutative polynomial in p variables of order n can be defined by

$$\mathscr{P}_n(\mathbf{A}_1, \ldots, \mathbf{A}_p) \triangleq \left\{ \sum_{i=1}^{\infty} \alpha_i \prod_{k=1}^{n} \prod_{q=1}^{p} \mathbf{A}_q^{\ell_{i,k,q}} : \right.$$

$$\left. \sum_{k=1}^{n} \ell_{i,k,q} \in \{0, 1, \ldots, n\} \wedge \alpha_i \in \mathbb{R} \forall i, q \right\}. \quad (31)$$

Note that the number of terms can be considerably large even for small values of n and p.

4.2.2 Definition of Freeness

In literature [29, 30, 31, 7], freeness is defined in terms of algebras and sub-algebras. Here we avoid referring to algebras, and define freeness in terms of non-commutative polynomials.

Definition 4.1 *Let $s_k \in \{1, 2, \ldots, r\}$ be a sequence of integers such that*

$$s_k - s_{k-1} \neq 0. \quad (32)$$

Then, the sets $\mathscr{Q}_1 \triangleq \{\mathbf{A}_1, \ldots, \mathbf{A}_a\}$, $\mathscr{Q}_2 \triangleq \{\mathbf{B}_1, \ldots, \mathbf{B}_b\}$, ..., $\mathscr{Q}_r$ form a free family $(\mathscr{Q}_1, \ldots, \mathscr{Q}_r)$ if, for every sequence s_k obeying (32), any sequence of polynomials $\mathbf{Q}_k$ such that $\mathbf{Q}_k \in \mathscr{P}_\infty(\mathscr{Q}_{s_k})$, and any positive integer n,

$$\text{Tr}(\mathbf{Q}_1) = \ldots = \text{Tr}(\mathbf{Q}_n) = 0 \Longrightarrow \text{Tr}(\mathbf{Q}_1\mathbf{Q}_2\cdots\mathbf{Q}_n) = 0. \quad (33)$$

Note that due to (32) adjacent factors in the product $\mathbf{Q}_1\mathbf{Q}_2\cdots\mathbf{Q}_n$ must be polynomials of different sets of the family. This reflects the non-commutative nature in the definition of freeness.

4.3 Free Random Matrices

Random matrices are a very popular and practically relevant example of non-commutative random variables. However, not all sets of statistically independent random matrices are capable of forming free families. So far, only a few examples of random matrices are known which form free families as their dimensions grow large. Most of them were discovered by Voiculescu [32, 29]. His results were strengthened and extended in [33, 7].

4.3.1 Gaussian Random Matrices

Let the random matrices $\mathbf{H}_i, \forall i$, be square $N \times N$ with independent identically complex Gaussian distributed proper entries with zero mean and variance $1/N$. Moreover, let $\mathbf{X}_j, \forall j$, be an $N \times N$ matrices with upper bounded norm and a limit distribution as $N \to \infty$. Then the family

$$\left(\left\{ \mathbf{X}_1, \mathbf{X}_1^{\text{H}}, \mathbf{X}_2, \mathbf{X}_2^{\text{H}}, \ldots \right\}, \left\{ \mathbf{H}_1, \mathbf{H}_1^{\text{H}} \right\}, \left\{ \mathbf{H}_2, \mathbf{H}_2^{\text{H}} \right\}, \ldots \right) \quad (34)$$

is asymptotically free as $N \to \infty$ almost surely [7, 33].

4.3.2 Hermitian Random Matrices

Let the random matrices $\mathbf{H}_i, \forall i$, be $N \times K$ with independent identically complex Gaussian distributed proper entries with zero mean and variance $1/N$. Moreover, let the matrices $\mathbf{X}_j, \forall j$, be as in Section 4.3.1 and let $\mathbf{S}_i = \mathbf{H}_i\mathbf{H}_i^{\text{H}}, \forall i$. Then the family

$$\left(\left\{ \mathbf{X}_1, \mathbf{X}_1^{\text{H}}, \mathbf{X}_2, \mathbf{X}_2^{\text{H}}, \ldots \right\}, \{\mathbf{S}_1\}, \{\mathbf{S}_2\}, \ldots \right) \quad (35)$$

is almost surely asymptotically free as $N, K \to \infty$ with $\beta = K/N$ fixed [7, 33].

