RANDOM MATRIX METHODS FOR DESIGN OF MULTIUSER COMMUNICATION SYSTEMS*

Ralf R. Müller

Norwegian University of Science & Technology
N–7491 Trondheim, Norway

(Received August 9, 2005)

2436

A short review on multiuser communication systems is given. System design for iterative multiuser decoding is improved by means of large system result from statistical physics and random matrix theory. With the application of multiuser detection for wireless communications in mind, it is shown how a system of linear equations with random coefficients can be solved efficiently exploiting the asymptotic convergence of its eigenvalue spectrum. In addition, the conditional convergence of the diagonal elements of a power of a random matrix drawn from a Matchenko–Pastur ensemble is established.

PACS numbers:

1. Introduction

Wireless communication systems are designed to work in environments with as few infrastructure as possible. They shall provide the users with the freedom to communicate with whomever they want regardless wherever they are. Since electromagnetic waves, the most popular carriers of digital communications, propagate to almost any place, each user, though communicating with only a single other user, interacts, in principle, with all other users in the network. Such a setting is hard to press in formulas, in particular if the environment is arbitrary.

The failure of the so-called 3rd generation of wireless technology is, to a large extent, due to a lack of understanding of the fundamental principle governing wireless communications in the presence of many users operating simultaneously. This knowledge gap has become a severe obstacle to the further penetration of wireless communication devices into modern society and lifestyle and, therefore, must be overcome.

* Presented at the Conference on Applications of Random Matrices to Economy and Other Complex Systems, Kraków, Poland, May 25–28, 2005.

Research on the behavior of systems where many bodies mutually interact with many others has been driven forward by physicists for more than a century studying the interactions of particles in gases, fluids, and solids. Statistical physics and multi-user communications show strong analogies from a conceptual point of view. In both cases many objects interact with each other through variables that are constrained in a certain way. These interdisciplinary analogies can be exploited to advance the understanding and design of future wireless communication systems. Though the analogies between the two fields do not extend too far and, in real-world communication systems, statistical physics results cannot be applied directly, the engineering community can strongly benefit from the analytical toolboxes developed by physicists. So far random matrix theory, originally studied to describe spacings of nuclear energy levels, has received the most attention in wireless system analysis and design. In addition, the replica method developed in statistical physics has entered wireless communication to cope with the often binary nature of wireless communication signals.

In wireless communications, random matrix theory and statistical mechanics tools have overwhelmingly used for performance analysis, see *e.g.* [1–11]. Only few works [12–14] have used these large system tools for actual design of communication systems. The content of paper [13] and of [12] and its continuation [14] will be reviewed in this paper after a short introduction into wireless communication systems. In addition to analysis and design, references [15, 16] have used random matrix theory for *modeling* of wireless communication channels.

2. Communication systems

Though our daily life is full of multi-dimensional communication systems, both natural and artifical ones, we are still lacking a comprehensive theory describing their capabilities of carrying information. Even the seemingly simple channel depicted in Fig. 1, called an *interference channel*, with only two inputs and two outputs is not fully understood in information theory. Upper and lower bounds are known on its capabilities of carrying information from input #1 to output #1 while input #2 is simultaneously communicating to output #2. Only in some special cases, the two bounds co-incide. The search for a theory of characterizing precisely the capabilities of information flow in whole networks seems to be hopeless, considering that there has been no significant progress even on the interference channel within the last two decades. The interference channel perfectly describes the most common situation in everyday's communication. Person #1 talks to listener #1 while person #2 is talking to listener #2. Nevertheless, technical communication systems are not designed as interference channels — mostly

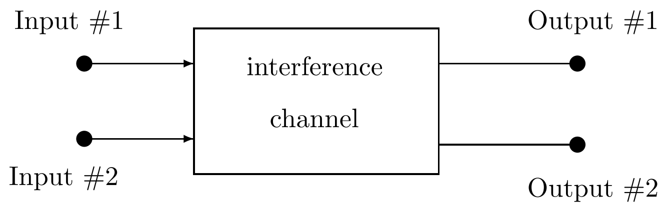

Fig. 1. Interference channel with two inputs and two outputs.

for the reason of our lack of understanding on how to deal with them — but as the concatenation of a *multiple-access channel* and a *broadcast channel* as shown in Fig. 2. A typical example for the concatenated approach is a cellular phone network with the base station taking the part of the relay.

