can be incorporated into a real-time system, the other two schemes may restrict their applications within those admitting unbounded coding/decoding delays. We have analyzed the schemes on the assumption that they produce multisets of integers, but actually they produce ordered linear lists of integers. Thus, there still remains much work for future exploration in considering orders of their elements.

ACKNOWLEDGMENT

This correspondence is an extension of our previous work [9], which has been performed jointly with Shingo Mamada. His contribution to the work should be appreciated. Also, the author would like to thank referees for their constructive comments.

REFERENCES

[1] M. Arimura and H. Yamamoto, "Asymptotic redundancy of the MTF scheme for stationary ergodic sources," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3742–3752, Nov. 2005.
[2] Z. Arnavut, "Inversion coding," *Comput. J.*, vol. 47, no. 1, pp. 46–57, 2004.
[3] J. L. Bentley, D. D. Sleator, R. E. Tarjan, and V. K. Wei, "A locally adaptive data compression scheme," *Commun. ACM*, vol. 29, no. 4, pp. 320–330, 1986.
[4] M. Burrows and D. J. Wheeler, "A block-sorting lossless data compression algorithm," *SRC Res.Report*, vol. 124, May 1994.
[5] T. M. Cover, "Enumerative source encoding," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 1, pp. 73–77, 1973.
[6] S. Deorowicz, "Second step algorithms in the Burrows-Wheeler compression algorithm," *Software–Practice and Experience*, vol. 32, no. 2, pp. 99–111, 2002.
[7] P. Elias, "Interval and recency rank source coding: Two on-line adaptive variable-length schemes," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 1, pp. 3–10, 1987.
[8] J. A. Fill, "An exact formula for the move-to-front rule for self organizing lists," *J. Theor. Probab.*, vol. 9, no. 1, pp. 113–159, 1996.
[9] S. Mamada and H. Yokoo, "A new interval source coding scheme over an ordered alphabet," in *Proc. 2003 IEEE Int. Symp. Inf. Theory*, Yokohama, 2003, p. 21.
[10] B. Y. Ryabko, "Data compression by means of a 'book stack'," *Probl. Info. Transm.*, vol. 16, no. 4, pp. 265–269, 1980.
[11] J. P. M. Schalkwijk, "An algorithm for source coding," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 3, pp. 395–399, 1972.
[12] F. M. J. Willems, "Universal data compression and repetition times," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 54–58, 1989.
[13] A. D. Wyner, J. Ziv, and A. J. Wyner, "On the role of pattern matching in information theory," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2045–2056, 1998.
[14] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 3, pp. 337–343, 1977.

Weak Convergence and Rate of Convergence of MIMO Capacity Random Variable

Vasanthan Raghavan, *Student Member, IEEE*, and Akbar M. Sayeed, *Senior Member, IEEE*

***Abstract*—Recent works on the distribution function of the capacity of independent and identically distributed (i.i.d.) and semicorrelated narrowband channels show that the outage capacity computed using a Gaussian approximation is close to the true outage capacity even for few antennas. Motivated by physical scattering considerations, we study a multi-antenna channel model with independent entries that are not necessarily identically distributed and show the weak convergence of capacity to a Gaussian random variable. The channel model considered in this paper subsumes well-studied cases like the i.i.d. and separable correlation models and thus we generalize previous results on weak convergence of multi-antenna capacity. Using recent results from random matrix theory, we also study the rate of convergence of ergodic capacity of i.i.d. channels to its limit value. Employing a well-known conjecture from random matrix theory, we establish tight results for the rate of convergence and show a dependence of this rate on signal-to-noise ratio.**

***Index Terms*—Canonical statistical model, empirical eigenvalue distribution, ergodic capacity, multiple-input–multiple-output (MIMO) capacity, outage capacity, random matrix theory, scattering, virtual representation.**

I. INTRODUCTION

The seminal works of Telatar [1] and Foschini and Gans [2] unveiled the benefits of using multiple antennas to push high data rates through the bandlimited fading channel. Recent results [3]–[6] generalize the linear scaling of *ergodic* (or average) capacity of multiple-input–multiple-output (MIMO) systems under different correlation assumptions. These results are limited by two constraints. First, these results are restricted to the ergodic capacity and do not provide insights on the distribution of the capacity random variable. Second, these results are true in the asymptotic sense, that is, when the ratio of transmitters to receivers is fixed and antenna dimensions tend to infinity.

Gaussian approximations to the *outage* capacity, which depends on the distribution of the capacity random variable, are shown to be close to the true values in the context of independent and identically distributed (i.i.d.) narrowband channels [7], semicorrelated[1] narrowband channels [8]–[10], and spatially uncorrelated, but temporally correlated frequency selective channels [11]. The capacity random variable of a channel with i.i.d. Rayleigh fading between antenna pairs is shown to converge in distribution to a Gaussian random variable in [12] and [13]. Guess *et al.* [14] compute in closed form the outage capacity of i.i.d. channels employing the BLAST architecture. On the other hand, many papers [1], [11], [12], [15], [16] have also reported fast convergence of the ergodic capacity to its limit value. However very little is known

Manuscript received May 1, 2005; revised March 1, 2006. This work was supported in part by the NSF under Grants CCR-0113385 and CCF-0431088. This paper was presented in part at the IEEE Global Telecommunications Conference 2003 (GLOBECOM 2003), San Fransisco, CA, December 2003.

The authors are with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: vasanthan_raghavan@ieee.org; akbar@engr.wisc.edu).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2006.878216

[1]A semi-correlated channel is a channel with a separable correlation model where either the transmitter or the receiver is uncorrelated.

about the rate of convergence of finite antenna results to the asymptotic capacity theorems reported in [1], [4], etc. and this problem has not been addressed rigorously to our knowledge so far.

The main purpose of this correspondence is twofold: 1) to prove the weak convergence of the capacity random variable for a correlated narrowband channel under very general assumptions on channel modeling and transmitter channel knowledge, and 2) to obtain lower bounds on the rate of convergence of the ergodic capacity for the i.i.d. channel. Thus our result generalizes the weak convergence of capacity random variable, reported for i.i.d. [12], [13] and semi-correlated [8]–[10] narrowband channels, to a very general channel modeling framework. We also show that the ergodic capacity of i.i.d. channels converges to its limit at a rate inversely proportional to the antenna dimensions. Employing a well-known conjecture in the random matrix theory (RMT) literature, we are able to show that even though convergence is assured for all signal-to-noise ratios (SNRs) to RMT-based asymptotics, this rate behaves differently at low and high SNRs.

The workhorse of our analysis is a recently proposed canonical decomposition of multi-antenna channels along the transmit and receive eigenspaces where the decomposition is characterized by independent complex Gaussian entries that are not necessarily identically distributed [17]–[19]. When uniform linear arrays (ULAs) are used at the transmitter and the receiver end, the transmit and receive eigenspaces reduce to the Fourier bases, and the resultant decomposition has been previously reported as the *virtual representation* of multi-antenna channels [20]. Recent measurement campaigns studying the accuracy and fitness of statistical channel models have shown that the canonical statistical model is quite accurate in capturing the scattering features of the underlying physical environment [18], [21]. Moreover, the canonical statistical model incorporates as special cases many well-known channel models like the i.i.d. model, the separable correlation model and the virtual representation.

Our results are based on spectral characterization of large dimensional random matrices with independent entries. We use this asymptotic spectral characterization to show that the capacity random variable centered about its mean and normalized by the square-root of its variance converges in distribution to a standard Gaussian random variable. The canonical statistical decomposition is the bridge connecting the results we prove for random matrices with independent entries and realistic physical channels. We also provide a short summary of the random matrix theory results over which we build our main conclusions. This is done for two reasons: 1) self-containment of the paper and 2) to expose in more detail some elements of the proof technique and machinery that are necessary in further developments in this paper.

This correspondence is organized as follows. Section II introduces the system model and the notations that are used henceforth in this work. Section III discusses the weak convergence of the capacity random variable of correlated channels. We study the rate of convergence of the ergodic capacity of an i.i.d. channel in Section IV. Conclusions are drawn in Section V. We use the following standard notations in this paper: bold-faced fonts for random quantities, x^H and x^T to denote Hermitian and regular transpose of x respectively, $\mathrm{Tr}(\cdot)$ to denote the trace of a matrix, and $\Pr(A)$ and $\chi_A(\cdot)$ to represent the probability and the indicator function of an event A. Weak and almost sure convergence are denoted by $\xrightarrow{w}$ and $\xrightarrow{a.s.}$, respectively. We use the bold-face $\mathbf{E}$ to denote the expectation operator, λ for eigenvalues in general, N_T, N_R and N for antenna dimensions, and I_N to denote the $N \times N$ identity matrix. We remind the readers that: 1) $f(N) = \mathcal{O}(g(N))$ as $N \to \infty$ implies that $\lim_{N\to\infty} \left|\frac{f(N)}{g(N)}\right| \leq K < \infty$, and 2) $f(N) = o(g(N))$ as $N \to \infty$ implies that $\lim_{N\to\infty} \left|\frac{f(N)}{g(N)}\right| = 0$.

II. System Model

A. Channel Model

We consider a single user narrowband communication system and assume that signaling is done using N_T transmit and N_R receive antennas. The N_R-dimensional received signal $\mathbf{y}$ and the N_T-dimensional transmitted signal $\mathbf{x}$ are related by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{n}$ is the noise and $\mathbf{H}$ is the channel matrix coupling the transmit and the receive antennas. The statistics of $\mathbf{H}$ depend on antenna geometry and separation, physical scattering environment, frequency of operation etc. Ideal channel modeling assumes that the entries of $\mathbf{H}$ are i.i.d. Gaussian random variables [1], [2]. Analysis of an i.i.d. channel is tractable, but the i.i.d. channel assumption, however, maybe unrealistic for wireless applications where large antenna spacings or a rich scattering environment are not possible.

