Mutual Information of IID Complex Gaussian Signals on Block Rayleigh-faded Channels

Fredrik Rusek, Angel Lozano and Nihar Jindal

Abstract—**We present a method to compute, quickly and efficiently, the mutual information achieved by an IID (independent identically distributed) complex Gaussian signal on a block Rayleigh-faded channel without side information at the receiver. The method accommodates both scalar and MIMO (multiple-input multiple-output) settings. Operationally, this mutual information represents the highest spectral efficiency that can be attained using Gaussian codebooks. Examples are provided that illustrate the loss in spectral efficiency caused by fast fading and how that loss is amplified when multiple transmit antennas are used. These examples are further enriched by comparisons with the channel capacity under perfect channel-state information at the receiver, and with the spectral efficiency attained by pilot-based transmission.**

Keywords: Wireless communications; Mutual information; MIMO; Multiantenna; Non-coherent, Spectral efficiency

I. Introduction

IID (independent identically distributed) complex Gaussian inputs are highly relevant in channels impaired by Gaussian noise. Some of the arguments for this relevance are that, with side information in the form of perfect CSI (channel state information) at the receiver:

- These are the unique capacity-achieving inputs.
- Their mutual information represents very well the mutual information of proper complex discrete constellations (e.g., QAM) commonly used in wireless systems. (This holds up to some power level that depends on the cardinality of the constellation [1].)

Indeed, expressions for the perfect-CSI capacity achieved by IID complex Gaussian inputs are available (cf. Section IV) and thus such capacity can be easily evaluated.

Remove now the side information. No expressions are available for the mutual information achieved by IID complex Gaussian inputs save for the very special case of memoryless channels [2]. Moreover, straight Monte-Carlo computation is not feasible because it would entail large-dimensional histograms. Only bounds [3], [4] and asymptotic low- and high-power expansions [5]–[8] are available for such mutual information. And yet, although no longer capacity-achieving without perfect CSI [9], [10], IID complex Gaussian inputs remain highly relevant. Operationally, the mutual information they achieve represents the highest spectral efficiency that can be attained using Gaussian codebooks. In fact, the capacity-achieving inputs in the absence of perfect-CSI become largely unfeasible in certain cases (in the low-power regime, for example, they become unacceptably peaky) and thus the capacity is sometimes less relevant to system designers than the mutual information of IID complex Gaussian inputs.

This paper presents analytical expressions for the output distributions of a block Rayleigh-faded channel fed with IID complex Gaussian inputs. Then, a simple outer Monte-Carlo in conjunction with these distributions yields a semi-analytical method that allows evaluating, quickly and efficiently, the mutual information. The method accommodates not only scalar channels, but also MIMO (multiple-input multiple-output) settings. Altogether, this allows answering questions such as:

- What is the impact of assuming side information?
- How close to the true channel capacity (without side information) can IID complex Gaussian inputs operate?
- How suboptimal are pilot-based schemes, i.e., schemes that form explicit channel estimates on the basis of pilot observations at the receiver and subsequently apply them to detect the data?
- At which power level is the power efficiency maximized (i.e., the energy per bit minimized)?

Not surprisingly, the answers to these questions end up being a function of the fading rate and the numbers of antennas. With the method presented, these relationships can be precisely established, and some examples of this are provided in the paper.

II. Channel Model

Consider n_{T} transmit and n_{R} receive antennas and let the $n_{\mathrm{R}} \times n_{\mathrm{T}}$ matrix $\boldsymbol{H}$ represent the discrete-time fading channel. Under block Rayleigh-fading, the channel entries are drawn from a zero-mean unit-variance complex Gaussian distribution at the beginning of each fading block and they remain constant for the n_{b} symbols within that block, where n_{b} is the coherence time in symbols (or the coherence bandwidth if it is the frequency domain being modeled). This process is repeated for

Fredrik Rusek is with Lund University, 221 00 Lund, Sweden.

Angel Lozano is with Universitat Pompeu Fabra (UPF), Barcelona 08018, Spain. His work is supported by the Spanish Ministry of Science and Innovation (Refs. TEC2009-13000 and CONSOLIDER-INGENIO CSD2008-00010 ”COMONSENS”) and by AGAUR (Catalan Agency for Administration of University and Research Grants).

Nihar Jindal is with the University of Minnesota, Minneapolis, MN 55455.

every block in an IID fashion. There is no antenna correlation and thus the entries of $\boldsymbol{H}$ are also independent.

Assembling into matrices the input, the output, and the noise for the n_{b} symbols within each block, their relationship becomes

$$\boldsymbol{Y} = \sqrt{\frac{\mathsf{SNR}}{n_{\mathrm{T}}}}\boldsymbol{H}\boldsymbol{X} + \boldsymbol{N} \tag{1}$$

where the input $\boldsymbol{X}$ is an $n_{\mathrm{T}} \times n_{\mathsf{b}}$ matrix while the output $\boldsymbol{Y}$ and the noise $\boldsymbol{N}$ are $n_{\mathrm{R}} \times n_{\mathsf{b}}$ matrices. Both $\boldsymbol{X}$ and $\boldsymbol{N}$ have IID zero-mean unit-variance complex Gaussian entries. With that, SNR indicates the average signal-to-noise ratio per receive antenna.

Each codeword spans a large number of fading blocks, in the time and/or frequency domains, which endows ergodic quantities with operational meaning.

Although a block-fading structure admittedly represents a drastic simplification of reality, it does capture the essential nature of fading and generally yields results that are remarkably similar to those obtained with continuous-fading models [11]. In fact, for a rectangular Doppler spectrum, an exact correspondence in terms of the estimation of $\boldsymbol{H}$ can be established between block- and continuous-fading models [12] whereby

$$n_{\mathsf{b}} = \frac{1}{2 f_{\mathsf{m}} T_{\mathrm{s}}} \tag{2}$$

where f_{m} and T_{s} are the maximum Doppler frequency and the symbol period, respectively. Typically, $f_{\mathsf{m}} = (v/c) f_{\mathsf{c}}$ with v the velocity and f_{c} the carrier frequency. The mapping in (2) is in terms of the minimum mean-square error in the estimation of $\boldsymbol{H}$, and thus it is exact for pilot-based schemes that rely on such explicit estimation, but more broadly we take it as indicative of the fading rate represented by a given value of n_{b}.

III. Computation of the Mutual Information

The mutual information under investigation can be expressed as

$$\bar{\mathcal{I}} = \frac{1}{n_{\mathsf{b}}} \left[\mathfrak{h}(\boldsymbol{Y}) - \mathfrak{h}(\boldsymbol{Y}|\boldsymbol{X}) \right] \tag{3}$$

where $\mathfrak{h}(\cdot)$ denotes the differential entropy operator.

Our first result leverages the derivations in [13] to obtain a closed-form expression for $\mathfrak{h}(\boldsymbol{Y}|\boldsymbol{X})$.

Proposition 1 *Let $E_q(\cdot)$ denote the exponential integral, i.e., $E_q(\zeta) = \int_1^\infty t^{-q} e^{-\zeta t} dt$. Then,*

$$\mathfrak{h}(\boldsymbol{Y}|\boldsymbol{X}) = n_{\mathrm{R}} \log_2(e)\, e^{n_{\mathrm{T}}/\mathsf{SNR}} \sum_{i=0}^{n_{\mathrm{T}}-1} \sum_{j=0}^{i} \sum_{\ell=0}^{2j} \Bigg[\binom{2i-2j}{i-j} \cdot \binom{2j+2n_{\mathsf{b}}-2n_{\mathrm{T}}}{2j-\ell} \frac{(-1)^\ell (2j)!\,(n_{\mathsf{b}}-n_{\mathrm{T}}+\ell)!}{2^{2i-\ell}\, j!\,\ell!\,(n_{\mathsf{b}}-n_{\mathrm{T}}+j)!} \sum_{q=0}^{n_{\mathsf{b}}-n_{\mathrm{T}}+\ell} E_{q+1}\left(\frac{n_{\mathrm{T}}}{\mathsf{SNR}}\right) \Bigg] + n_{\mathrm{R}} n_{\mathsf{b}} \log_2(\pi e) \tag{4}$$

Proof: *See Appendix.*

Having expressed $\mathfrak{h}(\boldsymbol{Y}|\boldsymbol{X})$ in (3), we now turn our attention to $\mathfrak{h}(\boldsymbol{Y})$. The output $\boldsymbol{Y}$ is affected by a combination of multiplicative and additive noise, and thus $\boldsymbol{Y}$ is not Gaussian distributed. The unconditional output distribution $p(\boldsymbol{Y})$, which had not been reported in the literature to the best of our knowledge, constitutes the central result in this paper. While the formula we obtain appears too involved to allow for a closed-form expression of $\mathfrak{h}(\boldsymbol{Y})$, the formula is easy to evaluate and it thereby enables computing

$$\mathfrak{h}(\boldsymbol{Y}) = -\int p(\boldsymbol{Y}) \log_2 p(\boldsymbol{Y}) \mathrm{d}\boldsymbol{Y} \tag{5}$$

$$= -\mathbb{E}\left[\log_2 p(\boldsymbol{Y})\right]. \tag{6}$$

through straight Monte-Carlo averaging. Our formula for $p(\boldsymbol{Y})$ is given in the next proposition, where we use the standard notation $[z]^+ = \max\{z, 0\}$.

