Asymptotic Performance of MMSE Receivers for Large Systems Using Random Matrix Theory

Ying-Chang Liang, *Senior Member, IEEE*, Guangming Pan, and Z. D. Bai

***Abstract*—Random matrix theory is used to derive the limit and asymptotic distribution of signal-to-interference-plus-noise ratio (SIR) for a class of suboptimal minimum mean-square-error (MMSE) receivers applied to large random systems with unequal-power users. We prove that the limiting SIR converges to a deterministic value when K and N go to infinity with $\lim K/N = y$ being a positive constant, where K is the number of users and N is the number of degrees of freedom. We also prove that the SIR of each particular user is asymptotically Gaussian for large N and derive the closed-form expressions of the variance for the SIR variable under real-spreading and complex-spreading channel environments. It is revealed that for a given (K, N) pair, under certain mild conditions, the variance of the SIR for complex-spreading channels is half of that for the corresponding real-spreading channels. Since the suboptimal MMSE receiver becomes optimal for the case when the users are equally powered, our results show that the conjecture made by Tse and Zeitouni for the complex-spreading case is not affirmative. We also derive the asymptotic distribution for SIR in decibels which provides better description when N is small. Numerical results and computer simulations are provided to evaluate the accuracy of various limiting and asymptotic results obtained in this paper.**

***Index Terms*—Code-division multiple access (CDMA), large systems, minimum mean-square-error (MMSE), multiple-input multiple-output (MIMO), random matrix theory, random spreading.**

I. Introduction

RANDOM matrix theory (RMT) has found wide applications in wireless communications [3], [12], [14]. Using limiting spectral properties of large random matrices, it was shown that for frequency-flat, synchronous code-division multiple access (CDMA) uplink with random spreading codes, the output signal-to-interference-plus-noise ratios (SIRs) using well-known linear receivers such as matched filter, decorrelator, and minimum mean-square-error (MMSE) receiver, converge to deterministic values for large systems, i.e., when both spreading gain and number of users go to infinity, and their ratio goes to a deterministic constant. Another scenario of interest is the multiple-input multiple-output (MIMO) antenna systems. For a rich multipath environment, the channel responses between the transmitters and the receivers are independent and identically distributed (i.i.d.), thus the propagation channel can be modeled as a random matrix. Inspired by the fundamental discoveries, in recent years, researchers have exploited the applications of RMT in various aspects of wireless communications, including, e.g., i) limiting capacity and asymptotic capacity distribution for random MIMO channels; ii) limiting SIR analysis for linearly precoded systems, such as the multicarrier CDMA using linear receivers [2]; iii) limiting SIR analysis for random channels with interference cancellation receivers [7], [8], and iv) limiting SIR analysis for coded multiuser systems [17]. Other applications include the design of receivers, such as the reduced-rank MMSE receiver [9], the asymptotic normality study for multiple-access interference (MAI) [16], and linear receiver output [4]. An excellent overview of applications of RMT in wireless communications is given by Tulino and Verdú in [11].

In this paper, we are concerned with the limiting and asymptotic performance of linear receivers for large random systems with unequal-power users. One typical example is the CDMA uplink where each user has different received power at the base station due to multipath fading or shadowing and imperfect power control. Another example is the MIMO systems in which soft interference cancellation (SIC)-based MMSE receiver [15] is applied. While each user may have identical average received powers, since the soft estimates of the transmitted symbols may vary from one user to another, for the purpose of detecting one particular user, the powers of the interferers after SIC will usually not be identical. We consider two classes of MMSE receivers: i) the optimal MMSE receiver which requires the knowledge of the instantaneous power profile for all users, and ii) the suboptimal MMSE receiver which treats the instantaneous powers of all users as being equal. While the optimal MMSE receiver yields the highest SIR among all linear receivers, in practice, it could be difficult to obtain the instantaneous power profile. Furthermore, for SIC-based MMSE receiver, if the power profile changes, the MMSE weighting vector has to be recalculated. Therefore, the suboptimal MMSE receiver is of great importance from the viewpoint of practical applicability.

In this paper, we derive the limiting SIRs and asymptotic SIR distributions for a class of suboptimal MMSE receivers. We prove that the limiting SIR converges to a deterministic value when the number of users and number of degrees of freedom go to infinity, with their ratio being a positive constant. We also prove that the SIR of each particular user is asymptotically Gaussian when the number of degrees of freedom is relatively

Manuscript received May 26, 2005; revised April 10, 2007. The work of Z. D. Bai was supported in part by CNSF 1057-1020 and NUS under Grant R-155-000-056-112.

Y.-C. Liang is with the Institute for Infocomm Research, Singapore 119613 (e-mail: ycliang@i2r.a-star.edu.sg).

G. Pan is with the Department of Statistics and Finance, University of Science and Technology of China, Hefei, China (e-mail: stapgm@gmail.com).

Z. D. Bai is with the KLASMOE, School of Mathematics and Statistics, Northeast Normal University, Changchun, 130024, China (e-mail: baiz@nenu.edu.cn).

Communicated by X. Wang, Associate Editor for Detection and Estimation.

Color versions of Figures 5 and 6 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2007.907497

large, and derive the closed-form expressions for the variance of the SIR variable under real-spreading and complex-spreading channel environments. Based on the limiting SIR results obtained for the general suboptimal MMSE receivers, we derive the optimal noise variance with which the suboptimal MMSE receiver provides the highest limiting SIR within this category of the suboptimal MMSE receivers. We also quantify the mismatch effect when the noise power is overestimated or underestimated in designing the suboptimal MMSE receiver. Furthermore, based on the asymptotic distribution of SIR derived in this paper and using the Taylor series expansion, we also establish the asymptotic normality of SIR in decibels, denoted as SIRdB, and derive the closed-form expressions for the mean and variance of this distribution. This derivation is useful as it provides better description on the distribution especially when the number of degrees of freedom is relatively small.

In [13], Tse and Zeitouni have proven the asymptotic normality of the output SIR for the optimal MMSE receiver for large systems when the users are equally powered. They have also derived rigorously the closed-form expression for the variance of SIR distribution using the central limit theorem for the real-spreading channel scenario, and provided a conjecture for the expression of variance for the complex-spreading channel scenario. Since the system model considered in this paper includes the equal-power case as a special case, based on the results obtained in this paper, under certain mild conditions, it is recognized that the asymptotic SIR variance for the complex-spreading channel scenario is just half of that for the corresponding real-spreading channel scenario. Our derived results thus show that the conjecture made in [13] is not affirmative.

Unequal-power large systems are considered in [17] where the limiting SIR of the mismatched MMSE receiver has been derived. While the suboptimal MMSE receiver considered in this paper can be considered as a special case of the mismatched MMSE receiver, the asymptotic SIR distribution of the receiver has not been studied in [17]. In [18], the authors use generalized Gamma distribution to describe the output SIR of the optimal MMSE receiver. This is derived by decoupling the output SIR of the receiver into two independent parts: one part related to the zero-forcing (ZF) receiver and the other related to the difference between the SIRs for the MMSE and ZF receivers. The unequal-power systems have also been studied in [4] using the random matrix method for various linear polynomial receivers, and in [6] using the replica method. The work in [6] extends the results of [5] where equal-power systems are considered.

This paper is organized as follows. Section II presents the system model and reviews the optimal MMSE receiver. Section III summarizes the main results derived in this paper for the suboptimal MMSE receiver, including the limiting SIR, optimal noise variance selection, asymptotic normality of the SIR, and the closed-form expressions for the variances of the asymptotic distributions for both real-spreading and complex-spreading channel cases. We also establish the asymptotic normality of SIRdB in this section. Numerical results and simulation comparisons are given in Section IV. Finally, conclusions are drawn in Section V.

In this paper, we use the following notations: $(\cdot)^T$ for transpose; $(\cdot)^\dagger$ for conjugate and transpose; $\mathsf{E}[\cdot]$ for expectation; $\mathsf{Tr}(\cdot)$ for trace; $\mathrm{Var}(\cdot)$ for variance; $\|\cdot\|$ for the spectral norm of a matrix or the Euclidean norm of a vector; $\mathrm{diag}(\boldsymbol{x})$ for a diagonal matrix with diagonal elements being vector $\boldsymbol{x}$.

II. Channel Model and MMSE Receiver

Let us consider the following system model:

$$\boldsymbol{x} = \sum_{i=1}^{K} \boldsymbol{h}_i s_i + \boldsymbol{\epsilon} \tag{1}$$

where K is the number of users, s_i is the transmitted symbol of user i, $\boldsymbol{h}_i \in \mathcal{C}^{N\times 1}$ is the channel vector of user i, $\boldsymbol{\epsilon} \in \mathcal{C}^{N\times 1}$ is the additive white Gaussian noise vector, and $\boldsymbol{x} \in \mathcal{C}^{N\times 1}$ is the received signal vector. The following assumptions are made throughout this paper.

(AS1) The channel vectors $\boldsymbol{h}_i$'s can be represented as

$$\boldsymbol{h}_i = \frac{1}{\sqrt{N}}[X_{1i}, X_{2i}, \ldots, X_{Ni}]^T \tag{2}$$

for $i = 1, \ldots, K$, where X_{ki}'s are i.i.d. variables with zero mean and unit variance, i.e., $\mathsf{E}[|X_{ki}|^2] = 1$ for all k's and i's.

(AS2) The transmitted symbols s_i's are i.i.d. variables with zero mean and average power $\mathsf{E}\left[|s_i|^2\right] = P_i$, for $i = 1, \ldots, K$.

(AS3) The noise vector $\boldsymbol{\epsilon}$ is i.i.d., zero mean, circularly symmetric complex Gaussian and with covariance matrix $\mathsf{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\dagger] = \tilde{\sigma}^2 I$.

Remark 2.1: For frequency-flat CDMA uplink with perfect power control, s_i and $\boldsymbol{h}_i$ denote the transmitted symbol and spreading sequence of user i, respectively. Thus, K and N denote the number of active users and processing gain, respectively. For third-generation wideband CDMA, the uplink spreading codes use random codes, thus the propagation channel can be modeled as a random matrix channel. The channel model (1) can also be used to represent MIMO antenna systems, where s_i and $\boldsymbol{h}_i$ denote, respectively, the transmitted symbol through the ith transmit antenna, and the channel responses from the ith transmit antenna to all receive antennas. In this case, K and N represent the transmit antenna number and receive antenna number, respectively. When there are rich local scatters around the transmitter and receiver sides, $\boldsymbol{h}_i$ can be modeled as an i.i.d. vector, thus channel (1) becomes a random matrix channel.

Remark 2.2: The channel coefficients can be real or complex. For *complex-spreading case*, we assume that the real and imaginary parts of each channel coefficient are independent, zero mean and with equal variance.

Remark 2.3: The use of normalization factor $1/\sqrt{N}$ in (2) implies that the signal-to-noise ratio (SNR), $\gamma_i = \frac{P_i}{\tilde{\sigma}^2}$, is defined as the average received SNR for user i over all degrees of freedom.

Rewrite (1) as

$$\boldsymbol{x} = H\boldsymbol{s} + \boldsymbol{\epsilon} \tag{3}$$

where $H = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]$ and $\boldsymbol{s} = [s_1, \ldots, s_K]^T$. We are then interested in recovering the transmitted symbols $s_1, \ldots, s_K$ from the received signal vector $\boldsymbol{x}$.

For linear receivers, the equalization output of user k is given by

$$\tilde{s}_k = \boldsymbol{w}_k^\dagger \boldsymbol{x} \tag{4}$$

where $\boldsymbol{w}_k \in \mathcal{C}^{N\times 1}$ is the weighting vector for user k. Using (1), $\tilde{s}_k$ can be represented as

$$\tilde{s}_k = \boldsymbol{w}_k^\dagger \boldsymbol{h}_k s_k + \sum_{j\neq k} \boldsymbol{w}_k^\dagger \boldsymbol{h}_j s_j + \boldsymbol{w}_k^\dagger \boldsymbol{\epsilon}. \tag{5}$$

Thus, the equalization output consists of two components: i) the desired signal component $\boldsymbol{w}_k^\dagger \boldsymbol{h}_k s_k$ and ii) the interference-plus-noise component $\sum_{i\neq k} \boldsymbol{w}_k^\dagger \boldsymbol{h}_i s_i + \boldsymbol{w}_k^\dagger \boldsymbol{\epsilon}$.

Denote $\alpha_k = \boldsymbol{w}_k^\dagger \boldsymbol{h}_k$, and write σ_k^2 as the variance of the interference-plus-noise component. One measure for quantifying the receiver performance is the output SIR of the equalizer, which is given by

$$\mathrm{SIR}_k = \frac{|\alpha_k|^2 P_k}{\sigma_k^2}. \tag{6}$$

Notice that

$$\sigma_k^2 = \boldsymbol{w}_k^\dagger \left(H_k \tilde{D}_k H_k^\dagger + \tilde{\sigma}^2 I \right) \boldsymbol{w}_k \tag{7}$$

where $\tilde{D}_k = \mathrm{diag}(P_1, \ldots, P_{k-1}, P_{k+1}, \ldots, P_K)$, and $H_k = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{k-1}, \boldsymbol{h}_{k+1}, \ldots, \boldsymbol{h}_K]$. We derive

$$\mathrm{SIR}_k = \frac{|\boldsymbol{w}_k^\dagger \boldsymbol{h}_k|^2 P_k}{\boldsymbol{w}_k^\dagger \left(H_k \tilde{D}_k H_k^\dagger + \tilde{\sigma}^2 I \right) \boldsymbol{w}_k}. \tag{8}$$

Define the mean-square-error (MSE) function for user k as

$$J(\boldsymbol{w}_k) = \mathsf{E}\left[\left| s_k - \boldsymbol{w}_k^\dagger \boldsymbol{x} \right|^2 \right].$$

The MMSE receiver, which minimizes the MSE function $J(\boldsymbol{w}_k)$, is represented as

$$\boldsymbol{w}_k = (H\tilde{D}H^\dagger + \tilde{\sigma}^2 I)^{-1} \boldsymbol{h}_k P_k \tag{9}$$

where $\tilde{D} = \mathrm{diag}(P_1, \ldots, P_K)$. Let $A = H\tilde{D}H^\dagger + \tilde{\sigma}^2 I$ and $C_k = H_k \tilde{D}_k H_k^\dagger + \tilde{\sigma}^2 I$. Using matrix inversion lemma and noticing that $A = C_k + P_k \boldsymbol{h}_k \boldsymbol{h}_k^\dagger$, we have

$$\boldsymbol{w}_k = \frac{C_k^{-1} \boldsymbol{h}_k}{1/P_k + \boldsymbol{h}_k^\dagger C_k^{-1} \boldsymbol{h}_k}. \tag{10}$$

The MMSE receiver also maximizes the output SIR which is given by

$$\mathrm{SIR}_k^{(o)} = \boldsymbol{h}_k^\dagger C_k^{-1} \boldsymbol{h}_k P_k. \tag{11}$$

III. Main Results

The MMSE receiver shown in (9) is referred to as the optimal MMSE receiver which produces the highest output SIR among all linear receivers. The optimal MMSE receiver however requires the instantaneous power profile of the transmitted symbols. If the power profile changes, the MMSE weighting vector has to be recalculated. A practical solution is to treat all transmitted symbols as being equally powered when designing the receiver. This receiver is called suboptimal MMSE receiver in this work. The weighting vector of suboptimal MMSE receiver for user k is then given by

$$\boldsymbol{w}_k = \left(H_k H_k^\dagger + \mu^2 I \right)^{-1} \boldsymbol{h}_k. \tag{12}$$

Note that if $P_1 = P_2 = \cdots = P_N = P$ and $\mu^2 = \tilde{\sigma}^2/P$, then the weighting vector in (12) is actually a scaler version of the weighting vector given in (10). For unequal-power systems, the selection of parameter μ^2 will be discussed later.

