

# Capacity Scaling and Spectral Efficiency in Wide-Band Correlated MIMO Channels

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

**Abstract**—The dramatic linear increase in ergodic capacity with the number of antennas promised by multiple-input multiple-output (MIMO) wireless communication systems is based on idealized channel models representing a rich scattering environment. Is such scaling sustainable in realistic scattering scenarios? Existing physical models, although realistic, are intractable for addressing this problem analytically due to their complicated nonlinear dependence on propagation path parameters, such as the angles of arrival and delays. In this paper, we leverage a recently introduced *virtual representation* of physical models that is essentially a Fourier series representation of wide-band MIMO channels in terms of fixed virtual angles and delays. Motivated by physical considerations, we propose a  $D$ -connected model for correlated channels defined by a virtual spatial channel matrix consisting of  $D$  nonvanishing diagonals with independent and identically distributed (i.i.d.) Gaussian entries. The parameter  $D$  provides a meaningful and tractable measure of the richness of scattering. We derive general bounds for the coherent ergodic capacity and investigate capacity scaling with the number of antennas and bandwidth. In the large antenna regime, we show that linear capacity scaling is possible if  $D$  scales linearly with the number of antennas. This, in turn, is possible if the number of resolvable paths grows quadratically with the number of antennas. The capacity saturates for linear growth in the number of paths (fixed  $D$ ). The ergodic capacity does not depend on frequency selectivity of the channel in the wide-band case. Increasing bandwidth tightens the bounds and hastens the convergence of scaling behavior. For large bandwidth, the capacity scales linearly with the signal-to-noise ratio (SNR) as well. We also provide an explicit characterization of the wide-band slope recently proposed by Verdú. Numerical results are presented to illustrate the key theoretical results.

**Index Terms**—Beamforming, empirical eigenvalue distribution, ergodic capacity, Fourier series, frequency selectivity, ray tracing, scattering, spectral efficiency.

## I. INTRODUCTION

THE use of multiple-element antenna arrays has emerged as a promising technology for dramatically increasing the spectral efficiency of wireless communication systems. Initial studies have indicated linear increase in the capacity of

narrow-band multiple-input multiple-output (MIMO) systems with the number of antennas (see [1], [2]). However, these studies are based on an idealized channel model, representing a rich scattering environment, that assumes independent and identically distributed (i.i.d.) Gaussian entries for the channel matrix. Such idealized channels seldom, if ever, occur in practice, particularly with practically feasible antenna spacings. Several experimental and analytical studies have shown that the capacity of realistic MIMO channels can be significantly less than that of i.i.d. models (see [3]–[6]).

Idealized statistical models, such as those used in [1], [2], represent one extreme in existing modeling approaches. On the other extreme are detailed physical (ray-tracing) models that describe the channel via signal propagation over multiple paths, each path associated with an angle of departure (AoD), an angle of arrival (AoA), a delay and a path gain (see, e.g., [7]–[9], [4], [10]). While quite accurate, such models depend on physical AoAs, AoDs, and delays in a nonlinear fashion making it rather difficult to incorporate them in system design or analytical calculations. Indeed, most capacity studies based on physical models have relied on numerical simulation for capacity assessment (see, e.g., [4], [7], [8]).

The goal of this paper is to investigate capacity scaling in realistic correlated MIMO channels. We consider both narrow-band and wide-band channels and explore scaling behavior of ergodic capacity as a function of both the number of antennas and bandwidth in a Rayleigh-fading environment. There are three main objectives of our work: 1) to provide a characterization of physical MIMO channel models that is analytically tractable, 2) to obtain rigorous mathematical results that characterize capacity scaling behavior, and 3) to relate the scaling results to the physical characteristics of realistic MIMO channels. Our focus is on systems that use uniform linear arrays (ULAs) of  $N$  antennas at both the transmitter and receiver. We assume that the channel is unknown at the transmitter but perfectly known at the receiver.

The workhorse of our analysis is a recently introduced *virtual representation* of narrow-band [6] and wide-band MIMO channels [11], [12] that provides an intuitive and tractable characterization of realistic physical models and yields useful insights into the effects of scattering characteristics on channel capacity. The virtual representation is based on the simple but fundamental observation that detailed channel modeling without regard to signal space characteristics is unnecessary from a communication-theoretic viewpoint—an *effective channel representation* that captures the interaction between the physical channel and the finite-dimensional space–time signal space is all that is needed. In wide-band MIMO channels, the signal space is

Manuscript received November 1, 2002; revised June 22, 2003. This work was supported in part by the National Science Foundation under CAREER Grant CCR-9875805 and ITR Grant CCR-0113385, and by the ONR under the YIP Grant N00014-01-1-0825. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Yokohama, Japan, June/July 2003 and will be presented in part at the IEEE Global Communications Conference, San Francisco, CA, December 2003.

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

Communicated by B. Hassibi, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2003.817446

characterized by the number of antennas  $N$  (for a given antenna spacing) and bandwidth  $W$ . The virtual representation is a Fourier series for the channel frequency response matrix that corresponds to sampling the angle-delay space at *fixed virtual* AoAs, AoDs, and delays. In particular, it induces a virtual partitioning of propagation paths in angle-delay space that exposes their contribution to channel capacity, and plays a key role in relating the scaling results to physical scattering characteristics.

Our capacity scaling analysis is based on a *D-connected model* for the narrow-band virtual matrix that consists of  $D$  nonvanishing diagonals with i.i.d. Gaussian entries. The *D-connected model* is motivated by physical considerations—it represents a scattering environment in which each virtual transmit angle couples with  $D$  virtual receive angles and *vice versa*. We call  $D$  the *channel connectivity* as it provides a meaningful and tractable measure of the richness of scattering. For example,  $D = 1$  corresponds to a loosely connected (highly correlated) channel, whereas  $D = N$  represents a rich scattering environment. In effect, the *D-connected model* provides a mathematical construct that greatly facilitates capacity analysis of correlated channels, analogous to the role of i.i.d. channel matrix in the idealized model.

To our knowledge, the most recent work addressing the issue of capacity scaling in correlated channels is [13]. The channel model used in [13] is a generalization of the i.i.d. model and still predicts linear capacity growth with the number of antennas, albeit with a smaller slope compared to i.i.d. channels. In Section VII, we interpret the channel model in [13] in the context of our framework. In particular, the results in this paper make a direct connection with physical models and precisely identify the situations in which capacity scaling can or cannot occur.

*Summary of Results:* We investigate capacity scaling in both the low-power and large antenna ( $N$ ) regimes for both narrow-band and wide-band channels. Most of our analysis is based on general lower and upper bounds on the ergodic capacity of the *D-connected model*. First, consider the narrow-band case. In the low-power regime, we show that  $C/P$  scales precisely as  $D$  where  $C$  denotes the capacity and  $P$  the total transmit power. The analysis in the large antenna regime is greatly facilitated by a fortuitous connection between the *D-connected model* and some earlier work of Grenander and Silverstein [14] on the limiting empirical eigenvalue distribution of a class of random matrices. We show that for large  $N$ ,  $C$  exhibits linear growth with  $N$  if  $D$  grows linearly with  $N$  as well. This, in turn, implies that linear capacity growth is sustainable in physical channels if the number of resolvable paths grows quadratically with  $N$ . For fixed  $D$ , which corresponds to linear growth in the number of paths, the capacity saturates. For a finite number of paths, there is no gain in increasing  $N$  beyond the number of paths.

In the wide-band case, we show that frequency selectivity does not affect *ergodic* capacity, which is consistent with known results in the single-antenna case [15]. In fact, wide-band capacity is solely governed by spatial channel characteristics via appropriate scaling with  $W$ . The most conspicuous effect of increasing bandwidth is that it tightens the capacity bounds and hastens the convergence of scaling behavior. In particular,

for a large bandwidth, we get linear capacity growth with transmit power as well. We also investigate spectral efficiency of wide-band MIMO channels and provide explicit characterizations of the minimum energy per bit (required for reliable communication) and the wide-band slope recently proposed by Verdú [16].

The rest of this paper is organized as follows. In the next section, we present a general physical model for wide-band MIMO channels, review the virtual representation and its relation to the physical model. We also discuss channel statistics and virtual path partitioning to motivate the *D-connected channel model*. In Section III, we formally define the *D-connected model* and obtain lower and upper bounds for its capacity. Section IV presents our capacity scaling results for narrow-band MIMO channels. Section V contains a parallel set of results for wide-band channels. Section VI investigates spectral efficiency issues. In Section VII, we provide a physical interpretation of the scaling results and also provide illustrative numerical results. Section VIII closes the paper with concluding remarks. Several of the proofs are relegated to the Appendixes.

## II. WIDE-BAND MIMO CHANNEL MODELING

In this section, we review the virtual representation for both narrow-band [6] and wide-band MIMO channels [11], [12] that plays a key role in connecting the scaling results in this paper to the structure of physical MIMO channels. We focus on the aspects of the virtual representation that are particularly relevant to this paper. The reader is referred to [6], [11], [12] for details. Throughout this paper, we consider MIMO systems with ULAs of  $N$  antennas at both the transmitter and receiver and assume that far-field conditions apply.

### A. A General Physical Model for Wide-Band MIMO Channels

We are interested in representing the MIMO channel over a two-sided bandwidth  $W$ . In the absence of noise, the transmitted and received signals are related as

$$\mathbf{x}(t) = \int \mathbf{H}_{c,\text{imp}}(\tau) \mathbf{s}(t - \tau) d\tau \quad (1)$$

$$\mathbf{X}(f) = \mathbf{H}_c(f) \mathbf{S}(f), \quad -W/2 \leq f \leq W/2 \quad (2)$$

where  $\mathbf{s}(t)$  is the  $N$ -dimensional transmitted signal in time,  $\mathbf{x}(t)$  is the  $N$ -dimensional received signal in time, and  $\mathbf{S}(f)$  and  $\mathbf{X}(f)$  are Fourier transforms of  $\mathbf{s}(t)$  and  $\mathbf{x}(t)$ , respectively

$$\mathbf{S}(f) = \int \mathbf{s}(t) e^{-j2\pi ft} dt. \quad (3)$$

The  $N \times N$  matrix  $\mathbf{H}_{c,\text{imp}}(\tau)$  represents the *impulse response matrix* and  $\mathbf{H}_c(f)$  is the corresponding *frequency response matrix* (the Fourier transform of  $\mathbf{H}_{c,\text{imp}}(\tau)$ ) coupling the transmitter and receiver elements. We will primarily work with  $\mathbf{H}_c(f)$  and we index the entries of  $\mathbf{H}_c(f)$  as  $H(i, k; f)$ :  $i = 0, 1, \dots, N - 1$ ,  $k = 0, 1, \dots, N - 1$ .<sup>1</sup>

<sup>1</sup>The subscript “c” denotes the actual channel matrix, as opposed to the virtual channel matrix.

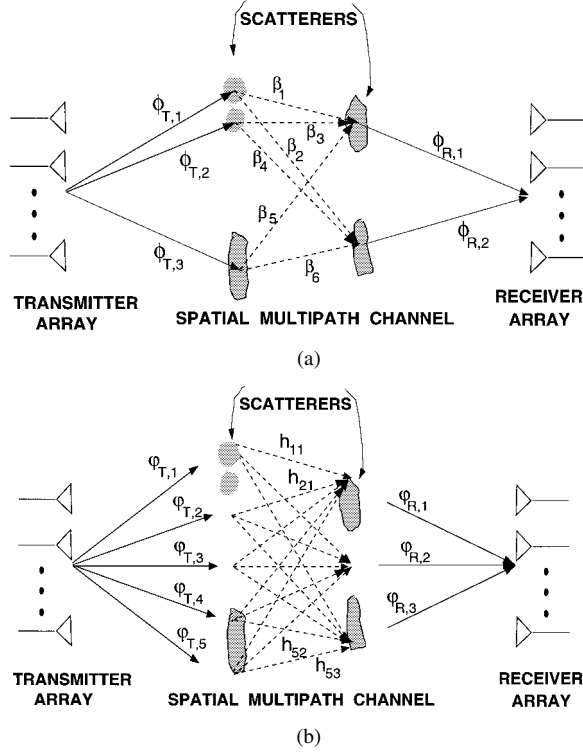


Fig. 1. A schematic illustrating *physical modeling* versus *virtual representation* in the spatial dimension. (a) *Physical Modeling*: Each scattering path is associated with a fading gain ( $\beta_n$ ) and a unique pair of transmit and receive angles ( $\phi_{T,n}, \phi_{R,n}$ ). (b) *Virtual Representation* of the scattering environment depicted in (a). The virtual angles are fixed *a priori* and their spacing defines the spatial resolution. The channel is characterized by the virtual coefficients  $\{H_V(q, p) = h_{q,p}\}$  that couple the  $N$  virtual transmit angles  $\{\varphi_{T,p}\}$  with the  $N$  virtual receive angles  $\{\varphi_{R,q}\}$ .

Let  $d_T$  and  $d_R$  denote the antenna spacings at the transmitter and receiver, respectively. The channel matrix for ULAs can be described via the array steering and response vectors

$$\begin{aligned} \mathbf{a}_T(\theta_T) &= \frac{1}{\sqrt{N}} \left[ 1, e^{-j2\pi\theta_T}, \dots, e^{-j2\pi(N-1)\theta_T} \right]^T \\ \mathbf{a}_R(\theta_R) &= \frac{1}{\sqrt{N}} \left[ 1, e^{-j2\pi\theta_R}, \dots, e^{-j2\pi(N-1)\theta_R} \right]^T \end{aligned} \quad (4)$$

where  $\theta$  is related to the AoA/AoD variable  $\phi$  (measured with respect to the horizontal axis—see Fig. 1) as  $\theta = d \sin(\phi)/\lambda = \alpha \sin(\phi)$ ,  $\lambda$  is the wavelength of propagation, and  $\alpha = d/\lambda$  is the normalized antenna spacing. We will primarily work with the spatial variable  $\theta$ . We restrict ourselves to critical ( $\lambda/2$ ) spacing:  $\alpha_T = \alpha_R = 0.5$ . In this case, there is a one-to-one mapping between  $\theta \in [-0.5, 0.5]$  and  $\phi \in [-\pi/2, \pi/2]$ . The effect of larger antenna spacing on capacity and diversity is discussed in detail in [6].

The channel matrix  $\mathbf{H}_c(f)$  can be generally modeled as

$$\mathbf{H}_c(f) = \sum_{n=1}^{N_{\text{path}}} \alpha_n \mathbf{a}_R(\theta_{R,n}) \mathbf{a}_T^H(\theta_{T,n}) e^{-j2\pi\tau_n f} \quad (5)$$

which corresponds to signal propagation along  $N_{\text{path}}$  paths, where  $\{\theta_{T,n}\}$  and  $\{\theta_{R,n}\}$  represent the AoDs and AoAs, respectively,  $\{\tau_n\}$  the delays, and  $\{\beta_n\}$  the corresponding complex path gains. The physical model is illustrated in Fig. 1(a).

A narrow-band MIMO system corresponds to  $W = 0$  in which case (5) reduces to

$$\mathbf{H}_c = \mathbf{H}_c(0) = \sum_{n=1}^{N_{\text{path}}} \alpha_n \mathbf{a}_R(\theta_{R,n}) \mathbf{a}_T^H(\theta_{T,n}). \quad (6)$$

Define

$$\begin{aligned} \theta_{R-} &= \min_n \theta_{R,n}, & \theta_{R+} &= \max_n \theta_{R,n} \\ \theta_{T-} &= \min_n \theta_{T,n}, & \theta_{T+} &= \max_n \theta_{T,n}. \end{aligned} \quad (7)$$

Then  $R_{\text{AS}} = [\theta_{R-}, \theta_{R+}]$  and  $T_{\text{AS}} = [\theta_{T-}, \theta_{T+}]$  represent the *angular spreads* seen by the receiver and transmitter, respectively. The *delay spread* is denoted by

$$\tau_{\text{DS}} = \max_n \tau_n - \min_n \tau_n. \quad (8)$$

Without loss of generality, we assume  $\min_n \tau_n = 0$  so that  $\tau_n \in [0, \tau_{\text{DS}}]$ .

A continuous version of (5), corresponding to a continuum of propagation paths, is insightful in relating the channel matrix to the scattering environment

$$\begin{aligned} \mathbf{H}_c(f) &= \int_{\theta_{R-}}^{\theta_{R+}} \int_{\theta_{T-}}^{\theta_{T+}} \int_0^{\tau_{\text{DS}}} G(\theta_R, \theta_T; \tau) \\ &\quad \cdot \mathbf{a}_R(\theta_R) \mathbf{a}_T^H(\theta_T) e^{-j2\pi\tau f} d\theta_R d\theta_T d\tau \end{aligned} \quad (9)$$

where  $G(\theta_R, \theta_T; \tau)$  denotes the *angle-delay spreading function* that characterizes the scattering environment. For the discrete model (5), it reduces to

$$G(\theta_R, \theta_T; \tau) = \sum_{n=1}^{N_{\text{path}}} \alpha_n \delta(\theta_R - \theta_{R,n}) \delta(\theta_T - \theta_{T,n}) \delta(\tau - \tau_n) \quad (10)$$

where  $\delta(\theta)$  denotes the Dirac delta function. For the narrow-band case, (9) reduces to

$$\mathbf{H}_c = \mathbf{H}_c(0) = \int_{\theta_{R-}}^{\theta_{R+}} \int_{\theta_{T-}}^{\theta_{T+}} G(\theta_R, \theta_T) \mathbf{a}_R(\theta_R) \mathbf{a}_T^H(\theta_T) d\theta_R d\theta_T \quad (11)$$

$$\begin{aligned} G(\theta_R, \theta_T) &= \int_0^{\tau_{\text{DS}}} G(\theta_R, \theta_T; \tau) d\tau \\ &= \sum_{n=1}^{N_{\text{path}}} \alpha_n \delta(\theta_R - \theta_{R,n}) \delta(\theta_T - \theta_{T,n}) \end{aligned} \quad (12)$$

where the second equality in (12) corresponds to the discrete model (6).

### B. A Virtual Representation for Wide-Band MIMO Channels

In (5), each propagation path is associated with an arbitrary AoD, AoA, and delay distributed within the angular and delay spreads. The virtual representation replaces the physical paths with virtual ones corresponding to *fixed* AoDs, AoAs, and delays that are determined by the spatial and temporal resolution of the array. The notion of virtual angles is illustrated in Fig. 1(b). Without loss of generality, assume that  $N$  is odd and define  $\tilde{N} = (N - 1)/2$ .

