Large System Performance of Linear Multiuser Receivers in Multipath Fading Channels

Jamie Evans, *Member, IEEE,* and David N. C. Tse, *Member, IEEE*

***Abstract*—A linear multiuser receiver for a particular user in a code-division multiple-access (CDMA) network gains potential benefits from knowledge of the channels of all users in the system. In fast multipath fading environments we cannot assume that the channel estimates are perfect and the inevitable channel estimation errors will limit this potential gain. In this paper, we study the impact of channel estimation errors on the performance of linear multiuser receivers, as well as the channel estimation problem itself. Of particular interest are the scalability properties of the channel and data estimation algorithms: what happens to the performance as the system bandwidth and the number of users (and hence channels to estimate) grows? Our main results involve asymptotic expressions for the signal-to-interference ratio of linear multiuser receivers in the limit of large processing gain, with the number of users divided by the processing gain held constant. We employ a random model for the spreading sequences and the limiting signal-to-interference ratio expressions are independent of the actual signature sequences, depending only on the system loading and the channel statistics: background noise power, energy profile of resolvable multipaths, and channel coherence time. The effect of channel uncertainty on the performance of multiuser receivers is succinctly captured by the notion of *effective interference*.**

***Index Terms*—Code-division multiple access (CDMA), effective interference, linear receivers, multipath fading channels, multiuser detection, random spread- ing.**

I. INTRODUCTION

WIDE-BAND code-division multiple access (CDMA) has been selected for the air interface of third-generation wireless systems [1]–[3]. A significant thrust of research in this area has focused on receiver design for signals contaminated not only by background noise but also by structured interference from other users of the multiaccess channel. This fundamental problem has led to an explosion of research activity over the past decade which can be grouped under the title of *multiuser detection* [4]–[6]. In particular, the design of linear multiuser detectors has received considerable attention including decorrelating receivers [7], [8] and the linear minimum-mean square error (LMMSE) receiver [9]–[12].

Most of these earlier works do not explicitly consider communication over a multipath fading channel. Initial investigations of multiuser receivers in fading channels assumed that the channel was perfectly known to the receiver (see [13] for single-path fading and [14]–[16] for multipath fading). In these papers, the focus was on extending performance measures such as asymptotic efficiency and near–far resistance, for application in time-varying conditions. Investigations which drop the assumption that the fading channel is perfectly known can be roughly grouped into three classes: 1) **decorrelator-based receivers**, that are noncoherent and do not require channel information, although the signature sequences of all users are assumed known [17], [15], [18]; 2) **coherent multiuser receivers**, that incorporate channel estimates in addition to knowledge of the signature sequences [19]–[23]; and 3) **fully adaptive receivers**, that do not explicitly estimate the channel nor require the knowledge of the signature sequences of the interferers [11], [18], [21], [24]–[28].

At present, it is very difficult to obtain any clear engineering insights on the performance comparison of these various approaches or to characterize performance limits of linear multiuser receivers in fading environments. Performance analysis leads to expressions for signal-to-interference ratio (SIR) or average bit-error rate (BER) in terms of the particular set of signature sequences employed. Channel estimation errors are often expressed in terms of the Kalman filter recursion for the error covariance [29], [22]; a solution which does not readily lend itself to an understanding of the scalability properties of the resultant multiuser receivers. Simulations are relied upon to convey some insight into the properties of the receivers and yet almost always, the simulations are based on small-scale systems; it is simply too computationally intensive to analyze the large-scale systems we are interested in here, and which are perhaps more relevant to future wireless systems.

One solution, upon which much of the multiuser detection literature has focused, is to employ simpler performance measures such as asymptotic efficiency and near–far resistance. While these measures have been very useful for understanding and comparing multiuser receiver structures, especially with regard to performance under worst case conditions, they are often too crude to provide insights. As an example, we note that the asymptotic efficiency and near–far resistance of the decorrelator and LMMSE receiver are equal [4]. While the difference in performance between these linear receivers is usually small under ideal channel conditions and low system

Manuscript received March 20, 1999; revised December 17, 1999. This work was supported by the Air Force Office of Scientific Research under Grant F49620-96-1-0199, and by an NSF CAREER Award under Grant NCR-9734090. The material in this paper was presented in part at the IEEE Information Theory Workshop, Kruger National Park, South Africa, June 1999, and at GLOBECOM (Communications Theory Symposium), Rio de Janeiro, Brazil, December, 1999. This work was performed when J. Evans was visiting the University of California at Berkeley.

J. Evans is with the School of Electrical and Information Engineering, University of Sydney, NSW 2006, Australia (e-mail: jevans@ee.usyd.edu.au).

D. N. C. Tse is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA (e-mail: dtse@eecs.berkeley.edu).

Communicated by M. L. Honig, Associate Editor for Communications.

Publisher Item Identifier S 0018-9448(00)06995-9.

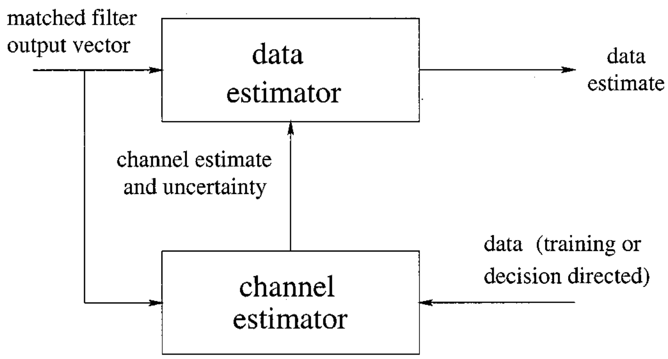


Fig. 1. Structure of multiuser receiver.

loading, such is no longer the case in a highly loaded system with multipaths and channel uncertainty. Thus for the problems addressed here, these measures are not *fine-grained* enough to draw interesting conclusions.

In [30] one alternative approach is presented. In that paper, signature sequences were modeled as random sequences leading to expressions for SIR which were random quantities. In the asymptotic limit of a large number of users and a large spreading bandwidth, it was shown that the random SIR expression *converges* in probability to a deterministic quantity, independent of the realization of the random sequences. More importantly, the resulting limit is shown to have a very nice form from which the concepts of effective interference and effective bandwidth emerge (see [30] for details). The same modeling paradigm of random spreading and large systems was extended to asynchronous systems in [31]. (Also see [32]–[35] for some parallel work based on random spreading sequences.)

In this paper we extend the philosophy and the techniques of [30] to the situation where the channel is time-varying. We drop the assumption that the channels of all users are perfectly known and assess receiver performance as a function of the uncertainty in the channel estimates, although the signature sequences are assumed to be known. For simplicity, we will focus on a symbol-synchronous channel but extension to the asynchronous situation along the lines of [31] is possible.

The (suboptimal) receiver structure shown in Fig. 1 is of central importance. The receiver and corresponding analysis can be decoupled into two parts, the data estimator and the channel estimator. The data estimator is a linear multiuser receiver which obtains estimates of the data of each user based on observation of the received signal over a single symbol interval, along with information supplied by the channel estimator. The data estimator is thus a *one-shot* linear estimator which incorporates information from other symbol intervals only through the coupling with the channel estimator. The design and analysis of the data estimator begins with the assumption that the channel is statistically characterized by the mean and covariance structure supplied by the channel estimator. The performance of the receiver is examined through the SIR attained and emphasis is placed on the LMMSE receiver which maximizes SIR over all linear receivers, however, results are also derived for the decorrelator and the conventional matched filter as a basis for comparison.

A. Summary of Results

The main results can be summarized through the notion of *effective interference*, first introduced in [30]. Suppose we have a synchronous CDMA system with spreading gain N, K users with received powers $|a_1|^2, \ldots, |a_K|^2$, and background noise power σ^2. In a large system (both N and K large), with random spreading, the SIR attained by the LMMSE receiver for user 1, is approximately

$$\text{SIR} = |a_1|^2 \beta_d$$

where β_d is the unique solution of the fixed point equation

$$\beta_d = \left[\sigma^2 + \frac{1}{N} \sum_{k=2}^{K} I(|a_k|^2, \beta_d)\right]^{-1}$$

where

$$I(p, \beta) = \frac{p}{1 + p\beta}.$$

Observe that the only term that involves user k is $I(|a_k|^2, \beta_d)$ and that these terms simply add across interfering users. We call this term the effective interference of user k on user 1 when the normalized SIR for user 1 is β_d.

Now consider a single (resolvable) path fading model where the channel gains of each user are not known perfectly to the data estimator but rather, a_k is specified by its estimate $\overline{a}_k$ and error variance ξ_k^2 (supplied by the channel estimator). In a large system the SIR for user 1 is approximately

$$\text{SIR} = \frac{|\overline{a}_1|^2 \beta_d}{1 + \xi_1^2 \beta_d}$$

where

$$\beta_d = \left[\sigma^2 + \frac{1}{N} \sum_{k=2}^{K} I(|\overline{a}_k|^2 + \xi_k^2, \beta_d)\right]^{-1}.$$

Interferer k looks like an interferer in the perfectly known channel case with power $|\overline{a}_k|^2 + \xi_k^2$.

Now consider a multiple-path fading model where each user appears as L resolvable paths or components at the receiver. If the gain for path l of user k is characterized by the estimate $\overline{a}_{kl}$ and error variance ξ_k^2 (we confirm later that this variance does not depend on l when the average power per path is equal), then in a large system, the SIR for the LMMSE receiver of user 1 is approximately

$$\text{SIR} = \frac{\sum_{l=1}^{L} |\overline{a}_{1l}|^2 \beta_d}{1 + \xi_1^2 \beta_d}$$

where

$$\beta_d = \left[\sigma^2 + \frac{1}{N} \sum_{k=2}^{K} \left((L-1) I(\xi_k^2, \beta_d) + I\left(\sum_{l=1}^{L} |\overline{a}_{kl}|^2 + \xi_k^2, \beta_d\right)\right)\right]^{-1}.$$

The overall effect of interferer k is given by the term

$$(L-1)I(\xi_k^2,\beta_d)+I\left(\sum_{l=1}^{L}|\overline{a}_{kl}|^2+\xi_k^2,\beta_d\right)$$

which is the same as the interference that would result in the single-path fading case from $L-1$ users with power ξ_k^2 and one user with power $\sum_{l=1}^{L}|\overline{a}_{kl}|^2+\xi_k^2$. As the uncertainty increases, an interferer moves from looking like a single high-power interferer, to looking like L separate interferers with power reduced by a factor of L.

In parallel with data estimation, we also analyze the performance of channel estimators, which jointly estimate the channel parameters of all users conditioned on knowledge of the data of all users for all symbols within the channel estimation window. This model applies directly to systems employing training sequences and provides a bound on the performance of *decision-directed* channel estimators (since we do not account for errors in the bits fed back to the channel estimator).

Combining the results in data and channel estimation yields the performance limits of linear multiuser receivers in fast-fading environments, as a function of key system parameters such as the length of the channel estimation window, number of resolvable multipaths, and system loading. Numerical examples based on the theoretical results provide interesting engineering insights and comparison of receivers in different parameter regimes.

All of the results in this paper are asymptotic in nature, yielding limiting SIR and channel estimation error expressions which are independent of the particular realization of the random signature sequences. This abstraction is valuable in obtaining insights of general applicability. It is important at this point to note that nowhere do we average with respect to the sequences, rather, the SIR expression for finite-size systems which is a function of the sequence realization, converges almost surely or in probability to a value which is independent of the sequences. Just as in [30], the results obtained here are based on the analysis of the spectrum of large random matrices.

II. Signal Model

Our starting point is the equation for the chip-matched filter output vector at time m

$$y(m)=\sum_{k=1}^{K}b_k(m)\sum_{l=1}^{L}a_{kl}(m)s_{kl}(m)+n(m) \qquad (1)$$

where $k\in\{1,\ldots,K\}$ indexes the multiple users, and $l\in\{1,\ldots,L\}$ indexes the paths of each user, $a_{kl}(m)$ is the channel gain for path l of user k over symbol period m, $b_k(m)$ is the data symbol of user k over period m, $s_{kl}(m)$ is the signature sequence for path l of user k over symbol period m, and $n(m)$ is an additive noise. We assume throughout that the delay spread of the channel is small compared to the symbol time so that intersymbol interference can be neglected. Note that the assumption that we know s_{kl} means that we implicitly assume knowledge of the timing of resolvable path l of user k.

The channel gain process for each path of each user is a circularly symmetric complex Gaussian random process and the processes for each path are independent with $\boldsymbol{E}[a_{kl}(m)]=0$ and $\boldsymbol{E}[a_{kl}(m)a_{kl}^*(m)]=\overline{p}_{kl}$. Our channel model can be considered conditioned on the much slower fading that effectively acts to determine the user powers. We assume that there is a time-scale separation in effect which makes it reasonable to assume that each $\overline{p}_{kl}$ is known perfectly and does not change over the time period of interest. For simplicity, we will assume that the average received power of all paths of all users is the same so that $\overline{p}_{kl}=\overline{p}/L$.

Each data symbol $b_k(m)$ is assumed to be of the form $e^{j\theta_i}$ where

$$\theta_i\in\left\{0,\frac{2\pi}{M},\ldots,\frac{2(M-1)\pi}{M}\right\}$$

with every data symbol independent of all others. We are thus assuming that the transmitter uses M-ary phase shift keying (PSK) modulation. We restrict ourselves to M-ary PSK modulation because the property that $|b_k(m)|=1$ greatly simplifies the performance analysis.

The signature sequence $s_{kl}(m)$ is assumed to be an N–dimensional column vector with independent and identically distributed (i.i.d.) elements each being a circularly symmetric complex Gaussian random variable with zero mean and variance $1/N$. The choice of this distribution allows a unified and compact treatment of many of the technical results, however almost all the results we present are actually insensitive to the distribution, requiring only that the elements are zero-mean and have variance $1/N$. The random sequences are independent across users, paths, and symbols. Thus this model is directly applicable to systems using *long* pseudorandom sequences. However, some of our results go beyond this basic model and will also be extended to systems using *repeated sequences*, i.e., each user repeats the same (random) signature sequence over different symbols. It should also be noted that the assumption of independence across paths is an unrealistic one, as the sequences along different paths are really shifted replicas of the same transmitted sequence. The assumption is made here solely to simplify the analysis of our main results; extensions of some of our results to the shifted case will be presented.

The additive noise is a circularly symmetric complex white Gaussian noise with $\boldsymbol{E}[n(m)]=0$ and $\boldsymbol{E}[n(m)n^H(m)]=\sigma^2I$.

