

Multi-Antenna Capacity of Sparse Multipath Channels

Vasanthan Raghavan and Akbar M. Sayeed

Department of Electrical and Computer Engineering

University of Wisconsin-Madison

1415 Engineering Drive, Madison, WI 53706

vasanthan_raghavan@ieee.org, akbar@engr.wisc.edu

Corresponding author: Akbar M. Sayeed - Phone: (608) 265-4731, Fax: (608) 265-4623

Submitted to the *IEEE Transactions on Information Theory*, August 2006.

This work was partly supported by the NSF under grant #CCF-0431088. This paper was presented in part at the IEEE Information Theory Workshop 2004 (ITW 2004), San Antonio, TX and the IEEE International Symposium on Information Theory 2006 (ISIT 2006), Seattle, WA.

Abstract

Existing results on multi-input multi-output (MIMO) channel capacity implicitly assume a rich scattering environment in which the channel power scales quadratically with the number of antennas, resulting in linear capacity scaling with the number of antennas. While this assumption may be justified in systems with few antennas, it leads to violation of fundamental power conservation principles in the limit of large number of antennas. Furthermore, recent measurement results have shown that physical MIMO channels exhibit a sparse multipath structure, even for relatively few antenna dimensions. Motivated by these observations, we propose a framework for modeling sparse channels and study the coherent capacity of sparse MIMO channels from two perspectives: 1) capacity scaling with the number of antennas, and 2) capacity as a function of transmit SNR for a fixed number of antennas. The statistically independent degrees of freedom (DoF) in sparse channels are less than the number of signal-space dimensions and, as a result, sparse channels afford a fundamental new degree of freedom over which channel capacity can be optimized: the distribution of the DoF's in the available signal-space dimensions. Our investigation is based on a family of sparse channel configurations whose capacity admits a simple and intuitive closed-form approximation and reveals a new tradeoff between the multiplexing gain and the received SNR. We identify an *ideal* channel configuration that yields the fastest capacity scaling at any SNR. For fixed number of antennas, we show that the capacity maximizing configuration depends on the operating SNR and optimizes the multiplexing gain-received SNR tradeoff. Surprisingly, only three such configurations suffice for near-optimal performance over the entire SNR range. Different channel configurations can be realized in practice by appropriately adjusting the antennas' spacings at the transmitter and the receiver.

Keywords

Antenna arrays, fading channels, MIMO systems, random matrix theory, reconfigurable arrays

I. INTRODUCTION

Multiple-input multiple-output (MIMO) systems that employ antenna arrays at the transmitter and the receiver have emerged as one of the most promising technologies to increase the spectral efficiency of high speed wireless communication systems. The intense research on MIMO systems was inspired by seminal works by Telatar [1] and Foschini and Gans [2] that showed a dramatic linear increase in channel capacity with the number of antennas. However, these initial results were based on an idealized channel model representing a rich scattering environment – the channel matrix entries, representing coupling between pairs of transmit and receive antennas, were assumed to be independent and identically distributed (i.i.d.) Gaussian random variables (Rayleigh fading). Since then researchers have studied capacity scaling for more realistic spatially correlated channel models. For example, using a Kronecker (separable in transmit and

receive statistics) channel model, it was shown in [3] that correlated channels would also exhibit linear capacity scaling, albeit with a smaller slope compared to i.i.d. channels. In [4], [5], capacity scaling was studied for non-separable channels, using the virtual channel representation for uniform linear arrays (ULA's), and it was shown that the number of resolvable scattering paths must scale quadratically with the number of antennas for linear capacity scaling. Furthermore, for slower path growth, the capacity may exhibit a sub-linear growth or may saturate. In all these works, linear capacity scaling is associated with the implicit assumption that the channel power $\rho_c \triangleq \mathbf{E} [\text{Tr}(\mathbf{H}\mathbf{H}^H)]$ scales quadratically with the number of antennas, N ; that is, $\rho_c \sim \mathcal{O}(N^2)$.

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 and is directly reflected in the quadratic channel power scaling assumption in existing results. While this assumption may be justified for small N , such scaling in ρ_c is *not* sustainable indefinitely from an energy conservation viewpoint since it implies that the received signal power increases linearly (and indefinitely) with the number of antennas for a given fixed transmit power (see also Section II). The second effect primarily governs channel capacity as a function of SNR for a fixed number of antennas. The capacity of a correlated MIMO channel depends on the statistically independent degrees of freedom (DoF)¹ in the channel matrix, which in turn depend on the number of resolvable paths in the scattering environment [6], [4]. The notion of $\mathcal{O}(N^2)$ channel power scaling implies that the dominant, independent DoF also scale with antenna dimensions as $\mathcal{O}(N^2)$. However, many recent measurement campaigns [7], [8] and channel modeling efforts under diverse sets of assumptions [9], [10], [11], [12], [13] indicate that such a possibility is an exception rather than the norm and most realistic propagation environments are accurately characterized by fewer dominant DoF. That is, realistic physical channels exhibit *sparse* DoF reflecting a sparse multipath environment.

Overview of Contributions: Motivated by the above considerations, we propose a framework for modeling sparse multipath channels and study the implications of sparsity for coherent²

¹The notion of independent DoF is loosely defined as those entries in the matrix channel that are sufficiently large enough to significantly impact information transfer. A more formal definition is provided in Section II.

²By coherent capacity, we mean that the receiver has perfect channel state information, whereas the transmitter only knows channel statistics.

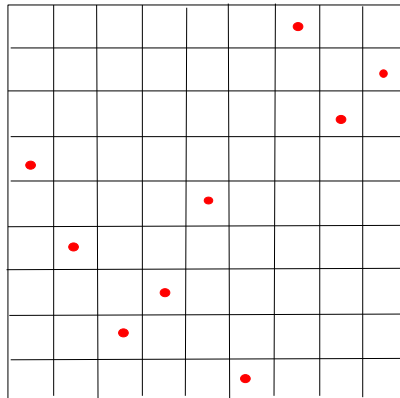


Fig. 1. A 9×9 virtual channel matrix representing sparse multipath. The 9 dominant, independent DoF in the channel matrix are denoted by the occupied bins.

MIMO channel capacity – both ergodic and outage. Our development is based on the virtual channel representation [6] that provides an accurate and analytically tractable model for physical MIMO channels induced by ULA's. A key property of the virtual channel matrix is that it is a unitarily equivalent representation of the actual channel matrix and the virtual channel entries are approximately independent since they reflect contributions from approximately disjoint sets of propagation paths [6]. As a result, spatial correlation in the antenna domain is captured by the *variation* in the powers of the virtual coefficients. Furthermore, the dominant DoF in the channel are characterized by the dominant virtual channel coefficients – the non-vanishing coefficients with sufficiently large power.

Our approach to modeling sparse channels is an alternative to existing approaches for capturing spatial correlation in the antenna domain. Consider an $N \times N$ virtual channel matrix \mathbf{H} corresponding to N transmit and receive antennas. For fixed N , our definition of sparsity implies that the DoF in the channel, $D < N^2$, is as illustrated in Fig. 1; that is, not all of N^2 virtual channel coefficients are dominant. From a scaling perspective, our definition of sparsity implies that the DoF and the channel power scale at a sub-quadratic rate with N ; that is, $D \sim \rho_c \sim \mathcal{O}(N^\gamma)$, $\gamma \in (0, 2]$. In contrast to conventional rich scattering assumptions, sparsity affords a new degree of freedom over which the capacity can be optimized: *the spatial distribution or configuration of the D dominant DoF within the available N^2 channel dimensions* (see Fig. 1). From a capacity scaling perspective, we address the following question: *Given a channel power (and DoF) scaling law $\rho_c \sim D \sim \mathcal{O}(N^\gamma)$, $\gamma \in (0, 2]$, what is the fastest achievable capacity scaling law,*

and what channel configuration achieves it? In the context of fixed antenna systems, we address the following question: *Does channel capacity as a function of SNR depend on the channel configuration, and if so, what channel configuration maximizes capacity at any given SNR?*

For the scaling question, we first show that no channel configuration with channel power scaling $\rho_c(N)$ can beat an ergodic capacity scaling of $\mathcal{O}(\sqrt{\rho_c(N)})$. We then address the question of whether this fundamental limit is achievable. This goal is facilitated by considering a structured family of channels for any given N corresponding to different canonical distributions in which the D non-vanishing virtual coefficients are factored as $D = pq$, where p denotes the multiplexing gain (or the number of parallel channels) and q denotes the number of DoF per parallel channel. We show that the capacity of all channels in the family is accurately approximated by the following simple and intuitive formula reminiscent of the Shannon formula for AWGN capacity:

$$C \approx p \log(1 + \rho_{rx}) = p \log\left(1 + \rho \frac{q}{p}\right) \quad (1)$$

where ρ_{rx} denotes the received SNR per parallel channel for a uniform power input³; that is, $\rho_{rx} \triangleq \frac{\mathbf{E}[\|\mathbf{H}\mathbf{x}\|^2]}{p} = \rho \frac{D}{p^2} = \rho \frac{q}{p}$. The above formula reveals a fundamental new tradeoff in sparse channels between the multiplexing gain p and the received SNR, ρ_{rx} : increasing p comes at the cost of decreasing ρ_{rx} and *vice versa*. Optimization of this fundamental tradeoff governs the optimal channel configuration from a capacity scaling viewpoint as well as from the viewpoint of maximizing capacity at any operating transmit SNR for a fixed number of antennas. It is worth noting that while existing works have obtained closed-form expressions for capacity only in the low- or high-SNR regimes, the above formula accurately approximates capacity over the entire SNR range.

We now present a simple example to illustrate our approach and main results. Consider the $\gamma = 1$ case such that $\rho_c = D = N$. Three canonical channel configurations from the structured family are illustrated in Fig. 2 for the $N = 9$ case. On one extreme is the *beamforming channel* for which \mathbf{H}_{bf} is an $N \times 1$ matrix with i.i.d. entries: $p = 1, q = N$ and the DoF are distributed to maximize $\rho_{rx} = \rho \frac{q}{p} = \rho N$. On the other extreme is the *multiplexing channel* for which \mathbf{H}_{mux} is an $N \times N$ diagonal matrix with i.i.d. diagonal entries: $p = N, q = 1$ and the DoF are distributed to maximize the multiplexing gain with $\rho_{rx} = \rho/N$. In between the two extremes is the *ideal*

³As we will see, the uniform power input is optimal for the family of channels considered.

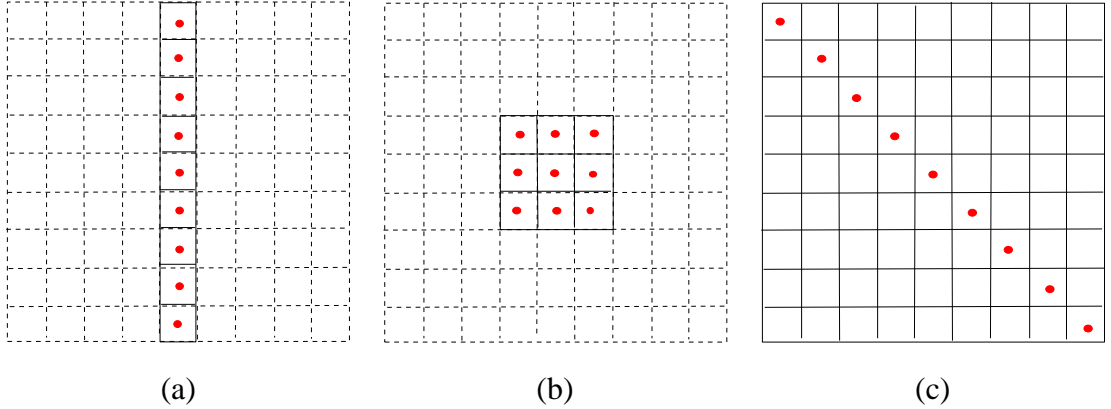


Fig. 2. A schematic illustrating three canonical sparse channel configurations: (a) the beamforming, (b) the ideal, and (c) the multiplexing channels.

channel⁴ for which \mathbf{H}_{id} is a $\sqrt{N} \times \sqrt{N}$ matrix with i.i.d. entries: $p = q = \sqrt{N}$ and the DoF are distributed to sustain a stable receive SNR for all N ; that is, $\rho_{rx} = \rho$.

From a scaling perspective, the capacities of the three channels are given by

$$C_{\text{bf}} \approx \log(1 + \rho N), \quad C_{\text{id}} \approx \sqrt{N} \log(1 + \rho) \quad \text{and} \quad C_{\text{mux}} \approx N \log(1 + \rho/N) \rightarrow \rho. \quad (2)$$

It is easy to check that $C_{\text{bf}} \sim o(C_{\text{id}})$ and $C_{\text{mux}} \sim o(C_{\text{id}})$. In fact, as we show in this paper, among all channels in the family, the ideal channel exhibits the fastest capacity scaling and also achieves the fundamental capacity scaling limit regardless of the SNR: $C_{\text{id}} = \mathcal{O}(\sqrt{N}) \sim \mathcal{O}(\sqrt{D}) \sim \mathcal{O}(\sqrt{\rho_c})$. We note that for $\gamma = 2$ ($\rho_c = N^2$), the ideal channel reduces to the $N \times N$ i.i.d. channel (rich multipath) and achieves the well-known capacity scaling of $\mathcal{O}(N)$ with a maximum multiplexing gain of N .

Fig. 3 shows the capacity of the three canonical channel configurations as a function of transmit SNR, ρ , for $N = 25$ and $\gamma = 1$. The curves are based on (1). As evident, in the low SNR regime $C_{\text{bf}} > C_{\text{id}} > C_{\text{mux}}$ whereas in the high SNR regime $C_{\text{bf}} < C_{\text{id}} < C_{\text{mux}}$. Thus, while the beamforming and multiplexing channels exchange roles and are optimal in the low- and high-SNR regimes, respectively, the ideal channel is a robust choice whose capacity is always between the two extremes. Furthermore, for every SNR, there exists an optimal channel configuration (corresponding to an optimal value of p , illustrated by the dotted curves) that optimizes the multiplexing gain-received SNR tradeoff and yields maximum capacity at that SNR.

⁴Assume for simplicity that \sqrt{N} is an integer.

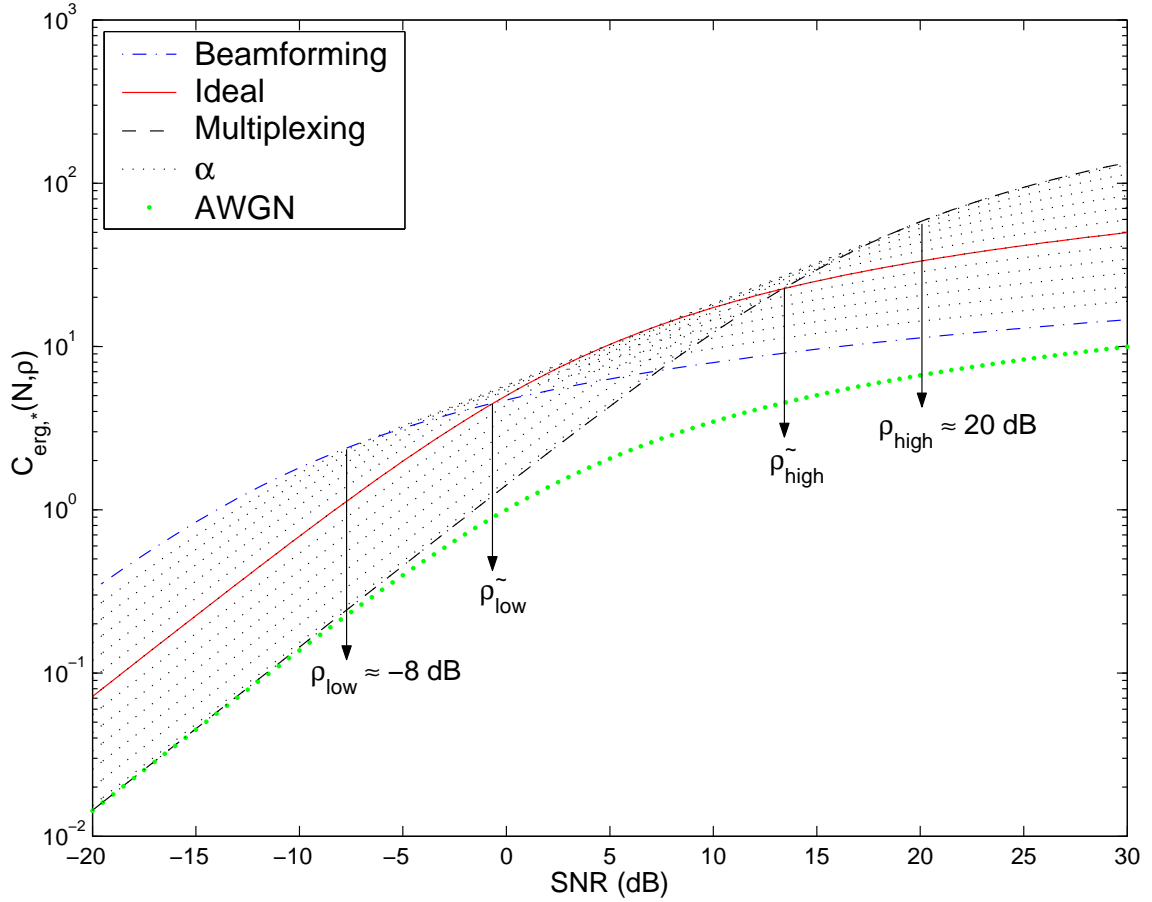


Fig. 3. Capacity as a function of transmit SNR, ρ , for the three channel configurations for $N = 25$ and $\gamma = 1$. The dotted curves correspond to 10 equally spaced values of α where $p = N^\alpha$ and $q = \frac{D}{p} = \frac{N}{p}$ in (1). Also shown is single-antenna AWGN capacity at the same ρ for comparison.

The ideal channel maximizes capacity in the medium SNR range. Surprisingly, for all practical purposes, the three canonical channel configurations – beamforming, ideal and multiplexing – suffice to accurately approximate the optimal channel configurations over the entire transmit SNR range, as evident from Fig. 3. The importance of the SNR quantities ρ_{low} , ρ_{high} , $\tilde{\rho}_{\text{low}}$ and $\tilde{\rho}_{\text{high}}$, shown in Fig. 3, will be addressed later in the sequel.

Building on the above results on ergodic capacity, we also study the outage capacities of the three channel configurations in closed-form and show that similar trends hold true. In the low- and the high-SNR regimes, the beamforming and the multiplexing channels are outage-optimal. When the SNR is high, the system is operating at a point where enhancing the reliability of signal reception can only marginally improve overall performance and hence maximizing the rate

of communication (multiplexing gain) is outage-optimal. At the low-SNR extreme, reliability of signal reception dominates and the beamforming channel is optimal. In the medium-SNR regime, however, the outage-optimal configuration depends on the desired level of operational reliability (outage probability) and a rate-reliability tradeoff is observed. When a low level of channel outage is acceptable, a channel configuration that enhances ρ_{rx} is optimal whereas if higher levels of channel outage are tolerable, then the multiplexing gain (rate) can be increased up to the point where the targeted outage level is met.

We also present an interpretation of our results in the light of recent connections between mutual information and minimum mean-squared error (MMSE) estimation [14], [15]. For all channel configurations in the family, the received SNR per parallel channel, ρ_{rx} , equals the SNR at the output of a linear MMSE receiver for the signal transmitted from any transmit dimension. Moreover, the mean-squared error (MSE) at the receiver is also identical for all channels and equals $\frac{1}{1+\rho_{rx}}$. The invariance of these quantities across the different transmit dimensions is due to *regularity* of the family of channels considered.⁵ Thus the beamforming configuration (which maximizes ρ_{rx}) trades off the number of data-streams for the MSE of the individual data streams. The multiplexing configuration corresponds to the other extreme in this tradeoff while the ideal configuration leads to a robust choice in this *rate-distortion* tradeoff between the number of data-streams and the MSE of the individual data-streams.

Finally, while in this paper we only deal with “thought experiments” where the impact of different channel configurations on capacity is studied, we point out that all channel configurations in the family can be realized in practice by systematically adapting the antenna spacings at the transmitter and the receiver to the level of sparsity and the operating transmit SNR [12]. In particular, when the paths are randomly distributed over the angular spreads, the multiplexing channel corresponds to maximum antenna spacings at both ends, the ideal channel corresponds to medium spacings at both ends, and the beamforming channel can be realized by closely spaced antennas at the transmitter and maximally spaced antennas at the receiver [12]. This connection with reconfigurable antenna arrays is beyond the scope of this paper and the readers are referred to [12] for details.