Physical channel modeling, on the other hand, adopts a ray-tracing approach and explicitly models multipath signal copies from different directions [22]. The channel matrix is represented as a superposition of rank-one matrices where each rank-one matrix corresponds to a propagation path and is determined by the array steering and response vectors pointing in the direction from the transmitter to the scatterer and the receiver to the scatterer, and the random path gains. Despite being highly accurate in capturing the scattering features, most of these models are not analytically tractable due to the nonlinear dependence of the channel matrix on propagation parameters. They do not offer insights into designing effective signaling strategies or space-time code construction in the general case. The need for channel modeling schemes that capture most of the statistical information of the underlying physical environment, and yet stay within the realms of analytical tractability is therefore paramount.

The separable correlation model which represents the correlation of channel entries in the form of a Kronecker product of the transmit and receive covariance matrices has been one of the first models proposed to meet this goal [4]. The separable correlation model, despite its mathematical tractability in terms of simplification of performance analysis of MIMO systems, suffers from deficiencies acquired by the separability property of channel correlation [23], [24]. Measurement campaigns have shown that the separable correlation model is accurate in capturing the underlying channel statistics under certain conditions [25], but in general, the model results in misleading estimates of the measured channel capacity [23], [17] and other system parameters.

Various statistical models have been proposed to overcome the structural deficiency associated with the separable correlation model. When the transmitter and receiver antennas form a ULA, the virtual representation [20] has been shown to capture the channel characteristics quite accurately. The virtual representation essentially samples the physical scattering environment at fixed, virtual angles and is a unitarily equivalent transformation of $\mathbf{H}$. Moreover, an uniform antenna spacing implies that the Fourier spatial bases functions approximately decorrelate the entries of the matrix decomposed on these basis functions. The readers are referred to [20], [26], [5], [21] for a detailed treatment of the virtual representation and measurement campaigns that support this mathematical model.

The generalization of the virtual representation to the non-ULA case, the canonical statistical model [17], [18], has also been studied extensively[2] of late, both mathematically and via measurement campaigns [21]. This model assumes that the auto and cross-correlation matrices

[2]This model has also been studied in [19] from an ergodic capacity viewpoint.

on both transmitter and receiver sides have the same eigen-basis, and exploits this redundancy to decompose the channel as

$$\mathbf{H} = \mathrm{U}_R \mathbf{H}_{\mathrm{ind}} \mathrm{U}_T^H \tag{2}$$

where U_R and U_T correspond to receive and transmit eigen-matrices respectively, and $\mathbf{H}_{\mathrm{ind}}$ is a random matrix with independent entries that are zero mean, proper complex Gaussian and variance σ_{ij}^2, which are not necessarily equal.

It is not difficult to see that the separable correlation model and the virtual representation are special cases of the canonical statistical model. The canonical statistical model has also been shown to be quite accurate in predicting performance metrics of measured channels. Accuracy in fitting the canonical statistical model to observed measurement data and its impact on predicting measured channel capacity and other performance metrics is reported in [21], [18].

B. Ergodic and Outage Capacity

Through the rest of the sequel we assume that the receiver has perfect channel state information (CSI). We now assume that no CSI is available at the transmitter with extensions to statistical knowledge at the transmitter provided in Section III. It is shown in [2] that the mutual information random variable (in bps/Hz) of a narrowband MIMO channel with transmit SNR ρ, and power-constrained[3] transmit covariance matrix $\mathbf{Q}$ to be

$$\begin{aligned} I(\mathbf{H}) &= \log_2 \det\left[I + \mathbf{H}\mathbf{Q}\mathbf{H}^H\right] \\ &= \log_2 \det\left[I + \mathrm{U}_R\mathbf{H}_{\mathrm{ind}}\mathrm{U}_T^H\mathbf{Q}\mathrm{U}_T\mathbf{H}_{\mathrm{ind}}^H\mathrm{U}_R^H\right] \\ &= \log_2 \det\left[I + \mathbf{H}_{\mathrm{ind}}\tilde{\mathbf{Q}}\mathbf{H}_{\mathrm{ind}}^H\right] \end{aligned} \tag{3}$$

where $\tilde{\mathbf{Q}} = \mathrm{U}_T^H\mathbf{Q}\mathrm{U}_T$. Due to the similarity in structure of the average mutual information expression as a function of $\mathbf{H}$ and $\mathbf{H}_{\mathrm{ind}}$ (which is a consequence of the unitary equivalence of $\mathbf{H}$ and $\mathbf{H}_{\mathrm{ind}}$ in (2)), when no confusion can arise, we will use $\mathbf{H}$ and $\mathbf{Q}$ to represent the underlying $\mathbf{H}_{\mathrm{ind}}$ and $\tilde{\mathbf{Q}}$ respectively. When the transmitter has no CSI, it is shown in [27] that $\mathbf{Q} = \frac{\rho}{N_T}I$ is a robust choice that maximizes the average mutual information. The mutual information random variable with this optimizing $\mathbf{Q}$ is denoted by $C(\mathbf{H})$ and referred to as the capacity random variable.

Without loss of generality, we also assume throughout this correspondence that $t \leq 1$ where $t \triangleq \frac{N_R}{N_T}$ (see [1], [28] etc. for more details on this assumption). The empirical eigenvalue distribution (EED) of $\frac{\mathbf{H}\mathbf{H}^H}{N_T}$ is defined as

$$F_{N_T,N_R}(\lambda) \triangleq \frac{1}{N_R}\sum_{i=1}^{N_R}\chi_{\left[\frac{\lambda_i(\mathbf{H}\mathbf{H}^H)}{N_T},\infty\right)}(\lambda) \tag{4}$$

where $\lambda_i(\mathbf{H}\mathbf{H}^H)$ denote the eigenvalues of $\mathbf{H}\mathbf{H}^H$. Note that $F_{N_T,N_R}(\lambda)$ is a random variable determined by the realizations of $\mathbf{H}$. If the entries of $\mathbf{H}$ are i.i.d. with zero mean and unit variance, then $F_{N_T,N_R}(\lambda)$ converges pointwise in probability as $\{N_T, N_R\} \to \infty$ to the Marčenko–Pastur law, $F(\lambda)$, given by [29]

$$F(\lambda) = \int_{(-\infty,\lambda]} \frac{1}{2\pi t}\sqrt{\left(\frac{\mathrm{t}_{\max} - x}{x}\right)\left(\frac{x - \mathrm{t}_{\min}}{x}\right)} \times \chi_{[\mathrm{t}_{\min},\mathrm{t}_{\max}]}(x)dx \tag{5}$$

where $\mathrm{t}_{\min} \triangleq (1-\sqrt{t})^2$ and $\mathrm{t}_{\max} \triangleq (1+\sqrt{t})^2$.

The ergodic capacity of the channel, C_{N_R}, is the mean of the random variable $C(\mathbf{H})$ where the expectation operation is over the realizations of $\mathbf{H}$. The normalized ergodic capacity, $\tilde{C}_{N_R}$, which is defined as capacity per receive antenna dimension, can then be shown to be [1]

$$\begin{aligned} \tilde{C}_{N_R} \triangleq \frac{C_{N_R}}{N_R} &= \frac{\mathbf{E}\left[\log_2 \det\left(I + \frac{\rho}{N_T}\mathbf{H}\mathbf{H}^H\right)\right]}{N_R} \\ &= \mathbf{E}\left[\frac{1}{N_R}\sum_{i=1}^{N_R}\log_2\left(1 + \rho\frac{\lambda_i\left(\mathbf{H}\mathbf{H}^H\right)}{N_T}\right)\right] \\ &= \mathbf{E}\left[\int_{[0,\infty)}\log_2(1+\rho x)\mathrm{d}F_{N_T,N_R}(x)\right] \\ &\to \mu_{\rho,t} \triangleq \int_{[\mathrm{t}_{\min},\mathrm{t}_{\max}]}\log_2(1+\rho x)\mathrm{d}F(x). \end{aligned}$$

The quantity $\mu_{\rho,t}$ can be computed in closed form [28] and is given by

$$\mu_{\rho,t} = \log_2(1+\rho+\rho t-\rho h) - \log_2(e)\frac{h}{t} + \left(1-\frac{1}{t}\right)\log_2(1-h) \tag{6}$$

where

$$h = \frac{1}{2}\left[1 + t + \frac{1}{\rho} - \sqrt{\left(1+t+\frac{1}{\rho}\right)^2 - 4t}\right]. \tag{7}$$

The insufficiency of the ergodic capacity as the sole measure of the information rate that a MIMO channel can support has been well documented [2], for if the path-gains of the channel were to remain static, there is an irreducible outage probability associated with every possible information transfer rate. To characterize without any ambiguity the information rate a time-varying MIMO channel can support, the need for a quantity like outage capacity is of prime importance. The outage capacity at an outage level of $q\%$ is defined as

$$C_q = \sup_{R\geq 0} R \quad \text{s.t.} \quad \Pr\left(\log_2\det\left[I + \frac{\rho}{N_T}\mathbf{H}\mathbf{H}^H\right] < R\right) \leq \frac{q}{100}. \tag{8}$$

The outage capacity is the maximum rate of communication that is guaranteed for at least $(100-q)\%$ of the channel realizations. The outage capacity of a channel is significantly more difficult to compute than the ergodic capacity and closed form results/approximations for outage capacity are known only for a few special cases [7], [8], [11]–[13]. The difficulty in computing the outage capacity is due to the lack of knowledge of the distribution of capacity random variable of a correlated channel.