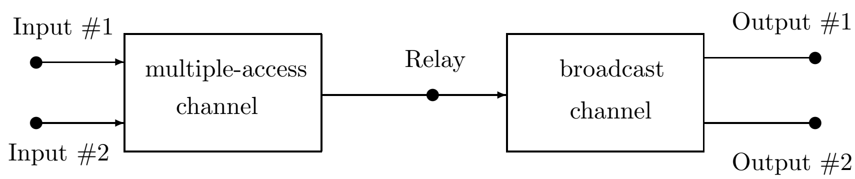

Fig. 2. Concatenation of multiple-access channel and broadcast channel.

The multiple-access channel is characterized by having only a single output though having multiple inputs while the broadcast channel is the dual to the multiple-access channel in terms of inputs and outputs. Both multiple-access channel and broadcast channel (provided that it is at least stochastically degraded[1]) are well-understood in information theory literature [17].

The principles summarized in this work 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 illustrative.

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

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

with

[1] See [17] for definition of stochastically degraded broadcast channels.

- the $K \times 1$ vector of transmitted symbols $\boldsymbol{x}[\nu]$,
- the $N \times 1$ vector of received symbols $\boldsymbol{y}[\nu]$,
- the $N \times K$ channel matrix $\boldsymbol{H}[\nu]$,
- the $N \times 1$ vector of additive white Gaussian noise $\boldsymbol{n}[\nu]$ with zero mean ($\mathrm{E}\,\boldsymbol{n} = \mathbf{0}$) and variance ($\mathrm{E}\,\boldsymbol{n}\boldsymbol{n}^{\mathrm{H}} = \sigma_0^2\mathbf{I}$),
- 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.

The vector-valued additive white Gaussian noise channel can be also written in an alternative manner: It is well known in literature [18] that the signal

$$\begin{aligned} \boldsymbol{r}[\nu] &= \boldsymbol{H}^{\mathrm{H}}[\nu]\boldsymbol{y}[\nu] \\ &= \boldsymbol{H}^{\mathrm{H}}[\nu]\boldsymbol{H}[\nu]\boldsymbol{x}[\nu] + \boldsymbol{H}^{\mathrm{H}}[\nu]\boldsymbol{n}[\nu] \end{aligned} \tag{2}$$

provides sufficient statistics for the estimation of the signal $\boldsymbol{x}[\nu]$. This means that all information about $\boldsymbol{x}[\nu]$ that could be extracted from the received signal $\boldsymbol{y}[\nu]$ can also be extracted from the signal $\boldsymbol{r}[\nu]$. 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 the forward link of a cellular CDMA system, the components of the vector $\boldsymbol{r}$ are regarded as the signals of K individual users while the vector y is the single input to the channel by the base station. In this case, (2) describes a broadcast channel. In the reverse link (uplink) of a cellular CDMA system, the components of the vector $\boldsymbol{x}$ are regarded as the signals of K individual users while the vector y is the single output of the channel observed by the base station. In this case, (1) describes a multiple-access channel. In both cases the matrix $\boldsymbol{H}$ contains the spreading sequences of the users as columns.

- In antenna array communications, the components of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ represent the signals sent and received by the K transmit and N receive antenna elements, respectively. Multiple antenna elements are employed to boost the data rate of one-to-one communication links. In this case, (1) describes a single-input single-output channel, though each input and output is a vector-valued observation and in literature often referred to as multiple-input multiple-output (MIMO) *system*, but not MIMO *channel*.