Proposition 2 *For $1 \le k \le \min(n_{\mathrm{R}}, n_{\mathrm{T}})$, define the functions*

$$f_k(x) = \int_0^\infty \exp\left\{ \frac{x\,\mathsf{SNR}\,z}{z\mathsf{SNR} + n_{\mathrm{T}}} - z \right\} \frac{z^{k-1+[n_{\mathrm{T}}-n_{\mathrm{R}}]^+}}{(z\mathsf{SNR}/n_{\mathrm{T}}+1)^{n_{\mathsf{b}}+1-n_{\mathrm{R}}}} \mathrm{d}z. \tag{7}$$

Let $\boldsymbol{d} = [d_1, \ldots, d_{n_{\mathrm{R}}}]$ be the eigenvalues of $\boldsymbol{Y}\boldsymbol{Y}^\dagger$ and define the $n_{\mathrm{R}} \times n_{\mathrm{R}}$ matrix $\boldsymbol{Z}$ with entries

$$Z_{ij} = \begin{cases} \left(\dfrac{n_{\mathrm{T}}}{\mathsf{SNR}}\right)^{j-1} f_j(d_i), & 1 \le i \le n_{\mathrm{R}},\ 1 \le j \le s \\ d_i^{\,j-n_{\mathrm{T}}-1}, & 1 \le i \le n_{\mathrm{R}},\ s+1 \le j \le n_{\mathrm{R}} \end{cases} \tag{8}$$

where $s = \min\{n_{\mathrm{T}}, n_{\mathrm{R}}\}$. Then

$$p(\boldsymbol{Y}) = \frac{\pi^{-n_{\mathsf{b}} n_{\mathrm{R}}} e^{-\|\boldsymbol{Y}\|^2}}{\prod_{1 \le i,j \le n_{\mathrm{R}}} (d_j - d_i) \prod_{k=[n_{\mathrm{T}}-n_{\mathrm{R}}]^+}^{n_{\mathrm{T}}-1} k!} \det \boldsymbol{Z}. \tag{9}$$

Proof: *See Appendix.*

In the special case of memoryless channels, i.e., for $n_{\mathsf{b}} = 1$, the solution in Proposition 2 reduces to the one in [2].

We also note that, due to the rotational invariance of $\boldsymbol{H}$, only the eigenvalues of $\boldsymbol{Y}\boldsymbol{Y}^\dagger$ are relevant to the distribution in (9).

Using (4) and (9), an algorithm to compute $\bar{\mathcal{I}}$ can be put forth as follows.

Algorithm 1: Evaluation of $\bar{\mathcal{I}}$.
1. Pre-compute $f_k(x)$, $1 \le k \le n_{\mathrm{R}}$, on a discrete set $\mathcal{X}$ with a suitable stepsize $\Delta x = x_k - x_{k-1}$.
2. Generate a sufficiently large number of input and output vectors according to (1).
3. For each input and output pair, apply (9) to obtain $p(\boldsymbol{Y})$.
4. Compute the sample average of $-\log_2 p(\boldsymbol{Y})$ via Monte Carlo, thereby obtaining $\mathfrak{h}(\boldsymbol{Y})$.
5. Compute $\mathfrak{h}(\boldsymbol{Y}|\boldsymbol{X})$ from (4) and apply (3).

The accuracy can be made as high as desired by averaging over more input/output sample pairs and by increasing the precision in Step 1. For the results presented in Section V, the number of samples and the value of Δx were chosen such that two decimal digits are correct with 90% probability. With a standard workstation, the entire computation process is a matter of seconds.

As a final remark, we mention that Proposition 2 is easily extendable to include all input distributions $\boldsymbol{X}$ that are rotationally invariant and where the eigenvalue distribution of $\boldsymbol{X}^\dagger \boldsymbol{X}$ (or of $\boldsymbol{X}\boldsymbol{X}^\dagger$) is of the form

$$p(\lambda) = \det{}^2 \boldsymbol{V}(\boldsymbol{\lambda}) \prod_{k=1}^{s} g_k(\lambda_k),$$

where $\boldsymbol{V}(\cdot)$ denotes a Vandermonde matrix and the functions $g_k(\cdot)$ are arbitrary. Further details are given in the Appendix.

IV. Baselines

Before exemplifying the method described in Section III, we introduce the perfect-CSI capacity, a lower bound to $\bar{\mathcal{I}}$, and the spectral efficiency achievable with pilot-based communication, all of which serve as baselines.

A. Capacity with Perfect CSI

If the receiver is provided with perfect CSI on the side, the ergodic capacity, in bits/s/Hz, equals

$$C(\mathsf{SNR}) = \mathbb{E}\left[\log_2 \det\left(\boldsymbol{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}} \boldsymbol{H}\boldsymbol{H}^\dagger\right)\right] \tag{10}$$

closed forms for which can be found in [13], [14].

B. Mutual Information Lower Bound

A simple application of Jensen's inequality to the bound in [15, Theorem 2] yields the following.

Proposition 3 *The mutual information achieved by IID Gaussian inputs satisfies* $\bar{\mathcal{I}}(\mathsf{SNR}) \geq \bar{\mathcal{I}}_{\mathsf{lower}}(\mathsf{SNR})$ *with*

$$\bar{\mathcal{I}}_{\mathsf{lower}}(\mathsf{SNR}) = C(\mathsf{SNR}) - \frac{n_{\mathrm{T}} n_{\mathrm{R}}}{n_{\mathsf{b}}} \log_2\left(1 + \mathsf{SNR}\frac{n_{\mathsf{b}}}{n_{\mathrm{T}}}\right). \tag{11}$$

C. Pilot-Based Communication

In pilot-based communication, n_{p} pilot symbols are inserted within each fading block, leaving $n_{\mathsf{b}} - n_{\mathsf{p}}$ symbols available for data. The channel is estimated on the basis of the pilot observations at the receiver, and this estimate is subsequently utilized to detect the data. We analyze here the spectral efficiency achievable with separate processing of the pilots and the data symbols, which refers to estimating the channel on the basis of only the received pilots and then decoding the data (through nearest neighbor decoding) as if that estimate was perfect [16].

During the transmission of pilot symbols,

$$\boldsymbol{Y}_{\mathsf{p}} = \sqrt{\frac{\mathsf{SNR}}{n_{\mathrm{T}}}} \boldsymbol{H}\boldsymbol{P} + \boldsymbol{N}_{\mathsf{p}} \tag{12}$$

where the output, $\boldsymbol{Y}_{\mathsf{p}}$, and the noise, $\boldsymbol{N}_{\mathsf{p}}$, are $n_{\mathrm{R}} \times n_{\mathsf{p}}$ matrices. The entries of $\boldsymbol{N}_{\mathsf{p}}$ are IID zero-mean unit-variance complex Gaussian while $\boldsymbol{P}$ is deterministic and satisfies $\boldsymbol{P}\boldsymbol{P}^\dagger = n_{\mathsf{p}} \boldsymbol{I}$ [17].

During the transmission of data symbols, in turn, (1) applies with $\boldsymbol{X}$ and $\boldsymbol{N}$ of dimension $n_{\mathrm{T}} \times (n_{\mathsf{b}} - n_{\mathsf{p}})$ and $n_{\mathrm{R}} \times (n_{\mathsf{b}} - n_{\mathsf{p}})$, respectively.

The value of n_{p}, which can be optimized by solving a convex problem, depends on SNR, n_{b} and n_{T}. This optimization, and the ensuing spectral efficiency, has been studied extensively, e.g., [3], [5], [17]–[22]. In bits/s/Hz, such spectral efficiency equals

$$\max_{n_{\mathsf{p}}:1 \leq n_{\mathsf{p}} \leq n_{\mathsf{b}}} \left(1 - \frac{n_{\mathsf{p}}}{n_{\mathsf{b}}}\right) C\left(\frac{\mathsf{SNR}^2 n_{\mathsf{p}}/n_{\mathrm{T}}}{1 + \mathsf{SNR}\,(1 + n_{\mathsf{p}}/n_{\mathrm{T}})}\right) \tag{13}$$

where $C(\cdot)$ is the perfect-CSI capacity in (10).

If the pilot and data symbols are not required to have the same power, i.e., if pilot power-boosting is allowed, then it is optimal to set $n_{\mathsf{p}} = n_{\mathrm{T}}$ and to optimize only over the relative powers of pilots and data. This results in a different convex optimization, which in this case can be solved explicitly [17] leading to[1]

$$\left(1 - \frac{n_{\mathrm{T}}}{n_{\mathsf{b}}}\right) C\left(\frac{n_{\mathsf{b}}\mathsf{SNR}}{n_{\mathsf{b}} - 2\,n_{\mathrm{T}}} \left(\sqrt{\gamma} - \sqrt{\gamma - 1}\right)^2\right) \tag{14}$$

in bits/s/Hz, and with

$$\gamma = \frac{n_{\mathsf{b}}\mathsf{SNR} + n_{\mathrm{T}}}{n_{\mathsf{b}}\mathsf{SNR}\,\frac{n_{\mathsf{b}} - 2\,n_{\mathrm{T}}}{n_{\mathsf{b}} - n_{\mathrm{T}}}}. \tag{15}$$

The spectral efficiency in (14) is superior to that in (13). However, pilot power boosting increases the peakiness of the overall signal distribution, rendering it less amenable to efficient amplification.