With the suboptimal receiver, from (6), the output SIR for user k can be written as (13) shown at the bottom of the page, where $p_k = P_k/\bar{P}_k$,

$$D_k = \mathrm{diag}(P_1/\bar{P}_k, \ldots, P_{k-1}/\bar{P}_k, P_{k+1}/\bar{P}_k, \ldots, P_K/\bar{P}_k)$$

with $\bar{P}_k = \frac{1}{K-1}\sum_{i\neq k} P_i$ and $\sigma^2 = \tilde{\sigma}^2/\bar{P}_k$. Note that $(1/(K-1))\mathsf{Tr}(D_k) = 1$, and p_k denotes the ratio between the desired user's power to the average value of the interference powers. In the sequel, we will derive the limiting SIR and asymptotic SIR distribution for the suboptimal MMSE receiver. Without loss of generality, we only consider the detection of user 1, and assume that $p_1 = 1$. If $p_1 \neq 1$, the limiting SIR will be scaled by a factor of p_1, while the variance of the asymptotic distribution of the SIR will be scaled by a factor of p_1^2.

A. Limiting SIR Analysis

In this subsection, we derive the limiting SIR for the suboptimal MMSE receiver when $N \to \infty$ and $K \to \infty$ in such a way that $\frac{K}{N}$ approaches a positive constant y. To simplify the notation, we denote $B_1 = H_1 H_1^\dagger + \mu^2 I$. Before stating the main theorems for the suboptimal MMSE receiver, we briefly state the limiting SIR results for equal-power systems where $D_k = I$.

Proposition 3.1 [12], [14]: Suppose that

1) for all i and j, X_{ij}'s are i.i.d. random variables with $\mathsf{E}[X_{11}] = 0$ and $\mathsf{E}[|X_{11}|^2] = 1$; and
2) $\frac{K}{N} \to y > 0$ (y is a constant) as $N \to \infty$.

Then, with probability 1, the following SIR:

$$\gamma = \boldsymbol{h}_1^\dagger B_1^{-1} \boldsymbol{h}_1 \tag{14}$$

converges to the deterministic value

$$\beta = \int \frac{dF_y(x)}{x + \mu^2} \tag{15}$$

$$\mathrm{SIR}_k^{(s)} = \frac{p_k \left| \boldsymbol{h}_k^\dagger \left(H_k H_k^\dagger + \mu^2 I \right)^{-1} \boldsymbol{h}_k \right|^2}{\boldsymbol{h}_k^\dagger \left(H_k H_k^\dagger + \mu^2 I \right)^{-1} \left(H_k D_k H_k^\dagger + \sigma^2 I \right) \left(H_k H_k^\dagger + \mu^2 I \right)^{-1} \boldsymbol{h}_k} \tag{13}$$

where $F_y(x)$ is the limiting spectral distribution function of the random matrix $HH^\dagger$.

According to [12], [14], and [31], β is the Stieltjes transform of the Marčhenko–Pastur law at $-\mu^2$, that is, the positive solution of the equation

$$y - 1 - \frac{y}{1+\beta} + \mu^2\beta = 0. \tag{16}$$

Equation (16) is also equivalent to the following equation which will be used frequently in the sequel:

$$\frac{1}{\beta} = y - 1 + \mu^2(1+\beta) = \frac{y}{1+\beta} + \mu^2. \tag{17}$$

From (16), one can readily derive that $F_y(x)$ has a density function

$$F_y'(x) = \frac{1}{2\pi x}\sqrt{(b-x)(x-a)}$$

where $b = (1+\sqrt{y})^2$ and $a = (1-\sqrt{y})^2$. Also, $F_y'(x)$ has a point mass $(1-y)$ at the origin if $y < 1$.

We now provide the limiting theorem for the output SIR of the suboptimal MMSE receiver.

Theorem 3.1: Suppose the following conditions hold.

1) For all i and j, X_{ij}'s are i.i.d. random variables with $\mathsf{E}[X_{11}] = 0$, $\mathsf{E}[|X_{11}|^2] = 1$, and $\mathsf{E}[|X_{11}|^4] < \infty$;
2) $\frac{K}{N} \to y > 0$ as $N \to \infty$; and
3) the powers of all users are i.i.d. with a fixed distribution and are uniformly bounded, that is, there exists a constant M such that $p_i \le M$.

Then the $\mathrm{SIR}_1^{(s)}$ in (13), denoted by $\gamma_1^{(s)}$, converges almost surely to a deterministic constant $\beta^{(s)}$ given by

$$\beta^{(s)} = \beta\mathcal{K}, \tag{18}$$

where

$$\beta = \frac{-\mu^2 - y + 1 + \sqrt{(\mu^2+y+1)^2 - 4y}}{2\mu^2} \tag{19}$$

and

$$\mathcal{K} = \frac{y + \mu^2(1+\beta)^2}{y + \sigma^2(1+\beta)^2}. \tag{20}$$

Proof: The proof is given in Appendix A.

Theorem 3.1 tells us that the limiting SIR is independent of the actual power distribution if the suboptimal MMSE receiver assumes equal powers for all users. The parameter $\mathcal{K}$ in (20) can be thought of as a mismatch factor which defines the difference between the true noise variance σ^2 and selected noise variance μ^2 in designing the receiver. For equal-power systems, the weighting vector in (12) becomes optimal if we choose $\mu^2 = \sigma^2$. In this case, $\mathcal{K} = 1$ and $\beta^{(s)} = \beta$. This result is consistent with the limiting result derived in [12] using the random matrix method, and those derived in [5] and [6] using the replica method.

Differentiating $\beta^{(s)}$ in (18) with respect to the variable μ^2, and setting $\frac{d\beta^{(s)}}{d\mu^2} = 0$, we get

$$\left[\frac{y}{(1+\beta)^2} + \mu^2\right]\left[\frac{y}{(1+\beta)^2} + \sigma^2\right] + \frac{2y\beta}{(1+\beta)^3}(\mu^2 - \sigma^2) + \frac{\beta}{\beta'}\left[\frac{y}{(1+\beta)^2} + \sigma^2\right] = 0. \tag{21}$$

From (16), the derivative of β over μ^2 can be calculated as follows:

$$-\frac{\beta}{\beta'} = \frac{y}{(1+\beta)^2} + \mu^2. \tag{22}$$

Applying (22) into the first term of (21), we observe that the extreme value of $\beta^{(s)}$ is taken only at point $\mu^2 = \sigma^2$. From (22) it is also noticed that β' is negative, thus the maximum limiting SIR is achieved when μ^2 is chosen as σ^2. Based on these arguments, we have the following theorem.

Theorem 3.2: Under the conditions of Theorem 3.1, the limiting SIR $\beta^{(s)}$ is maximized when $\mu^2 = \sigma^2$.

B. Asymptotic Distribution of the SIR

We next establish the asymptotic normality of $\gamma_1^{(s)}$ for which the following lemmas are needed.

Lemma 3.1: Under the assumptions of Theorem 3.1, we have

$$\max_{j\le N}\left|\left[(HH^\dagger + \mu^2 I)^{-1}\right]_{jj} - \beta\right| \to 0 \quad \text{a.s.} \tag{23}$$

and

$$\max_{j\le N}\left|\left[(HH^\dagger + \mu^2 I)^{-1}(HDH^\dagger + \sigma^2 I)(HH^\dagger + \mu^2 I)^{-1}\right]_{jj} - \frac{\beta}{\mathcal{K}}\right| \to 0 \quad \text{a.s.} \tag{24}$$

where $(A)_{jj}$ denotes the jth diagonal element of matrix A, and a.s. denotes "almost surely."

Proof: See Appendix B.

Lemma 3.2: Under the assumptions of Theorem 3.1, with probability 1, we get equation (25)–(26) at the bottom of the page and (27) at the top of the following page.

$$\max_{j\le N}\left|\frac{1}{N}\sum_{j=1}^{N}\left\{\left[(HH^\dagger + \mu^2 I)^{-1}\right]_{jj}\right\}^2 - \beta^2\right| \to 0 \tag{25}$$

$$\max_{j\le N}\left|\frac{1}{N}\sum_{j=1}^{N}\left\{\left[(HH^\dagger + \mu^2 I)^{-1}(HDH^\dagger + \sigma^2 I)(HH^\dagger + \mu^2 I)^{-1}\right]_{jj}\right\}^2 - \frac{\beta^2}{\mathcal{K}^2}\right| \to 0 \tag{26}$$

$$\max_{j\le N}\left|\frac{1}{N}\sum_{j=1}^{N}\left\{\left[(HH^\dagger+\mu^2I)^{-1}\right]_{jj}\left[(HH^\dagger+\mu^2I)^{-1}(HDH^\dagger+\sigma^2I)(HH^\dagger+\mu^2I)^{-1}\right]_{jj}\right\}-\frac{\beta^2}{\mathcal{K}}\right|\to 0. \tag{27}$$

Proof: Lemma 3.2 comes immediately from Lemma 3.1.

Lemma 3.3: In addition to the conditions of Theorem 3.1, we assume that the empirical distribution of $\{p_2,\dots,p_K\}$ tends to a limiting distribution $F_p(x)$. Then, as $N\to\infty$

$$\frac{1}{N}\mathsf{Tr}\left[H_1D_1H_1^\dagger B_1^{-2}H_1D_1H_1^\dagger B_1^{-2}\right]\xrightarrow{a.s.} m \tag{28}$$

where

$$\begin{aligned} m=&\frac{y^2}{(1+\beta)^4}\left[\frac{4}{(1+\beta)^2}\left(\int\frac{dF_y(x)}{(x+\mu^2)^2}\right)^3\right.\\ &-\frac{4}{(1+\beta)}\int\frac{dF_y(x)}{(x+\mu^2)^2}\int\frac{dF_y(x)}{(x+\mu^2)^3}\\ &\left.+\int\frac{dF_y(x)}{(x+\mu^2)^4}\right]\\ &+\frac{y\int x^2dF_p(x)}{(1+\beta)^4}\left(\int\frac{dF_y(x)}{(x+\mu^2)^2}\right)^2\\ =&4y^2\Gamma_2^3-4y^2\Gamma_2\Gamma_3+y^2\Gamma_4+y\Gamma_2^2\int x^2dF_p(x), \end{aligned} \tag{29}$$

with $\Gamma_j=(1+\beta)^{-j}\int\frac{dF_y(x)}{(x+\mu^2)^j}$ for $j\ge 2$.

Proof: See Appendix C.

Lemma 3.4: Suppose the assumptions of Theorem 3.1 hold, we then have

$$\limsup_{N\to\infty}\mathrm{Var}\left\{\mathsf{Tr}\left[B_1^{-2}H_1D_1H_1^\dagger\right]\right\}<\infty$$

and

$$\limsup_{N\to\infty}\mathrm{Var}\left\{\mathsf{Tr}\left[B_1^{-2}\right]\right\}<\infty. \tag{30}$$

Proof: See Appendix D.

Before proceeding further, we introduce the following notations:

$$\begin{aligned} y_N&=\frac{K}{N},\quad \beta_N=\int\frac{dF_{y_N}(x)}{x+\mu^2}\\ d_N&=\int\frac{dF_{y_N}(x)}{(x+\mu^2)^2}\left(\frac{y_N}{(1+\beta_N)^2}+\sigma^2\right)\\ \beta_N^{(s)}&=\frac{\beta_N^2}{d_N},\quad Y_N=\boldsymbol{h}_1^\dagger B_1^{-1}\left(H_1D_1H_1^\dagger+\sigma^2I\right)B_1^{-1}\boldsymbol{h}_1 \end{aligned}$$

where $F_{y_N}(x)$ denotes the distribution function by substituting y_N for y in $F_y(x)$.

Lemma 3.5: In addition to the assumptions of Theorem 3.1, assume that

$$\lim_{N\to\infty}\sqrt{N}\|F_{D_1}(x)-F_p(x)\|\to 0 \tag{31}$$

where $F_{D_1}(x)$ denotes the empirical distribution function of D_1. We then get

$$\limsup_{N\to\infty}N\left|\frac{1}{N}\mathsf{E}\left[\mathsf{Tr}\left(B_1^{-2}\right)\right]-\int\frac{dF_{y_N}(x)}{(x+\mu^2)^2}\right|<\infty \tag{32}$$

and

$$\sqrt{N}\left|\frac{1}{N}\mathsf{E}\left[\mathsf{Tr}\left(B_1^{-2}H_1D_1H_1^\dagger\right)\right]-\int\frac{dF_{y_N}(x)}{(x+\mu^2)^2}\frac{y_N}{(1+\beta_N)^2}\right|\to 0. \tag{33}$$

Proof: See Appendix E.

Theorem 3.3: Further to the conditions of Theorem 3.1, assume that the empirical moment of the users' powers tends to a limit ϑ_2, i.e.,

$$\lim_{K\to\infty}\frac{1}{K}\sum_{k=2}^{K}p_k^2=\vartheta_2. \tag{34}$$

Then, as $N\to\infty$

$$\sqrt{N}\left(\gamma_1^{(s)}-\beta_N^{(s)}\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,l^2) \tag{35}$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, and $\mathcal{N}(0,l^2)$ denotes the Gaussian distribution with zero mean and variance l^2. The variance l^2 can be calculated as follows.

i) If X_{11} is real, then

$$l^2=l_1\mathsf{E}\left[X_{11}^4-3\right]+2l_2\triangleq l_r^2. \tag{36}$$

ii) If X_{11} is complex and $\mathsf{E}\left(X_{11}^2\right)=0$, then the variance becomes

$$l^2=l_1\mathsf{E}[|X_{11}|^4-2]+l_2\triangleq l_c^2 \tag{37}$$

where

$$l_1=\mathcal{K}^2\beta^2 \tag{38}$$

$$\begin{aligned} l_2=&\mathcal{K}^2\Big[(1+\beta)^2\Gamma_2+2y^2(\mathcal{K}-1)^2\Gamma_2^3\\ &+2y(\mathcal{K}-1)\beta\Gamma_2^2+y\Gamma_2^2(\mathcal{K}^2\vartheta_2-1)\Big]. \end{aligned} \tag{39}$$

Proof: See Appendix F.

