High-SNR Capacity of Wireless Communication Channels in the Noncoherent Setting: A Primer

Giuseppe Durisi
Department of Signals and Systems
Chalmers University of Technology, 41296 Gothenburg, Sweden
E-mail: {durisi}@chalmers.se
Helmut Bölcskei
Department of Information Technology and Electrical Engineering
ETH Zurich, 8092 Zurich, Switzerland
E-mail: {boelcskei}@nari.ee.ethz.ch

***Abstract*—This paper, mostly tutorial in nature, deals with the problem of characterizing the capacity of fading channels in the high signal-to-noise ratio (SNR) regime. We focus on the practically relevant *noncoherent setting*, where neither transmitter nor receiver know the channel realizations, but both are aware of the channel law. We present, in an intuitive and accessible form, two tools, first proposed by Lapidoth & Moser (2003), of fundamental importance to high-SNR capacity analysis: the *duality approach* and the *escape-to-infinity property* of capacity-achieving distributions. Furthermore, we apply these tools to refine some of the results that appeared previously in the literature and to simplify the corresponding proofs.**

I. Introduction

Most wireless communication systems operate in the *noncoherent setting* where neither transmitter nor receiver have *a priori* information on the realization of the underlying fading channel. As channel state information is typically acquired by allocating transmission time and/or bandwidth to channel estimation (a typical example is the use of pilot symbols [1]), a problem of significant practical relevance is to determine the optimal amount of resources to be used for this task. This problem can be addressed in a fundamental fashion by determining the Shannon capacity (i.e., the ultimate limit on the rate of reliable communication [2]) in the noncoherent setting. Unfortunately, corresponding analytical results are exceedingly difficult to obtain, even for simple channel models [3]; nevertheless, significant progress has been made during the past few years by studying the capacity behavior in the asymptotic regimes of high and low signal-to-noise ratio (SNR). Throughout this paper, we shall deal exclusively with the high-SNR regime. The capacity behavior at high SNR turns out to be very sensitive to the channel model used [4]–[6]. In this paper, we shall focus on a channel model—the *correlated block-fading* model [7], [8]—that is simple and yet rich enough to illustrate some of the possible asymptotic dependencies of capacity on SNR, namely, logarithmic with different *pre-log* factors [7], [5], [9], [10], or double-logarithmic [4]. The aim of this tutorial paper is two-fold:

- We present, in an intuitive and accessible manner, two tools that turn out to be exceedingly useful in the characterization of capacity at high SNR: the *duality approach* and the *escape-to-infinity property* of capacity-achieving distributions. These tools were first introduced in [4].
- We use these tools to refine a result that appeared previously in [7] and to provide an alternative and much simpler proof of a result in [9], [10]. Furthermore, we develop insights into the use of duality by exploiting the geometry of the correlated block-fading model.

Notation

Uppercase boldface letters denote matrices and lowercase boldface letters designate vectors. Uppercase sans-serif letters (e.g., Q) denote probability distributions,[1] while lowercase sans-serif letters (e.g., r) are reserved for probability density functions. The superscripts ^{T} and ^{H} stand for transposition and Hermitian transposition, respectively. We denote the identity matrix of dimension $N \times N$ by $\mathbf{I}_N$; $\operatorname{diag}\{\mathbf{a}\}$ is the diagonal square matrix whose main diagonal contains the entries of the vector $\mathbf{a}$, and $\lambda_q(\mathbf{A})$ stands for the qth largest eigenvalue of the Hermitian positive-semidefinite matrix $\mathbf{A}$. For a random vector $\mathbf{x}$ with distribution Q, we write $\mathbf{x} \sim \mathsf{Q}$. We denote expectation by $\mathbb{E}[\cdot]$, and use the notation $\mathbb{E}_{\mathbf{x}}[\cdot]$ or $\mathbb{E}_{\mathsf{Q}}[\cdot]$ to stress that expectation is taken with respect to $\mathbf{x} \sim \mathsf{Q}$. We write $D(\mathsf{Q}(\cdot)\|\mathsf{R}(\cdot))$ for the relative entropy between the distributions Q and R [2, Sec. 8.5]. Furthermore, $\mathcal{CN}(\mathbf{0}, \mathbf{R})$ stands for the distribution of a circularly-symmetric [11, Def. 24.3.2] complex Gaussian random vector with covariance matrix $\mathbf{R}$. For two functions $f(x)$ and $g(x)$, the notation $f(x) = \mathcal{O}(g(x))$, $x \to \infty$, means that $\limsup_{x\to\infty} \bigl|f(x)/g(x)\bigr| < \infty$, and $f(x) = o(g(x))$, $x \to \infty$, means that $\lim_{x\to\infty} \bigl|f(x)/g(x)\bigr| = 0$. Finally, $\log(\cdot)$ indicates the natural logarithm.

A. The Channel Model

In our quest for simplicity of exposition, we chose to focus on the correlated block-fading channel model [7], [8]. In this model, the channel changes in an independent fashion across blocks of N discrete-time samples and exhibits correlated fading within each block (with the same fading statistics for all blocks). The

[1] We will refer to probability distributions simply as distributions in the remainder of the paper.

input-output (IO) relation corresponding to one such block is given by:

$$\mathbf{y} = \mathrm{diag}\{\mathbf{h}\}\mathbf{x} + \mathbf{w}. \tag{1}$$

Here, $\mathbf{x} = [x_1 \dots x_N]^T \in \mathbb{C}^N$ is the (random) input vector, which we assume to satisfy the average-power constraint

$$\frac{1}{N}\,\mathbb{E}\left[\|\mathbf{x}\|^2\right] \le \rho. \tag{2}$$

The vector $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ represents additive white Gaussian noise (AWGN), and $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$ contains the fading channel coefficients. The vectors $\mathbf{x}$, $\mathbf{h}$, and $\mathbf{w}$ are mutually independent. We assume that $\mathbf{R}$ has rank Q ($1 \le Q \le N$) and that the main-diagonal entries of $\mathbf{R}$ are all equal to 1. Throughout the paper, we consider the noncoherent setting where transmitter and receiver know the statistics of $\mathbf{h}$, but not its realizations.

