Multi-Antenna Channel-Hardening and its Implications for Rate Feedback and Scheduling

BERTRAND M. HOCHWALD
Bell Laboratories
Lucent Technologies
600 Mountain Ave., Rm. 2C-363
Murray Hill, NJ 07974
hochwald@lucent.com

THOMAS L. MARZETTA
Bell Laboratories
Lucent Technologies
600 Mountain Ave., Rm. 2C-373
Murray Hill, NJ 07974
tlm@research.bell-labs.com

VAHID TAROKH
Division of Eng. and App. Sci.
Harvard University
33 Oxford St., Rm. MD-131
Cambridge, MA 02138
vahid@deas.harvard.edu

March 25, 2004

Abstract

Wireless data traffic is expected to grow over the next few years and the technologies that will provide data services are still being debated. One possibility is to use multiple antennas at basestations and terminals to get very high spectral efficiencies in rich scattering environments. Such multiple-input multiple-output (MIMO) channels can then be used in conjunction with scheduling and rate-feedback algorithms to further increase channel throughput. This paper provides an analysis of the expected gains due to scheduling and bits needed for rate feedback.

Our analysis requires an accurate approximation of the distribution of the MIMO channel mutual information. Because the exact distribution of the mutual information in a Rayleigh fading environment is difficult to analyze, we prove a central limit theorem for MIMO channels with a large number of antennas. While the growth in average mutual information (capacity) of a MIMO channel with the number of antennas is well understood, it turns out that the variance of the mutual information can grow very slowly or even shrink as the number of antennas grows. We discuss implications of this "channel-hardening" result for data and voice services, scheduling and rate feedback. We also briefly discuss the implications when shadow fading effects are included.

Index Terms—Wireless communications, transmit diversity, receive diversity, fading channels

1 Introduction and Model

Multiple antennas are being considered for high data rate services on scattering-rich wireless channels because they promise high spectral efficiencies [1, 2, 3]. Much recent work has concentrated on the design of codes to achieve some of these high point-to-point rates, but there are also many important multiple-access network design issues that are influenced by the presence of multiple antennas. For example, in data sys-

tems it can be advantageous to schedule delay-tolerant transmissions to users whose channels happen to be good at the moment [4, 5]. Such scheduling techniques require the receiver to measure the channel and feed information about channel conditions back to the transmitter. The transmitter can then adapt its transmission rate to the conditions.

The success of scheduling and rate feedback techniques depends on timely natural or artificially induced fluctuations in the channel. If the channels to all the users are very similar and do not fluctuate much, the scheduling and rate feedback advantages would probably not justify the overhead needed for the receiver to send channel information back to the transmitter. It therefore becomes important to estimate the fluctuations that can be expected in a given scenario.

This paper provides simple but accurate approximations of the scheduling gains that can be expected in a multi-antenna multi-user environment. We obtain closed-form formulas that show how the gains vary as functions of the number of antennas and users. We also obtain closed-form formulas for the number of bits needed for rate control with feedback.

The results in the paper rely on central-limit theorems for the channel mutual information, when it is treated as a random variable, and are therefore asymptotic in the number of antennas or users. But as we show, the results are generally very accurate for a wide range of numbers of antennas or users. The mathematical results include Theorems 1–3 in Section 2, which give the asymptotic distribution of the mutual information with large numbers of transmit antennas, receive antennas, or both. These theorems are then applied to outage probabilities, scheduling gains, and rate feedback in Section 3. Some discussions and extensions appear in Section 4.

We show that as the number of antennas grows, the channel quickly "hardens," in the sense that the mutual information fluctuation as measured by its variance, decreases rapidly relative to its mean. Thus, the gains due to scheduling decrease rapidly with the number of antennas.

1.1 Multiple-Input/Multiple-Output channel

Let s be an $M \times 1$ vector of transmitted symbols, and let y be an $N \times 1$ vector of received signals related by

$$y = Hs + n, \tag{1}$$

where H is an $N \times M$ complex matrix, known perfectly to the receiver, and n is a vector of independent zero-mean complex Gaussian noise entries with variance $1/2$ per real component. Many narrowband flat-fading

space-time transmission schemes can be written in this form. The matrix H is the multiple-input/multiple-output (MIMO) matrix channel, and is assumed to have independent zero-mean complex-Gaussian entries with variance $1/2$ per real component.

In our information-theoretic calculations, we assume that the antennas all transmit independent Gaussian distributed signals with equal average power, and such that the signal-to-noise ratio (SNR) at any receive antenna is ρ, no matter how many transmit antennas M there are. The transmitter does not know H and is therefore not allowed to adapt its transmission strategy in response to H.

The mutual information between a white Gaussian input s and the output y for the MIMO model (1) with a given H is [3]

$$\mathcal{I} = \log\det\left(I_M + \frac{\rho}{M} H^* H\right), \qquad (\text{bits/channel use}) \tag{2}$$

where ρ is the signal-to-noise ratio as physically measured at each receive antenna. In general, because H is a random matrix, the mutual information $\mathcal{I}$ is a random variable measuring bits per channel-use. (In this paper $\log(\cdot)$ denotes the base-two logarithm and $\ln(\cdot)$ denotes the natural logarithm.)

The mean of this random variable $\mu = \mathrm{E}\,\mathcal{I}$ (where the expectation is over the entries of H) is the *ergodic capacity* of the channel. The ergodic capacity μ has operational meaning if H fluctuates significantly during a transmission/reception interval, for then μ can theoretically be attained with a sufficiently long channel code. In some situations, though, H is approximately constant during a transmission interval and our ability to achieve a given rate $\mathcal{I}_0$ depends on whether $\mathcal{I}$ happens to exceed $\mathcal{I}_0$ at the time. In this case, we have the notion of an *outage* [6], where the probability that $\{\mathcal{I} < \mathcal{I}_0\}$ (which we define to be an *outage event*) can be determined from the probability distribution function of $\mathcal{I}$.

There are other applications of the probability distribution of $\mathcal{I}$. Recent researches in scheduling of delay-tolerant traffic in a multi-user network propose using either artificial [5] or naturally occurring [4] channel fluctuations to decide when to schedule transmissions. As these papers show, it can be advantageous to schedule transmissions only to users whose channels happen to be good at the moment. One measure of "channel goodness" is $\mathcal{I}$. The scheduling advantage is then proportional to the fluctuations in $\mathcal{I}$ that are typically encountered.

Channel fluctuations also play a role in the efficacy of "rate feedback," where the transmission rate is adapted to the channel conditions by having the receiver use transmitted training signals to compute $\mathcal{I}$, which is then sent back to the transmitter. The more the dynamic range of $\mathcal{I}$, the more bandwidth that is

needed to quantize and relay it, and the more critical it is for the transmitter to adapt its rate to the condition of the channel.

In this paper, we are interested in computing the distribution of $\mathcal{I}$ and using this distribution to compute scheduling gains and the number of bits needed in rate feedback. Since the distribution is generally rather complicated, we confine our attention principally to Gaussian approximations. When M or N (or both) is large, central-limit theorem arguments show that the distribution of $\mathcal{I}$ approaches a Gaussian distribution. Our simulations show that even when M and N are small, the Gaussian approximation is very accurate. We give accurate estimates of the anticipated gains of scheduling according to the channel fluctuations, and the number of bits needed to implement rate feedback. We are also able to show how to easily predict and compute outage probabilities.

2 Background and Principal Theoretical Results

The ergodic capacity $\mu = \mathrm{E}\,\mathcal{I}$ plays an important role in our study because it represents the mean of the distribution of $\mathcal{I}$. This mean has been studied extensively and we summarize some known results.

Although [1] argues that μ generally grows linearly as the number of transmitter and receiver antennas increases, it is not until [3] that μ is computed explicitly. From (2), we have

$$\begin{aligned} \mu &= \mathrm{E}\,\log\det\left(I_M + \frac{\rho}{M}H^*H\right) \\ &= \mathrm{E}\,\sum_{m=1}^{M}\log\left(1+\rho\lambda_m\right) \\ &= M\mathrm{E}\,\log\left(1+\rho\lambda\right) \end{aligned} \tag{3}$$

where $\lambda_m = \lambda_m(H^*H/M)$ is the mth eigenvalue of the (normalized) *Wishart matrix* H^*H/M [8]. (In equation (3), we implicitly assume that $N \geq M$; if $M > N$, then

$$\mu = \mathrm{E}\,\log\det(I_N + \frac{\rho}{M}HH^*) = N\mathrm{E}\,\log(1+\rho\lambda). \tag{4}$$

The analysis that follows then applies by interchanging M and N.) The expectation in equation (3) is over λ, which has probability density [3]

$$p(\lambda) = \sum_{m=1}^{M}\phi_m^2(M\lambda)(M\lambda)^{N-M}e^{-M\lambda}, \tag{5}$$

where $\phi_{m+1}(\lambda) = \sqrt{m!/(m+N-M)!}L_m^{N-M}(\lambda)$ and $L_m^{N-M}(\lambda) = \frac{1}{m!}e^{\lambda}\lambda^{M-N}\frac{d^m}{d\lambda^m}(e^{-\lambda}\lambda^{N-M+m})$ is the associated Laguerre polynomial of order m [11].

Various asymptotic analyses of the distribution of $\mathcal{I}$ are possible. This paper considers three cases:

(i) large N and fixed M

(ii) large M and fixed N

(iii) large M and N; also specifically with fixed ratio $\beta = N/M$

The three corresponding asymptotic values of μ are known in these cases. We present these now, and then present the corresponding asymptotic distributions of $\mathcal{I}$ computed in this paper.

In case (i), we note that $H^*H/M = H^*H/N \cdot (N/M)$ and the $M \times M$ matrix H^*H/N converges to I_M as $N \to \infty$. (Technical arguments about the nature of convergence are omitted for the sake of brevity; since the elements of H are well-behaved independent Gaussian random variables, the various conditions for strong convergence are readily satisfied.) Therefore, for fixed M, the M eigenvalues of H^*H/N approach one; that is $p(\lambda(H^*H/N))$ approaches $\delta(\lambda - 1)$, and (3) becomes

$$\mu \sim M \log(1 + \frac{N\rho}{M}) \tag{6}$$

as $N \to \infty$ (in the sense that the difference between the two sides tends to zero).

For case (ii), a similar argument when $M \to \infty$ for fixed N shows that $p(\lambda(HH^*/M))$ approaches $\delta(\lambda - 1)$ and (4) becomes

$$\lim_{M\to\infty} \mu = N \log(1 + \rho). \tag{7}$$

For case (iii), where M and N both simultaneously grow large, the analysis is more difficult since the eigenvalues of the matrix H^*H/M do not converge to deterministic quantities. But these eigenvalues do converge in distribution to a known distribution. When $\beta = N/M \geq 1$, [12] (see also [3]) shows that

$$p(\lambda) = \begin{cases} \frac{1}{\pi}\sqrt{\frac{\beta}{\lambda} - \frac{1}{4}\left(1 + \frac{\beta-1}{\lambda}\right)^2} & (\sqrt{\beta}-1)^2 \leq \lambda \leq (\sqrt{\beta}+1)^2 \\ 0 & \text{otherwise} \end{cases}. \tag{8}$$

The expectation in (3) can then be computed in closed form. The result is

$$\mathrm{E}\ \log(1+\rho\lambda) = \frac{1}{\pi}\int_{(\sqrt{\beta}-1)^2}^{(\sqrt{\beta}+1)^2} d\lambda\ \log(1+\rho\lambda)\sqrt{\frac{\beta}{\lambda} - \frac{1}{4}\left(1 + \frac{\beta-1}{\lambda}\right)^2} = F(\beta,\rho), \tag{9}$$

where

$$\begin{aligned} F(\beta,\rho) &= \log\left(1+\rho(\sqrt{\beta}+1)^2\right) + (\beta+1)\log\left(\frac{1+\sqrt{1-a}}{2}\right) \\ &\quad -(\log e)\sqrt{\beta}\frac{1-\sqrt{1-a}}{1+\sqrt{1-a}} + (\beta-1)\log\left(\frac{1+\alpha}{\alpha+\sqrt{1-a}}\right), \end{aligned} \tag{10}$$

and $a = \frac{4\rho\sqrt{\beta}}{1+\rho(\sqrt{\beta}+1)^2}$ and $\alpha = \frac{\sqrt{\beta}-1}{\sqrt{\beta}+1}$. Therefore, from (9) and (3),

$$\mu \sim MF(\beta,\rho). \tag{11}$$

When $\beta < 1$, an analogous result holds

$$\mu \sim NF\left(\frac{1}{\beta},\beta\rho\right) = MF(\beta,\rho), \tag{12}$$

because $F(\beta,\rho) = \beta F(1/\beta,\beta\rho)$. See, for example, [13] or [14] for these derivations.

