

Applications of Large Random Matrices in Communications Engineering

Ralf R. Müller

Abstract—This work 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 $y = Hx + z$ where the number of components in each vector is large. Tools from probability theory, operator algebra, and statistical physics are reviewed. The survey of analytical tools is complemented by examples of applications in communications engineering.

Asymptotic eigenvalue distributions of many classes of random matrices are given. The treatment includes the problem of moments and the introduction of the Stieltjes transform.

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, the S-transform, and the free central limit theorem.

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.

These analytic tools are applied to the design and the analysis of linear multiuser detectors, the modelling of scattering in communication channels with dual antennas arrays, and the analysis of optimal detection for communication via code-division multiple-access and/or dual antenna array channels.

I. 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. This situation is almost as old as mankind and appears whenever more than two people start discussing with each other. 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 con-

trol 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 a theory 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 should occur. Indeed recent research has found theoretical and numerical evidence for the occurrence of phase transitions in large code-division multiple-access (CDMA) systems, cf. Section V, which correspond to the hysteresis behavior of magnetic materials.

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. Nowadays, it has become a well-established method in quantum physics to describe the energy-levels of heavy nuclei by the eigenvalue distributions of large dimensional random matrices [1]. Random matrices and their applications in communications engineering are discussed in Section III.

The interest in *free probability theory* was driven by mathematicians in the beginning. Particularly, those ones who were interested in operator algebras and the Riemann hypothesis. However, 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 became interested in the subject hoping 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. A detailed discussion on free probability and its applications in communications engineering is given in Section IV.

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 many 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 a powerful tool to analyze the thermodynamics

of magnetic materials 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. A detailed discussion of the replica method is given in Section V.

Mathematical formulas are used to simplify if variables approach either zero or infinity. This work aims to show that also engineering can get easier as system size becomes either large or small. Decades ago communication engineers already made use of this effect summarizing all kinds of additive distortion in the universal additive Gaussian noise. The author is confident that Gaussian noise was just the beginning and that the semicircle distribution and the R-transform will become tools in multi-user communications which are just as common as the Gaussian distribution and the Fourier transform are in single user digital communications.

II. COMMUNICATION SYSTEMS

The principles summarized in this thesis apply to a broad class of communication channels. We do not aim to cover all or even most of them, but restrict ourselves to the discrete vector-valued additive white Gaussian noise channel. It is general enough to develop rich examples for the application of the theory to be introduced, and simple enough to keep equations illustrating.

In vector notation the vector-valued additive white Gaussian noise channel is given by

$$\mathbf{y}[\nu] = \mathbf{H}[\nu]\mathbf{x}[\nu] + \mathbf{n}[\nu] \quad (1)$$

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 $\mathbf{n}[\nu]$ with zero mean ($\mathbf{E} \mathbf{n} = \mathbf{0}$) and variance ($\mathbf{E} \mathbf{n} \mathbf{n}^H = \sigma_0^2 \mathbf{I}$),
- and discrete time ν .

In order to simplify notation, 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 [2] that the signal

$$\mathbf{r}[\nu] = \mathbf{H}^H[\nu]\mathbf{y}[\nu] \quad (2)$$

$$= \mathbf{H}^H[\nu]\mathbf{H}[\nu]\mathbf{x}[\nu] + \mathbf{H}^H[\nu]\mathbf{n}[\nu] \quad (3)$$

provides sufficient statistics for the estimation of the signal $\mathbf{x}[\nu]$. This means that all information about $\mathbf{x}[\nu]$ that could be extracted from the received signal $\mathbf{y}[\nu]$ can also be extracted from the signal $\mathbf{r}[\nu]$. Thus, the two channels (1) and (2) are actually equivalent in terms of all performance measures such as bit error rate, signal-to-noise ratio, channel capacity, etc.

These 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 $\mathbf{H}^H \mathbf{H}$ 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 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 $\mathbf{H}^H \mathbf{H}$ accounts for inter-carrier interference.

Regardless of the application one has in mind, the performance of digital communication via the channel (1) 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 [2], [3] and we do not try to make any effort to be comprehensive in this thesis. In the following, we summarize only those of the known results which are needed in subsequent sections.

A. Channel Capacity

Channel capacity is the maximum data rate which can be transmitted reliably over a channel. For the channel (1) and many other channels with continuous input and output alphabet, channel capacity is infinite unless a constraint on the input alphabet is imposed. There are many ways to definite such constraints. The most widely used ones are upper bounds on the moments of the amplitude of the input signal, particularly on the second moment

$$\mathbf{E}_{\nu} \mathbf{x}^H \mathbf{x} \leq KP \quad (4)$$

which is actually an average power constraint. Note that P is the average power per component of \mathbf{x} .

Channel capacity crucially depends on the availability of channel state information to either receiver or transmitter or both. For the additive white Gaussian noise channel (1), perfect channel state information means full knowledge about \mathbf{H} . Assuming perfect channel state information at the receiver, but no channel state information at the transmitter, channel capacity of (1) is given by [4]

$$C = \mathbf{E}_{\mathbf{H}} \log \det \left(\mathbf{I} + \frac{P}{\sigma_0^2} \mathbf{H}^H \mathbf{H} \right). \quad (5)$$

With the eigenvalue decomposition

$$\mathbf{H}^H \mathbf{H} = \mathbf{U}^H \mathbf{A} \mathbf{U} \quad (6)$$

where \mathbf{U} is a unitary matrix containing the eigenvectors of $\mathbf{H}^H \mathbf{H}$ and $\mathbf{A} = \text{diag}(\lambda_1, \dots, \lambda_K)$, channel capacity can also be written as

$$C = \mathbf{E}_{\mathbf{A}} \log \det \left(\mathbf{I} + \frac{P}{\sigma_0^2} \mathbf{A} \right) \quad (7)$$

$$= \sum_{k=1}^K \mathbb{E}_{\lambda_k} \log \left(1 + \frac{P}{\sigma_0^2} \lambda_k \right) \quad (8)$$

$$= K \mathbb{E}_{\lambda} \log \left(1 + \frac{P}{\sigma_0^2} \lambda \right) \quad (9)$$

where λ denotes a randomly chosen eigenvalue of $\mathbf{H}^H \mathbf{H}$. Therefore, channel capacity is determined by the eigenvalues of the channel, the power constraint, and the noise variance.

B. Uncoded Error Probability

The purpose of a detector is to decide for a transmitted signal $\hat{\mathbf{x}}$ given the observation of the received signal \mathbf{y} (or equivalently the sufficient statistics \mathbf{r}). The detector is the better, the larger the overlap between the estimated signals and the signals actually sent.

In case the estimated signal and the true signal do not fully overlap, an error occurs. Two kinds of errors are distinguished:

Symbol errors occur if the estimated vector differs from the transmitted vector, i.e. $\hat{\mathbf{x}} \neq \mathbf{x}$.

Bit errors occur if one component of the estimated vector differs from the respective component of the transmitted vector, i.e. $\hat{x}_k \neq x_k$.

In literature [2], the detectors minimizing symbol and bit error probability and are called *jointly* and *individually* optimum detectors, respectively. The only difference between the cost functions of the two optimization criteria is that the individually optimum detector counts multiple differences within the components of the same symbol, while the jointly optimum detector does not do so.

Both detectors are special cases of the *marginal posterior mode* detector. The marginal posterior mode detector tries to individually minimize bit error probability under the, in general, false assumption that the variance of the noise is σ^2 instead of σ_0^2 . The assumed noise variance σ^2 is a control parameter of the marginal posterior mode detector. If the assumed noise variance is the true noise variance, i.e. $\sigma = \sigma_0$, it becomes the individually optimum detector. If the assumed noise variance tends to zero, i.e. $\sigma \rightarrow 0$, it becomes the jointly optimum detector [2].

The distribution of the received signal \mathbf{y} is Gaussian conditioned on the transmitted signal \mathbf{x} and the channel realization \mathbf{H} :

$$p_{\mathbf{y}|\mathbf{x},\mathbf{H}}(\mathbf{y}, \mathbf{x}, \mathbf{H}) = \frac{e^{-\sigma_0^{-2}(\mathbf{y}-\mathbf{H}\mathbf{x})^H(\mathbf{y}-\mathbf{H}\mathbf{x})}}{(\pi\sigma_0)^K}. \quad (10)$$

By Bayes' law, we find

$$p_{\mathbf{x}|\mathbf{y},\mathbf{H}}(\mathbf{x}, \mathbf{y}, \mathbf{H}) = \frac{p_{\mathbf{y}|\mathbf{x},\mathbf{H}}(\mathbf{y}, \mathbf{x}, \mathbf{H}) p_{\mathbf{x}|\mathbf{H}}(\mathbf{x}, \mathbf{H})}{p_{\mathbf{y}|\mathbf{H}}(\mathbf{y}, \mathbf{H})}. \quad (11)$$

Thus, the marginal posterior mode detector with perfect channel state information at receiver side and parameter σ is given by

$$\hat{x}_k(\sigma) = \underset{\xi}{\operatorname{argmax}} \sum_{\mathbf{x}:x_k=\xi} e^{-\sigma^{-2}(\mathbf{y}-\mathbf{H}\mathbf{x})^H(\mathbf{y}-\mathbf{H}\mathbf{x})} p_{\mathbf{x}|\mathbf{H}}(\mathbf{x}, \mathbf{H}). \quad (12)$$

For general matrices \mathbf{H} , the optimization problem (12) cannot be solved with polynomial complexity in K [2]. Bit error probability is given by

$$P_b(\sigma) = \frac{1}{K} \sum_{k=1}^K \Pr(\hat{x}_k(\sigma) \neq x_k). \quad (13)$$

Unlike channel capacity bit error probability does also depend on the eigenvectors of the channel.

C. Linear Detection

Since the complexity of the optimum detectors is exponential in K , it is often infeasible in practice. In order to circumvent this obstacle, suboptimum approaches were proposed, see e.g. [5], [2] for a survey of literature. The most common family of suboptimum detectors are the linear detectors. They implement the detection rule

$$\hat{\mathbf{x}} = \operatorname{quant}_{\mathcal{X}}(\mathbf{L}\mathbf{y}) \quad (14)$$

where \mathbf{L} is a $K \times N$ matrix whose construction characterizes the linear detector and the function $\operatorname{quant}_{\mathcal{X}}(\cdot)$ quantizes a vector component-wise towards the nearest element of the symbol alphabet \mathcal{X} . For binary antipodal transmission, it is the sign function.

The best linear detector in terms of bit error probability is very difficult to construct [2]. Minimizing the mean-squared error

$$\mathbb{E}_{\mathbf{x},\mathbf{n}|\mathbf{H}} (\mathbf{x} - \mathbf{L}\mathbf{y})^H (\mathbf{x} - \mathbf{L}\mathbf{y}) \quad (15)$$

instead results in a linear detector performing also very well. This detector results from applying the optimum decision rule under the, in general, false assumption that the symbol alphabet is Gaussian distributed with true mean and true variance. It can be expressed explicitly in terms of the channel matrix, the noise variance, and the signal power as [2]

$$\hat{\mathbf{x}} = \operatorname{quant}_{\mathcal{X}} \left(\sqrt{P} \left(\sigma_0^2 \mathbf{I} + P\mathbf{H}^H \mathbf{H} \right)^{-1} \mathbf{H}^H \mathbf{y} \right). \quad (16)$$

The detector (16) also maximizes the ratio between useful signal power P and interference and noise power. For the k^{th} component of the vector \mathbf{x} , it is given by [2]

$$\text{SINR}_k = P \mathbf{h}_k^H \left(\sigma_0^2 \mathbf{I} + P\mathbf{H}^H \mathbf{H} - P \mathbf{h}_k \mathbf{h}_k^H \right)^{-1} \mathbf{h}_k. \quad (17)$$

where \mathbf{h}_k denotes the k^{th} column of \mathbf{H} . With the singular value decomposition

$$\mathbf{H} = \mathbf{V}^H \sqrt{\Lambda} \mathbf{U} \quad (18)$$

and the matrix inversion lemma, the SINR can also be written as

$$\begin{aligned} \text{SINR}_k &= P \mathbf{u}_k^H \left(\sigma_0^2 \Lambda^{-1} + P \mathbf{I} - P \mathbf{u}_k \mathbf{u}_k^H \right)^{-1} \mathbf{u}_k \quad (19) \\ &= \frac{1}{1 - P \mathbf{u}_k^H \left(\sigma_0^2 \Lambda^{-1} + P \mathbf{I} \right)^{-1} \mathbf{u}_k} - 1. \quad (20) \end{aligned}$$

Note that the SINR depends on all eigenvalues of the matrix $\mathbf{H}^H \mathbf{H}$, but only on the k^{th} column of the eigenvector matrix \mathbf{U} . Note, however, that due to definition (6), the eigenvectors are not the columns but the rows of \mathbf{U} .

III. RANDOM MATRIX THEORY

The channel matrix \mathbf{H} introduced 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

$$\mathbf{H} = \begin{bmatrix} H_1[0] & H_1[1] & H_1[2] & \cdots \\ H_2[0] & H_2[1] & H_2[2] & \cdots \\ H_3[0] & H_3[1] & H_3[2] & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (21)$$

If we let the dimension $N, K \rightarrow \infty$, 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 $\text{rowsum} : \mathbf{X} \in \mathbb{C}^{N \times K} \mapsto K^{-\frac{1}{2}} \mathbf{X} \mathbf{1} \in \mathbb{C}^N$ with $\mathbf{1}$ denoting the all one vector. It simply sums up the rows of its argument and normalizes the result. As $K \rightarrow \infty$, the value of this function

$$\mathbf{h} = \text{rowsum}(\mathbf{H}) \quad (22)$$

is an N -dimensional Gaussian random vector due to the central limit theorem. As $N \rightarrow \infty$, the empirical distribution function of its components h_i

$$P_{\mathbf{h}}(x) = \frac{1}{N} |\{h_i : h_i < x\}| \quad (23)$$

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 $N \rightarrow \infty$. 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 $\text{rowsum}(\cdot)$, the projection follows a Gaussian distribution. For non-linear mappings, however, a rich plurality of other limit distributions occurs.

A. Convergence of Eigenvalues

The eigenvalues of random matrices were found to be particularly important to characterize some performance measures in communications engineering, cf. Section II. Calculating the eigenvalues of (a function of) a random matrix, is a projection similar to the rowsum function in (22). The main difference between the two is the, in general, non-linear nature of the projection.

Calculating the eigenvalues λ_k of a matrix \mathbf{H} is a non-linear operation, in general. However, calculation of the moments of the eigenvalue distribution is conveniently done by a normalized trace since

$$\frac{1}{N} \sum_{k=1}^N \lambda_k^m = \text{tr}(\mathbf{H}^m) \quad (24)$$

with

$$\text{tr}(\mathbf{H}) \triangleq \frac{1}{N} \text{trace}(\mathbf{H}). \quad (25)$$

In the following, we also use

$$\text{Tr}(\mathbf{H}) \triangleq \lim_{N \rightarrow \infty} \text{tr}(\mathbf{H}). \quad (26)$$

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.

A.1 Full 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 \mathcal{L} denote the set containing the eigenvalues of \mathbf{H} . Then, the empirical distribution of the eigenvalues

$$P_{\mathbf{H}}(z) = \frac{1}{N} |\{\lambda \in \mathcal{L} : \Re \lambda < \Re z \wedge \Im \lambda < \Im z\}| \quad (27)$$

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

$$p_{\mathbf{H}}(z) = \begin{cases} \frac{1}{\pi} & |z| < 1 \\ 0 & \text{elsewhere} \end{cases}. \quad (28)$$

The eigenvalues are uniformly distributed within the complex unit circle.

A.2 Semi-Circle Law

Let the random matrix \mathbf{H} fulfill the same conditions as needed for the full circle law. Let

$$\mathbf{K} = \frac{\mathbf{H} + \mathbf{H}^H}{\sqrt{2}}. \quad (29)$$

Then, the empirical distribution of the eigenvalues of \mathbf{K} converge to a non-random distribution function as $N \rightarrow \infty$ whose density is given by

$$p_{\mathbf{K}}(x) = \begin{cases} \frac{1}{2\pi} \sqrt{4 - x^2} & |x| < 2 \\ 0 & \text{elsewhere} \end{cases}. \quad (30)$$

The semicircle distribution plays a central role in free probability theory, cf. Section IV, where it serves as the equivalent to the Gaussian distribution in classical probability theory. Its even moments are the *Catalan* numbers C_n

$$\text{Tr}(\mathbf{K}^m) = \frac{1}{2\pi} \int_{-2}^{+2} x^m \sqrt{4 - x^2} dx \quad (31)$$

$$= C_{m/2} \quad \forall m \text{ even} \quad (32)$$

$$\triangleq \frac{1}{m/2 + 1} \binom{m}{m/2} \quad (33)$$

which play a crucial role in combinatorics, see [6] for further reading. Since the semicircular density is symmetric, the odd moments vanish.

