

# Design of Reduced-Rank MMSE Multiuser Detectors Using Random Matrix Methods

Linbo Li, Antonia M. Tulino, *Member, IEEE*, and Sergio Verdú, *Fellow, IEEE*

**Abstract**—Reduced-rank minimum mean-squared error (MMSE) multiuser detectors using asymptotic weights have been shown to reduce receiver complexity while maintaining good performance in long-sequence code-division multiple-access (CDMA) systems. In this paper, we consider the design of reduced-rank MMSE receivers in a general framework which includes fading, single and multiantenna receivers, as well as direct-sequence CDMA (DS-CDMA) and multicarrier CDMA (both uplink and downlink). In all these cases, random matrix results are used to obtain explicit expressions for the asymptotic eigenvalue moments of the interference autocorrelation matrix and for the asymptotic weights used in the reduced-rank receiver.

**Index Terms**—Code-division multiple access (CDMA), fading channels, minimum mean-squared error (MMSE) receivers, multiantenna systems, multicarrier CDMA, multiuser detection, random matrix theory.

## I. INTRODUCTION

THE analysis of randomly-spread direct-sequence code-division multiple access (DS-CDMA) in the asymptotic regime of number of users  $K$  and spreading gain  $N$  going to infinity with a constant ratio has received much attention because it helps gain insight into the behavior of multiuser receivers for large DS-CDMA systems employing pseudonoise spreading sequences (e.g., [13], [33], [38], [39]). Particular attention has been paid to the linear minimum mean-squared error (MMSE) receiver, which maximizes the output signal-to-interference-plus-noise ratio (SINR). For single and multiantenna DS-CDMA, the output SINR of the MMSE receiver is shown to converge [18], [33] in the asymptotic regime to the solution of a fixed-point equation by using results on the eigenvalue distribution of large random matrices.

To maximize the output SINR, the MMSE receiver needs to invert a matrix whose dimensionality is equal to either the number of users or the spreading gain. In large systems, this is a computationally intensive operation. It is therefore desirable to find a receiver which approaches the optimal output SINR of the MMSE receiver at a lower computational cost. Multistage MMSE receivers, studied in [21], [28], approximate the MMSE receiver by a linear operator on a signal subspace and,

hence, reduce the computational complexity. The multistage Wiener (MSW) filter, originally proposed in [10], [11], is analyzed in [21] for unfaded DS-CDMA in the asymptotic regime. It is shown in [21] that the MSW filter is equivalent to the reduced-rank MMSE receiver, which approximates the matrix inversion through a polynomial expansion. It is further demonstrated in [21] and [26], numerically and analytically, respectively, that the number of stages  $D$  needed in the reduced-rank receiver to achieve a desired output SINR does not scale with the system size; in fact,  $D = 8$  is sufficient to achieve near-full-rank output SINR regardless of the dimension of the signal space. However, the weights of the reduced-rank receiver depend on the spreading sequences. Therefore, for long-sequence CDMA, where the period of the spreading sequences is usually a large multiple of the symbol duration, they need to be re-evaluated and updated from symbol to symbol, which hampers real-time implementation.

To overcome this problem under the random spreading model, [29], [35], and [25] proposed the asymptotic reduced-rank MMSE receiver, which replaces the optimal weights by their limiting values calculated in the asymptotic regime ( $K, N \rightarrow \infty$  with  $\frac{K}{N} \rightarrow \beta$ ). In contrast to the exact weights, the asymptotic weights depend neither on the realization of the spreading sequences nor on the realization of the channel. Since the asymptotic weights depend only on  $\beta$  and the channel fading distribution, they do not need to be updated from symbol to symbol. Therefore, real-time implementation becomes feasible. In order to compute these asymptotic weights explicitly, explicit expressions for the asymptotic moments of the received spreading sequence matrix have been derived using random matrix theory. The scenario considered in [29] is that of DS-CDMA with equal-power users and no fading. In turn, [25] considers the more general scenario of DS-CDMA with unequal-power users and flat-fading.<sup>1</sup> In addition, [25] extends the analysis to multiantenna receivers (with the constraint that the received amplitudes at those antennas be equal).

Although, theoretically, the difference between the asymptotic weights and the optimal weights only vanishes in the limit of infinite system sizes, the numerical results in [25] show that the asymptotic weights work well for even moderately large effective spreading gain (the spreading gain multiplied by the number of antennas). For instance, for a system with effective spreading gain equal to 32, the performance loss due to the use of the asymptotic weights is at most 1 dB for the input signal-to-noise ratio (SNR) considered.

Section II gives several definitions and general results on asymptotic reduced-rank receivers used throughout the paper.

<sup>1</sup>Similar results were later reported in [5].

Manuscript received September 7, 2002; revised December 2, 2003.

L. Li was with Princeton University, Princeton, NJ 08544 USA. He is now with Qualcomm Inc, San Diego, CA 92121 USA (e-mail: linbol@qualcomm.com).

A. M. Tulino is with Università Federico II, 81125 Napoli, Italy (e-mail: atulino@ee.princeton.edu).

S. Verdú is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: verdu@ee.princeton.edu).

Communicated by G. Caire, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2004.828076

In addition to random matrix theory, we import results from noncrossing partition theory and their connection with the minimal factorization of a cyclic permutation. Using these tools, we obtain recursive expressions (which may be of independent interest) for the asymptotic moments of random matrices with independent nonidentically distributed entries. These expressions depend on the variance of the entries of the random matrix and generalize the relationship relating moments and free cumulants found in [32]. In the special case in which the matrix of variances can be written as an outer product, we prove an explicit expression for the asymptotic eigenvalue moments which had been independently conjectured in [42] in the context of the analysis of least squares algorithms for DS-CDMA. We also compute the asymptotic moments of some classes of random matrices whose entries are correlated across the columns but independent across the rows. This result encompasses the case of a product of a matrix with independent and identically distributed (i.i.d.) entries and an independent Hermitian matrix analyzed by [43].

Applying the technical results developed in Section II, we go beyond the design methodology put forth in [25] for DS-CDMA channels subject to flat fading. While adhering to the setting of spreading matrices with independent entries, Section III extends the analysis of reduced-rank MMSE by [25] in two different directions. First, DS-CDMA channels subject to frequency-selective fading are considered for both uplink and downlink scenarios<sup>2</sup> and, then, DS-CDMA flat-faded channels with multiple receive antennas are studied (removing the constraint, imposed in [25], of equal amplitude at each such antenna). Unlike previous works that deal exclusively with the downlink, our uplink analysis requires finer random matrix tools and is general enough to encompass non-Gaussian chip distributions.

In Section IV, we generalize the channel models to embrace multicarrier CDMA (MC-CDMA) subject to frequency-selective fading. In particular, Sections IV-A and IV-B derive the asymptotic reduced-rank receiver for uplink and downlink MC-CDMA, respectively. Further, in Section IV-C, the full-rank MMSE receiver is analyzed for both uplink and downlink MC-CDMA. A fixed-point equation for the asymptotic maximum output SINR is derived for both cases,<sup>3</sup> which particularizes to the Tse–Hanly equation for the special case of DS-CDMA subject to frequency-flat fading [33].

Our numerical results in Section V show that, for practical processing gains, the performance loss due to the use of asymptotic—rather than exact—weights is very small for input SNR levels of interest.

## II. ASYMPTOTIC EIGENVALUE MOMENTS

Throughout the paper,  $(\mathbf{A})_{i,j}$  indicates the  $(i, j)$ th entry of the matrix  $\mathbf{A}$ ,  $(\mathbf{A})_j$  indicates its  $j$ th column,  $\mathbf{A}_j$  indicates the submatrix obtained by removing from  $\mathbf{A}$  the  $j$ th column, and, finally,  $(\mathbf{x})_j$  indicates the  $j$ th entry of the vector  $\mathbf{x}$ .

<sup>2</sup>For the downlink, the full-rank MMSE receiver for isometric spreading sequences is studied in [3], [4] under the assumption that the users are partitioned into a finite number of classes, each having a distinct power.

<sup>3</sup>For downlink MC-CDMA in the special case of equal-power users, a fixed-point equation for the asymptotic maximum output SINR was independently derived in [6] for both isometric and i.i.d. spreading sequences.

The basic model used throughout the paper is a general linear model that accommodates most single and multiuser communication scenarios. Such multiple-input multiple-output (MIMO) model is

$$\mathbf{r} = \mathbf{T}\mathbf{b} + \mathbf{n} \quad (1)$$

where  $\mathbf{b}$  is the  $n_I$ -dimensional input vector whose entries are modeled as zero-mean unit variance and uncorrelated,  $\mathbf{r}$  is the  $n_o$ -dimensional output vector, and the  $n_o$ -dimensional vector  $\mathbf{n}$  is the additive Gaussian noise, whose components are independent complex Gaussian random variables with zero mean and independent real and imaginary parts each with variance  $\sigma^2/2$  (i.e., circularly symmetric).  $\mathbf{T} = [(\mathbf{T})_1, \dots, (\mathbf{T})_{n_I}]$ , in turn, is the  $n_o \times n_I$  complex random matrix describing the channel and it is independent of  $\mathbf{n}$ .

Model (1) encompasses a variety of channels of interest in wireless communications such as multiaccess channels, linear channels with frequency-selective and/or frequency-dispersive fading, multidimensional channels (multisensor reception, multicellular system with cooperative detection, etc), crosstalk in digital subscriber lines, signal space diversity, etc. In each of these cases,  $n_o$ ,  $n_I$  and  $\mathbf{T}$  take different meanings. For example,  $n_I$  and  $n_o$  may indicate i) the number of transmit and receive antennas, respectively, with  $\mathbf{T}$  describing the fading between each pair of transmit and receive antennas, ii) the spreading gain and the number of users with  $\mathbf{T}$  the signature matrix, or iii) time/frequency slots with  $\mathbf{T}$  the tone matrix. Naturally, the same linear model applies to problems that incorporate more than one of the above features; in these cases, it may be more convenient to use matrices or higher dimensional objects to represent the input and output quantities (for example, in a multi-antenna DS-CDMA system with processing gain  $N$ ,  $n_o = NL$  with  $L$  the number of receive antennas.)

In the remainder of the paper, we always consider a linear MMSE receiver that has perfect knowledge of the channel matrix. Consequently, the decision statistic adopted at the receiver to decode the  $k$ th input signal is  $(\hat{\mathbf{b}})_k = \mathbf{c}^\dagger \mathbf{r}$  where<sup>4</sup>

$$\mathbf{c} = \left( \mathbf{T}\mathbf{T}^\dagger + \sigma^2 \mathbf{I} \right)^{-1} (\mathbf{T})_k \quad (2)$$

$$\propto \left( \mathbf{T}_k \mathbf{T}_k^\dagger + \sigma^2 \mathbf{I} \right)^{-1} (\mathbf{T})_k \quad (3)$$

where  $\mathbf{T}_k$  is the  $n_o \times (n_I - 1)$  matrix defined as

$$\mathbf{T}_k = [(\mathbf{T})_1, \dots, (\mathbf{T})_{k-1}, (\mathbf{T})_{k+1}, \dots, (\mathbf{T})_K]. \quad (4)$$

The complexity of computing the inverse of the  $n_o \times n_o$  matrix  $(\mathbf{T}_k \mathbf{T}_k^\dagger + \sigma^2 \mathbf{I})$  can be lowered by approximating the matrix inversion with the following polynomial expansion:

$$\mathbf{c}_D = \sum_{m=0}^{D-1} (\mathbf{w})_m \left( \mathbf{T}_k \mathbf{T}_k^\dagger + \sigma^2 \mathbf{I} \right)^m (\mathbf{T})_k \quad (5)$$

<sup>4</sup>For convenience, the explicit dependence of  $\mathbf{c}$  on  $k$  is omitted.

where  $D$  is the rank (or the number of stages) of the receiver. The weights, chosen to minimize the mean-squared error (MSE), are given by [21], [25]

$$\mathbf{w} = \begin{bmatrix} \mathcal{H}_1 + \mathcal{H}_0 \mathcal{H}_0 & \cdots & \mathcal{H}_D + \mathcal{H}_{D-1} \mathcal{H}_0 \\ \vdots & \ddots & \vdots \\ \mathcal{H}_D + \mathcal{H}_{D-1} \mathcal{H}_0 & \cdots & \mathcal{H}_{2D-1} + \mathcal{H}_{D-1} \mathcal{H}_{D-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{H}_0 \\ \vdots \\ \mathcal{H}_{D-1} \end{bmatrix} \quad (6)$$

where the  $(i, j)$ th entry of the matrix in (6) is  $\mathcal{H}_{i+j-1} + \mathcal{H}_{i-1} \mathcal{H}_{j-1}$  with

$$\mathcal{H}_m = (\mathbf{T})_k^\dagger \left( \mathbf{T}_k \mathbf{T}_k^\dagger + \sigma^2 \mathbf{I} \right)^m (\mathbf{T})_k. \quad (7)$$

$D = n_o$  corresponds to the full-rank MMSE receiver. From (7), we note that the optimal weights depend on the channel matrix  $\mathbf{T}$ . Therefore, for applications where  $\mathbf{T}$  changes at every symbol interval (like in the case of aperiodic DS-CDMA with long spreading sequences), the weights must be re-evaluated and updated at every symbol interval. To lift the burden of computing the weights from the received spreading sequences for every symbol interval, [29], [35], and [25] proposed the asymptotic reduced-rank MMSE receiver which replaces the optimal weights with their limiting values in the asymptotic regime ( $n_I, n_o \rightarrow \infty$  with  $\frac{n_I}{n_o} \rightarrow \beta$ ) with the intuition that their dependence on the received spreading sequences would vanish asymptotically. For the scenarios considered in [25], namely, single and multiantenna (with the constraint of equal received amplitudes at those antennas) receiver for DS-CDMA subjected to flat fading, with and without linear conjugate transformations, as for scenarios analyzed in [29] and [35], this intuition was validated.

In this paper, following the philosophy in [25], we also calculate the limits of the optimal weights (6) in the asymptotic regime for more general and diverse scenarios whose mathematical model can be described by (1). From (6), the asymptotic weights boil down to the asymptotic value of  $\mathcal{H}_m$ , namely

$$\mathcal{H}_m^\infty = \lim_{n_I \rightarrow \infty} \mathcal{H}_m \quad (8)$$

from which it follows that the asymptotic weights  $(\mathbf{w}^\infty)_m$ 's are given by (6) where each  $\mathcal{H}_m$  is replaced by its asymptotic counterpart  $\mathcal{H}_m^\infty$ , i.e.,

$$\mathbf{w}^\infty = \begin{bmatrix} \mathcal{H}_1^\infty + \mathcal{H}_0^\infty \mathcal{H}_0^\infty & \cdots & \mathcal{H}_D^\infty + \mathcal{H}_{D-1}^\infty \mathcal{H}_0^\infty \\ \vdots & \ddots & \vdots \\ \mathcal{H}_D^\infty + \mathcal{H}_{D-1}^\infty \mathcal{H}_0^\infty & \cdots & \mathcal{H}_{2D-1}^\infty + \mathcal{H}_{D-1}^\infty \mathcal{H}_{D-1}^\infty \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{H}_0^\infty \\ \vdots \\ \mathcal{H}_{D-1}^\infty \end{bmatrix}. \quad (9)$$

The calculation of the asymptotic weights (9) is closely related to the evaluation of the asymptotic eigenvalue moments of  $\mathbf{T}\mathbf{T}^\dagger$ . Several models for the random matrix  $\mathbf{T}$  are adopted in the various problems studied in this paper.

- A matrix with zero-mean independent entries, arbitrarily distributed.
- A matrix of the form  $\mathbf{T} = [\mathbf{A}_1(\mathbf{S})_1, \dots, \mathbf{A}_{n_I}(\mathbf{S})_{n_I}] \bar{\mathbf{A}}$  where  $\bar{\mathbf{A}}$  is an  $n_I \times n_I$  arbitrary deterministic diagonal matrix,  $\mathbf{A}_k$  for  $k = 1, \dots, n_I$  is an arbitrary sequence of  $n_o \times n_o$  (possibly random) Toeplitz matrices, while  $\mathbf{S}$  is an  $n_o \times n_I$  zero-mean random matrix with i.i.d. entries and variance  $1/n_o$ .

We always assume that

$$\beta = \lim_{n_o \rightarrow \infty} \frac{n_I}{n_o} \quad (10)$$

is finite and strictly positive.

Let us now introduce some convenient definitions.

*Definition 1:* The empirical distribution  $F_{\mathbf{A}}^{(n)}(a)$  of  $\mathbf{A} = (a_1, a_2, \dots, a_n)$ , is defined as

$$F_{\mathbf{A}}^{(n)}(a) = \frac{1}{n} \sum_{i=1}^n 1\{a_i \leq a\}, \quad (11)$$

where  $1\{\cdot\}$  denotes the indicator function. For an  $n \times p$  matrix  $\mathbf{A}$ , the empirical distribution (of the square of its singular values),  $F_{\mathbf{A}}^{(n)}(\lambda)$ , is defined as

$$F_{\mathbf{A}}^{(n)}(\lambda) = \frac{1}{n} \sum_{i=1}^n 1\{\lambda_i \leq \lambda\} \quad (12)$$

where  $\lambda_i$  is the  $i$ th eigenvalue of  $\mathbf{A}\mathbf{A}^\dagger$ .

If (11) and/or (12) converge as  $n \rightarrow \infty$ , then the corresponding limits (asymptotic empirical distributions) are denoted by  $F_{\mathbf{A}}(a)$  and  $F_{\mathbf{A}}(\lambda)$ , respectively.

Let us first focus on the case that  $\mathbf{T}$  is an  $n_o \times n_I$  matrix whose entries are zero-mean independent complex random variables arbitrarily distributed with variances

$$E[|(\mathbf{T})_{i,j}|^2] = \frac{(\mathbf{V})_{i,j}}{n_o} \quad (13)$$

where  $\mathbf{V}$  is an  $n_o \times n_I$  deterministic matrix whose entries are uniformly bounded. For each  $n_o$ , let

$$v^{(n_o)} : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$$

be a function given by

$$v^{(n_o)}(x, y) = (\mathbf{V})_{i,j}, \quad i, j \text{ satisfying } \frac{i-1}{n_o} \leq x < \frac{i}{n_o}, \frac{j-1}{n_I} \leq y < \frac{j}{n_I} \quad (14)$$

and assume that the sequence  $v^{(n_o)}(x, y)$  converges uniformly to a limiting bounded measurable function  $v(x, y)$ , which will be referred to as *variance profile* of  $\mathbf{T}$ .

Denoting by

$$\mu_n^{(n_o)} = \frac{1}{n_o} \text{Tr} \left( (\mathbf{T}\mathbf{T}^\dagger)^n \right) \quad (15)$$

the  $n$ th moment of  $\mathbf{T}\mathbf{T}^\dagger$ , we give the following theorem.

*Theorem 1:* The  $n$ th moment of  $\mathbf{T}\mathbf{T}^\dagger$  converges almost surely (a.s.) to

$$\lim_{n_o \rightarrow \infty} \mu_n^{(n_o)} \stackrel{\text{a.s.}}{=} \mu_n = E[m_n(\mathbf{X})] \quad (16)$$

with  $m_n(x)$  obtained through the following recursive equation:

$$m_n(x) = \beta \sum_{\ell=1}^n m_{\ell-1}(x) E[v(x, \mathbf{Y}) \sum_{\substack{n_1+\dots+n_i=n-\ell \\ 1 \leq i \leq n-\ell}} E[v(\mathbf{X}, \mathbf{Y}) m_{n_1-1}(\mathbf{X}) | \mathbf{Y}] \dots E[v(\mathbf{X}, \mathbf{Y}) m_{n_i-1}(\mathbf{X}) | \mathbf{Y}]] \quad (17)$$

where  $m_0(x) = 1$  and where, in the second summation, the  $n_k$ 's are strictly positive integers. Finally,  $\mathbf{X}$  and  $\mathbf{Y}$  are two independent continuous random variables uniformly distributed in  $[0, 1]$ .

