Linear Multiuser Receivers in Random Environments

David N. C. Tse, *Member, IEEE*, and Ofer Zeitouni, *Fellow, IEEE*

***Abstract*—We study the signal-to-interference (SIR) performance of linear multiuser receivers in *random environments*, where signals from the users arrive in "random directions." Such random environment may arise in a DS-CDMA system with random signature sequences, or in a system with antenna diversity where the randomness is due to channel fading. Assuming that such random directions can be tracked by the receiver, the resulting SIR performance is a function of the directions and therefore also random. We study the asymptotic distribution of this random performance in the regime where both the number of users K and the number of degrees of freedom N in the system are large, but keeping their ratio fixed. Our results show that for both the decorrelator and the minimum mean-square error (MMSE) receiver, the variance of the SIR distribution decreases like $1/N$, and the SIR distribution is asymptotically Gaussian. We compute closed-form expressions for the asymptotic means and variances for both receivers. Simulation results are presented to verify the accuracy of the asymptotic results for finite-sized systems.**

***Index Terms*—Decorrelator, MMSE receiver, multiple antennas, multiuser detection, random matrices, random signature sequences.**

I. Introduction

IN A direct-sequence code-division multiple access (DS-CDMA) system, each user micapodulates the information symbols onto its unique signature (or spreading) sequence. This spreading of information provides additional degrees of freedom for communication. To fully exploit the available degrees of freedom, linear *multiuser receivers* have been proposed to reduce or suppress the interference from other users. Prominent among these receivers are the decorrelator [9], [10] and the minimum mean-square error (MMSE) receiver [27], [11], [15], [16].

A common performance measure for these linear receivers is the output signal-to-interference ratio (SIR). Clearly, the performance of these linear receivers depends on the signature sequences of the users. We focus on the common situation when the signature sequences of the users are *randomly* and *independently* chosen. This model is relevant in several scenarios: Users may employ pseudorandom spreading sequences, or the transmitted signals from the users are distorted by independent multipath fading channels which randomize the received signature sequences. While the sequences are random, we assume in this paper that they are known perfectly at the receiver. In practice, this knowledge is obtained through adaptation, channel measurements, or an initialization protocol. However, since the SIR performance of the users is a function of the signature sequences, it is also random. We are interested in characterizing their distributions.

The random sequence DS-CDMA model is an example of a system where multiuser receivers operate in a *random environment*. Another common example is a system with multiple antennas at the receiver. If the fading from a user to each of the receive antennas is independent, then diversity is achieved using multiple antennas to combat against the possibility of deep fade at any single antenna. Moreover, by tracking the fading of different users, linear multiuser receivers can be employed to suppress interference from other users while demodulating one particular user. The SIR performance is again random, being a function of the channel fading. This is very similar to the random signature sequence scenario since in both cases, signals from different users arrive from "random directions," where the directions are given by the signature sequences in the DS-CDMA system and by the fading patterns at the different antennas in the multiantenna system. Indeed, one can think of the random sequence model considered in this paper as a canonical model for a multiuser system with diversity.

In this paper, we analyze the performance of both the decorrelator and the MMSE receiver in a random environment. The MMSE receiver is particularly interesting as it maximizes the SIR among all linear receivers. While it is known [11] that both receivers have the same near–far resistance (ability to reject worst case interference), the MMSE receiver, by its definition, performs strictly better in terms of SIR when the powers of the interferers are controlled or when they are relatively weak (such as from out of cell). The performance of the decorrelator, on the other hand, does not depend on the interferers' powers. The SIR performance of the MMSE receiver under power control will be studied.

In recent independent studies [22], [25], it was shown that in a random environment, the SIR of a user under both the MMSE receiver and the decorrelator converges to a deterministic limit in a large system. The scaling is by letting the processing gain and the number of interferers go to infinity, keeping the number of interferers per unit processing gain fixed. In a finite system, however, the attained SIR will fluctuate around this limit. Such fluctuations determine important performance measures such as average probability of error and outage probability, i.e., the probability that the SIR of a user drops below

Manuscript received October 1, 1998; revised March 16, 1999. The work of D. N. C. Tse 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 work of O. Zeitouni was supported in part by the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, under Grant IRI 9310670.

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

O. Zeitouni is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: zeitouni@ee.technion.ac.il).

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

Publisher Item Identifier S 0018-9448(00)00072-9.

a certain threshold. The main goal of this paper is to characterize such fluctuations in various scenarios. We provide central limit theorems which show that under appropriate scaling, the fluctuations are asymptotically Gaussian. Moreover, we give closed-form formulas for the variances of the fluctuations in terms of system parameters. Our results are obtained using techniques from random matrix theory. While the analysis of the deterministic limit involves only the asymptotic *eigenvalue* distribution of certain random matrices, the characterization of the SIR fluctuation requires understanding the asymptotic distribution of the *eigenvectors* as well as the fluctuation of the eigenvalue distribution around the asymptotic limit. Both of these are current research topics in random matrix theory and indeed our proofs exploit several recent results.

In related work, Honig and Veerakachen [7] have studied the problem of performance variability of linear multiuser detection under random signature sequences. They derived a heuristic approximation of the SIR performance of the decorrelator, and provided simulation results for the MMSE receiver. In contrast, our analytical results are justified by limit theorems and they apply both to the MMSE receiver and to the decorrelator. In the context of systems with antenna diversity, Winters *et al.* [26] have obtained related results on the performance of the decorrelator under flat Rayleigh fading. In this paper, our results apply to general fading distributions, not necessarily Rayleigh, which are of particular interest for distributed antenna systems, where the antennas can be placed at different locations of a room or a floor. In this scenario, the fading experienced consists of both small-scale (multipath) and large-scale effects, and cannot be accurately modeled as Rayleigh distributed. It turns out that relaxing the Rayleigh assumption complicates the analysis considerably.

Most of our results make only very weak assumptions on the distribution of the randomness and are therefore transparent to the specific random environment. For concreteness, we will focus on the DS-CDMA system with random signature sequences throughout most of the paper. In Section II, we introduce the model. We analyze the performance of the decorrelator and the MMSE receiver in Sections III and IV, respectively, with our main results being Theorems 3.3 and 4.5. Section V contains simulations validating the accuracy of our asymptotic results. In Section VI, we briefly comment on the application of our results to systems with antenna diversity. Section VII contains our conclusions. The proofs of the results are presented in the Appendices.

During the final stage of the preparation of this paper, we were informed of independent work by Müller *et al.* [14] on the performance of the decorrelator. The relationship between their results and ours will be discussed in Section III. We were also informed of independent work by Kim and Honig [8] who have presented a heuristic approximation for the variance of the SIR under the MMSE.

II. Linear Receivers for DS-CDMA Systems

In a DS-CDMA system, each of the user's information or coded symbols is spread onto a much larger bandwidth via modulation by its own *signature* or *spreading sequence*. The following is a sampled discrete-time model for a symbol-synchronous DS-CDMA system:

$$\boldsymbol{y} = \sum_{i=1}^{K} b_i \boldsymbol{s}_i + \boldsymbol{z} \tag{1}$$

where $b_i \in \Re$ and $\boldsymbol{s}_i \in \Re^N$ are the transmitted symbol and signature sequence of user i, respectively, and $\boldsymbol{z}$ is $N(0, \sigma^2 I)$ background Gaussian noise. The length of the signature sequences is N, which is the number of degrees of freedom, and K is the number of users. The received vector is $\boldsymbol{y} \in \Re^N$. We assume the b_i's are independent and that $E[b_i] = 0$ and $E[b_i^2] = P_i$, where P_i is the received power of user i (energy per symbol).

We view multiuser receivers as *demodulators*, extracting good estimates of the (coded) symbols of each user as soft decisions to be used by the channel decoder [16]. From this point of view, the relevant performance measure is the signal-to-interference ratio (SIR) of the estimates. We shall now focus without loss of generality on the demodulation of user 1, assuming that the receiver has already acquired the knowledge of the spreading sequences. For user 1, the MMSE receiver $\boldsymbol{c}_1$ generates a soft decision $\hat{b}_1 \equiv \boldsymbol{c}_1^t \boldsymbol{y}$ which maximizes the output signal-to-interference ratio (SIR)

$$\frac{(\boldsymbol{c}_1^t \boldsymbol{s}_1)^2 P_1}{(\boldsymbol{c}_1^t \boldsymbol{c}_1)\sigma^2 + \sum_{i=2}^{K} (\boldsymbol{c}_1^t \boldsymbol{s}_i)^2 P_i}$$

(see [11], [15], and [16]).

The formulas for the MMSE demodulator and its performance are well known [11]

$$\hat{b}_{\text{mmse}}(\boldsymbol{y}) = \frac{P_1}{1 + P_1 \boldsymbol{s}_1^t (S_1 T S_1^t + \sigma^2 I)^{-1} \boldsymbol{s}_1} \boldsymbol{s}_1^t (S_1 T S_1^t + \sigma^2 I)^{-1} \boldsymbol{y} \tag{2}$$

and the signal-to-interference ratio β for user 1 is

$$\beta = P_1 \boldsymbol{s}_1^t (S_1 T S_1^t + \sigma^2 I)^{-1} \boldsymbol{s}_1 \tag{3}$$

where $S_1 := [\boldsymbol{s}_2, \cdots, \boldsymbol{s}_K]$ and $T := \operatorname{diag}(P_2, \cdots, P_K)$.

We observe that the MMSE receiver depends on the received powers of the interferers. The decorrelator is a simpler but suboptimal linear receiver that operates without the need of knowing the received powers of the interferers. It simply nulls out the interference from other users by projecting the received signal onto the subspace orthogonal to the span of their signature sequences. The vector of symbol estimates $\hat{\boldsymbol{b}}_{\text{dec}}$ generated by the decorrelator for all users is given by

$$\hat{\boldsymbol{b}}_{\text{dec}} := (S^t S)^{-1} S^t \boldsymbol{y}$$

where $S := [\boldsymbol{s}_1, \cdots, \boldsymbol{s}_K]$. Here, the inverse is replaced by the pseudo-inverse if $S^t S$ is not invertible. Observe that if there were no noise, the estimates would be exactly the original symbols, and hence it would be the multiuser analog of the zero-forcing equalizer. Assuming that $S^t S$ is invertible, the SIR γ of user 1 under the decorrelator is given by

$$\gamma = \frac{P_1}{\sigma^2 [(S^t S)^{-1}]_{11}} \tag{4}$$

where the denominator is the $(1, 1)$ entry of the matrix $(S^t S)^{-1}$. Note that the performance of the decorrelator does not depend on the powers of the interferers.

The formulas shown above for the SIR performance of various receivers can be numerically calculated given specific choices of the signature sequences. In this paper, however, we focus on the scenario when the sequences are *randomly* and independently chosen. In this case, the SIR performance of a receiver is a random variable, since it is a function of the spreading sequences, and we are interested in analyzing its statistics. We will assume that though the sequences are randomly chosen, they are known to the receiver once they are picked. In practice, this means that the change in the spreading sequences is at a much slower time scale than the symbol rate so that the receiver has the time to acquire the sequences. (There are known adaptive algorithms for which this can even be done blindly; see [6].) However, the *performance* of linear receivers depends on the initial choice of the sequences and hence is random.

The model for random sequences: let

$$\boldsymbol{s}_j = (1/\sqrt{N})(v_{1j}, \cdots, v_{Nj})^t, \qquad j = 1, \cdots, K.$$

The random variables v_{ij}'s are independent and identically distributed (i.i.d.), having zero mean, variance 1, and a distribution symmetric about 0. The normalization by $1/\sqrt{N}$ ensures that $E[||\boldsymbol{s}_j||^2] = 1$, i.e., we maintain a constant average power. In practice, it is quite common for the entries of the spreading sequences to be 1 or -1, but our results hold for general distributions, which are useful when we look at other random environments such as systems with antenna diversity. We will also make the technical assumption that $E[v_{ij}^8] < \infty$. This last assumption can be relaxed (it is probably enough to assume the fourth moment is finite), but we chose this slightly stronger assumption in order to simplify the proofs.

III. Performance of the Decorrelator

We shall begin by studying the performance of the decorrelator, before proceeding to the MMSE receiver, which requires a more sophisticated analysis. The following result shows that in a system with large processing gain and many users, the random SIR of a user converges to a deterministic limit. It is proved independently in [22] and [25].

Theorem 3.1: Let $\gamma^{(N)}$ be the (random) SIR of the decorrelating receiver for user 1 when the spreading length is N and the number of users $K = \lfloor \alpha N \rfloor$, where $\alpha > 0$ is a fixed constant. Then $\gamma^{(N)}$ converges to γ^* in probability as $N \to \infty$, where γ^* is given by

$$\gamma^* = \begin{cases} \dfrac{P_1(1-\alpha)}{\sigma^2}, & \alpha < 1 \\ 0, & \alpha \geq 1. \end{cases}$$

In the scaling considered, the number of users per degree of freedom (or, equivalently, per unit bandwidth) α is fixed while the number of degrees of freedom grows. This scaling makes sense as more users can be supported by a larger bandwidth. The parameter α can be thought of as the *system load*. Observe also that the above result holds regardless of the powers of the interferers, as the decorrelator nulls out all interferers, its performance does not depend on the interferers' powers. Intuitively, this result says that for random signature sequences, the loss in SIR due to interference from other users is proportional to the number of interferers per degree of freedom.