A.3 Quarter Circle Law

Let the random matrix \mathbf{H} fulfill the same conditions as needed for the full circle law. Then, the empirical distribution of the singular values of \mathbf{H} , i.e. the eigenvalues of

$$\mathbf{Q} = \sqrt{\mathbf{H}\mathbf{H}^H}, \quad (34)$$

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

$$p_{\mathbf{Q}}(x) = \begin{cases} \frac{1}{\pi} \sqrt{4-x^2} & 0 \leq x < 2 \\ 0 & \text{elsewhere} \end{cases}. \quad (35)$$

This distribution is called the quarter circle distribution. Obviously, its even moments are identical to those of the semicircular distribution. However, its odd moments do not vanish. They are given by

$$\text{Tr}(\mathbf{Q}^m) = \frac{1}{\pi} \int_0^2 x^m \sqrt{4-x^2} dx \quad (36)$$

$$= \frac{2^{2m}}{\pi m \left(\frac{m}{2} + 1\right) \binom{m-1}{\frac{m-1}{2}}} \quad \forall m \text{ odd.} \quad (37)$$

With standard methods for the transformation of probability densities, see [7], the asymptotic eigenvalue distribution of $\mathbf{Q}^2 = \mathbf{H}\mathbf{H}^H$ can be derived. It reads

$$p_{\mathbf{Q}^2}(x) = \begin{cases} \frac{1}{2\pi} \sqrt{\frac{4-x}{x}} & 0 < x < 4 \\ 0 & \text{elsewhere} \end{cases}. \quad (38)$$

Its m^{th} moments (even and odd) are the Catalan numbers C_m . This distribution is a special case of the Marčenko-Pastur distribution which also plays a central role in free probability theory.

A.4 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

$$\mathbf{R} = \sqrt{\mathbf{H}\mathbf{H}^H} \quad (39)$$

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

$$p_{\mathbf{R}}(x) = \begin{cases} \frac{\sqrt{4\beta - (x^2 - 1 - \beta)^2}}{\pi x} & |1 - \sqrt{\beta}| < x < 1 + \sqrt{\beta} \\ [1 - \beta]^+ \delta(x) & \text{elsewhere} \end{cases}. \quad (40)$$

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

$$p_{\mathbf{R}^2}(x) = \begin{cases} \frac{\sqrt{4\beta - (x-1-\beta)^2}}{2\pi x} & (1 - \sqrt{\beta})^2 < x < (1 + \sqrt{\beta})^2 \\ [1 - \beta]^+ \delta(x) & \text{elsewhere} \end{cases}. \quad (41)$$

This distribution is known as the Marčenko-Pastur distribution and its moments are given by

$$\text{Tr}(\mathbf{R}^{2m}) = \frac{1}{m} \sum_{i=1}^m \binom{m}{i} \binom{m}{i-1} \beta^i. \quad (42)$$

It has been used in [8] to calculate channel capacity (5) for CDMA with independent identically distributed random spreading.

A.5 Haar Distribution

Let the random matrix \mathbf{H} be square $N \times N$ with independent identically complex Gaussian distributed proper¹ entries with zero mean and finite positive variance. Then, the empirical distribution of the eigenvalues of the unitary random matrix

$$\mathbf{T} = \mathbf{H} \left(\mathbf{H}^H \mathbf{H} \right)^{-\frac{1}{2}} \quad (43)$$

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

$$p_{\mathbf{T}}(z) = \frac{1}{2\pi} \delta(|z| - 1). \quad (44)$$

Obviously, all of its moments $\mathbf{E}|z|^m = 1$ equal unity.

It is obvious that all eigenvalues lie on the complex unit circle, since the matrix \mathbf{T} is unitary. The essential statement of (44) is that the eigenvalues are uniformly distributed.

Note that unlike the circle laws in previous subsections, (44) demands for Gaussian distributed entries in the random matrix \mathbf{H} . This condition does not seem to be necessary, but allowing for any complex distribution with zero mean and finite variance is not sufficient [10].

A.6 Inverse Semi-Circle Law

Let the random matrix \mathbf{T} fulfill the same conditions and be defined in the same way as in Section III-A.5. Let

$$\mathbf{Y} = \mathbf{T} + \mathbf{T}^H. \quad (45)$$

Then, the empirical distribution of the eigenvalues of \mathbf{Y} converge to a non-random distribution function as $N \rightarrow \infty$ whose density is given by

$$p_{\mathbf{Y}}(x) = \begin{cases} \frac{1}{\pi} \frac{1}{\sqrt{4-x^2}} & |x| < 2 \\ 0 & \text{elsewhere} \end{cases}. \quad (46)$$

B. Stieltjes Transform

There are only few kinds of random matrices for which the corresponding asymptotic eigenvalue distributions are known explicitly. For a wider class of random matrices, however, explicit calculation of the moments turned out to be feasible, see Section IV-F for an example.

The task of finding an unknown probability distribution given its moments is known as the *problem of moments*. It

¹A complex random variable is said to be *proper* if real and imaginary part are independent and identically distributed [9].

was addressed by Stieltjes in 1894 [11] using the integral transform

$$G(s) \triangleq \int \frac{dP(x)}{x-s} \quad (47)$$

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

A simple Taylor series expansion of the kernel of the Stieltjes transform

$$-\lim_{s \rightarrow 0} \frac{d^m}{ds^m} \frac{G(s^{-1})}{s} = m! \int x^m dP(x) \quad (48)$$

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 simply taking the limit

$$p(x) = \lim_{y \rightarrow 0^+} \frac{1}{\pi} \Im G(x + jy) \quad (49)$$

which is often called the *Stieltjes inversion formula* [12].

B.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 $\mathbf{X} = \mathbf{X}^H$ 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 $P_{\mathbf{X}}(x)$ as $N \rightarrow \infty$. Then, almost surely, the eigenvalue distribution of the matrix product

$$\mathbf{P} = \mathbf{H}\mathbf{H}^H \mathbf{X} \quad (50)$$

converges in distribution, as $K, N \rightarrow \infty$, but $\beta = K/N$ fixed, to a nonrandom distribution function whose Stieltjes transform satisfies

$$G_{\mathbf{P}}(s) = \int \frac{dP_{\mathbf{X}}(x)}{x(1 - \beta - \beta s G_{\mathbf{P}}(s)) - s} \quad (51)$$

for $\Im s > 0$.

This result was proven in its present form by Silverstein [13]. Under less general conditions on the statistics of \mathbf{H} and \mathbf{X} , it can be found in the earlier work of Yin [14].

B.2 Sums of Random Matrices

Let the random matrix \mathbf{H} fulfill the same conditions as needed for the deformed quarter circle law. Let $\mathbf{X} = \mathbf{X}^H$ be an $N \times N$ Hermitian matrix with an eigenvalue distribution function converging weakly to $P_{\mathbf{X}}(x)$ almost surely. Let $\mathbf{Y} = \text{diag}(y_1, \dots, y_K)$ be a $K \times K$ diagonal matrix and the empirical distribution function of $\{y_1, \dots, y_K\} \in \mathbb{R}^K$ converge almost surely in distribution to a probability distribution function $P_{\mathbf{Y}}(x)$ as $K \rightarrow \infty$. Moreover, let the matrices $\mathbf{H}, \mathbf{X}, \mathbf{Y}$ be jointly independent. Then, almost surely, the empirical eigenvalue distribution of the random matrix

$$\mathbf{S} = \mathbf{X} + \mathbf{H}\mathbf{Y}\mathbf{H}^H \quad (52)$$

converges weakly, as $K, N \rightarrow \infty$, but $\beta = K/N$ fixed, to a nonrandom distribution function whose Stieltjes transform satisfies

$$G_{\mathbf{S}}(s) = G_{\mathbf{X}} \left(s - \beta \int \frac{y dP_{\mathbf{Y}}(y)}{1 + y G_{\mathbf{S}}(s)} \right) \quad (53)$$

for $\Im s > 0$.

This result was proven in its present form by Silverstein and Bai [15]. Under less general conditions on the statistics of \mathbf{H} and \mathbf{X} , it can be found in the earlier work of Marčenko and Pastur [16]. It was used by Tse and Hanly [17] to derive asymptotic results for the SINR of linear multiuser receivers. Subsequently, it was used by Shamai and Verdú [18] to derive the capacity of the flat fading Gaussian CDMA channel with several types of receivers.

B.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

$$\frac{i}{N} \leq x \leq \frac{i+1}{N} \quad \text{and} \quad \frac{j}{N} \leq y \leq \frac{j+1}{N} \quad (54)$$

converges to a bounded joint limit distribution $w(x, y)$ as $K = \beta N \rightarrow \infty$. Then, for each $a, b \in [0, 1], a < b$, and $\Im(s) > 0$

$$\frac{1}{N} \sum_{i=[aN]}^{[bN]} \left(\mathbf{H}\mathbf{H}^H - s\mathbf{I} \right)_{ii}^{-1} \rightarrow \int_a^b u(x, s) dx \quad (55)$$

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

$$u(x, s) = \left[-s + \int_0^{\beta} \frac{w(x, y) dy}{1 + \int_0^1 u(x', s) w(x', y) dx'} \right]^{-1} \quad (56)$$

for every $x \in [0, 1]$. The solution to (56) exists and is unique in the class of functions $u(x, s) \geq 0$, analytic for $\Im(s) > 0$ and continuous on $x \in [0, 1]$.

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

$$G_{\mathbf{H}\mathbf{H}^H}(s) = \int_0^1 u(x, s) dx. \quad (57)$$

This theorem is due to Girko [19]. It has been used by Hanly and Tse [20] to proof resource pooling of chips and receive antennas in CDMA systems with antenna diversity.

TABLE I
TABLE OF STIELTJES TRANSFORMS ($\Im s > 0$).

$$\begin{aligned} G_{\alpha\mathbf{I}}(s) &= \frac{1}{\alpha - s} \\ G_{\mathbf{K}}(s) &= \frac{s}{2} \sqrt{1 - \frac{4}{s^2}} - \frac{s}{2} \\ G_{\mathbf{Q}}(s) &= \sqrt{1 - \frac{4}{s^2}} \left(\frac{s}{2} - s \arcsin \frac{2}{s} \right) - \frac{s}{2} - \frac{1}{2\pi} \\ G_{\mathbf{Q}^2}(s) &= \frac{1}{2} \sqrt{1 - \frac{4}{s}} - \frac{1}{2} \\ G_{\mathbf{R}^2}(s) &= \sqrt{\frac{(1-\beta)^2}{4s^2} - \frac{1+\beta}{2s} + \frac{1}{4}} - \frac{1}{2} - \frac{1-\beta}{2s} \\ G_{\mathbf{Y}}(s) &= \frac{-\text{sign}(\Re s)}{\sqrt{s^2 - 4}} \end{aligned}$$

TABLE II
PROPERTIES OF THE STIELTJES TRANSFORM (FOR ANY $N \times \beta N$ MATRIX \mathbf{X}).

$$\begin{aligned} G_{\lambda^2}(s) &= \frac{G_{\lambda}(\sqrt{s}) - G_{\lambda}(-\sqrt{s})}{2\sqrt{s}} \\ G_{\mathbf{X}\mathbf{X}^H}(s) &= \beta G_{\mathbf{X}^H\mathbf{X}}(s) + \frac{\beta - 1}{s} \\ \Im G(s) &\geq 0 \end{aligned}$$

B.4 Stieltjes Transforms of Some Distributions

The Stieltjes transforms of the distributions in Table I can be found either via the defining integral or (often easier) via solving (51). The Stieltjes transforms in Table I hold for all $s \in \mathbb{C} \setminus \mathbb{R}$. Due to the multiple branches of the complex square root, some formulas can be further simplified in the local neighborhoods of particular s .

C. 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

$$\mathbf{U}^T \mathbf{A} \mathbf{U} = \mathbf{H}^T \mathbf{H}. \quad (58)$$

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

$$\mathbf{y} = \mathbf{U} \mathbf{x}. \quad (59)$$

Then, as $N, K \rightarrow \infty$, but $\beta = K/N$ fixed,

$$\sum_{k=1}^{\lceil tN \rceil} y_k^2 \longrightarrow t \quad (60)$$

almost surely for every $t \in [0; 1]$ with $\lceil x \rceil$ denoting the nearest integer to x which is not smaller than x [10].

This result is like a law of large numbers for the components of any linear combination of the components of the eigenvectors of $\mathbf{H}^T \mathbf{H}$. It is not obvious to hold, since the elements of the eigenvector matrix \mathbf{U}^T are not statistically independent. However, this theorem shows that, for the purpose of summing its squared elements, we can assume they were statistically independent in the large matrix limit.

The convergence in (60) straightforwardly implies the following extension: Let the real-valued sequence α_k be uniformly bounded from below by a constant larger than zero. Then, under the same conditions as required for (60) to hold, we have almost surely

$$\sum_{k=1}^{\lceil tN \rceil} \alpha_k y_k^2 \longrightarrow \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^{\lceil tN \rceil} \alpha_k. \quad (61)$$

This result can be used to show the almost sure convergence of the SINRs of linear MMSE detectors, see Section III-D.

D. Applications

In the following, some examples are given to demonstrate the usefulness of the previous results from random matrix theory to communications engineering.

D.1 Convergence of the SINR of the Linear MMSE Detector

Consider the linear MMSE detector studied in Section II-C with a real-valued channel matrix \mathbf{H} . Its SINR is given by (20) as

$$\text{SINR}_k = \frac{1}{1 - P \mathbf{u}_k^T (\sigma_0^2 \mathbf{A}^{-1} + P \mathbf{I})^{-1} \mathbf{u}_k} - 1. \quad (62)$$

In general, it is different for each user k . However, if the channel matrix \mathbf{H} is composed of independent identically random entries with zero mean, variance $1/N$ and all other moments finite, we can use the deformed quarter circle law and (61) to show almost sure convergence of the SINR of all users to the same deterministic limit, as the matrix size grows large.

If we let $x_k = 1 \Rightarrow x_{\neq k} = 0$ in (59), we find from (61) the almost sure convergence

$$\begin{aligned} 1 - P \mathbf{u}_k^T (\sigma_0^2 \mathbf{A}^{-1} + P \mathbf{I})^{-1} \mathbf{u}_k \\ \longrightarrow 1 - P \text{Tr} (\sigma_0^2 \mathbf{A}^{-1} + P \mathbf{I})^{-1} \end{aligned} \quad (63)$$

$$= \frac{\sigma_0^2}{P} \text{Tr} \left(\frac{\sigma_0^2}{P} \mathbf{I} + \mathbf{A} \right)^{-1}. \quad (64)$$

From (47) and the deformed quarter circle law, we get the almost sure identity

$$\text{Tr} \left(\frac{\sigma_0^2}{P} \mathbf{I} + \mathbf{A} \right)^{-1} = G_{\mathbf{H}^T \mathbf{H}} \left(-\frac{\sigma_0^2}{P} \right). \quad (65)$$

With Table II, the Stieltjes transform $G_{\mathbf{H}^T \mathbf{H}}(s)$ can be expressed in terms of $G_{\mathbf{H} \mathbf{H}^T}(s)$. The latter is given explicitly in Table I. Thus, combining (62), (64), and (65) gives after some trivial algebra the large system limit for the SINR of user k

$$\text{SINR}_k \longrightarrow \frac{(1-\beta)P}{2\sigma_0^2} - \frac{1}{2} + \sqrt{\frac{(1-\beta)^2 P^2}{4\sigma_0^4} + \frac{(1+\beta)P}{2\sigma_0^2} + \frac{1}{4}}. \quad (66)$$

This derivation was limited to real-valued channel matrices due to technical reasons (a convergence property for eigenvectors of complex matrices similar to (61) has not been established so far). However, it can be shown (via more involved methods) that it does hold for complex channel matrices, as well.

If the powers of the users differ, an explicit expression for the asymptotic SINRs is not possible. However, the SINRs still converge to asymptotic limits as shown by Tse and Hanly [17]. They turn out to be not identical, but linear functions of the power of the user of interest. Denoting the power of user k by P_k and assuming the empirical distribution of powers over the user population to converge to a limiting distribution $P_P(p)$ as $K = \beta N \rightarrow \infty$, we find [17]

$$\text{SINR}_k \longrightarrow P_k \eta \quad (67)$$

with

$$\eta = \frac{1}{\sigma_0^2 + \beta \int_0^\infty \frac{p dP_P(p)}{1+p\eta}}. \quad (68)$$

Tse and Hanly's asymptotic result was used by Caire and Müller in [21] to optimize power control in iteratively decoded CDMA systems which a large number of users.