⁵Let \mathbf{H} be an $N_r \times N_t$ random matrix with independent entries and let the variance of $\mathbf{H}(m, n)$ be given by $\Psi(m, n)$. A channel is called column-regular if $\{\sum_{m=1}^{N_r} \Psi(m, n)\}$ 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 [16].

Connections to Related Works: Many recent works point to the optimality of adaptive array configurations in [12] proposed in achieving the fundamental limits characterized in this paper. Particular among these works are [17] and [18]. It is shown in [17] that in the low-SNR setting, adapting the antenna spacings to excite only one channel eigenmode is optimal while in the high-SNR regime, adapting the arrays to excite all the channel eigenmodes is optimal from a capacity perspective. The same authors conjecture that the number of eigenmodes excited by the optimal array configuration is a monotonic function of SNR and leave the achievability of such an optimal scaling law as an open problem. We note that the beamforming and multiplexing configurations are the precise characterization of the low- and the high-SNR solutions of [17]. The ideal MIMO channel concept we develop resolves the conjecture posed in [17] for the intermediate SNR's. Furthermore, our work differs from [17] on two main counts: 1) we provide a physically motivated, systematic and constructive methodology for characterizing the optimal channel configurations as a function of SNR, and 2) we employ fundamentally different techniques, based on random matrix theory, to study the impact of channel configurations (which are realizable with reconfigurable arrays) on capacity.

Studies on error exponents for correlated channels in the low-SNR regime [18] also indicate that closely spaced antennas, corresponding to the beamforming configuration, are optimal from a reliability viewpoint. Furthermore, the notion of channel sparsity can be extended to time-varying and wideband channels as well [19], [20], [13]. In particular, a fundamental tradeoff between channel learnability and achievable diversity is reported in [20] for time-varying, wideband multipath channels from an error exponent (reliability) viewpoint.

Organization: Section II introduces the sparse channel modeling framework and the precise problem formulation is stated in Section III. Section IV characterizes a fundamental limit on capacity scaling in sparse channels and characterizes the (ideal) channel configuration that achieves this limit. In Section V, we study capacity as a function of transmit SNR for different channel configurations and characterize the optimal configuration at any operating SNR. Section VI discusses relations to other recent works. Concluding remarks and directions for future research are discussed in Section VII.

Notation: We use upper- and lower-case symbols to denote matrices and vectors respectively. $X(m, n)$ denotes the entry in the m -th row and n -th column of \mathbf{X} . We will use the simpler nota-

tion $\mathbf{X}(m)$ for the m -th diagonal entry of \mathbf{X} . \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 $\text{Tr}(\mathbf{X})$ and $\det(\mathbf{X})$, respectively. The upper-case $\mathbf{E}[\cdot]$ stands for the expectation operator and $\text{Pr}(\cdot)$ stands for the probability of an event. $\mathcal{CN}(\mu, \sigma^2)$ denotes the complex normal distribution with mean μ and variance σ^2 . The symbols λ , ρ and N are in general used for eigenvalues, SNR and antenna dimensions, respectively while the subscripts *erg* and *out* stand for ergodic and outage, respectively. \mathbf{I} stands for the identity matrix and we also use the standard big-O (\mathcal{O}) and little-o (o) notations.

II. MIMO CHANNEL MODELING FOR SPARSE MULTIPATH

We consider a single-user narrowband MIMO communication system equipped with ULA's of N antennas at the transmitter and the receiver. The N -dimensional received signal \mathbf{y} and the N -dimensional transmitted signal \mathbf{x} are related by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (3)$$

where $\mathbf{n} \in \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the additive white Gaussian noise (AWGN) and \mathbf{H} is the channel matrix coupling the transmit and the receive antennas. The transmit power is denoted by $\rho = \mathbf{E}[\|\mathbf{x}\|^2]$ which also denotes the total transmit SNR. Initial studies on multi-antenna systems use the so-called *i.i.d. model* where the entries of \mathbf{H} are assumed to be i.i.d. Gaussian random variables [2], [1]. However, these models are not suitable for applications where large antenna spacings or a rich scattering environment are not possible. Furthermore, physical arguments and recent experimental results indicate that rich multipath is an exception in practice rather than the norm. We next provide a brief review of physical and virtual channel modeling, and related modeling approaches, to motivate and develop a model for sparse multipath channels.

A. Physical and Virtual Channel Modeling

A physical multipath MIMO channel can be accurately modeled as

$$\mathbf{H} = \sum_{\ell=1}^{L(N)} N\beta_{\ell} \mathbf{a}_r(\theta_{r,\ell}) \mathbf{a}_t^H(\theta_{t,\ell}) \quad (4)$$

where the transmitter and receiver arrays are coupled through $L(N)$ propagation paths with complex path gains $\{\beta_{\ell}\}$, Angles of Departure (AoD) $\{\theta_{t,\ell}\}$ and Angles of Arrival (AoA) $\{\theta_{r,\ell}\}$. In

(4), $\mathbf{a}_r(\theta_r)$ and $\mathbf{a}_t(\theta_t)$ denote the receiver response and transmitter steering vectors for receiving/transmitting in the normalized direction θ_r/θ_t and $\mathbf{a}_r(\theta_r)$ is defined as

$$\mathbf{a}_r(\theta_r) = \frac{1}{\sqrt{N}} [1, e^{-j2\pi\theta_r}, \dots, e^{-j2\pi(N-1)\theta_r}]^T \quad (5)$$

and $\mathbf{a}_t(\theta_t)$ is defined similarly. The $1/\sqrt{N}$ factor in the definition of the steering/response vectors ensures that they are of unit norm, and the factor N in (4) accounts for this normalization. The angle θ is related to the physical angle (in the plane of the arrays) $\phi \in [-\pi/2, \pi/2]$ as $\theta = d \sin(\phi)/\lambda$ where d is the antenna spacing and λ is the wavelength of propagation. We consider $d = \lambda/2$ antenna 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 the scattering paths (AoA's and AoD's) remains 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 environments, it is rather difficult and cumbersome for capacity analysis due to the non-linear dependence of \mathbf{H} on the propagation parameters in (4). Instead, we resort to the virtual channel representation [6], [4] that characterizes a physical channel via the coupling between spatial beams in fixed virtual transmit and receive directions in the far-field. The virtual representation of \mathbf{H} is given by

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

where $\{\tilde{\theta}_{r,m} = \frac{m}{N}\}$ and $\{\tilde{\theta}_{t,n} = \frac{n}{N}\}$ are fixed virtual receive and transmit angles that uniformly sample the unit θ period and result in unitary discrete Fourier matrices \mathbf{A}_t and \mathbf{A}_r in (6). Thus, \mathbf{H} and \mathbf{H}_v are unitarily equivalent: $\mathbf{H}_v = \mathbf{A}_r^H \mathbf{H} \mathbf{A}_t$. The virtual representation is linear and is characterized by the matrix \mathbf{H}_v .

A key property of the virtual representation is that it induces a partitioning of propagation paths [6], [4]: each $\mathbf{H}_v(m, n)$ is associated with a set of physical paths – paths whose AoD's and AoA's lie within the intersection of the n -th transmit beam and m -th receive beam – and is approximately equal to the sum of the gains of the corresponding paths. Thus, distinct virtual channel coefficients correspond to approximately disjoint subsets of paths and as a result the

virtual channel coefficients are approximately independent due to the statistical independence between the complex path gains. For simplicity, we will assume that the virtual channel coefficients are statistically independent zero-mean Gaussian random variables (Rayleigh fading). This assumption has been validated with experimental measurement results [7], [21].

The virtual representation applies to ULA's at the transmitter and the receiver. The concept of the virtual representation can be extended to more general array geometries if [22], [21]: i) the auto- and cross-correlation matrices of the columns of \mathbf{H} share the same set of eigenvectors, and ii) the auto- and cross-correlation matrices of the rows of \mathbf{H} share the same set of eigenvectors. Under these assumptions, it is shown in [22], [21] that \mathbf{H} admits the following canonical decomposition analogous to the virtual representation: $\mathbf{H} = \mathbf{U}_r \mathbf{H}_{\text{ind}} \mathbf{U}_t^H$, where \mathbf{U}_r and \mathbf{U}_t are unitary matrices defined by the eigenvectors of the receive and transmit covariance matrices ($\mathbf{E}[\mathbf{H}\mathbf{H}^H]$ and $\mathbf{E}[\mathbf{H}^H\mathbf{H}]$), respectively, and \mathbf{H}_{ind} is a random matrix with statistically independent entries. \mathbf{U}_r , \mathbf{U}_t and \mathbf{H}_{ind} serve the roles of \mathbf{A}_r , \mathbf{A}_t and \mathbf{H}_v in the virtual representation, respectively.

Another commonly used model for correlated MIMO channels is the separable correlation model, also known as the ‘‘Kronecker product’’ model (see, e.g., [3]), where \mathbf{H} is defined as in the relationship $\mathbf{H} = \mathbf{U}_r \mathbf{D}_r^{1/2} \mathbf{H}_{\text{iid}} \mathbf{D}_t^{1/2} \mathbf{U}_t^H$ with \mathbf{H}_{iid} , the i.i.d. matrix and \mathbf{D}_r and \mathbf{D}_t , the diagonal matrices consisting of the eigenvalues of the receive and transmit covariance matrices, respectively. The Kronecker model is a special case of the canonical model (and the virtual representation when ULA's are used): it models the channel based on the marginal (transmit and receive) statistics, whereas the canonical and virtual models characterize the channel based on joint transmit-receive statistics. As a result, the Kronecker model is generally inadequate in accurately capturing channel statistics and leads to biases in predicting performance metrics in realistic scattering environments (see, e.g., [23], [24], [22], [21].)

While the mathematical development in the rest of the paper could be interpreted in terms of the canonical model, we will focus on the virtual representation due to the intuitive physical interpretation associated with it (e.g., the spatial eigenfunctions are beams in the virtual directions).

B. Channel Statistics and Degrees of Freedom

The statistics of \mathbf{H} are characterized by the virtual channel power matrix Ψ : $\Psi(m, n) = \mathbf{E}[|\mathbf{H}_v(m, n)|^2]$ since $\mathbf{H}_v(m, n)$ are statistically independent. The matrices \mathbf{A}_t and \mathbf{A}_r consti-

tute 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 (correlation matrices in the virtual domain). We can interpret Ψ as the joint distribution of channel power as a function of transmit and receive virtual angles. $\mathbf{\Lambda}_t$ and $\mathbf{\Lambda}_r$ serve as the corresponding marginal distributions: $\mathbf{\Lambda}_r(m) = \sum_n \Psi(m, n)$ and $\mathbf{\Lambda}_t(n) = \sum_m \Psi(m, n)$.

We next introduce the notion of degrees of freedom (DoF) in the MIMO channel to motivate and develop a model for sparse multipath channels.

Definition 1: Independent DoF and Channel Power. Let \mathbf{H} , \mathbf{H}_v and Ψ denote the channel matrix, its virtual channel matrix and the virtual channel power matrix, respectively. We define D , the number of independent DoF afforded by \mathbf{H} as the number of entries in \mathbf{H}_v with non-vanishing power

$$D = |\{(m, n) : \Psi(m, n) > 0\}|. \quad (7)$$

We also define the *channel power*, ρ_c , as

$$\rho_c \triangleq \mathbf{E}[\text{Tr}(\mathbf{H}\mathbf{H}^H)] = \mathbf{E}[\text{Tr}(\mathbf{H}_v\mathbf{H}_v^H)] = \sum_{m=1}^N \sum_{n=1}^N \Psi(m, n) = \sum_{\ell=1}^{L(N)} N^2 \mathbf{E}|\beta_\ell|^2 \quad (8)$$

where the last equality corresponds to the physical model (4). Due to the path partitioning property of the virtual representation, the DoF reflect the number of *resolvable* sets of paths that contribute to channel power. ■

We note that in practice, the threshold of 0 in (7) could be replaced by an appropriate pseudo-noise level $\epsilon > 0$, wherein channel entries with variances larger than ϵ contribute significantly to channel power and hence channel capacity. Furthermore, $L(N)$ is in general an increasing function of N in (4) and (8) since for a fixed antenna spacing, as N increases, the number of physical paths captured by the arrays increases due to larger array apertures. The following lemma relates channel power scaling with the scaling of D and $L(N)$.

Lemma 1: Assume that the non-vanishing entries in $\Psi(m, n)$ and $N^2 \mathbf{E}|\beta_\ell|^2$, which represent the contribution to channel power of resolvable sets of paths and physical paths, respectively, are $\mathcal{O}(1)$. Then, the number of independent DoF, $D(N)$, and the number of physical paths, $L(N)$,

scale with antenna dimensions at the same rate as the channel power:

$$\rho_c(N) \sim D(N) \sim L(N). \quad (9)$$

Proof: From (8) and the $\mathcal{O}(1)$ assumption above, it follows that

$$\left\{L(N), D(N)\right\} \cdot \mathcal{O}(1) \leq \rho_c(N) \leq \left\{L(N), D(N)\right\} \cdot \mathcal{O}(1). \quad (10)$$

■

Remark 1: We note that the assumption of $N^2 \mathbf{E}|\beta_\ell|^2 \sim \mathcal{O}(1)$ is a technical one related to the fact that the physical scatterers in (4) are modeled as point scatterers. In practice, scattering objects have a finite dimension and as the array aperture increases (with N), the spatial resolution of the arrays increases in direct proportion to N (the beamwidths decrease as $1/N$), and as a result the effective power contribution from each scatterer $\mathbf{E}|\beta_\ell|^2$ decreases as $1/N^2$ which is compensated by the N^2 transmit-receive array gain, resulting in $N^2 \mathbf{E}|\beta_\ell|^2 \sim \mathcal{O}(1)$. A more detailed discussion of this assumption is beyond the scope of this paper. Our main interest is in the scaling of the DoF, D , with N .

C. Sparse Channel Modeling

The implicit assumption that is prevalent in all existing works on channel capacity is that $\rho_c(N) \sim \mathcal{O}(N^2)$. This assumption is a legacy of the i.i.d. model used in initial results. A key motivation for sparse channel modeling is that this assumption is unrealistic in practice, both from a capacity scaling perspective (based on power conservation arguments) and from the perspective of finite dimensional systems (based on experimental measurement studies). We elaborate on this next.

From a capacity scaling viewpoint, if $\rho_c \sim \mathcal{O}(N^2)$, then for any signaling scheme that is efficient and does not waste power over the weakest spatial dimensions, the received signal power, $\text{Tr}(\mathbf{E}[\mathbf{y}\mathbf{y}^H])$, increases linearly with N even though the transmit power is fixed. In particular, the received power with an i.i.d. input that excites all dimensions satisfies

$$\begin{aligned} \text{Tr}(\mathbf{E}[\mathbf{y}\mathbf{y}^H]) &\stackrel{(a)}{\geq} \text{Tr}(\mathbf{E}[\mathbf{H}\mathbf{x}\mathbf{x}^H\mathbf{H}^H]) \\ &= \text{Tr}(\mathbf{Q}\mathbf{E}[\mathbf{H}^H\mathbf{H}]) \\ &\stackrel{(b)}{\geq} \rho \frac{\rho_c(N)}{N} = \mathcal{O}(N) \end{aligned} \quad (11)$$

where (a) follows from ignoring the noise contribution, ρ is the transmit power, and (b) follows from using an i.i.d. input⁶ \mathbf{x} and the channel power normalization of $\mathcal{O}(N^2)$. While this linear scaling of received power (with antenna dimensions) for a fixed transmit power ρ may be reasonable for small antenna dimensions⁷, it cannot be justified with increasing number of antennas since this would violate the power conservation principle: total average received power cannot exceed the total average transmit power. As a consequence, for sufficiently large N , $\rho_c(N) < \mathcal{O}(N^2)$ and hence $D(N) < N^2$ in view of Lemma 1.

From the viewpoint of capacity of fixed dimension systems, measurement campaigns and statistical characterizations of typical propagation environments, even for relatively small $N \sim 8$, show that the DoF in the channel tend to be fewer than the maximum allowable N^2 degrees of freedom; see, e.g., [7], [8], [9], [10]. Incidentally, this observation, coupled with the prevalent power normalization of $\rho_c \sim \mathcal{O}(N^2)$, leads to the misleading conclusion that “a correlated channel has higher capacity than an i.i.d. channel at sufficiently low SNR” that has been reported in some recent works [25], [26], [27]. In effect, the channel power is distributed in fewer DoF in a correlated channel than the N^2 DoF in an i.i.d. setting, resulting in a higher power per DoF in correlated channels which results in higher capacity at low SNR’s [22].

In view of the above observations, we abstract the notion of sparse multipath channels in the following definition.

Definition 2: Sparse Virtual Channels. For a given fixed N , an $N \times N$ virtual channel matrix \mathbf{H}_v is sparse if it contains $D < N^2$ non-vanishing coefficients (DoF) corresponding to resolvable sets of paths. For simplicity, we assume that each non-vanishing coefficient is $\mathcal{CN}(0, 1)$ reflecting the power contributed by the *unresolvable* paths associated with it. From a scaling perspective, the DoF and the channel power in a sparse channel scale at a sub-quadratic rate with N :

$$D(N) \sim \rho_c(N) \sim \mathcal{O}(N^\gamma) , \gamma \in (0, 2] \quad (12)$$

■

Sparse virtual channel matrices provide a model for spatial correlation in \mathbf{H} : in general, the sparser the \mathbf{H}_v in the virtual domain, the higher the correlation in the antenna domain \mathbf{H} . In

⁶It can be shown rigorously that for a channel with $\rho_c(N) = \mathcal{O}(N^2)$, the i.i.d. input that excites all dimensions is efficient from a capacity scaling perspective [5].

⁷For small antenna dimensions, the array gain compensates for the $1/r^2$ loss in power due to propagation.

(12), smaller values of γ reflect sparse channels, whereas $\gamma = 2$ corresponds to a rich scattering environment and yields the i.i.d. channel model. We also note that even though the assumption that all the non-vanishing virtual channel entries have a unit variance is simplistic, it does capture the underlying trends in capacity scaling with N and also in capacity as a function of SNR. We expect the results and conclusions of this paper to hold even if we assumed that the non-vanishing virtual coefficients satisfy $0 < \Psi(m, n) \leq K < \infty$. The important quantity from the viewpoint of sparsity is the number of DoF, D , relative to N^2 . Moreover, as will see, the 0-1 variance model in the definition of sparse channels is insightful and greatly facilitates closed-form capacity analysis.

Definition 3: Mask matrices. It is convenient to model a sparse $N \times N$ matrix \mathbf{H}_v with $D < N^2$ DoF as

$$\mathbf{H}_v = \mathbf{H}_{\text{iid}} \odot \mathbf{M} \quad (13)$$

where \odot denotes elementwise product, \mathbf{H}_{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 follows that the virtual channel power matrix $\Psi = \mathbf{M}$ and the entries of Λ_r and Λ_t represent the number of non-zero elements in the rows and columns of \mathbf{M} , respectively. ■

III. PROBLEM FORMULATION

Sparse multipath channels afford a fundamental new degree of freedom that is not available in rich channels: the spatial distribution or configuration of the $D < N^2$ DoF in the available N^2 transmit-receive dimensions in the $N \times N$ virtual channel matrix. While the different channel configurations can be created in practice with reconfigurable antenna arrays (see [12] and Section VI), the focus of this paper is to study the impact of different channel configurations on capacity. We assume perfect knowledge of \mathbf{H} at the receiver (coherent setting) while the transmitter only knows the channel statistics, $\{\Psi(m, n)\}$, from which the channel DoF, D , can be determined. Specifically, we study the impact of channel configuration from two perspectives: 1) capacity scaling as a function of antenna dimensions N , and 2) capacity optimization as a function of transmit SNR for a fixed N .

For fading channels, capacity is a random variable that depends on the channel realization \mathbf{H} and to stress this aspect we denote it here by $C(\mathbf{H})$. We now briefly review the two common

capacity metrics used in the literature: ergodic and outage capacities.

