Multiple-Antenna Capacity in the Low-Power Regime

Angel Lozano, *Senior Member, IEEE*, Antonia M. Tulino, *Member, IEEE*, and Sergio Verdú, *Fellow, IEEE*

***Abstract*—This paper provides analytical characterizations of the impact on the multiple-antenna capacity of several important features that fall outside the standard multiple-antenna model, namely: i) antenna correlation, ii) Ricean factors, iii) polarization diversity, and iv) out-of-cell interference; all in the regime of low signal-to-noise ratio. The interplay of rate, bandwidth, and power is analyzed in the region of energy per bit close to its minimum value. The analysis yields practical design lessons for arbitrary number of antennas in the transmit and receive arrays.**

***Index Terms*—Antenna correlation, channel capacity, cochannel interference, fading channels, low-power regime, multiple-antenna arrays, Rayleigh fading, Ricean fading.**

I. Introduction

A. The Multiple-Antenna Problem

THE use of multiple transmit and receive antennas can enable very large increases in capacity per unit of bandwidth. As a result, the multiple-antenna problem has been propelled to the research forefront in communication theory. The single-user capacity, in particular, has been thoroughly studied [1]–[12]. Most such studies, however, are restricted to a highly idealized canonical channel—uncorrelated zero-mean equal-variance transfer coefficients from each transmit to each receive antenna—impaired by additive white Gaussian noise (AWGN). Although an integral solution [2] and even a closed-form expression [3] for the capacity of this canonical channel exist, their complexity precludes analytical insight despite the simplicity of the model. With the aid of large random matrix theory, more explicit formulas for the capacity of this canonical channel can be obtained asymptotically in the number of antennas [13]–[15]. Since both the finite as well as the asymptotic capacities become particularly revealing in the high-power regime, design guidelines and insights into the capacity benefits that accrue as function of the number of antennas have been drawn mostly in that regime.

Beyond the canonical model, the capacity of more realistic channels has been studied mostly through Monte Carlo simulation [16]–[18]. In terms of analysis, random matrix theory has also been applied to the study of zero-mean channels with transmit and receive antenna correlations yielding an asymptotic fixed-point integral solution that, again, provides limited insight [19]. The impact of out-of-cell interference on the capacity, in turn, has also been studied both through simulation [20], [21] and asymptotically in the number of antennas [22]–[25].

In this paper, we study the single-user capacity in the low-power regime using models that characterize the channel and noise encountered in typical wireless systems. Focusing on those scenarios where the transmitter cannot (or chooses not to) track the channel, analytical insight is derived without the need to invoke a large number of antennas.

B. The Low-Power Regime

Since efficient bandwidth utilization requires aggressive frequency reuse across neighboring cells and sectors, high-capacity wireless systems are—by design—limited by their own interference. This has two immediate consequences.

- The noise is dominated by out-of-cell interference, potentially colored in space and subject to fading, rather than by thermal noise.
- Since, because of pure geometry, a majority of locations lie in the periphery of their cells, users must operate very often at low signal-to-noise ratio (SNR) while only rarely at high SNR.

This second point can be illustrated in the context of emerging third-generation data systems [26], [27], wherein almost 40% of geographical locations experience receiver SNR levels below 0 dB while less than 10% display levels above 10 dB.

Despite its relevance, the multiple-antenna low-SNR regime had not been analyzed in depth until [28] where, in contrast with most former multiple-antenna analyses, the figure of merit is not the SNR, but rather the normalized energy per information bit, $\frac{E_b}{N_0}$. As shown in [28], the analysis of the low-SNR capacity as function of per-symbol SNR may lead to misleading conclusions. Denoting the capacity by $\mathsf{C}(\frac{E_b}{N_0})$, a system with transmit power P, desired rate R (in bits/s), and bandwidth B must respect the fundamental limit

$$\frac{R}{B} \leq \mathsf{C}\left(\frac{P}{RN_0}\right). \tag{1}$$

Also shown in [28] is that the key performance measures in the low-SNR regime are $\frac{E_b}{N_0}_{\min}$ (the minimum energy per information bit required to convey any positive rate reliably) and S_0, the capacity slope therein in bits/s/Hz/(3 dB). These two quantities determine the first-order behavior of the capacity as a function of $\frac{E_b}{N_0}$ (in dB) via

$$\mathsf{C}\left(\frac{E_b}{N_0}\right) = S_0 \frac{\frac{E_b}{N_0}\Big|_{\mathrm{dB}} - \frac{E_b}{N_0\,\min}\Big|_{\mathrm{dB}}}{3\ \mathrm{dB}} + \epsilon \tag{1}$$

Manuscript received October 15, 2002; revised June 16, 2003. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Yokohama, Japan, June/July 2003.

A. Lozano is with Bell Laboratories (Lucent Technologies), Holmdel, NJ07733, USA (e-mail: aloz@lucent.com).

A. M. Tulino is with Universitá Degli Studi di Napoli (Federico II), 80125 Napoli, Italy (e-mail: atulino@ee.princeton.edu).

S. Verdú is with the Department of Electrical Engineering, Princeton University, Princeton, NJ08540, USA (e-mail: verdu@ee.princeton.edu).

Communicated by B. Hassibi, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2003.817429

with $\epsilon = o\left(\frac{E_b}{N_0} - \frac{E_b}{N_0}_{\min}\right)$. It follows that a system designed to achieve rate R with power P requires bandwidth

$$B = \frac{R}{\mathsf{C}\left(\frac{P}{RN_0}\right)} \approx \frac{R}{S_0} \frac{3\text{ dB}}{\frac{P}{RN_0}\Big|_{\text{dB}} - \frac{E_b}{N_0}_{\min}\Big|_{\text{dB}}}$$

where the approximation sharpens as $\frac{E_b}{N_0} \downarrow \frac{E_b}{N_0}_{\min}$.

C. Scope

In the remainder, we build on and expand the findings of [28] using channel and noise models that realistically describe the conditions found in typical multiple-antenna systems. Specifically, we consider the impact on capacity of the following features that fall outside the canonical model:

- antenna correlation (modeled by separate correlation matrices at the transmit and receive arrays);
- Ricean components (modeled by a deterministic matrix);
- antenna polarization (modeled by a polarization matrix that accounts for the different power transfers between co-polarized and orthogonally polarized antennas);
- spatially colored noise (modeled by nondiagonal noise covariance matrices);
- time-varying interference (modeled by noise covariance matrices subject to fading).

Section II discusses the models we use to incorporate the above nonideal features. Section III derives expressions for the minimum energy per bit and the low-SNR slope for the general model introduced in Section II. Section IV analyzes the effects of antenna correlation and Ricean components on capacity. Section V studies the capacity with cross-polarization diversity focusing primarily on the case of two-antenna arrays. Section VI investigates the impact on capacity of out-of-cell interference, which is a source of additive noise that is colored and subject to fading.

We relegate the proofs of the results to the Appendix, where we evaluate several moments of the trace of the product of certain random matrices which, in addition to serving our purposes, may be of independent interest.

II. Definitions and Models

A. Definitions

With n_{T} transmit and n_{R} receive antennas, a baseband discrete-time complex-valued model for the multiple-antenna channel with frequency-flat fading is

$$\boldsymbol{y} = \sqrt{\mathsf{g}}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}$$

where $\boldsymbol{x}$ is the n_{T}-dimensional transmit vector while $\boldsymbol{y}$ and $\boldsymbol{n}$ are the received and Gaussian noise n_{R}-dimensional vectors. The channel is represented by the $(n_{\mathrm{R}} \times n_{\mathrm{T}})$ random matrix $\sqrt{\mathsf{g}}\boldsymbol{H}$, independent of both $\boldsymbol{x}$ and $\boldsymbol{n}$. For convenience, we choose to factor out the scalar $\sqrt{\mathsf{g}}$ so as to yield a normalized matrix $\boldsymbol{H}$, the average power of whose entries is unity, i.e.,

$$E\left[\mathrm{Tr}\{\boldsymbol{H}\boldsymbol{H}^\dagger\}\right] = n_{\mathrm{T}}n_{\mathrm{R}}.$$

Hence, g can be interpreted as the average channel gain.[1]

When $n_{\mathrm{R}} = n_{\mathrm{T}}$ or when a particular equation applies to both transmitter and receiver, we may use n to refer to the number of antennas generically.

The single-sided spectral density of the noise is denoted by

$$N_0 = \frac{E[\|\boldsymbol{n}\|^2]}{n_{\mathrm{R}}}$$

and its normalized spatial covariance is defined as

$$\boldsymbol{\Phi}_{\boldsymbol{n}} \triangleq \frac{E[\boldsymbol{n}\boldsymbol{n}^\dagger]}{N_0}.$$

Most wireless systems are equipped with pilots that may be used to obtain an estimate of the channel. Indeed, in many low-SNR scenarios of interest, the realization of $\boldsymbol{H}$ can be tracked reliably as long as its coherence time (in symbols) is sufficiently large with respect to n_{T} [29]–[31]. We thus focus, throughout the paper, on the coherent regime where $\boldsymbol{H}$ is known at the receiver. As shown in [28], noncoherence incurs a severe bandwidth penalty.

Likewise, $\boldsymbol{\Phi}_{\boldsymbol{n}}$ can be estimated at the receiver if its coherence time is sufficient[2] or, otherwise, the receiver must assume that $\boldsymbol{\Phi}_{\boldsymbol{n}} = \boldsymbol{I}$.

The most common and practically appealing signaling strategies at the transmitter array satisfy

$$E[\boldsymbol{x}\boldsymbol{x}^\dagger] = \frac{E[\|\boldsymbol{x}\|^2]}{n_{\mathrm{T}}}\boldsymbol{I}. \tag{2}$$

When the transmitter has no knowledge of the channel realization, (2) is in fact required to achieve capacity for the canonical channel [2] and for other symmetric situations.

There are two distinct stages in the deployment of a wireless system.

- An early stage where widespread coverage is the main objective and, thus, the noise is basically thermal. In such conditions, the normalized covariance and density of the noise are simply given by

  $$\boldsymbol{\Phi}_{\boldsymbol{n}} = \boldsymbol{I}$$

  and

  $$N_0 = \gamma$$

  with γ the thermal spectral density per receive antenna.

- A mature stage, with high capacity as primordial goal, where the dominant form of noise is that of interference from adjacent cells. Unlike thermal noise, interference is spatially colored and subject to fading. The color tends to be particularly strong in the downlink, wherein the entire contribution received from another cell emanates from a

[1] As we shall see, when antennas with different polarizations are used the entries of $\boldsymbol{H}$ may not be identically distributed. For now, however, we may consider them to be identically distributed, in which case the power of each such entry is 1.

[2] In the downlink, the interferers (other base stations) are not moving and, thus, the coherence time of $\boldsymbol{\Phi}_{\boldsymbol{n}}$ is equal to that of $\boldsymbol{H}$. In the uplink, however, the coherence time of $\boldsymbol{\Phi}_{\boldsymbol{n}}$ is determined by the fastest contributor to the interference.

single localized source: its base station. We shall presume that, conditioned on its fading, this interference is still Gaussian.[3] More specifically, the aggregate noise adopts the form

$$\boldsymbol{n} = \sum_{\ell=1}^{L} \sqrt{\mathsf{g}_\ell}\, \boldsymbol{H}_\ell \boldsymbol{x}_\ell + \boldsymbol{n}_{\mathsf{th}} \tag{3}$$

with L the number of interferers, $\boldsymbol{x}_\ell$ the signal transmitted by the ℓth interferer, $\sqrt{\mathsf{g}_\ell}\,\boldsymbol{H}_\ell$ the channel from such interferer, and $\boldsymbol{n}_{\mathsf{th}}$ the underlying thermal noise, again with spectral density γ. The number of transmit antennas at the ℓth interferer is denoted by m_ℓ. Since the contribution of each interferer is subject to fading, the short-term covariance of $\boldsymbol{n}$ now depends on $\boldsymbol{H}_\ell$, $\ell = 1, \ldots, L$, which we assemble into $\overline{\boldsymbol{H}} = [\boldsymbol{H}_1 | \cdots | \boldsymbol{H}_L]$. With the various interferers mutually independent and signaling as in (2), the normalized covariance of (3) conditioned on $\overline{\boldsymbol{H}}$ becomes

$$\begin{aligned}\boldsymbol{\Phi}_{\boldsymbol{n}}\left(\overline{\boldsymbol{H}}\right) &= \frac{1}{N_0} E\left[\boldsymbol{n}\boldsymbol{n}^\dagger \middle| \overline{\boldsymbol{H}}\right] \\ &= \frac{1}{N_0}\left(\sum_{\ell=1}^{L} \mathsf{g}_\ell \frac{E[\|\boldsymbol{x}_\ell\|^2]}{m_\ell} \boldsymbol{H}_\ell \boldsymbol{H}_\ell^\dagger + \gamma \boldsymbol{I}\right). \end{aligned} \tag{4}$$

It is important to stress that the receiver does not have access to $\overline{\boldsymbol{H}}$, but only to $\boldsymbol{\Phi}_{\boldsymbol{n}}(\overline{\boldsymbol{H}})$. By conveniently defining

$$\mathcal{I}_\ell \triangleq \mathsf{g}_\ell \frac{E[\|\boldsymbol{x}_\ell\|^2]}{m_\ell} \frac{E[\mathrm{Tr}\{\boldsymbol{H}_\ell \boldsymbol{H}_\ell^\dagger\}]}{n_{\mathrm{R}}} \tag{5}$$

as the average energy per antenna received from the ℓth interferer, with expectation over both $\boldsymbol{x}_\ell$ and $\boldsymbol{H}_\ell$, we can write

$$N_0 = \sum_{\ell=1}^{L} \mathcal{I}_\ell + \gamma$$

noting that, in mature systems, γ tends to be small with respect to N_0.

Given that our noise is, in general, spatially colored, care must be exercised when computing the SNR. The noise power is not equal on every direction and thus we compute the SNR as the average of the signal-to-noise ratios along each of the principal directions of the noise space (rather than as the average signal power divided by the average noise power). The received SNR is thus found to be

$$\mathsf{SNR} = \mathsf{g} \frac{E[\|\boldsymbol{x}\|^2]}{N_0} \frac{E[\mathrm{Tr}\{\boldsymbol{H}\boldsymbol{H}^\dagger \boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}]}{n_{\mathrm{T}} n_{\mathrm{R}}} \tag{6}$$

with expectation over $\boldsymbol{x}$, $\boldsymbol{H}$, and $\overline{\boldsymbol{H}}$.[4] Note that, if $\boldsymbol{\Phi}_{\boldsymbol{n}} = \boldsymbol{I}$, (6) reduces to

$$\mathsf{SNR} = \mathsf{g} \frac{E[\|\boldsymbol{x}\|^2]}{N_0}.$$

[3]The use of multicode or multitone waveforms tends to render the interference Gaussian. Moreover, if the aggregate noise is not Gaussian, this assumption will yield a lower bound on the actual capacity.

[4]The SNR in (6) can also be interpreted as the average signal power divided by average noise power, both measured at the output of a noise whitening filter, that is, after the noise has been balanced across the various spatial dimensions.

The energy per bit required at a receiver operating at capacity satisfies[5]

$$\frac{E_b^r}{N_0} = \frac{\mathsf{SNR}}{C(\mathsf{SNR})}\, n_{\mathrm{R}}.$$

Although such received $\frac{E_b^r}{N_0}$ is of interest, the main metric throughout the paper is the *transmitted* $\frac{E_b}{N_0}$, given by

$$\begin{aligned}\frac{E_b}{N_0} &= \frac{E[\|\boldsymbol{x}\|^2]}{N_0 C(\mathsf{SNR})} \\ &= \frac{n_{\mathrm{T}} n_{\mathrm{R}}}{\mathsf{g} E[\mathrm{Tr}\{\boldsymbol{H}\boldsymbol{H}^\dagger \boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}]} \frac{\mathsf{SNR}}{C(\mathsf{SNR})}. \end{aligned} \tag{7}$$

To conclude this section, a final definition: given an $(n \times n)$ random matrix $\boldsymbol{A}$, we define its *dispersion* as

$$\zeta(\boldsymbol{A}) = n \frac{E[\mathrm{Tr}\{\boldsymbol{A}^2\}]}{E^2[\mathrm{Tr}\{\boldsymbol{A}\}]}. \tag{8}$$

Note that $\zeta(\boldsymbol{A})$ achieves its minimum value (equal to 1) when $\boldsymbol{A}$ is equal to the identity matrix. As we shall see, the dispersion serves to succinctly quantify how much the channel and noise depart from those in the canonical model.

B. Channel Model

The short-term fading encountered by wireless systems tends to be either Rayleigh or Ricean in nature and, thus, we model the entries of $\boldsymbol{H}$ as jointly Gaussian. With that, the characterization of $\boldsymbol{H}$ entails simply determining the mean and correlation between its entries. The main diversity mechanisms that impact such correlation are

- spatial diversity, based on antenna spacing;
- polarization diversity, based on the use of orthogonal polarizations.