D.2 Implementation of the Linear MMSE Detector

The linear MMSE detector, see Section II-C, reduces complexity of detection from exponential to polynomial dependency on the system dimension K . Nevertheless, it still requires a matrix inversion in (16) for performing its task. This matrix inversion is computationally very costly, if it is performed by algorithms which cannot be implemented on parallel hardware architecture such as Gauss elimination.

Parallel algorithms for matrix inversion operate iteratively. The classical Gauss-Seidel algorithm approximates the inverse matrix by a matrix-valued Taylor approximation

$$\mathbf{X}^{-1} = \alpha \sum_{i=0}^{\infty} (\mathbf{I} - \alpha \mathbf{X})^i \quad (69)$$

in the neighborhood of some multiple α^{-1} of the identity matrix which converges if all eigenvalues of the matrix \mathbf{X} lie within the interval $(0; \frac{2}{\alpha})$ [22]. By appropriate choice of the parameter α , convergence can always be ensured for positive definite matrices.

In principle, α can be chosen arbitrarily tiny ensuring convergence for any positive definite matrix. In practice, however, one would like to choose α not too small to avoid numerical inaccuracies due to quantization errors. For the latter purpose an upper bound on the eigenvalues of the

matrix \mathbf{X} is helpful. For asymptotically large random matrices such upper bounds are provided by random matrix theory. If the channel matrix is composed of independent identically distributed random entries, for instance, the eigenvalues of $\mathbf{H} \mathbf{H}^H$ are asymptotically upper bounded by

$$\lambda_k < \left(1 + \sqrt{\beta}\right)^2 \quad (70)$$

via (41). The matrix to be inverted for the linear MMSE detector is $\sigma_0^2 \mathbf{I} + P \mathbf{H} \mathbf{H}^H$. Since addition of identity matrices increases all eigenvalues by 1, convergence is ensured if

$$\sigma_0^2 + P \left(1 + \sqrt{\beta}\right)^2 < \frac{2}{\alpha}. \quad (71)$$

In practice, one would like to fulfill this condition with some margin to speed up convergence and to cope with deviations of the eigenvalue distributions of finite-size random matrices from their asymptotic behavior. A more comprehensive treatment of this matter can be found in the work of Trichard et al. [23].

D.3 Polynomial Expansion Detectors

The iterative method for matrix inversion presented in the previous subsection can be parallelized to run on up to K processors. Though, it may still require many iterations to converge. The convergence can be accelerated significantly making use of the asymptotic convergence of the eigenvalue distribution of the matrix to be inverted.

Assume you want to invert a $K \times K$ matrix \mathbf{X} whose eigenvalues $\mathcal{L} = \{\lambda_1, \dots, \lambda_K\}$ are known to you. Note that due to the Cayley-Hamilton Theorem [24] any matrix is a zero of its characteristic polynomial

$$\prod_{k=1}^K (\mathbf{X} - \lambda_k \mathbf{I})^k = \mathbf{0}. \quad (72)$$

Expanding the product into a sum, we find

$$\sum_{k=0}^K c_k(\mathcal{L}) \mathbf{X}^k = \mathbf{0} \quad (73)$$

with some coefficients c_k depending on the eigenvalues of \mathbf{X} . Solving this equation for $\mathbf{X}^0 = \mathbf{I}$ and multiplying both sides \mathbf{X}^{-1} gives the desired inverse matrix as a $(K-1)^{\text{st}}$ order polynomial in \mathbf{X}

$$\mathbf{X}^{-1} = - \sum_{k=0}^{K-1} \frac{c_{k+1}(\mathcal{L})}{c_0(\mathcal{L})} \mathbf{X}^k \triangleq \sum_{k=0}^{K-1} w_k(\mathcal{L}) \mathbf{X}^k. \quad (74)$$

Since the eigenvalue distribution depends only on the statistics of \mathbf{X} , the coefficients $w_k = -c_{k+1}/c_0$ can be pre-computed for large-dimensional random matrices.

While the Gauss-Seidel iteration (69) requires, in principle, the summation of an infinite number of terms to achieve arbitrary precision, the knowledge of the eigenvalues reduces the number of terms to be summed to the dimension of the matrix.

Evaluating a polynomial of degree $K-1$ can still be a task too complicated to perform in real-time. Though

polynomials with lower degrees can, in general, not equal the inverse of the matrix, they may be accurate approximations. Depending on the cost function for the approximation error, various designs for the coefficients of shorter polynomials are sensible.

Defining the total mean-squared error as cost function, the optimum coefficients for polynomial of order $L - 1$ are determined by a system of Yule-Walker equations [25]

$$\begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_L \end{bmatrix} = \begin{bmatrix} m_2 & m_3 & \dots & m_{L+1} \\ m_3 & m_4 & \dots & m_{L+2} \\ \vdots & \vdots & \ddots & \vdots \\ m_{L+1} & m_{L+2} & \dots & m_{2L} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{L-1} \end{bmatrix} \quad (75)$$

where

$$m_k \triangleq \frac{1}{K} \sum_{i=1}^K \lambda_i^k. \quad (76)$$

Moshavi et al. [25] suggested to track the empirical eigenvalue moments m_k adaptively and to solve the Yule-Walker equations (75) in real-time. Since these moments converge to non-random deterministic limits for a large class of random channels matrices, they can be computed analytically as functions of the channel statistics and so can the weights. This approach was proposed by Müller and Verdú [26] and explicit expressions for the optimum weights and the achievable SINRs were given for channel matrices with independent identically distributed entries. Their results can be generalized to channel matrices with other asymptotic statistics using results summarized in the next section.

The weights can also be calculated adaptively interpreting the polynomial expansion detector as a multistage Wiener filter [27], [28]. This approach was followed by Honig and Xiao [29] and large system SINRs were derived for equal power users in terms of continued fractions. A generalization for users with different powers can be found in [30]. Loubaton and Hachem [31] highlighted the connection between the Marčenko-Pastur distribution, continued fractions, and orthogonal polynomials for the analysis of polynomial expansion detectors. Such a connection is well established in mathematical literature [32], but found its way into the design and analysis of code-division multiple-access only recently.

IV. 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 why this causes problems for probability theory, if a random matrix is seen as a single (non-commutative) random

variable, consider the expectations

$$E \{(xy)^m\} = E \{x^m y^m\} \quad (77)$$

$$E_{\text{free}} \{(\mathbf{X}\mathbf{Y})^m\} \neq E_{\text{free}} \{\mathbf{X}^m \mathbf{Y}^m\} \quad (78)$$

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 which implies

$$E \{(xy)^m\} = E \{x^m y^m\} = E \{x^m\} E \{y^m\}. \quad (79)$$

For statistically independent random matrices \mathbf{X} and \mathbf{Y} , such a factorization cannot hold, 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 object². 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. Both approaches make sense and can enrich each other. 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.

A. 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

$$E_{\text{free}} \{\cdot\} \triangleq \text{Tr}(\cdot) \quad (80)$$

is the right definition for some random matrices to fit into the framework of free probability theory. We call those random matrices *converging* random matrices. All random matrices discussed in Section III are converging. For those random matrices, $\text{Tr}(\cdot)$ is indeed (almost surely) a deterministic quantity (as any expectation should be) due to the asymptotic convergence of their eigenvalues.

B. 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. The reason for this is due to the non-commutative nature of matrix multiplication and will soon become clear.

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

$$\text{Tr}(\mathbf{ABCD}) = \text{Tr}(\mathbf{AB})\text{Tr}(\mathbf{CD}) \quad (81)$$

$$\text{Tr}(\mathbf{ACBD}) \neq \text{Tr}(\mathbf{AB})\text{Tr}(\mathbf{CD}) \quad (82)$$

$$\text{Tr}(\mathbf{ACBD}) \neq \text{Tr}(\mathbf{AC})\text{Tr}(\mathbf{BD}). \quad (83)$$

²Note that the more common definition of statistical independence by factorization of densities fails even more obviously, since it is not possible to even define a joint density in a sensible way due to the lack of appropriate relation operators “ \leq ” and “ $>$ ”.

For classical random variables and classical expectation operators, (81) and (82) would contradict each other, since (81) implies that the random variables \mathbf{AB} and \mathbf{CD} are uncorrelated while (82) implies that they are not. For non-commutative multiplication, however, (81), (82), and (83) can be true at the same time. In fact, with an appropriate definition of freeness, we find that (81), (82), and (83) are implied by the assumption that the sets $\{\mathbf{A}, \mathbf{B}\}$ and $\{\mathbf{C}, \mathbf{D}\}$ form a family of free random variables. Note, however, that (81), (82), and (83) do not imply freeness of the family $(\{\mathbf{A}, \mathbf{B}\}, \{\mathbf{C}, \mathbf{D}\})$, but additional conditions need to be satisfied. In particular, the definition of freeness must respect the ordering of matrix factors in all possible products (joint moments).

B.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. On the one hand, let x, y be real numbers. The set of all n^{th} order polynomials in two variables x and y is given by

$$\mathcal{P}_n(x, y) \triangleq \left\{ \sum_{i=1}^{\infty} \alpha_i x^{\ell_i} y^{m_i} : \ell_i, m_i \in \{0, 1, \dots, n\} \wedge \alpha_i \in \mathbb{R} \forall i \right\}. \quad (84)$$

For order two for instance, it is canonically given by a sum with only nine terms

$$\alpha_1 x^2 y^2 + \alpha_2 x^2 y + \alpha_3 x y^2 + \alpha_4 x^2 + \alpha_5 x y + \alpha_6 y^2 + \alpha_7 x + \alpha_8 y + \alpha_9. \quad (85)$$

On the other hand, let \mathbf{A}, \mathbf{B} be real matrices. The set of all n^{th} order non-commutative polynomials in two variables \mathbf{A} and \mathbf{B} is given by

$$\mathcal{P}_n(\mathbf{A}, \mathbf{B}) \triangleq \left\{ \sum_{i=1}^{\infty} \alpha_i \prod_{k=1}^n \mathbf{A}^{\ell_{i,k}} \mathbf{B}^{m_{i,k}} : \sum_{k=1}^n \ell_{i,k}, \sum_{k=1}^n m_{i,k} \in \{0, 1, \dots, n\} \wedge \alpha_i \in \mathbb{R} \forall i \right\}. \quad (86)$$

For order two for instance, it is canonically given by a sum of 19 terms

$$\begin{aligned} & \alpha_1 \mathbf{A}^2 \mathbf{B}^2 + \alpha_2 \mathbf{A} \mathbf{B}^2 \mathbf{A} + \alpha_3 \mathbf{A} \mathbf{B} \mathbf{A} \mathbf{B} + \alpha_4 \mathbf{B} \mathbf{A} \mathbf{B} \mathbf{A} + \\ & \alpha_5 \mathbf{B} \mathbf{A}^2 \mathbf{B} + \alpha_6 \mathbf{B}^2 \mathbf{A}^2 + \alpha_7 \mathbf{A}^2 \mathbf{B} + \alpha_8 \mathbf{A} \mathbf{B} \mathbf{A} + \alpha_9 \mathbf{A} \mathbf{B}^2 + \\ & \alpha_{10} \mathbf{B} \mathbf{A}^2 + \alpha_{11} \mathbf{B} \mathbf{A} \mathbf{B} + \alpha_{12} \mathbf{B}^2 \mathbf{A} + \alpha_{13} \mathbf{A}^2 + \alpha_{14} \mathbf{A} \mathbf{B} + \\ & \alpha_{15} \mathbf{B} \mathbf{A} + \alpha_{16} \mathbf{B}^2 + \alpha_{17} \mathbf{A} + \alpha_{18} \mathbf{B} + \alpha_{19} \mathbf{I}. \end{aligned} \quad (87)$$

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

$$\mathcal{P}_n(\mathbf{A}_1, \dots, \mathbf{A}_p) \triangleq \left\{ \sum_{i=1}^{\infty} \alpha_i \prod_{k=1}^n \prod_{q=1}^p \mathbf{A}_q^{\ell_{i,k,q}} : \sum_{k=1}^n \ell_{i,k,q} \in \{0, 1, \dots, n\} \wedge \alpha_i \in \mathbb{R} \forall i, q \right\}. \quad (88)$$

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

B.2 Definition of Freeness

In literature [33], [34], [35], [12], 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 IV.1: Let $s_k \in \{1, 2, \dots, r\}$ be a sequence of integers such that

$$s_k - s_{k-1} \neq 0. \quad (89)$$

Then, the sets $\mathcal{Q}_1 \triangleq \{\mathbf{A}_1, \dots, \mathbf{A}_a\}$, $\mathcal{Q}_2 \triangleq \{\mathbf{B}_1, \dots, \mathbf{B}_b\}$, \dots , \mathcal{Q}_r form a free family $(\mathcal{Q}_1, \dots, \mathcal{Q}_r)$ if, for every sequence s_k obeying (89), any sequence of polynomials \mathbf{Q}_k such that $\mathbf{Q}_k \in \mathcal{P}_{\infty}(\mathcal{Q}_{s_k})$, and any positive integer n ,

$$\text{Tr}(\mathbf{Q}_1) = \dots = \text{Tr}(\mathbf{Q}_n) = 0 \implies \text{Tr}(\mathbf{Q}_1 \mathbf{Q}_2 \dots \mathbf{Q}_n) = 0. \quad (90)$$

Note that due to (89) adjacent factors in the product $\mathbf{Q}_1 \mathbf{Q}_2 \dots \mathbf{Q}_n$ must be polynomials of different sets of the family. This reflects the non-commutative nature in the definition of freeness.

B.3 Calculation of Expectations

Though the definition of freeness is very implicit, it can be used to recursively calculate cross-family joint moments of non-commutative random variables out of inter-family joint moments.

For instance, consider (82) and assume that the factors are chosen from the free family $(\{\mathbf{A}, \mathbf{B}\}, \{\mathbf{C}, \mathbf{D}\})$. Choose the non-commutative polynomials

$$\mathbf{Q}_1 = \mathbf{A} - \text{Tr}(\mathbf{A})\mathbf{I} \quad (91)$$

$$\mathbf{Q}_2 = \mathbf{C} - \text{Tr}(\mathbf{C})\mathbf{I} \quad (92)$$

$$\mathbf{Q}_3 = \mathbf{B} - \text{Tr}(\mathbf{B})\mathbf{I} \quad (93)$$

$$\mathbf{Q}_4 = \mathbf{D} - \text{Tr}(\mathbf{D})\mathbf{I}. \quad (94)$$

Note that polynomials with adjacent indices are built of matrices belonging to different sets of the family. Since

$$\text{Tr}(\mathbf{Q}_k) = \text{Tr}(\mathbf{X} - \text{Tr}(\mathbf{X})\mathbf{I}) = 0 \quad (95)$$

by the linearity of the expectation operator, Definition IV.1 implies

$$\text{Tr}(\mathbf{Q}_1 \mathbf{Q}_2 \mathbf{Q}_3 \mathbf{Q}_4) = 0. \quad (96)$$

Plugging (91) to (94) into (96) and expanding the products, we find with the linearity of the trace

$$\begin{aligned} \text{Tr}(\mathbf{ACBD}) &= \text{Tr}(\mathbf{B})\text{Tr}(\mathbf{ACD}) + \text{Tr}(\mathbf{D})\text{Tr}(\mathbf{ACB}) + \\ &+ \text{Tr}(\mathbf{A})\text{Tr}(\mathbf{CBD}) + \text{Tr}(\mathbf{C})\text{Tr}(\mathbf{ABD}) \\ &- \text{Tr}(\mathbf{B})\text{Tr}(\mathbf{D})\text{Tr}(\mathbf{AC}) \\ &- \text{Tr}(\mathbf{A})\text{Tr}(\mathbf{B})\text{Tr}(\mathbf{CD}) \\ &- \text{Tr}(\mathbf{A})\text{Tr}(\mathbf{D})\text{Tr}(\mathbf{CB}) \\ &- \text{Tr}(\mathbf{C})\text{Tr}(\mathbf{B})\text{Tr}(\mathbf{AD}) \\ &- \text{Tr}(\mathbf{C})\text{Tr}(\mathbf{D})\text{Tr}(\mathbf{AB}) \\ &- \text{Tr}(\mathbf{A})\text{Tr}(\mathbf{C})\text{Tr}(\mathbf{BD}) + \\ &+ 3\text{Tr}(\mathbf{A})\text{Tr}(\mathbf{B})\text{Tr}(\mathbf{C})\text{Tr}(\mathbf{D}). \end{aligned} \quad (97)$$

Now, we have broken down an expectation of four factors into sums of expectations of up to three factors.

