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

Fredrik Rusek
Lund University
221 00 Lund, Sweden
Email: fredrik.rusek@eit.lth.se

Angel Lozano
Universitat Pompeu Fabra
Barcelona 08003, Spain
Email: angel.lozano@upf.edu

Nihar Jindal
University of Minnesota
Minneapolis, MN 55455
Email: nihar@umn.edu

Abstract—We present a method to compute, quickly and efficiently, the mutual information achieved by an IID (independent identically distributed) complex Gaussian input 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, the mutual information thus computed represents the highest spectral efficiency that can be attained using standard Gaussian codebooks. Examples are provided that illustrate the loss in spectral efficiency caused by fast fading and how that loss is amplified by the use of multiple transmit antennas. 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.

I. INTRODUCTION

IID (independent identically distributed) complex Gaussian inputs are most relevant in wireless communication impaired by Gaussian noise. Some of the arguments for this relevance are that, if perfect CSI (channel state information) is available at the receiver, then:

- 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].)

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 perfect CSI. No expressions are available for the mutual information achieved by IID complex Gaussian inputs save for the special case of memoryless channels [2]. Moreover, straight Monte-Carlo computation is not feasible because it would entail large-dimensional histograms. Only asymptotic expansions of the mutual information in the low- and high-power regimes [3]–[6] have been derived. And yet, although no longer capacity-achieving [7], [8], IID complex Gaussian inputs remain highly relevant. Operationally, the mutual information they achieve represents the highest spectral efficiency that can be attained using standard Gaussian codebooks. In fact, the capacity-achieving inputs in the absence of perfect-CSI become largely impractical in many cases (in the low-power regime, for example, they become unacceptably peaky) and thus the capacity is arguably 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 the perfect-CSI assumption?
- How suboptimal are pilot-based schemes, i.e., those schemes that form explicit channel estimates on the basis of pilot observations at the receiver and subsequently apply these estimates 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, with $n_T \geq n_R$, and let the $n_R \times n_T$ 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 then 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 every block in an IID fashion. There is no antenna correlation and thus, within a given fading block, the entries of $\boldsymbol{H}$ are also independent.

Assembling on 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 $n_{\mathrm{T}} \times n_{\mathrm{b}}$ matrix while the output $\boldsymbol{Y}$ and the noise $\boldsymbol{N}$ are $n_{\mathrm{R}} \times n_{\mathrm{b}}$ 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 yields results that are remarkably similar to those obtained with continuous-fading models [9]. In fact, for a rectangular Doppler spectrum, an exact correspondence in terms of the error in the estimation of $\boldsymbol{H}$ can be established between block- and continuous-fading models [10] whereby

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

where f_{m} and T_{s} are the maximum Doppler frequency and the symbol period, respectively. Typically, $f_{\mathrm{m}} = (v/c) f_{\mathrm{c}}$ 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

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

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

Our first result leverages the derivations in [11] to obtain a closed-form expression for $\mathfrak{h}(\boldsymbol{Y}|\boldsymbol{X})$.

Proposition 1 *Let $E_q(\cdot)$ denote the exponential integral*

$$E_q(\zeta) = \int_1^\infty t^{-q} e^{-\zeta t} dt. \tag{4}$$

Then,

$$\mathfrak{h}(\boldsymbol{Y}|\boldsymbol{X}) = n_{\mathrm{R}} \log_2(e)\, e^{n_{\mathrm{T}}/\mathsf{SNR}} \sum_{i=0}^{n_{\mathrm{T}}-1} \sum_{j=0}^{i} \sum_{\ell=0}^{2j} \left[\binom{2i-2j}{i-j} \right.$$
$$\cdot \binom{2j + 2n_{\mathrm{b}} - 2n_{\mathrm{T}}}{2j - \ell} \frac{(-1)^\ell (2j)! (n_{\mathrm{b}} - n_{\mathrm{T}} + \ell)!}{2^{2i-\ell} j! \ell! (n_{\mathrm{b}} - n_{\mathrm{T}} + j)!}$$
$$\left. \sum_{q=0}^{n_{\mathrm{b}} - n_{\mathrm{T}} + \ell} E_{q+1}\!\left(\frac{n_{\mathrm{T}}}{\mathsf{SNR}}\right) \right] + n_{\mathrm{R}} n_{\mathrm{b}} \log_2(\pi e) \tag{5}$$

Proof: *See Appendix.*

Having expressed $\mathfrak{h}(\boldsymbol{Y}|\boldsymbol{X})$ in (3), we now turn our attention to $\mathfrak{h}(\boldsymbol{Y})$. Unfortunately, a closed-form solution for it seems out of reach. Denoting by $p(\boldsymbol{Y})$ the unconditional output distribution,

$$\mathfrak{h}(\boldsymbol{Y}) = -\int p(\boldsymbol{Y}) \log_2 p(\boldsymbol{Y}) \mathrm{d}\boldsymbol{Y} \tag{6}$$

$$= -\mathbb{E}\left[\log_2 p(\boldsymbol{Y}) \right] \tag{7}$$

Thus, if a formula for $p(\boldsymbol{Y})$ were at hand, $\mathfrak{h}(\boldsymbol{Y})$ could be evaluated by straightforward Monte Carlo averaging. Such a formula is the object of the next result.