III. Data Estimation

We consider the problem of forming an estimate of the mth data symbol of user 1 based only on the complex vector $y(m)$. We note immediately that basing an estimate of $b_1(m)$ on $y(m)$ alone is suboptimal since the channel fading process introduces memory into the system (see [20]). The above *one-shot* scheme has definite advantages in terms of computational complexity, however, and when the channel gains are known $y(m)$ is, in fact, a sufficient statistic for the estimation of $b_1(m)$. Without loss of generality (within the confines of the one-shot scheme) we can drop the time index from the terms in (1) leading to the observation equation

$$y = \sum_{k=1}^{K} b_k \sum_{l=1}^{L} a_{kl} s_{kl} + n.$$

We first note that if the channel gains are known, we can replace the above model by

$$y = \sum_{k=1}^{K} b_k\, s_k^e + n$$

where $s_k^e = \sum_{l=1}^{L} a_{kl} s_{kl}$ is the effective signature sequence for user k at the receiver. In this case, the problem is essentially reduced to a single-path fading problem. The key point though is that knowledge of the effective signature sequences requires knowledge of the channel.

In the single-path fading model, the direction in N-dimensional space of each user at the receiver (the effective signature sequence) does not depend on the channel. The channel gain affects only the energy in the direction of a user through scalar multiplication. This means that it is possible to design receivers which null out the interfering users without knowledge of the channel and this is precisely what the decorrelating multiuser detector does. There is clearly a complication in the multipath fading case because the signature sequence of a user at the receiver has both a direction and an energy which depend on the channel. If the channel is unknown then all we can say is that a particular user lies somewhere in an L-dimensional subspace. Suppose that estimates of each channel gain are available along with a measure of the confidence in these estimates. How should we design a multiuser receiver for such a system?

One approach suggested in the literature is the decorrelating, multipath-combining detector [15], [16], [18]. The received signal is initially processed by a decorrelator which treats all KL sequences s_{kl} as if they corresponded to interfering users. The L correlator outputs corresponding to each user are then combined using techniques well known for single-user multipath channels. The problem with this approach is that it is very wasteful of degrees of freedom. Whereas in the perfect knowledge case we know that each user takes up one degree of freedom, each user now occupies L directions. The decorrelating operation will thus become increasingly ill-conditioned as KL approaches N. This receiver is analyzed in Section III-B.

An alternative approach is the multipath-combining, decorrelating detector which forms estimates of the received signature sequence of each user $\overline{s}_k^e = \sum_{l=1}^{L} \overline{a}_{kl} s_{kl}$ and then performs a decorrelation operation. While such an approach makes much more efficient use of the available degrees of freedom, it is not clear how channel uncertainty would impact the performance. This receiver does not make use of the confidence measures supplied with the estimates and, in this sense, cannot be considered robust to channel estimation errors.

In the sequel we present and analyze an LMMSE receiver which combines the robustness properties of the decorrelating, multipath-combining detector (when channel uncertainty is high) with the superior performance of the multipath-combining, decorrelating detector (when the channel uncertainty is low). We do not assume that the a_{kl} are known perfectly at the receiver, but instead assume that the channel estimator provides the data estimator with channel estimates along with the mean-squared error (MSE) in these estimates. The data estimator is thus conditioned on the belief that a_{kl} is a random variable with mean $\overline{a}_{kl}$ and variance ξ_{kl}^2. Note that in the standard situation when the channel is assumed to be known perfectly, $\overline{a}_{kl} = a_{kl}$ and $\xi_{kl}^2 = 0$. All results we derive reduce to the perfectly known (slow-fading) case upon making these substitutions.

All expectations in this section should be seen as conditional on the information that is assumed known at the data estimator, that is on the signature sequences, and the mean and variance supplied by the channel estimator. We use the notation $\boldsymbol{E}_d$ to denote expectation conditioned on this information.

Let $s_k = [s_{k1}, \ldots, s_{kL}]$ and $a_k = [a_{k1}, \ldots, a_{kL}]^T$ denote the sequences and channel gains corresponding to user k. Let $S = [s_1, \ldots, s_K]$, $A = \mathrm{diag}\,(a_1, \ldots, a_K)$, and $b = [b_1, \ldots, b_K]^T$. The signal model over the symbol of interest is then expressed compactly as

$$y = SAb + n.$$

We now look in turn at the LMMSE receiver, a decorrelating receiver, and a single-user matched filter.

A. LMMSE Receiver

First we must calculate $\boldsymbol{E}_d[yy^H]$. Using the fact that $n, b_1, \ldots, b_K$ and A are independent

$$\boldsymbol{E}_d[yy^H] = \sum_{k=1}^{K} \sum_{l_1=1}^{L} \sum_{l_2=1}^{L} \boldsymbol{E}_d[a_{kl_1} a_{kl_2}^*] s_{kl_1} s_{kl_2}^H + \sigma^2 I$$

Letting

$$D = \boldsymbol{E}_d[AA^H] = \mathrm{diag}\,(\boldsymbol{E}_d[a_1 a_1^H], \ldots, \boldsymbol{E}_d[a_K a_K^H])$$

we can write

$$\boldsymbol{E}_d[yy^H] = SDS^H + \sigma^2 I$$

We also have

$$\boldsymbol{E}_d[b_1^* y] = \sum_{l=1}^{L} \boldsymbol{E}_d[a_{1l}] s_{1l} = s_1 \overline{a}_1$$

so that the LMMSE receiver for user 1 is

$$c = (SDS^H + \sigma^2 I)^{-1} s_1 \overline{a}_1.$$

Remark 1: An interesting special case arises when we assume nothing is known about the channels of the interfering users other than the *a priori* statistical characterization and that user 1's channel is known perfectly. In this case, the LMMSE receiver developed above is in fact the optimal multiuser receiver in the sense that the output is a sufficient statistic for b_1. This is the Bayesian analog of a result for the decorrelator which says that the decorrelator results from the joint maximum-likelihood estimation of the data and the channel gains [7].

If we write

$$\boldsymbol{E}_d[a_1 a_1^H] = \overline{a}_1 \overline{a}_1^H + \Xi_1$$

(so that $\Xi_1 = \boldsymbol{E}_d[(a_1 - \overline{a}_1)(a_1 - \overline{a}_1)^H]$) we can alternatively express c as

$$c = (S_1 D_1 S_1^H + \sigma^2 I + s_1(\overline{a}_1\overline{a}_1^H + \Xi_1)s_1^H)^{-1}(s_1\overline{a}_1)$$
$$= \text{constant} \times (S_1 D_1 S_1^H + \sigma^2 I + s_1 \Xi_1 s_1^H)^{-1}(s_1\overline{a}_1)$$

where $S_1 = [s_2, \ldots, s_K]$ and

$$D_1 = \text{diag}\,(\boldsymbol{E}_d[a_2 a_2^H], \ldots, \boldsymbol{E}_d[a_K a_K^H]).$$

The SIR for the estimate $z = c^H y$ of b_1 is then

$$\text{SIR}_1 = \overline{a}_1^H s_1^H (S_1 D_1 S_1^H + \sigma^2 I + s_1 \Xi_1 s_1^H)^{-1} s_1 \overline{a}_1 \qquad (2).$$

Let $Z_1 = s_1^H(S_1 D_1 S_1^H + \sigma^2 I)^{-1} s_1$ and apply the matrix inversion lemma to give

$$\text{SIR}_1 = \overline{a}_1^H (Z_1 - Z_1(\Xi_1^{-1} + Z_1)^{-1} Z_1)\overline{a}_1.$$

From this point on we will assume that $\Xi_1 = \xi_1^2 I$ which means that the error variances of all paths are equal and that the path estimates are uncorrelated. This assumption will be supported by the channel estimation results of the following section (see also Lemma 2).

To say more about the SIR expression requires an analysis of the $L \times L$ matrix Z_1. Clearly, this term depends on the particular realization of the sequences and this makes it difficult to give any general measures of performance. The situation changes, however, if we consider a limiting regime where $N \to \infty$ with $K = \alpha N$ (with L fixed). If almost surely the empirical distribution of the eigenvalues of D_1 converges to a fixed nonrandom distribution $F(p)$, then we have the following result. (See Appendix I for the formal definition of almost sure convergence of empirical distributions.)

Theorem 1 (LMMSE Receiver: Data Estimation): SIR_1 converges to

$$\text{SIR}_1^* = \frac{\sum_{l=1}^{L} |\overline{a}_{1l}|^2 \beta_d}{1 + \xi_1^2 \beta_d}$$

almost surely as $N \to \infty$ where β_d is the unique solution to the equation

$$\beta_d = \left[\sigma^2 + \alpha L \int_0^\infty \frac{p}{1 + p\beta_d} dF(p)\right]^{-1}. \qquad (3)$$

Proof: We have immediately from Theorem 7 in Appendix I that Z_1 converges almost surely elementwise to $\beta_d I$. The convergence of SIR_1 to SIR_1^* follows since SIR_1 is a continuous function of the elements of the $(L \times L)$ matrix Z_1. □

Remark 2: It should be clear from Theorem 7 that this result holds regardless of the shape of the distribution of the elements of S. Convergence in probability of the SIR in the perfectly known nonfading channel was proved for general signature sequences in [30, Theorem 3.1].

Remark 3: Equation (3) and all other fixed-point equations in this paper are easy to solve numerically by simple iteration. Convergence to the solution is guaranteed from any initial positive value, and usually happens very rapidly.

We have assumed that Ξ_1 is invertible ($\xi_1^2 > 0$), however, if the channel of user 1 is known perfectly then this will not be case and instead we obtain directly from (2) that the SIR is $a_1^H s_1^H Z_1^{-1} s_1 a_1$ which converges in probability to $a_1^H a_1 \beta_d$. We see that β_d has an interpretation as the SIR for user 1 when its own channel is known perfectly and the total energy in all its paths is 1. We call β_d the normalized SIR of user 1 although the user index is unnecessary here because the normalized SIR will be equal for all users in the system. Observe that β_d contains all the information related to the interference suppression capabilities of the receiver and that we have a separation of the effects of the estimate of, and uncertainty in, the channel of user 1 and the corresponding quantities for interfering users. The normalized SIR, β_d, thus provides a measure which isolates the effect of other user channel uncertainty, and which fully captures the multiuser properties of various receivers.

To say more requires us to obtain information about the eigenvalues of D_1. In this direction assume that

$$\boldsymbol{E}_d[a_k a_k^H] = \overline{a}_k \overline{a}_k^H + \xi_k^2 I$$

where $\overline{a}_k$ is the estimate of the channel gains for the paths of user k and ξ_k^2 is a scalar representing the common variance in these estimates. We have assumed that the errors $a_{kl} - \overline{a}_{kl}$ for fixed k are uncorrelated and have equal variance and note briefly that this assumption will be supported by the channel estimation analysis of the following section. With this assumption we see that D_1 has one eigenvalue at $\overline{a}_k^H \overline{a}_k + \xi_k^2$ and $(L-1)$ eigenvalues at ξ_k^2 for $k = 2, \ldots, K$ so that in a large system we will have the approximate relation

$$\beta_d = \left[\sigma^2 + \frac{1}{N} \sum_{k=2}^{K} \left((L-1)I(\xi_k^2, \beta_d) + I\left(\sum_{l=1}^{L} |\overline{a}_{kl}|^2 + \xi_k^2, \beta_d\right)\right)\right]^{-1}$$

where

$$I(p, \beta) = \frac{p}{1 + p\beta}.$$

The overall effect of interferer k is given by the term

$$(L-1)I(\xi_k^2, \beta_d) + I\left(\sum_{l=1}^{L} |\overline{a}_{kl}|^2 + \xi_k^2, \beta_d\right)$$

which is the same as the interference that would result in the single-path fading case ($L = 1$) from $L - 1$ users with power ξ_k^2 and one user with power $\sum_{l=1}^{L} |\overline{a}_{kl}|^2 + \xi_k^2$. Two special cases emerge

- when the channel is known perfectly ($\xi_k^2 = 0$) then an interferer looks like a single interferer with power $\sum_{l=1}^{L} |a_{kl}|^2$;
- when nothing is known of the channel and ξ_k^2 is simply the *a priori* average power in each resolvable path then an interferer looks like L interferers with power $\overline{p}_k / L$.

Due to the convexity of the effective interference as a function of power, one high-power interferer is more benign than L interferers with the same total power especially when the background noise power is low and this is why there is so much potential gain from obtaining accurate channel estimates. The extent to which the uncertainty in the estimates causes a single interferer to spill into L dimensions is captured very neatly in our framework. We note that similar conclusions have been obtained by more heuristic dimensionality arguments in [26] and [36].

To say more we will need statistical descriptions of the mean and variance information that is used by the data estimator so that we can specify $F(p)$, the limiting empirical distribution function of the eigenvalues of D_1. We leave this until Section V after we have looked at the channel estimation problem in Section IV.

B. Decorrelating, Multipath-Combining Receiver

In this section we consider a decorrelating receiver which is variously known as the post-combining decorrelator [16], the multipath decorrelating detector [15], and the decorrelating multipath-combining detector [18]. (We also note [37] which looks at an LMMSE equivalent of these receivers.) All reference to the decorrelator in this section refers to the decorrelating, multipath-combining receiver.

The first stage of the receiver corresponds to a decorrelator which treats every path of every user as a unique interferer. The second stage is to combine the L outputs corresponding to the paths of a particular user either coherently, if channel information is available at the combiner, or noncoherently. Whether coherent or noncoherent combining is used it is clear that the receiver for user 1 does not require any information about the channels of the other users (apart from the timing of the various resolvable multipath components).

The first stage of the decorrelator is based on processing the received vector by S^+ the Moore–Penrose generalized inverse of the signature matrix S. In the case, when S has full column rank, $S^+ = (S^H S)^{-1} S^H$ and the output is given by

$$\overline{z} = Ab + \overline{n}$$

where $\overline{n}$ has covariance matrix $\sigma^2 (S^H S)^{-1}$.

The second stage (for user 1) takes the first L components of $\overline{z}$ and combines them taking into account the covariance structure of the noise and of the channel estimate. Let $\overline{z}_1$ represent the vector consisting of the L decorrelator outputs of user 1. Then

$$\overline{z}_1 = a_1 b + \overline{n}_1$$

where $\overline{n}_1$ has covariance R equal to the first $L \times L$ sub-block of $\sigma^2 (S^H S)^{-1}$. Consider an LMMSE combiner for estimating b_1 from $\overline{z}_1$. Note that such a combiner is optimal in the sense of producing a scalar sufficient statistic for b_1 if $a_1 - \overline{a}_1$ is a circularly symmetric complex Gaussian vector. Then the combiner forms the decision statistic

$$z = \frac{\overline{a}_1^H (\Xi_1 + R)^{-1} \overline{z}}{1 + \overline{a}_1^H (\Xi_1 + R)^{-1} \overline{a}_1}$$

with resultant SIR $\overline{a}_1^H (\Xi_1 + R)^{-1} \overline{a}_1$.

If $\alpha L = LK/N < 1$ then we have the following result as $N \to \infty$.