2.1 Principal theoretical results

The principal theoretical results derived in this paper on the asymptotic distribution of $\mathcal{I}$ in cases (i)–(iii) are now given. In summary, whether N or M or both are allowed to grow, $\mathcal{I}$ asymptotically has a normal distribution. This is not surprising, because the mathematical operation of $\log\det(\cdot)$ involves an extensive amount of averaging (see, for example, [8] for central limit theorems involving $\log\det(H^*H)$). What is perhaps surprising is the speed with which the variance, relative to the mean, of $\mathcal{I}$ decreases with increasing M or N.

From the considerations in the previous section, the mean of $\mathcal{I}$ in the cases where M or N or both grow have essentially already been computed; it remains to verify the conditions for a central limit theorem and to compute the variances. We state the results as three separate theorems.

Theorem 1 (large N, fixed M): As $N \to \infty$, the mutual information obeys

$$\sqrt{N}\left[\mathcal{I} - M\log\left(1+\frac{N\rho}{M}\right)\right] \xrightarrow{\mathrm{d}} \mathcal{N}(0, M\log^2 e), \tag{13}$$

(where the notation $\xrightarrow{\mathrm{d}}$ means convergence in distribution). Equivalently, for large N, a numerically accurate

approximation of the distribution is given by

$$\mathcal{I} \sim \mathcal{N}\left(M \log\left(1+\frac{N\rho}{M}\right), \frac{M \log^2 e}{N}\right). \tag{14}$$

Theorem 2 (large M, fixed N): As $M \to \infty$, the mutual information obeys

$$\sqrt{M}\left[\mathcal{I} - N\log(1+\rho)\right] \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \frac{N\rho^2 \log^2 e}{(1+\rho)^2}\right). \tag{15}$$

Equivalently, for large M, a numerically accurate approximation is given by

$$\mathcal{I} \sim \mathcal{N}\left(N\log(1+\rho), \frac{N\rho^2 \log^2 e}{M(1+\rho)^2}\right). \tag{16}$$

Theorem 3 (large N and M): (i) **(Low SNR)** As M and N go to infinity,

$$\lim_{\rho\to 0} \frac{1}{\rho \log e}\sqrt{\frac{M}{N}}\left[\mathcal{I} - N\rho\log e\right] \xrightarrow{\mathrm{d}} \mathcal{N}(0,1). \tag{17}$$

This statement says that for small ρ and large M and N,

$$\mathcal{I} \sim \mathcal{N}\left(N\rho\log e, \frac{N}{M}\rho^2 \log^2 e\right). \tag{18}$$

(ii) **(High SNR)** Let $K = \max(M,N)$ and $k = \min(M,N)$. As M and N go to infinity,

$$\lim_{\rho\to\infty} \frac{1}{\sigma_{MN}}\left[\mathcal{I} - \mu_{MN}\right] \xrightarrow{\mathrm{d}} \mathcal{N}(0,1), \tag{19}$$

where

$$\mu_{MN} = k\log(\rho/M) + k\log e\left(\sum_{i=1}^{K-k}\frac{1}{i} - \gamma\right) + \log e \sum_{i=1}^{k-1}\frac{i}{K-i}, \qquad \gamma = 0.5772\ldots \tag{20}$$

$$\sigma^2_{MN} = (\log e)^2\left(\sum_{i=1}^{k-1}\frac{i}{(K-k+i)^2} + k\left[\frac{\pi^2}{6} - \sum_{i=1}^{K-1}\frac{1}{i^2}\right]\right) \tag{21}$$

Note that Theorem 3 does not require M and N to have any specific relationship as they both go to infinity. The special case where $\beta = N/M$ is fixed is presented as a corollary to Theorem 3.

Corollary 1 (large N and M, fixed $\beta = N/M$): (i) **(Low SNR)** For small ρ and large N and M,

$$\mathcal{I} \sim \mathcal{N}(N\rho \log e, \beta\rho^2 \log^2 e). \tag{22}$$

(ii) **(High SNR)** For large ρ and large N and M,

$$\begin{aligned} \mathcal{I} &\sim \mathcal{N}(M\mu(\beta,\rho), \sigma^2(\beta)), \qquad \beta \geq 1 && (23)\\ &\sim \mathcal{N}\left(N\mu(\tfrac{1}{\beta}, \beta\rho), \sigma^2(\tfrac{1}{\beta})\right), \qquad \beta \leq 1 && (24) \end{aligned}$$

where for $\beta \geq 1$,

$$\begin{aligned} \mu(\beta,\rho) &= \log(\rho/e) + 2\log(\sqrt{\beta}+1) + (\beta+1)\log\left(\frac{\sqrt{\beta}}{\sqrt{\beta}+1}\right) \\ &\quad + (\beta-1)\log\left(\frac{\sqrt{\beta}}{\sqrt{\beta}-1}\right), \qquad \beta \neq 1 && (25)\\ \mu(1,\rho) &= \log(\rho/e) \\ \sigma^2(\beta) &= (\log e)^2 \ln\left(\frac{\beta}{\beta-1}\right), \qquad \beta \neq 1 && (26)\\ \sigma^2(1) &= (\log e)^2[\ln N + \gamma + 1] && (27) \end{aligned}$$

The variance of $\mathcal{I}$, along with higher moments, is also computed in a general form in [15] when M and N are simultaneously large. It is observed empirically in [15] that the mean and variance of $\mathcal{I}$ are generally enough to characterize the distribution of $\mathcal{I}$ accurately, even when M and N are "not so large."

Theorems 1–3 and Corollary 1 are proven in Appendix A. The principal applications of these theorems appear in Section 3. We now discuss the numerical implications of these theorems.

2.2 Discussion of Theorems

Theorems 1–3 say that whether M or N (or both) increase, the limiting distribution of $\mathcal{I}$ is Gaussian. We may use the means and variances predicted by these theorems, or, alternatively, we may numerically calculate the mean and variance for any given M, N, and ρ. Then, an excellent approximation of the distribution of $\mathcal{I}$ is given by a Gaussian with this numerical mean and variance. Figure 1 shows that the Gaussian approximation is very accurate for even small M and N. In this figure, the empirical distribution of $\mathcal{I}$ and its Gaussian approximation are overlaid for $M = N = 2$. The Gaussian approximation gets even better as M or N (or both) increase.

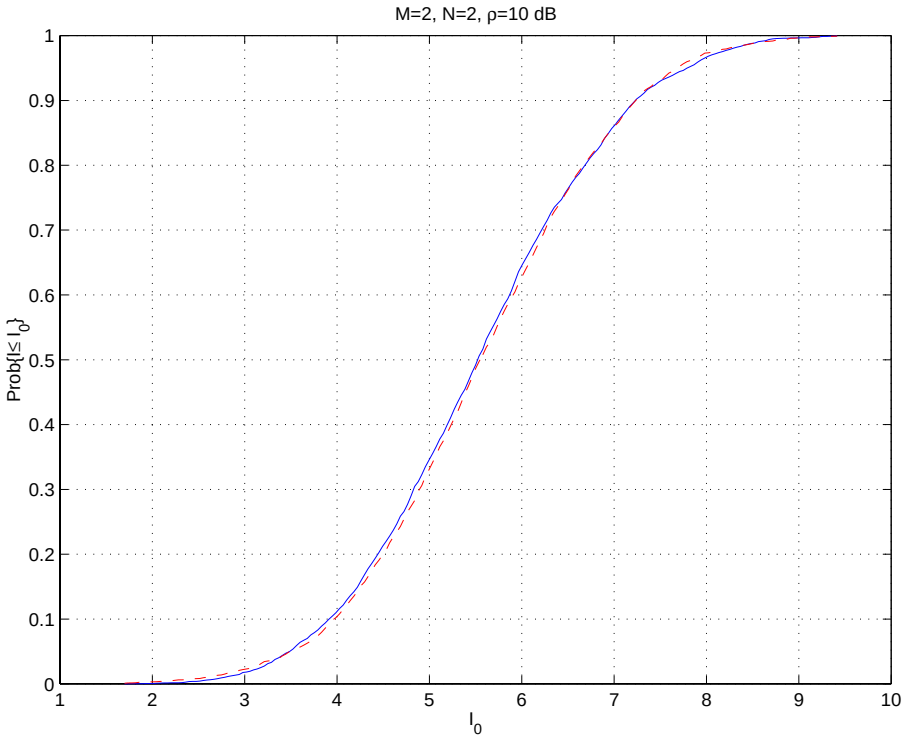


Figure 1: Empirical distribution of $\mathcal{I}$ (solid) versus Gaussian distribution with same mean and variance (dashed) showing how nearly-Gaussian the distribution of $\mathcal{I}$ is for even a relatively small number of antennas ($M = 2$, $N = 2$, $\rho = 10$ dB).

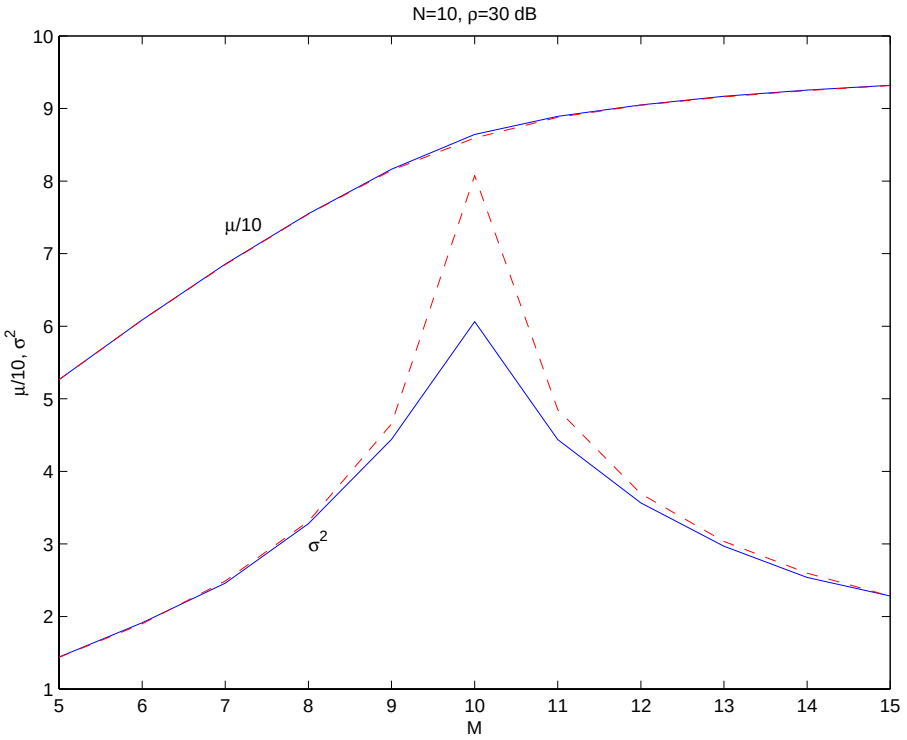


Figure 2: Mean (scaled by a factor of 10) and variance of $\mathcal{I}$ as a function of M for $N = 10$ and $\rho = 30$ dB. The solid lines are simulated values and the dashed lines are computed from equations (20) and (21).

Theorem 1 shows that, as the number of receiver antennas N increases, the average of $\mathcal{I}$ increases logarithmically but its variance decreases as $1/N$. Theorem 2 shows that as M increases, the average approaches a limit and the variance decreases as $1/M$. Theorem 3 and Corollary 1 show that the mean grows linearly as M and N grow simultaneously, but the behavior of the variance as a function of the number of antennas is highly variable. At low SNR, the variance depends linearly on the ratio $\beta = N/M$ and quadratically on ρ. At high SNR, the variance is independent of ρ, but is a constant function of β only when $\beta \neq 1$. When $\beta = 1$, the variance grows logarithmically with the number of antennas.

The accuracy of Theorem 3(ii) (high SNR) is shown in Figure 2, where we plot the mean and variance of $\mathcal{I}$ as a function of M for $N = 10$ and $\rho = 30$ dB. The solid curves represent the simulated mean and variance, and the dashed curves represent the formulas (20) and (21). Observe the "spike" in the variance of $\mathcal{I}$ when $M = 10$. The origin of this spike is made more apparent by Corollary 1(ii). The point $M = 10$ corresponds to $\beta = 1$, and we see from equations (26) and (27) that $\sigma^2(\beta)$ peaks at $\beta = 1$ for N and M sufficiently large. In fact, $\sigma^2(1) = (\log e)^2[\ln N + \gamma + 1]$, whereas $\sigma^2(\beta)$ is bounded for all $\beta \neq 1$.