*Definition 1—Virtual Channel Representation:* The virtual channel representation is defined by the Fourier series expansion [6], [11], [12]

$$\mathbf{H}_c(f) = \sum_{q=-\tilde{N}}^{\tilde{N}} \sum_{p=-\tilde{N}}^{\tilde{N}} \sum_{l=0}^{\infty} \cdot H[q, p; l] \mathbf{a}_R(q/N) \mathbf{a}_T^H(p/N) e^{-j2\pi lf/W} \quad (13)$$

corresponding to *fixed virtual* AoDs, AoAs, and delays

$$\tilde{\theta}_{T,p} = p/N, \quad \tilde{\theta}_{R,q} = q/N, \quad \tilde{\tau}_l = l/W. \quad (14)$$

The virtual (Fourier series) channel coefficients  $\{H[q, p; l]\}$  characterize the virtual representation and can be computed from  $\mathbf{H}_c(f)$  as

$$H[q, p; l] = \frac{1}{W} \int_{-W/2}^{W/2} \mathbf{a}_R^H(q/N) \mathbf{H}_c(f) \mathbf{a}_T(p/N) e^{j2\pi lf/W} df. \quad (15)$$

Combining (15) and (9) we get

$$\begin{aligned} H[q, p; l] &= \hat{G}(q/N, p/N; l/W) \\ &= \sum_{n=1}^{N_{\text{path}}} \alpha_n f_N(\theta_{R,n} - q/N) f_N^*(\theta_{T,n} - p/N) \\ &\quad \cdot \text{sinc}(\tau_n W - l) \end{aligned} \quad (16)$$

where the second equality corresponds to the discrete model and

$$\begin{aligned} \hat{G}(\theta_R, \theta_T; \tau) &= \int_{\theta_{R-}}^{\theta_{R+}} \int_{\theta_{T-}}^{\theta_{T+}} \int_0^{\tau_{\text{DS}}} \\ &\quad \cdot G(\theta_{R'}, \theta_{T'}; \tau') f_N(\theta_{R'} - \theta_R) f_N^*(\theta_{T'} - \theta_T) \\ &\quad \cdot \text{sinc}(W(\tau' - \tau)) d\theta_{R'} d\theta_{T'} d\tau' \quad (17) \\ f_N(\theta) &= \frac{1}{N} \sum_{i=0}^{N-1} e^{-j2\pi\theta i} = \frac{1}{N} e^{-j2\pi\theta\tilde{N}} \frac{\sin(\pi N\theta)}{\sin(\pi\theta)}, \\ \text{sinc}(x) &= \sin(\pi x)/(\pi x). \quad (18) \end{aligned}$$

Note that  $f_N(\theta)$  and  $\text{sinc}(W\tau)$  get peaky around the origin with increasing  $N$  and  $W$ . Thus, (16) states that the virtual channel coefficients are samples of a smoothed version of the delay-angle spreading function, and that the smoothing kernel gets narrower with increasing  $N$  and  $W$ .

The spacing between the transmit/receive virtual angles in (14) represents the spatial resolution of the array:  $\Delta\theta_T = \Delta\theta_R = 1/N$ . Similarly, the spacing between the virtual delays corresponds to the temporal resolution:  $\Delta\tau = 1/W$ . For sufficiently large  $N$  and  $W$ , most of the channel power is carried by a subset of the coefficients. The size of this subset of dominant  $\{H[q, p; l]\}$  is determined by the angular and delay spreads [6], [11], [12]

$$P_- \leq p \leq P_+ : P_- = \lfloor \theta_{T-} N \rfloor, P_+ = \lceil \theta_{T+} N \rceil \quad (19)$$

$$Q_- \leq q \leq Q_+ : Q_- = \lfloor \theta_{R-} N \rfloor, Q_+ = \lceil \theta_{R+} N \rceil \quad (20)$$

$$0 \leq l \leq L : L = \lceil W\tau_{\text{DS}} \rceil. \quad (21)$$

In the narrow-band case, virtual representation reduces to

$$\begin{aligned} \mathbf{H}_c &= \mathbf{H}_c(0) = \mathbf{A}_R \mathbf{H}(0) \mathbf{A}_T^H \\ &= \mathbf{A}_R \mathbf{H} \mathbf{A}_T^H = \sum_{q=-\tilde{N}}^{\tilde{N}} \sum_{p=-\tilde{N}}^{\tilde{N}} H(q, p) \mathbf{a}_R(q/N) \mathbf{a}_T^H(p/N) \end{aligned} \quad (22)$$

$$\begin{aligned} \mathbf{A}_R &= [\mathbf{a}_R(-\tilde{N}/N), \dots, \mathbf{a}_R(\tilde{N}/N)] \\ \mathbf{A}_T &= [\mathbf{a}_T(-\tilde{N}/N), \dots, \mathbf{a}_T(\tilde{N}/N)] \end{aligned} \quad (23)$$

where  $\mathbf{A}_R$  and  $\mathbf{A}_T$  are  $N \times N$  discrete Fourier transform (DFT) (unitary) matrices. The elements of the narrow-band virtual matrix  $\mathbf{H} = \mathbf{H}(0)$  are related to the discrete physical model as

$$\begin{aligned} H(q, p) &= \hat{G}(q/N, p/N) = \sum_l H[q, p; l] \\ &= \sum_{n=1}^{N_{\text{path}}} \alpha_n f_N(\theta_{R,n} - q/N) f_N^*(\theta_{T,n} - p/N). \end{aligned} \quad (24)$$

We note that the virtual representation is a unitary transformation of the actual channel matrix and, thus, all capacity-related issues can be equivalently investigated in the virtual domain.

### C. Virtual Path Partitioning

The virtual representation induces a partitioning of paths that is very insightful in relating physical scattering characteristics to channel statistics.

*Definition 2—Virtual Path Partitioning:* Define the following subsets of propagation paths:

$$S_{T,p} = \{n: -1/2N \leq (\theta_{T,n} - p/N) < 1/2N\} \quad (25)$$

$$S_{R,q} = \{n: -1/2N \leq (\theta_{R,n} - q/N) < 1/2N\} \quad (26)$$

$$S_{\tau,l} = \{n: -1/2W \leq (\tau_n - l/W) < 1/2W\} \quad (27)$$

corresponding to the spatial and delay resolutions. The above sets form a partition

$$\bigcup_q S_{R,q} = \bigcup_p S_{T,p} = \bigcup_l S_{\tau,l} = \{1, \dots, N_{\text{path}}\} \quad (28)$$

$$\bigcup_{q,p,l} [S_{R,q} \cap S_{T,p} \cap S_{\tau,l}] = \{1, \dots, N_{\text{path}}\}. \quad (29)$$

With the path partitioning, the virtual coefficients in (16) and (24) can be approximated as

$$H[q, p; l] \approx \sum_{n \in S_{q,p,l}} \alpha_n, \quad H(q, p) \approx \sum_{n \in S_{q,p}} \alpha_n \quad (30)$$

where  $S_{q,p,l} = S_{R,q} \cap S_{T,p} \cap S_{\tau,l}$  and  $S_{q,p} = S_{R,q} \cap S_{T,p}$ . Thus, in the narrow-band case, the paths are distributed in the virtual representation according to the spatial resolution. In the wide-band case, this distribution is further refined by the delay resolution.

### D. Statistics of Wide-Band Correlated MIMO Channels

In this paper, we are interested in modeling the channel over time scales over which the locations of scatterers, and hence

$\{\theta_{T,n}\}$ ,  $\{\theta_{R,n}\}$ , and  $\{\tau_n\}$ , do not change significantly relative to the transmitter and receiver. This is equivalent to considering time scales over which the channel *statistics* do not change appreciably. However, the channel realizations do vary over such time scales due to the phase variations in path gains.

We make the following (Rayleigh fading) assumption on physical scattering.

*Assumption 1—Independent Physical Scattering:* The physical channel parameters  $\{\theta_{T,n}\}$ ,  $\{\theta_{R,n}\}$ , and  $\{\tau_n\}$  are fixed over the time scales of interest. The path gains  $\{\alpha_n\}$  are independent zero-mean complex circular Gaussian random variables with variances  $\{\sigma_n^2\}$

$$\mathbb{E}[\alpha_n \alpha_{n'}] = 0, \quad \mathbb{E}[\alpha_n \alpha_{n'}^*] = \sigma_n^2 \delta_{n-n'}, \quad \forall n, n' \quad (31)$$

where  $\delta_n$  denotes the Kronecker delta function.

Under the above assumption, the elements of  $\{\mathbf{H}_c(f)\}$  are jointly complex circular Gaussian, and, consequently, so are the virtual coefficients  $\{H[q, p; l]\}$ . Assumption 1 implies uncorrelated statistics for the spreading function in (10) [6], [11]

$$\begin{aligned} \mathbb{E}[G(\theta_R, \theta_T; \tau) G^*(\theta_{R'}, \theta_{T'}; \tau')] \\ = M(\theta_R, \theta_T; \tau) \delta(\theta_R - \theta_{R'}) \delta(\theta_T - \theta_{T'}) \delta(\tau - \tau') \quad (32) \\ M(\theta_R, \theta_T; \tau) = \sum_{n=1}^{N_{\text{path}}} \sigma_n^2 \delta(\theta_R - \theta_{R,n}) \delta(\theta_T - \theta_{T,n}) \delta(\tau - \tau_n). \quad (33) \end{aligned}$$

The nonnegative function  $M(\theta_R, \theta_T; \tau)$  in (32) is called the *angle-delay scattering function* (or *angle-delay power profile*) and reflects the distribution of channel power in the  $(\theta_R, \theta_T; \tau)$  space; it is given by (33) for the discrete model.

An important property of the virtual representation is that  $\{H[q, p; l]\}$  are approximately uncorrelated (and, hence, approximately independent) under Assumption 1 [6], [11], [12]

$$\begin{aligned} \mathbb{E}[H[q, p; l] H^*[q', p'; l']] \\ \approx \sigma_{q,p,l}^2 \delta_{q-q'} \delta_{p-p'} \delta_{l-l'} = \left[ \sum_{n \in S_{q,p,l}} \sigma_n^2 \right] \delta_{q-q'} \delta_{p-p'} \delta_{l-l'} \quad (34) \end{aligned}$$

$$\begin{aligned} \sigma_{q,p,l}^2 &= \mathbb{E}[|H[q, p; l]|^2] \\ &= \int_{\theta_{R-}}^{\theta_{R+}} \int_{\theta_{T-}}^{\theta_{T+}} \int_0^{\tau_{\text{DS}}} M(\theta_R, \theta_T; \tau) |f_N(\theta_R - q/N)|^2 \\ &\quad \cdot |f_N(\theta_T - p/N)|^2 \text{sinc}^2(W\tau - l) d\theta_R d\theta_T d\tau \quad (35) \\ &= \hat{M}(q/N, p/N; l/W) \quad (36) \end{aligned}$$

and (using (4), (5), (13))

$$\begin{aligned} \mathbb{E}[H_c(i, k; f) H_c^*(i', k'; f')] \\ = \frac{1}{N^2} \int_{\theta_{R-}}^{\theta_{R+}} \int_{\theta_{T-}}^{\theta_{T+}} \int_0^{\tau_{\text{DS}}} M(\theta_R, \theta_T; \tau) e^{-j2\pi\theta_R(i-i')} \\ \cdot e^{j2\pi\theta_T(k-k')} e^{-j2\pi\tau(f-f')} d\theta_R d\theta_T d\tau \quad (37) \\ = \frac{1}{N^2} \sum_{n=1}^{N_{\text{path}}} \sigma_n^2 e^{-j2\pi\theta_{R,n}(i-i')} e^{j2\pi\theta_{T,n}(k-k')} e^{-j2\pi\tau_n(f-f')}. \quad (38) \end{aligned}$$

Equations (38) and (34) state that the elements of  $\{\mathbf{H}_c(f)\}$  form a three-dimensional stationary random field and the virtual coefficients  $\{H[q, p; l]\}$  are samples of a smoothed version of its underlying spectral representation and are hence approximately uncorrelated. The scattering function  $M(\theta_R, \theta_T; \tau)$  can be interpreted as the power spectral density<sup>2</sup> associated with  $\mathbf{H}_c(f)$ , and the dominant virtual coefficients

$$\{H[q, p; l] : Q_- \leq q \leq Q_+, P_- \leq p \leq P_+, 0 \leq l \leq L\}$$

approximately characterize the *independent degrees of freedom* in wide-band MIMO channels.

The following observation will be useful in wide-band capacity results.

*Proposition 1:* At any given frequency  $f$ , the spatial statistics of  $\mathbf{H}_c(f)$  are independent of  $f$ .

*Proof:* The proof directly follows from (38) by substituting  $f = f'$ .  $\square$

Thus, the *spatial* statistics of  $\mathbf{H}_c(f)$  are determined by the statistics of the narrow-band matrix  $\mathbf{H}_c = \mathbf{H}_c(0)$  (or, equivalently,  $\mathbf{H} = \mathbf{H}(0)$ ). The total channel power is distributed as [11], [12]

$$\begin{aligned} \sigma_H^2 &= \frac{1}{W} \int_{-W/2}^{W/2} \mathbb{E}[\text{tr}(\mathbf{H}_c^H(f) \mathbf{H}_c(f))] df = \mathbb{E}[\text{tr}(\mathbf{H}_c^H \mathbf{H}_c)] \\ &= \iiint M(\theta_R, \theta_T; \tau) d\theta_R d\theta_T d\tau = \sum_{n=1}^{N_{\text{path}}} \sigma_n^2 = \sum_{q,p,l} \sigma_{q,p,l}^2. \quad (39) \end{aligned}$$

#### E. *d*-Diagonal Virtual Model for Narrow-Band MIMO Channels

We now motivate a simple model—the *d*-diagonal virtual model—for correlated narrow-band MIMO channels that plays a key role in the scaling results. As we will see, the ergodic capacity of wide-band MIMO channels depends only on spatial statistics of the narrow-band virtual matrix  $\mathbf{H}$ . Spectral statistics only contribute to diversity and hence affect the outage capacity [11], [12].

Consider a single scattering cluster with maximum angular spreads at the transmitter and receiver:  $T_{AS} = R_{AS} = [-1/2, 1/2]$ . One source of correlation is limited angular spreads. However, the effective angular spread (in the  $\theta$  domain) can be maximized (and the channel decorrelated) in such cases by increasing the antenna spacing [6]. For maximum angular spreads, the nature of coupling between the scatterers within the cluster determines the channel correlation. On one extreme is “diagonal scattering” ( $\mathbf{H}$  diagonal), in which each virtual transmit angle couples with only a single corresponding virtual receive angle. This corresponds to a scattering environment consisting of a single line of scatterers (see [6, Fig. 7(a)]). In this case,  $\sigma_{q,p}^2$  is nonzero only for  $q = p$  and the channel exhibits significant correlation since only  $N$  out of  $N^2$  degrees of freedom are excited. On the other extreme is “maximally rich scattering” (all elements of  $\mathbf{H}$  nonzero) in which each virtual transmit angle couples with all virtual receive angles,

<sup>2</sup>The angular spreads represent the bandwidths associated with the stationary field in the spatial dimensions.

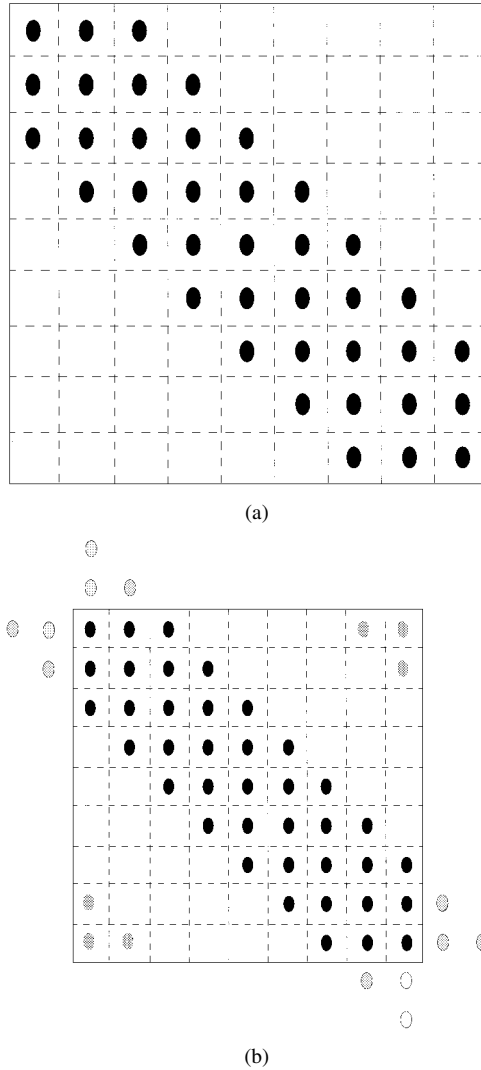


Fig. 2. A schematic illustrating the  $d$ -diagonal and circulant  $d$ -diagonal models for the virtual spatial matrix.  $N = 9$  and  $d = 2$  and each small square represents a spatial resolution bin of size  $\Delta\theta_R = \Delta\theta_T = 1/N$ . (a)  $d$ -diagonal model consisting of  $d$  nonvanishing diagonals above and below the main diagonal. Notice the truncation near the corners. (b) Circulant  $d$ -diagonal model. The truncated parts in (a) are wrapped around (aliased) and included in the matrix as the dark grey circles.

and *vice versa*. This corresponds to multiple lines of scatterers (see [6, Fig. 7(b)]) and the channel will exhibit minimal correlation since all  $\sigma_{q,p}^2$  are nonzero. In particular,  $\sigma_{q,p}^2 = \sigma^2$  corresponds to the i.i.d. model. Thus, we can capture a rich class of scattering environments, depicting varying levels of correlation, by imposing the following  $d$ -diagonal structure on  $\mathbf{H}$ :

$$\text{for } -\tilde{N} \leq q \leq \tilde{N} \quad (40)$$

$$H(q, p) = \begin{cases} \text{nonzero,} & \text{for } \max(-\tilde{N}, q-d) \\ & \leq p \leq \min(\tilde{N}, q+d) \\ 0, & \text{otherwise} \end{cases} \quad (41)$$

where  $0 \leq d \leq \tilde{N} - 1$  represents the number of nonvanishing diagonals above and below the main diagonal (see Fig. 2(a)). Diagonal scattering corresponds to  $d = 0$  and maximally rich scattering corresponds to  $d = \tilde{N} - 1$ .