A. Capacity Metrics

A.1 Coherent Capacity

Ergodic capacity is the mean of the capacity random variable where the expectation operation is over the statistics of \mathbf{H} . The ergodic capacity, $C_{\text{erg}}(N, \rho)$, of a MIMO channel at a transmit SNR of ρ is given by [1], [2]

$$\begin{aligned} C_{\text{erg}}(N, \rho) \triangleq \mathbf{E}_{\mathbf{H}} [C(\mathbf{H})] &= \max_{\mathbf{Q}: \text{Tr}(\mathbf{Q}) \leq \rho} \mathbf{E} [\log_2 \det(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^H)] \\ &= \max_{\mathbf{Q}: \text{Tr}(\mathbf{Q}) \leq \rho} \mathbf{E} [\log_2 \det(\mathbf{I} + \mathbf{A}_r \mathbf{H}_v \mathbf{A}_t^H \mathbf{Q} \mathbf{A}_t \mathbf{H}_v^H \mathbf{A}_r^H)] \\ &= \max_{\mathbf{Q}_v: \text{Tr}(\mathbf{Q}_v) \leq \rho} \mathbf{E} [\log_2 \det(\mathbf{I} + \mathbf{H}_v \mathbf{Q}_v \mathbf{H}_v^H)] \end{aligned} \quad (14)$$

where $\mathbf{Q} = \mathbf{E} [\mathbf{x}\mathbf{x}^H]$ is the transmit covariance matrix which is trace-constrained by ρ , and $\mathbf{Q}_v = \mathbf{A}_t^H \mathbf{Q} \mathbf{A}_t$ is the covariance matrix in the virtual domain.

When channel statistics are available at the transmitter, it is shown in [25], [22] that capacity is achieved by a diagonal \mathbf{Q}_v in the virtual domain; that is, independent signaling over different virtual beam directions is optimal from a capacity viewpoint. When the channel is regular (see Footnote 5) or when no information is available at the transmitter, uniform-power input, $\mathbf{Q}_v = \frac{\rho}{N} \mathbf{I}$, is optimal [16]. For general, non-regular correlated channels, full-rank uniform-power \mathbf{Q}_v is optimal at high-SNR's, whereas a rank-1 (beamforming) input is optimal at low-SNR's (the power is focused on the column of \mathbf{H}_v with the largest power). As ρ is increased from low- to high-SNR's, the rank of the optimal \mathbf{Q}_v increases from 1 to N . Note that at low- and medium-SNR's, the optimal input excites a subset of the transmit dimensions and hence couples only a fraction of the channel power to the receiver. This loss in channel power is significant at low-SNR's. Capacity optimization as a function of SNR studied in this paper, in which the configuration of the D DoF is also considered, is aimed at avoiding this channel power loss at lower SNR's.

A.2 Outage Capacity

The insufficiency of the ergodic capacity as the sole measure of the information rate that a MIMO channel can support has also been well-documented [28], [2]. The notion of outage

capacity was introduced in [28] to characterize the information rate that a MIMO channel can support with a certain guaranteed probability. Specifically, the outage capacity $C_{\text{out},q}(N, \rho)$ at an outage of $q\%$ is defined as

$$C_{\text{out},q}(N, \rho) \triangleq \sup_{R \geq 0} (R) \quad \text{s.t.} \quad \Pr(\log_2 \det [I + \mathbf{H}\mathbf{Q}\mathbf{H}^H] < R) \leq \frac{q}{100}. \quad (15)$$

The outage capacity is the maximum rate of communication that is guaranteed for at least $(100 - q)\%$ of the channel realizations.

While the outage capacity depends on the distribution of $C(\mathbf{H})$, Gaussian approximations of $C(\mathbf{H})$ have been shown to be close to the true values in the context of i.i.d. channels [29], semi-correlated⁸ channels [30], [31], [32], and spatially uncorrelated, but temporally correlated wide-band channels [33]. The weak convergence of the capacity random variable to a Gaussian (in the limit of antenna dimensions) is mathematically proved for the i.i.d. channel in [34] and [35], and under very general assumptions on \mathbf{H} in [36]. Furthermore, numerical and theoretical studies show that the rate of convergence to the asymptotic limits is on the order of the inverse of antenna dimensions, see e.g., [36]. Thus analysis of outage capacity reduces to computing the mean and the variance of the capacity random variable.

Given the mean and variance of $C(\mathbf{H})$ ($C_{\text{erg}}(N, \rho)$ and $\sigma^2(N, \rho)$, respectively), the outage capacity is given by [34]

$$C_{\text{out},q}(N, \rho) = C_{\text{erg}}(N, \rho) - x_q \sigma(N, \rho) + o(1) \quad (16)$$

where x_q is the unique solution of

$$\text{erfc}\left(x_q/\sqrt{2}\right) = 2q \quad (17)$$

with $\text{erfc}(\cdot)$ denoting the complementary error function. Closed-form results for $C_{\text{erg}}(N, \rho)$ are known in many cases while the corresponding results for $\sigma^2(N, \rho)$ (and hence $C_{\text{out},q}(N, \rho)$) are relatively rare.

B. Problem Formulation

Since the channel capacity can be equivalently cast in the virtual domain (due to unitary equivalence; see (14)), when there is no confusion, we will use \mathbf{H} and \mathbf{Q} to represent the virtual

⁸A semi-correlated channel is a channel with a separable correlation model where either the transmitter or the receiver is uncorrelated.

counterparts, \mathbf{H}_v and \mathbf{Q}_v , in the rest of the paper. As mentioned earlier, we are interested in studying the impact of different sparse channel configurations on capacity. Thus, for ergodic capacity, we want to solve the following optimization problem

$$C_{\text{opt, erg}}(N, \rho, D) = \max_{\mathcal{H}(D)} \max_{\mathbf{Q}: \text{Tr}(\mathbf{Q}) \leq \rho} \mathbf{E}_{\mathbf{H}} [\log_2 \det (\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^H)] \quad (18)$$

where $\mathcal{H}(D)$ is the set of all sparse virtual channel matrices with $D < N^2$ DoF, as defined in Definitions 2 and 3,

$$\mathcal{H}(D) = \{\mathbf{H} : \mathbf{H} = \mathbf{H}_{\text{iid}} \odot \mathbf{M}, \text{Tr}(\mathbf{M}^H \mathbf{M}) = D\}. \quad (19)$$

We address the following two fundamental questions in this paper.

1. *Given a channel power (and DoF) scaling law $\rho_c \sim D \sim \mathcal{O}(N^\gamma)$, $\gamma \in (0, 2]$, what is the fastest achievable capacity scaling law? Furthermore, which sparse channel configuration achieves the fastest capacity scaling?*

We show that for any fixed ρ , $C_{\text{opt, erg}}(N, \rho, D)$ cannot scale faster than $\mathcal{O}(\sqrt{\rho_c}) = \mathcal{O}(\sqrt{D})$. We also *explicitly* characterize the mask matrix \mathbf{M} corresponding to the channel configuration that achieves this scaling law. We call this configuration *the ideal MIMO channel*. The ideal channel corresponds to an optimal distribution of the channel DoF in the signal-space dimensions from a capacity scaling perspective.

2. *For a given fixed N , does the capacity at any operating SNR depend on the channel configuration? If so, what channel configuration maximizes capacity at any given operating SNR?*

For a given fixed N , we provide explicit constructions of \mathbf{M} as a function of the operating transmit SNR, ρ , that characterize the channel configurations that maximize capacity at the corresponding ρ . We also show that for all practical purposes, three channel configurations suffice to maximize capacity over the entire 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.

Note that the capacity formulation in (18) is fundamentally different from the conventional formulation (14) where the optimization is only over the input covariance matrix \mathbf{Q} . Furthermore, while closed-form expressions for capacity have only been obtained in the limit of low- or

high-SNR's in the conventional formulation, this new formulation enables closed-form capacity characterization at all SNR's.

C. A Structured Family of Channels

To address the above two questions, we introduce a structured family of mask matrices \mathbf{M} (and corresponding \mathbf{H} 's via Definition 3) that characterize systematically varying configurations of the $D < N^2$ DoF in the available N^2 transmit-receive dimensions in the virtual channel matrix \mathbf{H} . For any given D and N , the family is defined by two parameters (p, q) such that $D = pq$. The D DoF are configured to generate a channel matrix that can support p parallel channels and $q = D/p$ DoF per parallel channel. As we will show, the ergodic capacities of all channels in the family admit the following simple and intuitive expression

$$\begin{aligned} C_{\text{erg}}(N, \rho) &\approx p \log \left(1 + \rho \frac{q}{p} \right) \\ &= p \log (1 + \rho_{rx}) \end{aligned} \quad (20)$$

where the received SNR per parallel channel ρ_{rx} is defined as $\frac{\mathbf{E}[\|\mathbf{H}\mathbf{x}\|^2]}{p} = \rho \frac{D}{p^2} = \rho \frac{q}{p}$.

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

$$\mathbf{M}((n+m) \bmod r, n) = 1, \quad 1 \leq n \leq p, \quad q_- \leq m \leq q_+ \quad (21)$$

where $q_- = \lceil -(q-1)/2 \rceil$ and $q_+ = \lceil (q-1)/2 \rceil$. ■

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 diagonals⁹. Thus, the corresponding channel matrices \mathbf{H} defined via (13) are regular (see Footnote 5), with $\text{rank}(\mathbf{H}) = \min(r, p) = p$ for which the uniform-power input over the p parallel channels is the optimal transmission scheme [16].

The family of channels is illustrated in Fig. 4. The feasible range for p depends on the value of D relative to N^2 as formalized in the following parametric definitions for D , p and q .

⁹In the terminology of [4], \mathbf{M} is a q -connected p -dimensional matrix.

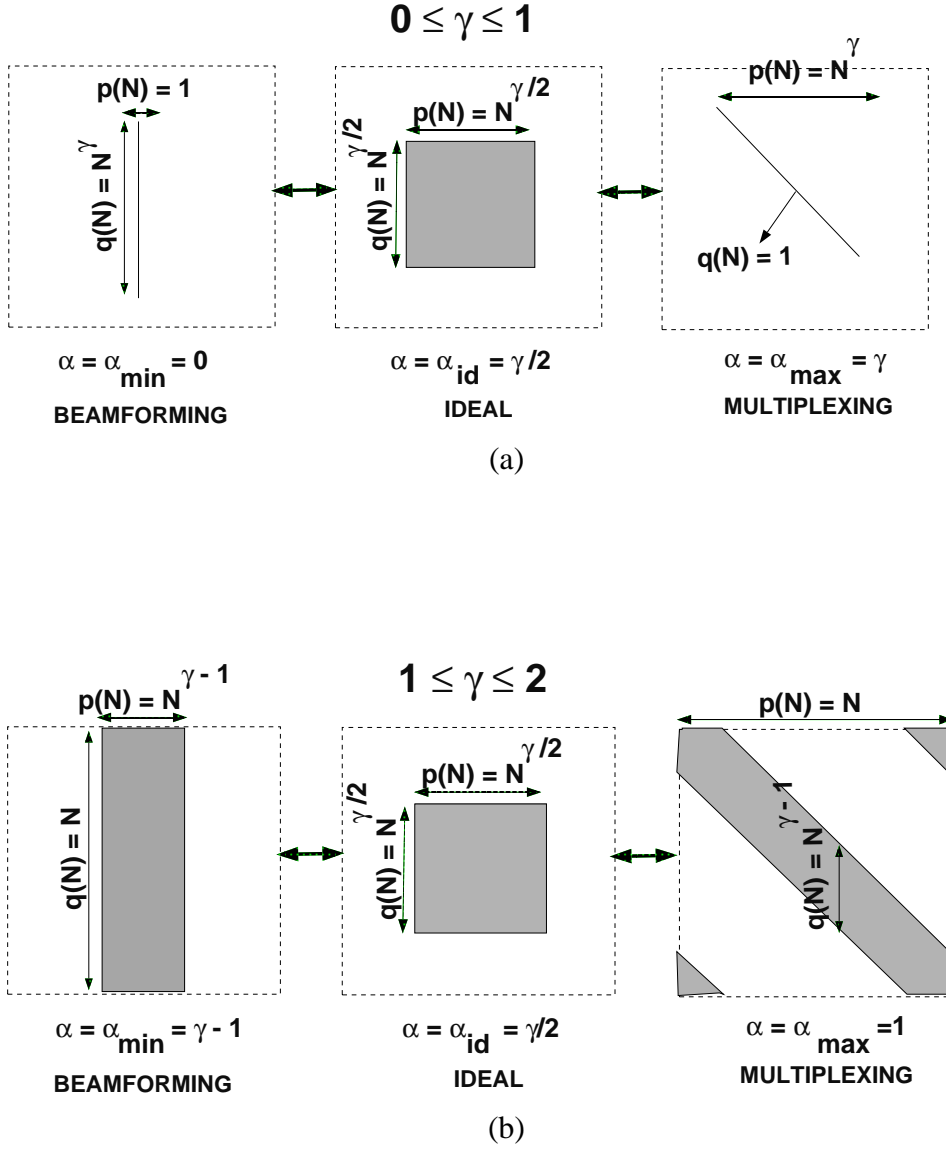


Fig. 4. A schematic illustrating the family of channels: (a) $\gamma \in [0, 1]$ and (b) $\gamma \in [1, 2]$.

Definition 5: Parameterization of D , p , and q . Let $D = \rho_c = N^\gamma$, $\gamma \in (0, 2]$; $p = \delta_p N^\alpha$, $\delta_p = \mathcal{O}(1)$, $\alpha \in [0, 1]$; and $q = \frac{1}{\delta_p} N^{\gamma-\alpha}$ in Definition 4. For a given γ , the feasible range for α is

$$\alpha_{\min} \triangleq \max(\gamma - 1, 0) \leq \alpha \leq \min(\gamma, 1) \triangleq \alpha_{\max}. \quad \blacksquare$$

Note that for $0 < \gamma \leq 1$, $\alpha_{\min} = 0$ and $\alpha_{\max} = \gamma$. For $1 \leq \gamma \leq 2$, $\alpha_{\min} = \gamma - 1 \geq 0$ and $\alpha_{\max} = 1$. We next identify three distinct regimes for p which correspond to three distinct canonical channel configurations and result in distinct capacity behaviors, both from a scaling perspective and also as a function of SNR.

Beamforming regime (\mathbf{H}_{bf}): $\alpha \in [\alpha_{\min}, \frac{\gamma}{2}) \iff p = o(q)$ and $\rho_{rx} \rightarrow \infty$ as $N \rightarrow \infty$.

Ideal regime (\mathbf{H}_{id}): $\alpha = \frac{\gamma}{2} \iff p = \mathcal{O}(q)$. Define $\kappa_0 = \lim_N \frac{q}{p} = \frac{1}{\delta_p^2} = \mathcal{O}(1)$. In this regime, $\rho_{rx} \rightarrow \kappa_0 \rho$ as $N \rightarrow \infty$.

Multiplexing regime (\mathbf{H}_{mux}): $\alpha \in (\frac{\gamma}{2}, \alpha_{\max}] \iff q = o(p)$ and $\rho_{rx} \rightarrow 0$ as $N \rightarrow \infty$.

The channels in the family morph from the beamforming regime to the multiplexing regime through the ideal regime as p is increased from $p_{\min}(\alpha_{\min})$ to $p_{\max}(\alpha_{\max})$ through $p_{\text{id}}(\alpha_{\text{id}})$, as illustrated in Fig. 4. Three canonical channel configurations, one from each regime, corresponding to $\alpha = \alpha_{\min}, \frac{\gamma}{2}$ and α_{\max} , will be referred to as *beamforming*, *ideal* and *multiplexing* channels and will be denoted by \mathbf{H}_{bf} , \mathbf{H}_{id} and \mathbf{H}_{mux} , respectively. We note that when $\gamma = 2$, $\alpha_{\min} = \alpha_{\max} = 1$, all channel configurations essentially reduce to $\mathbf{H}_{\text{id}} = \mathbf{H}_{\text{id}}$ in which $p = q = N$ and the resulting channel is the familiar $N \times N$ i.i.d. channel. In the rest of the paper, we will assume that p and q are integers for any feasible value of p . This is because, as shown in Appendix I, non-integer values of (p, q) for any given α and δ_p can be realized as a linear combination of four integer configurations: $(\lfloor p \rfloor, \lfloor q \rfloor)$, $(\lfloor p \rfloor, \lceil q \rceil)$, $(\lceil p \rceil, \lfloor q \rfloor)$, $(\lceil p \rceil, \lceil q \rceil)$. The desired channel configuration with non-integer (p, q) can be realized by appropriately time-sharing between the four integer (p, q) configurations.

IV. CAPACITY SCALING AS A FUNCTION OF ANTENNA DIMENSIONS

In this section we study the asymptotic behavior of ergodic and outage capacity as a function of the number of antennas for the family of sparse channels introduced in Sec. III-C. We first obtain a fundamental limit on ergodic capacity scaling for any given scaling of channel power or DoF. We then show that the asymptotic ergodic capacity of all channels in the family admits a simple closed-form approximation. In particular, the ideal channel configuration achieves the limit on ergodic capacity scaling and thus yields the fastest scaling for any power or DoF scaling. We then demonstrate the asymptotic accuracy of the approximate formula and also analyze the asymptotic outage capacity behavior of the different channel configurations.

A. Ergodic Capacity

A.1 Fundamental Limit on Capacity Scaling

We first compute a fundamental limit on capacity scaling for sparse channels with a given channel power or DoF scaling.

Theorem 1: Let $\mathcal{H}(D)$ denote the family of sparse virtual channel matrices with D DoF defined in (19). Then, $C_{\text{opt,erg}}(N, \rho, D)$ defined in (18) admits the following upperbound for any transmit SNR ρ

$$C_{\text{opt,erg}}(N, \rho, D) \leq \mathcal{O}(\sqrt{D}) = \mathcal{O}(\sqrt{\rho_c}). \quad (22)$$

In particular, with $\rho_c \sim D \sim \mathcal{O}(N^\gamma)$, the capacity cannot scale faster than $\mathcal{O}(N^{\frac{\gamma}{2}})$.

Proof: See Appendix II. ■

The importance of the above theorem is that no matter how the spatial DoF are distributed, the capacity of a sparse channel cannot exceed $\mathcal{O}(\sqrt{D}) = \mathcal{O}(\sqrt{\rho_c}) = \mathcal{O}(N^{\frac{\gamma}{2}})$, thus providing a benchmark to compare the capacities of different channel configurations in the family. In [5], we show that if the channel power scales as $\rho_c \sim \mathcal{O}(N^\gamma)$, then uniform power transmission leads to a mutual information scaling of $\mathcal{O}(N^{\gamma-1})$. Note that $\gamma - 1$ is strictly smaller than $\frac{\gamma}{2}$ if $\gamma < 2$. Thus it is of interest to study if the fundamental limit in Theorem 1 is indeed achievable. This is discussed next.

A.2 Capacity Analysis of Different Channel Configurations

We now analyze the capacity of the family of channels introduced in Section III-C. The proofs of the results have been relegated to Appendix III. We show that, remarkably, the ergodic capacity of all channels in the family admits the following simple approximation

$$C_{\text{erg}}(N) \approx p \log(1 + \rho_{rx}) = p \log\left(1 + \rho \frac{q}{p}\right) \triangleq C_{\text{appx}}(N). \quad (23)$$

In the above formula, capacity is controlled by the number of parallel channels $p = \delta_p N^\alpha$ (multiplexing gain) and the received SNR per parallel channel $\rho_{rx} = \rho \frac{q}{p} = \rho \frac{\rho_c}{p^2}$. \mathbf{H}_{bf} ($\alpha = \alpha_{\text{min}}$) and \mathbf{H}_{mux} ($\alpha = \alpha_{\text{max}}$) represent two extremes in which the DoF are distributed to emphasize ρ_{rx} and p , respectively, at the expense of the other quantity. \mathbf{H}_{id} ($\alpha = \alpha_{\text{id}}$) represents an ideal distribution in which ρ_{rx} is kept constant. Increasing α from α_{min} to $\alpha_{\text{id}} = \frac{\gamma}{2}$ converts \mathbf{H}_{bf} to

\mathbf{H}_{id} (p increases while ρ_{rx} decreases) while decreasing α from α_{max} to $\alpha_{\text{id}} = \frac{\gamma}{2}$ converts \mathbf{H}_{mux} to \mathbf{H}_{id} (p decreases and ρ_{rx} increases).

The goal of this sub-section is to establish the validity of this asymptotic approximation to ergodic capacity. However, before we do that, we note that the above approximate capacity expression implies that the ideal configuration with $p = q = \mathcal{O}(\sqrt{\rho_c}) = \mathcal{O}(\sqrt{D})$ achieves the fastest capacity scaling identified in Theorem 1; that is, $C_{\text{erg, id}}(N) \approx \sqrt{D} \log(1 + \rho\kappa_0) = \mathcal{O}(\sqrt{D})$. Furthermore, it is easy to see that since $p = o(q)$ in the beamforming regime and $q = o(p)$ in the multiplexing regime, we have $C_{\text{erg, bf}} = o(C_{\text{erg, id}})$ and $C_{\text{erg, mux}} = o(C_{\text{erg, id}})$. That is, the capacity of the ideal channel configuration dominates the capacity of the beamforming or multiplexing configurations and achieves the fastest scaling limit in Theorem 1.

