Upper and Lower Bounds on the Capacity of Amplitude-Constrained MIMO Channels

Alex Dytso[*], Mario Goldenbaum[*], Shlomo Shamai (Shitz)[†], and H. Vincent Poor[*]

[*]Department of Electrical Engineering, Princeton University
[†]Department of Electrical Engineering, Technion – Israel Institute of Technology

Abstract—In this work, novel upper and lower bounds on the capacity of channels with arbitrary constraints on the support of the channel input symbols are derived. As an immediate practical application, the case of multiple-input multiple-output channels with *amplitude constraints* is considered. The bounds are shown to be within a constant gap if the channel matrix is invertible and are tight in the high amplitude regime for arbitrary channel matrices. Moreover, in the high amplitude regime, it is shown that the capacity scales linearly with the minimum of the numbers of transmit and receive antennas, similarly to the case of average power-constrained inputs.

I. INTRODUCTION

While the capacity of a multiple-input multiple-output (MIMO) channel with an average power constraint is well understood [1], surprisingly, little is known about the capacity of the more practically relevant case in which the channel inputs are subject to *amplitude constraints*. The first major contribution to this problem was a seminal work of Smith [2] in which it was shown that, for the scalar Gaussian noise channel with an amplitude-constrained input, the capacity achieving inputs are discrete with finite support. In [3], this result was extended to peak-power-limited quadrature Gaussian channels. Using the approach of [3], in [4] the optimal input distribution was shown to be discrete for MIMO channels with an identity channel matrix and a Euclidian norm constraint on the input vector. Even though the optimal input distribution is known to be discrete, very little is known about the number or the optimal positions of the corresponding constellation points. To the best of our knowledge, the only exception is the work of [5] in which for a scalar Gaussian noise channel it was shown that two point masses are optimal for amplitude values smaller than 1.671 and three for amplitude values of up to 2.786.

Using a dual capacity expression, in [6] McKellips derived an upper bound on the capacity of a scalar amplitude constrained channel that is asymptotically tight in the high amplitude regime. By using a clever choice of an auxiliary channel output distribution in the dual capacity expression, the authors of [7] sharpened McKellips' bound and extended it to parallel MIMO channels with a Euclidian norm constraint on the input. The scalar version of the upper bound in [7] has

been further sharpened in [8] by yet another choice of auxiliary output distribution. In [9], asymptotic lower and upper bounds for a 2×2 MIMO system were presented and the gap between the bounds was specified.

In this work, we make progress on this open problem by deriving several new upper and lower bounds that hold for channels with *arbitrary constraints* on the support of the input distribution. We then apply them to the special case of MIMO channels with amplitude-constrained inputs.

A. Contributions and Paper Outline

Our contributions and paper outline are as follows. The problem is stated in Section II. In Section III, we derive upper and lower bounds on the capacity of a MIMO channel with an arbitrary constraint on the support of the input. In Section IV, we evaluate the performance of our bounds by studying MIMO channels with invertible channel matrices. In particular, in Theorem 6 it is shown that our upper and lower bounds are within $n \log(\rho)$ bits, where ρ is the packing efficiency and n is the number of antennas. For diagonal channel matrices, it is shown in Theorem 7 that the Cartesian product of pulse-amplitude modulation (PAM) constellations achieves the capacity to within $1.64n$ bits. Section V is devoted to MIMO channels with arbitrary channel matrices. It turns out that in the high amplitude regime, similarly to the average power constrained channel, it is shown that the capacity scales linearly with the minimum of the numbers of transmit and receive antennas. Section VI concludes the paper.

Due to space constraints, we omit the proofs of some of the results. All details, however, can be found in an extended version of this paper [10].

B. Notation

Vectors are denoted by bold lowercase letters, random vectors by bold uppercase letters, and matrices by bold uppercase sans serif letters (e.g., $\mathbf{x}$, $\mathbf{X}$, X). For any deterministic vector $\mathbf{x} \in \mathbb{R}^n$, $n \in \mathbb{N}$, we denote the Euclidean norm of $\mathbf{x}$ by $\|\mathbf{x}\|$. For some $\mathbf{X} \in \mathsf{supp}(\mathbf{X}) \subseteq \mathbb{R}^n$ and any $p > 0$ we define

$$\|\mathbf{X}\|_p^p := \frac{1}{n} \mathbb{E}[\|\mathbf{X}\|^p] , \tag{1}$$

where $\mathsf{supp}(\mathbf{X})$ denotes the support of $\mathbf{X}$. Note that for $p \geq 1$, the quantity in (1) defines a norm. The norm of a matrix $\mathbf{H} \in \mathbb{R}^{n \times n}$ is defined as

$$\|\mathbf{H}\| := \sup_{\mathbf{x} : \mathbf{x} \neq \mathbf{0}} \|\mathbf{H}\mathbf{x}\| / \|\mathbf{x}\| .$$