In this correspondence, we assume that the random channel matrix $\mathbf{H} = (\mathbf{H}_{ij})$ is an $N_R \times N_T$ matrix with independent, proper complex Gaussian entries that have zero mean and variance σ_{ij}^2, i.e., $\mathbf{E}\left[(\Re(\mathbf{H}_{ij}))^2\right] = \mathbf{E}\left[(\Im(\mathbf{H}_{ij}))^2\right] = \frac{\sigma_{ij}^2}{2}$. We also assume that the variances of $\mathbf{H}$ are uniformly bounded, i.e., $\sup_{N_R,N_T}\sup_{i,j}\sigma_{ij}^2 \leq 1$. Further, we assume that the number of trivial entries, corresponding to a zero variance, in each column is $\mathcal{O}(1)$. In this setting, we show that the capacity random variable converges weakly to a Gaussian random variable as antenna dimensions tend to infinity. We then assume that $\mathbf{H}$ is i.i.d. and study the rate of convergence of ergodic capacity to $N_R\mu_{\rho,t}$ as given by (6). We show that this rate is inversely proportional to the antenna dimensions.

[3]Power-constrained means $\mathrm{Tr}(\mathbf{Q}) \leq \rho$.

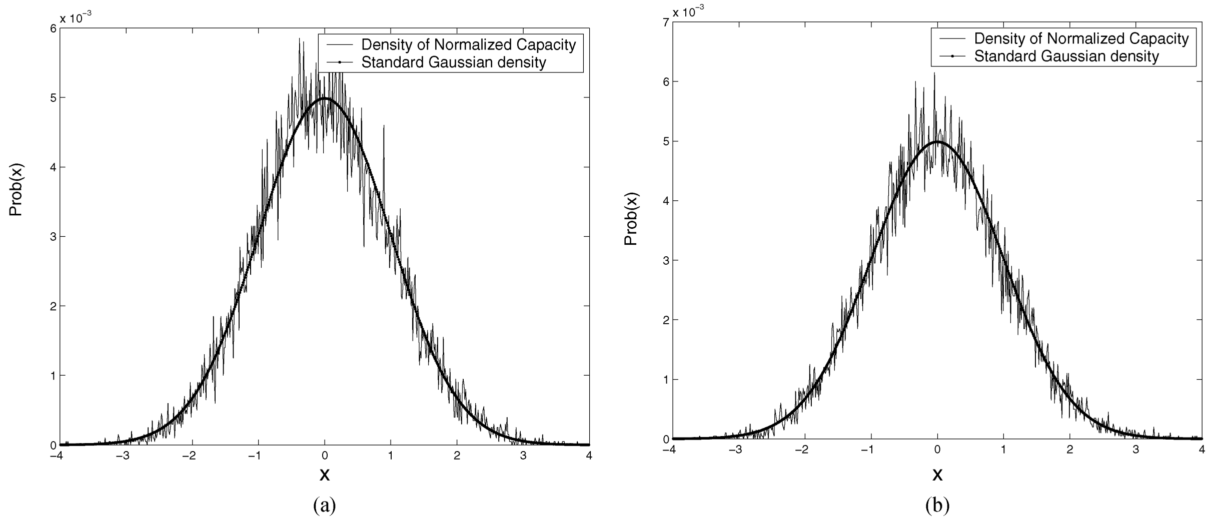

Fig. 1. The probability density of the normalized capacity random variable at an SNR of 10 dB with (a) an i.i.d. channel model for $N_T = N_R = 4$ and (b) canonical statistical model for $N_T = N_R = 6$ are plotted here.

III. Weak Convergence of Capacity Random Variable

In this section we study the weak convergence of the capacity random variable of a correlated narrowband MIMO channel. The MIMO channel matrix $\mathbf{H}$ is characterized by its canonical decomposition and we assume that $\mathbf{H}$ is a complex random matrix as described in the previous section. The main statement of this section stems from the following theorem.

Theorem 1: Let $(\mathbf{H}_{ij})$ be a random $N_R \times N_T$ matrix with independent complex Gaussian entries that are mean zero and variance σ_{ij}^2 and let

$$\lim_{N_R \to \infty} \sup_{ij} \left(\frac{\sigma_{ij}^2}{\sqrt{N_R}} \right) = 0$$
$$\sup_{N_R} \sup_{i} \left(\frac{\sum_{j=1}^{N_T} \sigma_{ij}^2}{N_T} \right) < \infty. \tag{9}$$

Also let E_{N_T,N_R} and V_{N_T,N_R} be defined by $E_{N_T,N_R} \triangleq \mathbf{E}\left[\log_2 \det(I + \frac{\rho}{N_T}\mathbf{H}\mathbf{H}^H)\right]$ and $V_{N_T,N_R} \triangleq \mathrm{Var}\left[\log_2 \det(I + \frac{\rho}{N_T}\mathbf{H}\mathbf{H}^H)\right]$ respectively. Then,

$$\frac{\log_2 \det[I + \frac{\rho}{N_T}\mathbf{H}\mathbf{H}^H] - E_{N_T,N_R}}{\sqrt{V_{N_T,N_R}}} \xrightarrow{w} \mathcal{N}(0,1) \tag{10}$$

where $\mathcal{N}(0,1)$ denotes the real normal random variable with mean zero and variance one.

Proof: See Appendix I. □

Fig. 1(a) and (b) illustrate the weak convergence of the capacity random variable centered about its mean and normalized by the square-root of its variance. In Fig. 1(a), an i.i.d. channel with $N_T = N_R = 4$ is considered. An SNR of 10 dB is assumed and 20000 independent realizations of the channel are used to obtain the probability density of the normalized capacity random variable. Also plotted is the density of the standard Gaussian random variable. On the other hand, Fig. 1(b) plots the probability density of capacity in the case of a channel with independent entries and $N_T = N_R = 6$. At the start of simulation, the variances of the entries of $\mathbf{H}_{\mathrm{ind}}$ are chosen as particular realizations from an ensemble with uniform density on $[0, 1]$. They then remain fixed through the simulation. The two figures described above illustrate Theorem 1.

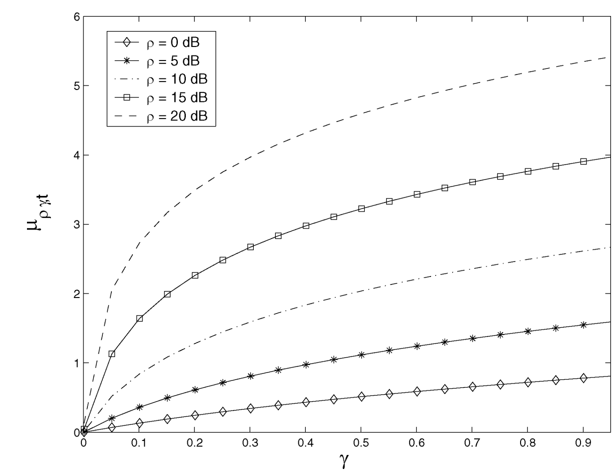

Fig. 2. This plot shows the limit of the normalized ergodic capacity of a D-connected channel, $\mu_{\rho\gamma,t}$, as a function of γ and at different SNRs. Here $t = 1$ is assumed.

1) Discussion 1: We now provide a straightforward extension of Theorem 1 to the case when channel statistics are known at the transmitter. In this setting, a diagonal $\tilde{\mathbf{Q}}$ achieves capacity [17]. Moreover, under the assumptions made in Section II, it is not difficult to see that the capacity-achieving $\tilde{\mathbf{Q}}$ is of the form $\tilde{\mathbf{Q}} = \mathrm{diag}(\tilde{\mathbf{Q}}_i)$, $\tilde{\mathbf{Q}}_i = \mathcal{O}(\frac{1}{N_T})$ for all i. The optimal $\tilde{\mathbf{Q}}$ being similar in form to $\frac{\rho}{N_T}I$, the proof of Theorem 1 can be easily adapted to this setting. While we have assumed a certain richness condition on $\mathbf{H}$ in proving Theorem 1, numerical studies show that the theorem maybe true under very general sparsity assumptions on $\mathbf{H}$.

The theorem that we have proved subsumes the central limit theorem of the capacity random variable of an i.i.d. channel shown in [12], [13]. Besides, Kamath *et al.* show the asymptotic tightness of the outage capacity of an i.i.d. channel as a consequence of the central limit theorem. The key to their result is to compute in closed form the limits of the mean and the variance of the capacity random variable. It is shown in [12] that the outage capacity, C_q, is given by

$$C_q = N_R \mu_{\rho,t} - t x_q \sigma_\rho + o(1) \tag{11}$$

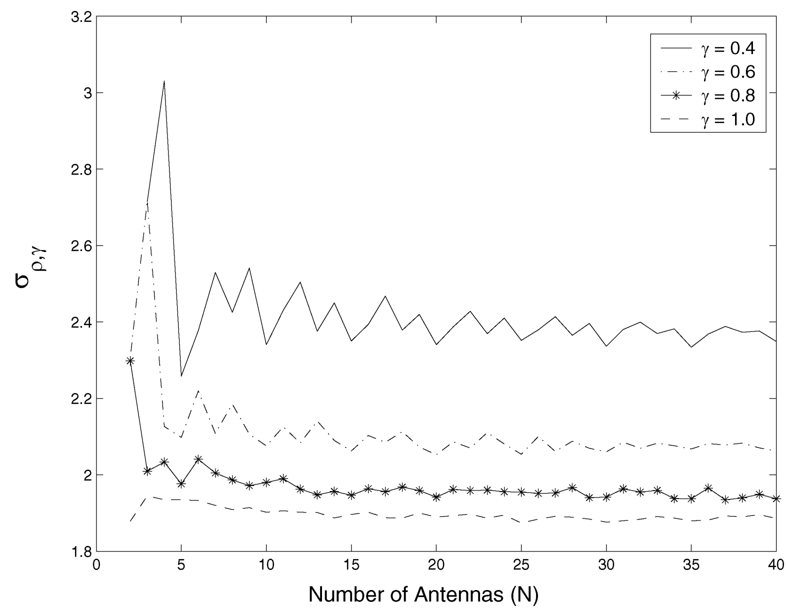

Fig. 3. Here the standard variation of capacity for the D-connected channels, $\sigma_{\rho,\gamma}$, is plotted for different N_T and N_R. An SNR of 20 dB and $N_T = N_R = N$ is assumed.

where $\mu_{\rho,t}$ is the limit of the normalized ergodic capacity as defined in (6), σ_ρ is the variance of the capacity random variable, and x_q is the unique solution of

$$\operatorname{erfc}\left(x_q/\sqrt{2}\right) = 2q \tag{12}$$

where $\operatorname{erfc}(\cdot)$ is the complementary error function.