*Proof:* See Appendix A.

Define  $\tilde{m}_n(x)$  through the following recursive equation:

$$\tilde{m}_n(x) = \sum_{\ell=1}^n c_\ell(x) \sum_{n_1+n_2+\dots+n_\ell=n} \tilde{m}_{n_1-1}(x) \dots \tilde{m}_{n_\ell-1}(x) \quad (18)$$

with

$$c_{\ell+1}(x) = \beta E[v(x, \mathbf{Y}) E^\ell[v(\mathbf{X}, \mathbf{Y}) | \mathbf{Y}]]$$

from which  $\tilde{m}_1(x) = c_1(x)$  and  $\tilde{m}_2(x) = c_2(x) + c_1^2(x)$ . In several interesting special cases, such as  $v(x, y) = v_o(x)v_I(y)$ , we can show  $E[m_n(\mathbf{X})] = E[\tilde{m}_n(\mathbf{X})]$ , with  $m_n(x)$  given by the recursive expression in (17). In other cases, there is strong numerical evidence that the relationship holds in general. However, a general proof has not been found.

Furthermore, for a given  $k$ , let  $v_k^{(n_o)}(x)$  be a function defined in  $[0, 1]$  such that

$$v_k^{(n_o)}(x) = (\mathbf{V})_{i,k}, \quad i \text{ satisfying } \frac{i-1}{n_o} \leq x < \frac{i}{n_o} \quad (19)$$

whose limit as  $n_o \rightarrow \infty$  is denoted by  $v_k(x)$ .

*Theorem 2:* Consider the following quadratic form:

$$\delta_{n,k}^{(n_o)} = (\mathbf{T})_k^\dagger \left( \mathbf{T}_k \mathbf{T}_k^\dagger \right)^n (\mathbf{T})_k. \quad (20)$$

As  $n_o$  and  $n_I$  go to infinity

$$\lim_{n_o \rightarrow \infty} \delta_{n,k}^{(n_o)} \stackrel{\text{a.s.}}{=} \delta_{n,k} = E[m_n(\mathbf{X}) v_k(\mathbf{X})] \quad (21)$$

where  $\mathbf{X}$  is a random variable uniformly distributed on  $[0, 1]$ , and  $m_n(x)$  is given by (17) in Theorem 1.

*Proof:* See Appendix B.

Corollary 1 follows from Theorems 1 and 2.

*Corollary 1:* The relationships between the moments  $E[m_n(\mathbf{X})]$ 's and the  $\delta_{n,k}$ 's are

$$E[m_n(\mathbf{X})] = \beta \sum_{\ell=1}^n E[\delta_{\ell-1}(\mathbf{Y}) \cdot \sum_{\substack{n_1+\dots+n_i=n-\ell \\ 1 \leq i \leq n-\ell}} \delta_{n_1-1}(\mathbf{Y}) \dots \delta_{n_i-1}(\mathbf{Y})] \quad (22)$$

and

$$\delta_{n,k} = \beta \sum_{\ell=1}^n E \left[ E[m_{\ell-1}(\mathbf{X}) v_k(\mathbf{X}) v(\mathbf{X}, \mathbf{Y}) | \mathbf{Y}] \cdot \sum_{\substack{n_1+\dots+n_i=n-\ell \\ 1 \leq i \leq n-\ell}} \delta_{n_1-1}(\mathbf{Y}) \dots \delta_{n_i-1}(\mathbf{Y}) \right] \quad (23)$$

with

$$\delta_q(y) = E[v(\mathbf{X}, y) m_q(\mathbf{X})].$$

Notice that  $\delta_q(y)$  represents the quadratic form given in (20) for  $k = \lfloor yn_I \rfloor$  with  $y$  denoting the normalized column index.

When the matrix of variances  $\mathbf{V}$  introduced in (13) has a special structure, a nonrecursive explicit expression for the asymptotic moments of  $\mathbf{T}$  can be derived by resorting to techniques of noncrossing partitions (see [32] for definition) and the complementation map (see [24] for definition). More concretely, let us consider now those cases where the variances of the entries of  $\mathbf{T}$  are such that the  $n_o \times n_I$  matrix  $\mathbf{V}$  is the outer product of two vectors with deterministic nonnegative entries

$$\mathbf{V} = \mathbf{v}_o \mathbf{v}_I^T. \quad (24)$$

Equivalently,  $\mathbf{T}$  is given by the product of two deterministic diagonal matrices that pre- and post-multiply a random matrix  $\mathbf{S}$  with zero-mean and  $\frac{1}{n_o}$  variance independent entries, i.e.,

$$\mathbf{T} = \sqrt{\text{diag}(\mathbf{v}_o)} \mathbf{S} \sqrt{\text{diag}(\mathbf{v}_I)}.$$

For this common structure, we prove the following result.

*Theorem 3:* Suppose that the empirical distribution of  $\mathbf{v}_o$  converges as  $n_o \rightarrow \infty$  to an asymptotic distribution  $F_o$  and the empirical distribution of  $\mathbf{v}_I$  converges to an asymptotic distribution  $F_I$  as  $n_I \rightarrow \infty$ , then as  $n_I, n_o \rightarrow \infty$  while  $n_I/n_o \rightarrow \beta$ , the  $m$ th moment of  $\mathbf{T}\mathbf{T}^\dagger$  converges a.s. to (25) at the bottom of the page, with

$$B(m_1, \dots, m_k, n_1, \dots, n_{m+1-k}) = \frac{m(m-k)!(k-1)!}{f(m_1, \dots, m_k) \cdot f(n_1, \dots, n_{m+1-k})} \quad (26)$$

where  $\forall 1 \leq \ell \leq m$

$$f(i_1, \dots, i_\ell) = i_1! \dots i_\ell! \quad (27)$$

$$\mu_m = \sum_{k=1}^m \beta^k \sum_{\substack{m_1+\dots+m_k=m \\ m_1 \leq \dots \leq m_k}} \sum_{\substack{n_1+\dots+n_{m+1-k}=m \\ n_1 \leq \dots \leq n_{m+1-k}}} B(m_1, \dots, m_k, n_1, \dots, n_{m+1-k}) \cdot E[\mathbf{V}_I^{m_1}] \dots E[\mathbf{V}_I^{m_k}] E[\mathbf{V}_o^{n_1}] \dots E[\mathbf{V}_o^{n_{m+1-k}}] \quad (25)$$

with  $f_i$  equal to the number of entries of the vector  $[i_1, \dots, i_\ell]$  equal to  $i$ .<sup>5</sup> The random variables  $V_o$  and  $V_I$  are independent and distributed according to  $F_o$  and  $F_I$ , respectively.

*Proof:* In Appendix C.

Note that  $B(m_1, \dots, m_k, n_1, \dots, n_{m+1-k})$  is the number of noncrossing partitions  $\pi$  on  $\{1, \dots, m\}$  satisfying the following conditions:

- i) the cardinalities of the subsets in  $\pi$  are (in increasing order)  $m_1, \dots, m_k$ ;
- ii) the cardinalities of the subsets in  $K(\pi)$ , which is the complementation map (see [24]) of  $\pi$ , are (in increasing order)  $n_1, \dots, n_{m+1-k}$ .

*Corollary 2:* If  $\mathbf{V} = \mathbf{v}_o \mathbf{v}_I^T$ , then as  $n_o, n_I \rightarrow \infty$  with  $n_I/n_o \rightarrow \beta$

$$\delta_{n,k} = (\mathbf{v}_I)_k E[V_o m_n(V_o)] = (\mathbf{v}_I)_k \xi_n \quad (28)$$

where  $\xi_n$  can be computed through the following recursive equation:

$$\xi_n = \beta \sum_{\ell=1}^n E[V_o^2 m_{\ell-1}(V_o)] \cdot \sum_{\substack{n_1+\dots+n_i=n-\ell \\ 1 \leq i \leq n-\ell}} E[V_I^{i+1}] \xi_{n_1-1} \dots \xi_{n_i-1} \quad (29)$$

with

$$m_n(r) = \beta r \sum_{\ell=1}^n m_{\ell-1}(r) \cdot \sum_{\substack{n_1+\dots+n_i=n-\ell \\ 1 \leq i \leq n-\ell}} E[V_I^{i+1}] \xi_{n_1-1} \dots \xi_{n_i-1}. \quad (30)$$

For the special case in which the variances of the random entries change only from column to column and are row-wise uniform (i.e., in (24)  $\mathbf{v}_o = \mathbf{1}$  with  $\mathbf{1}$  denoting the  $n_o$ -dimensional vector with all entries equal to one), the eigenvalue moments admit the following explicit expression.

*Theorem 4:* [25]<sup>6</sup> If the matrix of the variances of  $\mathbf{T}$  is  $\mathbf{V} = \mathbf{1} \mathbf{v}_I^T$ , the  $m$ th moment of  $\mathbf{T} \mathbf{T}^\dagger$  converges a.s. to

$$\sum_{k=1}^m \beta^k \sum_{\substack{m_1+\dots+m_k=m \\ m_1 \leq \dots \leq m_k}} c(m_1, \dots, m_k) \cdot E[V_I^{m_1}] \dots E[V_I^{m_k}] \quad (31)$$

with

$$c(m_1, \dots, m_k) = \frac{m!}{(m-k+1)! \cdot f(m_1, \dots, m_k)}. \quad (32)$$

*Theorem 5:* Consider an  $n_o \times n_I$  matrix  $\mathbf{T} = \mathbf{S} \circ \mathbf{B}$  with  $\circ$  denoting the Hadamard (element-wise) product and with  $\mathbf{S}$  and  $\mathbf{B}$  two independent  $n_o \times n_I$  random matrices where  $\mathbf{B}$  has independent columns. The entries of  $\mathbf{S}$ ,  $(\mathbf{S})_{i,j}$ , are modeled

<sup>5</sup>For example,  $f(1, 1, 4, 2, 1, 2) = 3! \cdot 2! \cdot 1!$ .

<sup>6</sup>A related result with an additional condition on  $\mathbf{T}$  is given in [43].

as zero-mean i.i.d. complex random variables arbitrarily distributed with variance  $1/N$ .<sup>7</sup> Denoting by  $\lfloor \cdot \rfloor$  the closest smaller integer, assume that for a given  $x \in [0, 1)$  the empirical distribution of

$$|(\mathbf{B})_{\lfloor xn_o \rfloor, 1}|^2, \dots, |(\mathbf{B})_{\lfloor xn_o \rfloor, n_I}|^2$$

converges a.s. to a nonrandom limiting cumulative distribution  $F_x$  with all moments finite and that, for a given  $y \in [0, 1)$ , the empirical distribution of

$$|(\mathbf{B})_{1, \lfloor yn_I \rfloor}|^2, \dots, |(\mathbf{B})_{n_o, \lfloor yn_I \rfloor}|^2$$

converges a.s. to a nonrandom limit  $F_y$  also with all moments finite.

Let  $\rho(x, y) : [0, 1]^2 \rightarrow \mathbb{R}$  be a two-dimensional function such that, if  $X$  is uniform in  $[0, 1]$ , the distribution of  $\rho(X, y)$  equals  $F_y$  and, if  $Y$  is uniform in  $[0, 1]$ , the distribution of  $\rho(x, Y)$  equals  $F_x$ . Analogously, define for every  $k$  the one-dimensional function  $\rho_k(x) : [0, 1] \rightarrow \mathbb{R}$  such that if  $X$  is uniformly distributed in  $[0, 1]$ , the distribution of  $\rho_k(X)$  equals the nonrandom asymptotic empirical distribution of  $|(\mathbf{B})_{1,k}|^2, \dots, |(\mathbf{B})_{n_o,k}|^2$ .<sup>8</sup> Then, as  $n_o, n_I \rightarrow \infty$  with  $n_I/n_o \rightarrow \beta$ , the results derived in Theorems 1 and 2 hold with  $v(x, y)$  replaced by  $\rho(x, y)$ . Moreover, if the elements of  $\mathbf{B}$  admit the factorization

$$|(B)_{i,j}|^2 = C_i P_j$$

with  $\{C_i\}$  and  $\{P_j\}$  nonnegative random variables— independent on  $\mathcal{S}$ —whose empirical distributions converge a.s. then

$$\lim_{n_o \rightarrow \infty} \mu_n^{(n_o)} \stackrel{\text{a.s.}}{=} \mu_n = E[m_n(C)] \quad (33)$$

where  $m_n(r)$  is obtained through the recursive equation (34) at the bottom of the page, where  $C$  and  $P$  are independent random variables with cumulative distribution equal to the asymptotic empirical distribution of  $C_1, \dots, C_{n_o}$  and  $P_1, \dots, P_{n_I}$ , respectively.<sup>9</sup> Furthermore

$$\lim_{n_o \rightarrow \infty} \delta_{n,k}^{(n_o)} \stackrel{\text{a.s.}}{=} P_k E[C m_n(C)]. \quad (35)$$

*Proof:* In Appendix D.

Next, let us focus on the case that the entries of  $\mathbf{T}$  are correlated. In particular, let

$$\mathbf{T} = \mathbf{A}_o \mathbf{S} \bar{\mathbf{A}} \quad (36)$$

where  $\mathbf{A}_o$  is an arbitrary deterministic Toeplitz matrix,  $\mathbf{S}$  is a zero-mean random matrix with i.i.d. entries with variance  $1/n_o$ , and  $\bar{\mathbf{A}}$  is an arbitrary deterministic diagonal matrix. Important scenarios where this matrix appears include unequal

<sup>7</sup>The assumption of identical distribution can be replaced by that of equal variance  $1/N$  if the entries have bounded fourth-order moments [36].

<sup>8</sup>For a given  $y$  (respectively, for a given  $x$ )  $\rho(x, y)$  is the generalized inverse of the distribution function  $F_y$  (respectively,  $F_x$ ). For a given  $y$  (respectively,  $x$ )  $\rho(x, y)$  is a left continuous and nondecreasing function. If for a given  $y$  (respectively, for a given  $x$ )  $\rho(x, y)$  and  $F_y$  (respectively,  $F_x$ ) are strictly increasing, then they are inverse to each other [27]. Analogously, considerations hold also for  $\rho_k(x)$ .

<sup>9</sup>An explicit expression for (33) is given in (25) with  $V_o$  and  $V_I$  replaced, respectively, by  $C$  and  $P$ .

$$m_n(r) = \beta r \sum_{\ell=1}^n m_{\ell-1}(r) \sum_{\substack{n_1+\dots+n_i=n-\ell \\ 1 \leq i \leq n-\ell}} E[P^{i+1}] E[C m_{n_1-1}(C)] \dots E[C m_{n_i-1}(C)] \quad (34)$$

power DS-CDMA subject to frequency-selective fading (the spreading sequences for different paths of a given user are linearly shifted versions of the original sequence) and unequal power MC-CDMA subject to frequency dispersive and/or selective fading.

If  $\mathbf{S}$  is a zero-mean complex Gaussian matrix<sup>10</sup> or a unitarily invariant matrix<sup>11</sup> with independent entries, then it follows immediately that the moments of  $\mathbf{T}\mathbf{T}^\dagger$  are given by (25) with  $V_o$  and  $V_I$  two random variables distributed according to the asymptotic distributions  $F_{\mathbf{A}_o}$  and  $F_{\bar{\mathbf{A}}}$ , respectively. Therefore, in terms of its moments and, consequently, its asymptotic distribution,  $\mathbf{T}$  is equivalent to a matrix with zero-mean independent unequally distributed entries whose variance matrix  $\mathbf{V}$  is the outer product of two vectors  $\lambda_o$  and  $\lambda_I$  representing the squared singular values of the matrices  $\mathbf{A}_o$  and  $\bar{\mathbf{A}}$ .

For arbitrarily distributed  $\mathbf{S}$ , we prove the following—more general—result.

*Theorem 6:* Consider the  $n_o \times n_I$  random matrix

$$\mathbf{T} = [\mathbf{A}_1(\mathbf{S})_1, \dots, \mathbf{A}_{n_I}(\mathbf{S})_{n_I}] \bar{\mathbf{A}} \quad (37)$$

where  $\mathbf{S}$  represents an  $n_o \times n_I$  matrix with zero-mean i.i.d. entries with variance  $\frac{1}{n_o}$ ;  $\bar{\mathbf{A}}$  defines an arbitrary deterministic diagonal matrix whose empirical distribution converges to  $F_{\bar{\mathbf{A}}}$ . The  $n_o \times n_o$  matrices  $\mathbf{A}_k$ ,  $k = 1, \dots, n_I$  are either finite order or infinite order absolutely summable Toeplitz<sup>12</sup> independent random matrices, independent of  $\mathbf{S}$ , whose empirical distribution converges a.s. to a nonrandom distribution function. Then, as  $n_I, n_o \rightarrow \infty$  while  $n_I/n_o \rightarrow \beta$ , for any  $n \geq 1$ , the  $n$ th moment of  $\mathbf{T}\mathbf{T}^\dagger$  converges a.s. to

$$\lim_{n_o \rightarrow \infty} \frac{1}{n_o} \text{Tr} \left( (\mathbf{T}\mathbf{T}^\dagger)^n \right) = E[m_n(\mathbf{X})] \quad (38)$$

where  $\mathbf{X}$  is a random variable uniformly distributed in  $[0, 1]$ , and  $m_n(x)$  is obtained through the recursive (17) with the function  $v(x, y)$  in (17) replaced by the two-dimensional function  $\rho(x, y)$ , such that denoting by  $\lambda_i(\mathbf{A}_j)$  the  $i$ th eigenvalue of  $\mathbf{A}_j \bar{\mathbf{A}}_j^\dagger$  and by  $\bar{A}_j$  the  $j$ th coefficient of the diagonal matrix  $\bar{\mathbf{A}}$ :

- if  $\mathbf{X}$  is uniformly distributed in  $[0, 1]$ , the distribution of  $\rho(\mathbf{X}, y)$  for a given  $y$  is equal to the asymptotic empirical distribution of

$$|\bar{A}_{\lfloor y n_I \rfloor}|^2 \lambda_1(\mathbf{A}_{\lfloor y n_I \rfloor}), \dots, |\bar{A}_{\lfloor y n_I \rfloor}|^2 \lambda_{n_o}(\mathbf{A}_{\lfloor y n_I \rfloor});$$

- if  $\mathbf{Y}$  is uniformly distributed in  $[0, 1]$ , the distribution of  $\rho(x, \mathbf{Y})$  for a given  $x$  is equal to the asymptotic empirical distribution of

$$|\bar{A}_1|^2 \lambda_{\lfloor x n_o \rfloor}(\mathbf{A}_1), \dots, |\bar{A}_{n_I}|^2 \lambda_{\lfloor x n_o \rfloor}(\mathbf{A}_{n_I}).$$

Analogously

$$\begin{aligned} \lim_{n_o \rightarrow \infty} E[\delta_{n,k}^{n_o}] &= \lim_{n_o \rightarrow \infty} E[(\mathbf{T})_k^\dagger (\mathbf{T}_k \mathbf{T}_k^\dagger)^n (\mathbf{T})_k] \\ &= E[m_n(\mathbf{X}) \rho_k(\mathbf{X})] \end{aligned} \quad (39)$$

with  $m_n(x)$  defined as in (38) and  $\rho_k(x) : [0, 1] \rightarrow \mathbb{R}$  such that if  $\mathbf{X}$  is uniformly distributed in  $[0, 1]$ , the distribution of

<sup>10</sup>A complex Gaussian matrix  $\mathbf{T}$  is a matrix whose entries are jointly Gaussian complex random variables.