- In cable transmission, the components of the vector $\boldsymbol{x}$ contains the signals sent on the bundled twisted pairs within a cable. The coefficients in the matrix $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ describe the electromagnetic crosstalk between the respective twisted pairs. In this case, (2) describes an interference channel.

- For block transmission over a dispersive channel, the components of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ contain the symbols sent and received consecutively in time. Discrete time ν counts blocks, and the matrix $\boldsymbol{H}$ is a circulant[2] matrix of the channel's discrete-time impulse response. In this case, (1) describes a single-input single-output channel.

- In orthogonal frequency-division multiple-access (OFDM), the components of the vectors $\boldsymbol{x}$ and $\boldsymbol{r}$ represent the K sub-carriers at transmitter and receiver site, respectively, and the matrix $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ accounts for inter-carrier interference. Depending on the purpose that is followed when applying OFDM, the channel (2) can be considered as either multiple-access, broadcast, interference, or single-input single-output channel.

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 $\boldsymbol{H}$. Numerous results are reported in literature [18, 20] and no effort is made here in trying to be comprehensive.

3. Iterative multiuser decoding

In order to achieve optimal performance for communication over a vector-valued channel (1), the signal must, in general, be processed jointly over the components of the vector and over time. This becomes a prohibitive task even for vectors with only a few dimensions and simple channel codes, since the number of states of the decoding trellis is exponential in the product of the dimension of the input vector and the block-length of the code [21].

Since optimal information processing is infeasible, suboptimum algorithms have to be used in practice. With the invention of turbo decoding for approaching channel capacity on scalar communication channels in [22], iterative decoding also became the method of choice for near-optimum multiuser communication with tolerable complexity and was studied by several works among them [23–27]. All these papers found, with methods of various kinds of sophistication, that iterative multiuser decoding, *cf.* Fig. 3, can closely approximate optimal multiuser decoding if the level of interference is

[2] Properties of circulant matrices are addressed in [19].

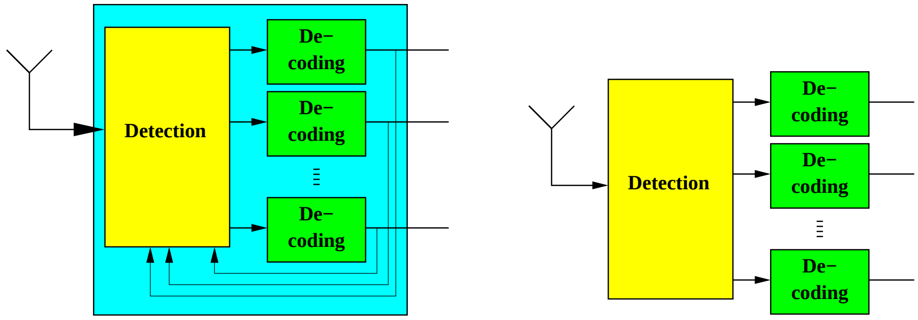

Fig. 3. Iterative multiuser decoding (left hand side) versus separated detection and decoding (right hand side).

low to moderately high and outperform all approaches separating detection and decoding. Nevertheless, it was observed that iterations fail to converge to correct decisions on the data if the interference level becomes to large.

Besides turbo codes, there is another class of codes for scalar communication channels called *low-density parity check codes* which are designed for iterative decoding. Understanding the iterative decoding algorithm as an instance of the *belief propagation* algorithm [28–30], an analysis tool for iterative decoding called *density evolution* was found [31]. It consists of tracking the empirical distribution of the decoder output from one iteration to the next. Instead of tracking the evolution of the exact distribution, tracking mean and variance of the distribution and imposing a Gaussian (mixture) distribution turned out to be a very accurate approximation [32, 33].

Inspired by the result that irregularity in the design of low-density parity check codes improves the convergence properties of the iterative decoding algorithm [34–36], dis-uniformizing the powers over the user population in iterative multiuser decoding was investigated in [13].

