

High-SNR Power Offset in Multiantenna Communication

Angel Lozano*, Antonia M. Tulino[†] and Sergio Verdú[‡]

August 16, 2005

Abstract

The analysis of the multiantenna capacity in the high-SNR regime has hitherto focused on the high-SNR *slope* (or maximum *multiplexing gain*), which quantifies the multiplicative increase as function of the number of antennas. This traditional characterization is unable to assess the impact of prominent channel features since, for a majority of channels, the slope equals the minimum of the number of transmit and receive antennas. Furthermore, a characterization based solely on the slope captures only the scaling but it has no notion of the power required for a certain capacity. This paper advocates a more refined characterization whereby, as function of $\text{SNR}|_{\text{dB}}$, the high-SNR capacity is expanded as an affine function where the impact of channel features such as antenna correlation, unfaded components, etc, resides in the zero-order term or *power offset*. The power offset, for which we find insightful closed-form expressions, is shown to play a chief role for SNR levels of practical interest.

Keywords: Channel capacity; Multiantenna arrays; High-SNR analysis; Fading channels; Antenna correlation; Ricean channels; Coherent communication;

*Angel Lozano is with Bell Laboratories (Lucent Technologies), Holmdel, NJ07733, USA.

[†]Antonia M. Tulino is with Universita Degli Studi di Napoli "Federico II", 80125 Napoli, Italy

[‡]Sergio Verdú is with Princeton University, Princeton, NJ08544, USA.

I Introduction

A Motivation

The observation was first made in [1, 2] that, in the high-power regime, the single-user multiantenna capacity with coherent reception behaves as

$$C(\text{SNR}) = \min(n_T, n_R) \log_2 \text{SNR} + O(1) \quad (1)$$

where n_T and n_R denote the number of transmit and receive antennas while SNR represents the signal-to-noise ratio. The linear scaling of the leading term with $\min(n_T, n_R)$ (variously referred to as the high-SNR *slope*, the number of *degrees of freedom* [3], the maximum *multiplexing gain* or the *pre-log*) was instrumental in firmly establishing the potential of multiantenna communication. Although the early analysis supporting (1) considered a canonical channel having zero-mean IID (independent identically distributed) Gaussian entries, it turns out that this behavior holds under rather general channel conditions [4, 5].¹

Some of the lessons that emerge from a traditional analysis of the coherent capacity at high SNR based on (1) are:

- Only $\min(n_T, n_R)$ is relevant. Any increase in $\max(n_T, n_R)$ is immaterial.
- In order to retain the multiplexing gain in the presence of known antenna correlations, it suffices to transmit nonzero power along every eigenvector outside the null space of the correlation matrix of the transmit array.
- The impact of correlation depends exclusively on the rank of the transmit and receive correlation matrices. Provided these matrices are nonsingular, as is usually the case in practice, correlation has no effect.
- The capacity is not altered by the existence of Ricean (unfaded) channel components.

In order to illustrate why these conclusions are unwarranted, Fig. 1 depicts the ergodic mutual information achieved by an isotropic Gaussian input on an IID Rayleigh-faded channel and on a Rayleigh-faded channel exhibiting antenna correlation. Although both channels have the same high-SNR slope, the power levels required to achieve a given mutual information at high SNR may be over 10 dB apart.

¹Coherent operation may be unfeasible for $\text{SNR} \rightarrow \infty$ when the noise becomes commensurate with the uncertainty in the knowledge of the precise state of the channel and the logarithmic scaling in (1) breaks down [6, 7]. For SNR levels of practical interest, however, the premise of coherent reception is valid in a vast majority of channels and an asymptotic characterization accurately captures the behavior of the capacity at those SNR levels.

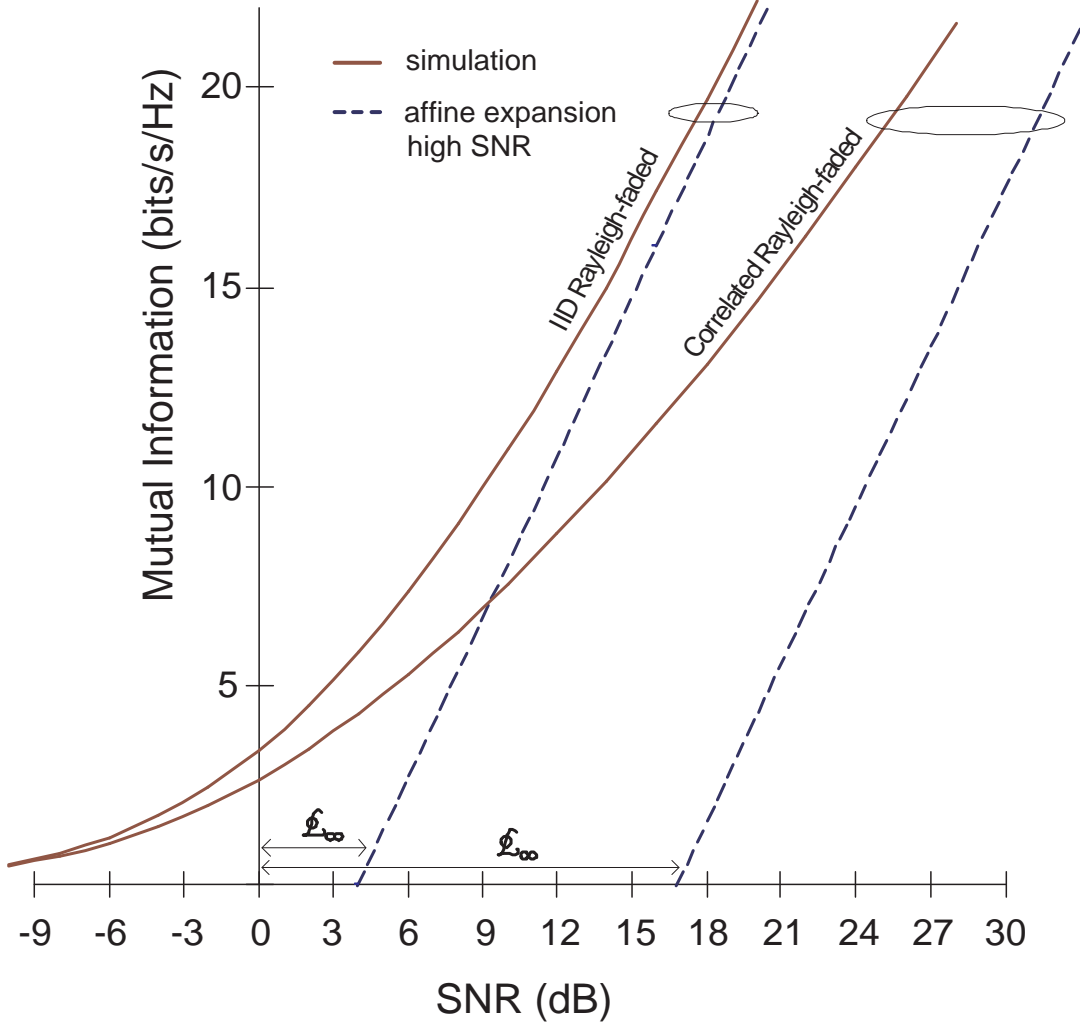


Figure 1: For $n_T = n_R = 4$, mutual information achieved by an isotropic input in an IID Rayleigh-faded channel and in a correlated Rayleigh-faded channel (cf. Example 3). The solid lines indicate simulation while the dashed lines indicate the high-SNR expansion with $S_\infty = 4$ bits/s/Hz/(3 dB) and with the power offset obtained from Proposition 4.

The objective of this paper is to show how each of the above statements is amended—in some cases quite drastically—by a more refined analysis. To that end, we evaluate the zero-order term in the high-SNR expansion of the capacity as an affine function of $\text{SNR}|_{\text{dB}}$, for which we coin the term *power offset*.² As will be shown, in the high-SNR regime most channels have the same slope yet their capacities may be very different because of sizeable disparities in the power offset.

The paper is organized as follows. The remainder of this section introduces the basic model and quantities to be utilized throughout. In Section II, the two fundamental high-SNR performance measures, namely the slope and the power offset, are defined. As an instructive exercise, in Section III these measures are evaluated for the canonical channel, for which a number of prior analyses are available. Section IV, in turn, deals with more general channels that incorporate antenna correlation. The analysis of the power offset on these channels yields valuable insight—that could not be drawn from the slope—on the impact of correlation and on the capacity-achieving input at high SNR. Finally, Sections V and VI present results for other channels of interest: Ricean channels and interference-limited channels. Various implications of the analysis, highlighted throughout the paper, are brought together in Section VII. Moreover, the applicability of the results to high-SNR levels of practical interest is discussed. Proofs are relegated to the appendices.

B Definitions and Notation

With frequency-flat fading,³ the baseband complex model we consider is

$$\mathbf{y} = \sqrt{g} \mathbf{H} \mathbf{x} + \mathbf{n} \quad (2)$$

where \mathbf{x} and \mathbf{y} are the input and output vectors while \mathbf{n} is AWGN (additive white Gaussian noise). The channel, both stationary and ergodic, is represented by the $n_{\text{R}} \times n_{\text{T}}$ random matrix $\sqrt{g} \mathbf{H}$ where the entries of \mathbf{H} are jointly Gaussian while the real-valued scalar g is such that

$$E[\text{Tr}\{\mathbf{H}\mathbf{H}^\dagger\}] = n_{\text{R}}n_{\text{T}}. \quad (3)$$

The receiver has access to the realization of \mathbf{H} whereas the transmitter has only access to its distribution. Therefore, temporally IID input vectors are optimum.

The spatial covariance of these input vectors, conveniently normalized by the energy per

²Notation: $x|_{\text{dB}} = 10 \log_{10} x$.

³If the fading process is frequency selective, the channel can be decomposed into a number of parallel noninteracting subchannels, each sufficiently narrow to experience frequency-flat fading and with the same ergodic capacity as the aggregate channel.

dimension, is denoted by

$$\mathbf{\Phi} = \frac{E[\mathbf{x}\mathbf{x}^\dagger]}{\frac{1}{n_T}E[\|\mathbf{x}\|^2]} \quad (4)$$

where the normalization ensures that $\text{Tr}\{\mathbf{\Phi}\}=n_T$. With that, an isotropic input maps onto $\mathbf{\Phi} = \mathbf{I}$. The ergodic mutual information (average rate per unit bandwidth) achieved by a zero-mean Gaussian input, in bits/s/Hz, is

$$\mathcal{I}(\text{SNR}, \mathbf{\Phi}) = E \left[\log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \mathbf{H}\mathbf{\Phi}\mathbf{H}^\dagger \right) \right] \quad (5)$$

with

$$\text{SNR} = g \frac{E[\|\mathbf{x}\|^2]}{\frac{1}{n_R}E[\|\mathbf{n}\|^2]} \quad (6)$$

The maximization of (5) over $\mathbf{\Phi}$ yields the ergodic capacity.

For notational compactness, we introduce specific symbols to indicate the smaller and larger of the numbers of transmit and receive antennas:

$$n_\downarrow = \min(n_T, n_R) \quad (7)$$

$$n_\uparrow = \max(n_T, n_R) \quad (8)$$

II High-SNR Measures: S_∞ and \mathcal{L}_∞

At high SNR, the ergodic mutual information achieved with input covariance $\mathbf{\Phi}$ behaves as

$$\mathcal{I}(\text{SNR}, \mathbf{\Phi}) = S_\infty(\mathbf{\Phi}) \left(\frac{\text{SNR}|_{\text{dB}}}{3 \text{ dB}} - \mathcal{L}_\infty(\mathbf{\Phi}) \right) + o(1) \quad (9)$$

where $3 \text{ dB} = 10 \log_{10} 2$ while $S_\infty(\mathbf{\Phi})$ denotes the high-SNR slope in bits/s/Hz/(3 dB),

$$S_\infty(\mathbf{\Phi}) = \lim_{\text{SNR} \rightarrow \infty} \frac{\mathcal{I}(\text{SNR}, \mathbf{\Phi})}{\log_2 \text{SNR}}. \quad (10)$$

The quantity $\mathcal{L}_\infty(\mathbf{\Phi})$, in turn, represents the zero-order term or *power offset*, in 3-dB units, with respect to a reference channel having the same high-SNR slope but with unfaded and orthogonal dimensions (i.e., such that $\frac{1}{n_T}\mathbf{H}\mathbf{H}^\dagger=\mathbf{I}$) whose expansion in (9) intersects the origin at $\text{SNR}|_{\text{dB}} = 0$. The power offset, which anchors the expansion, is given by

$$\mathcal{L}_\infty(\mathbf{\Phi}) = \lim_{\text{SNR} \rightarrow \infty} \left(\log_2 \text{SNR} - \frac{\mathcal{I}(\text{SNR}, \mathbf{\Phi})}{S_\infty} \right) \quad (11)$$

in 3-dB units.⁴ This performance measure was first introduced in [8], in the context of CDMA with random spreading, where it was shown to be equivalent to the similarly defined measure as function of the energy per information bit (instead of SNR).⁵

While $\mathcal{L}_\infty(\Phi)$ can be computed for any input covariance Φ , we distinguish the power offset for two particularly relevant inputs as follows:

- The power offset corresponding to the capacity-achieving input covariance is simply indicated by \mathcal{L}_∞ , obviating the dependence on Φ .
- The power offset corresponding to an isotropic input, whenever such input is not capacity-achieving, is indicated by \mathcal{L}_∞ .

III The IID Rayleigh-faded Channel

The high-SNR behavior of a scalar Rayleigh-faded channel is well established. The slope is $S_\infty = 1 \text{ bit/s/Hz}/(3 \text{ dB})$ while the power offset in 3-dB units equals [8, 9]

$$\mathcal{L}_\infty = \gamma \log_2 e \quad (12)$$

where γ is the Euler-Mascheroni constant

$$\gamma = \lim_{n \rightarrow \infty} \left(\sum_{\ell=1}^n \frac{1}{\ell} - \log_e n \right) \quad (13)$$

$$\approx 0.5772 \quad (14)$$

Thus, \mathcal{L}_∞ amounts to about 2.5 dB in a scalar Rayleigh-faded channel. In a scalar unfaded channel, in contrast, $\mathcal{L}_\infty = 0$. While the slope is unaffected by fading, the difference between the corresponding power offsets reflects the cost of fading in terms of power.

The characterization extends straightforwardly to the canonical multiantenna channel with IID Rayleigh-faded entries, for which the capacity-achieving input is isotropic and the slope is $S_\infty = n_\psi$ [1].

⁴Replacing the basis of the binary $\log(\cdot)$ in (11) and (5) with 10, we would obtain $\mathcal{L}_\infty(\Phi)$ in units of Bels. Further multiplication by 10 would yield dB. In fact, dB units are used to quantify the power offset in some numerical examples, but all of the analytical expressions in the paper are given in 3-dB units.

⁵The zero-order counterpart of \mathcal{L}_∞ for channels unknown to the receiver is the *fading number* introduced in [7]. In contrast to the coherent case that occupies us here, though, in the noncoherent case the leading term in the high-SNR expansion is not necessarily linear in $\text{SNR}|_{\text{dB}}$ and, as a result, the interpretation of the fading number does not mirror that of the power offset.

Proposition 1 For an IID Rayleigh-faded channel,

$$\mathcal{L}_\infty = \log_2 n_T + \left(\gamma - \sum_{\ell=1}^{n_\uparrow - n_\downarrow} \frac{1}{\ell} - \frac{n_\uparrow}{n_\downarrow} \sum_{\ell=n_\uparrow - n_\downarrow + 1}^{n_\uparrow} \frac{1}{\ell} + 1 \right) \log_2 e \quad (15)$$

which, for $n_T = n_R = n$, reduces to

$$\mathcal{L}_\infty = \log_2 n + \left(\gamma - \sum_{\ell=2}^n \frac{1}{\ell} \right) \log_2 e \quad (16)$$

Proof: See Appendix B.