The model we just described may seem contrived at first sight. Yet, it is of practical relevance for at least two reasons. First, it captures the essence of channel variations (in time) in an accurate but simple way: the rank Q of $\mathbf{R}$ corresponds to the minimum number of entries of $\mathbf{h}$ that need to be known to perfectly recover the whole vector (in the absence of noise); therefore, larger Q corresponds to faster channel variation. Second, when $\mathbf{R}$ is circulant, the IO relation (1) coincides with the IO relation—in the frequency domain—of a cyclic-prefix orthogonal frequency-division multiplexing system [12] operating over a frequency-selective channel with Q uncorrelated taps. In other words, the model in (1) can be thought of as the dual of the widely used intersymbol-interference channel model. Independence across blocks is a sensible assumption for systems employing time-division multiple access or frequency hopping [13]. Finally, we remark that for the special case $Q = 1$, the channel model in (1) reduces to the *piecewise-constant* block-fading channel model previously used in numerous papers such as [13], [10], [9].

B. Channel Capacity

The capacity of the channel in (1) is given by

$$C(\rho) = \frac{1}{N} \sup_{\mathsf{Q}} I(\mathbf{x}; \mathbf{y}). \tag{3}$$

Here, $I(\mathbf{x}; \mathbf{y})$ denotes the mutual information [2, Sec. 8.5] between $\mathbf{x}$ and $\mathbf{y}$ in (1), and the supremum is taken over all distributions Q on $\mathbf{x}$ that satisfy the average-power constraint (2). Because the variance of the entries of $\mathbf{h}$ and $\mathbf{w}$ is normalized to one, we can interpret ρ as the receive SNR.

The literature is essentially void of analytic expressions for $C(\rho)$, even for the simplest case $N = 1$. Nevertheless, as we shall see in the next section, the high-SNR behavior of $C(\rho)$ can be characterized fairly well.

C. Known Results and Our Contributions

For the *general case* $1 \le Q \le N$, Liang and Veeravalli showed that [7, Props. 3 and 4]

$$C(\rho) = \frac{N-Q}{N} \log \rho + \mathcal{O}(\log\log\rho), \quad \rho \to \infty. \tag{4}$$

This result is sufficient to characterize the capacity pre-log χ, defined as the asymptotic ratio between capacity and the logarithm of SNR as SNR goes to infinity:

$$\chi = \lim_{\rho\to\infty} \frac{C(\rho)}{\log\rho}.$$

The pre-log can be interpreted as the fraction of signal-space dimensions that can be used for communication. From (4) we find the pre-log to be given by the difference of two terms, i.e., $\chi = 1 - Q/N$. The first term can be thought of as the capacity pre-log when the channel is known perfectly at the receiver (in this case, $\chi = 1$ [14]); the second term quantifies the loss in signal-space dimensions due to the lack of channel knowledge. Note that Q/N is the smallest fraction of entries of the N-dimensional vector $\mathbf{h}$ that need to be known to reconstruct the whole vector in the absence of noise.[2] Hence, we can further interpret the penalty term Q/N as the fraction of signal-space dimensions in which *pilot symbols* need to be transmitted to allow the receiver to learn the channel.

When $Q = N$, i.e., the channel correlation matrix has full rank, (4) implies that the pre-log is equal to 0. It turns out that in this case the $\mathcal{O}(\log\log\rho)$ term in (4) is tight and capacity grows double-logarithmically in SNR. This surprising result was proven in [7, Lem. 5]. In Section III-B, we shall refine the result in [7, Lem. 5] by providing the following, more accurate, high-SNR capacity characterization:

$$C(\rho) = \log\log\rho - \gamma - 1 - \frac{1}{N}\sum_{q=1}^{N} \log \lambda_q(\mathbf{R}) + o(1), \quad \rho \to \infty. \tag{5}$$

This result characterizes capacity (for $Q = N$) up to a $o(1)$ term (i.e., a term that vanishes as $\rho \to \infty$). In contrast, the expression provided in [7, Lem. 5] agrees with capacity only up to a $\mathcal{O}(1)$ term (i.e., a term that is bounded as $\rho \to \infty$).

The most important tool in the proof of (5) is the *duality approach*, a technique first introduced in [4] to characterize the capacity of stationary ergodic fading channels with finite differential entropy rate. The essence of the duality approach is that it allows one to obtain tight upper bounds on $C(\rho)$ by choosing *appropriate distributions* on the output $\mathbf{y}$. Compared to the treatment in [4], our goal in Sections II-A and III-B is to provide the simplest and most accessible proofs for the main results underlying the duality approach. This comes at the cost of generality (in terms of noise and fading statistics).

While finding a capacity characterization that—like (5)—is tight up to a $o(1)$ term for all Q with $1 \le Q \le N$ is an interesting open problem, for the special case $Q = 1$ (with $Q < N$) the following result was reported in [10], [9]:

$$C(\rho) = \frac{N-1}{N}\left[\log\rho + \log N - \gamma - 1\right] - \frac{\log\Gamma(N)}{N} + o(1), \quad \rho \to \infty. \tag{6}$$

[2] As we shall see, neglecting additive noise in (1) yields useful insights on the capacity pre-log.

Here, γ denotes the Euler-Mascheroni constant, and $\Gamma(\cdot)$ is the Gamma function [4, Eq. (197)]. The proof of (6) provided in [10] is based on a rather technical argument and does not seem to explicitly exploit the geometry in the problem, i.e., the fact that $\mathbf{x}$ and $\mathbf{y}$ are collinear in the absence of noise. The proof in [9] does exploit this geometry through an apposite change of variables, and applies to the multiple-antenna setting as well.