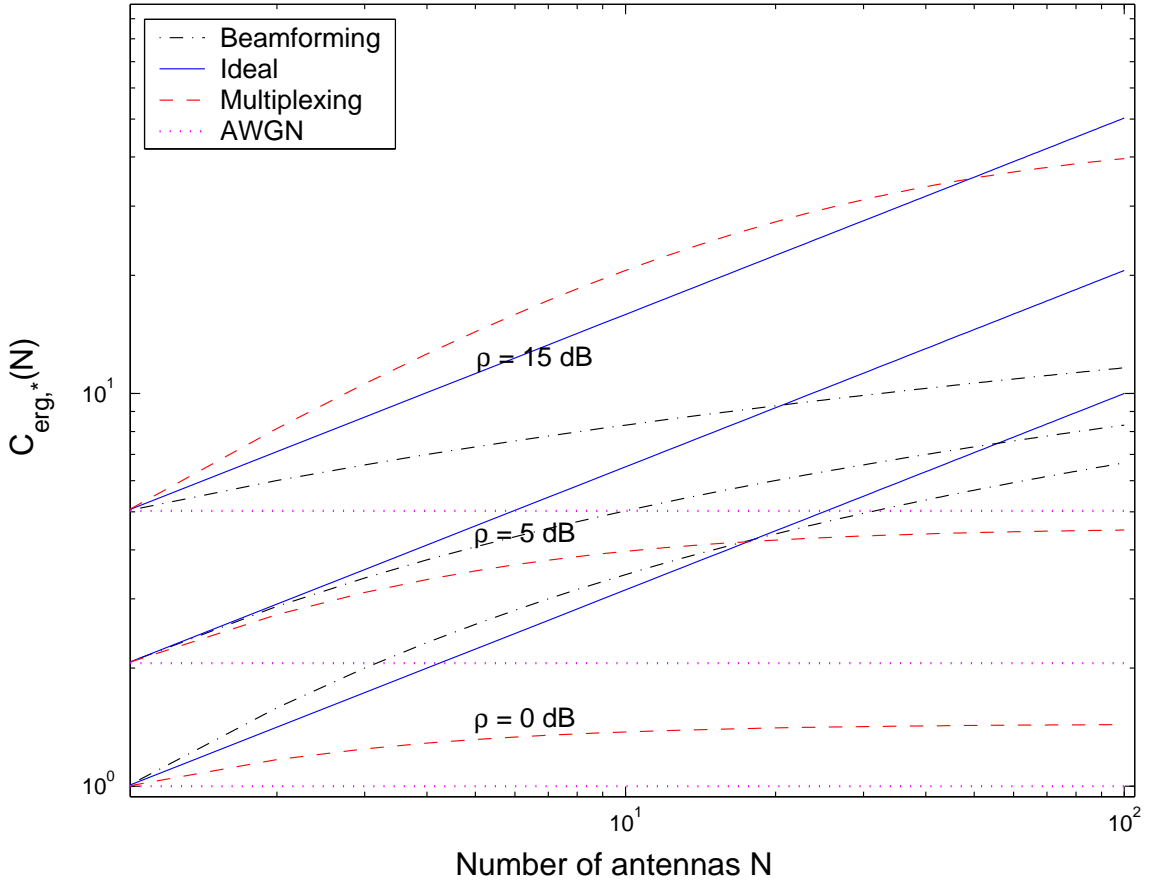


Fig. 5. The capacities of the beamforming, ideal, and multiplexing channel configurations as a function of N for three different values of ρ : 0, 5, and 15 dB. The plots are based on the approximate expression (23) and the single-antenna AWGN capacity is also plotted for comparison.

The optimality of the ideal channel configuration from a scaling perspective is illustrated in Fig. 5 for $D = N^\gamma$, $\gamma = 1$. We use $C_{\text{appx}}(N)$ in (23) to plot $C_{\text{erg, bf}}(N)$ ($p = 1$; $\alpha = \alpha_{\text{min}} = 0$), $C_{\text{erg, id}}(N)$ ($p = \sqrt{N}$; $\alpha = \alpha_{\text{id}} = 0.5$) and $C_{\text{erg, mux}}(N)$ ($p = N$; $\alpha = \alpha_{\text{max}} = 1$) on a log-log scale as a function of N for three SNR's: $\rho = 0, 5$ and 15 dB. The capacity of the single-antenna AWGN channel at the corresponding SNR's is also plotted for comparison. As evident, at low-SNR, the beamforming channel initially dominates for small N – $C_{\text{erg, mux}}(N, \rho) < C_{\text{erg, id}}(N, \rho) < C_{\text{erg, bf}}(N, \rho)$ – but the ideal channel eventually dominates as N gets large. On the other hand, at high-SNR, the multiplexing channel initially dominates – $C_{\text{erg, bf}}(N, \rho) < C_{\text{erg, id}}(N, \rho) < C_{\text{erg, mux}}(N, \rho)$ – but the ideal channel eventually dominates as N gets large. In Section V, we will explore this observation further when we analyze the capacity of the different configurations as a function of SNR for a finite N .

We now demonstrate the accuracy of the approximate capacity expression in (23) for the three regimes in the family of channels. We start with the beamforming regime.

Theorem 2: Let \mathbf{H} be a $q \times p$ channel in the beamforming regime ($p = o(q)$) with $D = pq$ and $\kappa_0 = \frac{q}{p}$. Define $t = \frac{\rho\sqrt{\kappa_0}}{1+\rho\kappa_0}$ and the normalized capacity $C_{\text{norm, bf}}(N) = \frac{C_{\text{erg, bf}}(N)}{C_{\text{appx}}(N)}$. If $\gamma \in (0, 1]$, in the asymptotics of N , the ergodic capacity $C_{\text{erg, bf}}(N)$ is given by

$$C_{\text{erg, bf}}(N) = p \log_2(1 + \rho \kappa_0) = C_{\text{appx}}(N). \quad (24)$$

However, if $\gamma \in (1, 2)$, the asymptotic ergodic capacity is given by

$$\begin{aligned} C_{\text{erg, bf}}(N) &= p \log_2(1 + \rho \kappa_0) + p \log_2 \left(\frac{1 + \sqrt{1 - 4t^2}}{2} \right) \\ &\quad + \frac{\log_2(e)p}{4t^2} \cdot \left(1 - \sqrt{1 - 4t^2} \right) - \frac{\log_2(e)p}{2}, \\ C_{\text{norm, bf}}(N) &= 1 + \Delta C_{\text{bf}}(N) \end{aligned} \quad (25)$$

where the correction term satisfies

$$\begin{aligned} |\Delta C_{\text{bf}}(N)| &\approx \frac{1}{\log_2(1 + \rho \kappa_0)} \cdot \left(\log_2(e) \frac{1}{2\kappa_0} + \log_2 \left(1 - \frac{1}{\kappa_0} - \frac{1}{\kappa_0^2} \right) \right) \\ &= \mathcal{O} \left(\frac{1}{\kappa_0 \log_2(1 + \rho \kappa_0)} \right) \rightarrow 0. \end{aligned} \quad (26)$$

■

Among all channels in the beamforming regime, the smallest value for the correction term is achieved by the beamforming channel that uses $\alpha = \alpha_{\text{min}}$. However, $\kappa_0 \rightarrow \infty$ as $N \rightarrow \infty$

for every channel in this regime and hence the correction term $\Delta C_{\text{bf}}(N) \rightarrow 0$ for all ρ . Thus, $C_{\text{appx}}(N)$ is a very good asymptotic approximation of the exact capacity at all SNR's. For a fixed N , note that $\Delta C_{\text{bf}}(N)$ takes the largest value in the low-SNR regime.

The capacity in the ideal regime is characterized next.

Theorem 3: Let \mathbf{H}_{id} be a $q \times p$ ideal channel ($p = \mathcal{O}(q)$) and let $\kappa_0 = \frac{q}{p}$. If $\kappa_0 \geq 1$, then in the limit of N , the capacity of \mathbf{H}_{id} is given by

$$\begin{aligned} C_{\text{erg, id}}(N) &= q \cdot \left(\log_2(1 + \rho - \rho h) + \frac{1}{\kappa_0} \log_2(1 - \rho h + \rho \kappa_0) - \log_2(e) \frac{h}{\kappa_0} \right) \\ h &= \frac{1}{2} \left[1 + \kappa_0 + \frac{1}{\rho} - \sqrt{\left(1 + \kappa_0 + \frac{1}{\rho}\right)^2 - 4\kappa_0} \right]. \end{aligned} \quad (27)$$

However, if $\kappa_0 < 1$, we have

$$\frac{C_{\text{erg, id}}(N)}{\log_2(e)} = p \cdot \left(\log_e \left(\frac{1 + 2\rho\kappa_0 + \sqrt{1 + 4\rho\kappa_0}}{2} \right) + \frac{\sqrt{1 + 4\rho\kappa_0} - (1 + 2\rho\kappa_0)}{2\rho\kappa_0} \right). \quad (28)$$

■

We now investigate the closeness of $C_{\text{erg, id}}(N)$ to $C_{\text{appx}}(N)$ as a function of SNR.

Proposition 1: Define the normalized capacity $C_{\text{norm, id}}(N) = \frac{C_{\text{erg, id}}(N)}{C_{\text{appx}}(N)} = 1 + \Delta C_{\text{id}}(N)$ where $\Delta C_{\text{id}}(N)$ is the correction term. Then, $\Delta C_{\text{id}}(N)$ can be well-approximated in the low- and the high-SNR extremes as

$$|\Delta C_{\text{id}}(N)| \approx \begin{cases} \text{as } \rho \rightarrow 0 \text{ by } \begin{cases} \frac{5}{2} \rho(1 + \kappa_0), & \kappa_0 > 1 \\ 5\rho\kappa_0, & \kappa_0 \leq 1. \end{cases} \\ \text{as } \rho \rightarrow \infty \text{ by } \begin{cases} \frac{\log\left(\frac{\kappa_0-1}{\kappa_0}\right)}{\log(\rho)}, & \kappa_0 > 1 \\ \frac{1}{\log(\rho\kappa_0)}, & \kappa_0 \leq 1 \end{cases} \end{cases}$$

Proof: See Appendix III. ■

It is easy to see that in either SNR extreme, the correction terms vanish and $C_{\text{appx}}(N)$ is accurate. While characterizing the accuracy of $C_{\text{appx}}(N)$ in the medium-SNR regime seems harder, numerical results (see the ensuing discussion) show that $C_{\text{appx}}(N)$ provides a reasonably good fit to the exact capacity value at all SNRs.

The capacity of channels in the multiplexing regime is characterized next.

Theorem 4: Let \mathbf{H} be a channel in the multiplexing regime ($q = o(p)$). In the asymptotics of

N , the ergodic capacity $C_{\text{erg, mux}}(N)$ is given by

$$\frac{C_{\text{erg, mux}}(N)}{\log_2(e)} = 2p \log_e \left(\frac{1 + \sqrt{1 + 4\rho\kappa_0}}{2} \right) - p \frac{\sqrt{1 + 4\rho\kappa_0} - 1}{\sqrt{1 + 4\rho\kappa_0} + 1} \quad (29)$$

The correction term to the normalized capacity $\Delta C_{\text{mux}}(N)$ satisfies $|\Delta C_{\text{mux}}(N)| \approx \frac{\rho\kappa_0}{2}$ where $\kappa_0 = \frac{q}{p} \rightarrow 0$ as $N \rightarrow \infty$. ■

Among all the channels in the multiplexing regime, the smallest value for the correction term is achieved by the channel with the largest p ($\alpha = \alpha_{\text{max}}$). However, in the asymptotics of N , κ_0 (and hence the correction term) vanishes for every channel in this regime. Thus, $C_{\text{appx}}(N)$ is a very 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.

In Fig. 6, we consider a sparse channel with $D = N$ ($\gamma = 1$) and plot the capacities of the beamforming ($p = 1$), ideal ($p = \sqrt{N}$) and multiplexing ($p = N$) configurations as a function of N for three values of ρ : 0, 5 and 15 dB. Three curves are plotted in each figure: the Monte Carlo estimates of capacity, the random matrix theory (RMT)-based expressions proved in Theorems 2 - 4 and the approximate capacity expression $C_{\text{appx}}(N)$. Fig. 6 illustrates the closeness of the RMT-based expressions to those obtained from Monte Carlo averaging for all the three channels. On the other hand, the plot also shows that the approximate capacity expressions are reasonably accurate at all SNRs for all the configurations. However, the mismatch between $C_{\text{appx}}(N)$ and the exact expressions is largest at low-, medium- and high-SNR's for the beamforming, ideal and multiplexing configurations, respectively, consistent with our theoretical results above. Furthermore, consistent with our theoretical results, the mismatch between exact and approximate capacity expressions vanishes in the limit of large N for both the beamforming and multiplexing channels; the mismatch vanishes for the ideal channel only in the limit of low- or high-SNR. Finally, observe that while $C_{\text{appx}}(N)$ typically overestimates the Monte Carlo capacity for all channel configurations and SNRs, the asymptotic exact expressions underestimate the Monte Carlo capacity, converging to it as N becomes larger.

The above results demonstrate the accuracy and validity of the approximate capacity expression in (23) even for finite values of N and we will use this approximate expression in the analysis of capacity as a function of SNR in Section V.

We have already noted that for a sparse channel with $\rho_c = D = N^\gamma$, the ideal channel configuration with $p = \mathcal{O}(N^{\gamma/2})$ achieves the fastest capacity scaling. The next result gives a

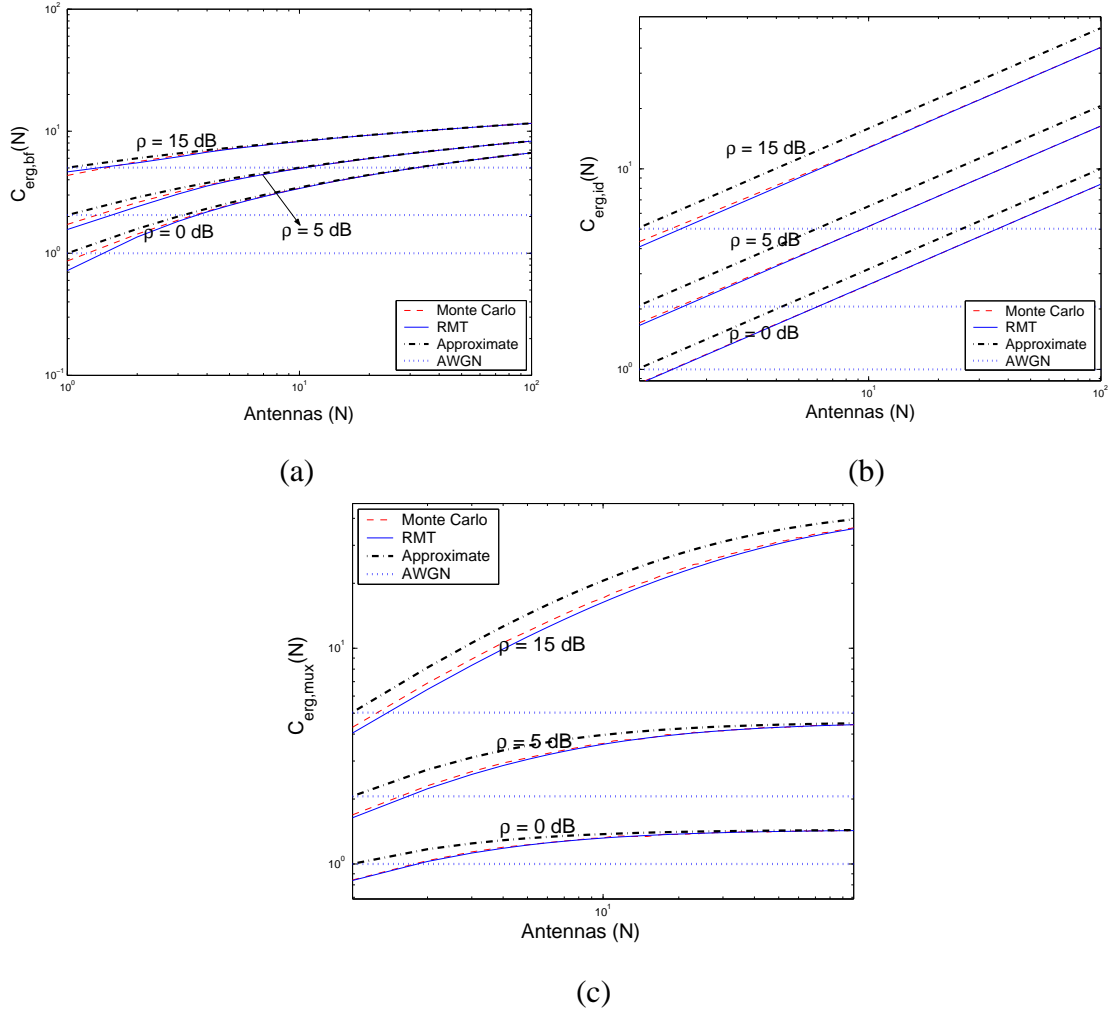


Fig. 6. Plots of the capacity of the beamforming, ideal and multiplexing channels using the exact capacity expressions, the approximate capacity expression in (23), and estimates based on a Monte Carlo simulation over many channel realizations.

precise characterization of the ideal channel configuration by characterizing the constant δ_p in $p = \delta_p N^{\frac{\gamma}{2}}$ as a function of the SNR ρ .

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

$$p_{\text{id}} \approx \sqrt{\frac{\rho D}{4}} = \frac{\sqrt{\rho}}{2} N^{\gamma/2}, \quad q_{\text{id}} \approx \sqrt{\frac{4D}{\rho}} = \frac{2}{\sqrt{\rho}} N^{\gamma/2} \quad (30)$$

maximizes the capacity among all possible channel configurations at the given ρ . Moreover, the ideal channel achieves the fundamental limit in Theorem 1 and is optimal from an ergodic

capacity scaling perspective

$$C_{\text{erg, id}}(N) = \mathcal{O}(\sqrt{\rho_c}) = \mathcal{O}(\sqrt{D}) = \mathcal{O}(N^{\gamma/2}). \quad (31)$$

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

$$\begin{aligned} \dot{C}_{\text{appx}}(N) &\triangleq \frac{dC_{\text{appx}}(N)}{dp} = \log_2(e) \cdot \left(\log_e(1+x) - \frac{2x}{1+x^2} \right) \\ \ddot{C}_{\text{appx}}(N) &\triangleq \frac{d\dot{C}_{\text{appx}}(N)}{dp} = \log_2(e) \cdot \frac{2x(1-x)}{(1+x)^2 p}. \end{aligned} \quad (32)$$

We note that $\dot{C}_{\text{appx}}(N) < 0$ for $x < x_0$ and $\dot{C}_{\text{appx}}(N) > 0$ for $x > x_0$ where $x_0 \approx 4$ while $\ddot{C}_{\text{appx}}(N) > 0$ for $x < 1$ and $\ddot{C}_{\text{appx}}(N) < 0$ for $x > 1$. Thus, the capacity is maximized at $x_0 = \frac{\rho N^\gamma}{p^2} = \frac{\rho \rho_c}{p^2} \approx 4$, from which the expressions for the pair $(p_{\text{id}}, q_{\text{id}})$ in (30) follow. ■

We note that in Fig. 5, the curves for $\rho = 5$ dB represent the optimal choice of $p_{\text{id}} = \sqrt{N}$ according to (30): the ideal configuration dominates the beamforming and multiplexing configurations for all values of N . The above result will also be useful in our analysis of capacity as a function of SNR. In particular, the proof of the above result also reveals an inflection point in the approximate capacity expression as a function of p (change in the sign of the second derivative).

