# Sub-Linear Capacity Scaling Laws for Sparse MIMO Channels

Vasanthan Raghavan* and Akbar M. Sayeed

***Abstract*— Recent attention on performance analysis of single-user multiple-input multiple-output (MIMO) systems has been on understanding the impact of the spatial correlation model on ergodic capacity. In most of these works, it is assumed that the statistical degrees of freedom (DoF) in the channel can be captured by decomposing it along a suitable eigen-basis and that the transmitter has perfect knowledge of the statistical DoF. With an increased interest in large-antenna systems in state-of-the-art technologies, these implicit channel modeling assumptions in the literature have to be revisited. In particular, multi-antenna measurements have showed that large-antenna systems are *sparse* where only a few DoF are dominant enough to contribute towards capacity. Thus, in this work, it is assumed that the transmitter can only afford to learn the dominant statistical DoF in the channel. The focus is on understanding ergodic capacity scaling laws in sparse channels. Unlike classical results, where linear capacity scaling is implicit, sparsity of MIMO channels coupled with a knowledge of only the dominant DoF is shown to result in a new paradigm of sub-linear capacity scaling that is consistent with experimental results and physical arguments. It is also shown that uniform-power signaling over all the antenna dimensions is wasteful and could result in a significant penalty over optimally adapting the antenna spacings in response to the sparsity level of the channel and transmit SNR.**

***Index Terms*— Antenna arrays, correlation, fading channels, information rates, MIMO systems, random matrix theory, reconfigurable arrays, sparse systems.**

## I. Introduction

***Background and Motivation:*** Multiple-input multiple-output (MIMO) systems that employ antenna arrays at the transmitter and receiver have emerged as a promising technology to increase the spectral efficiency of wireless communications. The intense research on MIMO systems has been inspired by seminal works due to Telatar [1] and Foschini and Gans [2] that showed a dramatic linear increase in channel capacity with the number of antennas. While these initial results were based on the *i.i.d. model* that represents a rich scattering environment and sufficiently large inter-element spacings at both ends, recent progress has been on realistic systems exhibiting correlated fading [3]. Using a Kronecker model for the spatial correlation that separates the impact of the transmit and receive arrays, Chuah *et al.* [4] showed that linear capacity scaling is realizable in correlated channels, albeit with a slope smaller than that possible in i.i.d. channels. Implicit in their work is the assumption that the transmitter can learn the marginal statistics of the channel (equivalent to the variance profile of the channel entries) perfectly.

The last few years of semiconductor and electromagnetics research has witnessed significant advances in the design of low-complexity, low-cost and miniature antennas and radio-frequency front-ends [5]. Thus, in contrast to current systems where 1-2 antennas is the norm, state-of-the-art MIMO technologies such as WiMAX, LTE-A, etc., envision $4$ to $8$ antennas [6], and short-range millimeter wave signaling in the $60$ GHz regime envisions[1] $16$ to $32$ antennas [7]–[11]. While the early theoretical attention has been mostly inspired by small-antenna systems, the dynamic of large-antenna systems brings to the fore issues that had not been studied in depth till recently.

In particular, with $N_t = 2$ transmit and $N_r = 2$ receive antennas, the Kronecker model can capture $N_t + N_r = 4$ of the $4$ marginal statistical parameters (the variances of the four entries in the channel matrix). Many measurement campaigns [12]–[17] have demonstrated the accuracy of a Kronecker fit to measured data in either the $2\times2$ or Manhattan-type urban settings. In general, as $N_t$ and/or $N_r$ increase(s), there is a mismatch between the linear increase ($N_t + N_r$) in the degrees of freedom (DoF) afforded by the Kronecker model and the quadratic increase ($N_tN_r$) in the DoF available in the true channel matrix (the variances of the entries). This is because the Kronecker model essentially enforces the joint power spectrum of the channel to be the product of the marginal power spectra at both ends, thereby producing artifact paths at the intersections of the real spectral peaks. These artifacts increase the diversity order but decrease the apparent capacity since they take away energy from all real paths that do not lie at the intersection points so that the overall power is kept constant. When larger antenna arrays are used (with improved angular resolution), the Kronecker

Manuscript received Dec. 29, 2009; accepted June 1, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Syed Ali Jafar.

This work has been supported in part by the NSF grant CCF-0431088. This paper was presented in part at the Allerton Conference on Communications, Control and Computing 2004, Allerton, IL, the IEEE Information Theory Workshop 2004, San Antonio, TX, and the IEEE International Symposium on Information Theory 2006, Seattle, WA.

The first author was with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA when this work was done. He is currently with the Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3052, Australia. The second author is with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA. Email: vasanthan_raghavan@ieee.org, akbar@engr.wisc.edu. *Corresponding author.

[1]Note that "60 GHz antennas have a smaller form factor than 1 GHz antennas, as antenna dimensions are inversely proportional to carrier frequency. Therefore, more antennas can be packed within a fixed area for an improved beamforming gain" [7]. Nevertheless, it is important to note that the sparsity assumptions of this paper may be violated in the 60 GHz regime. Further measurement studies are necessary to understand the applicability of this paper to the 60 GHz regime.

model's performance is thus significantly impaired [18]–[28].

To overcome the deficiencies of the Kronecker model, a non-separable correlation framework that models the $N_tN_r$ variances (after the channel has been decomposed along a suitable eigen-basis) has been introduced [27]–[30]. In this context, measurements [19, Figs. 9 and 11], [20]–[26], [31] as well as fundamental electromagnetic studies [32]–[39] show that not all of the $N_tN_r$ DoF in the true channel are made alike. In other words, a few entries in the true channel matrix have a large variance and capture a significant fraction of the channel power[2] whereas most of the entries are non-dominant. That is, large-MIMO systems are *sparse.*

As antenna dimensions increase, sparsity implies that not all the DoF in the channel can be learned at the transmitter with sufficient accuracy. It is often worthwhile for the transmitter to treat the non-dominant entries in the channel as noise. While we do not expect any difference in the classical linear capacity scaling law if the transmitter knew the statistics perfectly (and a significant fraction of the channel entries have a non-negligible variance), it is of interest in understanding the capacity scaling laws with statistical uncertainty.

***Contributions:***

- Towards this goal, we first develop a model for sparse MIMO systems. For this, we focus on the simpler setting where $N_t = N_r = N$. A family of channels of increasing dimension is sparse if the number of dominant DoF, $D$, satisfies[3] $D = \Theta(N^\gamma)$ for some $\gamma \in (0, 2)$ with $\gamma = 2$ corresponding to the non-sparse regime. Further, existing works are motivated by the small-antenna regime and hence, normalize the channel power as $\rho_c = \Theta\left(N^2\right)$. This assumption violates physical laws, more specifically energy conservation laws, since $\rho_c$ has to *eventually*[4] satisfy $\rho_c = \mathcal{O}(N)$. Similar justifications are also provided in [40]–[44]. With a primary focus on studying the impact of sparsity on capacity and as a means to accommodate a more realistic transition between the existing and asymptotic channel power normalizations, we assume that $\rho_c = \Theta(D) = \Theta(N^\gamma),\ \gamma \in (0, 2)$ in this work.
- With the above model, we study the rate of scaling of the average mutual information, $I_{\mathsf{unif}}(N,\, \rho)$, with an input constrained by power $\rho$ and excited uniformly from all antennas. Using asymptotic random matrix theory (RMT), for any fixed $\rho$, we show that $I_{\mathsf{unif}}(N,\, \rho) = \Theta\left(\rho_c/N\right) = \Theta(N^{\gamma-1})$. In other words, $I_{\mathsf{unif}}(N,\, \rho)$ scales sub-linearly with $N$ in sparse channels. For models described by [1], [2], [4], [30], [45], etc., where $\rho_c = \Theta(N^2)$ is assumed implicitly or explicitly, linear capacity scaling is observed.
- At this stage, it is not clear whether a linear capacity scaling is possible in sparse channels provided that the transmitter knew the statistics of the $D$ dominant DoF and hence, adapted the input in response to this knowledge. For this, we first establish a fundamental upper bound on the ergodic capacity: $C_{\mathsf{erg}}(N,\, \rho) = \mathcal{O}(N^{\frac{\gamma}{2}})$. Since $\frac{\gamma}{2} > \gamma - 1$ for $\gamma \in (0, 2)$, the upper bound is better than the scaling law for $I_{\mathsf{unif}}(N,\, \rho)$, but still *sub-linear* in sparse channels.
- We then show that the fundamental upper bound is achievable (from a scaling viewpoint) by constructing a structured family of channels corresponding to different canonical distributions in which the $D$ non-vanishing coefficients in the decomposed channel matrix can be factored as $D = pq$, where $p$ denotes the number of parallel channels and $q$ denotes the number of DoF per parallel channel. Using results from RMT, we show that the capacity of all channels from this family can be well-approximated as

$$C_{\mathsf{erg}}(N,\, \rho) \approx p \log\left(1 + \rho\, \frac{q}{p}\right). \tag{1}$$

  The above formula reveals a fundamental tradeoff in sparse channels between the number of parallel channels ($p$) and the received $\mathsf{SNR}$ per parallel channel ($\rho\, q/p$). Optimizing this tradeoff results in a channel configuration that is denoted as *the ideal MIMO channel* and whose capacity scaling law is in conformity with the upper bound. We also show how the sparse configuration achieving the scaling law of $C_{\mathsf{erg}}(N, \rho)$ can be realized in practice by adapting the antenna spacings at the transmitter and receiver in response to the sparsity level and $\rho$.
- We next turn our attention to capacity behavior of the structured family as a function of the $\mathsf{SNR}$. For a fixed (but large) $N$, we provide explicit constructions of channels that maximize capacity at a given $\mathsf{SNR}$. For all practical purposes, three channel configurations suffice to maximize capacity over the entire $\mathsf{SNR}$ range. We call these three configurations *the beamforming channel*, *the ideal channel* and *the multiplexing channel*, respectively. The ideal configuration in this setting is the same configuration that is optimal from a capacity scaling perspective.
- While the capacity analysis of correlated MIMO channels has received significant attention over the last few years [4], [30], [41], [45]–[57], much of the progress has relied on indirect characterizations of the limiting eigenvalue distribution function (EED) via the Stieltjes transform. Certain simplifications are possible in either the low- or the high-$\mathsf{SNR}$ extremes [28], [30], [47], [53], [58]–[61]. However, in general, asymptotic capacity formulas are based on solving certain complicated fixed-point equations that do not reveal the impact of channel statistics on capacity *transparently*. Our work applies new RMT techniques to study the capacity scaling problem that result in transparent capacity estimates for all $\mathsf{SNR}$s (see, e.g., (1)). While the structural form of (1) suggests that this formula can be guessed by using a "law of large numbers"-type argument, it must be pointed out that

[2]The channel power is a measure of the richness of multipath and is defined as $\rho_c \triangleq \mathbf{E}\left[\mathsf{Tr}(\mathbf{H}\mathbf{H}^H)\right]$ (see Definition 2 in Sec. II-B).

[3]We use the standard big-Oh ($\mathcal{O}$), little-oh ($o$) and $\Theta$ notations: That is, 1) $f(x) = \mathcal{O}(g(x))$ as $x \to \infty$ if $\lim\limits_{x\to\infty} \frac{f(x)}{g(x)} \leq K_u$ for some $K_u < \infty$, 2) $f(x) = o(g(x))$ if $\lim\limits_{x\to\infty} \frac{f(x)}{g(x)} = 0$, and 3) $f(x) = \Theta(g(x))$ if $K_l \leq \lim\limits_{x\to\infty} \frac{f(x)}{g(x)} \leq K_u$ for some $K_l > 0$ and $K_u < \infty$.

[4]While it is clear that what is "eventual" is debatable, the current channel power normalization is optimistic for large-antenna systems.

this argument is not rigorous and as much justified as if Telatar's seminal result [1] could be arrived at via the same argument instead of via a recourse to RMT.

***Organization:*** Section II introduces the channel model, the channel power, and discusses how it is normalized in this work. Section III formulates the problems studied in this paper. Section IV and V study capacity scaling in sparse channels as a function of antenna dimension $N$ and $\mathsf{SNR}$, respectively. Section VI discusses connections with some recent works. Concluding remarks and directions for future research are provided in Section VII.

***Notation:*** We use upper- and lower-case symbols to denote matrices and vectors, respectively. $\mathbf{X}(m,n)$ and $\mathbf{X}(m)$ denote the entry in the $m$-th row and $n$-th column, and the entry in $m$-th diagonal of $\mathbf{X}$, respectively. $\mathbf{X}^T$ and $\mathbf{X}^H$ denote the regular and the Hermitian transpose of $\mathbf{X}$ while its trace and determinant are denoted by $\mathsf{Tr}(\mathbf{X})$ and $\det(\mathbf{X})$. $\mathbf{I}_N$ stands for the identity matrix of order $N$. The upper-case $\mathbf{E}[\cdot]$ stands for the expectation operator and $\mathcal{CN}(\mu,\sigma^2)$ denotes the complex normal distribution with mean $\mu$ and variance $\sigma^2$. The symbols $\lambda$, $\rho$ and $N$ are in general reserved for eigenvalues, $\mathsf{SNR}$ and antenna dimensions, respectively while the subscripts $\mathsf{erg}$ and $\mathsf{opt}$ stand for ergodic and optimal, respectively.

## II. System Setup

Consider a Rayleigh-fading (zero-mean, complex Gaussian) MIMO system with $N_t$ antennas at the transmitter and $N_r$ antennas at the receiver modeled by the system equation

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2}$$

where $\mathbf{y}$ denotes the $N_r$-dimensional received vector, $\mathbf{x}$ denotes the $N_t$-dimensional transmitted vector, $\mathbf{H}$ denotes the $N_r \times N_t$ channel matrix, and $\mathbf{n}$ denotes the additive white Gaussian noise.

### A. Channel Model

A commonly-used physical model for the MIMO channel $\mathbf{H}$ is given as

$$\mathbf{H} \;=\; \sum_{\ell=1}^{L} \sqrt{N_t N_r}\cdot \beta_\ell \; \mathbf{a}_r(\theta_{r,\,\ell})\mathbf{a}_t^H(\theta_{t,\,\ell}) \tag{3}$$

where the transmitter and receiver arrays are coupled through $L$ propagation paths[5] with complex path gains $\{\beta_\ell\}$, Angles of Departure (AoD) $\{\theta_{t,\,\ell}\}$ and Angles of Arrival (AoA) $\{\theta_{r,\,\ell}\}$. In (3), $\mathbf{a}_t(\theta_t)$ and $\mathbf{a}_r(\theta_r)$ denote the transmit steering and receive response vectors for transmitting/receiving in the normalized direction $\theta_t/\theta_r$ where $\mathbf{a}_t(\theta_t)$ is defined as

$$\mathbf{a}_t(\theta_t) \;\triangleq\; \frac{1}{\sqrt{N_t}}\,[1, e^{-j2\pi\theta_t}, \cdots, e^{-j2\pi(N_t-1)\theta_t}]^T, \tag{4}$$

with $\mathbf{a}_r(\theta_r)$ defined similarly. The angle $\theta$ is related to the physical angle (in the plane of the arrays) $\phi \in [-\pi/2, \pi/2]$ as

$$\theta = d\sin(\phi)/\lambda \tag{5}$$

[5]In general, $L$ is an increasing function of $N_t$ and $N_r$ in (3) because as $\{N_t, N_r\}$ increases (with a fixed antenna spacing), the number of physical paths captured by the arrays increases due to larger array aperture.

where $d$ is the antenna spacing and $\lambda$ is the wavelength of propagation. We consider arrays with $d = \lambda/2$ spacing for which $\theta \in [-1/2, 1/2]$ and we assume that the paths are distributed over the entire angular spread. Finally, we assume that over the time-scales of interest, the location of scattering paths (AoDs and AoAs) remain fixed and the only source of randomness in the channel is due to the complex path gains $\{\beta_\ell\}$, in particular due to their random phases. Furthermore, the path gains are assumed to be statistically independent due to the independence between their random phases.

While the physical model accurately captures the scattering environment, it is cumbersome to use in capacity analysis due to the non-linear dependence of $\mathbf{H}$ on the propagation parameters in (3). Thus, statistical models have become important in capacity analysis of MIMO systems. Initial studies use the so-called *i.i.d. model* where the entries of $\mathbf{H}$ are assumed to be i.i.d. Gaussian random variables [1], [2]. However, this model is not suitable for applications where the assumptions of large antenna spacings or a rich scattering environment become invalid.

A decomposition of $\mathbf{H}$ (called the *virtual representation*) is proposed in [29], [45], [53] where

$$\mathbf{H} = \sum_{m=1}^{N_r}\sum_{n=1}^{N_t} \mathbf{H}_v(m,n)\mathbf{a}_r(\tilde{\theta}_{r,\,m})\mathbf{a}_t^H(\tilde{\theta}_{t,\,n}) = \mathbf{A}_r\mathbf{H}_v\mathbf{A}_t^H. \tag{6}$$

In (6), $\mathbf{H}$ is uniformly sampled over $\left\{\tilde{\theta}_{t,n} = n/N_t\right\}$ and $\left\{\tilde{\theta}_{r,m} = m/N_r\right\}$, and this choice results in unitary discrete Fourier matrices $\mathbf{A}_t$ and $\mathbf{A}_r$, respectively. Thus, $\mathbf{H}$ and

$$\mathbf{H}_v = \mathbf{A}_r^H\mathbf{H}\mathbf{A}_t \tag{7}$$

are unitarily equivalent. Distinct virtual channel coefficients correspond to approximately disjoint subsets of paths (leading to the notion of *path partitioning*) and hence, it is reasonable to model the virtual channel coefficients as independent Gaussian random variables [29].

The uniform sampling in the angular domain $\{\tilde{\theta}_{r,n}\}$ and $\{\tilde{\theta}_{t,m}\}$ for the virtual representation is justified with a uniform linear antenna array at both the transmit and receive ends. Extensions to more general array geometries have also been proposed in [27], [28], [30]. One particular case of this general channel model is the commonly-used Kronecker (separable) correlation model (see [28] and Appendix B for details). While the mathematical development in the rest of the paper could be interpreted in terms of this more general model, we will illustrate our results with the virtual representation due to the intuitive physical interpretation associated with it (e.g., the spatial eigenfunctions are beams in the virtual directions).