Theorem 3.1 can be viewed as a law of large numbers. Though it gives the asymptotic limit, this result does not provide any information about the fluctuation around the limit for finite-sized system. This is the main consideration in this section. It is of interest to consider only the case when the number of users is less than the number of degrees of freedom, i.e., $\alpha < 1$, because otherwise the limiting SIR is zero. Moreover, since the performance of the decorrelator does not depend on the powers of the interferers, we can just focus on the case when the interferers have equal received power P, i.e., $T = PI$.

The first step is to obtain a formula for the SIR performance under the decorrelator, equivalent to but more useful for analysis than (4). It is known [11] that for the same signature sequences, the asymptotic efficiency of the decorrelator and MMSE receivers are identical, i.e.,

$$\lim_{\sigma^2 \to 0} \gamma^{(N)} \sigma^2 = \lim_{\sigma^2 \to 0} \beta^{(N)} \sigma^2$$

where $\beta^{(N)}$ is the SIR under the MMSE receiver in a system with processing gain N. Using (3) and (4), we therefore get

$$\frac{1}{[(S^t S)^{-1}]_{11}} = \lim_{\sigma^2 \to 0} \sigma^2 \boldsymbol{s}_1^t (P S_1 S_1^t + \sigma^2 I)^{-1} \boldsymbol{s}_1.$$

Let $P S_1 S_1^t = O F O^t$ be the spectral decomposition of $P S_1 S_1^t$, where $F = \operatorname{diag}(\lambda_1, \cdots, \lambda_N)$ is a diagonal matrix with decreasing eigenvalues and O is an orthogonal matrix of the eigenvectors of $S_1 S_1^t$. Putting this in the above expression and evaluating the limit, we get

$$\frac{1}{[(S^t S)^{-1}]_{11}} = \boldsymbol{s}_1^t O D O^t \boldsymbol{s}_1 \tag{5}$$

where $D = \operatorname{diag}(0, \cdots, 0, 1, \cdots, 1)$ and the number of 1's in the diagonal of D is the number of zero eigenvalues of $S_1 S_1^t$.

To provide some background for our analysis of the random SIR performance for finite-sized systems, it is helpful to see first how Theorem 3.1 can be derived from the representation (5). The essence is based on the following lemma, proved in [13].

Lemma 3.2: Let $\boldsymbol{s} = (1/\sqrt{N})(v_1, \cdots, v_N)^t$ where v_i's are i.i.d. zero mean, unit variance random variable with finite fourth moment. Let A be a deterministic $N \times N$ symmetric positive-definite matrix. Then

$$E[\boldsymbol{s}^t A \boldsymbol{s}] = \frac{1}{N} \operatorname{tr} A$$

and

$$\operatorname{Var}\left[\boldsymbol{s}^t A \boldsymbol{s}\right] \leq \frac{1}{N} C_1 [\lambda_{\max}(A)]^2$$

for some constant C_1 which depends only on the fourth moment of v_1.

This lemma holds for any deterministic matrix A. Applying this Lemma by conditioning on $A = ODO^t$ and observing that A and $\boldsymbol{s}_1$ are independent, we obtain that

$$E[\boldsymbol{s}_1^t ODO^t \boldsymbol{s}_1] = \frac{1}{N} E[\operatorname{tr} D].$$

(Note that $E[\operatorname{tr} D]$ is just the average number of nonzero eigenvalues of $S_1 S_1^t$.) Also, $\lambda_{\max}(ODO^t) \le 1$ and an application of Chebychev's inequality yields

$$\boldsymbol{s}_1^t ODO^t \boldsymbol{s}_1 - \frac{1}{N} \operatorname{tr} D \xrightarrow{\mathcal{P}} 0. \tag{6}$$

(Here and in the sequel, the notation $\xrightarrow{\mathcal{P}}$ denotes convergence in probability, while the notation $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, or more generally weak convergence of probability measures.) Furthermore, Bai and Yin [1] showed that the smallest eigenvalue of the random matrix $S_1^t S_1$ converges almost surely to a positive number, when $\alpha < 1$. This implies that almost surely for large N, the signature sequences of the other users are linearly independent and the number of 1's in D is $N - K + 1$. This together with (5) and (6) immediately yields Theorem 3.1.

Geometrically, $\|\boldsymbol{s}_1 ODO^t \boldsymbol{s}_1\|^2$ can be interpreted as the amount of energy of $\boldsymbol{s}_1$ in V, and the above result says that in a large system, this amount is approximately proportional to the dimension of that space. This is what one would expect from the i.i.d. nature of the components of $\boldsymbol{s}_1$.

Observe that the above derivation of the asymptotic limit makes use of the convergence of $\operatorname{tr} D$ (i.e., the dimension of the subspace V) but not any properties of O, the eigenvectors of $S_1 S_1^t$. In fact, it depends only on the randomness of $\boldsymbol{s}_1$. However, when we are interested in characterizing the *fluctuations* of the SIR around the asymptotic limit, asymptotic properties of the eigenvectors are needed. The mathematical apparatus to deal with this is established in Appendix A. The solution depends on $\mu_D(\cdot)$, the asymptotic empirical distribution of the eigenvalues of ODO^t; this is given by

$$\mu_D(x) = \alpha\delta(x) + (1-\alpha)\delta(x-1).$$

Applying Corollary A.2 to this problem, we can then conclude that

$$\sqrt{N}\left[\boldsymbol{s}_1^t ODO^t \boldsymbol{s}_1 - \frac{1}{N}\operatorname{tr} D\right] \xrightarrow{\mathcal{D}} N(0, a)$$

where

$$\begin{aligned} a &= 2\int x^2 \mu_D(dx) + (E(v_{11}^4) - 3)\left(\int x\mu_D(dx)\right)^2 \\ &= 2(1-\alpha) + (E(v_{11}^4) - 3)(1-\alpha)^2. \end{aligned}$$

This together with the fact that $\operatorname{tr} D$ converges almost surely to $N - K + 1$ yields the following theorem.

Theorem 3.3: When the system load $\alpha < 1$, as $N \to \infty$

$$\sqrt{N}\left(\gamma^{(N)} - \frac{P_1}{\sigma^2}(1-\alpha)\right) q \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \left(\frac{P_1}{\sigma^2}\right)^2 a\right)$$

where

$$a = 2(1-\alpha) + (E[v_{11}^4] - 3)(1-\alpha)^2.$$

This theorem says that the fluctuation of the SIR around the limit is approximately Gaussian with variance $\frac{1}{N}(\frac{P_1}{\sigma^2})^2 a$, decreasing like $1/N$ and with a depending only on the system load α and the fourth moment of v_{11}. Observe also the variance increases with $E[v_{11}^4]$, and hence is minimized when the entries take on $+1$ or -1 values only. It should be noted that while the asymptotic limit depends only on the second moment of v_{ij}'s, the amount of fluctuation around the limit depends on the fourth moment, and thus varies from one distribution to another.

Since the truth of Theorem 3.3 depends entirely on the machinery developed in Appendix A and the proofs there are rather technical, we would like to give some intuition as to why it holds. Define

$$\boldsymbol{u} = \frac{1}{\sqrt{N}}(u_1, \cdots, u_N)^t := O^t \boldsymbol{s}_1.$$

Assuming that the signature sequences of the interfering users are linearly independent (which holds with probability 1 or close to 1 in a large system), we have

$$\gamma^{(N)} = \frac{P_1}{\sigma^2}\frac{1}{N}\sum_{i=K}^{N} u_i^2.$$

First consider the special case when the entries v_{1j} of the spreading sequences of user 1 are Gaussian. Then the u_i's are i.i.d. Gaussian $N(0, 1)$. In this case, $\gamma^{(N)}$ is Chi-square distributed. This is basically the main result of Winters *et al.* [26], except that they considered complex Gaussian v_{ij} for their Rayleigh fading model. For large N, a direct application of the central limit theorem yields Theorem 3.3, with $E[v_{11}^4] = 3$.

We observe that in the special case of Gaussian v_{ij}, the central limit approximation is actually not necessary as the explicit distribution of γ^N can be obtained for finite N. Moreover, the properties of the eigenvectors O play no role here, other than the fact that O is independent of $\boldsymbol{s}_1$. The key reason is that an i.i.d. Gaussian random vector is *isotropic*, i.e., its distribution is invariant to orthogonal transformations, so that whatever a deterministic O is, $O^t\boldsymbol{s}_1$ has the same distribution as $\boldsymbol{s}_1$. In particular, this means that $\boldsymbol{s}_1/\|\boldsymbol{s}_1\|$ is uniformly distributed on the $(N-1)$-sphere of radius 1.

Let us now consider the general case where the v_{ij}'s are not necessarily Gaussian so that the random vector $\boldsymbol{s}_1$ may be not isotropic. In this case, $O^t\boldsymbol{s}_1$ has a complicated distribution dependent on both the distribution of O and $\boldsymbol{s}_1$, and need not be isotropic. To analyze this problem, we need to exploit a special property of the eigenvectors of $S_1 S_1^t$. In particular, we show that even though $\boldsymbol{s}_1$ may not be isotropic, as $N \to \infty$, the random vector $\boldsymbol{v} := O^t(\boldsymbol{s}_1/\|\boldsymbol{s}_1\|)$ will be asymptotically uniformly distributed on the unit sphere and, moreover, independent of $\|\boldsymbol{s}_1\|$. (This fact is made precise in Theorem A.1 of Appendix A.) In essence, we show there that there is enough randomness in O to make $O^t(\boldsymbol{s}_1/\|\boldsymbol{s}_1\|)$ close to being uniformly distributed.

Any isotropic random vector with fixed norm 1 can be generated by an i.i.d. Gaussian random vector normalized to be on the unit sphere. It then follows that

$$\boldsymbol{v} \approx \frac{1}{||\boldsymbol{r}||}(r_1, \cdots, r_N)^t \tag{7}$$

where the r_i's are i.i.d. Gaussian $N(0, 1)$ and independent of $||\boldsymbol{s}_1||$, and $\approx$ means that the distributions of the random variables in both sides of (7) are "close" in a sense which will be made precise later. Thus

$$O^t\boldsymbol{s}_1 \approx \frac{||\boldsymbol{s}_1||}{||\boldsymbol{r}||}(r_1, \cdots, r_N)^t$$

and

$$\gamma^{(N)} \approx \frac{P_1}{\sigma^2}\frac{||\boldsymbol{s}_1||^2}{||\boldsymbol{r}||^2}\sum_{i=K}^{N} r_i^2. \tag{8}$$

Using the central limit theorem, it can be seen that

$$N||\boldsymbol{s}_1||^2 \approx 1 + \frac{1}{\sqrt{N}}\phi_1 \tag{9}$$

$$||\boldsymbol{r}||^2 \approx 1 + \frac{1}{\sqrt{N}}\phi_2 \tag{10}$$

$$\frac{1}{N}\sum_{i=K}^{N} r_i^2 \approx (1-\alpha) + \frac{1}{\sqrt{N}}\phi_3 \tag{11}$$

where the ϕ_i's are zero-mean jointly Gaussian and ϕ_1 independent of ϕ_2 and ϕ_3. The second moments of these random variables can be calculated as

$$E[\phi_1^2] = E[v_{11}^4] - 1 \qquad E[\phi_2^2] = 2 \qquad E[\phi_3^2] = 2(1-\alpha)^2$$

and

$$E[\phi_2\phi_3] = 2(1-\alpha).$$

Using (9)–(11), we can perform a Taylor-series expansion of (8), keeping the first- and second-order terms only, and obtain

$$\gamma^{(N)} \approx \frac{P_1}{\sigma^2}\left[1 - \alpha + \frac{1}{\sqrt{N}}((1-\alpha)\phi_1 - (1-\alpha)\phi_2 + \phi_3)\right].$$

Direct computation reveals that the variance of the Gaussian fluctuation $(1-\alpha)(\phi_1 - \phi_2) + \phi_3$ is precisely the value a given in Theorem 3.3.

The essence of the above argument is based on the fact that the eigenvector matrix O of $S_1S_1^t$ itself is in some sense asymptotically *isotropic*. A version of this phenomenon has been proved by Silverstein [18]: He showed that given any deterministic vector $\boldsymbol{s}_1$ whose entries are either $+1/\sqrt{N}$ or $-1/\sqrt{N}$, the random vector $O^t\boldsymbol{s}_1$ is asymptotically isotropic, to the accuracy of the central-limit approximation. We show that this is true also when $\boldsymbol{s}_1$ is a random vector with i.i.d. elements of general distribution, but independent of O. This fact is made precise in Theorem A.1 in Appendix A, using the theory of weak convergence.

An interesting observation from the above heuristic derivation is that the asymptotic distribution of the SIR under the decorrelator depends on the distribution of v_{ij}'s only through that of $||\boldsymbol{s}_1||^2$, i.e., the fluctuation of the received energy of the signal from user 1. In the special case when the signature sequence entries takes on $+1/\sqrt{N}$ or $-1/\sqrt{N}$, $||\boldsymbol{s}_1||^2 = 1$, and (8) simplifies to

$$\gamma^{(N)} \approx \frac{P_1}{\sigma^2}\frac{1}{||\boldsymbol{r}||^2}\sum_{i=K}^{N} r_i^2.$$

A similar approximation was proposed independently in [14]. However, the assumption of $O^t\boldsymbol{s}_1$ being asymptotically isotropic was made without justification. As was pointed out, this matter is rather subtle as the property depends both on the distributions of O and $\boldsymbol{s}_1$.