Although polarization diversity is known to ensure low levels of correlation [32]–[38] while enabling more compact arrays, the number of orthogonal polarization states is very small [39] and thus polarization diversity remains an auxiliary mechanism to that of spatial diversity. We shall explore the impact of polarization diversity but, by and large, our focus is on spatial diversity.

1) Rayleigh Channel: In the Rayleigh case, the mean of the entries of $\boldsymbol{H}$ is zero. The correlation between those entries, in turn, may be represented—in its most general form—as a four-dimensional tensor. In most cases, though, it is possible to obtain such correlation starting from the local correlation between transmit antennas and the local correlation between receive antennas, separately.[6] This separation turns the problem of determining the correlation between entries of $\boldsymbol{H}$ into the much more conventional problem of determining the correlation between antennas within a given array.[7]

The correlation coefficients between the n_{T} transmit antennas can be assembled into an $(n_{\mathrm{T}} \times n_{\mathrm{T}})$ matrix $\boldsymbol{\Theta}_{\mathrm{T}}$ while

[5]The use of C and C avoids the abuse of notation of assigning the same symbol to capacity functions of SNR and $\frac{E_b}{N_0}$.

[6]This separation applies if the immediate surroundings to each array are responsible for the correlation between its antennas but have no impact on the correlation between the antennas at the other end of the link [40], [41].

[7]Correlation arises from exposure to the same multipath environment [42]–[45]. If the antennas are very tightly spaced, electromagnetic coupling also plays a role (it tends to reduce correlation [46], [47]).

the correlation coefficients between the n_{R} receive antennas can be assembled into a corresponding $(n_{\mathrm{R}} \times n_{\mathrm{R}})$ matrix $\boldsymbol{\Theta}_{\mathrm{R}}$. The diagonal elements of $\boldsymbol{\Theta}_{\mathrm{T}}$ and $\boldsymbol{\Theta}_{\mathrm{R}}$ are, by definition, equal to one. These matrices, deterministic and Hermitian positive semidefinite, can be used to generate properly correlated channels $\boldsymbol{H}$ via [16], [48], [49]

$$\boldsymbol{H} = \boldsymbol{\Theta}_{\mathrm{R}}^{1/2} \boldsymbol{W} \boldsymbol{\Theta}_{\mathrm{T}}^{1/2}$$

with $\boldsymbol{W}$ denoting, throughout the paper, a matrix with independent zero-mean unit-variance complex Gaussian random entries. The covariance matrix of each row of $\boldsymbol{H}$ is thus given by $\boldsymbol{\Theta}_{\mathrm{T}}$ while the covariance matrix of each column is given by $\boldsymbol{\Theta}_{\mathrm{R}}$. The validity of this correlation model for spatial diversity has been experimentally confirmed [50]–[56].

2) Cross-Polar Discrimination: The use of orthogonal polarizations creates asymmetries: the average power transfer between copolarized antennas differs from the average power transfer between cross-polarized antennas and, as a result, the entries of $\boldsymbol{H}$ become nonidentically distributed. These imbalances can be quantified through the cross-polar discrimination (XPD), denoted by $\mathcal{X} \in [0, 1]$. We can define an $(n_{\mathrm{R}} \times n_{\mathrm{T}})$ matrix $\boldsymbol{P}$ containing the power of each entry of $\boldsymbol{H}$ and write

$$\boldsymbol{H} = \boldsymbol{Q} \circ \left(\boldsymbol{\Theta}_{\mathrm{R}}^{1/2} \boldsymbol{W} \boldsymbol{\Theta}_{\mathrm{T}}^{1/2} \right)$$

with $\circ$ indicating Hadamard (element-wise) multiplication and with $\boldsymbol{Q}$ a matrix whose entries are the square roots of the entries of $\boldsymbol{P}$, i.e., $\boldsymbol{P} = \boldsymbol{Q} \circ \boldsymbol{Q}$. From the normalization imposed on $\boldsymbol{H}$, the sum of the entries of $\boldsymbol{P}$ is constrained to equal

$$\sum_{j=1}^{n_{\mathrm{T}}} \sum_{i=1}^{n_{\mathrm{R}}} (\boldsymbol{P})_{i,\,j} = n_{\mathrm{T}} n_{\mathrm{R}}. \tag{9}$$

For example, if $n_{\mathrm{T}} = n_{\mathrm{R}} = 2$

$$\boldsymbol{P} = \frac{2}{1+\mathcal{X}} \begin{bmatrix} 1 & \mathcal{X} \\ \mathcal{X} & 1 \end{bmatrix}$$

with $\mathcal{X}$ equal to 1 if the antennas at either end of the link are copolarized.

Notice that (9) puts channels with different power averages on an equal footage, i.e., any difference in average power is removed from $\boldsymbol{P}$ and $\boldsymbol{H}$ and absorbed into g.

3) Ricean Channel: The channel model can be easily made Ricean by incorporating an additional deterministic matrix $\boldsymbol{H}_0$ containing unit-magnitude entries [57], [21], [37]

$$\boldsymbol{H} = \boldsymbol{Q} \circ \left(\sqrt{\frac{1}{\mathsf{K}+1}}\, \boldsymbol{\Theta}_{\mathrm{R}}^{1/2} \boldsymbol{W} \boldsymbol{\Theta}_{\mathrm{T}}^{1/2} + \sqrt{\frac{\mathsf{K}}{\mathsf{K}+1}}\, \boldsymbol{H}_0 \right) \tag{10}$$

so that

$$E[\boldsymbol{H}] = \sqrt{\frac{\mathsf{K}}{\mathsf{K}+1}}\, \boldsymbol{H}_0$$

with the Ricean K-factor quantifying the ratio between the deterministic (cohesive) and the random (scattered) energies [58].[8] The term $\boldsymbol{H}_0$ is typically associated with a line of sight or a diffracted component and thus

$$\boldsymbol{H}_0 = \boldsymbol{a}_{\mathrm{R}} \boldsymbol{a}_{\mathrm{T}}^{\dagger}$$

[8]Although different XPD factors may apply to the deterministic and random terms [37], the model in (10) suffices for our purposes.

where the vectors $\boldsymbol{a}_{\mathrm{T}}$ and $\boldsymbol{a}_{\mathrm{R}}$ are the transmit and receive array responses to a plane wave satisfying

$$\|\boldsymbol{a}_{\mathrm{T}}\|^2 = n_{\mathrm{T}} \qquad \|\boldsymbol{a}_{\mathrm{R}}\|^2 = n_{\mathrm{R}}.$$

In an n-antenna uniform linear array, for instance

$$\boldsymbol{a} = \left[1,\, e^{j2\pi d \cos\theta_0},\, \ldots,\, e^{j2\pi d(n-1)\cos\theta_0}\right]^T$$

with d the spacing between adjacent antennas, in wavelengths, and θ_0 the angle between the array and the Ricean component.

4) Properties: Since a correlation matrix—either transmit or receive—has unit diagonal elements, its dispersion (8) particularizes to

$$\zeta(\boldsymbol{\Theta}) = \frac{\mathrm{Tr}\{\boldsymbol{\Theta}^2\}}{n}$$

which we shall refer to as the *correlation number*. For example, if $n = 2$

$$\zeta\left(\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) = 1 + \rho^2.$$

Some useful properties of such correlation number are as follows.

Property 1: The correlation number is bounded by

$$1 \le \zeta(\boldsymbol{\Theta}) \le n$$

with the lower bound achieved if and only if the antennas are uncorrelated and the upper bound achieved if and only if the antennas are fully correlated.

Property 2: If a correlation matrix is Toeplitz (as is the case when the array topology follows a uniform regular pattern), the (k, ℓ)th entry of $\boldsymbol{\Theta}$ equals $\rho(k-\ell)$ with $\rho(i)$ denoting the correlation coefficient between antennas whose indexes differ by i. The correlation number is then given by

$$\zeta(\boldsymbol{\Theta}) = 1 + \frac{2}{n} \sum_{i=1}^{n-1} (n-i) |\rho(i)|^2 \tag{11}$$

and the eigenvalues of $\boldsymbol{\Theta}$ become, as $n \to \infty$, samples of the discrete Fourier transform of any of its rows [59]. Furthermore, if the correlation decays faster than $1/i$ across the array, (11) converges to a finite quantity

$$\lim_{n\to\infty} \zeta(\boldsymbol{\Theta}) = 1 + 2 \sum_{i=1}^{\infty} |\rho(i)|^2 < \infty$$

whereas if the correlation decays no faster than $1/i$ (as in the classic Jakes model [42], for example) the above sum diverges.

Property 3: Let $\boldsymbol{\Theta}_{\mathrm{B}}$ be an $(n_{\mathrm{B}} \times n_{\mathrm{B}})$ Toeplitz correlation matrix and let $\boldsymbol{\Theta}_{\mathrm{A}}$ be an $(n_{\mathrm{A}} \times n_{\mathrm{A}})$ principal submatrix of $\boldsymbol{\Theta}_{\mathrm{B}}$. (These may be the correlation matrices of an array whose number of antennas has grown from n_{A} to n_{B} preserving its original pattern.) Then

$$\zeta(\boldsymbol{\Theta}_{\mathrm{B}}) \ge \zeta(\boldsymbol{\Theta}_{\mathrm{A}})$$

with equality only if $\rho(i) = 0\ \forall i$. In fact, strict inequality holds even if $\rho(i) = 0$ for $i \geq n_{\mathrm{A}}$. Hence, for a given antenna spacing and polarization, the correlation number increases with the number of antennas.

III. Capacity in the Low-Power Regime

The ergodic capacity, or more precisely, the maximal achievable rate per unit bandwidth under the constraint in (2), can be expressed as a function of the SNR via

$$C(\mathsf{SNR}) = E\left[\log_2 \det\left(\boldsymbol{I} + \mathsf{SNR}\frac{\boldsymbol{H}\boldsymbol{H}^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})}{\frac{1}{n_{\mathrm{R}}}E\left[\mathrm{Tr}\{\boldsymbol{H}\boldsymbol{H}^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}\right]}\right)\right] \tag{12}$$

where the expectations are over the distributions of $\boldsymbol{H}$ and $\overline{\boldsymbol{H}}$. Notice that our fading model is purposely frequency-flat in nature. In frequency-selective environments, the channel can always be decomposed into a number of parallel noninteracting subchannels, each experiencing frequency-flat fading and having the same ergodic capacity as the overall channel.

From $C(\mathsf{SNR})$, the capacity as a function of $\frac{E_b}{N_0}$ can be obtained through

$$\mathsf{C}\left(\frac{E_b}{N_0}\right) = C(\mathsf{SNR}) \tag{13}$$

with SNR the solution to (7). Unfortunately, an explicit expression for $\mathsf{C}(\frac{E_b}{N_0})$ cannot be obtained from (7), (12), and (13) except for scalar unfaded channels. Recall, however, that in the low-SNR regime its behavior can be captured through the measures $\frac{E_b}{N_0\,\min}$ and S_0 using (1), which in linear scale becomes

$$\mathsf{C}\left(\frac{E_b}{N_0}\right) \approx S_0 \log_2\left(\frac{\frac{E_b}{N_0}}{\frac{E_b}{N_0\,\min}}\right).$$

As shown in [28], $\frac{E_b}{N_0\,\min}$ and S_0 can be computed from the first and second derivatives of $C(\mathsf{SNR})$ at $\mathsf{SNR} = 0$ by means of

$$\frac{E_b}{N_0\,\min} = \frac{n_{\mathrm{T}} n_{\mathrm{R}}}{\mathsf{g} E[\mathrm{Tr}\{\boldsymbol{H}\boldsymbol{H}^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}]}\,\frac{1}{\dot{C}(0)}$$

and

$$S_0 = \frac{-2[\dot{C}(0)]^2}{\ddot{C}(0)}\,\log_e 2.$$

Note that the first-order approximation to $\mathsf{C}(\frac{E_b}{N_0})$ captures the second-order behavior in terms of $C(\mathsf{SNR})$. In fact, the first-order derivative of $C(\mathsf{SNR})$ is—by itself—unable to give any indication about the capacity other than the value of $\frac{E_b}{N_0\,\min}$.

It is also noteworthy to point out that, although the unique capacity-achieving distribution is Gaussian, the minimum energy per bit and optimum slope can be obtained with far simpler signaling. For example, a very practical and appealing choice is equal-power quaternary phase-shift keying (QPSK) on each antenna [28, Theorem 14].

Using the following properties [28] of the determinant of a square matrix $\boldsymbol{A}$:

$$\frac{\mathrm{d}}{\mathrm{d}u}\,\log_e \det[\boldsymbol{I} + u\boldsymbol{A}]|_{u=0} = \mathrm{Tr}\{\boldsymbol{A}\}$$
$$\frac{\mathrm{d}^2}{\mathrm{d}u^2}\,\log_e \det[\boldsymbol{I} + u\boldsymbol{A}]|_{u=0} = -\mathrm{Tr}\{\boldsymbol{A}^2\}$$

the $\frac{E_b}{N_0\,\min}$ can be found to be

$$\frac{E_b}{N_0\,\min} = \frac{\log_e 2}{\mathsf{g}}\,\frac{n_{\mathrm{T}}}{E[\mathrm{Tr}\{\boldsymbol{H}\boldsymbol{H}^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}]} \tag{14}$$

while the slope becomes

$$S_0 = \frac{2\,n_{\mathrm{R}}}{\zeta(\boldsymbol{H}\boldsymbol{H}^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}}))} \tag{15}$$

where we recall the definition of $\zeta(\cdot)$ as the dispersion of a matrix in (8). The received $\frac{E_b}{N_0\,\min}$, in turn, equals

$$\frac{E_b^r}{N_0\,\min} = \log_e 2$$

which is approximately -1.59 dB. This value, which represents a fundamental property of Gaussian noise [28, Theorem 1], is extremely robust.[9]

In the canonical case (copolarized antennas in a Rayleigh uncorrelated channel impaired by AWGN), (14) and (15) particularize to

$$\frac{E_b}{N_0\,\min} = \frac{\log_e 2}{\mathsf{g}}\,\frac{1}{n_{\mathrm{R}}}$$

and

$$S_0 = 2\,\frac{n_{\mathrm{T}} n_{\mathrm{R}}}{n_{\mathrm{T}} + n_{\mathrm{R}}} \tag{16}$$

from which we observe the following.

- The $\frac{E_b}{N_0\,\min}$ depends on n_{R}, but not on n_{T}. This is a direct consequence of the total transmit power being constrained while the total captured power increases with the number of receive antennas. Therefore, the number of transmit antennas is irrelevant in terms of $\frac{E_b}{N_0\,\min}$.
- The slope is symmetric with respect to n_{T} and n_{R}.
- The slope in the low-SNR regime is at least as large as the slope at high SNR [60], given by $\min(n_{\mathrm{T}}, n_{\mathrm{R}})$, with equality only if $n_{\mathrm{T}} = n_{\mathrm{R}}$.

As function of power, rate, and number of antennas, the required bandwidth in the canonical case is thus

$$B \approx \frac{R}{2}\left(\frac{1}{n_{\mathrm{T}}} + \frac{1}{n_{\mathrm{R}}}\right)\frac{3\text{ dB}}{\left.\frac{P\mathsf{g}n_{\mathrm{R}}}{RN_0}\right|_{\mathrm{dB}} + 1.59\text{ dB}}$$

with the term $(P\mathsf{g}n_{\mathrm{R}})$ representing the total per-symbol power captured by the receiver.

Fig. 1 reproduces the exact capacity along with its first-order expression in (1) for the canonical channel with $n_{\mathrm{T}} = n_{\mathrm{R}} = 1$

[9] It holds regardless of whether the channel is known at the receiver and/or the transmitter.

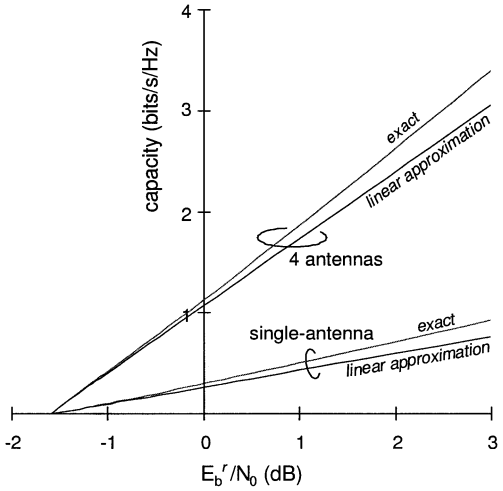

Fig. 1. Comparison between the exact capacity as a function of received energy per bit and its low-SNR first-order expression for single-antenna and four-antenna architectures.

TABLE I
$\frac{E_b^r}{N_0}$ BELOW WHICH FIRST-ORDER APPROXIMATION DEPARTS BY NO MORE THAN 10% FROM ACTUAL CAPACITY

antennas	E_b^r/N_0	locations
1	0.1 dB	24%
2	1.9 dB	60%
4	3 dB	72%
8	3.4 dB	76%

and $n_{\mathrm{T}} = n_{\mathrm{R}} = 4$. The first-order expression is tight for rather ambitious levels of $\frac{E_b}{N_0}$. Precisely, the levels below which the difference is less than 10% of the exact capacity are listed in Table I, parameterized by the number of transmit and receive antennas. Referring back to the emerging data systems described in [26], [27], Table I also shows that the first-order expression of a Rayleigh channel is over 90% accurate for a very large fraction of cell locations, increasingly so as the number of antennas grows. More importantly, these are precisely the locations wherein multiple-antenna techniques are more likely to be relevant.