The asymptotic freeness of some random covariance matrices has been used in [34] to analyze multiuser channel estimation in large CDMA systems, in [35] to calculate error rates of space-time codes, in [36] to descibe scattered wave propagation between antenna arrays, and in [37, 38, 39, 40, 41] to design multiuser detectors with sub-cubic complexity.

4.3.3 Unitary Random Matrices

Let the random matrices $\mathbf{T}_i, \forall i$, be $N \times N$ Haar distributed random matrices as defined in Section 3.1.3. Moreover, let the matrices $\mathbf{X}_j, \forall j$, be as in Section 4.3.1. Then, the family

$$\left(\left\{ \mathbf{X}_1, \mathbf{X}_1^{\text{H}}, \mathbf{X}_2, \mathbf{X}_2^{\text{H}}, \ldots \right\}, \left\{ \mathbf{T}_1, \mathbf{T}_1^{\text{H}} \right\}, \left\{ \mathbf{T}_2, \mathbf{T}_2^{\text{H}} \right\}, \ldots \right) \quad (36)$$

is almost surely asymptotically free as $N \to \infty$ [7, 33].

The asymptotic freeness of such unitary random matrices has been used in [42] for the analysis of multi-carrier systems.

4.4 R-Transform

Let $\mathbf{A}$ and $\mathbf{B}$ be two non-commutative random variables belonging to different sets of a free family. Further, let

$$\mathbf{C} \triangleq \mathbf{A} + \mathbf{B}. \tag{37}$$

Then, we call the probability measure (asymptotic eigenvalue distribution) $\mathrm{p}_{\mathbf{C}}(x)$ the *additive free convolution* of the probability measures $\mathrm{p}_{\mathbf{A}}(x)$ and $\mathrm{p}_{\mathbf{B}}(x)$. Unlike classical convolution which provides the distribution of a sum of independent commutative random variables, additive free convolution is a highly non-linear operation.

In principle, the moments of $\mathrm{p}_{\mathbf{C}}(x)$ could be found from the moments of $\mathrm{p}_{\mathbf{A}}(x)$ and $\mathrm{p}_{\mathbf{B}}(x)$ via the definition of freeness. Then, the distributions could be recovered from the moments solving the problem of moments via the Stieltjes transform. However, this is a very tedious task. Significant simplification is achieved via the *R-transform*. The R-transform is defined in terms of the Stieltjes transform as

$$\mathrm{R}(w) \triangleq \mathrm{G}^{-1}(-w) - w^{-1} \tag{38}$$

where $G^{-1}(\cdot)$ denotes the inverse function of the Stieltjes transform with respect to composition (this should not be confused with the inverse Stieltjes transform).

The R-transform linearizes additive free convolution of two non-commutative probability measures [29]. Thus, we have

$$\mathrm{R}_{\mathbf{C}}(w) = \mathrm{R}_{\mathbf{A}}(w) + \mathrm{R}_{\mathbf{B}}(w). \tag{39}$$

Then, the distribution of $\mathbf{C}$ can be recovered inverting (38) and, then, using the Stieltjes inversion formula (17).

Tse [43] discovered that the additivity of the R-transform is responsible for the decoupling of interference powers in the SINRs of asymptotically large random CDMA with linear multiuser receivers.

4.5 S-Transform

In analogy to additive free convolution, we define

$$\mathbf{D} \triangleq \mathbf{A}\mathbf{B} \tag{40}$$

and call the probability measure $\mathrm{p}_{\mathbf{D}}(x)$ the *multiplicative free convolution* of the two probability measures $\mathrm{p}_{\mathbf{A}}(x)$ and $\mathrm{p}_{\mathbf{B}}(x)$, again under the restriction that $\mathbf{A}$ and $\mathbf{B}$ belong to different sets of a free family of non-commutative random variables. Though, the factors are non-commutative operators, multiplicative free convolution is commutative [29].