A.3 Remarks

- Information (either direct or indirect) on the convergence and the limit of the empirical eigenvalue distribution of an appropriately normalized version of $\mathbf{H}^H \mathbf{H}$ played a very crucial role in evaluating the capacities of the different channel configurations in closed-form. A striking commonality of the three regimes is that when normalized by $C_{\text{appx}}(N)$, the empirical eigenvalue distributions of the different channel configurations converge. The difference, however, is that the limit of convergence critically depends on the ratio $\kappa_0 = \frac{q}{p}$, which is different in the three cases. The limiting distributions could be very different as elucidated by Lemmas 2 - 4 in Appendix III. The critical difference is that when $\frac{q}{p} \rightarrow \infty$ with N in the beamforming regime, the matrix $\mathbf{H}^H \mathbf{H}$ to its ensemble average $\mathbf{E}[\mathbf{H}^H \mathbf{H}]$. Thus its eigenvalues converge to those of $\mathbf{E}[\mathbf{H}^H \mathbf{H}]$. On the other hand, while the eigenvalues of channel configurations from the multiplexing and ideal regimes converge in the limit, the matrices themselves do not converge (under

any distance metric). This is a very important distinction between the beamforming regime and the other channel configurations.

- The optimality of the ideal channel can also be illustrated by the following heuristic, albeit non-rigorous, approach, which is very similar to that in [17]. Consider a channel with a channel power ρ_c and assume that signaling is done by exciting k of the N possible dimensions equally¹⁰.

The average mutual information $I(N)$ achieved by this signaling is

$$\begin{aligned}
I(N) &= \mathbf{E} \left[\sum_{i=1}^k \log_2 (1 + \lambda_i (\mathbf{H}\mathbf{Q}\mathbf{H}^H)) \right] \\
&\stackrel{(a)}{\approx} k \log_2 (1 + \lambda_{\text{avg}} (\mathbf{H}\mathbf{Q}\mathbf{H}^H)) \\
&= k \log_2 \left(1 + \frac{\mathbf{E} [\text{Tr} (\mathbf{H}\mathbf{Q}\mathbf{H}^H)]}{k} \right) \\
&= k \log_2 \left(1 + \frac{\rho \mathbf{E} [\text{Tr} (\mathbf{H}_k \mathbf{H}_k^H)]}{k^2} \right) \\
&\stackrel{(b)}{=} k \log_2 \left(1 + \frac{\rho \rho_c}{k^2} \right) \tag{33}
\end{aligned}$$

where $\lambda_{\text{avg}} (\mathbf{H}\mathbf{Q}\mathbf{H}^H)$ is the average eigenvalue of $\mathbf{H}\mathbf{Q}\mathbf{H}^H$, (a) follows from the fact that even though the largest and smallest eigenvalues of $\mathbf{H}\mathbf{Q}\mathbf{H}^H$ could exhibit significantly different statistical behavior, a bulk of the eigenvalues are statistically similar to the average eigenvalue, \mathbf{H}_k is the matrix formed from \mathbf{H} by retaining the k excited columns alone and (b) follows if $\mathbf{E} [\text{Tr} (\mathbf{H}_k \mathbf{H}_k^H)] = \rho_c$. It is not difficult to see that if \mathbf{H} has only k non-vanishing columns, then $k = \mathcal{O} (\sqrt{\rho_c})$, which corresponds to the ideal channel, maximizes (33).

B. Outage Capacity

We now study the asymptotic behavior of outage capacity of the family of channels introduced in Section III-C. The following proposition characterizes the variances of capacity of channels in the beamforming, ideal and multiplexing regimes in the limit of antenna dimensions.

Proposition 3: Let $\sigma_{\text{bf}}^2(N)$, $\sigma_{\text{id}}^2(N)$ and $\sigma_{\text{mux}}^2(N)$ be the variances of capacity of the channel configurations in the beamforming, ideal and multiplexing regimes in the limit of antenna

¹⁰For simplicity, we assume that \mathbf{H} is permuted so that the columns of \mathbf{H} are arranged in order of decreasing column power. Then, the input covariance matrix that we use is given by $\mathbf{Q} = \frac{\rho}{k} \mathbf{D}$ where \mathbf{D} is a diagonal matrix with ones in the first k diagonal entries.

dimensions. We then have

$$\begin{aligned}
\sigma_{\text{bf}}^2(N) &= \begin{cases} \frac{(\log_2(e))^2}{\kappa_0} \rightarrow 0 & \text{if } p \text{ is finite,} \\ \frac{(\log_2(e))^{4\kappa_0}}{(1+\kappa_0+\frac{1}{p})^2} \rightarrow 0 & \text{if } p \rightarrow \infty \end{cases} \\
\sigma_{\text{id}}^2(N) &= \begin{cases} (\log_2(e))^2 \cdot \log_e \left(\frac{1}{1-\frac{h^2}{\kappa_0}} \right) & \text{if } \kappa_0 \geq 1, \\ (\log_2(e))^2 \cdot \log_e \left(\frac{1}{1-h_1^2} \right) & \text{if } \kappa_0 < 1 \end{cases} \\
\sigma_{\text{mux}}^2(N) &= (\log_2(e))^2 \cdot \rho^2 \kappa_0^2 \rightarrow 0
\end{aligned} \tag{34}$$

where h is given in (27), $h_1 = \frac{1+2\rho\kappa_0-\sqrt{1+4\rho\kappa_0}}{2\rho\kappa_0}$ and $\kappa_0 = \frac{q}{p}$.

Proof: See Appendix IV. ■

Given that the means and variances of capacity are computed in closed form in Theorems 2–4 and Proposition 3, the formulas for outage capacity $C_{\text{out},q,\bullet}(N)$ follow from (16):

$$C_{\text{out},q,\bullet}(N) = C_{\text{erg},\bullet}(N) - x_q \sigma_{\bullet}(N) + o(1) \tag{35}$$

where x_q satisfies $\text{erfc}(x_q/\sqrt{2}) = 2q$. Note that the variance of each channel configuration satisfies $\frac{\sigma_{\bullet}(N)}{C_{\text{erg},\bullet}(N)} \rightarrow 0$ in the asymptotics of N , apart from the fact that $\sigma_{\text{bf}}(N)$ and $\sigma_{\text{mux}}(N)$ converge to 0. The trends of outage capacities in the antenna asymptotics are therefore very similar to those of the ergodic capacities and hence we do not plot them separately.

V. CAPACITY AS A FUNCTION OF SNR

In this section, we consider a fixed number of antennas and study the impact of the different sparse channel configurations defined in Section III-C on capacity as a function of SNR. Our starting point is the asymptotic capacity analysis in Section IV based on random matrix theory which demonstrated the accuracy of the approximate closed-form capacity expression (23) in the limit of large antenna dimension N . It has been shown by several recent results that the convergence of finite antenna capacity to asymptotic estimates is usually fast. In particular, numerical as well as theoretical results show that this convergence is on the order of the inverse of antenna dimension, see e.g., [36] and Fig. 6. Thus, we assume that N is sufficiently large and base our analysis on the approximate capacity expression (23). We first address the behavior of ergodic capacity and then discuss outage capacity as a function of SNR.

A. Ergodic Capacity

We first reproduce the approximate capacity expression (23) here to emphasize the dependence on parameters relevant to this section

$$C_{\text{erg}}(N, D, p, \rho) \approx p \log(1 + \rho_{rx}) = p \log\left(1 + \rho \frac{q}{p}\right) \triangleq C_{\text{appx}}(N, D, p, \rho). \quad (36)$$

The above formula reveals a fundamental new **multiplexing gain–received SNR trade-off** in sparse channels with $\rho_c = D < N^2$. There exists a tension between the multiplexing gain p and the received SNR, $\rho_{rx} = \rho \frac{q}{p} = \rho \frac{\rho_c}{p^2}$. Increasing p comes at the cost of ρ_{rx} and *vice versa*. The impact of this tradeoff on ergodic capacity is vividly illustrated in Fig. 3 in which $C_{\text{appx}}(N, D, p, \rho)$ is plotted as a function of ρ for $N = 25$ and $D = N$ ($\gamma = 1$). The plot shows the curves for the three canonical configurations – beamforming ($p = p_{\min} = 1$), ideal ($p = p_{\text{id}} = \sqrt{N}$), and multiplexing ($p = p_{\max} = N$) – as well as the capacities of 10 channel configurations corresponding to equally spaced values of $\alpha \in [0, 1]$ which span the whole range of p between p_{\min} and p_{\max} via $p = N^\alpha$. Fig. 3 shows that there exist SNR's ρ_{low} and ρ_{high} so that the beamforming configuration yields highest capacity for $\rho \leq \rho_{\text{low}}$ and the multiplexing configuration yields highest capacity for $\rho \geq \rho_{\text{high}}$. For the intermediate SNR range, $\rho_{\text{low}} < \rho < \rho_{\text{high}}$, the optimal (capacity-maximizing) channel configuration continuously transitions from beamforming to multiplexing channel through the ideal configuration. Furthermore, while the beamforming and multiplexing configurations exchange roles in the low- and high-SNR regimes, the ideal configuration is a robust choice whose capacity lies between the two extremes. The optimal channel configuration at any SNR optimizes the multiplexing gain– ρ_{rx} trade-off for maximizing capacity. At low-SNR's, the beamforming configuration is optimal and maximizes ρ_{rx} at the cost of p . At high-SNR's, the multiplexing configuration is optimal and maximizes p at the cost of ρ_{rx} . The ideal configuration is a robust intermediate choice that yields $\rho_{rx} = \rho$.

The next result precisely characterizes the optimal capacity-maximizing channel configuration as a function of SNR for a given N and $\rho_c = D < N^2$ using the approximate expression (36). It builds on Proposition 2, that characterizes the optimal (ideal) channel configuration from a capacity scaling viewpoint: $p_{\text{id}} = \sqrt{\frac{\rho D}{4}}$, $q_{\text{id}} = \sqrt{\frac{4D}{\rho}}$. Reinterpreting this result for a fixed N and D yields the capacity maximizing value of p as a function of SNR.

Proposition 4: Capacity-maximizing configuration as a function of SNR. For a given fixed

N , let $\rho_c = D = N^\gamma$, $\gamma \in (0, 2]$ and define $\alpha^* \triangleq \min(\gamma, 2 - \gamma)$, $\alpha_{\min} = \max(\gamma - 1, 0)$, $\alpha_{\max} = \min(\gamma, 1)$. The optimal capacity-maximizing channel configuration at any given operating transmit SNR ρ is characterized by p_{opt} given by

$$p_{\text{opt}}(\rho) \approx \begin{cases} N^{\alpha_{\min}}, & \rho < \rho_{\text{low}} \\ \frac{\sqrt{\rho D}}{2} = \frac{\sqrt{\rho}}{2} N^{\gamma/2}, & \rho \in [\rho_{\text{low}}, \rho_{\text{high}}] \\ N^{\alpha_{\max}}, & \rho > \rho_{\text{high}} \end{cases} \quad (37)$$

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

Proof: Recall the two different feasible ranges for p depending on γ . 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 $C_{\text{appx}}(N, \rho)$ vanishes when

$$\frac{\rho \rho_c}{p^2} = \frac{\rho N^\gamma}{p^2} \approx 4 \quad (38)$$

which characterizes the capacity-maximizing point. Solving $\rho_{\text{low}} = 4p_{\min}^2/N^\gamma$, it is easy to check that a common representation for ρ_{low} in both cases above is $\rho_{\text{low}} = \frac{4}{N^{\alpha^*}}$. Similarly, solving $\rho_{\text{high}} = 4p_{\max}^2/N^\gamma$ yields the value of ρ_{high} stated in the proposition in both cases. In the intermediate regime, p_{opt} directly follows from (38). \blacksquare

The capacity-maximizing $p_{\text{opt}}(\rho)$ characterized in the above result optimizes the multiplexing gain–received SNR trade-off at ρ .

Using Proposition 4, we get the theoretical estimates $\rho_{\text{low}} \approx -8$ dB and $\rho_{\text{high}}(N) \approx 20$ dB for the channel configurations ($N = 25$, $\gamma = 1$) illustrated in Fig. 3. In the figure, we observe that the beamforming channel is indeed optimal for $\rho < -8$ dB and the multiplexing channel is optimal for $\rho > 19$ dB. The value of p_{opt} characterized in Prop. 4 for the intermediate SNR range, $\rho \in [\rho_{\text{low}}, \rho_{\text{high}}]$, is illustrated by the dotted curves in the figure which plot

$$C_{\text{appx}, \alpha}(N, \rho) \triangleq N^\alpha \log(1 + \rho N^{\gamma-2\alpha}) \quad (39)$$

corresponding to $p = N^\alpha$ and $q = N^{\gamma-\alpha}$ for 10 equally spaced values of $\alpha \in [0, 1]$. For each intermediate-SNR, there is a $C_{\text{appx}, \alpha}(N, \rho)$ curve that yields the maximum capacity.

For the same case considered above, Fig. 7 shows capacity as a function of SNR for the beamforming, ideal and multiplexing channels. Capacity estimates obtained via Monte Carlo methods

are seen to be close to the RMT-based results proved earlier and the approximate capacity formula $C_{\text{appx}}(N, \rho)$ is close to these values for all channel configurations and SNRs.

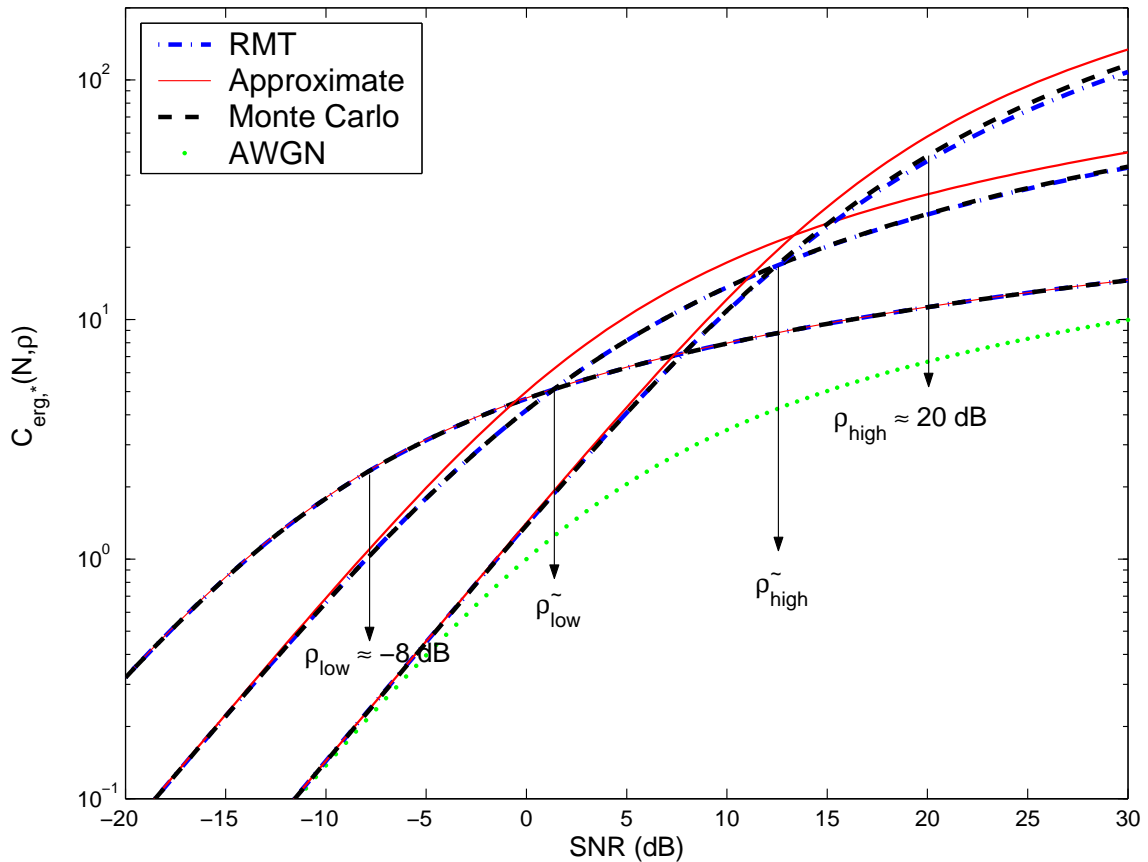


Fig. 7. Estimates for capacity of beamforming, ideal and multiplexing channels for $N = 25$ and $\gamma = 1$ based on Monte Carlo simulations, Theorems 2-4 and approximate capacity formula. Also shown are ρ_{low} , ρ_{high} , $\tilde{\rho}_{\text{low}}$ and $\tilde{\rho}_{\text{high}}$.

A.1 Remarks

- While Proposition 4 states the precise value of p_{opt} for any $\rho \in [\rho_{\text{low}}, \rho_{\text{high}}]$, from Fig. 3 we note that the ideal channel configuration with $p_{\text{id}} = N^{\gamma/2}$ serves as a robust *fixed* configuration in this intermediate range. Thus, for all practical purposes, the three canonical configurations – beamforming ($p = p_{\text{min}}$), ideal ($p = p_{\text{id}}$) and multiplexing ($p = p_{\text{max}}$) – accurately approximate

the capacity-maximizing configuration over the entire range. That is,

$$p_{\text{opt}}(\rho) = \begin{cases} p_{\text{min}} = N^{\alpha_{\text{min}}} & \rho < \tilde{\rho}_{\text{low}} \\ p_{\text{id}} = N^{\gamma/2} & \tilde{\rho}_{\text{low}} \leq \rho \leq \tilde{\rho}_{\text{high}} \\ p_{\text{max}} = N^{\alpha_{\text{max}}} & \rho > \tilde{\rho}_{\text{high}} \end{cases} \quad (40)$$

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

$$N^{\alpha_{\text{min}}} \log(1 + \tilde{\rho}_{\text{low}} N^{\gamma-2\alpha_{\text{min}}}) = N^{\gamma/2} \log(1 + \tilde{\rho}_{\text{low}}) \quad (41)$$

$$N^{\alpha_{\text{max}}} \log(1 + \tilde{\rho}_{\text{high}} N^{\gamma-2\alpha_{\text{max}}}) = N^{\gamma/2} \log(1 + \tilde{\rho}_{\text{high}}). \quad (42)$$

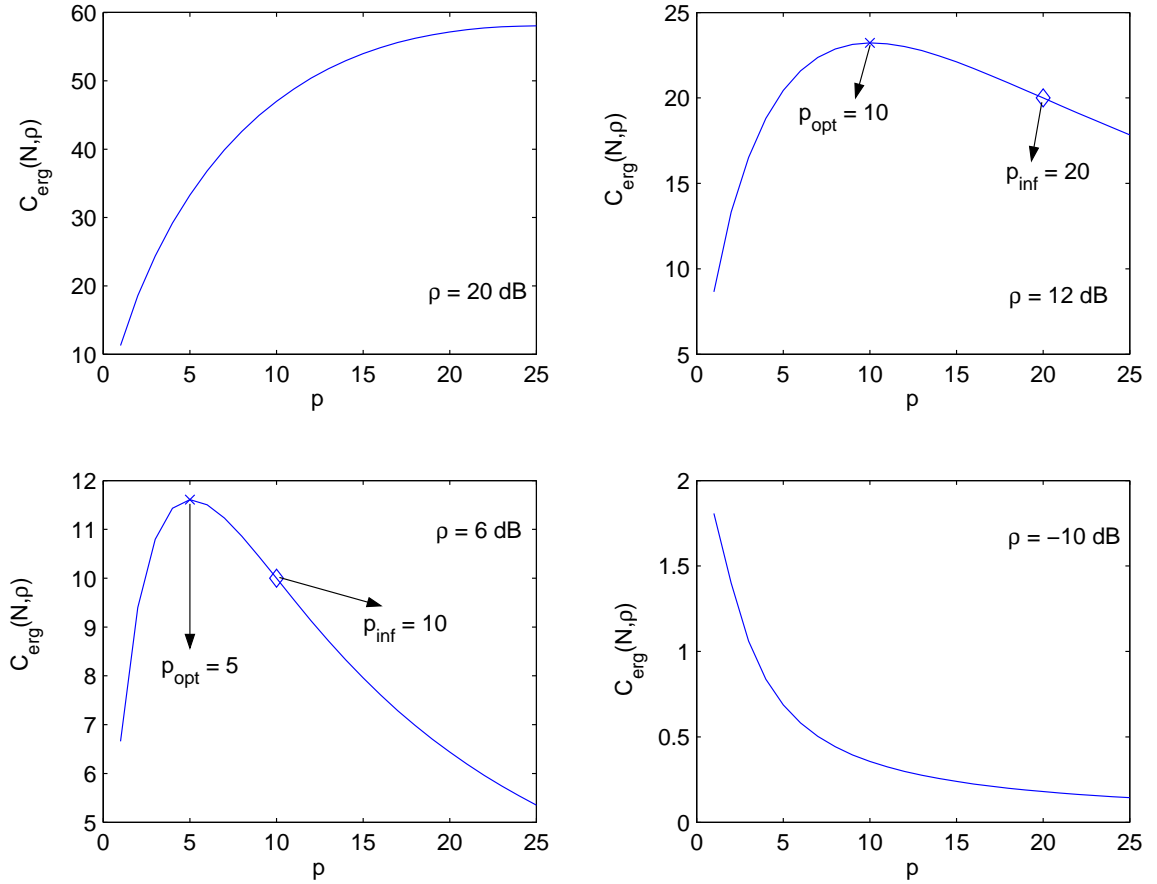


Fig. 8. Illustration of concave-to-convex transition of $C_{\text{appx}}(N, D, p, \rho)$ as a function of p for $N = 25$ and $\gamma = 1$ and four different SNR's.