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

$$f_k(x) = \int_0^\infty \exp\left\{ \frac{x\, \mathsf{SNR}\, z}{z\, \mathsf{SNR} + n_{\mathrm{T}}} - z \right\} \frac{z^{k-1+n_{\mathrm{R}}-n_{\mathrm{T}}} (n_{\mathrm{T}}/\mathsf{SNR})^{k-1}}{(z\, \mathsf{SNR}/n_{\mathrm{T}} + 1)^{n_{\mathrm{b}}+1-n_{\mathrm{R}}}} \mathrm{d}z \tag{8}$$

Let $\boldsymbol{d} = [d_1, \dots, d_{n_{\mathrm{R}}}]$ be the eigenvalues of $\boldsymbol{Y}^\dagger \boldsymbol{Y}$ and define the $n_{\mathrm{R}} \times n_{\mathrm{R}}$ matrix $\boldsymbol{Z}$ as $Z_{ij} = f_i(d_j)$, $1 \le i, j \le n_{\mathrm{R}}$. Then

$$p(\boldsymbol{Y}) = \frac{K(\boldsymbol{Y})}{\prod_{1 \le i,j \le n_{\mathrm{R}}} (d_j - d_i) \prod_{k=0}^{n_{\mathrm{R}}-1} (n_{\mathrm{T}} - k - 1)!} \det \boldsymbol{Z} \tag{9}$$

where

$$K(\boldsymbol{Y}) = \frac{e^{-\|\boldsymbol{Y}\|^2}}{\pi^{n_{\mathrm{b}} n_{\mathrm{R}}}}. \tag{10}$$

Proof: *See Appendix.*

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

We also note that, due to the rotational invariance of $\boldsymbol{H}$, the phase of the entries of $\boldsymbol{Y}$ carries no information about $\boldsymbol{X}$. Only the magnitude of the entries of $\boldsymbol{Y}$ is relevant to the distribution in (9).

Using (5) 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 (5) 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.

IV. BASELINES

Before exemplifying the method described in Section III, and in order to use them as baselines, we introduce the perfect-CSI capacity and the pilot-based communication boundaries.

A. Capacity with Perfect CSI

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

$$C(\mathsf{SNR}) = \mathbb{E}\left[\log_2 \det\left(\boldsymbol{I} + \frac{\mathsf{SNR}}{n_\mathrm{T}}\boldsymbol{H}\boldsymbol{H}^\dagger\right)\right] \quad (11)$$

a closed form for which can be found in [11]. For $n_\mathrm{T} = n_\mathrm{R} = 1$, this closed form reduces to

$$C(\mathsf{SNR}) = e^{1/\mathsf{SNR}}E_1\left(\frac{1}{\mathsf{SNR}}\right)\log_2 e \quad (12)$$

where $E_1(\cdot)$ is the exponential integral

$$E_1(\xi) = \int_1^\infty t^{-1}e^{-\xi t}dt. \quad (13)$$

At high SNR, the slope of $C(\mathsf{SNR})$ is $S_\infty = \min\{n_\mathrm{T}, n_\mathrm{R}\}$ bits/s/Hz/(3 dB).

B. 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 transmission. The channel is estimated on the basis of the pilot observations at the receiver, and this estimate is subsequently utilized to detect the data.

During the transmission of pilot symbols,

$$\boldsymbol{Y}_\mathsf{p} = \sqrt{\frac{\mathsf{SNR}}{n_\mathrm{T}}}\boldsymbol{H}\boldsymbol{P} + \boldsymbol{N}_\mathsf{p} \quad (14)$$

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 must satisfy $\boldsymbol{P}\boldsymbol{P}^\dagger = n_\mathsf{p}\boldsymbol{I}$ [12].

During the transmission of data symbols, in turn, (1) applies only 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], [12]–[18]. In bits/s/Hz, such spectral efficiency applying the channel estimate as if it were the true channel [19] equals

$$\max_{n_\mathsf{p}:1\leq n_\mathsf{p}\leq n_\mathsf{b}}\left(1 - \frac{n_\mathsf{p}}{n_\mathsf{b}}\right)C\left(\frac{\mathsf{SNR}^2 n_\mathsf{p}/n_\mathrm{T}}{1 + \mathsf{SNR}\left(1 + n_\mathsf{p}/n_\mathrm{T}\right)}\right) \quad (15)$$

where $C(\cdot)$ is the perfect-CSI capacity in (11) and (12). A more elaborate receiver that decoded on the basis of the joint distribution of the channel estimate and the

true channel, rather than by assuming that the channel estimate equals the true channel, would exceed (15) but only slightly [9].