<sup>11</sup>A unitarily invariant random matrix  $\mathbf{T}$  is a random matrix whose distribution is invariant to unitary transformation [20].

<sup>12</sup>The assumption of *finite order* or *infinite order absolutely summable* Toeplitz matrix is made in order to simplify the proof. For the extension of the results to summable Toeplitz matrices, the reader is referred to [36]

$\rho_k(\mathbf{X})$  equals the nonrandom asymptotic empirical distribution of  $|\bar{A}_k|^2 \lambda_1(\mathbf{A}_k), \dots, |\bar{A}_k|^2 \lambda_{n_o}(\mathbf{A}_k)$ .

*Proof:* In Appendix E.

From Theorem 6 it follows that the empirical distribution (of the squared singular values) of  $\mathbf{T}$  converges a.s. to a nonrandom limiting distribution  $F_{\mathbf{T}}(\lambda)$ , whose transform  $\mathcal{S}(z)$  is [36]

$$\mathcal{S}(z) = \int_0^\infty \frac{1}{z\lambda + 1} dF_{\mathbf{T}}(\lambda) = E[u(\mathbf{X}, z)] \quad (40)$$

where  $u(x, z)$  satisfies the equation

$$u(x, z) = \frac{1}{1 + z\beta E \left[ \frac{\rho(x, \mathbf{Y})}{1 + zE[u(\mathbf{X}, z)\rho(\mathbf{X}, \mathbf{Y})|\mathbf{Y}]} \right]} \quad (41)$$

for every  $x \in [0, 1]$  with  $\mathbf{X}$  and  $\mathbf{Y}$  defined as in Theorem 6.

The following corollary is a direct consequence of Theorem 6.

*Corollary 3:* When  $\mathbf{A}_k$ 's, defined in Theorem 6, are such that  $\mathbf{A}_k = \mathbf{A}_o$  for every  $k$ , then the  $m$ th moment of  $\mathbf{T}\mathbf{T}^\dagger$  converges a.s. to (25) where the two random variables  $V_o$  and  $V_I$  are distributed according to  $F_{\mathbf{A}_o}$  and  $F_{\bar{\mathbf{A}}}$ , respectively, and with  $\bar{\mathbf{A}}$  as defined in Theorem 6.

A special case of this corollary, where  $\mathbf{S}$  is a unitarily invariant random matrix with independent entries and  $\mathbf{A}_k = \mathbf{A}_o$ ,  $k = 1, \dots, n_I$ , has been analyzed in [16], [17]. Results for the asymptotic distribution and, consequently, for the moments of random matrices with more general correlation structures than the model considered in (37) have been obtained in [34] and in [36].

In the following sections, we consider a variety of wireless communication problems corresponding to the models for  $\mathbf{T}$  analyzed in this section.

### III. DIRECT-SEQUENCE CDMA

We begin by considering DS-CDMA and expanding the analysis of [25] in order to incorporate frequency-selective fading as well as multiple receive antennas. Both uplink and downlink are analyzed and, in fact, it is shown that the downlink reduces to a special case of the uplink.

#### A. Single-Antenna DS-CDMA With Frequency-Selective Fading

1) *Uplink:* Let us consider a synchronous DS-CDMA uplink with  $K$  active users employing random spreading codes and operating over a frequency-selective fading channel. The receiver as well as each of the transmitters is equipped with a single antenna.

Denote by  $\mathbf{p}(t) = [p_1(t), \dots, p_K(t)]^T$  the  $K$ -dimensional vector representing the baseband equivalent of the transmitted signal with

$$p_k(t) = A_k \sum_n b_k(n) s_k(t - nT_s)$$

where  $A_k$ ,  $b_k(n)$ , and  $s_k(t - nT_s)$  represent the amplitude,  $n$ th symbol of the  $k$ th user, and signature waveform of the  $k$ th user in the  $n$ th symbol interval, and  $T_s$  is the symbol duration. Denoting by  $q(t)$  the chip pulse, the received signal over a linear (in

general time-varying) channel, at the output of the chip-matched filter is

$$r(t) = \left( \int_{-\infty}^{\infty} \mathbf{c}(t, \tau)^T \mathbf{p}(t - \tau) d\tau \right) * q(-t) \quad (42)$$

where the  $k$ th entry of the  $K$ -dimensional vector  $\mathbf{c}(t, \tau)$  represents the impulse response of the channel for the  $k$ th user. Assuming that the symbol duration  $T_s$  is much larger than the delay spread, we can disregard the intersymbol interference. Stacking the samples corresponding to  $r(t)$  observed in the  $n$ th interval into an  $N$ -dimensional vector we obtain

$$\mathbf{r} = \sum_{k=1}^K A_k \mathbf{C}_k(\mathbf{S})_k b_k + \mathbf{n} = \mathbf{T} \mathbf{b} + \mathbf{n} \quad (43)$$

where  $(\mathbf{S})_k$  is the transmitted signature sequence of user  $k$  while

$$\mathbf{T} = [\mathbf{C}_1(\mathbf{S})_1, \dots, \mathbf{C}_K(\mathbf{S})_K] \mathbf{A}$$

with  $\mathbf{A}$  being a  $K \times K$  deterministic diagonal matrix containing the amplitudes of the users. In turn,  $\mathbf{C}_k$  is an  $N \times N$  Toeplitz matrix defined as

$$(\mathbf{C}_k)_{i,j} = \int \int c_k(t, \tau) q(\tau - (n-1)T_b - (j-1)T_c) \cdot q^*(t - (n-1)T_b - (i-1)T_c) dt d\tau$$

with  $c_k(t, \tau)$  denoting the impulse response of the channel for the  $k$ th user independent across the users.

The signature matrix  $\mathbf{S}$  is modeled as having i.i.d. complex random entries arbitrarily distributed with zero mean and variance  $1/N$ . Furthermore,  $\mathbf{C}_k$  is assumed to be either a *finite order* or an *infinite order absolutely summable* Toeplitz matrix.<sup>13</sup> It is convenient to define a two-dimensional variance profile function on the unit square

$$\rho(x, y), \quad (x, y) \in [0, 1]^2 \quad (44)$$

such that denoting by  $\lambda_i(\mathbf{C})$ , the  $i$ th eigenvalue of  $\mathbf{C}\mathbf{C}^\dagger$

- if  $X$  is uniformly distributed in  $[0, 1]$ , the distribution of  $\rho(X, y)$  for a given  $y$  is equal to the asymptotic empirical distribution of

$$|A_{[yK]}|^2 \lambda_1(\mathbf{C}_{[yK]}), \dots, |A_{[yK]}|^2 \lambda_N(\mathbf{C}_{[yK]});$$

- if  $Y$  is uniformly distributed in  $[0, 1]$  the distribution of  $\rho(x, Y)$  for a given  $x$  is equal to the asymptotic empirical distribution of

$$|A_1|^2 \lambda_{[xN]}(\mathbf{C}_1), \dots, |A_K|^2 \lambda_{[xN]}(\mathbf{C}_K).$$

Analogously, we also introduce a one-dimensional variance profile for a given  $k$  as the one-dimensional function  $\rho_k(x) : [0, 1] \rightarrow \mathbb{R}$  such that if  $X$  is uniformly distributed in  $[0, 1]$ , the distribution of  $\rho_k(X)$  is equal to the asymptotic empirical distribution of  $|A_k|^2 \lambda_1(\mathbf{C}_k), \dots, |A_k|^2 \lambda_N(\mathbf{C}_k)$ . Assuming that the channel behaves ergodically,<sup>14</sup> a.s. neither  $\rho_k(x)$  nor  $\rho(x, y)$  depend on the channel realization.

<sup>13</sup>The results derived in this section can be extended to the case of summable Toeplitz matrices [36].

<sup>14</sup>In the absence of ergodicity, the results derived in this section remain valid but become dependant on the specific realization of the channel.

Let us now evaluate the asymptotic reduced-rank- $D$  MMSE receiver for the  $k$ th user, which is given by (5) with the asymptotic weight vector defined as in (9). Through a binomial expansion of (7), the calculation of  $\mathcal{H}_m^\infty$  boils down to the limiting value of the quadratic form

$$\delta_{n,k} = \lim_{N \rightarrow \infty} (\mathbf{T})_k^\dagger (\mathbf{T}_k \mathbf{T}_k^\dagger)^n (\mathbf{T})_k \quad (45)$$

$$= |A_k|^2 (\mathbf{S})_k^\dagger \mathbf{C}_k^\dagger \left( \sum_{\ell \neq k} |A_\ell|^2 \mathbf{C}_\ell(\mathbf{S})_\ell (\mathbf{S})_\ell^\dagger \mathbf{C}_\ell^\dagger \right)^n \mathbf{C}_k(\mathbf{S})_k. \quad (46)$$

Straight application of Theorem 6 yields

$$\begin{aligned} \mathcal{H}_m^\infty &= \lim_{N \rightarrow \infty} |A_k|^2 (\mathbf{S})_k^\dagger \mathbf{C}_k^\dagger \\ &\quad \cdot \left( \sum_{\ell \neq k} |A_\ell|^2 \mathbf{C}_\ell(\mathbf{S})_\ell (\mathbf{S})_\ell^\dagger \mathbf{C}_\ell^\dagger + \sigma^2 \mathbf{I} \right)^m \mathbf{C}_k(\mathbf{S})_k \\ &= \sum_{n=0}^m \binom{m}{n} \sigma^{2m-2n} \delta_{n,k} \\ &= \sum_{n=0}^m \binom{m}{n} \sigma^{2m-2n} E[m_n(X) \rho_k(X)] \end{aligned} \quad (47)$$

with  $m_n(x)$  obtained through the recursive (17) given in Theorem 1 with the function  $v(x, y)$  in (17) replaced by  $\rho(x, y)$  defined in (44). Therefore, the asymptotic reduced-rank- $D$  MMSE receiver is characterized in very general conditions.

Specializing this result to the case that the signal transmitted by each user propagates through  $L$  discrete chip-spaced paths (where  $L$  does not grow with  $N$ ), the channel matrix for the  $k$ th user  $\mathbf{C}_k$  becomes

$$\mathbf{C}_k = \begin{bmatrix} H_{k,1} & 0 & 0 & \dots & 0 \\ H_{k,2} & H_{k,1} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ H_{k,L} & H_{k,L-1} & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & H_{k,L} & \dots & H_{k,1} \end{bmatrix}$$

where  $H_{k,\ell}$  represents the channel fading coefficient for path  $\ell$  of user  $k$ , modeled as a Gaussian random variable with zero mean and variance  $V_\ell = E[|H_{k,\ell}|^2]$  i.i.d. across both users and paths. As  $N \rightarrow \infty$ , since the  $\mathbf{C}_k$ 's are finite-order Toeplitz matrices, their singular values are bounded. Using the explicit expression for the eigenvalues of the circulant approximation [12, Theorems 4.1 and 4.4], we have that the  $m$ th moment is equal to

$$\begin{aligned} \mu_m(\mathbf{C}_k) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N (\lambda_n(\mathbf{C}_k))^m \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \sum_{\ell=1}^L H_{k,\ell} e^{i2\pi \ell n/N} \right|^{2m} \end{aligned} \quad (48)$$

where  $\lambda_n(\mathbf{C}_k)$  are the eigenvalues of  $\mathbf{C}_k \mathbf{C}_k^\dagger$ . Let  $P$  be a random variable whose distribution is given by the asymptotic empirical distribution of

$$\left\{ |A_1|^2 \sum_{\ell=1}^L |H_{1,\ell}|^2, \dots, |A_K|^2 \sum_{\ell=1}^L |H_{K,\ell}|^2 \right\}$$

then we have the following.

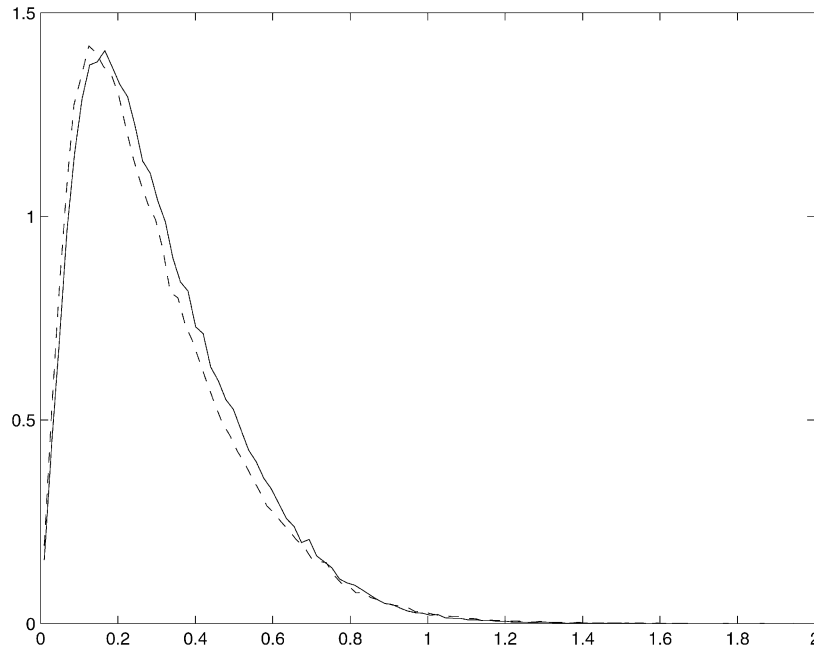


Fig. 1. The solid line is the spectrum of  $\tilde{\mathbf{T}}\tilde{\mathbf{T}}^\dagger$  while the dashed line is the density of  $\mathbf{T}\mathbf{T}^\dagger$  with  $N = 64$ ,  $K = 15$ , and  $L = 4$ .

*Corollary 4:* As  $K, N \rightarrow \infty$  with  $K/N \rightarrow \beta$ , while  $L$  is kept constant, the  $m$ th expected moment of  $\mathbf{T}\mathbf{T}^\dagger$ ,  $\mu_m$ , converges to

$$\sum_{k=1}^m \beta^k \sum_{\substack{m_1 + \dots + m_k = m \\ m_1 \leq \dots \leq m_k}} c(m_1, \dots, m_k) \cdot E[P^{m_1}] \cdots E[P^{m_k}] \quad (49)$$

with

$$c(m_1, \dots, m_k) = \frac{m!}{(m-k+1)! \cdot f(m_1, \dots, m_k)}. \quad (50)$$

Moreover, denoting by  $P_k = |A_k|^2 \sum_{\ell=1}^L |H_{k,\ell}|^2$ , we obtain

$$\lim_{n_o \rightarrow \infty} \frac{\delta_{m,k}^{(n_o)}}{P_k} \stackrel{\text{p.r.}}{=} \mu_m. \quad (51)$$

*Proof:* Appendix F.

Corollary 4 shows that with perfect knowledge of the fading coefficients at the receiver, the asymptotic expected moments for multipath fading channels have the same expression as those for a single-path fading channel with the received power of each user equal to the total received power of that user over all the paths. From the equivalence of the moments and using the same technique as in [25], the equivalence of the asymptotic full-rank MMSE output SINR between a multipath and a single-path system can also be established. Therefore, we conclude that, asymptotically as  $N \rightarrow \infty$ , each multipath interferer with a fixed number of resolvable paths acts like a single-path interferer with received power equal to the total received power of that user from all the paths. Fig. 1 depicts the eigenvalue density function of  $\mathbf{T}\mathbf{T}^\dagger$  with  $N = 64$ ,  $K = 15$ , and  $L = 4$  and compares it with the corresponding eigenvalue density function of  $\tilde{\mathbf{T}}\tilde{\mathbf{T}}^\dagger$  where  $\tilde{\mathbf{T}} = \mathbf{S}\mathbf{D}$  with  $\mathbf{D} = \text{diag}(P_1, \dots, P_K)$  and  $P_k = |A_k|^2 \sum_{\ell=1}^L |H_{k,\ell}|^2$ . Notice that, even for these moderate numbers of dimensions, both nonasymptotic distributions are already very close.

Let us now consider the case when, as  $N$  grows large, the chips become shorter and, equivalently, the system bandwidth scales up. The temporal resolution thus increases and hence the number of resolved paths  $L$  grows with  $N$ . In accordance with the classic wide-sense stationary uncorrelated scattering propagation model [15], these paths remain uncorrelated and thus we model them as zero-mean Gaussian random variables with variance  $V_\ell = E[|H_{k,\ell}|^2]$ , i.i.d. across users and paths. In this case, if the matrix  $\mathbf{C}_k$  satisfies the hypothesis of Theorem 6 then  $\eta^2 = \sum_{\ell=1}^\infty V_\ell < \infty$ ; moreover, using steps analogous to those in Corollary 4, we have that as  $K, N \rightarrow \infty$  with  $K/N \rightarrow \beta$ , and  $L/N \rightarrow \alpha$  the  $m$ th moment of the limiting eigenvalue distribution of  $\mathbf{T}\mathbf{T}^\dagger$  is given by (49) where  $P$  is a random variable whose distribution equals the limiting distribution of  $\eta^2 |A_1|^2, \dots, \eta^2 |A_K|^2$ .

2) *Downlink:* In the downlink, all users experience the same frequency-selective fading i.e.,  $\mathbf{C}_k = \mathbf{C} \forall k$ . Consequently, (43) particularizes to

$$\mathbf{r}(n) = \sum_{l=1}^K A_k \mathbf{C}(\mathbf{S})_k b_k + \mathbf{n} = \mathbf{C} \mathbf{S} \mathbf{A} \mathbf{b} + \mathbf{n} \quad (52)$$

which, contrasted with the linear general model in (1), corresponds to  $\mathbf{T} = \mathbf{C} \mathbf{S} \mathbf{A}$ . The assumptions on  $\mathbf{S}, \mathbf{C}, \mathbf{A}$  are as in Section III-A1. This implies that the empirical distribution of  $\mathbf{C}$  converges a.s. to a nonrandom limit.

Applying Corollary 3, we obtain that the expression given in (47) simplifies to<sup>15</sup>

$$\mathcal{H}_m^\infty = \sum_{n=0}^m \binom{m}{n} \sigma^{2m-2n} |A_k|^2 E[\Lambda m_n(\Lambda)] \quad (53)$$

where  $\Lambda$  is a random variable whose distribution equals the asymptotic eigenvalue distribution of  $\mathbf{C}\mathbf{C}^\dagger$ ,  $P$  is a random vari-

<sup>15</sup>A special case of this result can be found in [16], [17], where the asymptotic weights are calculated using free probability theory for CDMA with unitarily invariant i.i.d. spreading in frequency-selective fading; the case of isotropic spreading is also studied in [16], [17].

able whose distribution is given by the asymptotic distribution of  $|A_1|^2, \dots, |A_K|^2$ , and  $m_n(r)$  satisfies (54) at the bottom of the page.

### B. Multiantenna DS-CDMA With Flat Fading

In this subsection, we study the impact of having a base station equipped with multiple receive antennas. The focus here is entirely on the uplink, as the downlink was analyzed in [25].

