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 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.

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 \( n_R \times n_T \) matrix \( H \) 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 every block in an IID fashion. There is no antenna correlation and thus the entries of \( H \) are also independent.
Assembling into matrices the input, the output, and
the noise for the $n_b$ symbols within each block, their
relationship becomes

$$ Y = \sqrt{\frac{\text{SNR}}{n_t}} H X + N $$

(1)

where the input $X$ is an $n_r \times n_b$ matrix while the output
$Y$ and the noise $N$ are $n_r \times n_b$ matrices. Both $X$ and
$N$ 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 repre-
sents 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 $H$ can be established between block-
and continuous-fading models [12] whereby

$$ n_b = \frac{1}{2f_m T_s} $$

(2)

where $f_m$ and $T_s$ are the maximum Doppler frequency
and the symbol period, respectively. Typically, $f_m = (v/c)f_c$
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 $H$, 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

$$ I = \frac{1}{n_b} [\hat{h}(Y) - \hat{h}(Y|X)] $$

(3)

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

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

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

$$ \log h(Y|X) = n_r \log_2(e) \frac{e^{nt/\text{SNR}}}{n_t}\sum_{i=0}^{n_t-1} \sum_{j=0}^{n_t-1} \sum_{l=0}^{n_t-1} \left( \frac{2i - 2j}{i - j} \right)
\left( -1 \right)^{j} (2j)! (n_b - n_t + l)! 2^{2i-\ell} 2^{j} (2j)! (n_b - n_t + j)! \left( n_b - n_t + \ell \right) \sum_{q=0}^{n_t-1} E_q(\frac{nt}{\text{SNR}}) + n_r n_b \log_2(\pi e) $$

(4)

Proof: See Appendix.

Having expressed $h(Y|X)$ in (3), we now turn our
attention to $h(Y)$. The output $Y$ is affected by a com-
bination of multiplicative and additive noise, and thus
$Y$ is not Gaussian distributed. The unconditional output
distribution $p(Y)$, 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 $h(Y)$, the formula is easy to evaluate and
it thereby enables computing

$$ h(Y) = -\int p(Y) \log_2 p(Y) dY $$

(5)

$$ = -E \{ \log_2 p(Y) \}. $$

(6)

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

Proposition 2 For $1 \leq k \leq \min(n_r, n_t)$, define the func-
tions

$$ f_k(z) = \int_{0}^{\infty} \exp \left\{ \frac{x z}{\text{SNR}} - x \right\} \frac{z^{k-1} + [n_t - n_k]^{+}}{(n_r/\text{SNR}) + 1} \frac{n_b^{n_t - n_r}}{n_t^{n_r}} \frac{d}{dz} \left( \frac{dz}{n_t + 1} \right) dz. $$

(7)

Let $d = [d_1, \ldots, d_{n_r}]$ be the eigenvalues of $YY^\dagger$ and define
the $n_r \times n_r$ matrix $Z$ with entries

$$ Z_{ij} = \left\{ \begin{array}{ll}
\left( \frac{n_t}{\text{SNR}} \right)^{j-1} f_j(d_i), & 1 \leq i \leq n_r, 1 \leq j \leq s \\
d_i^{j-n_r-1}, & 1 \leq i \leq n_r, s + 1 \leq j \leq n_r
\end{array} \right. $$

(8)

where $s = \min\{n_r, n_t\}$. Then

$$ p(Y) = \frac{\pi^{-n_r n_t/2} |Y|^2}{\prod_{i,j=1}^{n_r} (d_{ij} - d_{ij}) \prod_{k=[n_t - n_r]^{+}}^{n_r-1} k!} \det Z. $$

(9)

Proof: See Appendix.

In the special case of memoryless channels, i.e., for
$n_b = 1$, the solution in Proposition 2 reduces to the one
in [2].

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

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

Algorithm 1: Evaluation of $I$.