IV. Performance of MMSE Receiver

We now turn to analyzing the performance of the MMSE receiver. In [22], it is shown that in a large system, the SIR under the MMSE receiver converges to a deterministic limit. While the results there apply to the general setting of users with unequal received powers, we focus here on the case when the users are controlled to equal received power. This would be the case when users are all in a single cell and have the same SIR requirements. In this case, the limiting SIR has a simple closed-form expression, which is also obtained independently in [25].

Theorem 4.1 [22]: Let $\beta^{(N)}$ be the (random) SIR of the MMSE receiver for user 1 when the spreading length is N. Suppose the received powers of the users are all equal to P. Then β^N converges to β^* in probability as $N \to \infty$, where β^* is given by

$$\beta^* = \frac{(1-\alpha)P}{2\sigma^2} - \frac{1}{2} + \sqrt{\frac{(1-\alpha)^2P^2}{4\sigma^4} + \frac{(1+\alpha)P}{2\sigma^2} + \frac{1}{4}}. \tag{12}$$

To provide some background in understanding our approach to analyzing the random performance in a finite-sized system, it helps to first give the basic intuition behind the proof of Theorem 4.1. Recall from (3) that the SIR of user 1 under the MMSE receiver is given by

$$\beta = P\boldsymbol{s}_1^t(S_1TS_1^t + \sigma^2 I)^{-1}\boldsymbol{s}_1$$

where $S_1 = [\boldsymbol{s}_2, \cdots, \boldsymbol{s}_K]$ and $T = PI$. In terms of the spectral decomposition $PSS^t = OFO^t$ introduced in the previous section, where $F = \operatorname{diag}(\lambda_1, \cdots, \lambda_N)$, we have

$$\beta = P\boldsymbol{s}_1^t O(F + \sigma^2 I)^{-1}O^t\boldsymbol{s}_1.$$

Comparing this to the performance of the decorrelator

$$\gamma = \frac{P}{\sigma^2}\boldsymbol{s}_1^t ODO^t\boldsymbol{s}_1$$

where $D = \operatorname{diag}(0, \cdots, 0, 1, \cdots, 1)$, we see that the expression for the MMSE receiver is more complicated as it depends on the random eigenvalues of $PS_1S_1^t$ as well. This reflects the fact that the MMSE receiver attains a better performance by taking into account the strength of the interferers rather than just nulling them out.

Nevertheless, Theorem 4.1 can be proved by, first, using Lemma 3.2 to show that for large N

$$\beta^{(N)} \approx \frac{P}{N}\,\mathrm{tr}(F+\sigma^2 I)^{-1}.$$

Second, using results from random matrix theory [13], [17], it can further be deduced that the empirical distribution of the eigenvalues of $S_1 T S_1^t$ converges to some limiting distribution G^*. Combining these facts, we obtain that $\beta^{(N)}$ converges in probability to

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

In [22] it is further shown how this limit can be explicitly computed to be (12).

Following this train of thought, the random fluctuation of $\beta^{(N)}$ around the limit β^* can be dealt with by decomposing into three terms

$$\beta^{(N)} - \frac{P}{N}\,\mathrm{tr}\,(F+\sigma^2 I)^{-1} \tag{13}$$

and

$$\frac{P}{N}\,\mathrm{tr}\,(F+\sigma^2 I)^{-1} - E[\beta^{(N)}] \tag{14}$$

and

$$E[\beta^{(N)}] - \beta^*. \tag{15}$$

Note that the first term depends on $\boldsymbol{s}_1$ and the eigenvector matrix O, while the second and third terms depend only on the fluctuation of the empirical eigenvalue distribution of $S_1 T S_1$ around the limiting distribution G^*. Just as for the decorrelator, the first fluctuation can be characterized using the theory developed in Appendix A. Applying Corollary A.2 there, we obtain:

Lemma 4.2:

$$\sqrt{N}\left(\beta^{(N)} - \frac{P}{N}\,\mathrm{tr}(F+\sigma^2 I)^{-1}\right) \xrightarrow{\mathcal{D}} N(0,\, b)$$

where

$$b = 2\int \left[\frac{P}{(\lambda+\sigma^2)}\right]^2 dG^*(\lambda) + (E(v_{11}^4)-3) \cdot \left[\int \frac{P}{\lambda+\sigma^2}\,dG^*(\lambda)\right]^2. \tag{16}$$

Note that

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

and

$$\int \frac{P}{(\lambda+\sigma^2)^2}\,dG^*(\lambda) = -\frac{d\beta^*}{d(\sigma^2)}.$$

Thus to compute the second integral, we need only to differentiate (12) with respect to σ^2. We therefore get

$$b = \frac{2\beta^*(1+\beta^*)^2}{\frac{\sigma^2}{P}(1+\beta^*)^2+\alpha} + (E[v_{11}^4]-3)(\beta^*)^2.$$

The above lemma says that the fluctuation of the first term (13) is of the order of $1/\sqrt{N}$. Regarding the fluctuation of $(P/N)\,\mathrm{tr}\,(F+\sigma^2 I)^{-1}$, we have the following result, the proof of which can be found in Appendix B.

Lemma 4.3:

$$\limsup_{N\to\infty} \mathrm{Var}\,\mathrm{tr}\left(F+\sigma^2 I\right)^{-1} < \infty.$$

This says that the fluctuation of $\frac{P}{N}\,\mathrm{tr}\,(F+\sigma^2 I)^{-1}$ around its mean is of the order at most $1/N$, negligible compared to the first source of fluctuation(13).

Finally, concerning the deviation of the mean SIR from the limit β^*, we have the following result, proved in Appendix C.

Lemma 4.4:

$$\limsup_{N\to\infty} N\left(E[\beta^{(N)}]-\beta^*\right) < \infty.$$

This shows that the mean SIR is of order at most $1/N$ from the limit β^*. Combining Lemmas 4.2, 4.3, and 4.4, we have the following main result characterizing the asymptotic distribution of the performance under the MMSE receiver.

Theorem 4.5: As $N \to \infty$

$$\sqrt{N}(\beta^{(N)}-\beta^*) \xrightarrow{\mathcal{D}} N(0,\, b)$$

where

$$b = \frac{2\beta^*(1+\beta^*)^2}{\frac{\sigma^2}{P}(1+\beta^*)^2+\alpha} + \left(E[v_{11}^4]-3\right)(\beta^*)^2$$

and

$$\beta^* = \frac{(1-\alpha)P}{2\sigma^2} - \frac{1}{2} + \sqrt{\frac{(1-\alpha)^2P^2}{4\sigma^4} + \frac{(1+\alpha)P}{2\sigma^2} + \frac{1}{4}}.$$

Moreover

$$\limsup_{N\to\infty} N\left(E[\beta^{(N)}]-\beta^*\right) < \infty.$$

This theorem says that while the asymptotic limit β^* can be expected to be a very accurate approximation of the mean SIR for reasonably sized system (difference of order $1/N$), the fluctuations can be significantly larger (of order $1/\sqrt{N}$). This will be validated by the simulation results in the next section.

We would like to give some intuition underlying the proof of this result. This is similar to our heuristic discussion on the decorrelator. Because $O^t\boldsymbol{s}_1$ is asymptotically isotropic, we can write

$$O^t\boldsymbol{s}_1 \approx \frac{\|\boldsymbol{s}_1\|}{\|\boldsymbol{r}\|}\,(r_1,\,\cdots,\,r_N)^t$$

where the r_i's are i.i.d. Gaussian $N(0,\,1)$ and independent of $\|\boldsymbol{s}_1\|$. Thus

$$\beta^{(N)} \approx \frac{\|\boldsymbol{s}_1\|^2}{\|\boldsymbol{r}\|^2}\sum_{i=1}^{N}\frac{P}{\lambda_i+\sigma^2}\,r_i^2. \tag{17}$$

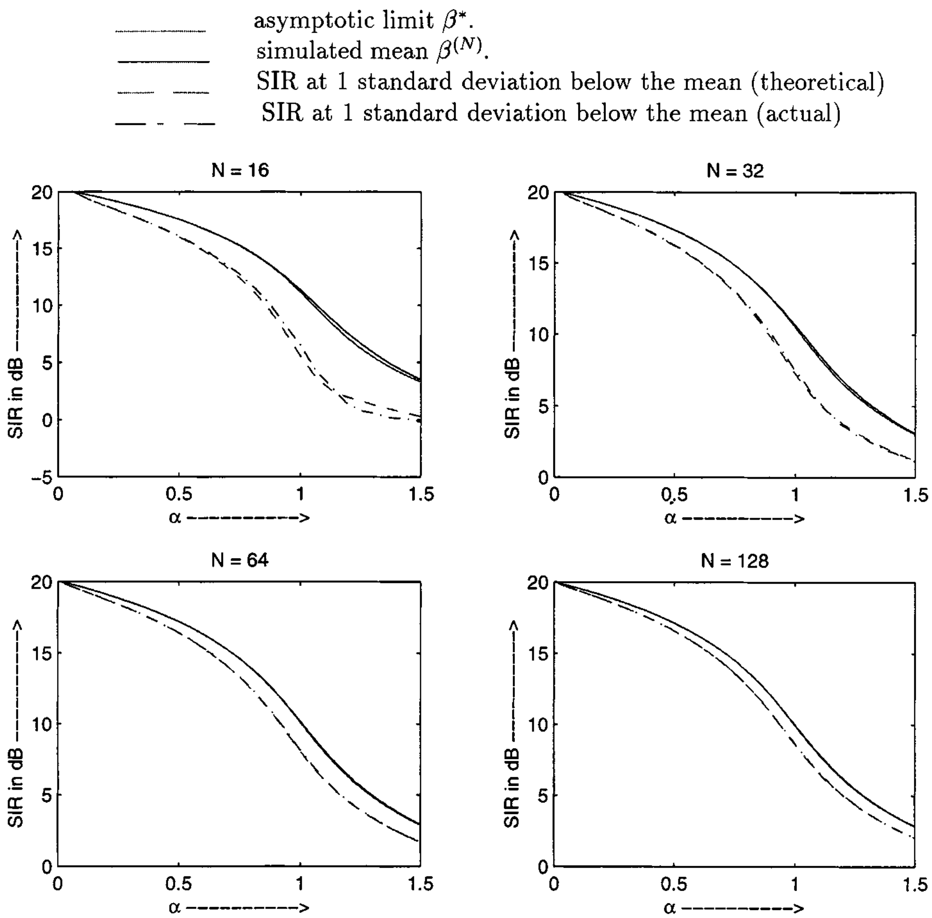

Fig. 1. Asymptotic limit, mean, and SIR at one standard deviation below the mean, theory and actual, for the MMSE receiver.

Using a process-level central limit argument together with the fact that the random eigenvalue fluctuation is small, it can be shown that

$$N\|\boldsymbol{s}_1\|^2 \approx 1 + \frac{1}{\sqrt{N}}\,\phi_1 \tag{18}$$

$$\|\boldsymbol{r}\|^2 \approx 1 + \frac{1}{\sqrt{N}}\,\phi_2 \tag{19}$$

$$\frac{1}{N}\sum_{i=1}^{N} \frac{P}{\lambda_i + \sigma^2}\, r_i^2 \approx \beta^* + \frac{1}{\sqrt{N}}\,\phi_3 \tag{20}$$

where the ϕ_i's are zero-mean jointly Gaussian and ϕ_1 is independent of ϕ_2 and ϕ_3. The second moments of these random variables can be calculated as

$$E[\phi_1^2] = E[v_{11}^4] - 1 \qquad E[\phi_2^2] = 2$$
$$E[\phi_3^2] = 2\int_0^\infty \left[\frac{P}{(\lambda+\sigma^2)}\right]^2 dG^*(\lambda)$$

and

$$E[\phi_2\phi_3] = 2\beta^*.$$

Using (18)–(20), we can expand (17) in the Taylor-series expansion, keeping the first- and second-order terms only, and obtain the Gaussian approximation given in the main theorem.

It was mentioned in Section III that the exact (nonasymptotic) distribution of the SIR under the decorrelator can be computed to be Chi-square when the chip distribution is *Gaussian*. Due to the dependency on the eigenvalue distribution, however, no such simple nonasymptotic expression seems possible for the SIR under the MMSE receiver. Even in the Gaussian case, the marginal distribution of the eigenvalues of $S_1 S_1^t$ is quite complicated, expressed in terms of Laguerre polynomials [21].