The full-rank MMSE receiver for such a synchronous uplink is analyzed in [18] and the asymptotic maximum output SINR of the desired user has been found as the solution to a fixed-point equation. In [25], the asymptotic reduced-rank MMSE receiver is proposed and the asymptotic weights are calculated for the multiantenna receiver. In [25], the received amplitudes of each user at different antennas are constrained to be equal, a restriction that we lift in this section.

Let us denote the amplitude of user  $k$  by  $A_k > 0$ . The output of the chip-matched filter at the  $l$ th antenna is

$$\mathbf{r}_l = \sum_{k=1}^K A_k H_{k,l} b_k(\mathbf{S})_k + \mathbf{n}_l, \quad l = 1, \dots, L \quad (55)$$

where  $H_{k,l}$  indicates the fading of user  $k$  at the  $l$ th antenna,  $b_k$  is the transmitted symbol of user  $k$ , assumed to be i.i.d. across users and to have unit energy as in the previous section, and  $\mathbf{n}_l$  is the additive white Gaussian noise vector distributed as  $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$  that is independent across  $l$ .  $(\mathbf{S})_k$ , the signature sequence of user  $k$ , has independent zero mean entries with variance  $1/N$  that are independent across  $k$ . With rich scattering and a sufficient separation between antennas, the fading coefficients  $H_{k,l}$  can be modeled as independent across  $k$  and  $l$  with uniformly distributed phases on  $[0, 2\pi)$ . For a given user  $k$ ,  $H_{k,1}, \dots, H_{k,L}$  are identically distributed. However, due to the different geographic location of the users, their fading distributions are different. This models the usual microdiversity scenarios (see, e.g., [18]) in which all the receive antennas are placed at the same base station.

Equation (55) can be written as

$$\mathbf{r} = \bar{\mathbf{S}} \mathbf{b} + \mathbf{n} \quad (56)$$

where

$$\begin{aligned} \bar{\mathbf{S}} &= [(\bar{\mathbf{S}})_1, \dots, (\bar{\mathbf{S}})_K] \\ (\bar{\mathbf{S}})_k &= [A_k H_{k,1}(\mathbf{S})_k^T, \dots, A_k H_{k,L}(\mathbf{S})_k^T]^T \\ \mathbf{n} &= [\mathbf{n}_1^T, \dots, \mathbf{n}_L^T]^T \end{aligned}$$

and

$$\mathbf{b} = (b_1, \dots, b_K)^T.$$

This scenario corresponds to the general linear model in (1) where  $\mathbf{T} = \bar{\mathbf{S}}$ . The rank- $D$  reduced-rank MMSE receiver for user  $k$  is found in (5) with the weights given by (6) and

$$\mathcal{H}_m = (\bar{\mathbf{S}})_k^\dagger (\bar{\mathbf{S}}_k \bar{\mathbf{S}}_k^\dagger + \sigma^2 \mathbf{I})^m (\bar{\mathbf{S}})_k \quad (57)$$

with

$$\bar{\mathbf{S}}_k = [(\bar{\mathbf{S}})_1, \dots, (\bar{\mathbf{S}})_{k-1}, (\bar{\mathbf{S}})_{k+1}, \dots, (\bar{\mathbf{S}})_K].$$

As discussed in [25] and in the previous section, in a long-sequence DS-CDMA system we can use the asymptotic weights instead of the actual weights to reduce the complexity of the receiver. Therefore, we need to calculate the large system limit of  $\mathcal{H}_m$ .

Let  $B_{k,l} = A_k |H_{k,l}|$ . We note that the  $B_{k,l}$ 's are again independent across  $k$  and  $l$  and, for a given user  $k$ ,  $B_{k,1}, \dots, B_{k,L}$  are identically distributed.

*Proposition 1:* Let  $P_k = \sum_{l=1}^L B_{k,l}^2$  be the total received power of user  $k$  and  $F(P)$  be the asymptotic empirical distribution of  $P_1, \dots, P_K$  as  $K \rightarrow \infty$ . Then, as  $K, N \rightarrow \infty$  with  $K/N \rightarrow \beta$ ,  $\mathcal{H}_m/P_k$  converges in probability to

$$\frac{\mathcal{H}_m}{P_k} \xrightarrow{pr.} \int (\lambda + \sigma^2)^m dG(\lambda)$$

with  $G(\lambda)$  a cumulative distribution function with  $m$ th moment

$$\begin{aligned} & \int \lambda^m dG(\lambda) \\ &= \sum_{k=1}^m (\beta')^k \sum_{\substack{m_1 + \dots + m_k = m \\ m_1 \leq \dots \leq m_k}} c(m_1, \dots, m_k) \cdot E[P^{m_1}] \cdots E[P^{m_k}] \end{aligned} \quad (58)$$

and

$$c(m_1, \dots, m_k) = \frac{m!}{(m-k+1)! \cdot f(m_1, \dots, m_k)} \quad (59)$$

$$\beta' = \frac{\beta}{L}, \quad (60)$$

with  $f(m_1, \dots, m_k)$  defined as in the previous section.

We note here that  $\mathcal{H}_m^\infty$  is the same as that in a single-antenna system with processing gain  $LN$  while user  $k$ 's received power is given by the total received power  $P_k$ . This point can be proved using [18, Theorem 17] or, alternatively, generalizing the combinatorial approach of [25, Proposition 2]. In the latter case, the condition of bounded  $P_k$  is not needed (as already conjectured in [18, Theorem 17]).

For the sake of brevity, we will not make explicit the results that consider the impact of frequency selectivity and multiple receive antennas on the DS-CDMA uplink. They can be obtained using the results in Sections III-A and III-B.

## IV. MULTICARRIER CDMA

MC-CDMA is one of the most promising technologies for wireless high data rate transmission [19]. In MC-CDMA, there are  $N$  subcarriers shared by all the users. As in DS-CDMA, the information symbol to be transmitted by each user is spread onto the entire bandwidth available to the system. First, the spreading

$$m_n(r) = \beta r \sum_{\ell=1}^n m_{\ell-1}(r) \sum_{\substack{n_1 + \dots + n_i = n - \ell \\ 1 \leq i \leq n - \ell}} E[P^{i+1}] E[\Lambda_{m_{n_1-1}}(\Lambda)] \cdots E[\Lambda_{m_{n_i-1}}(\Lambda)]. \quad (54)$$

sequence of each user is modulated by its information symbol. Then, the  $N$  resulting chips are modulated onto the  $N$  subcarriers and transmitted in parallel. Therefore,  $N$  plays a similar role as the spreading gain does for DS-CDMA. In the remainder of this section, we present our results for both uplink and downlink MC-CDMA.

### A. Uplink

Given  $K$  synchronous users, the received vector at the output of the front-end is

$$\mathbf{r} = \sum_{k=1}^K A_k b_k (\tilde{\mathbf{S}})_k + \mathbf{n} \quad (61)$$

$$= \tilde{\mathbf{S}} \mathbf{A} \mathbf{b} + \mathbf{n} \quad (62)$$

where  $(\tilde{\mathbf{S}})_k = \mathbf{H}_k (\mathbf{S})_k$  and  $(\mathbf{S})_k$  is the spreading sequence of user  $k$  modeled as a vector with independent zero-mean random complex entries with variance  $1/N$ . Furthermore,  $\mathbf{H}_k = \text{diag}\{H_k^1, \dots, H_k^N\}$  with  $H_k^i$  ( $1 \leq i \leq N$ ) the fading of the  $i$ th subcarrier of user  $k$ ,  $A_k > 0$  is the amplitude of user  $k$ ,  $b_k$  is the transmitted symbol of user  $k$ , assumed to be i.i.d. across users and to have unit energy as in the previous sections, the noise  $\mathbf{n}$  is complex and distributed as  $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ ,  $\tilde{\mathbf{S}} = ((\tilde{\mathbf{S}})_1, \dots, (\tilde{\mathbf{S}})_K)$ ,  $\mathbf{A} = \text{diag}\{A_1, \dots, A_K\}$ , and  $\mathbf{b} = (b_1, \dots, b_K)^T$ . Thus, this scenario corresponds to  $(\mathbf{T})_k = A_k (\mathbf{S})_k$  and  $\mathbf{T} = \tilde{\mathbf{S}} \mathbf{A}$ , which is assumed to be known at the receiver.

In what follows, our results on the asymptotic moments of the autocorrelation matrix of the received vector and the asymptotic weights of the reduced-rank receiver are presented.

Define the variance profile  $\rho(x, y)$  as the two-dimensional function on the unit square  $[0, 1]^2$  such that, if  $X$  is uniform in  $[0, 1]$ , the distribution of  $\rho(X, y)$  equals the asymptotic empirical distribution of

$$|H_{[yK]}^1 A_{[yK]}|^2, \dots, |H_{[yK]}^N A_{[yK]}|^2$$

and, if  $Y$  is uniform in  $[0, 1]$ , the distribution of  $\rho(x, Y)$  equals the asymptotic empirical distribution of

$$|H_1^{[xN]} A_1|^2, \dots, |H_K^{[xN]} A_K|^2.$$

Analogously, we also define a one-dimensional variance profile for a given  $k$  as the one-dimensional function  $\rho_k(x) : [0, 1] \rightarrow \mathbb{R}$  such that if  $X$  is uniformly distributed in  $[0, 1]$ , the distribution of  $\rho_k(X)$  equals the asymptotic empirical distribution of  $|H_k^1 A_k|^2, \dots, |H_k^N A_k|^2$ . The channel is assumed to be ergodic in the sense that a.s. none of these asymptotic empirical distributions for any given  $y$  and  $x$  depend on the actual channel

realizations.<sup>16</sup> Under these hypotheses, applying Theorem 5 to the binomial expansion of (7), we have that

$$\begin{aligned} \mathcal{H}_m^\infty &= \lim_{N \rightarrow \infty} |A_k|^2 (\tilde{\mathbf{S}})_k^\dagger (\tilde{\mathbf{S}}_k \tilde{\mathbf{S}}_k^\dagger + \sigma^2 \mathbf{I})^m (\tilde{\mathbf{S}})_k \\ &= \sum_{n=0}^m \binom{m}{n} \sigma^{2m-2n} \delta_{n,k} \end{aligned} \quad (63)$$

where

$$\delta_{n,k} = E[m_n(X) \rho_k(X)] \quad (64)$$

with  $m_n(x)$  obtained through the recursive equation

$$\begin{aligned} m_n(x) &= \beta \sum_{\ell=1}^n m_{\ell-1}(x) E[\rho(x, Y)] \\ &\cdot \sum_{\substack{n_1 + \dots + n_i = n - \ell \\ 1 \leq i \leq n - \ell}} E[\rho(X, Y) m_{n_1-1}(X) | Y] \\ &\dots E[\rho(X, Y) m_{n_i-1}(X) | Y] \end{aligned} \quad (65)$$

with  $m_0(x) = 1$ . Plugging (63) into (9), the asymptotic weight vector and, therefore, the asymptotic reduced-rank- $D$  MMSE receiver are derived in very general conditions. Notice that here we have not imposed independence of the fading coefficients  $H_k^i$ , but just ergodicity. If ergodicity is not satisfied, then the results depend on the specific realization of the channel.

Let us next specialize the general result in (63) to the relevant case that the fading experienced by the  $K$  users is independent across  $k = 1, \dots, K$  but statistically identical with

$$E[|H_k^i|^2] = V_i. \quad (66)$$

As a result, the empirical distribution of  $|H_k^1|^2, \dots, |H_k^N|^2$  converges to a nonrandom limit that does not depend on  $k$ . While the fading at different subcarriers is usually statistically identical, in the presence of interference, the noise may not be and such differences can be accounted for by scaling the fading. Thus, we keep the model general and allow for  $V_i$  to depend on  $i$ . Furthermore, the fading experienced by each user  $k$  across different subcarriers is zero-mean and asymptotically (in the carrier separation) independent.<sup>17</sup> Denote by  $F_{\mathbf{P}}(P)$  the limiting distribution  $P_1 = A_1^2, \dots, P_K = A_K^2$  as  $K \rightarrow \infty$ . Further, denote by  $F_{\mathbf{C}}(C)$  the limit to which the empirical distribution of  $V_1, \dots, V_N$  converges for  $N \rightarrow \infty$ .<sup>18</sup> Applying Theorem 5, we have that as  $K, N \rightarrow \infty$  with  $K/N \rightarrow \beta$ , the  $m$ th moment of  $\tilde{\mathbf{S}} \mathbf{A} \mathbf{A}^\dagger \tilde{\mathbf{S}}^\dagger$  converges a.s. to (67) at the bottom of the page, where  $C$  and  $P$  are random variables whose distributions are  $F_{\mathbf{C}}(C)$  and  $F_{\mathbf{P}}(P)$ , respectively, while the coefficient  $B(m_1, \dots, m_k, n_1, \dots, n_{m+1-k})$  is defined as in (26).

<sup>16</sup>A special case where this assumption holds is when the fading  $H_k^i$  ( $1 \leq i \leq N, 1 \leq k \leq K$ ) is independent across  $i$  and  $k$ .

<sup>17</sup>This would be the case if the total bandwidth were scaled up while the width of each subcarrier were kept constant.

<sup>18</sup>If  $V_i = 1 \forall i$ , then (67) and (69) reduce to their counterparts derived in [25].

$$\sum_{k=1}^m \beta^k \sum_{\substack{m_1 + \dots + m_k = m \\ m_1 \leq \dots \leq m_k}} \sum_{\substack{n_1 + \dots + n_{m+1-k} = m \\ n_1 \leq \dots \leq n_{m+1-k}}} B(m_1, \dots, m_k, n_1, \dots, n_{m+1-k}) \cdot E[P^{m_1}] \dots E[P^{m_k}] E[C^{n_1}] \dots E[C^{n_{m+1-k}}] \quad (67)$$

For example, let  $m = 5, k = 3, m_1 = 1, m_2 = 2, m_3 = 2$ , and  $n_1 = 1, n_2 = 1, n_3 = 3$

$$B(1, 2, 2, 1, 1, 3) = \frac{5(5-3)!(3-1)!}{f(1, 2, 2) \cdot f(1, 1, 3)} = 5. \quad (68)$$

In this case, since  $m$  and  $k$  are small, it is even possible to enumerate all five noncrossing partitions that satisfy conditions i) and ii):

$\pi$	$K(\pi)$
$\{(1), (2, 3), (4, 5)\}$	$\{(2), (4), (1, 3, 5)\}$
$\{(2), (3, 4), (1, 5)\}$	$\{(3), (5), (1, 2, 4)\}$
$\{(3), (4, 5), (1, 2)\}$	$\{(4), (1), (2, 3, 5)\}$
$\{(4), (1, 5), (2, 3)\}$	$\{(5), (2), (1, 3, 4)\}$
$\{(5), (1, 2), (3, 4)\}$	$\{(1), (3), (2, 4, 5)\}$

Because the asymptotic distributions of  $|H_k^1|^2, \dots, |H_k^N|^2$  and  $|H_1^i|^2, \dots, |H_k^i|^2$  do not depend on  $k$  or on  $i$ , (64) particularizes to

$$\delta_{n,k} = P_k E[Cm_n(C)] = P_k \xi_n \quad (69)$$

with

$$\xi_n = \beta \sum_{\ell=1}^n E[m_{\ell-1}(C)C^2] \cdot \sum_{\substack{n_1+\dots+n_i=n-\ell \\ 1 \leq i \leq n-\ell}} E[P^{i+1}] \xi_{n_1-1} \dots \xi_{n_i-1} \quad (70)$$

and

$$m_n(r) = \beta r \sum_{\ell=1}^n m_{\ell-1}(r) \cdot \sum_{\substack{n_1+\dots+n_i=n-\ell \\ 1 \leq i \leq n-\ell}} E[P^{i+1}] \xi_{n_1-1} \dots \xi_{n_i-1}. \quad (71)$$

In turn, (63) simplifies to (53) where the random variables  $\Lambda$  and  $P$  are replaced by  $C$  and  $P$ , respectively. From (69), explicit

expressions for the first six  $\xi_m$ 's are listed in (72)–(77) at the bottom of the page.

To close our analysis of the MC-CDMA uplink, we further specialize it to the case of equal power users where the fading is not only independent across users, but also asymptotically independent across subcarriers and marginally equally distributed with  $E[|H_k^i|^2] = 1$ . In this case, the asymptotic weights and the corresponding asymptotic rank- $D$  MMSE receiver are obtained plugging

$$\mathcal{H}_m^\infty = \sum_{n=0}^m \binom{m}{n} \sigma^{2m-2n} \frac{1}{n} \sum_{i=1}^n \binom{n}{i} \binom{n}{i-1} \beta^i \quad (78)$$

into (9).

### B. Downlink MC-CDMA

In the downlink, the fading experienced by the  $N$  subcarriers is common to all users. If these fading coefficients are  $H_1, \dots, H_N$ , the received vector at the output of the multicarrier front end is

$$\mathbf{r} = \sum_{k=1}^K A_k b_k \mathbf{H}(\mathbf{S})_k + \mathbf{n} \quad (79)$$

$$= \mathbf{H} \mathbf{S} \mathbf{A} \mathbf{b} + \mathbf{n} \quad (80)$$

which corresponds to  $(\mathbf{T})_k = A_k \mathbf{H}(\mathbf{S})_k$  and  $\mathbf{T} = \mathbf{H} \mathbf{S} \mathbf{A}$  in the general linear model (1). The assumptions on  $\mathbf{S}, \mathbf{H}, \mathbf{A}$  are as in Section IV-A. Since the downlink can be seen as a special case of the synchronous uplink when  $\mathbf{H}_k = \mathbf{H}$  for all  $k$ , it is immediate that in terms of the asymptotic moments of

$$\mathbf{T} \mathbf{T}^\dagger = \mathbf{H} \mathbf{S} \mathbf{A} \mathbf{A}^\dagger \mathbf{S}^\dagger \mathbf{H}^\dagger$$

and of the limiting value of  $\delta_{n,k}^{(N)}$  as defined in (20), the downlink scenario is a special case of the uplink.