3 Applications of the Theorems

3.1 Probability of Outage

In cases where the channel H is constant long enough to allow an entire coding/decoding interval, the notion of an outage event is meaningful [6]. A simple application of Theorems 1–3 gives an approximation of the outage probability in terms of the $Q(\cdot)$ function. Assume that we are in a regime where we have an approximate distribution $\mathcal{I} \sim \mathcal{N}(\mu, \sigma^2)$ for some μ and σ^2. Define

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty dt\, e^{-t^2/2}. \tag{28}$$

Let p_0 denote a desired "outage probability" and $\mathcal{I}_0$ denote the corresponding rate where the probability of the outage event $\{\mathcal{I} < \mathcal{I}_0\}$ is p_0. Then

$$p_0 = \int_{-\infty}^{\mathcal{I}_0} d\mathcal{I}\, p(\mathcal{I}) = 1 - Q\left(\frac{\mathcal{I}_0 - \mu}{\sigma}\right), \tag{29}$$

implying that

$$\mathcal{I}_0 = \mu + \sigma Q^{-1}(1 - p_0). \tag{30}$$

Our definition of p_0 differs from [3] where the transmitter has the freedom to use a subset of the given M antennas with an input covariance that is not necessarily diagonal in an attempt to minimize p_0.

The accuracy of the Gaussian approximation for computing outage probabilities can be seen in Figure 3. The figure shows the outage probability as a function of M for $N = 5$ and $\rho = -7$ dB, and for two different rates: $\mathcal{I}_0 = 1.47$ (upper curves) and $\mathcal{I}_0 = 1.14$ (lower curves). The solid and dashed lines almost touching each other across the whole figure are the actual (solid) and Gaussian approximations (dashed) to the outage probabilities. The means and variances of the Gaussian approximations were computed numerically from the simulations. The two remaining solid curves that converge at large M (but diverge for small M) also are Gaussian approximations, but using the mean and variance given in Theorem 2 (large M). As $M \to \infty$, the mutual information $\mathcal{I}$ is approaching the deterministic quantity $N \log(1 + \rho) = 1.32$ bits/channel-use. Therefore, a rate greater than 1.32 will have an outage probability that approaches unity as $M \to \infty$. An rate less than 1.32 will have an outage probability approaching zero.

The number of receive antennas $N = 5$ and the SNR $\rho = -7$ dB in Figure 3 are chosen to exihibit what can happen for "not-so-large" M. For example, we see that there is a "dip" in the outage probability

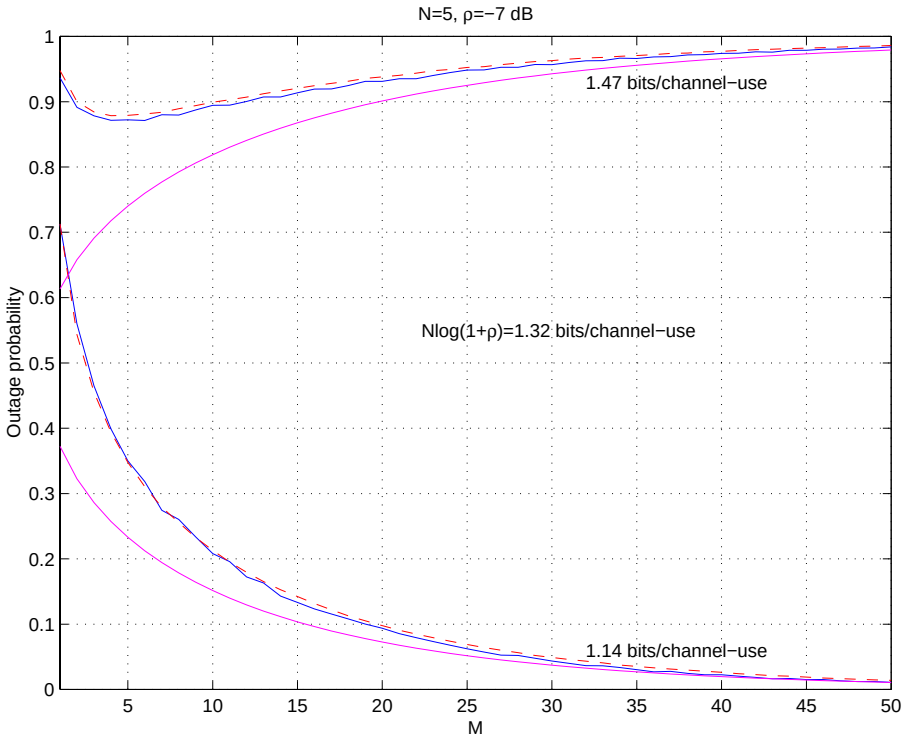


Figure 3: Outage probability as a function of number of transmit antennas M, for $N = 5$ and $\rho = -7$ dB. The two sets of solid and dashed curves following each other closely are the simulated (solid) and Gaussian approximation (dashed) to the outage probability for $\mathcal{I}_0 = 1.47$ (upper) and $\mathcal{I}_0 = 1.14$ (lower) bits/channel-use; the Gaussian approximations used numerically computed means and variances. The two remaining solid curves are also Gaussian approximations, but using the mean and variance of Theorem 2. As $M \to \infty$, the mutual information $\mathcal{I}$ is approaching the deterministic quantity $N \log(1 + \rho) = 1.32$ bits/channel-use. Therefore, $\mathcal{I}_0 > 1.32$ will have an outage probability approaching unity, and $\mathcal{I}_0 < 1.32$ will have an outage probability approaching zero.

for $\mathcal{I}_0 = 1.47$ when $M \approx 5$. Thus, outage probabilities do not necessarily behave monotonically. Yet, as Theorem 2 shows, the monotonically growing mean approaches a limit, while the variance shrinks, as the number of transmit antennas M grows. Therefore, we can infer that for rates less than the limiting mean, the outage probability generally decreases monotonically as more transmit antennas are used. For rates greater than this limiting mean, the outage probability may fluctuate with M, but eventually approaches one as M grows. We therefore answer, in part, the question of whether we can reduce the probability of an outage if we are given an opportunity to use fewer than all M antennas [3]: for rates less than $N \log(1+\rho)$, use all M antennas. Note that data rates of interest are generally much less than $N \log(1+\rho)$ since the outage probabilities of interest are generally much less than $1/2$.

We comment that the accuracy of outage event calculations with our central limit theorem approximations generally decreases as the outage probability decreases because central limit theorems are often not very accurate when used to approximate the tails of an empirical distribution.

3.2 Scheduling Gains

There has been recent interest in applying scheduling algorithms to boost the throughput of data in a multi-user environment. In one possible algorithm, the basestation transmitter obtains feedback from the various users on the conditions of their respective channels. The basestation then transmits to the user with the best channel. This algorithm generally needs the data to be delay-tolerant since it cannot guarantee that a particular user will be served at any given time. The algorithm also requires natural [4] or artificially induced [5] fluctuations in the channel so that bad channels eventually become good, and all users enjoy fair treatment.

We analyze the algorithm by assuming the following: i) the channels between the basestation and the various users have the same number of antennas and are independent; ii) the basestation transmits to the channel with the largest $\mathcal{I}$. We begin with a lemma on the maximum of a sequence of independent Gaussian random variables.

Lemma 1 (maximum of sequence of independent Gaussian random variables): Let $Z_1, \ldots, Z_K$ be a sequence of independent Gaussian random variables with mean μ and variance σ^2. Define $M_K = \max(Z_1, \ldots, Z_K)$. Then

$$\frac{M_K - \mu}{\sqrt{2\sigma^2 \ln K}} \to 1$$

in probability as $K \to \infty$. The proof of this is standard [9, pg. 76] and we omit it.

This lemma says that $M_K \sim \mu + \sqrt{2\sigma^2 \ln K}$ for large K. Applying the lemma to the scheduling algorithm described above, we conclude that the algorithm that transmits to the best user boosts the average data rate per user from μ, which would be obtained from simple round-robin, to a new value which is given approximately by S_K, defined as

$$S_K = \mu + \sqrt{2\sigma^2 \ln K} \tag{31}$$

(assuming everybody's channel is Gaussian and eventually becomes the best). We see that the difference between S_K and μ is significant if either σ^2 is large or K is very large.

Equation (31) is only an approximation of the scheduling gain because it applies asymptotically in K and the distribution of $\mathcal{I}$ is only asymptotically (in the number of antennas) Gaussian. There is potential danger in applying Lemma 1 to a Gaussian distribution obtained via a central limit theorem because the result in the lemma relies on the tail behavior of $Z_1, \ldots, Z_K$, which is outside the domain of most central limit theorems. Since our Gaussian approximations improve with M and N, Lemma 1 can be applied with increasing confidence as M and N grow. Nevertheless, it would be incorrect to draw conclusions by letting $K \to \infty$ without letting M and N go to infinity first. We show later in numerical examples that Lemma 1 may be applied successfully to $\mathcal{I}$ in all of our cases of interest.

3.2.1 Fractional scheduling gain

Define α to be the scheduling gain as a fraction of the mean. Then

$$\alpha = \frac{\sqrt{2\sigma^2 \ln K}}{\mu} \tag{32}$$

The gain of the scheduling algorithm is then given approximately by

$$S_K = \mu(1 + \alpha). \tag{33}$$

We can compute α as a function of the number of antennas and SNR using Theorems 1–3.

We apply Theorem 1 (large N), where $\mu = M \log(1 + N\rho/M)$ and $\sigma^2 = (M/N) \log^2 e$, to (32) to obtain

$$\alpha = \sqrt{\frac{2 \ln K}{NM \ln^2(1 + N\rho/M)}}. \qquad (\text{large } N) \tag{34}$$

Let $K = 4$, $N = 4$, $M = 1$ and $\rho = 10$ dB (four users, each with four receive antennas, and one basestation

antenna, all operating at 10 dB SNR). Then $\alpha \approx 22\%$. Increasing K to 32 yields $\alpha \approx 35\%$. Increasing N to 8 (with $K = 32$) yields $\alpha \approx 21\%$. Note that our confidence in the accuracy of the computed values of α is higher for larger values of M and N because (32) and Theorems 1–3 apply to large M or N.

If we apply Theorem 2 (large M) to (32), then

$$\alpha = \sqrt{\frac{2\rho^2 \ln K}{NM(1+\rho)^2 \ln^2(1+\rho)}}. \qquad \text{(large } M\text{)} \tag{35}$$

Let $K = 4$, $N = 1$, $M = 4$, and $\rho = 10$ dB (four users, each with one receive antenna, and four basestation antennas). Then $\alpha \approx 32\%$. Increasing K to 32 yields $\alpha \approx 50\%$. Increasing N to 8 (with $K = 32$) yields $\alpha \approx 35\%$.

The scheduling gains are especially pronounced at low SNR and for a small number of antennas. Figure 4 shows the data rate per user for the round-robin, and scheduled system for $K = 4$, $N = 1$, $M = 1, \ldots, 16$ and $\rho = -10$ dB (we choose these values to match Figure 4 in [4]). The lower solid curve is the rate obtained by a round-robin service of the four users, each with one receive antenna, and a basestation with M antennas. This round-robin service rate is given by the ergodic capacity, which for large M is approximately $N \log(1+\rho) \approx 0.138$ (see Theorem 2). The upper solid curve shows the increased rate obtained (through simulation) with the scheduling algorithm. The dashed curve is (33) applied to (35)

$$S_K = N \log(1+\rho) + \sqrt{\frac{2N\rho^2 \log^2 e}{M(1+\rho)^2} \ln K} \tag{36}$$

$$\approx 0.138 + \sqrt{\frac{0.0477}{M}} \tag{37}$$

We see in Figure 4 that (37) predicts that the scheduled rate *decreases* as $1/\sqrt{M}$ as M increases. This gives an analytical basis to the observation in [4] that more transmit antennas M are detrimental when the system is scheduled according to largest $\mathcal{I}$. The overall data rate in this scheduled system falls with M because the variance of $\mathcal{I}$ decreases rapidly with M.

Figure 5, which is the same as Figure 4 except with $K = 32$, shows that the difference between S_K and the true scheduling gain decreases as K increases. In this case,

$$S_K \approx 0.138 + \sqrt{\frac{0.119}{M}} \tag{38}$$

and the scheduling gain is almost 250% for small M, but much less for large M.