The expectations of three factors can be broken down into sums of expectations of two two factors. This is demonstrated at the example of $\text{Tr}(\mathbf{ACD})$. Since the neighboring factors \mathbf{C} and \mathbf{D} belong to the same set of the free family, we must define the non-commutative polynomials in a different way as in (91) to (94). Now, an appropriate definition is

$$\mathbf{Q}_1 = \mathbf{A} - \text{Tr}(\mathbf{A})\mathbf{I} \quad (98)$$

$$\mathbf{Q}_2 = \mathbf{CD} - \text{Tr}(\mathbf{CD})\mathbf{I}. \quad (99)$$

Note that with definitions (98) and (99), adjacent polynomials in $\text{Tr}(\mathbf{Q}_1\mathbf{Q}_2)$ belong to different sets of the free family. Proceeding this way for all remaining matrix products involving factors belonging to different sets of the family $(\{\mathbf{A}, \mathbf{B}\}, \{\mathbf{C}, \mathbf{D}\})$, we finally arrive at

$$\begin{aligned} \text{Tr}(\mathbf{ACBD}) &= \text{Tr}(\mathbf{A})\text{Tr}(\mathbf{B})\text{Tr}(\mathbf{CD}) + \\ &\quad + \text{Tr}(\mathbf{C})\text{Tr}(\mathbf{D})\text{Tr}(\mathbf{AB}) \\ &\quad - \text{Tr}(\mathbf{A})\text{Tr}(\mathbf{B})\text{Tr}(\mathbf{C})\text{Tr}(\mathbf{D}). \end{aligned} \quad (100)$$

The procedure for products of more than four factors is obvious, but can be very tedious. The key point to succeed with this procedure is to define the non-commutative polynomials in an appropriate way which simply consists of collecting all factors belonging to identical sets of the free family and subtract its expectation.

C. 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 [36], [33]. His results were strengthened and extended by Thorbjørnsen and Hiai and Petz [37], [12] later on.

C.1 Gaussian Random Matrices

Let the random matrices $\mathbf{H}_i, \forall i$, be square $N \times N$ with independent identically complex Gaussian distributed proper³ 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 $N \rightarrow \infty$. Then the family

$$\left(\left\{ \mathbf{X}_1, \mathbf{X}_1^H, \mathbf{X}_2, \mathbf{X}_2^H, \dots \right\}, \left\{ \mathbf{H}_1, \mathbf{H}_1^H \right\}, \left\{ \mathbf{H}_2, \mathbf{H}_2^H \right\}, \dots \right) \quad (101)$$

is asymptotically free as $N \rightarrow \infty$ almost surely [12], [37].

C.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,

³A complex random variable is said to be *proper* if real and imaginary part are independent and identically distributed [9].

let the matrices $\mathbf{X}_j, \forall j$, be as in Section IV-C.1 and let $\mathbf{S}_i = \mathbf{H}_i\mathbf{H}_i^H, \forall i$. Then the family

$$\left(\left\{ \mathbf{X}_1, \mathbf{X}_1^H, \mathbf{X}_2, \mathbf{X}_2^H, \dots \right\}, \left\{ \mathbf{S}_1 \right\}, \left\{ \mathbf{S}_2 \right\}, \dots \right) \quad (102)$$

is almost surely asymptotically free as $N, K \rightarrow \infty$ with $\beta = K/N$ fixed [12], [37].

The asymptotic freeness of some random covariance matrices has been used by Evans and Tse [38] to analyze multiuser channel estimation in large CDMA systems and by Biglieri et al. [39] to calculate error rates of space-time codes. The analysis in [38] shows that the entries of the random matrices \mathbf{H}_i need not necessarily be jointly independent. Allowing for some structure for the statistics of the random matrices \mathbf{H}_i which reflects the nature of mobile radio channels, Evans and Tse could still prove freeness. Their results were used and extended by Cottatellucci and Müller [40], [41], [42] and Li et al. [43], [44] to design polynomial expansion receivers, see Section III-D.3, for more general channel conditions.

C.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 III-A.5. Moreover, let the matrices $\mathbf{X}_j, \forall j$, be as in Section IV-C.1. Then, the family

$$\left(\left\{ \mathbf{X}_1, \mathbf{X}_1^H, \mathbf{X}_2, \mathbf{X}_2^H, \dots \right\}, \left\{ \mathbf{T}_1, \mathbf{T}_1^H \right\}, \left\{ \mathbf{T}_2, \mathbf{T}_2^H \right\}, \dots \right) \quad (103)$$

is almost surely asymptotically free as $N \rightarrow \infty$ [12], [37].

The asymptotic freeness of such unitary random matrices has been used by Debbah et. al. [45] for the analysis of multi-carrier systems.

D. R-Transform

The calculation of distributions of functions of several free random variables via the definition of freeness is often a very tedious task. However, for some operations such a summation, significant simplifications are possible.

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

$$\mathbf{C} \triangleq \mathbf{A} + \mathbf{B}. \quad (104)$$

Then, we call the probability measure (asymptotic eigenvalue distribution) $\mathbf{p}_\mathbf{C}(x)$ the *additive free convolution* of the probability measures $\mathbf{p}_\mathbf{A}(x)$ and $\mathbf{p}_\mathbf{B}(x)$. Unlike classical convolution which provides the distribution of a sum of independent commutative random variables, additive free convolution is a highly non-linear operation and, thus, cannot be performed by simple integration.

In principle, the moments of $\mathbf{p}_\mathbf{C}(x)$ could be found from the moments of $\mathbf{p}_\mathbf{A}(x)$ and $\mathbf{p}_\mathbf{B}(x)$ 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

$$\mathbf{R}(w) \triangleq \mathbf{G}^{-1}(-w) - w^{-1} \quad (105)$$

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 [33]. Thus, we have

$$R_{\mathbf{C}}(w) = R_{\mathbf{A}}(w) + R_{\mathbf{B}}(w). \quad (106)$$

Then, the distribution of \mathbf{C} can be recovered from the R-transform via

$$G_{\mathbf{C}}(R_{\mathbf{C}}(-w) - w^{-1}) = w \quad (107)$$

which follows directly from (105) and the Stieltjes inversion formula (49).

Tse [46] 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.

D.1 Tables of R-Transforms

In Table III, R-transforms of some of the random matrices introduced in Section III-A are listed. Table IV shows

TABLE III
TABLE OF R-TRANSFORMS.

$R_{\alpha \mathbf{I}}(w)$	$= \alpha$
$R_{\mathbf{K}}(w)$	$= w$
$R_{\mathbf{Q}^2}(w)$	$= \frac{1}{1-w}$
$R_{\mathbf{R}^2}(w)$	$= \frac{\beta}{1-w}$
$R_{\mathbf{Y}}(w)$	$= \frac{-1 + \sqrt{1 + 4w^2}}{w}$

two general properties of the R-transform.

D.2 Additive Free Convolution of Binary Measures

Assume that the two free random variables in (104) are binary distributed

$$p_{\mathbf{A}}(x) = p_{\mathbf{B}}(x) = \frac{1}{2} \delta(x-1) + \frac{1}{2} \delta(x+1). \quad (108)$$

Then, by (47), the Stieltjes transforms of their distributions are

$$G_{\mathbf{A}}(s) = G_{\mathbf{B}}(s) = \frac{s}{1-s^2}. \quad (109)$$

TABLE IV
PROPERTIES OF THE R-TRANSFORM.

$R_{\alpha \mathbf{X}}(w)$	$= \alpha R_{\mathbf{X}}(\alpha w)$
$\lim_{w \rightarrow 0} R(w)$	$= \int x dP(x)$

With definition (105), we find that the R-transform is given by

$$R_{\mathbf{A}}(w) = R_{\mathbf{B}}(w) = \frac{-1 + \sqrt{1 + 4w^2}}{2w}. \quad (110)$$

Here, we have discarded the other branch of the square root, since $\lim_{w \rightarrow 0} R_{\mathbf{A}}(w) = 0$ by Table IV.

The R-transform of the distribution of \mathbf{C} is the sum of the R-transforms of the distributions of the matrices \mathbf{A} and \mathbf{B} . Thus, we find

$$R_{\mathbf{C}}(w) = \frac{-1 + \sqrt{1 + 4w^2}}{w}. \quad (111)$$

Though it differs only by a factor of 2 from the R-transforms in (110), it corresponds to an essentially different distribution function. Using Table III, we find that the additive free convolution of two binary measures gives the *inverse* semicircle law

$$p_{\mathbf{C}}(x) = \begin{cases} \frac{1}{\pi} \frac{1}{\sqrt{4-x^2}} & |x| < 2 \\ 0 & \text{elsewhere} \end{cases}. \quad (112)$$

This procedure can be used, in principle, for the additive free convolution of any distributions. Practical problems occur, when calculating the inverse functions in (105) and (107), since these function can happen to not allow for an inverse function in closed form, e.g. since they are polynomials of fifth or higher order. In some of these cases, it helps to apply the Stieltjes inversion formula (49) to an implicit equation for the Stieltjes transform $G_{\mathbf{C}}(s)$. This results in solving a system of two equations for the real and imaginary part of $G_{\mathbf{C}}(s)$. In the worst case, one might succeed to develop those functions which cause trouble into power series. The latter approach, however, is rather complicated. It can be simplified, making use of the Kreweas complement (see e.g. [12]).

The additive free convolution of two binary distributions has turned out to be a continuous distribution. This is in sharp contrast to what we are used from adding (classical) commutative random variables.

D.3 Free Central Limit Theorem

In classical (commutative) probability theory, the sum of an infinite number of independent identically distributed zero-mean terms is Gaussian distributed. This is well-known as the central limit theorem. It seems obvious that there should be some counterpart in free probability as well. It is intuitive that the notion of statistical independence should be replaced by freeness, but what is the counterpart of the Gaussian distribution?

Consider again the additive free convolution of free binary measures as in the previous subsection. However, let there be now n terms to be summed and the sum to be normalized to unit variance

$$\mathbf{C}_n \triangleq \frac{1}{\sqrt{n}} \sum_{k=1}^n \mathbf{A}_k. \quad (113)$$

We know from (110) and Table IV that

$$R_{\mathbf{C}_n}(w) = \frac{\sqrt{n^2 + 4nw^2} - n}{2w}. \quad (114)$$

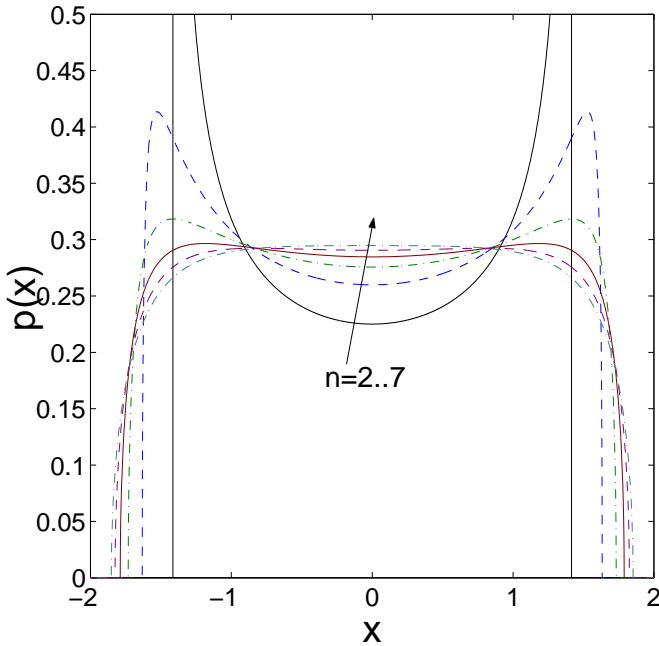


Fig. 1. Additive free convolution of several binary measures.

In the Stieltjes domain, this reads

$$G_{C_n}(s) = \frac{1}{2} \frac{(n-2)s - \sqrt{n^2 s^2 - 4n^2 + 4n}}{s^2 - n} \quad (115)$$

which, for $n > 1$, corresponds to the density

$$p_{C_n}(x) = \begin{cases} \frac{1}{2\pi} \frac{\sqrt{4n^2 - 4n - n^2 x^2}}{n - x^2} & |x| < 2\sqrt{1 - 1/n} \\ 0 & \text{elsewhere} \end{cases} \quad (116)$$

The densities for $n < 8$ are shown in Fig. 1. One can observe that they approach the semicircular distribution. This can be easily verified from (114) since

$$\lim_{n \rightarrow \infty} R_{C_n}(w) = w \quad (117)$$

and w is the R-transform of the semicircle law by Table III. It can also be seen, taking directly the limit $n \rightarrow \infty$ in (116).

The semicircular distribution is not only the limit distribution for the additive free convolution of many binary measures, but any appropriately scaled zero-mean measures with finite moments, as proven in the following.

The R-transform is an analytic function within the neighborhood of $w = 0$. Thus, we can write it as

$$R_{A_k}(w) = \sum_{\ell=0}^{\infty} \alpha_{\ell,k} w^\ell \quad (118)$$

where $\alpha_{0,k} = 0$ since we assume $\text{Tr}(A_k) = 0$. From (110), (113), and Table IV, we find

$$\begin{aligned} \lim_{n \rightarrow \infty} R_{C_n}(w) &= \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \sum_{k=1}^n R_{A_k} \left(\frac{w}{\sqrt{n}} \right) \quad (119) \\ &= \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \sum_{k=1}^n \sum_{\ell=1}^{\infty} \alpha_{\ell,k} \left(\frac{w}{\sqrt{n}} \right)^\ell \quad (120) \end{aligned}$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \alpha_{1,k} w. \quad (121)$$

Using Tables III and IV, this can be identified as a scaled semicircle distribution regardless of the higher order moments of the distributions of the terms to be summed.

E. S-Transform

In analogy to additive free convolution, we define

$$D \triangleq AB \quad (122)$$

and call the probability measure $p_D(x)$ the *multiplicative free convolution* of the two probability measures $p_A(x)$ and $p_B(x)$, again under the restriction that A and 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 [33].

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

$$S_D(z) = S_A(z) S_B(z) \quad (123)$$

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

$$\begin{aligned} \Upsilon(s) &\triangleq \int \frac{sx}{1-sx} dP(x) \quad (124) \\ &= -s^{-1} G(s^{-1}) - 1 \quad (125) \end{aligned}$$

which can be obtained either directly by integrating an appropriate kernel with respect to the measure of interest (124) or in terms of the Stieltjes transform (125). Since the definition in terms of the Stieltjes transform is rather straightforward, we do not provide a table for the auxiliary transform $\Upsilon(\cdot)$. Calculating the inverse with respect to composition of this auxiliary transform, the S-transform is given as

$$S(z) \triangleq \frac{1+z}{z} \Upsilon^{-1}(z). \quad (126)$$

In order to return to the probability distribution, you return to the Stieltjes domain via (126) and (125) and then apply the Stieltjes inversion formula (49).

S-transforms of some of the random matrices introduced in Section III-A are listed in Table V.

It was shown by Müller [47] that Silverstein's formula (51) 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 III-B.1.

F. Example: Channel Modeling

Asymptotic freeness of certain large random matrices can be used to characterize the properties of communication

TABLE V
TABLE OF S-TRANSFORMS.

$S_{\alpha\mathbf{I}}(z)$	$= \frac{1}{\alpha}$
$S_{\mathbf{K}}(z)$	$= \text{undefined}$
$S_{\mathbf{Q}^2}(z)$	$= \frac{1}{1+z}$
$S_{\mathbf{R}^2}(z)$	$= \frac{1}{\beta+z}$
$S_{\mathbf{Y}}(z)$	$= \text{undefined}$

channels with antenna arrays at both ends of the communication link in terms of the singular values of the channel's transfer matrix. This approach was studied by Müller [48] for the first time decomposing the channel matrix into a product of three asymptotically free matrices: the steering matrix at transmitter side, a scattering matrix, and a steering matrix at receiver side. The approach was generalized to matrix products of more factors by Müller in [47] and dispersive channels with decaying power-delay profile in [49] and confirmed by measurements in [50]. As an example the asymptotic eigenvalue distribution associated with a channel composed of an arbitrary number of free matrix factors is calculated by means of the S-transform.

Consider a communication channel with K_0 transmitting and K_N receiving antennas grouped into a transmitter and a receiver array, respectively. Let there be $N - 1$ clusters of scatterers each with $K_n, 1 \leq n \leq N - 1$, scattering objects. Assume that the vector-valued transmitted signal propagates from the transmitter array to the first cluster of scatterers, from the first to the second cluster, and so on, until it is received from the $(N - 1)^{\text{st}}$ cluster by the receiver array. Such a channel model is discussed and physical motivation is given in [51, Sec. 3]. Indoor propagation between different floors, for instance, may serve as an environment where multifold scattering can be typical, cf. [52, Sec. 13.4.1].