For the purpose of applying density evolution to iterative multiuser decoding an analytic expression for the uncoded error probability as a function of the *apriori* information of the detector was found. For the optimum multiuser detector, such a formula was derived in the large system limit making use the replica method. It reads

$$\Pr(\hat{x}_k \neq x_k) = \frac{1}{\sqrt{2\pi}} \int\limits_{\sqrt{\gamma_k \eta}}^{\infty} e^{-z^2/2} dz \qquad (3)$$

with the parameter η being determined by the fixed-point equation

$$\frac{1}{\eta} = 1 + \beta \int \gamma(1-t^2) \int\limits_{\mathbb{R}} \frac{1-\tanh\left(z\sqrt{\gamma\eta}+\gamma\eta\right)}{1-t^2\tanh^2\left(z\sqrt{\gamma\eta}+\gamma\eta\right)} \frac{e^{-z^2/2}}{\sqrt{2\pi}}\, dz\, d\mathrm{P}_{\gamma,t}\left(\gamma,t\right) . \tag{4}$$

Here, γ, $t = 2\Pr(x=1)-1$, and $\mathrm{P}_{\gamma,t}\left(\gamma,t\right)$ denote the signal-to-noise ratio, the bias of the prior, and the joint distribution of signal-to-noise ratio and bias of the prior over the user population, respectively. A recent generalization of this result is given in [37].

In order to solve the problem of optimal power assignment to the users, it is shown that under practical assumptions on the choice of the codes and target bit error rates, the iterations of the multiuser decoder lead to almost complete elimination of the multi-access interference if

$$\begin{aligned}
&\frac{1}{\eta+\varepsilon_1} - 1 \\
&\geq \beta \int \gamma \int\limits_{\mathbb{R}^2} \frac{\left(1-\tanh^2\left(y\sqrt{\mu(\gamma\eta)}+\mu(\gamma\eta)\right)\right)\left(1-\tanh\left(z\sqrt{\gamma\eta}+\gamma\eta\right)\right)}{1-\tanh^2\left(y\sqrt{\mu(\gamma\eta)}+\mu(\gamma\eta)\right)\tanh^2\left(z\sqrt{\gamma\eta}+\gamma\eta\right)} \\
&\times \frac{e^{-\frac{z^2+y^2}{2}}}{2\pi}\, dz\, dy\, d\mathrm{P}_{\gamma}\left(\gamma\right)
\end{aligned} \tag{5}$$

for all $\eta \in (0; 1-\varepsilon_2)$, with the scalar function $\mu(\cdot)$ describing the code characteristics, and $\varepsilon_1, \varepsilon_2$ being some small margins required for implementation. The function $\mu(\cdot)$ was found partially by simulation and partially by union bounds depending on the range of the argument. With the help of two properties of the condition (5), *i.e.* the linearity of the implicit equation with respect to the load β and the fact that (5) is a fixed point equation, the power optimization problem was formulated as a linear program and solved numerically. For the convolutional codes studied in the paper, all optimal power distributions were step functions with a finite number of steps.

The theoretical predictions obtained by asymptotic analysis and density evolution were confirmed by simulations. With the optimized power distribution, the iterations always converged to correct decisions on the data, even for very high interference levels. No upper bound on the possible number of users for a given spreading factor could be found. Moreover, with optimized power distribution, the total data rate of the system could be more than doubled.

For practical implementation, the optimum multiuser detector which is a sub-block in the iterative multiuser decoder still has too high complexity. It can be replaced, among other methods, by the conditional or unconditional linear minimum mean-squared error (MMSE) detector. The performances of both linear detector can be analyzed in the large system limit by means of random matrix theory. The respective equations that are counterparts to (4) read

$$\frac{1}{\eta} = 1 + \beta \int \frac{\gamma(1-t^2)}{1+\eta\gamma(1-t^2)}\, d\mathrm{P}_{\gamma,t}\left(\gamma, t\right) \tag{6}$$