Such a closed form expression for the mean and the variance of capacity of a correlated channel is, however, not possible without imposing some structure on the variances, $\{\sigma_{ij}^2\}$. In this context, we analyze the D-connected model which was introduced in [5] to study capacity scaling in correlated channels. The D-connected model is a simple channel model that provides fundamental insights into the capacity scaling phenomenon. A D-connected channel is the Hadamard product of an i.i.d. channel matrix and a zero-one mask matrix M. The nonzero entries of the mask matrix occur along the D principal diagonals and the antidiagonal extremities such that the total number of ones in each row and column is equal to D. Essentially, a D-connected matrix M has D nonvanishing diagonals. For example, with $N_T = N_R = 5$ and $D = 3$, M is given by

$$\mathrm{M} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix}. \tag{13}$$

The readers are referred to [5] for a more detailed treatment of the D-connected model and its consequences on capacity scaling.

We consider a D-connected channel with the ratio $\frac{D}{N_T} = \gamma > 0$. It is shown in [5] that the limit of the normalized ergodic capacity of a D-connected channel at a transmitted SNR of ρ is $\mu_{\rho\gamma,t}$, that is, the impact of connectivity on ergodic capacity is equivalent to reducing the transmitted SNR by the connectivity ratio, γ. Fig. 2 shows that $\mu_{\rho\gamma,t}$ is an increasing function of γ. An intuitive explanation for this behavior is that the eigenvalue spread of $\mathbf{HH}^H$ increases with increasing connectivity [5].

Computation of the variance of capacity of a D-connected channel in closed-form, however, is rather difficult. But numerical evidence (see Fig. 3) points out that similar to the mean of the capacity random variable, the variance is a function only of γ and ρ. To make this dependence explicit, we will henceforth use the notation $\sigma_{\rho,\gamma}^2$ to denote the variance of the capacity random variable. As Fig. 3 shows, $\sigma_{\rho,\gamma}$ increases as γ decreases. This trend of $\sigma_{\rho,\gamma}$ corresponds to increasing uncertainty of the capacity random variable as γ decreases. Note that a small value of γ corresponds to a sparse physical environment with fewer scatterers. Equation (11) indicates then that the gap between ergodic capacity and outage capacity at any level q would increase as $\frac{D}{N_T}$ decreases. This conclusion is consistent with the fact that for correlated channels (channels with fewer scatterers), outage capacity is a much more important information-theoretic measure than it is for near-i.i.d. channels.

Based on the preceding discussion, we use a Gaussian approximation for the outage capacity of a 4×4 MIMO channel with independent zero mean Gaussian entries that have unequal variances. The variances are chosen randomly from a uniform distribution on [0, 1] at the start of simulations and remain fixed thereafter. Fig. 4 illustrates the closeness of outage capacity computed using this Gaussian approximation to the actual outage capacity. This plot suggests that the weak convergence which we proved earlier has a very fast rate of convergence. This serves as a motivation prior to the discussion on convergence, albeit in a different setting, in the ensuing section.

IV. Convergence to Asymptotic Results

Telatar showed that the normalized ergodic capacity of i.i.d. channels with $\frac{N_R}{N_T} = t$ converges to $\mu_{\rho,t}$ as $N_T \to \infty$ [1]. Subsequent studies of the ergodic capacity of MIMO channels use results from RMT extensively [4]–[6], [12]. The attribute that is common to all of these works is that they assume that the antenna dimensions tend to infinity with the ratio of transmitters to receivers fixed. This constrains the applicability of RMT results to a very theoretical realm since rarely is the number of antennas greater than 4 or 5 in a practical setting.

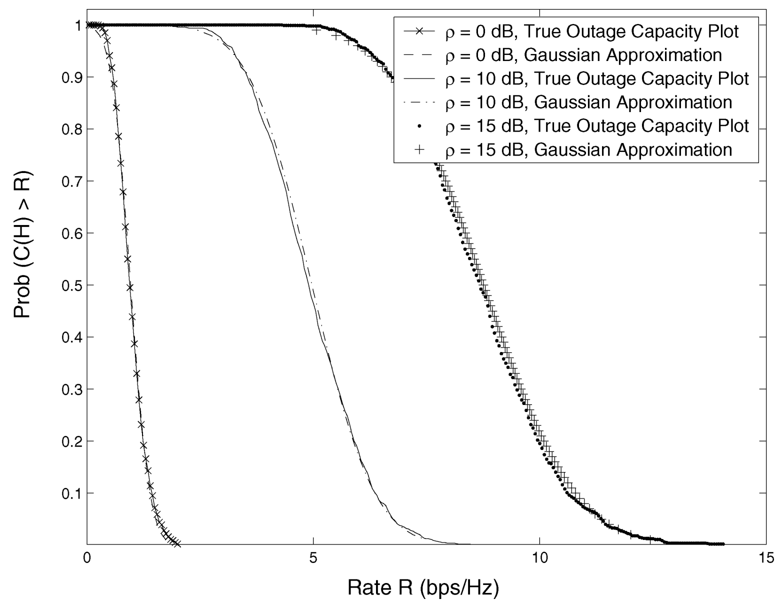

Fig. 4. A Gaussian approximation to the capacity random variable of a 4×4 MIMO channel is used to plot outage capacities at different outage probability levels. Here three SNRs, 0 dB, 10 dB, and 15 dB, are assumed and the Gaussian approximations are seen to be very close to the true outage plots.

We address the following question in this section. Given that we have N_T transmit and N_R receive antennas with i.i.d. Rayleigh fading between antenna pairs, how close is the ergodic capacity to $N_R\mu_{\rho,t}$ for finite N_T and N_R? There have been many results in the literature which suggest that the ergodic capacity is close to $N_R\mu_{\rho,t}$ even for a few antennas [1], [12], [15], [16]. But this problem has not been addressed rigorously to our knowledge so far. We address this question from a theoretical perspective by studying the rate of convergence of RMT asymptotics. Our main conclusion is stated as

Theorem 2: Let $(\mathbf{H}_{ij})$ be a random $N_R \times N_T$ matrix with i.i.d. random variables of mean zero and variance one. Also let $t = \frac{N_R}{N_T}$ be less than 1. The ergodic capacity of the channel at a transmitted SNR ρ is within $x\%$ of $N_R\mu_{\rho,t}$ if

$$N_R > \left[\frac{100}{x} \cdot \frac{c\log_2(1+\rho k)}{\mu_{\rho,t}}\right]^2 \tag{14}$$

where k and c are constants. The constant k is lower bounded[4] by $\mathrm{t}_{\max}$. The constant c is such that

$$\|\mathbf{E}\left[F_{N_T,N_R}(\lambda)\right] - F(\lambda)\| \le \frac{c}{2}\left(N_R^{-\frac{1}{2}}\right) \tag{15}$$

in Lemma 2.

Proof: See Appendix II for Lemmas 1 and 2 and a proof of the theorem. □

Remark: Our assumption of $t < 1$ means that the support of the limiting empirical eigenvalue density function is a subset of the compact interval $[0, 4]$ [29], [40]. But for finite N_T and N_R, the EED function which is a random variable, dependent on the realization of $\mathbf{H}$, could have a support region which is very different from the support of the limit function. Moreover, the fact that the distribution functions converge at a particular rate does not imply that the density functions converge at the same rate. The density functions at finite N_T and N_R are sums of Dirac mass functions and in particular the fact that the empirical eigenvalue density functions even converge to the density function associated with the Marčenko-Pastur law is not an obvious one. Our proof of the above theorem heavily hinges on rewriting the ergodic capacity in terms of the EED which is possible by a recourse to Lemma 3, the Lebesgue-Stieltjes integral formula. A proof without invoking the Lebesgue–Stieltjes integral formula would proceed along the lines of [41], but it looks like such a line of attack may involve additional moment constraints on the entries of $\mathbf{H}$. □

1) Discussion 2: We note that the exponent in (14) comes from Lemma 2. It is to be noted that this estimate is on the very high side as the convergence rate in Lemma 2 is slow. It has been conjectured in the RMT literature that $\|\mathbf{E}\left[F_{N_T,N_R}(\lambda)\right] - F(\lambda)\| \le \frac{1}{2N_R}$. Numerical and theoretical evidence abounds to suggest that this conjecture is true. These include but are not limited to: the rate of convergence of the EED of a related family of random matrices, the Wigner matrix family, has been conjectured[5] to be $\mathcal{O}(N_T^{-1})$ and shown to be $\mathcal{O}(N_T^{-\frac{2}{3}})$ [43], the rate of convergence of the Stieltjes transform of the EED of matrices of the form $\frac{\mathbf{H}\mathbf{H}^H}{N_T}$ has been shown to be $\mathcal{O}(N_T^{-1})$ [44] and indirect weak convergence results (see [34, Comments by J. W. Silverstein, p. 670], [37] and references therein).