1. Pre-compute $f_k(x)$, $1 \leq k \leq n_r$, on a discrete set $X$ 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 $p(Y)$.
4. Compute the sample average of $-\log_2 p(Y)$ via
Monte Carlo, thereby obtaining $\hat{h}(Y)$.
5. Compute $h(Y|X)$ 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 $\Delta x$ were chosen such that two decimal digits are correct with 90 percent 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 $X$ that are rotationally invariant and where the eigenvalue distribution of $X^T X$ (or of $XX^T$) is of the form

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

where $V(\cdot)$ 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 $\bar{I}$, 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

$$C(\text{SNR}) = E \left[ \log_2 \det \left( I + \frac{\text{SNR}}{n_T} HH^H \right) \right]$$

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 $\bar{I}(\text{SNR}) \geq \bar{I}_{\text{lower}}(\text{SNR})$ with

$$\bar{I}_{\text{lower}}(\text{SNR}) = C(\text{SNR}) - \frac{n_T n_p}{n_p} \log_2 \left( 1 + \frac{n_b}{n_T} \right).$$

C. Pilot-Based Communication

In pilot-based communication, $n_p$ pilot symbols are inserted within each fading block, leaving $n_b - n_p$ 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,

$$Y_p = \sqrt{\frac{\text{SNR}}{n_T}} HP + N_p$$

where the output, $Y_p$, and the noise, $N_p$, are $n_b \times n_p$ matrices. The entries of $N_p$ are IID zero-mean unit-variance complex Gaussian while $P$ is deterministic and satisfies $PP^\dagger = n_p I$ [17].

During the transmission of data symbols, in turn, (1) applies with $X$ and $N$ of dimension $n_T \times (n_b - n_p)$ and $n_H \times (n_b - n_p)$, 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

$$\max_{n_p:1 \leq n_p \leq n_b} \left( 1 - \frac{n_p}{n_b} \right) C \left( \frac{\text{SNR}^2 n_p/n_T}{1 + \text{SNR} (1 + n_p/n_T)} \right)$$

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 $n_p = n_T$ 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

$$\left( 1 - \frac{n_T}{n_b} \right) C \left( \frac{n_b SNR}{n_b - 2 n_T} \left( \sqrt{\gamma} - \sqrt{\gamma - 1} \right)^2 \right)$$

in bits/s/Hz, and with

$$\gamma = \frac{n_b SNR + n_T}{n_b SNR - 2 n_T}.$$  

The spectral efficiency in (14) is superior to that in (13). However, pilot power boosting increases the peakness 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 $T_s \approx 100 \mu s$. However, it could be shortened to $T_s \approx 10$–20 $\mu s$ 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.)
- 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.

1Eq. (14) requires that $n_b > 2 n_T$; variations thereof are also available for $n_b \leq 2 n_T$ [17].
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 \( \text{SNR} \).

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 \( \text{SNR} \). 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 \( \text{SNR} \), although this share diminishes with the \( \text{SNR} \). We further observe that, by optimizing the pilot overhead or the pilot power boost at every \( \text{SNR} \), 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 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 \( \text{SNR} \). This is indeed the case, and the \( \text{SNR} \) 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 \( \text{SNR} \) and \( n_b \).

Next, we see the impact of varying \( n_b \) and/or \( \text{SNR} \) 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 \( \text{SNR} \) 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 \( \text{SNR} = 3 \text{ dB} \) 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

$$\frac{E_b}{N_0} = \frac{\text{SNR}}{R/B} R$$

(16)

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

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

$$\frac{E_b}{N_0_{\text{min}}} = \frac{1}{\log_2 e}$$

(17)

which equals $-1.59$ dB. Without perfect CSI on the side, $\bar{I}(\text{SNR})$ with IID complex Gaussian inputs is convex below some (low) $\text{SNR}$ and concave above it [5]–[8]. It follows that $E_b/N_0$ is minimized at some finite $\text{SNR}$. However, this minimizing $\text{SNR}$, and the corresponding $E_b/N_0_{\text{min}}$, cannot be obtained using the low-$\text{SNR}$ expansions available in the literature because only the convex behavior of $\bar{I}(\text{SNR})$ is captured therein. Specifically, the most refined low-$\text{SNR}$ expansion available is [7]

$$\bar{I}(\text{SNR}) = \frac{n_b n_r}{2} \text{SNR}^2 + o(\text{SNR}^2)$$

(18)

based on which $E_b/N_0_{\text{min}}$ indeed cannot be obtained.\(^2\)

Applying the method presented in Section III, the correct $E_b/N_0_{\text{min}}$ and the corresponding $\text{SNR}$ can be calculated.