- We showed in the proof of Proposition 2 that the second derivative $\ddot{C}_{\text{appx}}(N, D, p, \rho)$ (w.r.t. p) vanishes at $x = \frac{\rho \rho_c}{p^2} = 1$ and is positive for $x < 1$ and negative for $x > 1$. That is, $C_{\text{appx}}(N, \rho)$

is convex for $x < 1$ (or $p > p_{\text{inf}}$), while it is concave for $x > 4$ (or $p < p_{\text{opt}}$) with a positive first derivative. In the intermediate regime of $x \in [1, 4]$ ($p \in [p_{\text{opt}}, p_{\text{inf}}]$), $C_{\text{appx}}(N, \rho)$ is concave, but with a negative first derivative. Thus the inflection point $x = 1$ ($p = p_{\text{inf}}$) signals the transition from convex-to-concave behavior. This transition is illustrated in Fig. 8 for $N = 25$ and $\gamma = 1$. Note that p_{inf} can be written in terms of p_{opt} as $2p_{\text{opt}}$.

- The ratio $\frac{\rho_{\text{high}}}{\rho_{\text{low}}}$ attains its largest value, N^2 , for $\gamma = 1$ ($\rho_c = N$), whereas it achieves its minimum value of unity for $\gamma = 0$ ($\rho_c = D = 1$) or $\gamma = 2$ ($\rho_c = N^2$). Thus, the multiplexing gain- ρ_{rx} tradeoff that determines p_{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 spatial power distributions lead to identical \mathbf{H} . On the other hand, the impact of this tradeoff on capacity is maximum for $\gamma = 1$ corresponding to $\rho_c = D = N$.

B. Outage Capacity

We now focus on the outage capacity as a function of SNR of the channel configurations in the three different regimes configurations. The computation of the variance of capacity in closed-form for all SNR values seems difficult. However, in the low- and the high-SNR extremes, we can compute the variances. This is the content of the following proposition.

Proposition 5: The variances $\sigma_{\bullet}^2(N, \rho)$ as a function of SNR are given in the following table:

Channel	Low-SNR	High-SNR
Beamforming, p finite	$\frac{(\log_2(e))^2}{\kappa_0}$	$\frac{(\log_2(e))^2}{\kappa_0}$
Beamforming, $p \rightarrow \infty$	$4\kappa_0\rho^2(\log_2(e))^2$	$(\log_2(e))^2 \frac{4\kappa_0}{(1+\kappa_0)^2}$
Ideal, $\kappa_0 > 1$	$(\log_2(e))^2 \log_e(1 + \rho^2\kappa_0)$	$(\log_2(e))^2 \log_e\left(\frac{\kappa_0}{\kappa_0-1}\right)$
Ideal, $\kappa_0 \leq 1$	$(\log_2(e))^2 \log_e(1 + \rho^2\kappa_0^2)$	$(\log_2(e))^2 \log_e\left(\frac{\sqrt{\rho\kappa_0}}{2}\right)$
Multiplexing	$(\log_2(e))^2 \rho^2\kappa_0^2$	$(\log_2(e))^2 \rho^2\kappa_0^2$

Proof: The proof follows by performing the limiting process in the corresponding variances as given in Proposition 3, see [35], [34] for details. ■

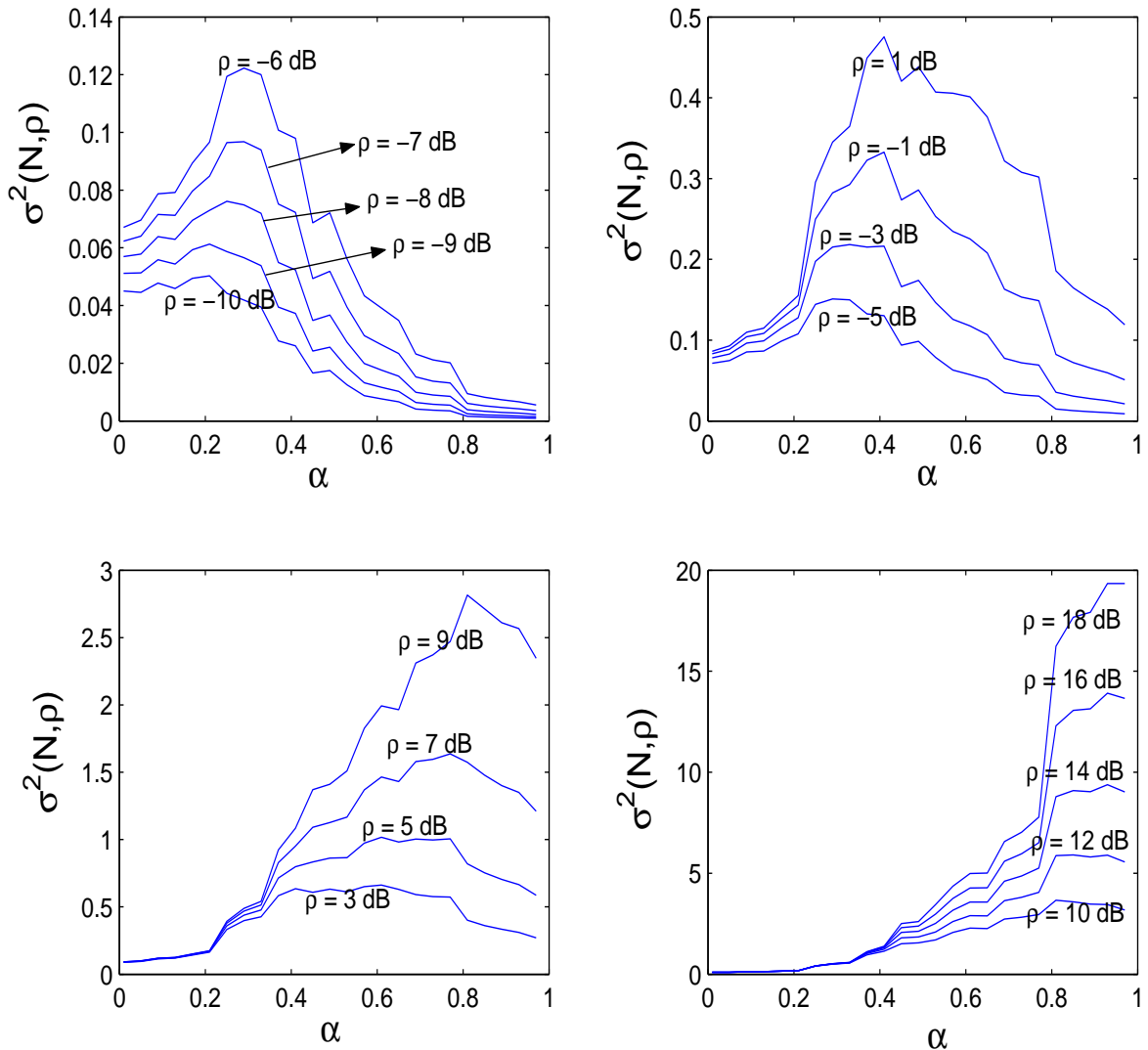


Fig. 9. Trends of the variance of capacity, $\sigma^2(N, \rho)$, as a function of α for different SNR with $N = 25$ and $\gamma = 1$.

In the intermediate-SNR regime, we have to resort to numerical studies to compute the variances $\sigma^2(N, \rho)$. Fig. 9 plots this quantity as a function of α for different SNR's assuming $N = 25$ and $\gamma = 1$. The non-smoothness of the variances in the plot is due to the flooring operations that generate $p = \lceil N^\alpha \rceil$. Non-integer antenna numbers corresponding to $p = N^\alpha$ can in fact be obtained via a time-sharing of channels as detailed in Appendix I. Fig. 9 however does not illustrate this smoothing effect.

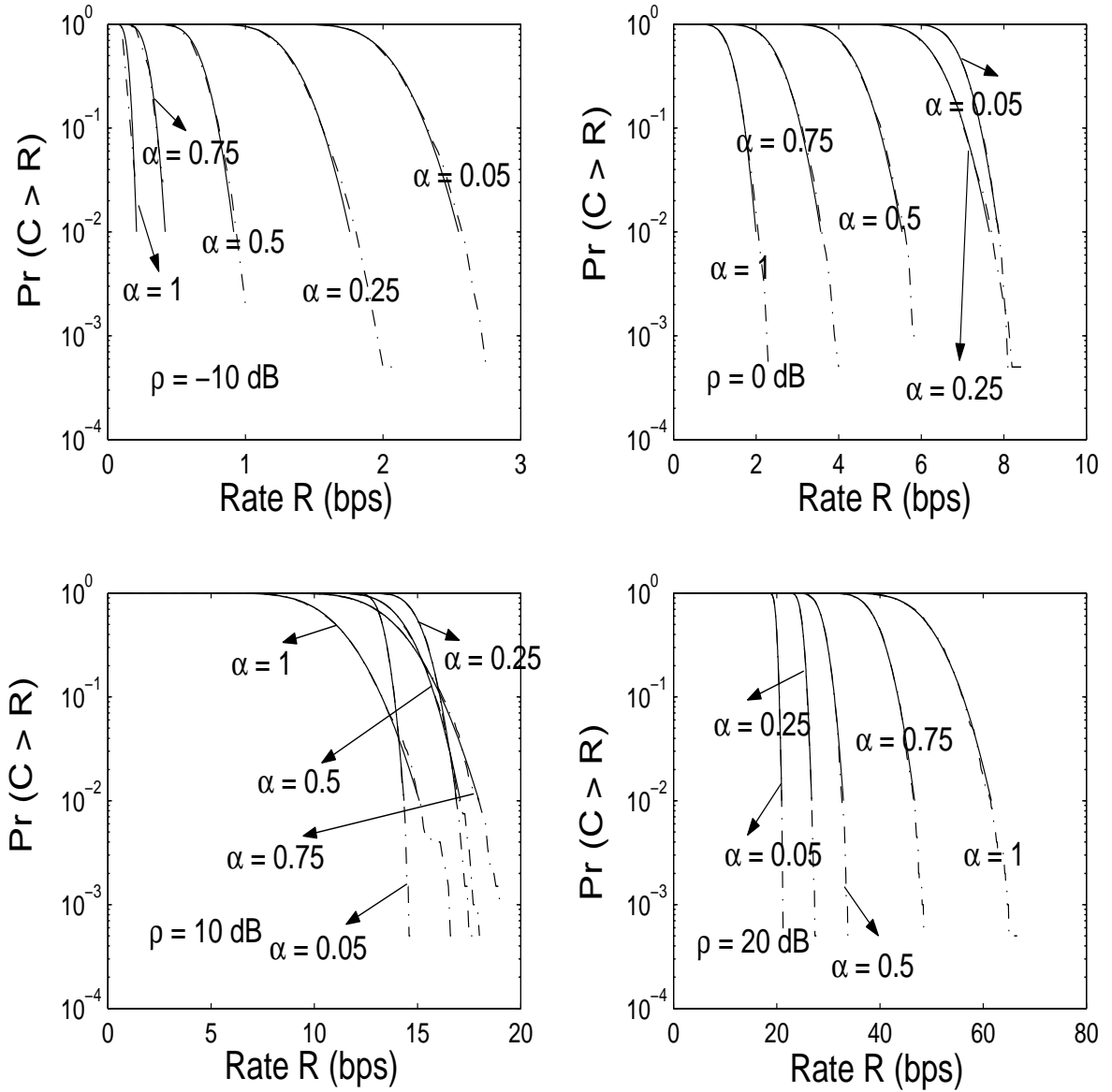


Fig. 10. Outage capacity of the family of channels for $N = 25$ and $\gamma = 1$. The capacities are plotted for $p = N^\alpha$ and for four different values of ρ : $-10, 0, 10, 20$ dB. The dotted lines correspond to the exact outage capacity while the solid lines correspond to the Gaussian approximation.

From Prop. 5, we see that at low-SNR, channels in the beamforming regime (with small values of α) tend to have the largest variance among all configurations. Similarly, at high-SNR, channels in the multiplexing regime tend to have the largest variance. Fig. 9 captures these trends apart from the dominance of the ideal channel in the medium-SNR range. Thus for any ρ , the largest variation in the capacity random variable is seen for the channel configuration that

is optimal for that SNR from an ergodic capacity perspective.

In Fig. 10, a Gaussian approximation to the capacity random variable from the mean and variance values is used to reproduce the outage capacities. The figure shows the closeness of the Gaussian approximation to the actual outage capacity for the three canonical channel configurations at different SNR's, thereby illustrating the accuracy of Theorems 2 - 4 and Propositions 3 - 5.

Figs. 9 and 10 also reveal a **rate–reliability tradeoff** associated with the family of channels. Since rate is associated with multiplexing gain and reliability with received SNR, the rate-reliability tradeoff can also be viewed as a manifestation of the multiplexing– ρ_{rx} tradeoff identified in the context of ergodic capacity. At low-SNR, communication is inherently constrained by the reliability of signal reception. Thus maximizing reliability by providing a higher level of diversity per parallel channel enhances performance. Note that q reflects the level of diversity (DoF) associated with each parallel channel. Thus, the beamforming channel that maximizes q yields the highest outage capacity at low-SNR, as seen from Fig. 10. When the SNR is high, the system is operating at a point where enhancing the reliability of signal reception by increasing q can only marginally improve performance and hence maximizing the rate of communication by increasing the multiplexing gain (p) is better. Thus a multiplexing channel is outage optimal at high-SNR as seen from Fig. 10. In the intermediate-SNR regime, the optimal configuration is governed by the desired level of reliability. When only a low level of channel outage is acceptable, a channel configuration with small α corresponding to higher diversity is better. When more reasonable outage levels can be tolerated, a channel configuration with a higher multiplexing (larger α) gain yields higher outage capacity. However, a robust solution in the medium-SNR range is the ideal channel configuration with $\alpha = \frac{1}{2}$.

VI. CONNECTIONS TO OTHER WORKS

A. Realizing the Channel Configurations in Practice

In this paper we have analyzed the impact of the distribution of the $D < N^2$ channel DoF in sparse channels on ergodic and outage capacity. Our results provide information theoretic benchmarks on optimal channel configurations from the viewpoint of capacity scaling as well as capacity as a function of SNR for finite N . In practice, however, the configuration of the channel

DoF is determined by the AoAs and the AoDs ($\theta_{r,\ell}$ and $\theta_{t,\ell}$ in (4)) of the 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 [12] where it is shown that in a sparse scattering environment with randomly distributed (within the angular spreads) paths, the antenna spacings at the transmitter and receiver can be systematically adapted to the level of sparsity to realize the different channel configurations discussed in the paper. In particular, the multiplexing configuration corresponds to maximal antenna spacings at both the transmitter and receiver, the ideal channel is realized via medium antenna spacings at both the transmitter and receiver, and the beamforming channel is realized via closely spaced antennas at the transmitter and large spacings at the receiver. The notion of small, medium and large spacings is quantified and the readers are referred to [12] for more details.

B. Recent Works on Closely Spaced Antenna Arrays

A recent work [17] provides insights into the problem of optimizing antenna spacing at the transmitter for maximizing the capacity in correlated MIMO channels. These insights are based on studying the variations in the channel's singular values as a function of the antenna spacings. It is argued in [17] 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 intermediate-SNR range 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 multiplexing gain p in our framework precisely reflects the number of channel eigenmodes excited at the transmitter and the beamforming and multiplexing channel configurations are characterizations of the low- and high-SNR analogues in [17]. Furthermore, our results characterize the optimal channel configuration at any operating SNR, thereby providing a systematic resolution of the conjecture in [17]. Indeed, the investigation in this paper was inspired by the need to reduce the loss in channel power inherent in the rank-1 capacity-maximizing input in fixed-spacing correlated MIMO channels.

Another recent work [37] discusses techniques for optimizing antenna locations in a fixed volume to maximize low-SNR performance metrics like spectral efficiency. Design guidelines offered in [37] provide indirect evidence to the optimality of closely spaced antennas in the low-SNR regime. Indirect evidence to the optimality of closely spaced transmit antennas in the low-SNR regime also comes from a recent study [18] 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. The results that we have presented in this paper could be useful in refining and extending the above works.

C. MMSE Estimation and Mutual Information

Random matrix theory tools that are used for capacity analysis of MIMO systems are of two types: 1) those that provide explicit information about the converging empirical spectral distributions like Lemma 3, and thus permit a closed-form expression for capacity, and 2) implicit characterizations of the empirical spectral distributions which include the Stieltjes and other transformation techniques. Both these techniques have their own advantages and insights into multi-antenna systems design. An underlying connection between MMSE estimation and mutual information, akin to [14], [15], is discernible by studying the implicit characterization of the converging empirical spectral distribution.

Proceeding via this route, it is shown in [38] that, given an $N_r \times N_t$ channel \mathbf{H} with independent entries, the ergodic capacity at an SNR of ρ under the assumption of no channel state information at the transmitter satisfies

$$C_{\text{erg}}(N_r, N_t, \rho) = N_t \cdot \mathbf{E}_T \left[\log_2 \left(1 + \rho \kappa \mathbf{E}_R [\mathcal{G}(R, T) \mathcal{D}(R) | T] \right) \right] + N_r \cdot \left(\log_2(e) \left(\mathbf{E}_R [\mathcal{D}(R)] - 1 \right) - \mathbf{E}_R [\log_2(\mathcal{D}(R))] \right) \quad (43)$$

in the limit of large antenna dimensions. In (43), κ denotes $\frac{N_r}{N_t}$, R and T are uniformly distributed random variables on $[0, 1]$ and the random variable over which averaging is performed is explicitly stated in the subscripts of the expectation operators $\mathbf{E}_R[\cdot]$ and $\mathbf{E}_T[\cdot]$. $\mathcal{G}(r, t)$ stands for the continuous representation of the channel power matrix supported on $[0, 1] \times [0, 1]$, that is,

$$\mathcal{G}(r, t) = \mathbf{E} [|\mathbf{H}_{ij}|^2], \quad \frac{i}{N_r} \leq r \leq \frac{i+1}{N_r}, \quad \frac{j}{N_t} \leq t \leq \frac{j+1}{N_t}. \quad (44)$$

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

$$\mathcal{D}(r) \left(1 + \rho \mathbf{E}_T \left[\frac{\mathcal{G}(r, T)}{1 + \rho \kappa \mathbf{E}_R [\mathcal{D}(R) \mathcal{G}(R, T) | T]} \right] \right) = 1. \quad (45)$$

Defining $\Gamma(t)$ and $\Upsilon(t)$ as

$$\Gamma(t) \triangleq \kappa \mathbf{E}_R [\mathcal{G}(R, t) \mathcal{D}(R)] \stackrel{(a)}{=} \kappa \mathbf{E}_R \left[\frac{\mathcal{G}(R, t)}{1 + \mathbf{E}_T [\mathcal{G}(R, T) \Upsilon(T) | R]} \right] \quad (46)$$

$$\Upsilon(t) \triangleq \frac{\rho}{1 + \rho \Gamma(t)} \quad (47)$$

the quantity $\rho \Gamma(t)$ can be seen to be the SNR at the output of a linear MMSE receiver for the signal transmitted from the corresponding transmit antenna. The corresponding MSE is seen to be $\frac{\Upsilon(t)}{\rho}$. The equality in (a) follows from the analysis in [38].

In our setting, the regularity of the family of channels under consideration can be exploited to obtain more insights on the trade-off between the number of data-streams (multiplexing gain) and the MSE of the individual data-streams.