Let $\mathcal{S}$ be a subset of $\mathbb{R}^n$. Then,

$$\mathrm{Vol}(\mathcal{S}) := \int_{\mathcal{S}} \mathrm{d}\mathbf{x}$$

denotes its volume.

Let $\mathbb{R}_+ := \{x \in \mathbb{R} : x \geq 0\}$. We define an n-dimensional ball or radius $r \in \mathbb{R}_+$ centered at $\mathbf{x} \in \mathbb{R}^n$ as the set

$$\mathcal{B}_{\mathbf{x}}(r) := \{\mathbf{y} : \|\mathbf{x} - \mathbf{y}\| \leq r\} .$$

Recall that for any $\mathbf{x} \in \mathbb{R}^n$ and $r \in \mathbb{R}_+$,

$$\mathrm{Vol}\big(\mathcal{B}_{\mathbf{x}}(r)\big) = \frac{\pi^{\frac{n}{2}}}{\Gamma\big(\frac{n}{2}+1\big)} r^n .$$

For any matrix $\mathbf{H} \in \mathbb{R}^{k \times n}$ and some $\mathcal{S} \subset \mathbb{R}^n$ we define

$$\mathbf{H}\mathcal{S} := \{\mathbf{y} : \mathbf{y} = \mathbf{H}\mathbf{x}, \, \mathbf{x} \in \mathcal{S}\} .$$

Note that for an invertible $\mathbf{H} \in \mathbb{R}^{n \times n}$ we have

$$\mathrm{Vol}(\mathbf{H}\mathcal{S}) = |\det(\mathbf{H})|\mathrm{Vol}(\mathcal{S}) .$$

We define the maximum and minimum radius of a set $\mathcal{S} \subset \mathbb{R}^n$ that contains the origin as

$$r_{\mathsf{max}}(\mathcal{S}) := \min\{r \in \mathbb{R}_+ : \mathcal{S} \subset \mathcal{B}_{\mathbf{0}}(r)\} ,$$
$$r_{\mathsf{min}}(\mathcal{S}) := \max\{r \in \mathbb{R}_+ : \mathcal{B}_{\mathbf{0}}(r) \subseteq \mathcal{S}\} .$$

For a given vector $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}_+^n$ we define

$$\mathrm{Box}(\mathbf{a}) := \{\mathbf{x} \in \mathbb{R}^n : |x_i| \leq a_i, i = 1, \ldots, n\}$$

and the smallest box containing a given set $\mathcal{S} \subset \mathbb{R}^n$ as

$$\mathrm{Box}(\mathcal{S}) := \inf\{\mathrm{Box}(\mathbf{a}) : \mathcal{S} \subseteq \mathrm{Box}(\mathbf{a})\} ,$$

respectively. Finally, all logarithms are taken to the base 2, $\log^+(x) := \max\{\log(x), 0\}$, $Q(x)$, $x \in \mathbb{R}$, denotes the Q-function, and $\delta_{\mathbf{x}}(\mathbf{y})$ the Kronecker delta, which is one for $\mathbf{x} = \mathbf{y}$ and zero otherwise.

II. PROBLEM STATEMENT

Consider a MIMO system with $n_t \in \mathbb{N}$ transmit and $n_r \in \mathbb{N}$ receive antennas. The corresponding n_r-dimensional channel output for a single channel use is of the form

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z} ,$$

for some fixed channel matrix $\mathbf{H} \in \mathbb{R}^{n_r \times n_t}$.[1] Here and hereafter, we assume $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n_r})$ is independent of the channel input $\mathbf{X} \in \mathbb{R}^{n_t}$ and $\mathbf{H}$ is known to both the transmitter and the receiver, where $\mathbf{I}_{n_r}$ denotes the $n_r \times n_r$ identity matrix.