Remark 3.1: In Theorem 3.3, we consider the asymptotic distribution of $\sqrt{N}(\gamma_1^{(s)}-\beta_N^{(s)})$ for a given (limited) size N, instead of that of $\sqrt{N}(\gamma_1^{(s)}-\beta^{(s)})$. This is because, in theory, the rate of instantaneous load y_N converging to y may be arbitrarily slow, thus $\sqrt{N}(\gamma_1^{(s)}-\beta^{(s)})$ may blow up. Please note that $\beta_N^{(s)}$ is the limit of $\beta^{(s)}$ with y being replaced by the instantaneous load y_N, and the subscript N in $\beta_N^{(s)}$ is used to denote the dependence on the instantaneous load only.

Remark 3.2: For equal-power systems, $\vartheta_2=1$, thus if we choose $\mathcal{K}=1$, we have

$$\begin{aligned} l_1&=\beta^2\\ l_2&=(1+\beta)^2\Gamma_2=\frac{\beta(1+\beta)^2}{y+\sigma^2(1+\beta)^2}. \end{aligned}$$

For the real-spreading channel case, the above results are consistent with the results derived in [13]. For the complex-spreading channel case, the variance expression of SIR distribution is different from the conjecture made in [13]. In fact, by extending the covariance results from real-spreading channels to complex-spreading channels, besides the term $\mathsf{E}\left[|X_{11}|^4 - 2\right]$ should be used to replace the term $\mathsf{E}\left[X_{11}^4 - 3\right]$, the term $2l_2$ in (36) should also be replaced with the term l_2.

Remark 3.3: Let us consider the relationship between the asymptotic SIR distributions for real-spreading and complex-spreading channel cases. For the real-spreading channel case, we denote X_{11} as $X_{11}^{(\mathrm{REAL})}$. For the corresponding complex-spreading channel case, X_{11} is denoted as

$$X_{11}^{(\mathrm{CMX})} = \frac{1}{\sqrt{2}}\left(X_{11}^{(r)} + jX_{11}^{(i)}\right)$$

where $X_{11}^{(r)}$ and $X_{11}^{(i)}$ are independent but both follow the same distribution of random variable $X_{11}^{(\mathrm{REAL})}$, which is with zero mean and unit variance, according to the assumptions of Theorem 3.1. It can be easily verified that

$$\mathsf{E}\left[\left|X_{11}^{(\mathrm{CMX})}\right|^4 - 2\right] = \frac{1}{2}\mathsf{E}\left[\left(X_{11}^{(\mathrm{REAL})}\right)^4 - 3\right].$$

Thus, from (34) and (35), we have $l_c^2 = \frac{l_r^2}{2}$.

If we further assume Gaussian distribution for $X_{11}^{(\mathrm{REAL})}$, then $l_{c,\mathrm{Gaussian}}^2 = \frac{l_{r,\mathrm{Gaussian}}^2}{2} = l_2$. On the other hand, comparing the real-spreading channel with binary phase-shift keying (BPSK) random codes and the complex-spreading channel with quaternary phase-shift keying (QPSK) random codes, then $l_{\mathrm{QPSK}}^2 = \frac{l_{\mathrm{BPSK}}^2}{2} = l_2 - l_1$. Computer simulations will be given in Section IV to compare the accuracy of the results derived in this paper and that of [13].

The accuracy of Theorem 3.3 in describing the distribution of SIR depends on the dimension N. The larger the N, the more accurate the real distribution approaches the theoretical distribution. Using Gaussian distribution to model the SIR distribution, one potential problem is that the SIR may appear to be negative, which is not feasible in practice. The negative SIR problem also applies to the results of [13]. In order to predict the SIR distribution more accurately for a small size system, and to avoid the SIR being negative, we turn to look at the distribution of SIR in decibels, denoted as SIRdB. The following theorem directly comes from Theorem 3.3 using the Taylor series expansion.

Theorem 3.4: Under the conditions of Theorem 3.3, as $N \to \infty$

$$\sqrt{N}\left(10\log_{10}\left(\gamma_1^{(s)}\right) - 10\log_{10}\left(\beta_N^{(s)}\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \left(\frac{10\log_{10}(e)}{\beta_N^{(s)}}\right)^2 l^2\right) \quad (40)$$

where the variance l^2 is given in Theorem 3.3 for both real-spreading and complex-spreading channels, and e is the natural logarithm base.

IV. Numerical Results and Computer Simulations

In this section, numerical results and computer simulations are presented to evaluate the accuracy of various limiting and asymptotic results derived in this paper for large random systems.

A. Evaluation of Theorem 3.2

For a given noise variance σ^2, the selection of noise variance in designing the suboptimal MMSE receiver will affect the limiting SIR. Fig. 1 illustrates the limiting SIRs versus the selected variable μ^2 for different parameters y and σ^2. The limiting SIR is maximized when $\mu^2 = \sigma^2$ for all cases. Also, it is seen that the SIR loss is less than 3 dB when the mismatch between μ^2 and σ^2 is within 10 dB.

B. Limit and Asymptotic Distribution of SIR

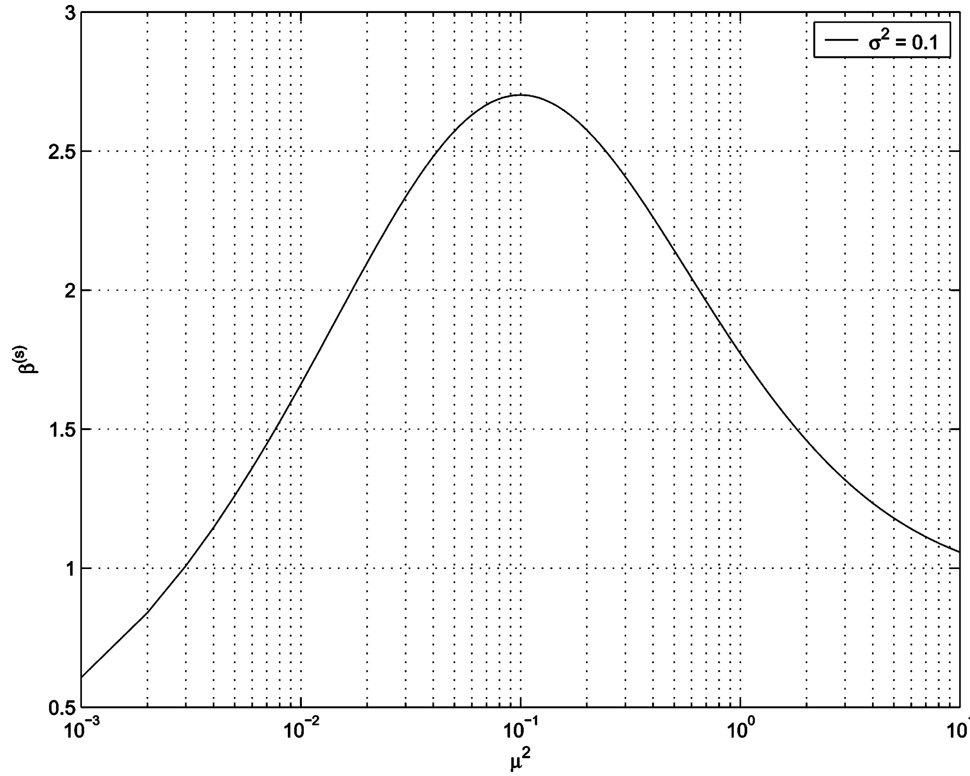
Computer simulation results are presented to evaluate the accuracy of the theoretical SIR distribution as compared to the simulated distribution for finite size systems. Two types of power distributions were simulated: equal power distribution and random exponential distribution. We also considered two types of distributions for the channel coefficients: the real Gaussian and the complex Gaussian. For each case, 10 000 Monte Carlo runs were carried out.

1) Equal Power Distribution: For equal power distribution, the suboptimal MMSE receiver is equivalent to the optimal MMSE receiver. The objective here is thus to compare the asymptotic SIR variances for real-spreading and complex-spreading channel cases. Fig. 2 shows the comparison between Tse and Zeitouni's formula (TZ formula) and the formula derived in this paper (LPB formula) with the simulated variances. Please note here the "variance of SIR (dB)" means "$10\log_{10}$ (variance of SIR)." For the real-spreading case, the theoretical results from TZ formula and LPB formula overlap. For the complex-spreading case, the formula derived in this paper is consistent with the simulation results, while the TZ formula is away from the simulated results.

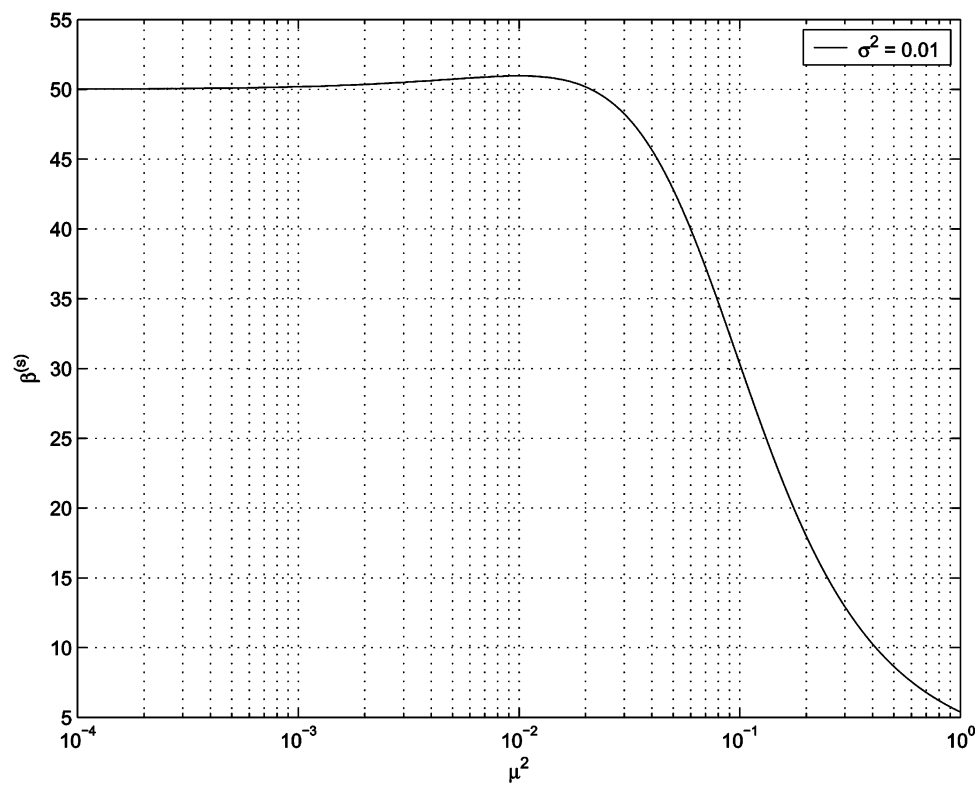
2) Exponential Power Distribution: For exponential power distribution, the results of this paper are needed in order to produce the theoretical mean and variance for the SIR variable. Fig. 3 compares the theoretical limiting SIR with the mean of the simulated SIRs, and Fig. 4 compares the theoretical variance with the variance of the simulated SIRs. It is seen that the simulated results are consistent with the theoretical results when the system load K/N is small $(K/N = 0.5)$, but with little bias when the system is fully loaded $K/N = 1$.

C. Comparison of Asymptotic Distribution for SIR and SIRdB

Finally, the tail behavior is of interest for the distributions of SIR and SIRdB. 1 000 000 Monte Carlo runs were simulated. We compare the simulated histogram, mean and 1% outage SIR or SIRdB from both theoretical and simulated results. Here the 1% outage SIR is defined as the value γ such that $\Pr(\mathrm{SIR} < \gamma) = 0.01$. Figs. 5 and 6 illustrate the comparisons for complex Gaussian channels and exponential power distribution with $K = 32$ and $N = 64$ and $K = N = 16$, respectively. Here, the "variance of SIRdB" means the variance



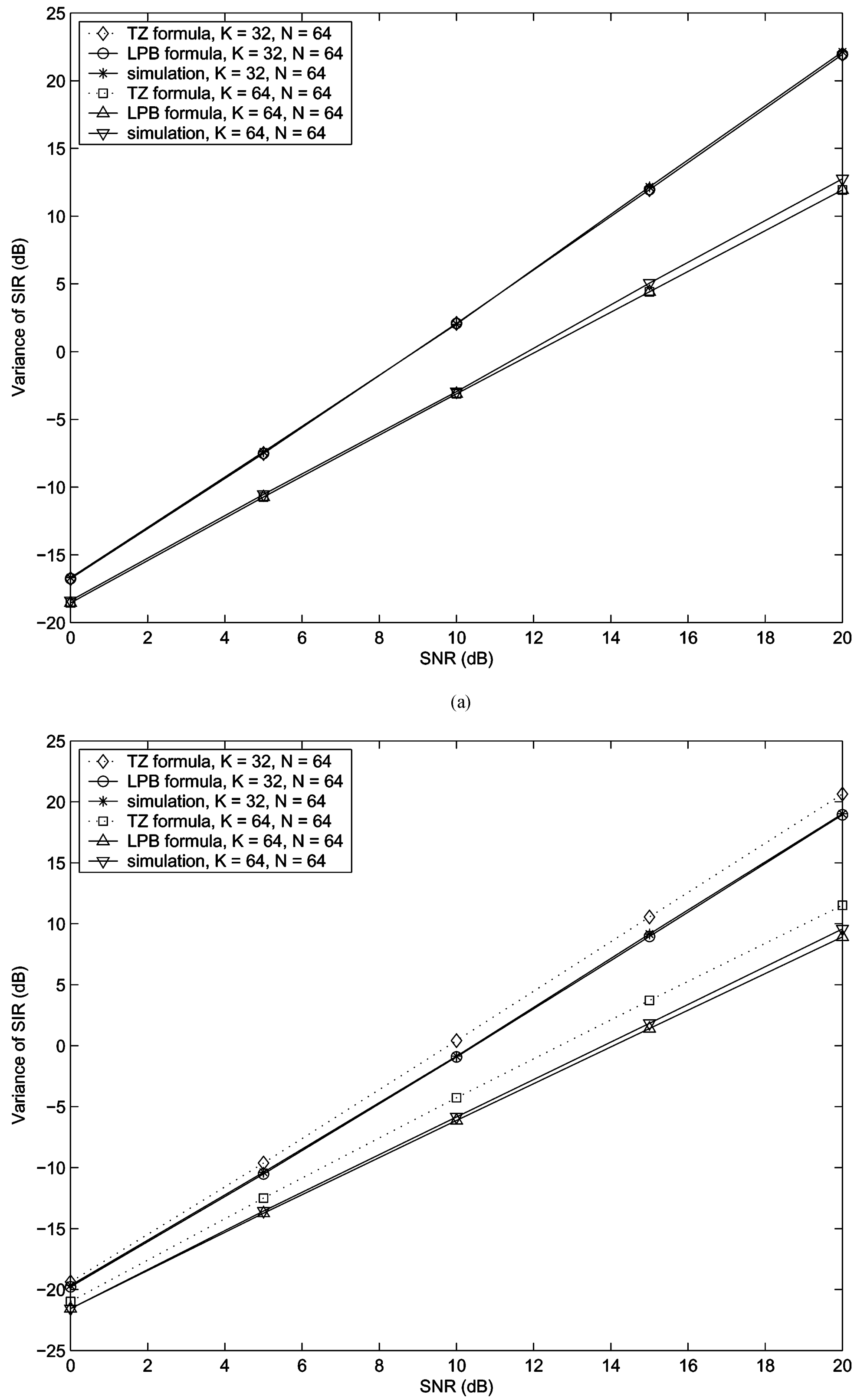
(a)