### B. Degrees of Freedom

Since $\{\mathbf{H}_v(m,n)\}$ are independent, the statistics of $\mathbf{H}$ are characterized by the virtual channel power matrix $\mathbf{\Psi}$ where

$$\mathbf{\Psi}(m,n) \triangleq \mathbf{E}\left[\left|\mathbf{H}_v(m,n)\right|^2\right]. \tag{8}$$

The matrices $\mathbf{A}_t$ and $\mathbf{A}_r$ constitute the matrices of eigenvectors for the transmit and receive covariance matrices, respectively: $\mathbf{E}[\mathbf{H}^H\mathbf{H}] = \mathbf{A}_t\mathbf{\Lambda}_t\mathbf{A}_t^H$ and $\mathbf{E}[\mathbf{H}\mathbf{H}^H] = \mathbf{A}_r\mathbf{\Lambda}_r\mathbf{A}_r^H$,

where $\mathbf{\Lambda}_t = \mathbf{E}[\mathbf{H}_v^H \mathbf{H}_v]$ and $\mathbf{\Lambda}_r = \mathbf{E}[\mathbf{H}_v \mathbf{H}_v^H]$ are the diagonal matrices of transmit and receive eigenvalues. We can interpret $\mathbf{\Psi}$ as the joint distribution of channel power as a function of transmit and receive virtual angles with $\mathbf{\Lambda}_t$ and $\mathbf{\Lambda}_r$ serving as the corresponding marginal distributions: $\mathbf{\Lambda}_t(n) = \sum_m \mathbf{\Psi}(m,n)$ and $\mathbf{\Lambda}_r(m) = \sum_n \mathbf{\Psi}(m,n)$.

We now introduce two notions that will be useful throughout this work.

*Definition 1:* ***Regular Channels [62].*** Let $\mathbf{H}$, $\mathbf{H}_v$ and $\mathbf{\Psi}$ denote the channel matrix, its virtual channel matrix and the virtual channel power matrix, respectively. A channel is called column-regular if $\sum_m \mathbf{\Psi}(m,n)$ is independent of $n$, row-regular if the above condition is true for $\mathbf{H}^T$, and regular if it is both row- and column-regular. ■

*Definition 2:* ***Channel Power and Independent DoF.*** The *channel power* $\rho_c$ is defined as the energy of the physical scattering environment captured by the transmit and receive antenna arrays. That is,

$$\rho_c \triangleq \mathbf{E}\left[\mathsf{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right] = \mathbf{E}\left[\mathsf{Tr}\left(\mathbf{H}_v\mathbf{H}_v^H\right)\right] \quad (9)$$

$$= \sum_{m=1}^{N_r}\sum_{n=1}^{N_t}\mathbf{\Psi}(m,n) = \sum_{\ell=1}^{L} N_t N_r \mathbf{E}\left|\beta_\ell\right|^2 \quad (10)$$

where the last equality follows from the physical model in (3). The channel power is a measure of the richness of multipath and is, in general, a function of array size (antenna spacing and number of antennas), geometry, etc.

We also define $D$, the number of independent DoF afforded by $\mathbf{H}$, as the number of dominant entries in $\mathbf{H}_v$:

$$D = \left|\{(m,n) \;:\; \mathbf{\Psi}(m,n) > \epsilon\}\right| \quad (11)$$

where the appropriately chosen[6] threshold $\epsilon$ is such that the dominant virtual channel entries contribute significantly to channel power and hence, channel capacity. ■

Initial measurement campaigns [16], [17] suggest that rich multipath models are reasonable models for multi-antenna systems. In particular, Gans *et al.* [16] present a measurement campaign at Lucent Bell Labs with a $7 \times 5$ array and show that a significant fraction of the i.i.d. MIMO capacity can be realized in practice with the BLAST scheme, thus providing evidence for a rich scattering model. While the authors introduce a degree of freedom parameter ($D_F$) [16, equations (4) and (5)] that could be viewed as *effective rank* of the channel, they neither study how many of the $N_tN_r$ channel coefficients are dominant as a function of the cell-site nor the behavior of $D_F$ as a function of antenna dimensions. More pertinent to the discussion here, the authors show that there exist cell-sites where the BLAST potential could not be reached, and they suggest [16, pp. 578-579] that this could be because of poor scattering (or fewer than $N_tN_r$ DoF in the channel). In [17], the authors report a campaign in Manhattan with a $16 \times 16$ array that suggests a close-fit for the channel with a rich multipath[7] model such as the Kronecker model. However, this work does not address settings where the terrain is typically different from that of downtown Manhattan (with transmit antennas usually on top of tall buildings). Another early work [15, Sec. VC] studies capacity scaling with up to $10$ transmit and receive antennas at the BYU campus, but it would be unfair to compare the setting in this work with [15] as they assume a fixed array aperture as antenna numbers increase.

Note that some of the early work from the Bell Labs (BLAST) project present evidence for sparsity. For example, Kyritsi *et al.* study capacity with a $15 \times 12$ array in an indoor setting and claim that " $\cdots$ the study showed that the system capacity does not scale linearly with the number of antenna elements, as one would have expected in a rich scattering environment. This indicates that the signals are highly correlated $\cdots$" [63, p. 1228]. Following these early works, MIMO channel modeling has attracted significant attention in the literature and extensive studies have been reported. In contrast with [16], [17], recent work provide evidence to suggest that many measured wireless channels are *sparse* reflecting a sparse multipath environment. We now present a small sampling of this extensive literature. Fundamental electromagnetic studies [32]–[39] as well as recent measurement campaigns [19, Figs. 9 and 11], [20]–[26], [31] and channel modeling efforts under diverse sets of assumptions [37]–[39] indicate that only a small subset of the available DoF (the $N_tN_r$ statistical parameters $\{\mathbf{\Psi}(m,n)\}$) in a MIMO channel are dominant enough to be leveraged towards reliable communications over practical $\mathsf{SNR}$ ranges.

A simpler communication-theoretic motivation for sparsity is that while there may be many channel coefficients whose energy levels are non-zero, their relative power is rather low and therefore are masked by noise It becomes impossible or too costly to estimate such coefficients accurately and thus from the transmitter's viewpoint, it is reasonable to treat their contributions as noise. A sparse virtual channel matrix provides a model for spatial correlation in $\mathbf{H}$: in general, the sparser the $\mathbf{H}_v$, the higher the correlation in $\mathbf{H}$ and *vice versa*.

It must also be pointed out that many works, apart from [16], propose a DoF metric that is related to the rank of $\mathbf{H}$ (multiplexing gain) which captures MIMO performance; see e.g., [19], [64]–[66] and references therein for both the single-user and multi-user contexts. In contrast, (11) proposes a DoF metric that captures the $N_tN_r$ statistical parameters of the channel matrix. While the rank of $\mathbf{H}$ is an important DoF metric, it does not accurately capture the impact of sparsity on capacity. In particular, the multiplexing gain may not change even if many coefficients in the channel cannot be learned reliably. In this setting, understanding the scaling of a metric such as in (11) can shed clear insights on capacity behavior. While the subsequent analysis in this work builds a case for (11), its suitability as a DoF metric merits further attention.

### C. Channel Power Normalization

It is common in the literature to normalize the channel so as to ensure that site-specific received-$\mathsf{SNR}$ biases caused by large-scale fading are removed. More specifically, if $\widetilde{\mathbf{H}}$ denotes the un-normalized channel matrix (inclusive of large-scale fading losses), the normalized channel matrix $\mathbf{H}$ is given

[6]The choice of this threshold needs further study and is beyond the scope of this paper. See [31] for one definition of $\epsilon$, as a function of the power of the dominant path.

[7]See App. B for why the Kronecker model is an example of a rich multipath model.

as

$$\mathbf{H} = \sqrt{L_p} \cdot \widetilde{\mathbf{H}} \tag{12}$$

where $L_p$ denotes the path loss factor. With the standard normalization (denoted as $\mathsf{N}_1$), we constrain $\mathbf{H}$ such that $\rho_c = N_t N_r$. While $\mathsf{N}_1$ is reasonable for systems with a small number of antennas, it fails to be consistent in asymptotic (in number of antennas) capacity analyses. Further, the use of $\mathsf{N}_1$ in sparse channels results in the unbounded growth of the variances of some (the dominant) channel coefficients as antenna dimensions increase. We identify the main problem with $\mathsf{N}_1$ and propose an alternate normalization in Sec. III to overcome it.

The total power captured by all the receiver antennas, $\rho_r$, is related to the transmit power constraint, $\rho_t = \rho$, by

$$\rho_r \triangleq \mathbf{E}\left[\mathbf{x}^H \widetilde{\mathbf{H}}^H \widetilde{\mathbf{H}} \mathbf{x}\right] = \mathbf{E}\left[\mathsf{Tr}\left(\widetilde{\mathbf{H}} \mathbf{x}\mathbf{x}^H \widetilde{\mathbf{H}}^H\right)\right] \tag{13}$$

$$\overset{(a)}{=} \frac{\rho_t}{N_t} \cdot \mathbf{E}\left[\mathsf{Tr}\left(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H\right)\right] \tag{14}$$

$$\overset{(b)}{=} \frac{\rho_t}{N_t} \cdot \frac{\mathbf{E}\left[\mathsf{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right]}{L_p} \tag{15}$$

where (a) follows from the assumption that the input is excited uniformly from the $N_t$ transmit antennas and (b) from (12). On the other hand, using the Friis transmission equation for MIMO systems with imperfect CSI at the transmitter and a coherent receiver, we have [44]

$$\frac{\rho_r}{N_r} = \frac{\rho_t}{N_t} \cdot \frac{N_t \cdot (1 + o(1))}{L_p} \tag{16}$$

where the $o(1)$ factor goes to zero as the operating frequency increases.

Comparing (15) and (16), we obtain the relationship:

$$\mathbf{E}\left[\mathsf{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right] = N_t N_r \cdot (1 + o(1)), \tag{17}$$

which is the basis for $\mathsf{N}_1$. However, equating (15) with (16) ignores the fact that $\rho_r$ cannot exceed $\rho_t$ (law of energy conservation), independent of the choice of $N_t$ or $N_r$. Incorporating this fact, we see that

$$\mathbf{E}\left[\mathsf{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right] \leq N_t L_p \tag{18}$$

for all choices of $\{N_t, N_r\}$ with

$$\mathbf{E}\left[\mathsf{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right] = N_t N_r \tag{19}$$

being a reasonable approximation[8] for $N_t$ and $N_r$, both small. In other words, for sufficiently large $N_t$ and $N_r$, the channel power has to satisfy $\rho_c = \mathcal{O}(N_t)$. Similar explanations for the unrealistic nature of $\mathsf{N}_1$, especially in the context of capacity scaling, can be found in [40]–[43].

While it may be argued that $L_p$ is so large in current/proposed wireless technologies that $\rho_c = \mathcal{O}(N_t)$ will kick in only for an exorbitantly large number of antennas, new technologies such as 60 GHz signaling not only envision the use of 16 to 32 antennas, but also short-range communications, which in turn implies that $L_p$ is bound to be small [7]–[11]. Nevertheless, applicability to 60 GHz is contingent on further measurement studies.

[8]The reason why we cannot claim that (19) is valid for all $N_r$ satisfying $N_r \leq \lfloor L_p \rfloor$ is that the far-field assumptions necessary for (16) to hold would be violated much earlier than $N_r = \lfloor L_p \rfloor$.

## III. Problem Formulation

In the classical setting where the receiver has perfect channel state information and the transmitter has complete information about the channel statistics, $\{\mathbf{\Psi}(m,n)\}$, the ergodic capacity is given as [1], [2]

$$C_{\mathsf{erg}}(N_t, N_r, \rho) = \max_{\mathbf{Q}\,:\,\mathsf{Tr}(\mathbf{Q}) \leq \rho} \mathbf{E}\left[\log_2 \det(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^H)\right] \tag{20}$$

$$= \max_{\mathbf{Q}_v\,:\,\mathsf{Tr}(\mathbf{Q}_v) \leq \rho} \mathbf{E}\left[\log_2 \det(\mathbf{I} + \mathbf{H}_v\,\mathbf{Q}_v\,\mathbf{H}_v^H)\right] \tag{21}$$

where $\mathbf{Q} = \mathbf{E}\left[\mathbf{x}\mathbf{x}^H\right]$ is the trace-constrained transmit covariance matrix and $\mathbf{Q}_v = \mathbf{A}_t^H \mathbf{Q} \mathbf{A}_t$. In this context, it is shown in [54], [62], [67], [68] that capacity is achieved by a diagonal $\mathbf{Q}_v$.

The mutual information achievable with a uniform-power input $\mathbf{Q}_v = \mathbf{Q} = \frac{\rho}{N_t}\mathbf{I}_{N_t}$ is denoted as $I_{\mathsf{unif}}(N_t, N_r, \rho)$. Note that a uniform-power input is optimal under certain conditions: i) i.i.d. channel [1], ii) regular channels [62], iii) maximin robust for Rayleigh fading channels when the transmitter has no statistical information [48], iv) high-SNR regime for non-degenerate correlated channels [3], [47], [49], etc. With $\mathbf{H}$ denoting an $N_r \times N_t$ channel with independent entries and under the assumption that $\{N_r, N_t\} \to \infty$, $\frac{N_r}{N_t} \to \kappa_{\mathsf{TLV}} \in (0, \infty)$, [30] computes $I_{\mathsf{unif}}(N_t, N_r, \rho)$ by using the Stieltjes transformational approach as

$$\frac{I_{\mathsf{unif}}(N_t, N_r, \rho)}{N_t} \overset{\{N_r, N_t\} \to \infty}{\to} \mathbf{E}_{\mathcal{T}}\left[\log_2\left(1 + \rho\,\kappa_{\mathsf{TLV}}\,\mathbf{E}_{\mathcal{R}}\left[\mathcal{G}(\mathcal{R},\mathcal{T})\mathcal{D}(\mathcal{R})|\mathcal{T}\right]\right)\right] + \kappa_{\mathsf{TLV}} \cdot \left(\log_2(e)\left(\mathbf{E}_{\mathcal{R}}\left[\mathcal{D}(\mathcal{R})\right] - 1\right) - \mathbf{E}_{\mathcal{R}}\left[\log_2(\mathcal{D}(\mathcal{R}))\right]\right). \tag{22}$$

In (22), $\mathcal{R}$ and $\mathcal{T}$ are uniformly distributed random variables on $[0,1]$, and $\mathbf{E}_{\mathcal{R}}$ and $\mathbf{E}_{\mathcal{T}}$ denote expectation with respect to these random variables, respectively. $\mathcal{G}(r,t)$ stands for the continuous representation of the virtual channel power matrix supported on $[0,1] \times [0,1]$, that is,

$$\mathcal{G}(r,t) = \mathbf{E}\left[|\mathbf{H}_v(i,j)|^2\right], \quad \frac{i}{N_r} \leq r \leq \frac{i+1}{N_r}, \quad \frac{j}{N_t} \leq t \leq \frac{j+1}{N_t}. \tag{23}$$

$\mathcal{D}(r)$ is the solution to the following fixed-point equation:

$$\mathcal{D}(r)\left(1 + \rho\,\mathbf{E}_{\mathcal{T}}\left[\frac{\mathcal{G}(r,\mathcal{T})}{1 + \rho\,\kappa_{\mathsf{TLV}}\,\mathbf{E}_{\mathcal{R}}\left[\mathcal{D}(\mathcal{R})\mathcal{G}(\mathcal{R},\mathcal{T})|\mathcal{T}\right]}\right]\right) = 1. \tag{24}$$

The above formula can also be extended to the case of $C_{\mathsf{erg}}(N_t, N_r, \rho)$ (see [30] for details).

In the setting where the transmitter does not have complete information about the variances of all the entries in $\mathbf{H}_v$, a fundamentally new paradigm arises due to the ability to spatially configure (or distribute) the $D < N_t N_r$ DoF in the

available $N_t N_r$ spatial dimensions of $\mathbf{H}_v$. In this context, the complicated nature of the fixed-point equation that describes asymptotic capacity in (22)-(24) implies that any intuition on the impact of sparsity on capacity difficult. To overcome this difficulty, we will make the following simplifying assumptions in this work.

- ***Assumption 1:*** We will henceforth assume that $N_t = N_r = N$. We will use $C_{\mathsf{erg}}(N,\rho)$ and $I_{\mathsf{unif}}(N,\rho)$ to denote the ergodic capacity and mutual information with a uniform-power input, respectively. Due to the similarity in structure between (20) and (21), we will henceforth use $\mathbf{H}$ and $\mathbf{Q}$ to denote $\mathbf{H}_v$ and $\mathbf{Q}_v$, respectively.
- ***Assumption 2:*** From Definition 2, it follows that the $D$ dominant entries of $\mathbf{\Psi}$ are bounded away from zero and hence,

$$\rho_c = \Theta(D) \tag{25}$$

  under the assumption that $\sup_N \max_{m,n} \mathbf{\Psi}(m,n) \leq K < \infty$. To facilitate closed-form capacity analysis, we will make the further assumption that each dominant coefficient of $\mathbf{H}$ is $\mathcal{CN}(0,1)$ and the non-dominant coefficients are zero. While this assumption is simplistic, it allows us to capture the underlying trends in capacity scaling and as a function of $\mathsf{SNR}$.
- ***Assumption 3:*** From the perspective of scaling laws, a family of virtual channel matrices of increasing dimension is said to be *sparse* if $D = o(N^2)$ as $N$ increases. More specifically, we will assume that

$$\rho_c = D = \Theta\left(N^{\gamma}\right) \ , \ \gamma \in (0,2] \tag{26}$$

  where the first equality follows from Assumption 2.
- ***Assumption 4:*** In the rest of the paper, without loss in generality, we will assume that $D$ is an integer when we write $D = N^{\gamma}$. This is because given a fixed $N$ and arbitrary $\gamma$, a non-integer value of $D$ is obtained by time-sharing an appropriate combination of channels of integer dimensions [69]. Similarly, we assume that $D$ can be decomposed as $D = pq$ where $p$ and $q$ are integers and use time-sharing to realize non-integer $\{p,q\}$ values.

*Definition 3:* ***Mask matrices.*** It is convenient to combine the above assumptions to generate a sparse family of matrices as

$$\mathbf{H} = \mathbf{H}_{\mathsf{iid}} \odot \mathbf{M} \tag{27}$$

where $\odot$ denotes elementwise (Hadamard) product, $\mathbf{H}_{\mathsf{iid}}$ is an i.i.d. matrix of $\mathcal{CN}(0,1)$ entries, and $\mathbf{M}$ is a mask matrix with $D$ unit entries and zeros elsewhere. It can be seen that the virtual channel power matrix $\mathbf{\Psi} = \mathbf{M}$ and the entries of $\mathbf{\Lambda}_t$ and $\mathbf{\Lambda}_r$ represent the number of non-zero elements in the columns and rows of $\mathbf{M}$, respectively. ■

We address the following fundamental questions in this paper.