Example 1 Let $n_T = n_R = 4$. The high-SNR expansion in (9), with $S_\infty = 4$ and with \mathcal{L}_∞ computed via Proposition 1, is plotted in Fig. 1 alongside a corresponding Montecarlo simulation of $\mathcal{I}(\text{SNR}, \mathbf{I})$. The asymptotic behavior is evident at rather modest SNR levels.

The high-SNR expansion of the capacity obtained using the power offset expressions in Proposition 1 had already been given by several authors, either as an approximation [10, Eq. 10]–[11, Prop. 2] or a lower bound [12, Eq. 9]. (A distinct upper bound was given in [13, Eq. 23].)

Proposition 2 Consider an IID channel whose entries are zero-mean and unit-variance but otherwise arbitrarily distributed. As the number of antennas grows with ratio $\beta = \frac{n_T}{n_R}$, the power offset converges to

$$\mathcal{L}_\infty \rightarrow \begin{cases} (\beta - 1) \log_2 \frac{\beta - 1}{\beta} + \log_2 e & \beta > 1 \\ \log_2 e & \beta = 1 \\ \frac{1 - \beta}{\beta} \log_2(1 - \beta) + \log_2(\beta e) & \beta < 1 \end{cases} \quad (17)$$

Proof: See Appendix C.

Note that, in contrast with Proposition 1, which is specific to Rayleigh fading, the validity of Proposition 2 extends to any fading distribution [14].

Using Propositions 1 and 2, it is simple to unveil the role played by n_\uparrow on the canonical channel. To this end, in the ensuing corollaries we write $\mathcal{L}_\infty(n_T, n_R)$ explicitly highlighting the dependence of the power offset on both n_T and n_R .

Corollary 1 Let $n_T = n_R = n$. Adding k receive antennas, while not altering S_∞ , would reduce the power offset as

$$\mathcal{L}_\infty(n, n + k) = \mathcal{L}_\infty(n, n) - \left(\sum_{\ell=n+1}^{n+k} \frac{1}{\ell} + \frac{k}{n} \sum_{\ell=k+1}^{n+k} \frac{1}{\ell} \right) \log_2 e \quad (18)$$

yielding a steady improvement with growing n_R because of the additional power captured by every new receive antenna. Once $n_R \gg n_T$, the gain approaches 3 dB for every doubling of n_R .

Corollary 2 Let $n_T = n_R = n$. Adding k transmit antennas, while not altering S_∞ , would reduce the power offset as

$$\mathcal{L}_\infty(n+k, n) = \mathcal{L}_\infty(n, n) + \log_2 \frac{n+k}{n} - \left(\sum_{\ell=n+1}^{n+k} \frac{1}{\ell} + \frac{k}{n} \sum_{\ell=k+1}^{n+k} \frac{1}{\ell} \right) \log_2 e \quad (19)$$

This improvement cannot exceed 1.5 dB for $k = 1$ and 4.3 dB for any k .

Relationships (18) and (19) are illustrated in Fig. 2.

Example 2 With respect to $n_T = n_R = 1$,

$$\mathcal{L}_\infty(1, 2) = \mathcal{L}_\infty(1, 1) - 4.3 \text{ dB} \quad (20)$$

$$\mathcal{L}_\infty(2, 1) = \mathcal{L}_\infty(1, 1) - 1.3 \text{ dB} \quad (21)$$

$$\mathcal{L}_\infty(\infty, 1) = \mathcal{L}_\infty(1, 1) - 2.5 \text{ dB} \quad (22)$$

Note that, as observed in [15], letting $n_T \rightarrow \infty$ with $n_R = 1$ exactly erases the high-SNR power penalty due to fading.

IV Correlated Rayleigh-faded Channels

Let us now extend the analysis to channels with transmit and receive correlations. We adhere to the widely used *separable* model whereby the correlation between the (i, j) and (k, ℓ) entries of \mathbf{H} is⁶

$$E [(\mathbf{H})_{i,j}(\mathbf{H})_{k,\ell}^*] = (\Theta_R)_{i,k}(\Theta_T)_{j,\ell}$$

where Θ_R and Θ_T are $n_R \times n_R$ and $n_T \times n_T$ correlation matrices whose entries indicate the correlation between receive antennas and between transmit antennas, respectively, while $(\cdot)_{i,j}$ denotes the (i, j) th entry of a matrix. A Rayleigh-faded channel whose correlation is separable can thus be expressed as

$$\mathbf{H} = \Theta_R^{1/2} \mathbf{W} \Theta_T^{1/2} \quad (23)$$

with \mathbf{W} having IID unit-variance Rayleigh-faded entries.

⁶For more general correlation structures, expressions for S_∞ and \mathcal{L}_∞ in the large-dimensional limit ($n_T, n_R \rightarrow \infty$) are given in [16].

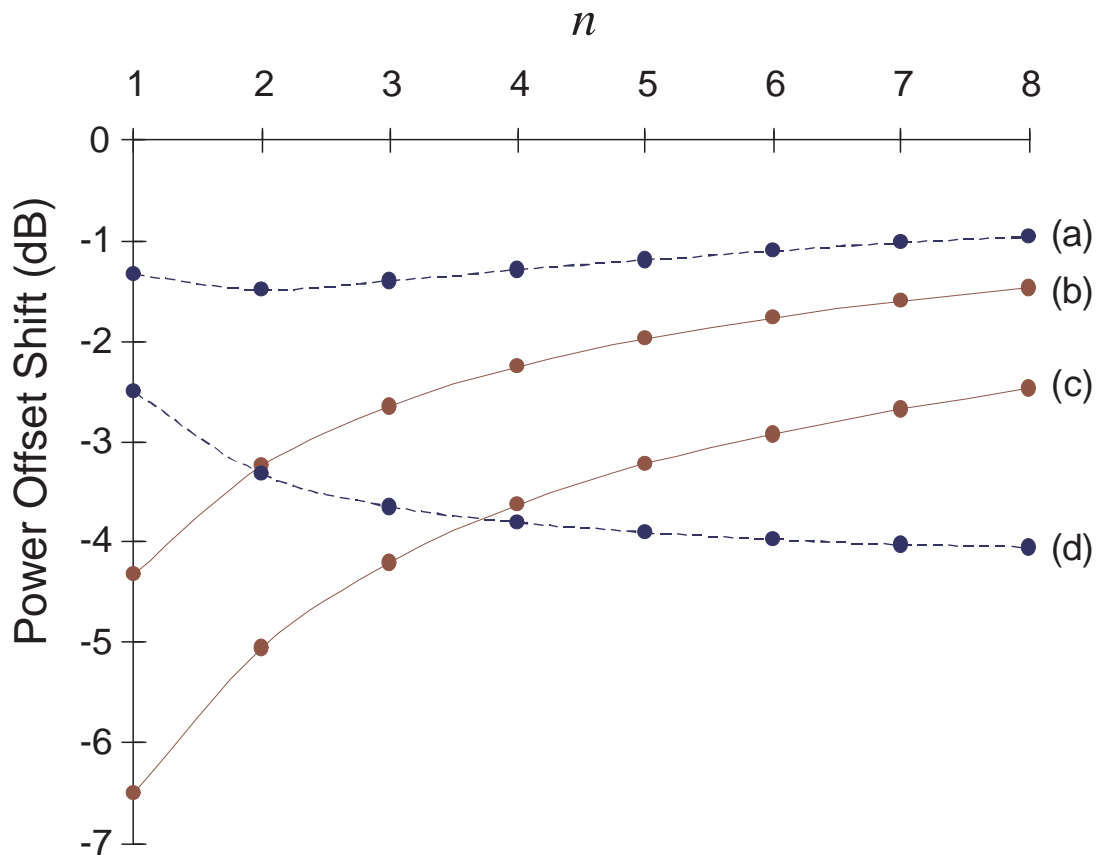


Figure 2: High-SNR power offset shift, in dB, obtained by adding either (a) one antenna to the transmitter, (b) one antenna to the receiver, (c) two antennas to the receiver, or (d) by letting $n_T \rightarrow \infty$.

For the channel in (23), achieving capacity when the transmitter knows the distribution but not the realization of \mathbf{H} requires that signalling take place on the eigenspace of Θ_{T} [17]–[22]. We therefore consider only input covariances of the form $\Phi = \mathbf{V}\mathbf{P}\mathbf{V}^\dagger$ where \mathbf{V} contains the eigenvectors of Θ_{T} while $\mathbf{P} = \text{diag}\{p_1, p_2, \dots, p_{n_{\text{T}}}\}$ where p_j denotes the power allocated to the j th such eigenvector.

We shall further make use of the following notion.

Definition 1 *An input covariance Φ is deemed admissible to the channel model in (23) if the null space of Φ is contained within the null space of Θ_{T} .*

Note that the identity is admissible to any channel that conforms to (23).

A The High-SNR Slope

Definition 2 *The ranks of Θ_{T} and Θ_{R} are denoted, respectively, by n'_{T} and n'_{R} .*

In terms of the high-SNR slope, n'_{T} and n'_{R} can be regarded as the number of effective antennas in a correlated channel. This point is formalized by the proceeding result.

Proposition 3 *For the Rayleigh-faded channel in (23),*

$$S_\infty(\Phi) \leq \min(n'_{\text{T}}, n'_{\text{R}}) \quad (24)$$

with equality if and only if Φ is admissible.

Proof: *See Appendix D.*

With this result, we can now formalize some of the points made at the outset of the paper:

- The largest achievable high-SNR slope is $\min(n'_{\text{T}}, n'_{\text{R}})$, irrespective of $\max(n'_{\text{T}}, n'_{\text{R}})$.
- This slope is achieved under the mild condition that the input covariance be admissible.
- Θ_{T} and Θ_{R} are reflected on the slope only through their rank.

As discussed in detail in the conclusions, some of these observations should be nuanced with a more pragmatic finite-SNR interpretation of the slope whereby n'_{T} and n'_{R} are defined as the number of nonnegligible (rather than strictly nonzero) eigenvalues of Θ_{T} and Θ_{R} . This interpretation, nonetheless, still lacks a connection between the correlations and the

required power for a certain capacity. The role of the power offset is precisely to establish such connection and to amend the assessments provided by the slope. Accordingly, we shall henceforth consider only admissible input covariances. Thus, $S_\infty = \min(n'_T, n'_R)$ and we drop its dependence on Φ .

B The High-SNR Power Offset

Only the nonzero eigenvalues of the correlation matrices turn out to be relevant to the power offset. Denoting by Λ_R and Λ_T the diagonal eigenvalue matrices of Θ_R and Θ_T ,⁷ we define corresponding $n'_R \times n'_R$ and $n'_T \times n'_T$ diagonal matrices Λ'_R and Λ'_T obtained by striking out the empty rows and columns of Λ_R and Λ_T , respectively.

It is also convenient to introduce an $m \times m$ matrix-valued function $\Psi_i(m, \Lambda)$, $i \in \{1, \dots, m\}$, whose (k, ℓ) th entry is

$$(\Psi_i)_{k, \ell} = \nu_{n-m+k} \lambda_{n-m+k}^{n-m-1+\ell} - \sum_{q=1}^{n-m} \nu_q \lambda_q^{n-m-1+\ell} \sum_{d=1}^{n-m} (\Upsilon^{-1})_{d, q} \lambda_{n-m+k}^{d-1} \quad (25)$$

where

$$\nu_q = \begin{cases} 1 & \ell \neq i \\ \sum_{u=1}^{\ell-1} \frac{1}{u} - \gamma + \log_e \lambda_q & \ell = i \end{cases} \quad (26)$$

while λ_j is the j th diagonal entry of the $n \times n$ matrix argument Λ . In turn, $\Upsilon(m, \Lambda)$ is the $(n-m) \times (n-m)$ principal submatrix of the Vandermonde matrix

$$\Omega(\Lambda) = \begin{bmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \cdots & \lambda_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_n & \cdots & \lambda_n^{n-1} \end{bmatrix} \quad (27)$$

Using these definitions, we can put forth the following result.

Proposition 4 Consider any input covariance $\Phi = \mathbf{V}\mathbf{P}\mathbf{V}^\dagger$ where \mathbf{V} is such that $\Theta_T = \mathbf{V}\Lambda_T\mathbf{V}^\dagger$ and denote by \mathbf{P}' the $n'_T \times n'_T$ matrix obtained by striking from \mathbf{P} the same rows and columns that

⁷Since the eigenspaces of Θ_T and Φ are common, the diagonal entries of Λ_T and \mathbf{P} should be ordered equally.

were removed from Λ_T to obtain Λ'_T . Then,

$$\mathcal{L}_\infty(\Phi) = \begin{cases} \log_2 n_T - \frac{1}{n'_R} \log_2 \det \Lambda'_R - \frac{\det \Upsilon(n'_R, \Lambda'_T \mathbf{P}')}{\det \Omega(\Lambda'_T \mathbf{P}')} \frac{\log_2 e}{n'_R} \sum_{i=1}^{n'_R} \det \Psi_i(n'_R, \Lambda'_T \mathbf{P}') & n'_T > n'_R \\ \log_2 n_T + \left(\gamma - \sum_{\ell=2}^{n'_T} \frac{1}{\ell} \right) \log_2 e - \frac{1}{n'_T} \log_2 \det(\Lambda'_R \Lambda'_T \mathbf{P}') & n'_T = n'_R \\ \log_2 n_T - \frac{1}{n'_T} \log_2 \det(\Lambda'_T \mathbf{P}') - \frac{\det \Upsilon(n'_T, \Lambda'_R)}{\det \Omega(\Lambda'_R)} \frac{\log_2 e}{n'_T} \sum_{i=1}^{n'_T} \det \Psi_i(n'_T, \Lambda'_R) & n'_T < n'_R \end{cases} \quad (28)$$

For an isotropic input, \mathcal{L}_∞ is given by the right-hand side of (28) with \mathbf{P}' replaced by the $n'_T \times n'_T$ identity matrix.

Proof: See Appendix E.

Note that, if any of the nonzero eigenvalues of either Θ_R or Θ_T has plural multiplicity, in some of the terms in (28) both numerator and denominator vanish but their ratio remains well defined.

An example of the applicability of Proposition 4 can be given by referring back to Fig. 1.

Example 3 Let $n_T = n_R = 4$ and consider the nonsingular transmit and receive correlation matrices

$$\Theta_T = \Theta_R \quad (29)$$

$$= \begin{bmatrix} 1 & 0.76 & 0.21 & -0.27 \\ 0.76 & 1 & 0.76 & 0.21 \\ 0.21 & 0.76 & 1 & 0.76 \\ -0.27 & 0.21 & 0.76 & 1 \end{bmatrix} \quad (30)$$

For an isotropic input, the high-SNR expansion of the capacity with $S_\infty = 4$ and with \mathcal{L}_∞ obtained from Proposition 4 is depicted in Fig. 1 alongside a corresponding Montecarlo simulation of $\mathcal{I}(\text{SNR}, \mathbf{I})$.

As we illustrate next, Proposition 4 leads to compact expressions when the number of effective antennas is modest.

Corollary 3 Let $n'_T = n'_R = 1$. Proposition 4 yields

$$\mathcal{L}_\infty(\Phi) = \gamma \log_2 e - \log_2(p'_1 n_R) \quad (31)$$

The power offset in (31) evidences the advantage of having multiple antennas even if they are fully correlated. This advantage would be overlooked by a high-SNR analysis that relied solely on \mathcal{S}_∞ .