The scaling results presented in subsequent sections are based on a  $D$ -connected model which corresponds to the following circulant definition of  $d$ -diagonal model:

$$\text{for } -\tilde{N} \leq q \leq \tilde{N} \quad (42)$$

$$H(q, p) = \begin{cases} \text{nonzero,} & \text{for } p = (q + i) \bmod N, \\ & \text{where } i = -d, \dots, d \\ 0, & \text{otherwise.} \end{cases} \quad (43)$$

The circulant modification is made for a technical reason—to make the number of nonvanishing elements in each column and each row to be the same—and does not affect the essential conclusions of our results. It can be shown that the circulant structure in (43) actually occurs in systems which employ larger than  $\lambda/2$  antenna spacing due to notion of *spatial aliasing* [6].<sup>3</sup> The difference between  $d$ -diagonal and circulant  $d$ -diagonal models is illustrated in Fig. 2. For the  $d$ -diagonal model in Fig. 2(a), notice the truncation near the corners of the matrix. In the circulant  $d$ -diagonal model in Fig. 2(b), the truncated parts in Fig. 2(a) are wrapped around (aliased) and included in the matrix as depicted by the dark grey circles.<sup>4</sup>

We assumed  $N$  to be odd in the above discussion. We will relax this assumption in the following sections. Furthermore, in the  $D$ -connected model introduced in Section III,  $D$  corresponds to the total number of diagonals;  $D = 2d + 1$  for  $D$  odd and  $D = 2d$  for  $D$  even.

#### 1) Alternative Interpretations for the $D$ -Connected Model:

Let  $p(\theta_R, \theta_T)$  denote the joint density of path angles; the angular spreads correspond to supports of the marginal densities. Consider maximum angular spreads:  $T_{AS} = R_{AS} = [-1/2, 1/2]$ . The path angles  $\{(\theta_{R,n}, \theta_{T,n})\}$  can be thought of as drawn independently according to  $p(\theta_R, \theta_T)$ . The  $D$ -connected model corresponds to the following structure on the conditional density of  $\theta_R$  given  $\theta_T$ :

$$p(\theta_R|\theta_T) = \begin{cases} \text{nonzero,} & \text{if } \max(\theta_T - \delta\theta, -1/2) \\ & \leq \theta_R \leq \min(\theta_T + \delta\theta, 1/2) \\ 0, & \text{otherwise} \end{cases} \quad (44)$$

where  $\delta\theta \in [0, 1/2]$ . Thus, even though the marginals  $p(\theta_R)$  and  $p(\theta_T)$  span the entire angular spreads, the conditional density exhibits a limited spread;  $R_{AS}(\theta_T) = [\theta_T - \delta\theta, \theta_T + \delta\theta]$ . Note that  $p(\theta_R, \theta_T)$  is a normalized version of power spectral density

$$M(\theta_R, \theta_T) = \int_0^{\tau_{DS}} M(\theta_R, \theta_T; \tau) d\tau$$

and may be estimated in practice from measurements (see, e.g., [4]). Finally, since the sampling resolution is  $\Delta\theta = 1/N$ ,  $D \approx 2\delta\theta N$  in (44).

<sup>3</sup>For larger than  $\lambda/2$  antenna spacing, the principal  $\theta$  range ( $[-1/2, 1/2]$ ) maps into a subset of the physical angle  $\phi$  range ( $[-\pi/2, \pi/2]$ ) (e.g., the black dots in Fig. 2(b)). However, due to the periodicity of steering and response vectors in  $\theta$ , scatterers outside the limited  $\phi$  range wrap into the principal  $\theta$  range (the grey dots in Fig. 2(b)). This is the notion of spatial aliasing.

<sup>4</sup>Note that only two light-grey colored truncated parts are included in the circulant  $d$ -diagonal matrix. In general, all four truncated parts may be aliased to yield the dark-grey parts. This would result in the dark-grey parts having twice as much power as the black parts. However, such a situation is less likely in practice and is not critical to the essence of our scaling results. Thus, this technical point is ignored in the definition of the  $D$ -connected model.

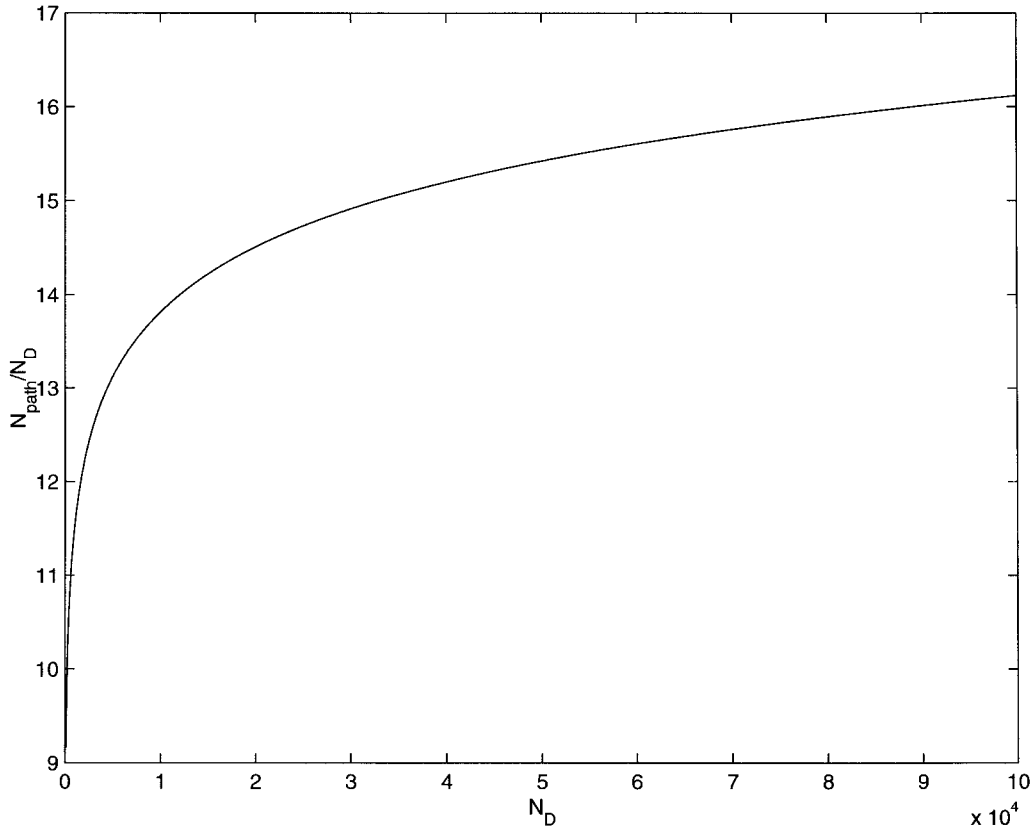


Fig. 3. Plot of  $N_{\text{path}}/N_D$  as a function of  $N_D$  for  $p_{\text{pop}} = 0.99$ .

2) *Number of Paths Needed to Populate a D-Connected Channel:* For given  $N$  and  $D$ , a natural question (that will be useful later) is: how many propagation paths are needed to populate a  $D$ -connected channel? The relation (30) states that we need at least as many paths as the number of nonvanishing entries,  $N_D = DN$  (see Fig. 2(b)), in the  $D$ -connected model. This applies to *resolvable paths* that lie in distinct virtual spatial bins of size  $\Delta\theta_R \times \Delta\theta_T = 1/N \times 1/N$ , as depicted in Fig. 2. However, since the path angles are randomly distributed, more than  $N_D$  paths will be needed to ensure with high probability that there is at least one path in each spatial bin. Assume that  $\{\theta_{R,n}, \theta_{T,n}\}$  are uniformly distributed over the  $(\theta_R, \theta_T)$  region in (44). Then, each path can land, with equal probability, in any of the  $N_D$  spatial bins and it can be shown that the probability that the  $D$ -connected model is fully populated satisfies

$$p_{\text{pop}} \geq 1 - N_D(1 - 1/N_D)^{N_{\text{path}}} \\ \Leftrightarrow N_{\text{path}} \geq \frac{\log((1 - p_{\text{pop}})/N_D)}{\log(1 - 1/N_D)}. \quad (45)$$

Fig. 3 plots the values of  $N_{\text{path}}/N_D$  as a function of  $N_D$  for  $p_{\text{pop}} = 0.99$ . Even though  $\lim_{N_D \rightarrow \infty} N_{\text{path}}/N_D = \infty$ , it is evident that  $N_{\text{path}}/N_D$  is bounded by a constant on the order of 10 for values of  $N_{\text{path}}$  of interest. Thus, we conclude that the number of paths needed to populate a  $D$ -connected channel with high probability satisfies

$$N_{\text{path}} \geq cN_D = cDN \quad (c \approx 10). \quad (46)$$

### III. $D$ -CONNECTED CHANNEL MODEL AND CAPACITY BOUNDS

In the following, we focus on narrow-band systems and impose a spatial structure on correlated MIMO channels via the virtual representation. The channel equation at any time instant can be written equivalently in the *virtual* domain as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \quad (47)$$

where  $\mathbf{x}$ ,  $\mathbf{y}$ , and  $\mathbf{z}$  are the  $N$ -dimensional transmitted signal vector, receive signal vector, and complex Gaussian noise vector, respectively. The noise vector is assumed to be white in space as well as over time. We assume that  $\mathbf{z}$  is zero-mean complex Gaussian with unit variance entries. The entries of the  $N \times N$  narrow-band virtual matrix  $\mathbf{H}$  are uncorrelated zero-mean complex Gaussian random variables whose variance may vary depending on the physical environment. In particular, many entries may be zero if the scattering is not rich enough to couple all the transmit and receive dimensions.

A concise way to describe such a pattern in  $\mathbf{H}$  is via the notion of the Hadamard product. Let  $\mathbf{A} = (a_{ij})$  and  $\mathbf{B} = (b_{ij})$  be  $N \times N$  matrices. The entries of the Hadamard product (written as  $\mathbf{C} = \mathbf{A} \odot \mathbf{B}$ ) are given by

$$c_{ij} = a_{ij}b_{ij}, \quad \forall 1 \leq i, j \leq N. \quad (48)$$

Let  $\mathbf{M} = (\sigma_{ij}^2)$  where  $\sigma_{ij}^2$  is the variance of the  $(i, j)$ th entry in  $\mathbf{H}$ . The virtual channel matrix can then be expressed as

$$\mathbf{H} = \mathbf{M} \odot \mathbf{H}' \quad (49)$$

where  $\mathbf{H}'$  consists of i.i.d. standard complex Gaussian random variables. Therefore, one can describe channel structure by specifying  $\mathbf{M}$ , which we call the *channel pattern mask*. We assume that the nonvanishing elements of  $\mathbf{H}$  have identical variance.

*Definition 3:* A  $D$ -connected channel with dimension  $N$ ,  $1 \leq D \leq N$ , is an  $N \times N$  MIMO channel whose channel pattern mask  $\mathbf{G} = (g_{ij})_{N \times N}$  is given by

$$g_{ij} = \begin{cases} 1, & \text{if } j = i + l \bmod N \\ & \text{where } l = \begin{cases} -(D-1)/2, \dots, (D-1)/2, & \text{if } D \text{ is odd} \\ -D/2 + 1, \dots, D/2, & \text{if } D \text{ is even.} \end{cases} \\ 0, & \text{otherwise.} \end{cases} \quad (50)$$

Note that the pattern mask matrix of  $D$ -connected channels is a *circulant* matrix with equal number ( $D$ ) of nonzero (unit) entries in each row and each column. The parameter  $D$  is called the *channel connectivity*. It models the degree of coupling between transmit–receive antenna pairs. When  $D = 1$ , the antenna array is loosely coupled (strongly correlated), while  $D = N$  represents a densely coupled rich scattering environment (completely uncorrelated). Fig. 4 illustrates a 3-connected channel of dimension 5. Its channel pattern mask is given by

$$\mathbf{M} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

*Proposition 2:* Let  $\mathbf{H}$  be the channel matrix of a  $D$ -connected channel of dimension  $N$ . Then

$$\mathbb{E}[\mathbf{H}\mathbf{H}^H] = \mathbb{E}[\mathbf{H}^H\mathbf{H}] = D\mathbf{I}_N. \quad (51)$$

*Proof:* It trivially follows from the definition of  $D$ -connected channels.  $\square$

The channel connectivity  $D$  has a significant effect on the statistics of  $\mathbf{H}$ . The special case of  $D$ -connected channel when  $D = N$  precisely gives rise to the commonly used i.i.d. channel model. As pointed out in [2], [17],  $\mathbf{H}$  in this case is *isotropically* distributed, that is,  $\mathbf{H}$  and  $\mathbf{H}\mathbf{U}$  have the same distribution for any unitary matrix  $\mathbf{U}$ . The exploitation of the isotropic property of i.i.d.  $\mathbf{H}$  has been the key to many elegant results regarding capacity and coding for i.i.d. channels (see, e.g., [2], [17]). However, if  $D \neq N$ ,  $\mathbf{H}$  is no longer isotropically distributed. In other words, isotropic property of  $\mathbf{H}$  is rather a rarity in correlated channels, such as  $D$ -connected channels. However, a weaker form of channel statistics turns out to be preserved as  $D$  scales. The following lemma connects the statistics of  $D$ -connected channels to a (scaled) doubly stochastic matrix (DSM), which plays a role analogous to that of the isotropic distribution in i.i.d. channels.

*Lemma 1—Scaled Doubly Stochastic Matrix:* Let  $\mathbf{H}$  represent a  $D$ -connected channel. For an arbitrary unitary matrix  $\mathbf{U}$ ,

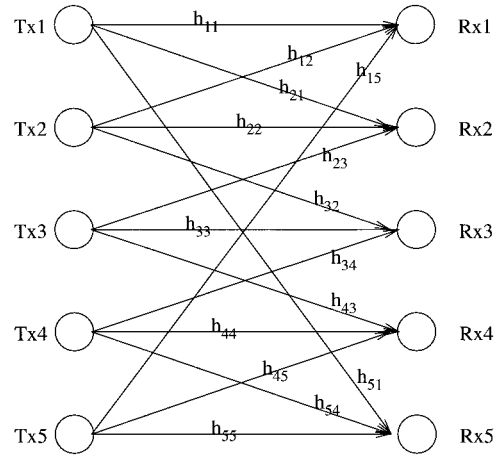


Fig. 4. A schematic illustrating  $D$ -connected channels. The dimension is 5 and the connectivity is 3. Observe each transmit dimension is coupled with three receive dimensions and *vice versa*.

denote  $\mathbf{A} = \mathbf{H}\mathbf{U}$  and  $\mathbf{B} = (\mathbb{E}|a_{ij}|^2)_{N \times N}$ . Then,  $\mathbf{B}$  is a scaled doubly stochastic matrix with scale  $D$ , that is,

$$\sum_{j=1}^N \mathbb{E}|a_{ij}|^2 = D, \quad \forall i = 1, \dots, N \quad (52)$$

$$\sum_{i=1}^N \mathbb{E}|a_{ij}|^2 = D, \quad \forall j = 1, \dots, N. \quad (53)$$

*Proof:* Since  $\mathbf{U}$  is a unitary matrix, (52) holds by noticing

$$\sum_{j=1}^N |a_{ij}|^2 = \sum_{j=1}^N |h_{ij}|^2, \quad \forall i = 1, \dots, N. \quad (54)$$

The verification of (53) requires the circular property of the  $D$ -connected channel. For given  $j$

$$\sum_{i=1}^N \mathbb{E}|a_{ij}|^2 = \sum_{i=1}^N \sum_{k \in \mathcal{I}_i} |u_{kj}|^2 \quad (55)$$

where  $u_{kj}$  is the  $(k, j)$ th entry in  $\mathbf{U}$  and  $\mathcal{I}_i$  is the set of column indexes of  $\mathbf{H}$  corresponding to nonzero entries in the  $i$ th row of the channel pattern mask matrix. Note that the  $D$ -connected channel structure implies that every row index of the  $j$ th column of  $\mathbf{U}$  is covered exactly  $D$  times, and hence the above sum is equal to  $D$ , which proves the lemma.  $\square$

Assuming perfect knowledge of  $\mathbf{H}$  at the receiver, the ergodic capacity of an  $N \times N$  Gaussian MIMO channel is given by [2]

$$C = \max_{\mathbf{Q}} \mathbb{E} \left[ \log \det \left( \mathbf{I}_N + \mathbf{H}\mathbf{Q}\mathbf{H}^H \right) \right] \quad (\text{nats/s}) \quad (56)$$

where the maximization is over a set of positive semidefinite Hermitian matrices  $\mathbf{Q}$  satisfying the power constraint  $\text{tr} \mathbf{Q} \leq P$ , and the expectation is with respect to random channel matrix  $\mathbf{H}$ . In the following, the base  $e$  is implicitly assumed for log unless specified otherwise.

An obstacle in analyzing capacity of realistic MIMO channels is the optimal input distribution  $\mathbf{Q}$  in (56). Except for diagonal channels ( $D = 1$ ), i.i.d. channels ( $D = N$ ), and a few other correlated cases [18]–[20], the optimal  $\mathbf{Q}$  is unknown, which limits the strength of capacity results for these channels. The



doubly stochastic structure in Lemma 1 greatly facilitates capacity analysis of  $D$ -connected channels. The following upper bound plays a vital role in our capacity analysis of  $D$ -connected channels. Although it can be alternatively proved by exploiting the DSM property (Lemma 1) as in Appendix I, we present a simpler proof.<sup>5</sup>

*Lemma 2—General Upper Bound:* The capacity of a  $D$ -connected channel of dimension  $N$  is upper-bounded by

$$C \leq N \log \left( 1 + \frac{D}{N} P \right) \quad (57)$$

where  $P$  is the total transmit power.