Consequently, the asymptotic weight of the rank- $D$  MMSE receiver in the downlink scenario can be easily derived computing

$$\mathcal{H}_m^\infty = |A_k|^2 \sum_{n=0}^m \binom{m}{n} \sigma^{2m-2n} E[Cm_n(C)] \quad (81)$$

with  $m_n(r)$  given by (71) and where  $C$  is now the random variable whose distribution is the asymptotic empirical distribution of  $\mathbf{H}$  while  $P$  is defined as in Section IV-A.

$$\xi_0 = E[C] \quad (72)$$

$$\xi_1 = \beta E[P] E[C^2] \quad (73)$$

$$\xi_2 = \beta E[P^2] E[C] E[C^2] + \beta^2 E[P]^2 E[C^3] \quad (74)$$

$$\xi_3 = \beta E[P^3] E[C]^2 E[C^2] + \beta^2 E[P] E[P^2] (2E[C] E[C^3] + E[C^2]^2) + \beta^3 E[P]^3 E[C^4] \quad (75)$$

$$\xi_4 = \beta E[P^4] E[C]^3 E[C^2] + \beta^2 E[P] E[P^3] (2E[C] E[C^2]^2 + 2E[C]^2 E[C^3]) + \beta^2 E[P]^2 (E[C] E[C^2]^2 + E[C]^2 E[C^3]) + \beta^3 E[P]^2 E[P^2] (3E[C] E[C^4] + 3E[C^2] E[C^3]) + \beta^4 E[P]^4 E[C^5] \quad (76)$$

$$\xi_5 = \beta E[P^5] E[C]^4 E[C^2] + \beta^2 E[P] E[P^4] (2E[C]^3 E[C^3] + 3E[C]^2 E[C^2]^2) + \beta^2 E[P]^2 E[P^3] (3E[C]^2 E[C^2]^2 + 2E[C]^3 E[C^3]) + \beta^4 E[P]^3 E[P^2] (4E[C^2] E[C^4] + 2E[C^3]^2 + 4E[C] E[C^5]) + \beta^3 E[P]^2 E[P^3] (6E[C] E[C^2] E[C^3] + 3E[C]^2 E[C^4] + E[C^2]^3) + \beta^5 E[P]^5 E[C^6] + \beta^3 E[P] E[P^2]^2 (6E[C] E[C^2] E[C^3] + 3E[C]^2 E[C^4] + E[C^2]^3). \quad (77)$$

### C. Asymptotic Maximum Output SINR

In this subsection, we analyze the asymptotic output SINR of the full-rank MMSE receiver for MC-CDMA. A comparison between the results obtained in Section IV-B for the downlink and the ones derived in Section IV-A for the uplink subject to (asymptotically) independent fading across subcarriers and users immediately yields that, if the asymptotic empirical distributions of  $\mathbf{H}$  and  $\mathbf{A}$  in (80) equal, respectively, the cumulative distributions  $F_{\mathcal{C}}(C)$  and  $F_{\mathcal{P}}(P)$  as defined in Section IV-A, then the asymptotic moments of  $\mathbf{T}$  in the (synchronous) uplink and in the downlink MC-CDMA (and consequently, the asymptotic reduced-rank receiver weights) are the same. From this equivalence, the equivalence of the respective asymptotic full-rank MMSE output SINRs follows. Our numerical results also show that the difference between the output MMSE SINRs of the two links is negligible for moderate system size ( $N = 64$ ,  $K = 33$ ).

For the uplink, the SINR at output of the MMSE receiver is

$$\text{SINR}_k = (\tilde{\mathbf{S}}_k^\dagger (\tilde{\mathbf{S}}_k \mathbf{A}_k \mathbf{A}_k^\dagger \tilde{\mathbf{S}}_k^\dagger + \sigma^2 \mathbf{I})^{-1} (\tilde{\mathbf{S}}_k)_k. \quad (82)$$

*Proposition 2:* Let  $\gamma_k$  be the signal-to-noise ratio (SNR) of the  $k$ th user  $|A_k|^2/\sigma^2$ . As  $N, K \rightarrow \infty$ , the SINR of the  $k = \lfloor yN \rfloor$  user,  $\text{SINR}_k$ , at the output of the MMSE receiver converges a.s. to a deterministic constant  $f(y, \gamma_k)$ , which can be obtained by solving the fixed-point equation

$$f(y, \gamma) = E \left[ \frac{\rho(X, y)}{1 + \gamma \beta E \left[ \frac{\rho(X, Y)}{1 + \gamma F(Y, \gamma)} \mid X \right]} \right] \quad (83)$$

where the two-dimensional variance profile  $\rho(x, y)$  is defined as in Section IV-A, and  $X$  and  $Y$  are independent and uniformly distributed in  $[0, 1]$ .

Proposition 2 is an immediate consequence of Lemma 1 in Appendix B whose proof can be found in [34, Lemma 7].

The downlink can be analyzed as a special case of Proposition 2.

*Proposition 3:* As  $N, K \rightarrow \infty$ , the SINR of the  $k$  user,  $\text{SINR}_k$ , at the output of the MMSE receiver converges a.s. to a deterministic constant  $|A_k|^2 F(\gamma_k)$ , which is the solution to the fixed-point equation

$$F(\gamma) = E \left[ \frac{C}{1 + \gamma \beta C E \left[ \frac{P}{1 + \gamma P F(\gamma)} \right]} \right] \quad (84)$$

where  $C$  and  $P$  are defined as in Section IV-B.

### D. Precoding

The assumption of perfect knowledge of  $\mathbf{T}$  at the receiver but not at the transmitter is very reasonable for the uplink. However, it becomes questionable if we focus our attention on a multiuser downlink where, at each receiver, only partial knowledge (specifically, the  $k$ th column) of  $\mathbf{T}$  will be available. For this case, [23] proposes a scheme where, assuming perfect knowledge of the channel matrix  $\mathbf{T}$  at the transmitter, the signal input is suitably precoded in order to manage the mutual interference between the signals being broadcasted to multiple users. The

precoder is such that the  $k$ th precoded input signal is given by the  $k$ th element of the vector

$$(\mathbf{x})_k = \left( \left( \frac{n_o}{n_I} \mathbf{T}_k^\dagger \mathbf{T}_k + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{b} \right)_k \quad (85)$$

with  $\mathbf{T}_k$  defined as in (4) and  $\mathbf{b}$  the information symbol vector. Consequently, the decision statistic will be obtained simply by projecting the received  $n_o$ -dimensional vector  $\mathbf{r}$  onto  $(\mathbf{T})_k$ . In order to avoid the complexity of computing  $(\frac{n_o}{n_I} \mathbf{T}_k^\dagger \mathbf{T}_k + \sigma^2 \mathbf{I})^{-1}$ , we could use the framework of this paper and replace (85) with a lower complexity input

$$(\mathbf{x})_k = \left( \sum_{m=0}^{D-1} (\mathbf{w}_T^\infty)_m \left( \frac{n_o}{n_I} \mathbf{T}_k^\dagger \mathbf{T}_k + \sigma^2 \mathbf{I} \right)^m \mathbf{b} \right)_k$$

where  $D$  is the rank (or the number of stages) of the precoder. Then the asymptotic weights  $\mathbf{w}_T^\infty$  can be evaluated using the results presented in this paper.

## V. NUMERICAL RESULTS

In this section, we compare the output SINR of two receivers: i) the reduced-rank MMSE receiver that uses the actual weights; and ii) the asymptotic reduced-rank MMSE receiver that uses the asymptotic weights. The output SINR of the matched filter and the full-rank MMSE receiver are also plotted for comparison. In every simulation, the channel is assumed perfectly known at the receiver. For each input SNR, the plotted output SINR is the average of the output SINRs of a large number of independent simulations. Therefore, it can be viewed as the time-average performance of the receivers in a long-sequence system.

For the implementation of the asymptotic reduced-rank receiver, the moments of  $F_A$ , the limiting empirical distribution of  $\mathbf{A} = (a_1, \dots, a_n)$ , are approximated for finite  $n$  by those of the empirical distribution of  $a_1, \dots, a_n$ , i.e.,

$$E[A^m] = \int a^m dF_A(a) \approx \frac{1}{n} \sum_{\ell=1}^n a_\ell^m. \quad (86)$$

### A. Frequency-Selective Fading Channels

Specializing the expression of  $\mathbf{C}_k$  to the case of three chip-spaced paths, we have

$$\mathbf{C}_k = \begin{bmatrix} H_{k,1} & 0 & \dots & \dots & 0 \\ H_{k,2} & H_{k,1} & \dots & \dots & 0 \\ H_{k,3} & H_{k,2} & H_{k,1} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & H_{k,3} & H_{k,2} & H_{k,1} \end{bmatrix}.$$

The channel-fading coefficient  $H_{k,\ell}$ , for path  $\ell$  of user  $k$ , is modeled as a Gaussian random variable with zero mean and variance  $V_\ell$  independent across both users and paths. The total average received power over the three paths of the interfering users is 5 dB above the desired user, which makes for a relatively challenging environment. The horizontal axis represents the input SNR  $P_1/\sigma^2$  where  $P_1$  is the total received power over the three paths of user 1. In the results,  $N = 64$ ,  $K = 15$ , and  $D = 6$  (number of stages of the reduced-rank receiver) are used.

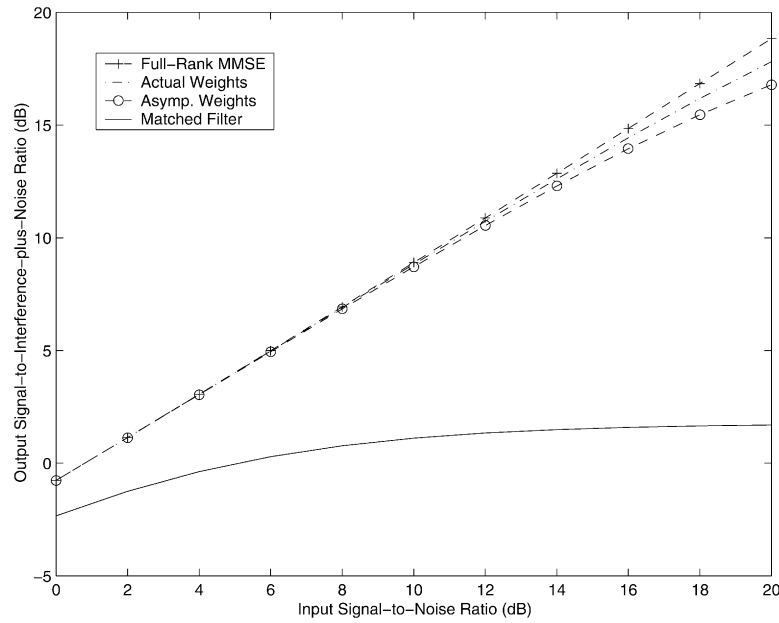


Fig. 2. Frequency-selective fading: the spreading sequences for each of the three paths of a given user are independent.  $N = 64$ ,  $K = 15$ , and  $D = 6$ .

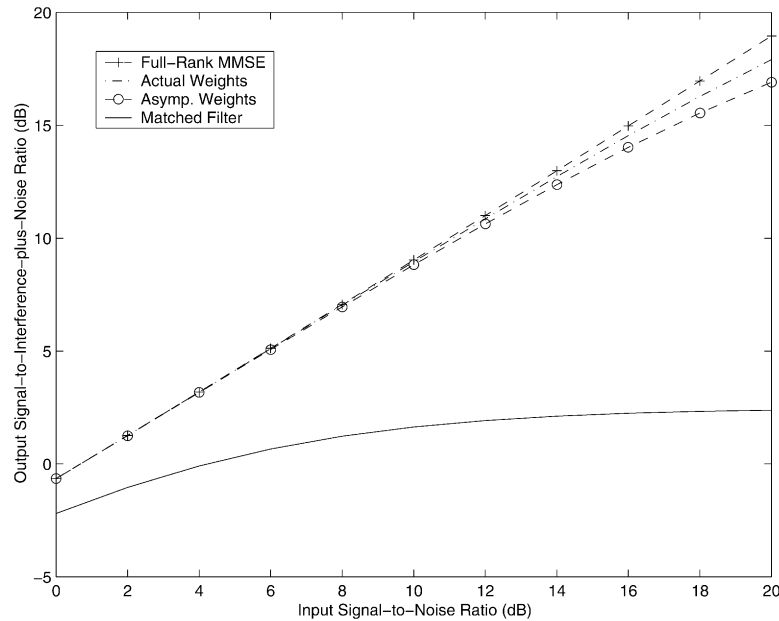


Fig. 3. Frequency-selective fading: the spreading sequences for each of the three paths of a given user are shifted with partial overlap.  $N = 64$ ,  $K = 15$ , and  $D = 6$ .

Fig. 2 shows the performance of the receivers when the spreading sequences for different paths of a given user are independent. We can see that the output SINR of the asymptotic reduced-rank receiver approximates very closely that of the receiver using actual weights for the input SNR considered, and both of them perform much better than the matched filter.

Fig. 3 plots the performance of the receivers when the spreading sequences for different paths of a given user are shifted with partial overlap (i.e., with zeros preceding the shifted sequence). The resemblance of Fig. 3 to Fig. 2 buttresses the analytical results in Section III-A1.

### B. Multiantenna

In these numerical results,  $N = 32$ ,  $K = 65$ ,  $L = 4$  (number of antennas), and  $D = 6$  (number of stages) are used. The received amplitudes  $B_{k,l}$ 's are independent Rayleigh random variables, and for a given  $k$ ,  $B_{k,1}, \dots, B_{k,L}$  are identically distributed. The distributions across users have different second moments (powers) and are chosen so that the powers of the interfering users are 5 dB higher than the desired user. From Fig. 4, we can see that the performance of the asymptotic reduced-rank receiver approximates that of the reduced-rank receiver using

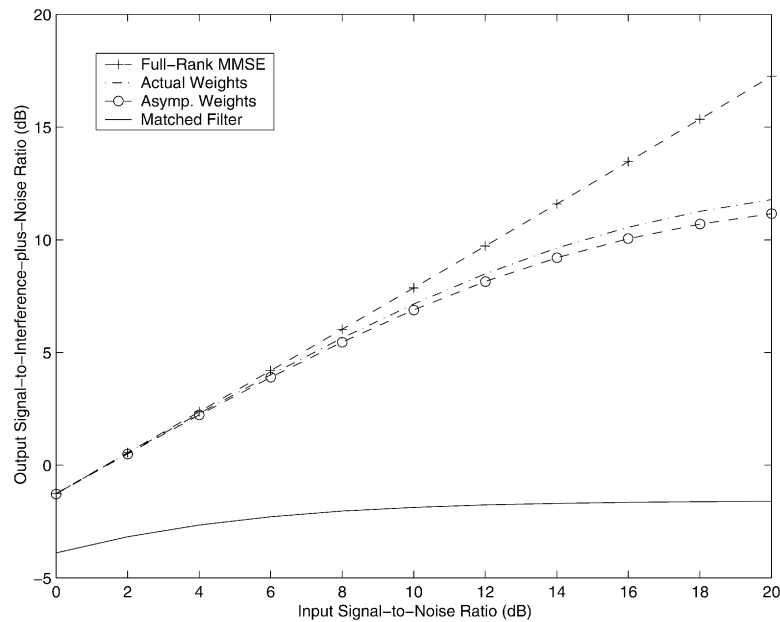


Fig. 4. Multiantenna:  $N = 32, K = 65, L = 4, D = 6$ .

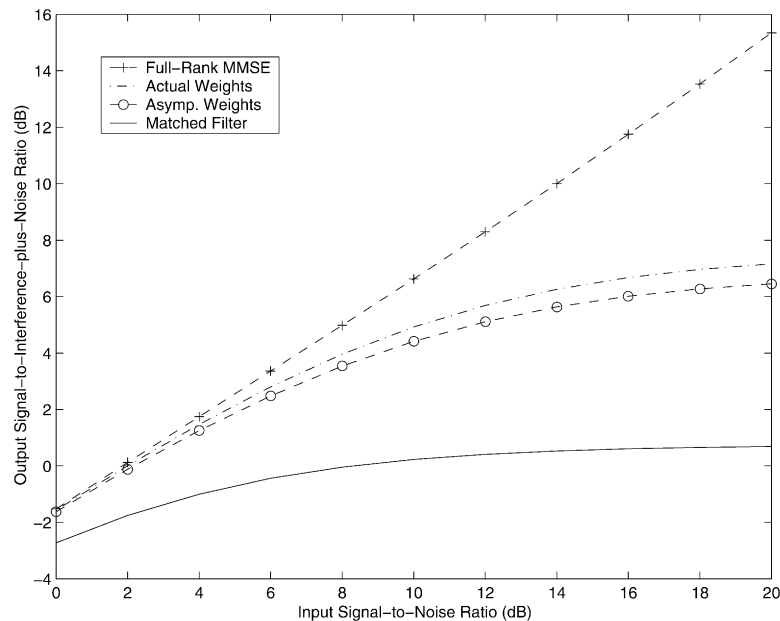


Fig. 5. Uplink MC-CDMA:  $N = 64, K = 33, \text{ and } D = 3$ .

actual weights and the maximum loss in the output SINR is less than 1 dB for the input SNR considered.

### C. Uplink MC-CDMA

In these results, the transmitted amplitudes of the interfering users are uniformly distributed and the interfering users have the same transmitted power as the desired user on average. The fading experienced by each user  $k$  is zero-mean independent across  $k = 1, \dots, K$  and across different subcarriers. The limiting empirical distribution of  $\sqrt{V_1}, \dots, \sqrt{V_N}$  is a Rayleigh distribution with unit second-order moment, and a three-stage reduced-rank receiver is used in the simulation. We point out that significant gain in the output SINR of the reduced-rank receivers

can be achieved if more stages are used. In Figs. 5 and 6 we see that the performance loss due to the use of asymptotic weights is small for the input SNR considered. Moreover, the performance gap between the upper two curves in Fig. 6 is smaller than that in Fig. 5. This suggests that the performance of the asymptotically weighted reduced-rank receiver converges to that of the reduced-rank receiver using the actual weights in a large CDMA system.

### D. Downlink MC-CDMA

The power profile of the interfering users and the number of stages are the same as in the uplink case. The fading coefficients

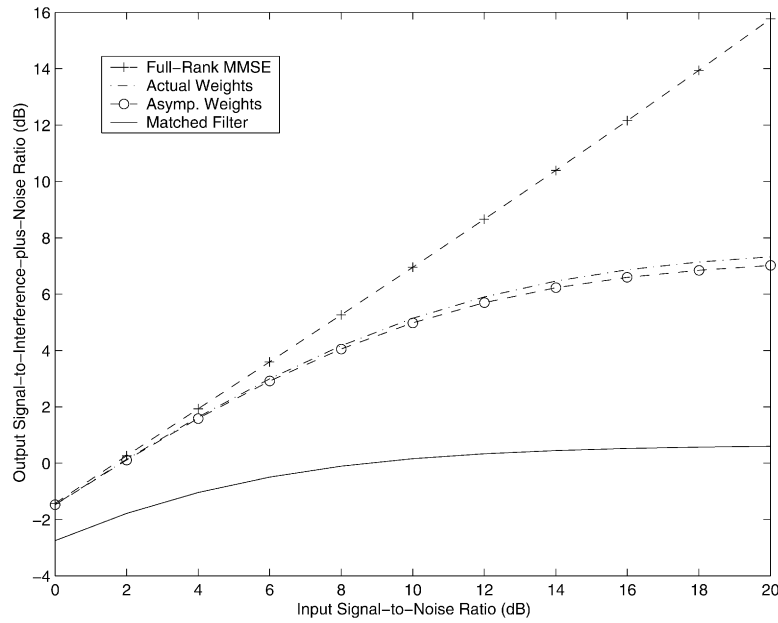


Fig. 6. Uplink MC-CDMA:  $N = 128$ ,  $K = 65$ , and  $D = 3$ .

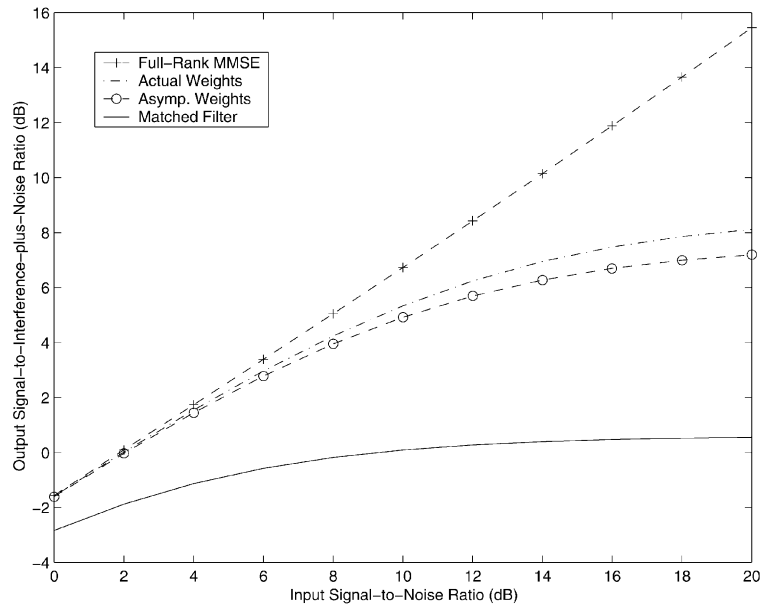


Fig. 7. Downlink MC-CDMA:  $N = 32$ ,  $K = 17$ ,  $D = 3$ .

are complex Gaussian zero-mean and independent across different subcarriers with  $E[|H_k^i|^2] = 1$ . The similarity of Fig. 5 to Fig. 8 is not surprising given the asymptotic equivalence between downlink and uplink. Figs. 7 and 8 present the same downlink scenario with different numbers of users and different processing gains, always such that  $K/N \simeq 1/2$ . A comparison of those figures also illustrates the rapid almost sure convergence of the  $\mathcal{H}_m$ 's to the same limit.