Proposition 1: (Decorrelating, Multipath-Combining Receiver: Data Estimation): The SIR for the decorrelator converges almost surely to the value

$$\text{SIR}_1^* = \frac{\sum_{l=1}^{L} |\overline{a}_{1l}|^2 \beta_d}{1 + \xi_1^2 \beta_d} \tag{4}$$

where the normalized SIR is given by

$$\beta_d = \frac{1 - \alpha L}{\sigma^2} = \left[\sigma^2 + \alpha L \frac{\sigma^2}{1 - \alpha L} \right]^{-1}. \tag{5}$$

Proof: With the assumption that the signature sequences of the different multipath components are independent, this result is fairly straightforward. The new aspect of the problem is the combining operation after decorrelation which means we have to worry about the off-diagonal elements of $(S^H S)^{-1}$. However, it can be shown (although it is not obvious) that the off-diagonal elements almost surely converge to zero so that R converges almost surely (elementwise) to $\frac{\sigma^2}{(1-\alpha L)} I$. The convergence of $(\Xi_1 + R)^{-1}$ to $(\Xi_1 + \frac{\sigma^2}{(1-\alpha L)} I)$ follows since the elements of the inverse matrix are continuous functions of the elements of the original $(L \times L)$ matrix. □

In this case, the SIR is independent of the powers of the interferers, since each interferer is nulled out. The effective interference is $\frac{L}{\beta_d}$ for each interferer, and does not depend on its power.

We note that the SIR for the decorrelator approaches 0 as αL approaches 1. The decorrelator takes an alternative form when $\alpha L > 1$ (the pseudo-inverse is still well-defined) and leads to a nonzero SIR. While the decorrelator still does not depend on the powers of the interfering users, the SIR does depend on the interference distribution when $\alpha L > 1$. Asymptotic results can be derived but in this paper we will give results for the decorrelator only when $\alpha L < 1$.

C. Single-User Matched Filter

We now consider a receiver which is based only on information about the desired user and which we call a matched filter. The matched filter we consider is simply $c = \sum_{l=1}^{L} \overline{a}_{1l} s_{1l}$ and leads to the following result for large systems (the proof is relatively straightforward and is omitted):

Proposition 2: (Single-User Matched Filter: Data Estimation): The SIR for the matched filter converges almost surely to the value

$$\text{SIR}_1^* = \frac{\sum_{l=1}^{L} |\overline{a}_{1l}|^2 \beta_d}{1 + \xi_1^2 \beta_d}$$

where the normalized SIR is given by

$$\beta_d = \left[\sigma^2 + \alpha L \int_0^\infty p \, dF(p) \right]^{-1} \tag{6}$$

where again F is the almost sure limiting empirical distribution function of the eigenvalues of D_1 (assuming such a limiting distribution exists).

In a large system we will have the approximate relation

$$\beta_d = \left[\sigma^2 + \frac{1}{N} \sum_{k=2}^{K} \left((L-1) I(\xi_k^2) + I\left(\sum_{l=1}^{L} |\overline{a}_{kl}|^2 + \xi_k^2 \right) \right) \right]^{-1}$$

where $I(p) = p$ is the effective interference of user k. Note that for the matched filter, the effective interference is linear in the interferer power which should be contrasted to the LMMSE receiver, for which $I(p) = \frac{p}{1+\beta p}$ at normalized SIR β. The performance of the matched filter will become arbitrarily bad as the power of an interferer is increased, while for fixed β, the LMMSE effective interference for a high-power user approaches $1/\beta$. For the matched filter, L low-power interferers have exactly the same impact as one interferer with the same total power.

IV. Channel Estimation

We now turn to the problem of estimating the channel fading process of each user in the system. Our eventual aim is to use these estimates and their corresponding MSEs as inputs to the LMMSE data estimator designed for the partially known channel. Note that our model assumes that the time delays of the resolvable multipath components of all users are known and that we consider estimation of the path gains only. We will perform the channel estimation conditional on the data, an assumption that is valid during a training period and that leads to performance bounds for the situation when the channel estimator is operating in a decision directed mode (since we assume that the data are perfectly known and do not allow for errors).

All expectations in this section should be seen as conditional on the information that is assumed known at the channel estimator, that is, on the signature sequences, the data and the average powers of the users. We use the notation $\boldsymbol{E}_c$ to denote expectation conditioned on this information.

To begin we recall the signal model under consideration

$$y(m) = \sum_{k=1}^{K} b_k(m) \sum_{l=1}^{L} a_{kl}(m) s_{kl}(m) + n(m)$$

Let

$$\overline{S}(m) = [b_1(m)s_{11}(m), \dots, b_1(m)s_{1L}(m), \dots, b_K(m)s_{K1}(m), \dots, b_K(m)s_{KL}(m)]$$

and

$$a(m) = [a_1(m)^T, \dots, a_K(m)^T]^T$$

where

$$a_k(m) = [a_{k1}(m), \dots, a_{kL}(m)]^T.$$

Then the observation vector can be expressed as

$$y(m) = \overline{S}(m)a(m) + n(m).$$

Note that conditional on the data and the signature sequences (i.e., conditional on $\overline{S}(m)$), the problem of channel estimation is one of Gaussian estimation for which LMMSE estimation and MMSE coincide (since $y(m)$ and $a(m)$ are jointly circularly symmetric). If a is a Markov process then the MMSE estimate of $a(m)$ based on $y(1), y(2), \dots, y(m-1)$ along with the error covariance can be recursively computed via the Kalman filter equations.

We will consider jointly estimating the channel parameters of all users over an *estimation window* of τ symbols and restrict attention to the situation where the channel coherence time, over which the channel is essentially constant, is greater than τ symbol intervals. In a system based on periodically sending training data (and estimating the channel over the training period), τ would typically be chosen to be a small fraction of the channel coherence time so that the training overhead is not too large.

Over the estimation window, we can thus drop the time dependence of the channel gains leading to the model

$$y(m) = \sum_{k=1}^{K} \sum_{l=1}^{L} a_{kl} b_k(m) s_{kl}(m) + n(m), \qquad m = 1, \dots, \tau$$

with $\boldsymbol{E}_c[a_{rs}a_{tu}^*] = \delta_{rt}\delta_{su}\overline{p}_{rs}$. Letting

$$y = [y(1)^T, \dots, y(\tau)^T]^T$$

and

$$n = [n(1)^T, \dots, n(\tau)^T]^T$$

we have

$$y = \tau^{1/2}\overline{S}a + n$$

where

$$\overline{S} = \tau^{-1/2}[\overline{S}(1)^T, \dots, \overline{S}(\tau)^T]^T.$$

We wish to estimate a based on observation of y which we observe to be a standard problem in Gaussian estimation. The resulting MMSE estimate (the conditional mean estimate) is given by

$$\overline{a} = \tau^{-1/2}\,\frac{\overline{p}}{L}\,\overline{S}^H \left(\frac{\overline{p}}{L}\,\overline{S}\,\overline{S}^H + \frac{\sigma^2}{\tau}I\right)^{-1} y \qquad (7)$$

and the error covariance is

$$\Xi = \frac{\overline{p}}{L}I - \frac{\overline{p}^2}{L^2}\,\overline{S}^H \left(\frac{\overline{p}}{L}\,\overline{S}\,\overline{S}^H + \frac{\sigma^2}{\tau}I\right)^{-1} \overline{S}. \qquad (8)$$

Now let

$$s_{kl} = \tau^{-1/2}[b_k(1)s_{kl}(1)^T, \dots, b_k(\tau)s_{kl}(\tau)^T]^T$$

and

$$\overline{S}_{11} = [\overline{s}_{12}, \dots, \overline{s}_{1L}, \dots, \overline{s}_{K1}, \dots, \overline{s}_{KL}].$$

The expression for the MSE of a_{11} (which is the $(1, 1)$ element of the covariance matrix Ξ) can be written as

$$\xi_{kl}^2 11 = \frac{\overline{p}/L}{1 + \overline{p}/L\beta_c^N}$$

where

$$\beta_c^N = \overline{s}_{11}^H \left(\frac{\overline{p}}{L}\,\overline{S}_{11}\overline{S}_{11}^H + \frac{\sigma^2}{\tau}I\right)^{-1} \overline{s}_{11}.$$

We now consider the asymptotic regime where $N \to \infty$ with $\alpha = K/N$ fixed.

Theorem 2: (MMSE Channel Estimation): The MSE for any path of any user converges almost surely as $N \to \infty$ to the nonrandom

$$\xi^2 = \frac{\frac{\overline{p}}{L}}{1 + \frac{\overline{p}}{L}\beta_c}$$

where β_c satisfies the equation

$$\beta_c = \left[\frac{\sigma^2}{\tau} + \frac{\alpha L}{\tau}\,\frac{\frac{\overline{p}}{L}}{1 + \frac{\overline{p}}{L}\beta_c}\right]^{-1} \qquad (9)$$

from which β_c can be solved explicitly

$$\beta_c = \frac{\tau - \alpha L}{2\sigma^2} - \frac{L}{2\overline{p}} + \left[\frac{(\tau - \alpha L)^2}{4\sigma^4} + L\frac{(\tau + \alpha)}{2\overline{p}\sigma^2} + \frac{L^2}{4\overline{p}^2} \right]^{1/2}$$

Proof: The result follows immediately from Theorem 7 in Appendix I upon observing that the elements of $\overline{S}$ remain i.i.d. in the presence of the data modulation. □

We see immediately from (9) that along with the reduction of background noise power by a factor of τ, the number of degrees of freedom (the processing gain) is increased by a factor of τ (α is reduced by a factor of τ). As τ increases, the contribution of the interference to β_c becomes negligible very quickly and the limiting MSE is well approximated by σ^2/τ, the value that would result for a single user in the absence of other users.

The multiple paths of each interferer look like separate interferers to the LMMSE channel estimator. As L is increased, with τ and the total power of each user constant, the system is equivalent to a single-path system with a large number of low-power interferers, and the performance of the LMMSE receiver would approach that of a matched-filter channel estimator. In this situation, we have that $\frac{\overline{p}}{L}\beta_c$ is small compared to 1 so that the ratio the estimation error to the energy per path is close to 1 and we essentially know very little other than the *a priori* statistics of the channel. As an example, a system with $\alpha = 0.5$, $L = 10$, $\tau = 1$, and $\overline{p}/\sigma^2 = 20$ dB will have an MSE to power per path ratio of 0.8 and channel estimation would buy only minimal performance gain. (We also draw the readers attention to [38] which treats some similar issues in the context of single-user fading channels from an information-theoretic point of view.)

The condition for other user interference to be negligible is that $\alpha L/\tau$ is small. If we consider increasing L but now with $\alpha L/\tau$ held fixed at a value significantly smaller than 1, then the ratio of MSE to power per path is roughly $\frac{\sigma^2\tau}{\overline{p}L}$. If $\tau/L = 5$ and $\overline{p}/\sigma^2 = 20$ dB then this ratio is 0.002.

A. Extension of Results to Repeated Sequences

While the effective increase in degrees of freedom from N to τN is intuitively what one would expect when the signature sequences are independently chosen from symbol to symbol, it is not immediately obvious that the effect would carry over to the case when the signature sequences are repeated. In some sense one would think that the sequence repetition would entail some loss in degrees of freedom. However, it turns out that sufficient randomness in the data is adequate to render the performance asymptotically equal in both cases.

Theorem 3: (MMSE Channel Estimation-Repeated Sequences): Under the additional assumption that the data symbols are zero-mean ($E[b_k(m)] = 0$), the conclusions of Theorem 2 also hold in a system using repeated sequences, i.e., $s_{kl}(m) = s_{kl}$ for all m.

Proof: See Appendix III. □

In Appendix III, we will give an explanation of this phenomenon using the concept of *freely independent* random matrices. Indeed, in the case of repeated sequences, the key random matrix $\overline{S}$ has dependent entries and this brings us beyond existing random matrix results. Techniques from free probability theory are appropriate for tackling this problem.

B. Extension of Results to Shifted Sequences

The fact that the multiple paths of each interferer look like separate interferers in Theorem 2 is a direct consequence of the assumption that the signature sequences along the different paths are independent. What happens if they are shifted replicas of the same transmitted sequence?

Theorem 4: (MMSE Channel Estimation-Shifted Sequences): Theorem 2 holds even if for each k and l, the signature sequence $\overline{s}_{kl}(m)$ is a cyclic shifted replica of the random sequence $\overline{s}_{k1}(m)$ by $l-1$ chips.

Proof: See Appendix III. □

The proof of the result in Appendix III gives an explanation of this curious phenomenon, again using the notion of freely independent random matrices. Basically, the shifting provides enough randomness even though there is sequence replication.

V. Estimator Coupling

In Sections III and IV, we looked at the performance of data and channel estimation, respectively. In this section, we couple these results together based on the receiver structure of Fig. 1. The theoretical result on which the coupling hinges, is the following result.

Theorem 5: When the channel estimates and error covariance are calculated using (7) and (8), respectively, then the empirical distribution of the eigenvalues of D_1 converges in probability to the fixed distribution

$$F(p) = \frac{L-1}{L}u(p - \xi^2) + \frac{1}{L}G(p) \tag{10}$$

where $G(p)$ is the distribution function of the random variable $\sum_{l=1}^{L} |x_l|^2 + \xi^2$, where each x_l is a circularly symmetric, zero-mean, complex Gaussian random variable with variance $\frac{\overline{p}}{L} - \xi^2$ and $x_1, \ldots, x_L$ are independent. In the above, ξ^2 is as given in Theorem 2.

Proof: See Appendix II, where the definition of convergence in probability of (random) empirical distributions is also given. □

To get some feel for this result we note that the $(K-1)L$ eigenvalues of D_1 are made up of the L eigenvalues of $\overline{a}_k\overline{a}_k^H + \Xi_k$ for each $k = 2, \ldots, K$. We also know (see Lemma 2) that Ξ_k converges to $\xi^2 I$ for each k so that we would expect the eigenvalues of $\overline{a}_k\overline{a}_k^H + \Xi_k$ to be close to those of $\overline{a}_k\overline{a}_k^H + \xi^2 I$, the latter matrix having $L-1$ eigenvalues at ξ^2 and one eigenvalue at $\overline{a}_k^H\overline{a}_k + \xi^2$. Now it is not difficult to show that the limiting (marginal) distribution of $\overline{a}_k^H\overline{a}_k$ is $G(p)$ for every user. Theorem 5 goes further stating that the empirical distribution across users converges to $G(p)$, a result which is not immediately obvious due to the dependence between users.