In pilot-based communication, the spectral efficiency is also a convex function of $\text{SNR}$ below some $\text{SNR}$ and concave above it [7]. Thus, $E_b/N_0_{\text{min}}$ is also achieved at some finite $\text{SNR}$, 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 $n_T = n_R = 1$ and let $n_b = 10$. Shown in Fig. 7 is the $E_b/N_0$ as function of $\text{SNR}$.

At this fading rate ($n_b = 10$), the most power-efficiency operating point is $\text{SNR} = -2.4$ dB. The corresponding $E_b/N_0_{\text{min}}$ 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 $E_b/N_0_{\text{min}}$ is suffered.

More generally, the method in Section III allows characterizing the power-efficient operating point and the corresponding $E_b/N_0_{\text{min}}$ as a function of the fading rate,

\(^2\)Eq. (18) would indicate that $E_b/N_0_{\text{min}}$ is achieved for $\text{SNR} \to \infty$, but (18) does not apply beyond the low-$\text{SNR}$ regime.

---

**VII. THE HIGH-SNR REGIME**

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

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

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

(19)

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-$\text{SNR}$ slope. Since $\bar{I}(\text{SNR})$ increases with the number of receive antennas (by the chain rule, additional outputs are never harmful), this implies that the optimal high-$\text{SNR}$ 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-$\text{SNR}$ 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 $\text{SNR}$. 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-$\text{SNR}$ mutual information. Also shown are $\bar{I}(\text{SNR})$ with perfect CSI, i.e., obtained from $C(\text{SNR})$.

---

**Fig. 6.** In solid, $\bar{I}(\text{SNR})$ for $\text{SNR} = 3$ 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.** $E_b/N_0$ as function of $\text{SNR}$ for $n_T = n_R = 1$ with $n_b = 10$. In solid, values obtained from $\bar{I}(\text{SNR})$ and also values corresponding to pilot-based communication, with and without pilot power boosting. For each curve, $(E_b^*/N_0)_{\text{min}}$ is explicitly indicated. In dashed, $E_b^*/N_0$ with perfect CSI, i.e., obtained from $C(\text{SNR})$.

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

**Fig. 9.** $(E_b^*/N_0)_{\text{min}}$ as function of $n_b$ for $n_T = n_R = 1$. Values obtained from $\bar{I}(\text{SNR})$ and also values corresponding to pilot-based communication, with and without pilot power boosting.
the perfect-CSI capacity and the high-SNR expansion of the true capacity, which for $n_T = n_R$ is given in [5] as

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

(20)

with

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

Although for $n_b \geq n_R + \min\{n_T, n_R\}$ the input $X$ that achieves the true capacity for $\text{SNR} \to \infty$ is an isotropically random unitary matrix $[9]$, [28]. IID Gaussian inputs seem to perform very well at high $\text{SNR}$, even in relatively fast fading. Non-asymptotically in the $\text{SNR}$, the optimum $X$ 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 $X$ is again the product of a unitary matrix and a nonnegative real diagonal matrix, even for $\text{SNR} \to \infty$. The high-SNR slope of the true capacity equals [5]

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

(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](http://www.dtic.upf.edu/~alozano/software).

ACKNOWLEDGMENTS

The authors thank Prof. Babak Hassibi (Caltech, USA) for pointing out some valuable references. The work of A. Lozano is supported by the Spanish Ministry of Science and Innovation (Refs. TEC2009-13000 and CONSOLIDER-INGENIO CSD2008-00010 "COMONSENS").

APPENDIX

A. Preliminaries

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

$$\int_{-\infty}^{\infty} \exp\{-x^2 A + x B\} \, dx = \exp\left\{\frac{B^2}{4A}\right\} \sqrt{\frac{\pi}{A}}.$$  (23)

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

$$\int e^{Tr(U DU^* Z)} \, dU = \prod_{m=1}^{M} (m-1)! \det E(d, b) \det V(d) \det V(b)$$

(24)

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

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

(25)

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

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

(26)

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

$$\lim_{b_{n+1}, \ldots, b_{M} \to 0} \frac{\det E(d, b)}{\det V(b)} = \prod_{k=1}^{N} b_k^{M-N-(k-1)!} \det V(b_{1 \ldots N})$$

(27)