- ***Question 1:*** We first study the impact of different sparse channel configurations on capacity. Given a channel power (and DoF) scaling law $\rho_c = D = \Theta\left(N^{\gamma}\right), \gamma \in (0,2]$, what is the fastest achievable capacity scaling law? Which sparse channel configuration achieves this fastest capacity scaling law? More precisely, we want to solve the following optimization:

$$C_{\mathsf{opt,\,erg}}(N,\rho) \triangleq \max_{\mathbf{H}\,\in\,\mathcal{H}(D)} \ \max_{\mathbf{Q}\,:\,\mathsf{Tr}(\mathbf{Q})\,\leq\,\rho} \mathbf{E}_{\mathbf{H}}\left[\log_2 \det\left(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^H\right)\right] \tag{28}$$

  where $\mathcal{H}(D)$ denotes the class of matrices

$$\mathcal{H}(D) = \Big\{\mathbf{H} : \mathbf{H} = \mathbf{H}_{\mathsf{iid}} \odot \mathbf{M},\ \mathbf{M}(i,j) = 0 \text{ or } 1, \text{ and } \mathsf{Tr}(\mathbf{M}^H\mathbf{M}) = D\Big\}. \tag{29}$$

- ***Question 2:*** We next ask: what is the structure of the channel configuration that maximizes capacity at a given $\mathsf{SNR}$?

Sec. IV studies the first question while Sec. V studies the second. Note that the capacity formulation in (28) is fundamentally different from the conventional formulation (21) where the optimization is only over the input covariance matrix. Furthermore, while closed-form expressions for capacity have been obtained under certain assumptions in the conventional formulation [28], [30], [47], [53], [58]–[61], this new formulation enables approximate capacity characterization at all $\mathsf{SNR}$'s.

## IV. Capacity Scaling as a Function of Antenna Dimensions

In our goal of understanding the capacity scaling laws of single-user MIMO systems, we start with $I_{\mathsf{unif}}(N,\rho)$.

### A. Scaling Law of $I_{\mathsf{unif}}(N,\rho)$

Note that $I_{\mathsf{unif}}(N,\rho)$ can be written as

$$I_{\mathsf{unif}}(N,\rho) \quad = \quad \mathbf{E}\left[\sum_{i=1}^{N} \log_2\left(1 + \frac{\rho\lambda_i}{N}\right)\right] \tag{30}$$

where $\lambda_i$ are the eigenvalues of $\mathbf{H}\mathbf{H}^H$. The parallel channel decomposition of (30) suggests that if $\Theta(f(N))$ eigenvalues of $\frac{\mathbf{H}\mathbf{H}^H}{N}$ are $\Theta(1)$ where $f(N)$ is a particular choice that grows sub-linearly with $N$, then $I_{\mathsf{unif}}(N,\rho)$ would scale as $\Theta(f(N))$ as $N$ increases. In other words, if

$$\mathsf{Tr}(\mathbf{H}\mathbf{H}^H) = \sum_{i=1}^{N} \lambda_i = \Theta(Nf(N)), \tag{31}$$

then the likelihood that $I_{\mathsf{unif}}(N,\rho) = \Theta(f(N))$ increases. The following theorem formalizes this heuristic.

*Theorem 1:* Let the $N \times N$ channel matrix $\mathbf{H}$ have independent entries with $\mathbf{H}(m,n) \sim \mathcal{CN}(0, \mathbf{\Psi}(m,n))$ and

$$\sup_N \max_{m,n} \mathbf{\Psi}(m,n) \leq K < \infty. \tag{32}$$

Leveraging results from the asymptotic spectral theory of random matrices, we have

$$\frac{\rho K}{1 + 4\rho K} \cdot \log_2(e) \leq \lim_N \frac{I_{\mathsf{unif}}(N,\rho) \cdot N}{\mathbf{E}\left[\mathsf{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right]} \leq \rho \cdot \log_2(e). \tag{33}$$

In other words, $I_{\mathsf{unif}}(N,\rho) = \Theta(f(N))$ if and only if $\mathbf{E}\left[\mathsf{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right] = \Theta(Nf(N))$.

*Proof:* See Appendix A. ∎

Theorem 1 states that $I_{\mathsf{unif}}(N,\rho)$ can be described completely by the channel power per antenna dimension which is a measure of the richness of multipath. Further, following Assumption 2, Theorem 1 implies that $I_{\mathsf{unif}}(N,\rho) = \Theta(f(N))$ if and only if the number of independent DoF in the channel, $D$, scales with $N$ as $\Theta(Nf(N))$.

Specific examples of Theorem 1 have been established in the literature in different forms.

- The famous linear capacity scaling result of Telatar [1] critically depends on the assumption that $\mathbf{H}$ has i.i.d. $\mathcal{CN}(0,1)$ entries resulting in $\rho_c = D = N^2$.
- In the correlated setting that can be modeled with the Kronecker model (a transmit and receive side correlation matrix), Chuah *et al.* [4] established a linear capacity scaling result. The critical assumption behind this result is the condition [4, p. 641] that the empirical eigenvalue distribution (EED) function (see Definition 5 in App. A) of both the transmit and receive covariance matrices converge. In Appendix B, we show that this condition is equivalent to $\rho_c = \Theta(N^2)$.
- In a prior work [45], a $\mathcal{D}$-connected model that ensures a systematic coupling between $\mathcal{D}$ of the $N$ transmit antennas with each receive antenna is studied. We had showed that the ratio $\frac{\mathcal{D}}{N}$ is decisive in controlling the capacity behavior. If $\frac{\mathcal{D}}{N} \to \kappa_{\mathsf{LRS}} \in (0,\infty)$, capacity is shown to scale linearly whereas if $\mathcal{D} = \Theta(1)$, capacity is shown to saturate. It can be seen that $\rho_c = \mathcal{D}N = \Theta(N^2)$ in the former case, while it scales as $\rho_c = \mathcal{D}N = \Theta(N)$ in the latter case, and thus in agreement with the conclusions of Theorem 1. In general, if $\rho_c = \Theta(N^\gamma)$ for some $\gamma \in [1,2]$, we have $I_{\mathsf{unif}}(N,\rho) = \Theta(N^{\gamma-1})$.
- A diagonal channel which corresponds to a scattering environment with a line of scatterers and satisfies $\rho_c = \Theta(N)$ leads to capacity saturation [70]. Note that the above conclusion critically depends on the twin assumptions that $\rho_c = \Theta(N)$ and $\mathbf{Q} = \frac{\rho}{N}\mathbf{I}_N$. However, if either assumption were to fail, capacity saturation need not be observed. For example, if $\mathbf{H}$ is rank 1, while a uniform-power input is wasteful and leads to $\Theta(1)$ capacity growth, beamforming along the particular dominant direction would achieve a capacity scaling of $\Theta(\log(N))$.

Theorem 1 can also be established via the implicit Stieltjes transformational approach. In [71, Chap. 4], we construct a necessary and sufficient condition for linear capacity scaling in terms of the EED function via a manipulation of the fixed-point equations in (22)-(24) that characterize capacity in the limit. We then show that linear capacity scaling is possible if and only if $D = \Theta(N^2)$. Theorem 1 can also be extended (see [71, Theorem 4.12] for details) to linear scaling of the average mutual information with the class of *almost full-rank* input covariance matrices, defined as,

$$\mathbf{Q} \text{ such that } \left|\left\{i:\, \mathbf{Q}(i) = \Theta\left(\frac{1}{N}\right)\right\}\right| = \Theta(N). \tag{34}$$

### B. Fundamental Upper Bound on Scaling Law of $C_{\mathsf{erg}}(N,\rho)$

Following the above discussion, it is important to understand how far from optimality a uniform-power input is with respect to a scaling viewpoint. In this pursuit, we now obtain a fundamental upper bound on the scaling law of $C_{\mathsf{erg}}(N,\rho)$.

*Theorem 2:* Given a constraint on the scaling of $D$, the fastest rate at which $C_{\mathsf{erg}}(N,\rho)$ can scale is bounded as

$$C_{\mathsf{opt,\,erg}}(N,\rho) = \mathcal{O}(\sqrt{D}). \tag{35}$$

*Proof:* See Appendix C. ∎

The importance of the above theorem is that in the special case where $\rho_c = D = \Theta(N^\gamma)$, no matter how the spatial DoF are distributed, $C_{\mathsf{erg}}(N,\rho)$ cannot exceed $\mathcal{O}(\sqrt{D}) = \mathcal{O}\big(N^{\frac{\gamma}{2}}\big)$. Thus, we have a benchmark to compare the capacities of different channel configurations in the family $\mathcal{H}(D)$. Note that in the same setting, $I_{\mathsf{unif}}(N,\rho) = \Theta(N^{\gamma-1})$. Since $\gamma - 1 < \frac{\gamma}{2}$ for $\gamma < 2$, we observe that the upper bound can be strictly larger than the rate achieved with a uniform-power scheme.

### C. A Structured Family of Channels to Achieve the Fundamental Upper Bound

To show that the upper bound in Theorem 2 is realizable, in an order sense, we now introduce a structured family of mask matrices $\mathbf{M}$ (and corresponding $\mathbf{H}$'s via Definition 3) that characterize systematically varying configurations of the $D$ DoF. For a given $D$ and $N$, the family is defined by two integer parameters $(p,\, q)$ such that $D = pq$ (see Assumption 4).

*Definition 4:* ***A family of mask matrices.*** Consider an $N \times N$ mask matrix $\mathbf{M}$ with $D = N^\gamma = pq$ non-zero entries distributed over $p$ columns and $q$ non-zero entries in each column such that $1 \le \{p,\, q\} \le N$ and $r = \max(p,q)$. The non-vanishing entries of $\mathbf{M}$ are contained in a $r \times p$ sub-matrix whose non-zero entries are given by

$$\mathbf{M}\big((n+m) \bmod r, n\big) = 1, \quad 1 \le n \le p,\ q_- \le m \le q_+ \tag{36}$$

where $q_- = \lceil -(q-1)/2 \rceil$ and $q_+ = \lceil (q-1)/2 \rceil$. In other words, the $D$ DoF are configured to generate a channel matrix that can support $p$ parallel channels and $q = D/p$ DoF per parallel channel. ∎

The feasible range of $p$ depends on the value of $D$ relative to $N^2$. Let $D = N^\gamma$ with $\gamma \in (0,2]$. Let $p = N^\alpha$, $q = N^{\gamma-\alpha}$, and $\alpha \in [0,1]$ in Definition 4. For a given $\gamma$, the feasible range for $\alpha$ is

$$\alpha_{\min} \triangleq \max(\gamma - 1, 0) \le \alpha \le \min(\gamma, 1) \triangleq \alpha_{\max}. \tag{37}$$

When $\gamma = 2$, $\alpha_{\min} = \alpha_{\max} = 1$, and all channel configurations essentially reduce to $\mathbf{H}_{\mathsf{iid}}$ in which $p = q = N$ and the resulting channel is the familiar $N \times N$ i.i.d. channel. The received[9] $\mathsf{SNR}$ per parallel channel is

$$\frac{\rho_r}{p} = \frac{\mathbf{E}[\|\mathbf{Hx}\|^2]}{p} = \rho\frac{D}{p^2} = \rho\frac{q}{p} \tag{38}$$

[9]We will henceforth define received $\mathsf{SNR}$ with respect to $\mathbf{H}$ and not bother about the path loss factor as in (13).

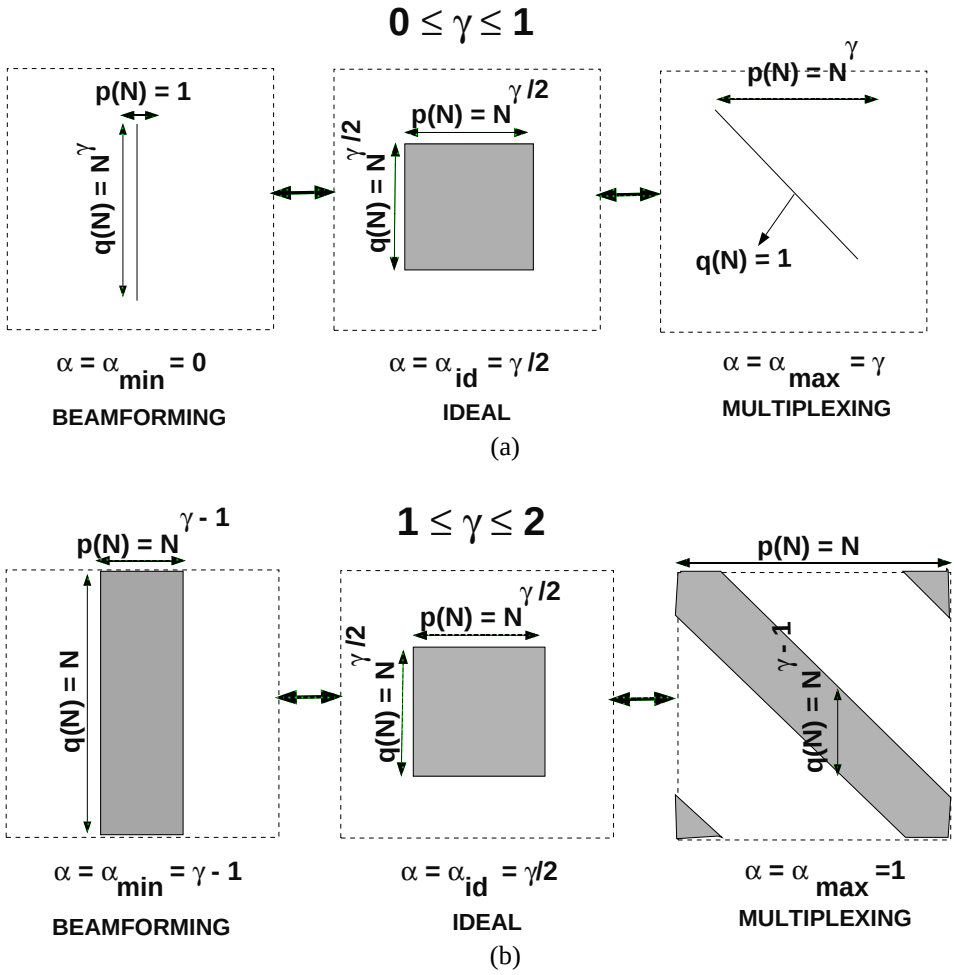


Fig. 1. A schematic illustrating the family of channels: (a) $\gamma \in [0,1]$ and (b) $\gamma \in [1,2]$.

where the second equality follows from the fact that

$$\mathbf{E}[\|\mathbf{Hx}\|^2] = \mathbf{E}\left[\mathsf{Tr}\left(\mathbf{Hxx}^H\mathbf{H}^H\right)\right] \quad (39)$$

$$= \frac{\rho}{p}\cdot\mathbf{E}\left[\mathsf{Tr}\left(\mathbf{HH}^H\right)\right] = \frac{\rho D}{p}. \quad (40)$$

To simplify the characterization in Definition 4, note that if $q \geq p$, the non-vanishing part of $\mathbf{M}$ is a $q \times p$ matrix of ones (the corresponding part of $\mathbf{H}$ is a $q \times p$ i.i.d. matrix), whereas if $q < p$, the non-vanishing part of $\mathbf{M}$ is a $p \times p$ matrix with essentially $q$ non-vanishing diagonals. In the terminology of [45], $\mathbf{M}$ is a $q$-connected $p$-dimensional matrix. Thus, the corresponding channel matrices $\mathbf{H}$ are regular (see Definition 1), with $\mathsf{rank}(\mathbf{H}) = \min(r,p) = p$ for which the uniform-power input over the $p$ parallel channels is optimal [62].

The family of channels is illustrated in Fig. 1. As can be seen from Fig. 1, note that the above definition is asymmetric as $p$ increases from $N^{\alpha_{\min}}$ to $N^{\alpha_{\max}}$. This is because if $p > q$, a $q \times p$ i.i.d. channel does not result in $p$ parallel channels with $q$ DoF per parallel channel. The only channel structure that ensures this condition is the $q$-connected $p$-dimensional matrix.

We also identify three distinct regimes for $p$ (and $\rho_r/p$) which result in distinct capacity behaviors.

***i) Beamforming regime:*** $\alpha \in \left[\alpha_{\min}, \frac{\gamma}{2}\right) \iff p = o(q)$ and $\frac{\rho_r}{p} \to \infty$ as $N \to \infty$.

***ii) Ideal regime:*** $\alpha = \frac{\gamma}{2} \iff \frac{q}{p} = \delta = \Theta(1)$ and $\frac{\rho_r}{p} = \rho\,\delta = \Theta(1)$ as $N \to \infty$.

***iii) Multiplexing regime:*** $\alpha \in \left(\frac{\gamma}{2}, \alpha_{\max}\right] \iff q = o(p)$ and $\frac{\rho_r}{p} \to 0$ as $N \to \infty$.

Three canonical channel configurations, one from each regime, corresponding to $\alpha = \alpha_{\min}, \frac{\gamma}{2}$ and $\alpha_{\max}$, will be referred to as *beamforming*, *ideal* and *multiplexing* channels and will be denoted by $\mathbf{H}_{\mathsf{bf}}, \mathbf{H}_{\mathsf{id}}$ and $\mathbf{H}_{\mathsf{mux}}$, respectively. Note that while the beamforming channel refers to a particular DoF configuration, the beamforming regime corresponds to a family of channels that satisfy $p = o(q)$.

### *D. Capacity Analysis of Sparse Family of Channels*

We now analyze the capacity of the family of channels introduced above. For this, we define $C_{\mathsf{appx}}(N,\rho)$ as

$$C_{\mathsf{appx}}(N,\rho) \triangleq p\log\left(1+\rho\,\frac{q}{p}\right). \quad (41)$$

The proofs of all statements in this section are relegated to Appendix D.

The capacity of channels in the beamforming regime follow from the asymptotic spectral characterization in Bai and Yin [72].