Corollary 4 Let $n'_T = 2$ and $n'_R = 1$ with λ'_1 and λ'_2 the nonzero eigenvalues of Θ_T . Then,

$$\mathcal{L}_\infty(\Phi) = 1 + \gamma \log_2 e - \log_2 n_R - \frac{p'_2 \lambda'_2 \log_2(p'_2 \lambda'_2) - p'_1 \lambda'_1 \log_2(p'_1 \lambda'_1)}{p'_2 \lambda'_2 - p'_1 \lambda'_1} \quad (32)$$

Corollary 5 Let $n'_T = 1$ and $n'_R = 2$ with λ'_1 and λ'_2 the nonzero eigenvalues of Θ_R . Then,

$$\mathcal{L}_\infty(\Phi) = \gamma \log_2 e - \log_2 p'_1 - \frac{\lambda'_2 \log_2 \lambda'_2 - \lambda'_1 \log_2 \lambda'_1}{\lambda'_2 - \lambda'_1} \quad (33)$$

Corollary 6 Let $n'_T = 1$ and $n'_R = n_R = 3$ with λ_1 , λ_2 and λ_3 the nonzero eigenvalues of Θ_R . Then,

$$\mathcal{L}_\infty(\Phi) = \gamma \log_2 e - \log_2 p'_1 - \frac{(\lambda_2 - \lambda_1) \lambda_3^2 \log_2 \lambda_3 + (\lambda_3 - \lambda_2) \lambda_1^2 \log_2 \lambda_1 + (\lambda_1 - \lambda_3) \lambda_2^2 \log_2 \lambda_2}{(\lambda_2 - \lambda_1) \lambda_3^2 + (\lambda_3 - \lambda_2) \lambda_1^2 + (\lambda_1 - \lambda_3) \lambda_2^2} \quad (34)$$

If $\lambda_2 = \lambda_3 = \lambda$, then (34) becomes

$$\mathcal{L}_\infty(\Phi) = \gamma \log_2 e - \log_2 p'_1 - \frac{\lambda_1^2 \log_2 \lambda_1 + \lambda^2 \log_2(\lambda e) - \lambda_1 \lambda \log_2(\lambda^2 e)}{(\lambda_1 - \lambda)^2} \quad (35)$$

Corollary 7 Let $n'_T = n'_R = 2$ with $\lambda'_j(\cdot)$ the j th nonzero eigenvalue of a matrix. Then,

$$\mathcal{L}_\infty(\Phi) = 1 + (\gamma - \frac{1}{2}) \log_2 e - \frac{\log_2(p'_1 \lambda'_1(\Theta_T) \lambda'_1(\Theta_R)) + \log_2(p'_2 \lambda'_2(\Theta_T) \lambda'_2(\Theta_R))}{2} \quad (36)$$

Although, in general, the expressions in Proposition 4 become progressively involved as the number of antennas grows large, it can be shown using random matrix theory that they crystalize into compact fixed-point equations.

Proposition 5 For $n_T, n_R \rightarrow \infty$ such that n'_T and n'_R also grow without bound, with ratio $\beta' = n'_T/n'_R$,

$$\mathcal{L}_\infty(\Phi) \rightarrow \begin{cases} -E \left[\log_2 \frac{\Lambda'_R}{\alpha \beta' e} \right] - \beta' E \left[\log_2 (1 + \alpha \Lambda') \right] & \beta' > 1 \\ -E \left[\log_2 \frac{\Lambda'_R \Lambda'}{e} \right] & \beta' = 1 \\ -E \left[\log_2 \frac{\varphi \Lambda'}{e} \right] - \frac{1}{\beta'} E \left[\log_2 \left(1 + \frac{\Lambda'_R}{\varphi} \right) \right] & \beta' < 1 \end{cases} \quad (37)$$

where α and φ are solutions to

$$E \left[\frac{1}{1 + \alpha \Lambda'} \right] = 1 - \frac{1}{\beta'} \quad (38)$$

$$E \left[\frac{1}{1 + \frac{\Lambda'_R}{\varphi}} \right] = 1 - \beta'. \quad (39)$$

The expectations are over Λ' and Λ'_R whose distributions equal the limiting ($n_T, n_R \rightarrow \infty$) empirical distributions of the diagonal entries of $\Lambda'_T \mathbf{P}'$ and Λ'_R , respectively.

The limiting value of \mathcal{L}_∞ is also given by (37), (38) and (39), except with Λ' distributed according to the limiting empirical distribution of the diagonal entries of Λ'_T .

Proof: See Appendix F.

The large-dimensional behavior in Proposition 5 settles rapidly and hence the corresponding solutions provide, as will be illustrated next, alternative means to evaluate the power offset that are accurate even for small numbers of antennas.

Example 4 The transmit array has $n_T = 4$ antennas arranged into a uniform linear array with d equal to the spacing (in wavelengths) between adjacent antennas. Given a broadside (truncated) Gaussian power azimuth spectrum with a 2° root-mean-square spread, the transmit antenna correlations are [23]

$$(\Theta_T)_{i,j} \approx e^{-0.05 d^2 (i-j)^2} \quad (40)$$

The receive array has $n_R = 2$ uncorrelated antennas. Fig. 3 depicts $\mathcal{I}(\text{SNR}, \mathbf{I})$ with an isotropic input and Rayleigh fading, computed via Monte Carlo simulation and parameterized by d . Also shown is the high-SNR expansion with $S_\infty = 2$ bits/s/Hz/(3 dB) and with \mathcal{L}_∞ computed using Proposition 4. A comparison between these exact values for \mathcal{L}_∞ and those obtained using Proposition 5, with the expectations replaced by arithmetic means over the nonzero eigenvalues of Θ_T , is presented in Table 1.

As in other mutual information analyses [24, 25, 26], we find an excellent agreement between the asymptotic ($n_T, n_R \rightarrow \infty$) results and the exact results for small numbers of antennas.

C Asymptotically Optimal Power Allocation

While the high-SNR slope is invariant to the power allocation as long as the input covariance is admissible, the power offset captures the sensitivity of the high-SNR capacity to the power allocation.

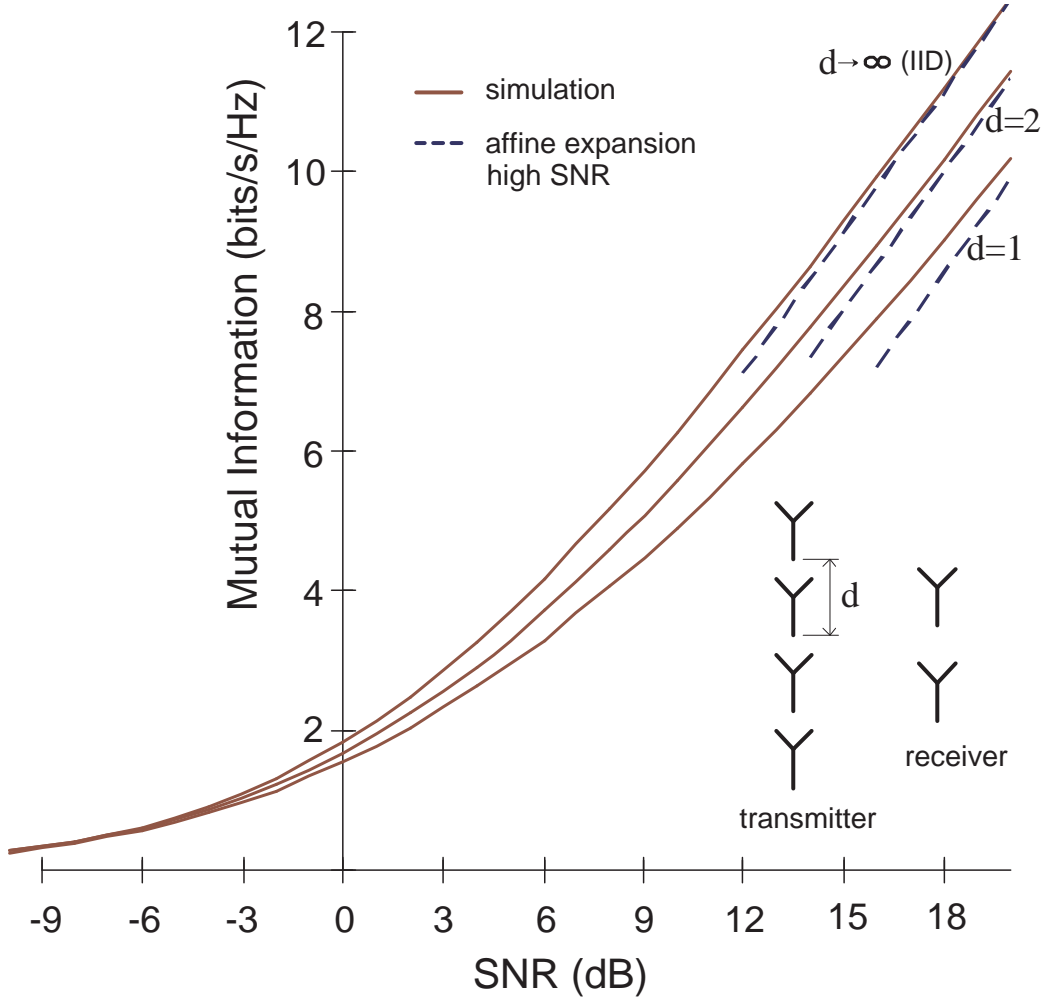


Figure 3: $\mathcal{I}(\text{SNR}, \mathbf{I})$ in a Rayleigh-faded channel with an isotropic input and with $n_T = 4$, $n_R = 2$, parameterized by the antenna separation at the transmitter (d in wavelengths). The transmit antennas are correlated with $(\Theta_T)_{i,j} = e^{-0.05 d^2 (i-j)^2}$, the receive antennas are uncorrelated. The solid lines indicate simulation (average of 50000 realizations) while the dashed lines indicate the high-SNR expansion with the power offset given by Proposition 4.

Table 1: High-SNR power offset in Example 4 as function of d . Exact values vs. asymptotic approximation

d	\mathcal{L}_∞ (exact)	\mathcal{L}_∞ (asymptotic)
$\rightarrow \infty$	1.3 dB	1.3 dB
2	3 dB	3.2 dB
1	5.2 dB	5.6 dB

Necessary and sufficient conditions for \mathbf{P} to achieve capacity at any SNR, as well as an iterative power allocation algorithm, are given in [27]. For $\text{SNR} \rightarrow \infty$, the asymptotically optimal power allocation is the one that, within the class of admissible input covariances, minimizes the power offset.

Proposition 6 *The null space of the asymptotically optimal power allocation coincides with the null space of Λ_{T} .*

Proof: See Appendix G.

It follows from Proposition 6 that, for the asymptotically optimal power allocation, the rows and columns removed from \mathbf{P} to obtain \mathbf{P}' are empty and thus

$$\text{Tr}\{\mathbf{P}'\} = \text{Tr}\{\mathbf{P}\} \quad (41)$$

$$= n_{\text{T}}. \quad (42)$$

Moreover, the powers along the diagonal of the resulting \mathbf{P}' are strictly positive. Let us denote by \mathbf{P}^* the asymptotically optimal \mathbf{P}' .

Proposition 7 *For $n'_{\text{T}} \leq n'_{\text{R}}$,*

$$\mathbf{P}^* = \frac{n_{\text{T}}}{n'_{\text{T}}} \mathbf{I} \quad (43)$$

while, for $n'_{\text{T}} > n'_{\text{R}}$,

$$\mathbf{P}^* = \arg \max_{\Xi: \text{Tr}\{\Xi\} = n_{\text{T}}} \left\{ \frac{\det \Upsilon(n'_{\text{R}}, \Lambda'_{\text{T}} \Xi)}{\det \Omega(\Lambda'_{\text{T}} \Xi)} \sum_{i=1}^{n'_{\text{R}}} \det \Psi_i(n'_{\text{R}}, \Lambda'_{\text{T}} \Xi) \right\} \quad (44)$$

with the maximization taken over all $n'_{\text{T}} \times n'_{\text{T}}$ positive-definite diagonal matrices with trace equal to n_{T} .

Proof: See Appendix H.

For $n'_{\text{T}} \leq n'_{\text{R}}$ the asymptotically optimal power allocation is uniform outside the null space of Λ_{T} . The suboptimality of a nonuniform power allocation is easily quantified as a power loss equal to

$$-\frac{1}{n'_{\text{T}}} \sum_{j=1}^{n'_{\text{T}}} \log_2 \left(\frac{n'_{\text{T}}}{n_{\text{T}}} p_j \right). \quad (45)$$

Example 5 *Consider arbitrary transmit and receive correlations, such that $n_{\text{T}} = n'_{\text{R}} = 4$ and $n'_{\text{T}} = 3$. From (45), the power allocation*

$$\mathbf{P}' = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 2.8 & 0 \\ 0 & 0 & 0.1 \end{bmatrix} \quad (46)$$

incurs a 6.4-dB power loss with respect to the asymptotically optimal allocation $\mathbf{P}' = \frac{4}{3} \mathbf{I}$.

For $n'_T > n'_R$, in contrast, \mathbf{P}^* is generally nonuniform, determined by the nonzero eigenvalues of Θ_T via the constrained maximization in (44). For specific numbers of antennas, (44) can be expatiated into more revealing optimality conditions.

Proposition 8 *Let $n_T = n'_T = 2$ and $n'_R = 1$ with λ_1 and λ_2 the nonzero eigenvalues of Θ_T . The optimum powers p_1^* and p_2^* satisfy*

$$p_1^* = \lambda_2 - \frac{\lambda_1 \lambda_2}{2} \log_e \frac{p_2^* \lambda_2}{p_1^* \lambda_1} \quad (47)$$

$$p_2^* = \lambda_1 + \frac{\lambda_1 \lambda_2}{2} \log_e \frac{p_2^* \lambda_2}{p_1^* \lambda_1}. \quad (48)$$

Example 6 *For $\lambda_1 = 1.5$ and $\lambda_2 = 0.5$, Proposition 8 yields $p_1^* = 1.85$ and $p_2^* = 0.15$.*

Note that, when the transmitter has access only to the distribution of \mathbf{H} (but not to its realization), \mathbf{P}^* does not correspond to a waterfill on any statistical measure of the channel [28, Section 8.3]. This is in contrast with the regime in which \mathbf{H} is known by the transmitter, in which case parallel noninteracting channels can be created and the power allocation does reduce to a waterfill [1, 4].

D Impact of Correlation

As long as Θ_T and Θ_R are nonsingular, $S_\infty = n_\downarrow$ as in an IID channel and hence correlation appears to be immaterial. A power-offset analysis, however, reveals its impact even when the correlation matrices are nonsingular.

Definition 3 *For any correlated Rayleigh-faded channel conforming to (23), we use $\Delta\mathcal{L}_\infty$ to denote the excess power offset relative to that of its IID Rayleigh-faded counterpart.*

D.1 Impact of Correlation: Isotropic Input

With an isotropic input, correlation within (23) is known to reduce the mutual information at every SNR [29] and thus, necessarily, $\Delta\mathcal{L}_\infty \geq 0$. This is indeed verified from our power offset expressions.

Table 2: Excess power offset in Example 7 as function of d .

d	4	2	1	0.75
$\Delta\mathcal{L}_\infty$	0.9 dB	5.4 dB	13 dB	16.1 dB

Corollary 8 Let $n_R = 1$ and $n_T = 2$ with nonzero transmit correlation eigenvalues $\lambda_1 < \lambda_2$. From (15) and (32),

$$\Delta\mathcal{L}_\infty = \left(1 - \frac{\lambda_2 \log_e \lambda_2 - \lambda_1 \log_e \lambda_1}{\lambda_2 - \lambda_1}\right) \log_2 e \quad (49)$$

$$\geq 0 \quad (50)$$

where (50) follows from the convexity of $x \log x$ and the chord lemma [30].

D.2 Impact of Correlation: Asymptotically Optimal Input

Propositions 1 and 4 also enable quantifying the impact of correlation on the power offset with optimum input covariances.

Corollary 9 For $n'_T = n'_R = n'$,

$$\Delta\mathcal{L}_\infty = \frac{1}{n'} \sum_{j=1}^{n'} \log_2 \frac{1}{\lambda'_j(\Theta_R) \lambda'_j(\Theta_T)} \quad (51)$$

This penalty, given as an upper bound in [13, Eq. 26] and as a large-dimensional limit in [4, Eq. 20], is hereby shown to be exact and valid for arbitrary n' .

Example 7 Let $n_T = n_R = 4$ with both arrays uniform and linear and with d the spacing (in wavelengths) between adjacent antennas at either end. Further let $\Phi = \mathbf{I}$. Consider, at the transmitter, a broadside (truncated) Gaussian power azimuth spectrum with a 2° root-mean-square spread and, at the receiver, a 360° uniform spectrum. The corresponding correlations are [23, 31]

$$(\Theta_T)_{i,j} = e^{-0.05 d^2 (i-j)^2} \quad (52)$$

$$(\Theta_R)_{i,j} = J_0(2\pi d |i-j|) \quad (53)$$

where $J_0(\cdot)$ is the zero-order Bessel function of the first kind. Table 2 lists the values obtained from (51), parameterized by d .