V. Simulations and Numerical Results

To see how accurate the limit theorems are for finite-sized system, we compare the theoretical results with actual values obtained by simulations. All simulation results are obtained by averaging over 10 000 independently generated samples, and will be considered as the actual values of the statistics. Users are received at equal power P, and the signal-to-noise ratio (SNR) P/σ^2 is set at 20 dB. The chips of the signature sequences have values $+1/\sqrt{N}$ or $-1/\sqrt{N}$. Figs. 1 and 2 display results for the MMSE receiver. In Fig. 1, we plot the limiting SIR β^* (given by (12)), the mean SIR $\overline{\beta}_1^{(N)}$, and the actual and theoretical SIR level at one standard deviation below the mean. These curves are plotted as a function of the system

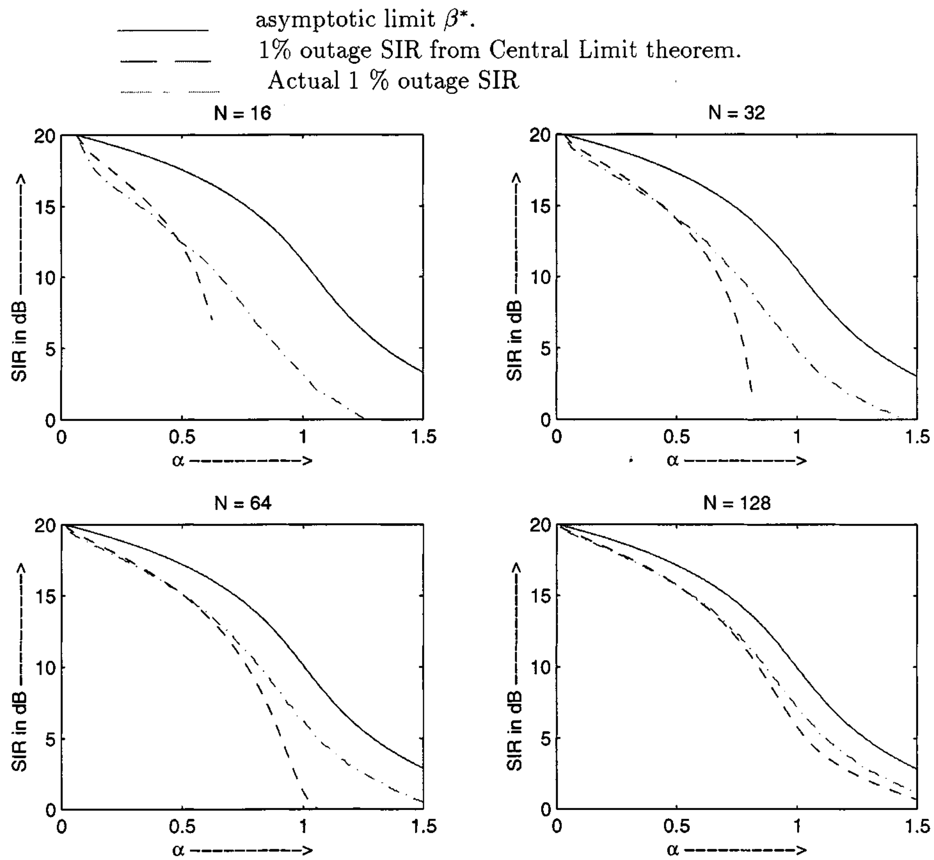

Fig. 2. Comparison of 1% outage SIR, theory and actual, for the MMSE receiver.

load α (the number of users per degree of freedom) for different system sizes: $N = 16, 32, 64, 128$. The theoretical SIR level 1 standard deviation from the mean is given by

$$\beta^* - \sqrt{\frac{b}{N}}$$

where b is given by (16).

We make several observations. First, the mean $\overline{\beta}^{(N)}$ is very close to β^*, and their difference is much smaller than one standard deviation. For $N = 32$ and greater, the β^* and $\overline{\beta}^{(N)}$ curves are almost indistinguishable. This confirms our theoretical results, which predict that $\beta^* - \overline{\beta}^{(N)}$ is going to zero at least as fast as $1/N$, while the standard deviation goes to zero like $1/\sqrt{N}$. Second, the theoretical prediction $\sqrt{b/N}$ of the standard deviation is quite close to the actual standard deviation. Again, the two corresponding curves are almost indistinguishable for $N \geq 32$. Third, the standard deviation compared to the mean SIR is small where there are few users per unit processing gain, but quite significant when there are many users. This is true even for $N = 64$.

Next, we investigate how accurate the central-limit results are in predicting the tail of the SIR distribution. In Fig. 2, we compare the actual 1% outage SIR with that predicted by Theorem 4.5. (The 1% outage level is the value x such that $\Pr(\text{SIR} < x) = 0.01$.) We see that while the theoretical result is accurate when the system load α is small (less than 0.5), it tends to be over-pessimistic for α larger, when the achieved SIR is small. The accuracy of the theoretical results becomes good for the entire range only when $N = 128$.

For the decorrelator, as we mentioned in Section III and as was also independently pointed out in [14], an alternative approximation is suggested by the heuristic (8). This does not assume a Gaussian approximation to the various sums, but is only based on the fact that $O^t\boldsymbol{s}_1$ is asymptotically isotropic. The random variable

$$\frac{\sum_{i=K}^{N} r_i^2}{\sum_{i=1}^{N} r_i^2}$$

follows a beta distribution, since the r_i's are i.i.d. $N(0, 1)$. (See, for example, [3].) Hence an approximation to $\gamma^{(N)}$ is a product of the independent random variable $||\boldsymbol{s}_1||^2$ and a beta distributed random variable. In the special case of $+1, -1$ sequences, $||\boldsymbol{s}_1||^2 = 1$, and this approximation simply becomes a beta distribution. The approximation is applied to calculate the 1% outage level for the decorrelator in Fig. 3. The result is compared to the actual 1% outage level, as well as the central-limit approximation provided by Theorem 3.3. We see that even for $N = 16$, the beta distribution approximation is very accurate, and in fact indistinguishable from the actual values for $N \geq 32$. On the other hand, the central limit approximation, while accurate for small α, tends to be over-pessimistic for α close to 1. This suggests that for moderate N, $O^t\boldsymbol{s}_1$ is

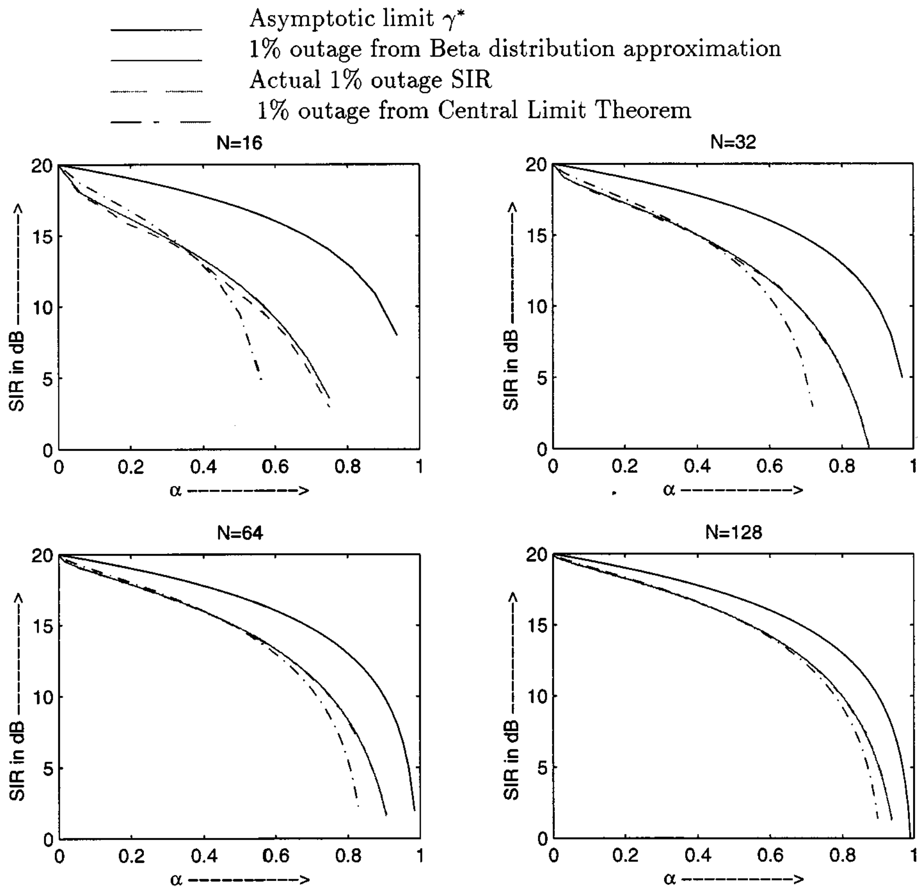

Fig. 3. Comparison of 1% outage SIR for the decorrelator, central limit approximation, beta distribution approximation, and actual.

already very close to perfectly isotropic. On the other hand, the Gaussian approximation to $\sum_{i=1}^{N} r_i^2$ and $\sum_{i=K}^{N} r_i^2$ introduces errors which are only negligible when N is quite large. Thus when $N = 128$, all three curves (actual, beta distribution approximation, central limit Theorem approximation) merge.

VI. Antenna Diversity

In the previous sections, we have focused on the DS-CDMA system with random signature sequences. Another example of a random environment in which linear multiuser receivers operate is a system with multiple antennas for providing spatial diversity. These antennas can be arranged in an array located at a single base station, or they can be distributed in geographically different locations in which case they provide *macrodiversity*. Antenna elements colocated in an array mainly serves to combat *multipath fading*, while a distributed antenna system can combat larger scale fading effects. In any case, performance can be improved by adaptive combination of signals received at the various antenna elements depending on the channel strengths. A general baseband model for such a system with flat fading is given by

$$\boldsymbol{y} = \sum_{i=1}^{K} b_i \boldsymbol{h}_i + \boldsymbol{z}$$

where b_i is the transmitted symbol of the ith user, and $\boldsymbol{y}$ is an N-dimensional vector of received symbols at the antennas. The vector $\boldsymbol{z}$ is i.i.d. complex circular symmetric Gaussian noise with variance per component σ^2. The vector $\boldsymbol{h}_i$ represents the (flat) fading of the ith user at each of the antennas. Let $\boldsymbol{h}_j = (v_{1j}/\sqrt{N}, \cdots, v_{Nj}/\sqrt{N})^t$. We will assume a fading model in which v_{ij} are i.i.d. circular symmetric random variables with variance $E[|v_{ij}|^2]$ normalized to be 1, to keep the total received energy at the antennas constant, irrespective of the number of antennas. The circular symmetry arises naturally when shifting from a high carrier frequency to the baseband. We will also assume the signal constellation is circular symmetric as well, so that b_i is circular symmetric. The average received power of all users are assumed to be the same, $E[|b_i|^2] = P$. We let $\alpha = \lfloor K/N \rfloor$ be the number of users per antenna element.

Assuming that the receiver can track the fading perfectly, the MMSE receiver is the optimal linear receiver in maximizing the SIR of each user 1. The decorrelator nulls out the interference from other users. The performance of both of these receivers is a function of the channel fading at the current time, and is therefore random.

The similarity of the multiantenna model with the DS-CDMA system is obvious, with the signature sequences replaced by the channel fading vectors. The only difference is that the entries of H are now complex rather than real as in the signature sequences. Rigorously speaking, Theorem A.1 which we used for analyzing the DS-CDMA problem is only proved for *real* v_{ij}. (The proofs of Lemmas 4.3 and 4.4 carry over verbatim to the complex case, cf. [2] for a similar argument.) The

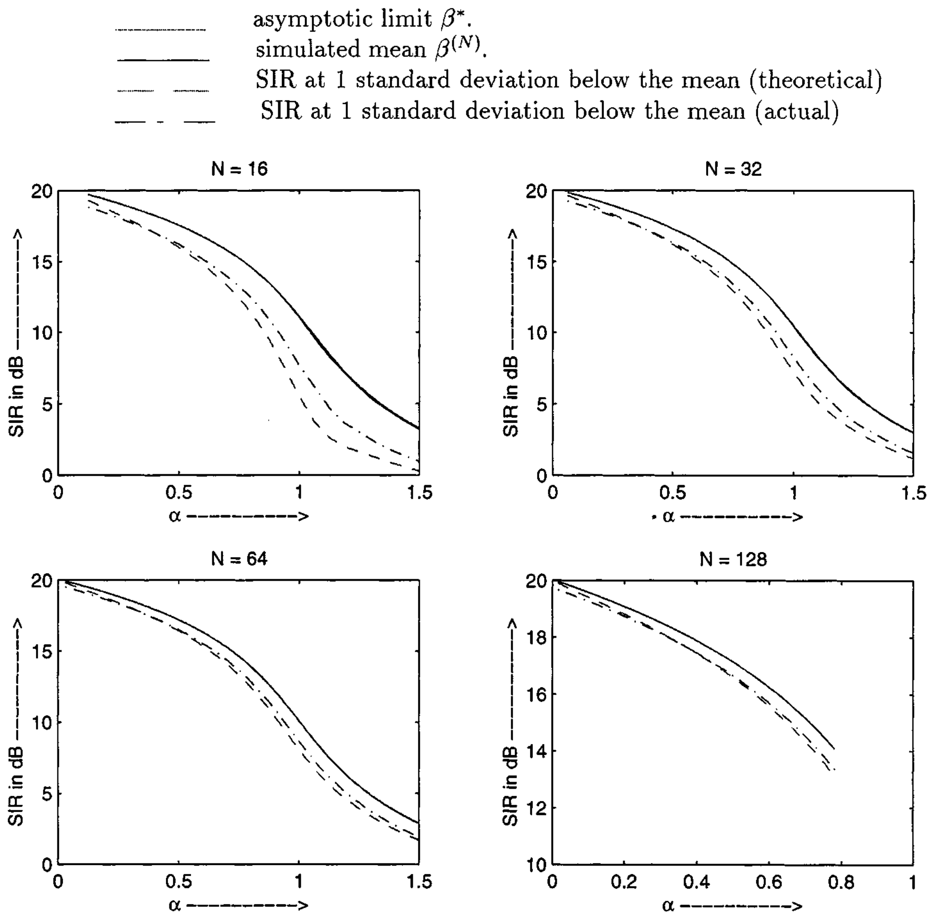

Fig. 4. Asymptotic limit, mean, and SIR at one standard deviation below the mean, theory and actual, for the MMSE receiver in a lognormal fading environment.

extension of Theorem A.1 to the complex case is straightforward, once [18, Theorem 4.1] is extended to the complex case, which is also straightforward, albeit tedious. Carrying out these extensions explicitly is, however, beyond the scope of this paper. Thus we believe that Theorem A.1 generalizes to the case when v_{ij} is complex circular symmetric, but we do not provide a detailed proof. Assuming this generalization, the performance of the decorrelator in the multiantenna system can be approximated by (in analogy to (8))

$$\gamma^{(N)} \approx \frac{P}{\sigma^2} \frac{\|\boldsymbol{h}_1\|^2}{\|\boldsymbol{r}\|^2} \sum_{i=K}^{N} |r_i|^2$$

where r_i's are i.i.d. zero-mean complex circular symmetric Gaussian random variables with $E[|r_i|^2] = 1$. This assumes $\alpha < 1$. Note that in the case of the Rayleigh fading model, v_{ij}'s are circular symmetric Gaussian and the approximation becomes exact, and this specializes to the result of [26]. For large N, applying the central limit theorem, $\gamma^{(N)}$ can be further approximated by a Gaussian random variable with mean $(1-\alpha)P/\sigma^2$ and variance

$$\frac{1}{N}\left(\frac{P}{\sigma^2}\right)^2 \left\{2(1-\alpha) + (E[|v_{11}|^4] - 3)(1-\alpha)^2\right\}.$$

For the MMSE receiver, the SIR performance can be approximated, for large N, by a Gaussian random variable with mean β^* and variance

$$\frac{1}{N}\left\{\frac{2\beta^*(1+\beta^*)^2}{\frac{\sigma^2}{P}(1+\beta^*)^2+\alpha} + \left(E[|v_{11}|^4]-3\right)(\beta^*)^2\right\}$$

where β^* is given by(12).