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

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

$$\int \cdots \int_{\mathcal{D}_{ord}} \det \Psi(x) \det \Phi(x) \prod_{m=1}^{M} \xi(x_m) \, dx$$



$$= \det \left( \left\{ \int_{a}^{b} \psi_i(x) \phi_j(x) \xi(x) \, dx \right\}_{i,j=} \right)$$

(28)

where the multiple integral is over the domain $\mathcal{D}_{ord} = \{ b \geq x_1 \geq x_2 \geq \ldots \geq x_M \geq a \}$.
B. Proof of Proposition 1

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

Let $y$ be an arbitrary row of $Y$. The conditional covariance of the $n_T$-dimensional column vector $y^\dagger$ equals
\[
\mathbb{E}[y^\dagger y|X] = I + \frac{\text{SNR}}{n_T} X^\dagger X
\]
and thus
\[
\mathbb{h}(y|X) = \mathbb{h}(y^\dagger|X)
\]
\[
= \mathbb{E}\left[ \log_2 \left( \frac{\pi e^{2n_b}}{\det \left( I + \frac{\text{SNR}}{n_T} X^\dagger X \right)} \right) \right]
\]
with expectation over the distribution of $X$. Factoring out the term $n_b \log_2(\pi e)$, what remains coincides with the perfect-CSI capacity of a MIMO channel, only with the role of the channel played by $X$. Since the entries of $X$ 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 $n_b$, we convert $\mathbb{h}(y|X)$ into $\mathbb{h}(Y|X)$ as desired.

C. Proof of Proposition 2

Define $\gamma = \text{SNR}/n_T$, $s = \min\{n_T, n_R\}$ and $r = \max\{n_T, n_R\}$. Then, denoting by $x_t$ the $t$th column of $X$,
\[
p(Y) = \mathbb{E}_H \left[ \int p(Y|H, X) p(X) \, dX \right]
\]
\[
= \frac{1}{(\pi \gamma)^{n_T n_R}} \frac{1}{(\pi \gamma)^{n_T n_R}} \mathbb{E}_H \left[ \prod_{t=1}^{n_T} \int \exp \left\{ -\|y_t - H x_t\|^2 \right\} \times \exp \left\{ -\|x_t\|^2/\gamma \right\} \, dx_t \right].
\]
Using the singular-value decomposition $H = U \Sigma V^\dagger$, absorbing $V$ into $X$ through the variable substitution $\hat{x}_t = V^\dagger x_t$, and assembling the diagonal entries of $\Sigma \Sigma^\dagger$ into $\lambda = [\lambda_1, \ldots, \lambda_r]$, we may write
\[
p(Y) = \mathbb{E}_H \left[ \prod_{t=1}^{n_T} \exp \left\{ -\|\hat{x}_t\|^2 \right\} \times \exp \left\{ 2 \text{Re} \{ y_t^\dagger U^\dagger \Sigma^\dagger \hat{x}_t \} \right\} \times \exp \left\{ -\|\hat{x}_t\|^2/\gamma \right\} \, d\hat{x}_t \right].
\]
Applying (23) to each variable $\hat{x}_{t,k}$ in (34) gives
\[
p(Y) = \frac{\exp\left\{ -\|Y\|^2 \right\}}{\pi^{n_R n_b}} \mathbb{E}_H \left[ \prod_{t=1}^{n_T} \prod_{k=1}^{n_b} \exp \left\{ \frac{\lambda_k \left( \text{Re} \{ y_{t,k}^\dagger u_k \} \right)^2 \left( \lambda_k + \frac{1}{\gamma} \right)}{\lambda_k + \frac{1}{\gamma}} \right\} \right.
\]
\[
\times \exp \left\{ -\frac{\|y_{t,k}^\dagger u_k \Sigma^\dagger \hat{x}_{t,k} \|^2}{\lambda_k + \frac{1}{\gamma}} \right\} \times \left( \prod_{k=n_R+1}^{n_T} \exp \left\{ -\frac{||\hat{x}_{t,k}||^2}{\gamma} \right\} \right) \, d\hat{x}_{t,k} \right].
\]
Let $A(\lambda)$ be a $n_R \times n_R$ diagonal matrix with $k$th diagonal entry $a_{k} = \lambda_k \gamma / (\lambda_k + 1)$, $1 \leq k \leq s$ and $a_k = 0$, $s < k \leq n_R$. It is known that $U$, $\Sigma$, and $V$ in the singular-value decomposition of $H$ are independent random matrices and that both $U$ and $V$ are isotropically distributed. Thus, we can express (37) as
\[
p(Y) = \frac{\exp\left\{ -\|Y\|^2 \right\}}{\pi^{n_R n_b}} \mathbb{E}_H \left[ \prod_{k=1}^{n_R} \left( \frac{1}{\lambda_k + 1} \right)^{n_b} \right.
\]
\[
\times \left[ \int p(U) \exp \left\{ \text{Tr} \left\{ A(\lambda)^\dagger \hat{U} \hat{Y}^\dagger \hat{Y} \hat{U} \right\} \right\} \, dU \right] \, d\lambda
\]
where the density distribution of the (ordered) eigenvalues in $\lambda$ equals
\[
p(\lambda) = \det^2 V(\lambda) \prod_{k=1}^{s} \frac{e^{-\lambda_k} \lambda_k^{s-k}}{(n_R - k)! (n_T - k)!},
\]
and scaling by $1/s!$ yields the corresponding density of the unordered eigenvalues.