*Theorem 3 (**Beamforming Regime):*** Let $\mathbf{H}$ be a $q \times p$ channel in the beamforming regime with $\kappa_{\mathsf{bf}} = \frac{q}{p} \to \infty$. The structure of the channel configuration in the beamforming regime depends on whether i) $p$ is finite as $\kappa_{\mathsf{bf}} \to \infty$, or ii) $p \to \infty$ as $\kappa_{\mathsf{bf}} \to \infty$. In the former case, we have

$$\frac{C_{\mathsf{erg,\,bf}}(N,\rho)}{C_{\mathsf{appx}}(N,\rho)} \to 1 + \frac{|\mathcal{O}_{\mathsf{prob.}}\left(1/\sqrt{q}\right)|}{\log_2\left(1+\rho\,\kappa_{\mathsf{bf}}\right)}. \tag{42}$$

In the latter case, $C_{\mathsf{erg,\,bf}}(N,\rho)$ is given by

$$\frac{C_{\mathsf{erg,\,bf}}(N,\rho)}{C_{\mathsf{appx}}(N,\rho)} \to 1+ \frac{\log_2\left(1+\sqrt{1-4t^2}\right) + \frac{\log_2(e)}{4t^2}\cdot\left(1-\sqrt{1-4t^2}\right) - \log_2(e)}{\log_2\left(1+\rho\,\kappa_{\mathsf{bf}}\right)} \tag{43}$$

where $t \triangleq \frac{\rho\sqrt{\kappa_{\mathsf{bf}}}}{1+\rho\,\kappa_{\mathsf{bf}}}$. In either case, the correction term $\Delta C_{\mathsf{bf}}(N,\rho)$, defined as,

$$\Delta C_{\mathsf{bf}}(N,\rho) \triangleq \frac{C_{\mathsf{erg,\,bf}}(N,\rho) - C_{\mathsf{appx}}(N,\rho)}{C_{\mathsf{appx}}(N,\rho)} \tag{44}$$

satisfies $|\Delta C_{\mathsf{bf}}(N,\rho)| \approx \mathcal{O}\left(\frac{1}{\log_2(1+\rho\kappa_{\mathsf{bf}})}\right)$. ∎

Among all channels in the beamforming regime, the smallest value for the correction term is achieved by the beamforming channel that corresponds to $\alpha = \alpha_{\min}$. However, $\kappa_{\mathsf{bf}} = \frac{q}{p} \to \infty$ as $N \to \infty$ for every channel in this regime and hence the correction term $\Delta C_{\mathsf{bf}}(N,\rho) \to 0$ for all $\rho$. Thus, $C_{\mathsf{appx}}(N,\rho)$ is a good asymptotic approximation of the exact capacity at all SNRs. For a fixed $N$, note that $|\Delta C_{\mathsf{bf}}(N,\rho)|$ takes the largest value in the low-SNR regime.

The capacity of channels in the ideal regime is essentially a reformulation of the well-known i.i.d. capacity formula due to Telatar [1], Rapajic and Popescu [73], and Verdú and Shamai [74].

*Proposition 1 (**Ideal Regime):*** Let $\mathbf{H}$ be a $q \times p$ channel and let $\kappa_{\mathsf{id}} = \frac{q}{p} = \Theta(1)$. If $\kappa_{\mathsf{id}} \in [1,\infty)$, the capacity of $\mathbf{H}$ is given by

$$\frac{C_{\mathsf{erg,\,id}}(N,\rho)}{q} \overset{\{p,q\}\to\infty}{\to} \log_2\left(1+\rho-\rho h\right) + \frac{1}{\kappa_{\mathsf{id}}}\log_2\left(1-\rho h+\rho\kappa_{\mathsf{id}}\right) - \log_2(e)\,\frac{h}{\kappa_{\mathsf{id}}} \tag{45}$$

$$h = \frac{1}{2}\left[1+\kappa_{\mathsf{id}}+\frac{1}{\rho} - \sqrt{\left(1+\kappa_{\mathsf{id}}+\frac{1}{\rho}\right)^2 - 4\kappa_{\mathsf{id}}}\right]. \tag{46}$$

If $\kappa_{\mathsf{id}} \in (0,1)$, we have

$$\frac{C_{\mathsf{erg,\,id}}(N,\rho)}{p\cdot\log_2(e)} \overset{\{p,q\}\to\infty}{\to} \log_e\left(\frac{1+2\rho\kappa_{\mathsf{id}}+\sqrt{1+4\rho\kappa_{\mathsf{id}}}}{2}\right) + \frac{\sqrt{1+4\rho\kappa_{\mathsf{id}}}-(1+2\rho\kappa_{\mathsf{id}})}{2\rho\kappa_{\mathsf{id}}}. \tag{47}$$

Define the correction term

$$\Delta C_{\mathsf{id}}(N,\rho) \triangleq \frac{C_{\mathsf{erg,\,id}}(N,\rho) - C_{\mathsf{appx}}(N,\rho)}{C_{\mathsf{appx}}(N,\rho)}. \tag{48}$$

Then, $\Delta C_{\mathsf{id}}(N,\rho)$ can be well-approximated in the low- and the high-SNR extremes as follows:

$$|\Delta C_{\mathsf{id}}(N,\rho)| \approx \begin{cases} \text{as}\ \ \rho\to 0\ \ \text{by}\ \begin{cases} \frac{5}{2}\,\rho(1+\kappa_{\mathsf{id}}), & \kappa_{\mathsf{id}} > 1 \\ 5\rho\kappa_{\mathsf{id}}, & \kappa_{\mathsf{id}} \le 1, \end{cases} \\ \text{as}\ \ \rho\to\infty\ \ \text{by}\ \begin{cases} \frac{\log\left(\frac{\kappa_{\mathsf{id}}-1}{\kappa_{\mathsf{id}}}\right)}{\log(\rho)}, & \kappa_{\mathsf{id}} > 1 \\ \frac{1}{\log(\rho\kappa_{\mathsf{id}})}, & \kappa_{\mathsf{id}} \le 1. \end{cases} \end{cases} \tag{49}$$

∎

Observe that the correction terms vanish and $C_{\mathsf{appx}}(N,\rho)$ is accurate in either SNR extreme. In the medium-SNR regime, $C_{\mathsf{appx}}(N,\rho)$ is only an approximate fit to $C_{\mathsf{erg}}(N,\rho)$.

The capacity of channels in the multiplexing regime follows from the regularity property coupled with an implicit spectral characterization in [75].

*Theorem 4 (**Multiplexing Regime):*** Let $\mathbf{H}$ be a channel in the multiplexing regime with $\kappa_{\mathsf{mux}} = \frac{q}{p} \to 0$. In the asymptotics of $N$, $C_{\mathsf{erg,\,mux}}(N,\rho)$ is given by

$$\frac{C_{\mathsf{erg,\,mux}}(N)}{\log_2(e)\cdot p} \to 2\log_e\left(\frac{1+\sqrt{1+4\rho\,\kappa_{\mathsf{mux}}}}{2}\right) - \frac{\sqrt{1+4\rho\,\kappa_{\mathsf{mux}}}-1}{\sqrt{1+4\rho\,\kappa_{\mathsf{mux}}}+1}. \tag{50}$$

The correction term $\Delta C_{\mathsf{mux}}(N,\rho)$, defined as,

$$\Delta C_{\mathsf{mux}}(N,\rho) \triangleq \frac{C_{\mathsf{erg,\,mux}}(N,\rho) - C_{\mathsf{appx}}(N,\rho)}{C_{\mathsf{appx}}(N,\rho)} \tag{51}$$

satisfies $|\Delta C_{\mathsf{mux}}(N)| \approx \frac{\rho\kappa_{\mathsf{mux}}}{2}$. ∎

Among all channels in the multiplexing regime, the smallest value for $|\Delta C_{\mathsf{mux}}(N,\rho)|$ is achieved by the multiplexing channel corresponding to $\alpha = \alpha_{\max}$. However, in the asymptotics of $N$, $\kappa_{\mathsf{mux}}$ and $\Delta C_{\mathsf{mux}}(N,\rho)$ vanish for every channel in this regime. Thus, $C_{\mathsf{appx}}(N,\rho)$ is a good asymptotic approximation of the exact capacity at all SNRs. In contrast to the beamforming channel, when $N$ is fixed, the correction term is largest at high-SNR.

### *E. Discussion*

- ***New Tricks or Reformulation of Existing Toolkit?:*** Since the literature on MIMO capacity is extensive, it is important to understand the fundamentally novel nature of our contributions that do not fall under the purview of existing MIMO capacity results/formulas. In particular, Telatar [1], Rapajic and Popescu [73], and Verdú and Shamai [74] study the regime in RMT where the transmit and receive dimensions ($p$ and $q$) satisfy $\frac{p}{q} = \Theta(1)$, which is the context of the ideal regime (Prop. 1). The cases where either $\frac{p}{q} \to 0$ (the context of Theorem 3) and $\frac{p}{q} \to \infty$ (the context of Theorem 4) require new tools from RMT and have not been studied in the MIMO literature. In fact, the limiting eigenvalue distributions in these two cases are different from the classical Marčenko-Pastur law that characterizes MIMO capacity in the context of Prop. 1. On the other hand, while [30], [55], [56], etc., has studied the capacity of random matrices with independent

entries (that are non-identically distributed) using the implicit Stieltjes transformation approach, Theorem 1 uses a new technique to capture capacity scaling as a function of channel power scaling. None of these results follow from a well-established heuristic or from a "law of large numbers"-type argument (that is not mathematically justified or rigorous) and require recourse to sophisticated RMT tools, as can be seen from App. D.

- ***Accuracy of*** $C_{\mathsf{appx}}(N,\rho)$***:*** Our analysis also produces a tractable capacity approximation, $C_{\mathsf{appx}}(N,\rho)$, and we now study its accuracy via numerical studies. In Fig. 2, we consider a sparse channel with $D = N$ ($\gamma = 1$) and plot the capacities of beamforming ($p = 1$), ideal ($p = \sqrt{N}$) and multiplexing ($p = N$) channels as a function of $N$ for $0, 5$ and $15$ dB. Three curves are plotted in each figure: i) the Monte Carlo estimates of capacity, ii) the RMT estimates proved in Sec. IV-D, iii) $C_{\mathsf{appx}}(N,\rho)$. Fig. 2 illustrates the closeness of RMT estimates with the Monte Carlo estimates. The plot also shows that while $C_{\mathsf{appx}}(N,\rho)$ is reasonably accurate across all SNRs, consistent with our theoretical results, the mismatch between $C_{\mathsf{appx}}(N,\rho)$ and $C_{\mathsf{erg}}(N,\rho)$ is largest at low-, medium- and high-SNR's for the beamforming, ideal and multiplexing channels, respectively. Further, as shown in Sec. IV-D, the mismatch vanishes in the limit of large $N$ for both beamforming and multiplexing channels; the mismatch vanishes for the ideal channel only in the limit of low- or high-SNR.
- ***Optimal Sparse Channel Configuration:*** While further study could be made rigorous by using the RMT estimates of capacity from Sec. IV-D, motivated by the reasonable accuracy of $C_{\mathsf{appx}}(N,\rho)$ across all channel configurations and SNRs and with an aim to keep the ensuing analysis simple, we will henceforth use $C_{\mathsf{appx}}(N,\rho)$ in our study of capacity.

*Proposition 2:* For any given channel power/DoF scaling law $\rho_c = D = N^\gamma$, $\gamma \in (0,2]$, and transmit SNR of $\rho$, the ideal channel characterized by

$$p_{\mathsf{id}} \approx \sqrt{\frac{\rho D}{4}} = \frac{\sqrt{\rho}}{2} N^{\gamma/2} \tag{52}$$

$$q_{\mathsf{id}} \approx \sqrt{\frac{4D}{\rho}} = \frac{2}{\sqrt{\rho}} N^{\gamma/2}\ ,\ \kappa_{\mathsf{id}} = \frac{q_{\mathsf{id}}}{p_{\mathsf{id}}} \tag{53}$$

maximizes the capacity among all possible channel configurations at the given $\rho$. Moreover, the ideal channel achieves the fundamental limit in Theorem 2 and is optimal from a capacity scaling perspective

$$\boxed{C_{\mathsf{erg,\,id}}(N,\rho) \approx \sqrt{D}\log(1+\rho\kappa_{\mathsf{id}})\mathcal{O}\big(\sqrt{D}\big) = \mathcal{O}\left(N^{\gamma/2}\right)}$$

*Proof:* Consider the first and second derivatives of $C_{\mathsf{appx}}(N,\rho)$ with respect to $p$. Defining $x \triangleq \frac{\rho\cdot\rho_c}{p^2} = \frac{\rho N^\gamma}{p^2}$, the derivatives turn out to be

$$\dot{C}_{\mathsf{appx}}(N,\rho) \triangleq \frac{d}{dp} C_{\mathsf{appx}}(N,\rho) \tag{54}$$

$$= \log_2(e)\cdot\left(\log_e(1+x) - \frac{2x}{1+x}\right) \tag{55}$$

$$\ddot{C}_{\mathsf{appx}}(N,\rho) \triangleq \frac{d}{dp}\dot{C}_{\mathsf{appx}}(N,\rho) \tag{56}$$

$$= \log_2(e)\cdot\frac{2x(1-x)}{(1+x)^2\ p}. \tag{57}$$

We note that $\dot{C}_{\mathsf{appx}}(N,\rho) < 0$ for $x < x_0$ and $\dot{C}_{\mathsf{appx}}(N,\rho) > 0$ for $x > x_0$ where $x_0 \approx 4$, while $\ddot{C}_{\mathsf{appx}}(N,\rho) > 0$ for $x < 1$ and $\ddot{C}_{\mathsf{appx}}(N,\rho) < 0$ for $x > 1$. Thus, the capacity is maximized at $x_0 = \frac{\rho N^\gamma}{p^2} = \frac{\rho D}{p^2} \approx 4$, from which the expressions for the pair ($p_{\mathsf{id}}$, $q_{\mathsf{id}}$) in (53) follow. Equation (54) directly follows from (53). ■

In particular, it is easy to see that since $p = o(q)$ in the beamforming regime and $q = o(p)$ in the multiplexing regime, we have

$$\frac{C_{\mathsf{erg,\,bf}}(N,\rho)}{C_{\mathsf{erg,\,id}}(N,\rho)} \overset{N\to\infty}{\to} 0, \quad \text{and} \tag{58}$$

$$\frac{C_{\mathsf{erg,\,mux}}(N,\rho)}{C_{\mathsf{erg,\,id}}(N,\rho)} \overset{N\to\infty}{\to} 0. \tag{59}$$

The optimality of the ideal channel configuration from a scaling perspective is illustrated in Fig. 3 for $D = N$ ($\gamma = 1$) via the approximate capacity expression. The capacity is plotted on a log-log scale as a function of $N$ for $\rho = 0, 5$ and $15$ dB. At low-SNR, the beamforming channel initially dominates for small $N$ as

$$C_{\mathsf{erg,\,mux}}(N,\rho) < C_{\mathsf{erg,\,id}}(N,\rho) < C_{\mathsf{erg,\,bf}}(N,\rho), \tag{60}$$

but the ideal channel eventually dominates as $N$ gets large. On the other hand, at high-SNR, the multiplexing channel initially dominates as

$$C_{\mathsf{erg,\,bf}}(N,\rho) < C_{\mathsf{erg,\,id}}(N,\rho) < C_{\mathsf{erg,\,mux}}(N,\rho), \tag{61}$$

but the ideal channel again eventually dominates. Note that the curves for $\rho = 5$ dB are close to the optimal choice of $p_{\mathsf{id}} = \sqrt{N}$ according to (53) and hence, the ideal channel dominates the beamforming and multiplexing channels for all values of $N$. In general, the value of $N$ at which the ideal channel becomes dominant is a function of $\rho$ and $\gamma$. The following section will explore this in more detail.

## V. Capacity as a function of SNR

We now study the impact of different sparse channel configurations defined in Sec. IV-C on capacity as a function of SNR. For this discussion, we assume that $N$ is large enough so that $C_{\mathsf{appx}}(N,\rho)$ captures $C_{\mathsf{erg}}(N,\rho)$ in all the three regimes sufficiently accurately.

The approximate capacity formula reveals a fundamental tradeoff between the number of parallel channels $p$ and the received SNR per parallel channel $\rho_r/p = \rho\cdot\frac{q}{p}$. The impact

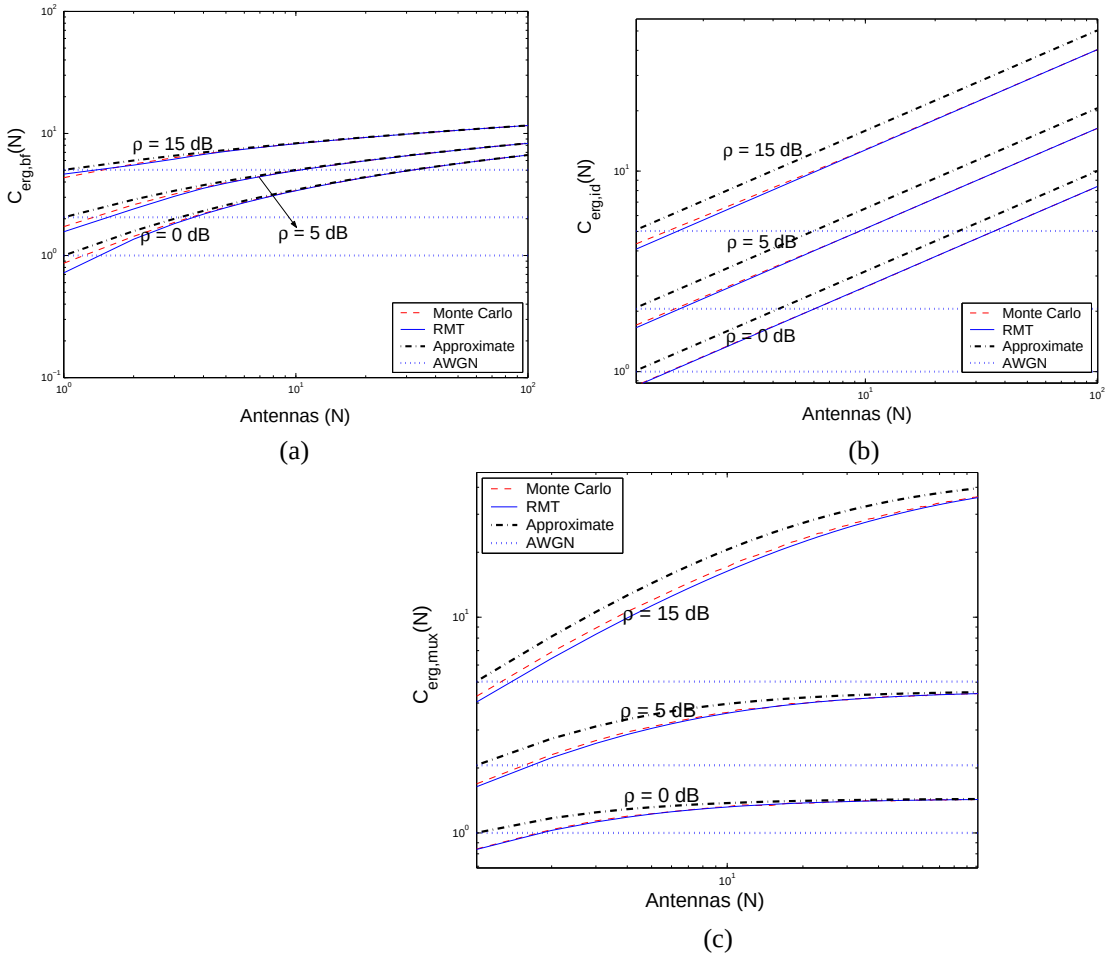


Fig. 2. Exact capacity expressions and estimates for the family of channels in Definition 4.

of this tradeoff on ergodic capacity is easily illustrated via a numerical study. Fig. 4 plots $C_{\mathsf{appx}}(N,\rho)$ as a function of $\rho$ for $N = 25$ and $D = N$ ($\gamma = 1$). The plot considers three configurations – beamforming ($p = p_{\min} = 1$), ideal ($p = p_{\mathsf{id}} = \sqrt{N}$), and multiplexing ($p = p_{\max} = N$) – as well as the capacities of 10 configurations corresponding to equally spaced values of $\alpha \in [0,1]$ via $p = N^{\alpha}$. That is,

$$C_{\mathsf{appx},\,\alpha}(N,\rho) \triangleq N^{\alpha}\log\left(1+\rho N^{\gamma-2\alpha}\right). \tag{62}$$

Fig. 4 shows that there exist SNRs, $\rho_{\mathsf{low}}$ and $\rho_{\mathsf{high}}$, so that the beamforming channel yields highest capacity for $\rho \le \rho_{\mathsf{low}}$ and the multiplexing channel yields highest capacity for $\rho \ge \rho_{\mathsf{high}}$. In the medium-SNR range, $\rho_{\mathsf{low}} < \rho < \rho_{\mathsf{high}}$, the optimal (capacity-maximizing) channel configuration continuously transitions from beamforming to multiplexing channel through the ideal configuration. Furthermore, while the beamforming and multiplexing channels exchange roles in the low- and high-SNR regimes, the ideal channel is a robust choice whose capacity lies between the two extremes. The optimal channel configuration at any SNR optimizes the $p$ vs. $\rho_r/p$ trade-off for maximizing capacity.