V. Some Examples

Recalling (2), and in order to calibrate the relevant values of n_{b}, the following observations can be made in the context of emerging systems such as 3GPP LTE [23] or IEEE 802.16 WiMAX [24]:

- The carrier frequency f_c typically lies between 1 and 5 GHz.
- The symbol period is $T_{\mathrm{s}} \approx 100$ μs. However, it could be shortened to $T_{\mathrm{s}} \approx$ 10-20 μs and the flat-faded model in (1) would still apply. (For wider bandwidths, a frequency-selective model would be required and the computation algorithm would have to be modified accordingly.)
- Vehicular velocities up to $v \approx 120$ Km/h are of interest, and for high-speed trains this extends to $v \approx 300$ Km/h.

[1]Eq. (14) requires that $n_{\mathsf{b}} > 2\,n_{\mathrm{T}}$; variations thereof are also available for $n_{\mathsf{b}} \leq 2\,n_{\mathrm{T}}$ [17].

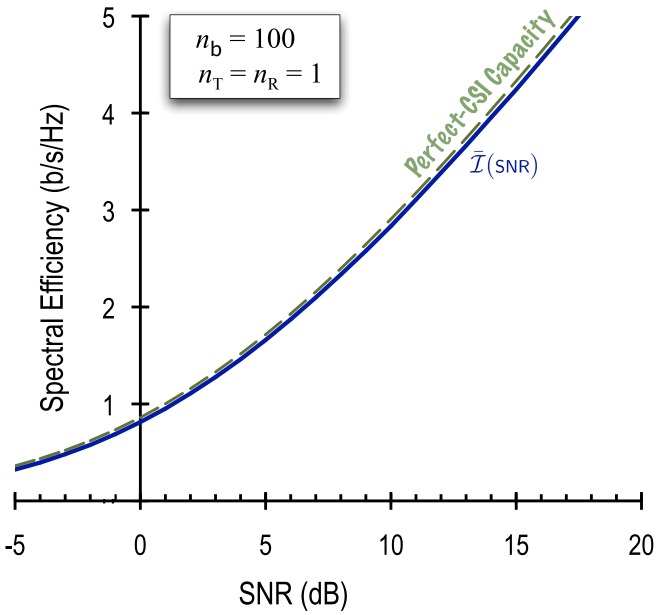

Fig. 1. In solid, $\bar{\mathcal{I}}(\mathsf{SNR})$ for $n_T = n_R = 1$ with $n_b = 100$. In dashed, the perfect-CSI capacity.

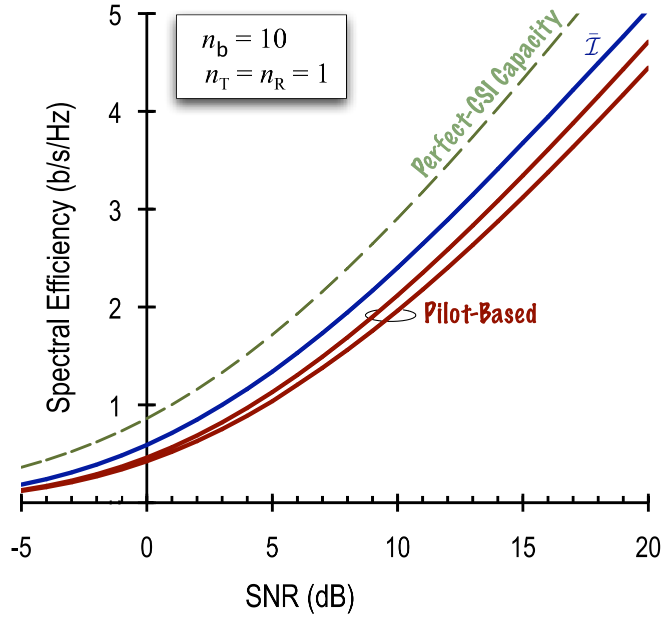

Fig. 2. In solid, $\bar{\mathcal{I}}(\mathsf{SNR})$ for $n_T = n_R = 1$ with $n_b = 10$. Also in solid, spectral efficiencies achieved by pilot-based communication, with and without pilot power boosting. In dashed, the perfect-CSI capacity.

With all of this taken into account, n_b can take values ranging from just over unity to several hundred. As the following example evidences, for large n_b the perfect-CSI capacity accurately represents the achievable mutual information.

Example 1 *Let $n_T = n_R = 1$ and let $n_b = 100$. Shown in Fig. 1 are the mutual information and the perfect-CSI capacity as function of* SNR.

For the remainder of this section, we shall thus focus on scenarios where n_b is small. Specifically, we shall use $n_b = 10$ and $n_b = 4$. These will tend to correspond to vehicular and high-speed-train velocities, possibly in conjunction with relatively long symbol periods and relatively high carrier frequencies.

Example 2 *Let $n_T = n_R = 1$ and let $n_b = 10$. Shown in Fig. 2 is the mutual information as function of* SNR. *Also shown are the spectral efficiencies achieved by pilot-based communication, with and without pilot power boosting, and the perfect-CSI capacity.*

We observe that a hefty share of the perfect-CSI capacity is achieved at high SNR, although this share diminishes with the SNR. We further observe that, by optimizing the pilot overhead or the pilot power boost at every SNR, pilot-based communication schemes can perform remarkably close to the fundamental communication limit of IID complex Gaussian inputs in this case.

Example 3 *Shown in Fig. 3 is a re-evaluation of Example 2 with $n_b = 4$.*

In this case, the relative gap between the perfect-CSI capacity and the achievable mutual information is

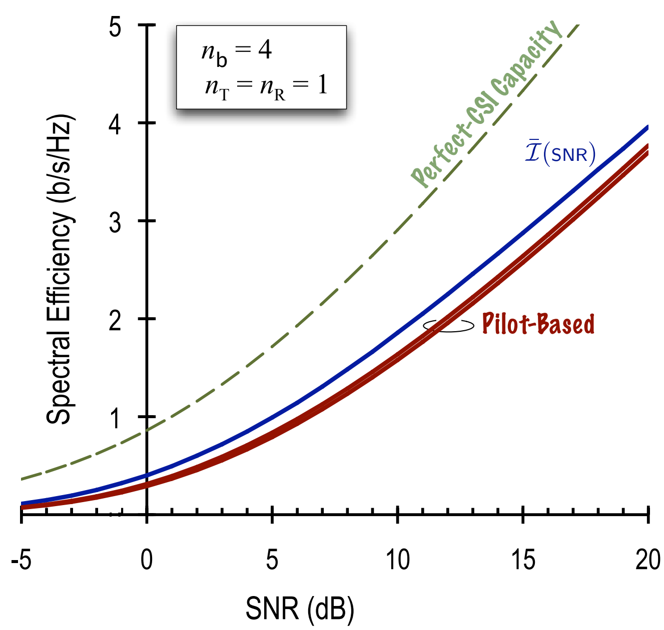

Fig. 3. In solid, $\bar{\mathcal{I}}(\mathsf{SNR})$ for $n_T = n_R = 1$ with $n_b = 4$. Also in solid, spectral efficiencies achieved by pilot-based communication, with and without pilot power boosting. In dashed, the perfect-CSI capacity.

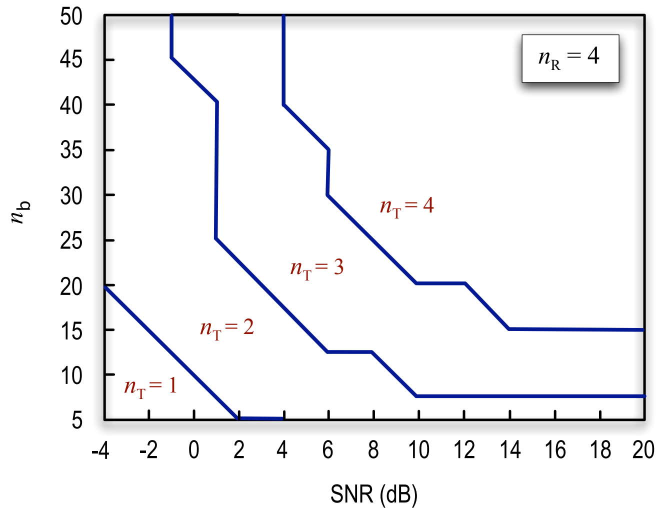

Fig. 4. Optimum n_T for $n_R = 4$ as function of SNR and n_b.

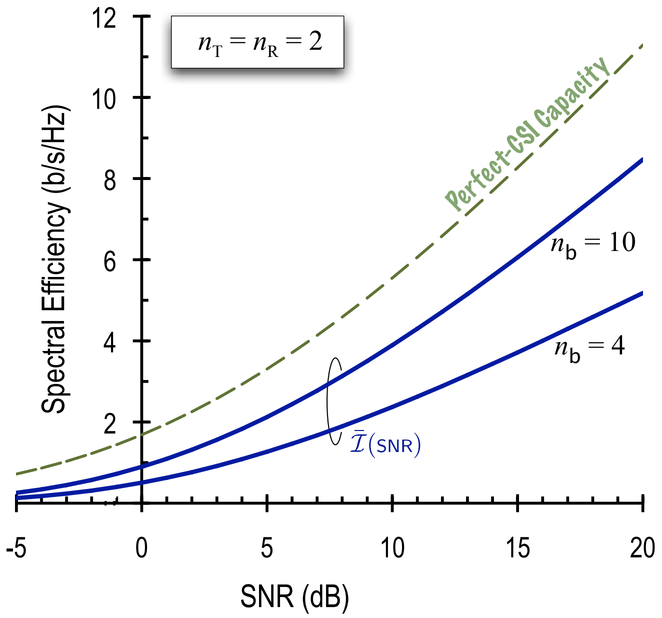

Fig. 5. In solid, $\bar{\mathcal{I}}(\text{SNR})$ for $n_T = n_R = 2$, with $n_b = 10$ and with $n_b = 4$. In dashed, the corresponding perfect-CSI capacity.

very substantial. (At 0 dB, less than half the perfect-CSI capacity can actually be achieved by IID complex Gaussian inputs.) The spectral efficiency of pilot-based schemes is similarly affected. Remarkably though, the performance of these schemes relative to the mutual information limit is essentially unaffected.