In Section III-A, we present a simple, alternative proof of (6) that, differently from the proofs in [10], [9], is based on duality and exploits the geometry in the problem to motivate the choice of the output distribution. Our proof needs another tool put forward in [4]: the *escape to infinity* property of the capacity achieving distribution. This property, which we review in Section II-B, allows one to restrict the maximization in (3) to a smaller set of distributions.

II. The Toolbox

A. The Duality Approach

To prove (5) and (6), we sandwich capacity between a lower and an upper bound that agree up to a $o(1)$ term. Establishing capacity lower bounds is, in principle, relatively simple: it suffices to evaluate the mutual information in (3) for an input distribution Q that satisfies the average-power constraint. Obviously, care must be exercised in choosing Q, so as to ensure that the resulting bound is tight in the limit $\rho \to \infty$ (see Section III-A2 for a concrete example).

Capacity upper bounds are more difficult to find because of the need for maximization over the set of eligible input distributions. To single out the main difficulty with this optimization problem, it is convenient to denote the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$ as $\mathsf{W}(\cdot\,|\,\mathbf{x})$ and to use the symbol QW to indicate the distribution induced on $\mathbf{y}$ by the input distribution Q and by the "channel" $\mathsf{W}(\cdot\,|\,\mathbf{x})$. By the definition of mutual information [2, Sec. 8.5] we have that

$$I(\mathbf{x};\mathbf{y}) = \mathbb{E}_{\mathsf{Q}}[D(\mathsf{W}(\cdot\,|\,\mathbf{x})\|(\mathsf{QW})(\cdot))]\,. \tag{7}$$

As the right-hand side (RHS) of (7) is a rather complicated function of Q, the maximization in (3) is difficult to carry out. The idea behind duality is to upper-bound the RHS of (7) by replacing QW by a distribution that does not depend on Q. Concretely, let R be an arbitrary distribution on $\mathbf{y}$. Then

$$\begin{aligned} I(\mathbf{x};\mathbf{y}) &\overset{(a)}{=} \mathbb{E}_{\mathsf{Q}}[D(\mathsf{W}(\cdot\,|\,\mathbf{x})\|\mathsf{R}(\cdot))] - D((\mathsf{QW})(\cdot)\|\mathsf{R}(\cdot)) \\ &\overset{(b)}{\leq} \mathbb{E}_{\mathsf{Q}}[D(\mathsf{W}(\cdot\,|\,\mathbf{x})\|\mathsf{R}(\cdot))]\,. \end{aligned} \tag{8}$$

Here, (a) follows from Topsøe's identity [15] and (b) is a consequence of the nonnegativity of relative entropy [2, Thm. 2.6.3]. The RHS of (8) is easier to deal with than $I(\mathbf{x};\mathbf{y})$. In fact, as we shall illustrate in Sections III-A and III-B, it is possible—for an appropriate choice of R—to find an asymptotically tight upper bound on $\mathbb{E}_{\mathsf{Q}}[D(\mathsf{W}(\cdot\,|\,\mathbf{x})\|\mathsf{R}(\cdot))]$ that holds for every Q satisfying the average-power constraint (2). By (3), this upper bound constitutes an upper bound on $C(\rho)$.

As a side remark, we note that the inequality (8) holds with equality when R coincides with QW. Hence, (8) yields the following expression for mutual information:

$$I(\mathbf{x};\mathbf{y}) = \inf_{\mathsf{R}} \mathbb{E}_{\mathsf{Q}}[D(\mathsf{W}(\cdot\,|\,\mathbf{x})\|\mathsf{R}(\cdot))]\,. \tag{9}$$

Through further manipulations (see [16], [4] for details), the identity (9) yields a *dual* expression for capacity, with the maximization over the input distribution in (3) replaced by a minimization over the output distribution. This is why the technique is referred to as duality approach.

An appropriate choice of the output distribution R is crucial for the bound in (8) to be tight. Throughout the paper, the output distribution R with density

$$\mathsf{r}(\mathbf{y}) = \frac{\Gamma(N)}{\pi^N \beta^\alpha \Gamma(\alpha)} \|\mathbf{y}\|^{2(\alpha-N)} e^{-\|\mathbf{y}\|^2/\beta}, \quad \mathbf{y} \in \mathbb{C}^N \tag{10}$$

will play a prominent role. Here, $\beta = N(\rho+1)/\alpha$, where α is a free parameter whose meaning will become clear later. This output distribution was put forward in [4] in a more general setting. The main features of this distribution are that $\mathbf{y}$ is isotropically distributed and that $\|\mathbf{y}\|^2$ is Gamma distributed with parameter α. In Section III-A, we will show that, for the piecewise-constant block-fading channel model (i.e., $Q=1$), this choice for the output distribution can be motivated through simple geometric intuition.

B. Escape-To-Infinity Property

Duality simplifies the maximization over the input distribution in (3), at the cost of getting an upper bound on capacity. This simplification, together with an appropriate choice of Q to obtain a matching capacity lower bound, is enough to establish (5), as we shall see in Section III-B. To prove (6), however, we need an additional tool. Specifically, we will make use of the fact that the asymptotic behavior of $C(\rho)$ does not change if we constrain the input distributions Q in (3) to be supported strictly outside a sphere of arbitrarily large radius. We formalize this result, which turns out to hold for almost all wireless channel models of practical interest [4], [17], in the following theorem. In view of (6), we focus on the case $Q=1$.

