

# Unified Large-System Analysis of MMSE and Adaptive Least Squares Receivers for a class of Random Matrix Channels

Matthew J. M. Peacock, *Member, IEEE*, Iain B. Collings, *Senior Member, IEEE*  
and Michael L. Honig, *Fellow, IEEE*

**Abstract**— We present a unified large-system analysis of linear receivers for a class of random matrix channels. The technique unifies the analysis of both the minimum-mean-squared-error (MMSE) receiver and the adaptive least-squares (ALS) receiver, and also uses a common approach for both random i.i.d. and random orthogonal precoding. We derive expressions for the asymptotic signal-to-interference-plus-noise (SINR) of the MMSE receiver, and both the transient and steady-state SINR of the ALS receiver, trained using either i.i.d. data sequences or orthogonal training sequences. The results are in terms of key system parameters, and allow for arbitrary distributions of the power of each of the data streams and the eigenvalues of the channel correlation matrix. In the case of the ALS receiver, we allow a diagonal loading constant and an arbitrary data windowing function. For i.i.d. training sequences and no diagonal loading, we give a fundamental relationship between the transient/steady-state SINR of the ALS and the MMSE receivers. We demonstrate that for a particular ratio of receive to transmit dimensions and window shape, all channels which have the same MMSE SINR have an identical transient ALS SINR response. We demonstrate several applications of the results, including an optimization of information throughput with respect to training sequence length in coded block transmission.

**Index Terms**— Large-System, MMSE, Recursive Least Squares, MIMO, CDMA.

## I. INTRODUCTION

**L**ARGE-system analysis of linear receivers for random matrix channels has attracted significant attention in recent years, and has proven to be a powerful tool in their understanding and design (e.g., see [1] and references therein). In particular, large-system analysis of the matched filter, decorrelator, and minimum-mean-squared-error (MMSE) receivers, which have knowledge of the channel state information, has been exhaustively studied for the downlink, using results such

Manuscript received Feb. 2, 2005; revised Aug. 26, 2005. This work was supported in part by the U.S. Army Research Office under DAAD19-03-1-0119, the National Science Foundation under grant CCR-0310809, and CSIRO Telecommunications and Industrial Physics, Australia. This paper was presented in part at IEEE ISIT, Adelaide SA, Australia, Sep. 2005.

M.J.M. Peacock and I.B. Collings are with the Telecommunications Laboratory, School of Electrical and Information Engineering, University of Sydney, NSW 2006, Australia. (e-mail: {mpeac, iain}@ee.usyd.edu.au).

I.B. Collings is with the CSIRO ICT Centre, Sydney, NSW 1710, Australia. (e-mail: iain.collings@csiro.au).

M.L. Honig is with the Department of Electrical & Computer Engineering, Northwestern University, Evanston, IL 60208, USA. (e-mail: mh@ece.northwestern.edu).

Communicated by R. Müller, Associate Editor for Communications Theory. Digital Object Identifier 00.0000/TIT.2006.000000

as the Silverstein-Bai theorem [2], Girko's law [3], and free probability [4]. In this paper we take a different approach to the problem, which allows us to consider random i.i.d. and orthogonal channels (or matrix of signatures) in the same treatment, in contrast to the existing separate analyses of i.i.d. [5] and orthogonally precoded [6, 7] channels. The new approach also allows us to consider a more general class of signal models, and a receiver which does not have the benefit of channel state information, namely the adaptive least-squares (ALS) receiver.

In this paper, we consider both the MMSE and ALS receivers. The ALS receiver approximates the MMSE receiver and requires training symbols [8]. In particular, the autocorrelation matrix of the received vector (an ensemble average), which is used in the MMSE receiver, is replaced in the ALS receiver by a sample autocorrelation matrix (a time average). Given sufficient training symbols, the performance of the ALS receiver approaches that of the MMSE receiver. In this paper, two types of adaptive training modes are considered, based on those presented in [9], where either the training sequence is known at the receiver, or a semi-blind method is employed. In a time-varying environment, weighting can be applied to the errors to create data windowing, which allows for tracking. When implemented online as a series of rank-1 updates with exponential windowing, this receiver is often referred to as the recursive least-squares (RLS) receiver [8]. To prevent ill-conditioning of the sample autocorrelation matrix with RLS filtering, diagonal loading can be employed, which refers to initializing this matrix with a small positive constant times the identity matrix.

Prior relevant work on ALS techniques (in particular, as applied to channel equalization and estimation) includes [10], where an approximate expression is derived for the transient excess mean-squared-error (MSE) of the ALS receiver with respect to the MMSE receiver, often referred to as self-noise, for a general channel model and windowing function. Also, an approximate expression is given for the convergence time constant of the receiver when tracking non-stationary signals. Steady-state and transient analysis of the RLS receiver was also considered in [11]. A comprehensive treatment of ALS techniques and variations is contained in [8].

The application of ALS to CDMA was considered in [12], where the convergence of a blind multiuser detector based on a stochastic gradient-descent adaptation rule is also ana-

lyzed. In [9], approximate expressions are given for the relationship between the MMSE signal-to-interference-plus-noise-ratio (SINR) and the steady-state ALS SINR with exponential windowing for DS-CDMA in flat-fading. This was extended in [13] to a steady-state analysis of two-stage algorithms based on RLS using decision-directed adaptation. A review of adaptive interference mitigation techniques is given in [14].

Large-system analysis for ALS receivers was first considered in [15, 16], which considered DS-CDMA in flat fading with i.i.d. training sequences, and an ALS receiver with diagonal loading and rectangular or exponential weighting. The transient SINR (i.e., after a given number of training symbols) and steady-state SINR (i.e., after an unlimited number of training symbols) of the ALS receiver was derived in the limit where the number of transmit dimensions, the number of multiplexed data streams, and the number of training symbols all tend to infinity with fixed ratios. With rectangular windowing and no diagonal loading, a relationship between the MMSE SINR and the transient ALS SINR was given. In other work, large-system analysis has been applied to so-called subspace-based blind ALS receivers in [17–19].

In this paper, we consider a more general complex-valued AWGN matrix-vector channel model of the form  $\mathbf{r} = \mathbf{H}\mathbf{S}\mathbf{A}\mathbf{b} + \mathbf{n}$ , where  $\mathbf{H}$  is an arbitrary matrix,  $\mathbf{A}$  is an arbitrary diagonal matrix, and  $\mathbf{S}$  is either an i.i.d. or orthogonal matrix. For example, the model applies explicitly to downlink synchronous direct sequence (DS) or multi-carrier (MC) CDMA in frequency-selective fading, as well as multi-input multi-output (MIMO) channels. In fact, if we additionally require that the data vector  $\mathbf{b}$  is unitarily invariant, then the MMSE and ALS results we obtain *apply to all AWGN matrix-vector systems* of the form  $\mathbf{r} = \mathbf{C}\mathbf{b} + \mathbf{n}$ , provided that the eigenvalues of  $\mathbf{C}\mathbf{C}^\dagger$  converge to a well-behaved deterministic distribution. As such, the results can be applied to systems not previously considered in large-system analysis, such as equalization of single-user finite-impulse-response (FIR) channels.

For this channel model, we derive the SINR of the MMSE receiver, with either i.i.d. or orthogonal  $\mathbf{S}$ , in the large-system limit where the number of transmit and receive dimensions, and the number of multiplexed data streams all tend to infinity with fixed ratios. The expression for the SINR is a function of these ratios and the received SNR, and allows for arbitrary asymptotic eigenvalue distributions (a.e.d.'s) of  $\mathbf{A}\mathbf{A}^\dagger$  and  $\mathbf{H}\mathbf{H}^\dagger$ . This result can also be derived under the same set of assumptions using the  $\mathcal{S}$ -transform from free probability. However, unlike the free probability technique, we show that the technique used to derive this result also applies to the ALS receiver.

For the ALS receiver, we extend the work of [16] to the general channel model described. We consider an arbitrary data windowing function and both i.i.d. and orthogonal training sequences. That is, we determine both the transient and steady-state ALS SINR in the limit described for the MMSE receiver, and also as the number of training symbols tend to infinity.

Also, we present an expression which relates the SINR of the MMSE receiver to the transient and steady-state SINRs of the ALS receiver for the case of i.i.d. training sequences and

no diagonal loading. We demonstrate that in this situation, for a particular ratio of receive to transmit dimensions and window shape, *all channels, which have the same MMSE SINR, have an identical transient ALS SINR response*. Since our results hold for all well-behaved matrix-vector systems for which the data vector is unitarily invariant (as previously discussed), the MMSE-ALS relationship is in fact seen to be a *fundamental property* of adaptive least-squares estimation.

It is interesting to compare our results to an approximate expression for the steady-state ALS SINR given in [9] for the special case of DS-CDMA in a flat-fading channel, with i.i.d. spreading and exponential weighting. A comparison of the expressions reveals the approximation in [9] to be excellent, particularly for large window sizes. Also, we note that for a general channel model, the study in [11] previously came to the conclusion that the ALS convergence rate is independent of the channel; however, their conclusion was based on making several approximations, and only exponential windowing (RLS) was considered. This conclusion was stated for DS-CDMA in [9]; however, an explicit relationship between the MMSE SINR and the transient ALS SINR, such as derived in this paper, is not given in either of those papers.

Unfortunately, we have not determined a simple relationship between the MMSE SINR and ALS SINR with orthogonal training sequences and/or diagonal loading. This remains an open problem. Inspection reveals that with orthogonal training sequences and no diagonal loading, the transient relationship will depend on the channel.

During the course of the analysis we solve for the Stieltjés (or Cauchy) transforms of the a.e.d.'s of both the autocorrelation matrix and the weighted sample autocorrelation matrix of the received signal. These expressions are found using matrix manipulations, which do not require the use of free probability. The first transform can also be derived using free probability; however, such techniques cannot be applied to derive the second transform, since the associated constituent matrices are not free. As such, the results are of independent mathematical interest.

Through numerical studies, we demonstrate the applicability of the large-system results to finite systems, and the benefits of orthogonal precoding and training is examined. We demonstrate an application of the results in the optimization of information throughput with respect to training sequence length in coded block transmission.

The paper is arranged as follows. Section II outlines the general transmission model, the receivers considered, and defines the large-system limit. Section III discusses the general approach we take for the analysis. Section IV reviews analytical approaches to the MMSE SINR problem, and presents the general solution based on the unified analytical approach. An alternate expression for the MMSE SINR, different from those presented in [1], is also presented which allows the relationship between the MMSE SINR and ALS SINR to be derived in certain cases in later sections. Section V presents the general result for the ALS SINR. A simple relationship between the ALS SINR and MMSE SINR with i.i.d. training and without diagonal loading is then presented in Section VI. Finally, numerical studies are presented in Section VII.

## II. SYSTEM MODEL

This paper considers a general matrix-vector transmission model. It applies to a wide range of practical data communication systems, including frequency-flat fading MIMO channels, and both frequency-flat and frequency-selective fading downlink DS- or MC-CDMA channels.

### A. General Transmission Model

In matrix notation<sup>1</sup>, the received signal in the  $m^{\text{th}}$  symbol period is

$$\mathbf{r}_m = \mathbf{H}\mathbf{S}\mathbf{A}\mathbf{b}_m + \mathbf{n}_m \quad (1)$$

where

- $\mathbf{H} \in \mathbb{C}^{M \times N}$  is a complex-valued channel matrix.
- $\mathbf{S} = [\mathbf{s}_1 \cdots \mathbf{s}_K] \in \mathbb{C}^{N \times K}$  contains either
  - random orthonormal columns, i.e., we assume that  $\mathbf{S}$  is obtained by extracting  $K \leq N$  columns from an  $N \times N$  Haar-distributed<sup>2</sup> unitary random matrix, or,
  - i.i.d. complex elements<sup>3</sup> with mean zero and variance  $\frac{1}{N}$ . For example, i.i.d. real & imaginary parts which are either  $\pm 1/\sqrt{2N}$  with equal probability, or i.i.d. Gaussian with zero mean and variance  $1/(2N)$ .
- We shall call the first case ‘isometric (iso.)  $\mathbf{S}$ ’, and the second case ‘i.i.d.  $\mathbf{S}$ ’, as is done in [6].
- $\mathbf{A} \in \mathbb{C}^{K \times K}$  is a diagonal matrix of transmit coefficients, i.e.  $\mathbf{A} = \text{diag}(A_1, \dots, A_K)$ . In fact, the results which follow depend only the values of  $P_k = |A_k|^2$ , and so to simplify notation, without lack of generality, we may assume  $A_k$ ,  $k = 1, \dots, K$ , is non-negative and real valued.
- $\mathbf{b}_m \in \mathbb{C}^{K \times 1}$  either contains transmit data, or training symbols (for an ALS receiver). Elements of  $\mathbf{b}_m$  can be either i.i.d. with zero mean and unit variance<sup>3</sup> (for data and i.i.d. training cases), or they can be drawn from a set of orthogonal sequences (for orthogonal training, as explained further in Section II-D).
- $\mathbf{n}_m \in \mathbb{C}^{M \times 1}$  contains i.i.d., zero mean, circularly symmetric, complex Gaussian entries with variance per dimension  $\sigma^2/2$ .
- $\mathbf{H}$ ,  $\mathbf{S}$ ,  $\mathbf{A}$ ,  $\mathbf{b}_m$ , and  $\mathbf{n}_m$  are mutually independent.

### B. Discussion

The general transmission model in (1) is widely applicable, and in particular includes the following systems.

<sup>1</sup>**Notation:** All vectors are defined as column vectors and designated with bold lower case; all matrices are given in bold upper case;  $(\cdot)^T$  denotes transpose;  $(\cdot)^\dagger$  denotes Hermitian (i.e. complex conjugate) transpose;  $(\cdot)^\ddagger$  denotes the operation  $\mathbf{X}^\ddagger = \mathbf{X}\mathbf{X}^\dagger$ ;  $\text{tr}[\cdot]$  denotes the matrix trace;  $|\cdot|$  and  $\|\cdot\|$  denote the Euclidian and induced spectral norms, respectively;  $\mathbf{I}_N$  denotes the  $N \times N$  identity matrix;  $\mathbb{R}^* = \{x \in \mathbb{R} : x \geq 0\}$ ; and, expectation is denoted  $\mathbf{E}[\cdot]$ .

<sup>2</sup> A unitary random matrix  $\mathbf{\Omega}$  is Haar distributed if its probability distribution is invariant to left or right multiplication by any constant unitary matrix. If  $\mathbf{X}$  is a square random matrix with i.i.d. complex Gaussian centered unit variance entries, then the unitary matrix  $\mathbf{X}(\mathbf{X}^\dagger\mathbf{X})^{-1/2}$  is Haar distributed.

<sup>3</sup>For technical reasons, we also require the elements have finite positive moments. Also, if  $\mathbf{H} \neq \mathbf{I}_N$ , we additionally require that  $\mathbf{S}$  is unitarily invariant, however, we believe the results apply more generally.

- *Downlink MC- or DS-CDMA:* In this case,  $\mathbf{S}$  represents the matrix of  $K$  signatures with spreading gain  $N$ . Typically, the number of output dimensions equals the number of input dimensions and hence  $\mathbf{H}$  is square.
  - For flat-fading DS-CDMA,  $\mathbf{H} = \mathbf{I}_N$  and  $\mathbf{A}$  represents the combined effect of each users’ transmit power and channel coefficient.
  - For MC-CDMA in frequency-selective fading,  $\mathbf{H}$  is the diagonal matrix of the channel frequency response in each subcarrier, and  $\mathbf{A}$  represents the transmit amplitude of each signature.
  - For DS-CDMA in frequency-selective fading,  $\mathbf{H}$  is a circulant or Toeplitz matrix constructed from the channel impulse response, and  $\mathbf{A}$  represents the transmit amplitude of each signature.
- *‘Rich’ MIMO:* The standard point-to-point flat fading MIMO channel model is given by (1), where  $K$  and  $N$  correspond to the number of transmit and receive antennas, respectively. The standard MIMO channel matrix with i.i.d. circularly symmetric complex Gaussian coefficients between each pair of transmit and receive antennas corresponds here to setting  $\mathbf{H} = \mathbf{I}_N$  and  $\mathbf{S}$  i.i.d. Gaussian. The matrix  $\mathbf{A}$  defines the transmit amplitudes on each antenna.
- *MIMO with richness parameter:* The MIMO channel model introduced in [20] can also be described by (1), with  $\mathbf{S}$  i.i.d. and  $\mathbf{H} = \mathbf{\Phi}\mathbf{C}$ , where  $\mathbf{\Phi}$  is i.i.d., and  $\mathbf{C}$  is diagonal. In this case,  $\mathbf{S}$  models the propagation from the transmitter to a ‘scattering array’, modeled by  $\mathbf{C}$ , and  $\mathbf{\Phi}$  models the propagation from the scattering array to the receiver. The rank of the scattering array matrix determines the richness of the MIMO channel.
- If we additionally require that the data vector  $\mathbf{b}$  is unitarily invariant, then the MMSE and ALS asymptotic SINR results we obtain *apply to all AWGN matrix-vector systems* of the form  $\mathbf{r} = \mathbf{C}\mathbf{b} + \mathbf{n}$ , under certain conditions on the channel matrix  $\mathbf{C}$ , and the data and noise vectors,  $\mathbf{b}$  and  $\mathbf{n}$ . A full explanation is given in Section II-E.

### C. MMSE Receiver

The output of the MMSE receiver with full channel state information (CSI) and knowledge of  $\mathbf{S}$  for stream  $k$  at symbol interval  $m$  is given by

$$\hat{\mathbf{b}}_m(k) = \mathbf{c}_k^\dagger \mathbf{r}_m \quad (2)$$

where

$$\mathbf{c}_k = \mathbf{R}^{-1} \mathbf{H}\mathbf{s}_k, \quad (3)$$

$$\mathbf{R} = (\mathbf{H}\mathbf{S}\mathbf{A})^\ddagger + \sigma^2 \mathbf{I}_M. \quad (4)$$

Now, identifying the signal and interference components of the received signal in (1), i.e.,  $\mathbf{r}_m = A_k \mathbf{H}\mathbf{s}_k \mathbf{b}_m(k) + \mathbf{r}_m^I$ , the corresponding output SINR is defined as

$$\text{SINR}_{k,N}^{\text{MMSE}} = \frac{\mathbf{E}[|\mathbf{c}_k^\dagger (\mathbf{r}_m - \mathbf{r}_m^I)|^2]}{\mathbf{E}[|\mathbf{c}_k^\dagger \mathbf{r}_m^I|^2]} \quad (5)$$

where the expectation in (5) is with respect to  $\mathbf{b}_m$  and  $\mathbf{n}_m$ . The subscript  $N$  indicates that this is a non-asymptotic quantity.

#### D. ALS Receiver

The output of the adaptive least-squares (ALS) receiver with  $i$  training symbols for stream  $k$  at symbol interval  $m > i$  is given by

$$\hat{\mathbf{b}}_m(k) = \hat{\mathbf{c}}_k^\dagger \mathbf{r}_m \quad (6)$$

where

$$\hat{\mathbf{c}}_k = \hat{\mathbf{R}}^{-1} \hat{\mathbf{s}}_k, \quad (7)$$

$$\hat{\mathbf{R}} = \frac{1}{i} \mathcal{R} \mathbf{W} \mathcal{R}^\dagger + \frac{\mu}{\eta} \mathbf{I}_M, \quad (8)$$

$$\mathcal{R} = \mathbf{H} \mathbf{S} \mathbf{A} \mathbf{B}^\dagger + \mathbf{N}, \quad (9)$$

$$\hat{\mathbf{s}}_k = \begin{cases} \frac{1}{i} \mathcal{R} \mathbf{W} \mathbf{b}_k & , \text{ with training,} \\ \mathbf{H} \mathbf{s}_k & , \text{ semi-blind.} \end{cases} \quad (10)$$

and where

- $\mathbf{B} \in \mathbb{C}^{i \times K}$  contains training data, where the  $m^{\text{th}}$  row of  $\mathbf{B}$  is  $\mathbf{b}_m^\dagger$ . The  $k^{\text{th}}$  column of  $\mathbf{B}$  will be denoted as  $\mathbf{b}_k$ . We consider both i.i.d. and orthogonal training sequences, i.e.,
  - $\mathbf{B}$  contains i.i.d. elements with zero mean, unit variance, and finite positive moments, or,
  - $\mathbf{B}$  contains either random orthogonal rows or columns. If  $K < i$  then  $\mathbf{B}^\dagger \mathbf{B} = i \mathbf{I}_K$ , and we assume that  $\frac{1}{\sqrt{i}} \mathbf{B}$  is obtained by extracting  $K$  columns from an  $i \times i$  Haar-distributed unitary random matrix. If  $K > i$ , then  $\mathbf{B} \mathbf{B}^\dagger = K \mathbf{I}_i$ , and we assume that  $\frac{1}{\sqrt{K}} \mathbf{B}^\dagger$  is obtained by extracting  $i$  columns from a  $K \times K$  Haar-distributed unitary random matrix.

We shall call the first case ‘i.i.d.  $\mathbf{B}$ ’, and the second case<sup>4</sup> ‘orthogonal (ort.)  $\mathbf{B}$ ’.

- $\mathbf{N} \in \mathbb{C}^{M \times i}$  contains observation noise, where the  $m^{\text{th}}$  column is  $\mathbf{n}_m$ .
- $\mu \in \mathbb{R}^*$  is a diagonal loading constant, and  $\eta = i/N$ .
- $\mathbf{W} \in \mathbb{R}^{i \times i}$  is a diagonal data windowing matrix, i.e.  $\mathbf{W} = \text{diag}(w_1, \dots, w_i)$  where  $w_m \in \mathbb{R}^*$  for  $m = 1, \dots, i$ . For example, with exponential weighting  $w_m = \epsilon^{i-m}$ , where  $\epsilon \in (0, 1]$ , or, without windowing  $\mathbf{W} = \mathbf{I}_i$ .

Note that, although strictly speaking the model in (1) applies only to time-invariant systems, we include windowing to allow for practical situations such as slowly time varying channels, or users entering/leaving the system. The term ‘semi-blind’ in (10) refers to the case when  $\mathbf{H}$  and  $\mathbf{s}_k$  are known, and there is no training data, whereas ‘with training’ refers to when just  $\mathbf{B}$  is known. For more details on the practical issues, see [9], where this ALS formulation is considered for DS-CDMA in flat-fading.

The SINR for the  $k^{\text{th}}$  stream at the output of the ALS receiver,  $\text{SINR}_{k,N}^{\text{ALS}}$ , is defined by the right-hand side of (5), however, with  $\mathbf{c}_k$  replaced by  $\hat{\mathbf{c}}_k$ .

#### E. Large-System Limit

We define  $\alpha = K/N$  and  $\beta = M/N$ , and for the ALS receiver,  $\eta = i/N$ .

<sup>4</sup>As the columns of iso.  $\mathbf{S}$  are orthonormal, while the columns/rows of ort.  $\mathbf{B}$  are just orthogonal, we do not refer to  $\mathbf{B}$  as isometric.

Throughout this paper we consider the asymptotic limit  $(M, N, K, i) \rightarrow \infty$  with  $K/N \rightarrow \alpha > 0$ ,  $i/N \rightarrow \eta > 0$ , and  $M/N \rightarrow \beta > 0$  constant. For isometric  $\mathbf{S}$ ,  $\alpha \leq 1$ .

With data windowing it is necessary to consider how  $\mathbf{W}$  is defined for each  $i$  so that the empirical distribution function (e.d.f.) of its diagonal values converges to something appropriate. Any finite window length becomes negligible in the large-system limit as  $i \rightarrow \infty$ , therefore it is necessary to scale the window shape with the system size. For example, as in [16], with exponential windowing we define  $L = \frac{1}{1-\epsilon}$  as the ‘average’ window length, and take  $L \rightarrow \infty$  with  $L/N \rightarrow \bar{L} > 0$  constant.

To facilitate the large-system analysis, we also require that  $\mathbf{H}^\dagger$ ,  $\mathbf{A}^2$ , and  $\mathbf{W}$  each have a uniformly bounded spectral norm,<sup>5</sup> that is, a bound which is independent of the system dimension  $N$ . Also, we require the empirical distribution functions of the eigenvalues of  $\mathbf{A}^2$ ,  $\mathbf{H}^\dagger$ , and  $\mathbf{W}$  to converge in distribution almost surely to non-random distributions on the non-negative real axis, which will have compact support due to the previous assumption. We also assume that the limiting distributions of  $\mathbf{A}^2$ ,  $\mathbf{H}^\dagger$ , and  $\mathbf{W}$  are non-trivial, i.e., do not have all mass at zero.

*Proposition 1:* The large-system MMSE and ALS SINRs corresponding to the transmission model in (1) are the same as the large-system MMSE and ALS SINRs, respectively, computed for any matrix-vector system of the form  $\mathbf{r} = \mathbf{C} \mathbf{b} + \mathbf{n}$  in which

- $\mathbf{C} \in \mathbb{C}^{M \times K}$ ,  $\mathbf{b} \in \mathbb{C}^{K \times 1}$ , and  $\mathbf{n} \in \mathbb{C}^{M \times 1}$  are mutually independent,
- $\mathbf{n}$  satisfies the same conditions as  $\mathbf{n}_m$  above,
- the data/training vector  $\mathbf{b}$  satisfies the same conditions as  $\mathbf{b}_m$  above, and *additionally is unitarily invariant*,<sup>6</sup>
- the channel matrix  $\mathbf{C}$  is such that the e.d.f. of  $\mathbf{C}^\dagger$  satisfies the conditions on  $\mathbf{H}^\dagger$  mentioned in the previous paragraph,

and corresponds to taking  $\mathbf{H} = \mathbf{C}$ ,  $\mathbf{S}$  to be  $K \times K$  isometric, and  $\mathbf{A} = \mathbf{I}_K$ , respectively.

*Proof:* Since  $\mathbf{b}$  is unitarily invariant, all data streams have identically distributed SINRs. Since  $\mathbf{b}$  has the same distribution as  $\mathbf{U} \mathbf{b}$ , where  $\mathbf{U}$  is a  $K \times K$  Haar-distributed random unitary matrix, we see that the distribution of the MMSE and ALS SINR associated with  $\mathbf{r}' = \mathbf{C} \mathbf{U} \mathbf{b} + \mathbf{n}$  are, respectively, the same as that for  $\mathbf{r}$ , and will share a common large-system limit, if it exists.  $\square$

Proposition 1 implies that our model encompasses the classic equalization model. Namely,  $\mathbf{r}$  represents  $N = M$  samples

<sup>5</sup>In particular, this condition is required for the derivations in the appendices, which frequently rely on Lemma 9 in Appendix 1 along with other key lemmas as a precursor to the asymptotic analysis contained in the remaining appendices.

<sup>6</sup>That is, the elements of  $\mathbf{U} \mathbf{b}$  have the same joint distribution as  $\mathbf{b}$  for any  $K \times K$  unitary matrix  $\mathbf{U}$ . This is an extra restriction on i.i.d. data vectors. For training vectors from Haar-distributed matrices, this condition is automatically satisfied, and is easily verified for  $\alpha \geq \eta$  (that is, when  $\mathbf{b}$  is a column from a Haar-distributed matrix). For  $\alpha < \eta$ , note that  $\mathbf{U} \mathbf{b}$  corresponds to the Hermitian transpose of a row of  $\mathbf{B} \mathbf{U}^\dagger$ , and  $\mathbf{B} \mathbf{U}^\dagger$  can be written as  $\Theta \mathbf{E}_K \mathbf{U}^\dagger$ , where  $\Theta$  is  $i \times i$  Haar, and  $\mathbf{E}_K = [\mathbf{I}_K, \mathbf{0}_{K, i-K}]^\dagger$ . Alternately, we have  $\mathbf{B} \mathbf{U}^\dagger = \Theta \tilde{\mathbf{U}} \mathbf{E}_K$ , where  $\tilde{\mathbf{U}}$  is the unitary matrix created by replacing the upper left  $K \times K$  sub-block of  $\mathbf{I}_i$  with  $\mathbf{U}^\dagger$ . Since  $\Theta$  is Haar,  $\Theta \tilde{\mathbf{U}}$  is also Haar, and hence  $\mathbf{B} \mathbf{U}^\dagger$  has the same distribution as  $\mathbf{B}$ .

at the output of a single-input/single-output (SISO) FIR channel  $\mathbf{h}$  of length  $L_c$ , i.e.,  $\mathbf{r}(n) = \sum_{\ell=0}^{L_c-1} \mathbf{h}(\ell)\mathbf{b}(n-\ell) + \mathbf{n}(n)$ . If a cyclic prefix of appropriate length is used, we set  $\mathbf{C}$  defined in Proposition 1 equal to the circulant channel matrix. Therefore, from Proposition 1, the corresponding model (1) takes  $\mathbf{A} = \mathbf{I}_N$ ,  $\mathbf{S}$  as  $N \times N$  isometric, and  $\mathbf{H}$  as  $\mathbf{C}$  or equivalently as an  $N \times N$  diagonal matrix with the  $N$ -point DFT of  $\mathbf{h}$  on the diagonal.

### III. UNIFIED LARGE-SYSTEM ANALYSIS

In Sections IV and V, we derive the asymptotic SINR for the model (1) with both MMSE and ALS receivers. The SINR in both cases is directly related to the Stieltjes transform<sup>7</sup> of the a.e.d. of the received signal correlation matrix  $\mathbf{R}$  for the MMSE receiver, and  $\hat{\mathbf{R}}$  for the ALS receiver. For the MMSE case, there are a number of existing methods for finding such transforms directly (see e.g., [1]). However, those methods do not extend to the ALS problem. We now discuss a general approach, which applies to both i.i.d. and isometric  $\mathbf{S}$  for both the MMSE and the ALS receiver.

The aim is to derive a set of equations for each ‘‘constituent’’ dimension in  $\mathbf{R}$  (or  $\hat{\mathbf{R}}$ ), which can be solved for the Stieltjes transform. For example,  $\mathbf{R}$  has three constituent dimensions ( $K$ ,  $N$ , and  $M$ ), while  $\hat{\mathbf{R}}$  has four, since it also includes  $i$ .

Each equation is based on expanding the simple identity  $\mathbf{R}^{-1}\mathbf{R} = \mathbf{I}_M$  (or  $\hat{\mathbf{R}}^{-1}\hat{\mathbf{R}} = \mathbf{I}_M$ ) in each constituent dimension. That is, since  $\mathbf{R}$  (or  $\hat{\mathbf{R}}$ ) is Hermitian, this term in the identity can be written as a sum of vector outer products, where the sum index runs up to the value of the dimension. Taking the normalized trace of both sides of the resulting equation can be simplified, and involves terms which have equivalent asymptotic forms, which can be evaluated using an asymptotic extension to the matrix inversion lemma. These equivalent forms are in terms of scalar variables, some of which are mixed matrix moments. Each of these moments can be expressed in terms of the other variables.

The result is a set of equations, which can be solved for the Stieltjes transform, and other unknowns (e.g., various matrix moments). Interestingly, in all cases we consider, the equations can be written in a form such that solving for the Stieltjes transform numerically amounts to zero-finding in at most two dimensions.

### IV. ANALYSIS OF MMSE RECEIVER

It has been shown that, for both i.i.d. and isometric  $\mathbf{S}$ , the asymptotic SINR for the  $k^{\text{th}}$  stream at the output of the full-CSI MMSE receiver in (3) satisfies [7, 21]

$$\max_{k \leq K} |\text{SINR}_{k,N}^{\text{MMSE}} - P_k \rho_1^N| \xrightarrow{a.s.} 0 \quad (11)$$

under the limit considered, where  $\xrightarrow{a.s.}$  denotes almost-sure convergence,

$$\rho_1^N = \begin{cases} \frac{1}{N} \text{tr}[\mathbf{H}^\dagger \mathbf{R}^{-1}] & , \text{ i.i.d. } \mathbf{S}, \\ \frac{1}{N-K} \text{tr}[(\mathbf{H}\mathbf{I}\mathbf{I})^\dagger \mathbf{R}^{-1}] & , \text{ iso. } \mathbf{S}, \end{cases} \quad (12)$$

<sup>7</sup>The Stieltjes (or Cauchy) transform of the distribution of a real-valued random variable  $X$  is the expected value of  $1/(X-z)$ , where  $z \in \mathbb{C}^+$  is the transform variable, and  $\mathbb{C}^+ = \{x \mid x \in \mathbb{C}, \text{Im}(x) > 0\}$  (e.g., see [2]).

and  $\mathbf{\Pi}^2 = \mathbf{I}_N - \mathbf{S}^\dagger$ .

We now discuss some existing methods for computing the limit of the moment  $\rho_1^N$  in (12) with the MMSE full-CSI receiver, and note that the methods do not extend to the ALS receiver. We then present the main result of this section, namely, a general SINR expression, which applies to both i.i.d. and isometric  $\mathbf{S}$ , derived using the approach discussed in Section III.

For i.i.d.  $\mathbf{S}$ , and square invertible  $\mathbf{H}$ , the SINR can be obtained in terms of the limiting distribution of  $\mathbf{H}^\dagger$  and  $\mathbf{A}^2$ , using the result of Silverstein and Bai [2] after writing

$$\rho_1^N = \frac{1}{N} \text{tr}[(\mathbf{S}\mathbf{A}^2\mathbf{S}^\dagger + \sigma^2(\mathbf{H}^\dagger\mathbf{H})^{-1})^{-1}] \quad (13)$$

as was done in [22], and for more general channel distributions in [6, Theorem 2]. More generally, a solution for arbitrary (non-square) channel models can be obtained for i.i.d.  $\mathbf{S}$  via Girko’s law (see e.g., [5, Lemma 1], or [23, Theorem IV.2] for just the Stieltjes transform of  $\mathbf{R}$ ), again in terms of the limiting distributions of  $\mathbf{H}^\dagger$  and  $\mathbf{A}^2$ . We note that neither of these techniques can be used to compute the output SINR for the MMSE receiver with isometric  $\mathbf{S}$ , or the ALS receiver.

For isometric  $\mathbf{S}$  and square  $\mathbf{H}$ , the asymptotic SINR was first presented in [6] with  $\mathbf{A} = \mathbf{I}_K$ , and was extended in [7] to include general  $\mathbf{A}$ . Although this approach could also be used to consider non-square  $\mathbf{H}$ , it does not extend to the ALS receiver since it relies on the particular structure of  $\mathbf{R}$ , which is not shared by  $\hat{\mathbf{R}}$ .

#### A. Asymptotic SINR for MMSE Receiver

The following theorem allows us to evaluate the limit of  $\rho_1^N$  (and hence the asymptotic MMSE SINR) for general channels and for either i.i.d. or isometric  $\mathbf{S}$ . The theorem is in terms of the Stieltjes transform of the e.d.f. of the eigenvalues of  $\mathbf{R}$  from (4) by replacing  $\sigma^2$  by a complex variable  $z \in \mathbb{C}^+$  (i.e.,  $\mathbf{R} = (\mathbf{H}\mathbf{S}\mathbf{A})^\dagger - z\mathbf{I}_M$ ), such that the Stieltjes transform of the e.d.f. of the eigenvalues of  $(\mathbf{H}\mathbf{S}\mathbf{A})^\dagger$  is given by  $G_{\mathbf{R}}^N(z) = \frac{1}{M} \text{tr}[\mathbf{R}^{-1}]$ . The theorem is given in terms of the two additional random sequences  $\rho_1^N \in \mathbb{C}^+$ , as defined in (12) using the redefinition of  $\mathbf{R}$ , and  $\tau_1^N \in \mathbb{C}^+$ . The variable  $\tau_1^N$  is defined in terms of matrix equations, and is given in Appendix III-A since the definition is lengthy and is not needed to state the following result.

*Theorem 1:* Under the assumptions in Section II-E, as  $(M, N, K) \rightarrow \infty$  with  $M/N \rightarrow \beta > 0$  and  $K/N \rightarrow \alpha > 0$  fixed, the Stieltjes transform of the e.d.f. of the eigenvalues of  $(\mathbf{H}\mathbf{S}\mathbf{A})^\dagger$ ,  $G_{\mathbf{R}}^N(z)$ ,  $z \in \mathbb{C}^+$ , along with  $\rho_1^N$  and  $\tau_1^N$  satisfy

$$|G_{\mathbf{R}}^N(z) - \gamma_1| \xrightarrow{a.s.} 0, \quad (14)$$

$$|\rho_1^N - \rho_1| \xrightarrow{a.s.} 0, \quad (15)$$

$$|\tau_1^N - \tau_1| \xrightarrow{a.s.} 0, \quad (16)$$

where  $\gamma_1, \rho_1, \tau_1 \in \mathbb{C}^+$  are solutions to

$$\gamma_1 = -z^{-1} \left( 1 - \frac{\alpha}{\beta} \rho_1 \mathcal{E}_{1,1} \right), \quad (17)$$

$$\rho_1 = \begin{cases} -z^{-1}\beta^*\mathcal{H}_{1,1} & , \text{ i.i.d. } \mathbf{S}, \\ \frac{-z^{-1}\beta^*\mathcal{H}_{1,1}}{1-\beta(1+z\gamma_1)} & , \text{ iso. } \mathbf{S}, \end{cases} \quad (18)$$

$$\tau_1 = \alpha\bar{p} - \begin{cases} \alpha\mathcal{E}_{1,1} & , \text{ i.i.d. } \mathbf{S}, \\ \frac{\alpha\mathcal{E}_{1,1}}{1-\beta(1+z\gamma_1)} & , \text{ iso. } \mathbf{S}, \end{cases} \quad (19)$$

where  $\beta^* = \min(\beta, 1)$ , and

$$\mathcal{E}_{p,1} = \mathbf{E} \left[ \frac{P^p}{1+P\rho_1} \right], \quad (20)$$

$$\mathcal{H}_{p,1} = \mathbf{E} \left[ \frac{H^p}{1+Hz^{-1}(\tau_1 - \alpha\bar{p})} \right], \quad (21)$$

for  $p \in \mathbb{Z}^*$ . The expectations in (20) and (21) are with respect to the scalar random variables  $P$  and  $H$ , respectively, and the distributions of  $P$  and  $H$  are the asymptotic eigenvalue distributions of  $\mathbf{A}^2$  and the first  $\beta^*N = \min(M, N)$  non-zero eigenvalues of  $\mathbf{H}^\dagger$ , respectively, and  $\bar{p} = \mathbf{E}[P]$ .