Let us now turn our attention to MIMO settings. A well-known feature of the perfect-CSI capacity is that it always increases with additional antennas, be it at the transmitter or at the receiver. However, [5] and [17] suggest that, without perfect CSI, activating too many transmit antennas would be detrimental at sufficiently high SNR. This is indeed the case, and the SNR above which a specific n_T becomes optimal depends on n_b as the following example illustrates.

Example 4 *Let $n_R = 4$. Shown in Fig. 4 is the optimum number of transmit antennas as function of both* SNR *and n_b.*

Next, we see the impact of varying n_b and/or SNR with fixed n_T and n_R.

Example 5 *Let $n_T = n_R = 2$. Shown in Fig. 5 is the mutual information as function of* SNR *with $n_b = 10$ and with $n_b = 4$. Also shown is the corresponding perfect-CSI capacity.*

Comparing Example 5 with Examples 2 and 3, notice how, at each fading rate, MIMO transmission suffers a more drastic loss relative to the perfect-CSI capacity. However, even with $n_b = 4$, the mutual information for $n_T = n_R = 2$ is larger than for $n_T = n_R = 1$ (Example 3). Thus, although additional transmit antennas should be activated only for sufficiently long n_b, additional transmit-receive pairs should be activated even for short n_b.

As a final and very illuminating example, we examine the scaling of the mutual information with the number of antennas for $n_T = n_R$.

Example 6 *Let $n_T = n_R$. Shown in Fig. 6 is the mutual information for* SNR $= 3$ *dB as function of $n_T = n_R$ with $n_b = 10$ and $n_b = 100$. Also shown is the perfect-CSI capacity.*

The linear scaling of the perfect-CSI capacity with $n_T = n_R$ is what fueled the early interest in MIMO. Without perfect CSI, the linear scaling is upheld approximately as long as the number of antennas is sufficiently small relative to n_b, but not otherwise. This is not just a limitation associated with the suboptimality of Gaussian inputs in the absence of perfect CSI, but rather a fundamental issue caused by channel uncertainty [9].

VI. The Low-SNR Regime

In power-limited conditions, power efficiency becomes relevant and the figure of merit that quantifies such efficiency is the energy per bit normalized by the noise spectral density. Measured at the receiver, this figure of merit equals

$$\frac{E_b^r}{N_0} = \frac{\text{SNR}}{R/B} n_R \tag{16}$$

where R/B is the spectral efficiency, i.e., $C(\text{SNR})$ with perfect CSI or $\bar{\mathcal{I}}(\text{SNR})$ without it.

With perfect CSI, it is known that $\frac{E_b^r}{N_0}$ is minimized for SNR $\to 0$ and that such minimum equals [25]

$$\frac{E_b^r}{N_0}_{\min} = \frac{1}{\log_2 e} \tag{17}$$

which equals -1.59 dB. Without perfect CSI on the side, $\bar{\mathcal{I}}(\text{SNR})$ with IID complex Gaussian inputs is convex below some (low) SNR and concave above it [5]–[8]. It follows that $\frac{E_b^r}{N_0}$ is minimized at some finite SNR. However, this minimizing SNR, and the corresponding

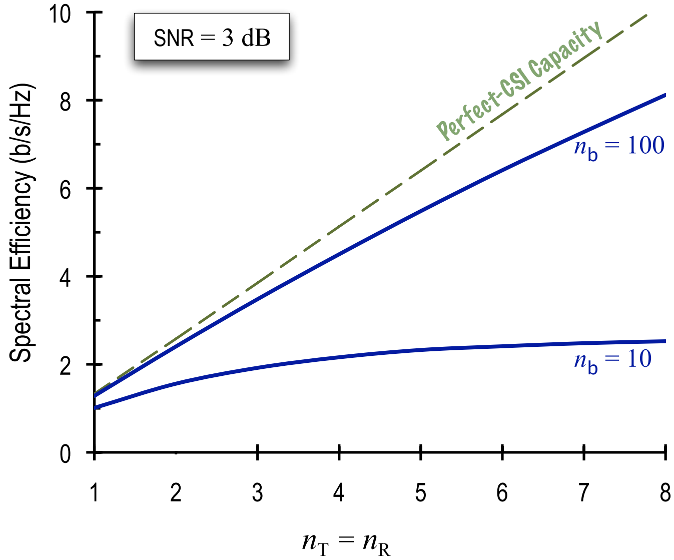

Fig. 6. In solid, $\bar{\mathcal{I}}(\text{SNR})$ for SNR $= 3$ dB as function of $n_T = n_R$ with $n_b = 10$ and $n_b = 100$. In dashed, the corresponding perfect-CSI capacity.

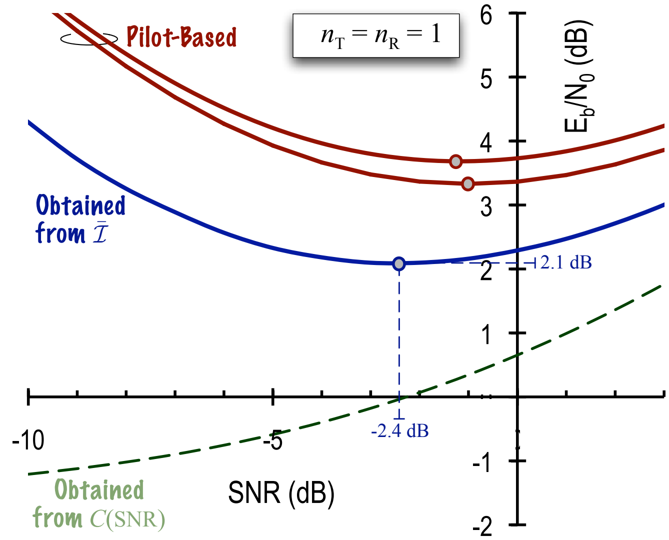

Fig. 7. E_b^r/N_0 as function of SNR for $n_T = n_R = 1$ with $n_b = 10$. In solid, values obtained from $\bar{\mathcal{I}}(\text{SNR})$ and also values corresponding to pilot-based communication, with and without pilot power boosting. For each curve, $(E_b^r/N_0)_{\min}$ is explicitly indicated. In dashed, E_b^r/N_0 with perfect CSI, i.e., obtained from $C(\text{SNR})$.

$\frac{E_b^r}{N_0}_{\min}$, cannot be obtained using the low-SNR expansions available in the literature because only the convex behavior of $\bar{\mathcal{I}}(\text{SNR})$ is captured therein. Specifically, the most refined low-SNR expansion available is [7]

$$\bar{\mathcal{I}}(\text{SNR}) = \frac{n_b n_R}{2\,n_T}\text{SNR}^2 + o(\text{SNR}^2) \tag{18}$$

based on which $\frac{E_b^r}{N_0}_{\min}$ indeed cannot be obtained.[2]

Applying the method presented in Section III, the correct $\frac{E_b^r}{N_0}_{\min}$ and the corresponding SNR can be calculated.

In pilot-based communication, the spectral efficiency is also a convex function of SNR below some SNR and concave above it [17]. Thus, $\frac{E_b^r}{N_0}_{\min}$ is also achieved at some finite SNR, which can be calculated numerically by solving for the spectral efficiencies in (13) and (14) and then using those results to minimize (16). Such calculations are conducted in [26].

Example 7 *Let $n_T = 1 = n_R = 1$ and let $n_b = 10$. Shown in Fig. 7 is the $\frac{E_b}{N_0}$ as function of* SNR.

At this fading rate ($n_b = 10$), the most power-efficiency operating point is SNR $= -2.4$ dB. The corresponding $\frac{E_b^r}{N_0}_{\min}$ equals 2.1 dB, almost 4 dB above what a perfect-CSI analysis would indicate. With pilot-based transmission, an additional penalty of over 1 dB in $\frac{E_b^r}{N_0}_{\min}$ is suffered.

More generally, the method in Section III allows characterizing the power-efficient operating point and the corresponding $\frac{E_b^r}{N_0}_{\min}$ as a function of the fading rate, n_b. This characterization is presented in Figs. 8 and 9 for $n_T = n_R = 1$, and similar results can be readily

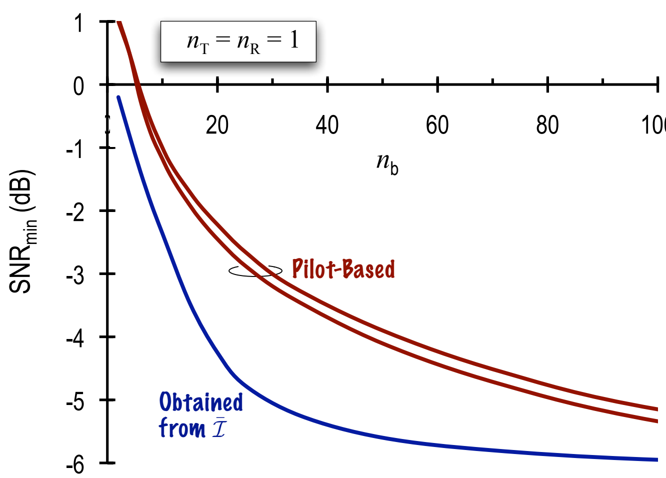

Fig. 8. SNR at which $(E_b^r/N_0)_{\min}$ is achieved as function of n_b for $n_T = n_R = 1$. Values obtained from $\bar{\mathcal{I}}(\text{SNR})$ and also values corresponding to pilot-based communication, with and without pilot power boosting.

obtained for MIMO. For typical vehicular scenarios, the operating points that maximize the power efficiency are substantially higher than what one might anticipate from a perfect-CSI analysis, and the corresponding $\frac{E_b}{N_0}_{\min}$ levels are markedly above the -1.59-dB floor.