IV. SPATIAL DIVERSITY IN AWGN

When the noise is white and unfaded, either because it is exclusively of thermal origin or else because of the presence of a large number of comparable-strength interferers,[10] we have $\boldsymbol{\Phi_n} = \boldsymbol{I}$. For copolarized antennas, we present the following central result.

[10]Also, if the noise is colored but its covariance cannot be estimated by the receiver, the capacity is lower-bounded by what it would be if the noise were white [61], [62].

Proposition 1: Consider a correlated Rayleigh/Ricean channel known by the receiver and given by

$$\boldsymbol{H} = \sqrt{\frac{1}{\mathsf{K}+1}}\,\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}^{1/2} + \sqrt{\frac{\mathsf{K}}{\mathsf{K}+1}}\,\boldsymbol{a}_{\mathrm{R}}\boldsymbol{a}_{\mathrm{T}}^{\dagger}.$$

In AWGN, the $\frac{E_b}{N_0}_{\min}$ is

$$\frac{E_b}{N_{0\,\min}} = \frac{\log_e 2}{\mathsf{g}}\,\frac{1}{n_{\mathrm{R}}} \tag{17}$$

while S_0 is as shown in (18) at the bottom of the page.

Proof: See Appendix B.

Remarkably, the $\frac{E_b}{N_0}_{\min}$ is unaffected by the existence of antenna correlation and a Ricean term and, consequently, a first-order analysis of $C(\mathsf{SNR})$ fails to reveal their impact. Only S_0 reflects the structure of the channel. In the remainder of this section, therefore, we study this slope in detail. Prior to that, however, we point out that the transmitter and the receiver play symmetric roles in (18) while (17) is inversely proportional to n_{R} and thus we have.

Corollary 1: If the capacity of a link is $\mathsf{C}(\frac{E_b}{N_0})$, that of the reverse link—up to first order—is $\mathsf{C}(\frac{n_{\mathrm{R}}}{n_{\mathrm{T}}}\frac{E_b}{N_0})$.

A. Rayleigh Fading

For Rayleigh fading, $\mathsf{K} = 0$ and the slope particularizes to

$$S_0 = \frac{2n_{\mathrm{T}}n_{\mathrm{R}}}{n_{\mathrm{T}}\zeta(\boldsymbol{\Theta}_{\mathrm{R}}) + n_{\mathrm{R}}\zeta(\boldsymbol{\Theta}_{\mathrm{T}})} \tag{19}$$

where the effects of transmit and receive antenna correlation appear only through the correlation numbers of the transmit and receive arrays. This is a powerful result for it indicates that a single scalar parameter uniquely quantifies the capacity impact of an entire correlation matrix. Any two matrices mapping onto the same correlation number are equivalent in terms of low-SNR capacity.

Using channel Property 1, it is straightforward to see that the slope in (19) satisfies

$$1 \le S_0 \le \frac{2n_{\mathrm{T}}n_{\mathrm{R}}}{n_{\mathrm{T}}+n_{\mathrm{R}}}$$

confirming the intuitive result that antenna correlation can only diminish the capacity. Although a first-order analysis of $C(\mathsf{SNR})$ indicates that the low-SNR capacity is unaffected by correlation [19], this is only true as far as the invariance of $\frac{E_b}{N_0}_{\min}$ is concerned. At any $\mathsf{SNR} > 0$, antenna correlation does reduce capacity, but such reduction is not revealed by a first-order analysis.

$$S_0 = \frac{2n_{\mathrm{T}}n_{\mathrm{R}}(\mathsf{K}+1)^2}{n_{\mathrm{T}}\zeta(\boldsymbol{\Theta}_{\mathrm{R}}) + n_{\mathrm{R}}\zeta(\boldsymbol{\Theta}_{\mathrm{T}}) + \mathsf{K}^2 n_{\mathrm{T}}n_{\mathrm{R}} + 2\mathsf{K}\left(n_{\mathrm{T}}\frac{\mathrm{Tr}\{\boldsymbol{a}_{\mathrm{R}}\boldsymbol{a}_{\mathrm{R}}^{\dagger}\boldsymbol{\Theta}_{\mathrm{R}}\}}{n_{\mathrm{R}}} + n_{\mathrm{R}}\frac{\mathrm{Tr}\{\boldsymbol{a}_{\mathrm{T}}\boldsymbol{a}_{\mathrm{T}}^{\dagger}\boldsymbol{\Theta}_{\mathrm{T}}\}}{n_{\mathrm{T}}}\right)}. \tag{18}$$

Corollary 2: The low-SNR slope is equivalent to that of $n_{\mathrm{T}}^{\mathrm{eq}}$ and $n_{\mathrm{R}}^{\mathrm{eq}}$ uncorrelated antennas given by

$$n_{\mathrm{T}}^{\mathrm{eq}} = \frac{n_{\mathrm{T}}}{\zeta(\mathbf{\Theta}_{\mathrm{T}})} \qquad n_{\mathrm{R}}^{\mathrm{eq}} = \frac{n_{\mathrm{R}}}{\zeta(\mathbf{\Theta}_{\mathrm{R}})}.$$

Hence, the reciprocals of the transmit and receive correlation numbers determine the fraction of n_{T} and n_{R} that would result in the same slope if they were uncorrelated.

Fixing rate and power, the bandwidth B required with correlation matrices $\mathbf{\Theta}_{\mathrm{T}}$ and $\mathbf{\Theta}_{\mathrm{R}}$ relative to the canonical bandwidth B_{can} required in the absence of correlation is

$$\frac{B}{B_{\mathsf{can}}} = \frac{n_{\mathrm{T}}\zeta(\mathbf{\Theta}_{\mathrm{R}}) + n_{\mathrm{R}}\zeta(\mathbf{\Theta}_{\mathrm{T}})}{n_{\mathrm{T}} + n_{\mathrm{R}}}. \tag{20}$$

Clearly, the impact of correlation in one of the arrays is diminished if the other array has fewer antennas, in which case the latter is already constraining the capacity. Conversely, if we fix the rate and the bandwidth, the power penalty that results from correlation is

$$\frac{P|_{\mathrm{dB}} - P_{\mathsf{can}}|_{\mathrm{dB}}}{3\ \mathrm{dB}} = \frac{R}{2B}\left(\frac{\zeta(\mathbf{\Theta}_{\mathrm{T}}) - 1}{n_{\mathrm{T}}} + \frac{\zeta(\mathbf{\Theta}_{\mathrm{R}}) - 1}{n_{\mathrm{R}}}\right) \tag{21}$$

which, in contrast with (20), depends on the operating point (R/B).

Example 1: Consider $n_{\mathrm{T}} = 2$ and $n_{\mathrm{R}} = 1$. The low-SNR capacity in a Rayleigh channel is

$$\mathsf{C}\left(\frac{E_b}{N_0}\right) \approx \frac{4}{3 + |\rho_{\mathrm{T}}|^2}\ \log_2\left(\frac{\mathsf{g}}{\log_e 2}\,\frac{E_b}{N_0}\right)$$

with ρ_{T} denoting the correlation between both transmit antennas. Remarkably, antenna correlation has a rather limited impact in this case: with full—unbeknownst to the transmitter—correlation, 75% of the canonical capacity can nevertheless be attained at any given $\frac{E_b}{N_0}$. The bandwidth expansion factor incurred because of correlation, in turn, is

$$\frac{B}{B_{\mathsf{can}}} = 1 + \frac{|\rho_{\mathrm{T}}|^2}{3}.$$

If the correlation matrices are Toeplitz, Property 2 indicates that, with rate R and power P, the bandwidth expansion factor due to correlation converges, as the number of antennas grows, to

$$\lim_{n_{\mathrm{R}}\to\infty} \frac{B}{B_{\mathsf{can}}} = 1 + \frac{2}{n_{\mathrm{T}} + n_{\mathrm{R}}} \sum_{i=1}^{\infty} \left(n_{\mathrm{T}}|\rho_{\mathrm{R}}(i)|^2 + n_{\mathrm{R}}|\rho_{\mathrm{T}}(i)|^2\right) \tag{22}$$

with $\rho_{\mathrm{T}}(i)$ and $\rho_{\mathrm{R}}(i)$ the transmit and receive correlation between antennas whose indexes differ by i. If correlation at either array does not decay faster than $1/i$, (22) diverges and thus the intended rate R cannot be achieved with power P. Otherwise, (22) provides a limiting value for the bandwidth expansion factor, a value that is approached as the number of antennas grows large.

Further insight on the capacity of correlated Rayleigh channels impaired by AWGN can be gathered from the slope given by (19) and the fact that the $\frac{E_b}{N_0}_{\min}$ is unaffected.

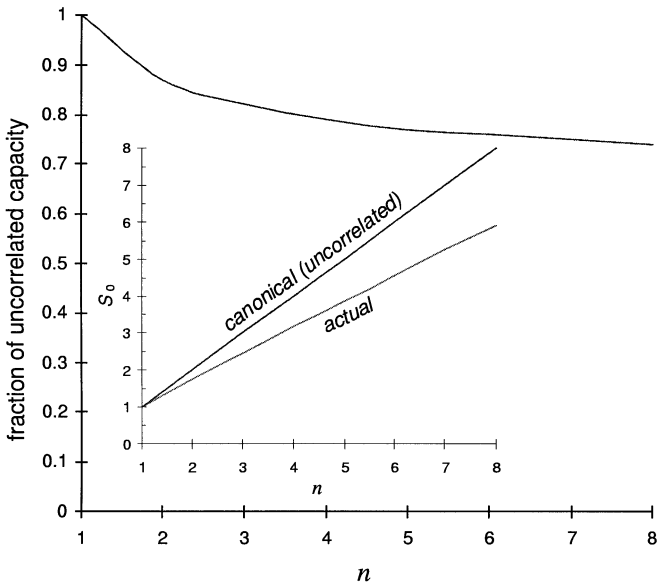

Fig. 2. Fraction of canonical capacity achievable with uniform linear arrays as function of the number of antennas ($n_{\mathrm{T}} = n_{\mathrm{R}} = n$). Antenna spacing is 4 and 0.5 wavelengths at base station and terminal, respectively. Angular spectrum is Gaussian with root-mean-square spread 2° at the base and uniform over 360° at the terminal.

Corollary 3: If $n_{\mathrm{T}} = n_{\mathrm{R}} = n$, the slope particularizes to

$$S_0 = \frac{2n}{\zeta(\mathbf{\Theta}_{\mathrm{R}}) + \zeta(\mathbf{\Theta}_{\mathrm{T}})}.$$

This result would seem to indicate—as found asymptotically in [19]—that the capacity still scales linearly with the number of antennas, as in the canonical channel, but with a reduced slope. However, since $\zeta(\mathbf{\Theta}_{\mathrm{T}})$ and $\zeta(\mathbf{\Theta}_{\mathrm{R}})$ are themselves functions of the number of antennas, in the presence of antenna correlation the capacity is, in general, no longer linear on n. Only for $n \to \infty$ may the capacity with correlation grow linearly with n. In the specific case of Toeplitz correlation matrices, this asymptotic linear scaling requires that (22) be finite [19]. For such Toeplitz matrices, we can formalize the nonasymptotic behavior as follows.

Corollary 4: If the correlation matrices are Toeplitz, the bandwidth expansion factor increases monotonically as the number of antennas grows with the antenna spacing preserved.

Hence, even though the canonical capacity increases linearly with the number of antennas, the penalty associated with antenna correlation compounds and thus only a diminishing fraction of such capacity can be attained. If (22) is finite, this achievable fraction of the canonical capacity approaches a nonzero asymptote as the number of antennas is driven to infinity [19].

Example 2: Let us consider a base station and a terminal equipped with uniform linear arrays with antenna spacing equal to 4 and 0.5 wavelenths, respectively. Consider a broadside Gaussian power spectrum at the base with a 2° root-mean-square angular spread and a 360° uniform spectrum at the terminal. As shown in Fig. 2, the attainable fraction of canonical capacity decreases from 100% with $n = 1$ down to 74% with $n = 8$. The comparison between the canonical and

actual capacity slopes is displayed in the inset. As anticipated by the analysis, the latter is sublinear in the number of antennas.

Experimental data [63], [64] confirms that this progressive decrease in slope occurs even with non-Toeplitz correlation matrices and beyond the low-SNR regime.

B. Ricean Fading

In Ricean conditions, the presence of a deterministic component compounds the effects of antenna correlation. Since the impact of correlation was studied in the previous section, we now evaluate the impact of this deterministic component by setting $\boldsymbol{\Theta}_{\mathrm{T}} = \boldsymbol{I}$ and $\boldsymbol{\Theta}_{\mathrm{R}} = \boldsymbol{I}$. According to Proposition 1, the low-SNR slope becomes

$$S_0 = \frac{2(\mathsf{K}+1)^2}{\mathsf{K}^2 + (2\mathsf{K}+1)\frac{n_{\mathrm{T}}+n_{\mathrm{R}}}{n_{\mathrm{T}}n_{\mathrm{R}}}}. \tag{23}$$

It is interesting to note that there is no dependence on either $\boldsymbol{a}_{\mathrm{T}}$ or $\boldsymbol{a}_{\mathrm{R}}$, that is, no dependence on either the geometry of the arrays or the direction of departure or arrival of the Ricean component. This is not the case in general: if $\boldsymbol{\Theta}_{\mathrm{T}}$ and $\boldsymbol{\Theta}_{\mathrm{R}}$ differ from the identity, their correlations may favor some directions differently.

For large K, we observe that

$$\lim_{\mathsf{K}\to\infty} S_0 = 2 \tag{24}$$

which is the low-SNR slope of a single-antenna unfaded channel. Together with (17), we can conclude that, in the low-SNR regime and in the presence of a strong Ricean component unknown to the transmitter, multiple transmit antennas are irrelevant and multiple receive antennas are only relevant in terms of $\frac{E_b}{N_0\,\min}$, but not in terms of slope. It is interesting to note—from (16) and (24)—that, if (and only if) one or both of the arrays has a single antenna, the capacity is higher in strong Ricean channels than in uncorrelated Rayleigh conditions, even though the transmitter is unaware of the Ricean statistics.

It can also be checked from (23) that, for $n_{\mathrm{T}} = n_{\mathrm{R}} = 2$, the slope equals $S_0 = 2$ irrespective of K. Hence we have the following.

Corollary 5: With two uncorrelated antennas at both transmitter and receiver

$$\mathsf{C}\left(\frac{E_b}{N_0}\right) \approx 2\log_2\left(\frac{2\mathsf{g}}{\log_e 2}\,\frac{E_b}{N_0}\right)$$

regardless of the possible presence of an unknown Ricean component.

V. Polarization Diversity

Before moving onto more elaborate types of noise, we briefly study the AWGN-limited capacity with polarization diversity and contrast it with what we learned for spatial diversity. Since, typically, two orthogonal polarizations are activated, we restrict our analysis to the case of two-antenna architectures. Readers interested in a more general characterization encompassing combinations of polarization and space diversity are referred to [65]. Straightforward application of (14) and (15) results in the following proposition.

Proposition 2: Consider two transmit and two receive antennas in a Rayleigh channel, known by the receiver, given by

$$\boldsymbol{H} = \boldsymbol{Q} \circ \left(\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}^{1/2}\right)$$

with

$$\boldsymbol{P} = \frac{2}{1+\mathcal{X}}\begin{bmatrix} 1 & \mathcal{X} \\ \mathcal{X} & 1 \end{bmatrix} \tag{25}$$

where $\mathcal{X}$ is the XPD. In AWGN, the $\frac{E_b}{N_0\,\min}$ and the slope therein in bits/s/Hz/(3 dB) are

$$\frac{E_b}{N_0\,\min} = \frac{\log_e 2}{\mathsf{g}}\,\frac{1}{2}$$

and

$$S_0 = \frac{2}{1 + \frac{2\mathcal{X}}{(1+\mathcal{X})^2}(|\rho_{\mathrm{T}}|^2 + |\rho_{\mathrm{R}}|^2)}$$

with ρ_{T} and ρ_{R} the transmit and receive correlation factors.

The following remarks can be made.