*Proof:* See Appendix III.  $\square$

Remarks:

- If (17)–(19) has a unique solution<sup>8</sup>  $\gamma_1, \rho_1, \tau_1 \in \mathbb{C}^+$  for any given  $z \in \mathbb{C}^+$ , then Theorem 1 additionally implies that the e.d.f. of the eigenvalues of  $(\mathbf{HSA})^\dagger$  almost surely converges in distribution to a deterministic distribution, whose Stieltjes transform is  $G_R(z)$ . Moreover, we have that  $\rho_1^N$  converges almost surely to the deterministic value  $\rho_1$  in the limit considered, and so, letting  $z = -\sigma^2 + \epsilon j$  and taking  $\epsilon \rightarrow 0$ , as suggested by (11), the MMSE SINR of the  $k^{\text{th}}$  data stream almost surely converges to  $P_k\rho_1$ .
- For i.i.d.  $\mathbf{S}$ , Theorem 1 can be obtained via Girko's law (see e.g., [23, Theorem IV.2]). For isometric  $\mathbf{S}$ , this result appears to be new. However, in both cases Theorem 1 can be derived (under the same set of assumptions) using the  $\mathbf{S}$ -transform from free probability, or the method of [24]. We give a different proof, relying only on elementary matrix manipulations. The primary reason for presenting this result is to lead into the ALS analysis, which will follow the same general procedure outlined in the proof of Theorem 1 in Appendix III.
- The following steps describe how to find  $G_R(z)$  numerically via Theorem 1 for specific distributions of  $P$  and  $H$ , and given values of  $z$ ,  $\alpha$ , and  $\beta$ .
  - Consider (18) as a scalar function of  $\rho_1$ , i.e.,  $X(\rho_1) = 0$ .
  - Numerically find the unique positive root of  $X(\rho_1)$  using standard techniques (e.g., using a routine such as *fzero* in Matlab), where, for a given value of  $\rho_1$ , the corresponding values of  $\tau_1$  and  $\gamma_1$  are directly evaluated using (19) for  $\tau_1$ , and (17) for  $\gamma_1$ .
- In fact, (17) is just one of many possible expressions which can be derived from the identity  $\frac{1}{M}\text{tr}[\mathbf{R}\mathbf{R}^{-1}] = 1$ . Other expressions involving  $\gamma_1$  derived in this manner

<sup>8</sup>Certainly, for i.i.d.  $\mathbf{S}$ , the equations of Theorem 1 have a unique solution, since the same result can be obtained via Girko's law, for which the solution is known to be unique. Although it is not proved, we believe that this is also true for isometric  $\mathbf{S}$ . Numerical studies support this.

include

$$\beta(1+z\gamma_1) = \alpha(1-\mathcal{E}_{0,1}) = \alpha\rho_1\mathcal{E}_{1,1} \quad (22)$$

$$= \beta^*(1-\mathcal{H}_{0,1}) = \beta^*z^{-1}(\tau_1 - \alpha\bar{p})\mathcal{H}_{1,1} \quad (23)$$

These expressions, derived in Appendix III, are used in the proof of Theorem 1.

- If  $\mathbf{H}$  is exponentially distributed with mean one (i.e., MC- or DS-CDMA in frequency-selective Rayleigh fading),  $\mathcal{H}_{1,1} = (1-f(x))/x$ , where  $x = z^{-1}(\tau_1 - \alpha\bar{p})$ , and  $f(x) = x^{-1}\exp(x^{-1})\text{Ei}(x^{-1})$  where  $\text{Ei}(x) = \int_1^\infty e^{-xt}t^{-1}dt$  is the first-order Exponential Integral.

## B. Alternate Representation of MMSE SINR

We now present an alternate expression for the asymptotic value of  $\text{SINR}_{k,N}^{\text{MMSE}}$ , which will allow us to determine the relationship between the asymptotic MMSE SINR and the asymptotic ALS SINR considered later in Sections V–VI. This expression depends on the additional random sequences  $\rho_j^N \in \mathbb{R}^*$ ,  $j = 2, 3, 4$ , which, along with the auxiliary random sequences  $\tau_j^N \in \mathbb{R}^*$ ,  $j = 2, 3$ , are defined in terms of matrix equations in Appendix III-A (The definitions of these variables are lengthy, and are not needed to state the following result; so to facilitate the flow of results they are not stated here.)

It is shown in Appendix V that, under the assumptions in Section II-E,

$$\max_{k \leq K} \left| \text{SINR}_{k,N}^{\text{MMSE}} - \frac{P_k |\rho_1^N|^2}{\rho_4^N + \sigma^2 \rho_2^N} \right| \xrightarrow{a.s.} 0 \quad (24)$$

as  $(M, N, K) \rightarrow \infty$  with  $M/N \rightarrow \beta > 0$  and  $K/N \rightarrow \alpha > 0$ . Moreover,  $|\rho_j^N - \rho_j| \xrightarrow{a.s.} 0$ ,  $j = 2, 3, 4$ , and  $|\tau_j^N - \tau_j| \xrightarrow{a.s.} 0$ ,  $j = 2, 3$ , in the limit considered, where  $\rho_j \in \mathbb{R}^*$ ,  $j = 2, 3, 4$ , and  $\tau_j \in \mathbb{R}^*$ ,  $j = 2, 3$ , are solutions to the following set of equations. For i.i.d.  $\mathbf{S}$ ,

$$\rho_j = \begin{cases} \beta^* |z|^{-2} (\mathcal{H}_{1,2} + \tau_2 \mathcal{H}_{2,2}) & , \quad j = 2, \\ \beta^* |z|^{-2} (1 + \tau_3) \mathcal{H}_{2,2} & , \quad j = 3, \\ \alpha \rho_3 \mathcal{E}_{1,2} & , \quad j = 4. \end{cases} \quad (25)$$

$$\tau_j = \alpha \rho_j \mathcal{E}_{2,2} \quad , \quad j = 2, 3. \quad (26)$$

while for isometric  $\mathbf{S}$ ,

$$\rho_j = \frac{\begin{cases} \beta^* |z|^{-2} (\mathcal{H}_{1,2} + \tau_2 \mathcal{H}_{2,2}) & , \quad j = 2, \\ \beta^* |z|^{-2} (1 + \tau_3) \mathcal{H}_{2,2} & , \quad j = 3, \\ \alpha (\rho_3 \mathcal{E}_{1,2} - |\rho_1|^2 \mathcal{E}_{1,2}) & , \quad j = 4. \end{cases}}{\alpha (\mathcal{E}_{0,2} - 1) + 1}, \quad (27)$$

$$\tau_j = \frac{\alpha \rho_j \mathcal{E}_{2,2} - \beta^* |z|^{-2} |\alpha\bar{p} - \tau_1|^2 \mathcal{H}_{j-1,2}}{\beta^* (\mathcal{H}_{0,2} - 1) + 1}, \quad j = 2, 3. \quad (28)$$

where

$$\mathcal{E}_{p,2} = \mathbf{E} \left[ \frac{P^p}{|1 + P\rho_1|^2} \right], \quad (29)$$

$$\mathcal{H}_{p,2} = \mathbf{E} \left[ \frac{H^p}{|1 + Hz^{-1}(\tau_1 - \alpha\bar{p})|^2} \right], \quad (30)$$

for  $p \in \mathbb{Z}^*$ , and  $\mathcal{E}_{q,1}, \mathcal{H}_{q,1}, q \in \mathbb{Z}^*$ ,  $\rho_1, \tau_1, \bar{p}, \beta^*, P$ , and  $H$  are determined from Theorem 1.

Again, assuming that (17)–(19) and also (25)–(28) have unique solutions, then we also have that the SINR of the  $k^{\text{th}}$  data stream converges almost surely to the deterministic quantity  $P_k |\rho_1|^2 / (\rho_4 + \sigma^2 \rho_2)$ .

## V. ANALYSIS OF ALS RECEIVER

In this section we derive the asymptotic SINR for the adaptive receiver, using the general approach discussed in Section III.

Firstly, we derive the asymptotic transient ALS SINR after a specified number of training intervals (either with training, or semi-blind). Our aim is to characterize the typical transient response of the receiver as a function of  $\eta$ , i.e., as the number of training symbols increases. The resulting expression is in terms of several large matrix variables involving the sample autocorrelation matrix. We present a theorem which gives the Stieltjes transform of the a.e.d. of the sample autocorrelation matrix, and fixed-point expressions for each variable required to compute the asymptotic SINR.

Secondly, from the transient SINR solution we determine the steady-state asymptotic ALS SINR, that is, the SINR as the number of training intervals (either with a training sequence, or semi-blind) goes to infinity (i.e.,  $\eta \rightarrow \infty$ ). Without data windowing, we verify that the solution for the ALS SINR converges to the MMSE SINR. Then we determine the steady-state SINR when an arbitrary windowing function is used.

### A. Transient ALS SINR

The following result relates the transient SINR of the ALS receiver to six auxiliary random sequences,  $\hat{\rho}_j^N$  and  $\hat{\psi}_j^N$ ,  $j = 1, 2, 4$ . The definitions of these variables are in terms of matrix traces and quadratic forms, and are quite lengthy. So, for clarity of presentation, and also since asymptotically equivalent values of these variables can be calculated from subsequent results, the definitions are given in Appendix VII-A.

*Theorem 2:* In the limit as  $(M, N, K, i) \rightarrow \infty$  with  $M/N \rightarrow \beta > 0$ ,  $K/N \rightarrow \alpha > 0$ , and  $i/N \rightarrow \eta > 0$  fixed,

$$\max_{k \leq K} \left| \text{SINR}_{k,N}^{\text{ALS}} - \frac{P_k |a_{k,1}|^2 |\hat{\rho}_1^N|^2}{|a_{k,1}|^2 (\hat{\rho}_4^N + \sigma^2 \hat{\rho}_2^N) + |a_{k,2}|^2 (\hat{\psi}_4^N + \sigma^2 \hat{\psi}_2^N)} \right| \xrightarrow{a.s.} 0 \quad (31)$$

where

$$(a_{k,1}, a_{k,2}) = \begin{cases} (1, -A_k \hat{\rho}_1^N) & , \text{ semi-blind LS,} \\ (A_k (\bar{w} - \hat{\psi}_1^N), 1) & , \text{ LS with training.} \end{cases} \quad (32)$$

$\bar{w}$  is the mean of the a.e.d. of  $\mathbf{W}$ , and the definitions of  $\hat{\rho}_j^N$  and  $\hat{\psi}_j^N$ ,  $j = 1, 2, 4$  are given in Appendix VII-A.

*Proof:* See Appendix VIII. Note that the definitions and derivations of Appendix VII necessarily precede Appendix VIII.  $\square$

Remarks:

- Expressions, which can be used to compute asymptotically equivalent values of  $\hat{\rho}_j^N$  and  $\hat{\psi}_j^N$ ,  $j = 1, 2, 4$ , are presented in Theorem 3 and Lemma 1.

- The preceding ALS SINR expression resembles the alternate MMSE SINR expression (24) derived in Section IV-B. However, a simplified expression for the ALS SINR, such as that presented for the MMSE SINR in Theorem 1, is not possible. This is due to the fact that a simplification of the interference power, as discussed in the proof of (24) in Appendix V, is not possible for the ALS receiver.

The following Theorem and Lemma give a sufficient number of relations to calculate the asymptotic moments required for the asymptotic SINR in (31) of Theorem 2.

In a similar manner to Section IV, firstly we determine expressions for the Stieltjes transform of the e.d.f. of the eigenvalues of the *sample* autocorrelation matrix. That is, we generalize the definition of  $\hat{\mathbf{R}}$  as follows,  $\hat{\mathbf{R}} = \frac{1}{i} \mathcal{R} \mathbf{W} \mathcal{R}^\dagger - z \mathbf{I}_M$ , where  $z \in \mathbb{C}^+$ , such that the Stieltjes transform of the e.d.f. of the eigenvalues of  $\frac{1}{i} \mathcal{R} \mathbf{W} \mathcal{R}^\dagger$  is given by  $G_{\hat{\mathbf{R}}}^N(z) = \frac{1}{M} \text{tr}[\hat{\mathbf{R}}^{-1}]$ . The result is necessarily stated in terms of the additional random sequences  $\hat{\nu}_1^N$ ,  $\hat{\rho}_1^N$ ,  $\hat{\tau}_1^N$ ,  $\hat{\psi}_1^N$ ,  $\hat{\omega}_1^N$ , and  $\hat{r}_1^N \in \mathbb{C}^+$ . As in Theorem 2, the definitions of these variables are in terms of matrix traces and quadratic forms, and are lengthy. To facilitate the presentation of results, the definitions of these variables are given in Appendix VII-A.

*Theorem 3:* Under the assumptions in Section II-E, as  $(M, N, K, i) \rightarrow \infty$ , with  $M/N \rightarrow \beta > 0$ ,  $K/N \rightarrow \alpha > 0$ , and  $i/N \rightarrow \eta > 0$  fixed, the Stieltjes transform of the e.d.f. of the eigenvalues of  $\frac{1}{i} \mathcal{R} \mathbf{W} \mathcal{R}^\dagger$ ,  $G_{\hat{\mathbf{R}}}^N(z)$ ,  $z \in \mathbb{C}^+$ , along with  $\hat{\nu}_1^N$ ,  $\hat{\rho}_1^N$ ,  $\hat{\tau}_1^N$ ,  $\hat{\psi}_1^N$ ,  $\hat{\omega}_1^N$ , and  $\hat{r}_1^N \in \mathbb{C}^+$  satisfy

$$|G_{\hat{\mathbf{R}}}^N(z) - \hat{\gamma}_1| \xrightarrow{a.s.} 0, \quad (33)$$

$$|\hat{\rho}_1^N - \hat{\rho}_1| \xrightarrow{a.s.} 0, \quad |\hat{\tau}_1^N - \hat{\tau}_1| \xrightarrow{a.s.} 0, \quad (34)$$

$$|\hat{\psi}_1^N - \hat{\psi}_1| \xrightarrow{a.s.} 0, \quad |\hat{\omega}_1^N - \hat{\omega}_1| \xrightarrow{a.s.} 0, \quad (35)$$

$$|\hat{\nu}_1^N - \hat{\nu}_1| \xrightarrow{a.s.} 0, \quad |\hat{r}_1^N - \hat{r}_1| \xrightarrow{a.s.} 0. \quad (36)$$

where  $\hat{\gamma}_1$ ,  $\hat{\rho}_1$ ,  $\hat{\tau}_1$ ,  $\hat{\psi}_1$ ,  $\hat{\omega}_1$ ,  $\hat{\nu}_1$ , and  $\hat{r}_1 \in \mathbb{C}^+$  are solutions to

$$\hat{\gamma}_1 = \hat{\nu}_1 \left( 1 + \frac{\alpha}{\beta} \hat{\rho}_1 (\hat{\psi}_1 - \bar{w}) \hat{\mathcal{E}}_{1,1} \right), \quad (37)$$

$$\hat{\nu}_1 = \frac{1}{\sigma^2 \mathcal{W}_{1,1} - z}, \quad (38)$$

$$\hat{\rho}_1 = \begin{cases} \hat{\nu}_1 \beta^* \hat{\mathcal{H}}_{1,1} & , \text{ i.i.d. } \mathbf{S}, \\ \frac{\hat{\nu}_1 \beta^* \hat{\mathcal{H}}_{1,1}}{1 - \beta(1 - \hat{\gamma}_1/\hat{\nu}_1)} & , \text{ iso. } \mathbf{S}, \end{cases} \quad (39)$$

$$\hat{\tau}_1 = \alpha \bar{p} \bar{w} + \begin{cases} \alpha (\hat{\psi}_1 - \bar{w}) \hat{\mathcal{E}}_{1,1} & , \text{ i.i.d. } \mathbf{S}, \\ \frac{\alpha (\hat{\psi}_1 - \bar{w}) \hat{\mathcal{E}}_{1,1}}{1 - \beta(1 - \hat{\gamma}_1/\hat{\nu}_1)} & , \text{ iso. } \mathbf{S}, \end{cases} \quad (40)$$

$\beta^* = \min(\beta, 1)$ , and

$$\hat{\psi}_1 = \bar{w} - \begin{cases} \mathcal{W}_{1,1} & , \text{ i.i.d. } \mathbf{B}, \\ \frac{\mathcal{W}_{1,1}}{1 - \frac{\beta}{\eta^*} (1 - \hat{\gamma}_1/\hat{\nu}_1)} & , \text{ ort. } \mathbf{B}, \end{cases} \quad (41)$$

$$\hat{\omega}_1 = \begin{cases} \frac{\alpha}{\eta} \hat{\rho}_1 \hat{\mathcal{E}}_{1,1} & , \text{ i.i.d. } \mathbf{B}, \\ \frac{\frac{\alpha}{\eta} \hat{\rho}_1 \hat{\mathcal{E}}_{1,1}}{1 - \frac{\beta}{\eta^*} (1 - \hat{\gamma}_1/\hat{\nu}_1)} & , \text{ ort. } \mathbf{B}, \end{cases} \quad (42)$$

$\eta^* = \max(\eta, \alpha)$ , and

$$\hat{\mathcal{E}}_{p,1} = \mathbf{E} \left[ \frac{P^p}{1 - P\hat{\rho}_1(\hat{\psi}_1 - \bar{w})} \right], \quad (43)$$

$$\hat{\mathcal{H}}_{p,1} = \mathbf{E} \left[ \frac{H^p}{1 - H\hat{\nu}_1(\hat{\tau}_1 - \alpha\bar{p}\bar{w})} \right], \quad (44)$$

$$\mathcal{W}_{p,1} = \mathbf{E} \left[ \frac{W^p}{1 + W\hat{r}_1} \right], \quad (45)$$

$$\hat{r}_1 = \hat{\omega}_1 + \frac{\beta}{\eta} \sigma^2 \hat{\gamma}_1, \quad (46)$$

for  $p \in \mathbb{Z}^*$ . The expectations in (43), (44), and (45) are with respect to the scalar random variables  $P$ ,  $H$ , and  $W$ , respectively, where the distributions of  $P$ ,  $H$ , and  $W$  are the a.e.d.s of  $\mathbf{A}^2$ , the first  $\beta^*N$  eigenvalues of  $\mathbf{H}^\dagger$ , and  $\mathbf{W}$ , respectively. Also,  $\bar{p} = \mathbf{E}[P]$  and  $\bar{w} = \mathbf{E}[W]$ .

*Proof:* See Appendix VII.  $\square$

Remarks:

- If (37)–(46) has a unique solution  $\hat{\gamma}_1, \hat{\nu}_1, \hat{\rho}_1, \hat{\tau}_1, \hat{\psi}_1, \hat{\omega}_1, \hat{r}_1 \in \mathbb{C}^+$  for any given  $z \in \mathbb{C}^+$ , then Theorem 3 additionally gives that the e.d.f. of the eigenvalues of  $\frac{1}{i} \mathcal{R} \mathbf{W} \mathcal{R}^\dagger$  almost surely converges in distribution to a deterministic distribution, whose Stieltjes transform is  $G_{\hat{\mathbf{R}}}(z)$ .
- Note that for i.i.d.  $\mathbf{B}$  and  $\mu \rightarrow 0$  (i.e., i.i.d. training sequences and no diagonal loading),  $\mathcal{W}_{1,1}\hat{\gamma}_1$ ,  $\mathcal{W}_{1,1}\hat{\rho}_1$ , and  $\mathcal{W}_{1,1}(\hat{\tau}_1 - \alpha\bar{p}\bar{w})$  satisfy the same equations as  $\gamma_1, \rho_1$ , and  $(\tau_1 - \alpha\bar{p})$  from Theorem 1. Moreover, due to (50), we have that  $\mathcal{W}_{1,1}$  is a function of only  $\beta, \eta$ , and the window shape. This observation, along with the alternate MMSE SINR expression of Section IV-B, are the key elements in determining the relationship between the ALS and MMSE SINRs, outlined later in Section VI.
- For exponential weighting with  $\bar{L} < \infty$  (where  $\bar{L}$  is the large-system window size defined in Section II-E), in Appendix VI the e.d.f. of  $\mathbf{W}$  is shown to converge in distribution to the fixed distribution

$$F_W(w) = 1 + \frac{\bar{L}}{\eta} \ln w, \quad e^{-\eta/\bar{L}} \leq w \leq 1 \quad (47)$$

which is the relevant distribution of  $W$  required in (45). Also, for  $z \rightarrow \text{Re}(z)$ , values of  $\mathcal{W}_{m,1}$  which are required can be evaluated using (47), and are given by

$$\mathcal{W}_{0,1} = \frac{\bar{L}}{\eta} \log \left( \frac{1 + e^{-\eta/\bar{L}} \hat{r}_1}{1 + \hat{r}_1} \right) + 1, \quad (48)$$

$$\mathcal{W}_{1,1} = \frac{1}{\hat{r}_1} (1 - \mathcal{W}_{0,1}). \quad (49)$$

- In fact, (37) is one of many possible expressions, which can be derived from the identity  $\frac{1}{M} \text{tr}[\hat{\mathbf{R}} \hat{\mathbf{R}}^{-1}] = 1$ . Other expressions involving  $\hat{\gamma}_1$  derived in this manner include

$$\beta(1 + z\hat{\gamma}_1) = \eta(1 - \mathcal{W}_{0,1}), \quad (50)$$

$$\beta(1 - \hat{\gamma}_1/\hat{\nu}_1) = \eta\hat{\omega}_1\mathcal{W}_{1,1} \quad (51)$$

$$= \alpha(1 - \hat{\mathcal{E}}_{0,1}) = -\alpha\hat{\rho}_1(\hat{\psi}_1 - \bar{w})\hat{\mathcal{E}}_{1,1} \quad (52)$$

$$= \beta^*(1 - \hat{\mathcal{H}}_{0,1}) = -\hat{\nu}_1(\hat{\tau}_1 - \alpha\bar{p}\bar{w})\beta^*\hat{\mathcal{H}}_{1,1}. \quad (53)$$

Note the similarity to the expressions (22)–(23), derived in a similar manner for the MMSE receiver. These expressions, derived in Appendix VII, are used in the proof of Theorem 3.

- Note that  $\hat{\mathcal{E}}_{p,1}$ ,  $\hat{\mathcal{H}}_{p,1}$ , and  $\mathcal{W}_{p,1}$  are all of the form  $\mathcal{X}_{p,1} = \mathbf{E}[X^p/(1 + xX)]$ , for which the following simple and useful identity holds.

$$\mathbf{E}[X^p] = \mathbf{E} \left[ X^p \frac{1 + xX}{1 + xX} \right] = \mathcal{X}_{p,1} + x\mathcal{X}_{p+1,1} \quad (54)$$

for  $p \in \mathbb{Z}^*$ . Observe also that the last equality in each of (22), (23), (52) and (53) follows from the identity (54), as does (49). This identity also relates (50) to (51), although in a less obvious way.

- The set of equations in Theorem 3 can be solved numerically in a similar manner to that discussed for the MMSE case after Theorem 1. Here, however, it is advantageous to consider (39) and (42) as a two-dimensional equation in the variables<sup>9</sup>  $\hat{\rho}_1$  and  $\hat{r}_1$ . During zero finding, given these values, the remaining variables  $\hat{\gamma}_1, \hat{\nu}_1, \hat{\tau}_1, \hat{\psi}_1, \hat{\omega}_1, \hat{\mathcal{E}}_{1,1}, \hat{\mathcal{H}}_{1,1}$ , and  $\mathcal{W}_{1,1}$  can be directly calculated.
- For reference, an independent treatment of the asymptotic distribution of sample covariance matrices of information-plus-noise signals is given in [25].

Using Theorem 3, we may now calculate  $\hat{\rho}_1$  and  $\hat{\psi}_1$ , which are asymptotically equivalent to  $\hat{\rho}_1^N$  and  $\hat{\psi}_1^N$ , two of the quantities which appear in the asymptotic SINR expression of Theorem 2.

The following Lemma gives expressions, which may be solved for quantities asymptotically equivalent to  $\hat{\rho}_j^N, \hat{\psi}_j^N \in \mathbb{R}^*$ ,  $j = 2, 4$ , and which occur in (31) of Theorem 2. This Lemma introduces more auxiliary random sequences, namely,  $\hat{\gamma}_j^N, \hat{\rho}_j^N, \hat{\psi}_j^N, \hat{\omega}_j^N, \hat{r}_j^N \in \mathbb{R}^*$ ,  $j = 2, 3, 4$ , plus  $\hat{\nu}_j^N, \tau_j^N \in \mathbb{R}^*$ ,  $j = 2, 3$ , which are defined in terms of matrix traces and quadratic forms in Appendix VII-A.

*Lemma 1:* In addition to the assumptions and definitions of Theorem 3, under the limit specified,  $|\hat{\gamma}_j^N - \hat{\gamma}_j| \xrightarrow{a.s.} 0$ ,  $|\hat{\rho}_j^N - \hat{\rho}_j| \xrightarrow{a.s.} 0$ ,  $|\hat{\psi}_j^N - \hat{\psi}_j| \xrightarrow{a.s.} 0$ ,  $|\hat{\omega}_j^N - \hat{\omega}_j| \xrightarrow{a.s.} 0$ ,  $|\hat{r}_j^N - \hat{r}_j| \xrightarrow{a.s.} 0$ ,  $j = 2, 3, 4$ , and  $|\hat{\nu}_j^N - \hat{\nu}_j| \xrightarrow{a.s.} 0$ ,  $|\tau_j^N - \tau_j| \xrightarrow{a.s.} 0$ ,  $j = 2, 3$ , where  $\hat{\gamma}_j, \hat{\rho}_j, \hat{\psi}_j, \hat{\omega}_j, \hat{r}_j \in \mathbb{R}^*$ ,  $j = 2, 3, 4$ , and  $\hat{\nu}_j, \tau_j \in \mathbb{R}^*$ ,  $j = 2, 3$ , are solutions to the following equations.

$$\hat{\gamma}_j = \begin{cases} z^{-\dagger} \left( \frac{\eta}{\beta} \hat{r}_2 \mathcal{W}_{1,2} - \hat{\gamma}_1 \right) & , \quad j = 2, \\ \frac{\beta^*}{\beta} (\hat{\nu}_2 \hat{\mathcal{H}}_{1,2} + |\hat{\nu}_1|^2 \hat{\tau}_2 \hat{\mathcal{H}}_{2,2}) & , \quad j = 3, \\ \frac{\alpha}{\beta} (\hat{\rho}_2 \hat{\mathcal{E}}_{1,2} + \hat{\psi}_2 |\hat{\rho}_1|^2 \hat{\mathcal{E}}_{2,2}) & , \quad j = 4, \end{cases} \quad (55)$$

$$\hat{\nu}_j = \begin{cases} |\hat{\nu}_1|^2 (1 + \sigma^2 \hat{r}_2 \mathcal{W}_{2,2}) & , \quad j = 2, \\ \sigma^2 |\hat{\nu}_1|^2 \hat{r}_3 \mathcal{W}_{2,2} & , \quad j = 3, \end{cases} \quad (56)$$

<sup>9</sup>That is, unless  $z \rightarrow 0$  is being considered (i.e., no diagonal loading), in which case it is necessary to instead consider  $\hat{\rho}_1$  and  $\hat{\omega}_1$  as the search variables, since  $\hat{r}_1$  depends only on  $\beta, \eta$ , and  $\mathcal{W}_{0,1}$  due to (50).

$$\hat{\rho}_j = \frac{\begin{cases} \beta^*(\hat{\nu}_2 \hat{\mathcal{H}}_{1,2} + \hat{\tau}_2 |\hat{\nu}_1|^2 \hat{\mathcal{H}}_{2,2}) - \alpha \hat{\psi}_2 |\hat{\rho}_1|^2 \hat{\mathcal{E}}_{1,2} & , \quad j = 2, \\ \beta^*(\hat{\nu}_3 \hat{\mathcal{H}}_{1,2} + (1 + \hat{\tau}_3) |\hat{\nu}_1|^2 \hat{\mathcal{H}}_{2,2}) - \alpha \hat{\psi}_3 |\hat{\rho}_1|^2 \hat{\mathcal{E}}_{1,2} & , \quad j = 3, \\ \alpha(\hat{\rho}_3 \hat{\mathcal{E}}_{1,2} + \hat{\psi}_3 |\hat{\rho}_1|^2 \hat{\mathcal{E}}_{2,2} - (1 + \hat{\psi}_4) |\hat{\rho}_1|^2 \hat{\mathcal{E}}_{1,2}) & , \quad j = 4, \end{cases}}{\alpha(\hat{\mathcal{E}}_{0,2} - 1) + 1}, \quad (59)$$

$$\hat{\tau}_j = \frac{\begin{cases} \alpha(\hat{\psi}_2 \hat{\mathcal{E}}_{1,2} + |\bar{w} - \hat{\psi}_1|^2 \hat{\rho}_2 \hat{\mathcal{E}}_{2,2}) - \beta^* \hat{\nu}_2 |\alpha \bar{w} - \hat{\tau}_1|^2 \hat{\mathcal{H}}_{1,2} & , \quad j = 2, \\ \alpha(\hat{\psi}_3 \hat{\mathcal{E}}_{1,2} + |\bar{w} - \hat{\psi}_1|^2 \hat{\rho}_3 \hat{\mathcal{E}}_{2,2}) - \beta^* |\alpha \bar{w} - \hat{\tau}_1|^2 (\hat{\nu}_3 \hat{\mathcal{H}}_{1,2} + |\hat{\nu}_1|^2 \hat{\mathcal{H}}_{2,2}) & , \quad j = 3, \end{cases}}{\beta^*(\hat{\mathcal{H}}_{0,2} - 1) + 1}, \quad (60)$$

For i.i.d.  $\mathbf{S}$ ,

$$\hat{\rho}_j = \begin{cases} \beta^*(\hat{\nu}_2 \hat{\mathcal{H}}_{1,2} + |\hat{\nu}_1|^2 \hat{\tau}_2 \hat{\mathcal{H}}_{2,2}) & , \quad j = 2, \\ \beta^*(\hat{\nu}_3 \hat{\mathcal{H}}_{1,2} + |\hat{\nu}_1|^2 (1 + \hat{\tau}_3) \hat{\mathcal{H}}_{2,2}) & , \quad j = 3, \\ \alpha(\hat{\rho}_3 \hat{\mathcal{E}}_{1,2} + \hat{\psi}_3 |\hat{\rho}_1|^2 \hat{\mathcal{E}}_{2,2}) & , \quad j = 4, \end{cases} \quad (57)$$

$$\hat{\tau}_j = \alpha(\hat{\psi}_j \hat{\mathcal{E}}_{1,2} + |\bar{w} - \hat{\psi}_1|^2 \hat{\rho}_j \hat{\mathcal{E}}_{2,2}) \quad , \quad j = 2, 3, \quad (58)$$

while for isometric  $\mathbf{S}$ , see (59)–(60). For i.i.d.  $\mathbf{B}$ ,

$$\hat{\psi}_j = \hat{r}_j \mathcal{W}_{2,2}, \quad (61)$$

$$\hat{\omega}_j = \begin{cases} \frac{\alpha}{\eta} (\hat{\rho}_j \hat{\mathcal{E}}_{1,2} + |\hat{\rho}_1| |\hat{\psi}_j \hat{\mathcal{E}}_{2,2}) & , \quad j = 2, 3, \\ \frac{\alpha}{\eta} (\hat{\rho}_4 \hat{\mathcal{E}}_{1,2} + |\hat{\rho}_1|^2 (1 + \hat{\psi}_4) \hat{\mathcal{E}}_{2,2}) & , \quad j = 4, \end{cases} \quad (62)$$

while for orthogonal  $\mathbf{B}$ ,

$$\hat{\psi}_j = \frac{\begin{cases} \hat{r}_j \mathcal{W}_{2,2} - \frac{\alpha}{\eta^*} |\bar{w} - \hat{\psi}_1|^2 \hat{\rho}_j \hat{\mathcal{E}}_{1,2} & , \quad j = 2, 3, \\ \hat{r}_4 \mathcal{W}_{2,2} - \frac{\alpha}{\eta^*} |\bar{w} - \hat{\psi}_1|^2 (\hat{\rho}_4 \hat{\mathcal{E}}_{1,2} + |\hat{\rho}_1|^2 \hat{\mathcal{E}}_{2,2}) & , \quad j = 4, \end{cases}}{\frac{\alpha}{\eta^*} (\hat{\mathcal{E}}_{0,2} - 1) + 1}, \quad (63)$$

$$\hat{\omega}_j = \frac{\begin{cases} \frac{\alpha}{\eta} (\hat{\rho}_j \hat{\mathcal{E}}_{1,2} + |\hat{\rho}_1|^2 \hat{\psi}_j \hat{\mathcal{E}}_{2,2}) - \sigma^2 \frac{\beta}{\eta^*} \hat{\gamma}_j |\hat{\omega}_1|^2 \mathcal{W}_{2,2} & , \quad j = 2, 3, \\ \frac{\alpha}{\eta} (\hat{\rho}_4 \hat{\mathcal{E}}_{1,2} + |\hat{\rho}_1|^2 (1 + \hat{\psi}_4) \hat{\mathcal{E}}_{2,2}) - \sigma^2 \frac{\beta}{\eta^*} \hat{\gamma}_4 |\hat{\omega}_1|^2 \mathcal{W}_{2,2} & , \quad j = 4, \end{cases}}{\frac{\eta}{\eta^*} (|\hat{\omega}_1|^2 \mathcal{W}_{2,2} - 2 \operatorname{Re}(\hat{\omega}_1 \mathcal{W}_{1,1})) + 1}. \quad (64)$$

Also,

$$\hat{\mathcal{E}}_{p,2} = \mathbf{E} \left[ \frac{P^p}{|1 - P \hat{\rho}_1 (\hat{\psi}_1 - \bar{w})|^2} \right], \quad (65)$$

$$\hat{\mathcal{H}}_{p,2} = \mathbf{E} \left[ \frac{H^p}{|1 - H \hat{\nu}_1 (\hat{\tau}_1 - \alpha \bar{w})|^2} \right], \quad (66)$$

$$\mathcal{W}_{p,2} = \mathbf{E} \left[ \frac{W^p}{|1 + W \hat{r}_1|^2} \right], \quad (67)$$

$$\hat{r}_j = \hat{\omega}_j + \sigma^2 \frac{\beta}{\eta} \hat{\gamma}_j \quad , \quad j = 1, \dots, 4, \quad (68)$$

for  $p \in \mathbb{Z}^*$ , where  $\hat{\gamma}_1, \hat{\nu}_1, \hat{\rho}_1, \hat{\tau}_1, \hat{\psi}_1, \hat{\omega}_1, \hat{r}_1, \hat{\mathcal{E}}_{q,1}, \hat{\mathcal{H}}_{q,1}$ , and  $\mathcal{W}_{q,1}$ ,  $q \in \mathbb{Z}^*$ , are determined by Theorem 3, and again  $z \rightarrow -\mu/\eta$ .

*Proof:* The proof of Lemma 1 follows the same approach as the derivation of Theorem 3. That is, the above expressions

for  $\hat{\nu}_j, \hat{\rho}_j, \hat{\tau}_j, \hat{\psi}_j, \hat{\omega}_j$ , and  $\hat{r}_j$  are derived in the same manner as the expressions for  $\hat{\nu}_1, \hat{\rho}_1, \hat{\tau}_1, \hat{\psi}_1, \hat{\omega}_1$ , and  $\hat{r}_1$  in Theorem 3, respectively. Also, as the expression for  $\hat{\gamma}_1$  in Theorem 3 is derived from the identity  $\hat{\mathbf{R}} \hat{\mathbf{R}}^{-1} = \mathbf{I}_M$ , so the expression for  $\hat{\gamma}_2$  in (55) is derived from the identity  $\hat{\mathbf{R}}^\dagger \hat{\mathbf{R}}^{-\dagger} \hat{\mathbf{R}}^{-1} = \hat{\mathbf{R}}^{-1}$ . A full derivation can be found in [26, Appendix D.7].  $\square$

Remarks:

- For exponential weighting with  $\bar{L} < \infty$ , as  $z \rightarrow \operatorname{Re}(z)$ , the required expressions for  $\mathcal{W}_{m,2}$  can be evaluated using (47), and are given by

$$\mathcal{W}_{1,2} = \frac{\bar{w}}{(1 + \hat{r}_1)(1 + e^{-\eta/\bar{L}} \hat{r}_1)}, \quad (69)$$

$$\mathcal{W}_{2,2} = \frac{1}{\hat{r}_1} (\mathcal{W}_{1,1} - \mathcal{W}_{1,2}), \quad (70)$$

where  $\bar{w} = \frac{\bar{L}}{\eta} (1 - e^{-\eta/\bar{L}})$ .

- Note that  $\hat{\mathcal{E}}_{p,2}, \hat{\mathcal{H}}_{p,2}$ , and  $\mathcal{W}_{p,2}$  are all of the form  $\mathcal{X}_{p,2} = \mathbf{E} \left[ \frac{X^p}{|1 + xX|^2} \right]$ , for which the following simple and useful identity holds.

$$\mathcal{X}_{p,1} = \mathbf{E} \left[ \frac{(1 + x^\dagger X) X^p}{|1 + xX|^2} \right] = \mathcal{X}_{p,2} + x^\dagger \mathcal{X}_{p+1,2} \quad (71)$$

for  $p \in \mathbb{Z}^*$ . This identity can be used to simplify the calculation of certain terms in Lemma 1, and also gives (70).