Theorem 1: Fix an arbitrary $\rho_0 > 0$ and let $\mathcal{K} = \{\mathbf{x} \in \mathbb{C}^N : \|\mathbf{x}\|^2 \leq \rho_0\}$. Denote by $C(\rho)$ the capacity of the channel with IO relation (1) (with $Q=1$) under the average-power constraint (2). Furthermore, denote by $C_{\mathcal{K}}(\rho)$ the capacity of the same channel under the additional constraint—besides (2)—that $\mathbf{x} \notin \mathcal{K}$ with probability one ($w.p.1$). Then

$$\lim_{\rho\to\infty}\Big[C(\rho) - \big(1-1/N\big)\log\rho\Big] = \lim_{\rho\to\infty}\Big[C_{\mathcal{K}}(\rho) - \big(1-1/N\big)\log\rho\Big].$$

Proof: The high-SNR capacity expansion (4) implies that the capacity pre-log is[3] $1-1/N$. The logarithmic growth of capacity in SNR allows us to invoke [17, Thm. 8] and conclude that the capacity-achieving input distribution must *escape to*

[3]It is worth mentioning that the proof of (4) does not make use of Theorem 1; so there is no cyclic argument here.

infinity [4, Def. 4.11], i.e., that for all $\rho_0 \geq 0$ there exists a family of input distributions $\{\mathsf{Q}_\rho\}_{\rho \geq 0}$ (parametrized with respect to ρ) satisfying $(1/N)\,\mathbb{E}_{\mathsf{Q}_\rho}\left[\|\mathbf{x}\|^2\right] \leq \rho$, such that, when $\mathbf{x} \sim \mathsf{Q}_\rho$,

$$\lim_{\rho\to\infty}\{C(\rho) - I(\mathbf{x};\mathbf{y})\} = 0$$

and

$$\lim_{\rho\to\infty}\mathbb{P}\{\|\mathbf{x}\|^2 \leq \rho_0\} = 0.$$

The proof is concluded by noting that the escape-to-infinity property is a sufficient condition for Theorem 1 to hold, as a consequence of [4, Thm. 4.12]. ■

III. High-SNR Capacity Asymptotics

A. The Rank-One Case

When $Q = 1$, we can rewrite (1) in the following (more convenient) form

$$\mathbf{y} = s\mathbf{x} + \mathbf{w} \tag{11}$$

where $s \sim \mathcal{CN}(0,1)$. The high-SNR capacity expansion (4) implies that the capacity pre-log of the channel in (11) is given by $1 - 1/N$. This is in agreement with the intuition we provided in Section I-C: one pilot symbol per block is enough to learn the channel in the absence of noise. We next provide a different interpretation of this result, which is of geometric nature and sheds light on how to select input and output distributions to get capacity bounds that are tight as $\rho \to \infty$.

1) Geometric Intuition: Let $\mathbf{x}$ be an arbitrary vector in $\mathbb{C}^N$. This vector can be specified by identifying i) the linear subspace spanned by $\mathbf{x}$, i.e., the complex line passing through the origin and $\mathbf{x}$ and ii) the point on that line corresponding to $\mathbf{x}$ (i.e., a complex number). If we neglect additive noise, the IO relation in (11) reduces to $\mathbf{y} = s\mathbf{x}$. As s varies, $\mathbf{y}$ spans the line $\mathbf{x}$ lies on. In other words—as pointed out in [9]—the random channel coefficient s destroys the information about $\mathbf{x}$ specified in the second step of our description above, but leaves the information about the linear subspace spanned by $\mathbf{x}$ unchanged. To summarize, when the random channel coefficient s is not known to the receiver, the information that the receiver can recover about the transmitted signal $\mathbf{x}$ is the line on which $\mathbf{x}$ lies. But a complex line in $\mathbb{C}^N$ is fully characterized by $N-1$ complex parameters.[4] Hence, the received signal "carries" $N-1$ parameters describing $\mathbf{x}$. This number, divided by N, coincides with the capacity pre-log.

2) A Capacity Lower Bound: The geometry unveiled in the previous section suggests to use the direction of $\mathbf{x}$, but not its magnitude, to convey information. This insight is helpful in choosing an input distribution that yields a tight capacity lower bound. Concretely, we take $\mathbf{x} = \sqrt{N\rho}\cdot\mathbf{u}_\mathbf{x}$ where $\mathbf{u}_\mathbf{x}$ is uniformly distributed on the unit sphere in $\mathbb{C}^N$. We use this input distribution, which trivially satisfies the average-power constraint (2), to lower-bound capacity as follows:

$$\begin{aligned} N\cdot C(\rho) &\geq I(\mathbf{x};\mathbf{y}) = \mathrm{h}(\mathbf{y}) - \mathrm{h}(\mathbf{y}\,|\,\mathbf{x}) \\ &\geq \mathrm{h}(\mathbf{y}\,|\,\mathbf{w}) - \mathrm{h}(\mathbf{y}\,|\,\mathbf{x}) \\ &= \mathrm{h}(\underbrace{s\mathbf{x}}_{\triangleq\mathbf{r}}) - \mathrm{h}(\mathbf{y}\,|\,\mathbf{x}). \end{aligned} \tag{12}$$

Here, $\mathrm{h}(\cdot)$ denotes differential entropy [2, Sec. 8.1], the second inequality follows because conditioning reduces differential entropy [2, Sec. 8.6], and the last equality follows because differential entropy is invariant to translations [2, Thm. 8.6.3] and $\mathbf{w}$ is independent of s and $\mathbf{x}$. To compute $\mathrm{h}(\mathbf{r})$, it is convenient to switch to polar coordinates, i.e., $\mathbf{r} \mapsto (\|\mathbf{r}\|, \mathbf{u}_\mathbf{r})$, where $\mathbf{u}_\mathbf{r} = \mathbf{r}/\|\mathbf{r}\|$. The change of variable theorem then yields [4, Lem. 6.17]:

$$\mathrm{h}(\mathbf{r}) = \mathrm{h}(\|\mathbf{r}\|) + \mathrm{h}_{\text{sphere}}(\mathbf{u}_\mathbf{r}\,|\,\|\mathbf{r}\|) + (2N-1)\,\mathbb{E}[\log\|\mathbf{r}\|]\,. \tag{13}$$