#### E. Downlink and Uplink Full-Rank MMSE

We compare the full-rank MMSE output SINR of the uplink and downlink MC-CDMA systems for  $N = 64$  and  $K = 33$ . The power profile of the interfering users and the fading profile

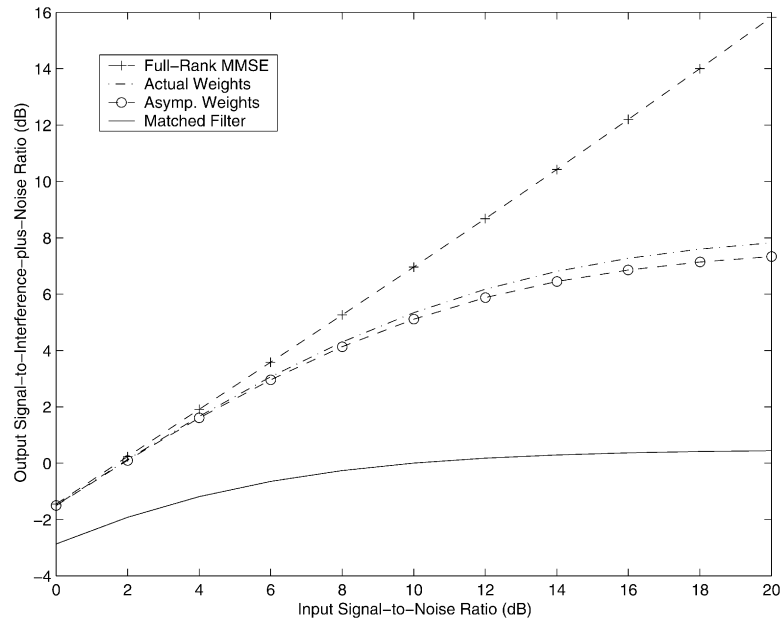
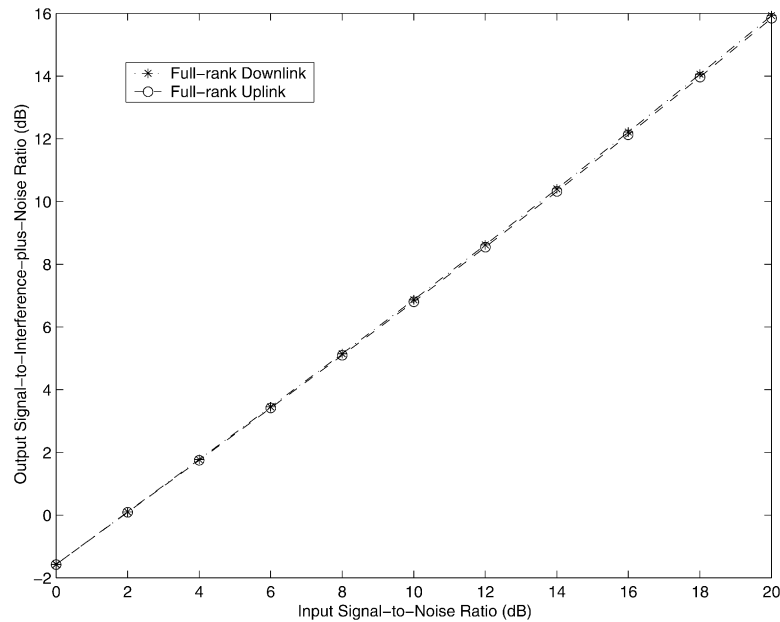
of the subcarriers are the same as before. As seen in Fig. 9, the difference between the two links is negligible in accordance with Section IV-C.

#### APPENDIX A PROOF OF THEOREM 1

The proof of Theorem 1 requires the following result [9, Corollary 10.1.2].

*Theorem 7:* Let  $\mathbf{T}$  be an  $n_o \times n_I$  random matrix with independent entries which are zero-mean and satisfy the condition

$$n_o \text{Var}(\mathbf{T}_{i,j}) < a \quad (87)$$


 Fig. 8. Downlink MC-CDMA:  $N = 64$ ,  $K = 33$ ,  $D = 3$ .

 Fig. 9. Full-rank MMSE:  $N = 64$ ,  $K = 33$ .

for some uniform bound  $a < \infty$ . Moreover, define for each  $n_o$  a function  $v_{n_o} : [0, 1]^2 \rightarrow \mathcal{R}$  as

$$v_{n_o}(x, y) = n_o \text{Var}(\mathbf{T}_{i,j}), \quad (88)$$

$$i, j \text{ satisfying } \frac{i}{n_o} \leq x < \frac{(i+1)}{n_o}, \frac{j}{n_I} \leq y < \frac{(j+1)}{n_I}$$

and suppose that  $v_{n_o}$  converges uniformly to a limiting bounded function  $v$  with a countable number of discontinuities.<sup>19</sup> Then, as  $n_o \rightarrow \infty, n_I \rightarrow \infty$  with  $n_I/n_o \rightarrow \beta$ , for each  $z \in \mathcal{C}^+$  the empirical distribution (of the squared singular values) of  $\mathbf{T}$  con-

verges a.s. to a nonrandom limiting distribution  $F_{\mathbf{T}}(\lambda)$ , whose transform  $\mathcal{S}(z)$  is

$$\mathcal{S}(z) = \int_0^\infty \frac{1}{z\lambda + 1} dF_{\mathbf{T}}(\lambda) = \int_0^1 u(x, z) dx. \quad (89)$$

where  $u(x, z)$  satisfies the equation

$$u(x, z) = \frac{1}{1 + z\beta \int_0^1 \frac{v(x,y)dy}{1+z \int_0^1 u(w,z)v(w,y)dw}} \quad (90)$$

for every  $x \in [0, 1]$ . The solution of (90) exists and is unique in the class of functions  $u(x, z) \geq 0$ , analytical on  $z$  and continuous on  $x \in [0, 1]$ .

<sup>19</sup>The continuous assumption in [9] can be dropped following the method of [14], [30].

Here we note that the almost sure convergence of the empirical distribution of  $\mathbf{T}$  can also be found in [14]. From this almost sure convergence it follows that

$$\mu_n^{(n_o)} = \frac{1}{n_o} \text{Tr}(\mathbf{T}\mathbf{T}^\dagger)^n \quad (91)$$

$$\xrightarrow{\text{a.s.}} \lim_{n_o \rightarrow \infty} \frac{1}{n_o} E[(\mathbf{T}\mathbf{T}^\dagger)^n]. \quad (92)$$

Consequently, to prove Theorem 1 we need to prove the following equality:

$$\lim_{n_o \rightarrow \infty} \frac{1}{n_o} E[(\mathbf{T}\mathbf{T}^\dagger)^n] = \mu_n \quad (93)$$

with  $\mu_n$  defined as in (16). To that end, let us notice that from (90)

$$u(x, \gamma) = 1 - \beta \gamma u(x, \gamma) E \left[ \frac{v(x, Y)}{1 + \gamma E[v(X, Y)u(X, \gamma)|Y]} \right] \quad (94)$$

with  $\gamma$  a nonnegative real number and with  $X$  and  $Y$  two fictitious independent random variables uniformly distributed in  $[0, 1]$ . If we assume that  $u(x, \gamma)$  also admits the following power series in  $\gamma$ :

$$\begin{aligned} u(x, \gamma) &= \sum_{n=0}^{\infty} m_n(x) (-\gamma)^n \\ &= 1 + \sum_{n=1}^{\infty} m_n(x) (-\gamma)^n \end{aligned} \quad (95)$$

where  $m_0(x) = 1$ , then, substituting (95) into (94), we obtain (96) at the bottom of the page. Identifying and equating the coefficients of every degree at either side of the equality, we obtain (97) and (98), also at the bottom of the page, where we recall that  $m_0(x) = 1$ . From (89) in Theorem 7 we know that [40, Remark 3.3.3]

$$\begin{aligned} \mathcal{S}(\gamma) &= 1 + \sum_{n=1}^{\infty} E[\lambda^n] (-\gamma)^n \\ &= E[u(X, \gamma)] \\ &= 1 + \sum_{n=1}^{\infty} E[m_n(X)] (-\gamma)^n \end{aligned} \quad (99)$$

where  $X$  is a fictitious random variable uniformly distributed in  $[0, 1]$ . Thus,  $E[m_n(X)] = E[\lambda^n]$  represents the  $n$ th moment of  $\mathbf{T}\mathbf{T}^\dagger$ .

## APPENDIX B PROOF OF THEOREM 2

The proof of Theorem 2 can be obtained following analogous steps to the ones described in Appendix A using as a starting point the following result.

*Lemma 1:* [34, Lemma 7] Let  $\mathbf{T}$  be an  $n_o \times n_I$  matrix as defined in Theorem 7. Further define

$$F^{(n_o)}(y, \gamma) = (\mathbf{T})_j^\dagger \left( \mathbf{I} + \gamma \sum_{q \neq j} (\mathbf{T})_q (\mathbf{T})_q^\dagger \right)^{-1} (\mathbf{T})_j, \quad \text{with } \frac{j-1}{n_I} \leq y < \frac{j}{n_I}.$$

As  $n_o, n_I$  grow large with their ratio kept constant,  $F^{(n_o)}$  converges a.s. to a function  $F(y, \gamma)$  that is the solution to the fixed-point equation

$$F(y, \gamma) = E \left[ \frac{v(X, y)}{1 + \gamma \beta E \left[ \frac{v(X, Y)}{1 + \gamma F(Y, \gamma)} | X \right]} \right] \quad (100)$$

with  $X$  and  $Y$  two fictitious independent random variables uniformly distributed in  $[0, 1]$ .

## APPENDIX C PROOF OF THEOREM 3

Applying again Theorem 7 we know that the empirical distribution of  $\mathbf{T} = \sqrt{\text{diag}(\mathbf{v}_o)} \mathbf{S} \sqrt{\text{diag}(\mathbf{v}_I)}$  converges a.s. to the nonrandom limiting distribution whose expression is given in (89), (90). Consequently

$$\mu_k^{(n_o)} = \frac{1}{n_o} \text{Tr}(\mathbf{T}\mathbf{T}^\dagger)^k \quad (101)$$

$$\xrightarrow{\text{a.s.}} \lim_{n_o \rightarrow \infty} \frac{1}{n_o} E[(\mathbf{T}\mathbf{T}^\dagger)^k] \quad (102)$$

---


$$\sum_{n=1}^{\infty} m_n(x) (-\gamma)^n = \beta \sum_{k=1}^{\infty} (-\gamma)^k m_{k-1}(x) E \left[ v(x, Y) \sum_{\ell=0}^{\infty} \left( \sum_{i=0}^{\infty} E[v(X, Y) m_i(X) | Y] (-\gamma)^{i+1} \right)^\ell \right]. \quad (96)$$


---

$$m_n(x) = \beta \sum_{\ell=1}^n m_{\ell-1}(x) E \left[ v(x, Y) \sum_{i=1}^{n-\ell} \sum_{n_1 + \dots + n_i = n-\ell} E[v(X, Y) m_{n_1-1}(x)] \dots E[v(X, Y) m_{n_i-1}(x)] \right] \quad (97)$$

$$= \beta \sum_{\ell=1}^n m_{\ell-1}(x) E \left[ v(x, Y) \sum_{\substack{n_1 + \dots + n_i = n-\ell \\ 1 \leq i \leq n-\ell}} E[v(X, Y) m_{n_1-1}(x)] \dots E[v(X, Y) m_{n_i-1}(x)] \right] \quad (98)$$

from which follows that in order to prove Theorem 3 we need to prove that the  $m$ th expected asymptotic moment of  $\mathbf{TT}^\dagger$  is given by (25). For simplicity of notation we let

$$\mathbf{C} = \text{diag}(\mathbf{v}_o) \quad \mathbf{P} = \text{diag}(\mathbf{v}_I). \quad (103)$$

Consequently, the  $m$ th expected asymptotic moment of  $\mathbf{TT}^\dagger$  is as given by (104)–(108) at the bottom of the following page, where, in (104)–(108),  $P_i$  and  $C_i$  are the  $i$ th coefficients of the diagonal matrices  $\mathbf{P}$  and  $\mathbf{C}$ , respectively, and  $W_{i,j}$  represents the  $(i, j)$ th entry of  $\mathbf{S}$ .

From [25], we know that the only nonzero contribution to (108) arises as follows: for a given partition  $\pi$  of indexes  $\{j_1, \dots, j_m\}$  into  $k$  disjoint subsets (recall that indexes assigned to the same subset are identical in the summation, whereas indexes assigned to different subsets are distinct), there exists a good partition  $\tau$  of  $\{i_1, \dots, i_m\}$  into  $(m+1-k)$  subsets such that the product  $W_{i_1, j_m}^* W_{i_1, j_1} \cdots W_{i_m, j_{m-1}}^* W_{i_m, j_m}$  is composed of  $m$  modulo squares of random variables. Then, we take expectation with respect to the random spreading sequences. Considering the fact that each random chip value has variance  $1/n_o$ , we have (109) and (110), also at the bottom of the page, where two partitions  $\pi$  and  $\tau$  are mutually good if in

$$\begin{pmatrix} (i_1, j_1) & (i_2, j_2) & \cdots & (i_m, j_m) \\ (i_1, j_2) & (i_2, j_3) & \cdots & (i_m, j_1) \end{pmatrix} \quad (111)$$

every ordered pair in the upper row is matched<sup>20</sup> with exactly one ordered pair in the lower row, and every ordered pair in the lower row is matched with exactly one ordered pair in the upper row.

<sup>20</sup>Two ordered pairs are matched iff the first indexes in both of them are in the same set, and so are the second indexes.

The above matching ensures that no random variable appears alone in the expectation, which would make the expectation zero. If such a partition of  $j_1, \dots, j_m$  exists, we call it a *good* partition for the given partition of  $i_1, \dots, i_m$ .

Suppose that the cardinalities of the subsets in  $\pi$  are  $m_1, \dots, m_k$ , and the cardinalities of the subsets in  $\tau$  are  $n_1, \dots, n_{m+1-k}$ , then (110) converges to

$$\sum_{\tau, \pi \text{ mutually good}} \beta^k E[\mathbf{V}_I^{m_1}] \cdots E[\mathbf{V}_I^{m_k}] E[\mathbf{V}_o^{n_1}] E[\mathbf{V}_o^{n_{m+1-k}}] \quad (112)$$

where  $\mathbf{V}_I$  and  $\mathbf{V}_o$  denote two random variables whose distribution is given by the asymptotic empirical distribution of  $\mathbf{v}_I$  and  $\mathbf{v}_o$ , respectively, while  $k$  is the number of disjoint subsets in the partition.

To avoid cumbersome notations,  $m = 4$  is used to illustrate the idea by which the convergence to (112) is proved. For general  $m$ , (112) can be proved by means of induction. Let  $m = 4$ ,  $k = 2$ ,  $\pi = \{(j_1, j_2), (j_3, j_4)\}$ , then  $\tau = \{(i_1), (i_3), (i_2, i_4)\}$ . According to (110) the contribution of  $\pi, \tau$  is

$$\lim \frac{1}{n_o^5} \sum_{(j_1=j_2) \neq (j_3=j_4)} P_{j_1} P_{j_2} P_{j_3} P_{j_4} \cdot \sum_{i_1 \neq i_3 \neq (i_2=i_4)} C_{i_1} C_{i_2} C_{i_3} C_{i_4} \quad (113)$$

$$= \lim \frac{1}{n_o^5} \sum_{j_1 \neq j_3} P_{j_1}^2 P_{j_3}^2 \sum_{i_1 \neq i_2 \neq i_3} C_{i_1} C_{i_2}^2 C_{i_3} \quad (114)$$

$$= \left( \lim \frac{1}{n_o^2} \sum_{j_1 \neq j_3} P_{j_1}^2 P_{j_3}^2 \right) \cdot \left( \lim \frac{1}{n_o^3} \sum_{i_1 \neq i_2 \neq i_3} C_{i_1} C_{i_2}^2 C_{i_3} \right). \quad (115)$$

$$E[\mu_m] = \lim \frac{1}{n_o} E \left[ \text{Tr} \left\{ (\mathbf{CSPS}^\dagger)^m \right\} \right] \quad (104)$$

$$= \lim \frac{1}{n_o} E \left[ \text{Tr} \left\{ \sum_{j_1, \dots, j_m=1}^{n_I} (\mathbf{C} \mathbf{s}_{j_1} \mathbf{s}_{j_1}^\dagger) \cdots (\mathbf{C} \mathbf{s}_{j_m} \mathbf{s}_{j_m}^\dagger) P_{j_1} \cdots P_{j_m} \right\} \right] \quad (105)$$

$$= \lim \frac{1}{n_o} E \left[ \sum_{j_1, \dots, j_m=1}^{n_I} \mathbf{s}_{j_m}^\dagger \mathbf{C} \mathbf{s}_{j_1} \cdots \mathbf{s}_{j_{m-1}}^\dagger \mathbf{C} \mathbf{s}_{j_m} P_{j_1} \cdots P_{j_m} \right] \quad (106)$$

$$= \lim \frac{1}{n_o} E \left[ \sum_{j_1, \dots, j_m=1}^{n_I} \sum_{i_1, \dots, i_m=1}^{n_o} W_{i_1, j_m}^* C_{i_1} W_{i_1, j_1} \cdots W_{i_m, j_{m-1}}^* C_{i_m} W_{i_m, j_m} P_{j_1} \cdots P_{j_m} \right] \quad (107)$$

$$= \lim \frac{1}{n_o} \sum_{j_1, \dots, j_m=1}^{n_I} \sum_{i_1, \dots, i_m=1}^{n_o} E \left[ W_{i_1, j_m}^* W_{i_1, j_1} \cdots W_{i_m, j_{m-1}}^* W_{i_m, j_m} \right] P_{j_1} \cdots P_{j_m} C_{i_1} \cdots C_{i_m} \quad (108)$$

$$E[\mu_m] = \lim \frac{1}{n_o} \sum_{j_1, \dots, j_m=1}^{n_I} \sum_{i_1, \dots, i_m=1}^{n_o} E \left[ |W_{i_1, j_1}|^2 \cdots |W_{i_m, j_m}|^2 \right] P_{j_1} \cdots P_{j_m} C_{i_1} \cdots C_{i_m} \quad (109)$$

$$= \lim \frac{1}{n_o^{m+1}} \sum_{\tau, \pi \text{ mutually good}} P_{j_1} \cdots P_{j_m} C_{i_1} \cdots C_{i_m}. \quad (110)$$