Note the severity of the penalty for small d . Since wireless systems are usually not designed to operate beyond SNR levels on the order of 20–25 dB, the power penalties in Table 2 are highly significant for antennas spaced by as much as two wavelengths.

For $n'_T \neq n'_R$, $\Delta\mathcal{L}_\infty$ can also be quantified exactly from Propositions 1 and 4, although not always by means of an explicit expression. The excess power offset does become explicit whenever correlation is present only at the side of the link with the fewest antennas, in which case

$$\Delta\mathcal{L}_\infty = \begin{cases} -\frac{1}{n'_R} \sum_{i=1}^{n'_R} \log_2 \lambda_i(\Theta_R) & n'_T > n'_R, \Theta_T = \mathbf{I} \\ -\frac{1}{n'_T} \sum_{j=1}^{n'_T} \log_2 \lambda_j(\Theta_T) & n'_T < n'_R, \Theta_R = \mathbf{I} \end{cases} \quad (54)$$

As an alternative to the exact expressions, $\Delta\mathcal{L}_\infty$ can be approximated by resorting to the large-dimensional results in Proposition 5 (with the expectations replaced by arithmetic means over the diagonal entries of $\Lambda'_T \mathbf{P}^*$ and Λ'_R) and Proposition 2.

It is worth summarizing the insight provided by the power offset on the impact of correlation on the capacity at high SNR, contrasting it with how it affects it at low SNR [16, 32]:

- Receive correlation reduces the capacity at every SNR. While the power loss caused by such correlation at high SNR, quantified by $\Delta\mathcal{L}_\infty$, is determined by a (weighted) average of the logarithm of the correlation eigenvalues, the low-SNR bandwidth penalty associated with receive correlation is determined by the average of the squared correlation eigenvalues.
- For $n'_T \leq n'_R$, transmit correlation reduces the capacity at high SNR while increasing it at low SNR.
- For $n'_T > n'_R$, transmit correlation increases the low-SNR capacity. Although assessing its exact impact at high SNR is sometimes complicated by the lack of explicit expressions for \mathbf{P}^* , transmit correlation may result in a capacity increase. Despite the fact that it lessens the diversity, transmit correlation has indeed the beneficial effect of enabling some degree of beamforming in the direction of the receiver

For $n'_T > 1$ and $n'_R = 1$, specifically, the capacity at any SNR is known to be Schur-convex with respect to the eigenvalues of the transmit correlation [33]. With the aid of the power offset, the high-SNR value of this benefit is easily quantified.

Example 8 Let $n_T = n'_T = 2$ and $n'_R = 1$. With the transmit correlation and asymptotically optimal powers given in Corollary 8, $\Delta\mathcal{L}_\infty = -0.52$ dB.

Additional results on the capacity impact of transmit correlation can be easily put forth.

Table 3: Measurement set-up

Carrier frequency	2.11 GHz
Base station height	100 m
Terminal height	1.5 m
Antenna gain	4 dBi
Mutual coupling	< -30 dB
Ricean factor	< -10 dB

Example 9 Let n_T be arbitrary while $n_R = 1$. As the transmit antennas become increasingly correlated, the transmit power concentrates on a single eigenvector and the excess power offset approaches

$$\Delta\mathcal{L}_\infty \rightarrow \log_2 e \sum_{\ell=1}^{n_T-1} \frac{1}{\ell} - 2 \log_2 n_T \quad (55)$$

which is always negative. For large n_T , (137) and (55) yield

$$\Delta\mathcal{L}_\infty \approx \gamma \log_2 e - \log_2 n_T. \quad (56)$$

As a complement to the examples provided thus far, in the following example we quantify the excess power offset using experimental data.

Example 10 Consider the correlation matrices obtained during an extensive measurement campaign conducted in New York City and reported in [34]: a narrowband transmit array, acting as a base station, was installed on a high-rise building while a receive array was mounted on a vehicle. As detailed in Table 3, each array contained 16 antennas, 8 on each polarization (vertical and horizontal). A total of 367 pairs of base station and terminal correlation matrices are processed with each polarization treated separately. Plugged in (51), each pair of correlation matrices yields an excess power offset, $\Delta\mathcal{L}_\infty$. For each polarization, a cumulative distribution of these excess power offsets is displayed in Fig. 4. The geometry of the arrays and antenna spacings (in wavelengths) is also depicted.

We observe that, in the dense urban environment in which the data was gathered, the excess power offset caused by correlation at high SNR, with an equal number of transmit and receive antennas, is on average modest: roughly 3 dB. In the worst 10% of locations, however, it is heftier: 6–9 dB.

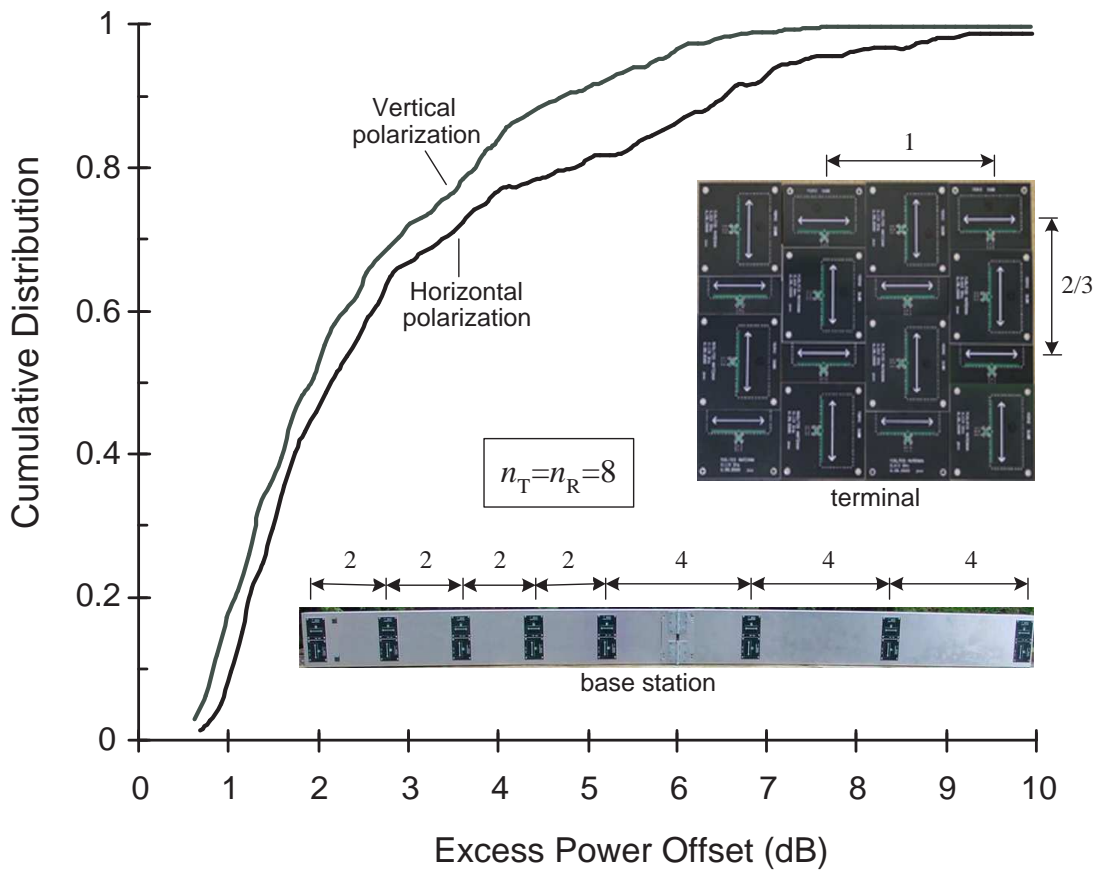


Figure 4: Cumulative distribution of excess power offsets experimentally measured (cf. Example 10). For each polarization, $n_T = n_R = 8$ with the transmit and receive arrays displayed within the figure. The indicated antenna spacings are in wavelengths.

V Ricean Channels

A Uncorrelated Ricean Channels

We now turn our attention to nonzero-mean channels of the kind

$$\mathbf{H} = \sqrt{\frac{K}{K+1}} \bar{\mathbf{H}} + \sqrt{\frac{1}{K+1}} \mathbf{W} \quad (57)$$

where $\bar{\mathbf{H}}$ is deterministic and arbitrary, normalized such that $\text{Tr}\{\bar{\mathbf{H}}\bar{\mathbf{H}}^\dagger\} = n_{\text{R}}n_{\text{T}}$, while the entries of \mathbf{W} are IID Rayleigh-faded and $K < \infty$ is the Ricean factor between the unfaded (deterministic) and fading (random) components.

For this class of channels, we circumscribe the analysis to isotropic inputs. If $n_{\text{T}} \leq n_{\text{R}}$, such inputs are asymptotically optimal for $\text{SNR} \rightarrow \infty$ [35] and hence we find \mathcal{L}_∞ . For $n_{\text{T}} > n_{\text{R}}$, we evaluate only \mathcal{L}_∞ .

It is easily verified that, with an isotropic input, the high-SNR slope is $S_\infty = n_\downarrow$. The power offset, in turn, is characterized by the following result. Since the key parameter in a Ricean channel is K , we explicitly write $\mathcal{L}_\infty(K)$.

Proposition 9 Denoting by $\phi_j, j \in \{1, \dots, n_\downarrow\}$, the nonzero eigenvalues of $K\bar{\mathbf{H}}\bar{\mathbf{H}}^\dagger$,

$$\mathcal{L}_\infty(K) = \mathcal{L}_\infty(0) + \log_2(K+1) - \frac{\log_2 e}{n_\downarrow} \frac{\sum_{i=1}^{n_\downarrow} \det \Xi_i}{\prod_{i < j} (\phi_i - \phi_j)} \quad (58)$$

where $\mathcal{L}_\infty(0)$ coincides with the power offset for an IID Rayleigh-faded channel in Proposition 1 while $\Xi_i, i \in \{1, \dots, n_\downarrow\}$, is an $n_\downarrow \times n_\downarrow$ matrix whose entries are

$$(\Xi_i)_{k,\ell} = \begin{cases} \phi_k^{n_\downarrow - \ell} & \ell \neq i \\ \phi_k^{n_\downarrow - \ell} F(k, \ell) & \ell = i \end{cases} \quad (59)$$

with

$$F(k, \ell) = \frac{{}_2F_2(1, 1; 2, n_\uparrow - \ell + 2; -\phi_k)}{n_\uparrow - \ell + 1} \phi_k \quad (60)$$

$$= \log_e \phi_k + E_1(\phi_k) + \gamma - \sum_{q=1}^{n_\uparrow - \ell} \frac{1}{q} + \sum_{q=1}^{n_\uparrow - \ell} (-\phi_k)^{-q} \left(e^{-\phi_k} (q-1)! - \frac{(n_\uparrow - \ell)!}{q (n_\uparrow - \ell - q)!} \right) \quad (61)$$

where ${}_2F_2(\cdot)$ is a hypergeometric function while

$$E_1(x) = \int_1^\infty \frac{e^{-x\xi}}{\xi} d\xi \quad (62)$$

is an exponential integral.

Proof: See Appendix I.

It is worth noting that, while the hypergeometric function ${}_2F_2(\cdot)$ corresponds—in general—to an infinite series, for the arguments in (60) we have been able to evaluate it in the closed-form (61) following the procedure in [7, Appendix X].

Corollary 10 For $n_T = n_R = 1$,

$$\mathcal{L}_\infty(K) = \mathcal{L}_\infty(K) \quad (63)$$

$$= \log_2 \frac{K+1}{K} - E_1(K) \log_2 e \quad (64)$$

which satisfies $0 \leq \mathcal{L}_\infty(K) \leq \mathcal{L}_\infty(0)$. This expression can also be found in [7].

Corollary 11 For $n_T = n_R = 2$,

$$\mathcal{L}_\infty(K) = \mathcal{L}_\infty(K) \quad (65)$$

$$= 1 + \log_2 \frac{K+1}{\sqrt{\phi_1 \phi_2}} - \frac{\log_2 e}{2} \left(E_1(\phi_1) + E_1(\phi_2) - \frac{e^{-\phi_2} - e^{-\phi_1}}{\phi_2 - \phi_1} \right). \quad (66)$$

The excess power offset induced by the Ricean component can be either positive or negative depending on the nature of such component and the numbers of antennas.

Often, the unfaded component is associated with a line-of-sight or a diffracted wave and thus it is essentially unit-rank, i.e., $\bar{\mathbf{H}} = \mathbf{a}_R \mathbf{a}_T^\dagger$ where the vectors \mathbf{a}_T and \mathbf{a}_R are the transmit and receive array responses to a plane wave.

Corollary 12 With a rank-1 unfaded component,

$$\mathcal{L}_\infty(K) = \mathcal{L}_\infty(0) + \log_2(K+1) - \frac{K}{\log_e 2} {}_2F_2(1, 1; 2, n_\uparrow + 1; -Kn_\downarrow n_\uparrow) \quad (67)$$

$$= \mathcal{L}_\infty(0) + \log_2 \frac{K+1}{(Kn_\downarrow n_\uparrow)^{1/n_\downarrow}} - \frac{\log_2 e}{n_\downarrow} \left(E_1(Kn_\downarrow n_\uparrow) + \gamma - \sum_{\ell=1}^{n_\uparrow-1} \frac{1}{\ell} + \sum_{\ell=1}^{n_\uparrow-1} \frac{e^{-Kn_\downarrow n_\uparrow} (\ell-1)! - \frac{(n_\uparrow-1)!}{\ell (n_\uparrow-1-\ell)!}}{(-Kn_\downarrow n_\uparrow)^\ell} \right) \quad (68)$$

The high-SNR mutual information in Ricean channels with rank-1 unfaded component is also analyzed in [36], where an approximate expression is given.

Corollary 13 *With $n_T = n_R = 2$ and rank-1 unfaded component,*

$$\mathcal{L}_\infty(K) = \mathcal{L}_\infty(K) \quad (69)$$

$$= 1 + \log_2 \left(\frac{K+1}{2\sqrt{K}} \right) - \frac{\log_2 e}{2} \left(\text{E}_1(4K) - \gamma - \frac{e^{-4K} - 1}{4K} \right) \quad (70)$$

which satisfies $\mathcal{L}_\infty(K) \geq \mathcal{L}_\infty(0)$.

While Corollary 10 affirms that the presence of a rank-1 unfaded component can only reduce the power offset on a scalar channel, Corollary 13 shows that this is no longer the case in a matrix channel. In fact, since the hypergeometric function in (67) decreases monotonically if either n_\downarrow or n_\uparrow grow, the behavior of Corollary 13 extends to any rank-1 Ricean channel with $n_\downarrow \geq 2$. It is worth contrasting this finding with the result put forth in [37, 38] (for arbitrary SNR), namely, that if the power received over the fading portion of the channel is held constant, then an unfaded component can only increase the mutual information. If the total received power is held constant, then this is not the case (at high SNR). The impact of Ricean components at high SNR is thus seen to depend critically on the coupling between the faded and unfaded components in terms of power gain.

As the number of antennas grows large, the excess power offset caused by a rank-1 unfaded component adopts a remarkably simple form:

$$\lim_{n_T, n_R \rightarrow \infty} \mathcal{L}_\infty(K) - \mathcal{L}_\infty(0) = \log_2(K+1) \quad (71)$$

indicating that, for sufficiently many antennas, the presence of a Ricean component has a deleterious effect at high SNR. Specifically, the mutual information achieved by an isotropic input in a channel whose unfaded component is rank-1 behaves, as the number of antennas grows large,⁸ as if only the fading portion of the channel were present. This behavior is a direct manifestation of the fact that only a single eigenvalue of $\mathbf{H}\mathbf{H}^\dagger$ is perturbed by the presence of the deterministic component. For $n_T, n_R \rightarrow \infty$, this perturbation is not reflected in the empirical eigenvalue distribution of $\mathbf{H}\mathbf{H}^\dagger$, which determines the power offset. Moreover, this is the case even if the rank of $\bar{\mathbf{H}}\bar{\mathbf{H}}^\dagger$ is $r > 1$ as long as [39, 40]

$$\lim_{n_T, n_R \rightarrow \infty} \frac{r}{n_\downarrow} = 0. \quad (72)$$

⁸In this case, the convergence of the mutual information and of related quantities such as the power offset to their limiting values may be significantly slower than in the case of zero-mean channels.