- Conditioned on g, the $\frac{E_b}{N_0\,\min}$ is unaffected.[11] Thus, in terms of $\frac{E_b}{N_0\,\min}$, the impact of using distinct polarizations can come only through a difference in average channel gain. Any decrease in the average power captured by the receive antennas translates directly onto an increase in $\frac{E_b}{N_0\,\min}$ and *vice versa*. With reference to all antennas being copolarized, the configuration in (25) with $\mathcal{X} \in [0, 1]$ may cause an increase of up to 3 dB in $\frac{E_b}{N_0\,\min}$. In many applications, however, the transmitter and/or receiver may be subject to a random orientation and thus spatial diversity configurations may also suffer a polarization power loss with respect to this reference [66]. Affecting all antennas at the same time, this loss may easily shift $\frac{E_b}{N_0\,\min}$ by more than 3 dB, in which case, polarization diversity appears preferable for it bounds the increase in $\frac{E_b}{N_0\,\min}$ while ensuring low correlation and, thus, a good slope. If, instead of thermal, the noise is dominated by interference, any power loss caused by XPD might apply also to the interference thereby reducing N_0 similarly to the desired signal. As a result, the low-correlation benefits of polarization diversity may be reaped without the associated power penalty, i.e., a good slope can be secured without an increase in $\frac{E_b}{N_0\,\min}$.
- Since orthogonally polarized antennas tend to be loosely correlated, the impact of $\mathcal{X}$ on the slope is minor. In fact, if the antennas are fully uncorrelated, then S_0 does not depend on $\mathcal{X}$. As shown in [65, Proposition 3], this reflects a more general property of certain channels with independent nonidentically distributed entries.

VI. Spatial Diversity in the Presence of Interference

When the noise contains out-of-cell interference, the covariance $\boldsymbol{\Phi}_{\boldsymbol{n}}$ comes into play and the expressions in Proposition 1 can be generalized. Recall that $\boldsymbol{\Phi}_{\boldsymbol{n}}$ is known to the receiver, but not to the transmitter. Recall also, from (4), that in the presence of L out-of-cell interferers each equipped with m_ℓ antennas,

[11] This finding in fact holds for arbitrary n_{T} and n_{R}.

$\ell = 1, \ldots, L$, the normalized conditional covariance of the noise is given by

$$\boldsymbol{\Phi_n}\left(\overline{\boldsymbol{H}}\right) = \frac{1}{N_0}\left(\sum_{\ell=1}^{L} \mathcal{I}_\ell \frac{\boldsymbol{H}_\ell \boldsymbol{H}_\ell^\dagger}{m_\ell} + \gamma \boldsymbol{I}\right) \tag{26}$$

with

$$N_0 = \sum_{\ell=1}^{L} \mathcal{I}_\ell + \gamma$$

so that

$$E[\mathrm{Tr}\{\boldsymbol{\Phi_n}\}] = n_{\mathrm{R}}. \tag{27}$$

Let us define the total number of interfering antennas as

$$n_{\mathrm{L}} \triangleq \sum_{\ell=1}^{L} m_\ell.$$

Concentrating on spatial diversity, we shall separately consider—as in the AWGN case—the instances where the channel (for the user of interest) is Rayleigh and Ricean.

A. Rayleigh Fading

Proposition 3: Consider a correlated Rayleigh channel known by the receiver and given by

$$\boldsymbol{H} = \boldsymbol{\Theta}_{\mathrm{R}}^{1/2} \boldsymbol{W} \boldsymbol{\Theta}_{\mathrm{T}}^{1/2}.$$

If the normalized conditional covariance of the noise is $\boldsymbol{\Phi_n}(\overline{\boldsymbol{H}})$, then

$$\frac{E_b}{N_{0\,\min}} = \frac{\log_e 2}{\mathsf{g}} \frac{1}{E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}} \boldsymbol{\Phi_n}^{-1}(\overline{\boldsymbol{H}})\}]} \tag{28}$$

and

$$S_0 = \frac{2 n_{\mathrm{T}} n_{\mathrm{R}}}{n_{\mathrm{T}} \zeta(\boldsymbol{\Theta}_{\mathrm{R}} \boldsymbol{\Phi_n}^{-1}(\overline{\boldsymbol{H}})) + n_{\mathrm{R}} \zeta(\boldsymbol{\Theta}_{\mathrm{T}}) \varphi(\boldsymbol{\Theta}_{\mathrm{R}} \boldsymbol{\Phi_n}^{-1}(\overline{\boldsymbol{H}}))} \tag{29}$$

with expectation over the fading of the interferers, $\overline{\boldsymbol{H}}$, and with the function

$$\varphi(\boldsymbol{A}) \triangleq \frac{E[\mathrm{Tr}^2\{\boldsymbol{A}\}]}{E^2[\mathrm{Tr}\{\boldsymbol{A}\}]}.$$

Proof: See Appendix B.

Using (27), it is easy to prove via Jensen's inequality that

$$E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}} \boldsymbol{\Phi_n}^{-1}\}] \geq n_{\mathrm{R}}$$

and, therefore, (28) is upper-bounded by the value it takes when $\boldsymbol{\Phi_n} = \boldsymbol{I}$. Hence we have the following.

Corollary 6: Noise fading and color diminish the minimum energy necessary to convey information reliably.

Proposition 3 merits some additional observations.

- The function $\varphi(\boldsymbol{\Theta}_{\mathrm{R}} \boldsymbol{\Phi_n}^{-1}(\overline{\boldsymbol{H}}))$ depends only on the noise fading, not on its color. If $\overline{\boldsymbol{H}}$ is deterministic, that is, if the interferers are unfaded, we have

  $$\varphi\left(\boldsymbol{\Theta}_{\mathrm{R}} \boldsymbol{\Phi_n}^{-1}\left(\overline{\boldsymbol{H}}\right)\right) = 1$$

  and $\frac{E_b}{N_{0\,\min}}$ and S_0 resemble those found for AWGN with $\boldsymbol{\Theta}_{\mathrm{R}}$ replaced by $\boldsymbol{\Theta}_{\mathrm{R}} \boldsymbol{\Phi_n}^{-1}(\overline{\boldsymbol{H}})$. This, of course, reflects the fact that the receiver can whiten the noise while correlating the received signal. The noise color appears then through the expected trace of $\boldsymbol{\Theta}_{\mathrm{R}} \boldsymbol{\Phi_n}^{-1}(\overline{\boldsymbol{H}})$ and its dispersion, which generalizes the receive correlation number encountered in AWGN.
- When the interferers fade, an additional mechanism is at play, one that was not present in AWGN. In a single-antenna scenario, it is easy to see from Jensen's inequality that noise fading can only improve the capacity

  $$E\left[\log_2\left(1 + \mathsf{g}\,\frac{E[|x|^2]}{\Phi_n N_0}\right)\right] > \log_2\left(1 + \mathsf{g}\,\frac{E[|x|^2]}{N_0}\right) \tag{30}$$

  where Φ_n is a random scalar representing the noise variance, normalized so that $E[\Phi_n] = 1$, and the outer expectation in the left-hand side of (30) is with respect to it. Nevertheless, as n_{L} and/or n_{R} grow, the impact of such fading subsides rapidly [67] and only the noise color remains relevant.
- For $\frac{n_{\mathrm{L}}}{n_{\mathrm{R}}} \to \infty$, the capacity behaves exactly as in AWGN-limited conditions. As its dimensionality grows, the interference appears white and unfaded to the receiver. Conversely, if n_{L} and n_{R} grow but their ratio does not, the noise color remains.

To proceed any further, we need to flesh out the structure of $\boldsymbol{\Phi_n}$, which is governed by the relative strength of the various interferers, $\mathcal{I}_\ell$, their fading, $\boldsymbol{H}_\ell$, their number of antennas, m_ℓ, and the relative weight of the underlying thermal noise, γ. We therefore make some basic considerations.

- Since we have already studied the AWGN-limited capacity in detail, in the remainder we seek new insight by concentrating on the converse scenario: interference-limited conditions, that is,

  $$\frac{1}{\gamma}\sum \mathcal{I}_\ell \to \infty.$$

- Given that the interferers are located in neighboring cells, their K-factor is usually negligible and it is thus reasonable to model each individual $\boldsymbol{H}_\ell$ as Rayleigh.
- In terms of transmit antenna correlation at each interferer, we postulate two limiting cases.
  - — Uncorrelated. In this case, the angular spread spanned at each of the interferers is assumed sufficiently large and, thus, all degrees of freedom therein are active.
  - — Fully correlated. This case models situations where the transmit antennas at each interferer are tightly correlated because of small angular spread, which is geometrically reasonable for distant interferers [68]. Consequently, each individual interferer contributes a single degree of freedom.

  In the general case of partial transmit correlation, these limiting cases serve as bounds.
- The receive antenna correlation experienced by each interference term is allowed to be arbitrary. The analysis,

however, simplifies if such correlations equal the receive correlation suffered by the desired signal.

1) Single Interferer: In many instances, particularly in strong shadow conditions, most of the interference may be contributed by a single neighbor. In the case of a single interferer, (26) becomes

$$\boldsymbol{\Phi}_{\boldsymbol{n}}(\boldsymbol{H}_1) = \frac{\mathcal{I}_1}{\mathcal{I}_1+\gamma}\,\frac{\boldsymbol{H}_1\boldsymbol{H}_1^{\dagger}}{m_1} + \frac{\gamma}{\mathcal{I}_1+\gamma}\,\boldsymbol{I}.$$

Let us scrutinize $\frac{E_b}{N_0}_{\min}$ and S_0 in the presence of a single interferer starting with the case where such interferer is uncorrelated.

Proposition 4: Consider the scenario of Proposition 3 with a single interferer transmitting from m_1 uncorrelated antennas. Let the fading of such interferer be Rayleigh with some arbitrary correlation $\boldsymbol{\Theta}_1$ at the receiver, that is,

$$\boldsymbol{H}_1 = \boldsymbol{\Theta}_1^{1/2}\boldsymbol{W}_1$$

with $\boldsymbol{W}_1$ distributed as $\boldsymbol{W}$. The interference-limited $\frac{E_b}{N_0}_{\min}$ satisfies

$$\lim_{(\mathcal{I}_1/\gamma)\to\infty} \frac{E_b}{N_0}_{\min} = \frac{\log_e 2}{\mathsf{g}}\,\frac{n_{\mathrm{R}}}{\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1}\}}\left[\frac{1}{n_{\mathrm{R}}} - \frac{1}{m_1}\right]^{+} \tag{31}$$

with $[x]^+ = x$ if $x \geq 0$ and $[x]^+ = 0$ otherwise. The corresponding slope is given by

$$\lim_{(\mathcal{I}_1/\gamma)\to\infty} S_0 = 2\left[\left(\frac{\Psi}{n_{\mathrm{R}}} + \frac{\zeta(\boldsymbol{\Theta}_{\mathrm{T}})\Upsilon}{n_{\mathrm{T}}}\right)\frac{n_{\mathrm{R}}^2 - \zeta(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1})}{n_{\mathrm{R}}^2-1} + \frac{\zeta(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1})-1}{n_{\mathrm{R}}^2-1}\left(n_{\mathrm{R}}\Upsilon + \frac{\zeta(\boldsymbol{\Theta}_{\mathrm{T}})\Psi}{n_{\mathrm{T}}}\right)\right]^{-1} \tag{32}$$

where

$$\Psi = \begin{cases} \dfrac{m_1(m_1-n_{\mathrm{R}})}{(m_1-n_{\mathrm{R}})^2-1}, & m_1 > n_{\mathrm{R}}+1 \\ \dfrac{n_{\mathrm{R}}}{n_{\mathrm{R}}-m_1}, & m_1 \leq n_{\mathrm{R}} \end{cases} \tag{33}$$

$$\Upsilon = \begin{cases} \dfrac{\Psi}{n_{\mathrm{R}}} + \dfrac{(n_{\mathrm{R}}-1)(m_1-n_{\mathrm{R}})}{n_{\mathrm{R}}(m_1-n_{\mathrm{R}}+1)}, & m_1 > n_{\mathrm{R}}+1 \\ 1, & m_1 \leq n_{\mathrm{R}}. \end{cases} \tag{34}$$

Proof: The key to this proposition is the closed-form evaluation of

$$E[\mathrm{Tr}\{(\boldsymbol{W}_1\boldsymbol{W}_1^{\dagger})^{-1}\}]$$
$$E[\mathrm{Tr}\{(\boldsymbol{W}_1\boldsymbol{W}_1^{\dagger})^{-2}\}]$$

and

$$E[\mathrm{Tr}^2\{(\boldsymbol{W}_1\boldsymbol{W}_1^{\dagger})^{-1}\}].$$

For these evaluations and the complete proof, see Appendix C.

This result reveals that, in the presence of a dominant interferer, the following holds.

- The receive antenna correlations experienced by signal and interference, respectively $\boldsymbol{\Theta}_{\mathrm{R}}$ and $\boldsymbol{\Theta}_1$, compound into an *effective* correlation given by $\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1}$. It is thus equivalent to have the signal experience such effective correlation while the interference experiences none. If the receiver correlations experienced by signal and interference coincide, the net result is no correlation for either.
- For $m_1 < n_{\mathrm{R}}$, the $\frac{E_b}{N_0}_{\min}$ vanishes as the interference-to-thermal ratio grows. This is clearly a direct consequence of the fact that interference with dimensionality inferior to the number of receive antennas can be completely suppressed, through mere linear processing, if only $\boldsymbol{\Phi}_{\boldsymbol{n}}$ is known [22], [69]. Such interference-limited conditions enable reliable communication at negligible levels of normalized energy. More precisely, when noise and interference experience identical corerlation at the receiver

$$\lim_{(\mathcal{I}_1/\gamma)\to\infty} \frac{E_b}{\gamma} = \frac{\log_e 2}{\mathsf{g}}\,\frac{1}{n_{\mathrm{R}}}$$

and thus, as $\frac{\mathcal{I}_1}{\gamma}$ grows, reliable communiation becomes feasible if only the received energy per bit is -1.59 dB above the thermal noise, irrespective of the strength of the interference. The corresponding slope, in turn, is

$$\lim_{(\mathcal{I}_1/\gamma)\to\infty} S_0 = \frac{2n_{\mathrm{T}}(n_{\mathrm{R}}-m_1)}{n_{\mathrm{T}}+\zeta(\boldsymbol{\Theta}_{\mathrm{T}})(n_{\mathrm{R}}-m_1)}$$

which, contrasted with (19), indicates that the equivalent of m_1 receive antennas have been spent suppressing the m_1-dimensional interference and, thus, the slope equals that which would be experienced with the remaining $(n_{\mathrm{R}}-m_1)$ receive antennas and the n_{T} transmit antennas in the presence of only the underlying thermal noise, γ.
- When $m_1 = n_{\mathrm{R}}$, both the $\frac{E_b}{N_0}_{\min}$ and S_0 vanish as the interference-to-thermal ratio grows without bound. In this case, however, the slope also approaching zero as the $\frac{E_b}{N_0}_{\min}$ decreases warns that nonnegligible capacity will require nonnegligible energy per bit.

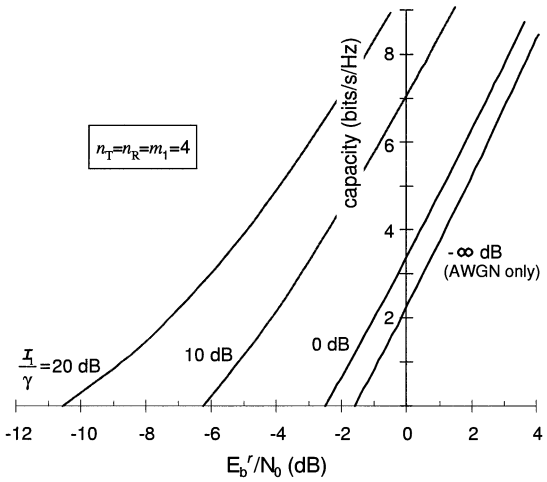

Fig. 3. Capacity versus $\frac{E_b^r}{N_0}$ in the presence of a single interferer plus thermal AWGN, with $n_{\mathrm{T}} = n_{\mathrm{R}} = m_1 = 4$, parameterized by the interference-to-thermal ratio.

Example 3: Displayed in Fig. 3 is the exact capacity with $n_{\mathrm{T}} = n_{\mathrm{R}} = 4$ in the presence of a four-antenna interferer $(m_1 = 4)$ plus thermal AWGN, with the various curves corresponding to different interference-to-thermal ratios. Both the desired user and the interferer experience Rayleigh fading and

the same antenna correlation at the receiver. The transmit antennas are uncorrelated. As the interference grows larger than the thermal AWGN, the capacity shows a clear displacement toward lower levels of $\frac{E_b}{N_0}$. The slope at $\frac{E_b}{N_0}_{\min}$, in turn, diminishes progressively and would approach zero asymptotically.

Clearly, the constraint that the transmit antennas at the interferer be uncorrelated maximizes the number of degrees of freedom within $\boldsymbol{\Phi_n}$. Any correlation therein can only enhance the noise color and thus the capacity. As shown by the next result, in the limit of full correlation, the interference exhibits a single degree of freedom and, therefore, it is as if the interferer were equipped with a single transmit antenna.

Proposition 5: Consider the same scenario of Proposition 4 but with the m_1 transmit antennas at the interferer fully correlated, that is,

$$\boldsymbol{H}_1 = \boldsymbol{\Theta}_1^{1/2} \boldsymbol{W}_1 \mathbf{1}^{1/2}$$

where $\mathbf{1}$ is an $(m_1 \times m_1)$ matrix whose entries are all unity. The interference-limited values for $\frac{E_b}{N_0}_{\min}$ and S_0 are given by their expressions in Proposition 4 with $m_1 = 1$.