Under the additional assumption that the probability measures of both factors have non-zero mean, i.e. $\mathrm{Tr}(\mathbf{A}) \neq 0 \neq \mathrm{Tr}(\mathbf{B})$, we can linearize multiplicative free convolutional via the definition of an appropriate transform such that

$$\mathrm{S}_{\mathbf{D}}(z) = \mathrm{S}_{\mathbf{A}}(z)\,\mathrm{S}_{\mathbf{B}}(z) \tag{41}$$

where $\mathrm{S}(\cdot)$ is called the *S-transform*. In order to define the S-transform explicitly, we first introduce an auxiliary transform

$$\Upsilon(s) \triangleq \int \frac{sx}{1-sx}\,\mathrm{dP}(x) = -s^{-1}\mathrm{G}\left(s^{-1}\right) - 1. \tag{42}$$

Calculating the inverse with respect to composition of this auxiliary transform, the S-transform is given as

$$\mathrm{S}(z) \triangleq \frac{1+z}{z}\,\Upsilon^{-1}(z). \tag{43}$$

In order to return to the probability distribution, you return to the Stieltjes domain via (43) and (42) and then apply the Stieltjes inversion formula (17).

It was shown in [18] that formula (18) for products of some asymptotic random matrices is equivalent to applying the S-transform. Thus, the S-transform is not restricted to free random matrices, but also applies to any asymptotic random matrices which obey the conditions of Section 3.2.1.

5. REPLICA METHOD

In the previous part of this work, considerations were restricted to the eigenvalues (and eigenvectors) of random matrices. In order to analyze and design large dimensional communication systems which cannot be described by eigenvalues and eigenvectors alone, but depend on more complicated functions of the channel matrix, such as minimum distances between signal points, a more powerful machinery than random matrix and free probability theory is needed. Such a machinery was developed in statistical physics for the analysis of some particular magnetic materials called spin glasses and is known as the *replica method* [44].

The replica method is also able to reproduce many of the results which were found by means of random matrix and free probability theory, but the calculations based on the replica method are often much more involved. Additionally, the replica method, in contrast to free probability theory, has not been developed into a mature theory, yet. Moreover, it is still lacking mathematical rigor in some respects. However, due to its success in explaining physical phenomena and its consistency with engineering results from random matrix and free probability theory [45], we can trust that its predictions in other engineering applications [46, 47, 48, 49, 50, 51, 52] are correct. Nevertheless, we should always exercise particular care when interpreting new results based on the replica method. Establishing a rigorous mathematical basis for the replica method is a topic of current research in mathematics and theoretical physics.

5.1 Self Average

While random matrix theory and recently also free probability theory [7, 33] prove the (almost sure) convergence of some random variables to deterministic values in the large matrix limit, statistical physics does not always do so. It is considered a fundamental principle of statistical physics that there are microscopic and macroscopic variables. Microscopic variables are physical properties of microscopically small particles, e.g. the speed of a gas molecule or the spin of an electron. Macroscopic variables are physical properties of compound objects that contain many microscopic particles, e.g. the temperature or pressure of a gas, the radiation of a hot object, or the magnetic field of a piece of ferromagnetic material. From a physics point of view, it is clear which variables are macroscopic and which ones are microscopic. An explicit proof that a particular variable is *self-averaging*, i.e. it converges to a deterministic value in the large system limit, is a nice result, if it is found, but it is not considerably important to the physics community. When applying the replica method, systems are often only assumed to be self-averaging. The replica method itself must be seen as a tool to enable the calculation of macroscopic properties by averaging over the microscopic properties.

5.2 Free Energy

The second law of thermodynamics demands the entropy of any physical system with conserved energy to converge to its maximum as time evolves. If the system is described by a density $\mathrm{p}_X(x)$ of states $X \in \mathbb{R}$, this means that in the thermodynamic equilibrium the (differential) entropy

$$\mathrm{H}(X) = -\int \log \mathrm{p}_X(x)\,\mathrm{dP}_X(x) \tag{44}$$

is maximized while keeping the energy

$$\mathrm{E}(X) = \int ||x||\,\mathrm{dP}_X(x) \tag{45}$$

constant. Hereby, the energy function $||x||$ can be any measure which is uniformly bounded from below.