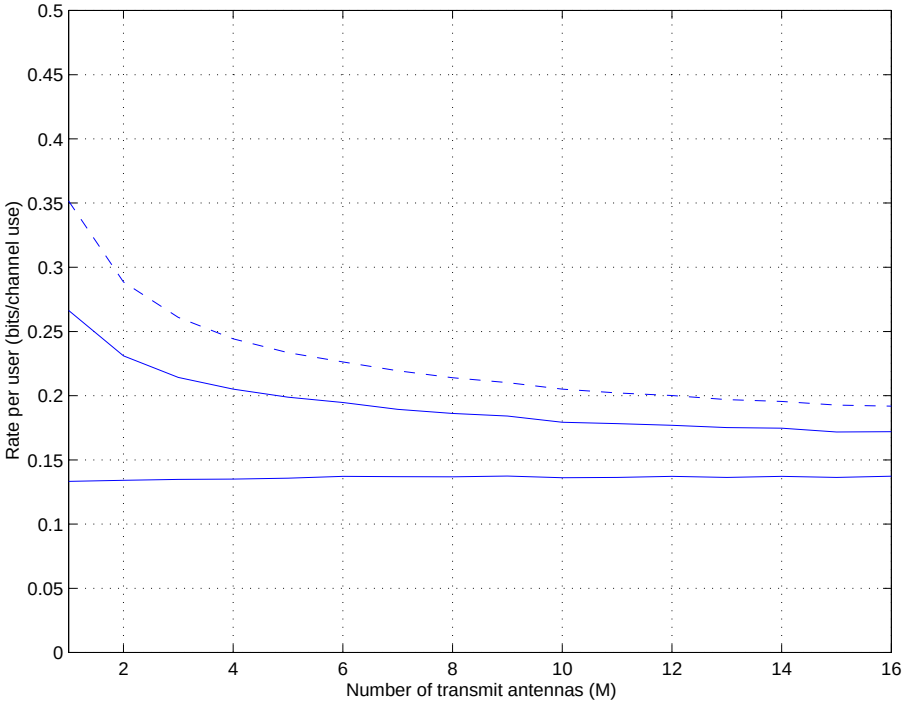


Figure 4: Simulated and computed rates per user for $K = 4$ users, $N = 1$, and $M = 1, \ldots, 16$ with $\rho = -10$ dB. The lower solid curve is the rate obtained by round-robin service, the upper solid curve is the rate obtained by scheduling, and the upper dashed curve is the closed-form approximation (37).

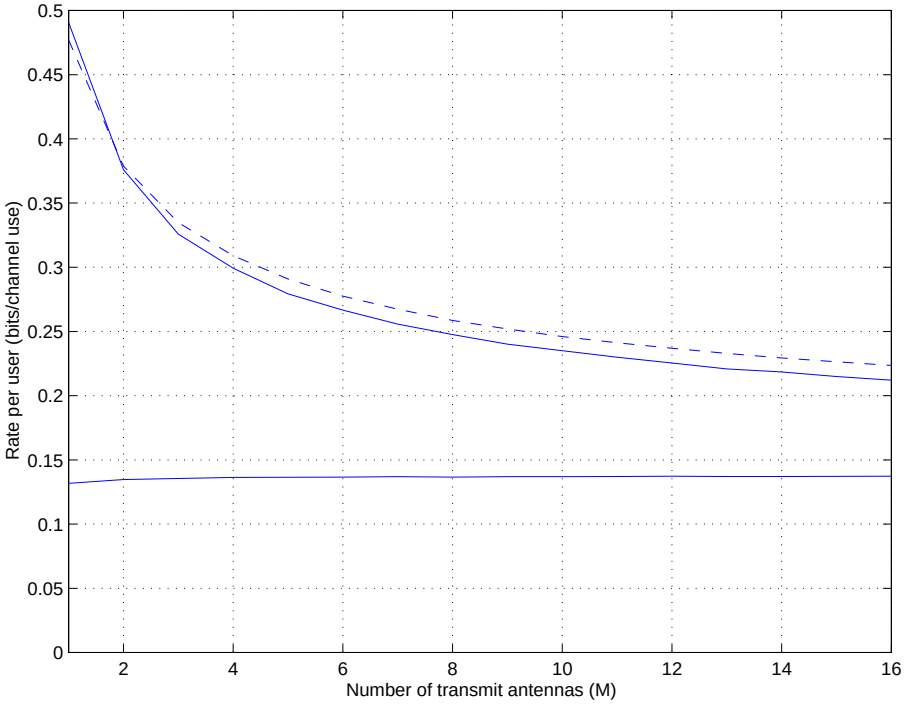


Figure 5: Simulated and computed rates per user for $K = 32$ users, $N = 1$, and $M = 1, \ldots, 16$ with $\rho = -10$ dB. The lower solid curve is the rate obtained by round-robin service, the upper solid curve is the rate obtained by scheduling, and the upper dashed curve is the closed-form approximation (38).

To find the dependence of S_K on the number of users K and number of transmit or receive antennas at low SNR we may apply Corollary 1(i) (low SNR) to (32) to obtain

$$\alpha = \sqrt{\frac{2 \ln K}{NM}}. \tag{39}$$

Observe that the gain decreases as the square-root of either the number of transmit or receive antennas, and is independent of ρ. The gain as a function of K is only as the square-root logarithm of K. This is indeed small!

Another region of interest is when M and N are both allowed to grow, and the SNR is high. In this case, we may apply Corollary 1(ii) (high SNR) to (32) to obtain

$$\begin{aligned} \alpha &= \frac{\sqrt{2(\log e)^2 \ln\left(\frac{1}{1-\beta}\right) \ln K}}{N\mu(1/\beta, \beta\rho)}, \qquad \beta < 1 && (40) \\ &= \frac{\sqrt{2(\ln N + \gamma + 1) \ln K}}{N(\ln \rho - 1)}, \qquad \beta = 1, && (41) \\ &= \frac{\sqrt{2(\log e)^2 \ln\left(\frac{\beta}{\beta-1}\right) \ln K}}{M\mu(\beta, \rho)}, \qquad \beta > 1, && (42) \end{aligned}$$

where $\mu(\beta, \rho)$ is defined in (25). For example, using (33) and (41), we obtain

$$S_K = N \log(\rho/e) + \sqrt{2(\log e)^2 (\ln N + \gamma + 1) \ln K} \tag{43}$$

for $\beta = 1$. A plot of the rate per user obtained by scheduling S_K in equation (43) is given in Figure 6 for $M = N = 1, \ldots, 12$, $K = 32$, and $\rho = 20$ dB. In this case, (43) becomes

$$S_K = 5.20N + \sqrt{14.43 \ln N + 22.75}. \tag{44}$$

The behavior of this scheduled system is markedly different from Figures 4 and 5, where the rate falls with increasing M. In Figure 6, the rate obtained by round-robin increases linearly with $M = N$. Equation (44) shows that the rate obtained by scheduling increases slightly faster than linearly with $M = N$. Observe that the dashed curve representing (44) becomes a more accurate approximation of the simulated scheduling gain when M is large.

In Figure 6 and equation (44), the rate obtained by scheduling at high SNR grows slightly faster than

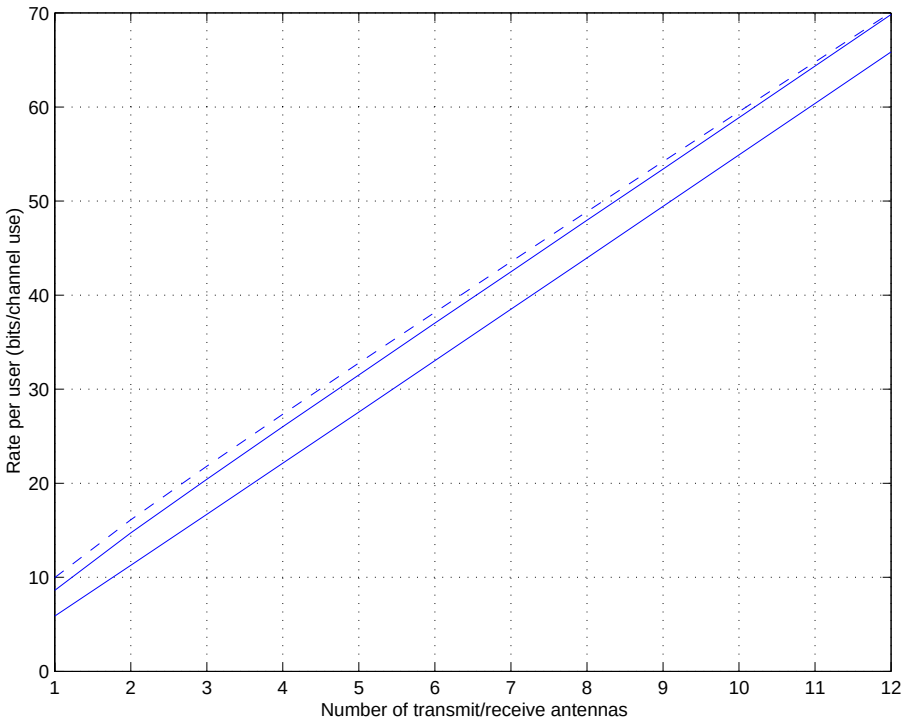


Figure 6: Simulated and computed rates per user for $K = 32$ users and $M = N = 1, \ldots, 12$ with $\rho = 20$ dB. The lower solid curve is the rate obtained by round-robin, the upper solid curve is the simulated rate obtained by scheduling, and the upper dashed curve is the closed-form approximation (44).

linearly, but this behavior is peculiar to $M = N$ (or $\beta = 1$). When $\beta \neq 1$, the scheduled and round-robin rates both grow linearly. This distinction between $\beta = 1$ and $\beta \neq 1$ is due to the behavior of the variance of $\mathcal{I}$ in Corollary 1(ii) (high SNR). The variance $\sigma^2(\beta)$ of $\mathcal{I}$ grows logarithmically with N when $\beta = 1$, and is bounded when $\beta \neq 1$. This effect is also seen in the equations for α in (40)–(42), where the logarithmic growth of $\sigma^2(\beta)$ when $\beta = 1$ appears in the numerator of (41).

3.2.2 Number of users needed to achieve prescribed scheduling gain

We may also ask how many users are needed to realize a certain gain α. Solving (32) for K yields the result

$$K = \exp\left(\frac{\alpha^2}{2}\frac{\mu^2}{\sigma^2}\right).$$

As we show, K grows very rapidly with the number of transmit or receive antennas.

We apply Theorem 1, for example, where $\mu = M\log(1 + N\rho/M)$ and $\sigma^2 = (M/N)\log^2 e$ for large N. Thus, for a given α,

$$K = \exp\left(\frac{\alpha^2}{2}NM\ln^2(1 + N\rho/M)\right).$$

Thus, K grows (slightly faster than) exponentially with N. Let $\alpha = 25\%$, $\rho = 10$ dB, $M = 1$ and $N = 4$. Then $K \approx 6$. Increase α to 50% and then $K \approx 1000$. Increase N to 8 ($\alpha = 25\%$) and then $K \approx 125$.

When we apply Theorem 2, the result is

$$K = \exp\left(\frac{\alpha^2}{2}NM\frac{(1+\rho)^2}{\rho^2}\ln^2(1+\rho)\right).$$

Let $\alpha = 25\%$, $\rho = 10$ dB, $M = 4$ and $N = 1$. Then $K \approx 2$; we add the caveat that computations that yield very small values of K should probably not be trusted too much because Lemma 1 applies most accurately to larger K. Increase α to 50% and then $K \approx 32$. Increase M to 8 ($\alpha = 25\%$) and then $K \approx 6$.