We use the conjectured value of convergence rate, $\frac{1}{2N_R}$, to obtain tighter results corresponding to (14), which is $N_R \ge N(\rho) \triangleq \frac{100}{x} \cdot \frac{\log_2(1+\rho k)}{\mu_{\rho,t}}$. We note that $N(\rho)$ tends to a limit at both the high SNR and the low SNR regimes, which is, $\lim_{\rho\to 0} N(\rho) = \frac{100k}{x}$ and $\lim_{\rho\to\infty} N(\rho) = \frac{100}{x}$. Since $k > \mathrm{t}_{\max} > 1$, the low SNR limit is larger than the high SNR limit.[6] Also in the vicinity of the low SNR regime, $N(\rho)$ shows an approximate $\mathcal{O}\left(\frac{1}{\rho}\right)$-type behavior. $N(\rho)$ is approximately between 10 and 15 at an SNR of 10 dB for a 10% error margin.

Fig. 5 plots the variation of $N(\rho)$, the number of antennas needed to be within $x = 1\%$, as a function of SNR. We consider a 1% difference between asymptotics-based estimates and the true ergodic capacities. $N(\rho)$ is normalized by $N_0 = \lim_{\rho\to 0} N(\rho)$. We note that convergence

[4]This is a consequence of Lemma 1 (see Appendix II).

[5]Recent work in [42] has proved this conjectured convergence rate.

[6]As $t \to 1$, $\mathrm{t}_{\max}$ converges to 4 and so does $\frac{\lim_{\rho\to 0} N(\rho)}{\lim_{\rho\to\infty} N(\rho)}$.

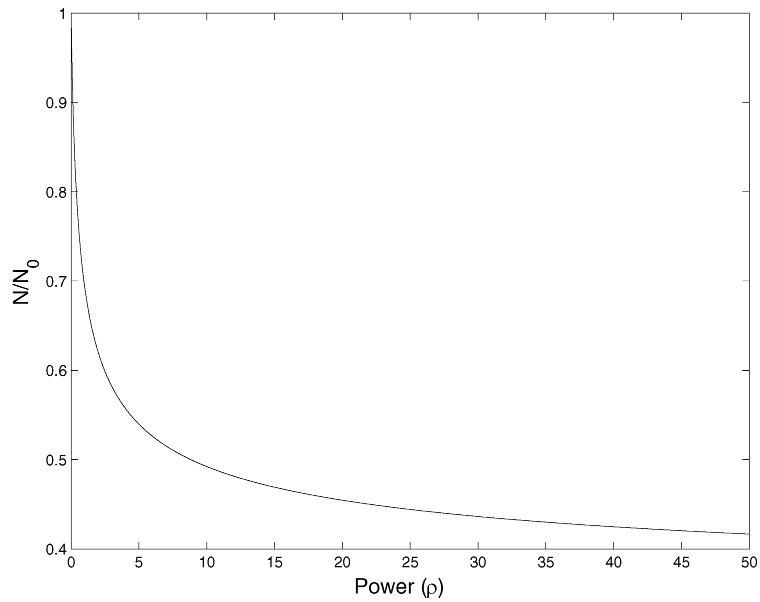

Fig. 5. The number of antennas needed for simulation results to be within $x = 1\%$ of asymptotic based predictions is plotted as a function of SNR. $N(\rho)$ is normalized by $N_0 = \lim_{\rho \to 0} N(\rho)$.

results of RMT asymptotics predict two entirely different kinds of behavior depending on the transmitted SNR even though convergence is assured as $N_R \to \infty$ for all SNR.

V. Conclusion

In this correspondence, we studied the capacity random variable (both with no CSI at the transmitter and with statistical knowledge) of a correlated MIMO channel described by its canonical decomposition. We showed that the capacity random variable centered about its mean and normalized by the square-root of its variance converges in distribution to a standard Gaussian random variable. Thus our results generalize the weak convergence of capacity random variable proved for i.i.d. narrowband channels in [12], [13] and semi-correlated narrowband channels in [8].

We also studied the D-connected model, introduced in [5] to investigate capacity scaling, and provided numerical evidence for the fact that the variance of the capacity random variable depends only on the connectivity ratio of the model, γ. We modeled the outage capacity of a correlated channel using a Gaussian distribution and showed that this Gaussian approximation is very close to the true outage capacity even with as few as four antennas.

We studied the rate of convergence of the ergodic capacity of an i.i.d. channel using RMT results on convergence of spectral statistics and showed that the error margin between RMT predicted ergodic capacity and true ergodic capacity is inversely proportional to the antenna dimensions. Employing a widely-believed conjecture in the RMT literature, we established tight rates of convergence and showed that there is a small difference in this rate in the high and low SNR regimes. Generalization of this result to correlated narrowband channels is not very obvious as this would require a better mathematical understanding of empirical eigenvalue distributions of correlated channels in their implicit Stieltjes transform form.

Appendix I
Proof of Theorem 1

Summary of Girko's Proof: We refer the reader to [30], [31, pp. 174–183] for an analogous theorem with $\mathbf{H} = (\mathbf{H}_{ij})$, where the $\mathbf{H}_{ij}$ are independent (but *not* necessarily Gaussian) *real* random variables with mean zero and variance σ_{ij}^2. The foundation on which this result is built is

a) Key Result [31, Theorem 5.4.3]: Let $\mathbf{H}$ be a random matrix defined as above and B_{N_R} be a positive definite matrix. Define Ξ and the random variables γ_k as follows:

$$\Xi \triangleq I_{N_R} + B_{N_R} + \frac{\rho}{N_T}\mathbf{H}\mathbf{H}^H$$
$$\gamma_k \triangleq \mathbf{E}\left[\log\det(\Xi)/\theta_k^{N_R}\right] - \mathbf{E}\left[\log\det(\Xi)/\theta_{k-1}^{N_R}\right]$$
$$k = 1, \ldots, N_T$$

where $\theta_k^{N_R}$ is the smallest σ-algebra with respect to which $\mathbf{H}_{ij}, i = 1, \ldots, N_R, j = k+1, \ldots, N_T$ are measurable. Then $\{\gamma_k\}$ forms a sequence of martingale differences with $\sum_{k=1}^{N_T} \gamma_k = \log\det(\Xi) - \mathbf{E}[\log\det(\Xi)]$ and $V_{N_T,N_R}^{B_{N_R}} \triangleq \sum_{k=1}^{N_T} \mathbf{E}\gamma_k^2 = \mathrm{Var}[\log\det(\Xi)]$. Further, if

$$\lim_{N_T,N_R} \frac{\sum_{k=1}^{N_T} \mathbf{E}\left|\gamma_k^2 - \mathbf{E}\left(\gamma_k^2/\theta_k^{N_R}\right)\right|}{V_{N_T,N_R}^{B_{N_R}}} = 0 \tag{16}$$

and for all $\tau > 0$

$$\lim_{N_T,N_R} \sum_{k=1}^{N_T} \int_{|x|>\tau} x^2 \mathrm{dPr}\left(\gamma_k < x\sqrt{V_{N_T,N_R}^{B_{N_R}}}\right) = 0 \tag{17}$$

then

$$\frac{1}{\sqrt{V_{N_T,N_R}^{B_{N_R}}}} \sum_k^{N_T} \gamma_k \xrightarrow{w} \mathcal{N}(0,1). \tag{18}$$

□

Under the assumption that $\mathbf{H}_{ij}$ are real random variables with uniformly bounded variance, the Dominated Convergence Theorem [32] yields the following:

$$\lim_{N_T,N_R} \sup_{ij} \int_{-\infty}^{\infty} \frac{\sqrt{N_T}x^2}{1+x^2} \mathrm{dPr}\left(\mathbf{H}_{ij} < x\sqrt{\frac{N_T}{\rho}}\right) = 0 \tag{19}$$

and

$$\lim_{h\downarrow 0} \limsup_{N_R} \sup_{j} \mathbf{E}\left[\log^2\left(1 + h\rho\frac{\sum_{i=1}^{N_R}\mathbf{H}_{ij}^2}{N_T}\right)\right] = 0. \tag{20}$$

Note that the conditions (19) and (20) are forms of the Lindeberg condition, particular to central limit theorem-type arguments. In addition to (19) and (20), assuming that

$$\liminf_{N_T,N_R} \frac{V_{N_T,N_R}^{B_{N_R}}}{N_T} > 0, \tag{21}$$

Girko shows in Theorem 6.2.1 that (16) and (17) hold. Thus the weak convergence follows from "Key Result" and by using $B_{N_R} = \epsilon I_{N_R}$ and letting $\epsilon \to 0$. This limit process also implies that $V_{N_T,N_R}^{B_{N_R}} \to \mathrm{Var}[\log\det(I_{N_R} + \frac{\rho}{N_T}\mathbf{H}\mathbf{H}^H)]$.

The condition (21) is used in [31, equation (6.2.2)] to prove (16) by showing that

$$\lim_{N_T,N_R} \sup_{k=1,\ldots,N_T} \mathbf{E}\left|\mathbf{E}\gamma_k^2 - \mathbf{E}(\gamma_k^2/\theta_k^{N_R})\right| = 0. \tag{22}$$

Given that (21) holds, (16) follows immediately as

$$\lim_{N_T,N_R} \frac{\sum_{k=1}^{N_T}\mathbf{E}\left|\gamma_k^2 - \mathbf{E}\left(\gamma_k^2/\theta_k^{N_R}\right)\right|}{V_{N_T,N_R}^{B_{N_R}}} \le \lim_{N_T,N_R} \frac{N_T}{V_{N_T,N_R}^{B_{N_R}}} \cdot \sup_k \mathbf{E}\left|\mathbf{E}\gamma_k^2 - \mathbf{E}(\gamma_k^2/\theta_k^{N_R})\right| = 0. \tag{23}$$

In the MIMO setting however, (21) does not hold and in fact $V_{N_T,N_R}^{B_{N_R}}$ is finite. We first work on relaxing (21) assumed in [31] exploiting the fact that the $\{\mathbf{H}_{ij}\}$ are Gaussian for all i and j. Under the relaxed assumption, we also need to check that (17) *still* holds.