*Proof:* The key is the fact that  $\det(\mathbf{A}) \leq (\text{tr}(\mathbf{A})/N)^N$  for an  $N \times N$  positive semidefinite matrix  $\mathbf{A}$ , which is a disguised form of geometric mean  $\leq$  arithmetic mean. The lemma is proved by the following chain of inequalities:

$$\begin{aligned} \mathbb{E} \left[ \log \det \left( \mathbf{I} + \mathbf{H} \mathbf{Q} \mathbf{H}^H \right) \right] &\leq \mathbb{E} \left[ \log \left( 1 + \frac{\text{tr}(\mathbf{H} \mathbf{Q} \mathbf{H}^H)}{N} \right)^N \right] \\ &\leq N \log \left( 1 + \frac{\mathbb{E}[\text{tr}(\mathbf{H} \mathbf{Q} \mathbf{H}^H)]}{N} \right) \\ &= N \log \left( 1 + \frac{\text{tr}(\mathbf{Q} \mathbb{E}[\mathbf{H}^H \mathbf{H}])}{N} \right) \\ &\leq N \log \left( 1 + \frac{D}{N} P \right). \end{aligned} \quad (58)$$

□

We next give two lower bounds for capacity of  $D$ -connected channels. The first (Lemma 3) follows from a quick observation that channel capacity is lower-bounded by the mutual information corresponding to the uniform power input distribution, that is,  $\mathbf{Q} = (P/N)\mathbf{I}_N$  in (56). The key to the second (Lemma 4) is to construct a suboptimal channel from the original MIMO channel and then to evaluate the capacity of the suboptimal channel, thus obtaining a looser but more tractable lower bound.

*Lemma 3—Uniform Power Lower Bound:* Given total transmit power  $P$ , the capacity of a  $D$ -connected channel of dimension  $N$  is lower-bounded by

$$C \geq \mathbb{E} \left[ \log \det \left( \mathbf{I}_N + \frac{P}{N} \mathbf{H} \mathbf{H}^H \right) \right] \quad (59)$$

where  $P$  is the total transmit power.

*Lemma 4—Rayleigh Subchannel Lower Bound:* Given total transmit power  $P$ , the capacity of a  $D$ -connected channel of dimension  $N$  is lower-bounded by

$$C \geq (N - D + 1) \mathbb{E} \left[ \log \left( 1 + \frac{P}{N} \chi_2^2 \right) \right] \quad (60)$$

where  $\chi_2^2$  is a unit variance chi-square random variable with two degrees of freedom.

*Proof:* See Appendix II. □

<sup>5</sup>Thanks to the input of an anonymous reviewer.

## IV. CAPACITY SCALING IN NARROW-BAND CHANNELS

### A. Finite Connectivity

We first study capacity saturation and scaling behavior when the channel connectivity  $D$  is finite.

#### 1) Large-Dimensional Asymptotics:

*Theorem 1—Capacity Saturation:* Channel capacity of an  $N \times N$  MIMO channel with fixed connectivity  $D$  at a given transmit power  $P$  is asymptotically bounded between  $P$  and  $DP$  as  $N$  approaches infinity.<sup>6</sup>

*Proof:* We write  $C(N)$  to emphasize capacity as a function of array dimension. The upper bound is an immediate corollary of the general upper bound lemma (Lemma 2) since

$$C(N) \leq N \log \left( 1 + \frac{D}{N} P \right) \rightarrow DP, \quad \text{as } N \rightarrow \infty. \quad (61)$$

In view of Lemma 4, channel capacity is lower-bounded by

$$C(N) \geq (N - D + 1) \mathbb{E} \left[ \log \left( 1 + \frac{P}{N} \chi_2^2 \right) \right] \quad (62)$$

where  $\chi_2^2$  is a unit variance chi-square random variable with two degrees of freedom. Note

$$N \log \left( 1 + \frac{P}{N} \chi_2^2 \right) \rightarrow P \chi_2^2 \quad \text{pointwise.}$$

Since

$$N \log \left( 1 + \frac{P}{N} \chi_2^2 \right) \leq P \chi_2^2$$

by  $\log(1+x) \leq x$  and  $\mathbb{E}[P \chi_2^2] = P < \infty$ , the dominated convergence theorem (DCT) [21] implies that

$$\begin{aligned} \lim_{N \rightarrow \infty} (N - D + 1) \mathbb{E} \left[ \log \left( 1 + \frac{P}{N} \chi_2^2 \right) \right] \\ = \lim_{N \rightarrow \infty} \frac{N - D + 1}{N} \mathbb{E} \left[ N \log \left( 1 + \frac{P}{N} \chi_2^2 \right) \right] = P. \end{aligned} \quad (63)$$

Therefore,

$$P \leq \liminf_{N \rightarrow \infty} C(N) \leq \limsup_{N \rightarrow \infty} C(N) \leq DP \quad (64)$$

which completes the proof. □

2) *Low-Power Regime:* We write  $C(P)$  to emphasize capacity as a function of transmit power  $P$ .

*Theorem 2—Capacity Scaling:* For fixed  $N$  and  $D$  ( $1 \leq D \leq N$ ), the capacity of a  $D$ -connected channel scales like  $DP$  as  $P$  becomes small. More precisely

$$\lim_{P \rightarrow 0} \frac{C(P)}{P} = D. \quad (65)$$

Moreover, any Gaussian input satisfying  $\text{tr} \mathbf{Q} = P$  achieves capacity asymptotically in the low-power regime. In particular, uniform power distribution, i.e.,  $\mathbf{Q} = \frac{P}{N} \mathbf{I}$  is asymptotically optimal.

<sup>6</sup>A series  $\{x_n\}$  is said to be asymptotically bounded between  $A$  and  $B$  iff  $A \leq \liminf_{n \rightarrow \infty} x_n \leq \limsup_{n \rightarrow \infty} x_n \leq B$ .

*Proof:* Again, an application of the general upper bound reads

$$\limsup_{P \rightarrow 0} \frac{C(P)}{P} \leq \lim_{P \rightarrow 0} \frac{N}{P} \log\left(1 + \frac{D}{N} P\right) = D. \quad (66)$$

For the other direction, denote by  $\lambda_i$  for  $1 \leq i \leq N$  the unordered eigenvalues of  $\mathbf{H}\mathbf{Q}\mathbf{H}^H$ , where  $\mathbf{Q}$  is the input power distribution satisfying the power constraint  $P$ . The unordered eigenvalues are obtained by random permutation of all eigenvalues of  $\mathbf{H}\mathbf{Q}\mathbf{H}^H$ . Note that all unordered eigenvalues have the same marginal distribution. It follows from  $\text{tr}(\mathbf{H}\mathbf{Q}\mathbf{H}^H) = \sum_{i=1}^N \lambda_i$  and  $\text{tr}\mathbf{Q} = P$  that  $\mathbb{E}\lambda = \frac{D}{N}P$ . Let  $I(P)$  denote the mutual information corresponding to power  $P$ . Similar to the proof of Theorem 1, the DCT can justify passing the limit in the following:

$$\liminf_{P \rightarrow 0} \frac{C(P)}{P} \geq \lim_{P \rightarrow 0} \frac{I(P)}{P} = \lim_{P \rightarrow 0} \frac{N\mathbb{E}[\log(1 + \lambda)]}{P} = D. \quad (67)$$

### B. Infinite Connectivity

A particularly interesting scenario is when the scattering environment is rich enough to sustain a growth in channel connectivity with antenna dimension. We study the asymptotic (large  $N$ ) capacity scaling behavior in such infinite connectivity environments.

1) *Empirical Spectral Distribution of Large-Dimensional Random Matrices:* The essential mathematical tool we will be using in studying the infinite connectivity case is the so-called *spectral analysis of random matrices*. Interested readers are referred to [22] for an excellent review on this subject. In the following, we make a brief introduction and clarify some common misconceptions in applying the random matrix theory.

*Definition 4:* Let  $\mathbf{A}$  be an  $N \times N$  Hermitian matrix and denote by  $\lambda_1 \leq \dots \leq \lambda_N$  its eigenvalues. The *empirical spectral distribution* (ESD) of  $\mathbf{A}$  is defined by

$$F_N^{\mathbf{A}}(\lambda) = \frac{\#\{k \leq N, \lambda_k \leq \lambda\}}{N} \quad (68)$$

where  $\#\{\cdot\}$  denotes the number of elements in the set indicated.

Note that  $F^{\mathbf{A}}$  itself is a *random variable* as it depends on outcomes of random matrix  $\mathbf{A}$ . A common practice is to regard  $F^{\mathbf{A}}$  simply as the distribution of eigenvalues of  $\mathbf{A}$ . In most applications, the quantities in interest, such as channel capacity, can be expressed as

$$S(N) = \mathbb{E} \int f_N(\lambda) dF_N^{\mathbf{A}_N}(\lambda) \quad (69)$$

where the expectation is with respect to random matrix  $\mathbf{A}_N$  of dimension  $N$ . For some type of random matrices, their ESDs tend to converge in a certain sense as the dimension gets large. A prevailing practice in engineering is to treat  $S(N)$  as convergent. However, the next theorem demonstrates the crucial difference between ESD and distribution of eigenvalues, which sets a rigorous ground for our channel capacity investigation.

*Theorem 3:* Let  $S(N)$  be as in (69). Assume the set of non-negative functions  $\{f_N(x)\}$  is *equicontinuous* [23], that is, if

for every  $\epsilon > 0$ , there is a neighborhood  $U_x$  of  $x$  such that  $|f_N(y) - f_N(x)| < \epsilon$  for all  $y \in U_x$  and all  $N$ . Suppose that  $f_N$  converges to  $f$  pointwise and that the ESDs  $\{F_N(\lambda)\}$  converge pointwise as  $N \rightarrow \infty$  to a deterministic  $F(\lambda)$  almost surely, written as  $F_N \rightarrow F$  a.s. Then

$$\liminf_{N \rightarrow \infty} S(N) \geq \int f(\lambda) dF(\lambda). \quad (70)$$

*Proof:* See Appendix III  $\square$

2) *Large Dimension Asymptotics:* Consider a series of  $D$ -connected channels  $\{\mathbf{H}_N\}$  with increasing dimension  $N$ . We study its capacity behavior when channel connectivity grows properly with  $N$ . More precisely

$$\lim_{n \rightarrow \infty} \frac{D(N)}{N} = \gamma \in [0, 1] \quad \text{and} \quad D(N) \rightarrow \infty \quad (71)$$

where we write  $D(N)$  to emphasize the dependency of connectivity on dimension and  $\gamma$  is called the *growth ratio* of connectivity.

A similar model has been considered in [14] when partially connected neural networks exhibit *limiting ESD* (LESD). Although the random network in [14] consists of  $\{-1, 1\}$ -valued random variables, its adaptation to our case is essentially straightforward. We state the following theorem and relegate its proof to Appendix IV.

*Theorem 4—Grenander and Silverstein'77:* Assume  $D(N) \rightarrow \infty$  as  $N \rightarrow \infty$ . Let  $F_N(\lambda)$  be the ESD of  $\frac{1}{D(N)}\mathbf{H}\mathbf{H}^H$ . Then,  $\{F_N(\lambda)\}$  converges pointwise in probability to

$$F(\lambda) = \begin{cases} 0, & \text{for } \lambda \leq 0 \\ \int_0^\lambda g(u) du, & \text{for } 0 \leq \lambda \leq 4 \\ 1, & \text{for } \lambda \geq 4 \end{cases} \quad (72)$$

where

$$g(u) = \frac{\sqrt{u(4-u)}}{2\pi u} \quad (73)$$

is the so-called Marčenko–Pastur law.

The ESDs of random  $D$ -connected channel matrices are shown in Fig. 5. As seen from the figure, the ESDs for  $N \geq 20$  are quite close to the limiting Marčenko–Pastur law.

With the aid of the above results, we come to the main theorem of this section.

*Theorem 5—Normalized Capacity Scaling:* For fixed transmit power  $P$ , if the channel connectivity grows properly with dimension  $N$ , that is, (71) is satisfied, then the capacity per dimension is asymptotically bounded as

$$\int \log(1 + \gamma P \lambda) dF(\lambda) \leq \frac{C(N)}{N} \leq \log(1 + \gamma P) \quad (74)$$

where  $F(\lambda)$  is the Marčenko–Pastur law given in (72) and  $\gamma$  is the growth ratio.

*Proof:* Applying Jensen's inequality and carrying out integration, one can verify, indeed, that

$$\int \log(1 + \gamma P \lambda) dF(\lambda) \leq \log(1 + \gamma P).$$

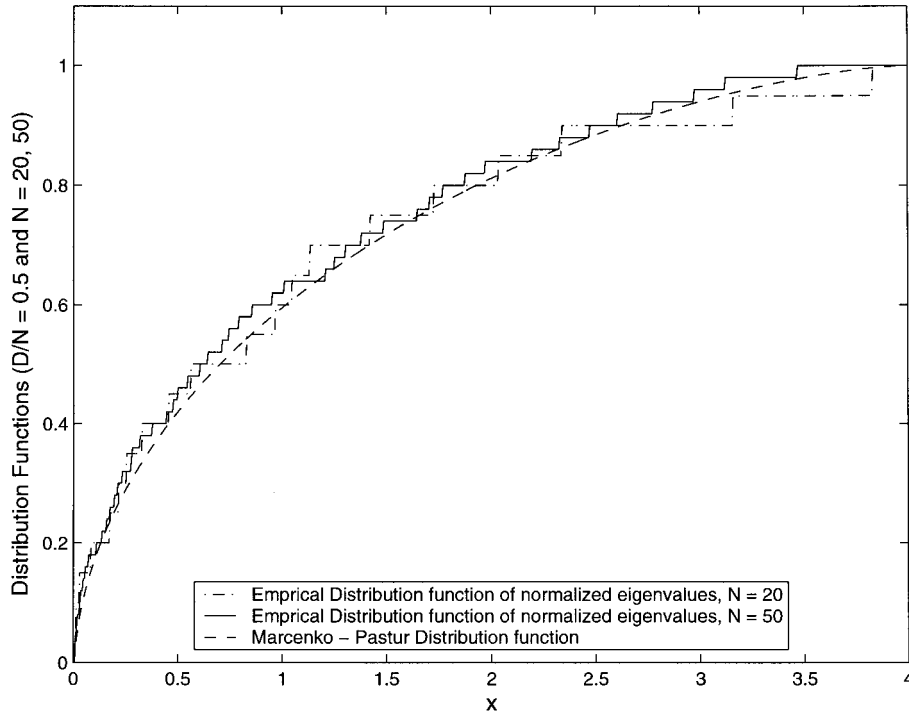


Fig. 5. The empirical spectral distributions of random  $D$ -connected channel matrices. The growth ratio is  $\gamma = 0.5$ . The Marcenko–Pastur law is also plotted.

The upper bound in (74) is again obtained from Lemma 2 by noting that

$$\frac{C(N)}{N} \leq \log\left(1 + \frac{D(N)}{N} P\right) \rightarrow \log(1 + \gamma P).$$

We resort to Theorem 3 for the lower bound. The key is to evaluate the mutual information for uniform power input. One has

$$\begin{aligned} \frac{I_{\text{uni}}(N)}{N} &= \mathbb{E} \frac{1}{N} \sum_{i=1}^N \log\left(1 + \frac{PD(N)}{N} \lambda_i\right) \\ &= \mathbb{E} \int \log\left(1 + \frac{PD(N)}{N} \lambda\right) dF_N(\lambda) \end{aligned} \quad (75)$$

where  $\lambda_i$ 's are the unordered eigenvalues of  $\frac{1}{D(N)} \mathbf{H}\mathbf{H}^H$  and  $F_N(\lambda)$  is the corresponding ESD. For  $a, x, y \geq 0$

$$\begin{aligned} &|\log(1 + ax) - \log(1 + ay)| \\ &= \left| \frac{d}{dt} \log(1 + at) \right|_{\xi} |x - y| \\ &= \frac{a}{1 + a\xi} |x - y| \leq a|x - y| \end{aligned} \quad (76)$$

for some  $\xi \in [\min(x, y), \max(x, y)]$ . Since  $\frac{D(N)}{N} \leq 1$

$$\sup_N \left| \log\left(1 + \frac{D(N)}{N} Px\right) - \log\left(1 + \frac{D(N)}{N} Py\right) \right| \leq P|x - y| \quad (77)$$

and, hence,  $\{\log(1 + \frac{D(N)}{N} P\lambda)\}$  are equicontinuous. Also, for each  $\lambda$

$$\log\left(1 + \frac{D(N)}{N} P\lambda\right) \rightarrow \log(1 + \gamma P\lambda).$$

In Theorem 4, the convergence of ESD is only in probability. However, there exists a subsequence  $\{N_k\} \rightarrow \infty$  such that the corresponding ESD converges almost surely [21]. For the investigation on capacity of a large antenna array, this seems to pose no serious constraint. So, we neglect this difficulty here.<sup>7</sup> Then, it follows from Theorem 3 that

$$\begin{aligned} \liminf_{N \rightarrow \infty} \frac{C(N)}{N} &\geq \liminf_{N \rightarrow \infty} \frac{I_{\text{uni}}(N)}{N} \\ &\geq \int \log(1 + \gamma P\lambda) dF(\lambda). \end{aligned} \quad (78)$$

Thus, the proof is complete.  $\square$

Fig. 6 shows plots of normalized capacity (upper and lower bounds) as a function of  $N$  for  $\gamma = 0.5$ . The asymptotic limits of the bounds are also plotted. It can be seen that the “capacity” corresponding to the uniform power distribution (lower bound) seems to converge exactly to the lower limit calculated from the Marcenko–Pastur law. We would like to leave it as a conjecture although we have shown a weaker result in Theorem 5.

## V. CAPACITY SCALING IN WIDE-BAND MIMO CHANNELS

In this section, we discuss the ergodic capacity of the wide-band MIMO channel characterized by the transfer function matrix  $\mathbf{H}(f)$  in the virtual spatial domain. Assuming perfect knowledge of  $\mathbf{H}(f)$  at the receiver, the ergodic capacity of the wide-band channel is given by

$$C_{\text{WB}} = \max_{\{\mathbf{Q}(f)\}} \int_{-W/2}^{W/2} \mathbb{E}[\log \det(\mathbf{I} + \mathbf{H}(f)\mathbf{Q}(f)\mathbf{H}^H(f))] df \quad (79)$$

<sup>7</sup>It seems from [22] that the convergence is almost surely.