The next result builds on Prop. 2 and characterizes the optimal capacity-maximizing channel configuration as a function of $\rho$.

*Proposition 3:* For a large $N$ such that $C_{\mathsf{appx}}(N,\rho)$ is sufficiently accurate, let $\rho_c = D = N^{\gamma}$, $\gamma \in (0,2]$ and define $\alpha^{\star} \triangleq \min(\gamma, 2-\gamma)$. The optimal capacity-maximizing channel configuration at any $\rho$ is given by

$$p_{\mathsf{opt}}(\rho) \approx \begin{cases} N^{\alpha_{\min}}, & \rho < \rho_{\mathsf{low}} \\ \frac{\sqrt{\rho D}}{2} = \frac{\sqrt{\rho}}{2}N^{\gamma/2}, & \rho \in [\rho_{\mathsf{low}}, \rho_{\mathsf{high}}] \\ N^{\alpha_{\max}}, & \rho > \rho_{\mathsf{high}} \end{cases} \tag{63}$$

where $\rho_{\mathsf{low}} \triangleq \frac{4}{N^{\alpha^{\star}}}$, and $\rho_{\mathsf{high}} \triangleq 4\,N^{\alpha^{\star}}$.

*Proof:* Recall the two different feasible ranges for $p$ depending on $\gamma$. For $\gamma \in (0,1)$, $p$ varies from $p_{\min} = 1$ to $p_{\max} = N^{\gamma}$. For $\gamma \in [1,2]$, $p$ varies from $p_{\min} = N^{\gamma-1}$ to $p_{\max} = N$. The first derivative of $C_{\mathsf{appx}}(N,\rho)$ with respect to $p$ vanishes when

$$\frac{\rho D}{p^2} = \frac{\rho N^{\gamma}}{p^2} \approx 4 \tag{64}$$

which characterizes the capacity-maximizing point. Solving $\rho_{\mathsf{low}} = 4p_{\min}^2/N^{\gamma}$, $p_{\min} = N^{\alpha_{\min}}$, it is easy to check that a common representation for $\rho_{\mathsf{low}}$ in both cases above is $\rho_{\mathsf{low}} = \frac{4}{N^{\alpha^{\star}}}$. Similarly, solving $\rho_{\mathsf{high}} = 4p_{\max}^2/N^{\gamma}$, $p_{\max} = N^{\alpha_{\max}}$, yields the value of $\rho_{\mathsf{high}}$ stated in the proposition in both cases. In the intermediate regime, $p_{\mathsf{opt}}$ directly follows from (64). ■

Using Prop. 3, we get the theoretical estimates $\rho_{\mathsf{low}} \approx -8$ dB and $\rho_{\mathsf{high}} \approx 20$ dB for the channel configurations ($N =$

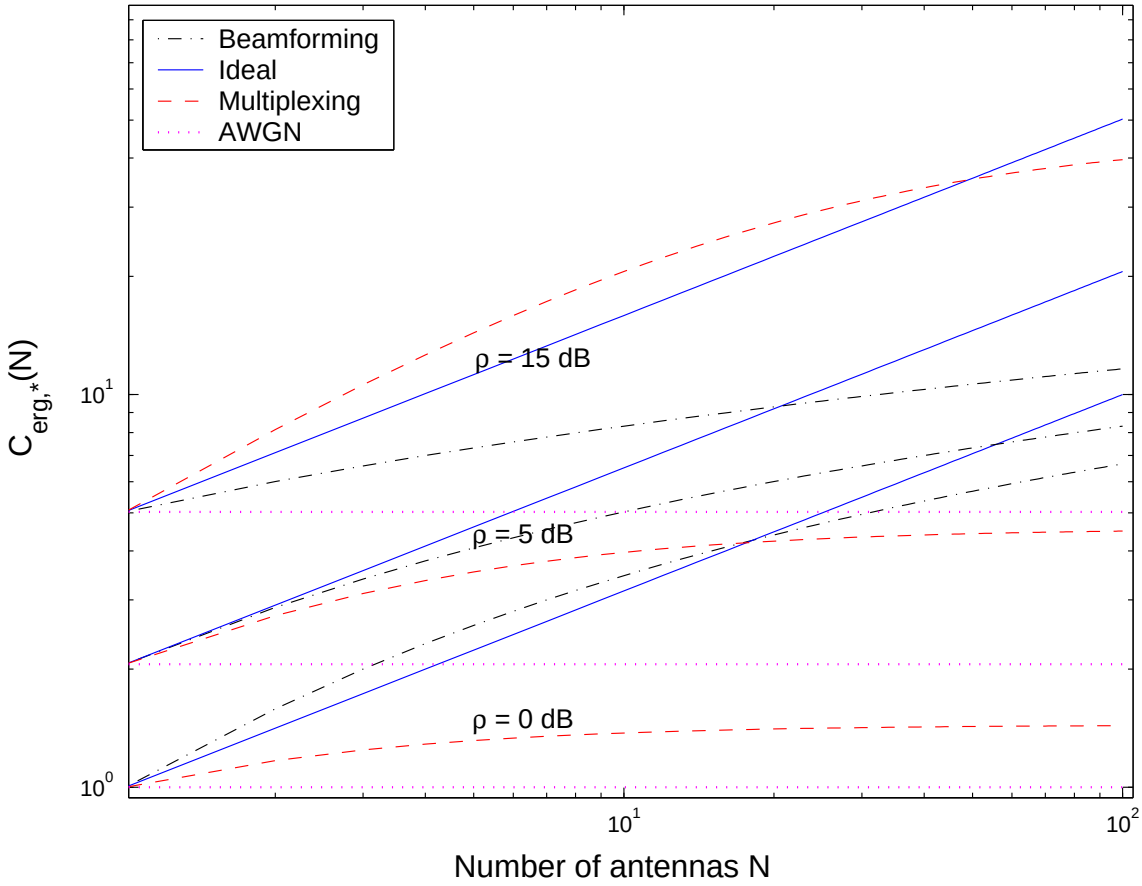


Fig. 3. Capacities of the beamforming, ideal, and multiplexing channel configurations as a function of $N$.

25, $\gamma = 1$) illustrated in Fig. 4. In the figure, we observe that the beamforming channel is indeed optimal for $\rho < -8$ dB and the multiplexing channel is optimal for $\rho > 20$ dB. The value of $p_{\mathsf{opt}}$ characterized in Prop. 3 for the medium-SNR range, $\rho \in [\rho_{\mathsf{low}}, \rho_{\mathsf{high}}]$, is illustrated by the dotted curves in the figure which plot $C_{\mathsf{appx},\,\alpha}(N, \rho)$ for 10 equally spaced values of $\alpha \in [0, 1]$. For each intermediate-SNR, there is a $C_{\mathsf{appx},\,\alpha}(N, \rho)$ curve that yields the maximum capacity.

While Prop. 3 states the precise value of $p_{\mathsf{opt}}$ for any $\rho \in [\rho_{\mathsf{low}}, \rho_{\mathsf{high}}]$, the ideal channel configuration with $p_{\mathsf{id}} = N^{\gamma/2}$ serves as a robust *fixed* configuration in this intermediate range. Thus, for all practical purposes, the three canonical configurations – beamforming ($p = p_{\min}$), ideal ($p = p_{\mathsf{id}}$) and multiplexing ($p = p_{\max}$) – accurately approximate the capacity-maximizing configuration over the entire range. That is,

$$\tilde{p}_{\mathsf{opt}}(\rho) = \begin{cases} p_{\min} = N^{\alpha_{\min}} & \rho < \tilde{\rho}_{\mathsf{low}} \\ p_{\mathsf{id}} = N^{\gamma/2} & \tilde{\rho}_{\mathsf{low}} \leq \rho \leq \tilde{\rho}_{\mathsf{high}} \\ p_{\max} = N^{\alpha_{\max}} & \rho > \tilde{\rho}_{\mathsf{high}} \end{cases} \tag{65}$$

where $\tilde{\rho}_{\mathsf{low}}$ and $\tilde{\rho}_{\mathsf{high}}$ are the solutions to the following equations

$$N^{\alpha_{\min}} \log(1 + \tilde{\rho}_{\mathsf{low}} N^{\gamma - 2\alpha_{\min}}) = N^{\gamma/2} \log(1 + \tilde{\rho}_{\mathsf{low}}) \tag{66}$$

$$N^{\alpha_{\max}} \log(1 + \tilde{\rho}_{\mathsf{high}} N^{\gamma - 2\alpha_{\max}}) = N^{\gamma/2} \log(1 + \tilde{\rho}_{\mathsf{high}}). \tag{67}$$

The SNRs $\tilde{\rho}_{\mathsf{low}}$ and and $\tilde{\rho}_{\mathsf{high}}$ are also illustrated in Fig. 4.

Note that the ratio $\frac{\rho_{\mathsf{high}}}{\rho_{\mathsf{low}}}$ attains its largest value, $N^2$, for $\gamma = 1$ ($D = N$), whereas it achieves its minimum value of unity for $\gamma = 0$ ($D = 1$) or $\gamma = 2$ ($D = N^2$). Thus, the $p$ vs. $\rho_r/p$ tradeoff that determines $p_{\mathsf{opt}}$ does not exist for the extreme cases of highly correlated ($\gamma = 0$) and i.i.d. ($\gamma = 2$) channels. Note that in either case all the three configurations reduce to the same $\mathbf{H}$. On the other hand, the impact of this tradeoff on capacity is maximum for $\gamma = 1$ corresponding to $\rho_c = D = N$.

## VI. Connections to Related Results

### A. Realizing the Channel Configurations in Practice

In this paper, we have analyzed the impact of distribution of the $D < N^2$ DoF in sparse channels on ergodic capacity. Our results provide information-theoretic benchmarks on optimal channel configurations from the viewpoint of scaling laws as well as capacity as a function of SNR for fixed, but large $N$. In practice, however, the configuration of the DoF is determined by the AoAs $\{\theta_{r,\,\ell}\}$ and AoDs $\{\theta_{t,\,\ell}\}$ of propagation paths in the scattering environment. Thus, a natural question is whether the theoretical results in this paper can be leveraged in practice for capacity gains in physically sparse channels. This question is addressed in our related paper [76] where we show that in a sparse scattering environment with randomly distributed (over the angular spreads) paths, the antenna spacings at

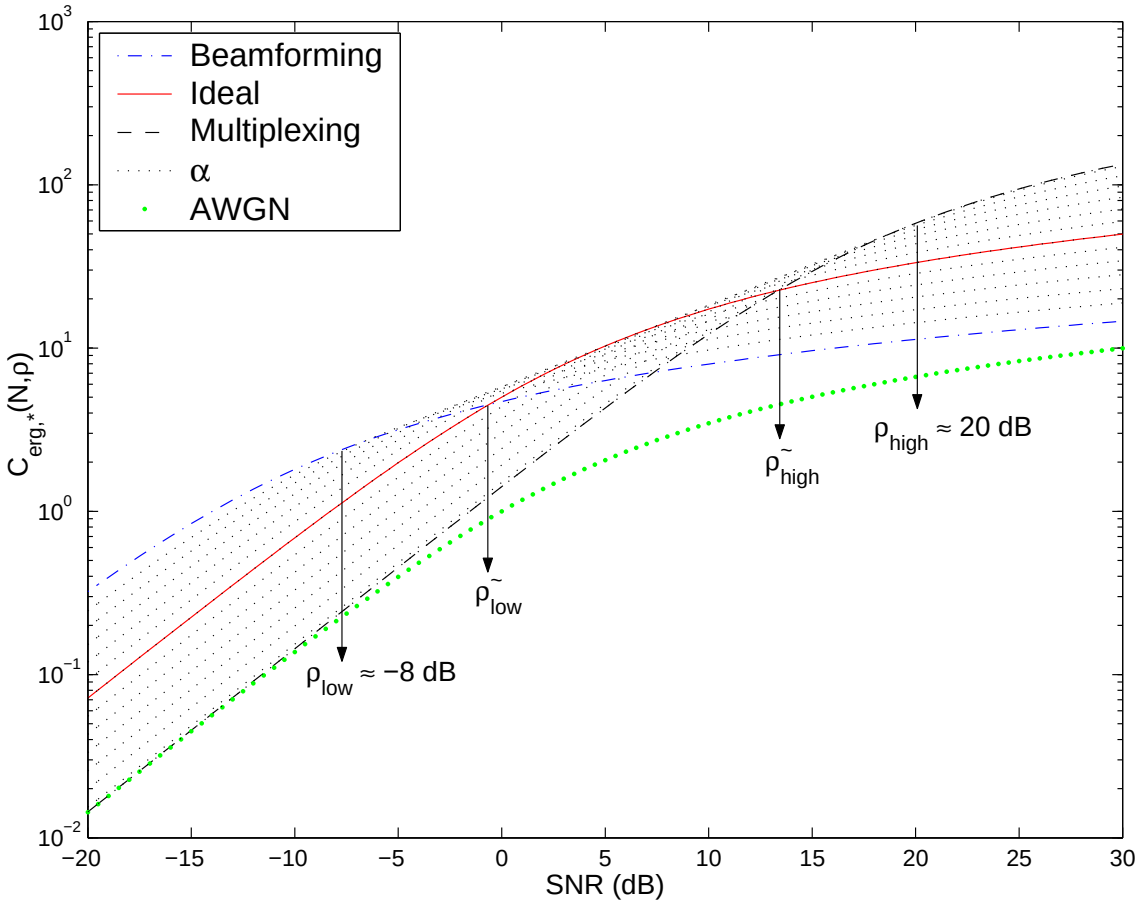


Fig. 4. Capacity as a function of $\rho$ for the different sparse channel configurations.

the transmitter and receiver can be systematically adapted to the level of sparsity and SNR to realize the different channel configurations discussed in this paper. In particular, the multiplexing channel corresponds to maximal antenna spacings at both transmitter and receiver, the ideal channel is realized via medium antenna spacings at both transmitter and receiver, and the beamforming channel is realized via closely spaced antennas at transmitter and large spacings at receiver. The notion of small, medium and large spacings is quantified in [76] and the readers are referred to it for more details.

### *B. Recent Works on Antenna Design and Large-Dimensional Systems*

A recent work by Barriac *et al.* [77] provides insight into the problem of optimizing antenna spacing at the transmitter for maximizing capacity in correlated MIMO channels. This insight is based on studying the variations in the channel's eigenvalues as a function of antenna spacings. It is argued in [77] that in the low-SNR setting, spacing the transmit antennas sufficiently close to each other so as to excite only one channel eigenmode is the optimal strategy while in the high-SNR regime, choosing the spacings sufficiently large so as to excite all the channel eigenmodes is optimal from a capacity perspective. The authors further conjecture that the optimal number of eigenmodes to be excited in the medium-SNR regime is a monotonic function of SNR. In this paper, we have provided a systematic investigation of this problem in the context of correlated channels in which the source of correlation is sparsity of multipath. The number of parallel channels $p$ is equivalent to the number of channel eigenmodes excited, and the beamforming and multiplexing channel configurations are characterizations of the low- and high-SNR analogues of [77].

Another recent work [78] discusses techniques for optimizing antenna locations in a fixed volume to maximize spectral efficiency in low-SNR. Design guidelines offered in [78] provide evidence to the optimality of closely spaced antennas in the low-SNR regime. Evidence for optimality of closely spaced transmit antennas in the low-SNR regime also comes from a recent study [79] on error exponents in correlated MIMO channels. This work shows that in the low-SNR setting, a fully correlated channel yields higher error exponents (reliability), thereby suggesting that closely spaced antennas may be more desirable in such scenarios.

A recent work [80] shows that in the multi-user setting and in the limit of an infinite number of base station antennas, powerful benefits such as: i) arbitrarily small transmit energy per bit, ii) disappearing effect of uncorrelated noise and fast fading, iii) total throughput per cell that is independent of cell-site, etc. can be accrued. It is important to note that most

of these benefits critically hinge on statistical knowledge at either end. However, as motivated in Sec. II-B, this is an optimistic assumption, especially in the large antennas regime. In contrast to [80], our work corresponds to the more realistic scenario of including the overhead of statistics estimation. The lack of knowledge of the independent DoF in $\mathbf{H}$ and their locations can also be addressed from a compressive sensing framework [81], [82].

### C. MMSE Estimation and Mutual Information

An underlying connection between MMSE estimation and mutual information is obtained in [30], [83] from the implicit capacity characterization of (22) by defining $\Gamma(t)$ and $\Upsilon(t)$:

$$\Gamma(t) \triangleq \kappa_{\mathsf{TLV}}\, \mathbf{E}_{\mathcal{R}}\left[\mathcal{G}(\mathcal{R},t)\mathcal{D}(\mathcal{R})\right] \quad (68)$$

$$\stackrel{(a)}{=} \kappa_{\mathsf{TLV}}\, \mathbf{E}_{\mathcal{R}}\left[\frac{\mathcal{G}(\mathcal{R},t)}{1+\mathbf{E}_{\mathcal{T}}\left[\mathcal{G}(\mathcal{R},\mathcal{T})\Upsilon(\mathcal{T})|\mathcal{R}\right]}\right] \quad (69)$$

$$\Upsilon(t) \triangleq \frac{\rho}{1+\rho\Gamma(t)} \quad (70)$$

where the equality in (a) follows from the analysis in [30]. It can be seen that the quantity $\rho\,\Gamma(t)$ corresponds to the SNR at the output of a linear MMSE receiver for the signal transmitted from the corresponding transmit dimension whereas the corresponding MSE is seen to be $\frac{\Upsilon(t)}{\rho}$. In our setting, the regularity of the family of channels in Definition 4 can be exploited to obtain more insights on the trade-off between number of data-streams and the MSE of the individual data-streams.