The communication link outlined above is a linear vector channel that is canonically described by a channel matrix

$$\mathbf{H}_N = \mathbf{M}_N \mathbf{M}_{N-1} \cdots \mathbf{M}_2 \mathbf{M}_1 \triangleq \prod_{n=1}^N \mathbf{M}_n \quad (127)$$

where the matrices \mathbf{M}_1 , $\mathbf{M}_{1 < n < N}$, and \mathbf{M}_N denote the subchannels from the transmitter array to the first cluster of scatterers, from the $(n - 1)^{\text{st}}$ cluster of scatterers to the n^{th} cluster, and from the $(N - 1)^{\text{st}}$ cluster to the receiving array, respectively. This means that \mathbf{M}_n is of size $K_n \times K_{n-1}$.

The performance of communication via linear vector channels described as in (127) is determined by the K_N eigenvalues of the covariance matrix $\mathbf{C}_N \triangleq \mathbf{H}_N \mathbf{H}_N^H$ for many practically relevant cases. In the following, the asymptotic distributions of these eigenvalues are calculated.

Assume that the family $(\{\mathbf{M}_1^H \mathbf{M}_1\}, \{\mathbf{M}_2^H \mathbf{M}_2\}, \dots, \{\mathbf{M}_N^H \mathbf{M}_N\})$ is asymptotically free as all sizes K_n tend to infinity with the ratios

$$\chi_n \triangleq \frac{K_{n-1}}{K_n}, \quad 1 \leq n \leq N, \quad (128)$$

remaining constant. Consider the random covariance matrices

$$\mathbf{C}_i \triangleq \mathbf{H}_i \mathbf{H}_i^H = \left(\prod_{n=1}^i \mathbf{M}_n \right) \left(\prod_{n=1}^i \mathbf{M}_n \right)^H. \quad (129)$$

In the following, the asymptotic eigenvalue distribution of \mathbf{C}_N is recursively calculated applying the S-transform.

For that purpose, note that the non-zero eigenvalues of the matrices

$$\tilde{\mathbf{C}}_i \triangleq \left(\prod_{n=1}^{i-1} \mathbf{M}_n \right) \left(\prod_{n=1}^{i-1} \mathbf{M}_n \right)^H \mathbf{M}_i^H \mathbf{M}_i. \quad (130)$$

and \mathbf{C}_i are identical and that

$$\tilde{\mathbf{C}}_{i+1} = \mathbf{C}_i \mathbf{M}_{i+1}^H \mathbf{M}_{i+1} \quad \forall i. \quad (131)$$

Since the non-zero eigenvalues of the matrices \mathbf{C}_i and $\tilde{\mathbf{C}}_i$ are identical, their empirical distributions differ only by a scaling factor and a point mass at zero. In the Stieltjes domain, this translates with Table II into

$$G_{\mathbf{C}_i}(s) + \frac{1}{s} = \chi_i G_{\tilde{\mathbf{C}}_i}(s) + \frac{\chi_i}{s}. \quad (132)$$

It is straightforward from (125) and (126) that (132) reads in terms of $\mathcal{Y}(s)$ and $S(z)$ as

$$\mathcal{Y}_{\mathbf{C}_i}(s) = \chi_i \mathcal{Y}_{\tilde{\mathbf{C}}_i}(s) \quad (133)$$

$$S_{\mathbf{C}_i}(z) = \frac{z+1}{z+\chi_i} S_{\tilde{\mathbf{C}}_i}\left(\frac{z}{\chi_i}\right), \quad (134)$$

respectively. Let the entries of $M_{1 \leq n \leq N}$ be independent and identically distributed with zero-mean and respective variances $1/K_n$. Then, Table V yields

$$S_{\mathbf{M}_i \mathbf{M}_i^H}(z) = \frac{1}{z + \chi_i} \quad (135)$$

$$S_{\mathbf{M}_i^H \mathbf{M}_i}(z) = \frac{1}{1 + z\chi_i}. \quad (136)$$

Now, we proof by induction:

Lemma IV.1: Define the ratios

$$\rho_n \triangleq \frac{K_n}{K_N}. \quad (137)$$

Let $K_n \rightarrow \infty$, but the ratios ρ_n remain fixed for all $0 \leq n \leq N$. Then,

$$S_{\mathbf{C}_N}(z) = \prod_{n=1}^N \frac{\rho_n}{z + \rho_{n-1}}. \quad (138)$$

Proof: Since the ratios ρ_n depend on N via (137), (138) is re-written into

$$S_{\mathcal{C}_N}(z) = \prod_{n=1}^N \frac{K_n}{zK_N + K_{n-1}}. \quad (139)$$

First, (139) is verified for $N = 1$. (135) directly gives

$$S_{\mathcal{C}_1}(z) = \frac{1}{z + \chi_1} = \frac{K_1}{zK_1 + K_0} \quad (140)$$

which proofs (139) for $N = 1$.

Second, assuming (139) holding for the i -fold product, (139) is shown to also hold for the $(i + 1)$ -fold product. Transforming (131) into the S-domain gives

$$S_{\tilde{\mathcal{C}}_{N+1}}(z) = S_{\mathcal{C}_N}(z) \frac{1}{1 + zK_N/K_{N+1}}. \quad (141)$$

The rotation formula (134) yields

$$S_{\mathcal{C}_{N+1}}(z) = \frac{z + 1}{z + K_N/K_{N+1}} S_{\tilde{\mathcal{C}}_{N+1}} \left(z \frac{K_{N+1}}{K_N} \right) \quad (142)$$

which gives with (141)

$$S_{\mathcal{C}_{N+1}}(z) = \frac{K_{N+1}}{zK_{N+1} + K_N} S_{\mathcal{C}_N} \left(z \frac{K_{N+1}}{K_N} \right). \quad (143)$$

We make use of the assumption that (139) is valid for the N -fold product and find

$$S_{\mathcal{C}_{N+1}}(z) = \frac{K_{N+1}}{zK_{N+1} + K_N} \prod_{n=1}^N \frac{K_n}{zK_{N+1} + K_{n-1}} \quad (144)$$

$$= \prod_{n=1}^{N+1} \frac{K_n}{zK_{N+1} + K_{n-1}}. \quad (145)$$

Hereby, the induction is complete. \square

The lemma yields with (126) and (125)

$$s(\mathcal{Y}_{\mathcal{C}_N}(s) + 1) \prod_{n=1}^N \frac{\mathcal{Y}_{\mathcal{C}_N}(s) + \rho_{n-1}}{\rho_n} = \mathcal{Y}_{\mathcal{C}_N}(s) \quad (146)$$

$$G_{\mathcal{C}_N}(s) \prod_{n=1}^N \frac{\rho_{n-1} - 1 - sG_{\mathcal{C}_N}(s)}{\rho_n} - sG_{\mathcal{C}_N}(s) = 1. \quad (147)$$

The Stieltjes transform of the asymptotic eigenvalue distribution of \mathcal{C}_N is determined in (147) by a polynomial equation of $(N + 1)$ st order. For $N > 3$, it cannot be resolved with respect to the Stieltjes transform, in general. However, it is shown in [47] how to obtain an infinite power series for $\mathcal{Y}_{\mathcal{C}_N}(s)$ and calculate the moments of the asymptotic eigenvalue distribution.

V. 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* [53].

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, we can trust that its predictions in other engineering applications 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.

A. Self Average

While random matrix theory and recently also free probability theory [12], [37] 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.

B. 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 probability distribution $p_{\mathbf{x}}(\mathbf{x})$ of states $\mathbf{x} \in \mathbb{R}^{K \times 1}$, this means that in the thermodynamic equilibrium the (nor-

malized) entropy

$$H(\mathbf{x}) = -\frac{1}{K} \sum_{\mathbf{x}} \Pr(\mathbf{x}) \log \Pr(\mathbf{x}) \quad (148)$$

is maximized while keeping the (normalized) energy

$$E(\mathbf{x}) = \frac{1}{K} \sum_{\mathbf{x}} \Pr(\mathbf{x}) \|\mathbf{x}\| \quad (149)$$

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

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

$$\Pr(\mathbf{x}) = \frac{e^{-\frac{1}{T}\|\mathbf{x}\|}}{\sum_{\mathbf{x}} e^{-\frac{1}{T}\|\mathbf{x}\|}} \quad (150)$$

and called the Boltzmann distribution. The parameter T is called the temperature of the system and determined by (149). 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*⁴ defined as

$$F(\mathbf{x}) \triangleq E(\mathbf{x}) - TH(\mathbf{x}) \quad (151)$$

$$= -\frac{T}{K} \log \left(\sum_{\mathbf{x}} e^{-\frac{1}{T}\|\mathbf{x}\|} \right). \quad (152)$$

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.

C. 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 (151). 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 from the validity of (151) 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 (151) 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 $p_{y|x}(y, x)$ and a source uniquely characterized by a prior density $p_x(x)$. Consider a detector for the output of this channel characterized by an assumed channel transition probability $\check{p}_{y|x}(y, x)$ and an assumed prior distribution $\check{p}_x(x)$. Let the detector minimize some kind of cost function, e.g. bit error probability, subject to its hypotheses on the channel transition probability $\check{p}_{y|x}(y, x)$ and the prior distribution $\check{p}_x(x)$. 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. Many popular detectors can be described with this framework, see Sections II-B and II-C for a few examples.

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 (150), we find

$$\check{p}_{\mathbf{x}|\mathbf{y}, \mathbf{H}}(\mathbf{x}, \mathbf{y}, \mathbf{H}) = \frac{e^{-\frac{1}{T}\|\mathbf{x}\|}}{\int e^{-\frac{1}{T}\|\mathbf{x}\|} d\mathbf{x}} \quad (153)$$

where the dependency on \mathbf{y} , \mathbf{H} , 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 (153).

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 (152):

$$F(\mathbf{x}) = F(\mathbf{x}|\mathbf{y}, \mathbf{H}) \quad (154)$$

$$= \mathbb{E}_{\mathbf{y}, \mathbf{H}} F(\mathbf{x}|\mathbf{y}, \mathbf{H}) \quad (155)$$

$$= \frac{T}{K} \mathbb{E}_{\mathbf{H}} \int \log \frac{1}{\int e^{-\frac{1}{T}\|\mathbf{x}\|} d\mathbf{x}} dP_{\mathbf{y}}(\mathbf{y}) \quad (156)$$

Note that, inside the logarithm, expectations are taken with respect to the assumed distribution via the definition of the energy function, while, outside the logarithm, expectations are taken 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

⁴The *free energy* is not related to *freeness* in free probability theory.

logarithm in (156) becomes $p_{\mathbf{y}}(\mathbf{y})$ up to a normalizing factor. Thus, the free energy becomes the entropy of \mathbf{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 in large communication systems which is purely based on asymptotic properties of large random matrices and does not exploit the tools provided by statistical mechanics has been very limited so far. One exception is the asymptotic SINR of linear MMSE multiuser detectors with erroneous assumptions on the powers of interfering users by Müller and Caire in [54].

D. 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 a random variable

$$\mathbb{E}_y \log(y). \quad (157)$$

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

$$\log(y) = \lim_{n \rightarrow 0} \frac{\partial}{\partial n} y^n. \quad (158)$$

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

$$\mathbb{E}_y \log(y) = \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \log \mathbb{E}_y y^n \quad (159)$$

and reduces the problem to the calculation of the n^{th} moment 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 $\mathbb{E} y^n$. 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

$$y = \int f(x, y) dx \quad (160)$$

for some function $f(x, y)$. Since the variable of integration is arbitrary, this implies

$$y^n = \left(\int f(x, y) dx \right)^n \quad (161)$$

$$= \prod_{a=1}^n \int f(x_a, y) dx_a. \quad (162)$$

Thus, instead of calculating the n^{th} power of 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.

In general, it is not easier to calculate the expectation of the right hand side of (162) than just the expectation over y^n . However, there are some functions $f(x, y)$ for which the replica method is indeed advantageous, particularly if there is no closed form solution for $\int f(x, y) dx$ and y can not be given explicitly. Then, it might help to substitute the set of variables (x_1, \dots, x_n, y) by some other variables which allow to solve the integral.

E. 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

$$\frac{1}{K} \log \int e^{Kf(x,y)} dx dy. \quad (163)$$

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

$$\lim_{K \rightarrow \infty} \frac{1}{K} \log \int e^{Kf(x,y)} dx dy = \max_{x,y} f(x, y). \quad (164)$$

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. $f(x, y) = f(y, x)$ for the bi-variate 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. $\max_x f(x, x)$ for the originally bi-variate case. This is the most critical assumption when applying the replica method. In fact, 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 [55], [53]. For the introductory character of this work, however, replica symmetry breaking is a far to advanced issue.

F. Example: Analysis of Large CDMA Systems

The replica method was introduced into multiuser communications by the landmark paper of Tanaka [56] for the purpose of studying the performance of the maximum a-posteriori detector. Subsequently his work was generalized and extended to other areas of communications by Guo and Verdú [57], [58], [59], [60] and Müller et al. [61], [62], [63], [64], [65], [66], [67]. The analysis of an asymptotically

large CDMA systems with arbitrary joint distribution of the variances of the random chips presented below, however, is yet unpublished. It includes the practically important case of multi-carrier CDMA transmission with users of arbitrary powers in frequency-selective fading as a special case.

Consider a vector-valued real additive white Gaussian noise channel characterized by the conditional probability distribution

$$p_{\mathbf{y}|\mathbf{x},\mathbf{H}}(\mathbf{y}, \mathbf{x}, \mathbf{H}) = \frac{e^{-\frac{1}{2\sigma_0^2}(\mathbf{y}-\mathbf{H}\mathbf{x})^\top(\mathbf{y}-\mathbf{H}\mathbf{x})}}{(2\pi\sigma_0^2)^{\frac{N}{2}}}. \quad (165)$$

Moreover, let the detector be characterized by the assumed conditional probability distribution

$$\check{p}_{\mathbf{y}|\mathbf{x},\mathbf{H}}(\mathbf{y}, \mathbf{x}, \mathbf{H}) = \frac{e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{H}\mathbf{x})^\top(\mathbf{y}-\mathbf{H}\mathbf{x})}}{(2\pi\sigma^2)^{\frac{N}{2}}} \quad (166)$$

and the assumed prior distribution $\check{p}_{\mathbf{x}}(\mathbf{x})$. Let the entries of \mathbf{H} be independent zero-mean with vanishing odd order moments and variances w_{ck}/N for row c and column k . Applying Bayes' law, we find

$$\check{p}_{\mathbf{x}|\mathbf{y},\mathbf{H}}(\mathbf{x}, \mathbf{y}, \mathbf{H}) = \frac{e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{H}\mathbf{x})^\top(\mathbf{y}-\mathbf{H}\mathbf{x}) + \log \check{p}_{\mathbf{x}}(\mathbf{x})}}{\int e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{H}\mathbf{x})^\top(\mathbf{y}-\mathbf{H}\mathbf{x})} d\check{P}_{\mathbf{x}}(\mathbf{x})}. \quad (167)$$

Since (150) holds for any temperature T , we set without loss of generality $T = 1$ in (150) and find the appropriate energy function to be

$$\|\mathbf{x}\| = \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{H}\mathbf{x})^\top(\mathbf{y} - \mathbf{H}\mathbf{x}) - \log \check{p}_{\mathbf{x}}(\mathbf{x}). \quad (168)$$

This choice of the energy function ensures that the thermodynamic equilibrium models the detector defined by the assumed conditional and prior distributions.

Applying successively (156) with (165) and (168), (159), and replica continuity (162), we find for the free energy (171) with $\sigma_a = \sigma, \forall a \geq 1$, $P_0(\mathbf{x}) = P_{\mathbf{x}}(\mathbf{x})$, and $P_a(\mathbf{x}) = \check{P}_{\mathbf{x}}(\mathbf{x}), \forall a \geq 1$.