The density at thermodynamic equilibrium is easily shown by the method of Lagrange multipliers to be

$$\mathrm{p}_X(x) = \frac{\mathrm{e}^{-\frac{1}{T}||x||}}{\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{T}||x||}\,\mathrm{d}x} \tag{46}$$

and called the Boltzmann distribution. The parameter T is called the temperature of the system and determined by (45). For a Euclidean energy measure, the Boltzmann distribution takes on the form of a Gaussian distribution which is well-known in information theory to maximize entropy for given average energy.

A helpful quantity in statistical mechanics is the (normalized) *free energy*[2] defined as

$$\mathrm{F}(X) \triangleq \mathrm{E}(X) - T\mathrm{H}(X) = -T \log \left(\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{T}||x||} \mathrm{d}x \right). \quad (47)$$

In the thermodynamic equilibrium, the entropy is maximized and the free energy is minimized since the energy is constant. The free energy normalized to the dimension of the system is a self averaging quantity.

5.3 The Meaning of the Energy Function

The free energy is clearly related in statistical mechanics to the entropy of the system at given energy due to (47). This establishes the usefulness of the free energy for information theoretic tasks like calculations of channel capacities. Moreover, the free energy is a tool to analyze various types of multiuser detectors. In fact, the free energy is such a powerful concept that it needs not any coding to be involved in the communication system to yield striking results. The only condition, it requires to be fulfilled, is the existence of macroscopic variables, microscopic random variables and the existence of an energy function. For communication systems, this requires, in practice, nothing more than their size growing above all bounds.

The broad applicability of the statistical mechanics approach to communication systems stems form the validity of (47) for any definition of the energy function. The energy function can be interpreted as the metric of a detector. Thus, any detector parameterized by a certain metric can be analyzed with the tools of statistical mechanics in the large system limit. There is no need that the performance measures of the detector depend only on the eigenvalues of the channel matrix in the large system limits. However, there is a practical limit to the applicability of the statistical mechanics framework to the analysis of large communication systems: The analytical calculations required to solve the equations arising from (47) are not always feasible. The replica method was introduced to circumvent such difficulties in certain cases. Many other cases, however, have remained intractable until present time.

Consider a communication channel uniquely characterized by a conditional probability density $\mathrm{p}_{Y|X}(y,x)$ and a source uniquely characterized by a prior density $\mathrm{p}_X(x)$. Consider a detector for the output of this channel characterized by an assumed channel transition probability $\check{\mathrm{p}}_{Y|X}(y,x)$ and an assumed prior distribution $\check{\mathrm{p}}_X(x)$. Let the detector minimize some kind of cost function, e.g. bit error probability, subject to its hypotheses on the channel transition probability $\check{\mathrm{p}}_{Y|X}(y,x)$ and the prior distribution $\check{\mathrm{p}}_X(x)$. If the assumed distributions equal the true distributions, the detector is optimum with respect to its cost function. If the assumed distributions differ from the true ones, the detector is mismatched in some sense.

The minimization of a cost function subject to some hypothesis on the channel transition probability and some hypothesis on the prior distribution defines a metric which is to be optimized. This metric corresponds to the energy function in thermodynamics and determines the distribution of the microscopic variables in the thermodynamic equilibrium. In analogy to (46), we find

$$\check{\mathrm{p}}_{X|Y}(x,y) = \frac{\mathrm{e}^{-\frac{1}{T}||x||}}{\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{T}||x||} \mathrm{d}x} \quad (48)$$

where the dependency on Y and the assumed distributions is implicit via the definition of the energy function $||\cdot||$. The energy function reflects the properties of the detector. Using Bayes' law, the appropriate energy function corresponding to particular hypotheses on the channel transition function and the prior distribution can be calculated via (48).