We now want to combine this with Theorem 1 to yield the limiting SIR for user 1. One technical point is that Theorem 5 only yields convergence *in probability* of the eigenvalue distribution of D_1 whereas Theorem 1 requires almost sure convergence of the eigenvalue distribution of D_1 to ensure almost sure convergence of the SIR for user 1. Corollary 2 in Appendix I shows that we can prove an analogous result to Theorem 1 where convergence in probability of the eigenvalue distribution of D_1 implies

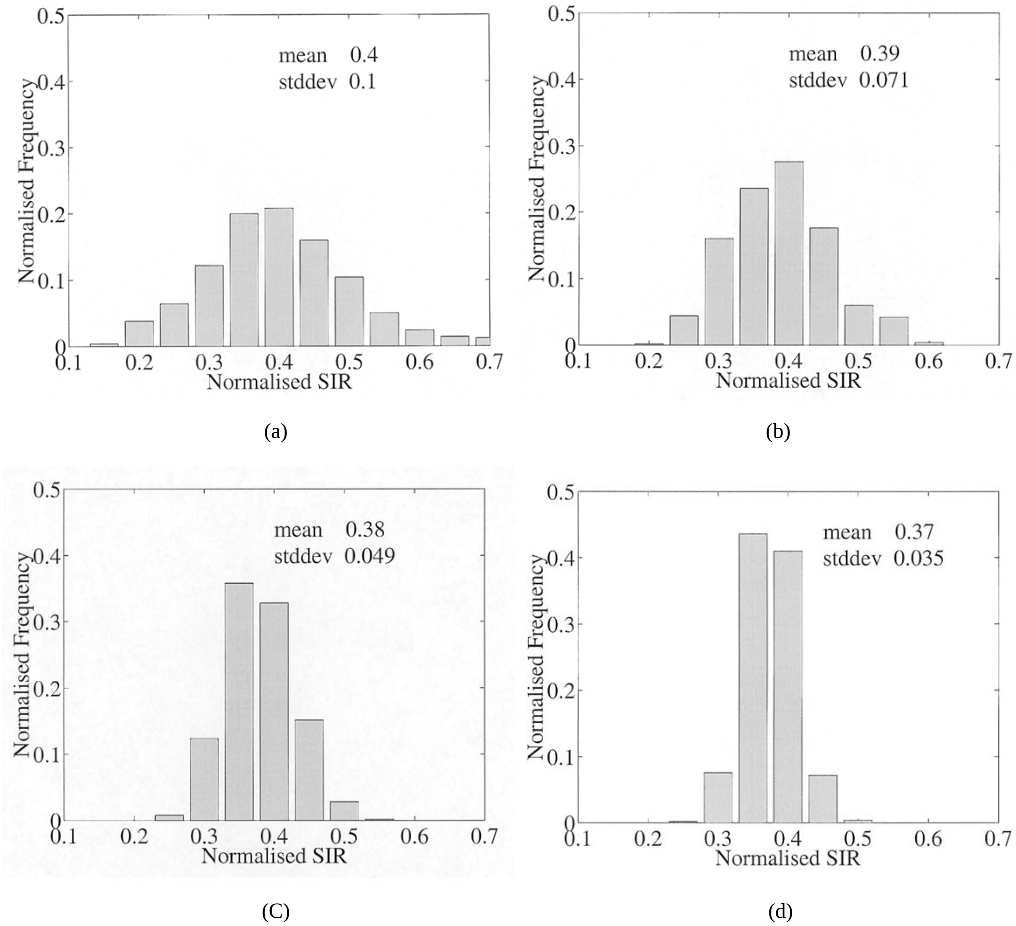


Fig. 2. Histograms of normalized SIR ($\beta_d^N \sigma^2$) for various N. In each case $K = N/2$, $\tau = 2$, $L = 2$, and $\overline{p}/\sigma^2 = 20$ dB. (a) $N = 32$. (b) $N = 64$. (c) $N = 128$. (d) $N = 256$.

convergence in probability of the SIR. Hence we can conclude that the limiting SIR (in probability) of user 1 is given by

$$\mathrm{SIR}_1 = \frac{\sum_{l=1}^{L} |\overline{a}_{1l}|^2 \beta_d}{1 + \xi_1^2 \beta_d} \tag{11}$$

and β_d is a constant satisfying the fixed-point equation

$$\beta_d = \left[\sigma^2 + \alpha(L-1) \frac{\xi^2}{1 + \xi^2 \beta_d} + \alpha \int_{\xi^2}^{\infty} \frac{p}{1 + p\beta_d} g(p)\, dp \right]^{-1} \tag{12}$$

where

$$g(p) = \frac{1}{(\frac{\overline{p}}{L} - \xi^2)^L (L-1)!} (p - \xi^2)^{L-1} \exp\left(- \frac{p - \xi^2}{\frac{\overline{p}}{L} - \xi^2} \right), \quad p \geq \xi^2 \tag{13}$$

and the $\overline{a}_{1l}$'s are i.i.d. circularly symmetric, zero-mean complex Gaussian random variables with variance $\frac{\overline{p}}{L} - \xi^2$ (with ξ^2 given in Theorem 2). Thus the SIR is asymptotically chi-square-distributed with $2L$ degrees of freedom.

We also have a coupling result for the matched filter which follows directly from Theorem 5 and Proposition 2. The SIR for the matched filter converges in probability to the value

$$\mathrm{SIR}_1^* = \frac{\sum_{l=1}^{L} |\overline{a}_{1l}|^2 \beta_d}{1 + \xi_1^2 \beta_d} \tag{14}$$

where the normalized SIR is given by

$$\beta_d = [\sigma^2 + \alpha \overline{p}]^{-1}. \tag{15}$$

Note that there is no need for a coupling result for the decorrelator as the SIR (as given in Proposition 1) is independent of the channels of the interfering users.

VI. Numerical Examples

In this section, we present some numerical examples to support the intuition gained from the preceding analysis. The main conclusion will be that in frequency-selective fading, there is much to be gained from knowledge of the channels of interfering users. Similar conclusions were also reached in [26], [36].

We first focus on the interference suppression capabilities of the linear receivers and assume the channel of the user of interest (user 1) is known perfectly ($\overline{a}_1 = a_1$ and $\xi_1^2 = 0$). We will compare receivers by the ratio of SIR/SNR where SNR =

$(\sum_{l=1}^{L} |a_{1l}|^2)/\sigma^2$. This measure is related to the previously defined normalized SIR β_d, through

$$\frac{\text{SIR}}{\text{SNR}} = \beta_d \sigma^2$$

and differs from β_d in that it depends on $\overline{p}$ and σ^2 only through their ratio. In this section, we reserve the term *normalized* SIR for the quantity $\beta_d \sigma^2$.

To begin, we present some simulation results to give some idea of the rate of convergence of the SIR values to their asymptotic limit. The results are presented in Fig. 2. The plots show histograms of 500 realizations of normalized SIR as the processing gain N is increased. The normalized SIR values are obtained from (2) with channel gain means and covariances calculated from (7) and (8). All plots use $\overline{p}/\sigma^2 = 20$ dB, $\tau = 2$, $\alpha = 0.5$ $(K = N/2)$ and $L = 2$. The frequency axis has been normalized by dividing by the number of realizations. Note that the results were generated using repeated sequences with multipath components modeled as cyclic-shifted versions of a common sequence. The asymptotic limit can be calculated from (12) with ξ^2 calculated using Theorem 2 and is $\beta_d = 0.38$.

From this point on all results will be based on our asymptotic SIR expressions. All results for LMMSE receivers are obtained from (11)–(13), while results for the decorrelator use (4) and (5) and results for the matched filter use (14) and (15).

A. Frequency-Flat Fading (Fig. 3)

Let us start by considering the special case of frequency-flat fading $(L = 1)$.

Fig. 3 shows plots of SIR/SNR versus α at various levels of average SNR $= \overline{p}/\sigma^2$. Plots are shown for LMMSE receivers with perfect channel estimation $(\xi^2 = 0)$ and with no channel estimation $(\xi^2 = \overline{p})$, the decorrelator (for which SIR/SNR = 1–α) and the matched filter (for which SIR/SNR$=(1+\alpha\frac{\overline{p}}{\sigma^2})^{-1}$).

The key observation is that the ideal LMMSE and the worst case LMMSE show very little performance difference (< 1 dB) over the range of α and $\overline{p}/\sigma^2$ covered with the performance gap increasing with both α and $\overline{p}/\sigma^2$. One should be aware that this observation is somewhat sensitive to the fading distribution (Rayleigh in this case). The gap between the ideal and worst case LMMSE receivers could be made very significant by careful choice of the fading distribution. In particular, suppose that the channel gain of each user took two values, one with large magnitude which occurs with small probability and one that is zero occurring with high probability. The ideal LMMSE is based on the small number of high-power interferers which it would simply null out, while the worst case LMMSE is based on a large number of interferers at the average power. Because of the convexity of the effective interference as a function of the power of the interferer, the ideal LMMSE would thus have significantly better performance.

B. Frequency-Selective Fading (Fig. 4)

In the frequency-flat fading channel we observed that, at least in terms of interference suppression, the price of knowing only the average power of other users or even knowing nothing at all about other users' powers, did not result in a very dramatic loss in performance relative to the ideal LMMSE receiver. The real interest in the multipath fading channel results because this observation no longer holds: to throw away channel information is very wasteful of degrees of freedom.

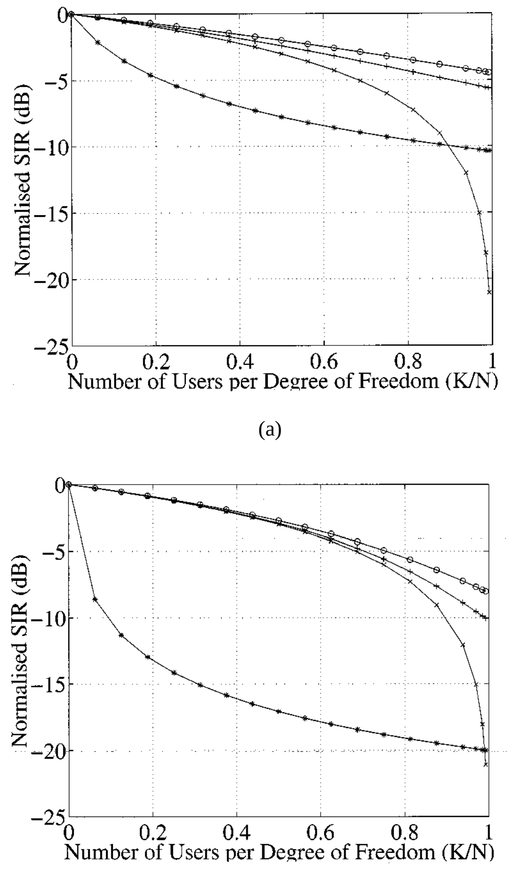
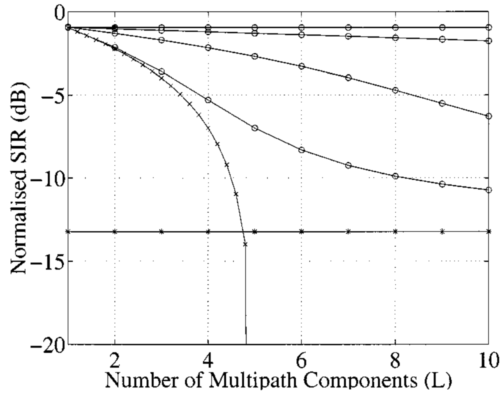


Fig. 3. Plots of normalized SIR $(\beta_d \sigma^2)$ versus the number of users per degree of freedom (α) for the ideal LMMSE $(\circ)$, the worst case LMMSE $(+)$, the decorrelator $(\times)$, and the matched filter $(*)$. (a) $\frac{\overline{p}}{\sigma^2} = 10$ dB. (b) $\frac{\overline{p}}{\sigma^2} = 20$ dB.

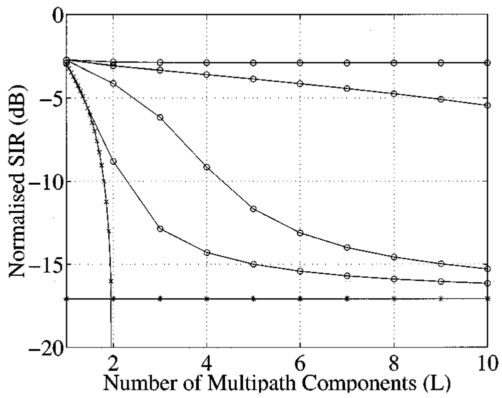
Fig. 4 shows plots of normalized SIR $(\beta_d \sigma^2)$ versus number of paths (L) for the LMMSE receiver, the decorrelator, and the matched filter. For the LMMSE receiver, results are shown for various length estimation windows which translate to various values of the path estimation error ξ^2.

The LMMSE receiver with worst performance corresponds to the receiver which knows nothing of the channels other than the average power. In this case, each user with total power $\overline{p}$ looks like L users with power $\overline{p}/L$. Note that even for high $\overline{p}$, as L is increased the power of each effective user decreases which leads the LMMSE performance to approach that of the matched filter. Note that unless the system is very lightly loaded (small α), so that αL is not too close to 1, the decorrelating, multipath-combining receiver is virtually useless.

We also observe that there is a significant difference between the ideal and worst case LMMSE receivers, even when there are only a few multipath components (10 dB for $\alpha = 0.5$ and $L = 3$). The question is, how good do our channel estimates need to be to achieve a significant improvement? For the parameter values considered here, provided the channel estimation window (in symbols) is at least as large as the number of multipath components, then the loss from ideal is less than 2 dB.

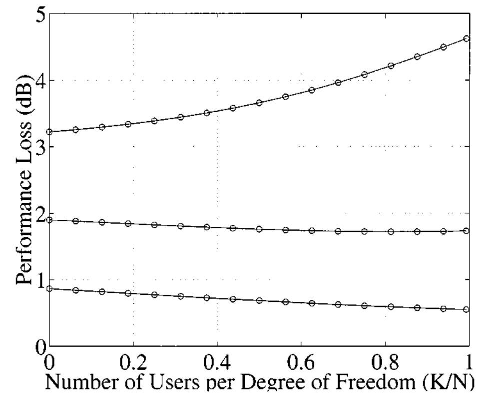


(a)

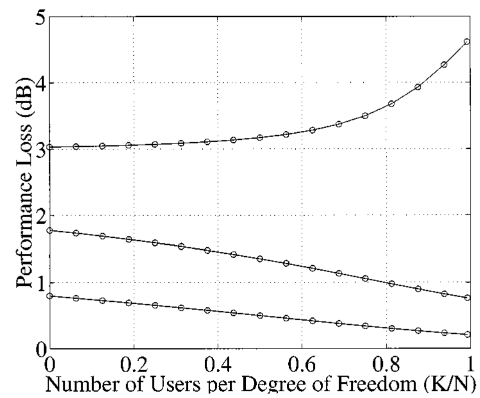


(b)

Fig. 4. Normalized SIR ($\beta_d \sigma^2$) for various α. In each case $\overline{p}/\sigma^2 = 20$ dB. Results are shown for the matched filter ($*$), the decorrelator ($\times$), and the LMMSE receiver ($\circ$). In the latter case, curves are shown for estimation window lengths of (from the top) $\tau = \infty$ (perfectly known channel), $\tau = 10, 2$, and finally for the case when nothing is known about the channels other than the *a priori* average power. (a) $\alpha = 0.2$. (b) $\alpha = 0.5$.



(a)



(b)

Fig. 5. Plots of performance loss for the LMMSE receiver versus the number of users per degree of freedom (α). In each case $L = 1$. Results are shown for channel estimator window lengths of (from the top) $\tau = 1, 2$, and 5. (a) $\frac{\overline{p}}{\sigma^2} = 10$ dB and (b) $\frac{\overline{p}}{\sigma^2} = 20$ dB.