VII. The High-SNR Regime

A. Case $n_b \geq n_R + \min\{n_T, n_R\}$

For such n_b, it is shown in [5] that the high-SNR slope of the true channel capacity (without side information at the receiver) is

$$\min\{n_T, n_R\}\left(1 - \frac{\min\{n_T, n_R\}}{n_b}\right) \tag{19}$$

[2]Eq. (18) would indicate that $\frac{E_b^r}{N_0}_{\min}$ is achieved for SNR $\to \infty$, but (18) does not apply beyond the low-SNR regime.

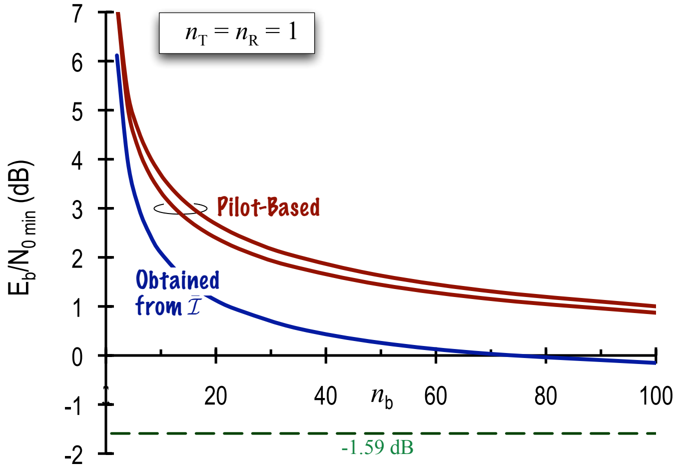

Fig. 9. $(E_b^r/N_0)_{\min}$ as function of n_b for $n_T = n_R = 1$. Values obtained from $\bar{\mathcal{I}}(\mathsf{SNR})$ and also values corresponding to pilot-based communication, with and without pilot power boosting.

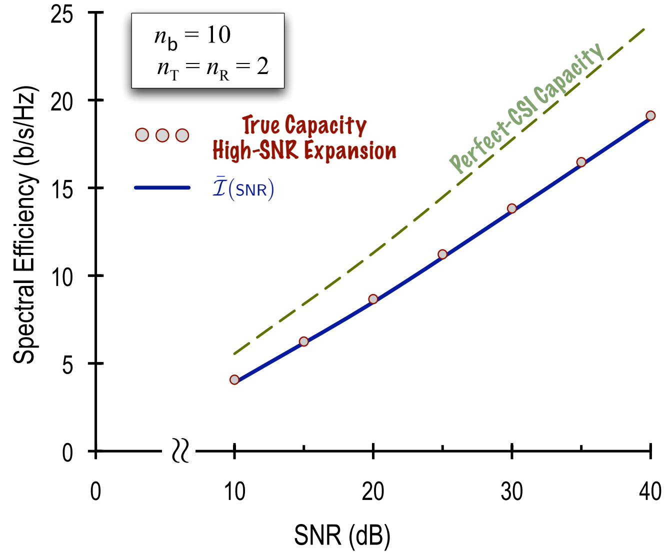

Fig. 10. In solid, $\bar{\mathcal{I}}(\mathsf{SNR})$ for $n_T = n_R = 2$ and $n_b = 10$. In dashed, the corresponding perfect-CSI capacity. In circles, the high-SNR expansion of the true capacity as per (20).

in bits/s/Hz/(3 dB). By activating $\min\{n_T, n_R\}$ transmit antennas and $\min\{n_T, n_R\}$ receive antennas, a straightforward computation shows that the lower bound in Proposition 3 achieves the same high-SNR slope. Since $\bar{\mathcal{I}}(\mathsf{SNR})$ increases with the number of receive antennas (by the chain rule, additional outputs are never harmful), this implies that the optimal high-SNR slope is achieved if IID Gaussian inputs are sent from $\min\{n_T, n_R\}$ transmit antennas and all available receive antennas are used. Based upon (13), it is also straightforward to confirm that pilot-based communication achieves the optimal high-SNR slope if $\min\{n_T, n_R\}$ transmit antennas are used with one pilot per antenna.

Thus, the slope is not a defining feature at high SNR. Rather, it is the power offset [27] that determines the performance in this regime, and the method in Section III can used to quantify it for IID Gaussian inputs.

Example 8 *Let $n_T = n_R = 2$ and let $n_b = 10$. Shown in Fig. 10 is the high-SNR mutual information. Also shown are the perfect-CSI capacity and the high-SNR expansion of the true capacity, which for $n_T = n_R$ is given in [5] as*

$$\left(1 - \frac{n_T}{n_b}\right)\left(C(\mathsf{SNR}) + n_T \log_2 \frac{n_b}{\pi e}\right) + \frac{1}{n_b}\log_2 G(n_b, n_T) + o(1) \tag{20}$$

with

$$G(t, n) = \frac{\prod_{i=t-n+1}^{t} \frac{2\pi^i}{(i-1)!}}{\prod_{i=1}^{n} \frac{2\pi^i}{(i-1)!}}. \tag{21}$$

Although for $n_b \geq n_R + \min\{n_T, n_R\}$ the input $\boldsymbol{X}$ that achieves the true capacity for $\mathsf{SNR} \to \infty$ is an isotropically random unitary matrix [9], [28], IID Gaussian inputs seem to perform very well at high SNR, even in relatively fast fading. Non-asymptotically in the SNR, the optimum $\boldsymbol{X}$ is no longer just a unitary matrix but rather the product of a unitary matrix and a nonnegative real diagonal matrix. No expressions are then available for the true capacity.

B. *Case $n_b < n_R + \min\{n_T, n_R\}$*

In this case, the optimum $\boldsymbol{X}$ is again the product of a unitary matrix and a nonnegative real diagonal matrix, even for $\mathsf{SNR} \to \infty$. The high-SNR slope of the true capacity equals [5]

$$\min\{n_T, n_R, \lfloor n_b/2 \rfloor\}\left(1 - \frac{\min\{n_T, n_R, \lfloor n_b/2 \rfloor\}}{n_b}\right) \tag{22}$$

but no further expressions are available for the true capacity. Our method to compute the mutual information of IID Gaussian inputs continues to apply.

VIII. Conclusion

We have presented a method (part analytical, part Monte Carlo) to compute the mutual information achieved by IID complex Gaussian inputs on block Rayleigh-faded channels, both scalar and MIMO. This mutual information is highly relevant as it represents the highest spectral efficiency attainable with Gaussian codebooks.

The method presented may be of further interest to other multivariate problems involving combinations of multiplicative and additive Gaussian noise, either with respect to the mutual information or to the constituting differential entropies.

A software routine that implements the described method in Matlab code is available for download at http://www.dtic.upf.edu/~alozano/software.

Acknowledgment

The authors thank Prof. Babak Hassibi (Caltech, USA) for pointing out some valuable references.

Appendix

A. Preliminaries

For subsequent use, we present four relevant identities. The first one, easily verified, is

$$\int_{-\infty}^{\infty} \exp\{-x^2 A + xB\}\,\mathrm{d}x = \exp\left\{\frac{B^2}{4A}\right\}\sqrt{\frac{\pi}{A}}. \tag{23}$$

The second one is an integral due to Itzykson and Zuber [29]. Given an $M \times M$ diagonal matrix $\boldsymbol{B}$ with diagonal entries $\boldsymbol{b}$, an arbitrary $M \times M$ matrix $\boldsymbol{D}$ with eigenvalues $\boldsymbol{d}$, and an $M \times M$ isotropically distributed unitary random matrix $\boldsymbol{U}$,

$$\int e^{\mathrm{Tr}\{\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\dagger \boldsymbol{Z}\}} p(\boldsymbol{U})\,\mathrm{d}\boldsymbol{U} = \prod_{m=1}^{M} \frac{(m-1)!\,\det \boldsymbol{E}(\boldsymbol{d},\boldsymbol{b})}{\det \boldsymbol{V}(\boldsymbol{d})\det \boldsymbol{V}(\boldsymbol{b})} \tag{24}$$

where the (i,j)th entry of the $M \times M$ matrix $\boldsymbol{E}(\boldsymbol{d},\boldsymbol{b})$ equals

$$E_{ij} = \exp\{d_i b_j\} \tag{25}$$

while $\boldsymbol{V}(\cdot)$ denotes a Vandermonde matrix, i.e., such that

$$\det \boldsymbol{V}(\boldsymbol{d}) = \prod_{1\le i<j\le M} (d_j - d_i). \tag{26}$$

The third identity treats (24) when $\boldsymbol{B}$ is of reduced rank. If $b_n = 0$ for $n > N$,

$$\begin{aligned}\lim_{b_{N+1},\ldots,b_M\to 0} &\frac{\det \boldsymbol{E}(\boldsymbol{d},\boldsymbol{b})}{\det \boldsymbol{V}(\boldsymbol{b})} \\ &= \frac{\det \tilde{\boldsymbol{E}}(\boldsymbol{d},\boldsymbol{b})}{\prod_{k=1}^{N} b_k^{M-N}\,(k-1)!}\frac{(-1)^{(M-N)(M-N-1)/2}}{\det \boldsymbol{V}(\boldsymbol{b}_{[1\ldots N]})}\end{aligned} \tag{27}$$

where the (i,j)th entry of $\tilde{\boldsymbol{E}}(\boldsymbol{d},\boldsymbol{z})$ equals $\exp\{d_i z_j\}$ for $j \le N$ and d_i^{j-N-1} for $N+1 \le j \le M$.