Figs. 4 and 5 show simulation results which support the theory, for the case when the channel gains v_{ij}'s are circular symmetric and magnitude distributed as lognormal with standard deviation 8 dB. This is a model for large-scale fading effects due to shadowing effects, as would be appropriate for a distributed antenna system where the antennas are physically spaced far apart. For a randomly located user, it is reasonable to model the fading to each of the antennas as independent. Similar to the results for the binary spreading sequences, the mean and variance approximations are quite accurate, even for small N, while the approximation for the tail tends to be conservative except for N large.

VII. Conclusions

In this paper, we studied the SIR performance of the decorrelator and the MMSE receiver in a random environment. Such random environments may arise in a DS-CDMA system with random signature sequences, or in a system with antenna diver-

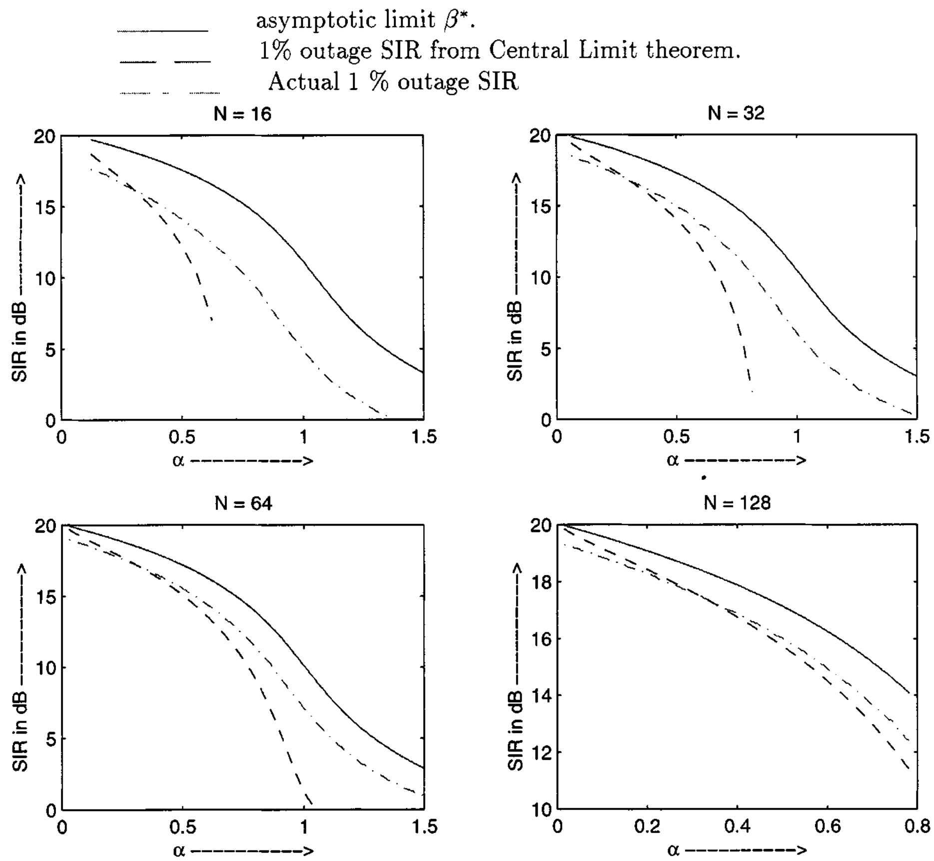

Fig. 5. Comparison of 1% outage SIR theory and actual, for the MMSE receiver in the lognormal fading environment.

sity where the randomness is due to channel fading. We showed that for the two receivers considered, the variance of the SIR distribution decreases like $1/N$, and the SIR distribution is asymptotically Gaussian. We computed closed-form expressions for the variances for both receivers, and observed that the relative amount of fluctuation is large when there are many users per degree of freedom and the achieved SIR is low.

Simulation results show that the asymptotic mean and variance computed from the theory are very accurate approximations for even moderate system size and for a wide range of α (the number of users per degree of freedom). On the other hand, when the achieved SIR is small and system size only moderate, the Gaussian approximation is not very good for approximating the tail of the SIR distribution (1% outage, for example.) Based on insights gained from the theory, an alternative approximation based on the beta distribution is derived for the performance of the decorrelator. This approximation, observed independently in [14], is very accurate for moderate system size and for the whole range of α.

There are several interesting directions for future work. One remedy to offset the random fluctuation of the SIR is through power control. The interesting question is then to characterize the distribution of power required to keep the SIR at a desired level. The problem is complicated by the fact that all users will vary their powers simultaneously to achieve their individual desired SIR. However, we conjecture that in the scaling considered in this paper, the performance of a user is insensitive to the power variations of other users and depends mainly on its own power. This would then imply that the power distribution can be computed as that of the reciprocal of the SIR calculated in this paper.

Another interesting question is to characterize the empirical distribution of SIR levels of the users across the system. Contrast this with the SIR distribution of a *particular* user, which is what we computed in this paper. We conjecture that in a large system, some kind of "weak asymptotic independence" between users will hold and with high probability the two distributions are very close.

APPENDIX A
ASYMPTOTICALLY ISOTROPIC EIGENVECTORS

In this section, we develop the machinery required to prove Theorem 3.3 and Lemma 4.2. Theorem A.1 quantifies precisely what it means to say that the eigenvectors of a random matrix are *asymptotically isotopic*. Its proof uses heavily ideas from Silverstein [18] and so we adopt his notations.

Notations: We let $\boldsymbol{x}_n = (x_1, \cdots, x_n)$ denote random vectors with

$$\|\boldsymbol{x}_n\| = \sqrt{\sum_{i=1}^{n} x_i^2} = 1$$

and let $\boldsymbol{z}_n = (z_1, \cdots, z_n)$ denote arbitrary random vectors in $\Re^n$. As in [18], we let

$$\{v_{ij}\}_{\substack{i=1,\cdots\\ j=1,\cdots}}$$

be i.i.d. random variables with $Ev_{ij}^2 = 1$, $Ev_{ij} = 0$, $Ev_{ij}^4 < \infty$ and symmetric distribution. With

$$V_n = \{v_{ij}\}_{\substack{i=1\\ j=1,\cdots,s(n)}}^{n}$$

let $M_n = \frac{1}{s(n)} V_n V_n^t$, and define

$$\frac{n}{s(n)} \underset{n\to\infty}{\rightarrow} y \in (0, \infty).$$

Let $M_n = O_n \Lambda_n O_n^t$ denote the spectral resolution of M_n, with Λ_n a diagonal matrix whose entries, the eigenvalues of M_n, are arranged in nondecreasing order, and O_n denoting an orthonormal matrix consisting of the eigenvectors of M_n.

Let $z_1, z_2, \cdots$ denote a sequence of i.i.d., independent of $\{v_i\}$ random variables, with $Ez_i = 0$, $Ez_i^2 = 1$, and $Ez_i^4 < \infty$, with symmetric distribution. Let

$$\eta_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (z_i^2 - 1) \xrightarrow{\mathcal{D}} \eta$$

where η denotes a Gaussian random variable with zero mean and variance $\sigma_\eta^w = Ez_1^4 - 1$, and the convergence, due to the CLT, is in the sense of distributions. Letting $W^o(\cdot)$ denote a Brownian bridge, that is the zero mean Gaussian process on $[0, 1]$ with covariance $E(W^o(s)W^o(t)) = \min(s, t) - st$, define $\boldsymbol{y}_n = (y_1, \cdots, y_n)^t = O_n^t \boldsymbol{z}_n$ where $\boldsymbol{z}_n = (z_1, \cdots, z_n)$. Introduce the process

$$Z_n(t) = \sqrt{\frac{1}{2n}} \sum_{i=1}^{[nt]} \left(y_i^2 - \frac{1}{n}\right).$$

The main result of this section is the following theorem. We refer to [5] for the basic definitions and properties of the Skorohod topology and of weak convergence of probability measures on $D[0, 1]$.

Theorem A.1:

$$\{Z_n(t)\}_{t\in[0,1]} \xrightarrow{\mathcal{D}} \left\{W^o + \frac{\eta}{\sqrt{2}} t\right\}_{t\in[0,1]} \tag{21}$$

where W^o and η are independent and the convergence is in the sense of distributions in $D[0, 1]$, the space of right continuous functions with left limits (RCLL), equipped with the Skorohod topology.

To see why this says that the vector $\boldsymbol{y}_n$ is asymptotically isotropic, consider the special case when $\boldsymbol{z}_n$ is an i.i.d. Gaussian random vector, i.e., isotropic to start with. Then $\boldsymbol{y}_n$ is also an i.i.d. zero-mean Gaussian vector which is clearly isotropic. It is not difficult in this case to verify by a standard functional central limit theorem that Theorem A.1 holds. What Theorem A.1 says is that $\boldsymbol{y}_n$ will be asymptotically isotropic even in the general case when the z_i's are not. Its truth depends on the asymptotic isotropic property of the eigenvector matrix O_n.

A consequence of Theorem A.1, of use to us in this paper, is the following. Suppose D_n is a diagonal matrix (possibly random) with entries monotone on the diagonal and eigenvalue distribution

$$L_n^D := \frac{1}{n} \sum_{i=1}^{n} \delta_{(D_n)_{ii}} \Rightarrow \mu_D$$

such that μ_D possesses a continuous cumulative distribution function (c.d.f.) $F_D(x)$ with support on some compact set $[0, a]$, with possibly a jump at 0. Denote by $M_n = O_n D_n O_n^t$, and let

$$\theta_n = \boldsymbol{y}_n D_n \boldsymbol{y}_n^t.$$

Here and in the sequel, we use $\mathbb{E}$ to denote expectation with respect to (w.r.t.) the randomness incurred in D_n, $\{z_i\}$ and M_n. Note that

$$\frac{1}{n}\mathbb{E}\theta_n = \frac{1}{n}\mathbb{E}\,\mathrm{tr}\, D_n \to_{n\to\infty} \int_0^a x\mu_D(dx).$$

Corollary A.2: As $n \to \infty$

$$\sqrt{\frac{1}{2n}}(\theta_n - \mathrm{tr}\, D_n) \xrightarrow{\mathcal{D}} N(0, \sigma_\theta^2) \tag{22}$$

where

$$\sigma_\theta^2 = \int_0^a x^2 \mu_D(dx) + \left(\frac{\sigma_\eta^2}{2} - 1\right)\left(\int_0^a x\mu_D(dx)\right)^2. \tag{23}$$

Proof of Corollary A.2: Let

$$F_n(x) = \int_0^{x^-} L_n^D(dy)$$

denote the number of eigenvalues of D_n not larger than x. Then (assuming D_n does not possess multiple eigenvalues except possibly at 0)

$$\begin{aligned}\theta_n - \mathrm{tr}\, D_n &= \sum_{i=1}^{n} \left(y_i^2 - \frac{1}{n}\right)(D_n)_{ii}\\ &= \sqrt{2n} \sum_{i=1}^{n} (D_n)_{ii} \left(Z_n((D_n)_{ii}^+) - Z_n((D_n)_{ii}^-)\right)\\ &= \sqrt{2n} \int_0^a x dZ_n(F_n(x))\\ &= -\sqrt{2n} \int_0^{-a} Z_n(F_n(x))\, dx.\end{aligned}$$

Since $F_n(x)$ converges to the continuous $F_D(x)$, the weak convergence of $Z_n(\cdot)$ is carried over to that of $Z_n(F_n(\cdot))$, and the latter converges in distribution to the Gaussian process

$$\left\{W^o(F_D(x)) + \frac{\eta}{\sqrt{2}} F_D(x)\right\}_{x\in[0,\infty)}.$$

Hence

$$\begin{aligned}\frac{1}{\sqrt{2n}}(\theta_n - \mathrm{tr}\, D_n) &\xrightarrow{\mathcal{D}} \int_0^a x\left[d(W^o(F_D(x)) + \frac{\eta}{\sqrt{2}} dF_D(x)\right]\\ &= \frac{\eta}{\sqrt{2}} \int_0^a (x dF_D(x) + \int_0^a d(W^o(F_D(x)).\end{aligned}$$

The convergence (22) and the value of the variance in (23) follow from evaluating the variance of the limiting Gaussian process. □

Proof of Theorem A.1: The proof is a modification of the argument presented in [18]. Let $L_n^{\Lambda_n} = \frac{1}{n}\sum_{i=1}^n \delta_{(\Lambda_n)_{ii}}$, and

$$F_n^\Lambda(x) = \int_0^{x^-} L_n^{\Lambda_n}(dy)$$

denote the number of eigenvalues of Λ_n smaller than x. It is well known that $L_n^{\Lambda_n} \xrightarrow{w} \mu_\Lambda$, in the sense of weak convergence of distributions, with

$$F^\Lambda(y) = \int_0^{x^-} \mu_\Lambda(dy)$$

denoting the appropriate distribution function. Then $F_n^\Lambda(x) \to F^\Lambda(x)$, uniformly, cf. the argument in [18, p. 1176].