Proof: See Appendix D.

More generally, partial transmit correlation at the interferer results in an *equivalent* number of virtual transmit antennas therein, uncorrelated and with possibly different powers. The number of such virtual antennas (degrees of freedom within $\boldsymbol{\Phi_n}$), ranging between 1 and m_1, equals the number of nonzero eigenvalues of said transmit correlation. Their powers, in turn, are determined by those nonzero eigenvalues.

2) Multiple Interferers: In terms of analysis, we restrict ourselves to scenarios where $\mathcal{I}_\ell = \mathcal{I}_1$, $\ell = 2, \ldots, L$. These scenarios boil down to a scaled version of the single-interferer case above. Specifically, the capacity in the presence of L equal-power interferers experiencing identical receive antenna correlation equals the capacity in the presence of a single *equivalent* interferer with that same receive correlation. The energy of this equivalent interferer is equal to the sum of energies of the L interferers and, if the transmit antennas within each of the L actual interferers are uncorrelated (resp., fully correlated), the equivalent interferer has n_{L} (resp., L) uncorrelated antennas.

Reference [64] considers the more general case of multiple unequal-power interferers and provides numerical examples from which the following messages emerge.

- The interference-limited capacity in a typical mobile wireless system can differ substantially from its AWGN-limited counterpart.
- The difference between such capacities grows as the number of receive antennas increases and/or the number of transmit antennas at the interferers diminishes.
- Transmit correlation is detrimental at the user of interest but beneficial when exhibited by the interferers.

B. Ricean Fading

When the channel for the desired user is Ricean, $\frac{E_b}{N_0}_{\min}$ and S_0 are given by (14) and (15) along with (45) and (47) in the Appendix, part B. Since the expressions are quite involved, we gain insight by focusing on the realm of large K-factors, where the following applies.

Proposition 6: Consider a channel known by the receiver and given by

$$\boldsymbol{H} = \boldsymbol{a}_{\mathrm{R}} \boldsymbol{a}_{\mathrm{T}}^{\dagger}.$$

If the normalized conditional covariance of the noise is $\boldsymbol{\Phi_n}(\overline{\boldsymbol{H}})$, then

$$\frac{E_b}{N_0}_{\min} = \frac{\log_e 2}{\mathsf{g}} \frac{1}{E[\mathrm{Tr}\{\boldsymbol{a}_{\mathrm{R}} \boldsymbol{a}_{\mathrm{R}}^{\dagger} \boldsymbol{\Phi_n}^{-1}(\overline{\boldsymbol{H}})\}]}$$

and

$$S_0 = \frac{2\, n_{\mathrm{R}}}{\zeta(\boldsymbol{a}_{\mathrm{R}} \boldsymbol{a}_{\mathrm{R}}^{\dagger} \boldsymbol{\Phi_n}^{-1}(\overline{\boldsymbol{H}}))}$$

with expectation over the fading of the interferers, $\overline{\boldsymbol{H}}$.

Proof: See Appendix B.

The low-SNR capacity is thus uniquely determined by the expected trace and dispersion of $\boldsymbol{a}_{\mathrm{R}} \boldsymbol{a}_{\mathrm{R}}^{\dagger} \boldsymbol{\Phi_n}^{-1}(\overline{\boldsymbol{H}})$. Also noteworthy is that, unlike in AWGN, the capacity does depend—through $\boldsymbol{a}_{\mathrm{R}}$—on the angle of arrival of the Ricean channel component. Even when the channel fading is uncorrelated across the receive antennas, the noise color may favor some directions of arrival with respect to others.

VII. Conclusion

For realistic channels and noise models, no insightful expressions for the capacity as function of the SNR had been found thus far. As shown in this paper, however, in the low-SNR regime it is possible to circumvent the computation of $C(\mathsf{SNR})$ entirely by posing the capacity as function of $\frac{E_b}{N_0}$. Since $\mathsf{C}(\frac{E_b}{N_0})$ is highly linear at low SNR, its characterization requires only two parameters, namely, the $\frac{E_b}{N_0}$ at which the capacity becomes zero and the slope therein. Throughout the paper, we have derived expressions for these parameters using realistic channel and noise models and, from these expressions, we have learned how the single-user capacity is affected by the existence of antenna correlation, different polarizations, a Ricean term, fading and correlation within the noise, etc. Moreover, these lessons have been learned without the need to invoke a large number of antennas and they are supported by experimental data [64]. We have found that at low SNR we can draw the following conclusions.

- The reduction in capacity caused by antenna correlation can be uniquely quantified through a scalar quantity for each of the arrays, transmit and receive. These quantities, which we refer to as the *correlation numbers* of the transmit $\zeta(\boldsymbol{\Theta}_{\mathrm{T}})$ and receive $\zeta(\boldsymbol{\Theta}_{\mathrm{R}})$ arrays can be easily computed.
- With correlated antennas, the capacity in Rayleigh channels does not grow—unlike in the canonical case—linearly with the number of antennas. In fact, the fraction of canonical capacity achievable in the presence of correlation diminishes, for most array structures, monotonically

with the number of antennas. (This reduction is sustained by empirical data even beyond the low-SNR regime.)

- The bandwidth expansion factor brought about by antenna correlation in Rayleigh fading equals

$$\frac{n_{\mathrm{T}}\zeta(\boldsymbol{\Theta}_{\mathrm{R}})+n_{\mathrm{R}}\zeta(\boldsymbol{\Theta}_{\mathrm{T}})}{n_{\mathrm{T}}+n_{\mathrm{R}}}.$$

  In particular, when one array has two antennas with correlation coefficient ρ, the bandwidth expansion factor is equal to $1+\rho^2\frac{n}{2+n}$ if the other array has n uncorrelated antennas, and equal to $1+\rho^2$ if the other array also has two ρ-correlated antennas.

- If either the transmitter or the receiver has a single antenna, the capacity is higher in strong Ricean conditions than in Rayleigh fading.

- If both transmitter and receiver have two antennas, the capacity is independent of the Ricean factor.

- Modeling outside interference (from other sectors or cells) as AWGN may lead to very inaccurate estimates of the actual capacity. With such interference being spatially colored and subject to fading, the capacity is always higher—sometimes much higher—than in the presence of an equivalent amount of AWGN.

- Both color and fading within the noise enhance the capacity. The impact of noise color relates to the ratio between the number of significant degrees of freedom in the interference and the number of receive antennas. The impact of noise fading, on the other hand, is only significant when both are small.

- Except to overcome thermal noise in systems wherein the orientation of both transmitter and receiver can be controlled, orthogonal polarizations should be used before resorting to spatial diversity.

It is worth pointing out that in no way do our conclusions hinge on the choice of defining the capacity slope on a logarithmic scale. Had we instead defined it on a linear scale, similar observations would have been drawn [28].

Our proof of the various formulas presented in the paper require the evaluation of the first- and second-order moments of the trace of the product of certain random matrices. To that end, we have found it useful to invoke auxiliary results on the mixed moments of Haar-distributed matrices, which have proven useful in the development of free-probability results in the theory of large random matrices.

APPENDIX

Given a matrix $\boldsymbol{A}$, we shall use $(\boldsymbol{A})_{i,j}$ to denote its (i,j)th element and $(\boldsymbol{A})_i$ to denote its ith row. Before proving the various properties and propositions contained in the paper, we present some auxiliary results that will be invoked throughout.

A. Auxiliary Results

Let $\boldsymbol{U}$ be an $(n\times n)$ matrix such that $\boldsymbol{U}\boldsymbol{U}^{\dagger}=\boldsymbol{I}$. $\boldsymbol{U}$ is uniquely specified by n^2 real parameters. As a subspace of $\mathbb{R}^{2n^2}$, these matrices form a submanifold $\boldsymbol{V}_{n,n}$ of dimension n^2. The uniform distribution over the submanifold $\boldsymbol{V}_{n,n}$ is called *Haar* (probability) measure. Matrices with this distribution are thus called Haar or standard unitaries. The most important property of the Haar probability measure is that it is left and right invariant under unitary transformation. This invariance implies that most mixed moments of the entries of a Haar unitary matrix are zero. In particular, we have the following.

Lemma 1 [70]: Consider the product of the k_pth power of $(\boldsymbol{U})_{i_p,j_p}$ and the m_pth power of $(\boldsymbol{U})^{*}_{i_p,j_p}$ with $\boldsymbol{U}$ a Haar matrix. Whenever $k_1,\ldots,k_\ell,m_1,\ldots,m_\ell>0$, and there exist some i or j such that

$$\sum_{p:\,i_p=i}(k_p-m_p)\neq 0 \quad \text{or} \quad \sum_{p:\,j_p=j}(k_p-m_p)\neq 0$$

then

$$E\left[\prod_{p=1}^{\ell}\left((\boldsymbol{U})_{i_p,j_p}\right)^{k_p}\left((\boldsymbol{U}^{\dagger})_{j_p,i_p}\right)^{m_p}\right]=0. \tag{35}$$

As a special case, (35) holds whenever the sum $\sum_{p=1}^{\ell}(k_p+m_p)$ is odd.

Lemma 2 [70]: If $1\le i_1,j_1,i_2,j_2\le n$, $i_1\neq i_2$, $j_1\neq j_2$, and $\boldsymbol{U}$ is a Haar matrix, then

$$\begin{aligned}
E[|(\boldsymbol{U})_{i_1,j_1}|^2]&=\frac{1}{n}\\
E[|(\boldsymbol{U})_{i_1,j_1}|^4]&=\frac{2}{n(n+1)}\\
E[|(\boldsymbol{U})_{i_1,j_1}|^2|(\boldsymbol{U})_{i_2,j_1}|^2]&=E[|(\boldsymbol{U})_{i_1,j_1}|^2|(\boldsymbol{U})_{i_1,j_2}|^2]\\
&=\frac{1}{n(n+1)}\\
E[|(\boldsymbol{U})_{i_1,j_1}|^2|(\boldsymbol{U})_{i_2,j_2}|^2]&=\frac{1}{n^2-1}\\
E[(\boldsymbol{U})_{i_1,j_1}(\boldsymbol{U})_{i_2,j_2}(\boldsymbol{U})_{i_1,j_2}(\boldsymbol{U})_{i_2,j_1}]&=-\frac{1}{n(n^2-1)}.
\end{aligned}$$

Using these properties of the Haar measure, we can prove the following result.

Lemma 3: Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be arbitrary Hermitian $(n\times n)$ random matrices, mutually independent. Let $\boldsymbol{U}$ be a random Haar matrix independent of $\boldsymbol{A}$ and $\boldsymbol{B}$ with $\boldsymbol{V}=\boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^{\dagger}$. Then

$$E[\mathrm{Tr}\{\boldsymbol{A}\boldsymbol{V}\}]=\frac{1}{n}E[\mathrm{Tr}\{\boldsymbol{A}\}]E[\mathrm{Tr}\{\boldsymbol{V}\}] \tag{36}$$

while

$$\begin{aligned}
&E[\mathrm{Tr}\{(\boldsymbol{A}\boldsymbol{V})^2\}]\\
&\quad=\frac{E[\mathrm{Tr}\{\boldsymbol{A}^2\}]}{n^2-1}\left(E[\mathrm{Tr}^2\{\boldsymbol{V}\}]-\frac{1}{n}E[\mathrm{Tr}\{\boldsymbol{V}^2\}]\right)\\
&\qquad+\frac{E[\mathrm{Tr}^2\{\boldsymbol{A}\}]}{n^2-1}\left(E[\mathrm{Tr}\{\boldsymbol{V}^2\}]-\frac{1}{n}E[\mathrm{Tr}^2\{\boldsymbol{V}\}]\right)
\end{aligned} \tag{37}$$

and

$$\begin{aligned}
&E[\mathrm{Tr}^2\{\boldsymbol{A}\boldsymbol{V}\}]\\
&\quad=\frac{E[\mathrm{Tr}\{\boldsymbol{A}^2\}]}{n^2-1}\left(E[\mathrm{Tr}\{\boldsymbol{V}^2\}]-\frac{1}{n}E[\mathrm{Tr}^2\{\boldsymbol{V}\}]\right)\\
&\qquad+\frac{E[\mathrm{Tr}^2\{\boldsymbol{A}\}]}{n^2-1}\left(E[\mathrm{Tr}^2\{\boldsymbol{V}\}]-\frac{1}{n}E[\mathrm{Tr}\{\boldsymbol{V}^2\}]\right).
\end{aligned} \tag{38}$$

Proof: Since $\boldsymbol{A}$ is independent of $\boldsymbol{B}$ and $\boldsymbol{U}$

$$E[\mathrm{Tr}\{\boldsymbol{AV}\}] = \mathrm{Tr}\{E[\boldsymbol{A}]\,E[\boldsymbol{UBU}^\dagger]\}. \tag{39}$$

$\boldsymbol{U}$ is Haar and, thus, the $(i,\,j)$th element of $E[\boldsymbol{UBU}^\dagger]$ is given by

$$\begin{aligned}(E[\boldsymbol{UBU}^\dagger])_{i,\,j} &= E[(\boldsymbol{U})_i \boldsymbol{B} (\boldsymbol{U})_j^\dagger] \\ &= \sum_{m=1}^{n}\sum_{p=1}^{n} E[(\boldsymbol{U})_{i,\,m}(\boldsymbol{B})_{m,\,p}(\boldsymbol{U})_{j,\,p}^*] \\ &= \delta_{i,\,j}\sum_{m=1}^{n} E[|(\boldsymbol{U})_{i,\,m}|^2]E[(\boldsymbol{B})_{m,\,m}] \\ &= \delta_{i,\,j}\,\frac{1}{n}\,\mathrm{Tr}\{E[\boldsymbol{B}]\} \end{aligned} \tag{40}$$

where (40) follows from Lemma 2 and where $\delta_{i,\,j}$ is the Kronecker delta. Thus, $E[\boldsymbol{UBU}^\dagger]$ is a multiple of the identity, i.e.,

$$\begin{aligned}E[\boldsymbol{UBU}^\dagger] &= \frac{1}{n}\,\mathrm{Tr}\{E[\boldsymbol{B}]\}\boldsymbol{I} \\ &= \frac{1}{n}\,E[\mathrm{Tr}\{\boldsymbol{V}\}]\boldsymbol{I}\end{aligned}$$

which, plugged into (39), proves (36).

Now, let $\boldsymbol{A} = \boldsymbol{Q_A}\boldsymbol{\Lambda}\boldsymbol{Q_A}^\dagger$ be the eigenvalue decomposition of $\boldsymbol{A}$. From the invariance of the Haar probability measure, we have that

$$\begin{aligned}E[\mathrm{Tr}\{(\boldsymbol{AV})^2\}] &= E[\mathrm{Tr}\{\boldsymbol{\Lambda V \Lambda V}\}] \\ &= E[\mathrm{Tr}\{\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}\boldsymbol{\Lambda}^{1/2}\}] \\ &= \sum_{ij} E[(\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}\boldsymbol{\Lambda}^{1/2})_{i,\,j}^2].\end{aligned}$$