- These equations can be solved numerically using three zero-finding routines, two of which are for two variables, while the third is for one variable. Specifically,

- 1) First solve the subset of equations given by (68), (55), (56), and depending on the type of  $\mathbf{S}$ , (57) and (58), or (59) and (60), and depending on the type of  $\mathbf{B}$ , (61) and (62), or (63) and (64), to find  $\hat{r}_2, \hat{\gamma}_2, \hat{\nu}_2, \hat{\rho}_2, \hat{\tau}_2, \hat{\psi}_2$ , and  $\hat{\omega}_2$ . This can be done numerically using a zero-finding routine for the two variables  $\hat{r}_2$  and  $\hat{\rho}_2$ .
- 2) Solve the subset of equations given by (68), (55), (56), and depending on the type of  $\mathbf{S}$ , (57) and (58), or (59) and (60), and depending on the type of  $\mathbf{B}$ , (61) and (62), or (63) and (64), to find  $\hat{r}_3, \hat{\gamma}_3, \hat{\nu}_3, \hat{\rho}_3, \hat{\tau}_3, \hat{\psi}_3$ , and  $\hat{\omega}_3$ . This can be done numerically using a zero-finding routine for the two variables  $\hat{\omega}_3$  and  $\hat{\rho}_3$ .
- 3) Now solve the subset of equations given by (68), (55), and depending on the type of  $\mathbf{S}$ , (57) or (59),

and depending of the type of  $\mathbf{B}$ , (61) and (62), or (63) and (64), to find  $\hat{r}_4$ ,  $\hat{\gamma}_4$ ,  $\hat{\rho}_4$ ,  $\hat{\psi}_4$ , and  $\hat{\omega}_4$ . This can be done numerically with a zero-finding routine for one variable, namely  $\hat{\psi}_4$ .

- The solution (55) is one of many possible expressions, which can be derived from the identity  $\frac{1}{M}\text{tr}[\hat{\mathbf{R}}^\dagger \hat{\mathbf{R}}^{-1} \hat{\mathbf{R}}^{-1}] = \frac{1}{M}\text{tr}[\hat{\mathbf{R}}^{-1}]$ . All possible expressions involving  $\hat{\gamma}_2$  derived in this manner include

$$\hat{\gamma}_1 + z^\dagger \hat{\gamma}_2 = \frac{\eta}{\beta} \hat{r}_2 \mathcal{W}_{1,2}, \quad (72)$$

$$\hat{\gamma}_2 \hat{\nu}_1 - \hat{\gamma}_1 \hat{\nu}_2 = \frac{\alpha}{\beta} ((\hat{\psi}_1 - \bar{w})^\dagger \hat{\rho}_2 + \hat{\rho}_1 \hat{\psi}_2) |\hat{\nu}_1|^2 \hat{\mathcal{E}}_{1,2} \quad (73)$$

$$= \frac{\beta^*}{\beta} ((\hat{\tau}_1 - \alpha \bar{p} \bar{w})^\dagger \hat{\nu}_2 + \hat{\nu}_1 \hat{\tau}_2) |\hat{\nu}_1|^2 \hat{\mathcal{H}}_{1,2}. \quad (74)$$

When solving the set of equations in Lemma 1, both of these expressions are more useful than (55) when considering  $z \rightarrow 0$  (i.e. no diagonal loading).

### B. Steady-State ALS SINR

We now determine the steady-state ALS SINR, that is, the SINR as the number of training intervals  $\eta \rightarrow \infty$  (either with a training sequence, or semi-blind), from the transient ALS SINR expressions in Section V-A. Of course, with no windowing (i.e.,  $\mathbf{W} = \mathbf{I}_i$ ), the output SINR converges to that of an MMSE receiver with full CSI. We first verify this result, and then turn to the more interesting case of data windowing and (optionally) diagonal loading. We will see that the steady-state response is the same for both i.i.d. and orthogonal training sequences, which matches intuition, since i.i.d. training sequences become orthogonal as  $\eta \rightarrow \infty$ .

An approximate analysis of the steady-state performance of the ALS receiver with exponential windowing was presented in [9] for DS-CDMA with flat fading. The large-system steady-state ALS performance is considered in [16, Corollary 2]. In [16], results from asymptotic analysis of reduced rank filters are used, which rely on arguments related to non-crossing partitions. Here we give a more direct derivation of the large-system steady-state performance of the ALS receiver for the general transmission model (1).

Strictly speaking, Theorems 1 and 3 require  $z \in \mathbb{C}^+$ , however for the following discussion we shall implicitly consider  $z \rightarrow -\sigma^2$  and  $z \rightarrow -\mu/\eta$ , respectively.

1) *No Windowing*: We first consider the limit of the equations in Theorem 3 as  $\eta \rightarrow \infty$ , and show that without windowing (i.e.,  $\mathbf{W} = \mathbf{I}_i$ ) the ALS SINR converges to the MMSE SINR.

First note that  $\hat{\omega}_1 \rightarrow 0$  for both i.i.d. and orthogonal  $\mathbf{B}$ , which means  $\mathcal{W}_{p,q} \rightarrow 1$ , and therefore  $\hat{\psi}_1 \rightarrow 0$  and  $\hat{\nu}_1 \rightarrow 1/\sigma^2$ . Moreover,  $\hat{\mathcal{E}}_{p,q} \rightarrow \mathcal{E}_{p,q}$  and  $\hat{\mathcal{H}}_{p,q} \rightarrow \mathcal{H}_{p,q}$  from Theorem 1. We see that the expressions for the ALS moments  $\hat{\gamma}_1$ ,  $\hat{\rho}_1$  and  $\hat{\tau}_1$  from Theorem 3 converge to the MMSE moments  $\gamma_1$ ,  $\rho_1$ , and  $\tau_1$ , respectively, in Theorem 1 at  $z \rightarrow -\sigma^2$  as  $\eta \rightarrow \infty$ .

Now consider the limit of the equations in Lemma 1 with no windowing as  $\eta \rightarrow \infty$ . Clearly,  $\hat{\omega}_2$ ,  $\hat{\omega}_3$ , and  $\hat{\omega}_4$  all  $\rightarrow 0$ , and hence also  $\hat{\psi}_2$ ,  $\hat{\psi}_3$ , and  $\hat{\psi}_4$  also  $\rightarrow 0$  for either i.i.d. or orthogonal  $\mathbf{B}$ . It follows that  $\hat{\nu}_2 \rightarrow (1/\sigma^2)^2$  and  $\hat{\nu}_3 \rightarrow 0$ .

Substituting the preceding limits into (57)–(60), we see immediately that as  $\eta \rightarrow \infty$ , the variables  $\hat{\rho}_2$ ,  $\hat{\rho}_3$ ,  $\hat{\rho}_4$ ,  $\hat{\tau}_2$ , and  $\hat{\tau}_3$  satisfy the same set of equations as  $\rho_2$ ,  $\rho_3$ ,  $\rho_4$ ,  $\tau_2$ , and  $\tau_3$  for the MMSE receiver, which appear in the SINR expression in Section IV-B, given by (25)–(28). Therefore,  $\hat{\rho}_4 + \sigma^2 \hat{\rho}_2 \rightarrow \rho_1$  as  $\eta \rightarrow \infty$ , and therefore the ALS SINR converges to the MMSE SINR as  $\eta \rightarrow \infty$ . The diagonal loading constant  $\mu/\eta$  disappears in this limit.

2) *Fixed-Length Data Windowing*: The following results apply to fixed-length windowing functions. That is, the large-system window size does scale with  $\eta$ .<sup>10</sup> For example, in the case of exponential windowing this corresponds to fixed  $\bar{L}$ . More precisely, we define fixed-length windowing as

$$\lim_{\eta \rightarrow \infty} \eta \mathbf{E}[W_\eta] > 0, \quad (75)$$

$$\lim_{\eta \rightarrow \infty} F_{W_\eta}(w) = \begin{cases} 0 & , w < 0, \\ 1 & , w \geq 0. \end{cases} \quad (76)$$

where, for given  $\eta$ ,  $W_\eta$  denotes a scalar r.v. with distribution  $F_{W_\eta}(w)$ , given by the (compactly supported) a.e.d. of  $\mathbf{W}$ . In other words,  $F_{W_\eta}$  converges in distribution to a delta-distribution at zero as  $\eta \rightarrow \infty$ , with the mean of  $W_\eta$  of order  $\eta^{-1}$ . For example, for exponential windowing, the mean window size is  $\bar{w} = \frac{\bar{L}}{\eta}(1 - e^{-\eta/\bar{L}})$ .

We have the following corollary to Theorem 2, Theorem 3, and Lemma 1, which specifies the steady-state ALS SINR with fixed-length windowing.

*Corollary 1*: Under the limit specified in Theorem 2, and also as  $\eta \rightarrow \infty$ , provided all asymptotic moments exist, the asymptotic steady-state SINR for the ALS receiver for stream  $k$  with fixed-length windowing is given by the asymptotic ALS SINR specified in Theorems 2, 3, and Lemma 1, where

- the i.i.d.  $\mathbf{B}$  relations are used for both i.i.d. and ort.  $\mathbf{B}$ ,
- $z$  is replaced by  $-\mu$ ,
- $\eta$  is replaced by one, and,
- the variables  $\bar{w}$  and  $\mathcal{W}_{p,q}$  are replaced by  $\tilde{w}$  and  $\tilde{\mathcal{W}}_{p,q}$ , where

$$\tilde{w} = \lim_{\eta \rightarrow \infty} \eta \bar{w}, \quad (77)$$

$$\tilde{\mathcal{W}}_{p,q} = \lim_{\eta \rightarrow \infty} \eta \mathcal{W}_{p,q}. \quad (78)$$

*Proof*: See Appendix IX.  $\square$

Remark:

- With exponential windowing,  $\tilde{w} = \bar{L}$ , and from (78) we have

$$\tilde{\mathcal{W}}_{1,1} = \frac{\beta(1 - \mu \tilde{\gamma}_1)}{\tilde{r}_1}, \quad (79)$$

$$\tilde{\mathcal{W}}_{1,2} = \frac{\bar{L}}{1 + \tilde{r}_1}, \quad (80)$$

$$\tilde{\mathcal{W}}_{2,2} = \frac{1}{\tilde{r}_1} \left( \tilde{\mathcal{W}}_{1,1} - \tilde{\mathcal{W}}_{1,2} \right), \quad (81)$$

where  $\tilde{r}_1 = \exp(\frac{\beta}{\bar{L}}(1 - \mu \tilde{\gamma}_1)) - 1$ , and we have used (50) to obtain (79).

<sup>10</sup>Of course, the actual window size should increase with  $N$  in order to define a meaningful large-system limit, as explained in Section II-E. Here we are referring to the large-system window size *after* the large-system limit has been determined. With exponential windowing this is the difference between  $L$  and  $\bar{L}$ .

## VI. RELATIONSHIP BETWEEN MMSE AND ALS RECEIVERS

In this section we present a simple relationship between the SINRs of the MMSE and ALS receivers given in the previous sections. We note that this relationship has recently been studied in the special case  $\mathbf{H} = \mathbf{I}_N$ , i.i.d.  $\mathbf{S}$ , i.i.d. training, and no diagonal loading. An approximate relation was given for rectangular windowing in [17], with a corresponding exact large-system expression given in [16, Corollary 1]. Also, [9] obtained approximate expressions for the steady-state SINR relationship ( $\eta \rightarrow \infty$ ) with exponential windowing.

### A. Transient Response

The following theorem applies to any  $\mathbf{H}$ , both i.i.d. and isometric  $\mathbf{S}$ , and any windowing shape. The only restrictions are that there is i.i.d. training and no diagonal loading (i.e.,  $\mu = 0$ ). The theorem relates the expressions in Theorem 3 and Lemma 1 to the alternate MMSE SINR expression of Section IV-B.

*Theorem 4:* For the  $k^{\text{th}}$  data stream, the asymptotic SINR of the full-CSI MMSE receiver  $\text{SINR}_k^{\text{MMSE}}$  is related to the asymptotic SINR of the ALS receiver with i.i.d. training sequences, data windowing, and no diagonal loading,  $\text{SINR}_k^{\text{ALS}}$ , according to

$$\text{SINR}_k^{\text{ALS}} = \frac{\text{SINR}_k^{\text{MMSE}}}{\zeta + \frac{\zeta - 1}{\text{SINR}_k^{\text{MMSE}}}} \quad (82)$$

with training sequences, and

$$\text{SINR}_k^{\text{ALS}} = \frac{\text{SINR}_k^{\text{MMSE}}}{\zeta + (\zeta - 1)\text{SINR}_k^{\text{MMSE}}} \quad (83)$$

for semi-blind training, with either i.i.d. or isometric  $\mathbf{S}$ , where

$$\zeta = \frac{\mathcal{W}_{1,1}}{\mathcal{W}_{1,2}} \quad (84)$$

which depends *only* on  $\eta$ ,  $\beta$ , and the window shape.

*Proof:* See Appendix X.  $\square$

Remarks:

- To calculate  $\zeta$ , note that from (50) we have  $\mathcal{W}_{0,1} = 1 - \frac{\beta}{\eta}$  for any window shape. Since  $\mathcal{W}_{0,1}$  is a fixed known function of  $\hat{r}_1$ , we can invert this equation to find  $\hat{r}_1$ . For example, with exponential windowing, we obtain from (48) that

$$\hat{r}_1 = \frac{e^{\beta/\bar{L}} - 1}{1 - e^{(\beta-\eta)/\bar{L}}} \quad (85)$$

and with rectangular windowing,  $\hat{r}_1 = \beta/(\eta - \beta)$ . Given  $\hat{r}_1$ , we can directly calculate  $\mathcal{W}_{p,q}$  from the definition given in (45), and  $\zeta$  from (84). The point here is that  $\hat{r}_1$ ,  $\mathcal{W}_{p,q}$ , and  $\zeta$  are essentially constants, depending only on  $\beta$ ,  $\eta$ , and the window shape.

- Remarkably, Theorem 4 implies that  $\text{SINR}_k^{\text{ALS}}$  only depends on  $\beta$ ,  $\eta$ , the window shape, and  $\text{SINR}_k^{\text{MMSE}}$ . That is, *the convergence rate of the ALS SINR to the steady-state value is independent of the channel* (of course, the steady-state value itself depends on the channel). Stating this another way, for a particular  $\beta$  and window shape,

all channels, which have the same MMSE SINR, have an identical transient ALS SINR response. ‘Channel’ here refers to the product  $\mathbf{H}\mathbf{S}\mathbf{A}$ . This has been observed in [9, 11], although a transient SINR relationship, such as that given in Theorem 4, has not previously been determined.

- Recall that due to Proposition 1, Theorem 4 also holds for the general AWGN model  $\mathbf{r} = \mathbf{C}\mathbf{b} + \mathbf{n}$  for which  $\mathbf{b}$  is unitarily invariant, and the eigenvalues of  $\mathbf{C}^\dagger$  are well behaved. In that sense, *Theorem 4 is a fundamental property of linear estimation*.
- In fact, our assumption that the additive noise  $\mathbf{n}_m$  is i.i.d. complex Gaussian distributed is an unnecessary restriction, as all results presented hold for any distribution such that  $\mathbf{n}_m$  is unitarily invariant, and the elements of  $\mathbf{n}_m$  are i.i.d. with zero mean and variance  $\sigma^2$ .
- With exponential windowing we have

$$\zeta = \frac{\beta(1 - e^{-\eta/\bar{L}})}{\bar{L}(1 - e^{(\beta-\eta)/\bar{L}})(1 - e^{-\beta/\bar{L}})} \quad (86)$$

and for rectangular windowing, we have  $\zeta = 1 + \beta/(\eta - \beta)$ . With rectangular windowing and  $\beta = 1$  (i.e., square  $\mathbf{H}$ ) this matches the expression derived in [16] for DS-CDMA with i.i.d. signatures in flat fading.

Unfortunately, we do not have a compact expression, analogous to Theorem 4, which relates the ALS and MMSE SINRs with orthogonal training sequences and/or diagonal loading, although it seems likely that such a relationship exists. What we can say is that with orthogonal training sequences and no diagonal loading the moments of Theorem 1 and Theorem 3 are related via

$$\hat{\rho}_1 = \frac{D_{\mathbf{B}}}{\mathcal{W}_{1,1}} \rho_1, \quad (87)$$

$$\hat{\gamma}_1 = \frac{1}{\mathcal{W}_{1,1}} \gamma_1, \quad (88)$$

$$\hat{\tau}_1 = \frac{\mathcal{W}_{1,1}}{D_{\mathbf{B}}} (\tau_1 - \alpha\bar{p}) + \alpha\bar{p}, \quad (89)$$

where  $D_{\mathbf{B}} = 1 - \frac{\alpha}{\eta} \rho_1 \mathcal{E}_{1,1}$ . Interestingly, these relationships depend on the channel through  $D_{\mathbf{B}}$  (which was not the case with i.i.d. training). Finding a complete compact relation for the SINRs with orthogonal training and/or diagonal loading remains an open problem.

### B. Steady-State Response

For the steady-state response ( $\eta \rightarrow \infty$ ) with fixed-length data windowing, Theorem 4 holds with

$$\zeta = \frac{\tilde{\mathcal{W}}_{1,1}}{\tilde{\mathcal{W}}_{1,2}} \quad (90)$$

where  $\tilde{\mathcal{W}}_{p,q} = \lim_{\eta \rightarrow \infty} \eta \mathcal{W}_{p,q}$ . This result is proved simply by letting  $\eta \rightarrow \infty$  in Theorem 4. Of course, this steady-state relationship also holds for orthogonal training sequences.

With exponential weighting, we have from (86) that  $\zeta = \frac{\beta}{\bar{L}(1 - e^{-\beta/\bar{L}})}$ . Note that as  $\bar{L} \rightarrow \infty$  (i.e., as we increase the window size)  $\zeta \rightarrow 1$  (using L’Hôpital’s rule), and  $\text{SINR}_k^{\text{ALS}} \rightarrow \text{SINR}_k^{\text{MMSE}}$ , as expected.

In [9], similar approximate relationships were derived for the steady state performance of the ALS receiver with exponential windowing for DS-CDMA in flat fading. The equivalent value of  $\zeta$  there is  $\zeta' = \frac{1-\epsilon}{2\epsilon}(N-1)$ , which converges to  $\zeta' \rightarrow \frac{1}{2\bar{L}} + 1$  as  $N \rightarrow \infty$  after substituting  $\epsilon = 1 - \frac{1}{N\bar{L}}$ . Fig. 1 shows a plot of this approximation, which is quite accurate when compared to the exact large-system value at  $\beta = 1$ , particularly for large  $\bar{L}$ .

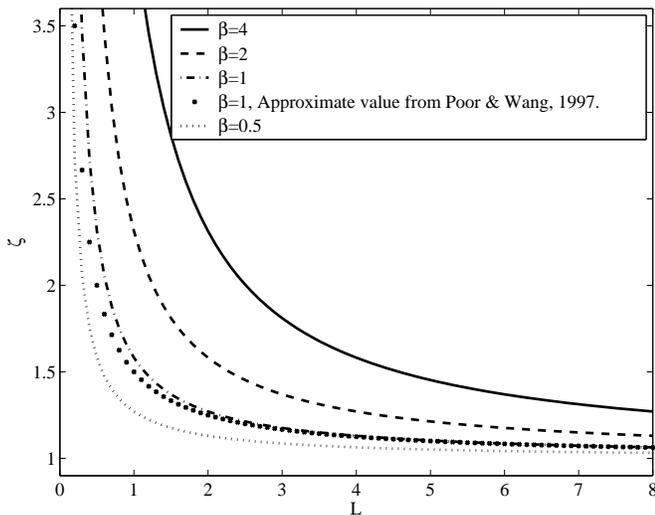


Fig. 1.  $\zeta$  vs.  $\bar{L}$  for  $\eta = \infty$  and a range of  $\beta$ , from (84).

### C. Capacity Relationship

Consider the difference in capacity per data stream<sup>11</sup> of the MMSE and the ALS receivers, defined as  $\Delta C_{\text{ALS}}^k = \log(1 + \text{SINR}_k^{\text{MMSE}}) - \log(1 + \text{SINR}_k^{\text{ALS}})$ . We have from (82) and (83)

$$\Delta C_{\text{ALS}}^k = \ln \left( 1 + \left( 1 - \frac{1}{\zeta} \right) \text{SINR}_k^{\text{MMSE}} \right) \quad (91)$$

with semi-blind training, and

$$\Delta C_{\text{ALS}}^k = \ln \left( 1 + \left( \frac{1 - \frac{1}{\zeta}}{\frac{\text{SINR}_k^{\text{MMSE}} - 1}{\zeta} + 1} \right) \text{SINR}_k^{\text{MMSE}} \right) \quad (92)$$

with training sequences.

As  $\zeta \rightarrow 1$ , the capacity difference approaches zero, whereas as  $\zeta \rightarrow \infty$ , the difference approaches  $\ln(1 + \text{SINR}_k^{\text{MMSE}})$ . Recall that  $\zeta$  only depends on  $\eta$  (the ratio of training symbols to transmit dimensions),  $\beta$  (the ratio of receive to transmit dimensions), and the window shape, and does not depend on the SNR, the (normalized) number of data streams  $\alpha$ , or the channel distribution. Nor does this value depend on whether  $\mathbf{S}$  is i.i.d. or isometric. Fig. 1 shows the steady-state value of  $\zeta$  vs.  $\bar{L}$  with exponential windowing and a range of  $\beta$  values.

<sup>11</sup>That is, we are assuming each data stream is independently coded and decoded. Also, we are assuming that the residual multi-access interference (MAI) is Gaussian, which one would expect to be valid in the asymptotic limit considered due to the central limit theorem.

## VII. NUMERICAL STUDIES

We now present various applications of the results presented in previous sections. We shall focus on three example systems:

- The first is the standard model of a MIMO channel with rich scattering, for which we set  $\mathbf{H} = \mathbf{I}_N$  and  $\mathbf{S}$  i.i.d., so that  $K$  and  $N$  represent the number of transmit and receive antennas, respectively.
- The second example system is CDMA in frequency-selective Rayleigh fading, for which  $\mathbf{S}$  contains either i.i.d. or isometric signatures, and  $\mathbf{H}$  is a square  $N \times N$  matrix (hence  $\beta = 1$ ), for which the a.e.d. of the channel correlation matrix  $\mathbf{H}^\dagger$  is exponential with mean one (i.e.,  $f_H(h) = \exp(-h)$  for  $h > 0$  is the density used to compute the  $\mathcal{H}_{p,q}$  values).
- The third example system is a SISO FIR channel with a cyclic prefix, as described after Proposition 1 in Section II-E, where  $\mathbf{h} = [0.227, 0.46, 0.688, 0.46, 0.227]^\dagger$  (i.e., where the ALS and MMSE receiver is used to equalize the so-called Proakis Channel-C [27, pp. 616]). That is, the empirical results will be obtained using  $\mathbf{H}$  given by the circulant matrix obtained from  $\mathbf{h}$ , and  $\mathbf{S} = \mathbf{A} = \mathbf{I}_N$ . As described in Proposition 1, the analytic results are obtained from the isometric  $\mathbf{S}$  equations with  $\alpha = 1$ , and  $H$  distributed according to the spectra of  $\mathbf{h}$ , i.e.,  $F_H(h) = \frac{1}{N} \sum_{n=1}^N u(h - Q(n))$ , where  $u(t)$  is the step function, and  $Q(n) = |\text{DFT}_N^n(\mathbf{h})|^2$ , where  $\text{DFT}_N^n(\mathbf{h})$  denotes the  $n^{\text{th}}$  element of the  $N$ -point discrete Fourier transform of  $\mathbf{h}$ .

Unless otherwise stated, we shall assume equal transmit power per data stream (i.e.,  $\mathbf{A} = \mathbf{I}_K$ ), and SNR= 10 dB, where SNR is defined as the energy transmitted per data stream in each symbol interval, divided by  $\sigma^2$ .

In the following plots we determine empirical values from averages of a size  $N = 32$  system with QPSK modulation for comparison with the large-system results. The asymptotic values for the MMSE curves have been determined from Theorem 1, and the asymptotic values for the ALS curves have been determined from Theorem 2, Theorem 3, and Lemma 1. The steady-state values of the ALS receiver have been determined from Corollary 1. Where possible, the ALS SINR has been determined from the MMSE SINR using Theorem 4 (i.e., any situation with i.i.d. training sequences and no diagonal loading).

### A. Transient ALS SINR response and comparison with empirical values

1) *MIMO example*: Firstly, we demonstrate the relevance of the large-system limit to practical finite systems. Fig. 2 shows both asymptotic and empirical values of MMSE and ALS SINR vs. training length for the example MIMO system with rich scattering. For the ALS receiver, the diagonal loading value is  $\mu = 0.1$ , and rectangular windowing is used. Clearly, the empirical (finite) values match the analytic (asymptotic) values very closely.

Note that the orthogonal training sequences clearly outperform the i.i.d. training sequences, particularly for ‘small’  $\eta$ . This gap also widens as the number of receive dimensions

decreases. Also, the performance of the semi-blind ALS receiver is comparable to the performance of the ALS receiver with training for the 2 to 1 transmit to receive antennas ratio case, but is significantly worse in the 1 to 2 transmit to receive antennas ratio case.

2) *CDMA in frequency-selective fading*: Fig. 3 shows empirical and asymptotic values of MMSE and ALS SINR vs.  $\eta$  for the example CDMA system in frequency-selective Rayleigh fading with  $\alpha = 0.50$ . The ALS receiver uses rectangular windowing and a diagonal loading constant  $\mu = 0.1$ . Curves are shown for both i.i.d. and isometric signatures, and i.i.d. and orthogonal training sequences. Again, the empirical (finite) values match the analytic (asymptotic) values.

Figure 3(b) shows the intuitively pleasing result that for a small number of training symbols (i.e., small  $\eta$ ), orthogonal training sequences improve the performance of the ALS receiver more than isometric signatures, and as  $\eta$  increases, this situation is quickly reversed. This is due to the fact that the  $K$  i.i.d. training sequences of length  $i$  become ‘more orthogonal’ as  $i$  increases, and also since isometric signatures consistently outperform i.i.d. signatures.

In subsequent plots, we shall omit the empirical values, and concentrate on applications of the analytical results.

3) *Equalization*: Fig. 4 shows empirical and asymptotic values of MMSE and ALS SINR vs.  $\eta$  for the example SISO FIR system at 20dB SNR. The ALS receiver uses exponential windowing and a diagonal loading constant  $\mu = 0.1$ . Curves are shown for both i.i.d. and orthogonal training sequences. Note that Proposition 1 requires that  $\mathbf{b}_m$  is unitarily invariant, whereas the empirical values in the figure are based on standard QPSK modulation (i.e.,  $\mathbf{b}_m$  is not unitarily invariant). Clearly, at least in this case, the asymptotic results are a very good approximation for non-unitarily invariant data vectors.

We remark that, in practice, an adaptive LS equalizer for a SISO FIR channel would process the inputs sequentially, as opposed to the block processing assumed here. Hence, the convergence to steady-state would require much less training than what is shown in Fig. 4. Characterizing the transient performance of such an adaptive equalizer remains an open problem.

### B. Capacity with exponential windowing

Now we examine the performance of the ALS receiver with both rectangular and exponential windowing, relative to the MMSE receiver. Fig. 5 shows the capacity difference per-signature as a function of the window size  $\bar{L}$  (determined from (91) and (92)) for the example CDMA system in frequency-selective Rayleigh fading with i.i.d. signatures, i.i.d. training, and a system load of  $\alpha = 0.75$ . Curves for the ALS receiver are shown with both rectangular and exponential windowing, and diagonal loading constant  $\mu = 0$ . Also,  $f_P(p) = \frac{3}{4}\delta(p - 1) + \frac{1}{4}\delta(p - \frac{1}{2})$ , that is, one quarter of the signatures are transmitted at half the power of the remaining signatures.

In this figure, we do not take into account the loss in rate due to the training. This is considered in the following subsection. Rather, for a single channel use at a certain SNR, we wish to see the relative capacity of the MMSE receiver (using full

CSI) and the ALS receiver, as a function of the number of training symbols used to generate the filter. Also, for the ALS receiver, we wish to compare exponential windowing with rectangular windowing at a given value of  $\eta$  as a function of the exponential windowing window size,  $\bar{L}$ .

Firstly, we see that for either type of windowing, increasing the number of training symbols is an exercise in diminishing returns. Also, we see that as the window size increases, exponential windowing asymptotes to rectangular windowing, as would be expected for the time-invariant system model (1). Of course, exponential windowing is included to allow for time-varying channels. As such, the curves for exponential windowing are a valid approximation for a time-varying system in which the coherence time of the system<sup>12</sup> is at least as large as the effective window size created by the exponential windowing. As such, the values of capacity or SINR obtained represent *best possible* values, which are only attained if the system remains static for the duration of the ALS training period. Extending these results to time-varying systems is an open problem, and is likely to be difficult.

### C. Application: Throughput Optimization

We now demonstrate how the results can be used to optimize the throughput with packet transmissions. More training symbols gives a higher ALS SINR, but leaves less room for data-carrying symbols in the packet. Clearly, there is an optimal ratio of training symbols to data-carrying symbols. Such an optimization has been considered for MIMO block fading channels and SISO FIR channels in [28,29] with an optimal (maximum-likelihood) receiver. In that work, the training symbols are used to estimate the channel directly. A lower bound on the capacity is derived, and is used to optimize the training length. Related work in [30] applies the large-system transient analysis in [16] for the MIMO i.i.d. channel to optimize the training length with an ALS receiver (without exponential windowing or diagonal loading). It is shown there that for large packet lengths ( $\ell$ ) the training length that maximizes capacity grows as  $O(\sqrt{\ell})$ . Optimization of power levels between the training and data symbols is also investigated.

Suppose we consider a packet containing  $T > i$  symbols, of which the first  $i$  are training symbols, and the remainder consists of data-carrying symbols. There are  $K$  equal power data streams, which are coded independently with capacity-achieving<sup>13</sup> codes with rate  $R_c = \log_2(1 + \text{SINR}^{\text{ALS}})$ . We focus on the ALS receiver with known training symbols. The number of information bits per block is therefore  $KR_c(T - i)$ , while the number of transmit dimensions per block is  $NT$ . Therefore, the number of information bits per transmit dimension (hereafter referred to as ‘normalized capacity’) is  $C = \alpha R_{\text{eff}}$ , where  $R_{\text{eff}} = R_c(1 - \eta/\ell)$  and  $\ell = T/N$ . We shall consider the additional limit  $T \rightarrow \infty$  with  $T/N \rightarrow \ell > 0$  in order to optimize  $C$  with respect to the normalized training

<sup>12</sup>‘Coherence time’ here refers to the number of symbols over which **HSA** and  $\sigma^2$  remain approximately constant.

<sup>13</sup>Here we assume that the residual interference at the receiver output is i.i.d. circularly symmetric complex Gaussian.

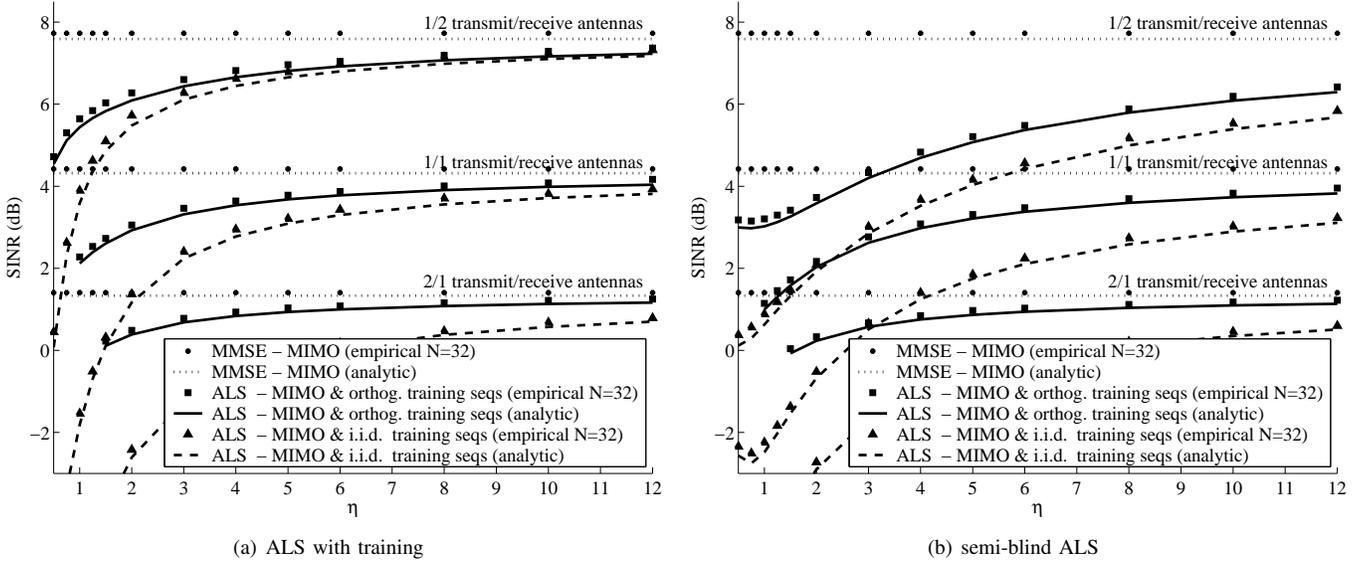


Fig. 2. SINR vs. training length: (Rich MIMO) equal transmit power per antenna, SNR= 10dB,  $\mu = 0.1$ , rectangular windowing. Comparison with empirical values  $N = 32$ , QPSK modulation.

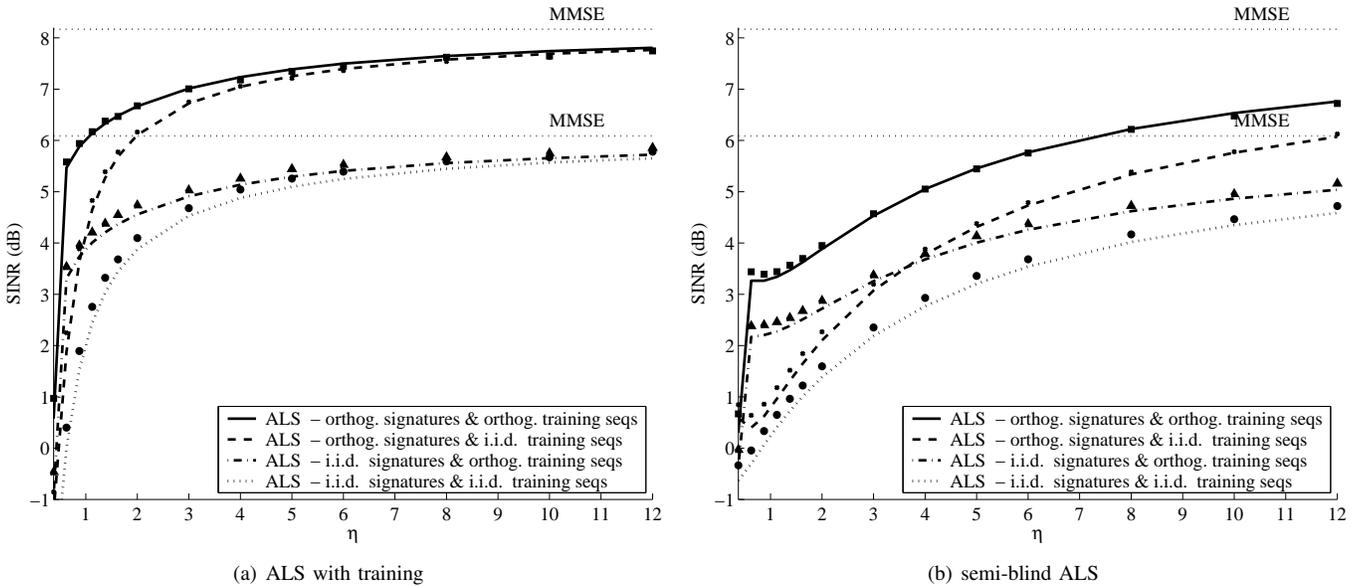


Fig. 3. SINR vs. training length: CDMA in frequency-selective fading, SNR=10dB,  $\alpha = 0.50$ ,  $\mu = 0.1$ , rectangular windowing, equal power per signature, exponential distribution for a.e.d. of  $\mathbf{H}^{\dagger}$ . Comparison with empirical values  $N = 32$ , QPSK modulation.

length  $\eta$ . We keep  $E_b/\sigma^2 = \text{SNR}/R_{\text{eff}}$  constant, and unless otherwise stated, in the numerical examples  $E_b/\sigma^2 = 10$  dB.

Fig. 6(a) shows the normalized capacity of the example CDMA system in frequency-selective fading as a function of  $\eta$  and  $\alpha$  for a normalized block length of  $\ell = T/N = 15$ . The ALS receiver uses rectangular windowing and no diagonal loading. Fig. 6(b) shows the additional normalized capacity obtained, relative to the results for i.i.d. training in Fig. 6(a), if orthogonal training sequences are used. Although not shown, plots analogous to Fig. 6 may also be obtained for iso. S.