Since

$$\begin{aligned} & \frac{1}{n_o} \sum_{j_1 \neq j_3} P_{j_1}^2 P_{j_3}^2 \\ &= \frac{1}{n_o} \sum_{j_1=1}^{n_1} P_{j_1}^2 \cdot \frac{1}{n_o} \sum_{j_3=1}^{n_1} P_{j_3}^2 - \frac{1}{n_o} \sum_{j_1=j_3=1}^{n_1} P_{j_1}^2 P_{j_3}^2 \quad (116) \end{aligned}$$

$$= \frac{n_1}{n_o} \frac{1}{n_1} \sum_{j_1=1}^{n_1} P_{j_1}^2 \cdot \frac{n_1}{n_o} \frac{1}{n_1} \sum_{j_3=1}^{n_1} P_{j_3}^2 - \frac{1}{n_o} \sum_{j_1=1}^{n_1} P_{j_1}^4 \quad (117)$$

$$\rightarrow \beta^2 E[V_I^2]^2 \quad (118)$$

because the first term in (117) converges to  $\beta^2 E[V_I^2]^2$  from the convergence of the empirical distribution of  $P_1, \dots, P_{n_1}$ , and the second term converges to zero because

$$\frac{1}{n_o} \sum_{j_1=1}^{n_1} P_{j_1}^4 \rightarrow \beta E[V_I^4]. \quad (119)$$

For the second factor in (115)

$$\begin{aligned} & \frac{1}{n_o^3} \sum_{i_1 \neq i_2 \neq i_3} C_{i_1} C_{i_2}^2 C_{i_3} \\ &= \frac{1}{n_o} \sum_{i_1=1}^{n_o} C_{i_1} \cdot \frac{1}{n_o} \sum_{i_2=1}^{n_o} C_{i_2}^2 \cdot \frac{1}{n_o} \sum_{i_3=1}^{n_o} C_{i_3} \\ & \quad - \frac{1}{n_o^3} \sum_{(i_1=i_2) \neq i_3} C_{i_1} C_{i_2}^2 C_{i_3} \\ & \quad - \frac{1}{n_o^3} \sum_{(i_1=i_3) \neq i_2} C_{i_1} C_{i_2}^2 C_{i_3} - \frac{1}{n_o^3} \sum_{(i_2=i_3) \neq i_1} C_{i_1} C_{i_2}^2 C_{i_3} \\ & \quad - \frac{1}{n_o^3} \sum_{i_1=i_2=i_3} C_{i_1} C_{i_2}^2 C_{i_3} \quad (120) \end{aligned}$$

$$\rightarrow E[V_o]^2 E[V_o^2]. \quad (121)$$

where (121) follows because the first term in (120) converges to  $E[V_o]^2 E[V_o^2]$ , and similar to (119) all the other terms converge to zero. So the overall contribution of  $\pi, \tau$  is  $\beta^2 E[V_I^2]^2 E[V_o]^2 E[V_o^2]$ , which agrees with (112). For  $m = 4$ , all the other terms in (112) can be obtained by following a similar argument.

Therefore, in terms of noncrossing partitions and the complementation map, (112) can be expressed as (122) at the bottom of the page, where the coefficient

$$B(m_1, \dots, m_k, n_1, \dots, n_{m+1-k})$$

is the number of noncrossing partitions  $\pi$  satisfying the following conditions:

- i) the cardinalities of the subsets in  $\pi$  are (in increasing order)  $m_1, \dots, m_k$ ;
- ii) the cardinalities of the subsets in  $K(\pi)$ , which is the complementation map (see [24]) of  $\pi$ , are (in increasing order)  $n_1, \dots, n_{m+1-k}$ .

For given small  $m$  and  $k$ , the coefficient

$$B(m_1, \dots, m_k, n_1, \dots, n_{m+1-k})$$

can be obtained by enumerating the noncrossing partitions that satisfy conditions i) and ii). To obtain an explicit expression for  $B(m_1, \dots, m_k, n_1, \dots, n_{m+1-k})$ , we need to solve the above combinatorial problem for general values of  $m$  and  $k$ . This problem can be solved by the embedding of noncrossing partitions into the permutations group. It was shown by Biane in [2] that  $B(m_1, \dots, m_k, n_1, \dots, n_{m+1-k})$  represents the number of factorizations of a cycle of length  $m$  into a product of two permutations of respective cycle type  $(m_1, \dots, m_k)$  and  $(n_1, \dots, n_{m+1-k})$ . An explicit expression for  $B(m_1, \dots, m_k, n_1, \dots, n_{m+1-k})$  is given by

$$\begin{aligned} & B(m_1, \dots, m_k, n_1, \dots, n_{m+1-k}) \\ &= \frac{m(m-k)!(k-1)!}{f(m_1, \dots, m_k) \cdot f(n_1, \dots, n_{m+1-k})} \quad (123) \end{aligned}$$

with  $f(m_1, \dots, m_k)$  and  $f(n_1, \dots, n_{m+1-k})$  defined as in (27).

#### APPENDIX D

##### PROOF OF THEOREM 5

Let us first consider the case  $(\mathbf{B})_{i,j} = \sqrt{C_i P_j}$ , whereby  $\mathbf{T}$  simplifies to

$$\mathbf{T} = \mathbf{S} \circ \mathbf{B} = \mathbf{B}_{n_o} \mathbf{S} \mathbf{B}_{n_1}$$

with  $\mathbf{B}_{n_o}$  and  $\mathbf{B}_{n_1}$  random diagonal matrices defined as

$$\begin{aligned} \mathbf{B}_{n_o} &= \text{diag}(\sqrt{C_1}, \dots, \sqrt{C_{n_o}}) \\ \mathbf{B}_{n_1} &= \text{diag}(\sqrt{P_1}, \dots, \sqrt{P_{n_1}}). \quad (124) \end{aligned}$$

In this case, it can be easily inferred from [31, Theorem 1] (for details see [36]) that the empirical distribution (of the squared singular value) of  $\mathbf{T}$  converges a.s. to a nonrandom limit  $F_{\mathbf{T}}$  whose transform is

$$\mathcal{S}(z) = \int_0^\infty \frac{1}{z\lambda + 1} dF_{\mathbf{T}}(\lambda) = E[u(C, z)] \quad (125)$$

where  $u(c, z)$  is the solution of the following fixed-point equation:

$$u(c, z) = \frac{1}{1 + z\beta c E\left[\frac{P}{1+zPE[Cu(C, z)]}\right]} \quad (126)$$

with  $C$  and  $P$  two independent random variables whose distributions equal the almost sure asymptotic empirical distributions of  $\mathbf{B}_{n_o}$  and  $\mathbf{B}_{n_1}$ , respectively.

Let us next focus on the more general case  $\mathbf{T} = \mathbf{S} \circ \mathbf{B}$ . First, we consider the scenario where the entries of  $\mathbf{B}$ ,  $(\mathbf{B})_{i,j}$ 's, are random variables compactly supported and such that

$$|(\mathbf{B})_{i,j}|^2 < a$$

for some uniform bound  $a < \infty$ . Then, applying Theorem 7 we have that, conditioned on  $\mathbf{B}$ , the empirical distribution of  $\mathbf{T}$  converges to (90) with  $v(x, y)$  defined as in Theorem 7. Denoting

$$\sum_{k=1}^m \beta^k \sum_{\substack{m_1 + \dots + m_k = m \\ m_1 \leq \dots \leq m_k}} \sum_{\substack{n_1 + \dots + n_{m+1-k} = m \\ n_1 \leq \dots \leq n_{m+1-k}}} B(m_1, \dots, m_k, n_1, \dots, n_{m+1-k}) \cdot E[V_I^{m_1}] \dots E[V_I^{m_k}] E[V_o^{n_1}] \dots E[V_o^{n_{m+1-k}}]$$

(122)

by  $X$  and  $Y$  two independent random variables uniformly distributed in  $[0, 1]$ , for almost all  $\mathbf{B}$  we can rewrite (90) as

$$\mathcal{S}_{\mathbf{T}}(z) = E[u(X, z)]. \quad (127)$$

with  $u(x, z)$  satisfying the equation

$$\begin{aligned} u(x, z) &= \frac{1}{1 + z\beta E \left[ \frac{v(x, Y)}{1 + zE[u(X, z)v(X, Y)|Y]} \right]} \\ &= \frac{1}{1 + z\beta E \left[ \frac{\rho(x, Y)}{1 + zE[u(X, z)\rho(X, Y)|Y]} \right]} \end{aligned} \quad (128)$$

where  $v(x, y)$  is the variance profile of  $\mathbf{T}$ , conditioned on  $\mathbf{B}$ , and  $\rho(x, y)$  is defined as in Theorem 5.

The scenario where the  $(\mathbf{B})_{i,j}$ 's are noncompactly supported random variables requires more care. Denoting by  $a$  a finite deterministic constant, let us consider the following truncated random variables:

$$(\tilde{\mathbf{B}})_{i,j} = (\mathbf{B})_{i,j} 1\{ |(\mathbf{B})_{i,j}| < \sqrt{a} \} \quad (129)$$

with  $1\{\cdot\}$  as the indicator function. Furthermore, let  $\hat{\mathbf{T}}$  be defined as  $\hat{\mathbf{T}} = \hat{\mathbf{S}} \circ \tilde{\mathbf{B}}$  with

$$(\hat{\mathbf{S}})_{i,j} = (\mathbf{S})_{i,j} 1\{ |(\mathbf{S})_{i,j}| < \frac{\log n_o}{\sqrt{n_o}} \}.$$

At this point, we invoke the following lemma.

*Lemma 2:* [41] Denoting by  $F_{\hat{\mathbf{T}}}^{n_o}$  the empirical distribution of  $\hat{\mathbf{T}}$ , the variance of  $\int x^k dF_{\hat{\mathbf{T}}}^{n_o}$ , conditioned on  $\mathbf{B}$ , satisfies

$$\text{Var} \left[ \int x^k dF_{\hat{\mathbf{T}}}^{n_o} | \mathbf{B} \right] < \frac{1}{n_o} \left( \mathcal{T}^{4k} + \frac{(\mathcal{T} \log n_o)^{4k}}{n_o} \right) (2k)^{2k+2}$$

with  $\mathcal{T} = \max \{ |(\tilde{\mathbf{B}})_{i,j}| : 1 \leq i \leq n_o, 1 \leq j \leq n_I \}$ . Moreover

$$\int x^k dF_{\hat{\mathbf{T}}}^{n_o} = \mathcal{R}(n_o, m) + \mathcal{R}' + \mathcal{R}'' \quad (130)$$

with

$$\begin{aligned} |\mathcal{R}'| &< \frac{m^m}{n_o} \mathcal{T} \\ E[\mathcal{R}'' | \tilde{\mathbf{B}}] &= 0 \end{aligned}$$

and with

$$\mathcal{R}(n_o, m) = \frac{1}{n_o} \sum (\hat{\mathbf{T}})_{i_1, j_m}^* (\hat{\mathbf{T}})_{i_1, j_1} \cdots (\hat{\mathbf{T}})_{i_m, j_{m-1}}^* (\hat{\mathbf{T}})_{i_m, j_m} \quad (131)$$

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

Denoting by  $\mu_k$  the expected  $k$ th moment of  $\mathbf{T}$  from Lemma 2 we have that

$$\begin{aligned} \lim_{n_o \rightarrow \infty} \frac{1}{n_o} E[\text{Tr}((\mathbf{T}\mathbf{T}^\dagger)^k)] &= \lim_{a \rightarrow \infty} \lim_{n_o \rightarrow \infty} E \left[ \int x^k dF_{\hat{\mathbf{T}}}^{n_o} \right] \\ &= \lim_{a \rightarrow \infty} \lim_{n_o \rightarrow \infty} E[\mathcal{R}(n_o, m)]. \end{aligned} \quad (132)$$

Furthermore, using again Lemma 2 and following the same steps as those given in the proof of [41, Theorem 1.1.4], we finally conclude that

$$\sum_{\ell=1}^{\infty} E \left[ \left( \int x^\ell dF_{\hat{\mathbf{T}}}^{n_o} - E \left[ \int x^\ell dF_{\hat{\mathbf{T}}}^{n_o} \right] \right)^2 \right] < \infty$$

from which it follows that

$$D(F_{\mathbf{T}}, F_{\hat{\mathbf{T}}}) \xrightarrow{\text{a.s.}} 0 \quad (133)$$

where  $D(\cdot, \cdot)$  is the so-called D-metric (or Dudley metric) [7].

With that, we have proved that the empirical distribution of  $\mathbf{T}$  converges a.s. to a nonrandom limit. Again, we can apply Theorem 7 and find that, conditioned on  $\tilde{\mathbf{B}}$ , the empirical distribution of  $\tilde{\mathbf{T}} = \mathbf{S} \circ \tilde{\mathbf{B}}$  converges a.s. to

$$\mathcal{S}_{\tilde{\mathbf{T}}}(z) = E[\tilde{u}(X, z)], \quad (134)$$

with  $\tilde{u}(x, z)$  satisfying the equation

$$\tilde{u}(x, z) = \frac{1}{1 + z\beta E \left[ \frac{\tilde{v}(x, Y)}{1 + zE[\tilde{u}(X, z)\tilde{v}(X, Y)|Y]} \right]} \quad (135)$$

where  $\tilde{v}(x, y)$  denotes the variance profile of  $\tilde{\mathbf{T}}$ , conditioned on  $\tilde{\mathbf{B}}$ , as defined in (13) and (14). Taking the limit as  $a \rightarrow \infty$  we have that, for almost all realizations of  $\mathbf{B}$

$$\begin{aligned} \mathcal{S}_{\mathbf{T}}(z) &= \lim_{a \rightarrow \infty} \mathcal{S}_{\hat{\mathbf{T}}}(z) \\ &= \lim_{a \rightarrow \infty} E[\tilde{u}(X, z)] \\ &= E \left[ \lim_{a \rightarrow \infty} \tilde{u}(X, z) \right] \\ &= E[u(X, z)] \end{aligned} \quad (136)$$

where  $u(x, z)$  is

$$\begin{aligned} u(x, z) &= \frac{1}{1 + z\beta E \left[ \frac{v(x, Y)}{1 + zE[u(X, z)v(X, Y)|Y]} \right]} \\ &= \frac{1}{1 + z\beta E \left[ \frac{\rho(x, Y)}{1 + zE[u(X, z)\rho(X, Y)|Y]} \right]}. \end{aligned} \quad (137)$$

This, in conjunction with Theorems 1 and 2, concludes the proof.

## APPENDIX E

### PROOF OF THEOREM 6

Since  $\mathbf{A}_j$  for  $j = 1, \dots, n_I$  are either *finite order* or *infinite order absolutely summable* Toeplitz matrices, each of them is asymptotically equivalent to a circulant matrix, which we denote by  $\mathbf{C}_j$  [12, Lemma 4.6]. Consequently, it follows from [43, eq. 3.4] that as  $n_o \rightarrow \infty$  the Dudley metric

$$D_2(F_{\mathbf{T}} - F_{\mathbf{T}_{\text{circ}}}) \xrightarrow{\text{a.s.}} 0 \quad (138)$$

with

$$\mathbf{T}_{\text{circ}} = [\mathbf{C}_1(\mathbf{S})_1, \dots, \mathbf{C}_{n_I}(\mathbf{S})_{n_I}] \bar{\mathbf{A}}. \quad (139)$$

Write  $\mathbf{C}_j$  in its spectral decomposition

$$\mathbf{C}_j^\dagger \mathbf{C}_j = \mathbf{U}_j \text{diag}((\boldsymbol{\Lambda})_{1,j}, \dots, (\boldsymbol{\Lambda})_{n_o,j}) \mathbf{U}_j^\dagger.$$

It is immediate that if  $\mathbf{S}$  is a complex zero-mean Gaussian matrix with i.i.d. entries, then the asymptotic empirical distribution of  $\mathbf{T}_{\text{circ}}$  equals the asymptotic empirical distribution of

$$\tilde{\mathbf{T}}_{\text{circ}} = [\sqrt{\text{diag}((\boldsymbol{\Lambda})_1)}(\tilde{\mathbf{S}})_1, \dots, \sqrt{\text{diag}((\boldsymbol{\Lambda})_{n_I})}(\tilde{\mathbf{S}})_{n_I}] \bar{\mathbf{A}}$$

where the  $n_o$ -dimensional vector  $\tilde{\mathbf{S}}_j = \mathbf{U}_j^\dagger(\mathbf{S})_j$  is still complex zero-mean Gaussian with i.i.d. entries. From Theorem 5 it follows that the moments of  $\tilde{\mathbf{T}}_{\text{circ}}$  and, consequently, of  $\mathbf{T}_{\text{circ}}$  depend on  $\rho(x, y)$  which is a function of the asymptotic empirical distribution of  $\mathbf{C}_j$  that equals the (nonrandom) limiting distribution of the squared singular values of  $\mathbf{A}_j$  [12, Theorem 4.2]. Theorem 6 then follows immediately if  $\mathbf{S}$  is Gaussian.

$$E[\mu_k^{(n_o)}] = \frac{1}{n_o} \sum_{j_1, \dots, j_m=1}^{n_1} \sum_{i_1, \dots, i_m=1}^{n_o} \sum_{d_1, \dots, d_m=1}^{n_o} \sum_{\ell_1, \dots, \ell_m=1}^{n_o} E \left[ (\mathbf{U})_{i_1, d_1}^* \hat{\mathbf{S}}_{d_1, j_m}^* (\mathbf{U})_{i_1, \ell_1} \hat{\mathbf{S}}_{\ell_1, j_1} \cdots (\mathbf{U})_{i_m, d_m}^* \cdot \mathbf{S}_{d_m, j_{m-1}}^* (\mathbf{U})_{i_m, \ell_m} \hat{\mathbf{S}}_{\ell_m, j_m} \right] \sqrt{(\boldsymbol{\Lambda})_{i_1, j_m}} \sqrt{(\boldsymbol{\Lambda})_{i_1, j_1}} \cdots \sqrt{(\boldsymbol{\Lambda})_{i_m, j_{m-1}}} \sqrt{(\boldsymbol{\Lambda})_{i_m, j_m}} P_{j_1} \cdots P_{j_m} \quad (148)$$

In the more general case of non-Gaussian  $\mathbf{S}$  the same arguments given in [31, Sec. 3] and [43] allow us to replace the assumptions on the entries of  $\mathbf{S}$  by standardized (i.e., zero-mean and unit variance) variables bounded in absolute value by a fixed multiple of  $\ln n_o/n_o$ . Furthermore, again the same arguments applied in [31] and [43] allow us, also, to replace  $\bar{\mathbf{A}}$  with the diagonal matrix  $\bar{\mathbf{A}}_\alpha$  whose diagonal coefficients are

$$(\bar{\mathbf{A}}_\alpha)_j = 1 \{ |(\bar{\mathbf{A}})_j| < \alpha \}$$

with  $\alpha = \ln n_o$ , from which  $D_2(F_{\mathbf{T}_{\text{circ}}} - F_{\hat{\mathbf{T}}_{\text{circ}}}) \xrightarrow{\text{a.s.}} 0$  follows, where  $F_{\hat{\mathbf{T}}_{\text{circ}}}$  is the asymptotic empirical distribution of

$$\hat{\mathbf{T}}_{\text{circ}} = [\mathbf{C}_1(\hat{\mathbf{S}})_1, \dots, \mathbf{C}_{n_1}(\hat{\mathbf{S}})_{n_1}] \bar{\mathbf{A}}_\alpha$$

with

$$(\hat{\mathbf{S}})_{i,j} = (\mathbf{S})_{i,j} 1 \left\{ |(\mathbf{S})_{i,j}| < \frac{\ln n_o}{\sqrt{n_o}} \right\} - E \left[ (\mathbf{S})_{i,j} 1 \left\{ |(\mathbf{S})_{i,j}| < \frac{\ln n_o}{\sqrt{n_o}} \right\} \right] \quad (140)$$

and  $F_{\mathbf{T}_{\text{circ}}}$  is the asymptotic empirical distribution of  $\mathbf{T}_{\text{circ}}$  as defined in (139). Therefore, to prove that the empirical distribution of  $\mathbf{T}$  converges a.s. to a nonrandom limiting distribution, we need only to prove that  $\hat{\mathbf{T}}_{\text{circ}}$  has an empirical distribution converging a.s. to a nonrandom limit. From (140), we have

$$|(\hat{\mathbf{S}})_{i,j}| < \frac{\ln n_o}{\sqrt{n_o}}, \quad E[(\hat{\mathbf{S}})_{i,j}] = 0.$$

We first prove that the  $m$ th expected moment of  $\hat{\mathbf{T}}_{\text{circ}}$  admits expression (38) before we prove almost sure convergence.