(b)

Fig. 1. Limiting SIRs versus parameter μ^2 for (a) $y = 1$ and $\sigma^2 = 0.1$ and (b) $y = 0.5$ and $\sigma^2 = 0.01$.

of "$10\log_{10}$ (SIR)." Comparing the histograms and the 1% outage SIR and 1% outage SIRdB, it is seen that for a small load $K/N = 0.5$ and reasonably large $N = 64$, the distribution of SIR is close to Gaussian, but this approximation is less

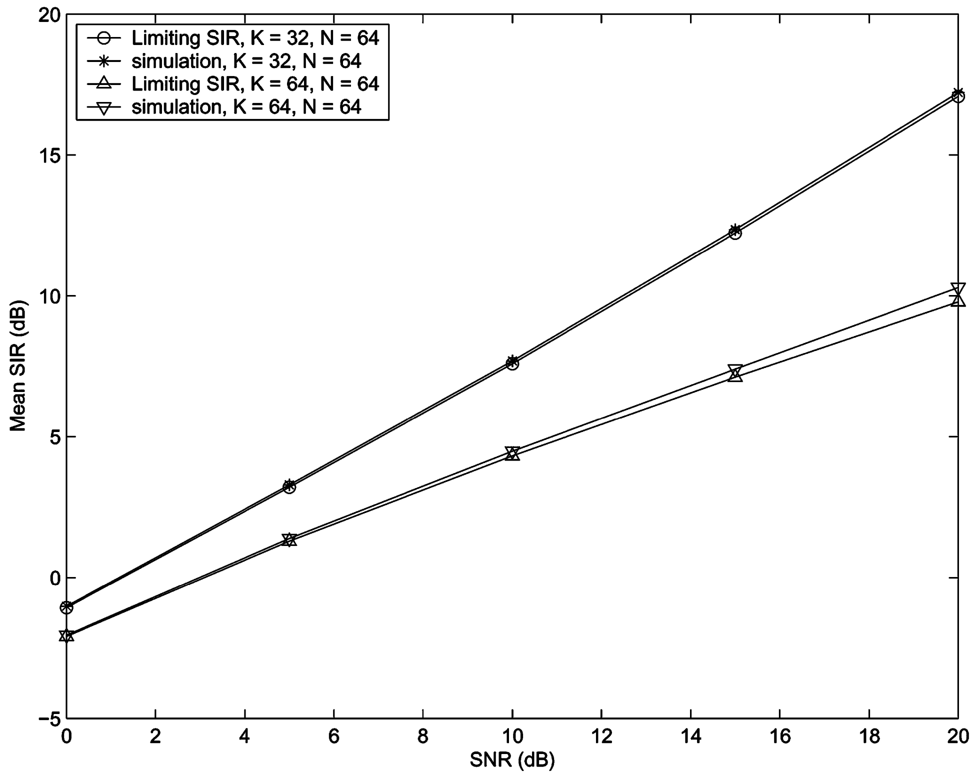


(a)

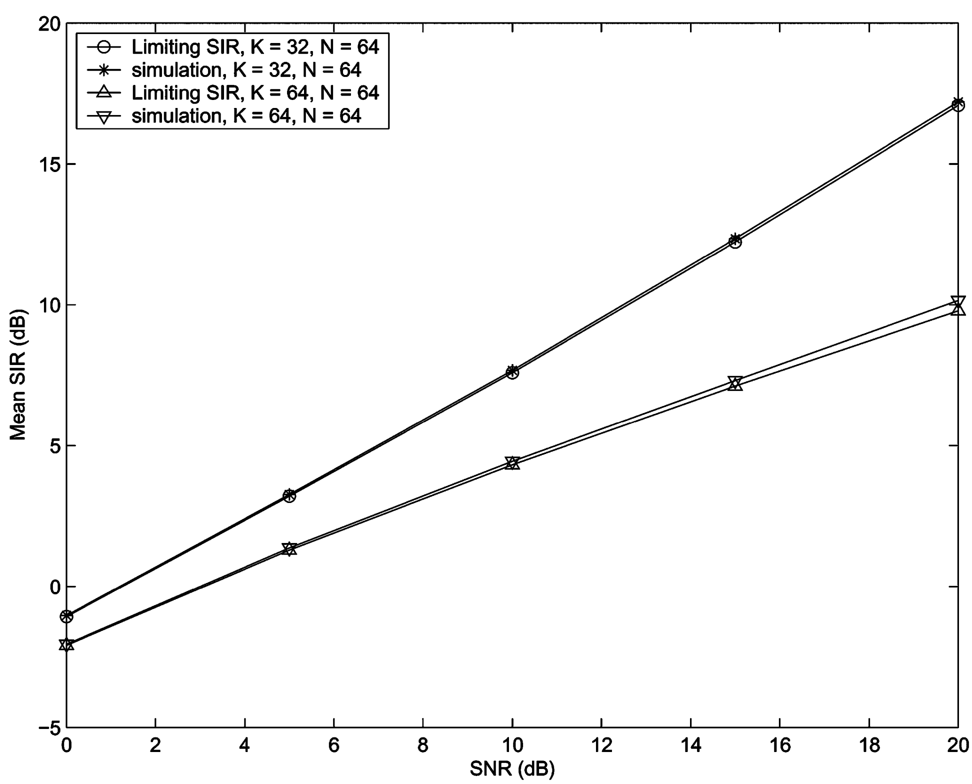
(b)

Fig. 2. Comparison of simulated variance and theoretical variance for systems with equal power distribution and (a) real-spreading channel; (b) complex-spreading channel.

accurate as compared to SIRdB. When $K = N = 16$ (small N and large load), the SIR distribution seems to be non-Gaussian, however, the SIRdB still follows the Gaussian distribution more closely.



(a)

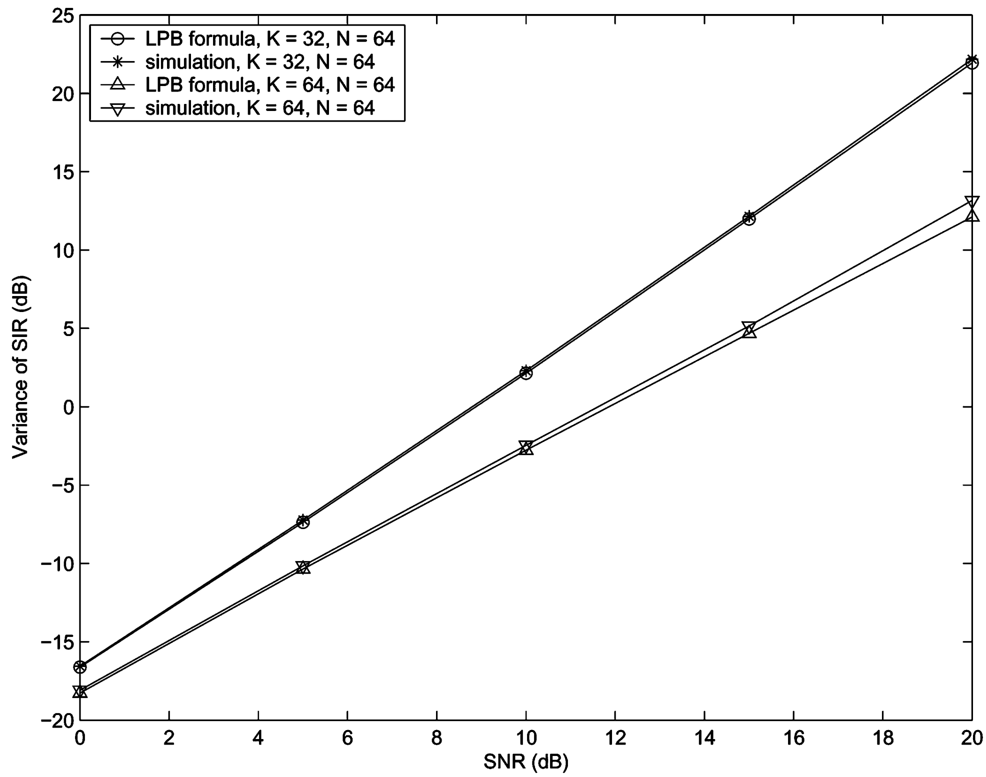


(b)

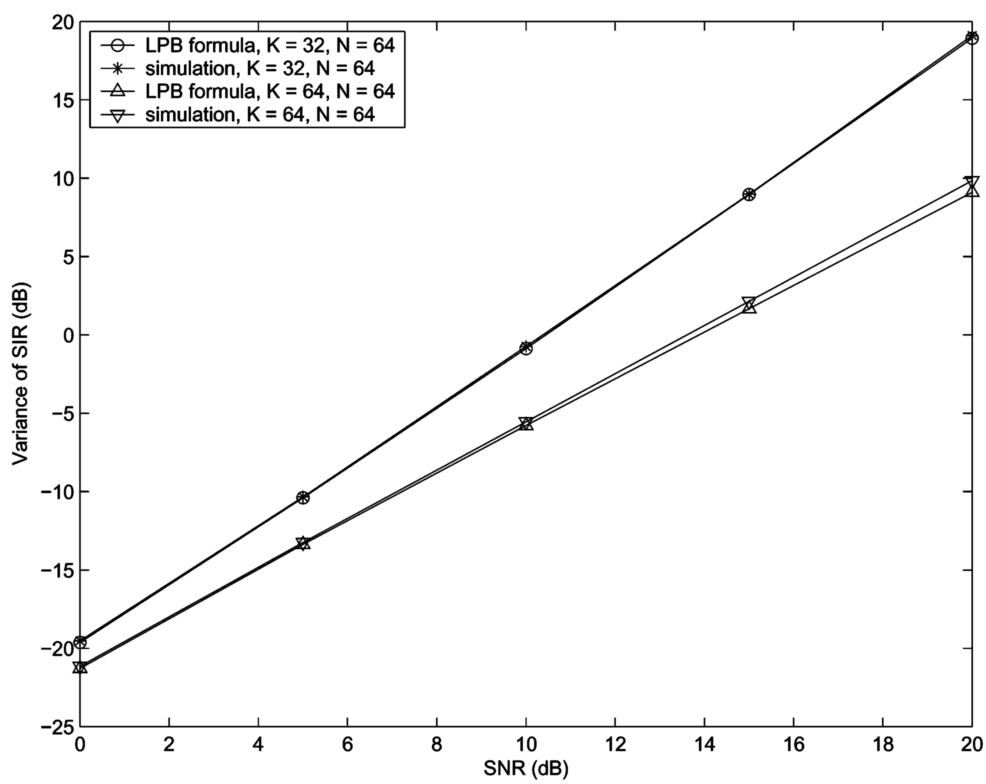
Fig. 3. Comparison of simulated mean and limiting SIR for systems with exponential power profile and (a) real-spreading channel; (b) complex-spreading channel.