Fig. 6(a) shows that there is an optimum value of  $\eta/\ell$ , i.e., the ratio of training length to block length, for each value of system load,  $\alpha$ . Fig. 7(a) shows the value of normalized

capacity at the optimum value of  $\eta$ , again for  $\ell = 15$ , as a function of the system load  $\alpha$ . Fig. 7(b) shows the corresponding value of  $\eta$  (expressed as a percentage of  $\ell$ ) which maximizes the normalized capacity. Also shown in Fig. 7(a) is the normalized capacity of the MMSE receiver at the same value of  $E_b/\sigma^2$ , for both types of signatures. Of course, the MMSE receiver assumes perfect CSI.

We now consider throughput optimization for the example MIMO system, and consider the growth in normalized capacity with respect to the normalized block length,  $\ell$ . In this example, since  $K$  represents the number of transmit dimensions, the number of transmit dimensions per block is  $KT$ , and hence the number of information bits per transmit dimension is  $C =$

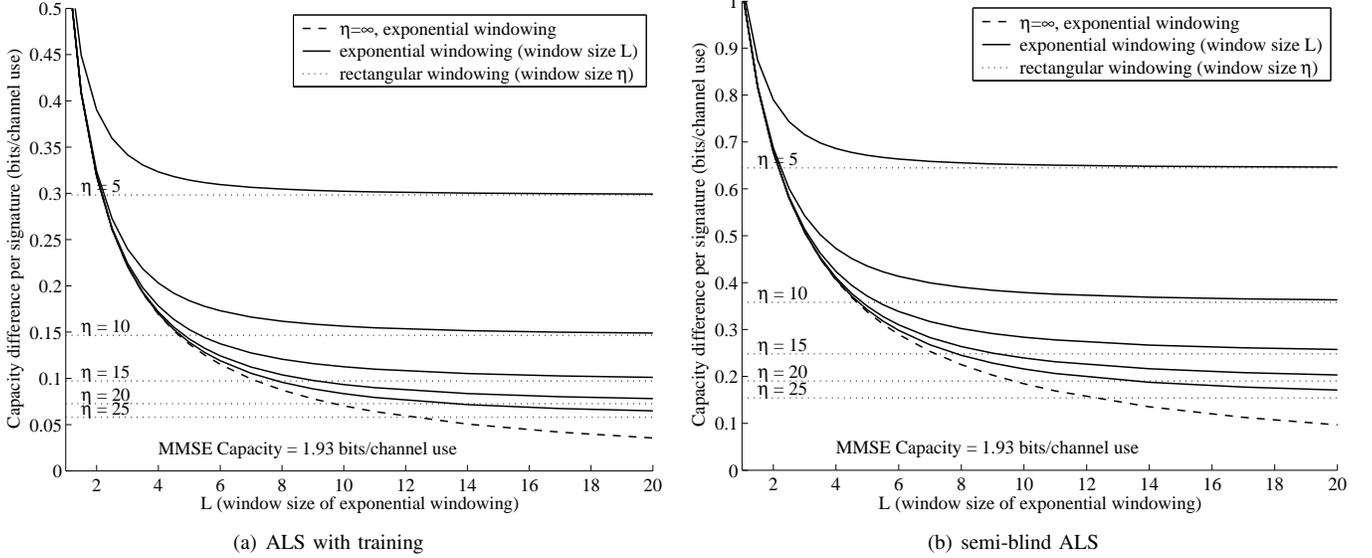


Fig. 5. Capacity difference between ALS and MMSE receivers vs. window size of exponential window: CDMA in frequency-selective fading, i.i.d. signatures, i.i.d. training, SNR=10dB,  $\alpha = 0.75$ ,  $\mu = 0$ , rectangular and exponential windowing. Note the window size for rectangular windowing is  $\eta$ . Also,  $f_P(p) = \frac{3}{4}\delta(p-1) + \frac{1}{4}\delta(p-\frac{1}{2})$ . Curves shown correspond to the first 3/4 of the signatures. Note that the scale of the vertical axis of Fig. 5(b) is twice that of Fig. 5(a).

$R_{\text{eff}}$ . Figure 8(a) shows the growth in normalized capacity, optimized with respect to  $\eta/\ell$ . These results show that the gain in using orthogonal training sequences appears to be more pronounced in situations where there is a high ratio of transmit antennas to receive antennas (i.e.,  $\alpha > 1$ ).

Figures 8(b) and 9 show the optimal value of training length  $\eta$  corresponding to Figure 8(a), for i.i.d. and orthogonal training sequences, respectively. In Figure 8(b),  $\eta$  is expressed as a percentage of the block length  $\ell$ . In Figure 9, it is interesting to note the case  $\alpha = 4$ , where we see that  $\eta < 4$

is never optimal. Recall that for  $\alpha > \eta$ , we have orthogonal rows of  $\mathbf{B}$  and for  $\alpha < \eta$ , we have orthogonal columns of  $\mathbf{B}$ . Clearly, orthogonal columns are preferable. If the axis were extended, we would see the same behavior in the other curves.

## VIII. CONCLUSIONS

Determining the transient behavior of ALS algorithms with random inputs is a classical problem, which is relevant to many communications applications, such as equalization and interference suppression. The large-system results presented here are the first set of *exact* results, which characterize the transient performance of ALS algorithms for a wide variety of channel models of interest. Namely, our results apply to any linear input-output model (see Proposition 1), where the input is unitarily invariant, and the channel matrix has a well-defined a.e.d. with finite moments. As such, these results can be used to evaluate adaptive equalizer performance in the context of space-time channels. This represents a significant generalization of the previous large-system results in [16], which apply only to an i.i.d. channel matrix. Furthermore, the analytical approach relies only on elementary matrix manipulations, and is general enough to allow for orthonormal spreading and/or training sequences, in addition to i.i.d. sequences. Numerical results were presented, showing that orthogonal training sequences can significantly outperform i.i.d. training sequences.

For the general ALS algorithm and model considered, the output SINR can be expressed as the solution to a set of nonlinear equations. These equations are complicated by the fact that they depend on a number of auxiliary variables, each of which is a particular large matrix moment involving the sample covariance matrix. Still, it is relatively straightforward to solve these equations numerically. Illustrative examples were presented showing the effect of training length on the capacity of a block fading channel.

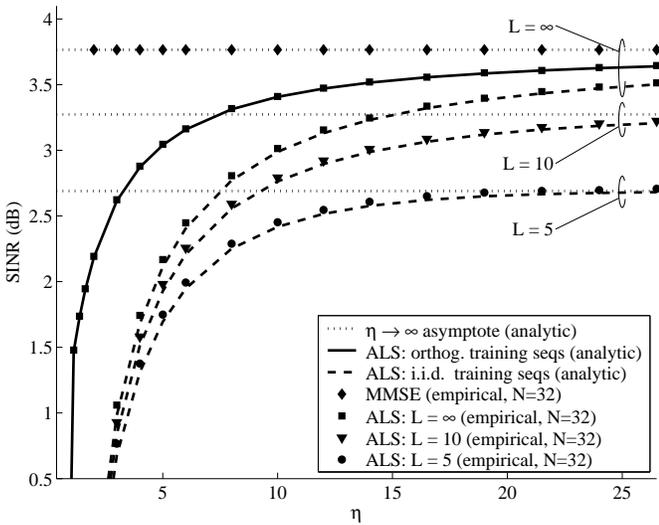


Fig. 4. SINR vs. training length: Equalization of Proakis C-channel (i.e.,  $\mathbf{H}$  circulant matrix constructed from impulse response, and  $\mathbf{S} = \mathbf{A} = \mathbf{I}_N$ ) at SNR=20dB. The ALS receiver uses  $\mu = 0.1$ , exponential windowing, with training using i.i.d. and orthogonal training sequences. Comparison with empirical values  $N = 32$ , QPSK modulation. Analytic values are obtained using isometric  $\mathbf{S}$  equations with  $\alpha = \beta = 1$ , as specified by Proposition 1.

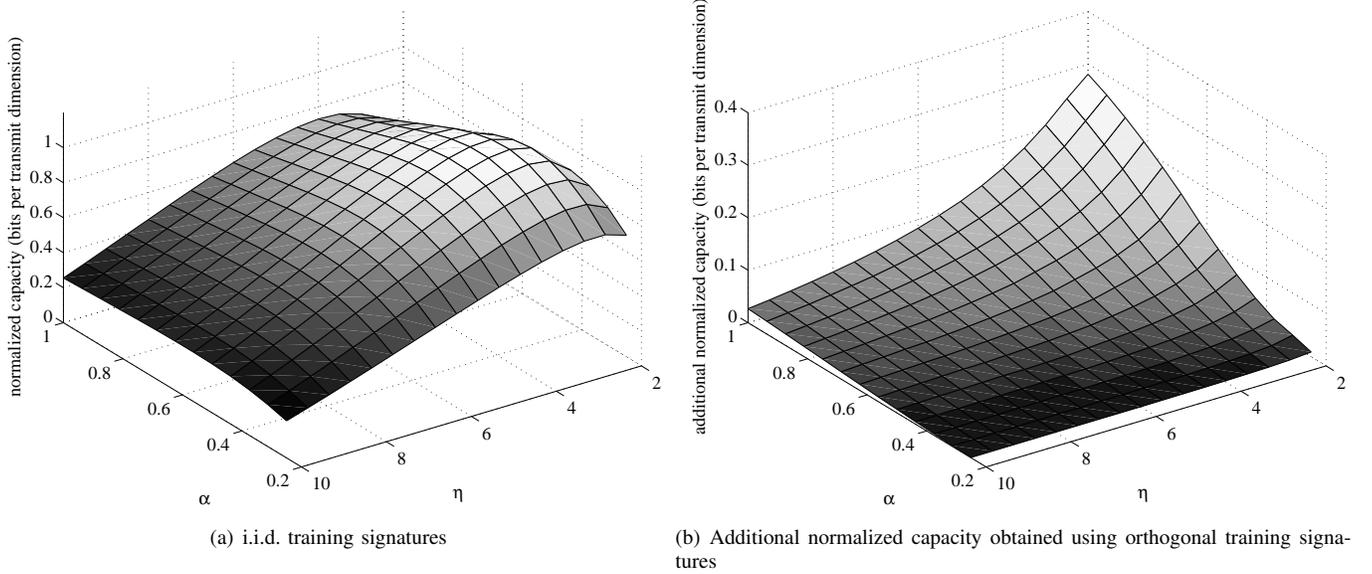


Fig. 6. Throughput optimization: CDMA in frequency-selective fading, i.i.d.  $\mathbf{S}$ , normalized block length  $T/N = 15$ , rectangular windowing, no diagonal loading,  $\frac{E_b}{\sigma^2} = 10$  dB. Fig. 6(b) shows the additional normalized capacity (with respect to Fig. 6(a)) obtained if orthogonal training sequences are used.

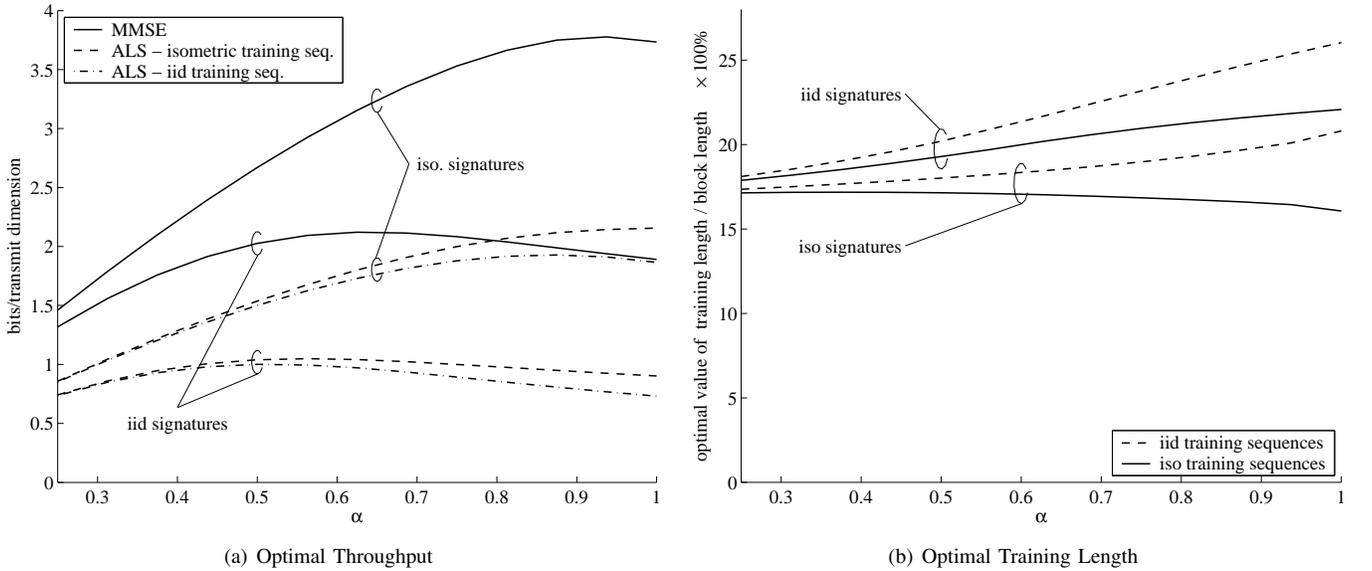


Fig. 7. Throughput optimization: CDMA in frequency-selective fading, normalized block length  $T/N = 15$ , rectangular windowing, no diagonal loading,  $\frac{E_b}{\sigma^2} = 10$  dB, equal power per signature, exponential distribution for a.e.d. of  $\mathbf{H}^\dagger$ .

In the case of i.i.d. training sequences and no diagonal loading, the set of equations for output SINR yields a simple relationship between the SINRs for ALS and MMSE receivers, which accounts for an arbitrary data shaping window. This relation shows that ALS performance depends on the channel matrix *only through the MMSE*. In other words, ALS performance is *independent* of the channel shape given a target output MMSE. Whether or not an analogous relation holds with orthogonal training sequences and/or diagonal loading remains an open problem. Application of the analysis presented here to more general channel models (e.g., multi-user/multi-antenna) is also a topic for further study.

## APPENDIX I PRECURSOR TO ASYMPTOTIC ANALYSIS

*Definition 1:* Let  $\{a_N\}_{N=1,\dots}$  and  $\{b_N\}_{N=1,\dots}$  denote a pair of infinite sequences of complex-valued random variables indexed by  $N$ . These sequences are defined to be *asymptotically equivalent*, denoted  $a_N \asymp b_N$ , iff  $|a_N - b_N| \xrightarrow{a.s.} 0$  as  $N \rightarrow \infty$ , where  $\xrightarrow{a.s.}$  denotes almost-sure convergence in the limit considered.  $\square$

Clearly  $\asymp$  is an equivalence relation, transitivity being obtained through the triangle inequality. We shall additionally define asymptotic equivalence for sequences of  $N \times 1$  vectors and  $N \times N$  matrices in an identical manner as above, where the absolute value is replaced by the Euclidean vector norm and

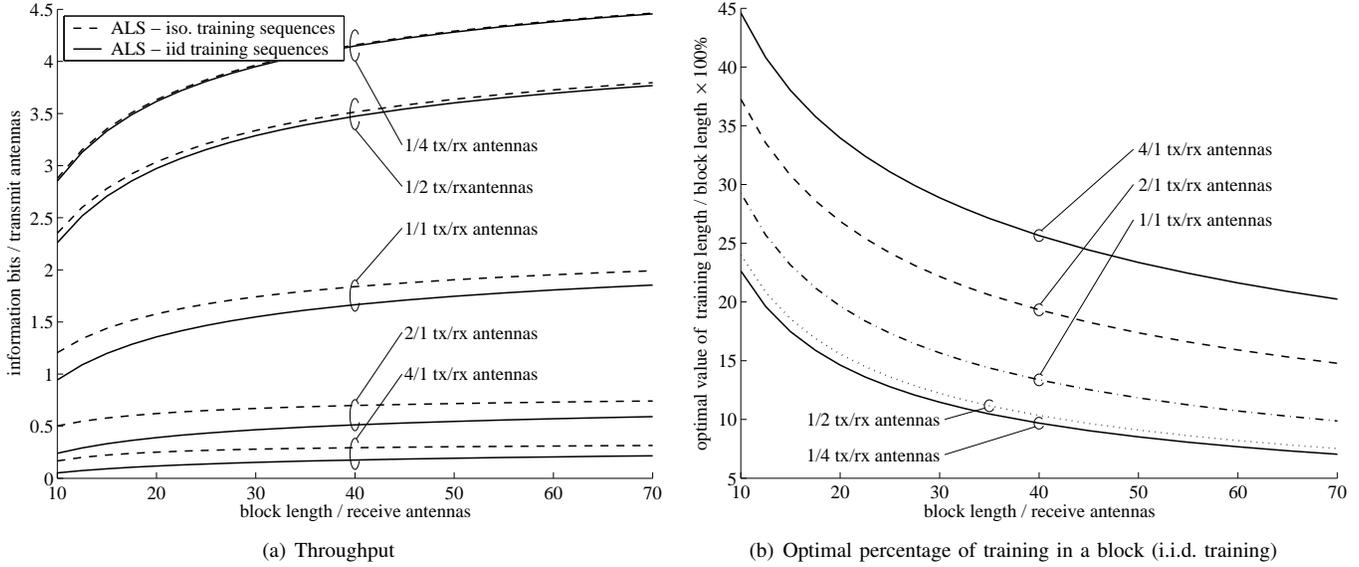


Fig. 8. Throughput optimization: MIMO channel, rectangular windowing, no diagonal loading,  $\frac{E_b}{\sigma^2} = 10$  dB, equal transmit power per antenna.

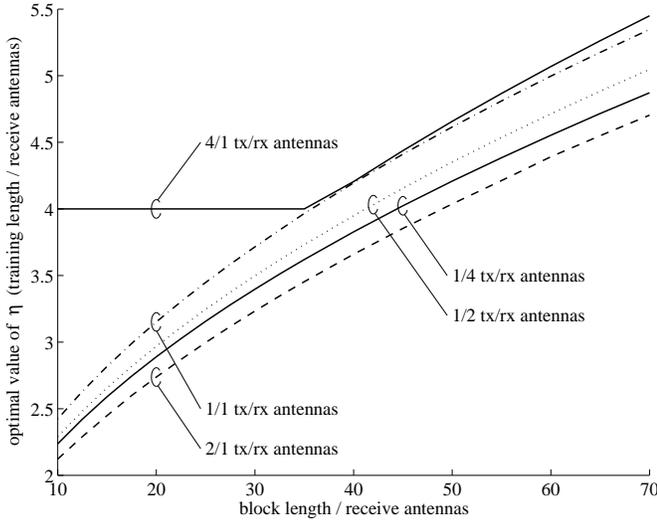


Fig. 9. Optimal value of normalized training length  $\eta$  for orthogonal training sequences corresponding to Figure 8(a).

the associated induced spectral norm, respectively.

If  $a_N \asymp b_N$  and  $x_N \asymp y_N$ , it follows that  $(a_N + x_N) \asymp (b_N + y_N)$ , again due to the triangle inequality. Multiplication, however, requires a little more care.

**Lemma 2:** If  $a_N \asymp b_N$  and  $x_N \asymp y_N$ , and if  $|a_N|$ ,  $|y_N|$  and/or  $|b_N|$ ,  $|x_N|$  are uniformly bounded above<sup>14</sup> over  $N$ , then  $a_N x_N \asymp b_N y_N$ . Similarly,  $a_N / x_N \asymp b_N / y_N$  if  $|a_N|$ ,  $|y_N|^{-1}$  and/or  $|b_N|$ ,  $|x_N|^{-1}$  are uniformly bounded above over  $N$ .

*Proof:* See Appendix II-A.  $\square$

Note that the multiplicative part of Lemma 2 holds for any mixture of matrices, vectors or scalars for which the dimensions of  $a_N$  and  $x_N$  are such that  $a_N x_N$  makes sense,

<sup>14</sup>A sequence  $\{a_N\}_{N=1,\dots}$  of complex-valued  $N \times 1$  vectors or scalars is uniformly bounded above over  $N$  if  $\sup_N |a_N| < \infty$ , or in the case of complex-valued  $N \times N$  matrices,  $\sup_N \|a_N\| < \infty$ .

due to the submultiplicative property of the spectral norm. The following definition and related results, however, are concerned with scalar complex sequences.

Also, it can be shown that any of the requirements for uniform upper bounds in Lemma 2 may be relaxed to either:

$$|z_N| = O(\ln^p(N)), \quad 0 \leq p < \infty, \quad (93)$$

or,

$$|z_N| \leq |z'_N| \xrightarrow{a.s.} c \quad \text{as } N \rightarrow \infty, \quad (94)$$

where  $c$  is a finite positive deterministic constant which does not depend on  $N$ , and  $\{z'_N\}_{n=1,\dots}$  is a judiciously chosen sequence which is used to bound  $|z_N|$ . Here,  $|z_N|$  is any of  $|a_N|$ ,  $|b_N|$ ,  $|x_N|$ ,  $|y_N|$ ,  $|x_N|^{-1}$ , or  $|y_N|^{-1}$ . Also,  $f(N) = O(g(N))$  means that there exists some finite positive constant  $C$  such that, for large enough  $N$ ,  $Cg(N)$  is an upper bound on  $f(N)$ .

The relaxed bound in (93) is a consequence of the fact that  $\sum_{N=1}^{\infty} \frac{\ln^m(N)}{N^n} = \frac{d}{dx^m} \zeta(x) \Big|_{x=n}$  is finite for any finite integers  $m \geq 0$  and  $n \geq 2$ , and where  $\zeta(x)$  is the Riemann Zeta function. Of course, the uniform upper bound condition is subsumed by (93) with  $p = 0$ .

Also, the relaxed bound in (94) is referred to as being bounded above almost surely. That is,  $|z_N| < 2c < \infty$  for all  $N > N'$  for  $N'$  sufficiently large, with probability one.

**Definition 2:** Let  $\{\{a_{N,n}\}_{n=1,\dots,N}\}_{N=1,\dots}$  and  $\{\{b_{N,n}\}_{n=1,\dots,N}\}_{N=1,\dots}$  denote a pair of infinite sequences, indexed by  $N$ . The  $N^{\text{th}}$  element is a complex-valued sequence of length  $N$ , indexed by  $n$ . These sequences are defined to be *uniformly asymptotically equivalent*, denoted  $a_{N,n} \stackrel{n}{\asymp} b_{N,n}$ , iff  $\max_{n \leq N} |a_{N,n} - b_{N,n}| \xrightarrow{a.s.} 0$  as  $N \rightarrow \infty$ .  $\square$

Also we define  $a_N$  and  $b_{N,n}$  (as defined in Definitions 1 and 2 above) as being uniformly asymptotically equivalent (denoted  $a_N \stackrel{n}{\asymp} b_{N,n}$ ), if  $a_{N,n} \stackrel{n}{\asymp} b_{N,n}$  where  $a_{N,n} = a_N$  for all  $n = 1, \dots, N$ .

As was the case with  $\asymp$ , if  $a_{N,n} \stackrel{n}{\asymp} b_{N,n}$  and  $x_{N,n} \stackrel{n}{\asymp} y_{N,n}$ , it follows that  $(a_{N,n} + x_{N,n}) \stackrel{n}{\asymp} (b_{N,n} + y_{N,n})$ . For multiplication of two sequences, we have the following result, analogous to Lemma 2.

**Lemma 3:** If  $a_{N,n} \stackrel{n}{\asymp} b_{N,n}$  and  $x_{N,n} \stackrel{n}{\asymp} y_{N,n}$ , and if  $|a_{N,n}|$ ,  $|y_{N,n}|$  and/or  $|b_{N,n}|$ ,  $|x_{N,n}|$  are uniformly bounded above over  $N$  and  $n$ , then  $a_{N,n}x_{N,n} \stackrel{n}{\asymp} b_{N,n}y_{N,n}$ . Similarly,  $a_{N,n}/x_{N,n} \stackrel{n}{\asymp} b_{N,n}/y_{N,n}$  if  $|a_{N,n}|$ ,  $|y_{N,n}|^{-1}$  and/or  $|b_{N,n}|$ ,  $|x_{N,n}|^{-1}$  are uniformly bounded above over  $N$  and  $n$ .  $\square$

As in the discussion after Lemma 2, the requirements for uniform upper bounds in Lemma 3 may be relaxed to either of (93) or (94), except that in this case  $|z_N|$  is any of  $\max_{n \leq N} |a_{N,n}|$ ,  $\max_{n \leq N} |b_{N,n}|$ , etc.

**Lemma 4:** If  $a_{N,n} \stackrel{n}{\asymp} b_{N,n}$ , then  $\frac{1}{N} \sum_{n=1}^N a_{N,n} \asymp \frac{1}{N} \sum_{n=1}^N b_{N,n}$ .

*Proof:* This follows immediately from

$$\left| \frac{1}{N} \sum_{n=1}^N (a_{N,n} - b_{N,n}) \right| \leq \max_{n \leq N} |a_{N,n} - b_{N,n}|.$$

$\square$

**Lemma 5:** For  $N = 1, \dots$ , let  $\mathbf{X}_N = \mathbf{M}_N - z\mathbf{I}_N$ , where  $\mathbf{M}_N$  is an  $N \times N$  Hermitian matrix and  $z \in \mathbb{C}^+$ , and suppose  $\mathbf{u}_N \in \mathbb{C}^N$ . Denote  $u_N = \mathbf{u}_N^\dagger \mathbf{X}_N^{-1} \mathbf{u}_N$ . If  $|\mathbf{u}_N|^{-1}$  and  $\|\mathbf{X}_N\|$  are uniformly bounded above by  $b$  and  $B$ , respectively, then

$$\text{Im}(u_N) \geq \text{Im}(z)(bB)^{-2}$$

and hence  $u_N \in \mathbb{C}^+$ .

*Proof:* See Appendix II-B.  $\square$

The following lemma is an asymptotic extension of the matrix inversion lemma, and is used extensively in the subsequent appendices to remove matrix dimensions as described in Section III. It is based on an approach in [16].

**Lemma 6:** Let  $\mathbf{Y}_N = \mathbf{X}_N + \mathbf{v}_N \mathbf{u}_N^\dagger + \mathbf{u}_N \mathbf{v}_N^\dagger + c_N \mathbf{u}_N \mathbf{u}_N^\dagger$ , where  $\mathbf{v}_N, \mathbf{u}_N \in \mathbb{C}^N$ ,  $c_N \in \mathbb{R}^*$ , and  $\mathbf{X}_N = \mathbf{M}_N - z\mathbf{I}_N$ , where  $\mathbf{M}_N$  is an  $N \times N$  Hermitian matrix and  $z \in \mathbb{C}^+$ . Denote

$$\begin{aligned} \epsilon_N &= \mathbf{u}_N^\dagger \mathbf{X}_N^{-1} \mathbf{v}_N, \\ u_N &= \mathbf{u}_N^\dagger \mathbf{X}_N^{-1} \mathbf{u}_N, \\ v_N &= \mathbf{v}_N^\dagger \mathbf{X}_N^{-1} \mathbf{v}_N. \end{aligned}$$

Assume that as  $N \rightarrow \infty$ ,

$$|\epsilon_N| \xrightarrow{a.s.} 0 \quad (95)$$

and

$$b = \sup_N |\mathbf{u}_N|^{-1} < \infty, \quad (96)$$

$$B = \sup_N \max \{ \|\mathbf{X}_N\|, |\mathbf{v}_N|, |\mathbf{u}_N|, |c_N| \} < \infty. \quad (97)$$

Then,

$$\left| \mathbf{Y}_N^{-1} \mathbf{u}_N - \frac{\mathbf{X}_N^{-1} (\mathbf{u}_N - u_N \mathbf{v}_N)}{1 - u_N (v_N - c_N)} \right| \xrightarrow{a.s.} 0, \quad (98)$$

$$\left| \mathbf{Y}_N^{-1} \mathbf{v}_N - \frac{\mathbf{X}_N^{-1} (-v_N \mathbf{u}_N + (1 + c_N u_N) \mathbf{v}_N)}{1 - u_N (v_N - c_N)} \right| \xrightarrow{a.s.} 0, \quad (99)$$

as  $N \rightarrow \infty$ , and

$$\delta = \inf_N |1 - u_N (v_N - c_N)| > 0 \quad (100)$$

where  $\delta$  depends only on  $b, B$ , and  $\text{Im}(z)$ .

*Proof:* See Appendix II-C.  $\square$

In addition, we remark that it can be shown the restrictive requirements of uniform upper bounds in (96)–(97) may be replaced by bounds of the form given in either (93) or (94).

**Lemma 7:** Let  $\mathbf{A}_N$  be an  $N \times N$  Hermitian matrix, and suppose  $\sup_N \|\mathbf{A}_N\| < \infty$ . Using the definitions and assumptions of Lemma 6, additionally define

$$\epsilon_N^{(1)} = \mathbf{u}_N^\dagger \mathbf{X}_N^{-1} \mathbf{A}_N \mathbf{X}_N^{-1} \mathbf{v}_N, \quad (101)$$

$$\epsilon_N^{(2)} = \mathbf{v}_N^\dagger \mathbf{X}_N^{-1} \mathbf{A}_N \mathbf{X}_N^{-1} \mathbf{u}_N, \quad (102)$$

$$\acute{u}_N = \mathbf{u}_N^\dagger \mathbf{X}_N^{-1} \mathbf{A}_N \mathbf{X}_N^{-1} \mathbf{u}_N, \quad (103)$$

$$\acute{v}_N = \mathbf{v}_N^\dagger \mathbf{X}_N^{-1} \mathbf{A}_N \mathbf{X}_N^{-1} \mathbf{v}_N. \quad (104)$$

Then,

$$\begin{aligned} \text{tr}[\mathbf{A}_N \mathbf{Y}_N^{-1}] &\asymp \text{tr}[\mathbf{A}_N \mathbf{X}_N^{-1}] \\ &+ \frac{u_N \acute{v}_N + (v_N - c_N) \acute{u}_N - \epsilon_N^{(1)} - \epsilon_N^{(2)}}{1 - u_N (v_N - c_N)} \end{aligned} \quad (105)$$

*Proof:* See Appendix II-D.  $\square$

**Lemma 8:** [2, Lemma 2.6] Let  $z \in \mathbb{C}^+$ ,  $\mathbf{A}$  and  $\mathbf{B}$   $N \times N$  Hermitian,  $\tau \in \mathbb{R}$ , and  $\mathbf{q} \in \mathbb{C}^{N \times 1}$ . Then,

$$|\text{tr} [((\mathbf{B} - z\mathbf{I}_N)^{-1} - (\mathbf{B} + \tau \mathbf{q} \mathbf{q}^\dagger - z\mathbf{I}_N)^{-1}) \mathbf{A}]| \leq \frac{\|\mathbf{A}\|}{\text{Im}(z)}.$$

$\square$

**Lemma 9:** [32, Lemma 1] Let  $\mathbf{C}_N$  be an  $N \times N$  complex-valued matrix with a uniformly bounded spectral norm, i.e.,  $\sup_N \|\mathbf{C}_N\| < \infty$ , and  $\mathbf{y} = [X_1, \dots, X_N]^\dagger / \sqrt{N}$ , where the  $X_i$ 's are i.i.d. complex random variables with mean zero, unit variance, and finite eighth moment. Then

$$\mathbf{E} \left[ \left| \mathbf{y}^\dagger \mathbf{C}_N \mathbf{y} - \frac{1}{N} \text{tr}[\mathbf{C}_N] \right|^4 \right] \leq \frac{c}{N^2}$$

where the constant  $c > 0$  does not depend on  $N$ ,  $\mathbf{C}$ , nor on the distribution of  $X_i$ .  $\square$

**Lemma 10:** Let  $\mathbf{S}$  be  $K < N$  columns of an  $N \times N$  Haar distributed random matrix, and suppose  $\mathbf{s}$  is a column of  $\mathbf{S}$ . Let  $\mathbf{X}_N$  be an  $N \times N$  complex-valued matrix, which is a non-trivial function of all columns of  $\mathbf{S}$  except  $\mathbf{s}$ , and has a uniformly bounded spectral norm. Then, provided  $K/N \rightarrow \alpha < 1$  as  $(N, K) \rightarrow \infty$ ,

$$\mathbf{E} \left[ \left| \mathbf{s}^\dagger \mathbf{X}_N \mathbf{s} - \frac{1}{N - K} \text{tr}[\Phi \mathbf{X}_N] \right|^4 \right] = O(N^{-2})$$

where  $\Phi = \mathbf{I}_N - (\mathbf{S}^\dagger - \mathbf{s}^\dagger)$ .

*Proof:* This result is a straightforward extension of [7, Proposition 4].  $\square$

Throughout the subsequent derivations, we shall use the fact that since we have assumed that the e.d.f.'s of  $\mathbf{A}^2, \mathbf{H}\mathbf{H}^\dagger$ ,

and  $\mathbf{W}$  converge in distribution almost surely to compactly supported non-random distributions on  $\mathbb{R}^*$ , we have [33]

$$\lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K f(P_k) = \mathbf{E}[f(P)], \quad (106)$$

$$\lim_{N \rightarrow \infty} \frac{1}{\beta^* N} \sum_{n=1}^{\beta^* N} f(d_n^2) = \mathbf{E}[f(H)], \quad (107)$$

$$\lim_{i \rightarrow \infty} \frac{1}{i} \sum_{m=1}^i f(w_m) = \mathbf{E}[f(W)], \quad (108)$$

almost surely, where  $d_n$  is the  $n^{\text{th}}$  singular value of  $\mathbf{H}$ , and  $f: \mathbb{R}^* \rightarrow \mathbb{R}^*$  is any fixed, bounded, continuous function on the support of the a.e.d. of  $\mathbf{A}^2$ , the first  $\beta^* N$  eigenvalues of  $\mathbf{H}\mathbf{H}^\dagger$ , and  $\mathbf{W}$ , respectively.

## APPENDIX II PROOFS FOR APPENDIX I

### A. Lemma 2

*Proof:* The fact that  $a_N x_N \asymp b_N y_N$  can be seen after writing  $a_N x_N - b_N y_N = a_N x_N - b_N y_N + a_N y_N - a_N y_N$  and hence  $|a_N x_N - b_N y_N| \leq |a_N| |x_N - y_N| + |y_N| |a_N - b_N|$ . Alternatively, we may add and subtract  $b_N x_N$  from  $a_N x_N - b_N y_N$  to obtain  $|a_N x_N - b_N y_N| \leq |x_N| |a_N - b_N| + |b_N| |x_N - y_N|$ . The division property,  $a_N/x_N \asymp b_N/y_N$ , can be shown in the same way

$$\left| \frac{a_N}{x_N} - \frac{b_N}{y_N} \right| \leq \begin{cases} \frac{|a_N| |x_N - y_N| + |a_N - b_N|}{|x_N| |y_N|} \\ \frac{|b_N| |x_N - y_N| + |a_N - b_N|}{|x_N| |y_N|} \end{cases}.$$

If we have that  $|y_N|$  is uniformly bounded below by some constant  $0 < \delta < \infty$ , then, given a realization for which  $|x_N - y_N| \rightarrow 0$ , we may take  $N'$  sufficiently large such that  $|x_N - y_N| \leq \delta/2$  and hence  $|x_N| \geq \delta/2$  for all  $N \geq N'$ . Using this fact, and the uniform upper bounds for  $|a_N|$  or  $|b_N|$ , we obtain the result. The same argument holds if we have that  $|x_N|$  is uniformly bounded below.  $\square$

### B. Lemma 5

*Proof:* First, we note the inequality (from the proof of [31, Lemma 16.5])

$$\frac{\mathbf{x}^\dagger \mathbf{x}}{\mathbf{x}^\dagger \mathbf{Y}^{-1} \mathbf{x}} \leq \frac{\mathbf{x}^\dagger \mathbf{Y} \mathbf{x}}{\mathbf{x}^\dagger \mathbf{x}} \quad (109)$$

for any  $\mathbf{x} \in \mathbb{C}^N$ ,  $\mathbf{x} \neq \mathbf{0}$ , and Hermitian positive definite  $N \times N$  complex-valued matrix  $\mathbf{Y}$ .

Now,

$$\begin{aligned} \text{Im}(u_N) &= \frac{1}{2j} \mathbf{u}_N^\dagger (\mathbf{X}^{-1} - \mathbf{X}^{-\dagger}) \mathbf{u}_N \\ &= \text{Im}(z) \mathbf{u}_N^\dagger (\mathbf{X}^\dagger \mathbf{X})^{-1} \mathbf{u}_N \end{aligned} \quad (110)$$

and hence using (109) and the bounds for  $|\mathbf{u}_N|^{-1}$  and  $\|\mathbf{X}_N\|$ , we obtain

$$\begin{aligned} \frac{1}{\text{Im}(u_N)} &= \frac{1}{\text{Im}(z) \mathbf{u}_N^\dagger (\mathbf{X}^\dagger \mathbf{X})^{-1} \mathbf{u}_N} \\ &\leq \frac{\mathbf{u}_N^\dagger \mathbf{X}^\dagger \mathbf{X} \mathbf{u}_N}{\text{Im}(z) (\mathbf{u}_N^\dagger \mathbf{u}_N)^2} \leq \frac{\|\mathbf{X}\|^2}{\text{Im}(z) |\mathbf{u}_N|^2} \end{aligned}$$

which, with  $z \in \mathbb{C}^+$  gives the result.  $\square$

### C. Lemma 6

*Proof:* We consider a realization for which (95) holds, and take  $N'$  sufficiently large such that  $|\epsilon_N| \leq 1/2$  so that

$$|1 + \epsilon_N| \geq 1/2, \quad \forall N \geq N'. \quad (111)$$

Now, note that  $u_N \in \mathbb{C}^+$  due to Lemma 5 and (96)–(97). Also, due to the definition of  $\mathbf{X}_N$ ,

$$\|\mathbf{X}_N^{-1}\| \leq \text{Im}(z)^{-1}. \quad (112)$$

Therefore,  $|u_N| \leq B^2 \text{Im}(z)^{-1}$  due to (97) and (112). Also,  $\text{Im}(-u_N^{-1}) = \text{Im}(u_N)/|u_N|^2$  and  $|u_N^{-1}| \leq \text{Im}(u_N)^{-1}$ . Additionally, note that  $\text{Im}(v_N) \geq 0$ , using an identical argument to (110). Using these facts we obtain

$$\begin{aligned} \left| \frac{1}{1 - u_N v_N} \right| &= \frac{|-u_N^{-1}|}{|-u_N^{-1} + v_N|} \leq \frac{|u_N^{-1}|}{\text{Im}(-u_N^{-1} + v_N)} \\ &\leq \frac{|u_N^{-1}|}{\text{Im}(-u_N^{-1})} \leq \frac{|u_N|}{\text{Im}(u_N)^2} \leq \left( \frac{bB^2}{\text{Im}(z)} \right)^4. \end{aligned} \quad (113)$$

And so,

$$\delta' = \inf_N |1 - u_N v_N| > 0. \quad (114)$$

In the same way, also using  $c_N \in \mathbb{R}^*$ , we obtain an identical uniform upper bound on  $|1 - u_N(v_N - c_N)|^{-1}$ , and hence obtain (100).