In order to study macroscopic properties of the system, we must calculate the free energy of the system. For that purpose, we make use of the self-averaging property of the thermodynamic equilibrium and (47):

$$\mathrm{F}(X) = \mathop{\mathsf{E}}_{Y} \mathrm{F}(X|Y) = -T \int \log \left(\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{T}||x||} \mathrm{d}x \right) \mathrm{dP}_Y(y) \quad (49)$$

Note that, inside the logarithm,the assumed distributions appear (implicitly) via the definition of the energy function, while, outside the logarithm, the integration is with respect to the true distribution.

In the case of matched detection, i.e. the assumed distributions equal the true distributions, the argument of the logarithm in (49) becomes $\mathrm{p}_Y(y)$ up to a normalizing factor. Thus, the free energy becomes the (differential) entropy of Y up to a scaling factor and an additive constant.

Statistical mechanics provides an excellent framework to study not only matched, but also mismatched detection. The analysis of mismatched detection based on asymptotic properties of large random matrices is difficult and has been very limited so far. One exception is the asymptotic SINR of linear MMSE detectors with erroneous assumptions on the powers of interfering users in [53].

5.4 Replica Continuity

The explicit evaluation of the free energy turns out to be very complicated in many cases of interest. One major obstacle is the occurrence of the expectation of the logarithm of some function $f(\cdot)$ of a random variable Y

$$\mathop{\mathsf{E}}_{Y} \log f(Y). \quad (50)$$

In order to circumvent this expectation which also appears frequently in information theory, the following identity is helpful

$$\log(Y) = \lim_{n \to 0} \frac{\partial}{\partial n} Y^n. \quad (51)$$

Under the assumption that limit and expectation can be interchanged, this gives

$$\mathop{\mathsf{E}}_{Y} \log f(Y) = \lim_{n \to 0} \frac{\partial}{\partial n} \log \mathop{\mathsf{E}}_{Y} [f(Y)]^n \quad (52)$$

and reduces the problem to the calculation of the n^{th} moment of the function of the random variable Y in the neighborhood of $n = 0$. Note that the expectation must be calculated for real-valued variables n in order to perform the limit operation.

At this point, it is customary to assume analytic continuity of the function $\mathop{\mathsf{E}}_{Y}[f(Y)]^n$. That is, the expectation is calculated for integer n only, but the resulting formula is trusted to hold for arbitrary real variables n in the neighborhood of $n = 0$. Note that analytic continuity is just an assumption. There is no mathematical theorem which states under which exact conditions this assumption is true or false. In fact, establishing a rigorous mathematical fundament for this step in the replica analysis is a topic of ongoing research.

Relying on the analytic continuity, let $f(Y) = \int f_Y(x)\,\mathrm{d}x$ for some function $f_Y(x)$. Since the variable of integration is arbitrary, this implies

$$[f(Y)]^n = \left(\int f_Y(x)\,\mathrm{d}x \right)^n = \prod_{a=1}^{n} \int f_Y(x_a)\,\mathrm{d}x_a. \quad (53)$$

Thus, instead of calculating the n^{th} power of $f(Y)$, replicas of x are generated. These replicated variables x_a are arbitrary and can be assigned helpful properties. Often they are assumed to be independent random variables.

[2] The *free* energy is not related to *free*ness in free probability theory.

In general, it is not easier to calculate the expectation of the right hand side of (53) than just the expectation over $[f(Y)]^n$. However, there are some functions $f_Y(x)$ for which the replica method is indeed advantageous, particularly if there is no closed form solution for $\int f_Y(x)\,\mathrm{d}x$, thus, $f(Y)$ cannot be given explicitly. Then, it might help to substitute the set of variables $(x_1,\ldots,x_n,y)$ by some other variables which allow to solve the integral.

5.5 Replica Symmetry

Typically, integrals arising from the replica ansatz are solved by saddle point integration. The general idea of saddle point integration is as follows: Consider an integral of the form

$$\frac{1}{K}\log\int \mathrm{e}^{Kf(x_1,x_2)}\,\mathrm{d}x_1\,\mathrm{d}x_2. \tag{54}$$

In the limit $K\to\infty$ the integral is dominated by that values x_1,x_2 which maximize the function $f(x_1,x_2)$. Thus, we have

$$\lim_{K\to\infty}\frac{1}{K}\log\int \mathrm{e}^{Kf(x_1,x_2)}\,\mathrm{d}x_1\,\mathrm{d}x_2 = \max_{x_1,x_2} f(x_1,x_2). \tag{55}$$

That means, the integral can be solved taking the derivative of the argument of the exponential function.