Here, $\mathrm{h}_{\text{sphere}}(\cdot)$ denotes the differential entropy computed with respect to the area measure on the unit sphere in $\mathbb{C}^N$ [4, p. 2457]. By the choice of the input distribution, we have $\|\mathbf{r}\| = \sqrt{N\rho}\,|s|$. Furthermore, because $\mathbf{u}_\mathbf{x}$ is uniformly distributed on the unit sphere and s is circularly symmetric (i.e., the phase of s is uniformly distributed on $[-\pi,\pi)$ and is independent of $|s|$ [11, Prop. 24.2.6]), it follows that $\mathbf{r}$ is isotropically distributed [4, Def. 6.19]. Hence, $\mathbf{u}_\mathbf{r}$ is uniformly distributed on the unit sphere and is independent of $\|\mathbf{r}\|$. Based on these observations, we can now simplify (13) as follows:

$$\begin{aligned} \mathrm{h}(\mathbf{r}) &\overset{(a)}{=} \mathrm{h}\big(\sqrt{N\rho}\,|s|\big) + \mathrm{h}_{\text{sphere}}(\mathbf{u}_\mathbf{r}) + (2N-1)\,\mathbb{E}[\log\|\mathbf{r}\|] \\ &\overset{(b)}{=} \log\sqrt{N\rho} + \mathrm{h}(|s|) + \log\frac{2\pi^N}{\Gamma(N)} \\ &\quad + (2N-1)\big[\log\sqrt{N\rho} + \mathbb{E}[\log|s|]\big] \\ &\overset{(c)}{=} N\log(N\rho) + \mathrm{h}\big(|s|^2\big) + \log\frac{\pi^N}{\Gamma(N)} \\ &\quad + (N-1)\,\mathbb{E}\left[\log|s|^2\right] \\ &\overset{(d)}{=} N\log(N\rho) + 1 + \log\frac{\pi^N}{\Gamma(N)} - (N-1)\gamma. \end{aligned} \tag{14}$$

Here, in (a) we used the independence of $\mathbf{u}_\mathbf{r}$ and $\|\mathbf{r}\|$ to drop conditioning in the second term on the RHS of (13). In (b) we used that $\mathrm{h}(ax) = \log a + \mathrm{h}(x)$ for x a real-valued random variable and a real and nonnegative; we also used that $\mathbf{u}_\mathbf{r}$ is uniformly distributed on the unit sphere in $\mathbb{C}^N$, and that, as a consequence, $\mathrm{h}_{\text{sphere}}(\mathbf{u}_\mathbf{r})$ is equal to the area of that sphere, i.e., $2\pi^N/\Gamma(N)$. In (c) we used that

$$\mathrm{h}(v) = \mathrm{h}(v^2) - \mathbb{E}[\log v] - \log 2$$

for every real nonnegative random variable v [4, Lem. 6.15], and (d) follows because $\mathbb{E}\left[\log|s|^2\right] = -\gamma$ and $\mathrm{h}(|s|^2) = 1$ for $s \sim \mathcal{CN}(0,1)$.

[4] More formally, the set of lines passing through the origin of $\mathbb{C}^N$ forms a manifold (the complex projective space $\mathbb{CP}^{N-1}$) of $N-1$ complex dimensions [18].

Since $\mathbf{h}$ is a circularly-symmetric complex Gaussian vector, $\mathrm{h}(\mathbf{y}\,|\,\mathbf{x})$ on the RHS of (12) admits the following closed-form expression:

$$\begin{aligned}\mathrm{h}(\mathbf{y}\,|\,\mathbf{x}) &= \log(\pi e)^N + \mathbb{E}\left[\log(1+\|\mathbf{x}\|^2)\right]\\ &= \log(\pi e)^N + \log(1+N\rho)\\ &= \log(\pi e)^N + \log(N\rho) + o(1), \quad \rho\to\infty. \qquad (15)\end{aligned}$$

Substituting (14) and (15) into (12), we get a lower bound on $C(\rho)$ that coincides with the RHS of (6).

3) A Matching Upper Bound: To obtain an upper bound that matches the lower bound we just found up to a $o(1)$ term, we use duality and the escape-to-infinity property of the capacity-achieving distribution. More specifically, as a consequence of Theorem 1, we can, without loss of generality, constrain the maximization of mutual information in (3) to input distributions Q that satisfy—besides the average-power constraint (2)—the additional constraint $\|\mathbf{x}\|^2 \geq \rho_0$ $w.p.1$. Here, $\rho_0 > 0$ is a parameter to be optimized later. We use duality with the density of R given by (10) with $\alpha = 1$. This choice is again motivated by the geometric considerations in Section III-A1: in the noiseless case, the density of the output distribution induced by the input distribution used in Section III-A2 (to derive a capacity lower bound), equals (10) with $\alpha = 1$. Fix $\rho_0 > 0$ and take an arbitrary input distribution Q such that $\mathbf{x}\sim\mathsf{Q}$ satisfies (2) and $\|\mathbf{x}\|^2 \geq \rho_0$ $w.p.1$. By duality, we have that

$$\begin{aligned}I(\mathbf{x};\mathbf{y}) &\leq \mathbb{E}_{\mathsf{Q}}[D(\mathsf{W}(\cdot\,|\,\mathbf{x})\|\mathsf{R}(\cdot))]\\ &= \mathbb{E}_{\mathsf{QW}}\left[\log\frac{1}{\mathsf{r}(\mathbf{y})}\right] - \mathrm{h}(\mathbf{y}\,|\,\mathbf{x})\\ &= \log\pi^N + \log[N(\rho+1)] - \log\Gamma(N)\\ &\quad + (N-1)\,\mathbb{E}_{\mathsf{QW}}\left[\log\|\mathbf{y}\|^2\right] + \frac{\mathbb{E}_{\mathsf{QW}}\left[\|\mathbf{y}\|^2\right]}{N(\rho+1)} - \mathrm{h}(\mathbf{y}\,|\,\mathbf{x}). \qquad (16)\end{aligned}$$