Proof: The original proof of (17) follows from a Lyapunov condition on the $2+\delta$-th moment of the martingale differences, γ_k. It is not obvious that the original Lyapunov condition holds in our setting. Equation (17) can however shown to be true with the assumption that the number of trivial entries in each column is $\mathcal{O}(1)$. This assumption constrains the variance of γ_k for all k to grow at a sub-linear rate. A simple application of the Dominated Convergence Theorem then implies that (17) follows immediately.

We now proceed to the proof of (16). A careful checking of the proofs of Lemmas (6.2.1), (6.2.2), (6.2.3), and (6.2.4) in [31] (which lead to proving (22)) do not use the condition (21), and hence it is still true that (p. 183)

$$\lim_{N_R\to\infty} \sup_{k=1,\ldots,N_T} \mathbf{E}\left|\mathbf{E}\gamma_k^2 - \mathbf{E}(\gamma_k^2/\theta_k^{N_R})\right| = 0. \tag{24}$$

We now use the assumption that $\mathbf{H}_{ij}$ are Gaussian to show that $\mathbf{E}\left|\mathbf{E}\gamma_k^2 - \mathbf{E}(\gamma_k^2/\theta_k^{N_R})\right|$ goes to 0 *exponentially fast* in N_R for all k. Therefore (16) still holds thus completing the first part of our program.

For completion sake, we reproduce (6.2.12) which bounds $\mathbf{E}\left|\mathbf{E}\gamma_k^2 - \mathbf{E}(\gamma_k^2/\theta_k^{N_R})\right|$ as

$$\mathbf{E}\left|\mathbf{E}\gamma_k^2 - \mathbf{E}(\gamma_k^2/\theta_k^{N_R})\right| \le I_1 + I_2 + I_3 \tag{25}$$

where

$$\begin{aligned}
I_1 &= 2\mathbf{E}\left|\mathbf{E}[\beta_1] - \mathbf{E}\left[\beta_1/\rho_k, \overline{\rho_k}\right]\right| \\
\beta_1 &= \int_1^A\int_1^A dx_1 dx_2 \int_0^\infty\int_0^\infty h_1(y_1,y_2)dy_1dy_2 \\
h_1(y_1,y_2) &= \exp\left(i\frac{\sum_{l=1}^{N_R}(\sqrt{y_1}\eta_l + \sqrt{y_2}\eta_l^{k-1})\sqrt{\rho}\mathbf{H}_{lk}}{\sqrt{N_T}} - y_1x_1 - y_2x_2\right) \\
I_2 &= 2\mathbf{E}\left|\mathbf{E}[\beta_2] - \mathbf{E}\left[\beta_2/\rho_k, \overline{\rho_k}\right]\right| \\
\beta_2 &= \int_1^A\int_1^A dx_1 dx_2 \int_0^\infty\int_0^\infty h_2(y_1,y_2)dy_1dy_2 \\
h_2(y_1,y_2) &= \exp\left(i\frac{\sum_{l=1}^{N_R}(\sqrt{y_1}\eta_l + \sqrt{y_2}\eta_l^{k})\sqrt{\rho}\mathbf{H}_{lk}}{\sqrt{N_T}} - y_1x_1 - y_2x_2\right) \\
I_3 &= 4\mathbf{E}\log^2\left[1 + \rho\frac{\sum_{j=1}^{N_R}\mathbf{H}_{jk}^2}{AN_T}\right]
\end{aligned} \tag{26}$$

where A is arbitrary.

It is a straightforward fact that I_3 can be made to converge exponentially in N_R to 0 for all k by an appropriate choice of A since (20) is true. The first two terms in (25) are similar and we therefore illustrate the exponential convergence of I_1. In the original proof, $I_1 \to 0$ follows from equation (6.2.22) in Lemma (6.2.4), which is

$$\begin{aligned}
&\mathbf{E}\left[\sum_{l=1}^{N_R}\left(\mathbf{E}\{f_l/R_k,\tilde{R}_k\} - \mathbf{E}f_l\right)\right]^2 \le I_4 + I_5 \\
I_4 &= 4\sum_{s\ne k}\mathbf{E}\left[\sum_{l=1}^{N_R}\left\{y_l\left|r_{ll}^k - r_{ll}^{ks}\right| + y_2\left|\tilde{r}_{ll}^k - \tilde{r}_{ll}^{ks}\right|\right\} \times \mathbf{E}\frac{\mu_{kl}^2}{1+\mu_{kl}^2}(\tau^2+1)\right]^2 \\
I_5 &= \left(1+\frac{1}{\tau^2}\right)\int_{|x|>\tau}\sum_{l=1}^{N_R}\frac{x^2}{1+x^2}\mathrm{dPr}\left(\mathbf{H}_{kl} < x\sqrt{\frac{N_T}{\rho}}\right).
\end{aligned} \tag{27}$$

Both the terms I_4 and I_5 go exponentially fast in N_R to 0 precisely because the entries of $\mathbf{H}$ are Gaussian with variances satisfying (9). The term $\mathbf{E}\left[\sum_{l=1}^{N_R}\left(\mathbf{E}\{f_l/R_k,\tilde{R}_k\} - \mathbf{E}f_l\right)\right]^2$ is the main contributor to I_1 and hence we are done.

The necessary conditions for the central limit theorem to hold are equations (6.2.2) and (6.2.3), which are

$$\lim_{N_R\to\infty}\left[\frac{\sum_{k=1}^{N_T}\mathbf{E}\left|\mathbf{E}\gamma_k^2 - \mathbf{E}\left(\gamma_k^2/\theta_k^{N_R}\right)\right|}{V_{N_T,N_R}^{B_{N_R}}}\right] = 0$$

$$\lim_{N_R\to\infty}\sum_{k=1}^{N_R}\int_{|x|>\tau}x^2\mathrm{dPr}\left(\gamma_k\left(V_{N_T,N_R}^{B_{N_R}}\right)^{-1/2} < x\right) = 0.$$

The above conditions are also necessary for the extension to the complex case [33, p. 43], [30]. This is so because a positive definite mapping from the vector space $\mathcal{C}$ to $\Re$ can be equivalently viewed as a map from $\Re^2$ to $\Re$. It is easy to check that the above conditions hold with real Gaussians replaced by complex Gaussians. The final step in the theorem is to use $B_{N_R} = \epsilon I_{N_R}$, and let $\epsilon \to 0$. □

APPENDIX II
PROOF OF THEOREM 2

Toward proving Theorem 2, we need a few technical results. We begin our convergence analysis with the following lemma on the spectral statistics of $\mathbf{H}\mathbf{H}^H$ when $\mathbf{H}$ is an i.i.d. matrix.

Lemma 1 (Yin, Bai, and Krishnaiah, 1988): Let $\mathbf{H}$ be an $N_R \times N_T$ random matrix with i.i.d. complex entries of mean 0 and variance 1 and $t = \frac{N_R}{N_T} \leq 1$. Also let the fourth moments of $\mathbf{H}_{ij}$ be uniformly bounded. Then

$$\limsup_{N_R, N_T} \frac{\lambda_{\max}(\mathbf{H}\mathbf{H}^H)}{N_T} = \mathrm{t}_{\max} = (1+\sqrt{t})^2 \text{ a.s.} \tag{28}$$

and for any $K > \mathrm{t}_{\max}$

$$\Pr\left(\frac{\lambda_{\max}(\mathbf{H}\mathbf{H}^H)}{N_T} \geq K\right) = o\left(\frac{1}{N_T^l}\right) \tag{29}$$

for all $l \geq 1$.

Proof: See Appendix III. □

The following lemma concerns the rate of convergence of expected spectral statistics of i.i.d. sample covariance matrices.

Lemma 2 (Bai, Miao, and Tsay, and Götze and Tikhomirov): Let $\mathbf{H}$ be a $N_R \times N_T$ random matrix with i.i.d. complex entries of mean 0 and variance 1. Then

$$\begin{aligned} \|\mathbf{E}\left[F_{N_T,N_R}(\lambda)\right] - F(\lambda)\| &= \mathcal{O}(N_R^{-\frac{1}{2}}) \\ &\quad \text{if } 0 < t < 1, 1 < t < \infty \\ \|\mathbf{E}\left[F_{N_T,N_R}(\lambda)\right] - F(\lambda)\| &= \mathcal{O}(N_R^{-\frac{5}{48}}) \\ &\quad \text{if } 1-\epsilon < t < 1+\epsilon, 0 < \epsilon < 1. \end{aligned}$$

Here $\|.\|$ refers to the supremum norm and $F(\lambda)$ is the Marčenko–Pastur law.

Proof: The readers are referred to [34], [36]–[38] for a proof of the above lemma. We note that the condition needed in [38] supersedes that of [36]. □

We also need the following lemma for the proof of Theorem 2.

Lemma 3 (Lebesgue, Stieltjes): If F and G are normalized functions of bounded variation mapping $\Re$ to $\Re$, and at least one of them is continuous, then for $-\infty < a < b < \infty$,

$$\int_{(a,b]} F dG + \int_{(a,b]} G dF = F(b)G(b) - F(a)G(a) \tag{30}$$

Proof: The readers are referred to [32, p. 107] for a proof of this lemma. □

We are now prepared to prove Theorem 2.

Proof: Our proof follows from integration by parts using the Lebesgue-Stieltjes integral formula. The relative difference in the capacities that we seek can be written as in (31)–(33) at the bottom of the page. We now bound (33) by quantities computed over a compact interval $[0,k]$ for some $k > \mathrm{t}_{\max}$ and an error quantity. The interval $[0,k]$ contains the support of the limiting empirical eigenvalue density and hence we have (34) in the bottom of the page.