By [18, Theorem 4.1], for any fixed sequence of vectors with

$$\|\boldsymbol{x}_n\| = 1, \quad \sum_{i=1}^n x_{n,i}^4 \underset{n\to\infty}{\to} 0$$

we have that

$$\left\{\sqrt{\frac{n}{2}}\left(\boldsymbol{x}_n^t M_n^r \boldsymbol{x}_n - \frac{1}{n}\operatorname{tr}(M_n^r)\right)\right\}_{r=1}^\infty \xrightarrow{\mathcal{D}} \left\{\int x^r\, dW^o\left(F^\Lambda(x)\right)\right\}_{r=1}^\infty$$

in the sense of convergence of laws in $\Re^\infty$. In particular, it follows that for any sequence $\varepsilon_n \to 0$

$$\overline{d} := \sup_{\{\boldsymbol{x}_n:\, \|\boldsymbol{x}_n\|=1, \sum_{i=1}^n x_{n,i}^4 < \varepsilon_n\}} d_{\mathcal{L}} \cdot \left(\left\{\sqrt{\frac{n}{2}}\left(\boldsymbol{x}_n^t M_n^r \boldsymbol{x}_n - \frac{1}{n}\operatorname{tr}(M_n^r)\right)\right\}_{r=1}^\infty, \left\{\int x^r\, dW^o\left(F^\Lambda(x)\right)\right\}_{r=1}^\infty\right) \underset{n\to\infty}{\to} 0 \tag{24}$$

where $d_{\mathcal{L}}(a, b)$ denote the distance between the laws of the random variables a, b in, say, the Lévy–Prohorov metric (see [5] for the definition; any metric between probability measures which is compatible with weak convergence can be used here).

Next, note that

$$\begin{aligned}
&\left\{\sqrt{\frac{1}{2n}}\left(\boldsymbol{z}_n^t M_n^r \boldsymbol{z}_n - \operatorname{tr}(M_n^r)\right)\right\}_{r=1}^\infty \\
&= \left\{\sqrt{\frac{1}{2n}}\left(\frac{\boldsymbol{z}_n^t}{\|\boldsymbol{z}_n\|} M_n^r \frac{\boldsymbol{z}_n}{\|\boldsymbol{z}_n\|} - \frac{\operatorname{tr}(M_n^r)}{n}\right)\right\}_{r=1}^\infty \|\boldsymbol{z}_n\|^2 \\
&\quad + \left\{\sqrt{\frac{1}{2n}}\frac{(\operatorname{tr} M_n^r)}{n}(\|\boldsymbol{z}_n\|^2)\right\}_{r=1}^\infty \\
&\triangleq Z_n^1 + Z_n^2
\end{aligned}$$

where Z_n^i are random variables taking values in $\Re^\infty$.

Let $\varepsilon_n = n^{-(1/2)}$, and set

$$A_n = \left\{\boldsymbol{x}_n:\, \|\boldsymbol{x}_n\| = 1, \sum_{i=1}^n x_{n,i}^4 < \varepsilon_n\right\}.$$

Note that

$$\sum_{i=1}^n \left(\frac{z_i}{\|\boldsymbol{z}_n\|}\right)^4 = \frac{\frac{1}{n}\sum_{i=1}^n z_i^4}{n \cdot \left(\frac{1}{n}\sum_{i=1}^n z_i^2\right)^2} \xrightarrow{P} 0$$

and further

$$P\left(\frac{\boldsymbol{z}_n}{\|\boldsymbol{z}_n\|} \notin A_n\right) \le C\frac{\varepsilon_n}{n} \underset{n\to\infty}{\to} 0.$$

Let $L_1^{\boldsymbol{z}_n}$ denote the law of Z_n^1 *conditioned* on $\boldsymbol{z}_n$, let L^W denote the law of $\{\int x^r\, dW^o(F^\Lambda(x))\}_{r=1}^\infty$. Then, using (24)

$$d_{\mathcal{L}}(L_1^{\boldsymbol{z}_n}, L^W) \le P\left(\frac{\boldsymbol{z}_n}{\|\boldsymbol{z}_n\|} \in A_n\right) + \overline{d} \underset{n\to\infty}{\to} 0. \tag{25}$$

This, together with the convergence $\operatorname{tr} M_n^r/n \to \int x^r\, dF_n^\Lambda(x)$ and $(\|\boldsymbol{z}_n\|^2 - n)/\sqrt{n} \xrightarrow{\mathcal{D}} N(0, \eta)$, imply that

$$Z_n^1 + Z_n^2 \xrightarrow{\mathcal{D}} \left\{\int x^r\, dW^o\left(F^\Lambda(x)\right)\right\}_{r=1}^\infty + \left\{\frac{\eta}{\sqrt{2}}\int x^r\, dF_n^\Lambda(x)\right\}_{r=1}^\infty \tag{26}$$

where η is a $N(0, \sigma_\eta)$ random variable independent of W^o.

The next step consists of inverting the time change in (26). In view of the argument in [18, p. 1191], (21) follows from (26) as soon as some tightness holds, that is, as soon as one shows that for some $C > 0$

$$\mathbb{E}(Z_n(F_n(0)))^4 \le C\, E(F_n(0))^2 \tag{27}$$

and for any $0 \le x_1 \le x_2$

$$\mathbb{E}(Z_n(F_n(x_2)) - Z_n(F_n(x_1)))^4 \le C\, E(F_n(x_2) - F_n(x_1))^2 \tag{28}$$

(compare with [18, Theorem 4.2]). In fact, the proof of these facts follows closely the proof in [18], whose notations we adopt here. Since the proof of (27) is similar, we consider below only the proof of (28). Let $P^n = \{P_{ij}\}$ denote the projection matrix on the subspace of $\Re^n$ spanned by the eigenvectors of M_n having eigenvalues in $[x_1, x_2]$. One checks immediately that

$$Z_n(F_n(x_2)) - Z_n(F_n(x_1)) = \sqrt{\frac{1}{2n}}(\boldsymbol{z}_n^t P^n \boldsymbol{z}_n - \operatorname{tr} P^n).$$

With $\gamma_{ij} = z_i z_j$, one sees that the left-hand side of (28) satisfies

$$\begin{aligned}
&\mathbb{E}(Z_n(F_n(x_2)) - Z_n(F_n(x_1)))^4 \\
&= \frac{1}{4n^2}\mathbb{E}\left(\sum_{i,j}\gamma_{ij}P_{ij} - \sum_i P_{ii}\right)^4 \\
&\le \frac{c}{n^2}\left(\mathbb{E}\left(\sum_{i\ne j}\gamma_{ij}P_{ij}\right)^4 + \mathbb{E}\left(\sum_{i=1}^n (z_i^2 - 1)P_{ii}\right)^4\right) \\
&\triangleq c(I_1 + I_2)
\end{aligned}$$

for some constant c independent of n.

Following the same argument that led to [18, eq. (4.10)], one finds that

$$\begin{aligned} I_1 = \frac{n-1}{n} \big(&12(n-2)Ez_1^4(Ez_1^2)^2E(P_{12}^2P_{13}^2) \\ &+ 3(n-2)(n-3)(Ez_1^2)^4E(P_{12}^2P_{34}^2) \\ &+ 12(n-2)(n-3)(Ez_1^2)^4E(P_{12}P_{23}P_{34}P_{14}) \\ &+ 2E(P_{12}^4)(Ez_1^4)^2\big). \end{aligned}$$

Using the fact that P is a projection matrix, we have that

$$P_{13} = \sum_{j\ge 4} P_{3j}P_{1j} + P_{31}P_{11} + P_{32}P_{12} + P_{33}P_{13}.$$

Using the fact that expectations are invariant with respect to permutations, we conclude, as in [18], that

$$\begin{aligned} &(n-2)(n-3)E(P_{12}P_{23}P_{34}P_{14}) \\ &\quad \le EP_{12}^2 + 2E(P_{11}P_{22}P_{12}^2) + 2EP_{11}^2P_{12}^2 \\ &\quad \le EP_{11}P_{22} + 2E(P_{11}^2P_{22}^2) + 2E(P_{11}P_{22}P_{11}^2) \\ &\quad \le EP_{11}P_{22} + 2E(P_{11}P_{22}) + 2E(P_{11}P_{22}) \\ &\quad = 5E(P_{11}P_{22}) \end{aligned}$$

where we made repeated use of $P_{12}^2 \le P_{11}P_{22}$ and $P_{11} \ge \max(P_{11}P_{22}, P_{11}^2)$. Similarly,

$$(n-2)E(P_{12}^2P_{13}^2) \le E(P_{11}P_{12}^2) \le E(P_{11}^2P_{22}) \le E(P_{11}P_{22})$$

and

$$(n-3)E(P_{12}^2P_{34}^2) \le E(P_{12}^2P_{33})$$

leading to

$$(n-2)(n-3)E(P_{12}^2P_{34}^2) \le (n-2)E(P_{12}^2P_{33}).$$

However, using the identity

$$P_{33}\left(\sum_{j\ne 1,3} P_{ij}P_{ij} + P_{11}P_{11} + P_{13}P_{13}\right) = P_{33}P_{11},$$

we obtain

$$(n-2)E(P_{33}P_{12}^2) + EP_{11}^2P_{33} + EP_{13}^2P_{33} = EP_{11}P_{33},$$

leading to

$$(n-2)EP_{33}P_{12}^2 \le EP_{11}P_{33} = EP_{11}P_{22}.$$

Hence

$$(n-2)(n-3)E(P_{12}^2P_{34}^2) \le EP_{11}P_{22}.$$

Combining the above, we conclude that

$$I_1 \le K_1E(P_{11}P_{22})$$

for some constant K_1 which depends on Ez_1^4 only. Similarly, for some constant K_2 independent of n

$$\begin{aligned} I_2 &\le \frac{1}{n}\,\mathbb{E}P_{11}^4 + \frac{3(n-1)}{n}(\mathbb{E}P_{11}^2P_{22}^2)\left(E(z_i^2-1)^2\right)^2 \\ &\le \frac{1}{n}\,\mathbb{E}P_{11}^4 + K_2\mathbb{E}(P_{11}P_{22}). \end{aligned}$$

Hence, for some K_3 independent of n

$$\begin{aligned} c(I_1+I_2) &\le K_3\left(\frac{1}{n}\,\mathbb{E}P_{11}^4 + \frac{(n-1)}{n}\,\mathbb{E}(P_{11}P_{22})\right) \\ &= K_3\mathbb{E}\left(\frac{1}{n}\sum_i P_{ii}\right)^2 = K_3(F_n(x_2)-F_n(x_1))^2 \end{aligned}$$

as required. □

Appendix B
Proof of Lemma 4.3

Throughout this proof, we follow the notations of Section IV, while for simplicity taking $P = 1$. That is, we consider the matrix $S_1 := [\boldsymbol{s}_2, \cdots, \boldsymbol{s}_K]$, and denote by $\{\lambda_i\}_{i=1}^N$ the eigenvalues of $S_1S_1^t$. The case for general P follows directly from a rescaling of σ^2.

We use various constants C, C_i, whose values may change from line to line and are always independent of N (but may depend on σ). We also use constants K_p, whose values may change from line to line, and which are independent of N and σ.

Before starting, we recall the Burkholder inequality, cf. [4]: If $\{\theta_i\}$ is a martingale difference sequence with respect to an increasing filtration $\mathcal{G}_i$, i.e., θ_i is $\mathcal{G}_i$ measurable and $E(\theta_i|\mathcal{G}_{i-1}) = 0$, then, for any $p > 1$,

$$E\left|\sum_{i=1}^k \theta_i\right|^p \le K_pE\left|\sum_{i=1}^k \theta_1^2\right|^{p/2}. \tag{29}$$

Using the fact that if $\{\theta_i\}$ is square integrable then $\theta_i^2 - E(\theta_i^2|\mathcal{G}_{i-1})$ is again a martingale difference sequence, and iterating $\lceil \log_2 p\rceil$ times this inequality, one also gets that for $p \ge 2$

$$E\left|\sum_{i=1}^k \theta_i\right|^p \le K_pE\sum_{i=1}^k |\theta_1^p| + K_pE\left(\sum_{i=1}^k E(\theta_i^2|\mathcal{G}_{i-1})\right)^{p/2}. \tag{30}$$

We emphasize that in (29) and (30), K_p does not depend on k.