If the function in the exponent is multivariate—typically all replicated random variables are arguments—one would need to find the extremum of a multivariate function for an arbitrary number of arguments. This can easily become a hopeless task, unless one can exploit some properties of the exponential argument.

Assuming *replica symmetry* means that one concludes from the symmetry of the exponent, e.g. $f(x_1,x_2)=f(x_2,x_1)$ for the bivariate case, that the extremum appears if all variables take on the same value. Then, the multivariate optimization problem reduces to a single variate one, e.g.

$$\max_{x_1,x_2} f(x_1,x_2) = \max_{x} f(x,x) \tag{56}$$

for the bi-variate case. This is the most critical assumption when applying the replica method. It is not always true, even in practically relevant cases. The general way to circumvent this trouble is to assume replica symmetry at hand and proof later, having found a replica symmetric solution, that it is correct.

There are also practically relevant cases without replica symmetric solutions. Such phenomena are labeled *replica symmetry breaking* and a rich theory in statistical mechanics literature exists to deal with them [54, 44]. For the introductory character of this work, however, replica symmetry breaking is a to advanced issue.

5.6 Phase Transitions

In thermodynamics, the occurrence of phase transitions, i.e. melting ice becomes water, is a well-known phenomenon. In digital communications, however, such phenomena are less known, though they do occur. Phase transitions in turbo decoding and detection of CDMA were found in [55] and [45], respectively.

Phase transitions in digital communications are similar to the hysteresis in magnetic materials. They occur if the equations determining the macroscopic parameters have multiple solutions. Then, it is the free energy to decide which of the solution corresponds to the thermodynamic equilibrium. If a system parameter, e.g. the load or the noise variance, changes, the free energy may shift its favor from the present to another solution. Since each solution corresponds to a different macroscopic behavior of the system, changing the valid solution means that a phase transition takes place.

In digital communications, a popular macroscopic property is the bit error probability. As an example, the bit error probability of a CDMA system with random spreading is depicted in Fig. 1. The thick curve shows the bit error probability of the individually optimum detector as a function of the load. The thin curves show alternative solutions for the bit error probability corresponding to

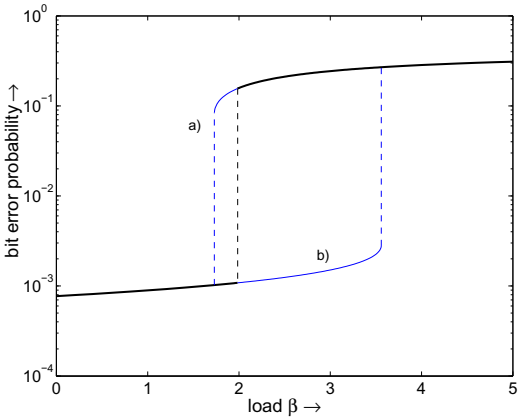

Figure 1: Bit error probability for the individually optimum detector with uniform binary prior distribution versus system load for $10\log_{10}(E_s/N_0)=6$ dB.

alternative solutions to the equations for the macroscopic variables. For a certain interval of the load, approximately $1.73\le\beta\le 3.56$ in Fig. 1, multiple solutions occur. The bit error probability increases with the load. At a load of approximately $\beta=1.986$ a phase transition occurs and lets the bit error probability jump. Unlike to ferromagnetic materials, there is no hysteresis effect for the bit error probability of the individually optimum detector, but only a phase transition.

Acknowledgment

The author would like to thank T. Tanaka, S. Verdú, G. Caire, S. Hanly, and L. Cottatellucci.

REFERENCES

[1] S. Verdú, *Multiuser Detection*, Cambridge University Press, New York, 1998.

[2] J. G. Proakis, *Digital Communications*, McGraw–Hill, New York, 4th ed., 2000.

[3] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw–Hill, New York, 3rd ed., 1991.