Since the scheduling gains tend to be most pronounced at low SNR, we apply Corollary 1(i) (low SNR), to obtain

$$K = \exp\left(\frac{\alpha^2}{2}NM\right). \tag{45}$$

This result is independent of the SNR. Since the number of users is integer, we should round computed values of K to the nearest integer. For example, when $\alpha = 50\%$ and $N = 1$ we find

$$K = \lceil \exp(M/8) \rceil. \tag{46}$$

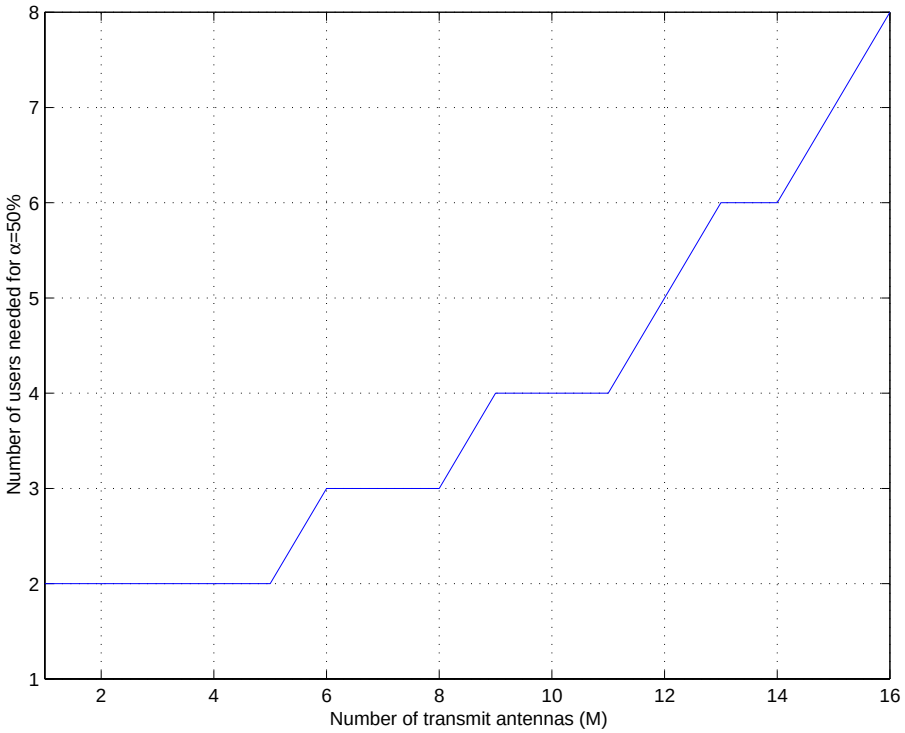


Figure 7: Number of users needed to achieve a gain of $\alpha = 50\%$ as a function of the number of transmit antennas M at low SNR, equation (46).

Figure 7 shows K as computed in (46) as a function of the number of transmit antennas M.

3.3 Rate Feedback

In situations such as the scheduling algorithm described in Section 3.2, the receiver measures the channel conditions and computes the transmission rate $\mathcal{I}$ that it knows the channel can support. But rather than feed this exact value of $\mathcal{I}$ back to the transmitter, the receiver, in practice, generally quantizes $\mathcal{I}$ and feeds this quantized rate back to the transmitter. The selection of quantized transmission rates is kept in a predetermined list; the receiver simply chooses the closest list entry that is less than $\mathcal{I}$.

Generally, rate feedback is most important when there are significant channel fluctuations that the transmitter needs to know about. We call a transmission rate fluctuation "significant" if it is some fixed percentage of the mean rate for the channel. In the spirit of Section 3.2, let α be the percentage of the mean rate that we consider significant (we can also call α the "granularity").

To provide the feedback, we generate a finite list of possible rates as

$$\mathcal{L}_n = \{\mu(1-n\alpha), \ldots, \mu(1-\alpha), \mu, \mu(1+\alpha), \ldots, \mu(1+(n-1)\alpha)\}. \tag{47}$$

The size of the table is $2n$; the receiver chooses rate $\mu(1+i\alpha)$ whenever $\mu(1+i\alpha) \leq \mathcal{I} < \mu(1+(i+1)\alpha)$. By making the upper end of the list one entry shorter than the lower end, we can declare $\mathcal{I}$ out of range if $|\frac{\mathcal{I}}{\mu} - 1| > n\alpha$. While an $\mathcal{I}$ that is too high poses no transmission problem, an $\mathcal{I}$ that is too low is an "outage event" that should be avoided. Therefore, in practice, n should be chosen such that, with some high probability p_0, the table covers the rates typically encountered. We assume some type of gray-code assignment of bits to the list entries is used. The number of bits needed is then $b = \log 2n$.

We show how to compute b for $\mathcal{L}_n$ as a function of the number of antennas, average SNR ρ, and granularity α. Let α be chosen. We employ the theorems of Section 2.1 and therefore assume that the random variable $\mathcal{I}$ has a Gaussian distribution. The list $\mathcal{L}_n$ (47) should be chosen large enough so that $\mathrm{P}\left(|\frac{\mathcal{I}}{\mu} - 1| \leq n\alpha\right) \geq p_0$. The quantity $1 - p_0$ is the probability that $\mathcal{I}$ exceeds the table's range (on either the upper or lower end), and therefore the quantity $\mu + \sigma Q^{-1}(\frac{1-p_0}{2})$ represents the upper limit of $\mathcal{I}$ that needs to be quantized, where $Q(\cdot)$ is defined in (28). Therefore, n must be chosen to solve

$$\mu + \sigma Q^{-1}\left(\frac{1-p_0}{2}\right) = \mu(1+n\alpha),$$

or

$$n = \frac{\sigma Q^{-1}\left(\frac{1-p_0}{2}\right)}{\mu\alpha}. \tag{48}$$

Let $p_0 = 0.95$, which ensures that 95% of the range of $\mathcal{I}$ is covered by $\mathcal{L}_n$. Then $Q^{-1}(\frac{1-p_0}{2}) \approx 2$. It follows from (48) that

$$n = \frac{2\sigma}{\mu\alpha}.$$

Because the number of bits is $b = \log 2n$,

$$b = \log\left(\frac{4\sigma}{\mu\alpha}\right). \tag{49}$$

The higher the required coverage probability p_0, the larger the constant multiplying σ in (49), and the larger the number of bits that are needed.

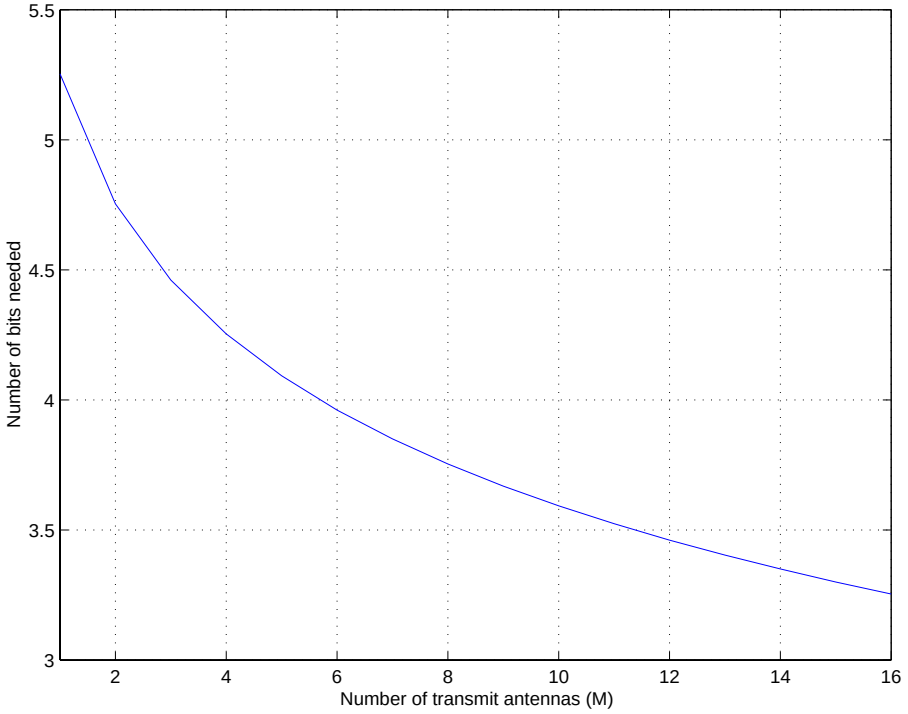


Figure 8: Number of bits needed for rate feedback as a function of the number of transmit antennas M, with $\rho = -10$ dB, and granularity $\alpha = 10\%$, equation (52).

We apply Theorem 2 (large M) to (49) to obtain

$$b = \log\left(\frac{4\rho}{\alpha(1+\rho)\ln(1+\rho)\sqrt{NM}}\right). \tag{50}$$

At low SNR, (50) becomes

$$b = \log\left(\frac{4}{\alpha\sqrt{NM}}\right). \tag{51}$$

As an example, we use the same environment that is used to generate Figures 4 and 5: one receive antenna $N = 1$, a range of transmit antennas $M = 1, \ldots, 16$, $\rho = -10$ dB, and $\alpha = 10\%$. Equation (50) becomes

$$b = 5.25 - \frac{1}{2}\log M \tag{52}$$

Figure 8 shows a plot of equation (52). In general, the number of bits needed as a function of M drops because the distribution of $\mathcal{I}$ gets more concentrated around its mean as M increases.

This effect is seen even more dramatically when the transmit and receive antennas are both allowed to

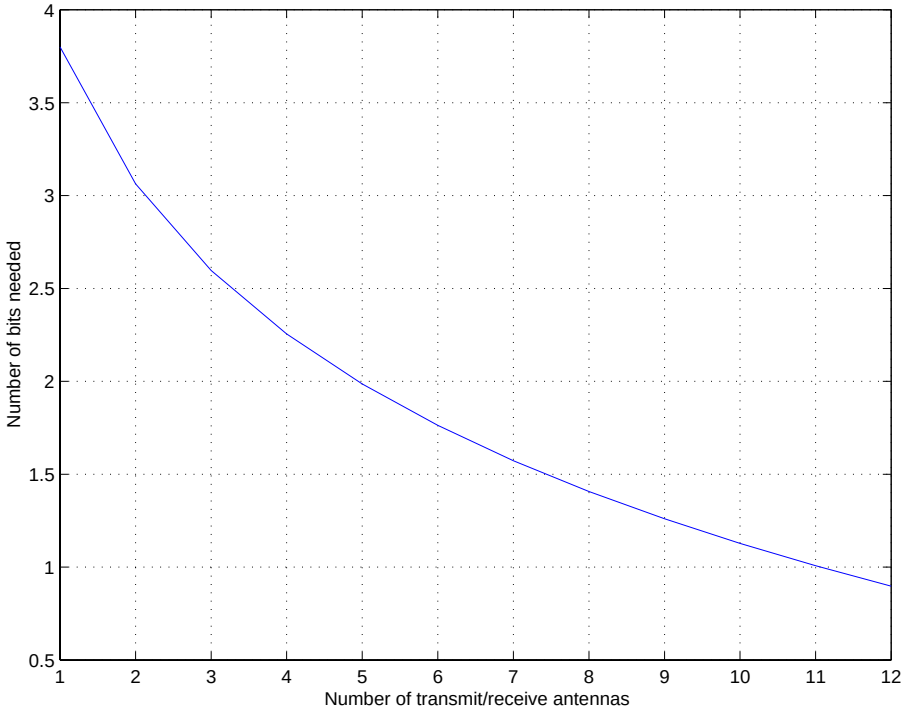


Figure 9: Number of bits needed for rate feedback as a function of the number of transmit/receive antennas, with $\rho = 20$ dB and granularity $\alpha = 10\%$, equation (54).

grow simultaneously. If we assume $M = N$, then applying Corollary 1(ii) (high SNR) to (49) yields

$$b = \log\left(\frac{4\sqrt{\ln N + \gamma + 1}}{\alpha N(\ln \rho - 1)}\right) \tag{53}$$

We apply (53) to the same environment that is used to generate Figure 6: $M = N = 1, \ldots, 12$, $\rho = 20$ dB, and $\alpha = 10\%$ and get

$$b = \log\left(\frac{11.1\sqrt{\ln N + 1.5772}}{N}\right) \tag{54}$$

This equation is plotted in Figure 9. Observe how rapidly the number of bits needed falls off as a function of the number of antennas.

4 Discussions and Extensions

Theorems 1–3 say that the variance of the limiting distribution of the channel mutual information $\mathcal{I}$ decreases relative to the mean as the number of antennas grows. This form of "channel hardening" is generally

welcome for voice and other traffic that is sensitive to channel fluctuations and delay. However, as the previous sections show, channel hardening limits the gains provided by scheduling of delay-tolerant traffic.

Nevertheless, if we modify the pure Rayleigh fading environment used in the previous sections by adding shadow fading, a qualitatively different picture emerges: in shadow fading, both scheduling and MIMO can simultaneously confer benefits.

The presence of shadow fading changes the signal model (1) to

$$y = \sqrt{z} H s + n \tag{55}$$

where z is a positive real random scalar which is independent for different users. Clearly, channel hardening—which for large number of antennas confers immunity to Rayleigh fading—does not extend to shadow fading since a small (bad) value of z affects all transmit-to-receiver gains.

Jakes [20] presents experimental evidence that the shadow fading z is log-normal distributed. Specifically the random variable $10 \log_{10} z$ is distributed as $\mathcal{N}(\mu', \sigma'^2)$. This model was obtained by comparing actual measurements with a *calculated* free-space attenuation. It is generally impractical to *measure* the free-space attenuation in a shadow-fading environment, because the obstacles that lead to shadow fading, such as buildings and mountains, cannot be easily removed!