This does not seem to be an unreasonable assumption, the result providing great motivation for acquiring and making use of the signature sequences of all users.

These results assume that the channel of the user of interest is perfectly known. It is also of interest to examine the impact of including the uncertainty in the estimate of the path gains for user 1. To do this we compare the value of SIR obtained in the cases where the channel is perfectly known and when the channel estimates have some associated uncertainty. The results are shown in Fig. 5 (frequency-flat fading) and Fig. 6 (frequency-selective fading). The vertical axis represents

$$10 \log \overline{\mathrm{SIR}}_1^* |_{\xi^2=0} - 10 \log \overline{\mathrm{SIR}}_1^* |_{\xi^2=\xi^2(\tau)}$$

where

$$\overline{\mathrm{SIR}}_1^* = \frac{\overline{p} - L\xi^2}{\xi^2 + \frac{1}{\beta_d}}$$

with β_d given by (12). The average SIR values are obtained by averaging the SIR over the channel estimates for user 1. We have written $\xi^2(\tau)$ for the MSE in the channel estimate resulting when the estimator window length (over which we assume the channel is constant) is τ as given in Theorem 2.

C. Frequency-Flat Fading (Fig. 5)

For medium-to-high values of $\overline{p}/\sigma^2$ we see that the performance loss is around 0.5 dB for $\tau = 5$, a value which does not seem unreasonable even in fast-fading environments (say, a Doppler frequency times symbol rate of 0.01). We could, of course, use a differentially coherent scheme to remove the need for obtaining a channel estimate for user 1 but it should be remembered that the cost is on the order of 3 dB. We also remark that most of the performance loss is due to the uncertainty associated with the estimate of the channel of user 1, not the uncertainty about the channels of the other interferers, as should be evident upon referring back to Fig. 3.

It may at first seem strange that the performance gap decreases in some cases as α increases. Surely ξ^2 should increase with α and result in performance degradation relative to the ideal case? The key to understanding this behavior is that we really have different effective values of α for channel and data estimation, namely, α/τ and α, respectively. For $\tau = 5$, for example, as α is increased from 0 to 1, the effective number of users per degree of freedom for channel estimation increases from 0 to 0.2 only. The value of ξ^2 increases only very slightly over this range and the overall performance gap is dominated by the impact of increasing α on the data estimation which in contrast to channel estimation is very significant.

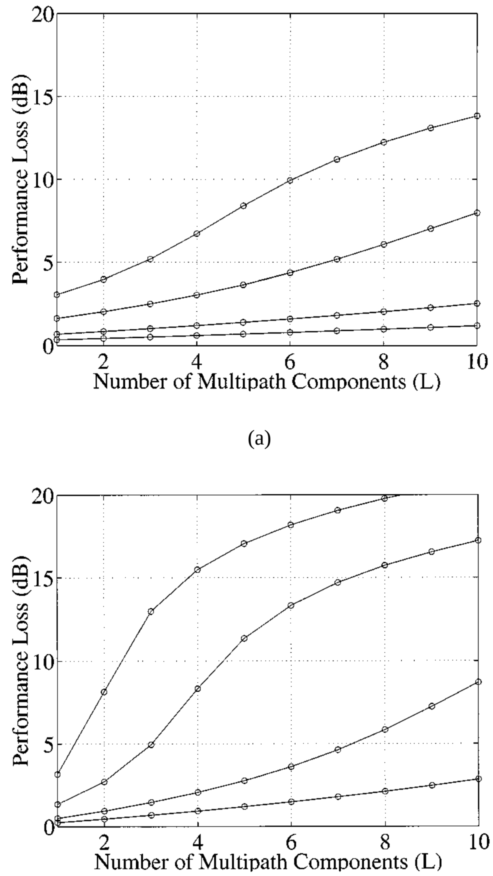


Fig. 6. Performance loss of LMMSE receiver relative to perfect knowledge case. In each case $\overline{p}/\sigma^2 = 20$ dB. Results are shown for channel estimator window lengths of (from the top) $\tau = 1, 2, 5$, and 10. (a) $\alpha = 0.2$. (b) $\alpha = 0.5$.

D. Frequency-Selective Fading (Fig. 6)

Perhaps the main observation that should be made is that again provided τ is at least as large as L, the loss from ideal performance is at most 3 to 4 dB for the range of parameters covered.

VII. Conclusions

We have attempted in this paper to elucidate fundamental properties and limitations of linear multiuser receivers operating in fading channels.

Through a combination of theoretical results and numerical examples, the main engineering insights to come to the surface were as follows.

- In single path fading, the penalty for not knowing the channels of interferers is not significant and the impact of a user not knowing his own channel tends to be the dominant effect.
- In multipath fading, the situation is very different, as we might have guessed from the discussion on effective interference. Receivers making use of accurate estimates of the channels of interferers will significantly outperform receivers designed to operate without channel knowledge or receivers that track only average power.
- But under what conditions could we expect to obtain estimates of sufficient accuracy to allow the potential rewards to be reaped? For the numerical examples considered, we have seen that provided the channel estimator window length (in symbol intervals) is at least as large as the number of resolvable multipath components, estimates of sufficient accuracy for near-optimal performance of the data estimator could potentially be obtained from a linear (optimal) channel estimator.

An important observation is that these insights are totally scalable with system size, in the sense that they are obtained in the limit as the spreading gain and the number of users grow large, as long as the number of users per degree of freedom is fixed.

On the theoretical side, addressing the questions in this paper prompts us to prove new results in the spectral analysis of random matrices, using techniques from free probability theory. These new results manifested in the interesting phenomenon that although there is repetition of random spreading sequences due to multipaths and the use of repeated sequences, in cases considered the asymptotic performance is exactly the same as though all sequences were independently chosen. This provides further evidence for the robustness and versatility of the random spreading sequence model for the performance analysis of multiuser receivers. While we have given rigorous proofs only for LMMSE channel estimation (Theorems 3 and 4), we conjecture that all of our results hold when the repeated and cyclic-shifted signature sequence model is used.

Appendix I
Some Key Results for LMMSE Estimation

Lemma 1: Let A be a deterministic $N \times N$ complex matrix with uniformly bounded spectral radius for all N.[1] Let $q = \frac{1}{\sqrt{N}}[q_1, \ldots, q_N]^T$ where the q_i's are i.i.d. complex random variables with zero mean, unit variance, and finite eighth moment. Let r be a similar vector independent of q. Then

$$\boldsymbol{E}\left[\left|q^H A q - \frac{1}{N}\operatorname{Tr} A\right|^4\right] \leq \frac{C_1}{N^2} \quad \text{and} \quad \boldsymbol{E}[|q^H A r|^4] \leq \frac{C_2}{N^2}$$

where the constants C_1 and C_2 do not depend on N or A.

Proof: The first result follows directly from [39, Lemma 2.7] while the second result can be proved in a similar manner. □

An immediate consequence of this lemma is that

Corollary 1:

$$q^H A q - \frac{1}{N}\operatorname{Tr} A \to 0 \quad \text{and} \quad q^H A r \to 0$$

almost surely as $N \to \infty$.

Proof: We prove the second limiting result and note that the first follows along similar lines. We have

$$P(|q^H A r| > \epsilon) \leq \frac{\boldsymbol{E}[|q^H A r|^4]}{\epsilon^4} \leq \frac{C_1 \epsilon^{-4}}{N^2}$$

[1] That is, there exists a real number, independent of N, which bounds the magnitudes of the eigenvalues of A for all N.

from Markov's inequality and our lemma. Thus

$$\sum_{N=1}^{\infty} P(|q^H A r| > \epsilon) < \infty$$

and the result follows from the first Borel–Cantelli lemma [40]. (Note the implicit dependence of the summands on N.) □

In the sequel, we will be dealing with the convergence of empirical distribution functions, so let us define things clearly.

Definition 1: We say that a sequence of (deterministic) cumulative distribution functions (cdfs) F_n converges to F if for every point of continuity x of F, $\lim_{n\to\infty} F_n(x) = F(x)$. If we identify a cdf with a measure, this is equivalent to *weak convergence* on the space of measures.

Empirical distribution functions are *random cdfs*, for which two notions of convergence can be defined.

Definition 2: A sequence of random cdfs $\boldsymbol{F_n}$ converges *almost surely* to F if almost all realizations F_n converge to F.

Definition 3: A sequence of random cdfs $\boldsymbol{F_n}$ converges **in probability** to a limit F if for every $\epsilon > 0$

$$\lim_{n\to\infty} P[d(\boldsymbol{F_n}, F) > \epsilon] = 0.$$

Here d is any metric which generates the weak topology on the space of measures to which F_n and F belongs.

We need the following theorem which is proved in [41]. For any square matrix with only real eigenvalues let F^A denote the empirical distribution function of the eigenvalues of A. Note that the Stieltjes transform of a distribution G is defined as the analytic function

$$m(z) = \int \frac{1}{\lambda - z} dG(\lambda), \qquad z \in \mathcal{C}^+ \equiv \{z \in \mathcal{C}: \operatorname{Im} z > 0\}.$$

Theorem 6: Let X be a $K \times N$ matrix of i.i.d. complex random variables with zero mean and variance $1/N$ and assume that $\lim_{N\to\infty} K/N = \alpha$. Let T be a $K \times K$ random Hermitian nonnegative-definite matrix independent of X such that almost surely F^T converges to a fixed distribution function F as $N \to \infty$. Then almost surely, F^{XTX^H} converges, as $N \to \infty$, to a (nonrandom) distribution function G whose Stieltjes transform $m(z)\,(z \in \mathcal{C}^+)$ satisfies

$$m = -\left[z - \alpha \int \frac{p}{1+pm}\, dF(p)\right]^{-1} \tag{16}$$

in the sense that for every $z \in \mathcal{C}^+$, $m = m(z)$ is the unique solution in $\mathcal{C}^+$ to (16).

Most of our results on LMMSE estimation follow directly from the following key theorem.

Theorem 7: Let X and T be as defined in Theorem 6 and let q and r be random vectors of length N, independent of X and T and as specified in Lemma 1. Then as $N \to \infty$, almost surely

$$q^H(XTX^H + \sigma^2 I)^{-1} r \to 0 \quad \text{and} \quad q^H(XTX^H + \sigma^2 I)^{-1} q \to \beta^*$$

where β^* is the unique positive solution to the fixed point equation

$$\beta^* = \left[\sigma^2 + \alpha \int \frac{p}{1+p\beta^*}\, dF(p)\right]^{-1}.$$

Proof: To begin we use Corollary 1 to note

$$q^H(XTX^H + \sigma^2 I)^{-1} r$$

and

$$q^H(XTX^H + \sigma^2 I)^{-1} q - \frac{1}{N} \operatorname{Tr}(XTX^H + \sigma^2 I)^{-1}$$

converge almost surely to zero. To apply Corollary 1 we condition on X and T and rely on the fact that q and r are independent of X and T. Note that the spectral radius of $(XTX^H + \sigma^2 I)^{-1}$ is uniformly bounded by $1/\sigma^2$. The first part of the theorem is already proved.

Next observe that

$$\frac{1}{N} \operatorname{Tr}(XTX^H + \sigma^2 I)^{-1} = \int \frac{1}{\lambda + \sigma^2}\, dF^{XTX^H}(\lambda)$$

and that the right-hand side converges almost surely to

$$\beta^* = \int \frac{1}{\lambda + \sigma^2}\, dG(\lambda)$$

since the integrand is bounded and continuous and according to Theorem 6, F^{XTX^H} almost surely converges to G. We have thus established that, almost surely

$$q^H(XTX^H + \sigma^2 I)^{-1} q \to \int \frac{1}{\lambda + \sigma^2}\, dG(\lambda).$$

Since the support of G is on the nonnegative real axis, the Stieltjes transform of G is continuous in the neighborhood of $z = -\sigma^2$ and it follows that

$$\lim_{z\to -\sigma^2} m(z) = \int \frac{1}{\lambda + \sigma^2}\, dG(\lambda) = \beta^*$$

and by the continuity of the right-hand side of (16) as a function of m, it follows that

$$\beta^* = \left[\sigma^2 + \alpha \int \frac{p}{1+p\beta^*}\, dF(p)\right]^{-1}. \qquad \square$$

Now suppose that F^T converges to a fixed distribution F in probability, rather than almost surely. Then we have the following corollary.

Corollary 2: $q^H(XTX^H + \sigma^2 I)^{-1} q$ converges in probability to β^*

Proof: Since F^T converges to F in probability, for every subsequence there exists a further subsequence on which F^T converges to F almost surely. On this subsequence we have from the previous theorem that $q^H(XTX^H + \sigma^2 I)^{-1} q$ converges almost surely to β^* and since the initial subsequence was arbitrary, our result is established. □

APPENDIX II
PROOF OF THEOREM 5

Recall that the channel estimate is given by (7)

$$\overline{a} = \tau^{-1/2} \frac{\overline{p}}{L} \overline{S}^H \left(\frac{\overline{p}}{L} \overline{S}\, \overline{S}^H + \frac{\sigma^2}{\tau} I\right)^{-1} y \tag{17}$$

with associated error covariance (8)

$$\Xi = \frac{\overline{p}}{L} I - \frac{\overline{p}^2}{L^2} \overline{S}^H \left(\frac{\overline{p}}{L} \overline{S}\,\overline{S}^H + \frac{\sigma^2}{\tau} I \right)^{-1} \overline{S}. \qquad (18)$$

Conditional on $\overline{S}$, $\overline{a}$ is a circularly symmetric, complex Gaussian random vector with covariance $\frac{\overline{p}}{L} I - \Xi$.

We have the following result on the elements of Ξ as $N \to \infty$.

Lemma 2:

$$\Xi(i,i) \to \xi^2 \quad \text{and} \quad \Xi(i,j) \to 0, \qquad i \neq j$$

where ξ^2 is given in Theorem 1, and convergence is almost sure.

Proof: The result for diagonal elements is just Theorem 1. For off-diagonal elements we note the equations at the bottom of this page. Now the denominator converges almost surely to $(1 + \frac{\overline{p}}{L}\beta_c)^2$ following Theorem 1 and the numerator converges almost surely to 0 from Corollary 1 so that $\Xi(i,j)$ converges to 0 almost surely as required. □

The following lemma will also be required.

Lemma 3: Suppose $F^{(N)}$ is the empirical distribution function for N random variables. If $F^{(N)}(p)$ converges in probability to $F(p)$ for all p, where F is a fixed distribution function, then the empirical distribution $F^{(N)}$ converges to F in probability.

Proof: Let

$$D^{(N)} = \sup_x |F^{(N)}(x) - F(x)|.$$

We will show that $D^{(N)} \to 0$ in probability from which the convergence of the empirical distributions $F^{(N)}$ to F in probability follows immediately.