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 [12] leading to[1]

$$\left(1 - \frac{n_\mathrm{T}}{n_\mathsf{b}}\right)C\left(\frac{n_\mathsf{b}\mathsf{SNR}}{n_\mathsf{b} - 2\,n_\mathrm{T}}\left(\sqrt{\gamma} - \sqrt{\gamma-1}\right)^2\right) \quad (16)$$

in bits/s/Hz, and with

$$\gamma = \frac{n_\mathsf{b}\mathsf{SNR} + n_\mathrm{T}}{n_\mathsf{b}\mathsf{SNR}\,\frac{n_\mathsf{b}-2\,n_\mathrm{T}}{n_\mathsf{b}-n_\mathrm{T}}}. \quad (17)$$

The spectral efficiency in (16) is superior to that in (15). However, pilot power boosting increases the peakiness of the overall signal distribution, rendering it less amenable to efficient amplification.

It can be easily verified that the high-SNR slope of $\bar{\mathcal{I}}_\mathsf{pb}(\mathsf{SNR})$ equals

$$\min\{n_\mathrm{T}, n_\mathrm{R}\}\left(1 - \frac{n_\mathrm{T}}{n_\mathsf{b}}\right) \quad (18)$$

bits/s/Hz/(3 dB), which for $n_\mathrm{T} \leq n_\mathrm{R}$ coincides with the high-SNR slope of the channel capacity without CSI at the receiver [3]:

$$\min\{n_\mathrm{T}, n_\mathrm{R}\}\left(1 - \frac{\min\{n_\mathrm{T}, n_\mathrm{R}\}}{n_\mathsf{b}}\right). \quad (19)$$

For $n_\mathrm{T} > n_\mathrm{R}$, (18) can also be made to equal (19) by using only a subset n_R of the transmit antennas. Thus, pilot-based communication need not a loss in high-SNR slope as long as pilot power boosting is allowed. Furthermore, it follows that the high-SNR slope of $\bar{\mathcal{I}}(\mathsf{SNR})$ also coincides with (19) and that, at high SNR, it is undesirable to activate $n_\mathrm{T} > n_\mathrm{R}$ transmit antennas. In the next section we shall examine whether this holds in a wider range of SNR.

V. SOME EXAMPLES

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 [20] or IEEE 802.16 WiMAX [21]:

- The carrier frequency f_c lies between 1 and 5 GHz.
- The symbol period is $T_\mathsf{s} \approx 100\ \mu\mathrm{s}$. However, it could be shortened to $T_\mathsf{s} \approx 10\text{-}20\ \mu\mathrm{s}$ 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.)

---

[1]Eq. (16) requires that $n_\mathsf{b} > 2\,n_\mathrm{T}$; variations thereof are also available for $n_\mathsf{b} \leq 2\,n_\mathrm{T}$ [12].
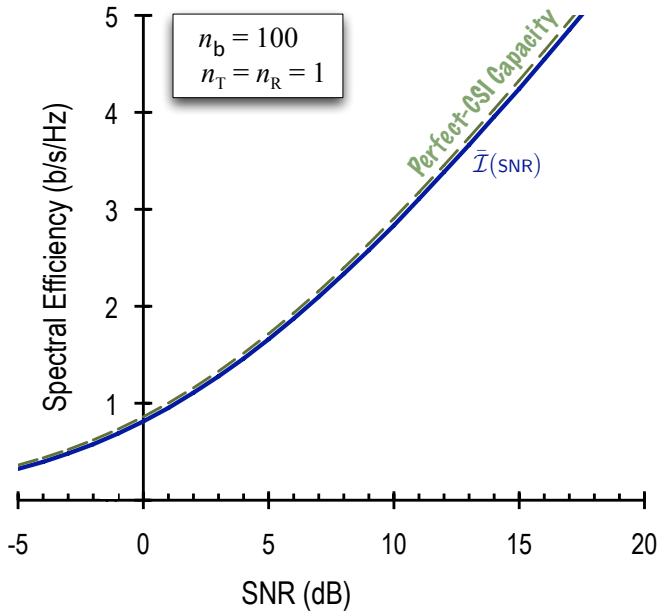
Fig. 1. In solid, mutual information $\bar{\mathcal{I}}(\mathsf{SNR})$ for $n_\mathrm{T} = n_\mathrm{R} = 1$ with $n_\mathrm{b} = 100$. In dashed, the perfect-CSI capacity.