Now, in all that follows let $\mathcal{X} \subset \mathbb{R}^{n_t}$ be a *convex and compact* channel input space that contains the origin (i.e., the length-n_t zero vector) and let $F_{\mathbf{X}}$ denote the cumulative distribution function of $\mathbf{X}$. As of the writing of this paper, the capacity

$$C(\mathcal{X}, \mathbf{H}) := \max_{F_{\mathbf{X}} : \mathbf{X} \in \mathcal{X}} I(\mathbf{X}; \mathbf{H}\mathbf{X} + \mathbf{Z}) \qquad (2)$$

of this channel is unknown and we are interested in finding novel lower and upper bounds. Even though most of the results

[1]Considering a real-valued channel model is without loss of generality.

in this paper hold for arbitrary $\mathcal{X}$, we are mainly interested in the two most important special cases:

(i) per-antenna amplitude constraints; that is, $\mathcal{X} = \mathrm{Box}(\mathbf{a})$ for some given $\mathbf{a} = (A_1, \ldots, A_{n_t}) \in \mathbb{R}_+^{n_t}$, and

(ii) n_t-dimensional amplitude constraint; that is, $\mathcal{X} = \mathcal{B}_{\mathbf{0}}(A)$ for some given $A \in \mathbb{R}_+$.

Remark 1. Note that determining the capacity of a MIMO channel with average per-antenna power constraints is also still an open problem and has been solved for some special cases only [11]–[13].

III. UPPER AND LOWER BOUNDS ON THE CAPACITY

A. Upper Bounds

To establish our first upper bound on (2), we need the following result [14, Th. 1]:

Lemma 1. (Maximum Entropy Under p-th Moment Constraint) *Let $n \in \mathbb{N}$ and $p \in (0, \infty)$ be arbitrary. Then, for any $\mathbf{U} \in \mathbb{R}^n$ such that $h(\mathbf{U}) < \infty$ and $\|\mathbf{U}\|_p < \infty$, we have*

$$h(\mathbf{U}) \leq n \log\Big(k_{n,p}\, n^{\frac{1}{p}} \|\mathbf{U}\|_p\Big) ,$$

where

$$k_{n,p} := \frac{\sqrt{\pi}\, \mathrm{e}^{\frac{1}{p}} \big(\frac{p}{n}\big)^{\frac{1}{p}} \Gamma\big(\frac{n}{p}+1\big)^{\frac{1}{n}}}{\Gamma\big(\frac{n}{2}+1\big)^{\frac{1}{n}}} .$$

Theorem 1. (Moment Upper Bound) *For any channel input space $\mathcal{X}$ and any fixed channel matrix $\mathbf{H}$, we have*

$$C(\mathcal{X}, \mathbf{H}) \leq \bar{C}_{\mathsf{M}}(\mathcal{X}, \mathbf{H}) := \inf_{p>0} n_r \log\bigg(\frac{k_{n_r,p}}{(2\pi\mathrm{e})^{\frac{1}{2}}} n_r^{\frac{1}{p}} \|\tilde{\mathbf{x}} + \mathbf{Z}\|_p\bigg),$$

where $\tilde{\mathbf{x}} \in \mathbf{H}\mathcal{X}$ is chosen such that $\|\tilde{\mathbf{x}}\| = r_{\mathsf{max}}(\mathbf{H}\mathcal{X})$.

Proof: Expressing (2) in terms of differential entropies results in

$$\begin{aligned}
C(\mathcal{X}, \mathbf{H}) &= \max_{F_{\mathbf{X}} : \mathbf{X} \in \mathcal{X}} h(\mathbf{H}\mathbf{X} + \mathbf{Z}) - h(\mathbf{Z}) \\
&\overset{a)}{\leq} \max_{F_{\mathbf{X}} : \mathbf{X} \in \mathcal{X}} n_r \log\bigg(\frac{k_{n_r,p}}{(2\pi\mathrm{e})^{\frac{1}{2}}} n_r^{\frac{1}{p}} \|\mathbf{H}\mathbf{X} + \mathbf{Z}\|_p\bigg) \\
&\overset{b)}{=} n_r \log\bigg(\frac{k_{n_r,p}}{(2\pi\mathrm{e})^{\frac{1}{2}}} n_r^{\frac{1}{p}} \max_{F_{\mathbf{X}} : \mathbf{X} \in \mathcal{X}} \|\mathbf{H}\mathbf{X} + \mathbf{Z}\|_p\bigg), \quad (3)
\end{aligned}$$

where $a)$ follows from Lemma 1 with the fact that $h(\mathbf{Z}) = \frac{n_r}{2}\log(2\pi\mathrm{e})$ and $b)$ from the monotonicity of the logarithm.