First define

$$\phi(u) = \inf[x: u \leq F(x)], \qquad 0 < u < 1$$

and let

$$x_{m,k} = \phi(k/m), \qquad m \geq 1,\ 1 \leq k \leq m.$$

Let

$$D_m^{(N)} = \max_{1 \leq k \leq m} \{ |F^{(N)}(x_{m,k}) - F(x_{m,k})|, \\ |F^{(N)}(x_{m,k}-) - F(x_{m,k}-)| \}.$$

It follows (see, for example, [40, Theorem 20.6 (Glivenko–Cantelli)]) that

$$D^{(N)} \leq D_m^{(N)} + m^{-1}.$$

Now we know that $F^{(N)}(x) \to F(x)$ in probability for all x. Then for each subsequence $\{F^{(N_l)}(x)\}$ there exists a further subsequence $\{F^{(N_{l(i)})}(x)\}$ such that $F^{(N_{l(i)})}(x) \to 0$ almost surely. Using the diagonal method starting from any initial subsequence we can extract a further subsequence such that $F^{(N')}(x_{m,k}) \to F(x_{m,k})$ almost surely for $m \geq 1$, $1 \leq k \leq m$. But this implies that $D_m^{(N')} \to 0$ almost surely and thus that $D^{(N')} \to 0$ almost surely. Since this is true for each initial choice of subsequence, we have that $D^{(N)} \to 0$ in probability as required. □

Now D_1 is a block-diagonal matrix with the $(k-1)$th diagonal block equal to the $L \times L$ matrix $\overline{a}_k \overline{a}_k^H + \Xi_k$. We need to look at the empirical eigenvalue distribution of D_1 which is the empirical distribution function of the eigenvalues of the matrices $\overline{a}_2 \overline{a}_2^H + \Xi_2, \ldots, \overline{a}_K \overline{a}_K^H + \Xi_K$.

Let $\lambda_k^{(N)} = [\lambda_{k1}^{(N)}, \ldots, \lambda_{kL}^{(N)}]^T$ denote the vector of L eigenvalues (counting multiplicities) of $\overline{a}_k \overline{a}_k^H + \Xi_k$, ordered from largest to smallest.

We know that Ξ_k converges almost surely elementwise to $\xi^2 I$ and we would thus expect that for large N, the eigenvalues of $\overline{a}_k \overline{a}_k^H + \Xi_k$ would be close to the eigenvalues of $\overline{a}_k \overline{a}_k^H + \xi^2 I$ and we know what the eigenvalues of the latter matrix look like: it has one eigenvalue at $\overline{a}_k^H \overline{a}_k + \xi^2$ and $(L-1)$ eigenvalues at ξ^2. To be precise, we have

Lemma 4:

$$\lambda_k^{(N)} - \overline{\lambda}_k^{(N)} \to 0$$

almost surely as $N \to \infty$, where

$$\overline{\lambda}_k^{(N)} = [\overline{a}_k^H \overline{a}_k + \xi^2, \xi^2, \ldots, \xi^2]^T.$$

Proof: The proof is a straightforward consequence of the perturbation theory of eigenvalues as given in [42, Ch. 2] (see, for example, the Wielandt–Hoffman Theorem), along with the fact that $\Xi_k - \xi^2 I$ converges almost surely elementwise to the $L \times L$ zero matrix (Lemma 2). □

Let us select a fixed number n of eigenvalues of D_1, $\lambda_{k_1 l_1}^{(N)}, \ldots, \lambda_{k_n l_n}^{(N)}$.

Lemma 5:

$$P(\lambda_{k_1 l_1}^{(N)} \leq \lambda_1, \ldots, \lambda_{k_n l_n}^{(N)} \leq \lambda_n) = \prod_{\{j:\, l_j > 1\}} u(\lambda_j - \xi^2) \prod_{\{j:\, l_j = 1\}} G(\lambda_j)$$

where $u(\lambda)$ is the unit step function and $G(\lambda)$ is as in Theorem 5.

Proof: Let

$$\overline{\lambda}_{kl}^{(N)} = \xi^2 + \overline{a}_k^H \overline{a}_k \delta_{l1}.$$

Lemma 4 tells us that $\lambda_{kl}^{(N)} - \overline{\lambda}_{kl}^{(N)}$ converges almost surely to 0. But this implies [40, Theorem 25.4] that $\lambda_{kl}^{(N)}$ and $\overline{\lambda}_{kl}^{(N)}$ have the

---

$$\Xi(i,j) = -\frac{\overline{p}^2}{L^2} \overline{s}_i^H \left(\frac{\overline{p}}{L} \overline{S}\,\overline{S}^H + \frac{\sigma^2}{\tau} I \right)^{-1} \overline{s}_j$$

$$= -\frac{\frac{\overline{p}^2}{L^2} \overline{s}_i^H \left(\frac{\overline{p}}{L} \overline{S}_{ij} \overline{S}_{ij}^H + \frac{\sigma^2}{\tau} I \right)^{-1} \overline{s}_j}{\left(1 + \frac{\overline{p}}{L} \overline{s}_i^H \left(\frac{\overline{p}}{L} \overline{S}_i \overline{S}_i^H + \frac{\sigma^2}{\tau} I \right)^{-1} \overline{s}_i \right) \left(1 + \frac{\overline{p}}{L} \overline{s}_j^H \left(\frac{\overline{p}}{L} \overline{S}_{ij} \overline{S}_{ij}^H + \frac{\sigma^2}{\tau} I \right)^{-1} \overline{s}_j \right)}$$

same limit in distribution (if such a limit exists). The Cramér–Wold device [40, Theorem 29.4] can then be used to show that the vectors $[\lambda_{k_1 l_1}^{(N)}, \ldots, \lambda_{k_n l_n}^{(N)}]^T$ and $[\overline{\lambda}_{k_1 l_1}^{(N)}, \ldots, \overline{\lambda}_{k_n l_n}^{(N)}]^T$ have the same limit in distribution in $\mathcal{R}^n$. We can thus focus on

$$P(\overline{\lambda}_{k_1 l_1}^{(N)} \le \lambda_1, \ldots, \overline{\lambda}_{k_n l_n}^{(N)} \le \lambda_n)$$

which immediately factors as

$$\prod_{\{j: l_j > 1\}} u(\lambda_j - \xi^2) P\left(\bigcap_{\{j: l_j = 1\}} \{\overline{a}_{k_j}^H \overline{a}_{k_j} + \xi^2 \le \lambda_j\} \right).$$

Recall that $\overline{a}_{k_j}$ when conditioned on Ξ_{k_j} is a circular-symmetric complex Gaussian random vector with covariance $(\overline{p}/L)I - \Xi_{k_j}$. It is relatively straightforward to show using Lemma 2 that in the limit, the $\overline{a}_{k_j}$ become independent circular- symmetric complex Gaussian random vectors with identical covariance $((\overline{p}/L) - \xi^2)I$, the result following immediately. □

Let

$$F^{(N)}(\lambda) = \frac{1}{K-1} \sum_{k=2}^{K} \left(\frac{1}{L} \sum_{l=1}^{L} 1(\lambda_{kl}^{(N)} \le \lambda) \right)$$

be the empirical distribution function of the eigenvalues of D_1.

Theorem 8: $F^{(N)}(\lambda)$ converges in probability to $F(\lambda)$ for all $\lambda \ge \xi^2$ where $F(\lambda)$ is defined in Theorem 5.

Proof: We will show that the variance of $F^{(N)}(\lambda) - F(\lambda)$ goes to zero as $K \to \infty$ which implies convergence in probability.

$$\begin{aligned}
\boldsymbol{E}[|F^{(N)}(\lambda) - F(\lambda)|^2] &= \boldsymbol{E}[(F^{(N)}(\lambda))^2] - F^2(\lambda) \\
&= \boldsymbol{E}\left[\frac{1}{(K-1)^2} \sum_{i=2}^{K} \sum_{j=2}^{K} \cdot \left(\frac{1}{L^2} \sum_{l=1}^{L} \sum_{m=1}^{L} 1(\lambda_{il}^{(N)} \le \lambda) 1(\lambda_{jm}^{(N)} \le \lambda) \right) \right] - F^2(\lambda) \\
&= \frac{1}{(K-1)^2} \sum_{i=2}^{K} \sum_{j=2}^{K} \left(\frac{1}{L^2} \sum_{l=1}^{L} \sum_{m=1}^{L} \cdot P(\lambda_{il}^{(N)} \le \lambda, \lambda_{jm}^{(N)} \le \lambda) - F^2(\lambda) \right).
\end{aligned}$$

The sum is then broken up into

$$\frac{1}{(K-1)^2} \sum_{i=2}^{K} \left(\frac{1}{L^2} \sum_{l=1}^{L} \sum_{m=1}^{L} \cdot P(\lambda_{il}^{(N)} \le \lambda, \lambda_{im}^{(N)} \le \lambda) - F^2(\lambda) \right)$$

and

$$\frac{1}{(K-1)^2} \sum_{i} \sum_{j \ne i} \left(\frac{1}{L^2} \sum_{l=1}^{L} \sum_{m=1}^{L} \cdot P(\lambda_{il}^{(N)} \le \lambda, \lambda_{jm}^{(N)} \le \lambda) - F^2(\lambda) \right).$$

Observing that the summands are exchangeable in the user indices, the first sum is

$$\frac{1}{K-1} \left(\frac{1}{L^2} \sum_{l=1}^{L} \sum_{m=1}^{L} P(\lambda_{2l}^{(N)} \le \lambda, \lambda_{2m}^{(N)} \le \lambda) - F^2(\lambda) \right)$$

which converges to zero, and the second sum is

$$\left(1 - \frac{1}{K-1}\right) \left(\frac{1}{L^2} \sum_{l=1}^{L} \sum_{m=1}^{L} \cdot P(\lambda_{2l}^{(N)} \le \lambda, \lambda_{3m}^{(N)} \le \lambda) - F^2(\lambda) \right)$$

which also converges to zero due to Lemma 5. □

The preceding theorem tells us that for any λ, the random variable $F^{(N)}(\lambda)$ converges in probability to the real constant $F(\lambda)$. The statement that the random distribution function $F^{(N)}$ converges to the fixed distribution function F in probability follows directly from Lemma 3.

APPENDIX III
PROOFS OF THEOREMS 3 AND 4

The proofs of these two theorems require results from *free probability theory*. Our treatment here is very brief; for more details please consult [43] or [44].

Definition 4: A noncommutative probability space $(\mathcal{A}, \varphi)$ is an algebra $\mathcal{A}$ over $\mathcal{C}$ with a unit element I and endowed with a linear functional $\varphi: \mathcal{A} \to \mathcal{C}$, $\varphi(I) = 1$. Elements of $\mathcal{A}$ are called (noncommutative) random variables. We shall also assume that φ has the *trace* property: $\varphi(XY) = \varphi(YX)$ for all $X, Y \in \mathcal{A}$.

An important example is the following. Let $\mathcal{A} = M_N$ be the algebra of complex N by N random matrices whose entries are scalar random variables defined on some underlying common probability space. For each random matrix $X \in M_N$, define

$$\varphi_N(X) := \frac{1}{N} \boldsymbol{E}[\mathrm{Tr}\, X].$$

Note that for $N = 1$, M_1 is simply the algebra of (standard) complex random variables, which are commutative. $\varphi_1(\cdot)$ is simply the expectation operator. For $N > 1$, the elements of M_N are noncommutative.

Definition 5: The *distribution* of $X \in \mathcal{A}$ is specified by the moments $\varphi(X^k)$, for $k \ge 1$. The *joint distribution* of a collection of random variables $X_1, \ldots, X_m \in \mathcal{A}$ is specified by all the joint moments $\varphi(X_{i_1} \ldots X_{i_p})$, $p \ge 1$.

For any N by N random Hermitian matrix $X \in M_N$ with random eigenvalues $\lambda_1, \ldots, \lambda_N$, the rth moment of X in the noncommutative probability space (M_N, φ_N) is given by

$$\varphi_N(X^r) = \frac{1}{N} \boldsymbol{E}[\mathrm{Tr}\, X^r] = \frac{1}{N} \boldsymbol{E}\left[\sum_{i=1}^{N} \lambda_i^r \right].$$

If we let $F_X(\cdot)$ be the expected empirical distribution of the eigenvalues of X

$$F_X(\lambda) := \frac{1}{N} \boldsymbol{E}[|\{i: \lambda_i \le \lambda\}|]$$

then the moments of the distribution F_X are precisely the moments of X as a noncommutative random variable.

Definition 6: A family of subalgebras containing I, $(\mathcal{A}_i)_{i\in I}$ in the noncommutative probability space $(\mathcal{A},\varphi)$ is *free* if $\varphi(X_1,\ldots,X_m) = 0$ whenever $\varphi(X_k) = 0$ for all $k = 1,\ldots,m$ and $X_k \in \mathcal{A}_{i(k)}$ where consecutive indices $i(k) \neq i(k+1)$ are distinct. A family of subsets in $(\mathcal{A},\varphi)$ is *free* if the subalgebras each one of them generates with I are free. Random variables $X_1,\ldots,X_m$ are *free* if the family of subsets $\{X_1\},\ldots,\{X_m\}$ is free.

One should think of the notion of freeness as the noncummutative analog of the notion of independence of (commutative) random variables. For independent random variables, the joint distribution can be specified completely by the marginal distributions. For free random variables, we have the analogous result, which can be proved directly from definition.

Proposition 3: The joint distribution (i.e., moments) of free random variables $X_1,\ldots,X_m$ can be completely specified by the moments of the individual X_i's.

In particular, if X and Y are free, then the moments $\varphi((X+Y)^n)$ of $X+Y$ can be completely specified by the moments of X and the moments of Y.

The notion of freeness is important for us because in a lot of cases of interest, large random matrices become *asymptotically free*.

Definition 7: A sequence of random matrices $X_1^{(N)},\ldots, X_m^{(N)} \in M_N$ is said to be asymptotically free if there exists a noncommutative probability space $(\mathcal{A},\varphi)$ and free random variables $X_1,\ldots,X_m \in \mathcal{A}$ such that all the joint moments of $X_1^{(N)},\ldots,X_m^{(N)}$ converge to the corresponding joint moments of $X_1,\ldots,X_m$ as $N\to\infty$. Analogous definition holds for a family of subsets of random matrices.

The following is the first important result establishing the connection between the asymptotic properties of large random matrices and free probability theory.

Theorem 9 ([45], [46]): Let $X_i^{(N)} \in M_N$ be a random matrix whose entries are complex circular symmetric Gaussian random variables with variance $1/N$ for the off-diagonal terms and $2/N$ for the diagonal terms. The matrix $X_i^{(N)}$ is Hermitian but otherwise the entries are independent. Consider now an independent family of such random matrices $X_i^{(N)}$, $i = 1,\ldots,m, N = 1,\ldots$. Let $\{D_1^{(N)},\ldots,D_k^{(N)}\}$ be a subset of constant Hermitian matrices in M_N such that for each k, the expected empirical eigenvalue distribution of $D_k^{(N)}$ converges as $N\to\infty$. Then the family of subsets $\{X_1^{(N)}\},\ldots,\{X_m^{(N)}\}$, $\{D_1^{(N)},\ldots,D_k^{(N)}\}$ are asymptotically free.

This result was first proved for diagonal matrices $D_k^{(N)}$ in [45] and then extended to general constant Hermitian matrices in [46]. The following is a corollary which does not follow directly but can be proved from Theorem 9.