*Proposition 4:* Consider the family of channels introduced in Definition 4 with $D = qp$ and $\kappa = \frac{q}{p}$. In the beamforming and the ideal regimes under the assumption that $\kappa > 1$, $\Gamma(t)$ satisfies

$$\left\{\Gamma_{\mathsf{bf}}(t),\, \Gamma_{\mathsf{id},\,\{\kappa>1\}}(t)\right\} = \frac{\rho\kappa-\rho-1+\sqrt{(\rho\kappa-\rho-1)^2+4\rho\kappa}}{2\rho}. \quad (71)$$

On the other hand, in the multiplexing and the ideal regimes under the assumption that $\kappa \leq 1$, $\Gamma(t)$ is given by

$$\left\{\Gamma_{\mathsf{mux}}(t),\, \Gamma_{\mathsf{id},\,\{\kappa\leq 1\}}(t)\right\} = \frac{\sqrt{1+4\rho\kappa}-1}{2\rho}. \quad (72)$$

*Proof:* See Appendix E. ∎

Since $\kappa \to 0$ in the multiplexing regime, $\rho\,\Gamma_{\mathsf{mux}}(t) \to \rho\kappa = \rho\,\frac{q}{p}$. That is, the SNR at the output of a linear MMSE receiver is the received SNR per parallel channel and the mean-squared error at the receiver is $\frac{1}{1+\rho_r/p}$. By letting $\kappa \to \infty$ in the beamforming regime, it is straightforward to see that $\rho\,\Gamma_{\mathsf{bf}}(t) \to \rho\,\frac{q}{p}$ for all $\rho$. Similarly, as $\rho \to 0$, a Taylor's series analysis of $\Gamma(t)$ in both cases of the ideal regime shows that $\rho\,\Gamma_{\mathsf{id}}(t) \to \rho\kappa$. The corresponding limits in the high-SNR regime are $\Gamma_{\mathsf{id},\,\{\kappa>1\}}(t) \to \rho(\kappa-1)$ and $\Gamma_{\mathsf{id},\,\{\kappa\leq 1\}}(t) \to 2\sqrt{\rho\kappa}$, respectively. The convergence of $\rho\Gamma(t)$ to $\rho_r/p$ for all $\rho$ in the beamforming and multiplexing regimes is related to the tightness of $C_{\mathsf{appx}}(N,\rho)$ in these regimes.

Thus the beamforming configuration (which maximizes $\rho_r/p$) trades off the number of data-streams for the MSE of the individual data streams and is optimal at low-SNR. The multiplexing configuration corresponds to the other extreme in this trade-off and is optimal at high-SNR while the ideal configuration optimizes this *rate-distortion* tradeoff at medium-SNR.

### D. Spectral Efficiency in the Low-SNR Regime

In the low-SNR regime, the seminal work by Verdú [58] shows that the minimum energy per bit necessary for reliable communication $\frac{E_b}{N_{\mathrm{o}}}_{\min}$ and the wideband slope $S_0$ are the key figures of merit. It is a straightforward exercise to show that (see [58] and [60] for related results) $\frac{E_b}{N_{\mathrm{o}}}_{\min}$ of all the three configurations is given by a common formula

$$\frac{E_b}{N_{\mathrm{o}}}_{\min,\,\star} = \frac{\log_e(2)}{\dot{C}_{\mathsf{erg},\,\star}(N,0)} = \frac{\log_e(2)}{q} \quad (73)$$

where $\dot{C}_{\mathsf{erg},\,\star}(N,0)$ and $\ddot{C}_{\mathsf{erg},\,\star}(N,0)$ denote the first and second derivatives (w.r.t. $\rho$) of the ergodic capacity (in nats/dimension) as $\rho \to 0$. Another straightforward calculation shows that

$$S_{0,\,\star} = \frac{2\left(\dot{C}_{\mathsf{erg},\,\star}(N,0)\right)^2}{-\ddot{C}_{\mathsf{erg},\,\star}(N,0)} = \frac{2\,q^2\,p^2}{\mathbf{E}\left[\mathrm{Tr}\left((\mathbf{H}\mathbf{H}^H)^2\right)\right]}, \quad (74)$$

which can be computed as

$$S_{0,\,\mathsf{bf}} = \frac{2qp}{q+p} \approx 2p, \quad S_{0,\,\mathsf{id}} = \frac{2qp}{q+p} = p, \quad S_{0,\,\mathsf{mux}} = p. \quad (75)$$

Note that $\frac{E_b}{N_{\mathrm{o}}}_{\min,\,\star}$ is the smallest for the beamforming channel. Even though the multiplexing channel has the largest $S_0$ among the three channels, the minimum energy per bit is the dominating figure of merit at low-SNR and the larger values of $S_0$ for the ideal and multiplexing channels become irrelevant. The low-SNR capacity gain of the beamforming channel relative to the ideal and multiplexing channels is precisely a manifestation of these gains in $\frac{E_b}{N_{\mathrm{o}}}_{\min}$.

## VII. Conclusion

MIMO capacity gains in i.i.d. channels relative to single antenna systems rest on two key effects: i) higher coupling between the transmitted and received signal energy due to the larger array apertures, and ii) statistical independence between the channel coefficients. The first effect primarily impacts capacity scaling with antenna dimensions and is directly reflected in the quadratic channel power scaling assumption in existing results. While this assumption may be justified for small antenna dimensions, such scaling in $\rho_c$ is *not* sustainable indefinitely from an energy conservation viewpoint. The second effect primarily governs channel capacity as a function of SNR for a fixed number of antennas. Thus the capacity scaling laws and capacity behavior as a function of SNR are critically dependent on the channel power normalization and the number of statistically independent DoF in the channel. Virtually all existing information-theoretic studies of MIMO capacity implicitly assume a rich scattering environment with a quadratic channel power scaling.

In this paper, we have made a case for studying sparse MIMO channels with a channel power scaling that is commensurate with the sparsity level. Sparse channels introduce a feature that affects capacity behavior and one that is not seen in non-sparse channels: the spatial configuration of the $D$ dominant DoF in the available $N^2$ spatial dimensions. While classical linear scaling behavior in ergodic capacity holds when the transmitter knows the statistics of the channel completely, in the case where the transmitter only knows the distribution of the dominant DoF in the channel, we show that the ergodic capacity scales sub-linearly with antenna dimensions. We also illustrate the sparse channel structure that achieves the optimal capacity scaling behavior, both as a function of antenna dimensions and SNR. For this, we develop a capacity characterization of sparse MIMO channels using new tools from RMT. This characterization has a heuristic interpretation (in hindsight) and reveals a trade-off between the number of parallel channels and the received SNR per parallel channel.

Complementing our results on ergodic capacity, we analyze the outage capacity of sparse channels in [69]. In this context, the $p$ vs. $\rho_r/p$ tradeoff manifests itself as a rate-reliability trade-off. Consistent with ergodic capacity results, the beamforming channel is outage optimal at low-SNR, the multiplexing channel at high-SNR, and the ideal channel at medium-SNR. One promising direction for future research is to investigate the potential of reconfigurable antenna arrays in cognitive wireless systems, as prompted by our related paper [76] in which it is shown that the different channel configurations can be realized in practice by systematically adapting the antenna spacings at the transmitter and receiver to the level of sparsity and operating SNR.

## ACKNOWLEDGMENT

The authors would like to thank Prof. Ada Poon for discussion on motivation of the paper, Prof. Marco Donald Migliore and Dr. Michel Ivrlac for discussion on channel power normalization issues, and Dr. Michail Matthaiou and the reviewers for their feedback on the paper.

## APPENDIX

### A. Proof of Theorem 1

***Preliminary results:*** We start with a trivial inequality.

*Lemma 1:* Let $z \geq 0$. We then have

$$\frac{z}{z+1} \;\leq\; \log_e(1+z) \;\leq\; z. \tag{76}$$

■

We define the empirical eigenvalue distribution (EED) function as follows.

*Definition 5:* The EED of an $n \times n$ Hermitian matrix $\mathbf{X}$ (denoted as $F_{\mathbf{X}}(\cdot)$) is defined as

$$F_{\mathbf{X}}(\lambda) \triangleq \frac{1}{n}\Big|\left\{i \,:\, \lambda_i(\mathbf{X}) \leq \lambda\right\}\Big|, \tag{77}$$

where $\big|\cdot\big|$ denotes the cardinality of the set under consideration. The density function corresponding to the EED is called the empirical eigenvalue density function. ■

We need the following spectral characterization from random matrix theory.

*Lemma 2:* ***Yin, Bai and Krishnaiah, 1988 [84].*** Consider a family of random matrices of increasing dimension. Let $\mathbf{H}$ be an $N \times N$ random matrix with independent entries satisfying $\mathbf{H}(m,n) \sim \mathcal{CN}(0, \mathbf{\Psi}(m,n))$ and $\max_{m,n} \mathbf{\Psi}(m,n) \leq K < \infty$, independent of $N$. Let $\lambda_{\max}$ denote the largest eigenvalue of $\mathbf{H}\mathbf{H}^H$. Then,

$$\mathsf{Pr}\left(\frac{\lambda_{\max}}{N} > 4K\right) = o\left(\frac{1}{N^\ell}\right) \text{ for all } \ell \geq 1. \tag{78}$$

*Remark 1:* Lemma 2 is a slight generalization of the following result that is more-often used in MIMO capacity analysis [1], [73], [74]: the largest eigenvalue of $\mathbf{H}\mathbf{H}^H$ where $\mathbf{H}$ is an i.i.d. random matrix, with entries that are mean zero and variance $\sigma^2$, is almost surely equal to $4\sigma^2$. The lemma can also be restated as: if $\mathbf{H}$ is a random matrix with independent entries having bounded variance, and if the EED function of $\frac{\mathbf{H}\mathbf{H}^H}{N}$ converges, then the limit distribution function has compact support.

*Proof:* We briefly summarize the proof from [84] to keep the flow of this paper self-contained. Without loss in generality, we will henceforth assume that $K = 1$. It is easy to see that for any $\mu > 0$ and all integers $m \geq 1$

$$\mathsf{Pr}\left(\frac{\lambda_{\max}}{N} > 4 + \mu\right) \leq \frac{\mathbf{E}\left[\mathrm{Tr}\left((\mathbf{H}\mathbf{H}^H)^m\right)\right]}{[(4+\mu)N]^m}. \tag{79}$$

The "Truncation lemma" of [84, p. 556] shows that there exists a sequence of positive numbers $\delta_N$ and $A_N$ such that $\delta_N \to 0$, $A_N \to \infty$, $m = A_N \log(N)$, $\delta_N N^{1/3} \to 0$, and $A_N \left(\delta_N\right)^{1/6} \to 0$ as $N \to \infty$. With this choice of $m$ and $\delta_N$, it can be shown that (see [84] for details),

$$\mathsf{Pr}\left(\frac{\lambda_{\max}}{N} > 4 + \mu\right) \leq N^2 \left(1 + 2\delta_N^{1/2}\right)^{2m} \left[\frac{4}{4+\mu}\right]^m \tag{80}$$

$$\to N^2 \exp\left(4m\delta_N^{1/2}\right) \left[\frac{4}{4+\mu}\right]^m. \tag{81}$$

With the above choice of $m$, it is easy to check that $m\delta_N^{1/2} \to 0$, and $N^{\ell+2}\left[\frac{4}{4+\mu}\right]^m \to 0$ for all $\ell$, and any $\mu > 0$. This completes the proof of Lemma 2. ■

Using Lemma 2, we now prove the following proposition.

*Proposition 5:* Let $\mathbf{H}$ be a random matrix as defined in Lemma 2. Then, $\mathbf{E}\left[\lambda_{\max}\, \chi\left(\frac{\lambda_{\max}}{N} > 4\right)\right]$ can be made arbitrarily small for sufficiently large $N$.

*Proof:* We use the fact [85, p. 50] that for a positive random variable $X$, the mean can be bounded as

$$\mathbf{E}\left[X\right] \leq \sum_{k=0}^{\infty} \mathsf{Pr}\left(X \geq k\right). \tag{82}$$

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

$$\begin{aligned}
&\mathbf{E}\left[\frac{\lambda_{\max}^2}{N^2}\right] \\
&\leq \sum_{k=0}^{35} \mathsf{Pr}\left(\frac{\lambda_{\max}^2}{N^2} > k\right) + \sum_{k=36}^{\infty} \mathsf{Pr}\left(\frac{\lambda_{\max}^2}{N^2} > k\right) && (83)\\
&\leq 36 + \sum_{k=6}^{\infty} N^2 \left(1 + 2\delta_N^{1/2}\right)^{2m} \left(\frac{4}{k}\right)^m (2k+1) && (84)\\
&\leq 36 + 2N^2 \left(1 + 2\delta_N^{1/2}\right)^{2m} 4^m \int_5^{\infty} \frac{\mathrm{d}x}{x^{m-1}} && (85)\\
&= 36 + 2N^2 \left(1 + 2\delta_N^{1/2}\right)^{2m} \frac{4^m}{5^{m-2}(m-2)}. && (86)
\end{aligned}$$

It is easy to see that for the choice of $m$ as in Lemma 2, the second term in the above bound tends to 0 as $N \to \infty$ and that we have proved $\mathbf{E}\left[\frac{\lambda_{\max}^2}{N^2}\right] \leq C < \infty$. The proof of Prop. 5 follows from the Schwarz's inequality,

$$\begin{aligned}
&\mathbf{E}\left[\lambda_{\max}\ \chi\left(\frac{\lambda_{\max}}{N} > 4\right)\right] \\
&\qquad \leq \left(\mathbf{E}\left[\lambda_{\max}^2\right]\right)^{1/2} \left[\mathsf{Pr}\left(\frac{\lambda_{\max}}{N} > 4\right)\right]^{1/2}, && (87)
\end{aligned}$$

and Lemma 2. ∎

***Proof of Theorem 1:*** We are now prepared to prove Theorem 1. For a given fixed $\delta > 0$, following Prop. 5, let $N(\delta)$ be such that

$$\mathbf{E}\left[\lambda_{\max}\ \chi\left(\frac{\lambda_{\max}}{N} > 4\right)\right] \leq \delta \text{ for all } N \geq N(\delta). \quad (88)$$

Note that from Lemma 1, we can bound $I_{\mathsf{unif}}(N,\rho)$ as follows:

$$\begin{aligned}
&\log_2(e) \cdot \sum_{i=1}^{N} \mathbf{E}\left[\frac{\frac{\rho\lambda_i}{N}}{1+\frac{\rho\lambda_i}{N}}\right] \leq I_{\mathsf{unif}}(N,\rho) \\
&\qquad \leq \log_2(e) \cdot \mathbf{E}\left[\sum_{i=1}^{N} \frac{\rho\lambda_i}{N}\right] && (89)
\end{aligned}$$

$$\begin{aligned}
&\log_2(e) \cdot \sum_{i=1}^{N} \mathbf{E}\left[\frac{\frac{\rho\lambda_i}{N}}{1+\frac{\rho\lambda_i}{N}}\right] \leq I_{\mathsf{unif}}(N,\rho) \\
&\qquad \leq \log_2(e) \cdot \frac{\rho}{N} \cdot \mathbf{E}\left[\mathsf{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right]. && (90)
\end{aligned}$$

For the lower bound, observe that if $N \geq N(\delta)$, we have

$$\begin{aligned}
&\sum_{i=1}^{N} \mathbf{E}\left[\frac{\frac{\rho\lambda_i}{N}}{1+\frac{\rho\lambda_i}{N}}\right] \\
&\geq \rho \cdot \sum_{i=1}^{N} \mathbf{E}\left[\frac{\frac{\lambda_i}{N}}{1+\frac{\rho\lambda_i}{N}}\ \chi\left(\frac{\lambda_{\max}}{N} \leq 4\right)\right] && (91)\\
&\geq \frac{\rho}{1+4\rho} \cdot \sum_{i=1}^{N} \mathbf{E}\left[\frac{\lambda_i}{N}\ \chi\left(\frac{\lambda_{\max}}{N} \leq 4\right)\right] && (92)\\
&= \frac{\rho}{1+4\rho} \cdot \sum_{i=1}^{N} \mathbf{E}\left[\frac{\lambda_i}{N}\left(1 - \chi\left(\frac{\lambda_{\max}}{N} > 4\right)\right)\right] && (93)\\
&\geq \frac{\rho}{1+4\rho} \cdot \left[\left(\sum_{i=1}^{N} \mathbf{E}\left[\frac{\lambda_i}{N}\right]\right) - \delta\right]. && (94)
\end{aligned}$$

Since $\delta > 0$ is arbitrary, Theorem 1 follows. ∎

### *B. Implicit Assumption of Quadratic Scaling of Channel Power in [4]*

With a general $N_t$ and $N_r$, an implicit assumption prevalent in existing works is that $\rho_c = \Theta(N_t N_r)$. We now show how one such popular result assumes this condition implicitly.