The following calculations are a generalization of the derivations by Tanaka [56], Guo and Verdú [57], and Müller et al. [62]. The integral in (171) is given by

$$\Xi_n = \int \prod_{c=1}^N \int_{\mathbb{R}} \frac{\int_{\mathbb{R}} \prod_{a=0}^n e^{-\frac{1}{2\sigma_a^2} \left(y_c - \sum_{k=1}^K h_{ck} x_{ak} \right)^2} dy_c}{\sqrt{2\pi}\sigma_0} \prod_{a=0}^n dP_a(\mathbf{x}_a) \quad (172)$$

with y_c , x_{ak} , and h_{ck} denoting the c^{th} component of \mathbf{y} , the k^{th} component of \mathbf{x}_a , and the $(c, k)^{\text{th}}$ entry of \mathbf{H} , respectively. The integrand depends on \mathbf{x}_a only through

$$v_{ac} \triangleq \frac{1}{\sqrt{\beta}} \sum_{k=1}^K h_{ck} x_{ak}, \quad a = 0, \dots, n. \quad (173)$$

Following [56], these quantities can be regarded, in the limit $K \rightarrow \infty$ as jointly Gaussian random variables with zero mean and covariances

$$Q_{ab}[c] = \mathbb{E}_{\mathbf{H}} v_{ac} v_{bc} = \frac{1}{K} \mathbf{x}_a \bullet^{(c)} \mathbf{x}_b \quad (174)$$

where we defined the following inner products

$$\mathbf{x}_a \bullet^{(c)} \mathbf{x}_b \triangleq \sum_{k=1}^K x_{ak} x_{bk} w_{ck}. \quad (175)$$

In order to perform the integration in (172), the $K(n+1)$ -dimensional space spanned by the replicas and the vector \mathbf{x}_0 is split into subshells

$$S\{Q[\cdot]\} \triangleq \left\{ \mathbf{x}_0, \dots, \mathbf{x}_n \mid \mathbf{x}_a \bullet^{(c)} \mathbf{x}_b = KQ_{ab}[c] \right\} \quad (176)$$

where the inner product of two different vectors \mathbf{x}_a and \mathbf{x}_b is constant.⁵ The splitting of the $K(n+1)$ -dimensional space is depending on the chip time c . With this splitting of the space, we find⁶

$$\Xi_n = \int_{\mathbb{R}^{N(n+1)(n+2)/2}} e^{K\mathcal{I}\{Q[\cdot]\}} \prod_{c=1}^N e^{\mathcal{G}\{Q[c]\}} \prod_{a \leq b} dQ_{ab}[c], \quad (177)$$

where

$$e^{K\mathcal{I}\{Q[\cdot]\}} = \int \left[\prod_{a \leq b} \prod_{c=1}^N \delta \left(\frac{\mathbf{x}_a \bullet^{(c)} \mathbf{x}_b}{N} - \beta Q_{ab}[c] \right) \right] \prod_{a=0}^n dP_a(\mathbf{x}_a) \quad (178)$$

denotes the probability weight of the subshell and

$$e^{\mathcal{G}\{Q[c]\}} = \frac{1}{\sqrt{2\pi}\sigma_0} \int_{\mathbb{R}} \mathbb{E}_{\mathbf{H}} \prod_{a=0}^n e^{-\frac{\beta}{2\sigma_a^2} \left(\frac{y_c}{\sqrt{\beta}} - v_{ac}\{Q[c]\} \right)^2} dy_c. \quad (179)$$

This procedure is a change of integration variables in multiple dimensions where the integration of an exponential function over the replicas has been replaced by integration over the variables $Q_{ab}[\cdot]$. In the following the two exponential terms in (177) are evaluated separately.

First, we turn to the evaluation of the measure $e^{K\mathcal{I}\{Q[\cdot]\}}$. Since for some $t \in \mathbb{R}$, we have the Fourier expansions of the Dirac measure

$$\begin{aligned} \delta \left(\frac{\mathbf{x}_a \bullet^{(c)} \mathbf{x}_b}{N} - \beta Q_{ab}[c] \right) &= \\ &= \frac{1}{2\pi j} \int_{\mathcal{J}} \exp \left[\tilde{Q}_{ab}[c] \left(\frac{\mathbf{x}_a \bullet^{(c)} \mathbf{x}_b}{N} - \beta Q_{ab}[c] \right) \right] d\tilde{Q}_{ab}[c] \end{aligned} \quad (180)$$

with $\mathcal{J} = (t - j\infty; t + j\infty)$, the measure $e^{K\mathcal{I}\{Q[\cdot]\}}$ can be expressed in the following way:

$$\begin{aligned} e^{K\mathcal{I}\{Q[\cdot]\}} &= \\ &= \int \left[\prod_{c=1}^N \prod_{a \leq b} \int_{\mathcal{J}} e^{\tilde{Q}_{ab}[c] \left(\frac{\mathbf{x}_a \bullet^{(c)} \mathbf{x}_b}{N} - \beta Q_{ab}[c] \right)} \frac{d\tilde{Q}_{ab}[c]}{2\pi j} \right] \end{aligned}$$

⁵The notation $f\{Q[\cdot]\}$ expresses the dependency of the function $f(\cdot)$ on all $Q_{ab}[c], 0 \leq a \leq b \leq n, 1 \leq c \leq N$.

⁶The notation $\prod_{a \leq b}$ is used as shortcut for $\prod_{a=0}^n \prod_{b=a}^n$.

$$F(\mathbf{x}) = -\frac{1}{K} \frac{E}{H} \int \int_{\mathbb{R}^N} \log \left(\int e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{H}\mathbf{x})^T(\mathbf{y}-\mathbf{H}\mathbf{x})} d\check{P}_{\mathbf{x}}(\mathbf{x}) \right) \frac{e^{-\frac{1}{2\sigma_0^2}(\mathbf{y}-\mathbf{H}\mathbf{x})^T(\mathbf{y}-\mathbf{H}\mathbf{x})}}{(2\pi\sigma_0^2)^{\frac{N}{2}}} d\mathbf{y} dP_{\mathbf{x}}(\mathbf{x}) \quad (169)$$

$$= -\frac{1}{K} \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \log \frac{E}{H} \int \int_{\mathbb{R}^N} \left(\int e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{H}\mathbf{x})^T(\mathbf{y}-\mathbf{H}\mathbf{x})} d\check{P}_{\mathbf{x}}(\mathbf{x}) \right)^n \frac{e^{-\frac{1}{2\sigma_0^2}(\mathbf{y}-\mathbf{H}\mathbf{x})^T(\mathbf{y}-\mathbf{H}\mathbf{x})}}{(2\pi\sigma_0^2)^{\frac{N}{2}}} d\mathbf{y} dP_{\mathbf{x}}(\mathbf{x}) \quad (170)$$

$$= -\frac{1}{K} \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \log \underbrace{\int_{\mathbb{R}^N} \frac{E}{H} \prod_{a=0}^n e^{-\frac{1}{2\sigma_a^2}(\mathbf{y}-\mathbf{H}\mathbf{x}_a)^T(\mathbf{y}-\mathbf{H}\mathbf{x}_a)} d\mathbf{y}}_{\triangleq \varepsilon_n} \prod_{a=0}^n dP_a(\mathbf{x}_a) \quad (171)$$

$$\prod_{a=0}^n dP_a(\mathbf{x}_a) \quad (181)$$

$$= \int_{\mathcal{J}^{N(n+2)(n+1)/2}} e^{-\beta \sum_{c=1}^N \sum_{a \leq b} \tilde{Q}_{ab}[c] Q_{ab}[c]} \left(\prod_{k=1}^K M_k \{ \tilde{Q}[\cdot] \} \right)$$

$$\prod_{c=1}^N \prod_{a \leq b} \frac{d\tilde{Q}_{ab}[c]}{2\pi^j} \quad (182)$$

with

$$M_k \{ \tilde{Q}[\cdot] \} = \int e^{\frac{1}{N} \sum_{a \leq b} \sum_{c=1}^N \tilde{Q}_{ab}[c] x_{ak} x_{bk} w_{ck}} \prod_{a=0}^n dP_a(x_{ak}). \quad (183)$$

In the limit of $K \rightarrow \infty$ one of the exponential terms in (177) will dominate over all others. Thus, only the maximum value of the correlation $Q_{ab}[c]$ is relevant for calculation of the integral.

At this point, we assume that the replicas within the dominant subshell are symmetric. This means, that we assume that the maximum values of the correlations $Q_{ab}[c]$ are identical for all positive $a \neq b$. An identical assumption is made for the maximum values of the correlations $Q_{a0}[c]$. Thus, we reduce the number of different correlation variables from $(n+1)(n+2)/2$ to four per chip time and can assume without loss of generality $Q_{00}[c] = p_{0c}$, $Q_{0a}[c] = m_c, \forall a \neq 0$, $Q_{aa}[c] = p_c, \forall a \neq 0$, $Q_{ab}[c] = q_c, \forall 0 \neq a \neq b \neq 0$. We apply the same idea to the correlation variables in the Fourier domain and set $\tilde{Q}_{00}[c] = G_{0c}/2$, $\tilde{Q}_{aa}[c] = G_c/2, \forall a \neq 0$, $\tilde{Q}_{0a}[c] = E_c, \forall a \neq 0$, and $\tilde{Q}_{ab}[c] = F_c, \forall 0 \neq a \neq b \neq 0$.

At this point the crucial benefit of the replica method becomes obvious. Assuming replica continuity, we have managed to reduce the evaluation of a continuous function to sampling it at integer points. Assuming replica symmetry we have reduced the task of evaluating infinitely many integer points to calculating 8 different correlations (4 of them in the original and 4 of them in the Fourier domain).

The assumption of replica symmetry leads to

$$\sum_{a \leq b} \tilde{Q}_{ab}[c] Q_{ab}[c] = nE_c m_c + \frac{n(n-1)}{2} F_c q_c + \frac{G_{0c} p_{0c}}{2} + \frac{n}{2} G_c p_c \quad (184)$$

and (185) where

$$\tilde{E}_k \triangleq \frac{1}{N} \sum_{c=1}^N E_c w_{ck} \quad (186)$$

$$\tilde{F}_k \triangleq \frac{1}{N} \sum_{c=1}^N F_c w_{ck} \quad (187)$$

$$\tilde{G}_k \triangleq \frac{1}{N} \sum_{c=1}^N G_c w_{ck} \quad (188)$$

$$\tilde{G}_{0k} \triangleq \frac{1}{N} \sum_{c=1}^N G_{0c} w_{ck}. \quad (189)$$

This result cannot be simplified further without an explicit assumption on the prior distribution.

For the evaluation of $e^{\mathcal{G}\{Q[c]\}}$ in (177), we can use the replica symmetry to construct the correlated Gaussian random variables v_{ac} out of independent zero-mean, unit-variance Gaussian random variables u_c, t_c, z_{ac} by

$$v_{0c} = u_c \sqrt{p_{0c} - \frac{m_c^2}{q_c}} - t_c \frac{m_c}{\sqrt{q_c}} \quad (190)$$

$$v_{ac} = z_{ac} \sqrt{p_c - q_c} - t_c \sqrt{q_c}, \quad a > 0. \quad (191)$$

With that substitution, we get (193) with the Gaussian measure $Dz = \exp(-z^2/2) dz / \sqrt{2\pi}$. Since the integral in (177) is dominated by the maximum argument of the exponential function, the derivatives of

$$\frac{1}{N} \sum_{c=1}^N \left(\mathcal{G}\{Q[c]\} - \beta \sum_{a \leq b} \tilde{Q}_{ab}[c] Q_{ab}[c] \right) \quad (194)$$

with respect to m_c, q_c, p_c and p_{0c} must vanish as $N \rightarrow \infty$. Taking derivatives after plugging (184) and (193) into

$$\begin{aligned}
M_k\{E, F, G, G_0\} &= \int_{\mathbb{R}^{n+1}} e^{\frac{1}{N} \sum_{c=1}^N w_{ck} \left(\frac{G_{0c}}{2} x_{0k}^2 + \sum_{a=1}^n E_c x_{0k} x_{ak} + \frac{G_c}{2} x_{ak}^2 + \sum_{b=a+1}^n F_c x_{ak} x_{bk} \right)} \prod_{a=0}^n dP_a(x_{ak}) \\
&= \int_{\mathbb{R}^{n+1}} e^{\frac{\tilde{G}_{0k}}{2} x_{0k}^2 + \sum_{a=1}^n \tilde{E}_k x_{0k} x_{ak} + \frac{\tilde{G}_k}{2} x_{ak}^2 + \sum_{b=a+1}^n \tilde{F}_k x_{ak} x_{bk}} \prod_{a=0}^n dP_a(x_{ak}) \tag{185}
\end{aligned}$$

$$e^{\mathcal{G}(m_c, q_c, p_c, p_{0c})} = \frac{1}{\sqrt{2\pi}\sigma_0} \int_{\mathbb{R}^2} \int_{\mathbb{R}} \exp \left[-\frac{\beta}{2\sigma_0^2} \left(u_c \sqrt{p_{0c} - \frac{m_c^2}{q_c}} - \frac{t_c m_c}{\sqrt{q_c}} - \frac{y_c}{\sqrt{\beta}} \right)^2 \right] Du_c \tag{192}$$

$$\begin{aligned}
&\times \left[\int_{\mathbb{R}} \exp \left[-\frac{\beta}{2\sigma^2} \left(z_c \sqrt{p_c - q_c} - t_c \sqrt{q_c} - \frac{y_c}{\sqrt{\beta}} \right)^2 \right] Dz_c \right]^n Dt_c dr_c \\
&= \sqrt{\frac{(1 + \frac{\beta}{\sigma^2}(p_c - q_c))^{1-n}}{1 + \frac{\beta}{\sigma^2}(p_c - q_c) + n \frac{\beta}{\sigma^2} \left(\frac{\sigma_0^2}{\beta} + p_{0c} - 2m_c + q_c \right)}} \tag{193}
\end{aligned}$$

(194), solving for E_c, F_c, G_c , and G_{0c} and letting $n \rightarrow 0$ yields for all c

$$E_c = \frac{1}{\sigma^2 + \beta(p_c - q_c)} \tag{195}$$

$$F_c = \frac{\sigma_0^2 + \beta(p_{0c} - 2m_c + q_c)}{[\sigma^2 + \beta(p_c - q_c)]^2} \tag{196}$$

$$G_c = F_c - E_c \tag{197}$$

$$G_{0c} = 0. \tag{198}$$

In order to proceed with the calculations, a prior distribution is specified. In the following, the calculations for Gaussian and binary priors are shown.

F.1 Gaussian Prior Distribution

Assume a Gaussian prior distribution

$$p_a(x_{ak}) = \frac{1}{\sqrt{2\pi}} e^{-x_{ak}^2/2} \quad \forall a. \tag{199}$$

Thus, the integration in (185) can be performed explicitly and we find with [56, (87)]

$$\begin{aligned}
M_k(\tilde{E}_k, \tilde{F}_k, \tilde{G}_k, \tilde{G}_{0k}) &= \tag{200} \\
&\sqrt{\frac{(1 + \tilde{F}_k - \tilde{G}_k)^{1-n}}{(1 - \tilde{G}_{0k})(1 + \tilde{F}_k - \tilde{G}_k - n\tilde{F}_k) - n\tilde{E}_k^2}}.
\end{aligned}$$

In the large system limit, the integral in (182) is also dominated by that value of the integration variable which maximizes the argument of the exponential function under some weak conditions on the variances w_{ck} . Thus, derivations of

$$-\beta \sum_{c=1}^N n E_c m_c + \frac{n(n-1)}{2} F_c q_c + \frac{G_{0c} p_{0c}}{2} + \frac{n}{2} G_c p_c +$$

$$+ \log \prod_{k=1}^K M_k(\tilde{E}_k, \tilde{F}_k, \tilde{G}_k, \tilde{G}_{0k}) \tag{201}$$

with respect to E_c, F_c, G_c, G_{0c} must vanish for all c as $N \rightarrow \infty$. An explicit calculation of these derivatives yields

$$m_c = \frac{1}{K} \sum_{k=1}^K w_{ck} \frac{\tilde{E}_k}{1 + \tilde{E}_k} \tag{202}$$

$$q_c = \frac{1}{K} \sum_{k=1}^K w_{ck} \frac{\tilde{E}_k^2 + \tilde{F}_k}{(1 + \tilde{E}_k)^2} \tag{203}$$

$$p_c = \frac{1}{K} \sum_{k=1}^K w_{ck} \frac{\tilde{E}_k^2 + \tilde{E}_k + \tilde{F}_k + 1}{(1 + \tilde{E}_k)^2} \tag{204}$$

$$p_{0c} = \frac{1}{K} \sum_{k=1}^K w_{ck} \tag{205}$$

in the limit $n \rightarrow 0$ with (197) and (198). Surprisingly, if we let the true prior to be binary and only the replicas to be Gaussian we also find (202) to (205). Note from Section II-C that this setting corresponds to linear MMSE detection.