and

$$\frac{1}{\eta} = 1 + \beta \int \frac{\gamma \int 1 - t^2 d\mathrm{P}_{t|\gamma}\left(\gamma, t\right)}{1+\eta\gamma \int 1 - t^2 d\mathrm{P}_{t|\gamma}\left(\gamma, t\right)}\, d\mathrm{P}_{\gamma}\left(\gamma\right) \tag{7}$$

for the conditional and the unconditional linear MMSE detector. While the complexity of the optimum detector is exponential in the number of users, complexity of these linear detectors is only cubic. In analogy to the procedure for the optimum multiuser detector, a counterpart to condition (5) can be found which shows the same nice properties, a fixed point equation that is linear in the load, and allows for optimization of the users' power profile by linear programming. Replacing the optimum multiuser detector by a simpler approach based on linear filters only a penalty of about 1 dB needs to be paid, *cf.* Fig. 4.

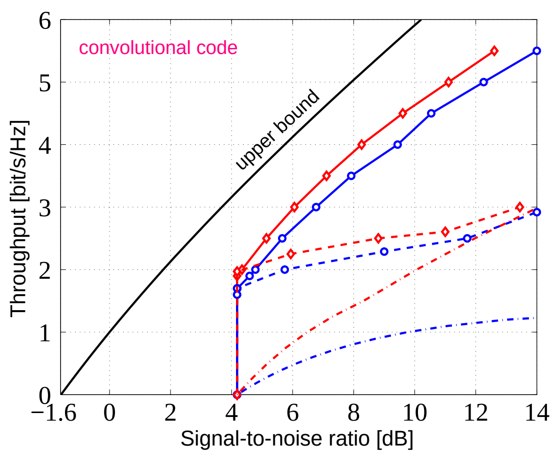

Fig. 4. Spectral efficiency of iterative multiuser decoding *vs.* signal-to-noise ratio for several methods of multiuser decoding at a bit error probability of 10^{-5}. Dash-dotted lines refer to separated detection and decoding, dashed lines refer to iterative multiuser decoding with equal powers for all users, solid lines refer to iterative multiuser decoding with optimized powers. The red and blue lines refer to optimal and linear MMSE detection, resp. The upper bound refers to the capacity (maximum data rate) of the channel. A standard convolutional code with 64 states was used.

Power optimization based on the asymptotic approximation turned out both accurate and simple enough to make real-time power optimization for iterative multiuser decoding possible in deployed mobile radio communication systems. In contrast, the exact optimization problem without large system approximation is, even for only a few users, still unsolved due to its prohibitive complexity.

4. Asymptotic design of multistage detectors

As mentioned in the previous section, even simplified multiuser detectors which are used in iterative multiuser decoding like the linear MMSE detectors have cubic complexity in the number of users. That is, as they require to solve a linear system of equations with one equation each per user. For hundreds of users this is not a trivial task to be performed in a mobile handset within microseconds. When it came to the frequency-division duplex (FDD) mode of Europe's UMTS[3], some believed that multiuser detection, though it would improve performance significantly, is infeasible with today's technology.

Motivated by the observation that the eigenvalues of large random matrices become more and more predictable the larger the matrices are reference [12] showed that multiuser detection should, in contrast to the common believe, not become more difficult but, at some point, simplify if there are enough users in the system. In [12], it is demonstrated how to simplify solving systems of linear equations given by a large random matrix drawn from a Marchenko–Pastur[4] ensemble exploiting the convergence properties of its eigenvalues.

Assume you want to invert a $K \times K$ matrix $\boldsymbol{X}$ whose eigenvalues $\mathcal{L} = \{\lambda_1, \ldots, \lambda_K\}$ are known to you. Note that due to the Cayley–Hamilton Theorem [38] any matrix is a zero of its characteristic polynomial

$$\prod_{k=1}^{K} (\boldsymbol{X} - \lambda_k \mathbf{I})^k = \mathbf{0} \,. \tag{8}$$

Expanding the product into a sum, we find

$$\sum_{k=0}^{K} c_k(\mathcal{L}) \boldsymbol{X}^k = \mathbf{0} \tag{9}$$

[3] Universal Mobile Telecommunications Standard (UMTS).