The final identity was proved in [30] by Chiani, Win and Zanella. Given two arbitrary $M \times M$ matrices $\boldsymbol{\Psi}(\boldsymbol{x})$ and $\boldsymbol{\Phi}(\boldsymbol{x})$ with (i,j)th entries $\Psi_i(x_j)$ and $\Phi_i(x_j)$, respectively, and an arbitrary function $\xi(\cdot)$,

$$\begin{aligned}\int\cdots\int_{\mathcal{D}_{\mathrm{ord}}} &\det \boldsymbol{\Psi}(\boldsymbol{x}) \det \boldsymbol{\Phi}(\boldsymbol{x}) \prod_{m=1}^{M} \xi(x_m)\mathrm{d}\boldsymbol{x} \\ &= \det\left(\left\{\int_a^b \Psi_i(x)\Phi_j(x)\xi(x)\mathrm{d}x\right\}_{i,j=1\ldots M}\right)\end{aligned} \tag{28}$$

where the multiple integral is over the domain $\mathcal{D}_{\mathrm{ord}} = \{b \ge x_1 \ge x_2 \ge \ldots \ge x_M \ge a\}$.

B. Proof of Proposition 1

Conditioned on $\boldsymbol{X}$, the output $\boldsymbol{Y}$ is complex Gaussian. Furthermore, the rows of $\boldsymbol{Y}$ are IID conditioned on $\boldsymbol{X}$. Hence, to obtain $\mathfrak{h}(\boldsymbol{Y}|\boldsymbol{X})$ it suffices to evaluate its value for an arbitrary row of $\boldsymbol{Y}$ and then scale it by the number of rows, i.e., by n_{R}.

Let $\boldsymbol{y}$ be an arbitrary row of $\boldsymbol{Y}$. The conditional covariance of the n_{b}-dimensional column vector $\boldsymbol{y}^\dagger$ equals

$$\mathbb{E}\left[\boldsymbol{y}^\dagger \boldsymbol{y}|\boldsymbol{X}\right] = \boldsymbol{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}}\boldsymbol{X}^\dagger \boldsymbol{X} \tag{29}$$

and thus

$$\begin{aligned}\mathfrak{h}(\boldsymbol{y}|\boldsymbol{X}) &= \mathfrak{h}(\boldsymbol{y}^\dagger|\boldsymbol{X}) &(30)\\ &= \mathbb{E}\left[\log_2\left((\pi e)^{n_{\mathsf{b}}}\det\left(\boldsymbol{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}}\boldsymbol{X}^\dagger \boldsymbol{X}\right)\right)\right] &(31)\end{aligned}$$

with expectation over the distribution of $\boldsymbol{X}$. Factoring out the term $n_{\mathsf{b}}\log_2(\pi e)$, what remains coincides with the perfect-CSI capacity of a MIMO channel, only with the role of the channel played by $\boldsymbol{X}$. Since the entries of $\boldsymbol{X}$ are IID complex Gaussian with zero mean and unit variance, we may directly apply the closed form in [13] with appropriate dimensioning. Scaling the end result by n_{R}, we convert $\mathfrak{h}(\boldsymbol{y}|\boldsymbol{X})$ into $\mathfrak{h}(\boldsymbol{Y}|\boldsymbol{X})$ as desired.

C. Proof of Proposition 2

Define $\gamma = \mathsf{SNR}/n_{\mathrm{T}}$, $s = \min\{n_{\mathrm{T}}, n_{\mathrm{R}}\}$ and $r = \max\{n_{\mathrm{T}}, n_{\mathrm{R}}\}$. Then, denoting by $\boldsymbol{x}_t$ the tth column of $\boldsymbol{X}$,

$$\begin{aligned}p(\boldsymbol{Y}) &= \mathbb{E}_{\boldsymbol{H}}\left[\int p(\boldsymbol{Y}|\boldsymbol{H},\boldsymbol{X})\,p(\boldsymbol{X})\,\mathrm{d}\boldsymbol{X}\right] &(32)\\ &= \frac{1}{(\pi\gamma)^{n_{\mathrm{T}}n_{\mathsf{b}}}}\frac{1}{\pi^{n_{\mathrm{R}}n_{\mathsf{b}}}}\mathbb{E}_{\boldsymbol{H}}\left[\prod_{t=1}^{n_{\mathsf{b}}}\int \exp\left\{-\|\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_t\|^2\right\}\right.\\ &\qquad \left.\times \exp\left\{-\|\boldsymbol{x}_t\|^2/\gamma\right\}\mathrm{d}\boldsymbol{x}_t\right]. &(33)\end{aligned}$$

Using the singular-value decomposition $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\dagger$, absorbing $\boldsymbol{V}$ into $\boldsymbol{X}$ through the variable substitution $\tilde{\boldsymbol{x}}_t = \boldsymbol{V}^\dagger \boldsymbol{x}_t$, and assembling the diagonal entries of $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\dagger$ into $\boldsymbol{\lambda} = [\lambda_1,\ldots,\lambda_s]$,

$$\begin{aligned}p(\boldsymbol{Y}) &= \frac{\exp\left\{-\|\boldsymbol{Y}\|^2\right\}}{(\pi\gamma)^{n_{\mathrm{T}}n_{\mathsf{b}}}\pi^{n_{\mathrm{R}}n_{\mathsf{b}}}}\mathbb{E}_{\boldsymbol{H}}\left[\int\prod_{t=1}^{n_{\mathsf{b}}}\exp\left\{-\tilde{\boldsymbol{x}}_t^\dagger \boldsymbol{\Lambda}\tilde{\boldsymbol{x}}_t\right\}\right.\\ &\qquad\left.\times\exp\left\{2\,\mathrm{Re}\{\boldsymbol{y}_t^\dagger \boldsymbol{U}\boldsymbol{\Sigma}\tilde{\boldsymbol{x}}_t\}\right\}\exp\left\{-\frac{\|\tilde{\boldsymbol{x}}_t\|^2}{\gamma}\right\}\mathrm{d}\tilde{\boldsymbol{x}}_t\right] &(34)\\ &= \frac{\exp\left\{-\|\boldsymbol{Y}\|^2\right\}}{(\pi\gamma)^{n_{\mathrm{T}}n_{\mathsf{b}}}\pi^{n_{\mathrm{R}}n_{\mathsf{b}}}}\mathbb{E}_{\boldsymbol{H}}\left[\int\prod_{t=1}^{n_{\mathsf{b}}}\left(\prod_{k=1}^{s}\right.\right.\\ &\qquad\left.\times\exp\left\{2\,\mathrm{Re}\{\boldsymbol{y}_t^\dagger \boldsymbol{u}_k\Sigma_k\tilde{x}_{t,k}\} - |\tilde{x}_{t,k}|^2\left(\lambda_k + \frac{1}{\gamma}\right)\right\}\right)\\ &\qquad\left.\times\left(\prod_{k=n_{\mathrm{R}}+1}^{n_{\mathrm{T}}}\exp\left\{-\frac{|\tilde{x}_{t,k}|^2}{\gamma}\right\}\right)\mathrm{d}\tilde{x}_{t,k}\right]. &(35)\end{aligned}$$

Applying (23) to each variable $\tilde{x}_{t,k}$ in (34) gives

$$p(\boldsymbol{Y}) = \frac{\exp\{-\|\boldsymbol{Y}\|^2\}}{\pi^{n_\mathrm{R} n_\mathsf{b}}} \mathbb{E}_{\boldsymbol{H}}\left[\prod_{t=1}^{n_\mathsf{b}}\prod_{k=1}^{s}\exp\left\{\frac{\lambda_k\left(\mathrm{Re}\{\boldsymbol{y}_t^\dagger \boldsymbol{u}_k\}\right)^2}{\left(\lambda_k+\frac{1}{\gamma}\right)}\right\}\right.$$
$$\left.\times \exp\left\{\frac{\lambda_k\left(\mathrm{Im}\{\boldsymbol{y}_t^\dagger \boldsymbol{u}_k\}\right)^2}{\left(\lambda_k+\frac{1}{\gamma}\right)}\right\}\left(\frac{1}{\lambda_k\gamma+1}\right)\right] \quad (36)$$
$$= \frac{\exp\{-\|\boldsymbol{Y}\|^2\}}{\pi^{n_\mathrm{R} n_\mathsf{b}}} \mathbb{E}_{\boldsymbol{H}}\left[\prod_{k=1}^{s}\exp\left\{\frac{\lambda_k\gamma\boldsymbol{u}_k^\dagger\boldsymbol{Y}\boldsymbol{Y}^\dagger\boldsymbol{u}_k}{(\lambda_k\gamma+1)}\right\}\right.$$
$$\left.\times\left(\frac{1}{\lambda_k\gamma+1}\right)^{n_\mathsf{b}}\right]. \quad (37)$$