As it stands the log-normal model is a crude approximation, especially when used to predict the performance of a best-of-K scheduler because of its heavy positive tails. We therefore model the shadow fading as truncated log-normal. Let $\zeta = 10 \log_{10} z$. Then ζ is a nonpositive random variable such that

$$p(\zeta) \propto \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-(\zeta-\mu')^2/(2\sigma'^2)}, & \zeta \leq 0 \\ 0, & \zeta > 0. \end{cases} \tag{56}$$

Figure 10 shows the probability density (in dB space) and the cumulative distribution of the shadow fading model as used in our simulations, where $\sigma' = 10.0$ dB (see [20]), and $\mu' = -6.86$ dB, which corresponds to a median of -10.0 dB.

Figure 11 shows the rate per user obtained, under shadow fading, for the round-robin and scheduled system for $K = 4$, $N = 1$, $M = 1, \ldots, 16$, and $\rho = -10.0$ dB (compare with Figure 4). The effect of shadow fading lowers the throughput for either strategy but there is significant scheduling gain regardless of the number of transmit antennas. This is due to the reduced channel hardening effects.

Figure 12 is similar to Figure 11 except that $K = 32$ (compare with Figure 5). The increased number of

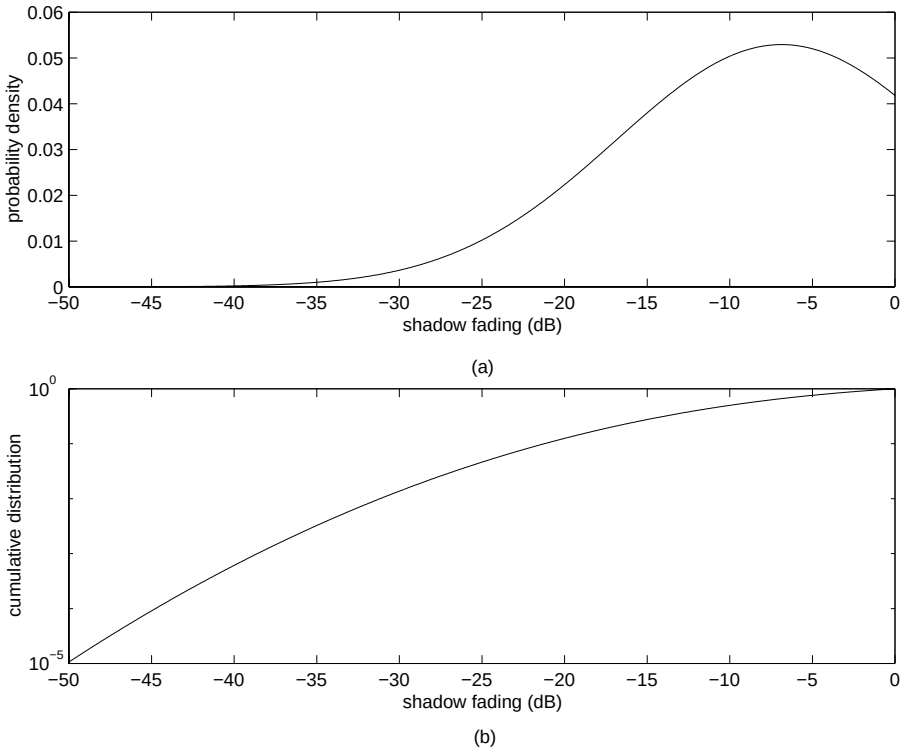


Figure 10: Truncated log-normal shadow fading, median -10.0 dB, untruncated Gaussian standard deviation 10.0 dB: (a) probability density (dB space), (b) cumulative distribution.

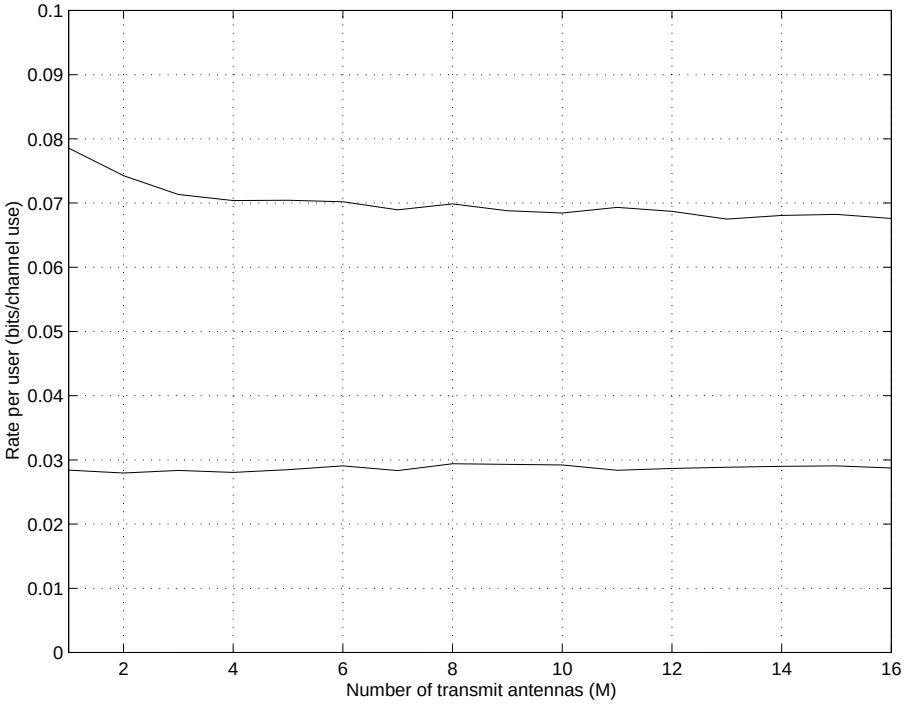


Figure 11: Simulated rate per user, under shadow fading, for $K = 4$ users, $N = 1$, $M = 1, \ldots, 16$, with $\rho = -10.0$ dB. The lower curve is the rate obtained by round-robin, and the upper curve is the rate obtained by scheduling. Compare this figure with Figure 4.

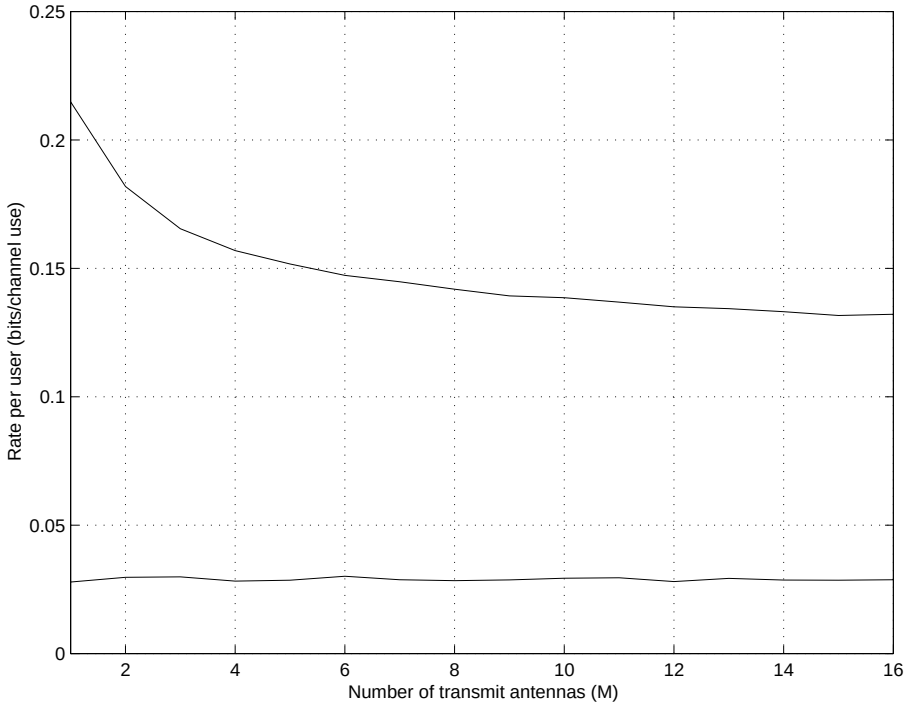


Figure 12: Simulated rate per user, under shadow fading, for $K = 32$ users, $N = 1$, $M = 1, \ldots, 16$, with $\rho = -10.0$ dB. The lower curve is the rate obtained by round-robin, and the upper curve is the rate obtained by scheduling. Compare this figure with Figure 5.

users yields a greater scheduling gain that is again maintained for large M.

Figure 13 shows the rates per user obtained by round-robin and by scheduling under shadow fading for $M = N = 1, \ldots, 12$, $K = 32$, and $\rho = 20.0$ dB (compare with Figure 6). Here we see a nice synergy between multiple antennas and scheduling: both simultaneously yield significant advantages. The rate obtained by scheduling under shadow fading approaches the rate for scheduling without shadow fading because the best of the $K = 32$ users typically suffers little attenuation from shadow fading.

A Proofs of the Theorems

We begin with a lemma that is used in the proofs.

Lemma A (mean and variance of eigenvalues of Wishart matrix): Let H be a $K \times k$ matrix of independent $\mathcal{CN}(0, 1)$ random variables. Define $W = H^*H$ as a $k \times k$ *Wishart* matrix; then W has k

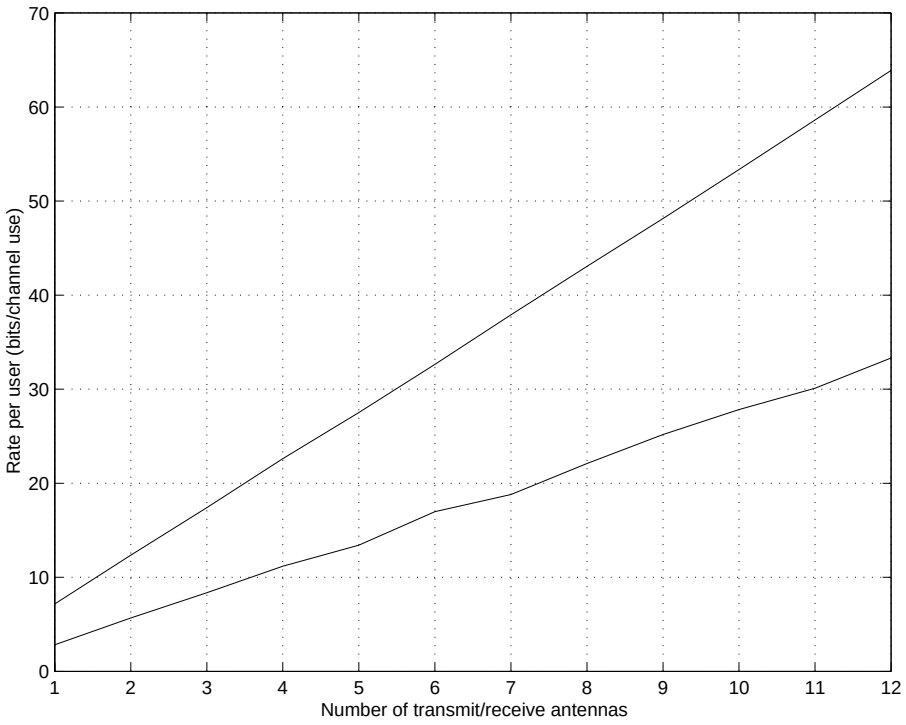


Figure 13: Simulated rate per user, under shadow fading, for $K = 32$ users, $M = N = 1, \ldots, 12$, with $\rho = 20.0$ dB. The lower curve is the rate obtained by round-robin, and the upper curve is the rate obtained by scheduling. Compare this figure with Figure 6.

random eigenvalues $\lambda_1, \ldots, \lambda_k$, and

$$\mathrm{E}\,\mathrm{tr}\,W = kK, \quad \mathrm{E}\,\lambda_i = K \tag{57}$$

$$\mathrm{E}\,\mathrm{tr}\,W^2 = kK(K+k), \quad \mathrm{E}\,\lambda_i^2 = K(K+k) \tag{58}$$

$$\mathrm{E}\,(\mathrm{tr}\,W)^2 = kK(kK+1), \quad \mathrm{E}\,\lambda_i\lambda_j = K(K-1) \qquad (i \neq j). \tag{59}$$

In general, we use this lemma for $K \geq k$, but this restriction is not essential.