In what follows, we will drop the dependence on  $N$  from  $u_N$ ,  $v_N$ ,  $c_N$ ,  $\mathbf{u}_N$ ,  $\mathbf{v}_N$ ,  $\mathbf{X}_N$ , and  $\mathbf{Y}_N$  to clarify the derivations. Define  $\mathbf{X}_1$  and  $\mathbf{X}_2$  according to the following equations

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}_2 + c \mathbf{u} \mathbf{u}^\dagger, \\ \mathbf{X}_2 &= \mathbf{X}_1 + \mathbf{u} \mathbf{v}^\dagger, \\ \mathbf{X}_1 &= \mathbf{X} + \mathbf{v} \mathbf{u}^\dagger. \end{aligned}$$

The matrix inversion lemma gives

$$\mathbf{Y}^{-1} = \mathbf{X}_2^{-1} - \frac{c \mathbf{X}_2^{-1} \mathbf{u} \mathbf{u}^\dagger \mathbf{X}_2^{-1}}{1 + c \mathbf{u}^\dagger \mathbf{X}_2^{-1} \mathbf{u}}, \quad (115)$$

$$\mathbf{X}_2^{-1} = \mathbf{X}_1^{-1} - \frac{\mathbf{X}_1^{-1} \mathbf{u} \mathbf{v}^\dagger \mathbf{X}_1^{-1}}{1 + \mathbf{v}^\dagger \mathbf{X}_1^{-1} \mathbf{u}}, \quad (116)$$

$$\mathbf{X}_1^{-1} = \mathbf{X}^{-1} - \frac{\mathbf{X}^{-1} \mathbf{v} \mathbf{u}^\dagger \mathbf{X}^{-1}}{1 + \mathbf{u}^\dagger \mathbf{X}^{-1} \mathbf{v}}. \quad (117)$$

First consider  $\mathbf{X}_1^{-1}$ , and note from (117), (95), Lemma 2, (97), (111), and (112) that

$$\mathbf{X}_1^{-1} \asymp \mathbf{X}^{-1} - \mathbf{X}^{-1} \mathbf{v} \mathbf{u}^\dagger \mathbf{X}^{-1}. \quad (118)$$

In fact, in the remainder of the proof, we shall repeatedly use (95), (97), (100), (111), (112) and (114) in order to apply Lemma 2, without explicitly stating this, however, it should be clear from the context.

From (118) we obtain

$$\mathbf{X}_1^{-1} \mathbf{u} \asymp \mathbf{X}^{-1} (\mathbf{u} - u \mathbf{v}), \quad (119)$$

$$\mathbf{X}_1^{-1} \mathbf{v} \asymp \mathbf{X}^{-1} \mathbf{v},$$

$$\mathbf{v}^\dagger \mathbf{X}_1^{-1} \asymp (\mathbf{v}^\dagger - v \mathbf{u}^\dagger) \mathbf{X}^{-1}, \quad (120)$$

$$\mathbf{u}^\dagger \mathbf{X}_1^{-1} \asymp \mathbf{u}^\dagger \mathbf{X}^{-1}.$$

From this, we obtain

$$\begin{aligned} \mathbf{u}^\dagger \mathbf{X}_1^{-1} \mathbf{u} &\asymp u, \\ \mathbf{v}^\dagger \mathbf{X}_1^{-1} \mathbf{v} &\asymp v, \\ \mathbf{v}^\dagger \mathbf{X}_1^{-1} \mathbf{u} &\asymp -uv, \\ \mathbf{u}^\dagger \mathbf{X}_1^{-1} \mathbf{v} &\asymp 0. \end{aligned} \quad (121)$$

Before we consider  $\mathbf{X}_2^{-1}$ , we first analyze the denominator of the second term in (116). Firstly, due to (121), we may take  $N'$  large enough such that  $|\mathbf{v}^\dagger \mathbf{X}_1^{-1} \mathbf{u} + uv| < \delta'/2$  for all  $N \geq N'$ , and hence with (114) we obtain  $|1 + \mathbf{v}^\dagger \mathbf{X}_1^{-1} \mathbf{u}| \geq \delta'/2$ . With this fact, we obtain  $(1 + \mathbf{v}^\dagger \mathbf{X}_1^{-1} \mathbf{u})^{-1} \asymp (1 - uv)^{-1}$ , since

$$\begin{aligned} \left| \frac{1}{1 + \mathbf{v}^\dagger \mathbf{X}_1^{-1} \mathbf{u}} - \frac{1}{1 - uv} \right| &= \frac{|\mathbf{v}^\dagger \mathbf{X}_1^{-1} \mathbf{u} + uv|}{|1 + \mathbf{v}^\dagger \mathbf{X}_1^{-1} \mathbf{u}| |1 - uv|} \\ &\leq \frac{2}{(\delta')^2} |\mathbf{v}^\dagger \mathbf{X}_1^{-1} \mathbf{u} + uv|, \quad \forall N \geq N'. \end{aligned} \quad (122)$$

Now consider  $\mathbf{X}_2^{-1}$ , for which from (116) and the preceding discussion we obtain

$$\begin{aligned} \mathbf{X}_2^{-1} \mathbf{v} &= \mathbf{X}_1^{-1} \mathbf{v} - \frac{\mathbf{X}_1^{-1} \mathbf{u} \mathbf{v}^\dagger \mathbf{X}_1^{-1} \mathbf{v}}{1 + \mathbf{v}^\dagger \mathbf{X}_1^{-1} \mathbf{u}} \\ &\asymp \mathbf{X}_1^{-1} \mathbf{v} - \frac{v \mathbf{X}_1^{-1} \mathbf{u}}{1 - uv} \\ &\asymp \frac{1}{1 - uv} \mathbf{X}^{-1} (-v \mathbf{u} + \mathbf{v}). \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbf{X}_2^{-1} \mathbf{u} &\asymp \frac{1}{1 - uv} \mathbf{X}^{-1} (\mathbf{u} - uv), \\ \mathbf{v}^\dagger \mathbf{X}_2^{-1} &\asymp \frac{1}{1 - uv} (-v \mathbf{u}^\dagger + \mathbf{v}^\dagger) \mathbf{X}^{-1}, \\ \mathbf{u}^\dagger \mathbf{X}_2^{-1} &\asymp \frac{1}{1 - uv} (\mathbf{u}^\dagger - uv \mathbf{v}^\dagger) \mathbf{X}^{-1}, \end{aligned} \quad (123)$$

$$\mathbf{u}^\dagger \mathbf{X}_2^{-1} \mathbf{v} \asymp \frac{1}{1 - uv} (\mathbf{u}^\dagger - uv \mathbf{v}^\dagger) \mathbf{X}^{-1} \mathbf{v}, \quad (124)$$

and so,

$$\begin{aligned} \mathbf{u}^\dagger \mathbf{X}_2^{-1} \mathbf{v} &\asymp \mathbf{v}^\dagger \mathbf{X}_2^{-1} \mathbf{u} \asymp -\frac{vu}{1 - uv}, \\ \mathbf{u}^\dagger \mathbf{X}_2^{-1} \mathbf{u} &\asymp \frac{u}{1 - uv}, \\ \mathbf{v}^\dagger \mathbf{X}_2^{-1} \mathbf{v} &\asymp \frac{v}{1 - uv}, \\ 1 + c \mathbf{u}^\dagger \mathbf{X}_2^{-1} \mathbf{u} &\asymp \frac{1 - u(v - c)}{1 - uv}. \end{aligned} \quad (125)$$

Before considering  $\mathbf{Y}^{-1}$ , we note that from (97), (100), (125), and similar arguments preceding (122), that  $|1 + c \mathbf{u}^\dagger \mathbf{X}_2^{-1} \mathbf{u}| \geq \delta/2$  for all  $N \geq N''$  for  $N''$  sufficiently large, and hence  $1/(1 + c \mathbf{u}^\dagger \mathbf{X}_2^{-1} \mathbf{u}) \asymp (1 - uv)/(1 - u(v - c))$ .

Considering  $\mathbf{Y}^{-1}$  using (115) and the preceding discussion, we obtain

$$\mathbf{Y}^{-1} \mathbf{u} = \frac{\mathbf{X}_2^{-1} \mathbf{u}}{1 + c \mathbf{u}^\dagger \mathbf{X}_2^{-1} \mathbf{u}} \asymp \frac{1}{1 - u(v - c)} \mathbf{X}^{-1} (\mathbf{u} - uv).$$

Similarly,

$$\begin{aligned} \mathbf{Y}^{-1} \mathbf{v} &= \mathbf{X}_2^{-1} \mathbf{v} - c \frac{\mathbf{u}^\dagger \mathbf{X}_2^{-1} \mathbf{v}}{1 + c \mathbf{u}^\dagger \mathbf{X}_2^{-1} \mathbf{u}} \mathbf{X}_2^{-1} \mathbf{u} \\ &\asymp \mathbf{X}_2^{-1} \mathbf{v} + c \frac{vu}{1 - u(v - c)} \mathbf{X}_2^{-1} \mathbf{u} \\ &\asymp \frac{1}{1 - u(v - c)} \mathbf{X}^{-1} (-v \mathbf{u} + (1 + uc) \mathbf{v}). \end{aligned}$$

□

#### D. Lemma 7

*Proof:* The proof continues from the proof of Lemma 6. Again, we drop the subscript  $N$  for convenience. We see from (95), (117), (111) and (101) that

$$\text{tr}[\mathbf{A} \mathbf{X}_1^{-1}] \asymp \text{tr}[\mathbf{A} \mathbf{X}^{-1}] - \varepsilon^{(1)}$$

while (116), (119), (120), (121), and (101)–(104) give

$$\begin{aligned} \text{tr}[\mathbf{A} \mathbf{X}_2^{-1}] &\asymp \text{tr}[\mathbf{A} \mathbf{X}_1^{-1}] - \frac{(\mathbf{v}^\dagger - v \mathbf{u}^\dagger) \mathbf{X}^{-1} \mathbf{A} \mathbf{X}^{-1} (\mathbf{u} - uv)}{1 - uv} \\ &= \text{tr}[\mathbf{A} \mathbf{X}_1^{-1}] + \frac{u \dot{v} + v \dot{u} - \varepsilon^{(2)} - uv \varepsilon^{(1)}}{1 - uv}. \end{aligned}$$

Similarly, (123), (124) and (101)–(104) give

$$\begin{aligned} \mathbf{u}^\dagger \mathbf{X}_2^{-1} \mathbf{A} \mathbf{X}_2^{-1} \mathbf{u} &\asymp \frac{(\mathbf{u}^\dagger - uv \mathbf{v}^\dagger) \mathbf{X}^{-1} \mathbf{A} \mathbf{X}^{-1} (\mathbf{u} - uv)}{(1 - uv)^2} \\ &= \frac{\dot{u} + u^2 \dot{v} - u(\varepsilon^{(1)} + \varepsilon^{(2)})}{(1 - uv)^2} \end{aligned}$$

and finally (115) and (125) yield

$$\text{tr}[\mathbf{A} \mathbf{Y}^{-1}] \asymp \text{tr}[\mathbf{A} \mathbf{X}_2^{-1}] - c \frac{1 - uv}{1 - u(v - c)} \mathbf{u}^\dagger \mathbf{X}_2^{-1} \mathbf{A} \mathbf{X}_2^{-1} \mathbf{u}.$$

Combining the above, we obtain (105). □

### APPENDIX III PROOF OF THEOREM 1

The analysis in these appendices is based on removing a single dimension from matrices and vectors, as described in Section III. The dimension removed will correspond to a particular data stream, transmit/receive dimension, or symbol interval. For example, in what follows,  $\mathbf{R}_{t_n}$  represents the matrix  $\mathbf{R}$  with the  $n^{\text{th}}$  transmit dimension removed. The symbol  $t_n$  is used in this case since the  $n^{\text{th}}$  transmit dimension is removed. We will use  $d_k$  when removing the  $k^{\text{th}}$  data stream, and  $r_m$  for removing the  $m^{\text{th}}$  received symbol interval.

Appendix I contains several important definitions and results required for the following analysis. In particular, we define  $\overset{k}{\asymp}$  and  $\overset{n}{\asymp}$  according to Definition 2 in Appendix I, where the maximum is over  $k \leq K$  and  $n \leq N$ , respectively, and the limit is as  $(M, N, K) \rightarrow \infty$  with  $K/N \rightarrow \alpha > 0$  and  $M/N \rightarrow \beta > 0$  constant, as described in Section II-E.

### A. Definitions

Let  $\mathbf{R} = (\mathbf{HSA})^\dagger - z\mathbf{I}_M$ ,  $z \in \mathbb{C}^+$ . The Stieltjés transform of the e.d.f. of the eigenvalues of  $(\mathbf{HSA})^\dagger$  is given by  $G_R^N(z) = \gamma_1^N$ , and the MMSE SINR in (5) is given by  $P_k \rho_{1,k}^N$ , where

$$\gamma_j^N = \frac{1}{M} \text{tr}[\mathbf{X}^{(j)}] \quad , \text{ for } j = 1, \dots, \quad (126)$$

$$\rho_{j,k}^N = \mathbf{h}_k^\dagger \mathbf{X}_{d_k}^{(j)} \mathbf{h}_k \quad , \text{ for } j = 1, \dots, 4, \quad 0 < k \leq K, \quad (127)$$

where  $\mathbf{h}_k = \mathbf{H}\mathbf{s}_k$ , and

$$\mathbf{X}^{(j)} = \begin{cases} \mathbf{R}^{-1} & , j = 1, \\ \mathbf{R}^{-\dagger} \mathbf{R}^{-1} & , j = 2, \\ \mathbf{R}^{-\dagger} \mathbf{H}^\dagger \mathbf{R}^{-1} & , j = 3, \\ \mathbf{R}^{-\dagger} (\mathbf{HSA})^\dagger \mathbf{R}^{-1} & , j = 4. \end{cases}$$

Furthermore,  $\mathbf{X}_{d_k}^{(j)}$  is defined by removing the  $k^{\text{th}}$  data stream,  $0 < k \leq K$ , from  $\mathbf{X}^{(j)}$ , by replacing  $\mathbf{R}$ ,  $\mathbf{S}$ , and  $\mathbf{A}$  by  $\mathbf{R}_{d_k}$ ,  $\mathbf{S}_{d_k}$ , and  $\mathbf{A}_{d_k}$ , respectively, where

$$\mathbf{R}_{d_k} = (\mathbf{H}\mathbf{S}_{d_k}\mathbf{A}_{d_k})^\dagger - z\mathbf{I}_M,$$

$\mathbf{S}_{d_k}$  is  $\mathbf{S}$  with the  $k^{\text{th}}$  column removed, and  $\mathbf{A}_{d_k}$  is  $\mathbf{A}$  with the  $k^{\text{th}}$  row and column removed. That is,  $\mathbf{R}_{d_k} = \mathbf{R} - P_k \mathbf{h}_k \mathbf{h}_k^\dagger$ .

The following proposition shows that we may substitute  $\mathbf{H}$  with an equivalent matrix, without lack of generality. This substitution is essential in the analysis which follows.

*Proposition 2:* For the model (1), the distribution of both the Stieltjés transform of the e.e.d. of  $(\mathbf{HSA})^\dagger$  and the MMSE SINR are invariant to the substitution of  $\mathbf{VD}$  for  $\mathbf{H}$ , where  $\mathbf{V}$  is an  $M \times M$  Haar-distributed random unitary matrix, and  $\mathbf{D}$  is a  $M \times N$  diagonal matrix containing the singular values of  $\mathbf{H}$ .

*Proof:* Let  $\mathbf{T}$  be an independent  $M \times M$  Haar-distributed random matrix. Now, note that the quantities of interest, namely  $\gamma_j^N$  and  $\rho_{j,k}^N$ , are unchanged by the substitution of  $\mathbf{TH}$  for  $\mathbf{H}$ . That is,

$$\begin{aligned} \gamma_1^N &= \frac{1}{N} \text{tr}[\mathbf{R}^{-1}] = \frac{1}{M} \text{tr}[\mathbf{T}^\dagger \mathbf{R}^{-1}] \\ &= \frac{1}{M} \text{tr}[(\mathbf{THSA})^\dagger - z\mathbf{I}_M]^{-1}. \end{aligned}$$

Writing  $\mathbf{THS} = (\mathbf{T}\mathbf{U}_M)\mathbf{D}(\mathbf{U}_N^\dagger\mathbf{S})$ , where  $\mathbf{U}_M\mathbf{D}\mathbf{U}_N^\dagger$  is the singular value decomposition of  $\mathbf{H}$ , the unitary invariance of  $\mathbf{T}$  and  $\mathbf{S}$  infers the result for the Stieltjés transform. A similar treatment of  $\rho_{j,k}^N$  gives the result for the MMSE SINR.  $\square$

Therefore, in the remainder of this appendix, we substitute  $\mathbf{H}$  with  $\mathbf{VD}$  everywhere.<sup>15</sup> We denote the  $n^{\text{th}}$  column of  $\mathbf{V}$  as  $\mathbf{v}_n$ , for  $0 < n \leq M$ , and define  $\mathbf{v}_n = \mathbf{0}$  for  $n > \beta^*N$ . Define  $\{d_1, \dots, d_{\beta^*N}\}$  as the diagonal elements of  $\mathbf{D}$ , note that  $\beta^*N = \min(M, N)$ , and define  $d_n = 0$  for  $n > \beta^*N$ .

For isometric  $\mathbf{S}$  only, we also require the following generalization of the system model and asymptotic limit.<sup>16</sup> We shall assume that  $\mathbf{D}$  and  $\mathbf{S}$  have dimensions  $M \times \bar{N}$  and  $\bar{N} \times K$ ,

<sup>15</sup>We stress that  $\mathbf{VD}$  is an equivalent matrix, as defined in Proposition 2, as opposed to a decomposition of  $\mathbf{H}$ .

<sup>16</sup>Without this step, it is not possible to show the uniform asymptotic equivalence in (130).

respectively, where  $0 < \bar{N} < N$ . Also, we shall include the limit  $\bar{N} \rightarrow \infty$  with  $\bar{N}/N \rightarrow \varsigma$  where  $\varsigma \in (0, 1)$ . That is,

$$\begin{aligned} \text{diag}(\mathbf{D}) &= \{d_1, \dots, d_{\beta^*N}\}, \\ \mathbf{S}^\dagger &= [\tilde{\mathbf{s}}_1 \cdots \tilde{\mathbf{s}}_{\bar{N}}], \end{aligned}$$

where the definition of  $\beta^*$  is extended to  $\beta^* = \min(\beta, \varsigma)$ . Note also that  $\tilde{\mathbf{s}}_k$  now has dimension  $\bar{N} \times 1$ , and the maximum involved in  $\overset{n}{\succ}$  shall be over  $n \leq \bar{N}$ . The desired result will be obtained by considering  $\varsigma \rightarrow 1^-$ . For i.i.d.  $\mathbf{S}$ ,  $\bar{N} = N$ , and hence  $\varsigma = 1$  and  $\beta^* = \min(\beta, 1)$ .

We can now define

$$\tau_{j,n}^N = \mathbf{u}_n^\dagger \mathbf{X}_{t_n}^{(j)} \mathbf{u}_n \quad , \text{ for } j = 1, 2, 3, \dots, \quad 0 < n \leq \beta^*N, \quad (128)$$

where  $\mathbf{u}_n = \mathbf{H}_{t_n} \mathbf{S}_{t_n} \mathbf{A}^2 \tilde{\mathbf{s}}_n$ . Also,  $\mathbf{X}_{t_n}^{(j)}$  denotes  $\mathbf{X}^{(j)}$  with the effect of the  $n^{\text{th}}$  transmit dimension removed,  $0 < n \leq \beta^*N$ , by replacing  $\mathbf{R}$ ,  $\mathbf{H}$ , and  $\mathbf{S}$  with  $\mathbf{R}_{t_n}$ ,  $\mathbf{H}_{t_n}$ , and  $\mathbf{S}_{t_n}$ , respectively, where

$$\begin{aligned} \mathbf{R}_{t_n} &= (\mathbf{H}_{t_n} \mathbf{S}_{t_n} \mathbf{A})^\dagger - z\mathbf{I}_M, \\ \mathbf{H}_{t_n} &= \mathbf{V}_{t_n} \mathbf{D}_{t_n}, \end{aligned}$$

and where  $\mathbf{V}_{t_n}$  and  $\mathbf{S}_{t_n}$  are  $\mathbf{V}$  and  $\mathbf{S}$  with their  $n^{\text{th}}$  column and row removed, respectively, and  $\mathbf{D}_{t_n}$  is  $\mathbf{D}$  with both the  $n^{\text{th}}$  column and row removed.

Returning to (127) and (128), note that these quadratic forms are uniformly asymptotically equivalent to the following expressions, derived in Appendix IV. These will be important in the subsequent analysis.

$$\rho_{j,k}^N \overset{k}{\succ} \rho_j^N = \begin{cases} \frac{1}{N} \text{tr}[\mathbf{H}^\dagger \mathbf{X}^{(j)}] & , \text{ i.i.d. } \mathbf{S}, \\ \frac{1}{N-K} \text{tr}[(\mathbf{H}\mathbf{I}\mathbf{I})^\dagger \mathbf{X}^{(j)}] & , \text{ iso. } \mathbf{S}, \end{cases} \quad (129)$$

$$\tau_{j,n}^N \overset{n}{\succ} \tau_j^N = \begin{cases} \frac{1}{N} \text{tr}[(\mathbf{HSA}^2)^\dagger \mathbf{X}^{(j)}] & , \text{ i.i.d. } \mathbf{S}, \\ \frac{1}{N-\bar{N}} \text{tr}[(\mathbf{HSA}^2\mathbf{\Xi})^\dagger \mathbf{X}^{(j)}] & , \text{ iso. } \mathbf{S}, \end{cases} \quad (130)$$

where

$$\mathbf{\Pi}^2 = \mathbf{I}_{\bar{N}} - \mathbf{S}\mathbf{S}^\dagger, \quad (131)$$

$$\mathbf{\Xi}^2 = \mathbf{I}_K - \mathbf{S}^\dagger\mathbf{S}. \quad (132)$$

Also, note that

$$\mathbf{s}_k^\dagger \mathbf{H}^\dagger \mathbf{H} \mathbf{s}_k \overset{k}{\succ} \beta^* \mathbf{E}[H] > 0 \quad (133)$$

from Lemma 9 or Lemma 10, the Borel-Cantelli lemma, and (107). The positivity of (133) is implied by  $\beta^* > 0$  and  $\mathbf{E}[H] > 0$ . Note that  $\mathbf{E}[H] > 0$  is implied by the assumption that the distribution of  $H$  has a compact support on  $\mathbb{R}^*$ , and does not have all mass at zero.

In addition, letting  $c_n = \tilde{\mathbf{s}}_n^\dagger \mathbf{A}^2 \tilde{\mathbf{s}}_n$  and  $\bar{p} = \mathbf{E}[P]$ , we have

$$c_n \overset{n}{\succ} \alpha \bar{p}. \quad (134)$$

This is shown in a similar manner to (133) using (106). For iso.  $\mathbf{S}$  it first requires  $\tilde{\mathbf{s}}_n$  written as described before (178).

We now give several bounds on particular matrix and vector norms which are required in order to apply Lemmas 2 and 3 later. Firstly, the assumption that  $z \in \mathbb{C}^+$  gives

$$\|\mathbf{R}^{-1}\| \leq \text{Im}(z)^{-1}. \quad (135)$$

Secondly, the assumptions on  $\mathbf{H}$ ,  $\mathbf{S}$ , and  $\mathbf{A}$  outlined in Section II imply

$$H_{\max} = \sup_N \|\mathbf{H}\|^2 < \infty, \quad (136)$$

$$\sup_N \|\mathbf{A}\|^2 < \infty, \quad (137)$$

$$\|\mathbf{S}\|^2 \asymp (1 + \sqrt{\alpha})^2, \quad (\text{i.i.d. } \mathbf{S}), \quad (138)$$

$$|\mathbf{s}_k|^2 \stackrel{k}{\asymp} \varsigma, \quad (139)$$

$$|\tilde{\mathbf{s}}_n|^2 \stackrel{n}{\asymp} \alpha, \quad (140)$$

where (138) is due to [34], while (139) and (140) are shown in an identical manner to (133) and (134), respectively. Of course,  $\|\mathbf{S}\| = 1$  for isometric  $\mathbf{S}$ . Moreover, (135)–(140) imply that all of

$$\|\mathbf{R}\|, \quad (141)$$

$$\max_{n \leq N} \max\{|\mathbf{u}_n|, |c_n|\}, \quad (142)$$

$$\max\{|\rho_j^N|, \max_{k \leq K} |\rho_{j,k}^N|\}, \quad \text{for } j = 1, \dots, 4, \quad (143)$$

$$\max\{|\tau_j^N|, \max_{n \leq N} |\tau_{j,n}^N|\}, \quad \text{for } j = 1, \dots, 3, \quad (144)$$

are bounded above almost surely, that is, have an upper bound of form given in (94). For this to be the case for (141), it also requires the assumption that  $|z| < \infty$ .

### B. Derivations

We start by applying Lemma 6 to  $\mathbf{R} = \mathbf{R}_{d_k} + P_k \mathbf{h}_k \mathbf{h}_k^\dagger$ , i.e., to extract the  $k^{\text{th}}$  data stream from  $\mathbf{R}$ , as described in Section III. Namely, we substitute  $\mathbf{R}$ ,  $\mathbf{R}_{d_k}$ ,  $\mathbf{0}$ ,  $\mathbf{h}_k$ , and  $P_k$  for  $\mathbf{Y}_N$ ,  $\mathbf{X}_N$ ,  $\mathbf{v}_N$ ,  $\mathbf{u}_N$ , and  $c_N$  in the Lemma, respectively. Condition (95) is satisfied trivially, while (133) and (136)–(141) satisfy (96) and (97), respectively, in the almost-sure sense. Hence,<sup>17</sup> Lemma 6 states that

$$\mathbf{R}^{-1} \mathbf{h}_k \asymp \frac{\mathbf{R}_{d_k}^{-1} \mathbf{h}_k}{1 + P_k \rho_{1,k}^N}, \quad (145)$$

$$\inf_N \max_{k \leq K} |1 + P_k \rho_{1,k}^N| > 0, \quad \text{a.s.} \quad (146)$$

This may be applied to the following identity to obtain

$$\begin{aligned} 1 &= \frac{1}{M} \text{tr}[\mathbf{R}\mathbf{R}^{-1}] = -z\gamma_1^N + \frac{1}{M} \sum_{k=1}^K P_k \mathbf{h}_k^\dagger \mathbf{R}^{-1} \mathbf{h}_k \\ &\asymp -z\gamma_1^N + \frac{\alpha}{\beta} \frac{1}{K} \sum_{k=1}^K \frac{P_k \rho_{1,k}^N}{1 + P_k \rho_{1,k}^N}. \end{aligned} \quad (147)$$

Now, due to (129), (137), (143), (146), and Lemma 3, it follows that

$$\frac{P_k \rho_{1,k}^N}{1 + P_k \rho_{1,k}^N} \stackrel{k}{\asymp} \frac{P_k \rho_1^N}{1 + P_k \rho_1^N}. \quad (148)$$

From (147), (148), and Lemma 4 we obtain

$$1 + z\gamma_1^N \asymp \frac{\alpha}{\beta} \rho_1^N \mathcal{E}_{1,1}^N = \frac{\alpha}{\beta} (1 - \mathcal{E}_{0,1}^N) \quad (149)$$

<sup>17</sup>Of course, (145) holds with = in place of  $\asymp$ , due to the matrix inversion lemma. We use Lemma 6 instead in order to obtain (146), which is needed to show (148).

where

$$\mathcal{E}_{p,1}^N = \frac{1}{K} \sum_{k=1}^K \frac{P_k^p}{1 + P_k \rho_1^N}.$$

For future reference, note that from (145), following the proof of (149) gives

$$\frac{1}{N} \text{tr}[(\mathbf{H}\mathbf{S}\mathbf{A}^2)^\dagger \mathbf{R}^{-1}] = \frac{1}{N} \sum_{k=1}^K \frac{P_k^2 \rho_{1,k}^N}{1 + P_k \rho_{1,k}^N} \asymp \alpha(\bar{p} - \mathcal{E}_{1,1}^N), \quad (150)$$

$$\frac{1}{N} \text{tr}[(\mathbf{H}\mathbf{S})^\dagger \mathbf{R}^{-1}] = \frac{1}{N} \sum_{k=1}^K \frac{\rho_{1,k}^N}{1 + P_k \rho_{1,k}^N} \asymp \alpha \rho_1^N \mathcal{E}_{0,1}^N. \quad (151)$$

To prove the remaining equations in Theorem 1, we consider another expansion of the correlation matrix  $\mathbf{R}$ , this time to remove the  $n^{\text{th}}$  transmit dimension,  $0 < n \leq \beta^* N$ , as described in Section III.

$$\begin{aligned} \mathbf{R} &= (\mathbf{H}_{t_n} \mathbf{S}_{t_n} + d_n \mathbf{v}_n \tilde{\mathbf{s}}_n^\dagger) \mathbf{A}^2 (\mathbf{H}_{t_n} \mathbf{S}_{t_n} + d_n \mathbf{v}_n \tilde{\mathbf{s}}_n^\dagger)^\dagger - z \mathbf{I}_M \\ &= \mathbf{R}_{t_n} + d_n \mathbf{u}_n \mathbf{v}_n^\dagger + d_n \mathbf{v}_n \mathbf{u}_n^\dagger + d_n^2 c_n \mathbf{v}_n \mathbf{v}_n^\dagger \end{aligned} \quad (152)$$

where  $\mathbf{u}_n$  and  $c_n$  are defined after (128) and above (134), respectively.

We now apply Lemma 6 to (152), where  $\mathbf{Y}_N$ ,  $\mathbf{X}_N$ ,  $\mathbf{v}_N$ ,  $\mathbf{u}_N$ , and  $c_N$  in the statement of the Lemma correspond to  $\mathbf{R}$ ,  $\mathbf{R}_{t_n}$ ,  $d_n \mathbf{u}_n$ ,  $\mathbf{v}_n$ , and  $c_n$ , respectively. We shall now verify that the conditions of the Lemma are satisfied. For any  $n \leq \beta^* N$ , since  $\mathbf{H}_{t_n}^\dagger \mathbf{v}_n = \mathbf{0}$  we have  $\mathbf{R}_{t_n} \mathbf{v}_n = -z \mathbf{v}_n$  and  $\mathbf{R}_{t_n}^\dagger \mathbf{v}_n = -z^\dagger \mathbf{v}_n$ , and moreover

$$\mathbf{v}_n^\dagger \mathbf{R}_{t_n}^{-1} \mathbf{v}_n = -z^{-1}, \quad (153)$$

$$\mathbf{u}_n^\dagger \mathbf{R}_{t_n}^{-1} \mathbf{v}_n = \mathbf{v}_n^\dagger \mathbf{R}_{t_n}^{-1} \mathbf{u}_n = 0, \quad (154)$$

where  $\mathbf{v}_n^\dagger \mathbf{R}_{t_n}^{-1} \mathbf{v}_n$  corresponds to  $u_N$  in the Lemma, and (154) satisfies condition (95) of the Lemma. Since  $|\mathbf{v}_n| = 1$ , condition (96) is satisfied, and along with (141)–(142) condition (97) is satisfied in the almost-sure sense. Note that  $\tau_{1,n}^N$ , defined in (128), corresponds to  $v_N$  in the Lemma. Therefore,

$$\mathbf{R}^{-1} \mathbf{v}_n \asymp \frac{\mathbf{R}_{t_n}^{-1} (\mathbf{v}_n + d_n z^{-1} \mathbf{u}_n)}{1 + d_n^2 z^{-1} (\tau_{1,n}^N - c_n)}, \quad (155)$$

$$\mathbf{R}^{-1} \mathbf{u}_n \asymp \frac{\mathbf{R}_{t_n}^{-1} (-d_n \tau_{1,n}^N \mathbf{v}_n + (1 - d_n^2 c_n z^{-1}) \mathbf{u}_n)}{1 + d_n^2 z^{-1} (\tau_{1,n}^N - c_n)}, \quad (156)$$

$$\inf_N \min_{n \leq \bar{N}} |1 + d_n^2 z^{-1} (\tau_{1,n}^N - c_n)| > 0, \quad \text{a.s.}, \quad (157)$$

which we shall now use to derive (18) and (19). Note that (157) will be required when using Lemma 3, to satisfy almost-sure uniform bound conditions.

With i.i.d.  $\mathbf{S}$ , we see from (130) that (150) gives an expression for  $\tau_1^N$ . For isometric  $\mathbf{S}$ , we see from (130) that in addition to (150) we require

$$\begin{aligned} \frac{1}{N} \text{tr}[(\mathbf{H}\mathbf{S}\mathbf{A}^2 \mathbf{S}^\dagger)^\dagger \mathbf{R}^{-1}] &= \frac{1}{N} \sum_{n=1}^{\bar{N}} \text{tr}[(\mathbf{H}\mathbf{S}\mathbf{A}^2 \tilde{\mathbf{s}}_n)^\dagger \mathbf{R}^{-1}] \\ &= \frac{1}{N} \sum_{n=1}^{\bar{N}} \text{tr}[(\mathbf{u}_n + d_n c_n \mathbf{v}_n)^\dagger \mathbf{R}^{-1}] \end{aligned} \quad (158)$$

where we have used  $\mathbf{HSA}^2 \tilde{\mathbf{s}}_n = \mathbf{u}_n + d_n c_n \mathbf{v}_n$ . Continuing with the preceding application of Lemma 6, we may use Lemma 7 to determine an equivalent asymptotic representation of the argument in the sum in (158), where  $\mathbf{A}_N$  in the statement of Lemma 7 corresponds to  $(\mathbf{u}_n + d_n c_n \mathbf{v}_n)^\dagger$ . That is, using (154), (155), and (156), we note that the terms corresponding to  $\varepsilon_N^{(1)}$  and  $\varepsilon_N^{(2)}$  are both asymptotically equivalent to  $-d_n^2 c_n z^{-1} \tau_{1,n}^N$ , while the terms corresponding to  $\hat{u}_N$ ,  $\hat{v}_N$ , and  $\text{tr}[\mathbf{A}_N \mathbf{X}_N^{-1}]$  are asymptotically equivalent to  $d_n^2 c_n^2 z^{-2}$ ,  $d_n^2 (\tau_{1,n}^N)^2$ , and  $\tau_{1,n}^N - d_n^2 c_n^2 z^{-1}$ , respectively. Therefore, after some algebra, we obtain from (105)

$$\text{tr}[(\mathbf{u}_n + d_n c_n \mathbf{v}_n)^\dagger \mathbf{R}^{-1}] \asymp c_n + \frac{\tau_{1,n}^N - c_n}{1 + d_n^2 z^{-1} (\tau_{1,n}^N - c_n)} \quad (159)$$

and from Lemma 3, (130), and (134) we obtain

$$c_n + \frac{\tau_{1,n}^N - c_n}{1 + d_n^2 z^{-1} (\tau_{1,n}^N - c_n)} \stackrel{n}{\asymp} \alpha \bar{p} + \frac{\tau_1^N - \alpha \bar{p}}{1 + d_n^2 z^{-1} (\tau_1^N - \alpha \bar{p})} \quad (160)$$

noting that the bounds required for the application of Lemma 3 are almost-surely satisfied by (144),  $\bar{p} < \infty$ , and (157).