[4] Since the spreading sequences in mobile radio standards are pseudo-random numbers, the channel matrix can be well-approximated by a random matrix.

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

$$\boldsymbol{X}^{-1} = -\sum_{k=0}^{K-1} \frac{c_{k+1}(\mathcal{L})}{c_0(\mathcal{L})} \boldsymbol{X}^k \triangleq \sum_{k=0}^{K-1} \tilde{w}_k(\mathcal{L}) \boldsymbol{X}^k. \tag{10}$$

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

While standard Gauss–Seidel iterations require, 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 a matrix polynomial of order $L-1$

$$\boldsymbol{X}^{-1} \approx \sum_{k=0}^{L-1} w_k \boldsymbol{X}^k. \tag{11}$$

are determined by a system of Yule–Walker equations [39]

$$\begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_L \end{bmatrix} = \begin{bmatrix} m_2 & m_3 & \dots & m_{L+1} \\ m_3 & m_4 & \dots & m_{L+2} \\ \vdots & \vdots & \ddots & \vdots \\ m_{L+1} & m_{L+2} & \dots & m_{2L} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{L-1} \end{bmatrix}, \tag{12}$$

where

$$m_k \triangleq \frac{1}{K} \sum_{i=1}^{K} \lambda_i^k. \tag{13}$$

Reference [39] suggested to track the empirical eigenvalue moments m_k adaptively and to solve the Yule–Walker equations (12) 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. The coefficients of the matrix-valued polynomial which lead to the smallest deviation of output signal of the approximated detector from the exact detector are given by the solution to the system of Yule–Walker equations (12). They depend only on the first $2L$ moments of the eigenvalue distribution of the matrix $\boldsymbol{X}$ which are well-known in literature. This approach to weight design was proposed in [12] and explicit expressions for the optimum weights and the achievable SINRs were given for channel matrices with independent identically distributed entries. It was generalized to channel matrices with more involved asymptotic statistics in [14].

The weights can also be calculated adaptively interpreting the polynomial expansion detector as a multistage Wiener filter [40,41]. This approach was followed by in [42] 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 [43]. Reference [44] highlighted the connection between the Marchenko–Pastur distribution, continued fractions, and orthogonal polynomials for the analysis of polynomial expansion detectors. Such a connection is well established in mathematical literature [45], but found its way into the design and analysis of code-division multiple-access only recently.

The weight design according to (12) with (13) minimizes the mean-squared error of the solution of the system of linear equations to be solved. For engineering purposes, however, minimizing the bit error probability of the users is a more sensible, tough related goal of optimization. As shown in [14], this leads to a different weight design, if the users' channels have different statistics. In that case the weight design is different for each user i. Thus, the weights actually turn into diagonal matrices. The Yule–Walker equations (12) stay valid, but (13) has to be replaced by

$$m_k = \left(\boldsymbol{X}^k\right)_{ii} \tag{14}$$

for user i. Apparently, the weights depend on the particular diagonal elements of the powers of the matrix, not only on its average, the trace. Similar to the convergence of the trace, the diagonal elements can also be shown to converge to deterministic limits for a given signal power of user i. For the reader's convenience, this statement is made more precise in the appendix.

A particularly pleasant feature of this way to iteratively solve a linear system of equations governed by a random matrix is the speed of convergence. In [44], it was shown that the mean-squared error of the approximation decays exponentially fast as long as the support of the eigenvalue density is a subset of a finite interval on the positive real axis.

As a consequences of all these particular results, the complexity of implementing multiuser detectors for a large number of users with pseudo-random spreading was found to be merely quadratic than cubic in the number of users. The fundamental principle used to demonstrate the feasibility of interference mitigation are not particular to code-division multiple-access, but rely on fundamental properties of random matrices and their convergence of its eigenvalues to deterministic limits. They can, therefore, be generalized to a variety of other applications which involve the vector channel (1).