Chuah *et al.* consider a Kronecker model for $\mathbf{H}$ where $\mathbf{H}$ is given as

$$\mathbf{H} = \mathbf{U}_r\ \mathbf{\Lambda}_r^{1/2}\ \mathbf{H}_{\mathsf{iid}}\ \mathbf{\Lambda}_t^{1/2}\ \mathbf{U}_t^H \quad (95)$$

with $\mathbf{H}_{\mathsf{iid}}$ being an i.i.d. matrix for some positive-definite diagonal matrices $\mathbf{\Lambda}_t$ and $\mathbf{\Lambda}_r$. They study MIMO capacity under the assumption that the EED functions of $\mathbf{\Lambda}_t$ and $\mathbf{\Lambda}_r$ converge in distribution (as dimensions increase) to a non-trivial distribution. From Definition 5, the assumptions on $\mathbf{\Lambda}_t$ and $\mathbf{\Lambda}_r$ imply that there exist positive numbers $\lambda_{t,0}$ and $\lambda_{r,0}$, $\{\epsilon_t, \epsilon_r\} > 0$, and integers $N_{t,0}$ and $N_{r,0}$ such that

$$\left|\{m : \mathbf{\Lambda}_r(m) > \lambda_{r,0}\}\right| > \epsilon_r N_r \text{ for all } N_r \geq N_{r,0}, \quad (96)$$

$$\left|\{n : \mathbf{\Lambda}_t(n) > \lambda_{t,0}\}\right| > \epsilon_t N_t \text{ for all } N_t \geq N_{t,0}. \quad (97)$$

Thus for all $N_t$ and $N_r$ sufficiently large, we have

$$\sum_m \mathbf{\Lambda}_r(m) > \lambda_{r,0} \cdot \epsilon_r \cdot N_r \quad \text{and} \quad (98)$$

$$\sum_n \mathbf{\Lambda}_t(n) > \lambda_{t,0} \cdot \epsilon_t \cdot N_t, \quad (99)$$

and hence,

$$\begin{aligned}
\rho_c &= \sum_{m,n} \mathbf{\Psi}(m,n) = \sum_{m,n} \mathbf{\Lambda}_r(m)\mathbf{\Lambda}_t(n) && (100)\\
&= \sum_m \mathbf{\Lambda}_r(m) \sum_n \mathbf{\Lambda}_t(n) = \Theta(N_t N_r). && (101)
\end{aligned}$$

### *C. Proof of Theorem 2*

By applying Jensen's inequality to (28) and using the fact that a diagonal $\mathbf{Q}$ $(= \mathsf{diag}\,(Q_i))$ achieves capacity for the class $\mathcal{H}(D)$ [62], [68], $C_{\mathsf{opt,\,erg}}(N,\rho)$ can be bounded as

$$\begin{aligned}
&C_{\mathsf{opt,\,erg}}(N,\rho) \\
&\quad \leq \max_{\mathbf{H} \in \mathcal{H}(D)} \max_{\mathsf{Tr}(\mathbf{Q}) \leq \rho} \log_2 \det\left(\mathbf{I} + \mathbf{Q} \cdot \mathbf{E}\left[\mathbf{H}^H\mathbf{H}\right]\right) && (102)\\
&\quad = \max_{\{P_i\}} \max_{\{Q_i\}} \sum_{i=1}^{N} \log_2\left(1 + P_i\, Q_i\right) && (103)
\end{aligned}$$

where $P_i = \sum_k \mathbf{E}\left[|\mathbf{H}(k,i)|^2\right] = \sum_k \mathbf{\Psi}(k,i)$ represent the column powers of $\mathbf{H}$. The optimization in (103) is subject to the twin constraints that

$$P_i \geq 0, \quad \textstyle\sum_{i=1}^{N} P_i = \rho_c, \quad (104)$$

$$Q_i \geq 0, \quad \textstyle\sum_{i=1}^{N} Q_i = \rho. \quad (105)$$

We now recast this optimization[10] as follows:

$$\max_{s} \max_{\{P_i\}} \max_{\{Q_i\}} \sum_{i=1}^{s} \log_2\left(1+P_i\,Q_i\right) \quad \text{s.t.} \quad 1 \le s \le N, \tag{106}$$

$$P_i > 0, \quad i = 1, \cdots, s \quad \text{and} \quad P_i = 0, \quad i > s$$

$$\text{with} \quad \sum_{i=1}^{s} P_i = \rho_c, \tag{107}$$

$$\text{and} \quad Q_i \ge 0, \quad \text{with} \quad \sum_{i=1}^{s} Q_i = \rho. \tag{108}$$

Given $s$ fixed between 1 and $N$, the inner optimization can be written in terms of Lagrange multipliers $\eta_1$ and $\eta_2$ as

$$\max_{\{P_i\}} \max_{\{Q_i\}} \left\{ \sum_{i=1}^{s} \log_2\left(1+P_i\,Q_i\right) + \eta_1 \left(\rho_c - \sum_{i=1}^{s} P_i\right) + \eta_2 \left(\rho - \sum_{i=1}^{s} Q_i\right)\right\}. \tag{109}$$

Given any $i$, the partial derivatives of the above argument with respect to $Q_i$ and $P_i$ have to vanish at the maxima. Thus we take partial derivatives of the above function with respect to $Q_i$ and $P_i$ and set the derivatives equal to 0. It is easy to check that $\frac{Q_i}{P_i}$ is independent of $i$. Using this relationship to meet $\sum_{i=1}^{s} Q_i = \rho$, we see that $Q_i = \frac{\rho}{\rho_c} \cdot P_i$. Now revisiting the partial derivative with respect to $Q_i$ under the constraint that $Q_i = \frac{\rho}{\rho_c} \cdot P_i$ and setting the derivative to 0, we get $P_i = \frac{\rho_c}{s}$. It is easy to verify that the second derivative is negative with this particular choice of $P_i$ and $Q_i$.

Using this choice of $Q_i$ in (102), we have

$$C_{\mathsf{opt,\,erg}}(N,\rho) \le \max_{s} \left\{ s \cdot \log_2\left(1+\rho\,\frac{\rho_c}{s^2}\right)\right\}. \tag{110}$$

The $s$ that achieves this maximum satisfies $s = \Theta\left(\sqrt{\rho_c}\right)$ which can be seen by setting the derivative w.r.t. $s$ to zero. This choice of $s$ leads to the statement of Theorem 2. ∎

### D. Asymptotic Ergodic Capacity Analysis

***Preliminary Results:*** We need a few asymptotic spectral characterizations to understand the ergodic capacity of the family of channels from Definition 4. The first two lemmas are needed to study capacity in the beamforming regime.

*Lemma 3:* Let $\mathbf{G}$ be a $p \times q$ complex random matrix with i.i.d. entries having mean zero and variance one. Let $\kappa_{\mathsf{bf}} = \frac{q}{p} \to \infty$ with $p$ finite. Then

$$\left\| \frac{\mathbf{G}\mathbf{G}^H}{q} - \mathbf{I}_p \right\| = \mathcal{O}_{\mathsf{prob.}}\left(\frac{1}{\sqrt{q}}\right) \tag{111}$$

where $\|\cdot\|$ refers to a well-defined matrix norm and the notation $\mathcal{O}_{\mathsf{prob.}}\left(\frac{1}{\sqrt{q}}\right)$ implies that the residual term converges in probability to zero with the dominant term being $\mathcal{O}\left(1/\sqrt{q}\right)$.

[10]Note that even though $s$ runs through 1 to $N$ in the following optimization, not all possible choices for $s$ may be feasible. For example, if $\rho_c = N^{\gamma}, \gamma < 1$, then $s$ is constrained by $1 \le s \le \lceil \rho_c \rceil$. We do not bother with this technicality here.

In other words, the eigenvalues of $\frac{\mathbf{G}\mathbf{G}^H}{q}$ converge to 1 almost surely. ∎

*Remark 2:* Note that the conclusion of Lemma 3 follows from the law of large numbers and depends critically on the fact that $p$ is finite while $q \to \infty$ [86]. If this condition fails to hold, the spectral behavior of $\mathbf{G}\mathbf{G}^H$ can be significantly different as shown below.

*Lemma 4:* ***Bai and Yin, 1988 [72], [87].*** Let $\mathbf{G}$ be a $p \times q$ complex random matrix with i.i.d. entries having mean zero, variance one and finite fourth moment. Assume that $\kappa_{\mathsf{bf}} = \frac{q}{p} \to \infty$ with $p \to \infty$. Then, the empirical eigenvalue density function of $\frac{\mathbf{G}\mathbf{G}^H - q\mathbf{I}}{\sqrt{pq}}$ converges with probability 1 to $f(\lambda)$ where

$$f(\lambda) = \frac{\sqrt{4-\lambda^2}}{2\pi}, \quad |\lambda| \le 2. \tag{112}$$

∎

We also need the following lemma which is relevant in the ideal regime.

*Lemma 5:* ***Grenander and Silverstein, 1977 [88].*** Let $q$ and $p$ be such that $\kappa_{\mathsf{id}} = \frac{q}{p} \in (0,1]$ as $p \to \infty$. Let $\mathbf{G}$ denote a $q$-connected $p$-dimensional channel [45] where the non-trivial entries of $\mathbf{G}$ are i.i.d. with mean zero and variance one. The empirical eigenvalue density function of $\frac{\mathbf{G}\mathbf{G}^H}{q}$ converges pointwise in probability to $f(\lambda)$ where

$$f(\lambda) = \frac{1}{2\pi}\sqrt{\frac{4-\lambda}{\lambda}}, \quad 0 \le \lambda \le 4. \tag{113}$$

∎

We are now prepared to prove the main statements of Sec. IV.

***Proof of Theorem 3:*** As mentioned in the statement of the theorem, we have two cases: i) $p$ is finite as $\kappa_{\mathsf{bf}} \to \infty$, or ii) $p \to \infty$ as $\kappa_{\mathsf{bf}} \to \infty$. In the former case, $\mathbf{H}^H\mathbf{H}$ is finite dimensional and a straightforward application of Lemma 3 results in (42). In the latter case, $\mathbf{H}^H\mathbf{H}$ is infinite-dimensional in the asymptotics of $N$ and it is not clear how to apply law of large numbers to an infinite number of entries. Taking a recourse to Lemma 4, we have

$$\frac{C_{\mathsf{erg,\,bf}}(N,\rho)}{p\log_2\left(1+\rho\,\frac{q}{p}\right)}$$

$$= \frac{\mathbf{E}\left[\sum_{i=1}^{p}\log_2\left(1+\rho\,\frac{\lambda_i(\mathbf{H}\mathbf{H}^H)}{p}\right)\right]}{p\log_2\left(1+\rho\,\frac{q}{p}\right)} \tag{114}$$

$$\overset{(a)}{=} \frac{\mathbf{E}\left[\sum_{i=1}^{p}\log_2\left(1+\rho\left(\mu_i\sqrt{\frac{q}{p}}+\frac{q}{p}\right)\right)\right]}{p\log_2\left(1+\rho\,\frac{q}{p}\right)} \tag{115}$$

$$\overset{(b)}{=} 1 + \frac{\mathbf{E}\left[\sum_{i=1}^{p}\log_2\left(1+t\mu_i\right)\right]}{p\log_2\left(1+\rho\,\frac{q}{p}\right)} \tag{116}$$

$$\overset{(c)}{\to} 1 + \frac{1}{\log_2\left(1+\rho\,\frac{q}{p}\right)} \cdot \int_{(-\infty,\infty)} \log_2\left(1+t\eta\right) f(\eta) \tag{117}$$

$$\overset{(d)}{=} 1 + \frac{\log_2(e)}{2\pi\cdot\log_2\left(1+\rho\,\frac{q}{p}\right)} \underbrace{\int_{[-2,2]} \log_e\left(1+t\eta\right)\sqrt{4-\eta^2}\,d\eta}_{I} \tag{118}$$

where $\mu_i$ represent the eigenvalues of $\frac{\mathbf{H}\mathbf{H}^H - q\mathbf{I}}{\sqrt{pq}}$ in (a), $t = \frac{\rho\sqrt{\frac{q}{p}}}{1+\rho\frac{q}{p}}$ in (b), and (c) follows from the convergence of $\mu_i$ to $f(\eta)$ in Lemma 4.

The integral in (d), $I$, can be rewritten by using the integration-by-parts formula as

$$I = 2\int_{-2}^{2} \frac{\log_e(1+t\eta)}{\sqrt{4-\eta^2}}d\eta + \frac{1}{2}\int_{-2}^{2}\frac{\sqrt{4-\eta^2}}{1+t\eta}d\eta - \frac{1}{2}\int_{-2}^{2}\sqrt{4-\eta^2}d\eta \quad (119)$$

$$\stackrel{(e)}{=} 2\pi\log_e\left(\frac{1+\sqrt{1-4t^2}}{2}\right) - \pi + \frac{1}{2t}\underbrace{\int_{-2}^{2}\frac{\sqrt{4-\eta^2}}{\eta+\frac{1}{t}}d\eta}_{I_1} \quad (120)$$

where (e) follows from [89, 4.292(3)]. $I_1$ can be computed using standard techniques as

$$I_1 \stackrel{(f)}{=} -\int_{-2}^{2}\frac{\eta}{\sqrt{4-\eta^2}}d\eta + \frac{1}{t}\int_{-2}^{2}\frac{d\eta}{\sqrt{4-\eta^2}} + \left(4-\frac{1}{t^2}\right)\int_{-2}^{2}\frac{d\eta}{\left(\eta+\frac{1}{t}\right)\sqrt{4-\eta^2}} \quad (121)$$

$$\stackrel{(g)}{=} \frac{1}{t}\int_{-2}^{2}\frac{d\eta}{\sqrt{4-\eta^2}} - \left(4-\frac{1}{t^2}\right)\int_{\frac{t}{1-2t}}^{\frac{t}{1+2t}}\frac{d\eta}{\sqrt{\eta^2\left(4-\frac{1}{t^2}\right)+\frac{2\eta}{t}-1}} \quad (122)$$

$$\stackrel{(h)}{=} \frac{\pi}{t} - \frac{\pi\sqrt{1-4t^2}}{t} \quad (123)$$

where (f) follows from [89, 2.282(1)], (g) from [89, 2.281] and (h) from [89, 2.261(3)]. Thus, we have

$$C_{\mathsf{erg,\,bf}}(N,\rho) = p\log_2\left(1+\rho\frac{q}{p}\right) + p\log_2\left(1+\sqrt{1-4t^2}\right) + \frac{\log_2(e)p}{4t^2}\cdot\left(1-\sqrt{1-4t^2}\right) - \log_2(e)p. \quad (124)$$

For the correction term $\Delta C_{\mathsf{bf}}(N,\rho)$, using a Taylor's series expansion, we have

$$|\Delta C_{\mathsf{bf}}(N,\rho)| \approx \frac{1}{\log_2\left(1+\rho\kappa_{\mathsf{bf}}\right)}\cdot\left[\log_2(e)\frac{1}{2\kappa_{\mathsf{bf}}} + \log_2\left(1-\frac{1}{\kappa_{\mathsf{bf}}}-\frac{1}{\kappa_{\mathsf{bf}}^2}\right)\right]. \quad (125)$$

■

***Proof of Proposition 1:*** When $\kappa_{\mathsf{id}} \in (1,\infty)$, $\mathbf{H}_{\mathsf{id}}$ reduces to a $q \times p$ i.i.d. channel. The i.i.d. channel capacity formula from [73], [74] is precisely the statement of the theorem. When $\kappa_{\mathsf{id}} \in (0,1]$, $\mathbf{H}_{\mathsf{id}}$ reduces to a $q$-connected $p$-dimensional channel. Using Lemma 5, it is not difficult to check that (47) holds (see [45, Theorem 5] for more details).

With respect to the correction terms, we need to compute the high- and low-SNR trends for the $\kappa_{\mathsf{id}} \leq 1$ and $\kappa_{\mathsf{id}} > 1$ cases separately. If $\kappa_{\mathsf{id}} > 1$ and $\rho \to \infty$, a Taylor's series analysis of $h$ in (45) shows that

$$h \to 1 - \frac{1}{\rho(\kappa_{\mathsf{id}}-1)} + \frac{\kappa_{\mathsf{id}}}{\rho^2(\kappa_{\mathsf{id}}-1)^3}. \quad (126)$$

Using this, it is easy to see that

$$\frac{C_{\mathsf{erg,\,id}}(N,\rho)}{\log_2(e)\cdot q} \to \frac{1}{\kappa_{\mathsf{id}}}\log_e\left(\frac{\kappa_{\mathsf{id}}}{\kappa_{\mathsf{id}}-1} + \rho(\kappa_{\mathsf{id}}-1) - \frac{\kappa_{\mathsf{id}}}{\rho(\kappa_{\mathsf{id}}-1)^3}\right) + \log_e\left(\frac{\kappa_{\mathsf{id}}}{\kappa_{\mathsf{id}}-1} - \frac{\kappa_{\mathsf{id}}}{\rho(\kappa_{\mathsf{id}}-1)^3}\right) - \left(\frac{1}{\kappa_{\mathsf{id}}} - \frac{1}{\rho\kappa_{\mathsf{id}}(\kappa_{\mathsf{id}}-1)}\right) \quad (127)$$

and this results in (49). As $\rho \to 0$, another Taylor's series expansion shows that $h \to \rho\kappa_{\mathsf{id}} - \rho^2\kappa_{\mathsf{id}}(1+\kappa_{\mathsf{id}})$ and

$$C_{\mathsf{erg,\,id}}(N,\rho) \to \log_2(e)\cdot\rho\cdot q\left(1-\frac{5}{2}\rho(1+\kappa_{\mathsf{id}})\right) \quad (128)$$

from which (49) follows immediately. For the $\kappa_{\mathsf{id}} \leq 1$ case, trivial computations yield

$$C_{\mathsf{erg,\,id}}(N,\rho) \to \log_2(e)\,p\left(\log_e(\rho\kappa_{\mathsf{id}}) - 1 - \frac{1}{2\rho\kappa_{\mathsf{id}}}\right) \quad (129)$$

as $\rho \to \infty$ and $C_{\mathsf{erg,\,id}}(N,\rho) \to \log_2(e)\cdot q\cdot\rho\,(1-5\rho\kappa_{\mathsf{id}})$ as $\rho \to 0$. The conclusions in (49) are again immediate. ■

***Proof of Theorem 4:*** In the multiplexing regime, $\mathbf{H}$ is a $q$-connected $p$-dimensional matrix. The definition of $\mathbf{H}$ leads to two possibilities: i) either $q$ is a constant and $\kappa_{\mathsf{mux}} = \frac{q}{p} \to 0$, or ii) $q \to \infty$ and $\kappa_{\mathsf{mux}} = \frac{q}{p} \to 0$. Note that Lemma 5 is applicable only when $\frac{q}{p} = \Theta(1)$. So in either case, we do not have knowledge (a closed-form expression) of the limiting EED.