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

- Velocities up to $v \approx 120$ Km/h are of interest, and for high-speed trains this extends to $v \approx 300$ Km/h. 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 $n_\mathrm{T} = n_\mathrm{R} = 1$ and let $n_\mathrm{b} = 100$. 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 focus on scenarios where n_b is small. Specifically, we shall use $n_\mathrm{b} = 10$ and $n_\mathrm{b} = 4$. 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 $n_\mathrm{T} = n_\mathrm{R} = 1$ and let $n_\mathrm{b} = 10$. 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 in this scenario, although this diminishes markedly 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 scenario.
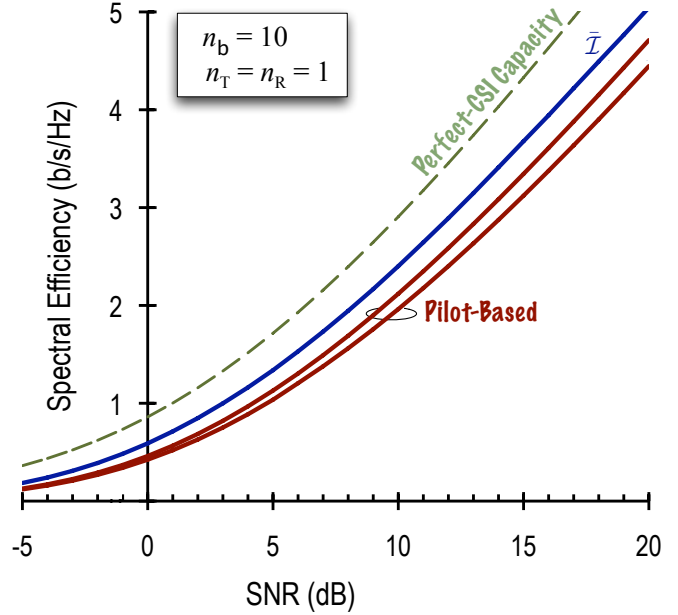

Fig. 3. In solid, mutual information $\bar{\mathcal{I}}(\mathsf{SNR})$ for $n_\mathrm{T} = n_\mathrm{R} = 1$ with $n_\mathrm{b} = 4$. Also in solid, spectral efficiencies achieved by pilot-based communication, with and without pilot power boosting. In dashed, the perfect-CSI capacity.
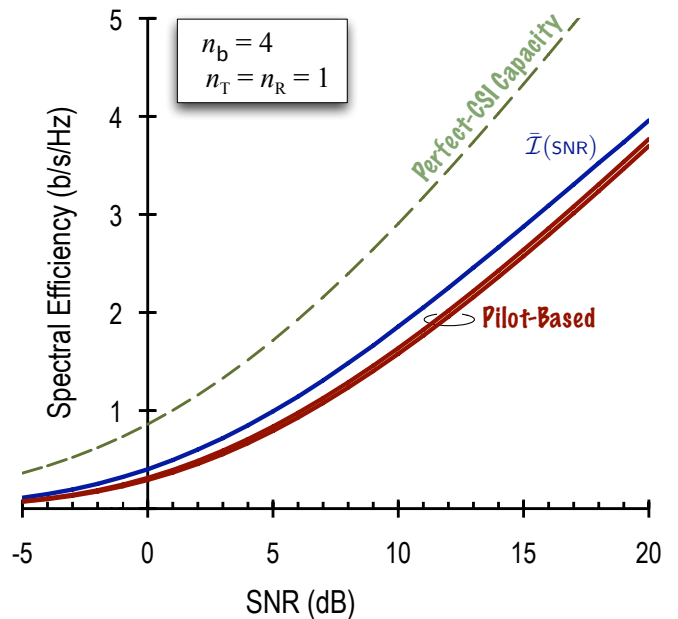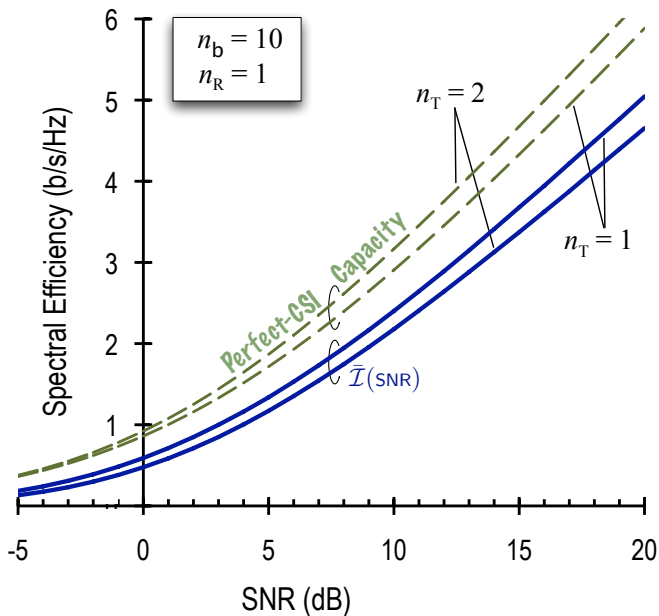
Fig. 4. In solid, mutual information $\bar{\mathcal{I}}(\text{SNR})$ for $n_{\text{T}} = n_{\text{R}} = 1$ and for $n_{\text{T}} = 2$, $n_{\text{R}} = 1$, with $n_{\text{b}} = 10$. In dashed, the respective perfect-CSI capacities.