For convenience of notation, in the following we denote

$$(\tilde{\mathbf{S}})_j = \mathbf{C}_j(\hat{\mathbf{S}})_j.$$

Consequently, the  $m$ th expected asymptotic moment  $\hat{\mathbf{T}}_{\text{circ}} \hat{\mathbf{T}}_{\text{circ}}^\dagger$  is as in (141)–(146) at the bottom of the page, where in these equations  $P_j = |(\bar{\mathbf{A}}_\alpha)_j|^2$  and  $\tilde{W}_{i,j}$  is the  $i$ th entry of  $(\tilde{\mathbf{S}})_j$ . Moreover, denoting by  $\boldsymbol{\Lambda}$  the  $n_o \times n_1$  matrix whose  $(i, j)$ th entry,  $(\boldsymbol{\Lambda})_{i,j}$  represents the  $i$ th eigenvalue of  $\mathbf{C}_j \mathbf{C}_j^\dagger$ , we have that

$$\tilde{W}_{i,j} = \sqrt{(\boldsymbol{\Lambda})_{i,j}} W_{i,j}$$

with  $W_{i,j}$  the  $i$ th entry of  $\mathbf{U}_j^\dagger(\hat{\mathbf{S}})_j$ .

Let us now investigate all possible ways to partition indexes  $i_1, \dots, i_m$  into  $k$  disjoint sets for  $1 \leq k \leq m$ . Indexes in the same set remain identical in the summation and indexes in different sets remain distinct in the summation. More precisely, we decompose the summation in (146) into three terms

$$\mu_m^{(n_o)} = \mathcal{R}(n_o, m) + \mathcal{R}' + \mathcal{R}'' \quad (147)$$

with

$$\mathcal{R}(n_o, m) = \frac{1}{n_o} \sum W_{i_1, j_m}^* W_{i_1, j_1} \cdots W_{i_m, j_{m-1}}^* W_{i_m, j_m} \cdot \sqrt{(\boldsymbol{\Lambda})_{i_1, j_m}} \sqrt{(\boldsymbol{\Lambda})_{i_1, j_1}} \cdots \sqrt{(\boldsymbol{\Lambda})_{i_m, j_{m-1}}} \sqrt{(\boldsymbol{\Lambda})_{i_m, j_m}} P_{j_1} \cdots P_{j_m}$$

where the summation is as in (131). Since  $W_{i,j}$  is the  $i$ th entry of  $\mathbf{U}_j(\hat{\mathbf{S}})_j$ , (148) follows (see the top of the page). Consequently,

$$E[\mu_m^{(n_o)}] = \frac{1}{n_o} E \left[ \text{Tr} \left\{ (\hat{\mathbf{T}}_{\text{circ}} \hat{\mathbf{T}}_{\text{circ}}^\dagger)^m \right\} \right] \quad (141)$$

$$= \frac{1}{n_o} E \left[ \text{Tr} \left\{ (\tilde{\mathbf{S}} \bar{\mathbf{A}}_\alpha \bar{\mathbf{A}}_\alpha^\dagger \tilde{\mathbf{S}}^\dagger)^m \right\} \right] \\ = \frac{1}{n_o} E \left[ \text{Tr} \left\{ \sum_{j_1=1}^{n_1} P_{j_1} (\tilde{\mathbf{S}})_{j_1} (\tilde{\mathbf{S}})_{j_1}^\dagger \cdots \sum_{j_m=1}^{n_1} P_{j_m} (\tilde{\mathbf{S}})_{j_m} (\tilde{\mathbf{S}})_{j_m}^\dagger \right\} \right] \quad (142)$$

$$= \frac{1}{n_o} E \left[ \sum_{j_1, \dots, j_m=1}^{n_1} (\tilde{\mathbf{S}})_{j_m}^\dagger (\tilde{\mathbf{S}})_{j_1} \cdots (\tilde{\mathbf{S}})_{j_{m-1}}^\dagger (\tilde{\mathbf{S}})_{j_m} P_{j_1} \cdots P_{j_m} \right] \quad (143)$$

$$= \frac{1}{n_o} E \left[ \sum_{j_1, \dots, j_m=1}^{n_1} \sum_{i_1, \dots, i_m=1}^{n_o} \tilde{W}_{i_1, j_m}^* \tilde{W}_{i_1, j_1} \cdots \tilde{W}_{i_m, j_{m-1}}^* \tilde{W}_{i_m, j_m} P_{j_1} \cdots P_{j_m} \right] \quad (144)$$

$$= \frac{1}{n_o} \sum_{j_1, \dots, j_m=1}^{n_1} \sum_{i_1, \dots, i_m=1}^{n_o} E \left[ \sqrt{(\boldsymbol{\Lambda})_{i_1, j_m}} W_{i_1, j_m}^* \sqrt{(\boldsymbol{\Lambda})_{i_1, j_1}} W_{i_1, j_1} \cdots \sqrt{(\boldsymbol{\Lambda})_{i_m, j_{m-1}}} W_{i_m, j_{m-1}}^* \sqrt{(\boldsymbol{\Lambda})_{i_m, j_m}} W_{i_m, j_m} \right] \cdot P_{j_1} \cdots P_{j_m} \quad (145)$$

$$= \frac{1}{n_o} \sum_{j_1, \dots, j_m=1}^{n_1} \sum_{i_1, \dots, i_m=1}^{n_o} E \left[ W_{i_1, j_m}^* W_{i_1, j_1} \cdots W_{i_m, j_{m-1}}^* W_{i_m, j_m} \right] \cdot \sqrt{(\boldsymbol{\Lambda})_{i_1, j_m}} \sqrt{(\boldsymbol{\Lambda})_{i_1, j_1}} \cdots \sqrt{(\boldsymbol{\Lambda})_{i_m, j_{m-1}}} \sqrt{(\boldsymbol{\Lambda})_{i_m, j_m}} P_{j_1} \cdots P_{j_m}, \quad (146)$$

using (147) and the fact that  $\hat{\mathbf{S}}_{d_\ell, j_p}$ 's are independent and zero-mean, it is easy to check by the following arguments analogous to the ones in [41, Lemma 1.2] that

$$|\mathcal{R}'| < \frac{m^m}{n_o} \sup\{|\langle \hat{\mathbf{T}}_{\text{circ}} \rangle_{i,j}|^{2m} : i \leq n_1, j \leq n_o\}$$

$$E[\mathcal{R}''] = 0.$$

The assumption on  $\mathbf{A}_j$  of either *finite order* or *infinite order absolutely summable* Toeplitz, implies that the eigenvalues of  $\mathbf{A}_j$  (and, consequently, of  $\mathbf{C}_j$ ) are bounded. From this, thanks to the truncation of  $(\hat{\mathbf{S}})_{i,j}$  and  $\hat{\mathbf{A}}$ , and thanks to the structure of  $\mathbf{U}_j$ , it follows that  $\lim_{n_o \rightarrow \infty} |\mathcal{R}'| \rightarrow 0$ . Therefore,

$$\begin{aligned} \lim_{n_o \rightarrow \infty} E[\mu_m^{(n_o)}] &= \lim_{n_o \rightarrow \infty} E[\mathcal{R}(n_o, m)] \\ &= \lim_{n_o \rightarrow \infty} E[\mu_m^{(n_o, G)}] \\ &= \mu_m \end{aligned} \quad (149)$$

with  $\mu_m$  defined as in Theorem 6 and  $\lim_{n_o \rightarrow \infty} E[\mu_m^{(n_o, G)}]$  denoting the  $m$ th expected moment of  $\hat{\mathbf{T}}_{\text{circ}} \hat{\mathbf{T}}_{\text{circ}}^\dagger$ . From (149) it clearly follows that, since for the case that  $\mathbf{S}$  is Gaussian the limiting distribution of  $\mathbf{T}$  exists and its moments satisfy the Carleman sufficiency condition, the same considerations hold for  $\mathbf{T}$  when the matrix  $\mathbf{S}$  is a matrix with i.i.d. entries arbitrarily distributed.

Finally, in order to prove almost sure convergence

$$\lim_{n_o \rightarrow \infty} \frac{1}{n_o} \text{Tr}(\mathbf{T} \mathbf{T}^\dagger)^m \xrightarrow{\text{a.s.}} \lim_{n_o \rightarrow \infty} E[\mu_m^{(n_o)}] = \mu_m$$

for each  $m$  as  $n_o \rightarrow \infty$ , we just need

$$\sum_{m=1}^{\infty} E[(\lim_{n_o \rightarrow \infty} \mu_m^{(n_o)} - \lim_{n_o \rightarrow \infty} E[\mu_m^{(n_o)}])^2] < \infty$$

which can be established following the same steps as those given in [43].

Summarizing, we have proved that the transform  $\mathcal{S}(z)$  of the nonrandom limiting distribution (of the squared singular values) of  $\mathbf{T}$  is given by (89) with  $u(x, z)$  satisfying

$$u(x, z) = \frac{1}{1 + z\beta \int_0^1 \frac{\rho(x, y) dy}{1 + z \int_0^1 u(w, z) \rho(w, y) dw}} \quad (150)$$

from which, following the same steps adopted in Theorem 7, the expression of  $\delta_{n,k}$  given in (39) follows immediately.

#### APPENDIX F PROOF OF COROLLARY 4

Applying Theorem 7, the transform  $\mathcal{S}_{\mathbf{T}}(z)$  of  $\mathbf{T}$  is

$$\begin{aligned} \mathcal{S}_{\mathbf{T}}(z) &= \lim_{N \rightarrow \infty} \frac{1}{N} E[\text{Tr}(\mathbf{I} + z\mathbf{T}\mathbf{T}^\dagger)^{-1}] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N E[u(n, z)] \end{aligned} \quad (151)$$

where

$$u(n, z) = \frac{1}{1 + z\beta \sum_{k=1}^K \frac{A_k^2 \lambda_n(\mathbf{C}_k)}{1 + z A_k^2 \frac{1}{N} \sum_{n=1}^N u(n, z) \lambda_n(\mathbf{C}_k)}} \quad (152)$$

with

$$\lambda_{n-1}(\mathbf{C}_k) = \left| \sum_{\ell=1}^L H_{k,\ell} e^{i2\pi \ell n/N} \right|^2$$

independent across the users. Since the asymptotic empirical distribution of  $\{\lambda_i(\mathbf{C}_1), \dots, \lambda_i(\mathbf{C}_K)\}$  does not depend on  $i$ , using Parseval equality we have that

$$\mathcal{S}_{\mathbf{T}}(z) = \lim_{K \rightarrow \infty} E \left[ \frac{1}{1 + z\beta \sum_{k=1}^K \frac{A_k^2 \sum_{\ell=1}^L |H_{k,\ell}|^2}{1 + z\mathcal{S}_{\mathbf{T}}(z) A_k^2 \sum_{\ell=1}^L |H_{k,\ell}|^2}} \right] \quad (153)$$

from which, invoking the fact that across users the fading coefficients are i.i.d., it follows that the right side of (153) converges to

$$\mathcal{S}_{\mathbf{T}}(z) = \frac{1}{1 + z\beta E \left[ \frac{P}{1 + z P \mathcal{S}_{\mathbf{T}}(z)} \right]} \quad (154)$$

with  $P$  a random variable whose distribution equals the non-random limiting distribution of

$$A_1^2 \sum_{\ell=1}^L |H_{1,\ell}|^2, \dots, A_K^2 \sum_{\ell=1}^L |H_{K,\ell}|^2.$$

#### ACKNOWLEDGMENT

The authors gratefully acknowledge suggestions by Prof. Philippe Biane, École Normale Supérieure, Paris, France; Prof. Roland Speicher, Queen's University, Kingston, ON, Canada; and Dr. Angel Lozano, Bell Laboratories.

#### REFERENCES

- [1] Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *Ann. Probab.*, vol. 26, pp. 316–345, 1998.
- [2] P. Biane, "Minimal factorizations of cycle and central multiplicative functions on the infinite symmetric group," *J. Combin. Theory*, ser. A 76, no. 2, pp. 197–212, 1996.
- [3] J. M. Chaufray, W. Hachem, and P. Loubaton, "Asymptotic analysis of optimum and sub-optimum CDMA downlink MMSE receivers," *IEEE Trans. Inform. Theory*, submitted for publication.
- [4] —, "Optimum and sub-optimum CDMA downlink MMSE receivers," in *Proc. 2002 IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, June/July 2002.
- [5] L. Cottarelli and R. R. Muller, "Asymptotic design and analysis of multistage detectors with unequal powers," in *Proc. 2002 IEEE Information Theory Workshop*, Bangalore, India, Oct. 2002.
- [6] M. Debbah, W. Hachem, P. Loubaton, and M. de Courville, "MMSE analysis of certain large isometric random precoded systems," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1293–1311, May 2003.
- [7] R. M. Dudley, "Convergence of Baire measures," *Studia Math.*, vol. 27, pp. 251–268, 1966.
- [8] J. S. Evans and D. N. C. Tse, "Linear multiuser receivers for multipath fading channels," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2059–2078, Sept. 2000.
- [9] V. L. Girko, *Theory of Random Determinants*. Boston, MA: Kluwer Academic, 1990.
- [10] J. S. Goldstein and I. S. Reed, "A new method of Wiener filtering and its application to interference mitigation for communications," in *Proc. IEEE MILCOM*, vol. 3, Monterey, CA, Nov. 1997, pp. 1087–1091.
- [11] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2943–2959, Nov. 1998.

- [12] R. M. Gray, "Toeplitz and circulant matrices: A Review," Stanford Univ., Stanford, CA, Tech. Rep., Aug. 2002.
- [13] A. J. Grant and P. D. Alexander, "Random sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2832–2836, Nov. 1998.
- [14] A. Guionnet and O. Zeitouni, "Concentration of the spectral measure for large matrices," *Electron. Commun. in Probability*, vol. 5, pp. 119–136, 2000.
- [15] W. C. Jakes, *Microwave Mobile Communications*. New York: IEEE Press, 1974.
- [16] W. Hachem, "Low complexity polynomial receivers for downlink CDMA," in *Proc. 36th Asilomar Conf. Systems, Signals and Computers*, Nov. 2002.
- [17] —, "Simple polynomial MMSE receiver for CDMA transmission on frequency selective channels," *IEEE Trans. Inform. Theory*, submitted for publication.
- [18] S. V. Hanly and D. N. C. Tse, "Resource pooling and effective bandwidths in CDMA networks with multiuser receivers and spatial diversity," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1328–1351, May 2001.
- [19] S. Hara and R. Prasad, "Overview of multicarrier CDMA," *IEEE Commun. Mag.*, vol. 35, pp. 126–133, Dec. 1997.
- [20] F. Hiai and D. Petz, "Asymptotic freeness almost everywhere for random matrices," *Acta Sci. Math. Szeged*, vol. 66, pp. 801–826, 2000.
- [21] M. L. Honig and W. Xiao, "Performance of reduced-rank linear interference suppression for DS-CDMA," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1928–1946, July 2001.
- [22] R. Horn and C. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1985.
- [23] C. Peel, B. Hochwald, and A. L. Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multi-Antenna Multi-User Communication," preprint.
- [24] G. Kreweeras, "Sur les partitions noncroisees d'un cycle," *Discr. Math.*, vol. 1, pp. 333–350, 1972.
- [25] L. Li, A. M. Tulino, and S. Verdú, "Asymptotic eigenvalue moments for linear multiuser detection," *Commun. Inform. Syst.*, vol. 1, no. 3, pp. 273–304, Sept. 2001.
- [26] P. Loubaton and W. Hachem, "Asymptotic analysis of reduced rank Wiener filter," in *Proc. 2003 IEEE Information Theory Workshop*, Paris, France, 2003, pp. 328–332.
- [27] V. A. Marčenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," in *Math USSR—Sbornik*, 1967, vol. 1, pp. 457–483.
- [28] S. Moshavi, E. G. Kanterakis, and D. L. Schilling, "Multistage linear receivers for DS-CDMA systems," *Int. J. Wireless Inform. Networks*, vol. 39, no. 1, pp. 1–17, 1996.
- [29] R. Mueller and S. Verdú, "Design and analysis of low-complexity interference mitigation on vector channels," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1429–1441, Aug. 2001.
- [30] D. Shlyankhtenko, "Random Gaussian band matrices and freeness with amalgamation," *Int. Math. Res. Note*, vol. 20, pp. 1013–1025, 1996.
- [31] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *J. Multivariate Anal.*, vol. 54, no. 2, pp. 175–192, 1995.
- [32] R. Speicher, "Free probability theory and non-crossing partitions," in *39e Seminaire Lotharingien de Combinatoire*, Apr. 1997.
- [33] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inform. Theory*, vol. 45, pp. 641–657, Mar. 1999.
- [34] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of Correlation on the Capacity of Multi-Antenna Channels," Bell Labs., Tech. Memo. ITD-03-44786F, 2003.
- [35] A. M. Tulino and S. Verdú, "Asymptotic analysis of improved linear receivers for BPSK-CDMA subject to fading," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1544–1555, Aug. 2001.
- [36] —, "Random matrix theory and wireless communications," *Foundation and Trends in Communications and Information Theory*, to be published.
- [37] S. Verdú, "Capacity region of Gaussian CDMA channels: The symbol-synchronous case," in *Proc. 24th Allerton Conf. Communications, Control and Computing*, Oct. 1986, pp. 1025–1034.
- [38] —, *Multiuser Detection*. New York: Cambridge Univ. Press, 1998.
- [39] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 622–640, Mar. 1999.
- [40] D. V. Voiculescu, K. J. Dykema, and A. Nica, "Free random variables: A noncommutative probability approach to free products with applications to random matrices, operator algebras, and harmonic analysis on free groups," in *CRM Monograph Series 1*. Providence, RI: Amer. Math. Soc., 1992.
- [41] K. W. Wachter, "The strong limits of random matrix spectra for sample matrices of independent elements," *Ann. Probab.*, vol. 6, no. 1, pp. 1–18, 1978.
- [42] W. Xiao and M. L. Honig, "Large system convergence analysis of adaptive reduced- and full-rank least squares algorithms," *IEEE Trans. Inform. Theory*, submitted for publication.
- [43] Y. Q. Yin, "Limiting spectral distribution for a class of random matrices," *J. Multivariate Anal.*, vol. 20, pp. 50–68, 1986.