5. Summary and outlook

Random matrix theory and other large system properties can be successfully used to come up with new designs of multiuser communication systems. However, not for all practically relevant channel statistics, results on the asymptotic eigenvalue distributions of channel matrices are available in mathematical and physics literature.

Appendix A

Theorem 1 *[14] Let $\boldsymbol{A}$ be a $K \times K$ diagonal matrix in $\mathbb{C}$ with bounded elements and such that the sequence of the eigenvalue distribution of $\boldsymbol{A}^H\boldsymbol{A}$ converges almost surely, as $K \to \infty$, to a non-random distribution function $F_{|\boldsymbol{A}|^2}(\lambda)$ with upper bounded support. Let $\boldsymbol{S} \in \mathbb{C}^{N\times K}$ with random i.i.d. zero mean entries with variance $\mathrm{E}\{|s_{ij}|^2\} = \frac{1}{N}$, and $\lim_{N\to\infty} \mathrm{E}\{N^3|s_{ij}|^6\} < +\infty$. Let $\boldsymbol{R} = \boldsymbol{A}^H\boldsymbol{S}^H\boldsymbol{S}\boldsymbol{A}$. Conditioned on a_{kk}, the k-th diagonal element of $\boldsymbol{A}$, $(\boldsymbol{R}^\ell)_{kk}$ converges almost surely, as $N, K \to \infty$ with $\frac{K}{N} \to \beta$, to the conditionally deterministic quantity $R^\ell_{kk,\infty}$*

$$R^\ell_{kk,\infty} = |a_{kk}|^2 \sum_{s=0}^{\ell-1} R^s_{kk,\infty} \beta m_{\boldsymbol{R}}^{\ell-s-1} \qquad \ell > 1 \tag{A.1}$$

for any $k, \ell \in \mathbb{Z}^+$. $m^s_{\boldsymbol{R}} = \mathrm{E}\left\{\frac{1}{K}\mathrm{tr}\left(\boldsymbol{R}^s\right)\right\}$. The initial values of the recursion are $R^0_{kk,\infty} = 1$ and $m^0_{\boldsymbol{R}} = \beta^{-1}$. □

A closed-form expression for the moments $m^s_{\boldsymbol{R}}$ can be found in [46].

REFERENCES

[1] S. Verdú, S. Shamai (Shitz), *IEEE Transactions on Information Theory* **45**, 622 (1999).

[2] D. Tse, S. Hanly, *IEEE Transactions on Information Theory* **45**, 641 (1999).

[3] J. Evans, D.N.C. Tse, *IEEE Transactions on Information Theory* **46**, 2059 (2000).

[4] S. Shamai (Shitz), S. Verdú, *IEEE Transactions on Information Theory* **47**, 1302 (2001).

[5] T. Tanaka, *IEEE Transactions on Information Theory* **48**, 2888 (2002).

[6] M. Debbah, W. Hachem, P. Loubaton, M. Courville, *IEEE Transactions on Information Theory* **49** (5), (2003).

[7] A.L. Moustakas, S.H. Simon, A.M. Sengupta, Technical Report, Bell-Laboratories, Murray-Hill, NJ, 2002.

[8] R.R. Müller, W.H. Gerstacker, *IEEE Transactions on Information Theory* **50**, 1769 (2004).

[9] R.R. Müller, *IEEE Transactions on Signal Processing*, **51**, 2821 (2003).

[10] A.M. Tulino, S. Verdú, *Foundations and Trends in Communications and Information Theory* **1** (1), June 2004.

[11] D. Guo, S. Verdú, *IEEE Transactions on Information Theory* **51**, 1983 (2005).

[12] R.R. Müller, S. Verdú, *IEEE Journal on Selected Areas in Communications* **19**, 1429 (2001).

[13] G. Caire, R.R. Müller, T. Tanaka, *IEEE Transactions on Information Theory* **50**, 1950 (2004).