We therefore obtain from (158), (159), (160), and Lemma 4 that

$$\frac{1}{N} \text{tr}[(\mathbf{HSA}^2 \mathbf{S}^\dagger)^\dagger \mathbf{R}^{-1}] \asymp \zeta \alpha \bar{p} + (\tau_1^N - \alpha \bar{p}) (\beta^* (\mathcal{H}_{0,1}^N - 1) + \zeta) \quad (161)$$

where

$$\mathcal{H}_{p,1}^N = \frac{1}{\beta^* N} \sum_{n=1}^{\beta^* N} \frac{d_n^{2p}}{1 + d_n^2 z^{-1} (\tau_1^N - \alpha \bar{p})}.$$

Moreover, from (130), (150), and (161) we have

$$\begin{aligned} \tau_1^N &= \frac{1}{N - \bar{N}} (\text{tr}[(\mathbf{HSA}^2)^\dagger \mathbf{R}^{-1}] - \text{tr}[(\mathbf{HSA}^2 \mathbf{S}^\dagger)^\dagger \mathbf{R}^{-1}]) \\ &\asymp \frac{1}{1 - \zeta} (\alpha (\bar{p} - \mathcal{E}_{1,1}^N) - \zeta \alpha \bar{p} - (\tau_1^N - \alpha \bar{p}) (\beta^* (\mathcal{H}_{0,1}^N - 1) + \zeta)) \end{aligned} \quad (162)$$

or equivalently,

$$\tau_1^N \asymp \alpha \bar{p} - \frac{\alpha \mathcal{E}_{1,1}^N}{\beta^* (\mathcal{H}_{0,1}^N - 1) + 1}. \quad (163)$$

For i.i.d.  $\mathbf{S}$ , using (129) and (154)–(155) in the same manner as the derivation of (161), we have that

$$\begin{aligned} \rho_1^N &= \frac{1}{N} \sum_{n=1}^{\beta^* N} d_n^2 \mathbf{v}_n^\dagger \mathbf{R}^{-1} \mathbf{v}_n \asymp \frac{1}{N} \sum_{n=1}^{\beta^* N} \frac{-z^{-1} d_n^2}{1 + d_n^2 z^{-1} (\tau_{1,n}^N - c_n)} \\ &\asymp -\beta^* z^{-1} \mathcal{H}_{1,1}^N \end{aligned} \quad (164)$$

and similarly from (129), (151), and (164) we obtain for isometric  $\mathbf{S}$

$$\rho_1^N \asymp \frac{1}{1 - \alpha} (-\beta^* z^{-1} \mathcal{H}_{1,1}^N - \alpha \rho_1^N \mathcal{E}_{0,1}^N). \quad (165)$$

We now simplify the preceding solution by noting that the identity  $\frac{1}{M} \text{tr}[\mathbf{R} \mathbf{R}^{-1}] = 1$  may also be expanded in the

dimension  $N$ , as opposed to the dimension  $K$  in (149). That is,

$$1 + z \gamma_1^N = \frac{1}{M} \sum_{n=1}^{\beta^* N} d_n \tilde{\mathbf{s}}_n^\dagger \mathbf{A}^2 (\mathbf{S}_{t_n}^\dagger \mathbf{H}_{t_n}^\dagger + d_n \tilde{\mathbf{s}}_n \mathbf{v}_n^\dagger) \mathbf{R}^{-1} \mathbf{v}_n. \quad (166)$$

Applying (154) and (155) to the argument to the above sum gives

$$\begin{aligned} &\tilde{\mathbf{s}}_n^\dagger \mathbf{A}^2 (\mathbf{S}_{t_n}^\dagger \mathbf{H}_{t_n}^\dagger + d_n \tilde{\mathbf{s}}_n \mathbf{v}_n^\dagger) \mathbf{R}^{-1} \mathbf{v}_n \\ &\asymp \frac{(\mathbf{u}_n^\dagger + d_n c_n \mathbf{v}_n^\dagger) \mathbf{R}_{t_n}^{-1} (\mathbf{v}_n + d_n z^{-1} \mathbf{u}_n)}{1 + d_n^2 z^{-1} (\tau_{1,n}^N - c_n)} \end{aligned} \quad (167)$$

$$\stackrel{n}{\asymp} 1 - \frac{1}{1 + d_n^2 z^{-1} (\tau_1^N - \alpha \bar{p})} \quad (168)$$

and so applying Lemma 4 to (166) with (168) gives

$$1 + z \gamma_1^N \asymp \frac{\beta^*}{\beta} (1 - \mathcal{H}_{0,1}^N). \quad (169)$$

We now use (169) and (149) to simplify (163) and (165) in the case of isometric  $\mathbf{S}$ . Combining (169) with (149) gives  $1 + \beta^* (\mathcal{H}_{0,1}^N - 1) \asymp 1 - \beta (1 + z \gamma_1^N)$ , which combined with (163) gives

$$\tau_1^N \asymp \alpha \bar{p} - \frac{\alpha \mathcal{E}_{1,1}^N}{1 - \beta (1 + z \gamma_1^N)}. \quad (170)$$

Similarly, combining (149) with (165) gives

$$\rho_1^N \asymp \frac{-\beta^* z^{-1} \mathcal{H}_{1,1}^N}{1 - \beta (1 + z \gamma_1^N)}. \quad (171)$$

In summary, the expressions (149), (150), (164), (170), and (171) may be expressed as

$$\mathbf{x}^{(N)}(z) = \Phi^{(N)} \left( \mathbf{x}^{(N)}(z) \right) + \boldsymbol{\epsilon}^{(N)}(z) \quad (172)$$

where  $\mathbf{x}^{(N)}(z)$  denotes the random vector  $[\gamma_1^N(z), \rho_1^N(z), \tau_1^N(z)]^T$  (Note that we have highlighted the dependence of  $\gamma_1^N$ ,  $\rho_1^N$ , and  $\tau_1^N$  on  $z$  for what follows). Also, each element of  $\boldsymbol{\epsilon}^{(N)}$  almost-surely converges to zero, and the vector-valued function  $\Phi^{(N)}$  is given by

$$\Phi^{(N)} \left( \mathbf{x}^{(N)}(z) \right) = \begin{bmatrix} z^{-1} \left( \frac{\alpha}{\beta} \rho_1^N(z) \mathcal{E}_{1,1}^N - 1 \right) \\ -\beta^* z^{-1} D_N(z) \mathcal{H}_{1,1}^N \\ \alpha (\bar{p} - D_N(z) \mathcal{E}_{1,1}^N) \end{bmatrix}$$

where  $D_N(z) = 1$  for i.i.d.  $\mathbf{S}$ , and  $D_N(z) = (1 - \beta (1 + z \gamma_1^N(z)))^{-1}$  for isometric  $\mathbf{S}$ .

We now show that  $\mathbf{x}^{(N)}(z)$  has a limit by the following steps.<sup>18</sup>

- 1) Consider a countable family  $(z_\kappa)_{\kappa \in \mathbb{N}}$  with an accumulation point contained in a compact subset of  $\mathbb{C}^+$ .
- 2) Note that on a set  $\Omega_1(\kappa)$  with probability one,  $\boldsymbol{\epsilon}^{(N)}(z_\kappa) \rightarrow \mathbf{0}$  for each  $\kappa$ .
- 3) Each element of the vector  $\mathbf{x}^{(N)}$  is analytic and bounded on  $\mathbb{C}^+$ , due to the definitions (126) and (129)–(130).
- 4) Due to point 3, we may apply Montel's theorem [36] to state that there exists a subsequence  $\mathbf{x}^{(N_\ell)}$  of  $\mathbf{x}^{(N)}$

<sup>18</sup>This follows essentially the same steps as the proof of [35, Theorem 2.5].

which converges uniformly on each compact subset of  $\mathbb{C}^+$  to a function  $\mathbf{x}^*$  analytic on  $\mathbb{C}^+$ . Denote  $\mathbf{x}^*(z) = [\gamma_1(z), \rho_1(z), \tau_1(z)]^T$ .

- 5) Due to the almost-sure convergence of the e.d.f.'s of  $\mathbf{H}\mathbf{H}^\dagger$  and  $\mathbf{A}^2$ , it follows from point 4, (106), and (107) that, on  $\{N_\ell\}$ ,  $\mathcal{H}_{p,1}^{N_\ell} \rightarrow \mathcal{H}_{p,1}$  and  $\mathcal{E}_{p,1}^{N_\ell} \rightarrow \mathcal{E}_{p,1}$  for  $p = 0, 1$ , where

$$\mathcal{E}_{p,1} = \mathbf{E} \left[ \frac{P^p}{1 + P\rho_1(z)} \right],$$

$$\mathcal{H}_{p,1} = \mathbf{E} \left[ \frac{H^p}{1 + Hz^{-1}(\tau_1(z) - \alpha\bar{p})} \right].$$

- 6) Now, due to (172) and points 2, 4, and 5, on the set  $\Omega = \bigcap_\kappa \Omega_1(\kappa)$  (which has probability one), we must have that

$$\mathbf{x}^*(z_\kappa) = \Phi^*(\mathbf{x}^*(z_\kappa)) \quad (173)$$

for each  $\kappa$ , where  $\Phi^*$  is the function

$$\Phi^*(\mathbf{x}^*(z)) = \begin{bmatrix} z^{-1} \left( \frac{\alpha}{\beta} \rho_1(z) \mathcal{E}_{1,1} - 1 \right) \\ -\beta^* z^{-1} D(z) \mathcal{H}_{1,1} \\ \alpha(\bar{p} - D(z) \mathcal{E}_{1,1}) \end{bmatrix} \quad (174)$$

where  $D(z) = 1$  for i.i.d.  $\mathbf{S}$ , and  $D(z) = (1 - \beta(1 + z\gamma_1(z)))^{-1}$  for isometric  $\mathbf{S}$ .

- 7) Moreover, on  $\Omega$ , since  $\mathbf{x}^*(z)$  is analytic, we have the analytic continuation  $\mathbf{x}^*(z) = \Phi^*(\mathbf{x}^*(z))$  for each  $z \in \mathbb{C}^+$  (due to the identity theorem [36]).
- 8) Assuming this set of equations has a unique solution  $\mathbf{x}(z)$ , we have  $\mathbf{x}^*(z) = \mathbf{x}(z)$ , and that on  $\Omega$ , every convergent subsequence of  $(\mathbf{x}^{(N)}(z))_{N \in \mathbb{N}}$  converges to the same point  $\mathbf{x}(z)$ . Moreover, this implies almost-sure convergence of the whole sequence, i.e., on  $\Omega$ ,  $\mathbf{x}^{(N)}(z)$  converges to  $\mathbf{x}(z)$  for each  $z \in \mathbb{C}^+$ , as stated in (14)–(16).

Finally, (173)–(174) gives (17)–(19) at  $\varsigma \rightarrow 1^-$ .

### C. Shortcut to (163).

Note that due to the conditions of Lemma 10, the uniform asymptotic equivalence in (130) does not hold for isometric  $\mathbf{S}$  if  $\bar{N}/N \rightarrow 1$ . This is the motivation for introducing the independent parameter  $\bar{N}$  and specifying  $\bar{N}/N \rightarrow \varsigma < 1$ . Moreover, this extra step yields the trace expression in (130), which was used in (162) to derive the relationship (163) for  $\tau_1^N$ . However, the following argument shows that it is possible to arrive at (163) directly, without introducing the parameter  $\bar{N}$ , and without relying on Lemma 10 to provide an expression, which does not depend on  $n$ .

First, we note that since  $\mathbf{S}^\dagger \mathbf{S} = \mathbf{I}_K$ , we may state

$$\frac{1}{N} \text{tr}[(\mathbf{H}\mathbf{S}\mathbf{A}^2)^\dagger \mathbf{R}^{-1}] = \frac{1}{N} \text{tr}[(\mathbf{H}\mathbf{S}\mathbf{A}^2 \mathbf{S}^\dagger)^\dagger \mathbf{R}^{-1}]. \quad (175)$$

Now we assume (without proof) that,  $\tau_{j,n}^N$  is uniformly asymptotically equivalent to another sequence, which does not depend on  $n$ . In that case, for isometric  $\mathbf{S}$  with  $\bar{N} = N$ , we can write

$$\tau_{j,n}^N \stackrel{n}{\asymp} \tau_j^N = \frac{1}{N} \sum_{n=1}^N \tau_{j,n}^N \quad (176)$$

With this assumption, (161) holds at  $\varsigma = 1$ , and combining this with (175) and (150) gives

$$\alpha(\bar{p} - \mathcal{E}_{1,1}^N) \asymp \alpha\bar{p} + (\tau_1^N - \alpha\bar{p})(\beta^*(\mathcal{H}_{0,1}^N - 1) + 1),$$

which indeed corresponds to (163) at  $\varsigma = 1$ . Moreover, establishing the convergence of (172) to (173) is unaffected by the parameter  $\varsigma$ .

In what follows, we will use this argument as a shortcut to derive analogous relations in Appendix VII. The assumption (176) can be proven by introducing  $\bar{N}$ , and taking the appropriate limits with  $\bar{N}/N \rightarrow \varsigma < 1$ . Because we cannot take  $\bar{N} = N$  when  $N$  is finite (see Lemma 10), this does not establish (176) for  $\bar{N}/N = 1$ . However, this assumption, combined with (175), leads to the same expressions in Theorem 1, which are obtained by introducing the parameter  $\bar{N}$ , and taking  $\varsigma \rightarrow 1^-$  after the large-system limit.

## APPENDIX IV

### PROOF OF (129) AND (130) IN APPENDIX III.

Here we show  $\max_{k \leq K} |\rho_{j,k}^N - \rho_j^N| \xrightarrow{a.s.} 0$  and  $\max_{n \leq N} |\tau_{j,n}^N - \tau_j^N| \xrightarrow{a.s.} 0$  for  $j = 1$  in the limit considered. The remaining cases  $j = 2, 3, 4$  are shown in an identical manner using the same results as outlined below.

For isometric  $\mathbf{S}$ , we need the following representation. Let  $\boldsymbol{\theta}_k$  and  $\tilde{\boldsymbol{\theta}}_n$  denote the  $k^{\text{th}}$  column and  $n^{\text{th}}$  row of the  $N \times N$  Haar distributed matrix  $\boldsymbol{\Theta}$  from which  $\mathbf{S}$  is obtained, respectively. Define  $\mathbf{E}_x = [\mathbf{I}_x, \mathbf{0}_{x, N-x}]$  for  $0 \leq x \leq N$ , so that  $\mathbf{S} = \mathbf{E}_{\bar{N}} \boldsymbol{\Theta} \mathbf{E}_K^\dagger$ ,  $\mathbf{s}_k = \mathbf{E}_{\bar{N}} \boldsymbol{\theta}_k$ ,  $\tilde{\mathbf{s}}_n = \mathbf{E}_K \tilde{\boldsymbol{\theta}}_n$ , and therefore

$$\rho_{1,k}^N = \text{tr}[(\mathbf{H}\mathbf{E}_{\bar{N}} \boldsymbol{\theta}_k)^\dagger \mathbf{R}_{d_k}^{-1}], \quad (177)$$

$$\tau_{1,n}^N = \text{tr}[(\mathbf{H}_{t_n} \mathbf{S}_{t_n} \mathbf{A}^2 \mathbf{E}_K \tilde{\boldsymbol{\theta}}_n)^\dagger \mathbf{R}_{t_n}^{-1}]. \quad (178)$$

Define

$$\rho_{1,k}^{N'} = \begin{cases} \frac{1}{\bar{N}} \text{tr}[\mathbf{H}^\dagger \mathbf{R}_{d_k}^{-1}] & , \text{ i.i.d. } \mathbf{S}, \\ \frac{1}{\bar{N}-K} \text{tr}[(\mathbf{H}\boldsymbol{\Pi})^\dagger \mathbf{R}_{d_k}^{-1}] & , \text{ iso. } \mathbf{S}, \end{cases}$$

$$\rho_{1,k}^{N''} = \frac{1}{N-K} \text{tr}[(\mathbf{H}\boldsymbol{\Pi})^\dagger \mathbf{R}_{d_k}^{-1}] \quad , \text{ iso. } \mathbf{S},$$

$$\boldsymbol{\Pi}_{d_k}^2 = \boldsymbol{\Pi}^2 + \mathbf{s}_k^\dagger.$$

For i.i.d.  $\mathbf{S}$ , we may apply Lemma 9 to (127). For isometric  $\mathbf{S}$ , we may we apply Lemma 10 to (177), noting that the appropriate choice of orthogonal projection matrix is  $\mathbf{I}_N - (\boldsymbol{\Theta} \mathbf{E}_K^\dagger)^\dagger + \boldsymbol{\theta}_k^\dagger$ , and simplify using

$$\mathbf{E}_{\bar{N}}^\dagger (\mathbf{I}_N - (\boldsymbol{\Theta} \mathbf{E}_K^\dagger)^\dagger) \mathbf{E}_{\bar{N}} = \boldsymbol{\Pi}^2$$

where  $\boldsymbol{\Pi}$  is defined in (131). Hence we have  $\rho_{1,k}^N \stackrel{k}{\asymp} \rho_{1,k}^{N'}$  for both i.i.d. and isometric  $\mathbf{S}$ , due to the Borel-Cantelli lemma.

Now, for isometric  $\mathbf{S}$ , we have  $|\rho_{1,k}^{N'} - \rho_{1,k}^{N''}| = \frac{1}{\bar{N}-K} |\rho_{1,k}^N|$ , which with (143) gives  $\rho_{1,k}^{N'} \stackrel{k}{\asymp} \rho_{1,k}^{N''}$ . Also, from Lemma 8 we have  $|\rho_1^N - \rho_{1,k}^{N''}| \leq \frac{H_{\max}}{\text{Im}(z)N}$  for i.i.d.  $\mathbf{S}$ , and  $\leq \frac{H_{\max}}{\text{Im}(z)(N-K)}$  for isometric  $\mathbf{S}$  ( $H_{\max}$  is defined in (136)), and hence  $\rho_{1,k}^{N''} \stackrel{k}{\asymp} \rho_1^N$ . Putting these together, we have  $\rho_{1,k}^N \stackrel{k}{\asymp} \rho_{1,k}^{N'} \stackrel{k}{\asymp} \rho_1^N$  for i.i.d.  $\mathbf{S}$ , and  $\rho_{1,k}^N \stackrel{k}{\asymp} \rho_{1,k}^{N'} \stackrel{k}{\asymp} \rho_{1,k}^{N''} \stackrel{k}{\asymp} \rho_1^N$  for isometric  $\mathbf{S}$ , as claimed in (129).

Turning our attention to  $\tau_{1,n}^N$ , for i.i.d.  $\mathbf{S}$  define

$$\begin{aligned}\tau_{1,n}^{N'} &= \frac{1}{N} \text{tr}[(\mathbf{H}_{t_n} \mathbf{S}_{t_n} \mathbf{A}^2)^\dagger \mathbf{R}_{t_n}^{-1}], \\ \tau_{1,n}^{N''} &= \frac{1}{N} \text{tr}[(\mathbf{H}_{t_n} \mathbf{S}_{t_n} \mathbf{A}^2)^\dagger \mathbf{R}^{-1}].\end{aligned}$$

Firstly,  $\tau_{1,n}^N \stackrel{v}{\asymp} \tau_{1,n}^{N'}$  from Lemma 9 and the Borel-Cantelli lemma. Now, applying Lemma 7 to

$$\left| \tau_{1,n}^{N'} - \tau_{1,n}^{N''} \right| = \frac{1}{N} \left| \text{tr}[(\mathbf{H}_{t_n} \mathbf{S}_{t_n} \mathbf{A}^2)^\dagger (\mathbf{R}^{-1} - \mathbf{R}_{t_n}^{-1})] \right|$$

and using (143)–(144), and (157) it is straightforward to show  $\tau_{j,n}^{N'} \stackrel{v}{\asymp} \tau_{j,n}^{N''}$ . Also,

$$\begin{aligned}\left| \tau_{1,n}^{N''} - \tau_1^N \right| &= \left| \frac{1}{N} \text{tr}[\mathbf{R}^{-1} ((\mathbf{H}_{t_n} \mathbf{S}_{t_n} \mathbf{A}^2)^\dagger - (\mathbf{HSA}^2)^\dagger)] \right| \\ &= \left| \frac{1}{N} \text{tr}[\mathbf{R}^{-1} (d_n \mathbf{u}_n \mathbf{v}_n^\dagger + d_n \mathbf{v}_n \mathbf{u}_n^\dagger + d_n^2 c_n \mathbf{v}_n \mathbf{v}_n^\dagger)] \right| \\ &\leq \frac{1}{N \text{Im}(z)} (2d_n |\mathbf{u}_n| + d_n^2 |c_n|)\end{aligned}\quad (179)$$

where  $\mathbf{u}_n$  and  $c_n$  are defined in Appendix III-A, and we have used  $|\mathbf{v}_n| = 1$  and (135). It is clear from (136) and (142) that the terms inside the bracket of (179) are almost-surely uniformly bounded above over  $N$  and  $n \leq N$ , so  $\tau_{1,n}^{N''} \stackrel{v}{\asymp} \tau_1^N$ . Moreover, as  $\tau_{1,n}^N \stackrel{v}{\asymp} \tau_{1,n}^{N'} \stackrel{v}{\asymp} \tau_{j,n}^{N''} \stackrel{v}{\asymp} \tau_1^N$ , and therefore we have completed the proof of (130) for i.i.d.  $\mathbf{S}$ .

Now turning to (130) for isometric  $\mathbf{S}$ , we may apply Lemma 10 to (178), where the appropriate choice of orthogonal projection matrix is  $\mathbf{I}_N - ((\mathbf{E}_N \Theta)^\dagger)^\dagger + \tilde{\boldsymbol{\theta}}_n^\dagger$ , and simplify using

$$\mathbf{E}_K (\mathbf{I}_N - ((\mathbf{E}_N \Theta)^\dagger)^\dagger) \mathbf{E}_K^\dagger = \boldsymbol{\Xi}^2.$$

where  $\boldsymbol{\Xi}$  is defined in (132). Then, using the same steps outlined above in showing  $\rho_{1,k}^N \stackrel{k}{\asymp} \rho_1^N$  and  $\tau_{1,n}^N \stackrel{v}{\asymp} \tau_1^N$  for i.i.d.  $\mathbf{S}$ , it can be shown that  $\tau_{1,n}^N \stackrel{v}{\asymp} \tau_1^N$  also for isometric  $\mathbf{S}$ .

#### APPENDIX V

##### ALTERNATE MMSE SINR OF SECTION IV-B

Note from (145) that the filter  $\mathbf{R}_{d_k}^{-1} \mathbf{h}_k$  has the same SINR as  $\mathbf{R}^{-1} \mathbf{h}_k$ . The associated signal and interference powers are  $P_k \left| \rho_{1,k}^N \right|^2$  and  $\rho_{4,k}^N + \sigma^2 \rho_{2,k}^N$  respectively, as defined in Appendix III-A. It is easily shown that the latter term simplifies to  $\rho_{1,k}^N$  (when  $z \rightarrow -\sigma^2$ ), and hence the MMSE SINR is  $P_k \rho_{1,k}^N$ . Namely,

$$\begin{aligned}\rho_{4,k}^N + \sigma^2 \rho_{2,k}^N &= \mathbf{h}_k^\dagger \mathbf{R}_{d_k}^{-\dagger} (\mathbf{R}_{d_k} + z \mathbf{I}_M) \mathbf{R}_{d_k}^{-1} \mathbf{h}_k + \sigma^2 \mathbf{h}_k^\dagger \mathbf{R}_{d_k}^{-\dagger} \mathbf{R}_{d_k}^{-1} \mathbf{h}_k \\ &\rightarrow \mathbf{h}_k^\dagger \mathbf{R}_{d_k}^{-1} \mathbf{h}_k \text{ as } z \rightarrow -\sigma^2.\end{aligned}$$

We now seek expressions for each of the variables which enter the interference power, without using the preceding simplification. Following on from the assumptions and results of Appendix III, firstly, note that in extension to (153)–(154),

$$\mathbf{v}_n^\dagger \mathbf{X}_{t_n}^{(j)} \mathbf{v}_n = \begin{cases} |z|^{-2} & , j = 2, \\ 0 & , j = 3, \end{cases}\quad (180)$$

$$\mathbf{v}_n^\dagger \mathbf{X}_{t_n}^{(j)} \mathbf{u}_n = \mathbf{u}_n^\dagger \mathbf{X}_{t_n}^{(j)} \mathbf{v}_n = 0 \quad , \quad j = 1, 2, 3, \quad (181)$$

for  $n = 1, \dots, \beta^* N$ . Also, define

$$\begin{aligned}\mathcal{E}_{p,2}^N &= \frac{1}{K} \sum_{k=1}^K \frac{P_k^p}{|1 + P_k \rho_1^N|^2}, \\ \mathcal{H}_{p,2}^N &= \frac{1}{\beta^* N} \sum_{n=1}^{\beta^* N} \frac{d_n^{2p}}{|1 + d_n^2 z^{-1} (\tau_1^N - \alpha \bar{p})|^2}.\end{aligned}$$

Considering terms which arise in  $\rho_j^N$ , defined in (129), we have from (155) and (145), and additionally using (180)–(181),

$$\begin{aligned}\frac{1}{N} \text{tr}[\mathbf{H}^\dagger \mathbf{X}^{(j)}] &= \begin{cases} \frac{1}{N} \sum_{n=1}^{\beta^* N} d_n^2 \mathbf{v}_n \mathbf{X}^{(j)} \mathbf{v}_n & , j = 2, 3, \\ \frac{1}{N} \sum_{k=1}^K P_k \mathbf{h}_k^\dagger \mathbf{R}^{-1} \mathbf{H}^\dagger \mathbf{R}^{-1} \mathbf{h}_k & , j = 4. \end{cases} \\ &\asymp \begin{cases} \beta^* |z|^{-2} (\mathcal{H}_{1,2}^N + \tau_2^N \mathcal{H}_{2,2}^N) & , j = 2, \\ \beta^* |z|^{-2} (1 + \tau_3^N) \mathcal{H}_{2,2}^N & , j = 3, \\ \alpha \rho_3^N \mathcal{E}_{1,2}^N & , j = 4. \end{cases}\end{aligned}\quad (182)$$

which corresponds to  $\rho_j^N$  for i.i.d.  $\mathbf{S}$ . Additionally note that

$$\begin{aligned}\frac{1}{N} \text{tr}[(\mathbf{HS})^\dagger \mathbf{X}^{(j)}] &= \frac{1}{N} \sum_{k=1}^K \mathbf{h}_k^\dagger \hat{\mathbf{X}}^{(j)} \mathbf{h}_k \\ &\asymp \begin{cases} \alpha \rho_j^N \mathcal{E}_{0,2}^N & , j = 2, 3, \\ \alpha (\rho_4^N \mathcal{E}_{0,2}^N + |\rho_1^N|^2 \mathcal{E}_{1,2}^N) & , j = 4. \end{cases}\end{aligned}\quad (183)$$

Combining (182) and (183) according to (129) for isometric  $\mathbf{S}$  gives

$$\rho_j^N \asymp \begin{cases} \frac{\beta^* |z|^{-2} (\mathcal{H}_{1,2}^N + \tau_2^N \mathcal{H}_{2,2}^N)}{\alpha (\mathcal{E}_{0,2}^N - 1) + 1} & , j = 2, \\ \frac{\beta^* |z|^{-2} (1 + \tau_3^N) \mathcal{H}_{2,2}^N}{\alpha (\mathcal{E}_{0,2}^N - 1) + 1} & , j = 3, \\ \frac{\alpha (\rho_3^N \mathcal{E}_{1,2}^N - |\rho_1^N|^2 \mathcal{E}_{1,2}^N)}{\alpha (\mathcal{E}_{0,2}^N - 1) + 1} & , j = 4. \end{cases}$$

Considering terms which arise in  $\tau_j^N$ , defined in (130), we have from (145)

$$\frac{1}{N} \text{tr}[(\mathbf{HSA}^2)^\dagger \mathbf{X}^{(j)}] = \frac{1}{N} \sum_{k=1}^K P_k^2 \mathbf{h}_k^\dagger \mathbf{X}^{(j)} \mathbf{h}_k \asymp \alpha \rho_j^N \mathcal{E}_{2,2}^N \quad (184)$$

for  $j = 2, 3$ , which corresponds to  $\tau_j^N$  for i.i.d.  $\mathbf{S}$ . Additionally, noting from (155)–(156) that

$$\mathbf{R}^{-1} (\mathbf{u}_n + d_n c_n \mathbf{v}_n) \asymp \frac{\mathbf{R}_{t_n}^{-1} (\mathbf{u}_n + d_n (c_n - \tau_{1,n}^N) \mathbf{v}_n)}{1 + d_n^2 z^{-1} (\tau_{1,n}^N - c_n)}$$

we obtain that

$$\begin{aligned}\frac{1}{N} \text{tr}[(\mathbf{HSA}^2 \mathbf{S}^\dagger)^\dagger \mathbf{X}^{(j)}] &= \frac{1}{N} \sum_{n=1}^{\beta^* N} \text{tr}[(\mathbf{u}_n + d_n c_n \mathbf{v}_n)^\dagger \mathbf{X}^{(j)}] \\ &\asymp \tau_j^N (\beta^* (\mathcal{H}_{0,2}^N - 1) + \varsigma) + \beta^* |z|^{-2} |\alpha \bar{p} - \tau_1^N|^2 \mathcal{H}_{j-1,2}^N\end{aligned}\quad (185)$$

for  $j = 2, 3$ , also using (180). Combining (184) and (185) according to (130) for isometric  $\mathbf{S}$ , and solving for  $\tau_j^N$  gives

$$\tau_j^N \asymp \frac{\alpha \rho_j^N \mathcal{E}_{2,2}^N - \beta^* |z|^{-2} |\alpha \bar{p} - \tau_1^N|^2 \mathcal{H}_{j-1,2}^N}{\beta^* (\mathcal{H}_{0,2}^N - 1) + 1}, \quad j = 2, 3. \quad (186)$$

It follows (using the same arguments outlined after (171) in the proof of Theorem 1, this time assuming that (25)–(28) has a unique solution) from Theorem 1, (106), (107), (182), (183), (184), and (186), that for all  $z \in \mathbb{C}^+$ ,  $|\rho_j^N - \rho_j| \xrightarrow{a.s.} 0$ ,  $j = 2, 3, 4$ , and  $|\tau_j^N - \tau_j| \xrightarrow{a.s.} 0$ ,  $j = 2, 3$ , where  $\rho_j$  and  $\tau_j \in \mathbb{R}^*$  are solutions to (25)–(28). In addition, since  $\rho_{4,k}^N + \sigma^2 \rho_{2,k}^N \stackrel{k}{\asymp} \rho_4^N + \sigma^2 \rho_2^N$  and  $\rho_{1,k}^N \stackrel{k}{\asymp} \rho_1^N$ , the SINR satisfies (24).

#### APPENDIX VI

##### PROOF OF (47): CONVERGENCE OF THE E.D.F. OF $\mathbf{W}$ FOR EXPONENTIAL WEIGHTING

With exponential weighting, the e.d.f. of the  $i \times i$  matrix  $\mathbf{W}$  corresponds to the distribution of a random variable,  $W_i$ , which is uniformly distributed on the set  $\{\epsilon^{i(1-j/i)} : j = 1, \dots, i\}$ , where  $\epsilon = (1 - \frac{\eta}{L})$  is the exponential weighting constant. The corresponding distribution function is  $F_{W_i}(w)$ . To prove convergence in distribution of the a.e.d. of  $\mathbf{W}$  to  $F_W(w)$ , we show that  $\lim_{i \rightarrow \infty} F_{W_i}(w) = F_W(w)$  where  $F_W(w)$  is given in (47), i.e., pointwise convergence. We have

$$\begin{aligned} F_{W_i}(w) &= \Pr(W_i \leq w) = \Pr(\epsilon^{i(1-J_i)} \leq w) \\ &= \Pr\left(J_i \leq 1 - \frac{\ln w}{i \ln \epsilon}\right) = \frac{1}{i} \sum_{j=1}^i u\left(\frac{j}{i} - 1 + \frac{\ln w}{i \ln \epsilon}\right) \end{aligned} \quad (187)$$

for  $\epsilon^i \leq w < 1$ , where  $J_i$  is a discrete random variable uniformly distributed on the set  $\{j/i : j = 1, \dots, i\}$ , and where  $u(t)$  is the step function, i.e.,  $u(t)$  is zero for  $t < 0$  and unity for  $t \geq 0$ . Now,

$$\begin{aligned} \lim_{i \rightarrow \infty} i \ln \epsilon &= \lim_{i \rightarrow \infty} \frac{\log(1 - \frac{\eta}{L})}{i^{-1}} = \frac{\lim_{i \rightarrow \infty} \frac{\partial}{\partial i} \log(1 - \frac{\eta}{L})}{\lim_{i \rightarrow \infty} \frac{\partial}{\partial i} i^{-1}} \\ &= \frac{\lim_{i \rightarrow \infty} \frac{\eta}{L i^2} \frac{\bar{L}}{L - \frac{\eta}{i}}}{\lim_{i \rightarrow \infty} -i^{-2}} = -\eta/\bar{L} \end{aligned}$$

and similarly, the limit of the lower bound on  $w$  simplifies to  $\lim_{i \rightarrow \infty} \epsilon^i = e^{-\eta/\bar{L}}$ . Taking the limit of the Riemann sum in (187) then gives

$$\lim_{i \rightarrow \infty} F_{W_i}(w) = \int_{1 + \frac{\bar{L}}{\eta} \ln w}^1 dt = 1 + \frac{\bar{L}}{\eta} \ln w$$

where we have also used  $\lim_{i \rightarrow \infty} F_{W_i}(w = 1) = 1$ . This establishes that as  $i \rightarrow \infty$ ,  $F_i(w)$  converges in distribution to  $F(w) = 1 + \frac{\bar{L}}{\eta} \ln w$  for  $e^{-\eta/\bar{L}} < w < 1$ .

#### APPENDIX VII

##### PROOF OF THEOREM 3

As in Appendix III, the analysis in this appendix is based on removing a single dimension from matrices and vectors, as described in Section III. We will use  $t_n$  when removing the  $n^{\text{th}}$  transmit dimension,  $d_k$  when removing the  $k^{\text{th}}$  data stream, and  $r_m$  for removing the  $m^{\text{th}}$  received symbol interval.

We define  $\stackrel{k}{\asymp}$ ,  $\stackrel{n}{\asymp}$ , and  $\stackrel{m}{\asymp}$  according to Definition 2 in Appendix I, where the maximum is over  $k \leq K$ ,  $n \leq N$ , and  $m \leq i$ , respectively, and the limit is as  $(M, N, K, i) \rightarrow \infty$  with  $K/N \rightarrow \alpha > 0$ ,  $i/N \rightarrow \eta > 0$ , and  $M/N \rightarrow \beta > 0$  constant, as described in Section II-E.

#### A. Definitions

As in Appendix III, throughout this appendix, we substitute  $\mathbf{H}$  with  $\mathbf{VD}$  without loss of generality, where  $\mathbf{V}$  is an  $M \times M$  Haar-distributed unitary matrix,  $\mathbf{D}$  is a diagonal  $M \times N$  matrix containing the singular values of  $\mathbf{H}$ . The justification for this substitution in this case will be established later in Proposition 3. We denote the  $n^{\text{th}}$  column of  $\mathbf{V}$  as  $\mathbf{v}_n$ , and define  $\mathbf{v}_n = \mathbf{0}$  for  $n > M$ . Also define  $d_n$  via  $\text{diag}(\mathbf{D}) = \{d_1, \dots, d_{\beta^* N}\}$ , and let  $d_n = 0$  for  $n > \beta^* N$ .

Define the following quantities

$$\hat{r}_{j,m}^N = \frac{1}{i} \mathbf{r}_m^\dagger \hat{\mathbf{X}}_{r_m}^{(j)} \mathbf{r}_m, \quad j = 1, \dots, 4, \quad 0 < m \leq i, \quad (188)$$

$$\hat{\gamma}_{j,m}^N = \tilde{\mathbf{n}}_m^\dagger \hat{\mathbf{X}}_{r_m}^{(j)} \tilde{\mathbf{n}}_m, \quad j = 1, \dots, 4, \quad 0 < m \leq i, \quad (189)$$

$$\hat{\omega}_{j,m}^N = \boldsymbol{\omega}_m^\dagger \hat{\mathbf{X}}_{r_m}^{(j)} \boldsymbol{\omega}_m, \quad j = 1, \dots, 4, \quad 0 < m \leq i, \quad (190)$$

$$\hat{\rho}_{j,k}^N = \mathbf{h}_k^\dagger \hat{\mathbf{X}}_{d_k}^{(j)} \mathbf{h}_k, \quad j = 1, \dots, 4, \quad 0 < k \leq K, \quad (191)$$

$$\hat{\psi}_{j,k}^N = \mathbf{q}_k^\dagger \hat{\mathbf{X}}_{d_k}^{(j)} \mathbf{q}_k, \quad j = 1, \dots, 4, \quad 0 < k \leq K, \quad (192)$$

$$\hat{\tau}_{j,n}^N = \boldsymbol{\tau}_n^\dagger \hat{\mathbf{X}}_{t_n}^{(j)} \boldsymbol{\tau}_n, \quad j = 1, 2, 3, \quad 0 < n \leq \beta^* N, \quad (193)$$

$$\hat{\nu}_{j,n}^N = \mathbf{v}_n^\dagger \hat{\mathbf{X}}_{t_n}^{(j)} \mathbf{v}_n, \quad j = 1, 2, 3, \quad 0 < n \leq \beta^* N, \quad (194)$$

where

$$\tilde{\mathbf{n}}_m = \frac{1}{\sqrt{i}} \mathbf{n}_m,$$

$$\boldsymbol{\omega}_m = \frac{1}{\sqrt{i}} \mathbf{H} \mathbf{S} \mathbf{A} \mathbf{b}_m,$$

$$\mathbf{h}_k = \mathbf{H} \mathbf{s}_k,$$

$$\mathbf{q}_k = \frac{1}{i} \mathcal{R}_{d_k} \mathbf{W} \mathbf{b}_k,$$

$$\boldsymbol{\tau}_n = \frac{1}{i} \mathcal{R}_{t_n} \mathbf{W} \mathbf{B} \tilde{\mathbf{s}}_n,$$

and

$$\hat{\mathbf{X}}^{(j)} = \begin{cases} \hat{\mathbf{R}}^{-1} & , j = 1, \\ \hat{\mathbf{R}}^{-\dagger} \hat{\mathbf{R}}^{-1} & , j = 2, \\ \hat{\mathbf{R}}^{-\dagger} \mathbf{H}^\dagger \hat{\mathbf{R}}^{-1} & , j = 3, \\ \hat{\mathbf{R}}^{-\dagger} (\mathbf{H} \mathbf{S} \mathbf{A})^\dagger \hat{\mathbf{R}}^{-1} & , j = 4. \end{cases}$$

Further, we define  $\hat{\mathbf{X}}_{r_m}^{(j)}$ ,  $\hat{\mathbf{X}}_{d_k}^{(j)}$ , and  $\hat{\mathbf{X}}_{t_n}^{(j)}$  by removing the contribution of the  $m^{\text{th}}$  training symbol, the  $k^{\text{th}}$  data stream, and the  $n^{\text{th}}$  transmit dimension, respectively, from  $\hat{\mathbf{X}}^{(j)}$  as follows:

- To remove the contribution of the  $m^{\text{th}}$  received training symbol interval, for some  $0 < m \leq i$ , replace  $\hat{\mathbf{R}}$  and  $\mathcal{R}$  with  $\hat{\mathbf{R}}_{r_m}$  and  $\mathcal{R}_{r_m}$ , respectively, where

$$\hat{\mathbf{R}}_{r_m} = \mathcal{R}_{r_m} \mathbf{W}_{r_m} \mathcal{R}_{r_m}^\dagger - z \mathbf{I}_M \quad (195)$$

and  $\mathcal{R}_{r_m}$  and  $\mathbf{W}_{r_m}$  are  $\mathcal{R}$  and  $\mathbf{W}$  with the  $m^{\text{th}}$  column, and  $m^{\text{th}}$  row and column removed, respectively. That is,  $\hat{\mathbf{R}}_{r_m} = \hat{\mathbf{R}} - \frac{w_m}{i} \mathbf{r}_m^\dagger$ .