Here, the first equality follows from straightforward algebraic manipulations; in the second equality we used (10) with $\alpha = 1$. We shall next evaluate or bound the terms on the RHS of (16) that depend on Q. First, note that

$$\mathbb{E}_{\mathsf{QW}}\left[\|\mathbf{y}\|^2\right] = \mathbb{E}_{\mathsf{QW}}\left[\|s\mathbf{x}+\mathbf{w}\|^2\right] \leq N(\rho+1). \qquad (17)$$

Here, we used independence of $\mathbf{x}$ and $\mathbf{w}$ and the power constraint (2). To evaluate $\mathbb{E}_{\mathsf{QW}}\left[\log\|\mathbf{y}\|^2\right]$, we proceed as follows: first note that

$$\mathbb{E}_{\mathsf{QW}}\left[\log\|\mathbf{y}\|^2\right] = \mathbb{E}_{\mathbf{x}}\left[\mathbb{E}_{s,\mathbf{w}}\left[\log\|\mathbf{y}\|^2 \,\middle|\, \mathbf{x}\right]\right].$$

We next use that, given $\mathbf{x}$, the random variable $\|\mathbf{y}\|^2$ is distributed as $\sum_{i=1}^{N-1}|z_i|^2 + (1+\|\mathbf{x}\|^2)\,|z_N|^2$, where the z_i, $i = 1,2,\dots,N$, are i.i.d. $\mathcal{CN}(0,1)$. This result follows by observing that, given $\mathbf{x}$, the output vector $\mathbf{y}$ has covariance matrix $\mathbf{x}\mathbf{x}^H + \mathbf{I}_N$ (whose eigenvalues are $1+\|\mathbf{x}\|^2$ and 1 with multiplicity $N-1$). Using Jensen's inequality with respect to the random variables $z_1,\dots,z_{N-1}$, we obtain the following bound:

$$\begin{aligned}&\mathbb{E}_{\mathbf{x}}\left[\mathbb{E}_{s,\mathbf{w}}\left[\log\|\mathbf{y}\|^2 \,\middle|\, \mathbf{x}\right]\right]\\ &= \mathbb{E}_{\mathbf{x}}\left[\mathbb{E}_{z_1,\dots,z_N}\left[\log\left(\sum_{i=1}^{N-1}|z_i|^2 + (1+\|\mathbf{x}\|^2)\,|z_N|^2\right)\middle|\,\mathbf{x}\right]\right]\\ &\leq \mathbb{E}_{\mathbf{x},z_N}\left[\log\left(N-1+(1+\|\mathbf{x}\|^2)\,|z_N|^2\right)\right]\\ &= \mathbb{E}_{\mathbf{x}}\left[\log(1+\|\mathbf{x}\|^2)\right] + \mathbb{E}_{\mathbf{x},z_N}\left[\log\left(\frac{N-1}{1+\|\mathbf{x}\|^2} + |z_N|^2\right)\right] \qquad (18)\\ &\leq \mathbb{E}_{\mathbf{x}}\left[\log(1+\|\mathbf{x}\|^2)\right]\\ &\quad + \sup_{\|\mathbf{x}\|^2\geq\rho_0}\mathbb{E}_{z_N}\left[\log\left(\frac{N-1}{1+\|\mathbf{x}\|^2} + |z_N|^2\right)\right]. \qquad (19)\end{aligned}$$

In the last step, we upper-bounded the second term on the RHS of (18) by replacing the expectation over $\mathbf{x}$ by the supremum over all vectors $\mathbf{x}$ satisfying $\|\mathbf{x}\|^2 \geq \rho_0$. This is the step where the escape-to-infinity property is used. Without this property, the supremum in (19) would be over all $\mathbf{x}$ satisfying $\|\mathbf{x}\|^2 \geq 0$ and the resulting bound would not match (up to a $o(1)$ term) the lower bound obtained in Section III-A2. To evaluate the second term on the RHS of (19) we use the following lemma:

Lemma 2: Let $z\sim\mathcal{CN}(0,1)$ and take $a > 0$. Then

$$\mathbb{E}_z\left[\log(a+|z|^2)\right] = \underbrace{e^a\Gamma(0,a) + \log a}_{\triangleq g(a)}.$$

Here $\Gamma(\cdot,\cdot)$ denotes the incomplete Gamma function [4, Eq. (200)]. The function $g(a)$ is monotonically increasing in a. Furthermore, $\lim_{a\to 0} g(a) = -\gamma$.