Corollary 3: Let $Y_i^{(N)} = V_i^{(N)} V_i^{(N)*}$, where $V_i^{(N)}$ is an N by $d_i(N)$ random matrix whose entries are i.i.d. complex circular symmetric Gaussian random variables with variance $1/N$. Consider now an independent family of such random matrices $Y_i^{(N)}$, $i = 1,\ldots,m, N = 1,\ldots$. If $d_i(N)/N \to \alpha_i$ for each i, then the family of subsets $\{Y_1^{(N)}\},\ldots,\{Y_m^{(N)}\}$, $\{D_1^{(N)},\ldots,D_k^{(N)}\}$ are asymptotically free. (The matrices $D_k^{(N)}$ are as specified in Theorem 9.)

We now have the machinery to prove Theorems 3 and 4. First we need the following simple lemma.

Lemma 6: Let $\{X\}$ and $\{U_1,W_1,\ldots,U_M,W_M\}$ be free subsets of random variables in $(\mathcal{A},\varphi)$. If

$$U_m W_m = W_m U_m = I$$

for all m and $\varphi(U_i W_j) = \delta_{ij}$, then the random variables $U_1XW_1,\ldots,U_MXW_M$ are free.

Proof: Consider any $m\geq 1$, and suppose for $k=1,\ldots,m$, Z_k is in the algebra generated by $U_{i(k)}XW_{i(k)}$ and 1 and satisfies $\varphi(Z_k)=0$. The indices satisfy $i(k)\neq i(k+1)$ for all k. Since $W_{i_m}U_{i_m}=1$, it holds that $Z_k = U_{i(k)}T_kW_{i(k)}$ for some T_k in the algebra generated by X and 1 and $\varphi(T_k)=0$. Now

$$\begin{aligned}\varphi(Z_1\ldots Z_m) &= \varphi\left(\prod_{k=1}^m U_{i(k)}T_kW_{i(k)}\right)\\ &= \varphi\left(\prod_{k=1}^m W_{i(k-1)}U_{i(k)}T_k\right)\end{aligned} \tag{19}$$

where $i(0) = i(m)$. This last step follows from the fact that $\varphi(AB) = \varphi(BA)$. Now by assumption $\varphi(W_{i(k-1)}U_{i(k)}) = 0$ for $k > 1$ and $\phi(T_k) = 0$ for $k \geq 1$. If $i(m) \neq i(1)$, then $\varphi(W_{i(0)}U_{i(1)}) = 0$ and by the freeness of $\{X\}$ and $\{U_1,W_1,\ldots,U_M,W_M\}$, it follows that (19) equals 0. If, on the other hand, $i(m)=i(1)$, then $W_{i(0)}U_{i(0)}=1$ and the same conclusion follows from freeness. Hence, we conclude that $U_1XW_1,\ldots,U_MXW_M$ are free. □

Repeated Sequences

We now analyze the spectrum of the matrix $\overline{S}\,\overline{S}^H$ where $\overline{S}$ is defined in Section IV. To simplify things we assume that $L = 1$.

Lemma 7: Assume that the entries of signature sequences are complex circular symmetric Gaussian random variables. Then irrespective of whether long or repeated random sequences are used, the expected empirical eigenvalue distribution of $\overline{S}\,\overline{S}^H$ converges to the same limit as $N,K\to\infty$ and $K/N\to\alpha$.

Proof: For the case of long sequences, it follows from the independence and circular symmetry of the signature sequence entries that the entries of $\overline{S}$ are i.i.d. random variables with variance $1/(\tau N)$. By existing random matrix results [47]–[49], it is known that the all the moments of $F_{\overline{S}\,\overline{S}^H}$, i.e., $\varphi_{\tau N}([\overline{S}\,\overline{S}^H]^r)$, converge as $N\to\infty$. Moreover, it is known that the limiting moments are those of a distribution with bounded support. Therefore, the convergence of the moments implies the convergence of the expected empirical eigenvalue distribution $F_{\overline{S}\,\overline{S}^H}$ to a limit F^*.

Let us now consider the case when repeated sequences are used, i.e., for all k, $s_k(1) = \cdots = s_k(\tau) = s_k$. In this case, there are dependencies in the entries of $\overline{S}$ and existing random matrix results cannot be used.

For all $r \geq 1$

$$\begin{aligned}\boldsymbol{E}[\mathrm{Tr}\,(\overline{S}\,\overline{S}^H)^r] &= \boldsymbol{E}[\mathrm{Tr}\,(\overline{S}^H\overline{S})^r] \\ &= \boldsymbol{E}\left[\left\{\frac{1}{\tau}\sum_{m=1}^{\tau}\overline{S}(m)^H\overline{S}(m)\right\}^r\right] \\ &= \boldsymbol{E}\left[\left\{\frac{1}{\tau}\sum_{m=1}^{\tau}B(m)^H S^H S B(m)\right\}^r\right]\end{aligned}$$

where $S := [s_1, \ldots, s_K]$ and

$$B(m) := \mathrm{diag}\,(b_1(m), \ldots, b_K(m)).$$

Thus the problem is equivalent to computing the limiting eigenvalue distribution of

$$\frac{1}{\tau}\sum_{m=1}^{\tau}B(m)^H S^H S B(m).$$

By Corollary 3, the two subsets $\{S^H S\}$ and $\{B(1), B(1)^H, \ldots, B(\tau), B(\tau)^H\}$ are asymptotically free, i.e., there exists $(\mathcal{A}, \varphi)$ and random variables $X, U_1, W_1, \ldots, U_\tau, W_\tau$ such that $\{X\}$ and $\{U_1, W_1, \ldots, U_\tau, W_\tau\}$ are free subsets and the joint distribution of $S^H S$, $B(1)$, $B(1)^H, \ldots, B(\tau)$, $B(\tau)^H$ converges to that of $X, U_1, W_1, \ldots, U_\tau, W_\tau$ as $N \to \infty$. Now since $\boldsymbol{E}[b_k(i)b_k^*(j)] = \delta_{ij}$ for all k, it follows that $\varphi_K(B(i)B(j)^H) = \delta_{ij}$, and also

$$\varphi(U_i W_j) = \varphi(W_j U_i) = \delta_{ij}.$$

Hence, from Lemma 6, the random variables $W_1 X U_1, \ldots, W_\tau X U_\tau$ are free. We can now conclude that $B(1)^H S^H S B(1), \ldots, B(\tau)^H S^H S B(\tau)$ are asymptotically free. But if $S(1), \ldots, S(\tau)$ are i.i.d. copies of S (corresponding to the long sequence case), then

$$B(1)^H S(1)^H S(1) B(1), \ldots, B(\tau)^H S(\tau)^H S(\tau) B(\tau)$$

are asymptotically free as well, and moreover each $B(m)^H S(m)^H S(m) B(m)$ has the same distribution as $B(m)^H S^H S B(m)$. Hence, by Proposition 3, the matrix

$$\frac{1}{\tau}\sum_{m=1}^{\tau}B(m)^H S^H S B(m)$$

has the same limiting distribution as

$$\frac{1}{\tau}\sum_{m=1}^{\tau}B(m)^H S(m)^H S(m) B(m).$$

It follows then that the limiting expected empirical eigenvalue distribution of $\overline{S}\,\overline{S}^H$ is the same regardless of whether long or repeated random sequences are used. □

Lemma 7 is a crucial step in explaining why repeated and long sequences make little difference asymptotically. Basically, although there is statistical dependency in the repeated sequence case, the randomness in the information symbols makes the relevant component matrices asymptotically free.

We can now give a Proof of Theorem 3, which concerns the performance of the channel estimator when repeated sequences are used.

Proof of Theorem 3: To simplify notation, we take $\overline{p} = 1$ and let $v := \sigma^2/\tau$. The general result follows from a simple rescaling.

Define

$$\begin{aligned}\overline{S}_i &:= [\overline{s}_1, \ldots, \overline{s}_{i-1}, \overline{s}_{i+1}, \ldots, \overline{s}_K] \\ \overline{S} &:= [\overline{s}_1, \ldots, \overline{s}_K]\end{aligned}$$

and

$$\beta_i^N := \overline{s}_i^H(\overline{S}_i\overline{S}_i^H + vI)^{-1}\overline{s}_i \tag{20}$$

which can be interpreted as the SIR achieved by user i under the MMSE receiver, when the spreading sequences are $\overline{s}_k$'s. In [50, Eq. (12)], a key equation relating the β_i^N's and the trace of $(\overline{S}\,\overline{S}^H + vI)^{-1}$ was derived

$$\frac{1}{\tau N}\sum_{i=1}^{K}\frac{\beta_i^N}{1+\beta_i^N} = 1 - \frac{v}{\tau N}\,\mathrm{Tr}\,(\overline{S}\,\overline{S}^H + vI)^{-1}.$$

Rearranging terms and then taking expectation with respect to the random sequences and information symbols, and observing that the β_i^N's are identically distributed, we get

$$\frac{K}{\tau N}\boldsymbol{E}\left[\frac{1}{1+\beta_1^N}\right] = \frac{K}{\tau N} - 1 + \boldsymbol{E}\left[\frac{v}{\tau N}\sum_{i=1}^{\tau N}\frac{1}{\lambda_i^N + v}\right] \tag{21}$$

where $\lambda_1^N, \ldots, \lambda_{\tau N}^N$ are the eigenvalues of $\overline{S}\,\overline{S}^H$. Let us investigate what happens as $N \to \infty$. Applying Lemma 7, we have

$$\begin{aligned}&\boldsymbol{E}\left[\frac{1}{\tau N}\sum_{i=1}^{\tau N}\frac{1}{\lambda_i^N + v}\right] \\ &\qquad = \int_0^\infty \frac{1}{\lambda + v}dF_{\overline{S}\,\overline{S}^H}(\lambda) \to \int_0^\infty \frac{1}{\lambda+v}dF^*(\lambda)\end{aligned} \tag{22}$$

as $N \to \infty$, where F^* is the limiting expected empirical eigenvalue distribution of $\overline{S}\,\overline{S}^H$. This convergence holds since $f(\lambda) = 1/(\lambda + v)$ is a bounded continuous function and $F_{\overline{S}\,\overline{S}^H}$ converges to F^* (in the weak topology). Since F^* is also the limiting distribution for the case with long sequences, it can be given as the solution β^* to the fixed point equation [47]–[49]

$$\beta^* = \left[v + \frac{\alpha}{\tau}\,\frac{1}{1+\beta^*}\right]^{-1}$$

or, equivalently,

$$\frac{\alpha}{\tau(1+\beta^*)} = \frac{\alpha}{\tau} - 1 + v\beta^*. \tag{23}$$

We observe that this β^* is exactly the same as the β_c in Theorem 2 (with $L = 1$) for the long sequence case.

From the facts that $\overline{s}_i$ is independent of $\overline{S}_i$ and that the entries of $\overline{s}_i$ are uncorrelated, zero-mean, variance $\frac{1}{\tau N}$, we get

$$\boldsymbol{E}[\beta_i^N] = \frac{1}{\tau N}\boldsymbol{E}[\mathrm{Tr}\,(\overline{S}_i\overline{S}_i^H + vI)^{-1}] \to \beta^* \tag{24}$$

as $N \to \infty$. Our goal now is to show from (21) that β_1^N in fact converges to β^* in probability as $N \to \infty$. To this end, let

$$\beta_1^N = \beta^* + \Delta_N.$$

Then, expanding about β^*

$$\begin{aligned}\frac{1}{1+\beta_1^N} &= \frac{1}{1+\beta^*+\Delta_N} \\ &= \frac{1}{1+\beta^*} - \frac{1}{(1+\beta^*)^2}\Delta_N + \frac{1}{(1+\xi_N)^3}\Delta_N^2\end{aligned}$$

for some ξ_N in between β_1^N and β^*. Substituting into (21) and taking limit as $N \to \infty$, and using (22) and (24), we get

$$\frac{\alpha}{\tau(1+\beta^*)} + \lim_{N\to\infty} \boldsymbol{E}\left[\frac{1}{(1+\xi_N)^3}\Delta_N^2\right] = \frac{\alpha}{\tau} - 1 + v\beta^*.$$

Comparing this to (23), we see that

$$\lim_{N\to\infty} \boldsymbol{E}\left[\frac{1}{(1+\xi_N)^3}\Delta_N^2\right] = 0.$$

Now

$$0 \le \beta_1^N \le \frac{1}{v}$$

so ξ_N falls in that bounded range as well. We thus conclude that

$$\boldsymbol{E}[(\Delta_N)^2] = \boldsymbol{E}[(\beta_1^N - \beta^*)^2] \to 0$$

as $N \to \infty$. Hence β_1^N converges to $\beta^* = \beta_c$ in probability, same as in the long sequence case. □

Shifted Sequences

We now turn to the Proof of Theorem 4 for the multipath case, focusing on the model defined in Section IV. We first analyze the spectrum of $\overline{S}\,\overline{S}^H$.

Lemma 8: Assume that the entries of the signature sequences are i.i.d. complex circular symmetric Gaussian random variables with variance $1/N$ and the sequences are independent across symbols and across users. Then the expected empirical eigenvalue distribution of $\overline{S}\,\overline{S}^H$ converge to the same limit as $N \to \infty$ irrespective of whether sequences along different multipaths of the same user are independent or cyclic shift of each other.

Proof:

$$\overline{S}\,\overline{S}^H = [\overline{s}_{11}, \dots, \overline{s}_{KL}]\,[\overline{s}_{11}, \dots, \overline{s}_{KL}]^H = \sum_{l=1}^{L} F_l F_l^H$$

where $F_l := [\overline{s}_{1l}, \dots, \overline{s}_{Kl}]$. For the model when the spreading sequences are independent along different paths, the matrices F_l are independent, and by Corollary 3, the random matrices $F_1F_1^H, \dots, F_LF_L^H$ are asymptotically free. For the model when the sequence $\overline{s}_{kl}$ is a cyclic shifted version of $\overline{s}_{k1}$ by $l-1$ chips, we can write

$$\overline{S}\,\overline{S}^H = \sum_{l=1}^{L} P_l F_1 F_1^H P_l^H$$

where P_l is the permutation matrix corresponding to a cyclic shift by $l-1$ chips. Observe that $P_lP_l^H = I$ and that for $l \ne 1 (\mathrm{mod}\, \tau N)$, $\mathrm{Tr}\, P_l = 0$, so for sufficiently large N

$$\mathrm{Tr}\,(P_iP_j^H) = \mathrm{Tr}\,(P_{i-j}) = \delta_{ij}.$$

It then follows from the asymptotic freeness of $\{F_1F_1^H\}$ and $\{P_1, P_1^H, \dots, P_L, P_L^H\}$ and Lemma 6 that

$$P_1F_1F_1^HP_1^H, \dots, P_LF_1F_1^HP_L^H$$

are asymptotic free. This together with the fact that each $P_lF_1F_1^HP_l^H$ has the same distribution as that of the corresponding matrix in the fully random model implies that the expected empirical eigenvalue distributions are asymptotically the same in both models. □

We can now give a Proof of Theorem 4.