Henceforth, we separately consider the cases $n_T \leq n_R$ and $n_T < n_R$.

1) Case $n_T \geq n_R$: The r.h.s. of (38) is precisely the setup in (24). Let $a(\lambda) = [a_1, \ldots, a_{n_R}]$ contain the diagonal elements of $A(\lambda)$ and let $d$ contain the eigenvalues of $Y Y^\dagger$. Then,
\[
p(Y) = \frac{\prod_{k=1}^{n_R} (k-1)! e^{-\|Y\|^2}}{\det V(d) \pi^{n_R n_b}} \int p(\lambda) \frac{\det E(a(\lambda), d)}{\det V(a(\lambda))} \times \prod_{k=1}^{n_R} \left( \frac{1}{\lambda_k + 1} \right)^{n_b} \, d\lambda
\]
with $E(\cdot, \cdot)$ and $V(\cdot)$ as per (25) and (26), respectively. Moreover,
\[
\frac{1}{\det V(a(\lambda))} = \prod_{k>\ell} \left( \frac{\lambda_k \gamma}{\lambda_k + 1 - \lambda_\ell \gamma} \right)^{-1} \\
= \prod_{k>\ell} \frac{1}{\gamma} \left( \frac{\lambda_k \gamma + 1(\lambda_\ell \gamma + 1)}{\lambda_k - \lambda_\ell} \right) \\
= \frac{1}{\gamma^{(n_R - n_T)/2}} \prod_{k=1}^{n_R} (\lambda_k \gamma + 1)^n \det V(\lambda) \\
= \frac{1}{\gamma^{(n_R - n_T)/2}} \prod_{k=1}^{n_R} \det V(\lambda). \tag{41}
\]
Plugging (39) and (41) into (40) yields
\[
p(Y) = \frac{\exp \left\{ -\|Y\|^2 \right\}}{\pi^{n_R n_T} \pi^{n_T n_R} \prod_{k=1}^{n_T} (n_T - k)!} \int \det V(\lambda) \\
\times \det E(a(\lambda), d) \prod_{k=1}^{n_T} (\lambda_k \gamma + 1)^n e^{-\lambda_k} d\lambda \\
= \frac{\exp \left\{ -\|Y\|^2 \right\}}{\pi^{n_R n_T} \pi^{n_T n_R} \prod_{k=1}^{n_T} (n_T - k)!} \int \det V(\lambda) \\
\times \det E(a(\lambda), d) \prod_{k=1}^{n_T} (\lambda_k \gamma + 1)^n e^{-\lambda_k} d\lambda. \tag{42}
\]
The multiple integral in (42) is an instance of (28). Simply identify
\[
\Phi(\lambda) = \det V(\lambda/\gamma) \\
\Psi(\lambda) = \det E(a(\lambda), d) \\
\xi(x) = \frac{e^{-x} x^{n_T - n_R}}{(x + 1)^n x^{n_T - n_R}} \tag{45}
\]
to obtain
\[
p(Y) = \exp \left\{ -\|Y\|^2 \right\} \frac{\det Z}{\pi^{n_T n_R} \prod_{k=1}^{n_T} (n_T - k)!} \\
\]
where
\[
Z_{ij} = \int_0^\infty (x/\gamma)^{i-1} \exp \left\{ j \gamma (x + 1) - x \right\} \frac{x^{n_T - n_R}}{(x + 1)^n} dx. \tag{47}
\]
\textbf{Remark:} The proposition can be generalized to also encompass non-Gaussian inputs. If $X$ has rotational invariance and the corresponding $p(\lambda)$ if of the form
\[
p(\lambda) = \det^2 V(\lambda) \prod_{k=1}^{n_T} \frac{g_k(\lambda_k)}{(n_R - k)! (n_T - k)!}, \tag{42}
\]
(42) equals
\[
p(Y) = \exp \left\{ -\|Y\|^2 \right\} \frac{\det Z}{\pi^{n_T n_R} \prod_{k=1}^{n_T} (n_T - k)!} \int \det V(\lambda/\gamma) \\
\times \det E(a(\lambda), d) \prod_{k=1}^{n_T} \frac{g_k(\lambda_k)}{(\lambda_k \gamma + 1)^n} d\lambda. \tag{48}
\]
In the case that the functions $g_k(x) = g(x)$ for all $k$, (48) is still an instance of (28), but with
\[
\xi(x) = \frac{g(x)}{(x + 1)^n x^{n_T - n_R}}. \tag{49}
\]
Hence, Proposition 2 applies, but with a different matrix $Z$. 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 $\det Z$ is replaced by a general tensor-operator). These observations for non-Gaussian inputs hold for the case $n_T < n_R$ as well.