V. CONCLUSION

In this paper, random matrix theory has been used to derive the limiting SIR and asymptotic SIR distribution for a class of suboptimal MMSE receivers applied to unequal-power large systems. We have proven that the limiting SIR converges to a deterministic value when K and N go to infinity with $\lim K/N =$



(a)



(b)

Fig. 4. Comparison of simulated variance and theoretical variance for systems with exponential power profile and (a) real-spreading channel; (b) complex-spreading channel.

y being a constant, where K is the number of users and N is the number of degrees of freedom. We have also proven that the output SIR is asymptotically Gaussian for large N and derived the closed-form expressions of the variance for both real-

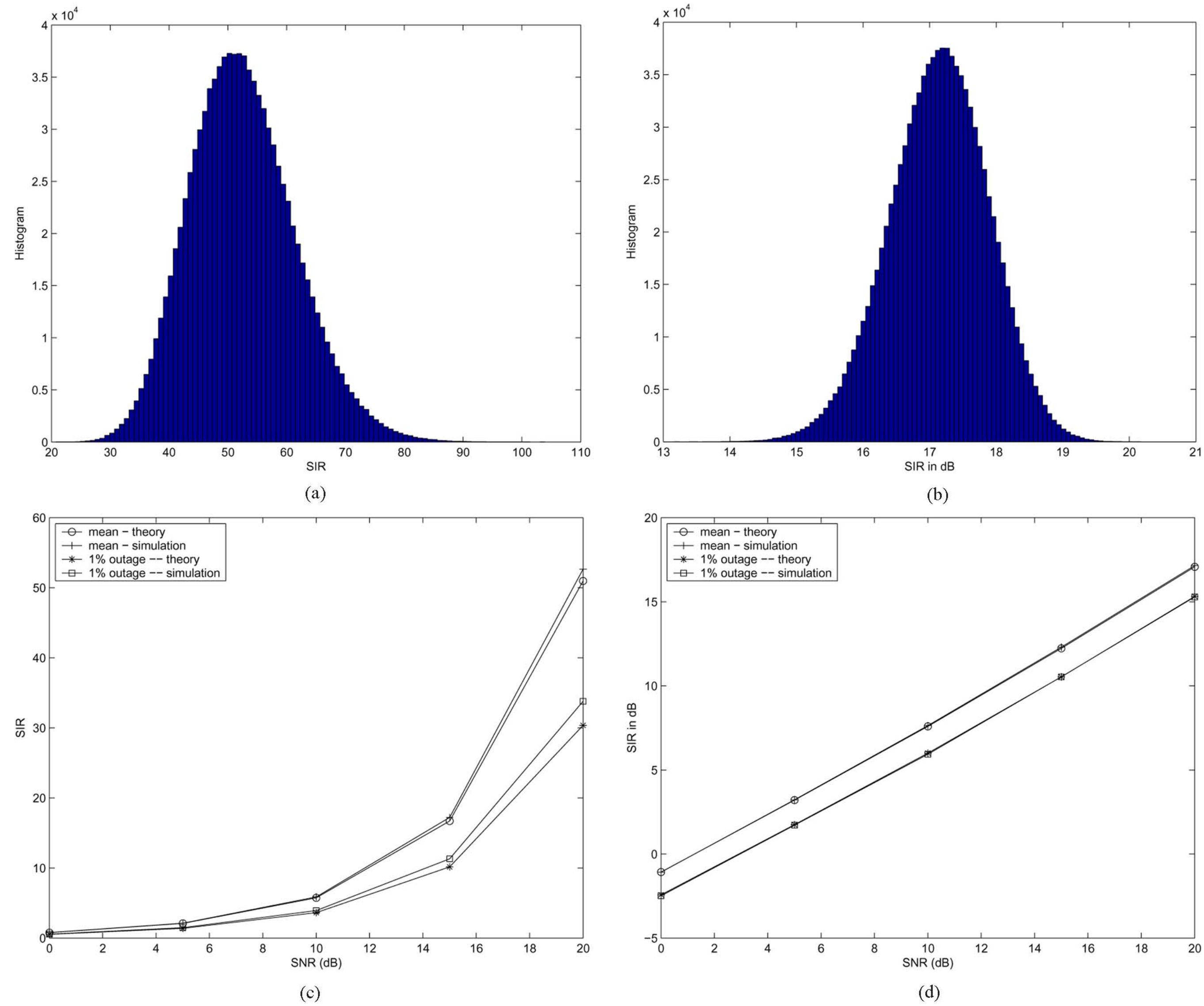


Fig. 5. Comparison of histograms for (a) SIR, (b) SIRdB when $SNR = 20$ dB; and mean and 1% outage of (c) SIR and (d) SIRdB under complex Gaussian channels with $K = 32$ and $N = 64$ and exponential power distribution.

spreading and complex-spreading channels. It is noticed that the complex channels produce half the variance of the real channels. We have also derived the asymptotic distribution of SIRdB, and obtained the closed-form expressions for its mean and variance for finite size N. Numerical results and computer simulations have evaluated the accuracy of the limiting results obtained in this paper. Specifically, the distribution of SIRdB provides a more accurate description than the distribution of SIR, especially when N is small and system load K/N is equal to 1.

Appendix A
Proof of Theorem 3.1

We first prove a lemma which will be frequently used in the following sections.

Lemma 5.1: In addition to the assumptions of Theorem 3.1, suppose that random matrices $\{B_{jk}, j \le K, k \le N^{\eta}\}$ are independent of $\{X_{1j}, \ldots, X_{Nj}\}$ for each j and have a uniform upper bound in the spectral norm, that is, $\|B_{jk}\| \le C_0$, where η and C_0 are two positive constants. Then

$$\max_{j,k} \left| \boldsymbol{h}_j^{\dagger} B_{jk} \boldsymbol{h}_j - \frac{1}{N} \mathsf{Tr}(B_{jk}) \right| \xrightarrow{a.s.} 0.$$

Proof: Define $\widehat{X}_{ij} = X_{ij}$ if $|x_{ij}| < \log N$ and 0 otherwise. Let $\hat{\boldsymbol{h}}_j = N^{-1/2}(\widehat{X}_{1j}, \ldots, \widehat{X}_{Nj})^T$ and $\tilde{\boldsymbol{h}}_j = \hat{\boldsymbol{h}}_j - \mathsf{E}(\hat{\boldsymbol{h}}_j) = N^{-1/2}(\widetilde{X}_{1j}, \ldots, \widetilde{X}_{Nj})^T$. Then by [25, Lemma 2] and the assumption of the finite fourth moment of the random variables X_{ij}'s, we have

$$\max_j \left| \|\boldsymbol{h}_j\|^2 - 1 \right| \xrightarrow{a.s.} 0 \tag{41}$$

$$\max_j \|\boldsymbol{h}_j - \tilde{\boldsymbol{h}}_j\|^2 \xrightarrow{a.s.} 0. \tag{42}$$

Here the second limit holds since for any large C

$$\|\mathsf{E}(\hat{\boldsymbol{h}}_j)\|^2 = |\mathsf{E}(\hat{X}_{11})|^2 \to 0$$

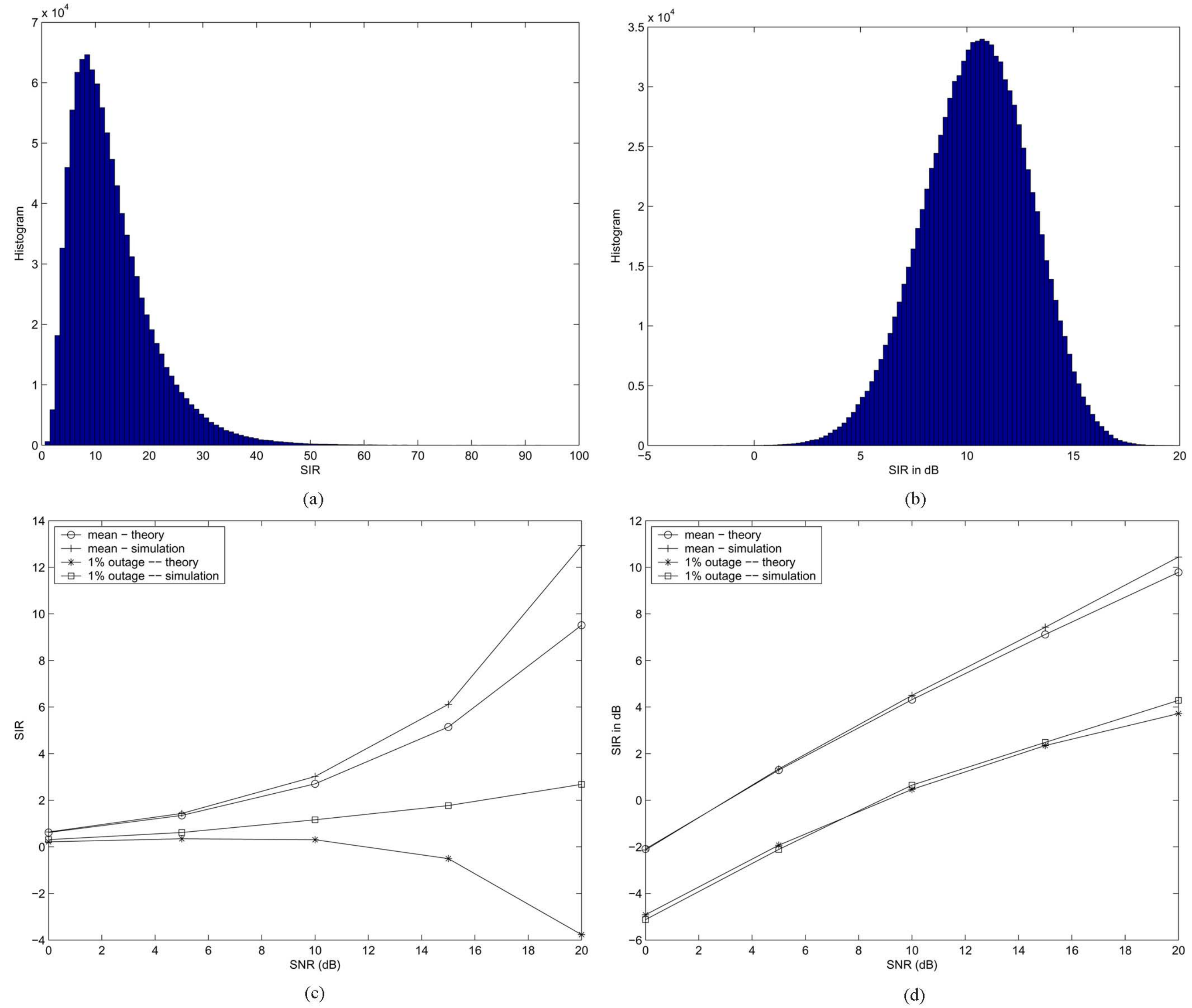


Fig. 6. Comparison of histograms for (a) SIR, (b) SIRdB when $SNR = 20$ dB; and mean and 1% outage of (c) SIR and (d) SIRdB under complex Gaussian channels with $K = 16$ and $N = 16$ and exponential power distribution.

$$
\begin{aligned}
\lim_{N\to\infty} \|\boldsymbol{h}_j - \hat{\boldsymbol{h}}_j\|^2 &= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} |X_{ij}|^2 I(|X_{ij}| \geq \log N) \\
&\leq \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} |X_{ij}|^2 I(|X_{ij}| \geq C) \\
&\overset{a.s.}{=} \mathsf{E}\left[|X_{11}|^2 I(|X_{11}| \geq C)\right]
\end{aligned}
$$

which can be made arbitrarily small by taking large C. In the above, $I(\cdot)$ is the indicator function which takes value 1 if the statement in the bracket is true, and 0 otherwise.

By the assumption of the lemma and using (41) and (42)

$$
\begin{aligned}
&\left|\boldsymbol{h}_j^\dagger B_{jk} \boldsymbol{h}_j - \tilde{\boldsymbol{h}}_j^\dagger B_{jk} \tilde{\boldsymbol{h}}_j\right| \\
&\qquad \leq C_0 \left[2\|\boldsymbol{h}_j\| \|\boldsymbol{h}_j - \tilde{\boldsymbol{h}}_j\| + \|\boldsymbol{h}_j - \tilde{\boldsymbol{h}}_j\|^2\right] \xrightarrow{a.s.} 0
\end{aligned}
$$

uniformly for $j \leq K$.

Now, by [19, Lemma 2.7], for any $p > 0$

$$
\mathsf{E}\left|\tilde{\boldsymbol{h}}_j^\dagger B_{jk} \tilde{\boldsymbol{h}}_j - \frac{\mathsf{E}\left(|\widetilde{X}_{11}|^2\right)}{N} \mathsf{Tr}(B_{jk})\right|^{2p} \leq C_p N^{-p} C_0^{2p} (\log N)^{4p}.
$$

Using the Borel–Cantelli lemma and choosing $p > \eta + 2$, we obtain

$$
\max_{jk} \left|\tilde{\boldsymbol{h}}_j^\dagger B_{jk} \tilde{\boldsymbol{h}}_j - \frac{\mathsf{E}(|\widetilde{X}_{11}|^2)}{N} \mathsf{Tr}(B_{jk})\right| \xrightarrow{a.s.} 0.
$$

The lemma then follows by noting that $\mathsf{E}(|\widetilde{X}_{11}|^2) \to 1$. $\square$

Now we turn to the proof of Theorem 3.1.

Proof: From the trivial inequality

$$B_1^{-1}\left(H_1D_1H_1^\dagger+\sigma^2 I\right)B_1^{-1}\le MB_1^{-1}+|\mu^2+\sigma^2|B_1^{-2}$$

one can easily show that the norm of the matrix $B_1^{-1}\left(H_1D_1H_1^\dagger+\sigma^2 I\right)B_1^{-1}$ is bounded by $(M+1)\sigma^{-2}+\mu^2\sigma^{-4}$, where M is a bound for the relative powers p_i of all users.

Using Lemma 5.1, we have

$$\boldsymbol{h}_1^\dagger B_1^{-1}\left(H_1D_1H_1^\dagger+\sigma^2 I\right)B_1^{-1}\boldsymbol{h}_1 -\frac{1}{N}\mathsf{Tr}\left[B_1^{-2}\left(H_1D_1H_1^\dagger+\sigma^2 I\right)\right]\xrightarrow{a.s.}0 \quad (43)$$

and

$$\max_{j\le K}\left|\boldsymbol{h}_j^\dagger B_{1j}^{-1}\boldsymbol{h}_j-\frac{1}{N}\mathsf{Tr}\left(B_{1j}^{-1}\right)\right|\to 0,\quad \text{a.s.} \quad (44)$$

Making use of the convergence of the Marčhenko–Pastur law (see [31, Theorem 2.5]), we obtain that

$$\frac{1}{N}\mathsf{Tr}(B^{-1})\xrightarrow{a.s.}\beta=\int\frac{1}{x+\mu^2}dF_y(x) \quad (45)$$

where $B=HH^\dagger+\mu^2 I$ and F_y is the Marčhenko–Pastur law defined earlier.

On the other hand, we have

$$\frac{1}{N}\left|\mathsf{Tr}(B^{-1})-\mathsf{Tr}\left(B_1^{-1}\right)\right|=\frac{1}{N}\frac{\boldsymbol{h}_1^\dagger B_1^{-2}\boldsymbol{h}_1}{1+\boldsymbol{h}_1^\dagger B_1^{-1}\boldsymbol{h}_1}\le\frac{1}{N\sigma^2}$$

and similarly $\left|\mathsf{Tr}\left(B_1^{-1}\right)-\mathsf{Tr}\left(B_{1j}^{-1}\right)\right|\le(1/N)\sigma^2$, where $B_{1j}=B_1-\boldsymbol{h}_j\boldsymbol{h}_j^\dagger$. Therefore, we obtain

$$\boldsymbol{h}_j^\dagger B_{1j}^{-1}\boldsymbol{h}_j\to\beta\quad\text{a.s.} \quad (46)$$

where the convergence is uniform in j.