We claim that if $k > \mathrm{t}_{\max}$, the second term in the numerator is $o\left(\frac{1}{N_T^l}\right)$ for all $l \geq 1$ and thus vanishes to zero at a rate much faster than the first term. The proof of this claim is the content of Proposition 1. We thus have (35) as in the bottom of the page. It is straightforward to see that the differences in distributions, $F_{N_T,N_R}(\lambda) - F(\lambda)$, and $h(\lambda)$ defined as

$$h(\lambda) = \begin{cases} 0 & \text{for } \lambda \leq 0 \\ \log_2(1+\rho\lambda) & \text{for } 0 < \lambda \leq k \\ \log_2(1+\rho k) & \text{for } \lambda > k \end{cases} \tag{36}$$

are normalized functions of bounded variation [32]. We can then use Lemma 3 to recast (35) using only the difference in distribution functions as follows:

$$\begin{aligned} \left|\tilde{C}_{N_R} - \mu_{\rho,t}\right| \leq& \left|\mathbf{E}\left[F_{N_T,N_R}(k) - F(k)\right]\log_2(1+\rho k)\right| \\ &+ \log_2 e \left|\mathbf{E}\left[\int_0^k \frac{\rho}{1+\rho\lambda}\left(F(\lambda) - F_{N_T,N_R}(\lambda)\right)\mathrm{d}\lambda\right]\right|. \end{aligned} \tag{37}$$

Both the terms in (37) can then be bounded by Lemma 2. The second term is to be bounded after using Fubini's theorem [32]. We thus have

$$\left|\frac{C_{N_R} - N_R\mu_{\rho,t}}{N_R\mu_{\rho,t}}\right| \leq \frac{2\log_2(1+\rho k)\|\mathbf{E}\left[F_{N_T,N_R}(\lambda)\right] - F(\lambda)\|}{\mu_{\rho,t}}$$

and if N_R is such that (14) holds, then the conclusion follows immediately. □

$$\frac{C_{N_R} - N_R\mu_{\rho,t}}{N_R\mu_{\rho,t}} = \frac{\tilde{C}_{N_R} - \mu_{\rho,t}}{\mu_{\rho,t}} \tag{31}$$

$$= \frac{\mathbf{E}\left[\int_0^\infty \log_2(1+\rho\lambda)\mathrm{d}F_{N_T,N_R}(\lambda)\right] - \int_{t_{\min}}^{t_{\max}} \log_2(1+\rho\lambda)\mathrm{d}F(\lambda)}{\mu_{\rho,t}} \tag{32}$$

$$= \frac{\mathbf{E}\left[\int_0^\infty \log_2(1+\rho\lambda)\mathrm{d}\left(F_{N_T,N_R}(\lambda) - F(\lambda)\right)\right]}{\mu_{\rho,t}} \tag{33}$$

$$= \frac{\mathbf{E}\left[\int_0^k \log_2(1+\rho\lambda)\mathrm{d}\left(F_{N_T,N_R}(\lambda) - F(\lambda)\right)\right] + \mathbf{E}\left[\int_k^\infty \log_2(1+\rho\lambda)\mathrm{d}F_{N_T,N_R}(\lambda)\right]}{\mu_{\rho,t}} \tag{34}$$

$$\left|\frac{C_{N_R} - N_R\mu_{\rho,t}}{N_R\mu_{\rho,t}}\right| \leq \left|\frac{\mathbf{E}\left[\int_0^k \log_2(1+\rho\lambda)\mathrm{d}\left(F_{N_T,N_R}(\lambda) - F(\lambda)\right)\right]}{\mu_{\rho,t}}\right| + o\left(\frac{1}{N_T^l}\right). \tag{35}$$

Proposition 1: Under the assumptions of Theorem 2, the second term in the numerator of (34), T_1, defined as

$$T_1 \triangleq \mathbf{E}\left[\int_k^\infty \log_2(1+\rho\lambda)\mathrm{d}F_{N_T,N_R}(\lambda)\right] = o\left(\frac{1}{N_T^l}\right) \quad (38)$$

for all $l \geq 1$.

Proof: We can write T_1 as

$$\begin{aligned} T_1 &= \mathbf{E}\left[\sum_i \log_2(1+\rho\lambda_i)\chi_{[\lambda_i \geq k]}\right] \\ &\leq N_R \mathbf{E}\left[\log_2(1+\rho\lambda_{\max})\chi_{[\lambda_{\max}\geq k]}\right] \\ &\overset{(a)}{\leq} \log_2(e) N_R \rho \mathbf{E}\left[\lambda_{\max}\chi_{[\lambda_{\max}\geq k]}\right] \\ &\overset{(b)}{\leq} \log_2(e) N_R \rho \left(\mathbf{E}\left[\lambda_{\max}^2\right]\right)^{\frac{1}{2}} \Pr\left(\lambda_{\max} \geq k\right)^{\frac{1}{2}} \end{aligned} \quad (39)$$

where λ_i and $\lambda_{\max}$ respectively refer to the i-th and the largest eigenvalue of $\frac{\mathbf{H}\mathbf{H}^H}{N_T}$, (a) follows from the log-inequality, and (b) follows from applying Cauchy–Schwarz inequality on $\lambda_{\max}$ and $\chi_{[\lambda_{\max}\geq k]}$. The conclusion follows from Lemma 1 and the fact that $\mathbf{E}\left[\lambda_{\max}^2\right] = \mathcal{O}(1)$. To show this we use the fact [39, p. 50] that for a positive random variable X, the following holds:

$$\mathbf{E}[X] \leq \sum_{j=0}^{\infty} \Pr(X \geq j). \quad (40)$$

The above fact along with Lemma 1 yields $\mathbf{E}\left[\frac{\lambda_{\max}(\mathbf{H}\mathbf{H}^H)^2}{N_T^2}\right] \leq C < \infty$. This follows from

$$\begin{aligned} &\mathbf{E}\left[\frac{\lambda_{\max}(\mathbf{H}\mathbf{H}^H)^2}{N_T^2}\right] \\ &\leq \sum_{k=0}^{35} \Pr\left(\frac{\lambda_{\max}(\mathbf{H}\mathbf{H}^H)^2}{N_T^2} > k\right) \\ &\quad + \sum_{k=36}^{\infty} \Pr\left(\frac{\lambda_{\max}(\mathbf{H}\mathbf{H}^H)^2}{N_T^2} > k\right) \\ &\overset{(c)}{\leq} 36 + \sum_{k=6}^{\infty} 2N_T N_R \left(1+\delta_{N_T}^{1/2}\right)^{2m} \\ &\quad \times \left(1 + C_1 \delta_{N_T} A_{N_T}^6\right)^m \left(\frac{4}{k}\right)^m (2k+1) \\ &\leq 36 + 4N_T N_R \left(1+\delta_{N_T}^{1/2}\right)^{2m} \left(1 + C_1 \delta_{N_T} A_{N_T}^6\right)^m \\ &\quad \times 4^m \int_5^\infty \frac{\mathrm{d}x}{x^{m-1}} \\ &= 36 + 4N_T N_R \left(1+\delta_{N_T}^{1/2}\right)^{2m} \left(1 + C_1 \delta_{N_T} A_{N_T}^6\right)^m \\ &\quad \times \frac{4^m}{5^{m-2}(m-2)} \end{aligned} \quad (41)$$

where (c) follows from (43) of Lemma 1. It is easy to see that for the choice of m as in Lemma 1, the second term in the above bound tends to 0 as $N_T \to \infty$ and that we have proved $\mathbf{E}\left[\frac{\lambda_{\max}(\mathbf{H}\mathbf{H}^H)^2}{N_T^2}\right] \leq C < \infty$. □

APPENDIX III
PROOF OF LEMMA 1

Summary of Proof: The readers are referred to [34, Th. 2.16, p. 635] for a proof of the first part of the lemma. The second part of the lemma follows from [35, Th. 3.1]. The basic idea behind the proof of [35] is centralization and truncation of random variables. Yin *et al.* show that if $\mathbf{H}$ is a random matrix with i.i.d. entries (zero mean, unit variance and bounded fourth moment), and $\lambda_{\max}$ and $\tilde{\lambda}_{\max}$ are the largest eigenvalues of $\frac{\mathbf{H}\mathbf{H}^H}{N_T}$ and $\frac{\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H}{N_T}$ respectively, where $\tilde{\mathbf{H}}_{ij} = \mathbf{H}_{ij}\chi_{[\mathbf{H}_{ij}\leq \delta_{N_T} N_T^{1/2}]} - \mathbf{E}\left[\mathbf{H}_{ij}\chi_{[\mathbf{H}_{ij}\leq \delta_{N_T} N_T^{1/2}]}\right]$, then $\lambda_{\max} \xrightarrow{a.s.} \tilde{\lambda}_{\max}$. In the above proof, δ_{N_T} can be chosen such that $\delta_{N_T} \to 0$ as $N_T \to \infty$ at *any* predetermined rate. The aim then is to compute $\Pr\left(\frac{\lambda_{\max}(\mathbf{H}\mathbf{H}^H)}{N_T} \geq K\right)$ for a $\mathbf{H}$ with i.i.d. entries that are zero mean and have higher moments bounded by powers of $\delta_{N_T} N_T^{1/2}$. The computation of $\mathbf{E}\left[\mathrm{Tr}\left((\mathbf{H}\mathbf{H}^H)^m\right)\right]$ is achieved by translating the problem to a graph theoretic setting and counting the number of edges that satisfy certain desired properties.