Returning to our initial goal, the evaluation of the free energy, and collecting our previous results, we find

$$\begin{aligned}
&-\frac{1}{K} \frac{\partial}{\partial n} \log \Xi_n = \\
&= \frac{1}{K} \frac{\partial}{\partial n} \sum_{c=1}^N \left[-\mathcal{G}(m_c, q_c, p_c, p_{0c}) + \beta n E_c m_c + \right. \\
&\quad \left. + \frac{\beta n(n-1)}{2} F_c q_c + \frac{\beta n}{2} G_c p_c \right] \\
&\quad - \sum_{k=1}^K \log M_k(\tilde{E}_k, \tilde{F}_k, \tilde{G}_k, 0) \tag{206}
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2K} \left[\sum_{c=1}^N \log \left(1 + \frac{\beta}{\sigma^2} (p_c - q_c) \right) + \right. \\
&\quad \left. + 2\beta E_c m_c + \beta(2n-1) F_c q_c + \beta G_c p_c \right. \\
&\quad \left. + \frac{\sigma_0^2 + \beta(p_{0c} - 2m_c + q_c)}{\sigma^2 + \beta(p_c - q_c) + n\sigma_0^2 + n\beta(p_{0c} - 2m_c + q_c)} \right] \\
&\quad + \frac{1}{2K} \sum_{k=1}^K \log \left(1 + \tilde{E}_k \right) - \frac{\tilde{E}_k^2 + \tilde{F}_k}{1 + \tilde{E}_k - n\tilde{E}_k^2 - n\tilde{F}_k} \\
&\xrightarrow{n \rightarrow 0} \frac{1}{2K} \left[\sum_{c=1}^N \log \left(1 + \frac{\beta}{\sigma^2} (p_c - q_c) \right) + \right. \\
&\quad \left. \frac{E_c}{F_c} + 2\beta E_c m_c - \beta F_c q_c + \beta G_c p_c \right] \\
&\quad + \frac{1}{2K} \sum_{k=1}^K \log \left(1 + \tilde{E}_k \right) - \frac{\tilde{E}_k^2 + \tilde{F}_k}{1 + \tilde{E}_k} \\
&= F(\mathbf{x}). \tag{209}
\end{aligned}$$

This is the final result for the free energy of the mismatched detector assuming noise variance σ^2 instead of the true noise variance σ_0^2 . The six macroscopic parameters $E_c, F_c, G_c, m_c, q_c, p_c$ are implicitly given by the simultaneous solution of the system of equations (195) to (197) and (202) to (204) with the definitions (186) to (188) for all chip times c . This system of equations can only be solved numerically.

Specializing our result to the matched detector assuming the true noise variance by letting $\sigma \rightarrow \sigma_0$, we have $F_c \rightarrow E_c$, $G_c \rightarrow G_{0c}$, $q_c \rightarrow m_c$, $p_c \rightarrow p_{0c}$. This makes the free energy simplify to

$$\begin{aligned}
F(\mathbf{x}) &= \frac{1}{2K} \left[\sum_{c=1}^N \log \left(1 + \frac{\beta}{\sigma_0^2} (p_{0c} - m_c) \right) + 1 + \beta E_c m_c \right] \\
&\quad + \frac{1}{2K} \sum_{k=1}^K \log \left(1 + \tilde{E}_k \right) - \tilde{E}_k \\
&= \frac{1}{2K} \left[\sum_{c=1}^N \sigma_0^2 E_c - \log(\sigma_0^2 E_c) \right] \\
&\quad + \frac{1}{2K} \sum_{k=1}^K \log \left(1 + \tilde{E}_k \right) \tag{211}
\end{aligned}$$

with

$$E_c = \frac{1}{\sigma_0^2 + \frac{\beta}{K} \sum_{k=1}^K \frac{w_{ck}}{1 + \tilde{E}_k}}. \tag{212}$$

This result is more compact and it requires only to solve (212) numerically which is conveniently done by fixed-point iteration.

It can be shown that the parameter \tilde{E}_k is actually the signal-to-interference and noise ratio of user k . It has been derived independently by Hanly and Tse [20] in context of CDMA with macro-diversity using known results from random matrix theory, cf. Section III-B.3.

Using the similarity of free energy and the entropy of the channel output mentioned at the end of Section V-C,

allows for a simple relationship

$$I(\mathbf{x}, \mathbf{y}) = F(\mathbf{x}) - \frac{1}{2\beta} \tag{213}$$

between free energy and the (normalized) mutual information between channel input signal \mathbf{x} and channel output signal \mathbf{y} given the channel matrix \mathbf{H} . Assuming that the channel is perfectly known to the receiver, but totally unknown to the transmitter, (213) gives the channel capacity per user.

F.2 Binary Prior Distribution

Now, we assume a non-uniform binary prior distribution

$$p_a(x_{ak}) = \frac{1+t_k}{2} \delta(x_{ak} - 1) + \frac{1-t_k}{2} \delta(x_{ak} + 1). \tag{214}$$

Plugging the prior distribution into (185), we find (217) where we used the following property of the Gaussian measure

$$\exp \left(\tilde{F}_k \frac{S^2}{2} \right) = \int \exp \left(\pm \sqrt{\tilde{F}_k} z S \right) D z \quad \forall S \in \mathbb{R} \tag{219}$$

to linearize the exponents. Since

$$f_n \triangleq \sum_{\{x_{ka}, a=1, \dots, n\}} e^{(z\sqrt{\tilde{F}_k} + \tilde{E}_k) \sum_{a=1}^n x_{ka}} \prod_{a=1}^n \Pr(x_{ka}) \tag{220}$$

$$= \sum_{x_{kn}} \Pr(x_{kn}) f_{n-1} e^{(z\sqrt{\tilde{F}_k} + \tilde{E}_k) x_{kn}} \tag{221}$$

$$= f_{n-1} \frac{\cosh \left[\lambda_k/2 + z\sqrt{\tilde{F}_k} + \tilde{E}_k \right]}{\cosh(\lambda_k/2)} \tag{222}$$

$$= \frac{\cosh^n \left[\lambda_k/2 + z\sqrt{\tilde{F}_k} + \tilde{E}_k \right]}{\cosh^n(\lambda_k/2)} \tag{223}$$

with $t_k \triangleq \tanh(\lambda_k/2)$, we find

$$\begin{aligned}
M_k \left(\tilde{E}_k, \tilde{F}_k, \tilde{G}_k, \tilde{G}_{0k} \right) &= \\
&= \frac{\int \frac{1+t_k}{2} \cosh^n \left(z\sqrt{\tilde{F}_k} + \tilde{E}_k + \frac{\lambda_k}{2} \right) D z}{\cosh^n \left(\frac{\lambda_k}{2} \right) \exp \left(\frac{n\tilde{F}_k - \tilde{G}_{0k} - n\tilde{G}_k}{2} \right)} + \\
&\quad + \frac{\int \frac{1-t_k}{2} \cosh^n \left(z\sqrt{\tilde{F}_k} + \tilde{E}_k - \frac{\lambda_k}{2} \right) D z}{\cosh^n \left(\frac{\lambda_k}{2} \right) \exp \left(\frac{n\tilde{F}_k - \tilde{G}_{0k} - n\tilde{G}_k}{2} \right)} \tag{224}
\end{aligned}$$

In the large system limit, the integral in (182) is dominated by that value of the integration variable which maximizes the argument of the exponential function under some weak conditions on the variances w_{ck} . Thus, derivations of

$$\begin{aligned}
&-\beta \sum_{c=1}^N n E_c m_c + \frac{n(n-1)}{2} F_c q_c + \frac{G_{0c} p_{0c}}{2} + \frac{n}{2} G_c p_c + \\
&\quad + \log \prod_{k=1}^K M_k \left(\tilde{E}_k, \tilde{F}_k, \tilde{G}_k, \tilde{G}_{0k} \right) \tag{225}
\end{aligned}$$

$$M_k \left(\tilde{E}_k, \tilde{F}_k, \tilde{G}_k, \tilde{G}_{0k} \right) = \int_{\mathbb{R}^{n+1}} e^{\frac{\tilde{G}_{0k} + n\tilde{G}_k}{2} + \sum_{a=1}^n \tilde{E}_k x_{0k} x_{ak} + \sum_{b=a+1}^n \tilde{F}_k x_{ak} x_{bk}} \prod_{a=0}^n dP_a(x_{ak}) \quad (215)$$

$$= e^{\frac{1}{2}(\tilde{G}_{0k} + n\tilde{G}_k)} \sum_{\{x_{ak}, a=1, \dots, n\}} \left\{ \frac{1+t_k}{2} \exp \left[\sum_{a=1}^n \tilde{E}_k x_{ak} + \sum_{b=a+1}^n \tilde{F}_k x_{ak} x_{bk} \right] + \frac{1-t_k}{2} \exp \left[\sum_{a=1}^n -\tilde{E}_k x_{ak} + \sum_{b=a+1}^n \tilde{F}_k x_{ak} x_{bk} \right] \right\} \prod_{a=1}^n \Pr(x_{ak}) \quad (216)$$

$$= e^{\frac{1}{2}(\tilde{G}_{0k} + n\tilde{G}_k - n\tilde{F}_k)} \sum_{\{x_{ak}, a=1, \dots, n\}} \left\{ \frac{1+t_k}{2} \exp \left[\frac{\tilde{F}_k}{2} \left(\sum_{a=1}^n x_{ak} \right)^2 + \tilde{E}_k \sum_{a=1}^n x_{ak} \right] + \frac{1-t_k}{2} \exp \left[\frac{\tilde{F}_k}{2} \left(\sum_{a=1}^n x_{ak} \right)^2 - \tilde{E}_k \sum_{a=1}^n x_{ak} \right] \right\} \prod_{a=1}^n \Pr(x_{ak}) \quad (217)$$

$$= e^{\frac{1}{2}(\tilde{G}_{0k} + n\tilde{G}_k - n\tilde{F}_k)} \sum_{\{x_{ak}, a=1, \dots, n\}} \int \frac{1+t_k}{2} \exp \left[\left(z\sqrt{\tilde{F}_k} + \tilde{E}_k \right) \sum_{a=1}^n x_{ak} \right] + \frac{1-t_k}{2} \exp \left[- \left(z\sqrt{\tilde{F}_k} + \tilde{E}_k \right) \sum_{a=1}^n x_{ak} \right] Dz \prod_{a=1}^n \Pr(x_{ak}) \quad (218)$$

with respect to E_c, F_c, G_c, G_{0c} must vanish for all c as $N \rightarrow \infty$. An explicit calculation of these derivatives gives

$$m_c = \frac{1}{K} \sum_{k=1}^K w_{ck} \int \frac{1+t_k}{2} \tanh \left(z\sqrt{\tilde{F}_k} + \tilde{E}_k + \frac{\lambda_k}{2} \right) + \frac{1-t_k}{2} \tanh \left(z\sqrt{\tilde{F}_k} + \tilde{E}_k - \frac{\lambda_k}{2} \right) Dz \quad (226)$$

$$q_c = \frac{1}{K} \sum_{k=1}^K w_{ck} \int \frac{1+t_k}{2} \tanh^2 \left(z\sqrt{\tilde{F}_k} + \tilde{E}_k + \frac{\lambda_k}{2} \right) + \frac{1-t_k}{2} \tanh^2 \left(z\sqrt{\tilde{F}_k} + \tilde{E}_k - \frac{\lambda_k}{2} \right) Dz \quad (227)$$

$$p_c = p_{0c} = \frac{1}{K} \sum_{k=1}^K w_{ck} \quad (228)$$

in the limit $n \rightarrow 0$. In order to obtain (227), note from (217) that the first order derivative of $M_k \exp(n\tilde{F}_k/2)$ with respect to \tilde{F}_k is identical to half of the second order derivative of $M_k \exp(n\tilde{F}_k/2)$ with respect to \tilde{E}_k .

Returning to our initial goal, the evaluation of the free energy, and collecting our previous results, we find

$$-\frac{1}{K} \frac{\partial}{\partial n} \log \Xi_n = \frac{1}{K} \frac{\partial}{\partial n} \sum_{c=1}^N \left[-\mathcal{G}(m_c, q_c, p_c, p_{0c}) + \beta n E_c m_c + \frac{\beta n(n-1)}{2} F_c q_c + \frac{\beta n}{2} G_c p_c \right] - \sum_{k=1}^K \log M_k \left(\tilde{E}_k, \tilde{F}_k, \tilde{G}_k, 0 \right) \quad (229)$$

$$\begin{aligned} & \xrightarrow{n \rightarrow 0} \frac{1}{2K} \sum_{c=1}^N \left[\log \left(1 + \frac{\beta}{\sigma^2} (p_c - q_c) \right) + \frac{E_c}{F_c} + 2\beta E_c m_c - \beta F_c q_c + \beta G_c p_c \right] \\ & - \frac{1}{K} \sum_{k=1}^K \int \frac{1+t_k}{2} \log \cosh \left(z\sqrt{\tilde{F}_k} + \tilde{E}_k + \frac{\lambda_k}{2} \right) + \frac{1-t_k}{2} \log \cosh \left(z\sqrt{\tilde{F}_k} + \tilde{E}_k - \frac{\lambda_k}{2} \right) Dz + \\ & + \frac{1}{2} \log(1-t_k^2) - \frac{\tilde{F}_k + \tilde{G}_k}{2} \quad (230) \\ & = \mathbf{F}(\mathbf{x}). \quad (231) \end{aligned}$$

This is the final result for the free energy of the mismatched detector assuming noise variance σ^2 instead of the true noise variance σ_0^2 . The six macroscopic parameters $E_c, F_c, G_c, m_c, q_c, p_c$ are implicitly given by the simultaneous solution of the system of equations (195) to (197) and (226) to (228) with the definitions (186) to (188) for all chip times c . This system of equations can only be solved numerically. Moreover, it can have multiple solutions. In case of multiple solutions, the correct solution is that one which minimizes the free energy, since in the thermodynamic equilibrium the free energy is always minimized, cf. Section V-B.

Specializing our result to the matched detector assuming the true noise variance by letting $\sigma \rightarrow \sigma_0$, we have $F_c \rightarrow E_c, G_c \rightarrow G_{0c}, q_c \rightarrow m_c, p_c \rightarrow p_{0c}$ which makes the free energy simplify to

$$\mathbf{F}(\mathbf{x}) = \frac{1}{2K} \sum_{c=1}^N \left[\log \left(1 + \frac{\beta}{\sigma_0^2} (p_{0c} - m_c) \right) + 1 + \beta E_c m_c \right]$$

$$\begin{aligned}
& -\frac{1}{K} \sum_{k=1}^K \frac{1}{2} \log(1 - t_k^2) - \frac{\tilde{E}_k}{2} \\
& + \int \frac{1+t_k}{2} \log \cosh \left(z \sqrt{\tilde{E}_k} + \tilde{E}_k + \frac{\lambda_k}{2} \right) \\
& + \frac{1-t_k}{2} \log \cosh \left(z \sqrt{\tilde{E}_k} + \tilde{E}_k - \frac{\lambda_k}{2} \right) Dz \quad (232) \\
= & \frac{1}{2K} \sum_{c=1}^N [\sigma_0^2 E_c - \log(\sigma_0^2 E_c)] \\
& -\frac{1}{K} \sum_{k=1}^K \frac{1}{2} \log(1 - t_k^2) - \tilde{E}_k \\
& + \int \frac{1+t_k}{2} \log \cosh \left(z \sqrt{\tilde{E}_k} + \tilde{E}_k + \frac{\lambda_k}{2} \right) + \\
& + \frac{1-t_k}{2} \log \cosh \left(z \sqrt{\tilde{E}_k} + \tilde{E}_k - \frac{\lambda_k}{2} \right) Dz \quad (233)
\end{aligned}$$

where the macroscopic parameters E_c are given by

$$\begin{aligned}
\frac{1}{E_c} = & \sigma_0^2 + \frac{\beta}{K} \sum_{k=1}^K w_{ck} \left[1 - \int \frac{1+t_k}{2} \tanh \left(z \sqrt{\tilde{E}_k} + \tilde{E}_k + \frac{\lambda_k}{2} \right) \right. \\
& \left. + \frac{1-t_k}{2} \tanh \left(z \sqrt{\tilde{E}_k} + \tilde{E}_k - \frac{\lambda_k}{2} \right) Dz \right] \quad (234)
\end{aligned}$$

$$\begin{aligned}
= & \sigma_0^2 + \frac{\beta}{K} \sum_{k=1}^K w_{ck} (1 - t_k^2) \\
& \int \frac{1 - \tanh \left(z \sqrt{\tilde{E}_k} + \tilde{E}_k \right)}{1 - t_k^2 \tanh^2 \left(z \sqrt{\tilde{E}_k} + \tilde{E}_k \right)} Dz. \quad (235)
\end{aligned}$$

Similar to the case of Gaussian priors, \tilde{E}_k can be shown to be a kind of signal-to-interference and noise ratio, in the sense that the bit error probability of user k is given by

$$\Pr(\hat{x}_k \neq x_k) = \int_{\frac{1}{\sqrt{\tilde{E}_k}}}^{\infty} Dz. \quad (236)$$

In fact, it can even be shown that in the large system limit, an equivalent additive white Gaussian noise channel can be defined to model the multiuser interference [64]. Constraining the input alphabet of the channel to follow the non-uniform binary distribution (214) and assuming channel state information being available only at the transmitter, channel capacity is given by (213) with the free energy given in (233).