Let

$$H_{1j}=(\boldsymbol{h}_2,\ldots,\boldsymbol{h}_{j-1},\boldsymbol{h}_{j+1},\ldots,\boldsymbol{h}_K)^T \text{ and } \alpha_j=\frac{1}{1+\boldsymbol{h}_j^\dagger B_{1j}^{-1}\boldsymbol{h}_j}.$$

Since $\boldsymbol{h}_j^\dagger B_1^{-1}=\alpha_j\boldsymbol{h}_j^\dagger B_{1j}^{-1}$, with probability 1, we get

$$\begin{aligned}\frac{1}{N}\mathsf{Tr}\left[B_1^{-1}H_1D_1H_1^\dagger\right]&=\frac{1}{N}\sum_{j=2}^{K}p_j\boldsymbol{h}_j^\dagger B_1^{-1}\boldsymbol{h}_j\\&=\frac{1}{N}\sum_{j=2}^{K}\alpha_jp_j\boldsymbol{h}_j^\dagger B_{1j}^{-1}\boldsymbol{h}_j\\&=\frac{1}{N}\sum_{j=2}^{K}p_j\frac{\beta}{1+\beta}+o(1)\\&\to\frac{y\beta}{1+\beta}.\end{aligned} \quad (47)$$

Therefore, by the Vitali theorem (refer to [22, p. 168]) and dominated convergence theorem, we conclude that

$$\begin{aligned}&\frac{1}{N}\mathsf{Tr}\left[B_1^{-2}\left(H_1D_1H_1^\dagger+\sigma^2 I\right)\right]\\&=-\frac{\partial}{\partial\mu^2}\frac{1}{N}\mathsf{Tr}\left[B_1^{-1}\left(H_1D_1H_1^\dagger+\sigma^2 I\right)\right]\\&\xrightarrow{a.s.}-\frac{\partial}{\partial\mu^2}\left(\frac{y\beta}{1+\beta}+\sigma^2\beta\right)\\&=\left(\frac{y}{(1+\beta)^2}+\sigma^2\right)\int\frac{dF_y(x)}{(x+\mu^2)^2}.\end{aligned} \quad (48)$$

In fact, it is easy to see that $\frac{1}{N}\mathsf{Tr}\left[B_1^{-1}\left(H_1D_1H_1^\dagger+\sigma^2 I\right)\right]$ is an analytic function on some bounded region of μ^2 and that

$$\begin{aligned}&\frac{1}{N}\mathsf{Tr}\left[B_1^{-1}\left(H_1D_1H_1^\dagger+\sigma^2 I\right)\right]\\&\le\frac{M}{N}\mathsf{Tr}\left(B_1^{-1}H_1H_1^\dagger\right)+\frac{\sigma^2}{N}\mathsf{Tr}\left(B_1^{-1}I\right)\\&=M-\frac{M\mu^2}{N}\mathsf{Tr}B_1^{-1}+\frac{\sigma^2}{N}\mathsf{Tr}\left(B_1^{-1}\right)\\&\le 2M+\sigma^2\mu^2\\&\le M_1,\end{aligned}$$

where M_1 denotes some constant independent of n and μ^2.

By similar arguments, we may prove that

$$\boldsymbol{h}_1^\dagger B_1^{-1}\boldsymbol{h}_1\to\beta\quad\text{a.s.} \quad (49)$$

Thus, by substituting the above two limits into the numerator and denominator of $\mathrm{SIR}_1^{(s)}$, we show that, with probability 1

$$\gamma_1^{(s)}\to\frac{\beta^2(1+\beta)^2}{\int\frac{dF_y(x)}{(x+\mu^2)^2}\left[\sigma^2(1+\beta)^2+y\right]}. \quad (50)$$

By (16), we have that

$$\frac{\beta}{\int\frac{dF_y(x)}{(x+\mu^2)^2}}=-\frac{\beta}{\beta'}=\frac{y}{(1+\beta)^2}+\mu^2$$

where $\beta'=\frac{d}{d\mu^2}\beta$. Substituting this into the right-hand side of (50), (18) immediately follows. Further, (19) is the positive solution of (16). □

APPENDIX B
PROOF OF LEMMA 3.1

Proof: Using the formula

$$(A^{-1})_{11}=\frac{1}{(A)_{11}-u_1'A_1^{-1}v_1}$$

where A and A_1 are nonsingular and $A=\begin{pmatrix}A_{11}&u_1'\\v_1&A_1\end{pmatrix}$. Then it follows that for any $j=1,\ldots,N$

$$\left[(HH^\dagger+\mu^2I)^{-1}\right]_{jj}=\frac{1}{(HH^\dagger)_{jj}+\mu^2-\hat{\boldsymbol{h}}_j^\dagger\hat{H}_1^\dagger\hat{B}_1^{-1}\hat{H}_1\hat{\boldsymbol{h}}_j} \quad (51)$$

where

$$\hat{\boldsymbol{h}}_j = \frac{1}{\sqrt{N}}(X_{j1},\ldots,X_{jK})^\dagger,$$
$$\hat{H}_j^\dagger = (\hat{\boldsymbol{h}}_1,\ldots,\hat{\boldsymbol{h}}_{j-1},\hat{\boldsymbol{h}}_{j+1},\ldots,\hat{\boldsymbol{h}}_N), \text{ and}$$
$$\hat{B}_j^{-1} = \left(\hat{H}_j\hat{H}_j^\dagger + \mu^2 I\right)^{-1}.$$

Here we would remind the readers to note the difference between $\hat{\boldsymbol{h}}_j$ and $\boldsymbol{h}_j$: the former is the complex conjugate transpose of the row vector while the latter is the column vector of the matrix H.

By applying [25, Lemma 2], we have

$$(HH^\dagger)_{jj} = \frac{1}{N}\sum_{i=1}^{K}|X_{ji}|^2 \to y \tag{52}$$

almost surely and uniformly for $j \le N$. Employing Lemma 5.1, it follows that

$$\max_{j\le N}\left|\hat{\boldsymbol{h}}_j^\dagger \hat{H}_j^\dagger \hat{B}_j^{-1}\hat{H}_j\hat{\boldsymbol{h}}_j - \frac{1}{N}\mathsf{Tr}\left[\hat{B}_j^{-1}\hat{H}_j\hat{H}_j^\dagger\right]\right| \xrightarrow{a.s.} 0. \tag{53}$$

It is obvious that

$$\frac{1}{N}\mathsf{Tr}\left[\hat{B}_j^{-1}\hat{H}_j\hat{H}_j^\dagger\right] = 1 - \frac{\mu^2}{N}\mathsf{Tr}\left[\hat{B}_j^{-1}\right] \xrightarrow{a.s.} 1-\mu^2\beta \tag{54}$$

uniformly for $j \le N$. Then (23) follows from (51)–(54) and (17).

Applying (23) and the Vitali theorem, one can get

$$\begin{aligned}\left[(HH^\dagger+\mu^2 I)^{-2}\right]_{jj} &= -\frac{d}{d\mu^2}\left[(HH^\dagger+\mu^2 I)^{-1}\right]_{jj}\\ &\xrightarrow{a.s.} -\frac{d}{d\mu^2}\frac{1}{y-1+\mu^2(1+\beta)} = -\frac{d}{d\mu^2}\beta\\ &= \int\frac{dF_y(x)}{(x+\mu^2)^2}\end{aligned} \tag{55}$$

uniformly for $j \le N$.

Let $e_j = (0,\ldots,0,1,0,\ldots,0)^\dagger$, the element 1 is only in the jth position. Then, using the arguments in proving (46), we have

$$\begin{aligned}&\left[(HH^\dagger+\mu^2 I)^{-1}HDH^\dagger(HH^\dagger+\mu^2 I)^{-1}\right]_{jj}\\ &= \sum_{k=1}^{K}p_k\alpha_k^2\left[\left(H_kH_k^\dagger+\mu^2 I\right)^{-1}\boldsymbol{h}_k\boldsymbol{h}_k^\dagger\left(H_kH_k^\dagger+\mu^2 I\right)^{-1}\right]_{jj}\\ &= \sum_{k=1}^{K}p_k\alpha_k^2\boldsymbol{h}_k^\dagger\left(H_kH_k^\dagger+\mu^2 I\right)^{-1}e_je_j^\dagger\left(H_kH_k^\dagger+\mu^2 I\right)^{-1}\boldsymbol{h}_k\\ &= \sum_{k=1}^{K}p_k\alpha_k^2\frac{1}{N}\left(H_kH_k^\dagger+\mu^2 I\right)^{-2}_{jj} + o_{a.s.}(1)\\ &\to \frac{y}{(1+\beta)^2}\int\frac{dF_y(x)}{(x+\mu^2)^2}.\end{aligned} \tag{56}$$

Hence, we conclude that

$$\begin{aligned}&\left[(HH^\dagger+\mu^2 I)^{-1}(HDH^\dagger+\sigma^2 I)(HH^\dagger+\mu^2 I)^{-1}\right]_{jj}\\ &\quad\xrightarrow{a.s.}\left(\frac{y}{(1+\beta)^2}+\sigma^2\right)\int\frac{dF_y(x)}{(x+\mu^2)^2}\end{aligned} \tag{57}$$

uniformly for $j \le N$.

Thus, we complete the proof of (24) by using the fact that $\beta' = -\int\frac{dF_y(x)}{(x+\mu^2)^2}$ and $-\frac{\beta}{\beta'} = \frac{y}{(1+\beta)^2}+\mu^2$. □

APPENDIX C
PROOF OF LEMMA 3.3

Proof: The following identity will be frequently used:

$$B_1^{-1} - B_{1j}^{-1} = -\alpha_j B_{1j}^{-1}\boldsymbol{h}_j\boldsymbol{h}_j^\dagger B_{1j}^{-1}.$$

Write $B_{1t} = H_1H_1^\dagger + \mu_t^2 I$, $t = 1,2$, and define $B_{1jt} = B_{1t} - \boldsymbol{h}_j\boldsymbol{h}_j^\dagger$ and $B_{1jkt} = B_{1jt} - \boldsymbol{h}_k\boldsymbol{h}_k^\dagger$. We decompose $\frac{1}{N}\mathsf{Tr}\left[H_1D_1H_1^\dagger B_{11}^{-1}H_1D_1H_1^\dagger B_{12}^{-1}\right]$ as

$$\begin{aligned}&\frac{1}{N}\mathsf{Tr}\left[H_1D_1H_1^\dagger B_{11}^{-1}H_1D_1H_1^\dagger B_{12}^{-1}\right]\\ &= \frac{1}{N}\sum_{j=2}^{K}p_j\boldsymbol{h}_j^\dagger B_{11}^{-1}H_1D_1H_1^\dagger B_{12}^{-1}\boldsymbol{h}_j\\ &= \frac{1}{N}\sum_{j=2}^{K}p_j\alpha_{j1}\alpha_{j2}\boldsymbol{h}_j^\dagger B_{1j1}^{-1}H_{1j}D_{1j}H_{1j}^\dagger B_{1j2}^{-1}\boldsymbol{h}_j\\ &\quad+\frac{1}{N}\sum_{j=2}^{K}p_j^2\alpha_{j1}\alpha_{j2}\left(\boldsymbol{h}_j^\dagger B_{1j1}^{-1}\boldsymbol{h}_j\right)\left(\boldsymbol{h}_j^\dagger B_{1j2}^{-1}\boldsymbol{h}_j\right)\end{aligned} \tag{58}$$

where

$$\alpha_{jt} = \frac{1}{1+\boldsymbol{h}_j^\dagger B_{1jt}^{-1}\boldsymbol{h}_j} \text{ and}$$
$$D_{1j} = \mathrm{diag}[p_2,\ldots,p_{j-1},p_{j+1},\ldots,p_K].$$

Applying Lemma 5.1, one can easily prove that

$$\max_{j\le N}\left|\boldsymbol{h}_j^\dagger B_{1jt}^{-1}\boldsymbol{h}_j - \beta_t\right| \to 0, \quad \text{a.s.}$$

where

$$\beta_t = \int\frac{dF_y(x)}{x+\mu_t^2}.$$

$$\begin{aligned}
&\frac{1}{N}\mathsf{Tr}[H_1D_1H_1^\dagger B_1^{-2}H_1D_1H_1^\dagger B_1^{-2}]\\
&\quad=\frac{\partial}{\partial\mu_1^2}\frac{\partial}{\partial\mu_2^2}\frac{1}{N}\mathsf{Tr}\left[H_1D_1H_1^\dagger B_{11}^{-1}H_1D_1H_1^\dagger B_{12}^{-1}\right]\Big|_{\mu_1=\mu_2=\mu}\\
&\quad\xrightarrow{a.s.}\frac{\partial}{\partial\mu_1^2}\frac{\partial}{\partial\mu_2^2}\left[\frac{y^2}{(1+\beta_1)^2(1+\beta_2)^2}\int\frac{dF_y(x)}{(x+\mu_1^2)(x+\mu_2^2)}+\frac{y\beta_1\beta_2\vartheta_2}{(1+\beta_1)(1+\beta_2)}\right]_{\mu_1=\mu_2=\mu}\\
&\quad=\frac{y^2}{(1+\beta)^4}\left[\frac{4}{(1+\beta)^2}\left(\int\frac{dF_y(x)}{(x+\mu^2)^2}\right)^3-\frac{4}{1+\beta}\int\frac{dF_y(x)}{(x+\mu^2)^2}\int\frac{dF_y(x)}{(x+\mu^2)^3}+\int\frac{dF_y(x)}{(x+\mu^2)^4}\right]+\frac{y\vartheta_2}{(1+\beta)^4}\left(\int\frac{dF_y(x)}{(x+\mu^2)^2}\right)^2.
\end{aligned}\tag{61}$$

Therefore

$$\frac{1}{N}\sum_{j=2}^{K}p_j^2\alpha_{j1}\alpha_{j2}\left(\boldsymbol{h}_j^\dagger B_{1j1}^{-1}\boldsymbol{h}_j\right)\left(\boldsymbol{h}_j^\dagger B_{1j2}^{-1}\boldsymbol{h}_j\right)\xrightarrow{a.s.}\frac{y\beta_1\beta_2\vartheta_2}{(1+\beta_1)(1+\beta_2)}\tag{59}$$

where $\vartheta_2=\int x^2dF_p(x)$.

To find the limit of the first term on the right-hand side of (58), we use again Lemma 5.1 and obtain

$$\begin{aligned}
&\frac{1}{N}\sum_{j=2}^{K}p_j\alpha_{j1}\alpha_{j2}\boldsymbol{h}_j^\dagger B_{1j1}^{-1}H_{1j}D_{1j}H_{1j}^\dagger B_{1j2}^{-1}\boldsymbol{h}_j\\
&\quad=\frac{1}{N^2}\sum_{j=2}^{K}p_j\alpha_{j1}\alpha_{j2}\mathsf{Tr}\left(B_{1j1}^{-1}H_{1j}D_{1j}H_{1j}^\dagger B_{1j2}^{-1}\right)\\
&\qquad+o_{a.s.}(1).\\
&\quad=\frac{1}{N^2}\sum_{j\neq k=2}^{K}p_jp_k\alpha_{j1}\alpha_{j2}\alpha_{jk1}\alpha_{jk2}\left(\boldsymbol{h}_j^\dagger B_{1j2}^{-1}B_{1j1}^{-1}\boldsymbol{h}_j\right)\\
&\qquad+o_{a.s.}(1)\\
&\quad=\frac{1}{N^3}\sum_{j\neq k=2}^{K}p_jp_k\alpha_{j1}\alpha_{j2}\alpha_{jk1}\alpha_{jk2}\mathsf{Tr}\left(B_{1j2}^{-1}B_{1j1}^{-1}\right)\\
&\qquad+o_{a.s.}(1)\\
&\quad\xrightarrow{a.s.}\frac{y^2}{(1+\beta_1)^2(1+\beta_2)^2}\int\frac{dF_y(x)}{(x+\mu_1^2)(x+\mu_2^2)}.
\end{aligned}\tag{60}$$

To complete the proof of this lemma, we employ the Vitali theorem twice (shown in (61) at the top of the page). □

APPENDIX D
PROOF OF LEMMA 3.4

Proof: Let $E_j(\cdot)$ for $j>1$ denote conditional expectation with respect to the σ-field generated by $h_2,\ldots,h_j$ and let E_1 denote expectation.