Fig. 5. In solid, mutual information $\bar{\mathcal{I}}(\text{SNR})$ for $n_{\text{T}} = n_{\text{R}} = 2$, with $n_{\text{b}} = 10$ and with $n_{\text{b}} = 4$. In dashed, the corresponding perfect-CSI capacity.

Example 3 *Fig. 3 re-evaluates Example 2 with $n_{\text{b}} = 4$.*

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

Example 4 *Let $n_{\text{R}} = 1$ and let $n_{\text{b}} = 10$. Shown in Fig. 4 is the mutual information as function of SNR with $n_{\text{T}} = 1$ and with $n_{\text{T}} = 2$. Also shown are the respective perfect-CSI capacities.*

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. Example 4 reflects this increase. However, the considerations in Section IV, inspired by [3], [12], suggest that in reality activating $n_{\text{T}} > n_{\text{R}}$ antennas would be detrimental at high SNR. Example 4 confirms that this is indeed the case, not only at high SNR but at every SNR of interest, when the level of mobility is sufficiently high. (For this particular example, up to around $n_{\text{b}} \approx 40$.)

Example 5 *Let $n_{\text{T}} = n_{\text{R}} = 2$. Shown in Fig. 5 is the mutual information as function of SNR with $n_{\text{b}} = 10$ and with $n_{\text{b}} = 4$. Also shown is the corresponding perfect-CSI capacity.*

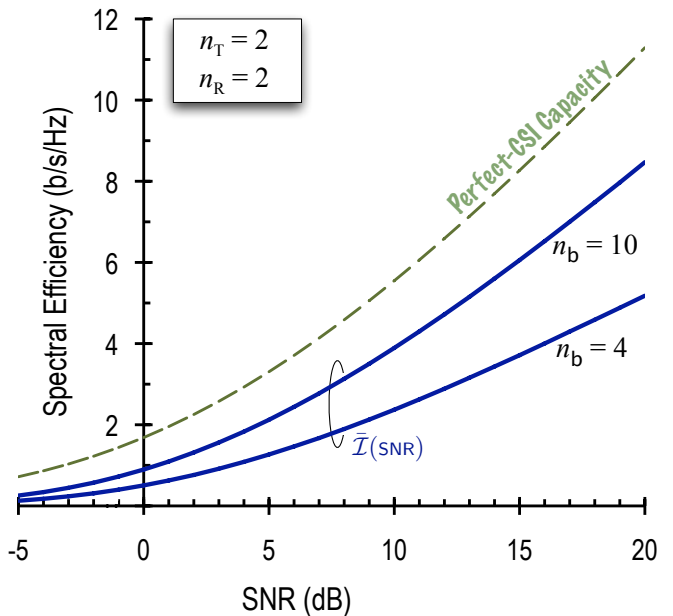Comparing Example 5 with Examples 2 and 3, notice how, at each level of mobility, MIMO transmission suffers a more drastic loss relative to the perfect-CSI capacity. Note that, with $n_{\text{b}} = 4$, the mutual information for $n_{\text{T}} = n_{\text{R}} = 2$ is larger thab for $n_{\text{T}} = n_{\text{R}} = 1$ (Example 3). Although Example 4 illustrated that an additional transmit antenna should be activated only for reasonably long n_{b}, this shows that an additional transmit-receive pair should be activated even for short n_{b}.

VI. CONCLUSION

We have presented a method (partly analytical, partly 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 standard Gaussian codebooks.

In prior work [22], some of the authors derived a lower bound to this mutual information (allowing but not requiring pilot symbols). Specialized to the case of no pilots, that bound is

$$\bar{\mathcal{I}}(\text{SNR}) \geq C(\text{SNR}) - \frac{n_{\text{T}} n_{\text{R}}}{n_{\text{b}}} \log_2 \left(1 + \text{SNR} \frac{n_{\text{b}}}{n_{\text{T}}} \right). \qquad (20)$$

The method presented in this paper also allows verifying the accuracy of this lower bound.

In addition, the method may be of 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 can be downloaded at

APPENDIX

A. Preliminaries

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

$$\int_{-\infty}^{\infty} \exp\{-x^2 A + xB\}\,\mathrm{d}x = \exp\left\{\frac{B^2}{4A}\right\}\sqrt{\frac{\pi}{A}}. \quad (21)$$

The second one is an integral due to Itzykson and Zuber [23]. Given an $M \times M$ diagonal matrix $\boldsymbol{Z}$ with diagonal elements $\boldsymbol{z}$, an arbitrary $M \times M$ matrix $\boldsymbol{D}$ with eigenvalues $\boldsymbol{d}$, and an $M \times M$ isotropically distributed unitary random matrix $\boldsymbol{U}$,

$$\int e^{\mathrm{Tr}\{\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\dagger \boldsymbol{Z}\}} p(\boldsymbol{U})\,\mathrm{d}\boldsymbol{U} = \prod_{m=1}^{M}\frac{(m-1)!\,\det\boldsymbol{E}(\boldsymbol{d},\boldsymbol{z})}{\det\boldsymbol{V}(\boldsymbol{d})\det\boldsymbol{V}(\boldsymbol{z})} \quad (22)$$

where the (i,j)th entry of the $M \times M$ matrix $\boldsymbol{E}(\boldsymbol{d},\boldsymbol{Z})$ equals $\exp\{d_i z_j\}$ while $\boldsymbol{V}(\cdot)$ denotes a Vandermonde matrix, i.e., such that

$$\det\boldsymbol{V}(\boldsymbol{d}) = \prod_{1 \le i,j \le M}(d_j - d_i). \quad (23)$$