Let $\boldsymbol{A}(\boldsymbol{\lambda})$ be a $n_\mathrm{R}\times n_\mathrm{R}$ diagonal matrix with kth diagonal entry $a_k = \lambda_k\gamma/(\lambda_k\gamma+1)$, $1\le k\le s$ and $a_k=0$, $s<k\le n_\mathrm{R}$. It is known that $\boldsymbol{U}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{V}$ in the singular-value decomposition of $\boldsymbol{H}$ are independent random matrices and that both $\boldsymbol{U}$ and $\boldsymbol{V}$ are isotropically distributed. Thus, we can express (37) as

$$p(\boldsymbol{Y}) = \frac{\exp\{-\|\boldsymbol{Y}\|^2\}}{\pi^{n_\mathrm{R} n_\mathsf{b}}}\int p(\boldsymbol{\lambda})\prod_{k=1}^{n_\mathrm{R}}\left(\frac{1}{\lambda_k\gamma+1}\right)^{n_\mathsf{b}}$$
$$\times\left[\int p(\boldsymbol{U})\exp\left\{\mathrm{Tr}\left\{\boldsymbol{A}(\boldsymbol{\lambda})\tilde{\boldsymbol{U}}^\dagger\boldsymbol{Y}\boldsymbol{Y}^\dagger\tilde{\boldsymbol{U}}\right\}\right\}\mathrm{d}\boldsymbol{U}\right]\mathrm{d}\boldsymbol{\lambda} \quad (38)$$

where the density distribution of the (ordered) eigenvalues in $\boldsymbol{\lambda}$ equals

$$p(\boldsymbol{\lambda}) = \det{}^2\boldsymbol{V}(\boldsymbol{\lambda})\prod_{k=1}^{s}\frac{e^{-\lambda_k}\lambda_k^{r-s}}{(n_\mathrm{R}-k)!\,(n_\mathrm{T}-k)!}, \quad (39)$$

and scaling by $1/s!$ yields the corresponding density of the unordered eigenvalues.

Henceforth, we separately consider the cases $n_\mathrm{T}\le n_\mathrm{R}$ and $n_\mathrm{T}<n_\mathrm{R}$.

1) Case $n_\mathrm{T}\ge n_\mathrm{R}$: The r.h.s. of (38) is precisely the setup in (24). Let $\boldsymbol{a}(\boldsymbol{\lambda})=[a_1,\ldots,a_{n_\mathrm{R}}]$ contain the diagonal elements of $\boldsymbol{A}(\boldsymbol{\lambda})$ and let $\boldsymbol{d}$ contain the eigenvalues of $\boldsymbol{Y}\boldsymbol{Y}^\dagger$. Then,

$$p(\boldsymbol{Y}) = \frac{\prod_{k=1}^{n_\mathrm{R}}(k-1)!\,e^{-\|\boldsymbol{Y}\|^2}}{\det\boldsymbol{V}(\boldsymbol{d})\,\pi^{n_\mathrm{R}n_\mathsf{b}}}\int p(\boldsymbol{\lambda})\frac{\det\boldsymbol{E}(\boldsymbol{a}(\boldsymbol{\lambda}),\boldsymbol{d})}{\det\boldsymbol{V}(\boldsymbol{a}(\boldsymbol{\lambda}))}$$
$$\times\prod_{k=1}^{n_\mathrm{R}}\left(\frac{1}{\lambda_k\gamma+1}\right)^{n_\mathsf{b}}\mathrm{d}\boldsymbol{\lambda} \quad (40)$$

with $\boldsymbol{E}(\cdot,\cdot)$ and $\boldsymbol{V}(\cdot)$ as per (25) and (26), respectively. Moreover,

$$\frac{1}{\det\boldsymbol{V}(\boldsymbol{a}(\boldsymbol{\lambda}))} = \prod_{k>\ell}\left(\frac{\lambda_k\gamma}{\lambda_k\gamma+1}-\frac{\lambda_\ell\gamma}{\lambda_\ell\gamma+1}\right)^{-1}$$
$$= \prod_{k>\ell}\frac{1}{\gamma}\frac{(\lambda_k\gamma+1)(\lambda_\ell\gamma+1)}{\lambda_k-\lambda_\ell}$$
$$= \frac{1}{\gamma^{(n_\mathrm{R}^2-n_\mathrm{R})/2}}\frac{\prod_{k>\ell}(\lambda_k\gamma+1)(\lambda_\ell\gamma+1)}{\det\boldsymbol{V}(\boldsymbol{\lambda})}$$
$$= \frac{1}{\gamma^{(n_\mathrm{R}^2-n_\mathrm{R})/2}}\frac{\prod_{k=1}^{n_\mathrm{R}}(\lambda_k\gamma+1)^{n_\mathrm{R}-1}}{\det\boldsymbol{V}(\boldsymbol{\lambda})}. \quad (41)$$

Plugging (39) and (41) into (40) yields

$$p(\boldsymbol{Y}) = \frac{\exp\{-\|\boldsymbol{Y}\|^2\}}{\det\boldsymbol{V}(\boldsymbol{d})\,\gamma^{\frac{n_\mathrm{R}^2-n_\mathrm{R}}{2}}\,\pi^{n_\mathrm{R}n_\mathsf{b}}\prod_{k=1}^{n_\mathrm{R}}(n_\mathrm{T}-k)!}\int\det\boldsymbol{V}(\boldsymbol{\lambda})$$
$$\times\det\boldsymbol{E}(\boldsymbol{a}(\boldsymbol{\lambda}),\boldsymbol{d})\prod_{k=1}^{n_\mathrm{R}}\frac{e^{-\lambda_k}\lambda_k^{n_\mathrm{T}-n_\mathrm{R}}}{(\lambda_k\gamma+1)^{n_\mathsf{b}+1-n_\mathrm{R}}}\mathrm{d}\boldsymbol{\lambda}$$
$$= \frac{\exp\{-\|\boldsymbol{Y}\|^2\}}{\det\boldsymbol{V}(\boldsymbol{d})\,\pi^{n_\mathrm{R}n_\mathsf{b}}\prod_{k=1}^{n_\mathrm{R}}(n_\mathrm{T}-k)!}\int\det\boldsymbol{V}(\boldsymbol{\lambda}/\gamma)$$
$$\times\det\boldsymbol{E}(\boldsymbol{a}(\boldsymbol{\lambda}),\boldsymbol{d})\prod_{k=1}^{n_\mathrm{R}}\frac{e^{-\lambda_k}\lambda_k^{n_\mathrm{T}-n_\mathrm{R}}}{(\lambda_k\gamma+1)^{n_\mathsf{b}+1-n_\mathrm{R}}}\mathrm{d}\boldsymbol{\lambda}. \quad (42)$$

The multiple integral in (42) is an instance of (28). Simply identify

$$\Phi(\boldsymbol{\lambda}) = \det\boldsymbol{V}(\boldsymbol{\lambda}/\gamma) \quad (43)$$
$$\Psi(\boldsymbol{\lambda}) = \det\boldsymbol{E}(\boldsymbol{a}(\boldsymbol{\lambda}),\boldsymbol{d}) \quad (44)$$
$$\xi(x) = \frac{e^{-x}\,x^{n_\mathrm{T}-n_\mathrm{R}}}{(x\gamma+1)^{n_\mathsf{b}+1-n_\mathrm{R}}} \quad (45)$$

to obtain

$$p(\boldsymbol{Y}) = \frac{\exp\left\{-\|\boldsymbol{Y}\|^2\right\}}{\pi^{n_\mathrm{R}n_\mathsf{b}}}\frac{\det\boldsymbol{Z}}{\det\boldsymbol{V}(\boldsymbol{d})\prod_{k=1}^{n_\mathrm{R}}(n_\mathrm{T}-k)!} \quad (46)$$

where

$$Z_{ij} = \int_0^\infty (x/\gamma)^{i-1}\exp\left\{d_j\frac{x\gamma}{x\gamma+1}-x\right\}\frac{x^{n_\mathrm{T}-n_\mathrm{R}}}{(x\gamma+1)^{n_\mathsf{b}+1-n_\mathrm{R}}}\,\mathrm{d}x. \quad (47)$$

Remark: The proposition can be generalized to also encompass non-Gaussian inputs. If $\boldsymbol{X}$ has rotational invariance and the corresponding $p(\boldsymbol{\lambda})$ if of the form

$$p(\boldsymbol{\lambda}) = \det{}^2\boldsymbol{V}(\boldsymbol{\lambda})\prod_{k=1}^{n_\mathrm{R}}\frac{g_k(\lambda_k)}{(n_\mathrm{R}-k)!(n_\mathrm{T}-k)!},$$

(42) equals

$$p(\boldsymbol{Y}) = \frac{\exp\{-\|\boldsymbol{Y}\|^2\}}{\det\boldsymbol{V}(\boldsymbol{d})\,\pi^{n_\mathrm{R}n_\mathsf{b}}\prod_{k=1}^{n_\mathrm{R}}(n_\mathrm{T}-k)!}\int\det\boldsymbol{V}(\boldsymbol{\lambda}/\gamma)$$
$$\times\det\boldsymbol{E}(\boldsymbol{a}(\boldsymbol{\lambda}),\boldsymbol{d})\prod_{k=1}^{n_\mathrm{R}}\frac{g_k(\lambda_k)}{(\lambda_k\gamma+1)^{n_\mathsf{b}+1-n_\mathrm{R}}}\mathrm{d}\boldsymbol{\lambda}. \quad (48)$$

In the case that the functions $g_k(x)=g(x)$ for all k, (48) is still an instance of (28), but with

$$\xi(x) = \frac{g(x)}{(x\gamma+1)^{n_\mathsf{b}+1-n_\mathrm{R}}}.$$

Hence, Proposition 2 applies, but with a different matrix $\boldsymbol{Z}$. If the functions $g_k(x)$ are distinct, (28) needs to be replaced with [30, Theorem 2]. This gives a similar, but not identical, result to Proposition 2 (the term $\det\boldsymbol{Z}$ is replaced by a general tensor-operator). These observations for non-Gaussian inputs hold for the case $n_\mathrm{T}<n_\mathrm{R}$ as well.