We shall frequently use the inequality

$$\mathsf{E}\left|\boldsymbol{h}_j^\dagger B\boldsymbol{h}_j-\frac{1}{N}\mathsf{Tr}(B)\right|^2\le C/N\tag{62}$$

if B is independent of $\boldsymbol{h}_j$ and $\|B\|$ is bounded by a nonrandom constant. In fact, the inequality is an easy consequence of [19, Lemma 3.3].

We use the following martingale decomposition:

$$\begin{aligned}
&\mathsf{Tr}\left[B_1^{-2}H_1D_1H_1^\dagger\right]-\mathsf{E}\mathsf{Tr}\left[B_1^{-2}H_1D_1H_1^\dagger\right]\\
&=\sum_{j=2}^{K}\left[\mathsf{Tr}\left[E_jB_1^{-2}H_1D_1H_1^\dagger\right]-\mathsf{Tr}\left[E_{j-1}B_1^{-2}H_1D_1H_1^\dagger\right]\right]\\
&=\sum_{j=2}^{K}\mathsf{Tr}\left[(E_j-E_{j-1})\left(B_1^{-2}H_1D_1H_1^\dagger-B_{1j}^{-2}H_{1j}D_{1j}H_{1j}^\dagger\right)\right]\\
&=\sum_{j=2}^{K}\mathsf{Tr}\left[(E_j-E_{j-1})\left(B_1^{-1}-B_{1j}^{-1}\right)B_1^{-1}H_1D_1H_1^\dagger\right]\\
&\quad+\sum_{j=2}^{K}\mathsf{Tr}\left[(E_j-E_{j-1})B_{1j}^{-1}\left(B_1^{-1}-B_{1j}^{-1}\right)H_1D_1H_1^\dagger\right]\\
&\quad+\sum_{j=2}^{K}\mathsf{Tr}\left[(E_j-E_{j-1})B_{1j}^{-2}\left(H_1D_1H_1^\dagger-H_{1j}D_{1j}H_{1j}^\dagger\right)\right]\\
&=U_1+U_2+U_3+U_4+U_5
\end{aligned}\tag{63}$$

where

$$\begin{aligned}
U_1&=-2\sum_{j=2}^{K}(E_j-E_{j-1})\left[\alpha_j\boldsymbol{h}_j^\dagger B_{1j}^{-2}H_{1j}D_{1j}H_{1j}^\dagger B_{1j}^{-1}\boldsymbol{h}_j\right]\\
U_2&=-2\sum_{j=2}^{K}(E_j-E_{j-1})\left[\alpha_jp_j\boldsymbol{h}_j^\dagger B_{1j}^{-2}\boldsymbol{h}_j\boldsymbol{h}_j^\dagger B_{1j}^{-1}\boldsymbol{h}_j\right]\\
U_3&=\sum_{j=2}^{K}(E_j-E_{j-1})\left[\alpha_j^2\boldsymbol{h}_j^\dagger B_{1j}^{-2}\boldsymbol{h}_j\boldsymbol{h}_j^\dagger B_{1j}^{-1}H_{1j}D_{1j}H_{1j}^\dagger B_{1j}^{-1}\boldsymbol{h}_j\right]\\
U_4&=\sum_{j=2}^{K}(E_j-E_{j-1})\left[\alpha_j^2p_j\boldsymbol{h}_j^\dagger B_{1j}^{-2}\boldsymbol{h}_j\left(\boldsymbol{h}_j^\dagger B_{1j}^{-1}\boldsymbol{h}_j\right)^2\right]\\
U_5&=\sum_{j=2}^{K}p_j(E_j-E_{j-1})\left(\boldsymbol{h}_j^\dagger B_{1j}^{-2}\boldsymbol{h}_j\right).
\end{aligned}$$

$$\begin{aligned}
&\mathsf{E}\left|(E_j - E_{j-1})\left[\alpha_j \boldsymbol{h}_j^\dagger B_{1j}^{-2} H_{1j} D_{1j} H_{1j}^\dagger B_{1j}^{-1} \boldsymbol{h}_j\right]\right|^2 \\
&\le 2\mathsf{E}\left[|\alpha_j|^2 \left|\boldsymbol{h}_j^\dagger B_{1j}^{-2} H_{1j} D_{1j} H_{1j}^\dagger B_{1j}^{-1} \boldsymbol{h}_j - \frac{1}{N}\mathsf{Tr}\left(B_{1j}^{-2} H_{1j} D_{1j} H_{1j}^\dagger B_{1j}^{-1}\right)\right|^2\right] \\
&\quad + 2\mathsf{E}\left[|\bar{\alpha}_j \epsilon_j - \alpha_j \bar{\alpha}_j \epsilon_j|^2 \left(\frac{1}{N}\mathsf{Tr}\left(B_{1j}^{-2} H_{1j} D_{1j} H_{1j}^\dagger B_{1j}^{-1}\right)\right)^2\right] \\
&\le C/N,
\end{aligned}$$

Now, it suffices to evaluate the variance of above items. To estimate $\mathsf{E}\left(|U_1|^2\right)$, using (62) we get the equation at the top of the page, where we have used the fact that $|\alpha_j| < 1$,

$$\left\|B_{1j}^{-2} H_{1j} D_{1j} H_{1j}^\dagger B_{1j}^{-1}\right\| \le M\mu^{-4},$$

and M is the bound for p_j's. This proves that $\mathsf{E}\left(|U_1|^2\right)$ is bounded by a constant.

Since $\alpha_j h_j^\dagger B_{1j}^{-1} h_j < 1$

$$\begin{aligned}
&\mathsf{E}\left|(E_j - E_{j-1})\left[\alpha_j \boldsymbol{h}_j^\dagger B_{1j}^{-2} \boldsymbol{h}_j \boldsymbol{h}_j^\dagger B_{1j}^{-1} \boldsymbol{h}_j\right]^2\right| \\
&\le 3\mathsf{E}\left|\boldsymbol{h}_j^\dagger B_{1j}^{-2} \boldsymbol{h}_j - \frac{1}{N}\mathsf{Tr}\left(B_{1j}^{-2}\right)\right|^2 \\
&\quad + 3\mu^{-4}\mathsf{E}\left|\boldsymbol{h}_j^\dagger B_{1j}^{-1} \boldsymbol{h}_j - \frac{1}{N}\mathsf{Tr}\left(B_{1j}^{-1}\right)\right|^2 + 3\mu^6 \mathsf{E}|\epsilon_j|^2 \\
&\le C/N
\end{aligned}$$

which shows that $\mathsf{E}(|U_2|^2)$ is bounded by a constant.

To verify that $\mathsf{E}(|U_3|^2)$ is bounded by a constant, we need only to note that

$$\left\|B_{1j}^{-1} H_{1j} D_{1j} H_{1j}^\dagger B_{1j}^{-1}\right\| \le M\mu^{-2}$$

and $\alpha_j \boldsymbol{h}_j^\dagger B_{1j}^{-2} \boldsymbol{h}_j < \mu^{-2}$. Similarly, we may verify that $\mathsf{E}(|U_4|^2)$ and $\mathsf{E}(|U_5|^2)$ are bounded.

As for estimating the variance of $\mathsf{Tr}\left[B_1^{-2}\right]$, we only need to repeat the previous steps for $\mathsf{Tr}\left[B_1^{-2} H_1 D_1 H_1^\dagger\right]$. Therefore

$$\limsup_{N\to\infty} \mathrm{Var}(\mathsf{Tr}[B_1^{-2}]) < \infty. \square$$

APPENDIX E
PROOF OF LEMMA 3.5

Proof: For any function $f(z)$ analytic in a region containing $[0, b+\epsilon]$ as inner points

$$\mathsf{E}\left[\mathsf{Tr}\left(f\left(H_1 H_1^\dagger\right)\right)\right] - (K-1)\int f(x) dF_{y_N}(x) = \frac{1}{2\pi i}\oint_{\mathcal{C}} M_N^2(z) f(z) dz$$

where $\mathcal{C}$ is a contour enclosing $[0, b+\epsilon]$

$$M_N^2(z) = \mathsf{E}[\mathsf{Tr}((H_1 H_1^\dagger - z)^{-1})] - (K-1)\int \frac{dF_{y_N}(x)}{x - z}$$

and $b = (1-\sqrt{y})^2$ which was defined in (19). Applying the result in [19, Sec. 4], we have

$$\mathsf{E}\left[\mathsf{Tr}\left(f\left(H_1 H_1^\dagger\right)\right)\right] - (K-1)\int f(x) dF_{y_N}(x) \to \text{a limit}. \tag{64}$$

Then the conclusion (32) follows from (64) with $f(x) = (x+\mu^2)^{-2}$.

Similarly, one can prove that for any positive integer t

$$\mathsf{E}\left[\mathsf{Tr}\left(\left(H_1 H_1^\dagger + \mu^2 I\right)^{-t}\right)\right] - (K-1)\int \frac{1}{(x+\mu^2)^t} dF_{y_N}(x) \to \text{a limit} \tag{65}$$

which is true when H_1 is replaced by H_{1j}.

By the Cauchy–Schwartz inequality and Lemma 3.4, for $t = 1, 2$, we have

$$\mathsf{E}\left|\boldsymbol{h}_2 B_{12}^{-t} \boldsymbol{h}_2 - \frac{1}{N}\mathsf{E}\left[\mathsf{Tr}\left(B_{12}^{-t}\right)\right]\right| = O(N^{-1/2}).$$

Now, let $\tilde{\alpha}_j = 1/\left(1 + \frac{1}{N}\mathsf{E}\left[\mathsf{Tr}\left(B_{12}^{-2}\right)\right]\right)$ and $\tilde{\epsilon}_2 = \boldsymbol{h}_2 B_{12}^{-1} \boldsymbol{h}_2 - \frac{1}{N}\mathsf{E}\left[\mathsf{Tr}\left(B_{12}^{-1}\right)\right]$. By the identity $\alpha_2 = \tilde{\alpha}_2 - \tilde{\alpha}_2^2 \epsilon_2 + \alpha_2 \tilde{\alpha}_2^2 \epsilon_2^2$ and the fact that $\alpha_2 \le 1$ and $\alpha_2 \boldsymbol{h}_2^\dagger B_{12}^{-2} \boldsymbol{h}_2 \le \mu^{-2}$, we have

$$\begin{aligned}
\mathsf{E}\left[\mathsf{Tr}\left(B_1^{-2}\left(H_1 D_1 H_1^\dagger\right)\right)\right] &= \sum_{j=2}^{K} p_j \mathsf{E}\left[\alpha_1^2 \boldsymbol{h}_j^\dagger B_{1j}^{-2} \boldsymbol{h}_j\right] \\
&= (K-1)\mathsf{E}\left[\alpha_2^2 \boldsymbol{h}_2 B_{12}^{-2} \boldsymbol{h}_2\right] \\
&= (K-1)\tilde{\alpha}_2^2 \frac{1}{N}\mathsf{E}\left[\mathsf{Tr}\left(B_{12}^{-2}\right)\right] \\
&\quad + O(1).
\end{aligned}$$

Applying (65) to the above estimate, we complete the proof of (33). $\square$

APPENDIX F
PROOF OF THEOREM 3.3

Proof: In the sequel, we will use the notation $Z_n = O_p(a_n)$ if for any $\varepsilon > 0$ there exists a positive number a so that $\lim_{n\to\infty} P\left(\left|a_n^{-1} Z_n\right| \ge a\right) < \varepsilon$ and the notation $Z_n = o_p(a_n)$ if $\lim_{n\to\infty} P(|a_n^{-1} Z_n| \ge \varepsilon) = 0$. Since the norms of the matrices involved are bounded, we have

$$\boldsymbol{h}_1^\dagger B_1^{-1} \boldsymbol{h}_1 - \frac{1}{N}\mathsf{Tr}\left(B_1^{-1}\right) = O_p(N^{-1/2})$$

and

$$Y_N - \frac{1}{N}\mathsf{Tr}\left(B_1^{-2}\left(H_1 D_1 H_1^\dagger + \sigma^2 I\right)\right) = O_p(N^{-1/2}).$$

Further, by Theorem 3.1 and Lemma 3.1

$$\begin{aligned}
\gamma_1^{(s)} &= \frac{\left(\frac{1}{N}\mathsf{Tr}\left(B_1^{-1}\right)\right)^2}{\frac{1}{N}\mathsf{Tr}\left[B_1^{-2}\left(H_1 D_1 H_1^\dagger + \sigma^2 I\right)\right]} \\
&\quad - \frac{\gamma_1^{(s)}\left(Y_N - \frac{1}{N}\mathsf{Tr}\left[B_1^{-2}\left(H_1 D_1 H_1^\dagger + \sigma^2 I\right)\right]\right)}{\frac{1}{N}\mathsf{Tr}\left[B_1^{-2}\left(H_1 D_1 H_1^\dagger + \sigma^2 I\right)\right]} \\
&\quad + 2\frac{\left(\frac{1}{N}\mathsf{Tr}\left(B_1^{-1}\right)\right)\left(\boldsymbol{h}_1^\dagger B_1^{-1}\boldsymbol{h}_1 - \frac{1}{N}\mathsf{Tr}\left(B_1^{-1}\right)\right)}{\frac{1}{N}\mathsf{Tr}\left[B_1^{-2}\left(H_1 D_1 H_1^\dagger + \sigma^2 I\right)\right]} \\
&\quad + O_p(1/N) \\
&= \frac{\left(\frac{1}{N}\mathsf{Tr}\left(B_1^{-1}\right)\right)^2}{\frac{1}{N}\mathsf{Tr}\left[B_1^{-2}\left(H_1 D_1 H_1^\dagger + \sigma^2 I\right)\right]} \\
&\quad - \mathcal{K}^2\left(Y_N - \frac{1}{N}\mathsf{Tr}\left[B_1^{-2}\left(H_1 D_1 H_1^\dagger + \sigma^2 I\right)\right]\right) \\
&\quad + 2\mathcal{K}\left(\boldsymbol{h}_1^\dagger B_1^{-1}\boldsymbol{h}_1 - \frac{1}{N}\mathsf{Tr}\left(B_1^{-1}\right)\right) + o_p(N^{-1/2}).
\end{aligned} \tag{66}$$

Thus, to complete the proof of the theorem, we only need to prove that

$$\frac{\left(\frac{1}{N}\mathsf{Tr}\left(B_1^{-1}\right)\right)^2}{\frac{1}{N}\mathsf{Tr}\left[B_1^{-2}\left(H_1 D_1 H_1^\dagger + \sigma^2 I\right)\right]} - \beta_N^{(s)} = o(N^{-1/2}), \text{ and} \tag{67}$$

$$\begin{aligned}
\sqrt{N}\Bigg[&2\mathcal{K}\left(\boldsymbol{h}_1^\dagger B_1^{-1}\boldsymbol{h}_1 - \frac{1}{N}\mathsf{Tr}\left(B_1^{-1}\right)\right) \\
&- \mathcal{K}^2\left(Y_N - \frac{1}{N}\mathsf{Tr}\left[B_1^{-2}\left(H_1 D_1 H_1^\dagger + \sigma^2 I\right)\right]\right)\Bigg] \\
&\to \mathcal{N}(0, l^2).
\end{aligned} \tag{68}$$

The conclusion (67) follows from Lemmas 3.4 and 3.5. To prove (68), we consider the real and complex cases separately.