With some algebraic manipulations and neglecting those terms whose expected value is zero, we find that, for $i \neq j$

$$\begin{aligned}&E[(\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}\boldsymbol{\Lambda}^{1/2})_{i,\,j}^2] \\ &\quad= E[(\boldsymbol{\Lambda})_{i,i}(\boldsymbol{\Lambda})_{j,j}(\boldsymbol{V})_{i,\,j}^2] \\ &\quad= E[(\boldsymbol{\Lambda})_{i,i}(\boldsymbol{\Lambda})_{j,j}]E\left[\left(\sum_{mp}(\boldsymbol{U})_{i,m}(\boldsymbol{B})_{m,p}(\boldsymbol{U})_{j,p}^*\right)^2\right] \\ &\quad= E[(\boldsymbol{\Lambda})_{i,i}(\boldsymbol{\Lambda})_{j,j}] \\ &\qquad\cdot E\left[\sum_{mp}\sum_{\ell q}(\boldsymbol{B})_{m,p}(\boldsymbol{B})_{\ell,q}^*(\boldsymbol{U})_{i,m}(\boldsymbol{U})_{j,p}^*(\boldsymbol{U})_{i,\ell}^*(\boldsymbol{U})_{j,q}\right] \\ &\quad= E[(\boldsymbol{\Lambda})_{i,i}(\boldsymbol{\Lambda})_{j,j}]E\left[\sum_{m}|(\boldsymbol{B})_{m,m}|^2|(\boldsymbol{U})_{i,m}|^2|(\boldsymbol{U})_{j,m}|^2\right] \\ &\qquad+ E[(\boldsymbol{\Lambda})_{i,i}(\boldsymbol{\Lambda})_{j,j}]E\left[\sum_{m\neq p}|(\boldsymbol{B})_{m,p}|^2|(\boldsymbol{U})_{i,m}|^2|(\boldsymbol{U})_{j,p}|^2\right] \\ &\qquad+ E[(\boldsymbol{\Lambda})_{i,i}(\boldsymbol{\Lambda})_{j,j}] \\ &\qquad\cdot E\left[\sum_{m\neq\ell}(\boldsymbol{B})_{m,m}(\boldsymbol{B})_{\ell,\ell}^*(\boldsymbol{U})_{i,m}(\boldsymbol{U})_{j,m}^*(\boldsymbol{U})_{i,\ell}^*(\boldsymbol{U})_{j,\ell}\right] \\ &\quad= E[(\boldsymbol{\Lambda})_{i,i}(\boldsymbol{\Lambda})_{j,j}]\left(\frac{E\left[\sum_m|(\boldsymbol{B})_{m,m}|^2\right]}{n(n+1)}\right. \\ &\qquad\left.+\frac{E\left[\sum_{m\neq p}|(\boldsymbol{B})_{m,p}|^2\right]}{n^2-1} - \frac{E\left[\sum_{m\neq\ell}(\boldsymbol{B})_{m,m}(\boldsymbol{B})_{\ell,\ell}^*\right]}{n(n^2-1)}\right) \\ &\quad= E[(\boldsymbol{\Lambda})_{i,i}(\boldsymbol{\Lambda})_{j,j}]\left(\frac{E[\mathrm{Tr}\{\boldsymbol{B}^2\}]}{n^2-1} - \frac{E[\mathrm{Tr}^2\{\boldsymbol{B}\}]}{n(n^2-1)}\right)\end{aligned}$$

while, for $i = j$

$$\begin{aligned}&E[(\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}\boldsymbol{\Lambda}^{1/2})_{i,i}^2] \\ &\quad= E[(\boldsymbol{\Lambda})_{i,i}^2(\boldsymbol{V})_{i,i}^2] \\ &\quad= E[(\boldsymbol{\Lambda})_{i,i}^2]E\left[\left(\sum_{mp}(\boldsymbol{U})_{i,m}(\boldsymbol{B})_{m,p}(\boldsymbol{U})_{i,p}^*\right)^2\right] \\ &\quad= E[(\boldsymbol{\Lambda})_{i,i}^2] \\ &\qquad\cdot E\left[\sum_{mp}\sum_{\ell q}(\boldsymbol{B})_{m,p}(\boldsymbol{B})_{\ell,q}^*(\boldsymbol{U})_{i,m}(\boldsymbol{U})_{i,p}^*(\boldsymbol{U})_{i,\ell}^*(\boldsymbol{U})_{i,q}\right] \\ &\quad= E[(\boldsymbol{\Lambda})_{i,i}^2]E\left[\sum_{m}|(\boldsymbol{B})_{m,m}|^2|(\boldsymbol{U})_{i,m}|^4\right] \\ &\qquad+ E[(\boldsymbol{\Lambda})_{i,i}^2]E\left[\sum_{m\neq p}|(\boldsymbol{B})_{m,p}|^2|(\boldsymbol{U})_{i,m}|^2|(\boldsymbol{U})_{i,p}|^2\right] \\ &\qquad+ E[(\boldsymbol{\Lambda})_{i,i}^2]E\left[\sum_{m\neq\ell}(\boldsymbol{B})_{m,m}(\boldsymbol{B})_{\ell,\ell}^*|(\boldsymbol{U})_{i,m}|^2|(\boldsymbol{U})_{i,\ell}|^2\right] \\ &\quad= E[(\boldsymbol{\Lambda})_{i,i}^2]\left(\frac{2E\left[\sum_m|(\boldsymbol{B})_{m,m}|^2\right]}{n(n+1)}\right. \\ &\qquad\left.+\frac{E\left[\sum_{m\neq p}|(\boldsymbol{B})_{m,p}|^2\right]}{n(n+1)} + \frac{E\left[\sum_{m\neq\ell}(\boldsymbol{B})_{m,m}(\boldsymbol{B})_{\ell,\ell}^*\right]}{n(n+1)}\right) \\ &\quad= \frac{E[(\boldsymbol{\Lambda})_{i,i}^2]}{n(n+1)}\left(E[\mathrm{Tr}\{\boldsymbol{B}^2\}] + E[\mathrm{Tr}^2\{\boldsymbol{B}\}]\right).\end{aligned}$$

Altogether, the equation at the top of the following page which, given that

$$\mathrm{Tr}\{\boldsymbol{B}\} = \mathrm{Tr}\{\boldsymbol{V}\} \quad \text{and} \quad \mathrm{Tr}\{\boldsymbol{B}^2\} = \mathrm{Tr}\{\boldsymbol{V}^2\}$$

proves (37). Finally

$$\begin{aligned}E[\mathrm{Tr}^2\{\boldsymbol{AV}\}] &= E[\mathrm{Tr}^2\{\boldsymbol{Q_A}\boldsymbol{\Lambda}\boldsymbol{Q_A}^\dagger\boldsymbol{V}\}] \\ &= E[\mathrm{Tr}^2\{\boldsymbol{\Lambda V}\}] \\ &= E[\mathrm{Tr}^2\{\boldsymbol{\Lambda U \Lambda_B U}^\dagger\}]\end{aligned}$$

$$
\begin{aligned}
E[\mathrm{Tr}\{(\boldsymbol{AV})^2\}] &= \sum_{ij} E[(\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}\boldsymbol{\Lambda}^{1/2})^2_{i,j}] \\
&= \sum_i E[(\boldsymbol{\Lambda})^2_{i,i}]\left(\frac{E[\mathrm{Tr}\{\boldsymbol{B}^2\}]+E[\mathrm{Tr}^2\{\boldsymbol{B}\}]}{n(n+1)}\right)+\sum_{i\neq j} E[(\boldsymbol{\Lambda})_{i,i}(\boldsymbol{\Lambda})_{j,j}]\left(\frac{E[\mathrm{Tr}\{\boldsymbol{B}^2\}]}{n^2-1}-\frac{E[\mathrm{Tr}^2\{\boldsymbol{B}\}]}{n(n^2-1)}\right) \\
&= E[\mathrm{Tr}\{\boldsymbol{B}^2\}]\left(\sum_i \frac{E[(\boldsymbol{\Lambda})^2_{i,i}]}{n(n+1)}+\sum_{i\neq j}\frac{E[(\boldsymbol{\Lambda})_{i,i}(\boldsymbol{\Lambda})_{j,j}]}{n^2-1}\right) \\
&\quad + E[\mathrm{Tr}^2\{\boldsymbol{B}\}]\left(\sum_i \frac{E[(\boldsymbol{\Lambda})^2_{i,i}]}{n(n+1)}-\sum_{i\neq j}\frac{E[(\boldsymbol{\Lambda})_{i,i}(\boldsymbol{\Lambda})_{j,j}]}{n(n^2-1)}\right) \\
&= \frac{E[\mathrm{Tr}\{\boldsymbol{B}^2\}]E[\mathrm{Tr}^2\{\boldsymbol{A}\}]}{n^2-1}+\frac{E[\mathrm{Tr}^2\{\boldsymbol{B}\}]E[\mathrm{Tr}\{\boldsymbol{A}^2\}]}{n(n+1)} \\
&\quad + \frac{E[\mathrm{Tr}\{\boldsymbol{A}^2\}](E[\mathrm{Tr}^2\{\boldsymbol{B}\}]-E[\mathrm{Tr}\{\boldsymbol{B}^2\}])-E[\mathrm{Tr}^2\{\boldsymbol{B}\}]E[\mathrm{Tr}^2\{\boldsymbol{A}\}]}{n(n^2-1)}
\end{aligned}
$$

---

with the eigenvalue decomposition $\bar{\boldsymbol{B}} = \boldsymbol{Q}_{\boldsymbol{B}}\boldsymbol{\Lambda}_{\boldsymbol{B}}\boldsymbol{Q}_{\boldsymbol{B}}^{\dagger}$. Further algebra yields

$$
\begin{aligned}
&E[\mathrm{Tr}^2\{\boldsymbol{AV}\}] \\
&= E\left[\left(\sum_i (\boldsymbol{\Lambda})_{i,i}(\boldsymbol{U})_i\boldsymbol{\Lambda}_{\boldsymbol{B}}(\boldsymbol{U})_i^{\dagger}\right)^2\right] \\
&= E\left[\sum_i\sum_j (\boldsymbol{\Lambda})_{i,i}(\boldsymbol{\Lambda})_{j,j}(\boldsymbol{U})_j\boldsymbol{\Lambda}_{\boldsymbol{B}}(\boldsymbol{U})_j^{\dagger}(\boldsymbol{U})_i\boldsymbol{\Lambda}_{\boldsymbol{B}}(\boldsymbol{U})_i^{\dagger})\right] \\
&= \sum_i E[(\boldsymbol{\Lambda})^2_{i,i}]\,E[(\boldsymbol{U})_i\boldsymbol{\Lambda}_{\boldsymbol{B}}(\boldsymbol{U})_i^{\dagger}(\boldsymbol{U})_i\boldsymbol{\Lambda}_{\boldsymbol{B}}(\boldsymbol{U})_i^{\dagger})] \\
&\quad + \sum_{i\neq j} E[(\boldsymbol{\Lambda})_{i,i}(\boldsymbol{\Lambda})_{j,j}]E[(\bar{\boldsymbol{U}})_j\boldsymbol{\Lambda}_{\boldsymbol{B}}(\bar{\boldsymbol{U}})_j^{\dagger}(\bar{\boldsymbol{U}})_i\boldsymbol{\Lambda}_{\boldsymbol{B}}(\bar{\boldsymbol{U}})_i^{\dagger})].
\end{aligned}
$$

For $i \neq j$

$$
\begin{aligned}
&E[(\boldsymbol{U})_j\boldsymbol{\Lambda}_{\boldsymbol{B}}(\boldsymbol{U})_j^{\dagger}(\boldsymbol{U})_i\boldsymbol{\Lambda}_{\boldsymbol{B}}(\boldsymbol{U})_i^{\dagger})] \\
&= E\left[\sum_p\sum_{\ell}(\boldsymbol{\Lambda}_{\boldsymbol{B}})_{p,p}(\boldsymbol{\Lambda}_{\boldsymbol{B}})_{\ell,\ell}|(\boldsymbol{U})_{j,p}|^2|(\boldsymbol{U})_{i,\ell}|^2\right] \\
&= \frac{E\left[\sum_p(\boldsymbol{\Lambda}_{\boldsymbol{B}})^2_{p,p}\right]}{n(n+1)}+\frac{E\left[\sum_{p\neq\ell}(\boldsymbol{\Lambda}_{\boldsymbol{B}})_{p,p}(\boldsymbol{\Lambda}_{\boldsymbol{B}})_{\ell,\ell}\right]}{n^2-1} \\
&= -\frac{E[\mathrm{Tr}\{\boldsymbol{B}^2\}]}{n(n^2-1)}+\frac{E[\mathrm{Tr}^2\{\boldsymbol{B}\}]}{n^2-1}
\end{aligned}
$$

while, for $i = j$,

$$
\begin{aligned}
&E[(\boldsymbol{U})_j\boldsymbol{\Lambda}_{\boldsymbol{B}}(\boldsymbol{U})_j^{\dagger}(\boldsymbol{U})_i\boldsymbol{\Lambda}_{\boldsymbol{B}}(\boldsymbol{U})_i^{\dagger})] \\
&= E\left[\sum_p\sum_{\ell}(\boldsymbol{\Lambda}_{\boldsymbol{B}})_{p,p}(\boldsymbol{\Lambda}_{\boldsymbol{B}})_{\ell,\ell}|(\boldsymbol{U})_{i,p}|^2|(\boldsymbol{U})_{i,\ell}|^2\right] \\
&= 2\,\frac{E\left[\sum_p(\boldsymbol{\Lambda}_{\boldsymbol{B}})^2_{p,p}\right]}{n(n+1)}+\frac{E\left[\sum_{p\neq\ell}(\boldsymbol{\Lambda}_{\boldsymbol{B}})_{p,p}(\boldsymbol{\Lambda}_{\boldsymbol{B}})_{\ell,\ell}\right]}{n(n+1)} \\
&= \frac{E[\mathrm{Tr}\{\boldsymbol{B}^2\}]}{n(n+1)}+\frac{E[\mathrm{Tr}^2\{\boldsymbol{B}\}]}{n(n+1)}.
\end{aligned}
$$

Altogether

$$
\begin{aligned}
E[\mathrm{Tr}^2\{\boldsymbol{AV}\}] &= E[\mathrm{Tr}\{\boldsymbol{A}^2\}]\left(\frac{E[\mathrm{Tr}\{\boldsymbol{B}^2\}]}{n(n+1)}+\frac{E[\mathrm{Tr}^2\{\boldsymbol{B}\}]}{n(n+1)}\right. \\
&\quad \left.+\frac{E[\mathrm{Tr}\{\boldsymbol{B}^2\}]}{n(n^2-1)}-\frac{E[\mathrm{Tr}^2\{\boldsymbol{B}\}]}{n^2-1}\right) \\
&\quad + E[\mathrm{Tr}^2\{\boldsymbol{A}\}]\left(\frac{E[\mathrm{Tr}^2\{\boldsymbol{B}\}]}{n^2-1}-\frac{E[\mathrm{Tr}\{\boldsymbol{B}^2\}]}{n(n^2-1)}\right)
\end{aligned}
$$

from which (38) follows easily.

Lemma 4: Given an $(n \times m)$ matrix $\boldsymbol{W}$ with independent zero-mean unit-variance complex Gaussian random entries

$$
\begin{aligned}
E[\mathrm{Tr}\{\boldsymbol{WW}^{\dagger}\}] &= mn \\
E[\mathrm{Tr}\{(\boldsymbol{WW}^{\dagger})^2\}] &= mn(m+n) \\
E[\mathrm{Tr}^2\{\boldsymbol{WW}^{\dagger}\}] &= mn(mn+1).
\end{aligned}
$$

Lemma 5: Given an $(n \times m)$ matrix $\boldsymbol{W}$ with independent zero-mean unit-variance complex Gaussian random entries as well as an $(m \times m)$ matrix $\boldsymbol{D}$ and an $(n \times n)$ matrix $\boldsymbol{S}$, the matrices $\boldsymbol{WDW}^{\dagger}$ and $\boldsymbol{W}^{\dagger}\boldsymbol{SW}$ admit the following eigenvalue decomposition:

$$
\begin{aligned}
\boldsymbol{WDW}^{\dagger} &= \boldsymbol{UMU}^{\dagger} \\
\boldsymbol{W}^{\dagger}\boldsymbol{SW} &= \tilde{\boldsymbol{U}}\boldsymbol{N}\tilde{\boldsymbol{U}}^{\dagger}
\end{aligned}
$$

where $\boldsymbol{U}$ and $\tilde{\boldsymbol{U}}$ are $(n \times n)$ and $(m \times m)$ Haar matrices and $\boldsymbol{M}$ and $\boldsymbol{N}$ are $(n \times n)$ and $(m \times m)$ random matrices independent of $\bar{\boldsymbol{U}}$ and $\tilde{\boldsymbol{U}}$, respectively.

Proof: This lemma is derived as part of the proof of [70, Theorem 3.2].