The final identity was proved in [24] 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 $\xi(\cdot)$,

$$\int\cdots\int_{\mathcal{D}_{\mathrm{ord}}}\det\boldsymbol{\Psi}(\boldsymbol{x})\det\boldsymbol{\Phi}(\boldsymbol{x})\prod_{m=1}^{M}\xi(x_m)\mathrm{d}\boldsymbol{x}$$
$$= \det\left(\left\{\int_a^b \Psi_i(x)\Phi_j(x)\xi(x)\mathrm{d}x\right\}_{i,j=1\ldots M}\right) \quad (24)$$

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

$$\mathbb{E}\left[\boldsymbol{y}^\dagger \boldsymbol{y}|\boldsymbol{X}\right] = \boldsymbol{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}}\boldsymbol{X}^\dagger \boldsymbol{X} \quad (25)$$

and thus

$$\mathfrak{h}(\boldsymbol{y}|\boldsymbol{X}) = \mathfrak{h}(\boldsymbol{y}^\dagger|\boldsymbol{X}) \quad (26)$$
$$= E\left[\log_2\left((\pi e)^{n_{\mathsf{b}}}\det\left(\boldsymbol{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}}\boldsymbol{X}^\dagger \boldsymbol{X}\right)\right)\right] \quad (27)$$

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 [11] 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 $\gamma = \mathsf{SNR}/n_{\mathrm{T}}$. Then, denoting by $\boldsymbol{x}_t$ the tth column of $\boldsymbol{X}$,

$$p(\boldsymbol{Y}) = \mathbb{E}_{\boldsymbol{H}}\left[\int p(\boldsymbol{Y}|\boldsymbol{H},\boldsymbol{X})\,p(\boldsymbol{X})\,\mathrm{d}\boldsymbol{X}\right] \quad (28)$$
$$= \frac{1}{(\pi\gamma)^{n_{\mathrm{T}}n_{\mathsf{b}}}}\frac{1}{\pi^{n_{\mathrm{R}}n_{\mathsf{b}}}}\mathbb{E}_{\boldsymbol{H}}\left[\prod_{t=1}^{n_{\mathsf{b}}}\int \exp\left\{-\|\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_t\|^2\right\}\right.$$
$$\left.\times\exp\left\{-\frac{\|\boldsymbol{x}_t\|^2}{\gamma}\right\}\mathrm{d}\boldsymbol{x}_t\right]. \quad (29)$$

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 $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_{n_{\mathrm{R}}}]$,

$$p(\boldsymbol{Y}) = \frac{\exp\left\{-\|\boldsymbol{Y}\|^2\right\}}{(\pi\gamma)^{n_{\mathrm{T}}n_{\mathsf{b}}}\pi^{n_{\mathrm{R}}n_{\mathsf{b}}}}\mathbb{E}_{\boldsymbol{H}}\left[\int\prod_{t=1}^{n_{\mathsf{b}}}\exp\left\{-\tilde{\boldsymbol{x}}_t^\dagger \boldsymbol{\Lambda}\tilde{\boldsymbol{x}}_t\right\}\right.$$
$$\left.\times\exp\left\{2\,\mathrm{Re}\{\boldsymbol{y}_t^\dagger \boldsymbol{U}\boldsymbol{\Sigma}\tilde{\boldsymbol{x}}_t\}\right\}\exp\left\{-\frac{\|\tilde{\boldsymbol{x}}_t\|^2}{\gamma}\right\}\mathrm{d}\tilde{\boldsymbol{x}}_t\right] \quad (30)$$
$$= \frac{\exp\left\{-\|\boldsymbol{Y}\|^2\right\}}{(\pi\gamma)^{n_{\mathrm{T}}n_{\mathsf{b}}}\pi^{n_{\mathrm{R}}n_{\mathsf{b}}}}\mathbb{E}_{\boldsymbol{H}}\left[\int\prod_{t=1}^{n_{\mathsf{b}}}\left(\prod_{k=1}^{n_{\mathrm{R}}}\right.\right.$$
$$\left.\times\exp\left\{2\,\mathrm{Re}\{\boldsymbol{y}_t^\dagger \boldsymbol{u}_k \Sigma_k \tilde{x}_{t,k}\} - |\tilde{x}_{t,k}|^2\left(\lambda_k + \frac{1}{\gamma}\right)\right\}\right)$$
$$\left.\times\left(\prod_{k=n_{\mathrm{R}}+1}^{n_{\mathrm{T}}}\exp\left\{-\frac{|\tilde{x}_{t,k}|^2}{\gamma}\right\}\right)\mathrm{d}\tilde{x}_{t,k}\right] \quad (31)$$