Nevertheless, [75, Corollary 3.1, p. 656] studies the limiting EED of a Gram matrix (a matrix of the form $\mathbf{H}\mathbf{H}^H$) with a given variance profile and shows that the Stieltjes transform of the EED shares the same structural form as that of (22)-(24). It is important to note that Hachem *et al.* allow the possibility of $\kappa_{\mathsf{mux}} \to 0$ in the structure of the variance profile; see [75, Assumption A-2, Remark 2.2, p. 652] for details. The structural similarity between the fixed-point equations in [75] and [30] implies that we can exploit the regularity of the multiplexing regime to solve for capacity in closed-form in either case as $N$ increases. It should not be surprising that the capacity formula is identical to that of the ideal channel with $\kappa_{\mathsf{id}} < 1$. The correction term follows from a routine application of Taylor's series expansion. ■

### *E. Proof of Proposition 4*

The fact that we have two distinct solutions for $\Gamma(t)$ in the ideal regime is a consequence of the structure of the family of channels studied. In the beamforming and ideal regimes with $\kappa > 1$, the channel $\mathbf{H}$ has the structure of a $q \times p$ channel with i.i.d. entries. On the other hand, in the ideal regime with $\kappa \leq 1$ and in the multiplexing regime, the channel $\mathbf{H}$ is a $q$-connected $p$-dimensional channel as illustrated in Fig. 1. In either case, the channel is both row- and column-regular and

hence, $\Gamma(t)$ is independent of the transmit dimension $t$. Thus we will denote it for simplicity by $\Gamma$ and the associated MSE quantity $\Upsilon$ satisfies the relationship $\Upsilon = \frac{\rho}{1+\rho\Gamma}$. Substituting for $\Upsilon$ in (68), we have

$$\Gamma = \kappa_{\mathsf{TLV}} \, \mathbf{E}_{\mathcal{R}} \left[ \frac{\mathcal{G}(\mathcal{R}, t)}{1 + \frac{\rho}{1+\rho\Gamma} \, \mathbf{E}_{\mathcal{T}}[\mathcal{G}(\mathcal{R}, \mathcal{T})|\mathcal{R}]} \right]. \tag{130}$$

In the first case, $\kappa_{\mathsf{TLV}}$ (the ratio of receive and transmit dimensions) reduces to $\kappa_{\mathsf{TLV}} = \kappa$. Note that $\mathbf{E}_{\mathcal{T}}[\mathcal{G}(r, \mathcal{T})]$ is independent of $r$ and equals 1. Thus, the conditional random variable $\mathbf{E}_{\mathcal{T}}[\mathcal{G}(\mathcal{R}, \mathcal{T})|\mathcal{R}] = 1$. Plugging this and simplifying (130), we have

$$\Gamma \left( 1 + \frac{\rho}{1 + \rho\Gamma} \right) = \kappa \, \mathbf{E}_{\mathcal{R}}[\mathcal{G}(\mathcal{R}, t)]. \tag{131}$$

Substitute $\mathbf{E}_{\mathcal{R}}[\mathcal{G}(\mathcal{R}, t)] = 1$ and solving for the quadratic equation in (131), we obtain (71). We proceed on similar lines in the second case. The following modifications are to be made: i) $\kappa_{\mathsf{TLV}} = 1$, ii) while $\mathbf{E}_{\mathcal{T}}[\mathcal{G}(r, \mathcal{T})]$ and $\mathbf{E}_{\mathcal{R}}[\mathcal{G}(\mathcal{R}, t)]$ are still independent of $r$ and $t$, respectively, they equal $\kappa$ here. Solving for the resultant quadratic, we obtain (72). ■

## REFERENCES


[1] Í. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–596, Nov. 1999.

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

[3] D. Gesbert, H. Bolcskei, D. A. Gore, and A. J. Paulraj, "Outdoor MIMO wireless channels: Models and performance prediction," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1926–1934, Dec. 2002.

[4] C.-N. Chuah, J. M. Kahn, and D. N. C. Tse, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[5] B. Razavi, *RF Microelectronics*, Prentice Hall, Upper Saddle River, NJ, 1998.

[6] S. Ahmadi, "An overview of next-generation mobile WiMAX technology," *IEEE Commun. Magaz.*, vol. 47, no. 6, pp. 84–98, June 2009.

[7] D. Cabric, M. S. W. Chen, D. A. Sobel, S. Wang, J. Yang, and R. W. Brodersen, "Novel radio architectures for UWB, 60 GHz, and cognitive wireless systems," *EURASIP Journ. Wireless Commun. and Networking*, vol. 2006, Article ID 17957, pp. 1–18, Jan. 2006.

[8] C. H. Doan, S. Emami, D. A. Sobel, A. M. Niknejad, and R. W. Brodersen, "Design considerations for 60 GHz CMOS radios," *IEEE Commun. Magaz.*, vol. 42, no. 12, pp. 132–140, Dec. 2004.

[9] S. K. Yong and C.-C. Chong, "An overview of multigigabit wireless through millimeter wave technology: Potentials and technical challenges," *EURASIP Journ. Wireless Commun. and Networking*, vol. 2007, Article ID 78907, 2007.

[10] B. Razavi, "Design of millimeter-wave CMOS radios: A tutorial," *IEEE Trans. Circuits and Systems - Part I*, vol. 56, no. 1, pp. 4–16, Jan. 2009.

[11] D. A. Sobel, "A baseband mixed-signal receiver front-end for 1 Gbps wireless communications at 60 GHz," Ph. D. dissertation, *University of California, Berkeley*, 2008.

[12] K. Yu, M. Bengtsson, B. Ottersten, P. Karlsson, and M. A. Beach, "Second order statistics of NLOS indoor MIMO channels based on 5.2 GHz measurements," *IEEE Global Telecommun. Conf.*, vol. 1, pp. 156–160, Nov. 2001.

[13] D. P. McNamara, M. A. Beach, and P. N. Fletcher, "Spatial correlation in indoor MIMO channels," *IEEE Intern. Symp. Pers. Ind. Mob. Radio Commun.*, vol. 1, pp. 290–294, Sept. 2002.

[14] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE Journ. Sel. Areas in Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[15] J. W. Wallace, M. A. Jensen, A. L. Swindlehurst, and B. D. Jeffs, "Experimental characterization of the MIMO wireless channel: Data acquisition and analysis," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 335–343, Mar. 2003.

[16] M. J. Gans, N. Amitay, Y. Yeh, H. Xu, T. Damen, R. A. Valenzuela, T. Sizer, R. Storz, D. Taylor, W. MacDonald, C. Tran, and A. Adamiecki, "Outdoor BLAST measurement system at 2.44 GHz: Calibration and initial results," *IEEE Journ. Sel. Areas in Commun.*, vol. 3, no. 20, pp. 570–583, Apr. 2002.

[17] D. Chizhik, J. Ling, P. W. Wolniansky, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple-input-multiple-output measurements and modeling in Manhattan," *IEEE Journ. Sel. Areas in Commun.*, vol. 3, no. 21, pp. 321–331, Apr. 2003.

[18] H. Ozcelik, M. Herdin, W. Weichselberger, J. Wallace, and E. Bonek, "Deficiencies of 'Kronecker' MIMO Radio Channel Model," *Electronics Letters*, vol. 39, no. 16, pp. 1209–1210, Aug. 2003.

[19] E. Bonek, "Experimental validation of analytical MIMO channel models," *Elektrotechnik und Informationstechnik (e&i)*, vol. 122, no. 6, pp. 196–205, 2005.

[20] N. Costa and S. Haykin, "A novel wideband channel model and experimental validation," *IEEE Trans. Antennas and Propagat.*, vol. 56, no. 2, pp. 550–562, Feb. 2008.

[21] L. Wood and W. S. Hodgkiss, "A reduced-rank eigenbasis MIMO channel model," *IEEE Wireless Telecommun. Symp.*, pp. 78–83, Apr. 2008.

[22] H. Ozcelik, N. Czink, and E. Bonek, "What makes a good MIMO channel model?," *IEEE Spring Veh. Tech. Conf.*, vol. 1, pp. 156–160, May 2005.

[23] S. Wyne, A. Molisch, P. Almers, G. Eriksson, J. Karedal, and F. Tufvesson, "Statistical evaluation of outdoor-to-indoor office MIMO measurements at 5.2 GHz," *IEEE Fall Veh. Tech. Conf.*, vol. 1, pp. 146–150, May 2005.

[24] T. A. Lamahewa, R. A. Kennedy, T. D. Abhayapala, and T. Betlehem, "MIMO channel correlation in general scattering environments," *Aus. Commun. Theory Workshop*, pp. 93–98, Feb. 2006.

[25] P. Almers, E. Bonek, and A. G. Burr *et al.*, "Survey of channel and radio propagation models for wireless MIMO systems," *EURASIP Journ. Wireless Commun. and Networking*, vol. 2007, no. 1, Jan. 2007.

[26] Y. Zhou, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental study of MIMO channel statistics and capacity via the virtual representation," Tech. Rep., University of Wisconsin-Madison, May 2006, Available: [Online]. `http://dune.ece.wisc.edu`.

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

[28] V. Raghavan, J. H. Kotecha, and A. M. Sayeed, "Why does the Kronecker model result in misleading capacity estimates?," *In print, IEEE Trans. Inf. Theory*, 2010.

[29] A. M. Sayeed, "Deconstructing multi-antenna fading channels," *IEEE Trans. Sig. Proc.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[30] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, July 2005.

[31] M. Matthaiou, A. M. Sayeed, and J. A. Nossek, "Sparse multipath MIMO channels: Performance implications based on measurement data," *IEEE 10th Workshop on Sig. Proc. Adv. in Commun. (SPAWC)*, pp. 364–368, June 2009.

[32] O. M. Bucci and G. Franceschetti, "On degrees of freedom of scattered fields," *IEEE Trans. Antennas and Propagat.*, vol. 37, no. 7, pp. 918–926, July 1989.

[33] M. D. Migliore, "On the role of the number of degrees of freedom of the field in MIMO channels," *IEEE Trans. Antennas and Propagat.*, vol. 54, no. 2, pp. 620–628, Feb. 2006.

[34] K. Chakraborty and M. Franceschetti, "Maxwell meets Shannon: Space-time duality in multiple antenna channels," *Allerton Conf. Commun., Cont., and Comput.*, pp. 761–770, Sept. 2006.

[35] J. Xu and R. Janaswamy, "Electromagnetic degrees of freedom in 2-D scattering environments," *IEEE Trans. Antennas and Propagat.*, vol. 54, no. 12, pp. 3882–3894, Dec. 2006.

[36] L. W. Hanlen, R. Timo, and R. Perera, "On dimensionality for sparse multipath," *Aus. Commun. Theory Workshop*, pp. 125–129, Feb. 2006.

[37] A. S. Y. Poon, R. W. Brodersen, and D. N. C. Tse, "Degrees of freedom in multiple-antenna channels: A signal space approach," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 523–536, Feb. 2005.

[38] R. A. Kennedy, P. Sadeghi, T. D. Abhayapala, and H. M. Jones, "Intrinsic limits of dimensionality and richness in random multipath fields," *IEEE Trans. Sig. Proc.*, vol. 55, no. 6, pp. 2542–2556, June 2007.

[39] N. A. Goodman, "MIMO channel rank via the aperture-bandwidth product," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2246–2254, June 2007.

[40] S. Loyka and G. Levin, "On physically-based normalization of MIMO channel matrices," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1107–1112, Mar. 2009.

[41] R. R. Muller, "Random matrix model of communication via antenna arrays," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2495–2506, Sept. 2002.

[42] A. G. Burr, "Capacity bounds and estimates for the finite scatterers MIMO wireless channel," *IEEE Journ. Sel. Areas in Commun.*, vol. 21, no. 5, pp. 812–818, June 2003.

[43] N. Chiurtu, B. Rimoldi, and Í. E. Telatar, "Dense multiple antenna systems," *IEEE Inf. Theory Workshop*, pp. 108–109, Sept. 2001.

[44] D. C. Jenn, "Transmission equation for multiple cooperative transmitters and collective beamforming," *IEEE Trans. Antennas and Wireless Propagat. Letters*, vol. 7, pp. 606–608, 2008.

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

[46] M. Skoglund and G. Jongren, "On the capacity of a multiple-antenna communication link with channel side information," *IEEE Journ. Sel. Areas in Commun.*, vol. 21, no. 3, pp. 395–405, Apr. 2003.

[47] A. J. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journ. Sel. Areas in Commun.*, vol. 21, no. 5, pp. 684–702, June 2003.

[48] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: A game-theoretic approach," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1707–1727, July 2003.

[49] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large N analysis," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.

[50] A. M. Tulino and S. Verdù, "Random matrices and wireless communications," *Foundations and Trends in Commun. and Inf. Theory*, vol. 1, no. 1, June 2004.

[51] C. Martin and B. Ottersten, "Asymptotic eigenvalue distributions and capacity for MIMO channels under correlated fading," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1350–1358, July 2004.

[52] E. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multi-antenna systems with different types of channel state information," *IEEE Trans. Sig. Proc.*, vol. 52, no. 12, pp. 3440–3453, Dec. 2004.

[53] V. V. Veeravalli, Y. Liang, and A. M. Sayeed, "Correlated MIMO Rayleigh fading channels: Capacity, optimal signaling and asymptotics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2058–2072, June 2005.

[54] L. Hanlen and A. J. Grant, "Optimal transmit covariance for MIMO channels with statistical transmitter side information," *IEEE Intern. Symp. Inf. Theory*, pp. 1818–1822, Sept. 2005.

[55] S. H. Simon, A. L. Moustakas, and L. Marinelli, "Capacity and character expansions: Moment-generating function and other exact results for MIMO correlated channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5336–5351, Dec. 2006.

[56] W. Hachem, O. Khorunzhiy, Ph. Loubaton, J. Najim, and L. Pastur, "A new approach for mutual information analysis of large dimensional multi-antenna channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3987–4004, Sept. 2008.

[57] X. Gao, B. Jiang, X. Li, A. B. Gershman, and M. R. McKay, "Statistical eigenmode transmission over jointly-correlated MIMO channels," *IEEE Trans. Inf. Theory*, vol. 8, no. 55, pp. 3735–3750, Aug. 2009.

[58] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[59] S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE Journ. Sel. Areas in Commun.*, vol. 21, no. 3, pp. 406–417, Apr. 2003.

[60] A. M. Tulino, A. Lozano, and S. Verdú, "Multiantenna capacity in the low-power regime," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2527–2544, Oct. 2003.

[61] A. Lozano, A. M. Tulino, and S. Verdú, "High-SNR power offset in multiantenna communication," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4134–4151, Dec. 2005.

[62] A. M. Tulino, A. Lozano, and S. Verdú, "Capacity-achieving input covariance for correlated multi-antenna channels," *Allerton Conf. Commun., Cont., and Comput.*, pp. 242–251, Oct. 2003.

[63] P. Kyritsi, D. C. Cox, R. A. Valenzuela, and P. W. Wolniansky, "Effect of antenna polarization on the capacity of a multiple element system in an indoor environment," *IEEE Journ. Sel. Areas in Commun.*, vol. 20, no. 6, pp. 1227–1239, Aug. 2002.

[64] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[65] S. A. Jafar and A. J. Goldsmith, "Isotropic fading vector broadcast channel: The scalar upperbound and loss in degrees of freedom," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 848–857, Mar. 2005.

[66] V. R. Cadambe and S. A. Jafar, "Interference alignment and the degrees of freedom for the K user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[67] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2632–2639, Sept. 2001.

[68] Y. Liang and V. V. Veeravalli, "Correlated MIMO Rayleigh fading channels: Capacity and optimal signaling," *IEEE Asilomar Conf. Signals, Systems and Computers*, vol. 1, pp. 1166–1170, Nov. 2003.

[69] V. Raghavan and A. M. Sayeed, "Multi-antenna capacity of sparse multipath channels," Tech. Rep. ECE-09-03, University of Wisconsin-Madison, Nov. 2009.

[70] V. Raghavan and A. M. Sayeed, "MIMO capacity scaling and saturation in correlated environments," *IEEE Intern. Conf. on Commun.*, vol. 5, pp. 3006–3010, May 2003.

[71] V. Raghavan, "Spatially correlated multi-antenna wireless channels: Capacity and limited feedback performance," Ph. D. dissertation, *University of Wisconsin-Madison*, Aug. 2006.

[72] Z. D. Bai and Y. Q. Yin, "Convergence to the semi-circle law," *Annals Prob.*, vol. 16, no. 2, pp. 863–875, 1988.

[73] P. B. Rapajic and D. Popescu, "Information capacity of a random signature multiple-input multiple-output channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 1245–1248, Aug. 2000.

[74] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.

[75] W. Hachem, Ph. Loubaton, and J. Najim, "The empirical distribution of the eigenvalues of a Gram matrix with a given variance profile," *Annales de l'Institut Henri Poincare, Probabilites et statistiques*, vol. 42, no. 6, pp. 649–670, 2006.

[76] A. M. Sayeed and V. Raghavan, "Maximizing MIMO capacity in sparse multipath with reconfigurable antenna arrays," *IEEE Journ. Special Topics in Sig. Proc.*, vol. 1, no. 1, pp. 156–166, June 2007.

[77] G. Barriac and U. Madhow, "Space-time communication for OFDM with implicit channel feedback," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3111–3129, Dec. 2004.

[78] T. Muharemovic, A. Sabharwal, and B. Aazhang, "Antenna packing in low-power systems: Communication limits and array design," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 429–440, Jan. 2008.

[79] X. Wu and R. Srikant, "MIMO channels in the low SNR regime: Communication rate, error exponent and signal peakiness," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1290–1309, Apr. 2007.

[80] T. L. Marzetta and A. Ashikhmin, "Beyond LTE: *Hundreds* of base station antennas," *IEEE Commun. Theory Workshop*, May 2010.

[81] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, June 2010.

[82] S. R. Bhaskaran, L. Davis, A. J. Grant, S. V. Hanly, and P. Tune, "Downlink scheduling using compressed sensing," *IEEE Inf. Theory Workshop*, pp. 201–205, June 2009.

[83] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

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

[85] R. A. Durrett, *Probability: Theory and Examples*, Duxbury Press, 2nd edition, 1995.

[86] Z. D. Bai and H. Saranadasa, "Effect of high dimension comparison of significance tests for a high dimensional two sample problem," *Statistica Sinica*, vol. 6, pp. 311–329, 1996.

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

[88] 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.

[89] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, NY, 4th edition, 1965.


**Vasanthan Raghavan (S'01–M'06)** received the B.Tech degree in Electrical Engineering from the Indian Institute of Technology at Madras, India in 2001, the M.S. and the Ph.D. degrees in Electrical and Computer Engineering in 2004 and 2006 respectively, and the M.A. degree in Mathematics in 2005, all from the University of Wisconsin-Madison, Madison, WI. He is currently a research fellow with The University of Melbourne, Parkville, Australia. He was with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL from 2006 to 2009. His research interests span multi-antenna communication techniques, quickest change detection, information theory, multihop networking, robust control, and random matrix theory.


**Akbar M. Sayeed (S'89–M'97–SM'02)** is currently Professor of Electrical and Computer Engineering at the University of Wisconsin-Madison. He received the B.S. degree from the University of Wisconsin-Madison in 1991, and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign in 1993 and 1996, all in Electrical Engineering. He was a postdoctoral fellow at Rice University from 1996 to 1997. His current research interests include wireless communications, statistical signal processing, multi-dimensional communication theory, information theory, learning theory, time-frequency analysis, and applications in wireless communication networks and sensor networks. Dr. Sayeed is a recipient of the Robert T. Chien Memorial Award (1996) for his doctoral work at Illinois, the NSF CAREER Award (1999), the ONR Young Investigator Award (2001), and the UW Grainger Junior Faculty Fellowship (2003). He is a Senior Member of the IEEE and is currently serving on the signal processing for communications technical committee of the IEEE Signal Processing Society. Dr. Sayeed also served as an Associate Editor for the IEEE Signal Processing Letters from 1999-2002, and as the technical program co-chair for the 2007 IEEE Statistical Signal Processing Workshop and the 2008 IEEE Communication Theory Workshop.