Lemma 6: Consider an $(n \times m)$ matrix $\boldsymbol{W}$ with independent zero-mean unit-variance complex Gaussian random entries. For $m > n$

$$
E\left[\mathrm{Tr}\left\{(\boldsymbol{WW}^{\dagger})^{-1}\right\}\right] = \frac{n}{m-n} \tag{41}
$$

and, for $m > n+1$

$$E\left[\mathrm{Tr}\left\{(\boldsymbol{W}\boldsymbol{W}^\dagger)^{-2}\right\}\right] = \frac{mn}{(m-n)^3-(m-n)}$$
$$E\left[\mathrm{Tr}^2\left\{(\boldsymbol{W}\boldsymbol{W}^\dagger)^{-1}\right\}\right] = \frac{n}{m-n}\left(\frac{m}{(m-n)^2-1} + \frac{n-1}{m-n+1}\right).$$

Proof:

$$E\left[\mathrm{Tr}\left\{(\boldsymbol{W}\boldsymbol{W}^\dagger)^{-1}\right\}\right] = n\,E\left[\frac{1}{\lambda}\right]$$

with λ an arbitrary eigenvalue of $(\boldsymbol{W}\boldsymbol{W}^\dagger)$, whose marginal density is given by the Wishart distribution [2]

$$f_\lambda(z) = \frac{1}{n}\sum_{k=0}^{n-1}\frac{k!(L_k^{m-n}(z))^2}{(k+m-n)!}\,z^{m-n}e^{-z}$$

with L_k^{m-n} the corresponding Laguerre polynomial of order k [71]. The expectation can be computed to yield

$$E\left[\frac{1}{\lambda}\right] = \int_0^\infty \frac{f_\lambda(z)}{z}\,\mathrm{d}z = \frac{1}{m-n}$$

which leads to (41). At the same time

$$E\left[\mathrm{Tr}\left\{(\boldsymbol{W}\boldsymbol{W}^\dagger)^{-2}\right\}\right] = nE\left[\frac{1}{\lambda^2}\right] = n\int_0^\infty \frac{f_\lambda(z)}{z^2}\,\mathrm{d}z = \frac{mn}{(m-n)^3-(m-n)}$$

and, from the joint density of any two different eigenvalues of $(\boldsymbol{W}\boldsymbol{W}^\dagger)$ given by

$$f_{\lambda\lambda'} = \frac{(\lambda\lambda')^{m-n}}{n(n-1)}\,e^{-(\lambda+\lambda')}\left[\left(\sum_{k=0}^{n-1}\frac{k!L_k^{m-n}(\lambda)L_k^{m-n}(\lambda')}{(k+m-n)!}\right)^2 + \sum_{k=0}^{n-1}\frac{k!(L_k^{m-n}(\lambda))^2}{(k+m-n)!}\sum_{k=0}^{n-1}\frac{k!(L_k^{m-n}(\lambda'))^2}{(k+m-n)!}\right]$$

we find that

$$\begin{aligned}
&E\left[\mathrm{Tr}^2\left\{(\boldsymbol{W}\boldsymbol{W}^\dagger)^{-1}\right\}\right]\\
&\quad=\sum_{i=1}^n\sum_{j=1}^n E\left[\frac{1}{\lambda_i\lambda_j}\right]\\
&\quad=\left(nE\left[\frac{1}{\lambda^2}\right]+(n^2-n)E\left[\frac{1}{\lambda\lambda'}\right]\right)\\
&\quad=n\left(\int_0^\infty\frac{f_\lambda(z)}{z^2}\,\mathrm{d}z+(n-1)\iint_0^\infty\frac{f_{\lambda\lambda'}(z,z')}{zz'}\,\mathrm{d}z\,\mathrm{d}z'\right)\\
&\quad=\frac{n}{m-n}\left(\frac{m}{(m-n)^2-1}+\frac{n-1}{m-n+1}\right).
\end{aligned}$$

□

B. Proof of Propositions 1, 3, and 6

To prove these results, we first derive general expressions for $\frac{E_b}{N_0}_{\min}$ and S_0 and then particularize them to each of the propositions. Notice from (14) and (15) that, in its most general form, expressing $\frac{E_b}{N_0}_{\min}$ and S_0 entails characterizing the terms

$$E[\mathrm{Tr}\{\boldsymbol{H}\boldsymbol{H}^\dagger\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}] \tag{42}$$

and

$$E[\mathrm{Tr}\{(\boldsymbol{H}\boldsymbol{H}^\dagger\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}}))^2\}]. \tag{43}$$

Equation (42) can be expanded, using our channel model, as

$$\begin{aligned}
&E[\mathrm{Tr}\{\boldsymbol{H}\boldsymbol{H}^\dagger\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}]\\
&\quad=\frac{1}{\mathsf{K}+1}E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}\boldsymbol{W}^\dagger\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}]\\
&\qquad+\frac{\mathsf{K}}{\mathsf{K}+1}E[\mathrm{Tr}\{\boldsymbol{H}_0\boldsymbol{H}_0^\dagger\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}].
\end{aligned}\tag{44}$$

From Lemma 5, we have that $\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}\boldsymbol{W}^\dagger$ admits the eigenvalue decomposition

$$\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}\boldsymbol{W}^\dagger = \boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^\dagger$$

where $\boldsymbol{U}$ is an $(n_{\mathrm{R}}\times n_{\mathrm{R}})$ Haar random matrix and $\boldsymbol{B}$ is the $(n_{\mathrm{R}}\times n_{\mathrm{R}})$ random eigenvalue matrix independent of $\boldsymbol{U}$. Furthermore, $\boldsymbol{H}$ and $\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})$ are statistically independent. Thus, $\boldsymbol{U}$ and $\boldsymbol{B}$ are mutually independent and both are independent of $\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})$. Hence, Lemma 3 applies and we can expand the first term in (44) as

$$\begin{aligned}
&E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}\boldsymbol{W}^\dagger\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}]\\
&\quad=E[\mathrm{Tr}\{\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}\boldsymbol{W}^\dagger\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\}]\\
&\quad=E[\mathrm{Tr}\{\boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^\dagger\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\}]\\
&\quad=\frac{1}{n_{\mathrm{R}}}E[\mathrm{Tr}\{\boldsymbol{B}\}]E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\}]\\
&\quad=\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{T}}\}E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}]\\
&\quad=n_{\mathrm{T}}E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}]
\end{aligned}$$

where we have used

$$E[\mathrm{Tr}\{\boldsymbol{B}\}] = E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{T}}\boldsymbol{W}^\dagger\boldsymbol{W}\}] = n_{\mathrm{R}}\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{T}}\}.$$

The second term in (44), in turn, expands as

$$\begin{aligned}
E[\mathrm{Tr}\{\boldsymbol{H}_0\boldsymbol{H}_0^\dagger\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}] &= E[\mathrm{Tr}\{\boldsymbol{a}_{\mathrm{R}}\boldsymbol{a}_{\mathrm{T}}^\dagger\boldsymbol{a}_{\mathrm{T}}\boldsymbol{a}_{\mathrm{R}}^\dagger\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}]\\
&= n_{\mathrm{T}}E[\mathrm{Tr}\{\boldsymbol{a}_{\mathrm{R}}\boldsymbol{a}_{\mathrm{R}}^\dagger\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}].
\end{aligned}$$

Altogether, (42) is given by

$$E[\mathrm{Tr}\{\boldsymbol{H}\boldsymbol{H}^\dagger\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}] = \frac{n_{\mathrm{T}}}{\mathsf{K}+1}\Big(E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}] + \mathsf{K}E[\mathrm{Tr}\{\boldsymbol{a}_{\mathrm{R}}\boldsymbol{a}_{\mathrm{R}}^\dagger\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}]\Big). \tag{45}$$

Let us now turn our attention to (43), which expands—disregarding those terms whose mean is zero—into

$$\begin{aligned}
&E[\mathrm{Tr}\{(\boldsymbol{H}\boldsymbol{H}^\dagger\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}}))^2\}]\\
&\quad=E[\mathrm{Tr}\{\boldsymbol{H}\boldsymbol{H}^\dagger\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{H}\boldsymbol{H}^\dagger\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}]
\end{aligned}$$

$$= \frac{1}{(\mathsf{K}+1)^2} E[\mathrm{Tr}\{(\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}\boldsymbol{W}^{\dagger}\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}}))^2\}] + \frac{\mathsf{K}^2}{(\mathsf{K}+1)^2}\Big(E[\mathrm{Tr}\{(\boldsymbol{H}_0\boldsymbol{H}_0^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}}))^2\}] + 2E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}\boldsymbol{W}^{\dagger}\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{H}_0\boldsymbol{H}_0^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}] + 2E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{T}}^{1/2}\boldsymbol{W}^{\dagger}\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{\Theta}_{\mathrm{R}}^{1/2} \cdot \boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}^{1/2}\boldsymbol{H}_0^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{H}_0\}]\Big). \quad (46)$$

Define $\boldsymbol{A} \triangleq \boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}$ and notice that

$$E[\mathrm{Tr}\{(\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}\boldsymbol{W}^{\dagger}\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}}))^2\}] = E[\mathrm{Tr}\{(\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}\boldsymbol{W}^{\dagger}\boldsymbol{A})^2\}] = E[\mathrm{Tr}\{(\boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^{\dagger}\boldsymbol{A})^2\}].$$

Applying Lemma 3 twice followed by Lemma 5 and Lemma 4, we find that

$$E[\mathrm{Tr}\{(\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}\boldsymbol{W}^{\dagger}\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}}))^2\}] = E[\mathrm{Tr}\{(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}}))^2\}]\mathrm{Tr}^2\{\boldsymbol{\Theta}_{\mathrm{T}}\} + E[\mathrm{Tr}^2\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}]\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{T}}^2\}.$$

Finally, focus on the last two terms in (46). Define

$$\boldsymbol{Q} \triangleq \boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{H}_0\boldsymbol{H}_0^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}$$

and apply yet again Lemma 3 to obtain

$$E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}\boldsymbol{W}^{\dagger}\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{H}_0\boldsymbol{H}_0^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}] = E[\mathrm{Tr}\{\boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^{\dagger}\boldsymbol{Q}\}] = \frac{1}{n_{\mathrm{R}}}E[\mathrm{Tr}\{\boldsymbol{B}\}E[\mathrm{Tr}\{\boldsymbol{Q}\}]$$

and further

$$E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}\boldsymbol{W}^{\dagger}\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{H}_0\boldsymbol{H}_0^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}] = n_{\mathrm{T}}^2 E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{a}_{\mathrm{R}}\boldsymbol{a}_{\mathrm{R}}^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}].$$

Similarly, invoking Lemma 5 we have

$$\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}^{1/2}\boldsymbol{H}_0^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{H}_0\boldsymbol{\Theta}_{\mathrm{T}}^{1/2}\boldsymbol{W}^{\dagger} = \boldsymbol{U}\boldsymbol{W}\boldsymbol{U}^{\dagger}$$

where $\boldsymbol{U}$ is an $(n_{\mathrm{R}} \times n_{\mathrm{R}})$ Haar random matrix and $\boldsymbol{W}$ is an $(n_{\mathrm{R}} \times n_{\mathrm{R}})$ random matrix independent of $\boldsymbol{U}$. Thus, we can write

$$E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{T}}^{1/2}\boldsymbol{W}^{\dagger}\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{\Theta}_{\mathrm{R}}^{1/2}\boldsymbol{W}\boldsymbol{\Theta}_{\mathrm{T}}^{1/2}\boldsymbol{H}_0^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{H}_0\}] = E[\mathrm{Tr}\{\boldsymbol{A}\boldsymbol{U}\boldsymbol{W}\boldsymbol{U}^{\dagger}\}] = \frac{1}{n_{\mathrm{R}}}E[\mathrm{Tr}\{\boldsymbol{A}\}]E[\mathrm{Tr}\{\boldsymbol{W}\}]$$

to yield

$$E[\mathrm{Tr}\{\boldsymbol{A}\}]E[\mathrm{Tr}\{\boldsymbol{W}\}] = n_{\mathrm{R}}\,E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}] \cdot E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{T}}\boldsymbol{a}_{\mathrm{T}}\boldsymbol{a}_{\mathrm{R}}^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{a}_{\mathrm{R}}\boldsymbol{a}_{\mathrm{T}}^{\dagger}\boldsymbol{\Theta}_{\mathrm{T}}\}].$$

Putting the pieces together, (43) is given by

$$E[\mathrm{Tr}\{(\boldsymbol{H}\boldsymbol{H}^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}}))^2\}] = \frac{1}{(\mathsf{K}+1)^2}(n_{\mathrm{T}}^2 E[\mathrm{Tr}\{(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}}))^2\}] + \mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{T}}^2\}E[\mathrm{Tr}^2\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}] + \mathsf{K}^2 n_{\mathrm{T}}^2 E[\mathrm{Tr}\{(\boldsymbol{a}_{\mathrm{R}}\boldsymbol{a}_{\mathrm{R}}^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}}))^2\}] + 2\mathsf{K}n_{\mathrm{T}}^2 E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{a}_{\mathrm{R}}\boldsymbol{a}_{\mathrm{R}}^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}] + 2\mathsf{K}E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\}] \cdot E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{T}}\boldsymbol{a}_{\mathrm{T}}\boldsymbol{a}_{\mathrm{R}}^{\dagger}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\overline{\boldsymbol{H}})\boldsymbol{a}_{\mathrm{R}}\boldsymbol{a}_{\mathrm{T}}^{\dagger}\boldsymbol{\Theta}_{\mathrm{T}}\}]). \quad (47)$$

Using (42) and (43), general expressions for both the $\frac{E_b}{N_0}_{\min}$ and slope can be assembled. From these expressions, Proposition 1 is obtained by setting $\boldsymbol{\Phi}_{\boldsymbol{n}} = \boldsymbol{I}$. Proposition 3, in turn, is obtained by setting $\mathsf{K} = 0$. Finally, Proposition 6 is obtained by letting $\mathsf{K} \to \infty$.

C. Proof of Proposition 4

Starting from Proposition 3, we need to elaborate on the expressions for $E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}\}]$, $\zeta(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1})$, and $\varphi(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1})$. The last two require $E[\mathrm{Tr}\{(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1})^2\}]$ and $E[\mathrm{Tr}^2\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}\}]$ in addition to the former. In the presence of a single interferer with uncorrelated transmit antennas

$$\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\boldsymbol{H}_1) = \left(1 + \frac{\gamma}{\mathcal{I}_1}\right)\left(\frac{\boldsymbol{\Theta}_1^{1/2}\boldsymbol{W}_1\boldsymbol{W}_1^{\dagger}\boldsymbol{\Theta}_1^{1/2}}{m_1} + \frac{\gamma}{\mathcal{I}_1}\boldsymbol{I}\right)^{-1}.$$

In the interference-limited regime, we have

$$\lim_{\mathcal{I}_1/\gamma\to\infty} \boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}(\boldsymbol{H}_1) = m_1\boldsymbol{\Theta}_1^{-1/2}\left(\boldsymbol{W}_1\boldsymbol{W}_1^{\dagger}\right)^{-1}\boldsymbol{\Theta}_1^{-1/2}. \quad (48)$$

From Lemma 5, we have that $(\boldsymbol{W}_1\boldsymbol{W}_1^{\dagger})$ admits an eigenvalue decomposition such that (48) conforms to the structure of $\boldsymbol{V}$ in Lemma 3, which can be applied to yield

$$\begin{aligned}
&\lim_{\mathcal{I}_1/\gamma\to\infty} E[\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}\}] \\
&= m_1\frac{\mathrm{Tr}\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1}\}}{n_{\mathrm{R}}}E[\mathrm{Tr}\{(\boldsymbol{W}_1\boldsymbol{W}_1^{\dagger})^{-1}\}] \\
&\lim_{\mathcal{I}_1/\gamma\to\infty} E[\mathrm{Tr}\{(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1})^2\}] \\
&= \frac{m_1^2 E[\mathrm{Tr}\{(\boldsymbol{W}_1\boldsymbol{W}_1^{\dagger})^{-2}\}]}{n_{\mathrm{R}}^2 - 1} \cdot \left(\mathrm{Tr}^2\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1}\} - \frac{1}{n_{\mathrm{R}}}\mathrm{Tr}\{(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1})^2\}\right) \\
&\quad + \frac{m_1^2 E[\mathrm{Tr}^2\{(\boldsymbol{W}_1\boldsymbol{W}_1^{\dagger})^{-1}\}]}{n_{\mathrm{R}}^2 - 1} \cdot \left(\mathrm{Tr}\{(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1})^2\} - \frac{1}{n_{\mathrm{R}}}\mathrm{Tr}^2\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1}\}\right) \\
&\lim_{\mathcal{I}_1/\gamma\to\infty} E[\mathrm{Tr}^2\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}\}] \\
&= \frac{m_1^2 E[\mathrm{Tr}\{(\boldsymbol{W}_1\boldsymbol{W}_1^{\dagger})^{-2}\}]}{n_{\mathrm{R}}^2 - 1} \cdot \left(\mathrm{Tr}\{(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1})^2\} - \frac{1}{n_{\mathrm{R}}}\mathrm{Tr}^2\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1})\}\right) \\
&\quad + \frac{m_1^2 E[\mathrm{Tr}^2\{(\boldsymbol{W}_1\boldsymbol{W}_1^{\dagger})^{-1}\}]}{n_{\mathrm{R}}^2 - 1} \cdot \left(\mathrm{Tr}^2\{\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1}\} - \frac{1}{n_{\mathrm{R}}}\mathrm{Tr}\{(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1})^2\}\right). \quad (49)
\end{aligned}$$

Defining

$$\Psi \triangleq \zeta\left((\boldsymbol{W}_1\boldsymbol{W}_1^{\dagger})^{-1}\right)$$

and

$$\Upsilon \triangleq \varphi\left((\boldsymbol{W}_1\boldsymbol{W}_1^{\dagger})^{-1}\right)$$

we can use the above expressions to obtain

$$\lim_{(\mathcal{I}_1/\gamma)\to\infty} \zeta(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}) = \frac{n_{\mathrm{R}}^2 - \zeta(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1})}{n_{\mathrm{R}}^2 - 1}\,\Psi + \frac{n_{\mathrm{R}}^2(\zeta(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1}) - 1)}{n_{\mathrm{R}}^2 - 1}\,\Upsilon$$

$$\lim_{(\mathcal{I}_1/\gamma)\to\infty} \varphi(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Phi}_{\boldsymbol{n}}^{-1}) = \frac{\zeta(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1}) - 1}{n_{\mathrm{R}}^2 - 1}\,\Psi + \frac{n_{\mathrm{R}}^2 - \zeta(\boldsymbol{\Theta}_{\mathrm{R}}\boldsymbol{\Theta}_1^{-1})}{n_{\mathrm{R}}^2 - 1}\,\Upsilon$$

from which the slope in (29) becomes (32). For $m_1 > n_{\mathrm{R}}$, Lemma 6 yields

$$E\left[\mathrm{Tr}\left\{(\boldsymbol{W}_1\boldsymbol{W}_1^\dagger)^{-1}\right\}\right] = \frac{n_{\mathrm{R}}}{m_1 - n_{\mathrm{R}}}$$

while, for $m_1 \le n_{\mathrm{R}}$

$$\lim_{(\mathcal{I}_1/\gamma)\to\infty} E\left[\mathrm{Tr}\left\{\left(\frac{\boldsymbol{H}_1\boldsymbol{H}_1^\dagger}{m_1} + \frac{\gamma}{\mathcal{I}_1}\boldsymbol{I}\right)^{-1}\right\}\right] \to \infty$$

and thus, altogether

$$E[\mathrm{Tr}\{(\boldsymbol{W}_1\boldsymbol{W}_1^\dagger)^{-1}\}] = \frac{1}{\left[\frac{1}{n_{\mathrm{R}}} - \frac{1}{m_1}\right]^+}$$

from which, given (49), the $\frac{E_b}{N_0}_{\min}$ in (31) follows.