Proof: First note that for any $K > \mathrm{t}_{\max}$ and for all integers m

$$\begin{aligned} \Pr\left(\frac{\lambda_{\max}(\mathbf{H}\mathbf{H}^H)}{N_T} \geq K\right) &= \Pr\left(\frac{\lambda_{\max}^m(\mathbf{H}\mathbf{H}^H)}{N_T^m} \geq K^m\right) \\ &\leq \Pr\left(\frac{\mathrm{Tr}\left((\mathbf{H}\mathbf{H}^H)^m\right)}{N_T^m} \geq K^m\right) \\ &\overset{(a)}{\leq} \frac{\mathbf{E}\left[\mathrm{Tr}\left((\mathbf{H}\mathbf{H}^H)^m\right)\right]}{(KN_T)^m} \end{aligned} \quad (42)$$

where (a) follows from Markov's inequality. The proof of "Truncation Lemma" [35] shows that there exists a sequence of positive numbers δ_{N_T} and A_{N_T} with $\delta_{N_T} \to 0$, $A_{N_T} \to \infty$, $m = A_{N_T}\log(N_T)$, $\delta_{N_T} {N_T}^{1/3} \to 0$, and $A_{N_T}\left(\delta_{N_T}\right)^{1/6} \to 0$ as $N_T \to \infty$ such that $\frac{\mathbf{E}\left[\mathrm{Tr}\left((\mathbf{H}\mathbf{H}^H)^m\right)\right]}{(KN_T)^m}$ is bounded according to (43)–(44) as follows:

$$\begin{aligned} &\frac{\mathbf{E}\left[\mathrm{Tr}\left((\mathbf{H}\mathbf{H}^H)^m\right)\right]}{(KN_T)^m} \\ &\leq 2N_T N_R \left(1+\delta_{N_T}^{1/2}\right)^{2m} \left(1 + C_1\delta_{N_T} A_{N_T}^6\right)^m \left[\frac{\mathrm{t}_{\max}}{K}\right]^m \quad (43) \\ &\xrightarrow{(b)} 2N_T N_R \exp\left(2m\delta_{N_T}^{1/2} + C_1 m \delta_{N_T} A_{N_T}^6\right)\left[\frac{\mathrm{t}_{\max}}{K}\right]^m. \quad (44) \end{aligned}$$

Here, C_1 is a positive constant independent of N_T and we have used the fact that $\lim_{x\to 0}(1+x)^{\frac{k}{x}} = \exp(k)$ in (b). The choice of m and δ_{N_T} from the Truncation Lemma also forces $m\delta_{N_T}^{1/2} \to 0$, $m\delta_{N_T} A_{N_T}^6 \to 0$ and $N_T^{l+2}\left[\frac{\mathrm{t}_{\max}}{K}\right]^m \to 0$ for all $l \geq 1$. This completes the proof of Lemma 1. □

ACKNOWLEDGMENT

The authors acknowledge the comments from the anonymous reviewers that helped in improving this exposition, the references provided by Prof. Vyacheslav L. Girko, and the correspondences with Prof. Jack W. Silverstein during the course of this work.

REFERENCES

[1] Í. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 2172–2178, 2000.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[3] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multi-element antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502–513, Mar. 2000.

[4] C.-N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valanzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, pp. 637–650, Mar. 2002.

[5] K. Liu, V. Raghavan, and A. M. Sayeed, "Capacity scaling and spectral efficiency in wideband correlated MIMO channels," *IEEE Trans. Inf. Theory,* , ser. Special Issue on MIMO Systems, vol. 49, no. 10, pp. 2504–2526, Oct. 2003.

[6] V. Raghavan and A. M. Sayeed, "Role of channel power in the sub-linear capacity scaling of MIMO channels," in *Proc. Allerton Conf. Commun., Contr. Comput.*, Allerton, IL, Sep. 2004.

[7] Z. Wang and G. B. Giannakis, "Outage mutual information rate of space-time MIMO channels," *IEEE Trans. Inf. Theory*, vol. 50, pp. 657–662, Apr. 2004.

[8] P. Smith, S. Roy, and M. Shafi, "Capacity of MIMO systems with semi-correlated flat fading," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2781–2788, Oct. 2003.

[9] M. Kang and M. S. Alouini, "Impact of correlation on the capacity of MIMO channels," in *Proc. IEEE Int. Conf. Commun.*, Anchorage, AK, May 2003, pp. 2623–2627.

[10] A. M. Tulino and S. Verdu, "Asymptotic outage capacity of multi-antenna channels," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Philadelphia, PA, Mar. 2005.

[11] A. Scaglione, "Statistical analysis of capacity of MIMO frequency selective Rayleigh fading channels with arbitrary number of inputs and outputs," in *Proc. IEEE ISIT 2002*, Lausanne, Switzerland, 2002.

[12] M. A. Kamath and B. L. Hughes, "The asymptotic capacity of multiple antenna Rayleigh fading channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4325–4333, Dec. 2005.

[13] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multi-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inf. Theory*, vol. 50, pp. 1893–1909, Sep. 2004.

[14] T. Guess, H. Zhang, and T. V. Kotchiev, "The outage capacity of BLAST for MIMO channels," in *Proc. Int. Conf. Commun.*, May 2003.

[15] E. Biglieri, G. Taricco, and A. Tulino, "How far away is infinity? Using asymptotic analyzes in multi-antenna systems," in *Proc. IEEE 7th Int. Symp. Spread Spectrum Techn. Applicat.*, Prague, Czech Rep., 2002.

[16] A. M. Tulino and S. Verdu, "Random matrices and wireless communications," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 1, Jun. 2004.

[17] J. H. Kotecha, V. Raghavan, and A. M. Sayeed, "Canonical statistical modeling and capacity analysis of correlated MIMO fading channels," *IEEE Trans. Commun.*, submitted for publication.

[18] W. Weichselberger, H. Ozcelik, M. Herdin, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 90–100, Jan. 2006.

[19] A. M. Tulino, A. Lozano, and S. Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 2491–2509, Jul. 2005.

[20] A. M. Sayeed, "Deconstructing multi-antenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, pp. 2563–2579, Oct. 2002.

[21] Y. Zhou, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental study of MIMO channel statistics and capacity via the virtual channel representation," *IEEE. Trans. Wireless Commun.*, Mar. 2006, submitted for publication.

[22] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Process. Mag.*, pp. 49–83, Nov. 1997.

[23] H. Ozcelik, M. Herdin, W. Weichselberger, J. Wallace, and E. Bonek, "Deficiencies of 'Kronecker' MIMO radio channel model," *Electron. Lett.*, vol. 39, no. 16, pp. 1209–1210, Aug. 2003.

[24] J. Wallace, H. Ozcelik, M. Herdin, E. Bonek, and M. Jensen, "Power and complex envelope correlation for modeling measured indoor MIMO channels: A beamforming evaluation," in *Proc. IEEE 58th Fall Veh. Technol. Conf., 2003*, Oct. 2003, vol. 1, pp. 363–367.

[25] D. P. McNamara, M. A. Beach, and P. N. Fletcher, "Spatial correlation in indoor MIMO channels," in *Proc. 13th IEEE Int. Symp. Pers. Indoor, Mobile Radio Commun., 2002*, 2002, vol. 1, pp. 290–294.

[26] A. M. Sayeed and V. V. Veeravalli, "Essential degrees of freedom in space-time fading channels," in *Proc. 13th IEEE Int. Symp. Pers. Indoor, Mobile Radio Commun.*, Sep. 2002, vol. 4, pp. 1512–1516.

[27] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: A game-theoretic approach," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1707–1727, Jul. 2003.

[28] P. B. Rapajic and D. Popescu, "Information capacity of a random signature multiple-input multiple-output channel," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1245–1248, Aug. 2000.

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

[30] V. L. Girko, "A central limit theorem for random determinants," *Theory of Prob. and Math. Stat.*, vol. 21, pp. 35–39, 1979.

[31] ——, *Theory of Random Determinants*. Reading, MA: Kluwer, 1990.

[32] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. New York: Wiley, 1999.

[33] M. Loeve, *Probability Theory*, 4th ed. Berlin, Germany: Springer-Verlag, 1977.

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

[35] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, "On limit of the largest eigenvalue of the large dimensional sample covariance matrix," *Prob. Th. Rel. Fields*, vol. 78, pp. 509–521, 1988.

[36] Z. D. Bai, B. Miao, and J. Tsay, "A note on the convergence rate of the spectral distributions of large random matrices," *Stat. Prob. Lett.*, vol. 34, pp. 95–101, 1997.

[37] Z. D. Bai, "Convergence rate of expected spectral distributions of large random matrices. Part I. Wigner matrices, Part II. Sample covariance matrices," *Ann. Prob.*, vol. 21, no. 2, pp. 625–672, Apr. 1993.

[38] F. Götze and A. N. Tikhomirov, "Rate of convergence to the Marčenko-Pastur law," *Bernoulli*, vol. 10, no. 3, pp. 503–548, 2004.

[39] R. A. Durrett, *Probability: Theory and Examples*, 2nd ed. Belmont, CA: Duxbury, 1995.

[40] J. W. Silverstein and S.-I. Choi, "Analysis of the limiting spectral distribution of large dimensional random matrices," *J. Mult. Var. Anal.*, vol. 54, pp. 295–309, 1995.

[41] S. Chatterjee and A. Bose, "A new method for bounding rates of convergence of empirical spectral distributions," *J. Theor. Prob.*, vol. 17, no. 4, pp. 1003–1019, 2004.

[42] B. Delyon and J.-F. Yao, "On the spectral distribution of Gaussian random matrices," Transl.:English Series *Acta Mathematicae Applicatae Sinica*, 2006, to be published.

[43] F. Götze and A. N. Tikhomirov, "Rate of convergence to the semicircle law for the Gaussian unitary ensemble," *Theory Prob. Appl.*, vol. 47, no. 2, pp. 323–330, 2003.

[44] J. W. Silverstein and Z. D. Bai, "CLT of linear spectral statistics of large dimensional sample covariance matrices," *Ann. Prob.*, vol. 32, no. 1A, pp. 553–605, 2004.