Write the left-hand side of (68) as

$$\frac{1}{\sqrt{N}}\left(\boldsymbol{X}_1^\dagger A_N \boldsymbol{X}_1 - \mathsf{Tr}(A_N)\right)$$

where

$$A_N = (a_{jm}) = 2\mathcal{K}B_1^{-1} - \mathcal{K}^2 B_1^{-1}\left(H_1 D_1 H_1^\dagger + \sigma^2 I\right)B_1^{-1}.$$

It is easy to verify that

$$\begin{aligned}
\mathsf{E}&\left[\frac{1}{\sqrt{N}}\left(\boldsymbol{X}_1^\dagger A_N \boldsymbol{X}_1 - \mathsf{Tr}(A_N)\right)\right]^2 \\
&= \frac{1}{N}\Bigg(\mathsf{Tr}\left(A_N^2\right) + |\mathsf{E}\left(X_{11}^2\right)|^2\mathsf{Tr}\left(A_N A_N^T\right) \\
&\qquad + \sum_{i=1}^{N}|a_{jj}|^2\left(\mathsf{E}\left(|X_{11}|^4\right) - 2 - |\mathsf{E}(X_{11})|^2\right)\Bigg).
\end{aligned} \tag{69}$$

In the real-spreading case, the second term is the same as the first term. For the complex-spreading case, $|\mathsf{E}X_{11}^2|$ may take any value in $[0,1]$. If it is not zero, $\frac{1}{N}\mathsf{Tr}\left(A_N A_N^T\right)$ may not have a limit. For this reason, we need an additional assumption to eliminate this term, that is, $\mathsf{E}\left(X_{11}^2\right) = 0$.

Our strategy is the following: given all A_N, by the result in an accepted paper [30], the conditional distribution of $\frac{1}{\sqrt{N}}(\boldsymbol{X}_1^\dagger A_N \boldsymbol{X}_1 - \mathsf{Tr}(A_N))$ tends to the normal distribution with mean zero and a variance which is independent of A_N, and Theorem 3.3 then follows by the Fubini theorem.

Before concluding this appendix, we shall derive the limits of the first and the third terms in (69) and thus the asymptotic variance of $\sqrt{N}\gamma_1^{(s)}$. Using the Vitali theorem and (48), we have

$$\begin{aligned}
\frac{1}{N}&\mathsf{Tr}\left[B_1^{-3}\left(H_1 D_1 H_1^\dagger\right)\right] \\
&= -\frac{1}{2}\frac{\partial}{\partial\mu^2}\frac{1}{N}\mathsf{Tr}\left[B_1^{-2}\left(H_1 D_1 H_1^\dagger\right)\right] \\
&\to -\frac{1}{2}\frac{\partial}{\partial\mu^2}\left(\frac{y}{(1+\beta)^2}\right)\int\frac{dF_y(x)}{(x+\mu^2)^2} \\
&= \frac{y}{(1+\beta)^2}\int\frac{dF_y(x)}{(x+\mu^2)^3} - \frac{y}{(1+\beta)^3}\left(\int\frac{dF_y(x)}{(x+\mu^2)^2}\right)^2 \\
&= y(1+\beta)\left[\Gamma_3 - \Gamma_2^2\right] \triangleq u
\end{aligned} \tag{70}$$

and furthermore

$$\begin{aligned}
\frac{1}{N}&\mathsf{Tr}\left[B_1^{-4}\left(H_1 D_1 H_1^\dagger\right)\right] \\
&\xrightarrow{a.s.} -\frac{1}{3}\frac{\partial}{\partial\mu^2}\Bigg\{\left(\frac{y}{(1+\beta)^2}\right)\int\frac{dF_y(x)}{(x+\mu^2)^3} \\
&\qquad\qquad - \frac{y}{(1+\beta)^3}\left(\int\frac{dF_y(x)}{(x+\mu^2)^2}\right)^2\Bigg\} \\
&\triangleq v
\end{aligned} \tag{71}$$

where

$$\begin{aligned}
v &= \frac{y}{(1+\beta)^2}\int\frac{dF_y(x)}{(x+\mu^2)^4} \\
&\quad - \frac{2y}{(1+\beta)^3}\int\frac{dF_y(x)}{(x+\mu^2)^2}\int\frac{dF_y(x)}{(x+\mu^2)^3} \\
&\quad + \frac{y}{(1+\beta)^4}\left(\int\frac{dF_y(x)}{(x+\mu^2)^2}\right)^3 \\
&= y(1+\beta)^2\left[\Gamma_4 - 2\Gamma_2\Gamma_3 + \Gamma_2^3\right].
\end{aligned}$$

Using Lemmas 3.2 and 3.3

$$\frac{1}{N}\mathsf{Tr}[B_1^{-2}] \xrightarrow{a.s.} \int\frac{dF_y(x)}{(x+\mu^2)^2} = (1+\beta)^2\Gamma_2 \tag{72}$$

and

$$\begin{aligned}
\frac{1}{N}&\mathsf{Tr}\left[\left(H_1 D_1 H_1^\dagger + \sigma^2 I\right)B_1^{-2}\left(H_1 D_1 H_1^\dagger + \sigma^2 I\right)B_1^{-2}\right] \\
&\xrightarrow{a.s.} m + 2\sigma^2 v + \sigma^4\int\frac{dF_y(x)}{(x+\mu^2)^4} \\
&= \frac{\beta^2\Gamma_4}{\Gamma_2^2\mathcal{K}^2} - \frac{4y\beta\Gamma_3}{\mathcal{K}} + 2y\left(y\Gamma_2^3 + \frac{\beta}{\mathcal{K}}\Gamma_2^2\right) + y\Gamma_2^2\vartheta_2.
\end{aligned} \tag{73}$$

Thus, we obtain

$$\begin{aligned}
\frac{1}{N}\mathsf{Tr}\left[A_N^2\right]&\\
\xrightarrow{a.s.} l_2 &= \int \frac{4\mathcal{K}^2 dF_y(x)}{(x+\mu^2)^2} - 4\mathcal{K}^3 u - \int \frac{4\mathcal{K}^3\sigma^2 dF_y(x)}{(x+\mu^2)^3}\\
&\quad + \mathcal{K}^4\left(m + 2\sigma^2 v + \int \frac{\sigma^4 dF_y(x)}{(x+\mu^2)^4}\right)\\
&= \mathcal{K}^2\bigg[4(1+\beta)^2\Gamma_2 - 4\beta(1+\beta)\Gamma_3\Gamma_2^{-1}\\
&\quad + 4\mathcal{K}y(1+\beta)\Gamma_2^2 + \beta^2\Gamma_4\Gamma_2^{-2} - 4\mathcal{K}y\beta\Gamma_3\\
&\quad + 2y^2\mathcal{K}^2\Gamma_2^3 + 2y\mathcal{K}\beta\Gamma_2^2 + y\mathcal{K}^2\Gamma_2^2\vartheta_2\bigg]. \qquad (74)
\end{aligned}$$

It is easy to see that if $D_1 = I$ and $\sigma^2 = \mu^2$, $\mathcal{K} = 1$ and the sum of the first three terms and that of the last six terms in (74) are equal to 0 and $(1+\beta)^2\Gamma_2$, respectively. Thus, we have

$$\Gamma_3\beta = y\Gamma_2^3 + (1+\beta)\Gamma_2^2$$

and

$$(1+\beta)^2\Gamma_2 = \beta^2\Gamma_4\Gamma_2^{-2} - 4y\beta\Gamma_3 + 2y^2\Gamma_2^3 + 2y\beta\Gamma_2^2 + y\Gamma_2^2.$$

Since these equalities are independent of D_1 and σ^2, they are true in the general case. Thus

$$\begin{aligned}
l_2 = \mathcal{K}^2\bigg[&(1+\beta)^2\Gamma_2 + 2y^2(\mathcal{K}-1)^2\Gamma_2^3 + 2y(\mathcal{K}-1)\beta\Gamma_2^2\\
&+ y\Gamma_2^2(\mathcal{K}^2\vartheta_2 - 1)\bigg]. \qquad (75)
\end{aligned}$$

Next, we need to determine the limit of the third term in (69), that is, the limit of $\sum |a_{jj}|^2$.

By Lemma 3.1, we have

$$\sum_{j=1}^{N}\left(\left(H_1H_1^\dagger + \mu^2 I\right)^{-1}\right)_{jj}^2 \xrightarrow{a.s.} \beta^2$$

$$\sum_{j=1}^{N}\left(B_1^{-1}\right)_{jj}\left(B_1^{-1}\left(H_1H_1^\dagger + \sigma^2 I\right)B_1^{-1}\right)_{jj} \xrightarrow{a.s.} \beta^2\mathcal{K}^{-1}$$

$$\sum_{j=1}^{N}\left(B_1^{-1}\left(H_1H_1^\dagger + \sigma^2 I\right)B_1^{-1}\right)_{jj}^2 \xrightarrow{a.s.} \beta^2\mathcal{K}^{-2}.$$

Therefore, we finally have

$$\frac{1}{N}\sum_{j=1}^{N}|a_{jj}|^2 \xrightarrow{a.s.} \mathcal{K}^2\beta^2 = l_1. \square$$

REFERENCES

[1] D. W. Bliss, K. W. Forsthe, and A. F. Yegulalp, "MIMO communication capacity using infinite dimension random matrix eigenvalue distributions," in *Proc. 35th Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2001, vol. 2, pp. 969–974.

[2] M. Debbah, W. Hachem, P. Loubaton, and M. Courvill, "MMSE analysis of certain large isometric random precoded systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1293–1311, May 2003.

[3] A. J. Grant and P. D. Alexander, "Random sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2832–2836, Nov. 1998.

[4] D. Guo, S. Verdú, and L. K. Rasmussen, "Asymptotic normality of linear multiuser receiver outputs," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3080–3095, Dec. 2002.

[5] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.

[6] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.

[7] M. L. Honig and R. Ratasuk, "Large-system performance of iterative multiuser decision-feedback detection," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1368–1377, Aug. 2003.

[8] L. G. Trichard, J. S. Evans, and I. B. Collings, "Large system analysis of linear multistage parallel interference cancellation," *IEEE Trans. Commun.*, vol. 50, no. 11, pp. 1778–1786, Nov. 2002.

[9] L. Li, A. M. Tulino, and S. Verdú, "Design of reduced-rank MMSE multiuser detectors using random matrix methods," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 986–1008, Jun. 2004.

[10] Y.-C. Liang, S. Sun, and C. K. Ho, "Block-iterative generalized decision feedback equalizers (BI-GDFE) for large MIMO systems: Algorithm design and asymptotic performance analysis," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2035–2048, Jun. 2006.

[11] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Found. Trends in Commun. and Inf. Theory*, vol. 1, no. 1, 2004.

[12] D. N. C. Tse and S. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 641–657, Mar. 1999.

[13] D. N. C. Tse and O. Zeitouni, "Linear multiuser receivers in random environments," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 171–188, Jan. 2000.

[14] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 622–640, Apr. 1999.

[15] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, Jul. 1999.

[16] J. Zhang, E. Chong, and D. N. C. Tse, "Output MAI distributions of linear MMSE multiuser receivers in DS-CDMA systems," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1028–1144, Mar. 2001.

[17] G. Caire, R. R. Müller, and T. Tanaka, "Iterative multiuser joint decoding: Optimal power allocation and low-complexity implementation," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1950–1973, Sep. 2004.

[18] P. Li, D. Paul, R. Narasimhan, and J. Cioffi, "On the distribution of SINR for the MMSE MIMO receiver and performance analysis," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 271–286, Jan. 2006.

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

[20] Y. Q. Yin, Z. D. Bai, and P. R. Krishanaiah, "On the limit of the largest eigenvalue of the large dimensional sample covariance matrix," *Probab. Theory Related Fields .*, vol. 78, pp. 509–521, 1988.

[21] J. W. Silverstein, "Strong convergence of the limiting distribution of the eigenvalues of large dimensional random matrices," *J. Multivariate Anal.*, vol. 55, pp. 331–339, 1995.

[22] E. C. Titchmarsh, *The Theory of Functions*, 2nd ed. London, U.K.: Oxford Univ. Press, 1939.

[23] G. M. Pan, B. Q. Miao, and B. S. Jin, "Some Limiting Theorems of Some Random Quadratic Forms", submitted for publication.

[24] Z. D. Bai, "Convergence rate of expected spectral distributions of large dimensional random matrices. Part I: Wigner matrices," *Ann. Probab.*, vol. 21, pp. 625–648, 1993.

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

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

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

[28] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[29] M. Marcus, "An eigenvalue inequality for the product of normal matrices," *Amer. Math. Monthly*, vol. 63, pp. 173–174, 1956.

[30] Z. D. Bai and J. F. Yao, "Some limit theorems for the eigenvalues of a kind of sample convariance matrix," *Ann. IHP Probab. Statist.*, 2007, to be published.

[31] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices: A review," *Statistica Sincia*, vol. 9, pp. 611–677, 1999.