Proof: Equation (57) is easily shown by looking at the expected trace of W,

$$\mathrm{E}\,\mathrm{tr}\,W = \sum_{i=1}^{K}\sum_{j=1}^{k}\mathrm{E}\,|h_{ij}|^2 = kK = \sum_{i=1}^{k}\mathrm{E}\,\lambda_i,$$

and therefore $\mathrm{E}\,\lambda_i = K$.

To prove equation (58), we look at the expected trace of W^2,

$$\mathrm{E}\,\mathrm{tr}\,W^2 = \sum_{i=1}^{k}\mathrm{E}\,\lambda_i^2.$$

To compute $\mathrm{E}\,\mathrm{tr}\,W^2$, we look at a typical diagonal entry of W^2. Observe that the matrix W has the structure

$$W = \begin{bmatrix} \sum_{i=1}^K |h_{i1}|^2 & \sum_{i=1}^K h_{i1}^* h_{i2} & \cdots & \\ \sum_{i=1}^K h_{i1} h_{i2}^* & \sum_{i=1}^K |h_{i2}|^2 & \cdots & \\ \vdots & & \ddots & \\ & & & \sum_{i=1}^K |h_{ik}|^2 \end{bmatrix}.$$

Therefore, the first diagonal entry of W^2 is

$$(\sum_{i=1}^{K}|h_{i1}|^2)^2 + \sum_{j=2}^{k}|\sum_{i=1}^{K}h_{i1}h_{ij}^*|^2. \tag{60}$$

The expected value of the first term in (60) is easily computed because $h_{i1}, \ldots, h_{K1}$ are independent of one another. Using the fact that $\mathrm{E}\,|h_{i1}|^4 = 2$, we obtain $\mathrm{E}\,(\sum_{i=1}^K |h_{i1}|^2)^2 = 2K + (K^2 - K) = K^2 + K$. The terms in the sum over j in (60) have identical expected value, so we concentrate on $j = 2$. We note that

$$\mathrm{E}\,|\sum_{i=1}^{K}h_{i1}h_{i2}^*|^2 = \mathrm{E}\,|h_{11}|^2|h_{12}|^2 + \mathrm{E}\,|h_{21}|^2|h_{22}|^2 + \ldots + \mathrm{E}\,h_{11}h_{12}^*h_{21}h_{22}^* + \mathrm{E}\,h_{11}^*h_{12}h_{21}^*h_{22} + \ldots$$

and $\mathrm{E}\,|h_{i1}|^2|h_{i2}|^2 = 1$ whereas the remaining terms have expected value zero. Hence, $\mathrm{E}\,|\sum_{i=1}^{K} h_{i1}h_{i2}|^2 = K$, and the expected value of (60) is $K^2 + K + (k-1)K = K^2 + kK$. The remaining diagonal entries of W^2 have the same expected value, and therefore,

$$\mathrm{E}\,\mathrm{tr}\,W^2 = \sum_{i=1}^{k} \mathrm{E}\,\lambda_i^2 = k(K^2 + kK) = kK(K+k). \tag{61}$$

We conclude that $\mathrm{E}\,\lambda_i^2 = K(K+k)$.

To prove (59), we look at

$$\mathrm{E}\,(\mathrm{tr}\,W)^2 = \mathrm{E}\,(\sum_{j=1}^{k}\sum_{i=1}^{K}|h_{ij}|^2)^2 = \mathrm{E}\,(\sum_{i=1}^{k}\lambda_i)^2.$$

This expected value is easily computed using the identities $\mathrm{E}\,|h_{ij}|^4 = 2$ (of which there are kK terms in the above sum) and $\mathrm{E}\,|h_{ij}|^2|h_{mn}|^2 = 1$ when $(i,j) \neq (m,n)$ (of which there are $k^2K^2 - kK$ terms). Thus, $\mathrm{E}\,(\mathrm{tr}\,W)^2 = 2kK + (k^2K^2 - kK) = k^2K^2 + kK$. We infer that

$$\begin{aligned}
\mathrm{E}\sum_{i\neq j}\lambda_i\lambda_j &= \mathrm{E}\,(\sum_i \lambda_i)^2 - \mathrm{E}\sum_i \lambda_i^2 \\
&= k^2K^2 + kK - kK(K+k) \\
&= kK(k-1)(K-1).
\end{aligned}$$

Because there are $k^2 - k$ terms in the sum $\sum_{i\neq j}\lambda_i\lambda_j$, it follows that $\mathrm{E}\,\lambda_i\lambda_j = K(K-1)$ when $i \neq j$. □

We note that this theorem implies that the eigenvalues have covariance

$$\mathrm{E}\,(\lambda_i - K)(\lambda_j - K) = K^2 - K - K^2 = -K. \tag{62}$$

Lemma B: Let H be a $K \times k$ matrix of independent $\mathcal{CN}(0,1)$ random variables. Then the trace of the Wishart matrix $W = H^*H$ has a normal distribution asymptotically in k or K,

$$\frac{1}{\sqrt{kK}}(\mathrm{tr}\,W - Kk) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1) \tag{63}$$

as $k \to \infty$ or $K \to \infty$ (or both).

Proof: This result is a straightforward application of the central limit theorem. Because $\mathrm{tr}\,W =$

$\sum_{i=1}^{K}\sum_{j=1}^{k}\mathrm{E}\,|h_{ij}|^2$ is a sum of independent identically distributed random variables having well-defined third moments, the distribution of the sum is asymptotically (in either k or K) normal [17]. All that is needed is to subtract the mean of $\mathrm{tr}\,W$ and normalize the difference by the square-root of the variance. The mean and variance of $\mathrm{tr}\,W$ are given by Lemma A. □

A.1 Proof of Theorem 1

We look at the quantity

$$\mathcal{I} = \log\det(I_M + \frac{\rho}{M}H^*H) = \sum_{m=1}^{M}\log\left(1+\frac{\rho N}{M}\lambda_m\right),$$

where $\lambda_1,\ldots,\lambda_M$ are the eigenvalues of H^*H/N. Because $N\to\infty$, we have that $H^*H/N \to I_M$ (strong law of large numbers [17]). Since $\lambda_1,\ldots,\lambda_M$ are continuous functions of the elements of H^*H/N, it follows that $\lambda_m \to 1$ for $m = 1,\ldots,M$. We therefore let $\lambda_m = 1+\tilde{\lambda}_m$, with the understanding that $\tilde{\lambda}_m \to 0$ (almost surely) as $N\to\infty$. Consider the following series of equalities:

$$\begin{aligned}
\sum_{m=1}^{M}\log\left(1+\frac{\rho N}{M}\lambda_m\right) &= M\log\left(1+\frac{\rho N}{M}\right) + \sum_{m=1}^{M}\log\left(\frac{1+\frac{\rho N}{M}\lambda_m}{1+\frac{\rho N}{M}}\right)\\
&= M\log\left(1+\frac{\rho N}{M}\right) + \sum_{m=1}^{M}\log\left(\frac{1+\frac{\rho N}{M}+\frac{\rho N}{M}\tilde{\lambda}_m}{1+\frac{\rho N}{M}}\right)\\
&= M\log\left(1+\frac{\rho N}{M}\right) + \sum_{m=1}^{M}\log\left(1+\frac{\frac{\rho N}{M}\tilde{\lambda}_m}{1+\frac{\rho N}{M}}\right)\\
&= M\log\left(1+\frac{\rho N}{M}\right) + \frac{\frac{\rho N}{M}\log e}{1+\frac{\rho N}{M}}\sum_{m=1}^{M}\tilde{\lambda}_m + O\left(\sum_{m=1}^{M}\tilde{\lambda}_m^2\right) \qquad (64)
\end{aligned}$$

as $N\to\infty$. In equation (64), we say that $x_N = O(y_N)$ if $|x_N| \le cy_N$ for some $c>0$ and all N sufficiently large.

Note that $\sum_{m=1}^{M}\tilde{\lambda}_m = \mathrm{tr}\,(H^*H)/N - M$ has mean zero. Lemma A implies that $\mathrm{E}\sum_{m=1}^{M}\tilde{\lambda}_m^2 = M^2/N$, and therefore the term $O(\sum_{m=1}^{M}\tilde{\lambda}_m^2)$ in (64) has expected value μ_N that is $O(1/N)$ as $N\to\infty$. The expected value of the right-hand side of (64) is therefore $M\log(1+\frac{\rho N}{M}) + O(1/N)$. By Lemma A, $\sum_{m=1}^{M}\tilde{\lambda}_m$ has variance M/N, and fourth-order moment calculations of the Wishart matrix (see [8, pg. 99]) show that the variance of $\sum_{m=1}^{M}\tilde{\lambda}_m^2$ is $O(1/N^2)$. Therefore, $\sum_{m=1}^{M}\tilde{\lambda}_m^2 - \mu_N$ is $O_p(1/N^2)$; this is a probabilistic statement, and we say that $x_N = O_p(y_N)$ if for any $\epsilon>0$, we can find a $c>0$ such that

$\mathrm{P}\left[|x_N| > cy_N\right] \leq \epsilon$ for N sufficiently large; see, for example, [18, pg. 255]. By Lemma B,

$$\sqrt{\frac{N}{M}} \sum_{m=1}^{M} \tilde{\lambda}_m \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$$

as $N \to \infty$. Thus, $\sum_{m=1}^{M} \tilde{\lambda}_m$ is $O_p(1/N)$ and is the dominant random term in (64). The asymptotic distribution of (64) is therefore also normal [18, pg. 254]. Using $\frac{\frac{\rho N}{M}}{1+\frac{\rho N}{M}} = 1 + O(1/N)$, we conclude that

$$\sqrt{N}\left[\sum_{m=1}^{M} \log\left(1 + \frac{\rho N}{M}\lambda_m\right) - M\log\left(1 + \frac{\rho N}{M}\right)\right] \xrightarrow{\mathrm{d}} \mathcal{N}(0, M\log^2 e).$$

This proves (13). □

A.2 Proof of Theorem 2

The proof of this theorem uses the same technique as the proof of Theorem 1, so we omit many details. We look at the quantity

$$\mathcal{I} = \log\det\left(I_N + \frac{\rho}{M}HH^*\right) = \sum_{n=1}^{N} \log\left(1 + \rho\lambda_n\right),$$

where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of HH^*/M. Because $HH^*/M \to I_N$ as $M \to \infty$, we let $\lambda_n = 1 + \tilde{\lambda}_n$ and then use the series of equalities

$$\begin{aligned}
\sum_{n=1}^{N} \log\left(1 + \rho\lambda_n\right) &= N\log(1+\rho) + \sum_{n=1}^{N} \log\left(\frac{1+\rho\lambda_n}{1+\rho}\right) \\
&= N\log(1+\rho) + \sum_{n=1}^{N} \log\left(1 + \frac{\rho\tilde{\lambda}_n}{1+\rho}\right) \\
&= N\log(1+\rho) + \frac{\rho\log e}{1+\rho}\sum_{n=1}^{N}\tilde{\lambda}_n + O\left(\sum_{n=1}^{N}\tilde{\lambda}_n^2\right) \qquad (65)
\end{aligned}$$

as $M \to \infty$.

The mean of $\sum_{n=1}^{N} \tilde{\lambda}_n$ is zero and its variance (by Lemma A) is N/M; the term $O(\sum_{n=1}^{N} \tilde{\lambda}_n^2)$ in (65) is asymptotically negligible. By Lemma B,

$$\sqrt{\frac{M}{N}} \sum_{n=1}^{N} \tilde{\lambda}_n \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$$

as $M \to \infty$. We conclude that

$$\sqrt{M}\left[\sum_{n=1}^{N}\log(1+\rho\lambda_n) - N\log(1+\rho)\right] \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \frac{N\rho^2\log^2 e}{(1+\rho)^2}\right).$$

This proves (15). □

A.3 Proof of Theorem 3

We provide a sketch of the proof of part (i) of this theorem (small ρ). We begin with

$$\mathcal{I} = \log\det\left(I_N + \frac{\rho}{M}HH^*\right) = \sum_{n=1}^{N}\log(1+\rho\lambda_n),$$

where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of HH^*/M. Unlike in Theorems 1 and 2, the matrix HH^*/M does not converge to a deterministic quantity, and likewise $\lambda_1, \ldots, \lambda_N$ also do not converge. Nevertheless, for small ρ we may use the expansion

$$\sum_{n=1}^{N}\log(1+\rho\lambda_n) = \rho\log e\sum_{n=1}^{N}\lambda_n + O(\rho^2). \tag{66}$$

The term $\sum_{n=1}^{N}\lambda_n$ has mean N and variance N/M. Lemma B implies that

$$\sqrt{\frac{M}{N}}\left[\sum_{n=1}^{N}\lambda_n - N\right] \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$$

as M and N simultaneously both go to infinity. This proves part (i).