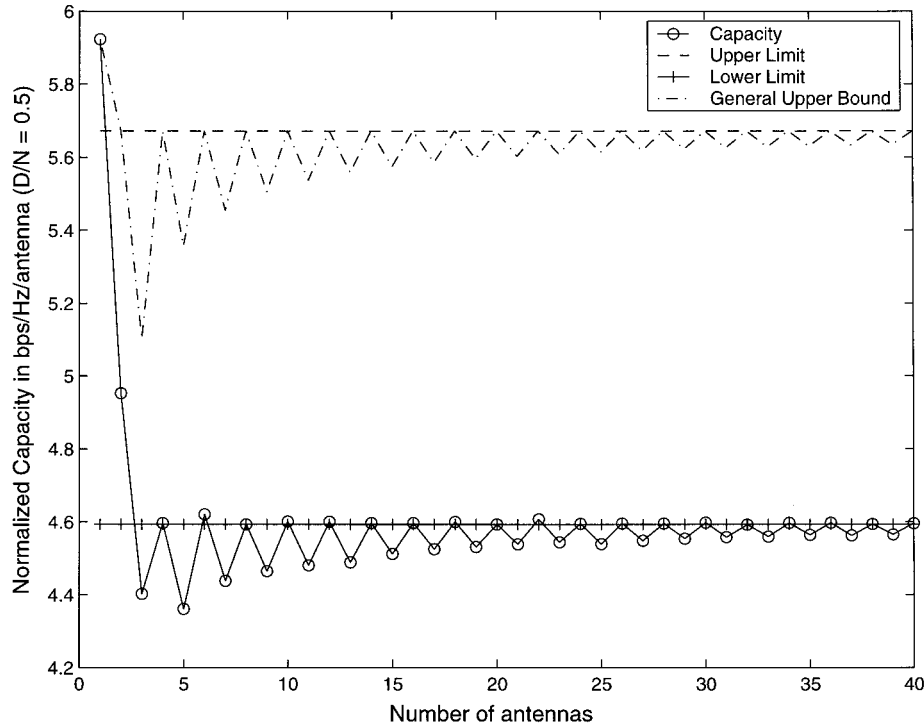


Fig. 6. Normalized narrow-band capacity  $C(N)/N$  as a function of  $N$  in an infinite connectivity channel. The growth ratio  $\gamma = 0.5$  and  $P = 20$  dB.

where the maximization is over a family of input Gaussian covariance matrices  $\{\mathbf{Q}(f), -W/2 \leq f \leq W/2\}$  that satisfy the total power constraint

$$\int_{-W/2}^{W/2} \text{tr}(\mathbf{Q}(f)) df \leq P. \quad (80)$$

The above definition of wide-band ergodic capacity is consistent with that obtained in [24] for single-antenna frequency selective additive white Gaussian noise (AWGN) channels.

Since the integrand in (79) is nonnegative, finding the optimal family  $\{\mathbf{Q}(f)\}$  is equivalent to finding the optimal  $\mathbf{Q}(f)$  at each  $f$ . Furthermore, at any  $f$ , the optimal  $\mathbf{Q}(f)$  only depends on the spatial statistics of  $\mathbf{H}(f)$ . From Proposition 1, we know that the spatial statistics of  $\mathbf{H}(f)$  are independent of  $f$ . Thus, the same  $\mathbf{Q}(f)$  is optimal for all  $f$  and the expression (79) for wide-band ergodic capacity reduces to

$$C_{\text{WB}} = \max_{\text{tr} \mathbf{Q} \leq P} W \mathbb{E} \left[ \log \det \left( \mathbf{I} + \frac{\mathbf{H} \mathbf{Q} \mathbf{H}^H}{W} \right) \right] \quad (81)$$

where  $\mathbf{H} = \mathbf{H}(f)|_{f=0}$  denotes the narrow-band MIMO channel matrix and  $\mathbf{Q} = \mathbf{Q}(f)/W$ . Since the wide-band ergodic capacity does not depend on channel correlation over frequency, we immediately have the following result.

**Theorem 6:** Frequency selectivity does not affect the ergodic capacity of wide-band MIMO channels.

Thus, whether we have a small or large delay spread does not affect the ergodic capacity, just as in the single-antenna case [15]. Note that some recent results suggested otherwise [25]. However, a correct interpretation of the results in [25] is con-

sistent with Theorem 6.<sup>8</sup> Recent experimental measurements support the conclusions of Theorem 6 as well.

*Remark 1:* Note that (81) is identical to the expression for narrow-band capacity (56) except for the linear capacity scaling due to bandwidth and the replacement of  $\mathbf{Q}$  by  $\mathbf{Q}/W$ . With the above connection between narrow-band and wide-band capacity, most of the narrow-band results directly carry over. In particular, the  $D$ -connected model for the narrow-band MIMO channel can be used in the wide-band case as well since the ergodic capacity is governed by spatial statistics of the narrow-band matrix  $\mathbf{H}$ . Thus, in all the following results, we assume a  $D$ -connected spatial structure for  $\mathbf{H}$ .

**Theorem 7—General Wide-Band Capacity Bounds:** For any fixed  $D$  and  $W$ , the wide-band ergodic capacity can be bounded as

$$W \mathbb{E} \left[ \log \det \left( \mathbf{I}_N + \frac{P}{NW} \mathbf{H} \mathbf{H}^H \right) \right] \leq C_{\text{WB}} \leq NW \log \left( 1 + \frac{D}{NW} P \right). \quad (82)$$

*Proof:* The result directly follows from Lemmas 2 and 3.  $\square$

From Lemma 4, we also have the following looser but more tractable lower bound for  $N > D - 1$ :

$$C_{\text{WB}} \geq W(N - D + 1) \mathbb{E} \left[ \log \left( 1 + \frac{P}{NW} \chi_2^2 \right) \right]. \quad (83)$$

<sup>8</sup>The numerical results in [25] show an increase in ergodic capacity with increased delay spread. However, under the modeling assumptions in [25], an increase in delay spread is associated with a corresponding increase in angular spread. Thus, the increase in capacity is actually due to the increase in angular spread, which is a well-understood effect (see, e.g., [6]). If the delay spread is changed, while keeping the angular spread constant, there is no change in ergodic capacity [11], [12], as in Theorem 6.

Note that both the lower and upper bounds above increase with  $W$ .

### A. Finite Connectivity Channels

Theorem 1 holds true for the wide-band channel as well.

*Theorem 8—Asymptotic Wide-Band Capacity for Finitely Connected Channels:* Channel capacity of an  $N \times N$  wide-band MIMO channel with fixed connectivity  $D$  at a given transmit power  $P$  is asymptotically bounded between  $P$  and  $DP$  as  $N$  approaches infinity.

Theorem 2 for the low-power regime also carries over unchanged.

*Theorem 9—Wide-Band Capacity Scaling in the Low-Power Regime:* For fixed  $N$ ,  $W$ , and  $D$  ( $1 \leq D \leq N$ ), the capacity of a  $D$ -connected wide-band channel scales like  $DP$  as  $P$  becomes small. More precisely

$$\lim_{P \rightarrow 0} \frac{C_{\text{WB}}(P)}{P} = D. \quad (84)$$

Moreover, any Gaussian input which satisfies the power constraint achieves capacity asymptotically in the low-power regime. In particular,  $\mathbf{Q} = \frac{P}{N} \mathbf{I}_N$  is asymptotically optimal.

### B. Infinite Connectivity Channels

The capacity scaling result (Theorem 5) for infinitely connected narrow-band channels also carries over, except for appropriate bandwidth scaling.

*Theorem 10—Normalized Wide-Band Capacity Scaling:* For fixed transmit power  $P$  and bandwidth  $W$ , if the spatial channel connectivity grows properly with dimension  $N$ , that is, (71) is satisfied, then the wide-band capacity per spatial dimension is asymptotically bounded as

$$W \int \log \left( 1 + \frac{\gamma P \lambda}{W} \right) dF(\lambda) \leq \frac{C_{\text{WB}}(N)}{N} \leq W \log \left( 1 + \frac{\gamma P}{W} \right) \quad (85)$$

where  $F(\lambda)$  is the Marčenko–Pastur law given in (72) and  $\gamma$  is the growth ratio.

### C. Infinite Bandwidth Channels

It can be shown that as  $W$  increases, both the lower and upper bounds in Theorem 7 converge to the same limit. The proof is similar to the more direct proof provided below.

*Theorem 11—Infinite Bandwidth Capacity:* For any given  $N$  and transmit power  $P$ , the infinite bandwidth capacity of a  $D(N)$ -connected MIMO channel is equal to  $D(N)P$

$$\lim_{W \rightarrow \infty} C_{\text{WB}}(N) = \max_{\text{tr} \mathbf{Q} \leq P} D(N) \text{tr}(\mathbf{Q}) = D(N)P. \quad (86)$$

Consequently, any Gaussian input satisfying the power constraint achieves capacity.

*Proof:* It follows from (81) that

$$\lim_{W \rightarrow \infty} C_{\text{WB}}(N) = \max_{\text{tr} \mathbf{Q} \leq P} \lim_{W \rightarrow \infty} W \mathbb{E} \left[ \log \det \left( \mathbf{I} + \frac{\mathbf{H} \mathbf{Q} \mathbf{H}^H}{W} \right) \right]. \quad (87)$$

Now we have

$$W \mathbb{E} \left[ \log \det \left( \mathbf{I} + \frac{\mathbf{H} \mathbf{Q} \mathbf{H}^H}{W} \right) \right] = \sum_{i=1}^N W \mathbb{E} \left[ \log \left( 1 + \frac{\lambda_i}{W} \right) \right] \quad (88)$$

where  $\lambda_i$  are the eigenvalues of  $\mathbf{H} \mathbf{Q} \mathbf{H}^H$ . By DCT we have

$$\begin{aligned} & \lim_{W \rightarrow \infty} \sum_{i=1}^N W \mathbb{E} \left[ \log \left( 1 + \frac{\lambda_i}{W} \right) \right] \\ &= \sum_{i=1}^N \mathbb{E}[\lambda_i] = \mathbb{E} \left[ \text{tr} \left( \mathbf{H} \mathbf{Q} \mathbf{H}^H \right) \right] \\ &= \text{tr} \left( \mathbf{Q} \mathbb{E} \left[ \mathbf{H}^H \mathbf{H} \right] \right) = D(N) \text{tr}(\mathbf{Q}) \end{aligned} \quad (89)$$

where the last equality follows from Proposition 2. Combining (87) and (89) completes the proof.  $\square$

If  $D(N)$  scales linearly with  $N$  then the infinite bandwidth capacity also scales linearly with  $N$ . Moreover, in the infinite bandwidth case, the capacity also scales linearly with the transmit power or the signal-to-noise ratio (SNR). The most conspicuous effect of large bandwidth is that capacity approaches the upper bound in the finite connectivity case and the upper and lower bounds converge in the infinite connectivity case.

Fig. 7 shows the upper and lower bounds for  $C_{\text{WB}}(N)/N$  as a function of  $W$  for  $N = 6$  and  $D = 3$ . We note that both bounds increase with  $W$  and are converging to  $DP \log_2(e)/N$  as predicted by Theorem 11. We have also plotted the large  $N$  bounds for  $\gamma = 0.5$  for comparison. It is worth noting that the plots for large  $N$  are nearly identical to those for  $N = 6$ , demonstrating the relatively fast convergence with  $N$ .

## VI. SPECTRAL EFFICIENCY IN WIDE-BAND MIMO CHANNELS

The tradeoff of spectral efficiency versus energy per information bit is the key measure of channel capacity in wide-band MIMO channels. Following [16], we investigate such a fundamental tradeoff for correlated MIMO channel via the virtual channel representation. Our results reveal an intrinsic link between channel structure and characteristics of the bandwidth–power tradeoff. We shall begin with an exposition that is tailored to our setup. Readers should consult the original work [16] for an elaborated treatment of spectral efficiency in the wide-band regime.

The wide-band *additive white Gaussian noise* (AWGN) channel is perhaps the best example to illustrate the tradeoff of spectral efficiency and energy per information bit. The capacity of wide-band AWGN with bandwidth  $W$  (Hz) is given by

$$C = W \log_2 \left( 1 + \frac{P}{W} \right) \quad (90)$$

where noise is assumed to have unit variance and  $P$  is the total transmit power. Denote by  $s = \frac{P}{W}$  the energy per dimension (J/s/Hz). Normalized by total bandwidth, the spectral efficiency or capacity per dimension (b/s/Hz) is

$$C = \frac{C}{W} = \log_2(1 + s). \quad (91)$$

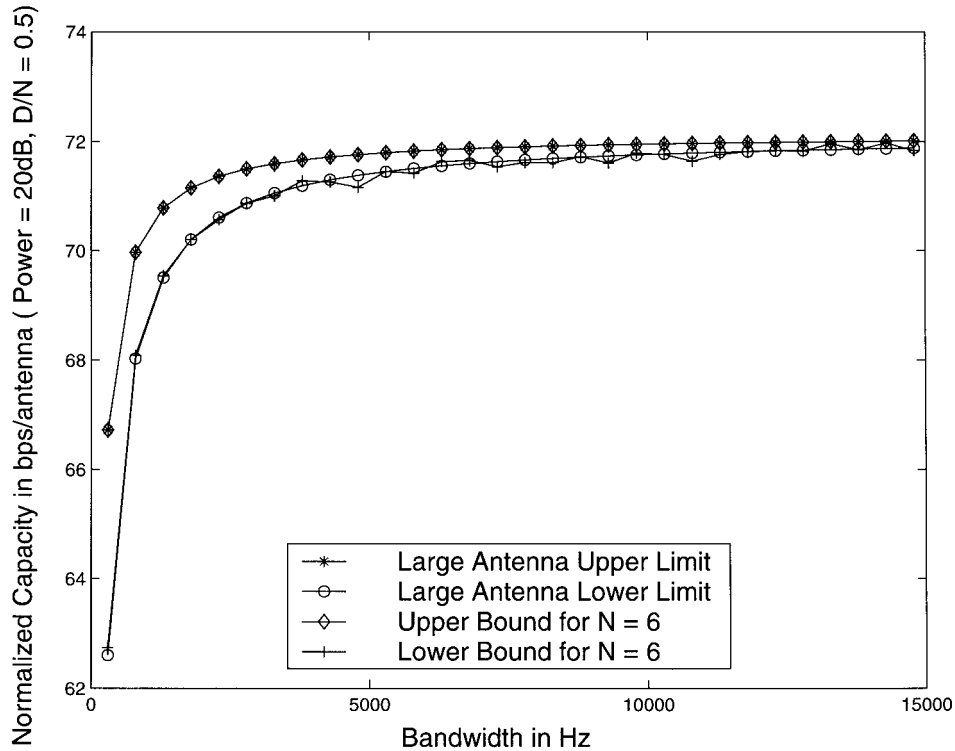


Fig. 7. Normalized wide-band capacity  $C_{\text{WB}}(N)/N$  as a function of bandwidth.  $N = 6$ ,  $D = 3$ ,  $P = 20$  dB. The large  $N$  limits correspond to  $\gamma = D(N)/N = 0.5$ .

Since the energy required to support  $C$  bits is  $s$  per dimension, the energy per information bit ( $J/b$ ) is simply given by

$$E_b = \frac{P}{W \log_2 \left(1 + \frac{P}{W}\right)} = \frac{s}{\log_2(1+s)}. \quad (92)$$

As the bandwidth  $W$  approaches infinity or, equivalently, as the energy per dimension  $s$  approaches 0, spectral efficiency converges to 0. However, the infinite bandwidth capacity is positive and

$$\begin{aligned} (E_b)_{\min} &= \lim_{W \rightarrow \infty} \frac{P}{W \log_2 \left(1 + \frac{P}{W}\right)} \\ &= \lim_{s \rightarrow 0} \frac{s}{\log_2(1+s)} = \log 2 = -1.59 \text{ dB} \end{aligned} \quad (93)$$

which is the *minimum energy per information bit* required for reliable communication [16].

Generally, let  $C(s)$  be the Shannon capacity per dimension where  $s$  is the energy per dimension. The spectral efficiency versus energy per information bit ( $C$ - $E_b$ ) tradeoff is a curve parameterized by the energy per dimension  $s$ :  $s \geq 0$  (see [16, eqs. (15) and (16)]) as

$$C = C(s) \quad (\text{b/s/Hz}) \quad (94)$$

$$E_b = \frac{s}{C(s)} = 10 \log_{10} s - 10 \log_{10} C(s) \quad (\text{dB-J/b}) \quad (95)$$

The region of  $C$ - $E_b$  curve near  $s = 0$  is of great interest for wide-band applications. As  $s$  approaches zero,  $E_b$  converges to the minimum energy per information bit  $(E_b)_{\min}$  required for reliable communications. Although spectral efficiency diminishes as bandwidth increases, its decay near  $s = 0$ , that is, the slope  $\mathcal{S}_0$  of  $C$ - $E_b$  at  $s = 0$ , is a key measure in assessing system capacity in the wide-band regime. The explicit parameterization

of the tradeoff curve with respect to  $s$  is quite convenient to compute  $(E_b)_{\min}$  and  $\mathcal{S}_0$ .

*Theorem 12*—See [16]: Assume  $\dot{C}(s)$  and  $\ddot{C}(s)$  exist on a neighborhood of 0.<sup>9</sup> Then

$$(E_b)_{\min} = \lim_{s \rightarrow 0} \frac{s}{C(s)} = \frac{1}{\dot{C}(0)} \quad (96)$$

$$\mathcal{S}_0 = \frac{dC}{dE_b} (E_b = (E_b)_{\min}) = \frac{2 [\dot{C}(0)]^2}{-\ddot{C}(0)} \frac{1}{10 \log_{10} e}. \quad (97)$$

*Proof:* We provide a proof based on (94) and (95). Apply L'Hospital's rule to get (96) as

$$(E_b)_{\min} = \lim_{s \rightarrow 0} \frac{s}{C(s)} = \frac{1}{\dot{C}(0)}.$$

Note that

$$\mathcal{S}_0 = \frac{dC}{dE_b} (E_b = (E_b)_{\min}) = \frac{\frac{dC}{ds}}{\frac{dE_b}{ds}} (s = 0). \quad (98)$$

Now applying L'Hospital's rule repeatedly on (94) and (95), one has

$$\begin{aligned} \mathcal{S}_0 &= \frac{\dot{C}(0)}{10 \log_{10} e} \lim_{s \rightarrow 0} \frac{sC(s)}{C(s) - s\dot{C}(s)} \\ &= \frac{\dot{C}(0)}{10 \log_{10} e} \lim_{s \rightarrow 0} \frac{C(s) + s\dot{C}(s)}{-s\ddot{C}(s)} = \frac{2 [\dot{C}(0)]^2}{-\ddot{C}(0)} \frac{1}{10 \log_{10} e} \end{aligned}$$

which is (97).  $\square$

<sup>9</sup>Note that the units of  $C(s)$  are in bits per dimension (as opposed to nats per dimension) which results in a constant offset compared with the corresponding formula in [16].