- To remove the contribution of the  $k^{\text{th}}$  data stream, for some  $0 < k \leq K$ , replace  $\hat{\mathbf{R}}$ ,  $\mathcal{R}$ ,  $\mathbf{S}$ ,  $\mathbf{A}$ , and  $\mathbf{B}$  with  $\hat{\mathbf{R}}_{d_k}$ ,  $\mathcal{R}_{d_k}$ ,  $\mathbf{S}_{d_k}$ ,  $\mathbf{A}_{d_k}$ , and  $\mathbf{B}_{d_k}$ , respectively, where

$$\hat{\mathbf{R}}_{d_k} = \frac{1}{i} \mathcal{R}_{d_k} \mathbf{W} \mathcal{R}_{d_k}^\dagger - z \mathbf{I}_M,$$

$$\mathcal{R}_{d_k} = \mathbf{H} \mathbf{S}_{d_k} \mathbf{A}_{d_k} \mathbf{B}_{d_k}^\dagger + \mathbf{N},$$

and  $\mathbf{S}_{d_k}$  and  $\mathbf{B}_{d_k}$  are  $\mathbf{S}$  and  $\mathbf{B}$  with their  $k^{\text{th}}$  columns removed, respectively, and  $\mathbf{A}_{d_k}$  is  $\mathbf{A}$  with the  $k^{\text{th}}$  column and row removed.

- To remove the contribution of the  $n^{\text{th}}$  transmit dimension, for some  $0 < n \leq \beta^* N$ , replace  $\hat{\mathbf{R}}, \mathcal{R}, \mathbf{H}$ , and  $\mathbf{S}$  with  $\hat{\mathbf{R}}_{t_n}, \mathcal{R}_{t_n}, \mathbf{H}_{t_n}$ , and  $\mathbf{S}_{t_n}$ , respectively, where

$$\begin{aligned}\hat{\mathbf{R}}_{t_n} &= \frac{1}{i} \mathcal{R}_{t_n} \mathbf{W} \mathcal{R}_{t_n}^\dagger - z \mathbf{I}_M, \\ \mathcal{R}_{t_n} &= \mathbf{H}_{t_n} \mathbf{S}_{t_n} \mathbf{A} \mathbf{B}^\dagger + \mathbf{N}, \\ \mathbf{H}_{t_n} &= \mathbf{V}_{t_n} \mathbf{D}_{t_n},\end{aligned}$$

and  $\mathbf{V}_{t_n}$  and  $\mathbf{S}_{t_n}$  are  $\mathbf{V}$  and  $\mathbf{S}$  with the  $n^{\text{th}}$  column and row removed, respectively. Also,  $\mathbf{D}_{t_n}$  is  $\mathbf{D}$  with the  $n^{\text{th}}$  column and row removed.

Considering (189)–(194) asymptotically, it can be shown using the same steps as in the proof of (129) and (130) in Appendix IV that

$$\frac{\eta}{\beta \sigma^2} \hat{\gamma}_{j,m}^N \stackrel{m}{\asymp} \hat{\gamma}_j^N = \frac{1}{M} \text{tr}[\hat{\mathbf{X}}^{(j)}], \quad (196)$$

$$\hat{r}_{j,m}^N \stackrel{m}{\asymp} \hat{r}_j^N = \hat{\omega}_j + \sigma^2 \frac{\beta}{\eta} \hat{\gamma}_j^N, \quad (197)$$

$$\hat{\omega}_{j,m}^N \stackrel{m}{\asymp} \hat{\omega}_j^N = \begin{cases} \frac{1}{i} \text{tr}[(\mathbf{HSA})^\dagger \hat{\mathbf{X}}^{(j)}] & , \text{i.i.d. } \mathbf{B}, \\ \frac{1}{i} \sum_{m=1}^i \hat{\omega}_{j,m}^N & , \text{ort. } \mathbf{B}, \alpha < \eta, \\ \frac{K}{i(K-i)} \text{tr}[(\mathbf{HSA}\mathbf{\Upsilon})^\dagger \hat{\mathbf{X}}^{(j)}] & , \text{ort. } \mathbf{B}, \alpha > \eta, \end{cases} \quad (198)$$

and

$$\hat{\rho}_{j,k}^N \stackrel{k}{\asymp} \hat{\rho}_j^N = \begin{cases} \frac{1}{N} \text{tr}[\mathbf{H}^\dagger \hat{\mathbf{X}}^{(j)}] & , \text{i.i.d. } \mathbf{S}, \\ \frac{1}{N-K} \text{tr}[(\mathbf{H}\mathbf{\Pi})^\dagger \hat{\mathbf{X}}^{(j)}] & , \text{iso. } \mathbf{S}, \end{cases} \quad (199)$$

$$\hat{\psi}_{j,k}^N \stackrel{k}{\asymp} \hat{\psi}_j^N = \begin{cases} \frac{1}{i^2} \text{tr}[(\mathcal{R}\mathbf{W})^\dagger \hat{\mathbf{X}}^{(j)}] & , \text{i.i.d. } \mathbf{B}, \\ \frac{1}{i(i-K)} \text{tr}[(\mathcal{R}\mathbf{W}\mathbf{\Upsilon})^\dagger \hat{\mathbf{X}}^{(j)}] & , \text{ort. } \mathbf{B}, \alpha < \eta, \\ \frac{1}{K} \sum_{k=1}^K \hat{\psi}_{j,k}^N & , \text{ort. } \mathbf{B}, \alpha > \eta, \end{cases} \quad (200)$$

and

$$\hat{r}_{j,n}^N \stackrel{n}{\asymp} \hat{r}_j^N = \begin{cases} \frac{1}{N^2} \text{tr}[(\mathcal{R}\mathbf{W}\mathbf{B}\mathbf{A})^\dagger \hat{\mathbf{X}}^{(j)}] & , \text{i.i.d. } \mathbf{S}, \\ \frac{1}{N} \sum_{n=1}^N \hat{r}_{j,n}^N & , \text{iso. } \mathbf{S}, \end{cases} \quad (201)$$

$$\hat{\nu}_{j,n}^N \stackrel{n}{\asymp} \hat{\nu}_j^N = \frac{1}{\beta^* N} \sum_{n=1}^{\beta^* N} \hat{\nu}_{j,n}^N, \quad (202)$$

where  $\mathbf{\Pi}$  is defined in (131) (with  $\bar{N} = N$  and hence  $\mathbf{\Pi}^2 = \mathbf{\Pi}$ ), and

$$\mathbf{\Upsilon} = \begin{cases} \mathbf{I}_i - \frac{1}{i} \mathbf{B}\mathbf{B}^\dagger & , \alpha < \eta, \\ \mathbf{I}_K - \frac{1}{K} \mathbf{B}^\dagger \mathbf{B} & , \alpha > \eta. \end{cases}$$

Although the derivations of the uniform asymptotic equivalence in (199)–(202) are not shown, note that they rely on expressions derived later in this appendix, namely (231)–(233) and (244)–(246).

Also, analogous to the discussion in Appendix III-C, we assume that  $\hat{\nu}_{j,n}^N$  is uniformly asymptotically equivalent to another sequence, which does not depend on  $n$ , in which case  $\hat{\nu}_{j,n}^N \stackrel{n}{\asymp} \hat{\nu}_j^N$ , as stated in (202). Applying the shortcut described

in Appendix III-C will give an asymptotic expression for  $\hat{\nu}_j^N$ , which, in the large-system limit, leads to (38) in Theorem 3. To make this rigorous, we could introduce a new independent parameter  $\bar{N}$ , corresponding to the number of columns in  $\mathbf{V}$ . Then, the large-system limit would also contain the condition  $\bar{N} \rightarrow \infty$  with  $\bar{N}/N \rightarrow \varsigma < 1$ , and (38) is obtained by letting  $\varsigma \rightarrow 1^-$  (after taking the large-system limit). A similar argument is applied to the other variables  $\hat{\omega}_{j,m}^N \stackrel{m}{\asymp} \hat{\omega}_j^N$  (for ort.  $\mathbf{B}$ ,  $\alpha < \eta$ ),  $\hat{\psi}_{j,k}^N \stackrel{k}{\asymp} \hat{\psi}_j^N$  (for ort.  $\mathbf{B}$ ,  $\alpha > \eta$ ), and  $\hat{r}_{j,n}^N \stackrel{n}{\asymp} \hat{r}_j^N$  (for iso.  $\mathbf{S}$ ).

Note that (133) again applies, and also

$$\bar{w}_k \stackrel{k}{\asymp} \bar{w}, \quad (203)$$

$$\hat{c}_n \stackrel{n}{\asymp} \alpha \bar{w} \bar{p}, \quad (204)$$

$$\frac{1}{i} \mathbf{r}_m^\dagger \mathbf{r}_m \stackrel{m}{\asymp} \alpha \beta^* \eta^{-1} \bar{p} \mathbf{E}[H] + \sigma^2 \beta \eta^{-1} > 0, \quad (205)$$

for  $0 < k \leq K$ ,  $0 < n \leq N$ , and  $0 < m \leq i$ , where

$$\bar{w}_k = \frac{1}{i} \mathbf{b}_k^\dagger \mathbf{W} \mathbf{b}_k,$$

$$\hat{c}_n = \frac{1}{i} \mathbf{s}_n^\dagger \mathbf{A} \mathbf{B}^\dagger \mathbf{W} \mathbf{B} \mathbf{A} \mathbf{s}_n,$$

$\bar{p} = \mathbf{E}[P]$ , and  $\bar{w} = \mathbf{E}[W]$ . (203) is shown in identical manner to (133) and (134), in this case using (108). (204) can be shown in an identical manner to (134), also using (203), i.e.,

$$\hat{c}_n \stackrel{n}{\asymp} \frac{1}{iN} \text{tr}[\mathbf{W} \mathbf{B} \mathbf{A}^2 \mathbf{B}^\dagger] = \frac{1}{N} \sum_{k=1}^K P_k \bar{w}_k \asymp \alpha \bar{w} \bar{p}.$$

Finally, (205) can be shown from  $\frac{1}{i} \mathbf{r}_m^\dagger \mathbf{r}_m \stackrel{m}{\asymp} \frac{1}{i} \text{tr}[(\mathbf{HSA})^\dagger + \sigma^2 \mathbf{I}_M]$  using (106), (133), and  $i |\mathbf{n}_m^\dagger \boldsymbol{\omega}_m|^2 \stackrel{m}{\asymp} \text{tr}[(\mathbf{HSA})^\dagger] \asymp \frac{\alpha \beta^*}{\eta} \mathbf{E}[P] \mathbf{E}[H]$ .

We now give several bounds on particular matrix and vector norms which are required in order to apply Lemmas 2 and 3 later. Firstly, the assumption that  $z \in \mathbb{C}^+$  gives

$$\|\hat{\mathbf{R}}^{-1}\| \leq \text{Im}(z)^{-1}. \quad (206)$$

Recall that assumptions on  $\mathbf{H}$ ,  $\mathbf{S}$ , and  $\mathbf{A}$  outlined in Section II again give (136)–(140), and additionally the assumptions on  $\mathbf{W}$ ,  $\mathbf{B}$  and  $\mathbf{N}$  give

$$\sup_N \|\mathbf{W}\| < \infty, \quad (207)$$

$$\frac{1}{i} \|\mathbf{B}\|^2 \asymp (1 + \sqrt{\alpha \eta^{-1}})^2, \quad (\text{i.i.d. } \mathbf{B}), \quad (208)$$

$$\frac{1}{i} \|\mathbf{N}\|^2 \asymp \sigma^2 (1 + \sqrt{\alpha \eta^{-1}})^2, \quad (209)$$

$$\frac{1}{i} |\mathbf{b}_k|^2 \stackrel{k}{\asymp} 1, \quad (210)$$

$$\frac{1}{i} |\mathbf{b}_m|^2 \stackrel{m}{\asymp} \alpha \eta^{-1}, \quad (211)$$

$$\frac{1}{i} |\mathbf{n}_m|^2 \stackrel{m}{\asymp} \beta \eta^{-1}, \quad (212)$$

where (208)–(209), like (138), is due to [34]. Of course,  $\frac{1}{N} \|\mathbf{B}\|^2 = \eta^*$  for orthogonal  $\mathbf{B}$ . Moreover, (136)–(140) and

(206)–(212) imply that all of

$$\|\hat{\mathbf{R}}\|, \quad (213)$$

$$\max_{k \leq K} \max\{|\mathbf{h}_k|, |\mathbf{q}_k|, \bar{w}_k\}, \quad (214)$$

$$\max_{n \leq N} \max\{|\boldsymbol{\tau}_n|, |\hat{c}_n|\}, \quad (215)$$

$$\max\{|\hat{r}_j^N|, \max_{m \leq i} |\hat{r}_{j,m}^N|\}, \quad \text{for } j = 1, \dots, 4, \quad (216)$$

$$\max\{|\hat{\rho}_j^N|, \max_{k \leq K} |\hat{\rho}_{j,k}^N|\}, \quad \text{for } j = 1, \dots, 4, \quad (217)$$

$$\max\{|\hat{\psi}_j^N|, \max_{k \leq K} |\hat{\psi}_{j,k}^N|\}, \quad \text{for } j = 1, \dots, 4, \quad (218)$$

$$\max\{|\hat{\tau}_j^N|, \max_{n \leq N} |\hat{\tau}_{j,n}^N|\}, \quad \text{for } j = 1, \dots, 3, \quad (219)$$

$$\max\{|\hat{\nu}_j^N|, \max_{n \leq N} |\hat{\nu}_{j,n}^N|\}, \quad \text{for } j = 1, \dots, 3, \quad (220)$$

are bounded above almost surely, that is, have an upper bound of form given in (94). For this to be the case for (213), it also requires the assumption that  $|z| < \infty$ .

### B. Derivations

Note that  $\hat{\gamma}_1^N = \frac{1}{M} \text{tr}[\hat{\mathbf{R}}^{-1}]$  is the Stieltjés transform of the e.d.f. of the eigenvalues of  $\frac{1}{i} \mathcal{R} \mathbf{W} \mathcal{R}^\dagger$ . The proof of Theorem 3 proceeds as described in Section III, by applying Lemma 6 to remove the effect of the  $m^{\text{th}}$  training interval,  $k^{\text{th}}$  data stream, and  $n^{\text{th}}$  transmit dimension from  $\hat{\mathbf{R}}$ ,  $0 < m \leq i$ ,  $0 < k \leq K$ ,  $0 < n \leq \beta^* N$ . The removal of each dimension generates a pair of variables; expressions for which are then derived. Additionally, these results are applied to expansions of the identity  $\hat{\mathbf{R}} \hat{\mathbf{R}}^{-1} = \mathbf{I}_M$  to yield relationships between  $G_{\hat{\mathbf{R}}}^N(z)$  and the variables generated.

1) *Expanding Dimension  $i$* : From Lemma 6, (188), (195), (205), and (213) we have that

$$\hat{\mathbf{R}}^{-1} \mathbf{r}_m \asymp R_m \hat{\mathbf{R}}_{r_m}^{-1} \mathbf{r}_m, \quad (221)$$

$$\inf_N \max_{m \leq i} |1 + w_m \hat{r}_{1,m}^N| > 0, \quad \text{a.s.}, \quad (222)$$

where  $R_m = 1/(1 + w_m \hat{r}_{1,m}^N)$ . In addition, since

$$\boldsymbol{\omega}_m^\dagger \hat{\mathbf{R}}_{r_m}^{-1} \tilde{\mathbf{n}}_m \stackrel{m}{\asymp} \tilde{\mathbf{n}}_m^\dagger \hat{\mathbf{R}}_{r_m}^{-1} \boldsymbol{\omega}_m \stackrel{m}{\asymp} 0 \quad (223)$$

due to  $i \left| \boldsymbol{\omega}_m^\dagger \hat{\mathbf{R}}_{r_m}^{-1} \tilde{\mathbf{n}}_m \right| \stackrel{m}{\asymp} \sigma^2 \hat{\omega}_2$ , we have from the matrix inversion lemma and (223) that

$$\hat{\mathbf{R}}^{-1} \boldsymbol{\omega}_m \asymp R_m \hat{\mathbf{R}}_{r_m}^{-1} \left( (1 + w_m \hat{\gamma}_{1,m}^N) \boldsymbol{\omega}_m - w_m \hat{\omega}_{1,m}^N \tilde{\mathbf{n}}_m \right), \quad (224)$$

$$\hat{\mathbf{R}}^{-1} \tilde{\mathbf{n}}_m \asymp R_m \hat{\mathbf{R}}_{r_m}^{-1} \left( -w_m \hat{\gamma}_{1,m}^N \boldsymbol{\omega}_m + (1 + w_m \hat{\omega}_{1,m}^N) \tilde{\mathbf{n}}_m \right). \quad (225)$$

Now, we expand the identity  $\hat{\mathbf{R}}^{-1} \hat{\mathbf{R}} = \mathbf{I}_M$  along dimension  $i$  using (221). We have that

$$1 = \frac{1}{M} \text{tr}[\hat{\mathbf{R}}^{-1} \hat{\mathbf{R}}] = -z \hat{\gamma}_1^N + \frac{1}{M} \sum_{m=1}^i \frac{w_m \hat{r}_{1,m}^N}{1 + w_m \hat{r}_{1,m}^N}. \quad (226)$$

It follows from Lemma 3, (197), (216), (207), and (222) that

$$\frac{w_m \hat{r}_{1,m}^N}{1 + w_m \hat{r}_{1,m}^N} \stackrel{m}{\asymp} \frac{w_m \hat{r}_1^N}{1 + w_m \hat{r}_1^N}. \quad (227)$$

Moreover, from (226), (227), and Lemma 4, we have

$$1 + z \hat{\gamma}_1^N \asymp \begin{cases} \frac{\eta}{\beta} (1 - \mathcal{W}_{0,1}^N) \\ \frac{\eta}{\beta} \hat{r}_1^N \mathcal{W}_{1,1}^N \end{cases} \quad (228)$$

where the expressions in the right hand side of (228) are related via (54), and

$$\mathcal{W}_{p,1}^N = \frac{1}{i} \sum_{m=1}^i \frac{w_m^p}{1 + w_m \hat{r}_1^N}.$$

2) *Expanding Dimension  $K$* : In this case, we write

$$\begin{aligned} \hat{\mathbf{R}} &= \frac{1}{i} (\mathcal{R}_{d_k} + A_k \mathbf{h}_k \mathbf{b}_k^\dagger) \mathbf{W} (\mathcal{R}_{d_k} + A_k \mathbf{h}_k \mathbf{b}_k^\dagger)^\dagger - z \mathbf{I}_M \\ &= \hat{\mathbf{R}}_{d_k} + A_k \mathbf{q}_k \mathbf{h}_k^\dagger + A_k \mathbf{h}_k \mathbf{q}_k^\dagger + \bar{w}_k P_k \mathbf{h}_k \mathbf{h}_k^\dagger. \end{aligned} \quad (229)$$

We now apply Lemma 6 to (229), where  $\mathbf{Y}_N$ ,  $\mathbf{X}_N$ ,  $\mathbf{u}_N$ ,  $\mathbf{v}_N$ , and  $c_N$  in the Lemma correspond to  $\hat{\mathbf{R}}$ ,  $\hat{\mathbf{R}}_{d_k}$ ,  $\mathbf{h}_k$ ,  $A_k \mathbf{q}_k$ , and  $P_k \bar{w}_k$ , respectively. Therefore,  $u_N$  and  $v_N$  in the Lemma correspond to  $\hat{\rho}_{1,k}^N$  and  $P_k \hat{\psi}_{1,k}^N$ , respectively. Note that

$$\mathbf{h}_k^\dagger \hat{\mathbf{R}}_{d_k}^{-1} \mathbf{q}_k \stackrel{k}{\asymp} \mathbf{q}_k^\dagger \hat{\mathbf{R}}_{d_k}^{-1} \mathbf{h}_k \stackrel{k}{\asymp} 0 \quad (230)$$

due to  $N \left| \mathbf{h}_k^\dagger \hat{\mathbf{R}}_{d_k}^{-1} \mathbf{q}_k \right|^2 \stackrel{k}{\asymp} \hat{\psi}_{3,k}^N$  and (218) for i.i.d.  $\mathbf{S}$ , and similarly with isometric  $\mathbf{S}$ , it can be shown  $(N - K) \left| \mathbf{h}_k^\dagger \hat{\mathbf{R}}_{d_k}^{-1} \mathbf{q}_k \right|^2$  has a finite, almost-sure, uniform upper bound. And so, (230) satisfies condition (95) of the Lemma. Additionally, (133) and (213)–(214) satisfy conditions (96) and (97) of the Lemma, respectively, in the almost-sure sense.

Therefore, from Lemma 6 we have

$$\hat{\mathbf{R}}^{-1} \mathbf{h}_k \asymp K_k \hat{\mathbf{R}}_{d_k}^{-1} (\mathbf{h}_k - A_k \hat{\rho}_{1,k}^N \mathbf{q}_k), \quad (231)$$

$$\hat{\mathbf{R}}^{-1} \mathbf{q}_k \asymp K_k \hat{\mathbf{R}}_{d_k}^{-1} (-A_k \hat{\psi}_{1,k}^N \mathbf{h}_k + (1 + \bar{w}_k P_k \hat{\rho}_{1,k}^N) \mathbf{q}_k), \quad (232)$$

$$\inf_N \min_{k \leq K} \left| 1 - P_k \hat{\rho}_{1,k}^N (\hat{\psi}_{1,k}^N - \bar{w}_k) \right| > 0, \quad \text{a.s.}, \quad (233)$$

where  $K_k = 1/(1 - P_k \hat{\rho}_{1,k}^N (\hat{\psi}_{1,k}^N - \bar{w}_k))$ .

Now, we expand the identity  $\hat{\mathbf{R}}^{-1} \hat{\mathbf{R}} = \mathbf{I}_M$  along dimension  $K$  using (231)–(232). We have

$$1 + z \hat{\gamma}_1^N = \frac{1}{iM} \text{tr}[\hat{\mathbf{R}}^{-1} \mathbf{H} \mathbf{S} \mathbf{A} \mathbf{B}^\dagger \mathbf{W} \mathcal{R}^\dagger] + \frac{1}{iM} \text{tr}[\hat{\mathbf{R}}^{-1} \mathbf{N} \mathbf{W} \mathcal{R}^\dagger]. \quad (234)$$

Expanding the first term in (234) with respect to dimension  $K$  using (231)–(232), we have

$$\begin{aligned} \frac{1}{iM} \text{tr}[\hat{\mathbf{R}}^{-1} \mathbf{H} \mathbf{S} \mathbf{A} \mathbf{B}^\dagger \mathbf{W} \mathcal{R}^\dagger] &= \frac{1}{iM} \sum_{k=1}^K A_k \mathbf{b}_k^\dagger \mathbf{W} \mathcal{R}^\dagger \hat{\mathbf{R}}^{-1} \mathbf{h}_k \\ &\asymp \frac{1}{M} \sum_{k=1}^K A_k K_k (A_k \bar{w}_k \mathbf{h}_k + \mathbf{q}_k)^\dagger \hat{\mathbf{R}}_{d_k}^{-1} (\mathbf{h}_k - A_k \hat{\rho}_{1,k}^N \mathbf{q}_k) \\ &\asymp \begin{cases} \frac{\alpha}{\beta} (1 - \hat{\mathcal{E}}_{0,1}^N) \\ -\frac{\alpha}{\beta} \hat{\rho}_1^N (\hat{\psi}_1^N - \bar{w}) \hat{\mathcal{E}}_{1,1}^N \end{cases} \end{aligned} \quad (235)$$

where we have used (199), (200), (203), (230), and (233), and define

$$\hat{\mathcal{E}}_{p,1}^N = \frac{1}{K} \sum_{k=1}^K \frac{P_k^p}{1 - P_k \hat{\rho}_1^N (\hat{\psi}_1^N - \bar{w})}.$$

Now considering the second term in (234), using (221), (223), and (196) we see

$$\begin{aligned} \frac{1}{iM} \text{tr}[\hat{\mathbf{R}}^{-1} \mathbf{N} \mathbf{W} \mathcal{R}^\dagger] &= \frac{1}{M} \sum_{m=1}^i \frac{w_m (\boldsymbol{\omega}_m + \tilde{\mathbf{n}}_m)^\dagger \hat{\mathbf{R}}_{r_m}^{-1} \tilde{\mathbf{n}}_m}{1 + w_m \hat{r}_{1,m}^N} \\ &\asymp \sigma^2 \hat{\gamma}_1^N \mathcal{W}_{1,1}^N. \end{aligned} \quad (236)$$

Combining (234), (235), and (236) gives

$$\beta(1 - \hat{\gamma}_1^N / \hat{\nu}_1^N) \asymp \alpha(1 - \hat{\mathcal{E}}_{0,1}^N) = -\alpha \hat{\rho}_1^N (\hat{\psi}_1^N - \bar{w}) \hat{\mathcal{E}}_{1,1}^N. \quad (237)$$

Note that up until this point, none of the analysis has relied on the fact that we have made the substitution of  $\mathbf{H}$  with  $\mathbf{VD}$ . Therefore, we are in a position to state the following proposition, which shows that this substitution is valid. Note that the analysis in the remainder of this appendix relies on this substitution having been made throughout.

*Proposition 3:* For the model (1), the distribution of both the Stieltjes transform of the e.e.d. of  $\frac{1}{i} \mathcal{R} \mathbf{W} \mathcal{R}^\dagger$  and the asymptotically equivalent ALS SINR given in (31) are invariant to the substitution of  $\mathbf{VD}$  for  $\mathbf{H}$ , where  $\mathbf{V}$  is an  $M \times M$  Haar-distributed random unitary matrix, and  $\mathbf{D}$  is a  $M \times N$  diagonal matrix containing the singular values of  $\mathbf{H}$ .

*Proof:* Using (231)–(232), we may derive the asymptotically equivalent form of the ALS SINR given in (31) (see Appendix VIII), which depends on  $\hat{\rho}_j^N$  and  $\hat{\psi}_j^N$ ,  $j = 1, \dots, 4$ . Now note that the distributions of  $\hat{\gamma}_1^N$ ,  $\hat{\rho}_j^N$  and  $\hat{\psi}_j^N$ ,  $j = 1, \dots, 4$  are unchanged by the substitution of  $\mathbf{H}$  by  $\mathbf{VD}$ . That is, let  $\mathbf{T}$  be an independent  $M \times M$  Haar-distributed random matrix. Then,

$$\begin{aligned} \hat{\gamma}_1^N &= \frac{1}{M} \text{tr}[\hat{\mathbf{R}}^{-1}] = \frac{1}{N} \text{tr}[\mathbf{T} \mathbf{T}^\dagger \mathbf{R}^{-1}] \\ &= \frac{1}{M} \text{tr}[(\mathbf{T} \mathcal{R}) \mathbf{W} (\mathbf{T} \mathcal{R})^\dagger - z \mathbf{I}_M]^{-1}. \end{aligned}$$

Note that  $\mathbf{T} \mathcal{R} = \mathbf{T} \mathbf{H} \mathbf{S} \mathbf{A} \mathbf{B}^\dagger + \mathbf{T} \mathbf{N}$ , and so, writing  $\mathbf{H} \mathbf{S} = (\mathbf{T} \mathbf{U}_M) \mathbf{D} (\mathbf{U}_N^\dagger \mathbf{S})$ , where  $\mathbf{U}_M \mathbf{D} \mathbf{U}_N^\dagger$  is the singular value decomposition of  $\mathbf{H}$ , the unitary invariance of  $\mathbf{T}$ ,  $\mathbf{S}$ , and  $\mathbf{N}$  implies the result for the Stieltjes transform of  $\hat{\mathbf{R}}$ . A similar treatment of  $\hat{\rho}_j^N$  and  $\hat{\psi}_j^N$  gives the result for the asymptotically equivalent form of the ALS SINR given in (31).  $\square$

3) *Expanding Dimension  $N$ :* Writing

$$\mathcal{R} = \sum_{n=1}^{\beta^* N} d_n \mathbf{v}_n (\mathbf{B} \mathbf{A} \tilde{\mathbf{s}}_n)^\dagger + \mathbf{N}$$

and substituting this into (8), in a similar manner to (229), we have for some  $1 \leq n \leq \beta^* N$ ,

$$\hat{\mathbf{R}} = \hat{\mathbf{R}}_{t_n} + d_n \mathbf{v}_n \boldsymbol{\tau}_n^\dagger + d_n \boldsymbol{\tau}_n \mathbf{v}_n^\dagger + d_n^2 \hat{c}_n \mathbf{v}_n \mathbf{v}_n^\dagger. \quad (238)$$

We now apply Lemma 6 to (238), where  $\mathbf{Y}_N$ ,  $\mathbf{X}_N$ ,  $\mathbf{u}_N$ ,  $\mathbf{v}_N$ , and  $c_N$  in the Lemma correspond to  $\hat{\mathbf{R}}$ ,  $\hat{\mathbf{R}}_{t_n}$ ,  $\mathbf{v}_n$ ,  $d_n \boldsymbol{\tau}_n$ , and  $d_n^2 \hat{c}_n$ , respectively. Therefore,  $u_N$  and  $v_N$  in the Lemma correspond to  $\hat{\nu}_{1,n}^N$  and  $d_n^2 \hat{r}_{1,n}^N$ , respectively. We have (213), (215), and  $|\mathbf{v}_n| = 1$ , which almost-surely satisfy conditions (96)–(97) of the Lemma. We now show

$$\mathbf{v}_n^\dagger \hat{\mathbf{R}}_{t_n}^{-1} \boldsymbol{\tau}_n \asymp \boldsymbol{\tau}_n^\dagger \hat{\mathbf{R}}_{t_n}^{-1} \mathbf{v}_n \asymp 0 \quad (239)$$

which, with (136) satisfies condition (95) of the Lemma. To see this, note that  $\mathbf{v}_n^\dagger \boldsymbol{\tau}_n \asymp 0$ , since

$$\begin{aligned} N |\mathbf{v}_n^\dagger \boldsymbol{\tau}_n|^2 &= \frac{N}{i^2} (\mathbf{v}_n^\dagger \mathbf{N} \mathbf{W} \mathbf{B} \mathbf{A} \tilde{\mathbf{s}}_n)^\dagger \asymp \frac{1}{i^2 M} \text{tr}[(\mathbf{N} \mathbf{W} \mathbf{B} \mathbf{A})^\dagger] \\ &\asymp \frac{\sigma^2 \alpha}{\eta} \mathbf{E}[P] \mathbf{E}[W^2]. \end{aligned} \quad (240)$$

Also, since  $\mathbf{H}_{t_n}^\dagger \mathbf{v}_n = \mathbf{0}$ , we have

$$\hat{\mathbf{R}}_{t_n} \mathbf{v}_n = \frac{1}{i} \mathcal{R}_{t_n} \mathbf{W} \mathbf{N}^\dagger \mathbf{v}_n - z \mathbf{v}_n. \quad (241)$$

From (241), we find

$$\begin{aligned} z \boldsymbol{\tau}_n^\dagger \hat{\mathbf{R}}_{t_n}^{-1} \mathbf{v}_n &\asymp \frac{1}{i} \boldsymbol{\tau}_n^\dagger \hat{\mathbf{R}}_{t_n}^{-1} \mathcal{R}_{t_n} \mathbf{W} \mathbf{N}^\dagger \mathbf{v}_n \\ &= \frac{1}{i} \sum_{m=1}^i w_m \mathbf{n}_m^\dagger \mathbf{v}_n \boldsymbol{\tau}_n^\dagger \hat{\mathbf{R}}_{t_n}^{-1} \mathbf{r}_m. \end{aligned} \quad (242)$$

Considering the argument in the preceding sum, note that

$$w_m \mathbf{n}_m^\dagger \mathbf{v}_n \boldsymbol{\tau}_n^\dagger \hat{\mathbf{R}}_{t_n}^{-1} \mathbf{r}_m \asymp \frac{w_m \sigma^2 \boldsymbol{\tau}_n^\dagger \hat{\mathbf{R}}_{t_n}^{-1} \mathbf{v}_n}{1 + w_m \hat{r}_{1,t_n}^N} \quad (243)$$

where we define  $\hat{\gamma}_{1,t_n}^N$ ,  $\hat{r}_{1,t_n}^N$ , and  $\hat{\omega}_{1,t_n}^N$  from  $\hat{\gamma}_1^N$ ,  $\hat{r}_1^N$ , and  $\hat{\omega}_1^N$  in (196), (197), and (198), respectively, with the contribution of transmit dimension  $n$  removed, as explained in Appendix VII-A. Returning to (242), and additionally using  $1 + z \hat{\gamma}_{1,t_n}^N \asymp \hat{r}_{1,t_n}^N \frac{1}{M} \sum_{m=1}^i \frac{w_m}{1 + w_m \hat{r}_{1,t_n}^N}$ , which is shown in an identical manner to (228), it follows that

$$\boldsymbol{\tau}_n^\dagger \hat{\mathbf{R}}_{t_n}^{-1} \mathbf{v}_n \frac{z \hat{\omega}_{1,t_n}^N}{\hat{r}_{1,t_n}^N} \underset{n}{\asymp} 0.$$

Now, since  $z$ ,  $\hat{\omega}_{1,t_n}^N$ , and  $\hat{r}_{1,t_n}^N$  are in  $\mathbb{C}^+$  almost surely (through an application of Lemma 5), and also since  $|\hat{r}_{1,t_n}^N|$  is almost-surely uniformly bounded above over  $N$  and  $n \leq N$  (which is shown identically to (216)), we have (239).

So we may apply Lemma 6 to (238), to obtain

$$\hat{\mathbf{R}}^{-1} \mathbf{v}_n \asymp V_n \hat{\mathbf{R}}_{t_n}^{-1} (\mathbf{v}_n - d_n \hat{\nu}_{1,n}^N \boldsymbol{\tau}_n), \quad (244)$$

$$\hat{\mathbf{R}}^{-1} \boldsymbol{\tau}_n \asymp V_n \hat{\mathbf{R}}_{t_n}^{-1} (-d_n \hat{r}_{1,n}^N \mathbf{v}_n + (1 + d_n^2 \hat{c}_n \hat{\nu}_1) \boldsymbol{\tau}_n), \quad (245)$$

$$\inf_{N} \min_{n \leq N} |1 - d_n^2 \hat{\nu}_{1,n}^N (\hat{r}_{1,n}^N - \hat{c}_n)| > 0 \quad \text{, a.s.,} \quad (246)$$

for  $1 \leq n \leq \beta^* N$ , where  $V_n = 1/(1 - d_n^2 \hat{\nu}_{1,n}^N (\hat{r}_{1,n}^N - \hat{c}_n))$ .

Now, we expand the identity  $\hat{\mathbf{R}}^{-1} \hat{\mathbf{R}} = \mathbf{I}_M$  along dimension  $N$  using (244)–(245). Continuing from (234), we may expand the first term along dimension  $N$  to obtain

$$\begin{aligned} \frac{1}{iM} \text{tr}[\hat{\mathbf{R}}^{-1} \mathbf{H} \mathbf{S} \mathbf{A} \mathbf{B}^\dagger \mathbf{W} \mathcal{R}^\dagger] &= \frac{1}{iM} \sum_{n=1}^{\beta^* N} d_n \tilde{\mathbf{s}}_n^\dagger \mathbf{A} \mathbf{B}^\dagger \mathbf{W} \mathcal{R}^\dagger \hat{\mathbf{R}}^{-1} \mathbf{v}_n \\ &\asymp \frac{1}{M} \sum_{n=1}^{\beta^* N} d_n V_n (d_n \hat{c}_n \mathbf{v}_n + \boldsymbol{\tau}_n)^\dagger \hat{\mathbf{R}}_{t_n}^{-1} (\mathbf{v}_n - d_n \hat{\nu}_{1,n}^N \boldsymbol{\tau}_n) \\ &\asymp \frac{1}{M} \sum_{n=1}^{\beta^* N} d_n^2 V_n \hat{\nu}_{1,n}^N (\hat{c}_n - \hat{r}_{1,n}^N) \\ &\asymp \begin{cases} -\frac{\beta^*}{\beta} (\hat{\mathcal{H}}_{0,1}^N - 1) \\ -\hat{\nu}_1 (\hat{r}_1^N - \alpha \bar{p} \bar{w}) \frac{\beta^*}{\beta} \hat{\mathcal{H}}_{1,1}^N \end{cases} \end{aligned} \quad (247)$$

where

$$\hat{\mathcal{H}}_{p,1}^N = \frac{1}{\beta^* N} \sum_{n=1}^{\beta^* N} \frac{d_n^{2p}}{1 - d_n^2 \hat{\nu}_1^N (\hat{\tau}_1^N - \alpha \bar{p} \bar{w})}$$

and we have also used (201), (202), (204), and (246).