2) Case $n_T < n_R$: Let $a(\lambda) = [a_1, \ldots, a_{n_T}]$ contain the diagonal elements of $A(\lambda)$ and let $d$ contain the eigenvalues of $YY^\dagger$. 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,
\[
p(Y) = \frac{\gamma^{n_T - n_T}}{\pi^{n_T n_R} \pi^{n_T n_R} \prod_{k=1}^{n_T} (n_T - k)!} \int \det V(\lambda/\gamma) \\
\times \det E(a(\lambda), d) \prod_{k=1}^{n_T} (\lambda_k \gamma + 1)^n e^{-\lambda_k} d\lambda \tag{49}
\]
where the density of unordered eigenvalues has been used and thus the integration limits are 0 to $\infty$ for all variables.

Eq. (49) is not an instance of (28) due to the structure of the matrix $E(a(\lambda), d)$. However, the same result as in (28) still applies,
\[
\int \det V(\lambda/\gamma) \det E(a(\lambda), d) \prod_{n=1}^{n_T} (\lambda_n \gamma + 1)^n e^{-\lambda_n} d\lambda \\
= \sum_{\sigma \in P_{n_T}} \sum_{\sigma' \in P_{n_R}} (-1)^{\sigma + \sigma'} \prod_{n=1}^{n_T} \int \frac{\lambda_n \sigma_n - 1}{\gamma} d\lambda_n \prod_{n=1}^{n_R} d\sigma_n^{n - n_T - 1} \\
= n_T! \sum_{\sigma \in P_{n_T}} (-1)^{\sigma} \prod_{n=1}^{n_T} f_n(d_{\sigma_n}) \prod_{n=1}^{n_R} d\sigma_n^{n - n_T - 1} \tag{50}
\]
where
\[
f_n(x) = \int_0^\infty (z/\gamma)^{n-1} \exp \left\{ x \gamma (z + 1) - z \right\} \frac{1}{(z + 1)^n x^{n_T - n_R}} dz,
\]
$(-1)^\sigma$ denotes the sign of the permutation $\sigma$, $P_N$ denotes the set of all permutations of the integers $1, \ldots, N$, and the $(i, j)$th entry of $Z$ equals $Z_{ij} = f_j(d_i)$ if $j \leq n_T$ and $Z_{ij} = d_i^{n_T - n_T}$ if $j > n_T$.

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, 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] 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.

[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, Sep. 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. 972–983, 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, Jun. 2007.

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

[24] J. G. Andrews, A. Ghosh, and R. 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, Jun. 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. Verdú, “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, Jun. 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.