We next characterize the fundamental  $C-E_b$  tradeoff in wide-band MIMO channels. We shall adopt the practice of normalizing capacity by antenna dimensions as in [16], that is, the capacity unit is b/s/Hz/antenna, and correspondingly the energy per dimension is given by

$$s = \frac{P}{NW} \quad (99)$$

where  $P$  is the total transmit power,  $N$  is the antenna dimension, and  $W$  is the bandwidth. Since it is hard to find the optimal input distribution in general, we use the capacity bounds.

*Theorem 13—Minimum Energy Per Information Bit:* Given a wide-band  $D$ -connected channel of dimension  $N$ , then

$$(E_b)_{\min} = \frac{1}{D} \log 2 = -1.59 - 10 \log_{10} D \quad \text{dB} \quad (100)$$

*Proof:* It follows from Theorem 7 that

$$\mathbb{E}[\log_2(1 + s\lambda)] \leq C(s) = \frac{C_{\text{WB}}}{NW} \leq \log_2(1 + sD) \quad (101)$$

where  $\lambda$  is an unordered eigenvalue of  $\mathbf{H}\mathbf{H}^H$  with  $\mathbb{E}\lambda = D$ . Trivially

$$\lim_{s \rightarrow 0} \frac{s}{\log_2(1 + sD)} = \frac{1}{D} \log 2.$$

Similar to the proof of Theorem 2, the DCT can justify the following:

$$\begin{aligned} \lim_{s \rightarrow 0} \frac{s}{\mathbb{E}[\log_2(1 + s\lambda)]} \\ = \frac{1}{\mathbb{E}[\lim_{s \rightarrow 0} (1/s) \log_2(1 + s\lambda)]} = \frac{1}{D} \log 2. \end{aligned}$$

Then, (100) results from taking  $s \rightarrow 0$  on both sides of

$$\frac{s}{\log_2(1 + s)} \leq \frac{s}{C(s)} \leq \frac{s}{\mathbb{E}[\log_2(1 + s\lambda)]}. \quad \square$$

The wide-band slope  $\mathcal{S}_0$  concerns local behavior of  $C-E_b$  tradeoff for small values of energy per dimension  $s$ . To facilitate exposition, we normalize the input (Gaussian) distribution.

*Definition 5—Normalized Input Distributions:* Let  $\mathcal{Q}$  be the set of all  $N \times N$  semipositive-definite Hermitian matrices corresponding to the covariance matrices of  $N$ -dimensional complex Gaussian input distributions. Given any  $\mathbf{Q} \in \mathcal{Q}$ , write

$$\mathbf{Q} = \frac{\text{tr} \mathbf{Q}}{N} \tilde{\mathbf{Q}} \quad (102)$$

where  $\text{tr} \tilde{\mathbf{Q}} = N$ , that is, write any  $\mathbf{Q}$  as a product of a scalar and a *normalized* positive-definite Hermitian matrix whose trace is fixed at  $N$ . Denote by  $\tilde{\mathcal{Q}}$  the set of such matrices  $\tilde{\mathbf{Q}}$ . Note that any  $\tilde{\mathbf{Q}} \in \tilde{\mathcal{Q}}$  has a fixed trace equal to  $N$ .

Using the notion of normalized input distributions, we can write the the power constraint as

$$\{\mathbf{Q}: \text{tr} \mathbf{Q} \leq P, \mathbf{Q} \in \mathcal{Q}\} = \left\{ \frac{\tilde{P}}{N} \tilde{\mathbf{Q}}: \tilde{P} \leq P, \tilde{\mathbf{Q}} \in \tilde{\mathcal{Q}} \right\} \quad (103)$$

where  $P$  is the total transmit power. Then, the channel capacity formula can be rewritten as

$$\begin{aligned} C_{\text{WB}} &= W \max_{\text{tr}(\mathbf{Q}) \leq P} \mathbb{E} \left[ \log \det \left( \mathbf{I} + \frac{1}{W} \mathbf{H}\mathbf{Q}\mathbf{H}^H \right) \right] \\ &= W \max_{\tilde{P} \leq P, \tilde{\mathbf{Q}}} \mathbb{E} \left[ \log \det \left( \mathbf{I} + \frac{\tilde{P}}{NW} \mathbf{H}\tilde{\mathbf{Q}}\mathbf{H}^H \right) \right] \\ &= NW \max_{\tilde{P} \leq P, \tilde{\mathbf{Q}}} \mathbb{E} \left[ \log \left( 1 + \frac{\tilde{P}}{NW} \tilde{\lambda} \right) \right] \\ &= NW \max_{\tilde{\mathbf{Q}}} \mathbb{E} \left[ \log \left( 1 + \frac{P}{NW} \tilde{\lambda} \right) \right] \quad (104) \end{aligned}$$

where the maximization is over all normalized input distributions and  $\tilde{\lambda}$  is the unordered eigenvalue of  $\mathbf{H}\tilde{\mathbf{Q}}\mathbf{H}^H$ . Therefore, the optimal  $C-E_b$  tradeoff curve is given by (94) and (95) with

$$C(s) = \max_{\tilde{\mathbf{Q}}} \mathbb{E} \left[ \log_2 \left( 1 + s\tilde{\lambda} \right) \right]. \quad (105)$$

The proof of Theorem 13 shows that uniform power distribution, that is,  $\tilde{\mathbf{Q}} = \mathbf{I}$ , asymptotically achieves  $(E_b)_{\min}$  and is thus *first-order optimal in the wide-band regime* in the terminology of [16]. Actually, the result can be strengthened, following a similar argument as in Theorem 2, so that any normalized  $\tilde{\mathbf{Q}}$  is first-order optimal. However, different signaling schemes may result in different wide-band slopes  $\mathcal{S}_0$ . The task of finding the maximal  $\mathcal{S}_0$  is complicated by the maximization in (105). Little is known about the optimal  $\tilde{\mathbf{Q}}$  for general  $D$ -connected channels. Furthermore, it is possible that the optimal  $\tilde{\mathbf{Q}}$  may depend on  $s$  as well. However, we have the following result whose proof relies on Lemma 1.

*Theorem 14—Maximal Wide-Band Slope:* Suppose that the normalized input distribution is kept unchanged as energy per dimension scales. Then, uniform input distribution, that is,  $\tilde{\mathbf{Q}} = \mathbf{I}$ , gives the maximal wide-band slope for a given wide-band  $D$ -connected channel of dimension  $N$  and its corresponding slope is

$$\begin{aligned} \mathcal{S}_0 &= \frac{1}{10 \log_{10} 2} \quad (\text{b/s/Hz/dB/antenna}) \\ &= 1 \text{ b/s/Hz/3 dB/antenna} \quad (106) \end{aligned}$$

which is independent of connectivity of the channel.

*Proof:* See Appendix V.  $\square$

*Remark 2:* It is consistent to see that connectivity has no effect on the maximal wide-band slope. If  $D = 1$ , the  $D$ -connected channel is essentially a parallel of Rayleigh-fading channels. Its slope is given via [16, Theorem 13] by setting  $m = n = 1$ , which is 1 b/s/Hz/3-dB/antenna. If  $D = N$ , the same theorem, specialized by  $m = n = N$ , says that the maximal slope is again 1 b/s/Hz/3 dB/antenna.

Fig. 8 illustrates the tradeoff between spectral efficiency and energy per information bit in the wide-band regime for  $D$ -connected channels. The  $C-E_b$  curves for different normalized input distributions are plotted. The channel dimension  $N$  is 10 and the connectivity  $D$  is 5 in the simulation. As evident from the figure, all tradeoff curves approach the same  $(E_b)_{\min}$  as

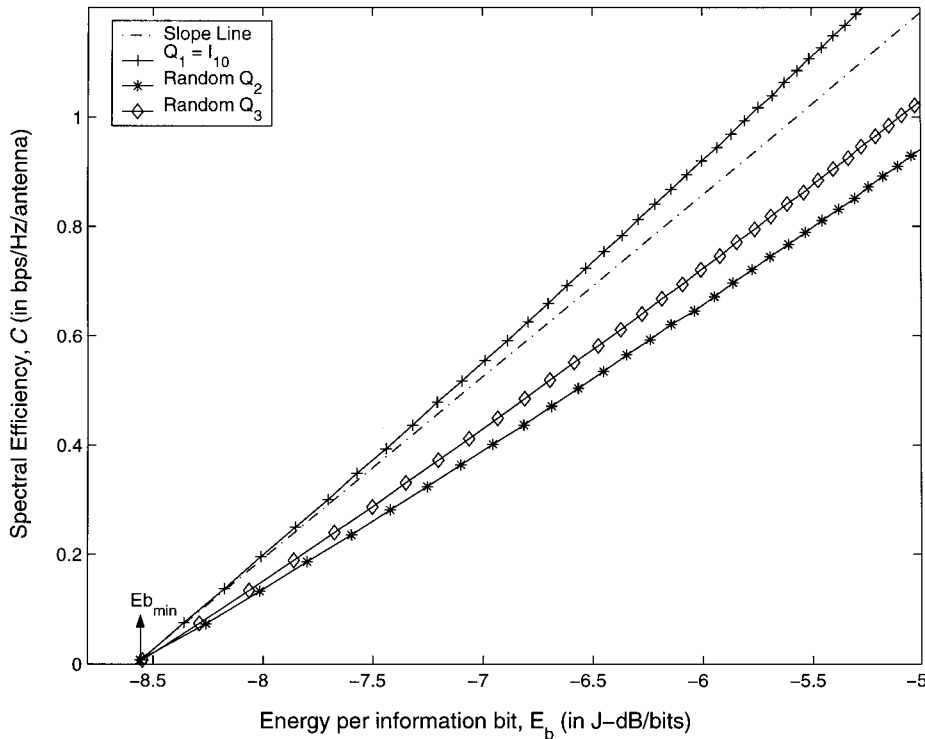


Fig. 8. Spectral efficiency as a function of energy per information bit in the wide-band regime for a 5-connected channel of dimension 10.

determined by Theorem 13, but the uniform distribution gives the best slope  $S_0$  as in Theorem 14.

## VII. PHYSICAL INTERPRETATION OF SCALING RESULTS

In this section, we provide a physical interpretation of the scaling results, particularly in terms of the number and spatial distribution of propagation paths. The most important scaling results are due to Theorems 1 and 5. For infinite connectivity, with a nonvanishing growth ratio  $\gamma \in (0, 1]$ , capacity scales linearly with  $N$  for large  $N$  with the slope between the bounds derived in the Theorem 5. For large  $N$ ,  $D(N) \approx \gamma N$  and thus connectivity must also scale linearly with  $N$  to sustain linear capacity growth. Since  $\gamma = 1$  corresponds to a fully populated i.i.d. channel, the bounds in Theorem 5 show that the only effect of  $\gamma < 1$  is to reduce the effective asymptotic received SNR (slope of capacity growth). This is consistent with the interpretation that  $\gamma = D(N)/N$  reflects the fraction of virtual receive angle that couple with each virtual transmit angle and *vice versa*. However, for fixed connectivity  $D$ , the capacity saturates to a value between  $D$  and  $DP$  because scattering is not rich enough for  $D$  to scale with  $N$ .

Recall from (46) that we need on the order of  $N_{\text{path}} = D(N)N$  resolvable paths to populate a  $D$ -connected channel with high probability. Thus, the number of paths must scale as  $N_{\text{path}} \sim \gamma N^2$  (quadratically) with  $N$  to support linear growth in capacity (and connectivity). On the other hand, the number of paths must scale as  $N_{\text{path}} \sim DN$  (linearly) with  $N$  to support a finite connectivity ( $D$ ) and nonvanishing but finite asymptotic capacity. In addition to the growth in the number of resolvable paths, the spatial distribution of the paths is also critical from a scaling viewpoint. To see this, it

is instructive to interpret the scaling results in terms *channel power per dimension* ( $\sigma_{\text{dim}}^2(N)$ ). Under the assumption that the power per path is constant ( $\sigma_n^2 = \sigma^2$ ), the total channel power is given by

$$\sigma_H^2 = \mathbb{E} \text{tr}(\mathbf{H}\mathbf{H}^H) = N_{\text{path}}\sigma^2$$

and from Proposition 2,  $\sigma_H^2 = \sigma_v^2 D(N)N$  for a  $D$ -connected channel, where  $\sigma_v^2$  is the power in each nonvanishing virtual coefficient ( $\sigma_v^2 = 1$  in the analysis). Then, we have

$$\sigma_{\text{dim}}^2(N) = \sigma_H^2/N = N_{\text{path}}\sigma^2/N = \sigma_v^2 D(N) \quad (107)$$

and the bounds in Theorem 5 show that for large  $N$  the capacity is given by

$$C(N) \approx N \log(1 + P\sigma_{\text{dim}}^2(N)/N) = N \log(1 + \text{SNR}(N)) \quad (108)$$

where  $N$  reflects the number of parallel channels and  $\text{SNR}(N) = P\sigma_{\text{dim}}^2(N)/N$  is the received SNR per parallel channel (dimension). For an infinite connectivity channel,  $\sigma_{\text{dim}}^2(N)$  increases linearly with  $N$  (from (107)) and  $\text{SNR}(N)$  remains constant, leading to linear capacity scaling (from (108)). On the other hand, for fixed connectivity  $D$ ,  $\sigma_{\text{dim}}^2(N)$  remains constant and  $\text{SNR}(N) = P\sigma_v^2 D/N$ , leading to capacity saturation.

The preceding discussion leads to a general and intuitive interpretation of the number and spatial distribution of paths required for capacity scaling. For an  $N$ -dimensional channel,  $\mathcal{O}(N)$  resolvable paths, uniformly distributed in the  $N$  diagonal spatial bins of  $\mathbf{H}$  (see Fig. 2), are sufficient to create  $N$  parallel channels. In order to keep  $\text{SNR}(N)$  constant, we need



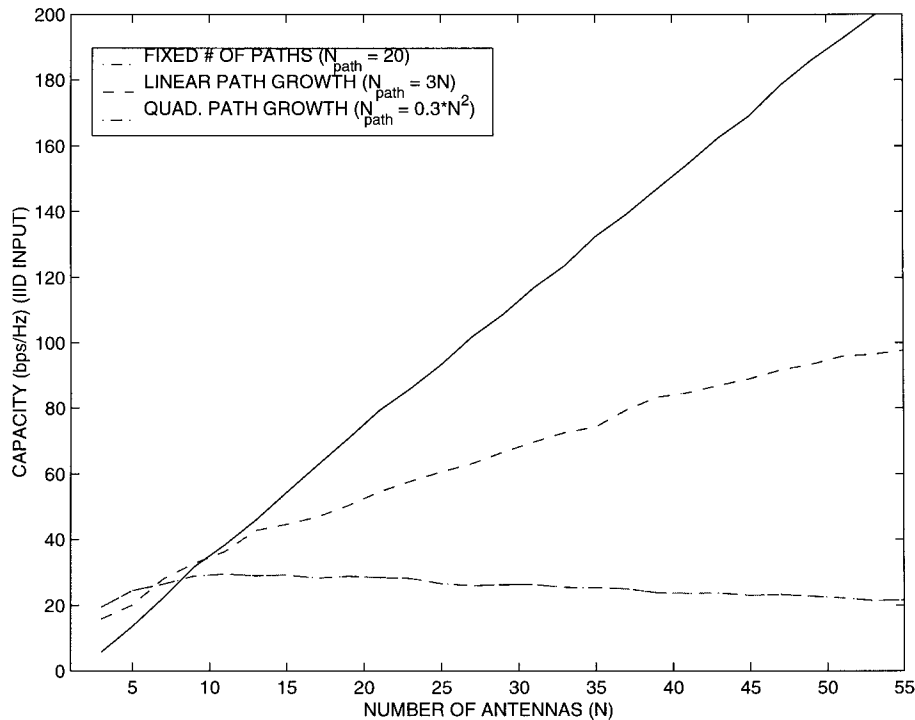


Fig. 9. Narrow-band capacity scaling versus the number of antennas for a channel simulated via the physical model. Three different physical scenarios are shown: fixed number of paths ( $N_{\text{path}} = 20$ ), linear growth in the number of paths ( $N_{\text{path}} = 3N$ ; finite connectivity  $D = 3$ ), and quadratic growth in the number of paths ( $N_{\text{path}} = 0.30 * N^2$ ; infinite connectivity  $\gamma = 0.3$ ).

additional  $\mathcal{O}(N)$  resolvable paths per parallel channel (diagonal element of  $\mathbf{H}$ ) to couple each virtual receive (transmit) angle with  $D(N) \approx \gamma N$  transmit (receive) angles, so that  $\sigma_{\text{dim}}^2(N)$  also increases linearly with  $N$ . Thus,  $\mathcal{O}(N^2)$  paths are needed overall and they should be distributed in a physical scattering environment with a *nonvanishing conditional angular spread* (associated with each transmit or receive angle) to yield  $D(N) = \gamma N$ . The scattering environment in (44) (corresponding to a  $D$ -connected model) with

$$R_{\text{AS}} = T_{\text{AS}} = [-0.5, 0.5]$$

and

$$R_{\text{AS}}(\theta_T) = [\theta_T - \delta\theta, \theta_T + \delta\theta], \quad \delta\theta > 0$$

would suffice ( $\gamma = 2\delta\theta$ ). However, a more realistic scattering environment with nonuniform but nonvanishing conditional angular spreads ( $R_{\text{AS}}(\theta_T) = [\theta_T - \delta\theta(\theta_T), \theta_T + \delta\theta(\theta_T)]$ ,  $\inf_{\theta_T} \delta\theta(\theta_T) > 0$ ) would also suffice. A finite connectivity channel essentially corresponds to a “diagonal” scattering environment with a *vanishing* conditional angular spread ( $\delta\theta = 0$  in (44)); it can be populated with  $\mathcal{O}(N)$  paths to yield linear scaling in the number of parallel channels, but no matter how many paths populate it, it cannot sustain a constant  $\text{SNR}(N)$  since  $D$  does not scale.<sup>10</sup> For a given fixed number of paths  $N_{\text{path}}$  (fixed  $\sigma_H^2$ ), we expect the capacity to increase linearly with  $N$  up to  $N \approx \sqrt{N_{\text{path}}}$ , saturate to a maximum around  $N \approx N_{\text{path}}$ , and then go to zero if we increase  $N$  beyond

that. Thus,  $N \approx N_{\text{path}}$  represents the optimal number of antennas—distributing power over more antennas is inefficient.