Applying (21) to each variable $\tilde{x}_{t,k}$ in (30) gives

$$p(\boldsymbol{Y}) = \frac{\exp\{-\|\boldsymbol{Y}\|^2\}}{\pi^{n_{\mathrm{R}}n_{\mathsf{b}}}}\mathbb{E}_{\boldsymbol{H}}\left[\prod_{t=1}^{n_{\mathsf{b}}}\prod_{k=1}^{n_{\mathrm{R}}}\exp\left\{\frac{\lambda_k\left(\mathrm{Re}\{\boldsymbol{y}_t^\dagger \boldsymbol{u}_k\}\right)^2}{\left(\lambda_k + \frac{1}{\gamma}\right)}\right\}\right.$$
$$\left.\times\exp\left\{\frac{\lambda_k\left(\mathrm{Im}\{\boldsymbol{y}_t^\dagger \boldsymbol{u}_k\}\right)^2}{\left(\lambda_k + \frac{1}{\gamma}\right)}\right\}\left(\frac{1}{\lambda_k\gamma + 1}\right)\right] \quad (32)$$
$$= \frac{\exp\left\{-\|\boldsymbol{Y}\|^2\right\}}{\pi^{n_{\mathrm{R}}n_{\mathsf{b}}}}\mathbb{E}_{\boldsymbol{H}}\left[\prod_{k=1}^{n_{\mathrm{R}}}\exp\left\{\frac{\lambda_k\gamma\boldsymbol{u}_k^\dagger \boldsymbol{Y}\boldsymbol{Y}^\dagger \boldsymbol{u}_k}{(\lambda_k\gamma + 1)}\right\}\right.$$
$$\left.\times\left(\frac{1}{\lambda_k\gamma + 1}\right)^{n_{\mathsf{b}}}\right]. \quad (33)$$

Let $\boldsymbol{A}(\boldsymbol{\lambda})$ be a diagonal matrix with kth diagonal entry $a_k = \lambda_k\gamma/(\lambda_k\gamma + 1)$. 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. Therefore, we can express (33) as

$$p(\boldsymbol{Y}) = \frac{\exp\left\{-\|\boldsymbol{Y}\|^2\right\}}{\pi^{n_\mathrm{R} n_\mathrm{b}}} \int p(\boldsymbol{\lambda}) \prod_{k=1}^{n_\mathrm{R}} \left(\frac{1}{\lambda_k \gamma + 1}\right)^{n_\mathrm{b}}$$
$$\times \left[\int p(\boldsymbol{U}) \exp\left\{\mathrm{Tr}\left\{\boldsymbol{A}(\boldsymbol{\lambda}) \tilde{\boldsymbol{U}}^\dagger \boldsymbol{Y} \boldsymbol{Y}^\dagger \tilde{\boldsymbol{U}}\right\}\right\} \mathrm{d}\boldsymbol{U}\right] \mathrm{d}\boldsymbol{\lambda}. \tag{34}$$

The r.h.s. of (34) is precisely the setup in (22). 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,

$$p(\boldsymbol{Y}) = \frac{\prod_{k=1}^{n_\mathrm{R}}(k-1)!\, e^{-\|\boldsymbol{Y}\|^2}}{\det \boldsymbol{V}(\boldsymbol{d})\, \pi^{n_\mathrm{R} n_\mathrm{b}}} \int p(\boldsymbol{\lambda}) \frac{\det \boldsymbol{E}(\boldsymbol{a}(\boldsymbol{\lambda}), \boldsymbol{d})}{\det \boldsymbol{V}(\boldsymbol{a}(\boldsymbol{\lambda}))}$$
$$\times \prod_{k=1}^{n_\mathrm{R}} \left(\frac{1}{\lambda_k \gamma + 1}\right)^{n_\mathrm{b}} \mathrm{d}\boldsymbol{\lambda} \tag{35}$$

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

$$p(\boldsymbol{\lambda}) = \det^2 \boldsymbol{V}(\boldsymbol{\lambda}) \prod_{k=1}^{n_\mathrm{R}} \frac{e^{-\lambda_k} \lambda_k^{n_\mathrm{T} - n_\mathrm{R}}}{(n_\mathrm{R} - k)!\,(n_\mathrm{T} - k)!}. \tag{36}$$