Large system results for binary prior (even for uniform binary prior) cannot be obtained using results on the asymptotic eigenvalue distributions of large random matrices. Only for the case of vanishing noise variance a result is reported by Tse and Verdú [68]. This is, as for binary priors, the dependence of performance measures, such as channel capacity or signal-to-interference-and-noise ratio, cannot solely be described by eigenvalues and eigenvectors of the covariance matrix of the channel, but they also depend on the distance profile of the signal points. Nevertheless, performance measures are macroscopic variables and, as such, have deterministic asymptotic limits.

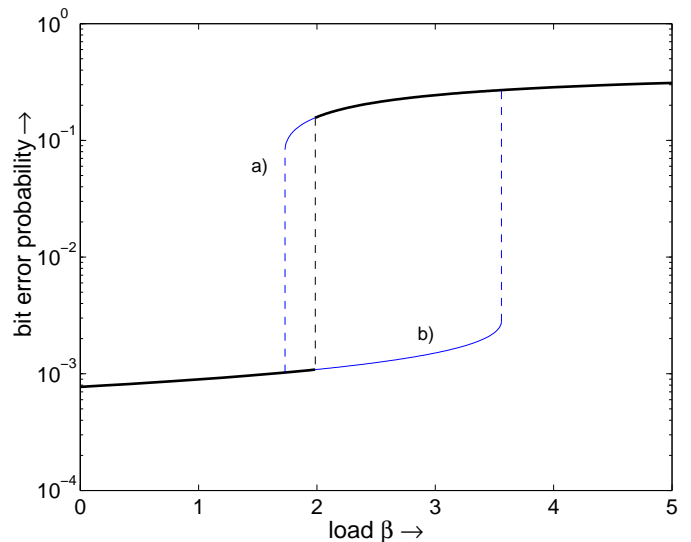


Fig. 2. Bit error probability for the individually optimum detector with uniform binary prior distribution versus system load for $10 \log_{10}(E_s/N_0) = 6$ dB.

G. 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. The similarity between thermodynamics and multiuser detection pointed out in Section V-F, should be sufficient to convince the reader that phase transitions in digital communications do occur. Phase transitions in turbo decoding and detection of CDMA were found in [69] and [56], respectively.

The phase transitions in digital communications are similar to the hysteresis in ferro-magnetic materials. They occur if the equations determining the macroscopic parameters, e.g. E_c determined by (235), 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 property 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. It is related to the macroscopic property \tilde{E}_k in (235) by (236) for the case considered in Section V-F. Numerical results are depicted in Fig. 2. 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 alternative solutions to the equations for the macroscopic variable \tilde{E}_k . Only for a certain interval of the load, approximately $1.73 \leq \beta \leq 3.56$ in Fig. 2, multiple solutions occur. As expected, 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.

In order to observe a hysteresis behavior, we can expand our scope to neural networks. Consider a Hopfield neural network [70] implementation of the individually optimum multiuser detector which is an algorithm based on non-linear gradient search maximizing the energy function associated with the detector. Its application to the problem of multiuser detection is discussed in [71]. With appropriate definition of the energy function, such a detector will achieve the performance of the upper curve in Fig. 2 in the large system limit. Thus, in the interval $1.73 \leq \beta \leq 1.986$ where the free energy favors the curve with lower bit error probability, the Hopfield neural network is suboptimum (labeled with a). The curve labeled with b can also be achieved by the Hopfield neural network, but only with the help of a genie. In order to achieve a point in that area, cancel with the help of a genie as many interferers as needed to push the load below the area where multiple solutions occur, i.e. $\beta < 1.73$. Then, initialize the Hopfield neural network with the received vector where the interference has been canceled and let it converge to the thermodynamic equilibrium. Then, slowly add one by one the interferers you had canceled with the help of the genie while the Hopfield neural network remains in the thermodynamic equilibrium by performing iterations. If all the interference suppressed by the genie has been added again, the targeted point on the lower curve in area b is reached. The Hopfield neural network follows the lower curve, if interference is added, and it follows the upper line, if it is removed.

It should be remarked that a hysteresis behavior of the Hopfield neural network detector does not occur for any definition of the energy function and any prior distribution of the data to be detected, but additional conditions on the microscopic configuration of the system need to be fulfilled.

ACKNOWLEDGMENT

The author would like to thank Antonia Tulino, Toshiyuki Tanaka, Sergio Verdú, Giuseppe Caire, Stephen Hanly, Laura Cottatellucci, and Nurul Mahmood.

REFERENCES

- [1] Madan Lal Mehta, *Random Matrices*, Academic Press, Boston, MA, 2nd edition, 1991.
- [2] Sergio Verdú, *Multiuser Detection*, Cambridge University Press, New York, 1998.
- [3] John G. Proakis, *Digital Communications*, McGraw-Hill, New York, 4th edition, 2000.
- [4] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 1991.
- [5] Shimon Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Communications Magazine*, vol. 34, no. 10, pp. 124-136, Oct. 1996.
- [6] Peter Hilton and Jean Pedersen, "Catalan numbers, their generalization, and their uses," *The Mathematical Intelligencer*, vol. 13, no. 2, pp. 64-75, 1991.
- [7] Athanasios Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 3rd edition, 1991.
- [8] Sergio Verdú and Shlomo Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 622-640, Mar. 1999.
- [9] Fredy D. Neeser and James L. Massey, "Proper complex random processes with applications to information theory," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1293-1302, July 1993.
- [10] Jack W. Silverstein, "Eigenvalues and eigenvectors of large dimensional sample covariance matrices," in *Random Matrices and Their Applications*, Joel E. Cohen, Harry Kesten, and Charles M. Newman, Eds., pp. 153-159. American Mathematical Society, Providence, RI, 1986.
- [11] T. J. Stieltjes, "Recherches sur les fractions continues," *Annales de la Faculté des Sciences de Toulouse*, vol. 8, pp. 1-122, 1894.
- [12] Fumio Hiai and Dénes Petz, *The Semicircle Law, Free Random Variables and Entropy*, American Mathematical Society, Providence, RI, 2000.
- [13] Jack W. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 55, pp. 331-339, 1995.
- [14] Y. Q. Yin, "Limiting spectral distribution for a class of random matrices," *Journal of Multivariate Analysis*, vol. 20, pp. 50-68, 1986.
- [15] Jack W. Silverstein and Z.D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, pp. 175-192, 1995.
- [16] 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.
- [17] David Tse and Stephen Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 641-657, Mar. 1999.
- [18] Shlomo Shamai (Shitz) and Sergio Verdú, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1302-1327, May 2001.
- [19] Vjačeslav L. Girko, *Theory of Random Determinants*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990.
- [20] Stephen V. Hanly and David N.C. Tse, "Resource pooling and effective bandwidth in CDMA networks with multiuser receivers and spatial diversity," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1328-1351, May 2001.
- [21] Giuseppe Caire and Ralf R. Müller, "The optimal received power distribution for IC-based iterative multiuser joint decoders," in *Proc. of 39th Annual Allerton Conference on Communications, Control, and Computing*, Monticello, IL, Oct. 2001.
- [22] Owe Axelsson, *Iterative Solution Methods*, Cambridge University Press, New York, 1994.
- [23] Louis G. F. Trichard, Jamie S. Evans, and Iain B. Collings, "Large system analysis of linear multistage parallel interference cancellation," *IEEE Transactions on Communications*, vol. 50, no. 11, pp. 1778-1786, Nov. 2002.
- [24] Roger A. Horn and Charles R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, U.K., 1985.
- [25] Shimon Moshavi, Emmanuel G. Kanterakis, and Donald L. Schilling, "Multistage linear receivers for DS-CDMA systems," *International Journal of Wireless Information Networks*, vol. 3, no. 1, pp. 1-17, Jan. 1996.
- [26] Ralf R. Müller and Sergio Verdú, "Design and analysis of low-complexity interference mitigation on vector channels," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 8, pp. 1429-1441, Aug. 2001.
- [27] J. Scott Goldstein and Irving S. Reed, "Reduced rank adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 42, no. 2, pp. 492-496, Feb. 1997.
- [28] J. Scott Goldstein, Irving S. Reed, and Louis L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Transactions on Information Theory*, vol. 44, no. 7, Nov. 1998.
- [29] Michael Honig and Weimin Xiao, "Performance of reduced-rank linear interference suppression," *IEEE Transactions on Information Theory*, vol. 47, no. 5, pp. 1928-1946, July 2001.
- [30] Louis G. F. Trichard, Jamie S. Evans, and Iain B. Collings, "Optimal linear multistage receivers and the large system SIR," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Lausanne, Switzerland, June/July 2002, p. 21.
- [31] Philippe Loubaton and Walid Hachem, "Asymptotic analysis of reduced rank Wiener filters," in *Proc. of IEEE Information Theory Workshop (ITW)*, Paris, France, Apr. 2003.
- [32] Percy Deift, *Orthogonal Polynomials and Random Matrices:*

- A Riemann–Hilbert Approach*, American Mathematical Society, Providence, RI, 1998.
- [33] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free Random Variables*, American Mathematical Society, Providence, RI, 1992.
- [34] Roland Speicher, “Freie Wahrscheinlichkeitstheorie,” Lecture Notes, 1997/98.
- [35] Dan Voiculescu, “Lectures on free probability theory,” Notes for a Course at the Saint-Flour Summer School on Probability Theory 1998, 1998.
- [36] Dan Voiculescu, “Limit laws for random matrices and free products,” *Inventiones Mathematicae*, vol. 104, pp. 201–220, 1991.
- [37] S. Thorbjørnsen, “Mixed moments of Voiculescu’s Gaussian random matrices,” *Journal of Functional Analysis*, vol. 176, pp. 213–246, 2000.
- [38] Jamie Evans and David N.C. Tse, “Large system performance of linear multiuser receivers in multipath fading channels,” *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 2059–2078, Sept. 2000.
- [39] Ezio Biglieri, Giorgio Taricco, and Antonia Tulino, “Performance of space-time codes for a large number of antennas,” *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1794–1803, July 2002.
- [40] Laura Cottatellucci and Ralf R. Müller, “Multiuser interference mitigation with multistage detectors: Design and analysis for unequal powers,” in *Proc. of Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2002.
- [41] Laura Cottatellucci and Ralf R. Müller, “Asymptotic analysis of multistage receivers for multipath fading channels,” in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Yokohama, Japan, June/July 2003.
- [42] Laura Cottatellucci and Ralf R. Müller, “A systematic approach to multistage detectors in multipath fading channels,” *Submitted to IEEE Transactions on Information Theory*, Feb. 2003.
- [43] Linbo Li, Antonia Tulino, and Sergio Verdú, “Asymptotic eigenvalue moments for linear multiuser detection,” *Communications in Information and Systems*, vol. 1, no. 3, pp. 273–304, Fall 2001.
- [44] Linbo Li, Antonia M. Tulino, and Sergio Verdú, “Asymptotic eigenvalue moments for linear multiuser detection,” in *Proc. of Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2001, pp. 1177–1181.
- [45] Mérouane Debbah, Walid Hachem, Philippe Loubaton, and Marc de Courville, “MMSE analysis of certain large isometric random precoded systems,” *IEEE Transactions on Information Theory*, vol. 49, no. 5, May 2003.
- [46] David Tse, “Multiuser receivers, random matrices and free probability,” in *Proc. of 37th Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, Sept. 1999.
- [47] Ralf R. Müller, “On the asymptotic eigenvalue distribution of concatenated vector-valued fading channels,” *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 2086–2091, July 2002.
- [48] Ralf R. Müller, “A random matrix model for communication via antenna arrays,” *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2495–2506, Sept. 2002.
- [49] Ralf R. Müller, “A random matrix model for the antenna array channel with decaying power delay profile,” in *Proc. of IEEE Information Theory Workshop (ITW)*, Cairns, Australia, Sept. 2001.
- [50] Ralf R. Müller and Helmut Hofstetter, “Confirmation of random matrix model for the antenna array channel by indoor measurements,” in *Proc. of IEEE Antennas & Propagation Soc. Intl. Symposium*, Boston, MA, July 2001, pp. 472–475.
- [51] J. Bach Andersen and Istvan Kovacs, “Power distributions revisited,” in *COST 273 TD (02) 004*, Jan. 2002.
- [52] Simon R. Saunders, *Antennas and Propagation for Wireless Communication Systems*, John Wiley & Sons, Chichester, U.K., 1999.
- [53] Hidetoshi Nishimori, *Statistical Physics of Spin Glasses and Information Processing*, Oxford University Press, Oxford, U.K., 2001.
- [54] Ralf R. Müller and Giuseppe Caire, “Efficient implementation of iterative multiuser decoding,” in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Lausanne, Switzerland, June/July 2002, p. 446.
- [55] Marc Mezard, Giorgio Parisi, and Miguel Angel Virasoro, *Spin Glass Theory and Beyond*, World Scientific, Singapore, 1987.
- [56] Toshiyuki Tanaka, “A statistical mechanics approach to large-system analysis of CDMA multiuser detectors,” *IEEE Transactions on Information Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.
- [57] Dongning Guo and Sergio 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.
- [58] Dongning Guo and Sergio Verdú, “Spectral efficiency of large-system CDMA via statistical physics,” in *Proc. of Conference on Information Science and Systems (CISS)*, Baltimore, MD, Mar. 2003.
- [59] Dongning Guo and Sergio Verdú, “Replica analysis of large-system CDMA,” in *Proc. of IEEE Information Theory Workshop (ITW)*, Paris, France, Mar./Apr. 2003.
- [60] Dongning Guo and Sergio Verdú, “CDMA spectral efficiency: Asymptotics via statistical physics,” *Submitted to IEEE Transactions on Information Theory*, 2003.
- [61] Ralf R. Müller, “On channel capacity, uncoded error probability, ML-detection, and spin glasses,” in *Proc. Workshop on Concepts in Information Theory*, Breisach, Germany, June 2002, pp. 79–81.
- [62] Ralf R. Müller, Giuseppe Caire, and Toshiyuki Tanaka, “Density evolution and power profile optimization for iterative multiuser decoders based on individually optimum multiuser detectors,” in *Proc. of 40th Annual Allerton Conference on Communication, Control & Computing*, Monticello, IL, Oct. 2002.
- [63] Ralf R. Müller and Wolfgang Gerstacker, “On the capacity loss due to separation of detection and decoding in large CDMA systems,” in *Proc. of IEEE Information Theory Workshop (ITW)*, Bangalore, India, Oct. 2002, p. 222.
- [64] Ralf R. Müller and Wolfgang H. Gerstacker, “On the capacity loss due to separation of detection and decoding,” *Submitted to IEEE Transactions on Information Theory*, Apr. 2002.
- [65] Ralf R. Müller, “Channel capacity and minimum probability of error in large dual antenna array systems with binary modulation,” *To appear in IEEE Transactions on Signal Processing*, 2003.
- [66] Ralf R. Müller and Werner Weichselberger, “Minimum probability of error and channel capacity in large dual antenna array systems,” in *Proc. of International Conference on Telecommunications (ICT)*, Papeete, French Polynesia, Feb. 2003.
- [67] Ralf R. Müller, “Performance analysis of large dual antenna array systems with binary modulation,” in *To appear in Proc. of IEEE International Symposium on Information Theory (ISIT)*, Yokohama, Japan, June/July 2003.
- [68] David N. C. Tse and Sergio Verdú, “Optimum asymptotic multiuser efficiency of randomly spread CDMA,” *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2718–2722, Nov. 2000.
- [69] Dakshi Agrawal and Alexander Vardy, “The turbo decoding algorithm and its phase trajectories,” *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 699–722, Feb. 2001.
- [70] J.J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” *Proceedings of the National Academy of Sciences of the USA*, vol. 79, pp. 2554–2558, Apr. 1982.
- [71] George I. Kechriotis and Elias S. Manolakos, “Hopfield neural network implementation of the optimal CDMA multiuser detector,” *IEEE Transactions on Neural Networks*, vol. 7, no. 1, pp. 131–141, Jan. 1996.