Combining (234), (247), and (236) gives

$$\beta(1 - \hat{\gamma}_1^N / \hat{\nu}_1^N) = \beta^*(1 - \hat{\mathcal{H}}_{0,1}^N) = -\hat{\nu}_1^N (\hat{\tau}_1^N - \alpha \bar{p} \bar{w}) \beta^* \hat{\mathcal{H}}_{1,1}^N. \quad (248)$$

4) *Identities via moment definitions:* In order to find expressions for  $\hat{\omega}_1^N$ ,  $\hat{\rho}_1^N$ ,  $\hat{\psi}_1^N$ ,  $\hat{\tau}_1^N$ , and  $\hat{\nu}_1^N$ , we now apply (221), (224)–(225), (231)–(232), and (244)–(245) to expansions of the definitions of these variables, or expressions related to them. After applying (221) or (224)–(225), we shall use (196), (197), (198), (223), and (227) in conjunction with Lemmas 3 and 4 to further simplify the resulting expression. Similarly, after applying (231)–(232), we shall use (199), (200), (217), (218), (203), (230), and (233) in conjunction with Lemmas 3 and 4 to further simplify the resulting expression. Finally, after applying (244)–(245), we shall use (201), (202), (219), (220), (204), (239), and (246) in conjunction with Lemmas 3 and 4 to further simplify the resulting expression.

Firstly, consider  $\hat{\omega}_1^N$  in (198). From (231)–(232) we obtain

$$\begin{aligned} \frac{1}{i} \text{tr}[(\mathbf{HSA})^\dagger \hat{\mathbf{R}}^{-1}] &\asymp \frac{1}{i} \sum_{k=1}^K P_k K_k \mathbf{h}_k^\dagger \hat{\mathbf{R}}_{d_k}^{-1} (\mathbf{h}_k - A_k \hat{\rho}_1^N \mathbf{q}_k) \\ &\asymp \frac{\alpha}{\eta} \hat{\rho}_1^N \hat{\mathcal{E}}_{1,1}^N \end{aligned} \quad (249)$$

which corresponds to  $\hat{\omega}_1^N$  for i.i.d.  $\mathbf{B}$ . For orthogonal  $\mathbf{B}$  with  $\alpha < \eta$ , we use (249) and  $\frac{1}{i} \mathbf{B}^\dagger \mathbf{B} = \mathbf{I}_K$  to obtain

$$\begin{aligned} \frac{\alpha}{\eta} \hat{\rho}_1^N \hat{\mathcal{E}}_{1,1}^N &\asymp \frac{1}{i^2} \text{tr}[(\mathbf{HSAB}^\dagger)^\dagger \hat{\mathbf{R}}^{-1}] \asymp \frac{1}{i} \sum_{m=1}^i \boldsymbol{\omega}_m^\dagger \hat{\mathbf{R}}^{-1} \boldsymbol{\omega}_m \\ &\asymp \hat{\omega}_1^N (1 - \hat{\omega}_1^N \mathcal{W}_{1,1}^N) \end{aligned} \quad (250)$$

which is proven in an identical manner to (228), also using (224), and may be simplified using (228) to give

$$\hat{\omega}_1^N \asymp \frac{\frac{\alpha}{\eta} \hat{\rho}_1^N \hat{\mathcal{E}}_{1,1}^N}{1 - \frac{\beta}{\eta} (1 - \hat{\gamma}_1^N / \hat{\nu}_1^N)}. \quad (251)$$

Now if  $\alpha > \eta$ , we have directly from (249) and (250)

$$\begin{aligned} \hat{\omega}_1^N &= \frac{\alpha}{\eta} \frac{1}{K - i} \text{tr}[\mathbf{Y} \mathbf{A}^\dagger \mathbf{S}^\dagger \mathbf{H}^\dagger \hat{\mathbf{R}}^{-1} \mathbf{HSA}] \\ &\asymp \frac{\alpha}{\alpha - \eta} \left( \frac{\alpha}{\eta} \hat{\rho}_1^N \hat{\mathcal{E}}_{1,1}^N - \frac{\eta}{\alpha} \hat{\omega}_1^N (1 - \hat{\omega}_1^N \mathcal{W}_{1,1}^N) \right). \end{aligned}$$

Combining this with (228) gives

$$\hat{\omega}_1^N = \frac{\frac{\alpha}{\eta} \hat{\rho}_1^N \hat{\mathcal{E}}_{1,1}^N}{1 - \frac{\beta}{\alpha} (1 - \hat{\gamma}_1^N / \hat{\nu}_1^N)}. \quad (252)$$

Now we consider  $\hat{\tau}_1^N$ , defined in (201), using (231)–(232). Firstly, note that  $\frac{1}{i} \mathcal{R} \mathbf{W} \mathbf{b}_k = \frac{1}{i} (\mathcal{R}_{d_k} + A_k \mathbf{h}_k \mathbf{b}_k^\dagger) \mathbf{W} \mathbf{b}_k = \mathbf{q}_k +$

$A_k \bar{w}_k \mathbf{h}_k$ , and hence

$$\begin{aligned} \frac{1}{N i^2} \text{tr}[(\mathcal{R} \mathbf{W} \mathbf{B} \mathbf{A})^\dagger \hat{\mathbf{R}}^{-1}] &= \frac{1}{N i^2} \sum_{k=1}^K P_k \text{tr}[(\mathcal{R} \mathbf{W} \mathbf{b}_k)^\dagger \hat{\mathbf{R}}^{-1}] \\ &= \frac{1}{N} \sum_{k=1}^K P_k \text{tr}[(\mathbf{q}_k + A_k \bar{w}_k \mathbf{h}_k)^\dagger \hat{\mathbf{R}}^{-1}] \\ &\asymp \alpha \frac{1}{K} \sum_{k=1}^K P_k K_k (\hat{\psi}_1^N - \bar{w} P_k \hat{\rho}_1^N (\hat{\psi}_1^N - \bar{w})) \\ &\asymp \alpha (\bar{w} \bar{p} + (\hat{\psi}_1^N - \bar{w}) \hat{\mathcal{E}}_{1,1}^N) \end{aligned} \quad (253)$$

which corresponds to  $\hat{\tau}_1^N$  for i.i.d.  $\mathbf{S}$ . To find an expression for  $\hat{\tau}_1^N$  with isometric  $\mathbf{S}$ , we use  $\mathbf{S}^\dagger \mathbf{S} = \mathbf{I}_K$  and (253) in reverse, i.e.,

$$\begin{aligned} \alpha (\bar{w} \bar{p} + (\hat{\psi}_1^N - \bar{w}) \hat{\mathcal{E}}_{1,1}^N) &\asymp \frac{1}{N i^2} \text{tr}[(\mathcal{R} \mathbf{W} \mathbf{B} \mathbf{A} \mathbf{S}^\dagger)^\dagger \hat{\mathbf{R}}^{-1}] \\ &\asymp \frac{1}{N} \sum_{n=1}^N V_n (\boldsymbol{\tau}_n + d_n \hat{c}_n \mathbf{v}_n)^\dagger \hat{\mathbf{R}}_{t_n}^{-1} (\boldsymbol{\tau}_n + d_n (\hat{c}_n - \hat{\tau}_{1,n}^N) \mathbf{v}_n) \\ &\asymp \frac{1}{N} \sum_{n=1}^N (\hat{c}_n + (\hat{\tau}_{1,n}^N - \hat{c}_n) V_n) \\ &\asymp \alpha \bar{w} \bar{p} + (\hat{\tau}_1^N - \alpha \bar{w} \bar{p}) (\beta^* (\hat{\mathcal{H}}_{0,1}^N - 1) + 1) \end{aligned} \quad (254)$$

where we have used  $\frac{1}{i} \mathcal{R} \mathbf{W} \mathbf{B} \mathbf{A} \tilde{\mathbf{s}}_n = \frac{1}{i} (\mathcal{R}_{t_n} + d_n \mathbf{v}_n \tilde{\mathbf{s}}_n) \mathbf{W} \mathbf{B} \mathbf{A} \tilde{\mathbf{s}}_n = \boldsymbol{\tau}_n + d_n \hat{c}_n \mathbf{v}_n$ . Moreover, using (248) to simplify (254) we obtain

$$\hat{\tau}_1^N \asymp \alpha \bar{w} \bar{p} + \frac{\alpha (\hat{\psi}_1^N - \bar{w}) \hat{\mathcal{E}}_{1,1}^N}{1 - \beta (1 - \hat{\gamma}_1^N / \hat{\nu}_1^N)}. \quad (255)$$

Now we consider  $\hat{\rho}_1^N$  defined in (199), using (244)–(245). Firstly,

$$\begin{aligned} \frac{1}{N} \text{tr}[\mathbf{H}^\dagger \hat{\mathbf{R}}^{-1}] &= \frac{1}{N} \sum_{n=1}^{\beta^* N} d_n^2 \mathbf{v}_n^\dagger \hat{\mathbf{R}}^{-1} \mathbf{v}_n \\ &\asymp \frac{1}{N} \sum_{n=1}^{\beta^* N} d_n^2 V_n \mathbf{v}_n^\dagger \hat{\mathbf{R}}_{t_n}^{-1} \mathbf{v}_n \asymp \hat{\nu}_1^N \beta^* \hat{\mathcal{H}}_{1,1}^N \end{aligned} \quad (256)$$

which corresponds to  $\hat{\rho}_1^N$  with i.i.d.  $\mathbf{S}$ . Note that  $\frac{1}{N} \text{tr}[(\mathbf{HS})^\dagger \hat{\mathbf{R}}^{-1}] \asymp \alpha \hat{\rho}_1^N \hat{\mathcal{E}}_{0,1}^N$  follows from an analogous derivation to (249). Therefore, combining this with (256), we obtain for isometric  $\mathbf{S}$ ,

$$\begin{aligned} \hat{\rho}_1^N &= \frac{1}{1 - \alpha} \left( \frac{1}{N} \text{tr}[\mathbf{H}^\dagger \hat{\mathbf{R}}^{-1}] - \frac{1}{N} \text{tr}[(\mathbf{HS})^\dagger \hat{\mathbf{R}}^{-1}] \right) \\ &\asymp \frac{1}{1 - \alpha} (\beta^* \hat{\nu}_1^N \hat{\mathcal{H}}_{1,1}^N - \alpha \hat{\rho}_1^N \hat{\mathcal{E}}_{0,1}^N). \end{aligned}$$

Moreover, combining this with (237) gives

$$\hat{\rho}_1^N = \frac{\beta^* \hat{\nu}_1^N \hat{\mathcal{H}}_{1,1}^N}{1 - \beta (1 - \hat{\gamma}_1^N / \hat{\nu}_1^N)}. \quad (257)$$

Now we consider  $\hat{\psi}_1^N$ , as defined in (200), using (221). Firstly,

$$\frac{1}{i^2} \text{tr}[(\mathcal{R} \mathbf{W})^\dagger \hat{\mathbf{R}}^{-1}] = \frac{1}{i} \sum_{m=1}^i \frac{w_m^2 \hat{r}_{1,m}^N}{1 + w_m \hat{r}_{1,m}^N} \asymp \bar{w} - \mathcal{W}_{1,1}^N \quad (258)$$

which corresponds to  $\hat{\psi}_1^N$  for i.i.d.  $\mathbf{B}$ . Note that  $\frac{1}{N i^2} \text{tr}[(\mathcal{R}\mathbf{W}\mathbf{B})^\dagger \hat{\mathbf{R}}^{-1}] \asymp \alpha(\bar{w} + (\hat{\psi}_1^N - \bar{w})\hat{\mathcal{E}}_{0,1}^N)$  from an analogous derivation to (253). Therefore, combining this with (258), we obtain for orthogonal  $\mathbf{B}$  and  $\alpha < \eta$ ,

$$\begin{aligned} \hat{\psi}_1^N &= \frac{\eta}{\eta - \alpha} \left( \frac{1}{i^2} \text{tr}[(\mathcal{R}\mathbf{W}\mathbf{B})^\dagger \hat{\mathbf{R}}^{-1}] - \frac{1}{i^3} \text{tr}[(\mathcal{R}\mathbf{W}\mathbf{B})^\dagger \hat{\mathbf{R}}^{-1}] \right) \\ &\asymp \frac{\eta}{\eta - \alpha} (\bar{w} - \mathcal{W}_{1,1}^N - \frac{\alpha}{\eta} (\bar{w} + (\hat{\psi}_1^N - \bar{w})\hat{\mathcal{E}}_{0,1}^N)). \end{aligned}$$

Moreover, combining with (237) gives

$$\hat{\psi}_1^N = \bar{w} - \frac{\mathcal{W}_{1,1}^N}{1 - \frac{\beta}{\eta}(1 - \hat{\gamma}_1^N/\hat{\nu}_1^N)}. \quad (259)$$

Finally, for orthogonal  $\mathbf{B}$  and  $\alpha > \eta$ , from (258) and  $\frac{1}{K}\mathbf{B}^\dagger\mathbf{B} = \mathbf{I}_i$ , we have

$$\begin{aligned} \bar{w} - \mathcal{W}_{1,1}^N &\asymp \frac{1}{K i^2} \text{tr}[(\mathcal{R}\mathbf{W}\mathbf{B})^\dagger \hat{\mathbf{R}}^{-1}] \\ &\asymp \bar{w} + (\hat{\psi}_1^N - \bar{w})\hat{\mathcal{E}}_{0,1}^N \end{aligned}$$

following an analogous derivation to (253). Moreover, combining with (237) gives

$$\hat{\psi}_1^N \asymp \bar{w} - \frac{\mathcal{W}_{1,1}^N}{1 - \frac{\beta}{\alpha}(1 - \hat{\gamma}_1^N/\hat{\nu}_1^N)}. \quad (260)$$

To derive an expression for  $\hat{\nu}_1$ , we start with (241) and obtain

$$1 + z\hat{\nu}_{1,n}^N = \frac{1}{i} \sum_{m=1}^i w_m \mathbf{n}_m^\dagger \mathbf{v}_n \mathbf{v}_n^\dagger \hat{\mathbf{R}}_{t_n}^{-1} \mathbf{r}_m \quad (261)$$

for which, like (243), we have

$$w_m \mathbf{n}_m^\dagger \mathbf{v}_n \mathbf{v}_n^\dagger \hat{\mathbf{R}}_{t_n}^{-1} \mathbf{r}_m \asymp \frac{w_m \sigma^2 \hat{\nu}_{1,n}^N}{1 + w_m \hat{r}_1^N} \quad (262)$$

where we have additionally used  $\hat{r}_{1,m,t_n}^N \asymp \hat{r}_{1,m}^N \asymp \hat{r}_1^N$ , which follows from Lemma 7, continuing on from the application of Lemma 6 in Appendix VII-B.3. Therefore, from (261) and (262), we obtain

$$1 + z\hat{\nu}_1^N \asymp \sigma^2 \hat{\nu}_1^N \mathcal{W}_{1,1}^N. \quad (263)$$

It follows (using the same arguments outlined after (171) in the proof of Theorem 1, this time assuming that the resulting equations have a unique solution) from (106)–(108), (197), (235), (249), (251), (252), (253), (255), (256), (257), (258), (259), (260), and (263) that, with probability one, for all  $z \in \mathbb{C}^+$ ,  $|G_{\hat{\mathbf{R}}}^N(z) - \hat{\gamma}_1| \rightarrow 0$ ,  $|\hat{\rho}_1^N - \hat{\rho}_1| \rightarrow 0$ ,  $|\hat{\tau}_1^N - \hat{\tau}_1| \rightarrow 0$ ,  $|\hat{\psi}_1^N - \hat{\psi}_1| \rightarrow 0$ ,  $|\hat{\omega}_1^N - \hat{\omega}_1| \rightarrow 0$ ,  $|\hat{\nu}_1^N - \hat{\nu}_1| \rightarrow 0$ , and  $|\hat{r}_1^N - \hat{r}_1| \rightarrow 0$ , where  $\hat{\gamma}_1$ ,  $\hat{\rho}_1$ ,  $\hat{\tau}_1$ ,  $\hat{\psi}_1$ ,  $\hat{\omega}_1$ ,  $\hat{\nu}_1$ , and  $\hat{r}_1 \in \mathbb{C}^+$  are solutions to (37)–(46).

## APPENDIX VIII

### PROOF OF THEOREM 2: ASYMPTOTIC SINR

This proof continues on from the proof of Theorem 3 in Appendix VII. Firstly, note that the steering vector (10) can be written as

$$\hat{\mathbf{s}}_k = \begin{cases} \mathbf{q}_k + A_k \bar{w}_k \mathbf{h}_k & , \text{ ALS with training,} \\ \mathbf{h}_k & , \text{ semi-blind ALS.} \end{cases} \quad (264)$$

That is, (231) is the equivalent asymptotic form for the semi-blind LS filter given in (7). Also, using (264), (231)–(232) we obtain the equivalent asymptotic form for the ALS filter with training. We can express the asymptotic form for both receivers as

$$\mathbf{c}_k \asymp K_k \hat{\mathbf{R}}_{d_k}^{-1} (a_{k,1}^N \mathbf{h}_k + a_{k,2}^N \mathbf{q}_k) \quad (265)$$

where

$$(a_{k,1}^N, a_{k,2}^N) = \begin{cases} (A_k(\bar{w}_k - \hat{\psi}_{1,k}^N), 1) & , \text{ ALS with training,} \\ (1, -A_k \hat{\rho}_{1,k}^N) & , \text{ semi-blind ALS.} \end{cases}$$

We now compute the large-system SINR for the filter (265) for stream  $k$ , and symbol interval  $m > i$ . For notational simplicity, we drop the subscript  $m$  in this appendix. Note therefore that in this appendix  $\mathbf{b}_{d_k}$  denotes  $\mathbf{b}_m$  with the  $k^{\text{th}}$  element removed. Since  $K_k$  cancels in the ratio when calculating the SINR, we ignore this constant. The signal component is

$$\begin{aligned} A_k \mathbf{c}_k^\dagger \mathbf{h}_k \mathbf{b}(k) &\asymp A_k (a_{k,1}^N \mathbf{h}_k + a_{k,2}^N \mathbf{q}_k)^\dagger \hat{\mathbf{R}}_{d_k}^{-\dagger} \mathbf{h}_k \mathbf{b}(k) \\ &\asymp A_k a_{k,1}^N \mathbf{b}(k) \mathbf{h}_k^\dagger \hat{\mathbf{R}}_{d_k}^{-\dagger} \mathbf{h}_k \end{aligned}$$

using (137), (217), and (230). The interference component is

$$\begin{aligned} &\mathbf{c}_k^\dagger (\mathbf{H}\mathbf{S}_{d_k} \mathbf{A}_{d_k} \mathbf{b}_{d_k} + \mathbf{n}) \\ &\asymp (a_{k,1}^N \mathbf{h}_k + a_{k,2}^N \mathbf{q}_k)^\dagger \hat{\mathbf{R}}_{d_k}^{-\dagger} (\mathbf{H}\mathbf{S}_{d_k} \mathbf{A}_{d_k} \mathbf{b}_{d_k} + \mathbf{n}) \\ &\asymp a_{k,1}^N \left( \mathbf{h}_k^\dagger \hat{\mathbf{R}}_{d_k}^{-\dagger} \mathbf{H}\mathbf{S}_{d_k} \mathbf{A}_{d_k} \mathbf{b}_{d_k} + \mathbf{h}_k^\dagger \hat{\mathbf{R}}_{d_k}^{-\dagger} \mathbf{n} \right) \\ &\quad + a_{k,2}^N \left( \mathbf{q}_k^\dagger \hat{\mathbf{R}}_{d_k}^{-\dagger} \mathbf{H}\mathbf{S}_{d_k} \mathbf{A}_{d_k} \mathbf{b}_{d_k} + \mathbf{q}_k^\dagger \hat{\mathbf{R}}_{d_k}^{-\dagger} \mathbf{n} \right). \end{aligned}$$

Therefore, the signal power is asymptotically equivalent to  $P_k |a_{k,1}^N|^2 |\hat{\rho}_{1,k}^N|^2$ . Also, the interference power, averaged over the data symbols and noise is asymptotically equivalent to

$$|a_{k,1}^N|^2 (\hat{\rho}_{4,k}^N + \sigma^2 \hat{\rho}_{2,k}^N) + |a_{k,2}^N|^2 (\hat{\psi}_{4,k}^N + \sigma^2 \hat{\psi}_{2,k}^N).$$

Moreover, from (191), (192), (199), (200), we have

$$\text{SINR}_{k,N}^{\text{ALS}} \asymp \frac{P_k |a_{k,1}^N|^2 |\hat{\rho}_1|^2}{|a_{k,1}^N|^2 (\hat{\rho}_4^N + \sigma^2 \hat{\rho}_2^N) + |a_{k,2}^N|^2 (\hat{\psi}_4^N + \sigma^2 \hat{\psi}_2^N)}. \quad (266)$$

## APPENDIX IX

### STEADY-STATE ALS SINR WITH WINDOWING

By inspection of Theorem 3 and Lemma 1, we note the following limits as  $\eta \rightarrow \infty$ :

$$\begin{aligned} \tilde{\gamma}_j &= \lim_{\eta \rightarrow \infty} \begin{cases} \hat{\gamma}_1/\eta & , j = 1, \\ \hat{\gamma}_j/\eta^2 & , j = 2, 3, 4. \end{cases} \\ \tilde{\omega}_j &= \lim_{\eta \rightarrow \infty} \begin{cases} \hat{\omega}_1 & , j = 1, \\ \hat{\omega}_j/\eta & , j = 2, 3, 4. \end{cases} \\ \tilde{r}_j &= \lim_{\eta \rightarrow \infty} \begin{cases} \hat{r}_1 & , j = 1, \\ \hat{r}_j/\eta & , j = 2, 3, 4. \end{cases} \end{aligned}$$

and

$$\tilde{\rho}_j = \lim_{\eta \rightarrow \infty} \begin{cases} \hat{\rho}_1/\eta & , j = 1, \\ \hat{\rho}_j/\eta^2 & , j = 2, 3, 4. \end{cases}$$

$$\tilde{\psi}_j = \lim_{\eta \rightarrow \infty} \begin{cases} \eta \hat{\psi}_1 & , j = 1, \\ \hat{\psi}_j & , j = 2, 3, 4. \end{cases}$$

and

$$\tilde{\nu}_j = \lim_{\eta \rightarrow \infty} \begin{cases} \hat{\nu}_1/\eta & , j = 1, \\ \hat{\nu}_j/\eta^2 & , j = 2, 3. \end{cases}$$

$$\tilde{\tau}_j = \lim_{\eta \rightarrow \infty} \begin{cases} \eta \hat{\tau}_1 & , j = 1, \\ \hat{\tau}_j & , j = 2, 3. \end{cases}$$

and

$$\tilde{w} = \lim_{\eta \rightarrow \infty} \eta \bar{w},$$

$$\tilde{\mathcal{W}}_{p,q} = \lim_{\eta \rightarrow \infty} \eta \mathcal{W}_{p,q},$$

$$\tilde{\mathcal{E}}_{p,q} = \lim_{\eta \rightarrow \infty} \hat{\mathcal{E}}_{p,q},$$

$$\tilde{\mathcal{H}}_{p,q} = \lim_{\eta \rightarrow \infty} \hat{\mathcal{H}}_{p,q}.$$

Moreover, the variables  $\tilde{\gamma}_j, \tilde{\rho}_j, \tilde{\tau}_j, \tilde{\psi}_j, \tilde{\omega}_j, \tilde{\nu}_j, \tilde{r}_j, \tilde{\mathcal{E}}_{p,q}, \tilde{\mathcal{H}}_{p,q}$  and  $\tilde{\mathcal{W}}_{p,q}$  satisfy the same set of equations as  $\hat{\gamma}_j, \hat{\rho}_j, \hat{\tau}_j, \hat{\psi}_j, \hat{\omega}_j, \hat{\nu}_j, \hat{r}_j, \hat{\mathcal{E}}_{p,q}, \hat{\mathcal{H}}_{p,q}$ , and  $\mathcal{W}_{p,q}$ , respectively (for appropriate values of  $j, m$ , and  $n$ ), if we set  $z = -\mu, \eta = 1$ , and assume that  $\mathbf{B}$  is i.i.d..

#### APPENDIX X

##### SKETCH OF PROOF OF THEOREM 4

We provide a sketch of the proof for i.i.d.  $\mathbf{S}$  only, mainly due to the fact that the equations are simpler to manipulate. However, the same approach is valid for isometric  $\mathbf{S}$ .

Firstly, according to the remark made after Theorem 3, with i.i.d.  $\mathbf{B}$  and  $\mu = 0$  we have  $\mathcal{W}_{1,1}\hat{\rho}_1 = \rho_1$ . That is, the term  $\hat{\rho}_1$  in the numerator of the asymptotic SINR in (31) can be written as  $\frac{1}{P_k \mathcal{W}_{1,1}} \text{SINR}_k^{\text{MMSE}}$ .

Consider the denominator of the alternate MMSE SINR expression in Section IV-B with i.i.d.  $\mathbf{S}$ . Solving (25)–(26) for  $\rho_4 + \sigma^2 \rho_2$  gives

$$\rho_4 + \sigma^2 \rho_2 = \frac{\beta^* (\alpha \mathcal{E}_{1,2} \mathcal{H}_{2,2} - z \mathcal{H}_{1,2})}{z^2 - \alpha \beta^* \mathcal{E}_{2,2} \mathcal{H}_{2,2}},$$

which from Appendix V equals  $\rho_1$ .

The next step is to simplify the  $\hat{\rho}_4 + \sigma^2 \hat{\rho}_2$  term in the denominator of the ALS SINR (31). To do this, we solve (57), (58), (61), and (56) to find  $\hat{\rho}_3, \hat{\tau}_3, \hat{\psi}_4, \hat{\nu}_3$  in terms of  $\hat{\rho}_4$  and  $\hat{\rho}_2$ , and then substitute into (57). We then solve (55), (58), and (61) for  $\hat{\gamma}_2, \hat{\tau}_2$ , and  $\hat{\psi}_2$  in terms of  $\hat{\rho}_4$  and  $\hat{\rho}_2$ , and substitute into (57). Combining these results gives

$$\hat{\rho}_4 + \sigma^2 \hat{\rho}_2 = \frac{\frac{\hat{\rho}_1}{\mathcal{W}_{1,2}} (1 - \alpha \frac{\mathcal{W}_{2,2}}{\mathcal{W}_{1,1}} \hat{\rho}_1 \hat{\mathcal{E}}_{1,1})}{1 - \alpha \frac{\mathcal{W}_{2,2}}{\mathcal{W}_{1,1}} \hat{\rho}_1 (\hat{\mathcal{E}}_{1,2} + \hat{\rho}_1 \mathcal{W}_{1,1} \hat{\mathcal{E}}_{2,2})}.$$

Now we use the identity (71) to obtain  $\hat{\mathcal{E}}_{1,1} = \hat{\mathcal{E}}_{1,2} + \hat{\rho}_1 \mathcal{W}_{1,1} \hat{\mathcal{E}}_{2,2}$ , and so we have that

$$\hat{\rho}_4 + \sigma^2 \hat{\rho}_2 = \frac{\hat{\rho}_1}{\mathcal{W}_{1,2}} = \frac{1}{P_k \mathcal{W}_{1,1} \mathcal{W}_{1,2}} \text{SINR}_k^{\text{MMSE}}. \quad (267)$$

Next we simplify the  $\hat{\psi}_4 + \sigma^2 \hat{\psi}_2$  term in the denominator of the ALS SINR (31). To do this, we solve (55), (62), (61), (61) and (62) for  $\hat{\gamma}_2, \hat{\omega}_2, \hat{\psi}_2, \hat{\psi}_4$ , and  $\hat{\omega}_4$ , and form the sum

$$\hat{\psi}_4 + \sigma^2 \hat{\psi}_2 = \frac{\mathcal{W}_{2,2} ((\hat{\rho}_4 + \sigma^2 \hat{\rho}_2) \hat{\mathcal{E}}_{1,2} + \hat{\rho}_1^2 \hat{\mathcal{E}}_{2,2}) \mathcal{W}_{1,2} \alpha + \sigma^2 \beta \hat{\gamma}_1}{\mathcal{W}_{1,2} (\eta - \alpha \hat{\rho}_1^2 \hat{\mathcal{E}}_{2,2} \mathcal{W}_{2,2})}.$$

Substituting (267) and  $\hat{r}_1 = \frac{1}{\eta} (\alpha \hat{\rho}_1 \hat{\mathcal{E}}_{1,1} + \beta \sigma^2 \hat{\gamma}_1)$ , and simplifying with (71) gives

$$\hat{\psi}_4 + \sigma^2 \hat{\psi}_2 = \frac{\mathcal{W}_{1,1}}{\mathcal{W}_{1,2}} - 1. \quad (268)$$

Substituting (267) and (268) into (31) along with the expressions for  $\hat{a}_{k,1}$  and  $\hat{a}_{k,2}$  from (32), and simplifying gives (82) and (83).

#### ACKNOWLEDGMENT

The authors thank Philippe Loubaton for his generous assistance, and in particular for pointing out certain shortcomings in some of our original derivations, along with helpful guidance in resolving those issues.

#### REFERENCES

- [1] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.
- [2] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995.
- [3] V. L. Girko, *Theory of Random Determinants*. Kluwer Academic, 1990.
- [4] F. Hiai and D. Petz, *The semicircle law, free random variables and entropy*. American Mathematics Society, Mathematical Surveys and Monographs, Vol. 77, 2000.
- [5] L. Li, A. M. Tulino, and S. Verdú, "Design of reduced-rank MMSE multiuser detectors using random matrix methods," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 986–1008, June 2004.
- [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, no. 5, pp. 1293–1311, May 2003.
- [7] J. M. Chaufray, W. Hachem, and P. Loubaton, "Asymptotic analysis of optimum and suboptimum CDMA downlink MMSE receivers," *IEEE Trans. Inform. Theory*, vol. 50, no. 11, pp. 2620–2638, Nov. 2004.
- [8] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Prentice Hall, 1996.
- [9] H. V. Poor and X. Wang, "Code-aided interference suppression for DS/CDMA communications Part II: Parallel blind adaptive implementations," *IEEE Trans. Commun.*, vol. 45, no. 9, pp. 1112–1122, Sept. 1997.
- [10] F. Ling and J. Proakis, "Nonstationary learning characteristics of least squares adaptive estimation algorithms," in *IEEE ICASSP*, vol. 9, Mar. 1984, pp. 118–121.
- [11] E. Eleftheriou and D. Falconer, "Tracking properties and steady-state performance of RLS adaptive filter algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 5, pp. 1097–1110, Oct. 1986.
- [12] M. L. Honig, U. Madhow, and S. Verdú, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, pp. 944–960, July 1995.
- [13] G. Caire, "Two-stage nondata-aided adaptive linear receivers for DS/CDMA," *IEEE Trans. Commun.*, vol. 48, no. 10, pp. 1712–1724, Oct. 2000.
- [14] M. L. Honig and H. V. Poor, "Adaptive Interference Mitigation" in *Wireless Communications: A Signal Processing Perspective*. Englewood Cliffs, NJ: Prentice-Hall, 1998, ch. 2, pp. 64–128, H. V. Poor and G. Wornell, eds.
- [15] W. Xiao and M. L. Honig, "Convergence analysis of adaptive reduced-rank linear filters for DS-CDMA," in *Conference on Information Sciences and Systems*, Princeton University, Mar. 2000, pp. WP2–6 – WP2–11.
- [16] —, "Large system transient analysis of adaptive least squares filtering," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2447–2474, July 2005.

- [17] J. Zhang and X. Wang, "Large-system performance analysis of blind and group-blind multiuser receivers," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2507–2523, Sept. 2002.
- [18] Z. Xu and X. Wang, "Large-sample performance of blind and group-blind multiuser detectors: a perturbation perspective," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2389–2401, Oct. 2004.
- [19] A. Host-Madsen, X. Wang, and S. Bahng, "Asymptotic analysis of blind multiuser detection with blind channel estimation," *IEEE Trans. Signal Processing*, vol. 52, no. 6, pp. 1722–1738, June 2004.
- [20] R. R. Müller, "A random matrix model of communication via antenna arrays," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2495–2506, Sept. 2002.
- [21] S. Verdú, *Multiuser Detection*. Cambridge University Press, 1998.
- [22] M. J. M. Peacock, I. B. Collings, and M. L. Honig, "Asymptotic analysis of MMSE multiuser receivers for multi-signature multicarrier CDMA in Rayleigh fading," *IEEE Trans. Commun.*, vol. 52, no. 6, pp. 964–972, June 2004.
- [23] C. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.
- [24] V. Vasilchuk, "On the law of multiplication of random matrices," *Mathematical Physics, Analysis and Geometry*, vol. 4, no. 1, pp. 1–36, 2001.
- [25] R. B. Dozier and J. W. Silverstein, "On the empirical distribution of eigenvalues of large dimensional information-plus-noise type matrices," May 2004, preprint: <http://www4.ncsu.edu/~jack/infoise.pdf>.
- [26] M. J. M. Peacock, "Random matrix theory analysis of fixed and adaptive linear receivers," Ph.D. dissertation, The University of Sydney, Aug. 2005.
- [27] J. Proakis, *Digital Communications*, 4th ed. McGraw Hill, 2001.
- [28] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [29] H. Vikalo, B. Hassibi, B. M. Hochwald, and T. Kailath, "On the capacity of frequency-selective channels in training-based transmission schemes," submitted to *IEEE Trans. Inform. Theory*, 2004.
- [30] Y. Sun, "Transmitter and receiver techniques for wireless fading channels," Ph.D. dissertation, Northwestern University, June 2003.
- [31] V. L. Girko, *Theory of stochastic canonical equations*. Kluwer Academic, 2001.
- [32] J. Evans and D. N. C. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *IEEE Trans. Inform. Theory*, vol. 46, no. 6, pp. 2059–2078, Sept. 2000.
- [33] K. L. Chung, *A Course in Probability Theory*, 3rd ed. Academic Press, 2001.
- [34] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, "On limit of the largest eigenvalue of the large dimensional sample covariance matrix," *Probability Theory Related Fields*, vol. 78, pp. 509–521, 1988.
- [35] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," submitted to *Annals of Applied Probability*, July 2005, preprint: <http://arxiv.org/abs/math/0507172>.
- [36] J. B. Conway, *Functions of One Complex Variable*. New York: Springer-Verlag, 1978.

**Matthew J. M. Peacock** (S'01-M'05) is from Willaura, Australia, and was born in 1978. He received the combined B.E. degree in Electrical Engineering and B.Sc. degree in Mathematics/Computer Science from the University of Melbourne in 2001. He has recently completed his Ph.D. degree in Electrical Engineering at The University of Sydney, working with the Telecommunications and Industrial Physics group at Australia's Commonwealth Scientific & Industrial Research Organisation (CSIRO) as a recipient of a Postgraduate scholarship.

His current research interests include turbo-equalization, large-system analysis, information theory, and random matrix theory.

**Iain B. Collings** (S'92-M'95-SM'02) was born in Melbourne, Australia, in 1970. He received the B.E. degree in Electrical and Electronic Engineering from the University of Melbourne in 1992, and the Ph.D. degree in Systems Engineering from the Australian National University in 1995.

Currently he is a Science Leader in the Wireless Technologies Laboratory at the CSIRO ICT Centre, Australia. Prior to this he was an Associate Professor at the University of Sydney (1999-2005); a Lecturer at the University of Melbourne (1996-1999); and a Research Fellow in the Australian Cooperative Research Centre for Sensor Signal and Information Processing (1995). His current research interests include mobile digital communications and broadband digital subscriber line communications. More specifically, synchronization, channel estimation, equalization, and multi-carrier modulation, for time-varying and frequency-selective channels.

Dr Collings currently serves as an Editor for the IEEE Transactions on Wireless Communications, and as a Guest Editor for the EURASIP Journal on Advanced Signal Processing. He has also served as the Vice Chair of the Technical Program Committee for IEEE Vehicular Technology Conf. (Spring) 2006, as well as serving on a number of other TPCs and organizing committees of IEEE conferences. He is also a founding organizer of the Australian Communication Theory Workshops 2000-06.

**Michael L. Honig** (S'80-M'81-SM'92-F'97) received a B.S. degree in Electrical Engineering from Stanford University in 1977, and M.S. and Ph.D. degrees in Electrical Engineering from the University of California, Berkeley, in 1978 and 1981, respectively. He subsequently joined Bell Laboratories in Holmdel, NJ, where he worked on local area networks and voiceband data transmission. In 1983 he joined the Systems Principles Research Division at Bellcore, where he worked on Digital Subscriber Lines and wireless communications. Since the Fall of 1994, he has been with Northwestern University where he is a Professor in the Electrical and Computer Engineering Department. He has held visiting positions at Princeton University, the University of Sydney, and the University of California, Berkeley.

He has served as an editor for the IEEE Transactions on Information Theory and the IEEE Transactions on Communications. He has also served as a member of the Digital Signal Processing Technical Committee for the IEEE Signal Processing Society, and as a member of the Board of Governors for the Information Theory Society. He is the co-recipient of the 2002 IEEE Communications Society and Information Theory Society Joint Paper Award.