Proposition 6: Consider the family of channels introduced in Section III-C with $D = qp$ and $\kappa_0 = \frac{q}{p}$. In the beamforming regime and the ideal regime under the assumption that $\kappa_0 > 1$, $\Gamma(t)$, defined as in (46), satisfies

$$2\rho \Gamma_{\text{bf, id}_1}(t) = \rho \kappa_0 - \rho - 1 + \sqrt{(\rho \kappa_0 - \rho - 1)^2 + 4\rho \kappa_0}. \quad (48)$$

On the other hand, in the multiplexing regime and the ideal regime under the assumption that $\kappa_0 \leq 1$, $\Gamma(t)$ is given by

$$2\rho \Gamma_{\text{mux, id}_2}(t) = \sqrt{1 + 4\rho \kappa_0} - 1. \quad (49)$$

Proof: See Appendix V. ■

Since $\kappa_0 \rightarrow 0$ in the multiplexing regime, $\rho \Gamma_{\text{mux}}(t) \rightarrow \rho \kappa_0 = \rho \frac{q}{p}$. That is, the SNR at the output of a linear MMSE receiver is the received SNR per parallel channel ρ_{rx} and the mean-squared error at the receiver is $\frac{1}{1 + \rho_{rx}}$. Similarly, by letting $\kappa_0 \rightarrow \infty$ in the beamforming regime, it is not difficult to see that $\rho \Gamma_{\text{bf}}(t) \rightarrow \rho \frac{q}{p}$ for all transmit SNR.

As $\rho \rightarrow 0$, a Taylor's series analysis of $\Gamma(t)$ in both cases of the ideal regime shows that $\rho \Gamma_{\text{id}}(t) \rightarrow \rho \kappa_0$. The corresponding limits for id_1 and id_2 in the high SNR regime are $\rho(\kappa_0 - 1)$ and $2\sqrt{\rho \kappa_0}$, respectively. The convergence of $\rho \Gamma(t)$ to ρ_{rx} for all ρ in the beamforming and

multiplexing regimes is related to the tightness of the capacity expressions obtained in Section IV in these regimes.

Thus the beamforming configuration (which maximizes ρ_{rx}) trades off the number of data-streams for the MSE of the individual data streams. The multiplexing configuration corresponds to the other extreme in this trade-off while the ideal configuration leads to a robust choice in this *rate-distortion* trade-off between the number of data-streams and the MSE of the individual data-streams.

D. Spectral Efficiency in the Low Power Regime

We now provide an alternate interpretation of the low-SNR (or equivalently, the wideband) results on ergodic capacity in terms of spectral efficiency. The minimum energy per bit necessary for reliable communication $\frac{E_b}{N_o \min}$ and the wideband slope S_0 have been shown to be the key figures of merit in the low-power/wideband regime [39]. We now compute $\frac{E_b}{N_o \min}$ and S_0 for the three channel configurations: beamforming, ideal and multiplexing. The regularity of the three channels and the resultant optimality of the uniform-power input ease this computation.

Let $\dot{C}_{\text{erg, bf}}(N, 0)$, $\ddot{C}_{\text{erg, bf}}(N, 0)$, $\dot{C}_{\text{erg, id}}(N, 0)$, $\ddot{C}_{\text{erg, id}}(N, 0)$, $\dot{C}_{\text{erg, mux}}(N, 0)$, and $\ddot{C}_{\text{erg, mux}}(N, 0)$ denote the first and second derivatives (w.r.t. ρ) of the ergodic capacity (in nats/dimension) of the beamforming, ideal and multiplexing channels respectively in the limit $\rho \rightarrow 0$. The minimum energy per bit for reliable communication is given by [39]

$$\begin{aligned} \frac{E_b}{N_o \min, \star} &= \frac{\log_e(2)}{\dot{C}_{\text{erg, } \star}(N, 0)} \\ &= \frac{\log_e(2) p}{\mathbf{E}[\text{Tr}(\mathbf{H}\mathbf{H}^H)]} \\ &= \frac{\log_e(2)}{q} \end{aligned} \tag{50}$$

where we have used \star in the subscript in (50) to denote that the channel could be either of the three possible types. The approximate capacity expression $p \log_2 \left(1 + \rho \frac{q}{p} \right)$ provides an alternate route to computing $\frac{E_b}{N_o \min}$ and a Taylor's series expansion of this approximation shows that $\frac{E_b}{N_o \min}$ is the same for the three channels and is equal to $\frac{\log_e(2)}{q}$. This also corroborates the fact that the approximation due to $C_{\text{appx}}(N, \rho)$ is close to the actual capacity in the low-SNR regime for all the channel configurations. Since the beamforming channel maximizes q , $\frac{E_b}{N_o \min}$ is the smallest for this configuration.

The wideband slope S_0 is given by [39]

$$\begin{aligned} S_{0,\star} &= \frac{2 \left(\dot{C}_{\text{erg},\star}(N, 0) \right)^2}{-\ddot{C}_{\text{erg},\star}(N, 0)} \\ &= \frac{2q^2p^2}{\mathbf{E} [\text{Tr}((\mathbf{H}\mathbf{H}^H)^2)]}. \end{aligned} \quad (51)$$

The quantity $\mathbf{E}[\text{Tr}((\mathbf{H}\mathbf{H}^H)^2)]$ can be computed using the Gaussian moment-factoring theorem [40] and equals $2q^2p$ in the multiplexing case [4], and $qp(q+p)$ for the beamforming and ideal channels [39]. Thus the wideband slope is $S_{0,\text{bf}} = \frac{2qp}{q+p} \approx 2p$, $S_{0,\text{id}} = \frac{2qp}{q+p} = p$, and $S_{0,\text{mux}} = p$. The approximate capacity expression yields the same wideband slope $2p$ for all the three channels and is equal to the exact expressions computed above up to an $\mathcal{O}(1)$ factor. Since the multiplexing gain is maximized by the multiplexing channel, the wideband slope is maximized for that 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. To illustrate this, consider the $\gamma = 1$ case. Then, we have $\frac{\frac{E_b}{N_0 \min_{\text{bf}}}}{\frac{E_b}{N_0 \min_{\text{id}}}} = \frac{1}{\sqrt{N}}$ and $\frac{\frac{E_b}{N_0 \min_{\text{bf}}}}{\frac{E_b}{N_0 \min_{\text{mux}}}} = \frac{1}{N}$; that is, $\frac{E_b}{N_0 \min_{\text{bf}}}$ is smaller than that of the ideal channel by a factor of \sqrt{N} and smaller than that of the multiplexing channel by a factor of N . Since, $\frac{E_b}{N_0 \min}$ reflects the SNR-axis intercept of the capacity versus SNR relation in the wideband limit, the larger values of S_0 for the ideal and multiplexing channels become irrelevant. The low-SNR capacity gains of the beamforming channel relative to the ideal and multiplexing channels in Fig. 3 are precisely a manifestation of these gains in $\frac{E_b}{N_0 \min}$, as also evident from (50).

VII. SUMMARY AND CONCLUSIONS

Virtually all existing information theoretic studies of multi-antenna capacity implicitly assume a rich scattering environment. This assumption is manifested in the quadratic channel power scaling or normalization with respect to the number of antennas, and is also prevalent in the capacity analysis of correlated MIMO channels. In this paper we have argued that this widespread assumption is not justified in view of physical power conservation principles and results from experimental studies. Motivated by this observation, we have proposed a framework for modeling MIMO channels in a sparse multipath environment and studied the implications of

sparsity on channel capacity. Our approach is based on the virtual representation [6] of physical wireless channels that quantifies the statistically independent degrees of freedom (DoF) in the channel as a function of the level of sparsity (or richness) in the scattering environment. An $N \times N$ sparse MIMO channel possesses $D < N^2$ DoF – fewer than the maximum allowable N^2 . From a capacity scaling perspective, the channel power and the DoF scale at a sub-quadratic rate, $\mathcal{O}(N^\gamma)$, $\gamma \in (0, 2]$, with the number of antennas.

Sparse channels afford a fundamental new degree of freedom that is not available in rich channels: the spatial distribution or configuration of the $D < N^2$ DoF in the available N^2 transmit-receive dimensions. The focus of this paper is to study the impact of this new degree of freedom (channel configuration) on capacity, both from a scaling perspective as well as from the viewpoint of capacity as function of transmit SNR for a finite N . From a scaling perspective, we show that the ergodic capacity of any sparse channel configuration with a channel power scaling $\rho_c(N)$ cannot grow faster than $\mathcal{O}(\sqrt{\rho_c(N)})$, thereby opening a new paradigm of sub-linear capacity scaling that is consistent with experimental results and physical arguments. We further show that the above fundamental limit can be achieved asymptotically by an optimal channel configuration, which we call the *ideal MIMO channel* configuration; the pre-constant of scaling depends on the transmit SNR.

The development of the ideal channel is based on a structured family of sparse channel matrices. For a given D , all channels in the family are parameterized by a single parameter p representing the multiplexing gain (number of parallel channels) afforded by the configuration. Remarkably, the capacity of all channels in the family admits a simple closed-form asymptotic approximation (see (1)), whose validity is established using results from random matrix theory. The capacity formula reveals a new *multiplexing gain versus received SNR trade-off* that is not available in rich channels. Using the formula for a fixed number of antennas, we show that for every operating transmit SNR there is an optimal channel configuration that optimizes the multiplexing gain-received SNR trade-off to yield the highest capacity (see Fig. 3). Surprisingly, three canonical channel configurations suffice for near-optimal performance over the entire SNR range: the beamforming configuration (that maximizes received SNR) in the low-SNR regime; the multiplexing configuration (that maximizes the multiplexing gain) in the high-SNR regime; and the ideal channel configuration (that maintains an $\mathcal{O}(\sqrt{D})$ multiplexing gain) in the

medium-SNR range.

Complementing our results on ergodic capacity, we also analyze the outage capacity of sparse channels. In this context, the multiplexing gain-received SNR 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. An interpretation of our results in view of recent connections between mutual information and MMSE estimation reveals a rate-distortion tradeoff that is optimized by the three canonical configurations as a function of transmit SNR. Finally, we demonstrate the superiority of the beamforming configuration from the viewpoint of spectral efficiency in the low-SNR/wideband limit.

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 [12] 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 the receiver to the level of sparsity and operating SNR. The implications for wideband communication are significant: the beamforming configuration is N -times more spectrally efficient compared to the multiplexing configuration that represents a fixed-array operation. Another promising direction is to investigate the implications of the results in this paper and [12] in the context of cooperative communication and distributed beamforming in wireless networks. Finally, a third promising direction is to explore the implications of multipath sparsity on non-coherent communication in time, frequency and space, as prompted by our recent results that reveal a learnability versus diversity tradeoff in wideband communication over time-varying multipath channels [20], [13].

APPENDIX

I. REALIZING NON-INTEGER p AND q

Non-integer values of (p, q) for a given α and δ_p can be realized by appropriate time sharing (linear combination) of four channels with dimensions given by $(\lfloor p \rfloor, \lfloor q \rfloor)$, $(\lfloor p \rfloor, \lceil q \rceil)$, $(\lceil p \rceil, \lfloor q \rfloor)$ and $(\lceil p \rceil, \lceil q \rceil)$. If η_1, η_2, η_3 and $\eta_4 = (1 - \eta_1 - \eta_2 - \eta_3)$ represent respectively the fraction of communication time that each of the four channels are used to construct an effective (p, q)

channel, then the following holds:

$$\begin{aligned} \lfloor p \rfloor (\eta_1 + \eta_2) + \lceil p \rceil (1 - \eta_1 - \eta_2) &= p \\ \lfloor q \rfloor (\eta_1 + \eta_3) + \lceil q \rceil (1 - \eta_1 - \eta_3) &= q. \end{aligned} \quad (52)$$

Using the fact that $\lceil p \rceil = \lfloor p \rfloor + 1$, we have $(1 - \eta_1 - \eta_2) = p - \lfloor p \rfloor$ and $(1 - \eta_1 - \eta_3) = q - \lfloor q \rfloor$. Parameterizing all the variables in terms of η_1 , we have $\eta_2 = 1 - \eta_1 - p + \lfloor p \rfloor$, $\eta_3 = 1 - \eta_1 - q + \lfloor q \rfloor$ and $\eta_4 = p - \lfloor p \rfloor + q - \lfloor q \rfloor + \eta_1 - 1$. Using the condition that $0 \leq \eta_i \leq 1$ for all i , we can obtain the following bound for η_1 :

$$\max((1 - p + \lfloor p \rfloor - q + \lfloor q \rfloor), 0) \leq \eta_1 \leq \min((1 - p + \lfloor p \rfloor), (1 - q + \lfloor q \rfloor)). \quad (53)$$

Note that the left-hand side of (53) follows from $\eta_1 \geq 0$ and $\eta_4 \geq 0$ while the right-hand side is a consequence of $\eta_2 \geq 0$ and $\eta_3 \geq 0$. The other constraints lead to weaker conditions subsumed by (53). A choice of η_1 that satisfies (53) would effectively create the (q, p) channel.

Let $C_{\text{erg}}(q, p, \rho)$ and $\sigma^2(q, p, \rho)$ denote the mean and variance of capacity of channel with dimensions (p, q) . Using the time-sharing concept described above, $C_{\text{erg}}(q, p, \rho)$ and $\sigma^2(q, p, \rho)$ for non-integer values of p and q are given by

$$\begin{aligned} C_{\text{erg}}(q, p, \rho) &= \eta_1 C_{\text{erg}}(\lfloor q \rfloor, \lfloor p \rfloor, \rho) + \eta_2 C_{\text{erg}}(\lceil q \rceil, \lfloor p \rfloor, \rho) \\ &\quad + \eta_3 C_{\text{erg}}(\lfloor q \rfloor, \lceil p \rceil, \rho) + \eta_4 C_{\text{erg}}(\lceil q \rceil, \lceil p \rceil, \rho), \\ \sigma^2(q, p, \rho) &= \eta_1 \sigma^2(\lfloor q \rfloor, \lfloor p \rfloor, \rho) + \eta_2 \sigma^2(\lceil q \rceil, \lfloor p \rfloor, \rho) \\ &\quad + \eta_3 \sigma^2(\lfloor q \rfloor, \lceil p \rceil, \rho) + \eta_4 \sigma^2(\lceil q \rceil, \lceil p \rceil, \rho). \end{aligned} \quad (54)$$

II. FUNDAMENTAL LIMIT OF CAPACITY SCALING

Proof of Theorem 1: By applying Jensen's inequality to (18) and noting that a diagonal \mathbf{Q} ($= \text{diag}(\mathbf{Q}_i)$) achieves capacity for the class $\mathcal{H}(D)$, the optimal capacity can be written as

$$\begin{aligned} C_{\text{opt,erg}}(N, \rho, D) &\leq \max_{\mathbf{H} \in \mathcal{H}(D)} \max_{\text{Tr}(\mathbf{Q}) \leq \rho} \log_2 \det(\mathbf{I} + \mathbf{Q} \mathbf{E}_{\mathbf{H}}[\mathbf{H}^H \mathbf{H}]) \\ &= \max_{\{P_i\}} \max_{\{\mathbf{Q}_i\}} \sum_{i=1}^N \log_2(1 + P_i \mathbf{Q}_i) \end{aligned} \quad (55)$$

where $P_i = \sum_k \mathbf{E} |\mathbf{H}(k, i)|^2 = \sum_k \Psi(k, i)$ represent the column powers of the channel \mathbf{H} and

the optimization is subject to the twin constraints

$$\begin{aligned} P_i &\geq 0, & \sum_{i=1}^N P_i &= \rho_c \\ Q_i &\geq 0, & \sum_{i=1}^N Q_i &= \rho. \end{aligned}$$

We now recast this optimization¹¹ as follows:

$$\begin{aligned} \max_p \max_{\{P_i\}} \max_{\{Q_i\}} \sum_{i=1}^p \log_2(1 + P_i Q_i) \quad \text{s.t.} \quad & 1 \leq p \leq N, \\ P_i > 0, \quad i = 1 \cdots p \quad \text{and} \quad P_i = 0, \quad i > p \quad \text{with} \quad & \sum_{i=1}^p P_i = \rho_c \\ \text{and} \quad Q_i \geq 0, \quad \text{with} \quad & \sum_{i=1}^p Q_i = \rho. \end{aligned}$$

Given p fixed between 1 and N , the inner optimization can be written in terms of Lagrange multipliers λ_1 and λ_2 as

$$\max_{\{P_i\}} \max_{\{Q_i\}} \sum_{i=1}^p \log_2(1 + P_i Q_i) + \lambda_1 \left(\rho_c - \sum_{i=1}^p P_i \right) + \lambda_2 \left(\rho - \sum_{i=1}^p Q_i \right). \quad (56)$$

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 to 0. It is then easy to check that $\frac{Q_i}{P_i}$ is independent of i . Using this relationship to meet $\sum_{i=1}^p Q_i = \rho$, we see that $Q_i = \frac{P_i}{\rho_c} \rho$. Now revisiting the partial derivative with respect to Q_i under the constraint that $Q_i = \frac{P_i}{\rho_c} \rho$ and setting the derivative to 0, we get $P_i = \frac{\rho_c}{p}$. It is easy to verify that the second derivative condition is satisfied by the maxima.

Using this choice of Q_i in (55), we have

$$C_{\text{opt, erg}}(N, \rho, D) \leq \max_p p \log_2 \left(1 + \rho \frac{\rho_c}{p^2} \right). \quad (57)$$

The p that achieves this maximum satisfies $p = \mathcal{O}(\sqrt{\rho_c})$. This choice of p leads to the statement of Theorem 1. ■

¹¹Note that even though p runs through 1 to N in the following optimization, not all possible choices for p may be feasible, for e.g., if $\rho_c = N^\gamma, \gamma < 1$, then p is constrained by $1 \leq p \leq \rho_c$. We do not bother with this technicality here.

III. ASYMPTOTIC ERGODIC CAPACITY ANALYSIS

We need a few eigenvalue characterizations [41] to understand the ergodic capacity of the family of channels studied here. The first two lemmas are needed to study capacity in the beamforming regime.

Lemma 2: Let \mathbf{H} be a $p \times q$ complex random matrix with i.i.d. entries having mean zero and variance one. Let $\frac{p}{q} \rightarrow 0$ with p being finite and $q \rightarrow \infty$. Then

$$\left\| \frac{\mathbf{H}\mathbf{H}^H}{q} - \mathbf{I}_p \right\| = \mathcal{O}_p\left(\frac{1}{\sqrt{q}}\right) \quad (58)$$

where $\|\cdot\|$ refers to any matrix norm and the subscript p corresponds to convergence in probability. Thus, the eigenvalues of $\frac{\mathbf{H}\mathbf{H}^H}{q}$ converge to 1 almost surely.

Remark 2: Note that the conclusion of Lemma 2 follows from the law of large numbers and depends critically on the fact that p is finite while $q \rightarrow \infty$ [42]. If this condition fails to hold, the spectral behavior of $\mathbf{H}\mathbf{H}^H$ can be significantly different as shown below.

Lemma 3 (Bai and Yin, 1988) : Let \mathbf{H} be a $p \times q$ complex random matrix with i.i.d. entries having mean zero, variance one and finite fourth moment. Assume that $p \rightarrow \infty$ with $\frac{p}{q} \rightarrow 0$. Then with probability 1 the empirical spectral density of $\mathbf{G} = \frac{\mathbf{H}\mathbf{H}^H - q\mathbf{I}}{\sqrt{pq}}$ converges to $f_{\mathbf{G}}(\lambda) = \frac{\sqrt{4-\lambda^2}}{2\pi}$, $|\lambda| \leq 2$.

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

Lemma 4 (Grenander and Silverstein, 1977) : Let q and p be such that $\frac{q}{p} \in (0, 1]$ as $p \rightarrow \infty$. The empirical spectral density of $\mathbf{G} = \frac{\mathbf{H}\mathbf{H}^H}{q}$ where \mathbf{H} is the ideal channel converges pointwise in probability to $f_{\mathbf{G}}(\lambda) = \frac{1}{2\pi} \sqrt{\frac{4-\lambda}{\lambda}}$, $0 \leq \lambda \leq 4$.

Proof of Theorem 2: The structure of the channel configurations in the beamforming regime depend on whether $\gamma \in (0, 1]$ or $\gamma \in (1, 2)$, as illustrated in Fig. 4. If $\gamma \in (0, 1]$, the dimensionality of \mathbf{H} is $\mathcal{O}(N^\gamma) \times 1$, while if $\gamma \in (1, 2)$, the dimensionality is $N \times \mathcal{O}(N^{\gamma-1})$. In either case, $\frac{p}{q} \rightarrow 0$. If $\gamma \in (0, 1]$, the result immediately follows from Lemma 2. This is because $\mathbf{H}^H\mathbf{H}$ is a finite dimensional matrix in the limit of N and the law of large numbers can be applied to each entry of $\mathbf{H}^H\mathbf{H}$. Thus, $\frac{p}{q} \mathbf{H}^H\mathbf{H} \rightarrow \frac{pq}{p} \mathbf{I}_p$.