B Correlated Ricean Channels

The power offset expressions for the standard Ricean channel in (57) can be easily extended to certain correlated Ricean channels of the form

$$\mathbf{H} = \sqrt{\frac{K}{K+1}} \bar{\mathbf{H}} + \sqrt{\frac{1}{K+1}} \boldsymbol{\Theta}_R^{1/2} \mathbf{W} \boldsymbol{\Theta}_T^{1/2} \quad (73)$$

where $\boldsymbol{\Theta}_R$ and $\boldsymbol{\Theta}_T$ are nonsingular receive and transmit correlation matrices.

It is easily verified that, for the channel in (73) with $n_T < n_R$ and $\boldsymbol{\Theta}_R = \mathbf{I}$, \mathcal{L}_∞ is given by (58), (59) and (61) except with:

- $\mathcal{L}_\infty(0)$ equal to the power offset of the channel $\mathbf{H} = \mathbf{W} \boldsymbol{\Theta}_T^{1/2}$, provided by Proposition 4 with $\mathbf{P}' = \mathbf{I}$.
- ϕ_j the nonzero eigenvalues of $K \bar{\mathbf{H}} \boldsymbol{\Theta}_T^{-1} \bar{\mathbf{H}}^\dagger$.

For the channel in (73) with $n_T > n_R$ and $\boldsymbol{\Theta}_T = \mathbf{I}$, \mathcal{L}_∞ is given by (58), (59) and (61) except with:

- $\mathcal{L}_\infty(0)$ equal to the power offset of the channel $\mathbf{H} = \boldsymbol{\Theta}_R^{1/2} \mathbf{W}$, provided by Proposition 4 with $\mathbf{P}' = \mathbf{I}$.
- ϕ_j the nonzero eigenvalues of $K \bar{\mathbf{H}} \bar{\mathbf{H}}^\dagger \boldsymbol{\Theta}_R^{-1}$.

For the channel in (73) with $n_T = n_R$, \mathcal{L}_∞ is given by (58), (59) and (61) except with:

- $\mathcal{L}_\infty(0)$ equal to the power offset of the channel $\mathbf{H} = \boldsymbol{\Theta}_R^{1/2} \mathbf{W} \boldsymbol{\Theta}_T^{1/2}$, provided by Proposition 4 with $\mathbf{P}' = \mathbf{I}$.
- ϕ_j the nonzero eigenvalues of $K \bar{\mathbf{H}} \boldsymbol{\Theta}_T^{-1} \bar{\mathbf{H}}^\dagger \boldsymbol{\Theta}_R^{-1}$.

VI Interference-limited Channels

In addition to quantifying the impact of channel features such as correlation and unfaded components, the power offset can also be used to reveal the impact on the capacity of certain noise attributes. An interesting scenario is that of intercell interference, which—like the desired signal itself—is subject to fading and correlation. Interference-limited scenarios, in fact, are very relevant to modern cellular systems, which tend to operate

precisely in such conditions. Let us thus consider an interference-limited model where the interference emanates from a single multiantenna array,

$$\mathbf{y} = \sqrt{g} \mathbf{H}\mathbf{x} + \sqrt{g_I} \mathbf{H}_I \mathbf{x}_I \quad (74)$$

where the underlying thermal noise has been neglected and \mathbf{x}_I is an n_I -dimensional isotropic Gaussian signal whose codebook is unknown to the receiver. The random channel \mathbf{H}_I , known at the receiver, is subject to the standard normalization, $E[\text{Tr}\{\mathbf{H}_I \mathbf{H}_I^\dagger\}] = n_I n_R$. The conditional covariance of the interference is then

$$E[\sqrt{g_I} \mathbf{H}_I \mathbf{x}_I | \mathbf{H}_I] = g_I \frac{E[\|\mathbf{x}_I\|^2]}{n_I} \mathbf{H}_I \mathbf{H}_I^\dagger. \quad (75)$$

In order to contrast the essence of the intercell interference scenario in (74) with that of AWGN, we model both \mathbf{H} and \mathbf{H}_I as IID Rayleigh-faded and thus the capacity-achieving input is isotropic and the ergodic capacity becomes

$$C(\text{SNR}) = E \left[\log_2 \det \left(\mathbf{I} + \text{SNR} \frac{\mathbf{H}\mathbf{H}^\dagger}{n_T} \left(\frac{\mathbf{H}_I \mathbf{H}_I^\dagger}{n_I} \right)^{-1} \right) \right] \quad (76)$$

with expectation over both \mathbf{H} and \mathbf{H}_I and with

$$\text{SNR} = \frac{g E[\|\mathbf{x}\|^2]}{g_I E[\|\mathbf{x}_I\|^2]}. \quad (77)$$

The dimensionality of the interference must be restricted to $n_I \geq n_R$ for otherwise some spatial dimensions would be interference-free and the capacity would be unbounded.⁹ With this restriction, $\mathbf{H}_I \mathbf{H}_I^\dagger$ is nonsingular with probability 1 and the high-SNR slope is the usual $S_\infty = n_\downarrow$. In order to assess the role of n_I , we must once again resort to the power offset.

Proposition 10 *For the interference-limited channel in (74),*

$$\mathcal{L}_\infty = \log_2 \frac{n_T}{n_I} + \frac{\log_2 e}{n_\downarrow} \sum_{k=1}^{n_\downarrow} \left(\binom{n_I - k - [n_R - n_T]^+}{\sum_{\ell=1}^{n_I - k} \frac{1}{\ell}} - \sum_{\ell=1}^{n_\uparrow - k} \frac{1}{\ell} \right) \quad (78)$$

where $[\cdot]^+ = \max(\cdot, 0)$. For $n_T = n_R = n$, (78) simplifies to

$$\mathcal{L}_\infty = \log_2 \frac{n}{n_I} + \frac{\log_2 e}{n} \sum_{k=1}^n \sum_{\ell=n-k+1}^{n_I-k} \frac{1}{\ell} \quad (79)$$

Proof: See Appendix J

⁹This limitation is an artifact of the model. In any actual situation, the inevitable thermal noise underlying the interference would prevent it from ever being singular. Nonetheless, for $n_I \geq n_R$ this is an adequate model for scenarios where the interference dominates.

For $n_I/n_R \rightarrow \infty$, the interference becomes white and the above expressions converge to the corresponding ones for AWGN (cf. Proposition 1). For finite n_I/n_R , however, the power offset is strictly lower than in AWGN. Its lowest value, found for $n_I = n_R$, is

$$\mathcal{L}_\infty = \log_2 \frac{n_T}{n_R} - \frac{\log_2 e}{n_\downarrow} \sum_{k=1}^{n_\downarrow} \sum_{\ell=n_\downarrow-k+1}^{n_\uparrow-k} \frac{1}{\ell} \quad (80)$$

If $n_T = n_R = n_I$, in particular, then $\mathcal{L}_\infty^* = 0$.

Also of interest is the power offset in the large-dimensional limit, for which it is again possible to obtain closed-form expressions.

Proposition 11 For $n_T, n_R, n_I \rightarrow \infty$ with $n_T/n_R = \beta$ and $n_I/n_R = \beta_I$ where $\beta_I \geq 1$,

$$\mathcal{L}_\infty \rightarrow \begin{cases} (\beta - 1) \log_2(1 - \frac{1}{\beta}) - (\beta_I - 1) \log_2(1 - \frac{1}{\beta_I}) & \beta > 1 \\ -(\beta_I - 1) \log_2(1 - \frac{1}{\beta_I}) & \beta = 1 \\ (\frac{1}{\beta} - 1) \log_2(1 - \beta) + \log_2 \frac{\beta + \beta_I - 1}{\beta_I/\beta} + \frac{\beta_I - 1}{\beta} \log_2(1 + \frac{\beta}{\beta_I - 1}) & \beta < 1 \end{cases} \quad (81)$$

Proof: See Appendix J.

For $\beta_I = 1$, the expressions in Proposition 11 reduce to

$$\mathcal{L}_\infty \rightarrow \begin{cases} (\beta - 1) \log_2(1 - \frac{1}{\beta}) & \beta > 1 \\ 0 & \beta = 1 \\ (\frac{1}{\beta} - 1) \log_2(1 - \beta) + 2 \log_2 \beta & \beta < 1 \end{cases} \quad (82)$$

which is the large-dimensional limit of (80).

In an interference-limited channel, the shift in power offset due to the interference being subject to fading is essentially the opposite of the shift caused by fading in the desired signal. If the number of interfering antennas equals the number of transmit antennas, then both effects cancel out and the high-SNR capacity equals that of an AWGN-channel whose dimensions were unfaded and orthogonal.

Although, throughout this section, we have considered a single multiantenna interferer, the exact same analysis holds for multiple equal-power interferers. Also, the characterization can be extended to correlated signal and interference [26], for which similar observations can be made: if the interference is correlated at the receiver, the effect is essentially the opposite of the desired signal being correlated.

The main finding of this section is that there is much to be gained by exploiting the spatial color of dominant interferers even if their codebooks are unknown. For example, when $n_T = n_R = 4$, the difference in power offset between background noise being spatially white or emanating from a 4-antenna interferer is 3.8 dB.

VII Conclusions

The capacity (or, more generally, the mutual information) regulates the interplay between power, bandwidth and rate. With transmit power P and bandwidth B , the rate R that can be conveyed with arbitrary reliability must satisfy $R/B \leq \mathcal{I}$ where, in the high-SNR regime,

$$\mathcal{I} \approx S_\infty \left(\frac{g \frac{P}{BN_0} |_{\text{dB}}}{3 \text{ dB}} - \mathcal{L}_\infty \right) \quad (83)$$

with N_0 the noise spectral density per receive antenna and with g the average path-gain. Given a desired rate and available bandwidth, the minimum required transmit power is therefore

$$P|_{\text{dB}} \approx \frac{BN_0}{g} |_{\text{dB}} + \left(\frac{R}{BS_\infty} + \mathcal{L}_\infty \right) 3 \text{ dB} \quad (84)$$

where the approximation sharpens as $\text{SNR} \rightarrow \infty$. The power required to sustain a certain rate R with a given bandwidth B is critically influenced, at high R/B , by \mathcal{L}_∞ . In this paper, we have derived analytical characterizations of \mathcal{L}_∞ from which the impact of various features (correlation, Ricean factors, interference, etc) can be insightfully assessed. While the focus has been on single-user channels, many of the results apply almost verbatim to multiple-access channels where a number of scalar transmitters access a common multi-antenna receiver.

Although S_∞ and \mathcal{L}_∞ are asymptotic (high-SNR) quantities, we have exemplified how the expansion they render can be utilized to approximate the mutual information at high but finite SNR (cf. Fig. 1 and Fig. 3). In order to increase the relevance of these quantities in the finite-SNR realm, it is helpful to acknowledge that, at any given SNR, some nonzero channel singular values may be “dormant” (i.e., well below the noise) and thus, at such SNR, the contribution of these singular is negligible [3]. In the case of a correlated channel at a certain SNR, for instance, this pragmatic interpretation states that those eigenvalues of the correlation matrices that are significantly smaller than $1/\text{SNR}$ should be regarded as zero in the computation of S_∞ and \mathcal{L}_∞ . Armed with this perspective, we can revisit the correlated Rayleigh-faded channel in Fig. 1. From the corresponding correlation matrices,

given in (30), we have the nonzero correlation eigenvalues

$$\Lambda'_T = \Lambda'_R \quad (85)$$

$$= \begin{bmatrix} 2.3 & 0 & 0 & 0 \\ 0 & 1.5 & 0 & 0 \\ 0 & 0 & 0.15 & 0 \\ 0 & 0 & 0 & 0.002 \end{bmatrix} \quad (86)$$

which, plugged into Proposition 4, yield \mathcal{E}_∞ . In conjunction with $S_\infty = 4$, this leads to the expansion for $\text{SNR} \rightarrow \infty$ depicted in Fig. 1. The pragmatic finite-SNR interpretation, however, indicates that the smallest of the correlation eigenvalues should be neglected below SNR levels on the order of $1/0.002$ (roughly 27 dB) and the second smallest correlation eigenvalue should be neglected below SNR levels on the order of $1/0.15$ (roughly 9 dB). These eigenvalue exclusions lead to different expansions, respectively having $S_\infty = 3$ and $S_\infty = 2$ and correspondingly different power offsets as given by Proposition 4, which are displayed in Fig. 5. Within the anticipated SNR ranges, these expansions do represent more accurately the behavior of the mutual information with an isotropic input thus buttressing the validity of this finite-SNR interpretation of our asymptotic results.

To conclude, it is worth remarking that the expressions presented in this paper have direct operational significance in ergodic channel conditions. In non-ergodic channels, the power offset becomes a random quantity (whose expected value is precisely given by the expressions in this paper). Although characterizing the outage distribution of such random quantity might be analytically intractable, recent results indicate that this distribution becomes rapidly Gaussian as the number of antennas grows [14, Sec. 3.3.8]. Thus, the expressions given herein, complemented by those for the variance, would suffice to provide accurate evaluations thereof. Also noteworthy in the realm of non-ergodic channels is the fact that it is sometimes convenient to sacrifice some of the degrees of freedom in exchange for diversity, according to a well established tradeoff [41]. Hence, for non-ergodic scenarios it would be of interest to study the power offset corresponding to other points in this tradeoff, particularly in light of recently proposed codes that achieve the optimum multiplexing-diversity tradeoff [42, 43] but that may nevertheless operate a distance away from capacity. The high-SNR power offset quantifies this distance.

We lastly note that, although our analysis has focused on single-user multiantenna channels, some of the formulas apply verbatim to multiaccess channels (with single-antenna transmitters and a multiantenna receiver) and, in a number of cases, to broadcast channels as well [44]. Additional characterizations of the power offset for multiantenna broadcast channels can be found in [44].