We now sketch a proof of part (ii). Assume that $M \geq N$ and therefore $K = \max(M,N) = M$ and $k = \min(M,N) = N$. Then

$$\begin{aligned}
\log\det\left(I_N + \frac{\rho}{M}HH^*\right) &= \log\det\frac{\rho}{M}\left(HH^* + \frac{M}{\rho}I_N\right) \\
&= N\log(\rho/M) + \log\det\left(HH^*\left(I + \frac{M}{\rho}(HH^*)^{-1}\right)\right) \\
&= N\log(\rho/M) + \log\det(HH^*) + \log\det\left(I + \frac{M}{\rho}(HH^*)^{-1}\right) \\
&= N\log(\rho/M) + \log\det(HH^*) + O_p(1/\rho). \qquad (67)
\end{aligned}$$

Hence, to first order, the statistics of (67) are determined by the statistics of $\log\det(HH^*)$.

We use the decomposition $HH^* = LL^*$ where L is a lower-triangular matrix whose diagonal entries $\ell_{ii} > 0$ are independent random variables with distribution $\ell_{ii}^2 \sim \frac{1}{2}\chi^2_{2(M-i+1)},\ i = 1,\ldots,N$ [8, Thm. 3.3.4]. Therefore,

$$\log\det(HH^*) = \sum_{i=1}^{N} \log\left(\frac{1}{2}\chi^2_{2(M-i+1)}\right), \tag{68}$$

where the summands are statistically independent. We compute the mean and variance of this sum by first computing the mean and variance of the logarithm of a $\frac{1}{2}\chi^2_{2(M-i+1)}$ random variable. Let $X \sim \frac{1}{2}\chi^2_{2m}$. Then X has density

$$p(x) = \frac{1}{(m-1)!}e^{-x}x^{m-1}.$$

The following moments are easily verified using the integral tables in [19, pp. 607,954,1101]:

$$\mu := \mathrm{E}\,\ln X = \frac{1}{(m-1)!}\int_0^\infty dx\, e^{-x}x^{m-1}\ln x \;=\; \psi(m) \tag{69}$$

$$\mathrm{Var}\,(\ln X) = \mathrm{E}\,(\ln X - \mu)^2 \;=\; \zeta(2,m) \tag{70}$$

$$\mathrm{E}\,(\ln X - \mu)^4 \;=\; 3\zeta^2(2,m) + 6\zeta(4,m) \tag{71}$$

where $\psi(\cdot)$ is the *digamma* function and $\zeta(\cdot,\cdot)$ is the *Riemann zeta* function given by

$$\psi(m) \;=\; -\gamma + \sum_{j=1}^{m-1}\frac{1}{j} \tag{72}$$

$$\zeta(n,m) \;=\; \sum_{j=0}^{\infty}\frac{1}{(m+j)^n}, \tag{73}$$

where $\gamma = 0.5772\ldots$.

Hence,

$$\begin{aligned}\mathrm{E}\,\log\det(HH^*) &= (\log e)\sum_{i=1}^{N}\psi(M-i+1) = (\log e)\sum_{i=1}^{N}\left(-\gamma + \sum_{j=1}^{M-i}\frac{1}{j}\right)\\ &= (\log e)\left[-N\gamma + N\sum_{i=1}^{M-N}\frac{1}{i} + \sum_{i=1}^{N-1}\frac{i}{M-i}\right].\end{aligned}$$

Combining this last equation with (67) gives the expression for μ_{MN} in (20) for $M \geq N$. (We omit the

computation of μ_{MN} for $N > M$.) Furthermore, (67) and (70) yield

$$\begin{aligned}
\sigma_{MN}^2 &= \operatorname{Var}\left(\log\det(HH^*)\right) = (\log e)^2 \sum_{i=1}^{N} \zeta(2, M-i+1) && (74)\\
&= (\log e)^2 \sum_{i=1}^{N}\sum_{j=0}^{\infty} \frac{1}{(M-i+1+j)^2} \\
&= (\log e)^2 \left[N \sum_{i=M}^{\infty} \frac{1}{i^2} + \sum_{i=1}^{N-1} \frac{i}{(M-N+i)^2} \right] \\
&= (\log e)^2 \left[N\zeta(2, M) + \sum_{i=1}^{N-1} \frac{i}{(M-N+i)^2} \right] && (75)\\
&= (\log e)^2 \left(N \left[\frac{\pi^2}{6} - \sum_{i=1}^{M-1} \frac{1}{i^2} \right] + \sum_{i=1}^{N-1} \frac{i}{(M-N+i)^2} \right), && (76)
\end{aligned}$$

because $\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$. Equation (76) coincides with (21) for $M \geq N$. We have therefore computed the mean and variance in Theorem 3(ii).

We have not yet verified that $\log\det(HH^*)$ is asymptotically normal. Denote the summands in (68) as $\ln X_{N1}, \ldots, \ln X_{NN}$ (we switch logarithm bases for convenience). These summands are independent but they are not identically distributed. Furthermore, as M and N grow in a fixed ratio, their distributions change. Fortunately, there are central limit theorems for so-called *triangular arrays* of independent random variables [17, pp. 368–371]. Asymptotic normality (as $N \to \infty$) of the sum $\frac{1}{\sigma_{MN}} \sum_{i=1}^{N} \ln X_{Ni}$ can be verified by confirming that the Lyapounov condition

$$\lim_{N\to\infty} \frac{1}{\sigma_{MN}^{2+\delta}} \sum_{i=1}^{N} \mathrm{E}\,|\ln X_{Ni} - \mu_{Ni}|^{2+\delta} = 0 \qquad (77)$$

is satisfied for some $\delta > 0$. To show that the Lyapounov condition is satisfied in our case, we choose $\delta = 2$ and, for simplicity, $M = N$.

When $M = N$, (75) becomes

$$\sigma_{NN}^2 = (\log e)^2 \left[N\zeta(2, N) + \sum_{i=1}^{N-1} \frac{1}{i} \right]. \qquad (78)$$

We use the expansion [19, pg. 1101]

$$\zeta(n,m) = \sum_{i=0}^{N} \frac{1}{(m+i)^n} + \frac{1}{(n-1)(N+m)^{n-1}} + R_N \tag{79}$$

for any N, where R_N is asymptotically as $N \to \infty$ negligible compared to the first two terms. Equation (79) applied to $\zeta(2,1)$ implies that

$$\zeta(2,N) = \zeta(2,1) - \sum_{i=0}^{N-2} \frac{1}{(i+1)^2} = \frac{1}{N} + R'_N, \tag{80}$$

where R'_N is negligible compared with $1/N$ as $N \to \infty$. Using the expansion $1 + \frac{1}{2} + \ldots + \frac{1}{N} = \ln N + \gamma + o(1)$, we find that (78) becomes

$$\begin{aligned} \sigma^2_{NN} &= (\log e)^2 \left[N \left(\frac{1}{N} + R'_N \right) + \ln N + \gamma + o(1) \right] \\ &= (\log e)^2 [\ln N + \gamma + 1] + o(1). \end{aligned} \tag{81}$$

Therefore, the variance grows logarithmically with N.

Equation (71) implies that (for $M = N$)

$$\mathrm{E}\,(\ln X_{Ni} - \mu_{Ni})^4 = 3\zeta^2(2, N-i+1) + 6\zeta(4, N-i+1), \qquad i = 1, \ldots, N.$$

Inserting (81) into the Lyapounov condition (77) for $\delta = 2$ yields the requirement

$$\lim_{N\to\infty} \frac{1}{\ln^2 N} \sum_{i=1}^{N} [3\zeta^2(2,i) + 6\zeta(4,i)] = 0. \tag{82}$$

Equation (79) applied to $\zeta(4,N)$ shows that

$$\zeta(4,N) = \zeta(4,1) - \sum_{i=0}^{N-2} \frac{1}{(i+1)^4} = \frac{1}{3N^3} + R''_N.$$

Combining this result with (80) implies that the sum $\sum_{i=1}^{N} [3\zeta^2(2,i) + 6\zeta(4,i)]$ converges, and therefore (82) is satisfied.

A.4 Proof of Corollary 1

Part (i) (Low SNR) follows immediately from Theorem 3; β is substituted for N/M in (18).

The proof of part (ii) (high SNR) uses (11) and (10) to find $\mu(\beta,\rho)$ and (74) to find $\sigma^2(\beta)$. We see from (11) that $\mathcal{I}$ has mean $MF(\beta,\rho)$ when N and M are large, where $F(\beta,\rho)$ is given in (10). The formula given for $\mu(\beta)$ in (25) is simply a high SNR approximation of $MF(\beta,\rho)$. For example, the first term in $MF(\beta,\rho)$ is

$$M\log(1+\rho(\sqrt{\beta}+1)^2) = M\left[\log\rho + 2\log(\sqrt{\beta}+1)\right] + O(1/\rho).$$

This expansion gives the first two terms in $\mu(\beta,\rho)$. The remaining terms in $MF(\beta,\rho)$ are handled similarly, and we omit the details.

To find $\sigma^2(\beta)$, we use a large N and M expansion of (74). When $\beta = 1$, we already have the expansion (81) as $N \to \infty$. When $\beta < 1$, we use the approximation (80) to obtain

$$\begin{aligned}
\sigma^2_{MN} &= (\log e)^2 \sum_{i=1}^{N} \zeta(2, M-i+1) \\
&= (\log e)^2 \sum_{i=1}^{\beta M} \left(\frac{1}{M-i+1} + R_{M-i+1}\right) \\
&= (\log e)^2 \left[\frac{1}{M} + \frac{1}{M-1} + \ldots + \frac{1}{M-\beta M+1}\right] + R \\
&= (\log e)^2 \left[\ln M - \ln(M-\beta M)\right] + R \\
&= (\log e)^2 \ln\left(\frac{1}{1-\beta}\right) + R
\end{aligned}$$

where in the fourth equality we used the expansion $1 + \frac{1}{2} + \ldots + \frac{1}{M} = \ln M + \gamma + o(1)$, and the remainder term $R = \sum_i R_{M-i+1}$ is negligible. The case $\beta > 1$ is handled similarly and is omitted.

References

[1] J. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *IEEE J. Sel. Areas Comm.*, vol. 5, pp. 871–878, June 1987.

[2] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[3] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom.*, vol. 10, pp. 585–595, Nov. 1999.

[4] S. Borst and P. Whiting, "The use of diversity antennas in high-speed wireless systems: Capacity gains, fairness issues, multi-user scheduling," *Bell Labs. Tech. Mem.*, 2001. Download available at `http://mars.bell-labs.com`.

[5] P. Viswanath, D. Tse and R. Laroia, "Opportunistic beamforming using dumb antennas," *Submitted to IEEE Trans. Info. Theory*, 2001.

[6] L. H. Ozarow, S. Shamai (Shitz) and A. D. Wyner, "Information theoretic considerations for cellular mobile radio", *IEEE Trans. Veh. Tech.*, vol. 43, pp. 359–378, May 1994.

[7] E. Biglieri, J. Proakis and S. Shamai (Shitz), "Fading Channels: Information-Theoretic and Communications Aspects," *IEEE Trans. Info. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.

[8] A. Gupta and D. Nagar, *Matrix Variate Distributions*, Boca Raton: Chapman & Hall, 2000.

[9] R. Durrett, *Probability: Theory and Examples*, Wadsworth and Brooks/Cole: Pacific Grove, 1991.

[10] H. David, *Order Statistics*, 1st ed., New York: Wiley, 1970.

[11] P. Beckmann, *Orthogonal Polynomials for Engineers and Physicists*, Boulder: Golem Press, 1973.

[12] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *J. Mult. Anal.*, vol. 54, pp. 175–192, 1995.

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

[14] B. Hochwald, T. Marzetta and B. Hassibi, "Space-time autocoding," *IEEE Trans. Info. Theory*, vol. 47, pp. 2761–2781, Nov. 2001.

[15] A. Moustakas, S. Simon and A. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large N analysis," *IEEE Trans. Info. Theory*, pp. 2545–2561, Oct. 2003. Download available at `http://mars.bell-labs.com`.

[16] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley: New York, 1991.

[17] P. Billingsley, *Probability and Measure*, 2nd ed., John Wiley & Sons: New York, 1986.

[18] Y. S. Chow and H. Teicher, *Probability Theory: Independence, Interchangeability, Martingales*, Springer-Verlag: New York, 1988.

[19] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed., Academic Press: San Diego, 1994.

[20] W. C. Jakes (ed.), *Microwave Mobile Communications*, IEEE Press: Piscataway, NJ, 1974.