However, if $\gamma \in (1, 2)$, $\mathbf{H}^H\mathbf{H}$ is an infinite-dimensional matrix in the asymptotics of N and it is not very clear how to apply law of large numbers to an infinite number of entries. Luckily, the empirical eigenvalue distribution of \mathbf{H} has been well studied in this case, as documented

by Lemma 3. Our proof takes a recourse to this result. The central theme that ties both the cases is that, in the limit of large dimensions, there is an eigen-hardening which has also been documented in recent works in different contexts [43], [44]. We now prove the theorem for the $\gamma \in (1, 2)$ case.

First, the capacity of the beamforming channel normalized by $C_{\text{appx}}(N)$ can be written as

$$\begin{aligned}
\frac{C_{\text{erg, bf}}(N)}{p \log_2 \left(1 + \rho \frac{q}{p}\right)} &= \frac{\mathbf{E} \left[\sum_{i=1}^p \log_2 \left(1 + \rho \frac{\lambda_i(\mathbf{H}\mathbf{H}^H)}{p}\right) \right]}{p \log_2 \left(1 + \rho \frac{q}{p}\right)} \\
&\stackrel{(a)}{=} \frac{\mathbf{E} \left[\sum_{i=1}^p \log_2 \left(1 + \rho \left(\mu_i \sqrt{\frac{q}{p}} + \frac{q}{p}\right)\right) \right]}{p \log_2 \left(1 + \rho \frac{q}{p}\right)} \\
&\stackrel{(b)}{=} 1 + \frac{\mathbf{E} \left[\sum_{i=1}^p \log_2 (1 + t\mu_i) \right]}{p \log_2 \left(1 + \rho \frac{q}{p}\right)} \\
&\stackrel{(c)}{\rightarrow} 1 + \frac{1}{\log_2 \left(1 + \rho \frac{q}{p}\right)} \cdot \int_{(-\infty, \infty)} \log_2 (1 + t\eta) dF_{\mathbf{G}}(\eta) \\
&\stackrel{(d)}{=} 1 + \frac{1}{\log_2 \left(1 + \rho \frac{q}{p}\right)} \cdot \frac{\log_2(e)}{2\pi} \cdot \underbrace{\int_{[-2, 2]} \log_e (1 + t\eta) \sqrt{4 - \eta^2} d\eta}_I
\end{aligned} \tag{59}$$

where μ_i represent the eigenvalues of $\mathbf{G} = \frac{\mathbf{H}\mathbf{H}^H - q\mathbf{I}}{\sqrt{pq}}$ in (a), $t = \frac{\rho\sqrt{\frac{q}{p}}}{1+\rho\frac{q}{p}}$ in (b) and (c) follows from the convergence of μ_i to $f_{\mathbf{G}}(\eta)$ in Lemma 3, and $F_{\mathbf{G}}(\eta)$ denotes the corresponding distribution.

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

$$\begin{aligned}
I &= 2 \int_{-2}^2 \frac{\log_e(1 + t\eta)}{\sqrt{4 - \eta^2}} d\eta + \frac{1}{2} \int_{-2}^2 \frac{\sqrt{4 - \eta^2}}{1 + t\eta} d\eta - \frac{1}{2} \int_{-2}^2 \sqrt{4 - \eta^2} d\eta \\
&\stackrel{(e)}{=} 2\pi \log_e \left(\frac{1 + \sqrt{1 - 4t^2}}{2} \right) - \pi + \frac{1}{2t} \underbrace{\int_{-2}^2 \frac{\sqrt{4 - \eta^2}}{\eta + \frac{1}{t}} d\eta}_{I_1}
\end{aligned} \tag{60}$$

where (e) follows from [45, 4.292(3)].

I_1 can now be computed using standard techniques as

$$\begin{aligned}
I_1 &\stackrel{(f)}{=} - \int_{-2}^2 \frac{\eta}{\sqrt{4-\eta^2}} d\eta + \frac{1}{t} \int_{-2}^2 \frac{d\eta}{\sqrt{4-\eta^2}} + \left(4 - \frac{1}{t^2}\right) \int_{-2}^2 \frac{d\eta}{\left(\eta + \frac{1}{t}\right) \sqrt{4-\eta^2}} \\
&\stackrel{(g)}{=} \frac{1}{t} \int_{-2}^2 \frac{d\eta}{\sqrt{4-\eta^2}} - \left(4 - \frac{1}{t^2}\right) \int_{\frac{t}{1+2t}}^{\frac{t}{1-2t}} \frac{d\eta}{\sqrt{\eta^2 \left(4 - \frac{1}{t^2}\right) + \frac{2\eta}{t} - 1}} \\
&\stackrel{(h)}{=} \frac{\pi}{t} - \frac{\pi\sqrt{1-4t^2}}{t}
\end{aligned} \tag{61}$$

where (f) follows from [45, 2.282(1)], (g) from [45, 2.281] and (h) from [45, 2.261(3)]. Thus $C_{\text{erg, bf}}(N)$ is

$$\begin{aligned}
C_{\text{erg, bf}}(N) &= p \log_2 \left(1 + \rho \frac{q}{p}\right) + p \log_2 \left(\frac{1 + \sqrt{1-4t^2}}{2}\right) \\
&\quad + \frac{\log_2(e)p}{4t^2} \cdot \left(1 - \sqrt{1-4t^2}\right) - \frac{\log_2(e)p}{2}
\end{aligned} \tag{62}$$

The conclusion leading to the correction term $\Delta C_{\text{bf}}(N)$ follows trivially by using a Taylor's series expansion for $\frac{1}{\kappa_0}$ which converges to 0 in the beamforming regime. This completes the proof of the theorem. \blacksquare

Proof of Theorem 3: When $\kappa_0 \in (1, \infty)$, the ideal channel reduces to a $q \times p$ i.i.d. channel according to Definition 4. The i.i.d. channel capacity formula from [46] is precisely the statement of the theorem. When $\kappa_0 \in (0, 1]$, the ideal channel definition reduces to a q -connected p -dimensional channel [4]. Using Lemma 4, it is not difficult to check that (28) holds (see [4, Theorem 5] for more details). \blacksquare

Proof of Proposition 1: We need to compute the high-SNR and the low-SNR trends for the $\kappa_0 \leq 1$ and $\kappa_0 > 1$ cases separately. First, we consider the $\kappa_0 > 1$ case. If $\kappa_0 > 1$ and $\rho \rightarrow \infty$, a Taylor's series analysis of h (as a function of $\frac{1}{\rho}$ around 0) in (27) shows that

$$h \rightarrow 1 - \frac{1}{\rho(\kappa_0 - 1)} + \frac{\kappa_0}{\rho^2(\kappa_0 - 1)^3}. \tag{63}$$

Using this, it is easy to see that

$$\begin{aligned}
\frac{C_{\text{erg, id}}(N)}{\log_2(e)q} &\rightarrow \left(\frac{1}{\kappa_0} \log_e \left(\frac{\kappa_0}{\kappa_0 - 1} + \rho(\kappa_0 - 1) - \frac{\kappa_0}{\rho(\kappa_0 - 1)^3} \right) \right. \\
&\quad \left. + \log_e \left(\frac{\kappa_0}{\kappa_0 - 1} - \frac{\kappa_0}{\rho(\kappa_0 - 1)^3} \right) - \left(\frac{1}{\kappa_0} - \frac{1}{\rho\kappa_0(\kappa_0 - 1)} \right) \right)
\end{aligned}$$

and this results in (29). As $\rho \rightarrow 0$, another Taylor's series expansion (as a function of ρ around 0) shows that $h \rightarrow \rho\kappa_0 - \rho^2\kappa_0(1 + \kappa_0)$ and

$$C_{\text{erg, id}}(N) \rightarrow \log_2(e) \rho q \left(1 - \frac{5}{2} \rho(1 + \kappa_0) \right) \quad (64)$$

from which (29) follows immediately. For the $\kappa_0 \leq 1$ case, trivial computations yield

$$C_{\text{erg, id}}(N) \rightarrow \log_2(e) p \left(\log_e(\rho\kappa_0) - 1 - \frac{1}{2\rho\kappa_0} \right) \quad (65)$$

as $\rho \rightarrow \infty$ and $C_{\text{erg, id}}(N) \rightarrow \log_2(e) q \rho (1 - 5\rho\kappa_0)$ as $\rho \rightarrow 0$. The conclusions in (29) are again immediate. ■

Proof of Theorem 4: The definition of \mathbf{H} in the multiplexing regime in Definition 4 leads to two possible cases: Either q is a constant and $\frac{q}{p} \rightarrow 0$ or $q \rightarrow \infty$ and $\frac{q}{p} \rightarrow 0$. Note that Lemma 4 applies only in the case $\frac{q}{p} > 0$. So in either of the two cases, listed above, we do not have a knowledge (closed-form expressions) of the converging empirical eigenvalue distribution.

However, the implicit characterization of capacity in [38] and the regularity of the multiplexing channel implies that we can obtain closed-form expressions for capacity in closed form in either case in the limit of N . In fact, after simplification, the capacity formula is seen to be identical to that of an ideal channel with $\kappa_0 < 1$. This lends credence to the fact that the converging distributions (in both cases) are the same. The correction term follows by applying a Taylor's series expansion of capacity as a function of ρ and $\frac{1}{\rho}$ around the point 0 in the low- and high-SNR regimes, respectively. ■

IV. ASYMPTOTIC OUTAGE CAPACITY ANALYSIS

Proof: As described in Appendix III, there are two possibilities in the beamforming regime: either 1) p is constant and $q \rightarrow \infty$, or 2) $\{p, q\} \rightarrow \infty$ with $\frac{p}{q} \rightarrow 0$. In the first case, the variance follows from [35, Theorem 1], and in the second case, by computing the variance of capacity for the channel with $\kappa_0 = \frac{q}{p}$ constant, and letting $\kappa_0 \rightarrow \infty$ [34].

In the ideal regime, we again have two cases: 1) \mathbf{H}_{id} is a $q \times p$ i.i.d. matrix with $\kappa_0 \geq 1$ and 2) \mathbf{H}_{id} is a q -connected p -dimensional matrix with $\kappa_0 < 1$. In the first case, $\sigma_{\text{id}}^2(N)$ follows from [34], while in the latter case, we note that the capacity random variable is the same as that of a $p \times p$ i.i.d. channel with effective transmit power $\rho\kappa_0$ [4]. In this case, the variance of capacity then follows by setting $\kappa_0 = 1$ and using the appropriate value for ρ .

For the multiplexing regime, we proceed via the same logic that governs the ideal channel with $\kappa_0 < 1$. Note that the effective transmit power $\rho\kappa_0 \rightarrow 0$, and thus the variance follows from the i.i.d. channel formula in the low-SNR regime [35]. In the low-SNR regime, $h_1 \rightarrow \rho\kappa_0$ and substitution in $\sigma_{\text{id}}^2(N)$ yields the result. ■

V. MUTUAL INFORMATION AND MMSE ESTIMATION

Proof of Proposition 6: The fact that we have two distinct solutions for $\Gamma(t)$ is a consequence of the structure of the family of channels studied. In the beamforming and the ideal regimes with $\kappa_0 > 1$, 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 $\kappa_0 \leq 1$ and in the multiplexing regime, the channel \mathbf{H} is a q -connected p -dimensional channel as illustrated in Fig. 4.

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 Γ and the associated MSE quantity Υ satisfies the relationship $\Upsilon = \frac{\rho}{1+\rho\Gamma}$. Substituting for Υ in (46), we have

$$\Gamma = \kappa \mathbf{E}_R \left[\frac{\mathcal{G}(R, t)}{1 + \frac{\rho}{1+\rho\Gamma} \mathbf{E}_T[\mathcal{G}(R, T)|R]} \right]. \quad (66)$$

With the setup as described, consider the first case. The quantity κ (the ratio of receive and transmit dimensions) reduces to κ_0 in this case. Next, note that $\mathbf{E}_T[\mathcal{G}(r, T)]$ is independent of r and is equal to 1. Thus, the conditional random variable $\mathbf{E}_T[\mathcal{G}(R, T)|R] = 1$. Plugging this and simplifying (66), we have

$$\Gamma \left(1 + \frac{\rho}{1 + \rho\Gamma} \right) = \kappa_0 \mathbf{E}_R[\mathcal{G}(R, t)]. \quad (67)$$

Substitute $\mathbf{E}_R[\mathcal{G}(R, t)] = 1$ and solving for the quadratic equation in (67), we get (48). We proceed on similar lines in the second case. The first difference is that κ takes the value 1 here. $\mathbf{E}_T[\mathcal{G}(r, T)]$ and $\mathbf{E}_R[\mathcal{G}(R, t)]$ are still independent of r and t , respectively, but they take a value of κ_0 . Now solving for the resultant quadratic, we get (49). ■

REFERENCES

- [1] Í. E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," *Eur. Trans. Telecomm.*, vol. 10, pp. 2172–2178, 2000.
- [2] G. J. Foschini and M. J. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

- [3] C. N. Chuah, J. M. Kahn, and D. N. C. Tse, "Capacity Scaling in MIMO Wireless Systems under Correlated Fading," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.
- [4] K. Liu, V. Raghavan, and A. M. Sayeed, "Capacity Scaling and Spectral Efficiency in Wideband Correlated MIMO Channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2504–2526, Dec 2003.
- [5] V. Raghavan and A. M. Sayeed, "Role of Channel Power in the Sub-Linear Capacity Scaling of MIMO Channels," *Allerton Conf. Commun. Cont. and Comp.*, 2004.
- [6] A. M. Sayeed, "Deconstructing Multi-Antenna Fading Channels," *IEEE Trans. Sig. Proc.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.
- [7] Y. Zhou, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental Study of MIMO Channel Statistics and Capacity via the Virtual Channel Representation," *Revised, IEEE Trans. Antennas Propagation*, 2006.
- [8] L. W. Hanlen, R. Timo, and R. Perera, "On Dimensionality for Sparse Multipath," Tech. Rep., NICTA, 2005.
- [9] 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. Inform. Theory*, vol. 51, no. 2, pp. 523–536, Feb. 2005.
- [10] H. M. Jones, R. A. Kennedy, and T. D. Abhayapala, "On Dimensionality of Multipath Fields: Spatial Extent and Richness," *Proc. ICASSP*, vol. 3, pp. 2837–2840, May 2002.
- [11] A. M. Sayeed, V. Raghavan, and J. H. Kotecha, "Capacity of Space-Time Wireless Channels: A Physical Perspective," *Proc. ITW*, 2004.
- [12] A. M. Sayeed and V. Raghavan, "The Ideal MIMO Channel: Maximizing Capacity in Sparse Multipath with Reconfigurable Arrays," *Proc. ISIT*, 2006.
- [13] A. M. Sayeed, "Sparse Multipath Wireless Channels: Modeling and Implications," *Proc. ASAP*, 2006.
- [14] G. D. Forney, Jr., "On the Role of MMSE Estimation in Approaching the Information-Theoretic Limits of Linear Gaussian Channels: Shannon Meets Wiener," *Allerton Conf. Commun. Cont. and Comp.*, 2003.
- [15] D. Guo, S. Shamai, and S. Verdú, "Mutual Information and Minimum Mean-Square Error in Gaussian Channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1261–1282, Apr 2005.
- [16] A. M. Tulino, A. Lozano, and S. Verdú, "Capacity Achieving Input Covariance for Correlated Multi-Antenna Channels," *Allerton Conf. Commun. Cont. and Comp.*, 2003.
- [17] G. Barriac and U. Madhow, "Space-time Communication for OFDM with Implicit Channel Feedback," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3111–3129, Dec. 2004.
- [18] X. Wu and R. Srikant, "MIMO Channels in the Low SNR Regime: Communication Rate, Error Exponent and Signal Peakiness," *Submitted, IEEE Trans. Inform Theory*, 2004.
- [19] V. Raghavan, G. Hariharan, and A. M. Sayeed, "Exploiting Time-Frequency Coherence to Achieve Coherent Capacity in Wideband Wireless Channels," *Allerton Conf. Commun. Cont. and Comp.*, 2005.
- [20] G. Hariharan and A. M. Sayeed, "On the Reliability of Wideband Communication: Impact of Channel Coherence and Sparsity," *To be submitted*, 2006.
- [21] W. Weichselberger, M. Herdin, H. Ozelik, 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.
- [22] J. H. Kotecha, V. Raghavan, and A. M. Sayeed, "Canonical Statistical Models for Correlated MIMO Fading Channels and Capacity Analysis," *Revised, IEEE Trans. Commun.*
- [23] H. Ozelik, 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.
- [24] J. Wallace, H. Ozelik, M. Herdin, E. Bonek, and M. Jensen, "Power and Complex Envelope Correlation for Modeling

- Measured Indoor MIMO Channels: A Beamforming Evaluation,” *IEEE 58th Fall Vehicular Technology Conference, 2003*, vol. 1, pp. 363–367, Oct. 2003.
- [25] V. Veeravalli, Y. Liang, and A. Sayeed, “Correlated MIMO Rayleigh Fading Channels: Capacity, Optimal Signaling and Asymptotics,” *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 2058–2072, June 2005.
- [26] 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. Inform. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.
- [27] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, “Capacity Limits of MIMO Channels,” *IEEE Journ. Selected Areas in Commun.*, vol. 21, no. 5, pp. 684–702, June 2003.
- [28] L. H. Ozarow, S. Shamai (Shitz), and A. Wyner, “Information Theoretic Considerations for Cellular Mobile Radio,” *IEEE Trans. Veh. Tech.*, vol. 43, pp. 359–378, May 1994.
- [29] Z. Wang and G. B. Giannakis, “Outage Mutual Information Rate of Space-Time MIMO Channels,” *IEEE Trans. Inform. Theory*, vol. 50, no. 4, pp. 657–662, Apr. 2004.
- [30] P. J. Smith, S. Roy, and M. Shafi, “Capacity of MIMO systems with Semicorrelated Flat Fading,” *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2781–2788, Oct. 2003.
- [31] M. Kang and M. S. Alouini, “Impact of Correlation on the Capacity of MIMO Channels,” *IEEE ICC*, pp. 2623–2627, 2003.
- [32] A. M. Tulino and S. Verdú, “Asymptotic Outage Capacity of Multiantenna Channels,” *IEEE ICASSP*, 2005.
- [33] A. Scaglione, “Statistical Analysis of Capacity of MIMO Frequency Selective Rayleigh Fading Channels with Arbitrary Number of Inputs and Outputs,” *IEEE ISIT*, 2002.
- [34] M. A. Kamath and B. L. Hughes, “The Asymptotic Capacity of Multiple Antenna Rayleigh Fading Channels,” *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4325–4333, Dec. 2005.
- [35] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, “Multiple-Antenna Channel Hardening and its Implications for Rate Feedback and Scheduling,” *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1893–1909, Sept 2004.
- [36] V. Raghavan and A. M. Sayeed, “Weak Convergence and Rate of Convergence of MIMO Capacity Random Variable,” *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3799–3809, Aug. 2006.
- [37] T. Muharemovic, A. Sabharwal, and B. Aazhang, “Antenna Packing in Low Power Systems: Communication Limits and Array Designs,” *Submitted, IEEE Trans. Inform. Theory*, 2005.
- [38] A. M. Tulino, A. Lozano, and S. Verdú, “Impact of Antenna Correlation on the Capacity of Multiantenna Channels,” *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2491–2509, July 2005.
- [39] S. Verdú, “Spectral Efficiency in the Wideband Regime,” *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.
- [40] I. S. Reed, “On a Moment Theorem for Complex Gaussian Processes,” *IRE Trans. Inform. Theory*, vol. 8, pp. 194–195, April 1962.
- [41] Z. D. Bai, “Methodologies in Spectral Analysis of Large Dimensional Random Matrices, A Review,” *Statistica Sinica*, vol. 9, pp. 611–677, 1999.
- [42] 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.
- [43] V. Raghavan, R. W. Heath, Jr., and A. M. Sayeed, “Systematic Codebook Designs for Quantized Beamforming in Correlated MIMO Channels,” *IEEE Jour. Select. Areas in Commun.*, *Under review*, 2006.
- [44] V. Raghavan and A. M. Sayeed, “Impact of Spatial Correlation on Statistical Precoding in MIMO Channels with Linear Receivers,” *Allerton Conf. Commun. Cont. and Comp.*, 2006.

- [45] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, NY, 4th edition, 1965.
- [46] P. B. Rapajic and D. Popescu, "Information Capacity of a Random Signature Multiple-Input Multiple-Output Channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 8, pp. 1245–1248, Aug. 2000.