Proof of Theorem 4: To simplify notation, we take $\overline{p}/L = 1$ and let $v := \sigma^2/\tau$. The general result follows from a simple rescaling.

Define $\overline{S}_{kl}$ as the τN by $KL-1$ matrix obtained from $\overline{S}$ by removing the column $\overline{s}_{kl}$ and let

$$\beta_{kl}^N := \overline{s}_{kl}^H(\overline{S}_{kl}\overline{S}_{kl}^H + vI)^{-1}\overline{s}_{kl}. \tag{25}$$

Consider again the identity

$$\frac{1}{\tau N}\sum_{i=1}^{KL}\frac{\beta_{kl}^N}{1+\beta_{kl}^N} = 1 - \frac{v}{\tau N}\,\mathrm{Tr}\,(\overline{S}\,\overline{S}^H + vI)^{-1}.$$

Rearranging terms and then taking expectation with respect to the random spreading sequences and information symbols, and observing that for each l, $\beta_{1l}^N, \dots, \beta_{Kl}^N$ have the same distribution, we have

$$\frac{K}{\tau N}\sum_{l=1}^{L}\boldsymbol{E}\left[\frac{1}{1+\beta_{1l}^N}\right] = \frac{KL}{\tau N} - 1 + \boldsymbol{E}\left[\frac{v}{\tau N}\,\mathrm{Tr}\,(\overline{S}\,\overline{S}^H + vI)^{-1}\right]. \tag{26}$$

Applying Lemma 8, we have

$$\boldsymbol{E}\left[\frac{v}{\tau N}\,\mathrm{Tr}\,(\overline{S}\,\overline{S}^H + vI)^{-1}\right] \to \int_0^\infty \frac{1}{\lambda + v}\,dF^*(\lambda) \tag{27}$$

as $N \to \infty$, where F^* is the limiting expected empirical eigenvalue distribution of $\overline{S}\,\overline{S}^H$. This limit is given as the solution β^* to the fixed-point equation [47]–[49]

$$\beta^* = \left[v + \frac{\alpha L}{\tau}\,\frac{1}{1+\beta^*}\right]^{-1}$$

or, equivalently,

$$\frac{\alpha L}{\tau(1+\beta^*)} = \frac{\alpha L}{\tau} - 1 + v\beta^*. \tag{28}$$

Expanding $1/(1+\beta_{1l}^N)$ about β^*

$$\frac{1}{1+\beta_{1l}^N} = \frac{1}{1+\beta^*} - \frac{1}{(1+\beta^*)^2}(\beta_{1l}^N - \beta^*) + \frac{1}{(1+\xi_l^N)^3}(\beta_{1l}^N - \beta^*)^2$$

for some ξ_l^N a function of β_{1l}^N and in between β_{1l}^N and β^*. Substituting into (21) and taking limit as $N \to \infty$, and using (27), we get

$$\frac{\alpha}{\tau(1+\beta^*)} + \lim_{N\to\infty}\sum_{l=1}^{L}\boldsymbol{E}\left[-\frac{1}{(1+\beta^*)^2}(\beta_{1l}^N - \beta^*) + \frac{1}{(1+\xi_l^N)^3}\,(\beta_{1l}^N - \beta^*)^2\right] = \frac{\alpha}{\tau} - 1 + v\beta^*.$$

Comparing this to (28), we see that

$$\lim_{N\to\infty}\sum_{l=1}^{L}\left\{\frac{1}{(1+\beta^*)^2}(\beta^* - \boldsymbol{E}[\beta_{1l}^N]) + \boldsymbol{E}\left[\frac{1}{(1+\xi_l^N)^3}(\beta_{1l}^N - \beta^*)^2\right]\right\} = 0. \tag{29}$$

By matrix inversion lemma, for $j \neq l$, we have

$$\begin{aligned}\beta_{1l}^N &= \bar{s}_{1l}^H(\overline{S}_{1l}\overline{S}_{1l}^H + vI)^{-1}\bar{s}_{1l} \\ &= \bar{s}_{1l}^H(\overline{S}_{1lj}\overline{S}_{1lj}^H + vI)^{-1}\bar{s}_{1l} \\ &\quad - \frac{|\bar{s}_{1l}^H(\overline{S}_{1lj}\overline{S}_{1lj}^H + vI)^{-1}\bar{s}_{1j}|^2}{1 + \bar{s}_{1j}^H(\overline{S}_{1lj}\overline{S}_{1lj}^H + vI)^{-1}\bar{s}_{1j}}\end{aligned}$$

where $\overline{S}_{1lj}$ is the matrix obtained by removing the column $\bar{s}_{1j}$ from $\overline{S}_{1l}$. We can then conclude that for all realization of $\overline{S}$

$$\beta_{1l}^N \leq \bar{s}_{1l}^H(\overline{S}_{1lj}\overline{S}_{1lj}^H + vI)^{-1}\bar{s}_{1l}.$$

Repeating this $L-2$ times, we get

$$\beta_{1l}^N \leq \bar{s}_{1l}^H(\overline{S}_1\overline{S}_1^H + vI)^{-1}\bar{s}_{1l}$$

where $\overline{S}_1$ is the matrix obtained by removing all the columns $\bar{s}_{11}, \ldots, \bar{s}_{1L}$ from $\overline{S}$. We observe that $\overline{S}_1$ is independent from $\bar{s}_{1l}$ and hence, as in (24)

$$\boldsymbol{E}[\bar{s}_{1l}^H(\overline{S}_1\overline{S}_1^H + vI)^{-1}\bar{s}_{1l}] = \frac{1}{\tau N}\boldsymbol{E}[\mathrm{Tr}\,(\overline{S}_i\overline{S}_i^H + vI)^{-1}] \to \beta^* \tag{30}$$

as $N \to \infty$. Hence

$$\limsup_{N\to\infty} \boldsymbol{E}[\beta_{1l}^N] \leq \beta^*$$

and combining this with (29), we can conclude that for every l $\lim_{N\to\infty} E(\beta_{1l}^N) = \beta^*$

$$\text{and} \quad \lim_{N\to\infty} \boldsymbol{E}\left[\frac{1}{(1+\xi_l^N)^3}(\beta_{1l}^N - \beta^*)^2\right] = 0.$$

Now

$$0 \leq \beta_{1l}^N \leq \frac{1}{v}$$

so ξ_l^N falls in that bounded range as well. We thus conclude that $\boldsymbol{E}[(\beta_{1l}^N - \beta^*)^2] \to 0$ as $N \to \infty$. Hence β_{1l}^N converges to $\beta^* = \beta_c$ in probability, same as for the fully random sequence model. □

ACKNOWLEDGMENT

The authors wish to thank Prof. D. Voiculescu of the University of California, Berkeley, for valuable discussions in free probability and for his help in formulating Lemma 6 in particular.

REFERENCES

[1] F. Adachi, M. Sawahashi, and H. Suda, "Wideband DS-CDMA for next-generation mobile communications systems," *IEEE Commun. Mag.*, vol. 36, pp. 56–69, Sept. 1998.
[2] E. Dahlman, B. Gudmundson, M. Nilsson, and J. Sköld, "UMTS/IMT-2000 based on wideband CDMA," *IEEE Commun. Mag.*, vol. 36, pp. 70–80, Sept. 1998.
[3] T. Ojanperä and R. Prasad, "An overview of air interface multiple access for IMT-2000/UMTS," *IEEE Commun. Mag.*, vol. 36, pp. 82–95, Sept. 1998.
[4] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[5] ——, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 85–96, Jan. 1986.
[6] ——, "Optimum multiuser asymptotic efficiency," *IEEE Trans. Commun.*, vol. COM-34, pp. 890–897, Sept. 1986.
[7] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 123–136, Jan. 1989.
[8] ——, "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Trans. Commun.*, vol. 38, pp. 496–508, Apr. 1990.
[9] Z. Xie, R. T. Short, and C. K. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 685–690, May 1990.
[10] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 42, pp. 3178–3188, Dec. 1994.
[11] P. Rapajic and B. Vucetic, "Adaptive receiver structures for asynchronus CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 685–697, May 1994.
[12] S. L. Miller, "An adaptive direct-sequence code-division multiple-access receiver for multiuser interference rejection," *IEEE Trans. Commun.*, vol. 43, pp. 1746–1755, Feb./Mar./Apr. 1995.
[13] Z. Zvonar and D. Brady, "Multiuser detection in single-path fading channels," *IEEE Trans. Commun.*, vol. 42, pp. 1729–1739, Feb./Mar./Apr. 1994.
[14] ——, "Optimum detection in asynchronous multiple-access multipath Rayleigh fading," in *Proc. Conf. Information Sciences and Systems*, Princeton Univ., Princeton, NJ, Mar. 1992, pp. 826–831.
[15] ——, "Suboptimal multiuser detector for frequency-selective Rayleigh fading synchronous CDMA channels," *IEEE Trans. Commun.*, vol. 43, pp. 154–157, Feb./Mar./Apr. 1995.
[16] M. K. Varanasi, "Parallel group detection for synchronous CDMA communication over frequency-selective Rayleigh fading channels," *IEEE Trans. Inform. Theory*, vol. 42, pp. 116–128, Jan. 1996.
[17] M. K. Varanasi and B. Aazhang, "Optimally near-far resistant multiuser detection in differentially coherent synchronous channels," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1006–1018, July 1991.
[18] H. C. Huang and S. Verdú, "Linear differentially coherent multiuser detection for multipath channels," *Wireless Personal Commun.*, vol. 6, pp. 113–136, 1998.
[19] M. K. Varanasi and S. Vasudevan, "Multiuser detectors for synchronous CDMA communication over nonselective Rician fading channels," *IEEE Trans. Commun.*, vol. 42, pp. 711–722, Feb./Mar./Apr. 1994.
[20] S. Vasudevan and M. K. Varanasi, "Achieving near-optimum asymptotic efficiency and fading resistancre over the time-varying Rayleigh-faded CDMA channel," *IEEE Trans. Commun.*, vol. 44, pp. 1130–1143, Sept. 1996.
[21] X. Wang and H. V. Poor, "Adaptive joint multiuser detection and channel estimation in multipath fading CDMA channels," *Wireless Networks*, vol. 4, no. 6, pp. 453–470, 1998.
[22] M. Stojanovic and Z. Zvonar, "Performance of linear multiuser detectors in time-varying multipath fading CDMA channels," in *Proc. GLOBECOM Communications Theory Mini-Conf.*, London, U.K., Nov. 1996, pp. 163–167.
[23] ——, "Performance of multiuser detection with adaptive channel estimation," in *Proc. Int. Conf. Communications*, Montreal, Que., Canada, June 1997, pp. 590–594.
[24] A. N. Barbosa and S. L. Miller, "Adaptive detection of DS/CDMA signals in fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 115–124, Jan. 1998.
[25] L. J. Zhu and U. Madhow, "Adaptive interference suppression for direct sequence CDMA over severely time-varying channels," in *Proc. Global Conf. Communications*, Phoenix, AZ, Nov. 1997, pp. 917–922.
[26] ——, "MMSE interference suppression for rapidly faded CDMA systems: A new cost function for differentially modulated signals," in *Proc. Int. Symp. Information Theory*, Cambridge, MA, Aug. 1998.
[27] T. J. Lim and S. Roy, "Adaptive filters in multiuser (MU) CDMA detection," *Wireless Networks*, vol. 4, pp. 307–318, 1998.
[28] M. L. Honig, S. L. Miller, M. J. Shensa, and L. B. Milstein, "Performance of adaptive linear interference suppression in the presence of dynamic fading," *IEEE Trans. Commun.*, to be published.
[29] Z. Zvonar and M. Stojanovic, "Performance of multiuser diversity reception in nonselective Rayleigh fading CDMA channels," in *Proc. GLOBECOM Communications Theory Mini-Conf.*, San Francisco, CA, Nov. 1994, pp. 171–175.
[30] D. N. C. Tse and S. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inform. Theory*, vol. 45, pp. 641–657, Mar. 1999.
[31] Kiran and D. N. C. Tse, "Effective bandwidths and effective interference for linear multiuser receivers in asynchronous systems," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1426–1447, July 2000.
[32] A. J. Grant and P. D. Alexander, "Random sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2832–2836, Nov. 1998.

[33] R. R. Müller, P. Scramm, and J. B. Huber, "Spectral efficiency of CDMA systems with linear interference suppression," in *Proc. Workshop Kommunikationstechnik*, Ulm, Germany, Jan. 1997, pp. 93–97.

[34] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 622–640, Mar. 1999.

[35] P. B. Rapajic, M. L. Honig, and G. Woodward, "Multiuser decision-feedback detection: Performance bounds and adaptive algorithms," in *Proc. Int. Symp. Information Theory*, Cambridge, MA, Aug. 1998, p. 34.

[36] S. L. Miller, M. L. Honig, and L. B. Milstein, "Performance analysis of MMSE receivers for DS-CDMA in frequency selective fading channels," IEEE Trans. Commun., to be published.

[37] M. Latva-aho, M. Juntti, and K. Kansanen, "LMMSE receivers for DS-CDMA systems in frequency-selective fading channels," in *Proc. FRAMES Workshop*, Delft, The Netherlands, Jan. 1999, pp. 196–204.

[38] E. Telatar and D. N. C. Tse, "Capacity and mutual information of wide-band multipath fading channels," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1384–1400, July 2000.

[39] Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *Ann. Probab.*, vol. 26, no. 1, pp. 316–345, 1998.

[40] P. Billingsley, *Probability and Measure*, 3rd ed. New York: Wiley, 1995.

[41] J. W. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *J. Multivariate Anal.*, vol. 55, no. 2, pp. 331–339, 1995.

[42] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*. Oxford, U.K.: Clarendon, 1965.

[43] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free Random Variables*, ser. CRM Monograph. Providence, RI: Amer. Math. Soc., 1992, vol. 1.

[44] D. Voiculescu, Lectures on Free Probability Theory, in unpublished Notes for a Course at the Saint-Flour Summer School on Proability, 1998.

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

[46] ——, "A strengthened asymptotic freeness result for random matrices with applications to free entropy," *Int. Math. Res. Notices*, vol. 1, pp. 41–63, 1998.

[47] U. Grenander and J. W. Silverstein, "Spectral analysis of networks with random topologies," *SIAM J. Appl. Math.*, vol. 32, pp. 499–519, Mar. 1977.

[48] K. W. Wachter, "The strong limits of random matrix spectra for sample matrices of independent elements," *Ann. Probab.*, vol. 6, no. 1, pp. 1–18, 1978.

[49] D. Jonsson, "Some limit theorems for the eigenvalues of a sample covariance matrix," *J. Multivariate Anal.*, vol. 12, pp. 1–38, 1982.

[50] P. Viswanath, V. Anantharam, and D. N. C. Tse, "Optimal sequences, power control and capacity of synchronous CDMA systems with linear MMSE multiuser receivers," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1968–1983, Sept. 1999.