Figs. 9 and 10 illustrate narrow-band capacity scaling with  $N$  in a channel simulated via the physical model (6). Three different scattering environments are simulated: i) fixed number of propagation paths ( $N_{\text{path}} = 20$ ), ii) linear growth in the number of paths ( $N_{\text{path}} = DN = 3N$ ) depicting finite connectivity, and iii) quadratic growth in the number of paths ( $N_{\text{path}} = \gamma N^2 = 0.30N^2$ ) depicting infinite connectivity. The SNR is  $\log_{10}(P) = 20$  dB and the power per path  $\sigma^2 = 1$  in all cases. The channel was simulated according to a uniform conditional angular density of the form (44) with  $\delta\theta = 0.15$ ; the paths angles were uniformly distributed over the support of the scattering function:

$$\{(\theta_R, \theta_T) : -0.5 \leq \theta_T \leq 0.5, \theta_T - 0.15 \leq \theta_R \leq \theta_T + 0.15\}.$$

Channel capacity was approximated with the mutual information for uniform input power distribution which corresponds to the lower bounds on capacity. As evident from Fig. 9, capacity is converging to zero for environment i), exhibiting saturation for environment ii), and is showing linear growth for environment iii). Note that the lower bound in Theorem 1 for ii) is 144 b/s/Hz. Fig. 10 plots the ratio  $C(N)/N$  for the three environments. As expected, the growth ratio is converging to zero for both i) and ii), whereas it is stabilizing to a value near 4 b/s/Hz/antenna for iii). Furthermore, the capacity growth rate in iii) closely agrees with the lower bound in (5) which yields a value of about 4 for  $\gamma = 0.3$ . It is also worth noting that the growth rate has stabilized to its asymptotic value around  $N = 20$  corresponding to  $N_{\text{path}} = 0.3 * N^2 = 120$  paths.

<sup>10</sup>This is related to the fact that  $M(\theta_R, \theta_T)$  is concentrated on a one-dimensional curve in diagonal scattering.

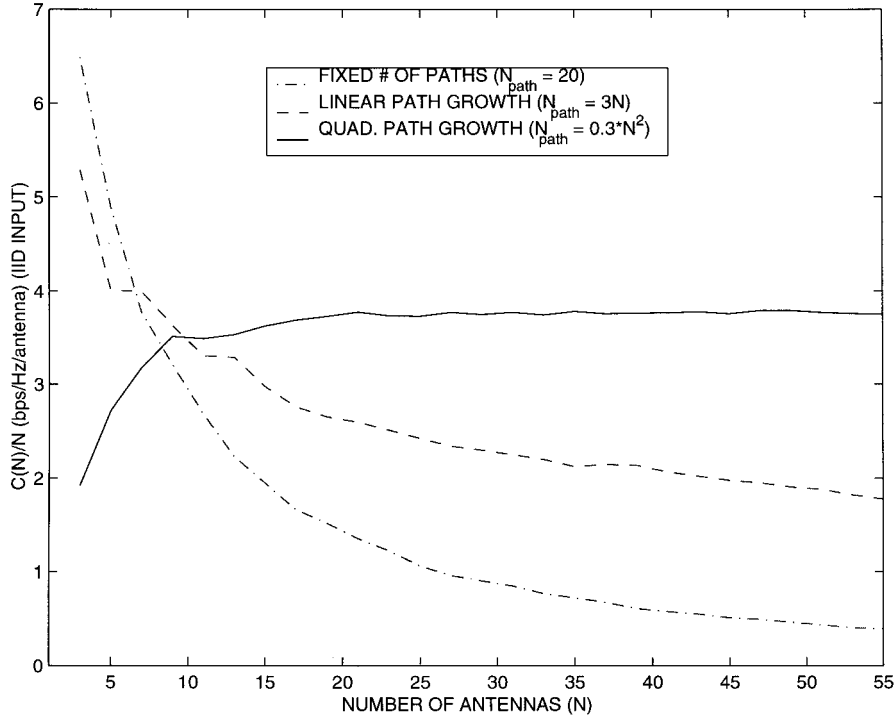


Fig. 10. Plots of  $C(N)/N$  as a function of  $N$  for a channel simulated via the physical model. Three different scenarios are shown: fixed number of paths ( $N_{\text{path}} = 20$ ), linear growth in the number of paths ( $N_{\text{path}} = 3N$ ; finite connectivity  $D = 3$ ), and quadratic growth in the number of paths ( $N_{\text{path}} = 0.3N^2$ ; infinite connectivity  $\gamma = 0.3$ ).

We now briefly relate our results to the scaling results reported by Chuah *et al.* in [13]. The channel model assumed in [13] is one of the product type, that is,

$$\mathbf{H}_c = \Psi_R^{1/2} \mathbf{W} \Psi_T^{1/2} \quad (109)$$

where  $\mathbf{W}$  has i.i.d. zero-mean complex circular Gaussian entries with variance 1. The matrices  $\Psi_R$  and  $\Psi_T$  represent the spatial correlation at the receiver and the transmitter, respectively, and are assumed to possess a Toeplitz structure consistent with the stationary spatial statistics for ULAs identified in Section II-D. It is well known that Toeplitz matrices are diagonalized by DFT matrices asymptotically [26]. Thus, for large  $N$

$$\Psi_R \approx \mathbf{A}_R \mathbf{D}_R \mathbf{A}_R^H \quad \text{and} \quad \Psi_T \approx \mathbf{A}_T \mathbf{D}_T \mathbf{A}_T^H \quad (110)$$

where  $\mathbf{D}_R$  and  $\mathbf{D}_T$  are diagonal matrices consisting of the non-negative eigenvalues of  $\Psi_R$  and  $\Psi_T$ . Substituting (110) in (109) we get

$$\mathbf{H}_c = \mathbf{A}_R \mathbf{D}_R^{1/2} \tilde{\mathbf{W}} \mathbf{D}_T^{1/2} \mathbf{A}_T^H \quad (111)$$

where  $\tilde{\mathbf{W}} = \mathbf{A}_R^H \mathbf{W} \mathbf{A}_T$  is also an i.i.d. matrix since  $\mathbf{A}_R$  and  $\mathbf{A}_T$  are unitary. Comparing (11) with (22), we can identify

$$\mathbf{H} = \mathbf{D}_R^{1/2} \tilde{\mathbf{W}} \mathbf{D}_T^{1/2} \quad (112)$$

as the narrow-band virtual channel matrix corresponding to the model (109) used in [13]. However, the above matrix is a special case of the class of virtual matrices which consists of all matrices with independent Gaussian entries with arbitrary variances. Furthermore, unlike the  $D$ -connected channel model, the conditional and marginal angular spreads/bandwidths are always the same in product correlation models, as evident from (112). As we argued above, the ratio of conditional-to-marginal

angular spreads is a key determinant of whether linear capacity scaling is possible or not. The product models will always predict linear scaling, as in [13].

## VIII. CONCLUSION

We have investigated capacity scaling and spectral efficiency in wide-band correlated MIMO channels using the virtual (Fourier) channel representation that provides an analytical framework for relating characteristics of physical (ray tracing) models to channel statistics and capacity. In particular, for ULAs, the virtual channel coefficients sample the physical scattering environment, are approximately uncorrelated regardless of the correlation exhibited by the physical channel, and characterize the degrees of freedom in correlated channels (which are fewer than i.i.d. channels). The key construct behind our analysis is a  $D$ -connected model for the virtual channel matrix that was motivated via physical considerations and provides a meaningful and tractable measure of the richness of scattering. Our scaling results show that linear capacity growth with the number of antennas ( $N$ ) is possible if the number of resolvable paths ( $N_{\text{path}}$ ) grows quadratically with  $N$  to sustain a rich scattering environment. For linear growth in  $N_{\text{path}}$ , the capacity eventually saturates. For a finite but large  $N_{\text{path}}$ , we expect the following approximate behavior: i) linear growth for  $1 \leq N \leq \sqrt{N_{\text{path}}}$ , ii) saturation between  $\sqrt{N_{\text{path}}} \leq N \leq N_{\text{path}}$ , and decay to zero for  $N > N_{\text{path}}$ .

We showed that frequency selectivity does not affect the ergodic capacity of wide-band channels. Thus, the wide-band capacity is essentially governed by the spatial structure of the narrow-band MIMO channel. In particular, the infinite bandwidth capacity scales linearly with transmit power. We studied

the spectral efficiency of wide-band  $D$ -connected MIMO channels and provided explicit characterizations of the minimum energy per bit and the wide-band slope.

We emphasize that the  $D$ -connected channel is a model for correlated channels based on two assumptions: i) spatial scattering function has a banded support (see (44)), and ii) uniform spatial power distribution.<sup>11</sup> However, it encompasses many existing models, including the product correlation model that has been used in several analytical studies. Furthermore, as argued in Section VII, it captures the essence of scaling in more general (and more realistic) scattering environments in which the richness of coupling between transmit and receive spatial dimensions scales appropriately with  $N$ . Currently, we are investigating scaling behavior under less stringent assumptions on the spatial scattering function.

In closing, we believe that the simple and intuitively appealing interpretation of physical scattering afforded by the virtual representation can be fruitfully exploited in many other aspects, including space-time code design [27], [28], channel estimation [29], and channel simulation.

#### APPENDIX I ALTERNATIVE PROOF OF LEMMA 2

*Proof:* Starting with arbitrary  $\mathbf{Q}$ , the eigen-decomposition of  $\mathbf{Q}$  is given by

$$\mathbf{Q} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \quad (113)$$

where  $\mathbf{U}$  is a unitary matrix and  $\mathbf{\Lambda}$  is a diagonal matrix with nonnegative diagonal entries. Denoting  $\mathbf{A} = \mathbf{H}\mathbf{U} = (a_{ij})$ , one has

$$\begin{aligned} \det(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^H) &= \det(\mathbf{I} + \mathbf{A}\mathbf{\Lambda}\mathbf{A}^H) \\ &\leq \prod_{i=1}^N \left( 1 + \sum_{j=1}^N \lambda_j |a_{ij}|^2 \right) \end{aligned} \quad (114)$$

where the last step follows from Hadamard's inequality [30] and  $\lambda_j$  is the  $j$ th diagonal entry in  $\mathbf{\Lambda}$ . Then it follows from Jensen's inequality that

$$\begin{aligned} \mathbb{E} \left[ \log \det(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^H) \right] &\leq \sum_{i=1}^N \log \left( 1 + \sum_{j=1}^N \lambda_j \mathbb{E}[|a_{ij}|^2] \right) \\ &= \sum_{i=1}^N \log \left( 1 + \sum_{j=1}^N \lambda_j b_{ij} \right) \end{aligned} \quad (115)$$

where  $\mathbf{B} = (\mathbb{E}[|a_{ij}|^2])_{N \times N} = (b_{ij})$ . Then  $\mathbf{B}$  is a scaled doubly stochastic matrix with scale  $D$  by Lemma 1. The general upper bound then follows from the following nonlinear programming:

$$\max \left\{ \sum_{i=1}^N \log \left( 1 + \sum_{j=1}^N \lambda_j b_{ij} \right) : \sum_{j=1}^N \lambda_j \leq P, \lambda_j \geq 0 \quad \forall 1 \leq j \leq N \right\}. \quad (116)$$

<sup>11</sup>We have recently begun experimental studies in collaboration with Prof. Ernst Bonek of FTW, Vienna, for experimental validation of the model.

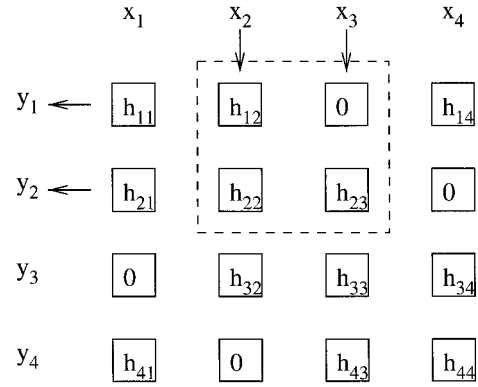


Fig. 11. A schematic illustrating Rayleigh subchannel construction from a four-dimensional 3-connected channel.  $x_i$ 's and  $y_i$ 's are input and output signals, respectively. The arrow indicates actual signals involved in construction. A two-dimensional subsystem is denoted by a dotted frame in the figure.

It is easy to see that the objective function is concave and the constraint domain is convex. The Lagrangian of the program is given by

$$\phi(\lambda, \mu) = \sum_{i=1}^N \log \left( 1 + \sum_{j=1}^N \lambda_j b_{ij} \right) - \mu \left( \sum_{j=1}^N \lambda_j - P \right). \quad (117)$$

Setting partial derivatives to zero we get

$$\begin{aligned} \sum_{i=1}^N \frac{b_{ij}}{1 + \sum_{j=1}^N \lambda_j b_{ij}} - \mu &= 0, \quad \forall 1 \leq j \leq N \\ \sum_{j=1}^N \lambda_j - P &= 0. \end{aligned} \quad (118)$$

Observe  $\lambda_1 = \dots = \lambda_N = P/N$  is a solution to (118). It is also easy to check that this solution together with associated  $\mu$  satisfies the Kuhn-Tucker condition [31], and thus achieves the maximum which turns out to be  $N \log(1 + \frac{D}{N}P)$ .  $\square$

#### APPENDIX II PROOF OF LEMMA 4

*Proof:* We illustrate the idea by an example shown in Fig. 11 where  $D = 3$  and  $N = 4$ . If the information is only transmitted at the second and third transmit antenna, the received signal at the first and second receive antenna can be written as

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} h_{12} & 0 \\ h_{22} & h_{23} \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}. \quad (119)$$

Note that the effective two-dimensional MIMO channel matrix is a lower triangular matrix with entries in the main diagonal being complex Gaussian distributed. Similar to BLAST-type processing [32], [33], successive decoding and interference cancellation can be used to construct two parallel Rayleigh-fading subchannels. More specifically, the first subchannel corresponding to the first received antenna has the following channel equation:

$$y_1 = h_{12}x_2 + z_1.$$

Assuming that the signal has been correctly decoded (assuming capacity-achieving codes are used), its interference toward the second receive antenna can be removed as

$$\tilde{y}_2 = y_2 - h_{22}x_2 = h_{23}x_3 + z_2,$$

which gives rise to the second Rayleigh-fading channel associated with the second receive antenna. Note that this method of constructing one-dimensional subchannels has been used in many works to analyze system capacity (see, e.g., [1]).

Generally, consider an  $N$ -dimensional  $D$ -connected channel. Without loss of generality, one can assume  $D$  to be an odd integer. A transmit antenna with index  $j$  is allowed to transmit if

$$\frac{D+1}{2} \leq j \leq N - \frac{D-1}{2}.$$

Hence, the number of effective transmit antennas is  $N - D + 1$ . After collecting signals from the  $i$ th receive antenna with  $1 \leq i \leq N - D + 1$ , the effective system has a lower triangular channel matrix with dimension  $N - D + 1$ . Similar to the successive decoding and interference cancellation method elaborated above, a total of  $N - D + 1$  parallel Rayleigh subchannels can be formed by successive interference cancellation. The processing begins with the decoding of information sent by the  $(D + 1)/2$ th transmit antenna at the first receive antenna. The corresponding first Rayleigh subchannel is given by

$$y_1 = h_{1, \frac{D+1}{2}} x_{\frac{D+1}{2}} + z_1.$$

Assuming perfect decoding, the interference of the  $(D + 1)/2$ th transmit antenna toward the next transmit antenna can be subtracted as

$$\tilde{y}_2 = y_2 - h_{2, \frac{D+1}{2}} x_{\frac{D+1}{2}} = h_{2, \frac{D+3}{2}} x_{\frac{D+3}{2}} + z_2$$

which gives rise to the second Rayleigh subchannel. This procedure continues until the last receive antenna  $y_{N-D+1}$  has been processed. Therefore, the mutual information of the  $N - D + 1$  parallel Rayleigh subchannels is

$$(N - D + 1) \mathbb{E} \left[ \log \left( 1 + \frac{P}{N} \chi_2^2 \right) \right]$$

which provides the desired lower bound for channel capacity.  $\square$

### APPENDIX III PROOF OF THEOREM 3

We need some results from real analysis and probability theory.

*Lemma 5—Fatou's Lemma, See [23]:* If  $f_n \geq 0$  then

$$\liminf_{n \rightarrow \infty} \int f_n d\mu \geq \int \left( \liminf_{n \rightarrow \infty} f_n \right) d\mu. \quad (120)$$

*Definition 6—Weak Convergence of Distributions:* A sequence of distribution functions  $\{F_n\}$  is said to converge weakly to a limit  $F$  (written  $F_n \Rightarrow F$ ) if  $F_n(\lambda) \rightarrow F(\lambda)$  for all  $\lambda$  that are continuity points of  $F$ .

*Theorem 15—(See [21]):* If  $F_n \Rightarrow F_\infty$ , then there are random variables  $Y_n$ ,  $1 \leq n \leq \infty$ , with distribution  $F_n$  so that  $Y_n \rightarrow Y_\infty$  a.s.

*Proposition 3:* Let  $\{f_n \geq 0\}$  be equicontinuous and  $f_n \rightarrow f$  pointwise. If  $F_n \Rightarrow F$  then

$$\liminf_{n \rightarrow \infty} \int f_n(\lambda) dF_n(\lambda) \geq \int f(\lambda) dF(\lambda). \quad (121)$$

*Proof:* By Theorem 15, there exist random variables  $Y_n$  and  $Y$  in some probability space  $(\Omega, \mathcal{B}, \nu)$ , with distribution  $F_n$  and  $F$ , respectively, and  $Y_n \rightarrow Y$  a.s. It is fairly straightforward to see that

$$f_n(Y_n(\omega)) \rightarrow f(Y(\omega)) \quad (122)$$

at all  $\omega$  for which  $Y_n(\omega) \rightarrow Y(\omega)$  by virtue of the equicontinuity of  $\{f_n\}$ . Then

$$\begin{aligned} \liminf_{n \rightarrow \infty} \int f_n(\lambda) dF_n(\lambda) &= \liminf_{n \rightarrow \infty} \int f_n(Y_n(\omega)) d\nu(\omega) \\ &\geq \int \left( \liminf_{n \rightarrow \infty} f_n(Y_n(\omega)) \right) d\nu(\omega) \\ &= \int f(Y(\omega)) d\nu(\omega) = \int f(\lambda) dF(\lambda) \end{aligned} \quad (123)$$

where we have used Fatou's lemma in passing  $\liminf$  inside the integral.  $\square$

We now give a proof of Theorem 3.