Moreover,

$$[\det \boldsymbol{V}(\boldsymbol{a}(\boldsymbol{\lambda}))]^{-1} = \prod_{k > \ell} \left(\frac{\lambda_k \gamma}{\lambda_k \gamma + 1} - \frac{\lambda_\ell \gamma}{\lambda_\ell \gamma + 1}\right)^{-1}$$
$$= \prod_{k > \ell} \frac{1}{\gamma} \frac{(\lambda_k \gamma + 1)(\lambda_\ell \gamma + 1)}{\lambda_k - \lambda_\ell}$$
$$= \frac{1}{\gamma^{(n_\mathrm{R}^2 - n_\mathrm{R})}} \frac{\prod_{k > \ell} (\lambda_k \gamma + 1)(\lambda_\ell \gamma + 1)}{\det \boldsymbol{V}(\boldsymbol{\lambda})}$$
$$= \frac{1}{\gamma^{(n_\mathrm{R}^2 - n_\mathrm{R})}} \frac{\prod_{k=1}^{n_\mathrm{R}} (\lambda_k \gamma + 1)^{n_\mathrm{R} - 1}}{\det \boldsymbol{V}(\boldsymbol{\lambda})}. \tag{37}$$

Plugging (36) and (37) into (35) yields

$$p(\boldsymbol{Y}) = \frac{\exp\left\{-\|\boldsymbol{Y}\|^2\right\}}{\det \boldsymbol{V}(\boldsymbol{d})\, \gamma^{(n_\mathrm{R}^2 - n_\mathrm{R})} \pi^{n_\mathrm{R} n_\mathrm{b}} \prod_{k=1}^{n_\mathrm{R}}(n_\mathrm{T} - k)!} \int \det \boldsymbol{V}(\boldsymbol{\lambda})$$
$$\times \det \boldsymbol{E}(\boldsymbol{a}(\boldsymbol{\lambda}), \boldsymbol{d}) \prod_{k=1}^{n_\mathrm{R}} \frac{e^{-\lambda_k} \lambda_k^{n_\mathrm{T} - n_\mathrm{R}}}{(\lambda_k \gamma + 1)^{n_\mathrm{b} + 1 - n_\mathrm{R}}} \mathrm{d}\boldsymbol{\lambda}$$
$$= \frac{\exp\left\{-\|\boldsymbol{Y}\|^2\right\}}{\det \boldsymbol{V}(\boldsymbol{d})\, \pi^{n_\mathrm{R} n_\mathrm{b}} \prod_{k=1}^{n_\mathrm{R}}(n_\mathrm{T} - k)!} \int \det \boldsymbol{V}(\boldsymbol{\lambda}/\gamma)$$
$$\times \det \boldsymbol{E}(\boldsymbol{a}(\boldsymbol{\lambda}), \boldsymbol{d}) \prod_{k=1}^{n_\mathrm{R}} \frac{e^{-\lambda_k} \lambda_k^{n_\mathrm{T} - n_\mathrm{R}}}{(\lambda_k \gamma + 1)^{n_\mathrm{b} + 1 - n_\mathrm{R}}} \mathrm{d}\boldsymbol{\lambda}. \tag{38}$$

The multiple integral in (38) is an instance of (24). Simply identify

$$\Phi(\boldsymbol{\lambda}) = \det \boldsymbol{V}(\boldsymbol{\lambda}/\gamma) \tag{39}$$

$$\Psi(\boldsymbol{\lambda}) = \det \boldsymbol{E}(\boldsymbol{a}(\boldsymbol{\lambda}), \boldsymbol{d}) \tag{40}$$

$$\xi(x) = \frac{e^{-x} x^{n_\mathrm{T} - n_\mathrm{R}}}{(x\gamma + 1)^{n_\mathrm{b} + 1 - n_\mathrm{R}}} \tag{41}$$

to obtain

$$p(\boldsymbol{Y}) = \frac{\exp\left\{-\|\boldsymbol{Y}\|^2\right\}}{\pi^{n_\mathrm{R} n_\mathrm{b}}} \frac{\det \boldsymbol{Z}}{\det \boldsymbol{V}(\boldsymbol{d}) \prod_{k=1}^{n_\mathrm{R}}(n_\mathrm{T} - k)!} \tag{42}$$

where

$$Z_{ij} = \int_0^\infty (x/\gamma)^{i-1} \exp\left\{d_j \frac{x\gamma}{x\gamma + 1} - x\right\} \frac{x^{n_\mathrm{T} - n_\mathrm{R}}}{(x\gamma + 1)^{n_\mathrm{b} + 1 - n_\mathrm{R}}} \mathrm{d}x. \tag{43}$$

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, Jul. 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] 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.

[4] 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.

[5] 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, Sep. 2004.

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

[7] 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.

[8] 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.

[9] 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.

[10] N. Jindal and A. Lozano, "Optimum pilot overhead in wireless communication: A unified treatment of continuous and block-fading channels," *arXiv:0903.1379v1*, 2009.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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, Jun. 2007.

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

[19] 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.

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

[21] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*. Prentice Hall PTR, 2007.

[22] N. Jindal, A. Lozano, and T. Marzetta, "What is the value of joint processing of pilots and data in block-fading channels?" *Proc. of ISIT*, 2009.

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

[24] 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.