Let $A := (S_1S_1^t + \sigma^2I)^{-1}$. Noting that

$$S_1S_1^t = \sum_{j=2}^K \boldsymbol{s}_j\boldsymbol{s}_j^t$$

we let $A_j := (S_1S_1^t + \sigma^2I - \boldsymbol{s}_j\boldsymbol{s}_j^t)^{-1}$. Since

$$\operatorname{tr} A = \sum_{i=1}^N \frac{1}{\lambda_i + \sigma^2}$$

we need only estimate $E(\operatorname{tr} A - E\operatorname{tr} A)^2$.

Let $\mathcal{F}_j = \sigma(\boldsymbol{s}_i, 2 \le i \le j)$, and write $E_j(\cdot) = E(\cdot|\mathcal{F}_j)$. Using the identity

$$\operatorname{tr} A_j - \operatorname{tr} A = \frac{\boldsymbol{s}_j^t A_j^2 \boldsymbol{s}_j}{1 + \boldsymbol{s}_j^t A_j \boldsymbol{s}_j}$$

we have that

$$\begin{aligned}\operatorname{tr} A - E \operatorname{tr} A &= \sum_{j=2}^{K} E_j \operatorname{tr} A - E_{j-1} \operatorname{tr} A \\ &= \sum_{j=2}^{K} (E_j - E_{j-1}) \operatorname{tr} A_j \\ &\quad + \sum_{j=2}^{K} (E_j - E_{j-1}) \frac{\boldsymbol{s}_j^t A_j^2 \boldsymbol{s}_j}{1 + \boldsymbol{s}_j^t A_j \boldsymbol{s}_j} \\ &= \sum_{j=2}^{K} (E_j - E_{j-1}) \frac{\boldsymbol{s}_j^t A_j^2 \boldsymbol{s}_j}{1 + \boldsymbol{s}_j^t A_j \boldsymbol{s}_j}.\end{aligned}$$

We now define

$$a_j = \frac{1}{N} \operatorname{tr}(A_j^2) \qquad \alpha_j = \boldsymbol{s}_j^t A_j^2 \boldsymbol{s}_j - a_j \qquad \omega_j = \frac{1}{1 + \boldsymbol{s}_j^t A_j \boldsymbol{s}_j}$$

$$b_N = \frac{1}{1 + N^{-1} E \operatorname{tr} A_j} \qquad \zeta_j = \boldsymbol{s}_j^t A_j \boldsymbol{s}_j - N^{-1} E \operatorname{tr} A_j$$

$$\hat{\zeta}_j = \boldsymbol{s}_j^t A_j \boldsymbol{s}_j - N^{-1} \operatorname{tr} A_j.$$

Using some algebra, one arrives at

$$\begin{aligned}&\sum_{j=2}^{K} (E_j - E_{j-1}) \frac{\boldsymbol{s}_j^t A_j^2 \boldsymbol{s}_j}{1 + \boldsymbol{s}_j^t A_j \boldsymbol{s}_j} \\ &\quad = b_N \sum_{j=2}^{K} E_j \alpha_j - b_N^2 \sum_{j=2}^{K} E_j a_j \hat{\zeta}_j \\ &\qquad - b_N^2 \sum_{j=2}^{K} (E_j - E_{j-1})(\alpha_j \zeta_j - \boldsymbol{s}_j^t A_j^2 \boldsymbol{s}_j \omega_j \zeta_j^2) \\ &\quad := W_1 - W_2 - W_3.\end{aligned} \tag{31}$$

Hence, since $0 < \alpha < \infty$ and b_N is uniformly bounded, it will be enough to estimate $E(W_1/b_N)^2$, $E(W_i/b_N^2)^2$, $i = 2, 3$.

Recall the following Lemma, which represents a variant of Lemma 3.2.

Lemma B.1 [2, Lemma 2.7]: There exists for each $p \ge 2$ a universal constant K_p such that, for any deterministic matrix D, and any vector of i.i.d. random variables $\boldsymbol{x} = (x_1, \cdots, x_n)$ with $Ex_1 = 0$ and $Ex_1^2 = 1$

$$\begin{aligned}E|\boldsymbol{x}^t D \boldsymbol{x} - \operatorname{tr} D|^p \le K_p (E(|x_1|^4 &\operatorname{tr} DD^t)^{p/2} \\ &+ E|x_1|^{2p} \operatorname{tr}(DD^t)^{p/2}).\end{aligned}$$

Turning to the first term in (31), note that $E_j(\alpha_j)$ is a martingale difference sequence, i.e., $E_{j-1}(E_j \alpha_j) = 0$. Hence, by Burkholder's inequality (30), for each $p \ge 2$, there exists some universal constant K_p such that

$$\begin{aligned}E(W_1/b_N)^2 &= E\left(\sum_{j=2}^{K} (E_j \alpha_j)\right)^2 \\ &\le K_p \left(E\left(\sum_{j=2}^{K} E_{j-1}(E_j \alpha_j)^2\right)^{p/2} \right. \\ &\qquad \left. + E\left(\sum_{j=2}^{K} (E_j \alpha_j)^p\right)\right).\end{aligned} \tag{32}$$

Then, since the eigenvalues of the matrices A_j are bounded above by σ^{-2}, we have for any $p > 2$ by Lemma B.1 (take $\boldsymbol{x} = \boldsymbol{s}_j \sqrt{N}$, $D = A_j^2$, and the expectation with respect to $\boldsymbol{s}_j$, and use the independence of A_j and $\boldsymbol{s}_j$ together with $\operatorname{tr} A_j^k \le N\sigma^{-2k}$)

$$E\left[\sum_{j=2}^{K} (E_j \alpha_j)^p\right] \le \sum_{j=2}^{K} E\alpha_j^p \le KC_5/N^{p/2} \tag{33}$$

and, similarly,

$$E_{j-1}(E_j \alpha_j)^2 \le C_6 N^{-1}.$$

Combining the above estimates, with $p = 2$, one gets that

$$E(W_1/b_N)^2 \le C_7.$$

The estimate for W_2 is similar, modulo the following auxiliary results, valid for $p = 2, 4$:

$$E(|\hat{\zeta}_j|^p) \le C_7 N^{-p/2} \qquad E(|\zeta_j - \hat{\zeta}_j|^p) \le C_7 N^{-p/2} \tag{34}$$

implying that

$$E(|\zeta_j|^p) \le C_8 N^{-p/2}. \tag{35}$$

To see (34), take in Lemma B.1 $p = 2, 4$, $\boldsymbol{x} = \boldsymbol{s}_j \sqrt{N}$, $D = A_j$, and the expectation with respect to $\boldsymbol{s}_j$, and use the independence of A_j and $\boldsymbol{s}_j$ together with $\operatorname{tr} A_j^k \le N\sigma^{-2k}$ to obtain that

$$E_{j-1}(|\hat{\zeta}_j|^p) \le C_9 N^{-p/2} \tag{36}$$

and hence $E(|\hat{\zeta}_j|^p) \le C_7 N^{-p/2}$. On the other hand, letting for any $j > 2$ $A_{2j} = (A_2^{-1} - \boldsymbol{s}_j^t \boldsymbol{s}_j)^{-1}$, and noting that still $\operatorname{tr} A_{2j}^k \le N\sigma^{-2k}$

$$\begin{aligned}&E|\zeta_j - \hat{\zeta}_j|^p \\ &\quad = E|\zeta_2 - \hat{\zeta}_2|^p \\ &\quad = E\left|\frac{1}{N} \sum_{j=3}^{K} (E_j \operatorname{tr} A_2 - E_{j-1} \operatorname{tr} A_2)\right|^p \\ &\quad = E\left|\frac{1}{N} \sum_{j=3}^{K} (E_j \operatorname{tr}(A_2 - A_{2j}) - E_{j-1} \operatorname{tr}(A_2 - A_{2j}))\right|^p \\ &\quad = E\left|\frac{1}{N} \sum_{j=3}^{K} (E_j - E_{j-1}) \frac{\boldsymbol{s}_j^t A_{2j}^2 \boldsymbol{s}_j}{1 + \boldsymbol{s}_j^t A_{2j} \boldsymbol{s}_j}\right|^p \\ &\quad \le K_p \frac{1}{N^p} E\left|\sum_{j=3}^{K} \left|(E_j - E_{j-1}) \frac{\boldsymbol{s}_j^t A_{2j}^2 \boldsymbol{s}_j}{1 + \boldsymbol{s}_j^t A_{2j} \boldsymbol{s}_j}\right|^2\right|^{p/2} \\ &\quad \le C_{10} \frac{1}{N^{p/2}}\end{aligned}$$

where the Burkholder inequality (29) was used in the next to last step.

Returning to the estimate on W_2/b_N^2, recall that all a_j's are deterministically bounded, uniformly in N. Therefore, using Burkholder's inequality (30) in the first step, and (36) in the third,

$$\begin{aligned} E\left|\sum_{j=2}^{K} E_j(a_j\hat{\zeta}_j)\right|^2 &\le C_{11}\left(E\left(\sum_{j=2}^{K} E_{j-1}(E_j(a_j\hat{\zeta}_j))^2\right)\right. \\ &\qquad \left. + \sum_{j=2}^{K} E|E_j(a_j\hat{\zeta}_j)|^2\right) \\ &\le C_{12}\left(E\left(\sum_{j=2}^{K} E_{j-1}(\hat{\zeta}_j)^2\right)\right. \\ &\qquad \left. + \sum_{j=2}^{K} |E(\hat{\zeta}_j)^2|\right) \\ &\le C_{13}(KN^{-1}) \end{aligned}$$

proving the desired estimate on W_2.

The argument involving W_3 is similar, only simpler: First, note that $\omega_j \boldsymbol{s}_j^t A_j^2 \boldsymbol{s}_j$ is bounded uniformly, and so is ω_j. Therefore, using again Burkholder's inequality (29) in the first step,

$$\begin{aligned} &E(W_3/b_N^2)^2 \\ &= E\left(\sum_{j=2}^{K}(E_j - E_{j-1})(\alpha_j\zeta_j - \boldsymbol{s}_j^t A_j^2 \boldsymbol{s}_j \omega_j \zeta_j^2)\right)^2 \\ &\le K_p E\left(\sum_{j=2}^{K}((E_j - E_{j-1})(\alpha_j\zeta_j - \boldsymbol{s}_j^t A_j^2 \boldsymbol{s}_j \omega_j \zeta_j^2))^2\right) \\ &\le C_{14}\sum_{j=2}^{K}(E(\alpha_j\zeta_j)^2 + E(\zeta_j^4)) \\ &= (K-1)(E(\alpha_2^2\zeta_2^2) + E(\zeta_2^4)). \end{aligned}$$

In view of (35), we need only to recall from (33) that $E(\alpha_2^4) \le C_{15}N^{-p}$. □

Remark: Another possible route to the proof of Lemma 4.3 is as follows. Note that the function $(1 + x/\sigma^2)^{-1}$ is analytic in the (open) disk $|x| < \sigma^2$. For $\sigma^2/P > (1 + \sqrt{\alpha})^2$, this disk includes the support of the limiting measure $G^*(\cdot)$, which is

$$[P(1 - \sqrt{(\alpha)}^2,\, P(1 + \sqrt{\alpha})^2].$$

Expanding in Taylor series the function $(1 + x/\sigma^2)^{-1}$ up to order $k = C \log N$, and controlling the reminder by using the analyticity, it follows that Lemma 4.3 holds as soon as, for some C_1 large enough, $k \le C_1(\log N)$

$$\mathrm{Var}\left(\sum_{i=1}^{N} \frac{\lambda_i^k}{(1 + \sqrt{\alpha})^k}\right) < C_2. \tag{37}$$

(All constants C_i are taken to be independent of N.) Such an estimate is the main result in [20], who under stronger moment assumptions deal with Wigner matrices and not with sample covariance matrices, and show that (37) actually holds for $k = o(\sqrt{N})$. But for sample covariance matrices, one may use the same construction as [20], except that instead of considering paths $i_0 \to i_1 \to i_2 \to \cdots \to i_k$ as in [20], one considers paths $i_0 \to j_0 \to i_1 \to j_1 \to \cdots \to i_k \to j_k$, where $i_m \in \{1, \cdots, N\}$ and $j_m \in \{1, \cdots, K\}$. The parameter s in [20] is replaced by a pair s_1, s_2 with $s_1 + s_2 = s$, keeping the random-walk parameterization as in [20]. We do not see, however, how to extend this argument to the range $\sigma^2 \le P(1 + \sqrt{\alpha})^2$. On the other hand, for $\sigma^2 > P(1 + \sqrt{\alpha})^2$, the argument in [20] actually shows that $N(\beta^{(N)} - \beta^*)$ is asymptotically normal.

We note that for $k = o(\log N/\log\log N)$, the cruder estimates contained in [19, Lemma 1] are enough to yield (37): indeed, for p finite and independent of N this is the content of the proof there, while by bounding there the number of set partitions of $\{i_1, \cdots, i_r, i_1', \cdots, i_r'\}$, $\{k_1, \cdots, k_r, k_1', \cdots, k_r'\}$ one extends the conclusion to k's as above. Unfortunately, this technique seems to break down when $k = O(\log N)$.