[4] P. Hilton and J. Pedersen, "Catalan numbers, their generalization, and their uses," *The Math. Intelligencer*, vol. 13, no. 2, pp. 64–75, 1991.

[5] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.

[6] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1293–1302, July 1993.

[7] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*, American Math. Soc., Providence, RI, 2000.

[8] J.W̃. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *J. Multivariate Analysis*, vol. 55, pp. 331–339, 1995.

[9] Y. Q. Yin, "Limiting spectral distribution for a class of random matrices," *J. Multivariate Analysis*, vol. 20, pp. 50–68, 1986.

[10] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *J. Multivariate Analysis*, vol. 54, pp. 175–192, 1995.

[11] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Mathematics of the U.S.S.R — Sbornik*, vol. 1, no. 4, pp. 457–483, 1967.

[12] D. Tse and S. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 641–657, Mar. 1999.

[13] S. Shamai (Shitz) and S. Verdú, "The impact of frequency–flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1302–1327, May 2001.

[14] D. Guo, S. Verdú, and L. K. Rasmussen, "Asymptotic normality of linear multiuser receiver outputs," *IEEE Trans. Inform. Theory*, vol. 48, no. 12, pp. 3080–3095, Dec. 2002.

[15] J. Zhang, E. K. P. Chong, and D. N. C. Tse, "Output MAI distributions of linear MMSE multiuser receivers in DS–CDMA systems," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 1128–1144, Mar. 2001.

[16] M. Honig and W. Xiao, "Performance of reduced–rank linear interference suppression," *IEEE Trans. Inform. Theory*, vol. 47, no. 5, pp. 1928–1946, July 2001.

[17] W. Phoel and M. L. Honig, "Performance of coded DS-CDMA with pilot–assisted channel estimation and linear interference suppression," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 822–832, May 2002.

[18] R. R. Müller, "On the asymptotic eigenvalue distribution of concatenated vector–valued fading channels," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 2086–2091, July 2002.

[19] J. Zhang and E. K. P. Chong, "Linear MMSE multiuser receivers: MAI conditional weak convergence and network capacity," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 2114–2122, July 2002.

[20] J. Boutros and G. Caire, "Iterative multiuser joint decoding: Unified framework and asymptotic analysis," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1772–1793, July 2002.

[21] J. Zhang and X. Wang, "Large-system performance analysis of blind and group-blind multiuser receivers," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2507–2523, Sept. 2002.

[22] H. Dai and H. V. Poor, "Asymptotic spectral efficiency of multicell MIMO systems with frequency-flat fading," *IEEE Trans. Sign. Proc.*, vol. 51, no. 11, pp. 2976–2988, Nov. 2003.

[23] V. L. Girko, *Theory of Random Determinants*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990.

[24] Kiran and D. N. C. Tse, "Effective interference and effective bandwidth of linear multiuser receivers in asynchronous cdma systems," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1426–1447, Sept. 2000.

[25] S. V. Hanly and D. N. C. Tse, "Resource pooling and effective bandwidth in CDMA networks with multiuser receivers and spatial diversity," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1328–1351, May 2001.

[26] C.-N. Chuah, D. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in dual–antenna–array wireless systems," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[27] J. W. Silverstein, "Eigenvalues and eigenvectors of large dimensional sample covariance matrices," in *Random Matrices and Their Applications*, J. E. Cohen, H. Kesten, and C. M. Newman, Eds., pp. 153–159. American Math. Soc., Providence, RI, 1986.

[28] D. Tse and O. Zeitouni, "Linear multiuser receivers in random environments," *IEEE Trans. Inform. Theory*, vol. 46, no. 1, pp. 171–188, Jan. 2000.

[29] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free Random Variables*, American Math. Soc., Providence, RI, 1992.

[30] R. Speicher, "Freie Wahrscheinlichkeitstheorie," Lecture Notes, 1997/98.

[31] D. Voiculescu, "Lectures on free probability theory," Notes for a Course at the Saint-Flour Summer School on Probability Theory 1998, 1998.