Proof: Let $v = |z|^2$. Then

$$\begin{aligned}\mathbb{E}_z\left[\log(a+|z|^2)\right] &= \int_0^\infty e^{-v}\log(a+v)\,dv\\ &= e^a\int_a^\infty e^{-t}\log t\,dt \qquad (20)\\ &= e^a\Gamma(0,a) + \log a. \qquad (21)\end{aligned}$$

Here, to obtain the second equality we used integration by parts. Now let us denote the RHS of (21) by $g(a)$. It is easy to verify that $g(a)$ is a monotonic function of $a \geq 0$. In fact,

$$\frac{dg(a)}{da} = e^a\Gamma(0,a)$$

which is nonnegative for all $a \geq 0$. Finally, the claim that $\lim_{a\to 0} g(a) = -\gamma$ follows from (20) by setting $a = 0$. ■

As a consequence of Lemma 2, we have that

$$\sup_{\|\mathbf{x}\|^2\geq\rho_0}\mathbb{E}_{z_N}\left[\log\left(\frac{N-1}{1+\|\mathbf{x}\|^2} + |z_N|^2\right)\right] = g\left(\frac{N-1}{1+\rho_0}\right).$$

Finally, for the conditional differential entropy term in (16) we have

$$\mathrm{h}(\mathbf{y}\,|\,\mathbf{x}) = \log(\pi e)^N + \mathbb{E}\left[\log(1+\|\mathbf{x}\|^2)\right].$$

To summarize, we proved that

$$I(\mathbf{x};\mathbf{y}) \le \log(N\rho + N) + (N-2)\,\mathbb{E}\left[\log(1+\|\mathbf{x}\|^2)\right] + (N-1)g\left(\frac{N-1}{1+\rho_0}\right) - \log\Gamma(N) - (N-1).$$

Now, using Jensen's inequality on $\mathbb{E}\left[\log(1+\|\mathbf{x}\|^2)\right]$, we obtain

$$\lim_{\rho\to\infty}\left[I(\mathbf{x};\mathbf{y}) - (N-1)\log\rho\right] \le (N-1)\left[\log N - 1 + g\left(\frac{N-1}{1+\rho_0}\right)\right] - \log\Gamma(N).$$

The proof is concluded by recalling that, by Theorem 1, the asymptotic behavior of $C(\rho)$ does not change if we constrain Q to satisfy $\|\mathbf{x}\|^2 \ge \rho_0\ w.p.1$, and by noting that, by Lemma 2, we can make the term $g((N-1)/(1+\rho_0))$ to be arbitrarily close to $-\gamma$ by taking ρ_0 sufficiently large.

B. The Full-Rank Case

Due to space constraints, we shall give an outline only of the proof of (5) and, furthermore, restrict ourselves to i.i.d. channels, i.e., $\mathbf{R} = \mathbf{I}_N$. We comment on the general case at the end of the section.

First, we note that $\mathbf{R} = \mathbf{I}_N$ implies that the channel is memoryless, and, hence, capacity is achieved by i.i.d. inputs. As a consequence,

$$\sup_{\mathsf{Q}} I(\mathbf{x};\mathbf{y}) = N \sup_{\widetilde{\mathsf{Q}}} I(x;y). \tag{22}$$

Here, $y = sx + w$ with $s, w \sim \mathcal{CN}(0,1)$ and the supremum is over the distributions $\widetilde{\mathsf{Q}}$ on x that satisfy the average-power constraint $\mathbb{E}_{\widetilde{\mathsf{Q}}}\left[|x|^2\right] \le \rho$. The capacity of the memoryless channel $y = sx + w$ was first proven to grow double-logarithmically in SNR in [19]. This result was then extended in [4, Thm. 4.2] to multiple-antenna channels with general stationary ergodic fading distribution (of finite differential entropy rate) and general noise distributions. The proof we provide here is based on the duality technique and is particularly simple, as it exploits the Gaussianity of the fading distribution. More specifically, we use duality with the density of R given in (10), with $N = 1$ and $\alpha = [1 + \log(1+\rho)]^{-1}$. The choice of α might appear unmotivated, and in fact, differently from the previous section, it is hard to find an intuitive explanation for this choice, besides the fact that it simplifies the proof. Consider an arbitrary $\widetilde{\mathsf{Q}}$ satisfying $\mathbb{E}_{\widetilde{\mathsf{Q}}}\left[|x|^2\right] \le \rho$; using (8) and (10), we obtain the following upper bound on $I(x;y)$:

$$\begin{aligned} I(x;y) &\le \log\pi + \alpha\log(1+\rho) - \alpha\log\alpha + \log\Gamma(\alpha) \\ &\quad + (1-\alpha)\,\mathbb{E}_{\widetilde{\mathsf{Q}}\mathsf{W}}\left[\log|y|^2\right] + \alpha\frac{\mathbb{E}_{\widetilde{\mathsf{Q}}\mathsf{W}}\left[|y|^2\right]}{1+\rho} - \mathrm{h}(y\,|\,x) \\ &\le \log\pi + \alpha[1+\log(1+\rho)] - \alpha\log\alpha + \log\Gamma(\alpha) \\ &\quad + \mathbb{E}_{\widetilde{\mathsf{Q}}\mathsf{W}}\left[\log|y|^2\right] - \mathrm{h}(y\,|\,x). \end{aligned} \tag{23}$$

The last step follows because $\mathbb{E}_{\widetilde{\mathsf{Q}}\mathsf{W}}\left[|y|^2\right] \le 1+\rho$ and $\alpha < 1$, by assumption, so that

$$(1-\alpha)\,\mathbb{E}_{\widetilde{\mathsf{Q}}\mathsf{W}}\left[\log|y|^2\right] \le \mathbb{E}_{\widetilde{\mathsf{Q}}\mathsf{W}}\left[\log|y|^2\right].$$

We continue by establishing that

$$\mathbb{E}_{\widetilde{\mathsf{Q}}\mathsf{W}}\left[\log|y|^2\right] - \mathrm{h}(y\,|\,x) = -\gamma - \log\pi - 1.$$