We finally note that besides the condition $E(v_{ij}^4) < \infty$, the assumption $E(v_{ij}^8) < \infty$ was used only in bounding W_3; a tightening of this argument, valid under the condition $E(v_{ij}^4) < \infty$, seems possible, following [2, Sec. 4], but we do not pursue this direction here.

Appendix C
Proof of Lemma 4.4

Define

$$S_i := [\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_{i-1}, \boldsymbol{s}_{i+1}, \cdots, \boldsymbol{s}_K]$$

and

$$S := [\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_K].$$

As in Appendix B, we will for simplicity take $P = 1$. The general result follows from a rescaling of σ^2. Now

$$\beta_i^{(N)} = \boldsymbol{s}_i^t(S_iS_i^t + \sigma^2 I)^{-1}\boldsymbol{s}_i \tag{38}$$

is the SIR attained by user i. In [22, Eq. (27)], a key equation relating the achieved SIR's of the users and the trace of $(SS^t + \sigma^2 I)^{-1}$ was derived

$$\frac{1}{N}\sum_{i=1}^{K}\frac{\beta_i^{(N)}}{1 + \beta_i^{(N)}} = 1 - \frac{\sigma^2}{N}\,\mathrm{tr}\,(SS^t + \sigma^2 I)^{-1}. \tag{39}$$

It follows from Lemma 3.2 that

$$\overline{\beta}_i^{(N)} := E(\beta_i^{(N)}) = E[N^{-1}\,\mathrm{tr}\,(S_iS_i^t + \sigma^2 I)^{-1}].$$

Let

$$\overline{\beta}^{(N)} := E[N^{-1}\,\mathrm{tr}(SS^t + \sigma^2 I)^{-1}].$$

It follows from Lemmas 4.2 and 4.3 that for large N, each of the $\beta_i^{(N)}$ is close to $\overline{\beta}_i^{(N)}$, which in turn is close to $\overline{\beta}^{(N)}$. Moreover, $N^{-1}\,\mathrm{tr}\,(SS^t + \sigma^2 I)^{-1}$ is also close to $\overline{\beta}^{(N)}$. Substituting these approximations into (39) gives us an approximate fixed-point equation in $\overline{\beta}^{(N)}$

$$\alpha\frac{\overline{\beta}^{(N)}}{1 + \overline{\beta}^{(N)}} \approx 1 - \sigma^2\overline{\beta}^{(N)}.$$

The *exact* fixed-point equation has a unique positive solution, which is precisely the limiting value β^*. (In fact, the formula (12) for β^* is obtained by solving this quadratic equation.) Thus to estimate how far $\overline{\beta}^{(N)}$ is from β^*, we need estimates on how far each of the $\beta_i^{(N)}$ deviates is from $\overline{\beta}^{(N)}$. This is the main idea of the following development.

One can write

$$\overline{\beta}_i^{(N)} = \overline{\beta}^{(N)} + \delta_i^{(N)}$$
$$\delta_i^{(N)} = \frac{1}{N} E[\mathrm{tr}\,(S_i S_i^t + \sigma^2 I)^{-1} - \mathrm{tr}\,(SS^t + \sigma^2 I)^{-1}].$$

By the matrix-inversion lemma

$$\begin{aligned}(S_i S_i^t + \sigma^2 I)^{-1} &- (SS^t + \sigma^2 I)^{-1} \\ &= (S_i S_i^t + \sigma^2 I)^{-1} \boldsymbol{s}_i \boldsymbol{s}_i^t (SS^t + \sigma^2 I)^{-1}\end{aligned}$$

so

$$\begin{aligned}|\delta_i^{(N)}| &= \left| \frac{1}{N} E[\mathrm{tr}\,((S_i S_i^t + \sigma^2 I)^{-1} \boldsymbol{s}_i \boldsymbol{s}_i^t (SS^t + \sigma^2 I)^{-1})] \right| \\ &= \frac{1}{N} E[\mathrm{tr}\, \boldsymbol{s}_i^t (SS^t + \sigma^2 I)^{-1} (S_i S_i^t + \sigma^2 I)^{-1} \boldsymbol{s}_i)] \\ &\le \frac{1}{N} E\left[\frac{1}{\sigma^4} \|\boldsymbol{s}_i\|^2 \right] \\ &= \frac{1}{\sigma^4 N}. \end{aligned} \tag{40}$$

Also

$$\begin{aligned}\mathrm{Var}\,[\beta_i^{(N)}] &= E\left[\left(\beta_i^{(N)} - N^{-1}\,\mathrm{tr}\,(S_i S_i^t + \sigma^2 I)^{-1} \right.\right. \\ &\qquad \left.\left. + N^{-1}\,\mathrm{tr}\,(S_i S_i^t + \sigma^2 I)^{-1} - \overline{\beta}_i^{(N)} \right) \right] \\ &= E\left[\left(\beta_i^{(N)} - \mathrm{tr}\,(S_i S_i^t + \sigma^2 I)^{-1} \right)^2 \right] \\ &\quad + 2E\left[\left(\beta_i^{(N)} - N^{-1}\,\mathrm{tr}\,(S_i S_i^t + \sigma^2 I)^{-1} \right) \right. \\ &\qquad \left. \cdot \left(N^{-1}\,\mathrm{tr}\,(S_i S_i^t + \sigma^2 I)^{-1} - \overline{\beta}_i^{(N)} \right) \right] \\ &\quad + E\left[\left(N^{-1}\,\mathrm{tr}\,(S_i S_i^t + \sigma^2 I)^{-1} - \overline{\beta}_i^{(N)} \right)^2 \right].\end{aligned}$$

By Lemma 3.2, the first term of the preceding expression is bounded by $C_1 \sigma^4 / N$ for some constant C_1 that depends only on the fourth moment of v_{11}, and the second term is 0. By Lemma 4.3, the third term is bounded by C_2/N^2 for some constant C_2 independent of N and i. Hence $\mathrm{Var}\,[\beta_i^{(N)}] \le C_3/N$ for some constant C_3 independent of N and i.

Combining this with (40), we can now write

$$\beta_i^{(N)} = \overline{\beta}^{(N)} + \Delta_i^{(N)} \tag{41}$$

where

$$|E[\Delta_i^{(N)}]| \le \frac{1}{\sigma^4 N} \tag{42}$$

and

$$E[(\Delta_i^{(N)})^2] \le \frac{C_4}{N}$$

for some constant C_4 independent of N and i.

Substituting (41) into (39)

$$\frac{K}{N} - \frac{1}{N} \sum_{i=1}^{K} \frac{1}{1 + \overline{\beta}^{(N)} + \Delta_i^{(N)}} = 1 - \frac{\sigma^2}{N}\,\mathrm{tr}\,(SS^t + \sigma^2 I)^{-1}. \tag{43}$$

Let $\nu_i^{(N)} := \Delta_i^{(N)} / (1 + \overline{\beta}^{(N)})$. Then

$$\begin{aligned}\frac{1}{1 + \overline{\beta}^{(N)} + \Delta_i^{(N)}} &= \frac{1}{1 + \overline{\beta}^{(N)}} \frac{1}{1 + \nu_i^{(N)}} \\ &= \frac{1}{1 + \overline{\beta}^{(N)}} \left(1 - \nu_i + \nu_i^2 \frac{1}{(1 + \xi_i)^3} \right)\end{aligned}$$

for some $\xi_i := \xi_i(\nu_i^{(N)})$ satisfying $\xi_i \in [0, \nu_i^{(N)}] \cup [\nu_i^{(N)}, 0]$. Note that

$$\begin{aligned}\nu_i^{(N)} &= \frac{\beta_i^{(N)} - \overline{\beta}^{(N)}}{1 + \overline{\beta}^{(N)}} \\ &\ge -1 + \frac{1}{1 + \overline{\beta}^{(N)}}, \qquad \text{since } \beta_i^{(N)} \ge 0 \\ &\ge -1 + \frac{\sigma^2}{1 + \sigma^2}, \qquad \text{since } \overline{\beta}^{(N)} \le 1/\sigma^2.\end{aligned}$$

Hence $(1 + \xi_i)^{-1} \le C_5$ for some deterministic constant C_5 independent of N. Substituting in (43) and taking expectations, using the fact from (42) that

$$|E[\Delta_i^{(N)}]| \le \frac{1}{\sigma^4 N}$$

and that

$$E[N^{-1}\,\mathrm{tr}\,(SS^t + \sigma^2 I)^{-1}] = \overline{\beta}^{(N)},$$

we get, for some C_6 independent of N

$$\left| \frac{K}{N} - \frac{K}{N(1 + \overline{\beta}^{(N)})} - 1 + \sigma^2 \overline{\beta}^{(N)} \right| \le \frac{C_6}{N}$$

and hence, for some C_7 independent of N

$$\left| \alpha - \frac{\alpha}{1 + \overline{\beta}^{(N)}} - 1 + \sigma^2 \overline{\beta}^{(N)} \right| \le \frac{C_7}{N}.$$

But $\beta^* := \int (\lambda + \sigma^2)^{-1}\, dG^*(\lambda)$ is the unique solution of the equation

$$\alpha - \frac{\alpha}{1 + \beta} - 1 + \sigma^2 \beta = 0$$

and, moreover, it can be easily seen that the solution of this equation is a differentiable function of the right-hand side at 0. Hence

$$\limsup_{N \to \infty} N \left(\overline{\beta}^{(N)} - \beta^* \right) < \infty.$$

The lemma now follows from (40). □

ACKNOWLEDGMENT

The authors wish to thank Kiran for help on the simulations. D. N. C. Tse would like to thank Prof. J. Salz for his discussions in the initial phase of this work. The authors would also like to thank Prof. J. Silverstein for pointing out that the proof of Lemma 4.3 is contained in [2], Prof. S. Shamai for the reference to [14], and P. Viswanath for detecting a mistake in an earlier version of this paper.

REFERENCES

[1] Z. D. Bai and Y. Q. Yin, “Limit of the smallest eigenvalue of a large dimensional sample covariance matrix,” *Ann. Probab.*, vol. 21, pp. 1275–1294, 1993.

[2] 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.

[3] P. J. Bickel and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*. San Francisco, CA: Golden Day, 1977.

[4] D. L. Burkholder, "Distribution function inequalities for martingales," *Ann. Probab.*, vol. 1, pp. 19–42, 1973.

[5] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*. New York, NY: Wiley, 1986.

[6] M. Honig, U. Madhow, and S. Verdú, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, pp. 944–960, July 1995.

[7] M. Honig and W. Veerakachen, "Performance variability of linear multiuser detection for DS-CDMA," in *Proc. Vehicular Technology Conf.* , 1996.

[8] J. B. Kim and M. Honig, "Outage Probability of Multicode DS-CDMA with Linear Interference Suppression," presented at MILCOM, to be published.

[9] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple access," *IEEE Trans. Inform. Theory*, vol. 35, pp. 123–136, Jan. 1989.

[10] ——, "Near–far resistance of multiuser detectors in asynchronous channels," *IEEE Trans. Commun.*, vol. 38, pp. 496–508, Apr. 1990.

[11] U. Madhow and M. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 42, pp. 3178–3188, Dec. 1994.

[12] ——, "On the average near–far resistance for MMSE detection of direct-sequence CDMA signalswith random spreading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2039–2045, Sept. 1999.

[13] V. A. Marcenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math. USSR-Sb*, vol. 1, pp. 457–483, 1967.

[14] R. R. Müller, P. Schramm, and J. B. Huber, "Spectral efficiency of CDMA systems with linear interference suppression," in *IEEE Workshop on Communication Engineering*, Ulm, Germany, Jan. 1997, pp. 93–97.

[15] P. Rapajic and B. Vucetic, "Adaptive receiver structures for asynchronous CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 685–697, May 1994.

[16] M. Rupf, F. Tarkoy, and J. Massey, "User-separating demodulation for code-division multiple access systems," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 786–795, June 1994.

[17] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *J. Multivariate Anal.*, vol. 54, no. 2, pp. 175–192, 1995.

[18] J. W. Silverstein, "Weak convergence of random functions defined by the eigenvectors of sample covariance matrices," *Ann. Probab.*, vol. 18, pp. 1174–1194, 1990.

[19] ——, "On the eigenvectors of large dimensional sample covariance matrices," *J. Multivariate Anal.*, vol. 30, pp. 1–16, 1989.

[20] Y Sinai and A. Soshnikov, "Central limit theorem for traces of large random symmetric matrices with independent matrix elements," *Bol. Soc. Mat. Brasil. (NS)*, vol. 29, pp. 1–24, 1998.

[21] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs., Memo, 1995.

[22] D. Tse and S. V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inform. Theory*, vol. 45, pp. 641–657, Mar. 1999.

[23] S. Verdú, "Minimum probability of error for asynchronous Gaussian channels," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 85–96, Jan. 1986.

[24] ——, "Optimum multiuser asymptotic efficiency," *IEEE Trans. Commun.*, vol. COM-34, pp. 890–897, Sept. 1986.

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

[26] J. H. Winters, J. Salz, and R. Gitlin, "The impact of antenna diversity on the capacity of wireless communications systems," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1740–1751, 1994.

[27] Z. Xie, R. Short, and C. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 683–690, May 1990.