[14] L. Cottatellucci, R.R. Müller, *IEEE Transactions on Information Theory* **51** (9), September 2005.

[15] R.R. Müller, *IEEE Transactions on Information Theory* **48**, 2495 (2002).

[16] M. Debbah, R. Müller, *IEEE Transactions on Information Theory* **51**, 1667 (2005).

[17] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York 1991.

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

[19] P.J. Davis, *Circulant Matrices*, John Wiley & Sons, New York 1979.

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

[21] T.R. Giallorenzi, S.G. Wilson, *IEEE Transactions on Communications* **44**, 997 (1996).

[22] C. Berrou, A. Glavieux, *IEEE Transactions on Communications* **44**, 1261 (1996).

[23] M. Moher, *IEEE Transactions on Communications* **46**, 870 (1998).

[24] P.D. Alexander, A. Grant, M. Reed, *European Transactions on Telecommunications* **9**, 419 (1999).

[25] X. Wang, H.V. Poor, *IEEE Transactions on Communications* **47**, 1046 (1999).

[26] H. El Gamal, E. Geraniotis, *IEEE Journal on Selected Areas in Communications* **18**, 30 (2000).

[27] J. Boutros, G. Caire, *IEEE Transactions on Information Theory* **48**, 1772 (2002).

[28] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Francisco, CA 1988.

[29] R.J. McEliece, D.J.C. MacKay, J.-F. Cheng, *IEEE Journal on Selected Areas in Communications* **16**, 140 (1998).

[30] F.R. Kschischang, B.J. Frey, H.-A. Loeliger, *IEEE Transactions on Information Theory* **47**, 498 (2001).

[31] T.J. Richardson, R.L. Urbanke, *IEEE Transactions on Information Theory* **47**, 599 (2001).

[32] H. El Gamal, A.R. Hammons, *IEEE Transactions on Information Theory* **47**, 671 (2001).

[33] S.-Y. Chung, T.J. Richardson, R.L. Urbanke, *IEEE Transactions on Information Theory* **47**, 657 (2001).

[34] M.G. Luby, M. Mitzenmacher, M.A. Shokrollahi, D.A. Spielman, *IEEE Transactions on Information Theory* **47**, 585 (2001).

[35] D.J.C. MacKay, S.T. Wilson, M.C. Davey, *IEEE Transactions on Communications* **47**, 1449 (1999).

[36] T.J. Richardson, M.A. Shokrollahi, R.L. Urbanke, *IEEE Transactions on Information Theory* **47**, 619 (2001).

[37] R.R. Müller, A.M. Tulino, In Proc. of IEEE Intern. Symp. on Spread Spectrum Techniques and Applications (ISSSTA), Sydney, Australia, September 2004.

[38] R.A. Horn, C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, U.K. 1985.

[39] S. Moshavi, E.G. Kanterakis, D.L. Schilling, *International Journal of Wireless Information Networks* **3**, 1 (1996).

[40] J.S. Goldstein, I.S. Reed, *IEEE Transactions on Signal Processing* **42**, 492 (1997).

[41] J.S. Goldstein, I.S. Reed, L.L. Scharf, *IEEE Transactions on Information Theory* **44** (7), November 1998.

[42] M. Honig, W. Xiao, *IEEE Transactions on Information Theory* **47**, 1928 (2001).

[43] L.G.F. Trichard, J.S. Evans, I.B. Collings, Proc. of IEEE International Symposium on Information Theory (ISIT), p. 21, Lausanne, Switzerland, June/July 2002.

[44] P. Loubaton, W. Hachem, Proc. of IEEE Information Theory Workshop (ITW), Paris, France, April 2003.

[45] P. Deift, *Orthogonal Polynomials and Random Matrices: A Riemann–Hilbert Approach*, American Mathematical Society, Providence, RI 1998.

[46] Y.Q. Yin, *Journal of Multivariate Analysis* **20**, 50 (1986).