*Proof:* Denote by  $(\Omega, \mathcal{B}, \nu)$  the probability space of random ESD  $\{F_N\}$ . Let

$$\mathcal{J} = \{\omega \in \Omega: F_N^\omega(\lambda) \text{ converges to } F(\lambda)\}.$$

By hypothesis,  $\Pr(\mathcal{J}) = 1$ . For all  $\omega \in \mathcal{J}$ , it is obvious that  $F_N^\omega \Rightarrow F$  and, hence, by Proposition 3

$$\liminf_{N \rightarrow \infty} \int f_N(\lambda) dF_N^\omega(\lambda) \geq \int f(\lambda) dF(\lambda). \quad (124)$$

Now we apply Fatou's lemma to get

$$\begin{aligned} \liminf_{N \rightarrow \infty} S(N) &= \liminf_{N \rightarrow \infty} \int \left( \int f_N(\lambda) dF_N^\omega(\lambda) \right) d\nu(\omega) \\ &\geq \int \left( \liminf_{N \rightarrow \infty} \int f_N(\lambda) dF_N^\omega(\lambda) \right) d\nu(\omega) \\ &\geq \int f(\lambda) dF(\lambda) \end{aligned} \quad (125)$$

which completes the proof.  $\square$

### APPENDIX IV PROOF OF THEOREM 4

*Proof:* We adapt Grenander and Silverstein's proof. Please refer to the original work [14] for notations. In our case,  $d = 1$  and  $P_{ij} = 1$ , that is, no random connectivity and, hence,  $C = D$  in our notation. We examine the LESD of  $\mathbf{W} = \frac{1}{C} \mathbf{V} \mathbf{V}^H$ . Reference [14, eq. (2, 2)] becomes

$$\mathbb{E} \left[ \frac{1}{N} \text{tr} \mathbf{W}^r \right] = \frac{1}{N C^r} \sum_{\substack{i_1 \dots i_r \\ k_1 \dots k_r}} \mathbb{E} [v_{i_1 k_1} v_{i_2 k_1}^* \dots v_{i_r k_r} v_{i_1 k_r}^*]. \quad (126)$$

The proof would be the same if one could show [14, Lemma 1] holds for the complex case. The key is to break the sum in (126) into two parts: one part contributes in the limit and the

other does not. Since all the moments of a complex Gaussian random variable are finite, the noncontributing terms diminish. Hence, only terms that exactly pair up  $v_{ik}$ 's are relevant, which is the content of Lemma 1. A similar argument can be used to show  $\text{var}((1/N)\text{tr } \mathbf{W}^r) \rightarrow 0$ . Therefore, the desired conclusion holds.  $\square$

APPENDIX V  
PROOF OF THEOREM 14

*Proof:* Since  $\tilde{\mathbf{Q}}$  is fixed during  $s$  scaling, the capacity (mutual information) per dimension is given by

$$C(s) = \mathbb{E} \left[ \log_2 \left( 1 + s\tilde{\lambda} \right) \right] \quad (127)$$

where  $\tilde{\lambda}$  is an unordered eigenvalue of  $\mathbf{H}\tilde{\mathbf{Q}}\mathbf{H}^H$ . The DCT can justify the interchange of expectation and derivatives to give

$$\dot{C}(0) = \mathbb{E} \left[ \left( \frac{d}{ds} \log_2 \left( 1 + s\tilde{\lambda} \right) \right) (s=0) \right] = \log_2 e \mathbb{E} \tilde{\lambda} \quad (128)$$

$$\ddot{C}(0) = \mathbb{E} \left[ \left( \frac{d^2}{ds^2} \log_2 \left( 1 + s\tilde{\lambda} \right) \right) (s=0) \right] = \log_2 e \mathbb{E} \tilde{\lambda}^2. \quad (129)$$

Applying Theorem 12,

$$\begin{aligned} S_0 &= \frac{2}{10 \log_{10} 2} \frac{(\mathbb{E} \tilde{\lambda})^2}{\mathbb{E} \tilde{\lambda}^2} \\ &= \frac{1}{10 \log_{10} 2} \frac{2}{N} \frac{\left( \text{tr} \left( \mathbb{E} \left[ \mathbf{H}\tilde{\mathbf{Q}}\mathbf{H}^H \right] \right) \right)^2}{\text{tr} \left( \mathbb{E} \left[ \left( \mathbf{H}\tilde{\mathbf{Q}}\mathbf{H}^H \right)^2 \right] \right)} \end{aligned} \quad (130)$$

which is [16, Theorem 13] specialized to our case. Our task is to evaluate  $S_0$  for a  $D$ -connected channel  $\mathbf{H}$ .

Let  $\tilde{\mathbf{Q}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$  be the eigen-decomposition of  $\tilde{\mathbf{Q}}$  and let  $\mathbf{A} = \mathbf{H}\mathbf{U}$ . One has

$$\mathbf{H}\tilde{\mathbf{Q}}\mathbf{H}^H = \mathbf{A}\mathbf{\Lambda}\mathbf{A}^H. \quad (131)$$

Then

$$\begin{aligned} \text{tr} \left( \mathbb{E} \left[ \mathbf{H}\tilde{\mathbf{Q}}\mathbf{H}^H \right] \right) &= \text{tr} \left( \mathbb{E} \left[ \mathbf{A}\mathbf{\Lambda}\mathbf{A}^H \right] \right) \\ &= \sum_{i=1}^N \sum_{j=1}^N \lambda_j \mathbb{E} |a_{ij}|^2 \\ &= D \sum_{j=1}^N \lambda_j = DN \end{aligned} \quad (132)$$

where we have used the scaled doubly stochastic matrix property of  $(\mathbb{E} |a_{ij}|^2)$  (Lemma 1) and the fact that

$$\sum_{j=1}^N \lambda_j = \text{tr} \tilde{\mathbf{Q}} = N. \quad (133)$$

Let  $\mathbf{B} = \mathbf{A}\mathbf{\Lambda}\mathbf{A}^H$ . We shall compute

$$\text{tr} \left( \mathbb{E} \left[ \left( \mathbf{H}\tilde{\mathbf{Q}}\mathbf{H}^H \right)^2 \right] \right) = \text{tr} \left( \mathbb{E} \left[ \mathbf{B}^2 \right] \right) = \sum_{i,j} \mathbb{E} |b_{ij}|^2 \quad (134)$$

where  $b_{ij}$  is the  $(i, j)$ th entry of  $\mathbf{B}$ . The matrix  $\mathbf{B}$  looks like

$$\begin{bmatrix} \lambda_1 |a_{11}|^2 + \cdots + \lambda_N |a_{1N}|^2 & \cdots & \\ & \vdots & \ddots \\ \lambda_1 a_{N1} a_{11}^* + \cdots + \lambda_N a_{NN} a_{1N}^* & \cdots & \\ & & \lambda_1 a_{11} a_{N1}^* + \cdots + \lambda_N a_{1N} a_{NN}^* \\ & & \vdots & \\ & & \lambda_1 |a_{N1}|^2 + \cdots + \lambda_N |a_{NN}|^2 & \end{bmatrix} \quad (135)$$

where generally

$$b_{ij} = \sum_{k=1}^N \lambda_k a_{ik} a_{jk}^*. \quad (136)$$

Note that  $\sum_{ij} \mathbb{E} |b_{ij}|^2$  is a quadratic polynomial of  $\lambda_1, \dots, \lambda_N$ . We take the first row ( $i=1$ ), for example, to illustrate the computation for the coefficients of this polynomial.

First, consider the terms like  $\lambda_k^2$  for  $1 \leq k \leq N$ . The coefficient of the  $\lambda_k^2$  term due to the first row in  $\mathbf{B}$  is

$$c_k^1 = \sum_{j=1}^N \mathbb{E} [a_{1k} a_{jk}^* a_{1k}^* a_{jk}]. \quad (137)$$

By  $\mathbf{A} = \mathbf{H}\mathbf{U}$ , different rows of  $\mathbf{A}$  are uncorrelated since entries of  $\mathbf{H}$  are uncorrelated. Thus, if  $j \neq 1$ , one has

$$\mathbb{E} [a_{1k} a_{jk}^* a_{1k}^* a_{jk}] = \mathbb{E} |a_{1k}|^2 \mathbb{E} |a_{jk}|^2.$$

Moreover, since the entries of  $\mathbf{H}$ , and, thus,  $\mathbf{A}$ , are from a *proper* complex Gaussian joint distribution, the *Gaussian moment-factoring theorem* (GMFT) [34] implies that

$$\mathbb{E} |a_{1k}|^4 = 2 \mathbb{E} |a_{1k}|^2 \mathbb{E} |a_{1k}|^2.$$

Therefore,

$$\begin{aligned} c_k^1 &= \mathbb{E} |a_{1k}|^4 + \sum_{j \neq 1} \mathbb{E} |a_{1k}|^2 \mathbb{E} |a_{jk}|^2 \\ &= \mathbb{E} |a_{1k}|^2 \mathbb{E} |a_{1k}|^2 + \mathbb{E} |a_{1k}|^2 \sum_{j=1}^N \mathbb{E} |a_{jk}|^2 \\ &= \mathbb{E} |a_{1k}|^2 \mathbb{E} |a_{1k}|^2 + D \mathbb{E} |a_{1k}|^2 \end{aligned} \quad (138)$$

where Lemma 1 is used in the last step.

Next, consider the terms like  $\lambda_r \lambda_t$  for  $1 \leq r < t \leq N$ . The coefficient due to the first row in  $\mathbf{B}$  is

$$c_{r,t}^1 = \sum_{j=1}^N \mathbb{E} [a_{1r} a_{jr}^* a_{1t}^* a_{jt}]. \quad (139)$$

Use GMFT to break the sum into two parts as

$$\begin{aligned} c_{r,t}^1 &= \sum_{j=1}^N \left( \mathbb{E} [a_{1r} a_{1t}^*] \mathbb{E} [a_{jr}^* a_{jt}] + \mathbb{E} [a_{1r} a_{jr}^*] \mathbb{E} [a_{1t}^* a_{jt}] \right) \\ &= \mathbb{E} [a_{1r} a_{1t}^*] \sum_{j=1}^N \mathbb{E} [a_{jr}^* a_{jt}] + \sum_{j=1}^N \mathbb{E} [a_{1r} a_{jr}^*] \mathbb{E} [a_{1t}^* a_{jt}]. \end{aligned} \quad (140)$$

Let  $\mathbf{u}_r$  and  $\mathbf{u}_t$  be the  $r$ th and  $t$ th column vectors of  $\mathbf{U}$ , respectively. The first sum in (140) vanishes because

$$\begin{aligned} \sum_{j=1}^N \mathbb{E}[a_{jr}^* a_{jt}] &= \mathbb{E}[\mathbf{u}_r^H \mathbf{H} \mathbf{H}^H \mathbf{u}_t] = \mathbf{u}_r^H \mathbb{E}[\mathbf{H} \mathbf{H}^H] \mathbf{u}_t \\ &= D \mathbf{u}_r^H \mathbf{u}_t = 0 \end{aligned} \quad (141)$$

where we have used Proposition 2 and  $\mathbf{u}_r^H \mathbf{u}_t = 0$  for  $r \neq t$ . Since different rows of  $\mathbf{A}$  are uncorrelated, the second term in (140) reduces to  $\mathbb{E}|a_{1r}|^2 \mathbb{E}|a_{1t}|^2$ . Therefore,

$$c_{r,t}^2 = \mathbb{E}|a_{1r}|^2 \mathbb{E}|a_{1t}|^2. \quad (142)$$

Combining (138) and (142), the polynomial due to the first row is

$$p^1(\lambda_1, \dots, \lambda_n) = \left( \sum_{j=1}^N \lambda_j \mathbb{E}|a_{1j}|^2 \right)^2 + D \sum_{j=1}^N \lambda_j^2 \mathbb{E}|a_{1j}|^2. \quad (143)$$

Similar calculation can be done for other rows. Adding all the polynomials, one has

$$\begin{aligned} \text{tr}(\mathbb{E}[\mathbf{B}^2]) &= \sum_{i=1}^N p^i(\lambda_1, \dots, \lambda_N) \\ &= \sum_{i=1}^N \left( \sum_{j=1}^N \lambda_j \mathbb{E}|a_{ij}|^2 \right)^2 + D \sum_{i=1}^N \sum_{j=1}^N \lambda_j^2 \mathbb{E}|a_{ij}|^2 \\ &= \sum_{i=1}^N \left( \sum_{j=1}^N \lambda_j \mathbb{E}|a_{ij}|^2 \right)^2 + D^2 \sum_{j=1}^N \lambda_j^2 \end{aligned} \quad (144)$$

where we again used Lemma 1.

Since  $\text{tr}(\mathbb{E}[\mathbf{H} \mathbf{Q} \mathbf{H}^H]) = DN$  is constant, it follows from (130) that maximizing  $\mathcal{S}_0$  is equivalent to minimizing  $\text{tr}(\mathbb{E}[\mathbf{B}^2])$ , which is (144) over the constraint set

$$\left\{ (\lambda_1, \dots, \lambda_N) : \sum_{j=1}^N \lambda_j = N, \quad \lambda_j \geq 0 \quad \forall 1 \leq j \leq N \right\}. \quad (145)$$

Similar to the alternative proof in Lemma 2 (Appendix I), the scaled doubly stochastic property of  $(\mathbb{E}|a_{ij}|^2)$  critically establishes that the uniform power distribution, that is,  $\lambda_1 = \dots = \lambda_N = 1$ , achieves the minimum and the corresponding minimum value is

$$\left( \text{tr}(\mathbb{E}[\mathbf{B}^2]) \right)_{\min} = 2D^2N. \quad (146)$$

Therefore, the maximal slope is

$$(\mathcal{S}_0)_{\max} = \frac{1}{10 \log_{10} 2} \frac{2D^2N^2}{N2D^2N} = \frac{1}{10 \log_{10} 2}. \quad (147)$$

□

#### ACKNOWLEDGMENT

The authors gratefully acknowledge Prof. Jack W. Silverstein of the Mathematics Department at the North Carolina State University for discussions on random matrix theory and assistance with Theorem 4.

#### REFERENCES

- [1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, pp. 311–335, 1998.
- [2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs., 1995.
- [3] G. German, Q. Spencer, L. Swindlehurst, and R. Valenzuela, "Wireless indoor channel modeling: Statistical agreement of ray tracing simulations and channel sounding measurements," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2001, pp. 2501–2504.
- [4] A. F. Molisch, M. Steinbauer, M. Toeltsch, E. Bonek, and R. S. Thomä, "Capacity of MIMO systems based on measured wireless channels," *IEEE J. Select. Areas Commun.*, vol. 20, pp. 561–569, Apr. 2002.
- [5] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502–513, Mar. 2000.
- [6] A. M. Sayeed, "Deconstructing multi-antenna fading channels," *IEEE Trans. Signal Processing*, vol. 50, pp. 2563–2579, Oct. 2002.
- [7] S. J. Fortune, D. H. Gay, B. W. Kernighan, O. Landron, R. A. Valenzuela, and M. H. Wright, "WiSE design of indoor wireless systems: Practical computation and optimization," *IEEE Comput. Sci. Eng.*, vol. 2, pp. 58–68, Mar. 1995.
- [8] G. J. Foschini and R. A. Valenzuela, "Initial estimation of communication efficiency of indoor wireless channels," *Wireless Networks*, vol. 3, no. 2, pp. 141–154, 1997.
- [9] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Mag.*, vol. 14, pp. 49–83, Nov. 1997.
- [10] J. Fuhl, A. F. Molisch, and E. Bonek, "Unified channel model for mobile radio systems with smart antennas," *IEE Proc. Radar, Sonar Navig.*, vol. 145, no. 1, pp. 32–41, Feb. 1998.
- [11] A. M. Sayeed and V. Veeravalli, "The essential degrees of freedom in space-time fading channels," in *Proc. 13th IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications (PIMRC'02)*, Lisbon, Portugal, Sept. 2002, pp. 1512–1516.
- [12] —, "A virtual representation for time- and frequency-selective MIMO channels: A bridge between physical and statistical models," manuscript. Available [Online] at <http://dune.ece.wisc.edu>; to be submitted to *IEEE Trans. Commun.*
- [13] V. V. C-N Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in mimo wireless systems under correlated fading," *IEEE Trans. Inform. Theory*, vol. 48, pp. 637–650, Mar. 2002.
- [14] U. Grenander and J. W. Silverstein, "Spectral analysis of networks with random topologies," *SIAM J. Appl. Math.*, vol. 32, no. 2, pp. 499–519, Mar. 1977.
- [15] E. Biglieri, J. Proakis, and S. Shamai(Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.
- [16] S. Verdú, "Spectral efficiency in the wide-band regime," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1319–1343, June 2002.
- [17] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.
- [18] S. A. Jafar, S. Vishwanath, and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *Proc. IEEE Int. Conf. Communications*, 2001, pp. 2266–2270.
- [19] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 48, pp. 637–650, Mar. 2001.
- [20] J. H. Kotecha and A. M. Sayeed, "On the capacity of correlated MIMO channels," Univ. Wisconsin-Madison, Tech. Rep., 2002.
- [21] R. Durrett, *Probability: Theory and Examples*, 2nd ed. Belmont, CA: Wadsworth, 1996.
- [22] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices, a review," *Statist. Sinica*, pp. 611–677, 1999.
- [23] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. New York: Wiley, 1999.
- [24] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. Inform. Theory*, vol. 34, pp. 380–388, May 1988.
- [25] H. Bolcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, pp. 225–234, Feb. 2002.
- [26] R. M. Gray, "On the asymptotic eigenvalue distribution of toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 725–730, Nov. 1972.

- [27] Z. Hong, K. Liu, R. Heath, and A. Sayeed, "Spatial multiplexing in correlated fading via the virtual channel representation," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 856–866, June 2003.
- [28] K. Liu and A. M. Sayeed, "Space-time D-block codes via the virtual MIMO channel representation," *IEEE Trans. Wireless Commun.*, also presented at Allerton 2002 Conf., to be published.
- [29] J. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Processing*, also presented at Allerton 2002 Conf., to be published.
- [30] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
- [31] O. L. Mangasarian, *Nonlinear Programming*. New York: McGraw-Hill, 1994.
- [32] 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, Autumn 1996.
- [33] K. Liu and A. M. Sayeed, "Improved layered space-time processing in multiple antenna systems," in *Proc. 39th Allerton Conf. Communication, Control and Computing*, Allerton, IL, Oct. 2001.
- [34] I. S. Reed, "On a moment theorem for complex Gaussian processes," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 194–195, Apr. 1962.