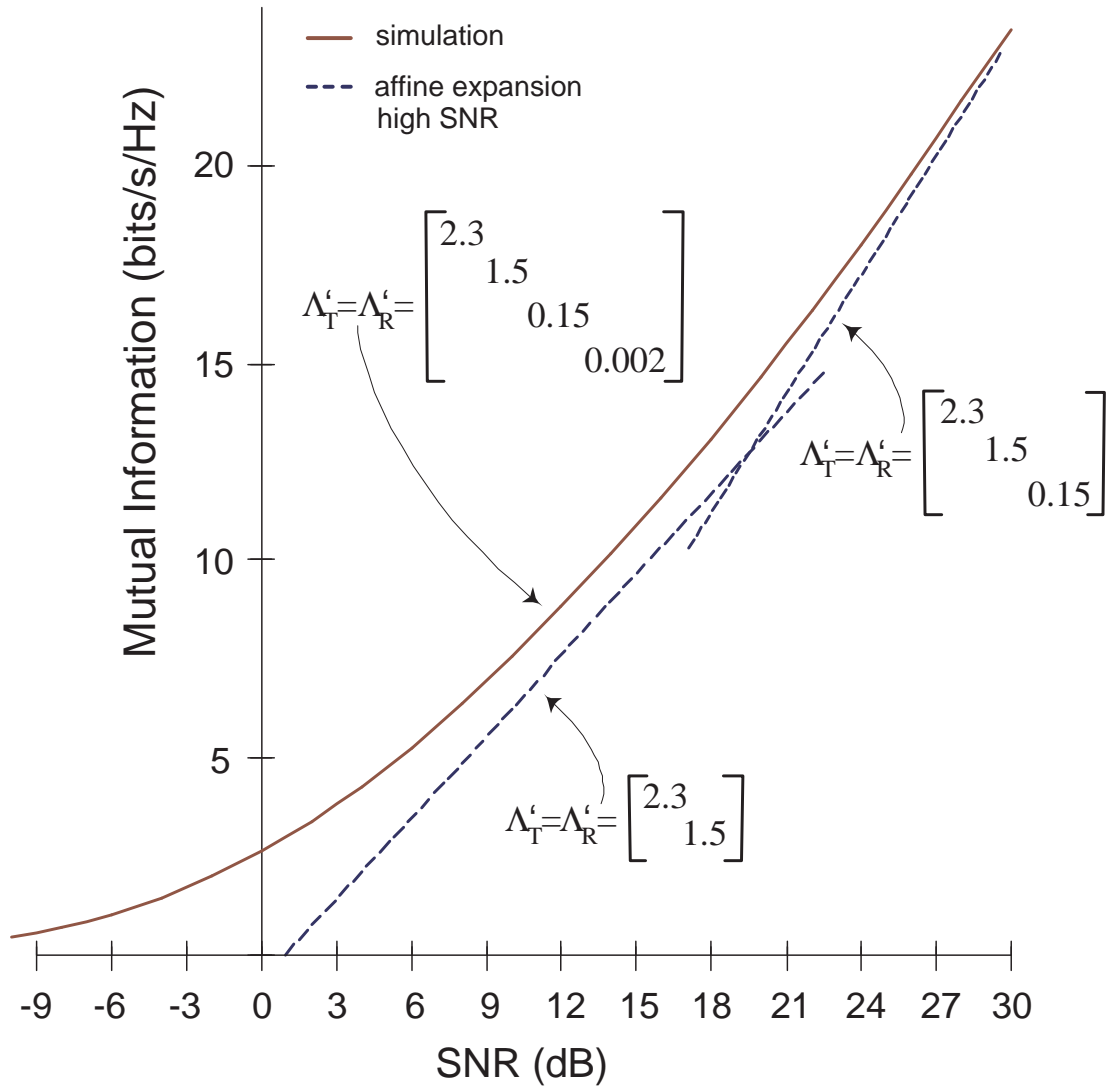


Figure 5: For $n_T = n_R = 4$, $\mathcal{I}(\text{SNR}, \mathbf{I})$ in a correlated Rayleigh-faded channel with identical transmit and receive correlation matrices (cf. Example 3). The solid line indicates simulation while the dashed lines indicate the high-SNR expansions obtained by neglecting, respectively, the smallest and the two smallest eigenvalues of the transmit and receive correlation matrices. The corresponding high-SNR expansion with all of the correlation eigenvalues accounted for is shown in Fig. 1.

Appendices

A Auxiliary Results

We shall make use of the several results presented in this appendix. Lemma 1, in particular, has already been applied to multiantenna analyses by other authors [13, 45].

Lemma 1 [46](see also [47]) *Let \mathbf{F} and \mathbf{G} be two $n \times n$ matrices whose (i,j) th entries are, respectively, $(\mathbf{F})_{i,j} = f_j(w_i)$ and $(\mathbf{G})_{i,j} = g_j(w_i)$ where f_j and g_j , $j = 1, \dots, n$, are functions defined on \mathbb{R}^+ . Then, for $b > a > 0$,*

$$\int_a^b \dots \int_a^b \det \mathbf{F} \det \mathbf{G} dw_1 \dots dw_n = n! \det \mathbf{A} \quad (87)$$

where \mathbf{A} is another $n \times n$ matrix whose (i,j) th entry is

$$\mathbf{A} = \int_a^b f_i(w) g_j(w) dw \quad (88)$$

Note that, in [46], the factor $n!$ does not appear because the variables w_1, \dots, w_n are in a specific order. As claimed in Lemma 1, in contrast, they are unordered.

Lemma 2 *Let \mathbf{W} be a $m \times n$ matrix with IID zero-mean unit-variance complex Gaussian random entries and with $m < n$. Let Θ be a $n \times n$ Hermitian positive-definite matrix. Then,*

$$E [\log_e \det (\mathbf{W}\Theta\mathbf{W}^\dagger)] = \frac{\det \Upsilon}{\det \Omega} \sum_{i=1}^m \det \Psi_i \quad (89)$$

where Ψ_i is an $m \times m$ matrix whose entries are

$$(\Psi_i)_{k,\ell} = \nu_{n-m+k} \lambda_{n-m+k}^{n-m-1+\ell} - \sum_{d=1, q=1}^{n-m} \nu_q (\Upsilon^{-1})_{d,q} \lambda_{n-m+k}^{d-1} \lambda_q^{n-m-1+\ell} \quad (90)$$

where λ_j is the j th eigenvalue of Θ while

$$\nu_q = \begin{cases} 1 & \ell \neq i \\ \Psi(\ell) + \log_e \lambda_q & \ell = i \end{cases} \quad (91)$$

In turn, Ω is the Vandermonde matrix

$$\Omega = \begin{bmatrix} 1 & \lambda_1 & \dots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \dots & \lambda_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_n & \dots & \lambda_n^{n-1} \end{bmatrix} \quad (92)$$

and Υ is the $(n - m) \times (n - m)$ principal submatrix of Ω . Euler's digamma function, $\psi(\cdot)$, is given in (136).

Proof: Denoting by $\xi_1 \geq \xi_2 \geq \dots \geq \xi_m$ the ordered eigenvalues of $\mathbf{W}\Theta\mathbf{W}^\dagger$, their joint density distribution is [48]

$$f_\xi(\xi_1, \dots, \xi_m) = \frac{\det \Gamma \det \Xi}{\det \Omega \prod_{\ell=1}^{m-1} \ell!} \quad (93)$$

where

$$\Xi = \begin{bmatrix} 1 & \xi_1 & \dots & \xi_1^{m-1} \\ 1 & \xi_2 & \dots & \xi_2^{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_m & \dots & \xi_m^{m-1} \end{bmatrix} \quad (94)$$

is also a Vandermonde matrix while

$$\Gamma = \begin{bmatrix} 1 & \lambda_1 & \dots & \lambda_1^{n-m-1} & \lambda_1^{n-m-1} e^{-\xi_1/\lambda_1} & \dots & \lambda_1^{n-m-1} e^{-\xi_m/\lambda_1} \\ 1 & \lambda_2 & \dots & \lambda_2^{n-m-1} & \lambda_2^{n-m-1} e^{-\xi_1/\lambda_2} & \dots & \lambda_2^{n-m-1} e^{-\xi_m/\lambda_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_n & \dots & \lambda_n^{n-m-1} & \lambda_n^{n-m-1} e^{-\xi_1/\lambda_n} & \dots & \lambda_n^{n-m-1} e^{-\xi_m/\lambda_n} \end{bmatrix} \quad (95)$$

which, recalling the definition of Υ , conforms to the block structure

$$\Gamma = \begin{bmatrix} \Upsilon & \mathbf{D} \\ \mathbf{C} & \mathbf{B} \end{bmatrix}. \quad (96)$$

The joint density distribution of the *unordered* eigenvalues of $\mathbf{W}\Theta\mathbf{W}^\dagger$, in turn, is given by $\frac{1}{m!} f_\xi(\cdot)$. The moment-generating function of $\log_e \det(\mathbf{W}\Theta\mathbf{W}^\dagger)$ is thus

$$M(t) = E \left[\det(\mathbf{W}\Theta\mathbf{W}^\dagger)^t \right] \quad (97)$$

$$= \frac{1}{\det \Omega \prod_{\ell=1}^m \ell!} \int_{\xi_1 \dots \xi_m} \left(\prod_{i=1}^m \xi_i \right)^t \det \Gamma \det \Xi \, d\xi_1, \dots, d\xi_m \quad (98)$$

$$= \frac{\det \Upsilon}{\det \Omega \prod_{\ell=1}^m \ell!} \int_{\xi_1 \dots \xi_m} \left(\prod_{i=1}^m \xi_i \right)^t \det \Xi \det(\mathbf{B} - \mathbf{C}\Upsilon^{-1}\mathbf{D}) \, d\xi_1, \dots, d\xi_m \quad (99)$$

$$= \frac{\det \Upsilon}{\det \Omega \prod_{\ell=1}^m \ell!} \int_{\xi_1 \dots \xi_m} \det(\text{diag}\{\xi_1^t, \dots, \xi_m^t\} \Xi) \det(\mathbf{B} - \mathbf{C}\Upsilon^{-1}\mathbf{D}) \, d\xi_1, \dots, d\xi_m \quad (100)$$

where, in (99), we have used the determinant expansion of a block matrix [49]. Applying Lemma 1,

$$M(t) = \frac{\det \Upsilon \det \mathbf{A}}{\det \Omega \prod_{\ell=1}^{m-1} \ell!} \quad (101)$$

with \mathbf{A} a $m \times m$ matrix whose (k, ℓ) th entry is

$$(\mathbf{A})_{k, \ell} = \int_0^\infty \xi^{t+\ell-1} \left(\lambda_{n-m+k}^{n-m-1} e^{-\frac{\xi}{\lambda_{n-m+k}}} - \sum_{d=1}^{n-m} \sum_{q=1}^{n-m} (\mathbf{\Upsilon}^{-1})_{d,q} \lambda_q^{n-m-1} \lambda_{n-m+k}^{d-1} e^{-\frac{\xi}{\lambda_q}} \right) d\xi \quad (102)$$

$$= \lambda_{n-m+k}^{n-m-1+\ell+t} \Gamma(\ell+t) - \sum_{d=1}^{n-m} \sum_{q=1}^{n-m} (\mathbf{\Upsilon}^{-1})_{d,q} \lambda_{n-m+k}^{d-1} \lambda_q^{n-m-1+\ell+t} \Gamma(\ell+t) \quad (103)$$

where, in (103), we have invoked the definition of the Gamma function [50]

$$\Gamma(\ell) = \int_0^\infty \xi^{\ell-1} e^{-\xi} d\xi \quad (104)$$

which, for integer ℓ , satisfies $\Gamma(\ell+1) = \ell!$ From $M(t)$,

$$E [\log_e \det(\mathbf{W}\Theta\mathbf{W}^\dagger)] = \frac{d}{dt} M(t)|_{t=0} \quad (105)$$

$$= \frac{\det \mathbf{\Upsilon}}{\det \mathbf{\Omega} \prod_{\ell=1}^{m-1} \ell!} \frac{d}{dt} \det \mathbf{A}|_{t=0} \quad (106)$$

$\frac{d}{dt} \det \mathbf{A}|_{t=0}$ equals the sum of the determinants of m matrices, the i th of which has entries that, except for the i th column, are given by the corresponding entries of \mathbf{A} evaluated at $t = 0$ whereas, for the i th column, are given by the derivative at $t = 0$ of the corresponding entries of \mathbf{A} . Applying this rule, using $\dot{\Gamma}(\ell) = \Gamma(\ell)\psi(\ell)$ and absorbing $\prod_{\ell=1}^{m-1} \ell!$, Lemma 2 follows from (106).

Lemma 3 Consider an $m \times n$ random matrix $\mathbf{H} = \bar{\mathbf{H}} + \mathbf{W}$, with $m \leq n$, where $\bar{\mathbf{H}}$ is deterministic while the entries of \mathbf{W} are zero-mean unit-variance IID complex Gaussian. Denoting by ϕ_j , $j \in \{1, \dots, m\}$, the eigenvalues of $\bar{\mathbf{H}}\bar{\mathbf{H}}$,

$$E[\log_e \det(\mathbf{H}\mathbf{H}^\dagger)] = \sum_{i=0}^{m-1} \psi(n-i) + \frac{\sum_{i=1}^m \det \mathbf{\Xi}_i}{\prod_{i < j} (\phi_i - \phi_j)} \quad (107)$$

with $\psi(\cdot)$ the digamma function while $\mathbf{\Xi}_i$, $i \in \{1, \dots, m\}$, is an $m \times m$ matrix whose entries are

$$(\mathbf{\Xi}_i)_{k, \ell} = \begin{cases} \phi_k^{m-\ell} & \ell \neq i \\ \phi_k^{m-\ell} F(k, \ell) & \ell = i \end{cases} \quad (108)$$

with

$$F(k, \ell) = \frac{{}_2F_2(1, 1; 2, n-\ell+2; -\phi_k)}{n-\ell+1} \phi_k \quad (109)$$

$$= \log_e \phi_k + E_1(\phi_k) + \gamma - \sum_{q=1}^{n-\ell} \frac{1}{q} + \sum_{q=1}^{n-\ell} (-\phi_k)^{-q} \left(e^{-\phi_k} (q-1)! - \frac{(n-\ell)!}{q(n-\ell-q)!} \right) \quad (110)$$

where $E_1(\cdot)$ is an exponential integral while ${}_2F_2(\cdot)$ is the hypergeometric function

$${}_2F_2(1, 1; 2, p; z) = \sum_{q=0}^{\infty} \frac{(p-1)!}{(p+q-1)!} \frac{z^q}{q+1}. \quad (111)$$

Proof: The density distribution of the $m \times m$ noncentral Wishart matrix $\mathbf{H}\mathbf{H}^\dagger$ is [51]

$$f_{\mathbf{H}\mathbf{H}^\dagger}(\mathbf{S}) = \frac{{}_0F_1(n, \bar{\mathbf{H}}\bar{\mathbf{H}}^\dagger\mathbf{S})}{\pi^{m(m-1)/2} \prod_{i=1}^m (n-i)!} \det(\mathbf{S})^{n-m} e^{-\text{Tr}\{\bar{\mathbf{H}}\bar{\mathbf{H}}^\dagger + \mathbf{S}\}} \quad (112)$$

from which the moment-generating function of $\log_e \det(\mathbf{H}\mathbf{H}^\dagger)$ can be obtained as

$$M(t) = E[\det(\mathbf{H}\mathbf{H}^\dagger)^t] \quad (113)$$

$$= \int_{\mathbf{S}>0} \det(\mathbf{S})^{t+n-m} \frac{{}_0F_1(n, \bar{\mathbf{H}}\bar{\mathbf{H}}^\dagger\mathbf{S})}{\pi^{m(m-1)/2} \prod_{i=1}^m (n-i)!} e^{-\text{Tr}\{\bar{\mathbf{H}}\bar{\mathbf{H}}^\dagger + \mathbf{S}\}} d\mathbf{S} \quad (114)$$

$$= {}_1F_1(n+t, n, \bar{\mathbf{H}}\bar{\mathbf{H}}^\dagger) e^{-\text{Tr}\{\bar{\mathbf{H}}\bar{\mathbf{H}}^\dagger\}} \prod_{i=1}^m \frac{(t+n-i)!}{(n-i)!} \quad (115)$$

where ${}_0F_1(\cdot, \cdot)$ and ${}_1F_1(\cdot, \cdot, \cdot)$ are hypergeometric functions of matrix argument and, in (115), we have used [51]

$$\frac{\pi^{-m(m-1)/2}}{\prod_{i=1}^m (a-i)!} \int_{\mathbf{S}>0} {}_pF_q(a_1, \dots, a_p, b_1, \dots, b_q, \mathbf{S}\mathbf{Z}) \frac{\det(\mathbf{S})^{a-m}}{e^{\text{Tr}\{\mathbf{S}\}}} d\mathbf{S} = {}_{p+1}F_q(a_1, \dots, a_p, a, b_1, \dots, b_q, \mathbf{Z}) \quad (116)$$

The hypergeometric function of matrix argument ${}_1F_1(\cdot, \cdot, \cdot)$ in (115) is given by [52, 53]

$${}_1F_1(n+t, n, \bar{\mathbf{H}}\bar{\mathbf{H}}^\dagger) = \frac{\det \mathbf{G}}{\prod_{i<j} (\phi_i - \phi_j)} \quad (117)$$

where the entries of the $m \times m$ matrix \mathbf{G} are

$$(\mathbf{G})_{k,\ell} = \phi_k^{m-\ell} {}_1F_1(n+t-\ell+1, n-\ell+1, \phi_k). \quad (118)$$

From $M(t)$,

$$E[\log_e \det(\mathbf{H}\mathbf{H}^\dagger)] = \frac{d}{dt} M(t)|_{t=0} \quad (119)$$

$$= \sum_{i=0}^{m-1} \psi(n-i) + \frac{\sum_{i=1}^m \det \Xi_i}{\prod_{i<j} (\phi_i - \phi_j)} \quad (120)$$

where the entries of Ξ_i are

$$(\Xi_i)_{k,\ell} = \begin{cases} \phi_k^{m-\ell} & \ell \neq i \\ \phi_k^{m-\ell} \left(-\psi(n-\ell+1) + e^{-\phi_k} \sum_{q=0}^{\infty} \frac{\phi_k^q}{q!} \psi(n-\ell+1+q) \right) & \ell = i \end{cases}$$

The expression for $(\Xi_i)_{k,i}$ can be further manipulated into the hypergeometric form in (109) and, using the equivalence in [7, Eqs. 414 and 418], also into the closed form in (110).