2) Case $n_\mathrm{T} < n_\mathrm{R}$: Let $\boldsymbol{a}(\boldsymbol{\lambda}) = [a_1, \ldots, a_{n_\mathrm{T}}]$ contain the diagonal elements of $\boldsymbol{A}(\boldsymbol{\lambda})$ and let $\boldsymbol{d}$ contain the eigenvalues of $\boldsymbol{Y}\boldsymbol{Y}^\dagger$. The r.h.s. of (38) is precisely the setup in (24), but in this case the limit in (27) must be taken. Taking that limit and plugging in the density (39) yields, after a few manipulations,

$$p(\boldsymbol{Y}) = \frac{\gamma^{n_\mathrm{R}-n_\mathrm{T}} \exp\left\{-\|\boldsymbol{Y}\|^2\right\}}{\det\boldsymbol{V}(\boldsymbol{d})\, \pi^{n_\mathrm{R} n_\mathsf{b}} \prod_{k=1}^{n_\mathrm{T}} k!} \int \det\boldsymbol{V}(\boldsymbol{\lambda}/\gamma) \times \det\tilde{\boldsymbol{E}}(\boldsymbol{a}(\boldsymbol{\lambda}), \boldsymbol{d}) \prod_{k=1}^{n_\mathrm{T}} \frac{e^{-\lambda_k}}{(\lambda_k\gamma+1)^{n_\mathsf{b}+1-n_\mathrm{R}}} \mathrm{d}\boldsymbol{\lambda} \quad (49)$$

where the density of unordered eigenvalues has been used and thus the integration limits are 0 to ∞ for all variables.

Eq. (49) is not an instance of (28) due to the structure of the matrix $\tilde{\boldsymbol{E}}(\boldsymbol{a}(\boldsymbol{\lambda}), \boldsymbol{d})$. However, the same result as in (28) still applies,

$$\begin{aligned}
&\int \det\boldsymbol{V}(\boldsymbol{\lambda}/\gamma) \det\tilde{\boldsymbol{E}}(\boldsymbol{a}(\boldsymbol{\lambda}), \boldsymbol{d}) \prod_{n=1}^{n_\mathrm{T}} \frac{e^{-\lambda_n}}{(\lambda_n\gamma+1)^{n_\mathsf{b}+1-n_\mathrm{R}}} \mathrm{d}\boldsymbol{\lambda} \\
&= \sum_{\sigma\in\mathcal{P}_{n_\mathrm{T}}} \sum_{\sigma'\in\mathcal{P}_{n_\mathrm{R}}} (-1)^{\sigma+\sigma'} \prod_{n=1}^{n_\mathrm{T}} \int {\frac{\lambda_n}{\gamma}}^{\sigma_n-1} \exp\left\{ d_{\sigma'_n} \frac{\lambda_n\gamma}{\lambda_n\gamma+1} \right\} \\
&\quad \times \frac{e^{-\lambda_n}}{(\lambda_n\gamma+1)^{n_\mathsf{b}+1-n_\mathrm{R}}} \mathrm{d}\lambda_n \prod_{n=n_\mathrm{T}+1}^{n_\mathrm{R}} d_{\sigma'_n}^{n-n_\mathrm{T}-1} \\
&= \sum_{\sigma\in\mathcal{P}_{n_\mathrm{T}}} \sum_{\sigma'\in\mathcal{P}_{n_\mathrm{R}}} (-1)^{\sigma+\sigma'} \prod_{n=1}^{n_\mathrm{T}} f_{\sigma_n}(d_{\sigma'_n}) \prod_{n=n_\mathrm{T}+1}^{n_\mathrm{R}} d_{\sigma'_n}^{n-n_\mathrm{T}-1} \\
&= n_\mathrm{T}! \sum_{\sigma'\in\mathcal{P}_{n_\mathrm{R}}} (-1)^{\sigma'} \prod_{n=1}^{n_\mathrm{T}} f_n(d_{\sigma'_n}) \prod_{n=n_\mathrm{T}+1}^{n_\mathrm{R}} d_{\sigma'_n}^{n-n_\mathrm{T}-1} \\
&= n_\mathrm{T}!\, \det\boldsymbol{Z},
\end{aligned} \quad (50)$$

where

$$f_n(x) = \int_0^\infty (z/\gamma)^{n-1} \exp\left\{ x\frac{z\gamma}{z\gamma+1} - z \right\} \frac{1}{(z\gamma+1)^{n_\mathsf{b}+1-n_\mathrm{R}}} \, \mathrm{d}z,$$

$(-1)^\sigma$ denotes the sign of the permutation σ, $\mathcal{P}_N$ denotes the set of all permutations of the integers $1, \ldots, N$, and the (i,j)th entry of $\boldsymbol{Z}$ equals $Z_{ij} = f_j(d_i)$ if $j \le n_\mathrm{T}$ and $Z_{ij} = d_i^{j-n_\mathrm{T}-1}$ if $j > n_\mathrm{T}$.

By combining both cases, the complete statement in the proposition is obtained.

References

[1] A. Lozano, A. M. Tulino, and S. Verdú, "Power allocation over parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inform. Theory*, vol. 52, no. 7, pp. 3033–3051, July 2006.

[2] R. R. Perera, T. S. Pollock, and T. D. Abhayapala, "Gaussian inputs: Performance limits over non-coherent SISO and MIMO channels," *European Transactions on Telecommunications*, vol. 18, pp. 235–244, 2007.

[3] X. Deng and A. M. Haimovich, "Achievable rates over time-varying Rayleigh fading channels," *IEEE Trans. Communications*, vol. 55, no. 7, pp. 1397–1406, July 2007.

[4] M. Dörpinghaus, M. Senst, G. Ascheid, and H. Meyr, "On the achievable rate of stationary Rayleigh flat-fading channels with Gaussian input distribution," *Int'l Symp. on Inform. Theory and its Applications (ISITA'08)*, pp. 1–6, Dec. 2008.

[5] L. Zheng and D. N. C. Tse, "Communication on the Grassman manifold: A geometric approach to the non-coherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[6] V. V. Prelov and S. Verdu, "Second-order asymptotics of mutual information," *IEEE Trans. Inform. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.

[7] C. Rao and B. Hassibi, "Analysis of multiple-antenna wireless links at low SNR," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 2123–2130, Sept. 2004.

[8] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.

[9] T. L. Marzetta and B. H. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, Jan. 1999.

[10] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1290–1301, Apr. 2001.

[11] A. Lozano, "Interplay of spectral efficiency, power and Doppler spectrum for reference-signal-assisted wireless communication," *IEEE Trans. Communications*, vol. 56, no. 12, Dec. 2008.

[12] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Trans. on Communications, to appear*, 2010.

[13] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering and keyhole," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2636–2647, Oct. 2003.

[14] H. Shin and M. Z. Win, "Gallager's exponent for MIMO channels: A reliability-rate tradeoff," *IEEE Trans. Communications*, vol. 57, no. 4, pp. 972–985, Apr. 2009.

[15] M. Godavarti, T. L. Marzetta, and S. Shamai, "Capacity of a mobile multiple-antenna wireless link with isotropically random Rician fading," *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3330–3334, Dec. 2003.

[16] A. Lapidoth and S. Shamai, "Fading channels: How perfect need 'perfect side information' be?," *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[17] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[18] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[19] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Communications*, vol. 4, no. 2, pp. 453–466, Mar. 2005.

[20] J. Baltersee, G. Fock, and H. Meyr, "An information theoretic foundation of synchronized detection," *IEEE Trans. Communications*, vol. 49, no. 12, pp. 2115–2123, Dec. 2001.

[21] S. Ohno and G. B. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," *IEEE Trans. Wireless Communications*, vol. 1, no. 4, pp. 712–720, Oct. 2002.

[22] S. Furrer and D. Dahlhaus, "Multiple-antenna signaling over fading channels with estimated channel state information: Capacity analysis," *IEEE Trans. Inform. Theory*, vol. 53, no. 6, pp. 2028–2043, June 2007.

[23] 3rd Generation Partnership Project (3GPP), *UTRA-UTRAN Long Term Evolution (LTE)*, Nov. 2004.

[24] Jeffrey G. Andrews, Arunabha Ghosh, and Rias Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*, Prentice Hall PTR, 2007.

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

[26] M. C. Gursoy, "On the capacity and energy efficiency of training-based transmissions over fading channels," *IEEE Trans. Inform. Theory*, vol. 55, pp. 4543–4567, Oct. 2009.

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

[28] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: The received signal density in closed form," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1473–1484, June 2002.

[29] C. Itzykson and J.-B. Zuber, "The planar approximation II," *J. Math. Phys.*, vol. 21, pp. 411–421, Mar. 1980.

[30] M. Chiani, M. Z. Win, and A. Zanella, "On the capacity of spatially correlated MIMO rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2363–2371, Oct. 2003.