[32] D. Voiculescu, "Limit laws for random matrices and free products," *Inventiones Mathematicae*, vol. 104, pp. 201–220, 1991.

[33] S. Thorbjørnsen, "Mixed moments of Voiculescu's Gaussian random matrices," *J. of Functional Analysis*, vol. 176, pp. 213–246, 2000.

[34] J. Evans and D. N. C. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *IEEE Trans. Inform. Th.*, vol. 46, no. 6, pp. 2059–2078, Sept. 2000.

[35] E. Biglieri, G. Taricco, and A. Tulino, "Performance of space–time codes for a large number of antennas," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1794–1803, July 2002.

[36] R. R. Müller, "A random matrix model for communication via antenna arrays," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2495–2506, Sept. 2002.

[37] L. Cottatellucci and R. R. Müller, "Multiuser interference mitigation with multistage detectors: Design and analysis for unequal powers," in *Proc. of Asilomar Conf.*, Pacific Grove, CA, Nov. 2002.

[38] L. Cottatellucci and R. R. Müller, "Asymptotic analysis of multistage receivers for multipath fading channels," in *Proc. of IEEE ISIT*, Yokohama, Japan, June/July 2003.

[39] L. Cottatellucci and R. R. Müller, "A systematic approach to multistage detectors in multipath fading channels," Submitted to *IEEE Trans. Inform. Theory*, Feb. 2003.

[40] L. Li, A. Tulino, and S. Verdú, "Asymptotic eigenvalue moments for linear multiuser detection," *Communications in Inform. and Systems*, vol. 1, no. 3, pp. 273–304, Fall 2001.

[41] L. Li, A. M. Tulino, and S. Verdú, "Asymptotic eigenvalue moments for linear multiuser detection," in *Proc. of Asilomar Conf.*, Pacific Grove, CA, Nov. 2001, pp. 1177–1181.

[42] M. Debbah, W. Hachem, P. Loubaton, and M. de Courville, "MMSE analysis of certain large isometric random precoded systems," *IEEE Trans. Inform. Th.*, vol. 49, no. 5, May 2003.

[43] D. Tse, "Multiuser receivers, random matrices and free probability," in *Proc. of 37th Ann. Allerton Conf.*, Monticello, IL, Sept. 1999.

[44] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing*, Oxford Univ. Press, Oxford, U.K., 2001.

[45] T. Tanaka, "A statistical mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inform. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.

[46] 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. Kluwer Academic Publishers, 2002.

[47] R. R. Müller, G. Caire, and T. Tanaka, "Density evolution and power profile optimization for iterative multiuser decoders based on individually optimum multiuser detectors," in *Proc. of 40th Ann. Allerton Conf.*, Monticello, IL, Oct. 2002.

[48] R. R. Müller and W. Gerstacker, "On the capacity loss due to separation of detection and decoding in large CDMA systems," in *Proc. of IEEE ITW*, Bangalore, India, Oct. 2002, p. 222.

[49] D. Guo and S. Verdú, "Spectral efficiency of large-system CDMA via statistical physics," in *Proc. of CISS*, Baltimore, MD, Mar. 2003.

[50] D. Guo and S. Verdú, "Decoupling of CDMA multiuser detection via the replica method," in *Proc. of 41st Ann. Allerton Conf.*, Monticello, IL, Oct. 2003.

[51] T. Tanaka and D. Saad, "A statistical-mechanics analysis of coded CDMA with regular LDPC codes," in *Proc. of IEEE ISIT*, Yokohama, Japan, June/July 2003, p. 444.

[52] Y. Kabashima, "A statistical mechanical approach to CDMA multiuser detection: propagating beliefs in a dense graph," in *Proc. of IEEE ISIT*, Yokohama, Japan, June/July 2003, p. 329.

[53] R. R. Müller and G. Caire, "Efficient implementation of iterative multiuser decoding," in *Proc. of IEEE ISIT*, Lausanne, Switzerland, June/July 2002, p. 446.

[54] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond*, World Scientific, Singapore, 1987.

[55] D. Agrawal and A. Vardy, "The turbo decoding algorithm and its phase trajectories," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 699–722, Feb. 2001.