This identity follows because

$$\mathrm{h}(y\,|\,x) = \mathbb{E}\left[\log(1+|x|^2)\right] + \log(\pi e)$$

and because, given x, the random variable $|y|^2$ is distributed as $(1+|x|^2)\,|z|^2$, where $z \sim \mathcal{CN}(0,1)$, so that

$$\begin{aligned} \mathbb{E}_{\widetilde{\mathsf{Q}}\mathsf{W}}\left[\log|y|^2\right] &= \mathbb{E}_x\left[\mathbb{E}_{s,w}\left[\log|y|^2\,\middle|\,x\right]\right] \\ &= \mathbb{E}_x\left[\mathbb{E}_z\left[\log\left[(1+|x|^2)\,|z|^2\right]\,\middle|\,x\right]\right] \\ &= \mathbb{E}_x\left[\log(1+|x|^2)\right] + \underbrace{\mathbb{E}_z\left[\log|z|^2\right]}_{-\gamma}. \end{aligned}$$

Finally, since $\alpha[1+\log(1+\rho)] = 1$ and since [4, Eq. (337)]

$$\log\Gamma(\alpha) - \alpha\log\alpha + \log\alpha = o(1), \quad \rho\to\infty,$$

we get

$$\begin{aligned} I(x;y) &\le \log\pi + \underbrace{\alpha[1+\log(1+\rho)]}_{1} \\ &\quad + \underbrace{\log\Gamma(\alpha) - \alpha\log\alpha + \log\alpha}_{o(1),\ \rho\to\infty} - \log\alpha \\ &\quad + \underbrace{\mathbb{E}_{\widetilde{\mathsf{Q}}\mathsf{W}}\left[\log|y|^2\right] - \mathrm{h}(y\,|\,x)}_{-\gamma-\log\pi-1} \qquad (24) \\ &\le -\log\alpha - \gamma + o(1), \quad \rho\to\infty \\ &= \log\log\rho - \gamma + o(1), \quad \rho\to\infty. \qquad (25) \end{aligned}$$

This upper bound, which suffices to conclude that capacity grows at most double-logarithmically in ρ, can actually be tightened. A more careful choice of the output distribution makes the term $\alpha[1+\log(1+\rho)]$ vanish as $\rho\to\infty$, so that $-\gamma$ in (25) gets replaced by $-\gamma-1$ (see [4, App. VII] for details). This modified upper bound is tight in the sense that one can find a capacity lower bound that matches it up to a $o(1)$ term (see [4, Thm. 4.16]). To summarize, for $\mathbf{R} = \mathbf{I}_N$, we have that

$$C(\rho) = \log\log\rho - \gamma - 1 + o(1), \quad \rho\to\infty. \tag{26}$$

We conclude by noting that the double-logarithmic growth of capacity in SNR holds for every full-rank channel covariance matrix $\mathbf{R}$. Correlation among the channel entries, however, results in a different constant term in (26). More specifically, the final result in (5) follows from [4, Lem. 4.5] and from an adaptation of [4, Thm. 4.41] to the block-fading setup considered here.

IV. Open Problems

Duality is the main tool we used to establish the novel capacity expansion (5) for the full-rank case and to provide an alternative, simple proof of (6) for the rank-1 case (i.e., the piecewise-constant block-fading channel model). For the latter case, in particular, we showed how the geometry of the communication problem at hand can be used to find an output distribution that yields an asymptotically tight upper bound. Finding a $o(1)$-accurate capacity characterization when $1 < Q < N$ is an interesting open problem.

Throughout the paper, we focused exclusively on the single-antenna setup. In the multiple-antenna case, not even a pre-log characterization is available when $1 < Q < N$. A pre-log lower bound for the single-input multiple-output (SIMO) case has been obtained recently in [8]. Surprisingly, the bound in [8] implies that the SIMO pre-log can be larger than the pre-log in the single-input single-output case. Establishing whether this bound is tight is an open problem. For two further channel models of practical interest, namely, the single-antenna frequency-selective and doubly-selective fading channel the state of affairs is similar: pre-log lower bounds have been reported in [20] and [21], respectively. Establishing whether these bounds are tight is again an open problem.

References

[1] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, U.S.A.: Wiley, 2006.

[3] I. C. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz), "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.

[4] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

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

[6] G. Durisi, V. I. Morgenshtern, H. Bölcskei, U. G. Schuster, and S. Shamai (Shitz), "Information theory of underspread WSSUS channels," in *Wireless Communications over Rapidly Time-Varying Channels*, F. Hlawatsch and G. Matz, Eds. Academic Press, Mar. 2011, ch. 2, pp. 65–115.

[7] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3095–3110, Dec. 2004.

[8] V. I. Morgenshtern, G. Durisi, and H. Bölcskei, "The SIMO pre-log can be larger than the SISO pre-log," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Austin, TX, U.S.A., Jun. 2010, pp. 320–324.

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

[10] B. M. Hochwald and T. L. Marzetta, "Unitary space–time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.

[11] A. Lapidoth, *A Foundation in Digital Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[12] A. Peled and A. Ruiz, "Frequency domain data transmission using reduced computational complexity algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 5, Denver, CO, U.S.A., Apr. 1980, pp. 964–967.

[13] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[14] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[15] F. Topsøe, "An information theoretical identity and a problem involving capacity," *Studia Scientiarum Math. Hung.*, vol. 2, pp. 291–292, 1967.

[16] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless systems*. Orlando, FL, U.S.A.: Academic Press, Inc., 1982.

[17] A. Lapidoth and S. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 437–453, Feb. 2006.

[18] W. M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, 2nd ed. Orlando, FL, U.S.A.: Academic Press, Inc., 1986.

[19] G. Taricco and M. Elia, "Capacity of fading channel with no side information," *Electron. Lett.*, vol. 33, no. 16, pp. 1368–1370, Jul. 1997.

[20] H. Vikalo, B. Hassibi, B. M. Hochwald, and T. Kailath, "On the capacity of frequency-selective channels in training-based transmission schemes," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2572–2583, Sep. 2004.

[21] A. P. Kannu and P. Schniter, "On the spectral efficiency of noncoherent doubly selective channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2829–2844, Jun. 2010.