Lemma 4 *Let \mathbf{X} and \mathbf{Y} be independent random matrices whose entries are zero-mean IID complex Gaussian, respectively $n \times m$ and $n \times p$, with $m \leq n \leq p$. Then,*

$$E \left[\log_e \det \left(\mathbf{X}^\dagger (\mathbf{Y}\mathbf{Y}^\dagger)^{-1} \mathbf{X} \right) \right] = \sum_{k=1}^m \left(\sum_{\ell=1}^{n-k} \frac{1}{\ell} - \sum_{\ell=1}^{m+p-n-k} \frac{1}{\ell} \right) \quad (121)$$

Proof: The density distribution of $\mathbf{Z} = \mathbf{X}^\dagger (\mathbf{Y}\mathbf{Y}^\dagger)^{-1} \mathbf{X}$ is [51, 54]¹⁰

$$f_{\mathbf{Z}}(\mathbf{S}) = \frac{\tilde{\Gamma}_n(m+p)}{\tilde{\Gamma}_n(p)\tilde{\Gamma}_m(n)} \det(\mathbf{I} + \mathbf{S})^{-(m+p)} \det(\mathbf{S})^{n-m} \quad (122)$$

where

$$\tilde{\Gamma}_a(b) = \pi^{a(a-1)/2} \prod_{i=1}^a (b-i)! \quad (123)$$

Denote by $M(t)$ the moment-generating function of $\log_e \det \mathbf{Z}$. From (122), for $t \in (-1, 1)$,

$$M(t) = E \left[\det(\mathbf{Z})^t \right] \quad (124)$$

$$= \frac{\tilde{\Gamma}_n(m+p)}{\tilde{\Gamma}_n(p)\tilde{\Gamma}_m(n)} \int_{\mathbf{Z}} \det(\mathbf{I} + \mathbf{Z})^{-(m+p)} \det(\mathbf{Z})^{n-m+t} d\mathbf{Z} \quad (125)$$

Using the following generalization to complex $m \times m$ matrices of the result given by real matrices in [54],

$$\int_{\mathbf{S} > 0} \det(\mathbf{I} + \mathbf{S})^{-(r+s)} \det(\mathbf{S})^{r-m} d\mathbf{S} = \frac{\tilde{\Gamma}_m(r)\tilde{\Gamma}_m(s)}{\tilde{\Gamma}_m(r+s)} \quad (126)$$

the moment-generating function turns out to be

$$M(t) = \frac{\tilde{\Gamma}_n(m+p)\tilde{\Gamma}_m(n+t)\tilde{\Gamma}_m(m+p-n-t)}{\tilde{\Gamma}_n(p)\tilde{\Gamma}_m(n)\tilde{\Gamma}_m(m+p)} \quad (127)$$

from which

$$E[\log_e \det \mathbf{Z}] = \frac{d}{dt} M(t) \Big|_{t=0} \quad (128)$$

$$= \sum_{i=0}^{m-1} (\psi(n-i) - \psi(m+p-n-i)) \quad (129)$$

which, using (136), leads to the claim.

¹⁰In fact, [54] gives the density distribution for matrices of the form $\mathbf{X}^\dagger (\mathbf{Y}\mathbf{A}\mathbf{Y}^\dagger)^{-1} \mathbf{X}$ where \mathbf{A} is a deterministic matrix. Thus, a more general version of this lemma could be derived with application to multi-antenna channels with correlated interference.

B Proof of Proposition 1

Let $n_T \geq n_R$. Using (5) with $\Phi = \mathbf{I}$, the capacity is

$$C(\text{SNR}) = n_R \log_2 \text{SNR} + E \left[\log_2 \det \left(\frac{1}{\text{SNR}} \mathbf{I} + \frac{1}{n_T} \mathbf{W} \mathbf{W}^\dagger \right) \right]. \quad (130)$$

where \mathbf{W} has IID unit-variance Rayleigh-faded entries. Recognizing $S_\infty(\mathbf{I}) = n_R$ and applying (11),

$$\mathcal{L}_\infty = \log_2 n_T - \frac{1}{n_R} E \left[\log_2 \det (\mathbf{W} \mathbf{W}^\dagger) \right] \quad (131)$$

where the Wishart matrix $\mathbf{W} \mathbf{W}^\dagger$, nonsingular with probability 1, satisfies [55, 56]

$$E \left[\log_e \det (\mathbf{W} \mathbf{W}^\dagger) \right] = \sum_{\ell=0}^{n_R-1} \psi(n_T - \ell) \quad (132)$$

with $\psi(\cdot)$ the digamma function in (136). Plugging (132) into (131) and invoking

$$1 + \frac{1}{n} \sum_{\ell=1}^n \psi(\ell) = \psi(n+1) \quad (133)$$

we obtain

$$\mathcal{L}_\infty = \log_2 n_T - \log_2 e \left(\frac{n_T}{n_R} \psi(n_T + 1) - \frac{n_T - n_R}{n_R} \psi(n_T - n_R + 1) - 1 \right) \quad (134)$$

from which, using (136), the final expression for \mathcal{L}_∞ is found.

For $n_T \leq n_R$, an analogous derivation can be conducted originating from

$$C(\text{SNR}) = n_T \log_2 \text{SNR} + E \left[\log_2 \det \left(\frac{1}{\text{SNR}} \mathbf{I} + \frac{1}{n_T} \mathbf{W}^\dagger \mathbf{W} \right) \right]. \quad (135)$$

In the case that $n_T = n_R$, both derivations become identical.

C Proof of Proposition 2

Proposition 1 can be written in terms of the digamma function [50]

$$\psi(n) = -\gamma + \sum_{\ell=1}^{n-1} \frac{1}{\ell} \quad (136)$$

which, for large n , behaves as

$$\psi(n) = \log_e n + O\left(\frac{1}{n}\right). \quad (137)$$

Thus, Proposition 1 leads directly to the limiting expressions in Proposition 2 in the case of an IID Rayleigh-faded channel. As it turns out, these limiting expressions are valid regardless of the fading distribution. For any channel with zero-mean unit-variance IID entries, as $n_T, n_R \rightarrow \infty$ with $\beta = n_T/n_R$ the capacity per antenna converges to [57]

$$\begin{aligned} \frac{1}{n_R} \mathcal{I}(\text{SNR}, \mathbf{I}) &\rightarrow \log_2 \left(1 + \text{SNR} - \frac{1}{4} \mathcal{F}(\beta, \frac{\text{SNR}}{\beta}) \right) \\ &+ \beta \log_2 \left(1 + \frac{\text{SNR}}{\beta} - \frac{1}{4} \mathcal{F}(\beta, \frac{\text{SNR}}{\beta}) \right) - \beta \frac{\log_2 e}{4 \text{SNR}} \mathcal{F}(\beta, \frac{\text{SNR}}{\beta}) \end{aligned} \quad (138)$$

given the auxiliary function

$$\mathcal{F}(x, y) \triangleq \left(\sqrt{1 + y(1 + \sqrt{x})^2} - \sqrt{1 + y(1 - \sqrt{x})^2} \right)^2. \quad (139)$$

Application of (11) to (138) leads again to the limiting expressions in Proposition 2, in this case for arbitrary fading distributions (cf. [25, Eq. 15–16]).

D Proof of Proposition 3

The high-SNR slope can be calculated as [8]

$$S_\infty(\Phi) = \lim_{\text{SNR} \rightarrow \infty} \frac{\text{SNR} \dot{\mathcal{I}}(\text{SNR}, \Phi)}{\log_2 e}. \quad (140)$$

where $\dot{\mathcal{I}}(\text{SNR}, \cdot)$ indicates the derivative with respect to SNR. Recalling (23), using the eigenvalue decompositions $\Theta_R = \mathbf{U} \Lambda_R \mathbf{U}^\dagger$, $\Theta_T = \mathbf{V} \Lambda_T \mathbf{V}^\dagger$ and $\Phi = \mathbf{V} \mathbf{P} \mathbf{V}^\dagger$, and exploiting the unitary invariance of \mathbf{W} ,

$$\dot{\mathcal{I}}(\text{SNR}, \Phi) = \sum_{i=1}^{n_R} E \left[\frac{\lambda_i(\tilde{\mathbf{H}} \tilde{\mathbf{H}}^\dagger)}{n_T + \text{SNR} \lambda_i(\tilde{\mathbf{H}} \tilde{\mathbf{H}}^\dagger)} \right] \log_2 e \quad (141)$$

where $\tilde{\mathbf{H}} = \left(\Lambda_R^{1/2} \mathbf{W} \Lambda_T^{1/2} \mathbf{P}^{1/2} \right)$ and $\lambda_i(\cdot)$ denotes the i th eigenvalue. Thus,

$$\lim_{\text{SNR} \rightarrow \infty} \text{SNR} \dot{\mathcal{I}}(\text{SNR}, \Phi) = n_R (1 - \text{Pr}\{\lambda=0\}) \log_2 e \quad (142)$$

where λ is a random variable whose distribution equals the marginal distribution of an unordered eigenvalue of $\tilde{\mathbf{H}} \tilde{\mathbf{H}}^\dagger$. Since the entries of $\tilde{\mathbf{H}}$ are independent and $\tilde{\mathbf{H}} \tilde{\mathbf{H}}^\dagger$ has n_R eigenvalues,

$$\text{Pr}\{\lambda=0\} = \max \left(1 - \frac{m}{n_R}, 1 - \frac{n'_R}{n_R} \right) \quad (143)$$

where, recall, n'_R is the number of nonzero entries in the diagonal matrix $\Lambda_R^{1/2}$ while m is introduced to indicate the number of nonzero entries in the diagonal matrix $\Lambda_T^{1/2} \mathbf{P}^{1/2}$. Combining (142) and (143),

$$S_\infty(\Phi) = \min(m, n'_R) \quad (144)$$

and, given that $m \leq n'_T$, the claim is proved.

If and only if the input covariance is admissible, then $m = n'_T$ and (144) is met with strict equality.

E Proof of Proposition 4

Exploiting the unitary invariance of the distribution of a matrix \mathbf{W} with IID Rayleigh-faded entries and the eigenvector commonality of Φ and Θ_T ,

$$\log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \Theta_R^{1/2} \mathbf{W} \Theta_T^{1/2} \Phi \Theta_T^{1/2} \mathbf{W}^\dagger \Theta_R^{1/2} \right) = \log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \mathbf{W} \Lambda_T \mathbf{P} \mathbf{W}^\dagger \Lambda_R \right) \quad (145)$$

Let \mathbf{W}' be a $n'_R \times n'_T$ matrix obtained by eliminating from \mathbf{W} the rows and columns whose indices correspond, respectively, with those of zero eigenvalues of Θ_R and Θ_T . Using the fact that, given a matrix \mathbf{A} , $\det(\mathbf{I} + \mathbf{A})$ is unaffected by the removal of any rows or columns of \mathbf{A} that are identically zero, the left-hand side of (145) can be further seen to equal

$$\log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \Theta_R^{1/2} \mathbf{W}' \Theta_T^{1/2} \Phi \Theta_T^{1/2} \mathbf{W}'^\dagger \Theta_R^{1/2} \right) = \log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \mathbf{W}' \Lambda'_T \mathbf{P}' (\mathbf{W}')^\dagger \Lambda'_R \right) \quad (146)$$

For $n'_T \geq n'_R$, we have that $S_\infty = n'_R$ and that $(\mathbf{W}' \Lambda'_T \mathbf{P}' (\mathbf{W}')^\dagger \Lambda'_R)$ is nonsingular with probability 1. From (11),

$$\mathcal{L}_\infty(\Phi) = \log_2 n_T - \frac{1}{n'_R} E \left[\log_2 \det (\mathbf{W}' \Lambda'_T \mathbf{P}' (\mathbf{W}')^\dagger \Lambda'_R) \right] \quad (147)$$

$$= \log_2 n_T - \frac{1}{n'_R} \log_2 e E \left[\log_e \det (\mathbf{W}' \Lambda'_T \mathbf{P}' (\mathbf{W}')^\dagger) \right] - \frac{1}{n'_R} \log_2 \det(\Lambda'_R) \quad (148)$$

from which, applying Lemma 2, the corresponding expression in Proposition 4 follows.

For $n'_T \leq n'_R$, we have that $S_\infty = n'_T$ and that $((\mathbf{W}')^\dagger \Lambda'_R \mathbf{W}' \Lambda'_T \mathbf{P}')$ is nonsingular with probability 1. From (11),

$$\mathcal{L}_\infty(\Phi) = \log_2 n_T - \frac{1}{n'_T} E \left[\log_2 \det ((\mathbf{W}')^\dagger \Lambda'_R \mathbf{W}' \Lambda'_T \mathbf{P}') \right] \quad (149)$$

$$= \log_2 n_T - \frac{1}{n'_T} \log_2 e E \left[\log_e \det ((\mathbf{W}')^\dagger \Lambda'_R \mathbf{W}') \right] - \frac{1}{n'_T} \log_2 \det(\Lambda'_T \mathbf{P}') \quad (150)$$

from which, applying again Lemma 2, the corresponding expression in Proposition 4 follows.

For $n'_T = n'_R$,

$$E [\log_2 \det (\mathbf{W}' \Lambda'_T \mathbf{P}' (\mathbf{W}')^\dagger \Lambda'_R)] = E [\log_2 \det (\mathbf{W}' (\mathbf{W}')^\dagger)] + \log_2 \det (\Lambda'_T \mathbf{P}' \Lambda'_R) \quad (151)$$

from which, applying (132), the corresponding expression in Proposition 4 is found.

F Proof of Proposition 5

As argued in Appendix E, the removal of the projection of the channel on the null spaces of Θ_T and Θ_R is inconsequential. Thus, we can restrict our consideration to $n'_R \times n'_T$ channel matrices whose correlation matrices are nonsingular.

From [5, Prop. 4] (see also [14]), as $n'_T, n'_R \rightarrow \infty$ with ratio $\beta' = n'_T/n'_R$,

$$\frac{1}{n'_R} \mathcal{I}(\text{SNR}, \Phi) \rightarrow \beta' E [\log_2(1 + \Lambda' \text{SNR } \Psi)] + E [\log_2(1 + \Lambda'_R \text{SNR } \Upsilon)] - \beta' \text{SNR } \Psi \Upsilon \log_2 e \quad (152)$$

with Ψ and Υ the solutions to

$$\Psi = \frac{1}{\beta'} E \left[\frac{\Lambda'_R}{1 + \Lambda'_R \text{SNR } \Upsilon} \right] \quad (153)$$

$$\Upsilon = E \left[\frac{\Lambda'}{1 + \Lambda' \text{SNR } \Psi} \right] \quad (154)$$

Let us study the behavior of Ψ for $\text{SNR} \rightarrow \infty$, to which end we define

$$\varphi \triangleq \lim_{\text{SNR} \rightarrow \infty} \Psi. \quad (155)$$

Under the hypothesis that $\varphi > 0$, from (153) we obtain

$$E \left[\frac{1}{1 + \frac{\Lambda'_R}{\varphi}} \right] = 1 - \beta' \quad (156)$$

Since φ is nonnegative:

- If $\beta' > 1$, then (156) does not admit a nonzero solution and thus the above hypothesis is contradicted. Thus, $\varphi = 0$.
- If $\beta' = 1$, then (156) only admits the solution $\varphi = 0$. Again, the hypothesis $\varphi > 0$ is contradicted confirming that indeed $\varphi = 0$.
- If $\beta' < 1$, a nonzero solution is admitted and (156) becomes the claimed equation for φ .