For $m_1 > n_{\mathrm{R}}+1$, (33), and (34) are obtained through Lemma 6, whereas, for $m_1 \le n_{\mathrm{R}}$, they result from

$$\lim_{(\mathcal{I}_1/\gamma)\to\infty} \frac{E\left[\mathrm{Tr}\left\{\left(\frac{\boldsymbol{H}_1\boldsymbol{H}_1^\dagger}{m_1} + \frac{\gamma}{\mathcal{I}_1}\boldsymbol{I}\right)^{-2}\right\}\right]}{E^2\left[\mathrm{Tr}\left\{\left(\frac{\boldsymbol{H}_1\boldsymbol{H}_1^\dagger}{m_1} + \frac{\gamma}{\mathcal{I}_1}\boldsymbol{I}\right)^{-1}\right\}\right]} = \frac{1}{n_{\mathrm{R}} - m_1}$$

and

$$\lim_{(\mathcal{I}_1/\gamma)\to\infty} \frac{E\left[\mathrm{Tr}^2\left\{\left(\frac{\boldsymbol{H}_1\boldsymbol{H}_1^\dagger}{m_1} + \frac{\gamma}{\mathcal{I}_1}\boldsymbol{I}\right)^{-1}\right\}\right]}{E^2\left[\mathrm{Tr}\left\{\left(\frac{\boldsymbol{H}_1\boldsymbol{H}_1^\dagger}{m_1} + \frac{\gamma}{\mathcal{I}_1}\boldsymbol{I}\right)^{-1}\right\}\right]} = 1.$$

D. Proof of Proposition 5

The channel for an interferer with fully correlated transmit antennas is given by

$$\begin{aligned}\boldsymbol{H}_1 &= \boldsymbol{\Theta}_1^{1/2}\boldsymbol{W}_1\mathbf{1}^{1/2} \\ &= \boldsymbol{\Theta}_1^{1/2}\frac{1}{\sqrt{n_{\mathrm{T}}}}\boldsymbol{W}_1\mathbf{1} \\ &= \boldsymbol{\Theta}_1^{1/2}\frac{1}{\sqrt{n_{\mathrm{T}}}}[\boldsymbol{h}\;\boldsymbol{h}\;\cdots\;\boldsymbol{h}]\end{aligned}$$

and thus, prior to applying receive correlations, the n_{T} columns of the channel matrix are identical and given by the vector $\boldsymbol{h}$. It is easy to see that the corresponding normalized conditional covariance $\boldsymbol{\Phi}_{\boldsymbol{n}}(\boldsymbol{H}_1)$ is identical to that of a single-antenna interferer with channel

$$\boldsymbol{H}_1 = \boldsymbol{\Theta}_1^{1/2}\boldsymbol{h}$$

and the same total power $\mathcal{I}_1$. The vector $\boldsymbol{h}$, in turn, is given by

$$\boldsymbol{h} = \left[\sum_{j=1}^{n_{\mathrm{T}}}(\boldsymbol{W}_1)_{1,j}\;\;\sum_{j=1}^{n_{\mathrm{T}}}(\boldsymbol{W}_1)_{2,j}\;\;\cdots\;\;\sum_{j=1}^{n_{\mathrm{T}}}(\boldsymbol{W}_1)_{n_{\mathrm{T}},j}\right]^T.$$

Since the entries of $\boldsymbol{W}_1$ are independent zero-mean unit-variance Gaussian random variables, the entries of $\boldsymbol{h}$ are likewise and Proposition 4 can be applied simply by setting $m_1 = 1$.

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On the limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 315–335, Mar. 1998.

[2] I. E. Telatar, "Capacity of multiple-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.

[3] M. Dohler and H. Aghvami, "A closed-form expression of MIMO capacity over ergodic narrowband channels," preprint, 2003.

[4] G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communications," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.

[5] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1273–1289, May 2001.

[6] S. Diggavi, "On achievable performance of spatial diversity fading channels," *IEEE Trans. Inform. Theory*, vol. 47, pp. 308–325, Jan. 2001.

[7] A. Grant, "Rayleigh fading multiple-antenna channels," *EURASIP J. Applied Signal Processing, Special Issue on Space-Time Coding (Part I)*, vol. 2002, no. 3, pp. 316–329, Mar. 2002.

[8] J. B. Andersen, "Array gain and capacity for known random channels with multiple element arrays at both ends," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 2172–2178, Nov. 2000.

[9] T. L. Marzetta and B. H. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.

[10] A. Lapidoth and S. M. Moser, "On the fading number of multiple-antenna systems," in *Proc. IEEE Intern. Symp. Inform. Theory*, Washington, DC, June 2001, pp. 110–111.

[11] P. J. Smith and M. Shafi, "On a Gaussian approximation to the capacity of a wireless MIMO link," in *Proc. Int. Conf. Communications (ICC'02)*, New York, Apr. 2002.

[12] M. Chiani, "Evaluating the capacity distribution of MIMO Rayleigh fading channels," in *Proc. IEEE Int. Symp. Advances in Wireless Commun.*, Vancouver, BC, Canada, 2002.

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

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

[15] A. Lozano, "Capacity-approaching rate function for layered multiple-antenna architectures," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 616–620, July 2003.

[16] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effects on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502–511, Mar. 2000.

[17] H. Bolcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, pp. 225–2234, Feb. 2002.

[18] M. A. Khalighi, J. M. Brossier, G. Jourdain, and K. Raoof, "On capacity of Ricean MIMO channels," in *Proc. Int. Conf. Personal, Indoor and Mobile Radio Communications (PIMRC'01)*, vol. A, San Diego, CA, Sept. 2001, pp. 150–154.

[19] C. Chuah, D. N. C. Tse, J. Kahn, and R. Valenzuela, "Capacity scaling in dual-antenna-array wireless systems," *IEEE Trans. Inform. Theory*, vol. 48, pp. 637–650, Mar. 2002.

[20] R. S. Blum, J. H. Winters, and N. R. Sollenberger, "On the capacity of cellular systems with MIMO," in *IEEE Vehicular Technology Conf. (VTC'2001 Fall)*, vol. 2, Atlantic City, NJ, Oct. 2001, pp. 1220–1224.

[21] F. Rashid-Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-optimal space-time processing with multiple transmit and receive antennas," *IEEE Commun. Lett.*, vol. 5, pp. 85–87, Mar. 2001.

[22] A. Lozano and A. M. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Trans. Inform. Theory*, vol. 48, pp. 3117–3128, Dec. 2002.

[23] D. Bliss, K. Forsythe, and A. Yegulalp, "MIMO communication capacity using infinite dimension random matrix eigenvalue distributions," in *Proc. 35th Asilomar Conf.*, vol. 2, 2001, pp. 969–974.

[24] R. S. Blum, "MIMO capacity with interference," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 793–801, June 2003.

[25] 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," Bell Labs Tech. Memo., TM-02-42984X, May 2002.

[26] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: A bandwidth-efficient high-speed wireless data service for nomadic users," *IEEE Commun. Mag.*, vol. 38, pp. 70–77, July 2000.

[27] 3rd Generation Partnership Project; Technical Specification Group Radio Access Newtork, "UTRA High Speed Downlink Packet Access," Tech. Rep., 3G TR 25.950, Mar. 2001.

[28] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1319–1343, June 2002.

[29] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high capacity space-time wireless," in *Proc. 37th Annual Allerton Conf. Communication, Control, Computing*, Monticello, IL, Sept. 1999.

[30] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, pp. 549–553, Apr. 2003.

[31] Q. Sun, D. C. Cox, H. C. Huang, and A. Lozano, "Estimation continuous fading BLAST channels," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 951–963, Oct. 2002.

[32] W. C. Y. Lee and Y. S. Yeh, "Polarization diversity system for mobile radio," *IEEE Trans. Commun.*, vol. COM-20, pp. 912–923, Oct. 1972.

[33] J. A. Lempiainen and J. K. Laiho-Steffens, "The performance of polarization diversity schemes at a base station in small/micro cells at 1800 MHz," *IEEE Trans. Veh. Technol.*, vol. 47, pp. 1087–1092, Aug. 1999.

[34] R. G. Vaughan, "Polarization diversity in mobile communications," *IEEE Trans. Veh. Technol.*, vol. 39, pp. 177–186, Aug. 1990.

[35] H. Bolcskei, R. Nabar, V. Erceg, D. Gesbert, and A. Paulraj, "Performance of multi-antenna signaling strategies in the presence of polarization diversity," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'01)*, vol. 4, Salt Lake City, UT, May 2001, pp. 2437–2440.

[36] D. Chizhik, J. Ling, P. Wolniansky, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple input multiple output measurements and modeling in Manhattan," *J. Select. Areas Commun. (Special Issue on MIMO Systems and Applications)*, vol. 21, pp. 321–331, Apr. 2001.

[37] P. Soma, D. S. Baum, V. Erceg, R. Krishnamoorthy, and A. Paulraj, "Analysis and modeling of multiple-input multiple-output (MIMO) radio channel based on outdoor measurements conducted at 2.5 GHz for fixed BWA applications," in *Proc. Int. Conf. Communications (ICC'02)*, New York, Apr. 2002.

[38] P. Kyritsi, D. C. Cox, R. A. Valenzuela, and P. W. Wolniansky, "Effect of antenna polarization on the capacity of a multiple element system in an indoor environment," *J. Select. Areas Commun.*, vol. 20, pp. 1227–1239, Aug. 2002.

[39] T. L. Marzetta, "Fundamental limitations on the capacity of wireless links that use polarimetric antenna arrays," in *Proc. IEEE Int. Symp. Infomation Theory (ISIT'2002)*, Lausanne, Switzeland, July 2002, p. 51.

[40] A. Moustakas, H. Baranger, L. Balents, A. Sengupta, and S. Simon, "Communication through a diffusive medium: Coherence and capacity," *Science*, vol. 287, pp. 287–290, Jan. 2000.

[41] D. Gesbert, H. Bolcskei, D. Gore, and A. J. Paulraj, "MIMO wireless channels: Capacity and performance prediction," in *Proc. IEEE GLOBECOM'2000*, San Francisco, CA, Dec. 2000, pp. 1083–1088.

[42] W. C. Jakes, *Microwave Mobile Communications*. New York: IEEE Press, 1974.

[43] J. Salz and J. H. Winters, "Effect of fading correlation on adaptive arrays in digital mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 1049–1057, Nov. 1994.

[44] K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, "Spatial channel characteristics in outdoor environments and their impact on BS antenna system performance," in *Proc. IEEE Vehicular Technology Conf. (VTC'98)*, vol. 2, 1998, pp. 719–723.

[45] J. Fuhl, A. F. Molisch, and E. Bonek, "Unified channel model for mobile radio systems with smart antennas," *Proc. Inst. Elec. Eng.–Radar, Sonar Navigation*, vol. 145, pp. 32–41, 1998.

[46] C. B. Dietrich Jr, K. Dietze, J. R. Nealy, and W. L. Stutzman, "Spatial, polarization, and pattern diversity for wireless handheld terminals," *IEEE Trans. Antennas Propagat.*, vol. 49, pp. 1271–1281, Sept. 2001.

[47] V. Jungnickel, V. Pohl, and C. von Helmolt, "Experiments on the element spacing in multi-antenna systems," in *Proc. IEEE Vehicular Technology Conf. (VTC'03 Spring)*, Jeju, Korea, 2003.

[48] D. Chizhik, F. R. Farrokhi, J. Ling, and A. Lozano, "Effect of antenna separation on the capacity of BLAST in correlated channels," *IEEE Commun. Lett.*, vol. 4, pp. 337–339, Nov. 2000.

[49] K. I. Pedersen, J. B. Andersen, J. P. Kermoal, and P. E. Mogensen, "A stochastic multiple-input multiple-output radio channel model for evaluations of space-time coding algorithms," in *Proc. IEEE Vehicular Technology Conf. (VTC'2000 Fall)*, vol. 2, Boston, MA, 2000, pp. 893–897.

[50] C. C. Martin, J. H. Winters, and N. R. Sollenberger, "Multiple-input multiple-output (MIMO) radio channel measurements," in *Proc. IEEE Vehicular Technology Conf. (VTC'2000)*, vol. 2, Boston, MA, Sept. 2000, pp. 774–779.

[51] J. P. Kermoal, P. E. Mogensen, S. H. Jensen, J. B. Andersen, F. Frederiksen, T. B. Sorensen, and K. I. Pedersen, "Experimental investigation of multipath richness for multi-element transmit and receive antenna arrays," in *Proc. IEEE Vehicular Technology Conf. (VTC'2000 Spring)*, 2000.

[52] H. Xu, M. J. Gans, N. Amitay, and R. A. Valenzuela, "Experimental verification of MTMR system capacity in a controlled propagation environment," *Electron. Lett.*, vol. 37, p. , July 2001.

[53] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO channel model with experimental validation," *IEEE J. Select. Areas Commun.*, vol. 20, pp. 1211–1226, Aug. 2002.

[54] D. Chizhik, G. J. Foschini, M. J. Gans, and R. A. Valenzuela, "Propagation and capacities of multi-element transmit and receive antennas," in *Proc. 2001 IEEE AP-S Int. Symp. USNC/URSI National Radio Science Meeting*, vol. 1, Boston, MA, July 2001, pp. 438–441.

[55] J. Ling, D. Chizhik, P. W. Wolniansky, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple transmit multiple receive (MTMR) capacity survey in Manhattan," *Electron. Lett.*, vol. 37, no. 16, pp. 1041–1042, Aug. 2001.

[56] V. Erceg, P. Soma, D. S. Baum, and A. Paulraj, "Capacity obtained from multiple-input multiple-output channel measurements in fixed wireless environments at 2.5 GHz," in *Proc. Int. Conf. Communications (ICC'02)*, vol. 1, New York, Apr. 2002, pp. 396–400.

[57] P. Driessen and G. J. Foschini, "On the capacity formula for multiple-input multiple-output channels: A geometric interpretation," *IEEE Trans. Commun.*, vol. 47, pp. 173–176, Feb. 1999.

[58] S. Rice, "Mathematical analysis of random noise," *Bell Syst. Tech. J.*, 1944.

[59] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 725–730, Nov. 1972.

[60] G. J. Foschini, "Layered space-time architecture for wireless communications in a fading environment when using multi-element antennas," *AT&T Bell Labs. Tech. J.*, pp. 41–59, 1996.

[61] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1990.

[62] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2148–2177, Oct. 1998.

[63] A. Lozano, A. M. Tulino, and S. Verdú, "Correlation number: A new design criterion in multi-antenna communication," in *Proc. IEEE Vehicular Technology Conf. (VTC'2003) Spring*, Jeju, Korea, Apr. 2003.

[64] ——, "Multi-antenna capacity in the low-power regime," Bell Labs. Tech. Memo., 10009652-020916-13TM, Sept. 2002.

[65] A. M. Tulino, S. Verdú, and A. Lozano, "Capacity of antenna arrays with space, polarization and pattern diversity," in *Proc. Information Theory Workshop (ITW'03)*, Paris, France, Mar. 2003.

[66] S. A. Bergmann and H. W. Arnold, "Polarization diversity in portable communications environment," *Electron. Lett.*, vol. 22, no. 11, pp. 609–610, May 1986.

[67] Z. D. Bai, B. Q. Miao, and J. Tsay, "Convergence rates of the spectral distributions of large Wigner matrices," *Int. Math. J.*, vol. 1, no. 1, pp. 65–90, 2002.

[68] L. Hanlen and M. Fu, "Multiple antenna wireless communication systems: Limits to capacity growth," in *Proc. Wireless Communications and Networking Conf. (WCNC'99)*, vol. 1, Mar. 2002, pp. 172–176.

[69] J. H. Winters, "Optimum combining in digital mobile radio with cochannel interference," *IEEE J. Select. Areas Commun.*, vol. JSAC-2, pp. 528–539, July 1984.

[70] F. Hiai and D. Petz, "Asymptotic freeness almost everywhere for random matrices," *Acta Sci. Math. (Szeged)*, vol. 66, pp. 801–826, 2000.

[71] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. New York: Academic Press, 1980. Corrected and enlarged edition.