Since, for $\beta' \geq 1$, φ vanishes, we need to evaluate the limiting value of $\text{SNR} \Psi$, to which end we define

$$\alpha \triangleq \lim_{\text{SNR} \rightarrow \infty} \text{SNR} \Psi. \quad (157)$$

For $\beta' \leq 1$, α diverges. For $\beta' > 1$, and under the hypothesis that $0 < \alpha < \infty$, we can again use (153) to obtain

$$E \left[\frac{1}{1 + \alpha \Lambda'} \right] = 1 - \frac{1}{\beta'} \quad (158)$$

which admits a unique nonzero solution.

Having characterized the behavior of Ψ and α for $\text{SNR} \rightarrow \infty$, we can finally turn to \mathcal{L}_∞ , which in the large-dimensional regime is obtained as

$$\mathcal{L}_\infty(\Phi) = \lim_{\text{SNR} \rightarrow \infty} \log_2 \text{SNR} - \frac{\mathcal{I}(\text{SNR}, \Phi)/n'_R}{\min(\beta', 1)} \quad (159)$$

application of which, together with (152), (156) and (158), yields the claimed expressions.

G Proof of Propositions 6

Since Φ is restricted to being admissible, the null space of Λ_T contains the null space of the optimum \mathbf{P} . In order to prove that these null spaces in fact coincide, we need to show that the null space of the optimum \mathbf{P} contains the null space of Λ_T .

From (23), and exploiting the unitary invariance of the distribution of \mathbf{W} and the eigenspace commonality of Φ and Θ_T , we have that

$$\log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \Theta_R^{1/2} \mathbf{W} \Theta_T^{1/2} \Phi \Theta_T^{1/2} \mathbf{W}^\dagger \Theta_R^{1/2} \right) = \log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \mathbf{W} \Lambda_T \mathbf{P} \mathbf{W}^\dagger \Lambda_R \right) \quad (160)$$

Consider a \mathbf{P} whose null space does not contain that of Λ_T . We can decompose $\mathbf{P} = \mathbf{P}_1 + \mathbf{P}_2$ with \mathbf{P}_2 the (non-empty) projection of \mathbf{P} over the null space of Λ_T . Thus, the null space of \mathbf{P}_1 does contain the null space of Λ_T . Then,

$$\log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \mathbf{W} \Lambda_T \mathbf{P} \mathbf{W}^\dagger \Lambda_R \right) = \log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \mathbf{W} \Lambda_T \mathbf{P}_1 \mathbf{W}^\dagger \Lambda_R \right) \quad (161)$$

while $\text{Tr}\{\mathbf{P}_1\} < \text{Tr}\{\mathbf{P}\}$. A strictly higher value can be obtained in the left-hand side of (161), abiding by the trace constraint, by using $\frac{n_T}{\text{Tr}\{\mathbf{P}_1\}} \mathbf{P}_1$ rather than \mathbf{P} . Hence, no power allocation can be optimal if its null space does not contain that of Λ_T . (Note that this optimality condition holds at every SNR while the condition that the input covariance be admissible holds only for $\text{SNR} \rightarrow \infty$.)

H Proof of Proposition 7

The sought \mathbf{P}^* is given by the matrix \mathbf{P}' that minimizes the power offset expressions in Proposition 4 while satisfying the trace constraint in (42). For $n'_T \leq n'_R$, it suffices to apply Jensen's inequality to $\log_2 \det(\mathbf{P}')$. For $n'_T > n'_R$, \mathbf{P}^* must minimize the uppermost expression. This leads to the equivalent maximization in (44).

I Proof of Proposition 9

For $n_T \geq n_R$, we have that $S_\infty = n_R$ and (11) yields

$$\mathcal{L}_\infty = \log_2 n_T - \frac{1}{n_R} \log_2 \det(\mathbf{H}\mathbf{H}^\dagger). \quad (162)$$

Applying Lemma 3 with proper scaling of the deterministic and random components and identifying (15), the claimed expression is found. For $n_T \leq n_R$, the same derivation applies starting from

$$\mathcal{L}_\infty = \log_2 n_T - \frac{1}{n_T} \log_2 \det(\mathbf{H}^\dagger \mathbf{H}) \quad (163)$$

To specialize the result to Corollary 12, it suffices to let $\phi_1 = K n_T n_R$ while $\phi_i \rightarrow 0 \forall i \neq 1$.

J Proof of Propositions 10 and 11

For $n_T \geq n_R$, from (11) and (76),

$$\mathcal{L}_\infty = \log_2 \frac{n_T}{n_I} - \frac{\log_2 e}{n_R} \left(E \left[\log_e \det(\mathbf{H}\mathbf{H}^\dagger) \right] - \left[\log_e \det(\mathbf{H}_I \mathbf{H}_I^\dagger) \right] \right) \quad (164)$$

whose expectations are given in (132), which can in turn be expanded using (136). For $n_T < n_R$,

$$\mathcal{L}_\infty = \log_2 \frac{n_T}{n_I} - \frac{\log_2 e}{n_T} E \left[\log_e \det \left(\mathbf{H}^\dagger \left(\mathbf{H}_I \mathbf{H}_I^\dagger \right)^{-1} \mathbf{H} \right) \right] \quad (165)$$

whose expectation is given by Lemma 4. Blending (164) and (165), Proposition 10 is found.

In turn, defining

$$\Sigma \triangleq \frac{\mathbf{H}_I \mathbf{H}_I}{n_I}, \quad (166)$$

the expressions in Proposition 11 can be obtained from those in Proposition 5 by recognizing that Σ^{-1} plays the role of a receive correlation matrix.¹¹ Hence, it suffices to specialize

¹¹The fact that Σ is random becomes asymptotically immaterial as the empirical distribution of its eigenvalues converges to a deterministic function.

Proposition 5 with $\Lambda' = 1$ and with Λ'_R replaced by another random variable, $\Lambda_{\Sigma^{-1}}$, whose distribution of equals that of an inverse Wishart matrix. This distribution satisfies [14]

$$E[\log_2 \Lambda_{\Sigma^{-1}}] = (\beta_I - 1) \log_2 \left(1 - \frac{1}{\beta_I}\right) + \log_2 e \quad (167)$$

$$E\left[\frac{1}{1 + \frac{\Lambda_{\Sigma^{-1}}}{\varphi}}\right] = \frac{\beta_I}{4\varphi} \mathcal{F}\left(\beta_I, \frac{\varphi}{\beta_I}\right) \quad (168)$$

$$E\left[\log_2 \left(1 + \frac{\varphi}{\Lambda_{\Sigma^{-1}}}\right)\right] = \log_2 \left(1 + \varphi - \frac{1}{4} \mathcal{F}\left(\frac{\varphi}{\beta_I}, \beta_I\right)\right) + \beta_I \log_2 \left(1 + \frac{\varphi}{\beta_I} - \frac{1}{4} \mathcal{F}\left(\frac{\varphi}{\beta_I}, \beta_I\right)\right) - \frac{\beta_I}{4\varphi} \mathcal{F}\left(\frac{\varphi}{\beta_I}, \beta_I\right) \log_2 e \quad (169)$$

where

$$\mathcal{F}(x, z) = \left(\sqrt{x(1 + \sqrt{z})^2 + 1} - \sqrt{x(1 - \sqrt{z})^2 + 1}\right)^2 \quad (170)$$

Using these identities, the claimed closed-form expressions emerge from Proposition 5.

References

- [1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom*, vol. 10, pp. 585–595, Nov. 1999.
- [2] G. J. Foschini and M. J. Gans, "On the limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 315–335, Mar. 1998.
- [3] D. Tse and P. Viswanath, *Fundamentals of Wireless communications*, Cambridge University Press, 2005.
- [4] C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, "Capacity scaling in dual-antenna-array wireless systems," *IEEE Trans. on Inform. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.
- [5] A. M. Tulino, S. Verdú, and A. Lozano, "Capacity of antenna arrays with space, polarization and pattern diversity," *Proc. IEEE Inform. Theory Workshop (ITW'03)*, pp. 324–327, Mar. 2003.
- [6] A. Lapidoth and S. Shamai, "Fading channels: How perfect need 'perfect side information' be?," *IEEE Trans. on Inform. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.
- [7] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multi-antenna systems on flat-fading channels," *IEEE Trans. on Inform. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

- [8] S. Shamai and S. Verdú, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. on Inform. Theory*, vol. 47, no. 5, pp. 1302–1327, May 2001.
- [9] W. C. Y. Lee, "Estimate of channel capacity in Rayleigh fading environments," *IEEE Trans. Veh. Technol.*, vol. 39, pp. 187–189, Aug. 1990.
- [10] A. M. Sengupta and P. P. Mitra, "Capacity of multivariate channels with multiplicative noise: I. Random matrix techniques and large-N expansions for full transfer matrices," *LANL arXiv:physics/0010081*, Oct. 2000.
- [11] Ö. Oyman, R. U. Nabar, H. Bölcskei, and A. J. Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels," *IEEE Trans. on Signal Processing*, vol. 51, no. 11, pp. 2784–2795, Nov. 2003.
- [12] A. Grant, "Rayleigh fading multiple-antenna channels," *EURASIP J. on Applied Signal Processing*, vol. 2002, no. 3, pp. 316–329, Mar. 2002.
- [13] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering and keyhole," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 2636–2647, Oct. 2003.
- [14] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.
- [15] E. Jorswieck and H. Boche, "Optimal power transmission with imperfect channel state information at the transmit antenna array," *Wireless Personal Communications*, vol. 27, pp. 33–56, Oct. 2003.
- [16] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of correlation on the capacity of multi-antenna channels," to appear in *IEEE Trans. on Inform. Theory*, 2005.
- [17] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.
- [18] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. on Inform. Theory*, vol. 47, pp. 2632–2639, Sept. 2001.
- [19] S. A. Jafar, S. Vishwanath, and A. J. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems with imperfect feedback," *IEEE Trans. on Wireless Communications*, vol. 3, pp. 1165–1175, July 2004.
- [20] S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 406–417, Apr. 2003.
- [21] E. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Communications*, vol. 3, no. 5, pp. 1543–1553, Sept. 2004.

- [22] J. Kotecha and A. Sayeed, "On the capacity of correlated MIMO channels," *Proc. of Intern. Symp. on Inform. Theory (ISIT'03)*, July 2003.
- [23] T.-S. Chu and L. J. Greenstein, "A semiempirical representation of antenna diversity gain at cellular and PCS base stations," *IEEE Trans. Commun.*, vol. 45–46, pp. 644–646, June 1997.
- [24] E. Biglieri, A. M. Tulino, and G. Taricco, "How far is infinity ? Using asymptotic analyses in multi-antenna systems," *Intern. Symp. on Spread Spectrum Techn. & Applic. (ISSSTA'02)*, vol. 1, pp. 1–6, Sept. 2002.
- [25] A. Lozano and A. M. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Trans. on Inform. Theory*, vol. 48, no. 12, pp. 3117–3128, Dec. 2002.
- [26] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: a (not so) large N analysis," *IEEE Trans. on Inform. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.
- [27] A. M. Tulino, A. Lozano, and S. Verdú, "Capacity-achieving input covariance for single-user multi-antenna channels," *IEEE Trans. on Wireless Communications (to appear)*, 2005.
- [28] A. Lozano, A. M. Tulino, and S. Verdú, "Multiantenna capacity: Myths and realities," *Space-Time Wireless Systems: From Array Processing to MIMO Communications (Editors: H. Bölcskei, D. Gesbert, C. Papadias and J. Van der Veen)*, Cambridge University Press, 2005.
- [29] E. Jorswieck and H. Boche, "Average mutual information in spatial correlated MIMO systems with uninformed transmitter," *Proc. Conf. on Inform. Sciences and Systems (CISS'04)*, pp. 30–35, Mar. 2004.
- [30] H. L. Royden, *Real Analysis*, Prentice Hall; 3rd Edition, 1988.
- [31] William C. Jakes, *Microwave Mobile Communications*, New York, IEEE Press, 1974.
- [32] A. Lozano, A. M. Tulino, and S. Verdú, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 2527–2544, Oct. 2003.
- [33] E. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information," *IEEE Trans. on Signal Processing*, vol. 52, no. 12, pp. 3440–3453, Dec. 2004.
- [34] J. Ling, D. Chizhik, P. W. Wolniansky, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple transmit multiple receive (MTMR) capacity survey in Manhattan," *IEE Electronic Letters*, vol. 37, no. 16, pp. 1041–1042, Aug. 2001.
- [35] D. Hoesli and A. Lapidotoh, "How good is an isotropic input on a MIMO Ricean channel?," *Proc. of IEEE Int. Symp. on Inform. Theory (ISIT'04)*, July 2004.
- [36] J. Hansen and H. Bolcskei, "A geometrical investigation of the rank-1 Ricean MIMO channel at high SNR," *Proc. of IEEE Int. Symposium on Inform. Theory (ISIT'04)*, p. 64, July 2004.

- [37] Y.-H. Kim and A. Lapidoth, "On the log-determinant of non-central Wishart matrices," *Proc. IEEE Int. Symp. on Inform. Theory*, pp. 54–54, July 2004.
- [38] D. Hoesli, Y.-H. Kim, and A. Lapidoth, "Monotonicity results for coherent MIMO Ricean channels," *to appear in IEEE Trans. on Inform. Theory*, 2005.
- [39] Y. Q. Yin, "Limiting spectral distribution for a class of random matrices," *Journal of Multivariate Analysis*, vol. 20, pp. 50–68, 1986.
- [40] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices," *Statistica Sinica*, vol. 9, no. 3, pp. 611–661, 1999.
- [41] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 1073–1096, May 2003.
- [42] H. El Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Trans. on Inform. Theory*, vol. 50, no. 6, pp. 968–985, June 2004.
- [43] S. Tavildar and P. Viswanath, "Permutation codes for the parallel fading channel: Achieving the diversity-multiplexing tradeoff," *Proc. Conf. on Information Sciences and Systems (CISS'04)*, Mar. 2004.
- [44] N. Jindal, "High SNR analysis of MIMO broadcast channels," *Proc. of IEEE Int. Symp. on Inform. Theory (ISIT'05)*, Sept. 2005.
- [45] M. Chiani, M. Z. Win, and A. Zanella, "On the capacity of spatially correlated MIMO Rayleigh-fading channels," *IEEE Trans. on Inform. Theory*, vol. 49, no. 10, pp. 2363–2371, Oct. 2003.
- [46] C. G. Khatri, "On the moments of traces of two matrices in three situations for complex multivariate normal populations," *Sankhya: The Indian Journal of Statistics*, vol. 32, series A, pp. 65–80, 1970.
- [47] M. C. Andréief, "Note sur une relation entre les intégrals définies des produits des fonctions," *Mém. de la Soc. Sci. de Boredeaux*, vol. 2, pp. 1–14, 1883.
- [48] H. Gao and P.J. Smith, "A determinant representation for the distribution of quadratic forms in complex normal vectors," *Journal of Multivariate Analysis*, vol. 73, no. 2, pp. 155–165, May 2000.
- [49] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [50] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, New York: Dover Publications, 1972.
- [51] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *Ann. Math. Statistics* 35, pp. 475–501, 1964.

- [52] K. I. Gross and D. S. P. Richards, "Total positivity, spherical series, and hypergeometric functions of matrix argument," *J. Approx. Theory*, vol. 59, no. 2, pp. 224–246, 1989.
- [53] A. Y. Orlov, "New solvable matrix integrals," *International Journal of Modern Physics*, vol. 19, pp. 276–293, 2004.
- [54] C. G. Khatri, "On certain distribution problems based on positive definite quadratic functions in normal vectors," *The Annals of Mathematical Statistics*, vol. 37, no. 2, pp. 467–479, April 1966.
- [55] R. J. Muirhead, *Aspects of multivariate statistical theory*, New York, Wiley, 1982.
- [56] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*, Chapman & Hall/CRC, 2000.
- [57] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. on Inform. Theory*, vol. 45, pp. 622–640, Mar. 1999.