Now, notice that $\|\mathbf{H}\mathbf{X} + \mathbf{Z}\|_p$ is linear in $F_{\mathbf{X}}$ and therefore it attains its maximum at an extreme point of the set $\mathcal{F}_{\mathbf{X}} := \{F_{\mathbf{X}} : \mathbf{X} \in \mathcal{X}\}$ (i.e., the set of all cumulative distribution functions of $\mathbf{X}$). As a matter of fact [15], the extreme points of $\mathcal{F}_{\mathbf{X}}$ are given by the set of degenerate distributions on $\mathcal{X}$; that is, $\{F_{\mathbf{X}}(\mathbf{y}) = \delta_{\mathbf{x}}(\mathbf{y}), \mathbf{y} \in \mathcal{X}\}_{\mathbf{x} \in \mathcal{X}}$. This allows us to conclude

$$\max_{F_{\mathbf{X}} : \mathbf{X} \in \mathcal{X}} \|\mathbf{H}\mathbf{X} + \mathbf{Z}\|_p = \max_{\mathbf{x} \in \mathcal{X}} \|\mathbf{H}\mathbf{x} + \mathbf{Z}\|_p .$$

Observe that the Euclidian norm is a convex function, which is therefore maximized at the boundary of the set $\mathbf{H}\mathcal{X}$. Combining this with (3) and taking the infimum over $p > 0$ completes the proof. ∎

The following theorem provides two alternative upper bounds that are based on duality arguments.

Theorem 2. (Duality Upper Bounds) *For any channel input space $\mathcal{X}$ and any fixed channel matrix $\mathbf{H}$*

$$C(\mathcal{X}, \mathbf{H}) \leq \bar{C}_{\mathsf{Dual},1}(\mathcal{X}, \mathbf{H}) := \log\left(c_{n_r}(d) + \frac{\mathrm{Vol}\big(\mathcal{B}_{\mathbf{0}}(d)\big)}{(2\pi\mathrm{e})^{\frac{n_r}{2}}}\right),$$
(4)

where

$$d := r_{\mathsf{max}}(\mathbf{H}\mathcal{X}), \; c_{n_r}(d) := \sum_{i=1}^{n_r-1} \binom{n_r-1}{i} \frac{\Gamma\big(\frac{n_r-1}{2}\big)}{2^{\frac{n_r}{2}}\Gamma\big(\frac{n_r}{2}\big)} d^i,$$

and

$$C(\mathcal{X}, \mathbf{H}) \leq \bar{C}_{\mathsf{Dual},2}(\mathcal{X}, \mathbf{H}) := \sum_{i=1}^{n_r} \log\left(1 + \frac{2A_i}{\sqrt{2\pi\mathrm{e}}}\right), \quad (5)$$

where $\mathbf{a} = (A_1, \ldots, A_{n_r})$ such that $\mathrm{Box}(\mathbf{a}) = \mathrm{Box}(\mathbf{H}\mathcal{X})$.

Proof: Using duality bounds, it has been shown in [7] that for any centered n-dimensional ball of radius $r \in \mathbb{R}_+$

$$\max_{F_{\mathbf{X}}: \mathbf{X} \in \mathcal{B}_{\mathbf{0}}(r)} I(\mathbf{X}; \mathbf{X} + \mathbf{Z}) \leq \log\left(c_n(r) + \frac{\mathrm{Vol}\big(\mathcal{B}_{\mathbf{0}}(r)\big)}{(2\pi\mathrm{e})^{\frac{n}{2}}}\right), \quad (6)$$

where $c_n(r) := \sum_{i=1}^{n-1}\binom{n-1}{i}\frac{\Gamma\big(\frac{n-1}{2}\big)}{2^{\frac{n}{2}}\Gamma\big(\frac{n}{2}\big)}r^i$.

Now, observe that

$$
\begin{aligned}
C(\mathcal{X}, \mathbf{H}) &= \max_{F_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} h(\mathbf{H}\mathbf{X} + \mathbf{Z}) - h(\mathbf{H}\mathbf{X} + \mathbf{Z}|\mathbf{H}\mathbf{X}) \\
&= \max_{F_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} I(\mathbf{H}\mathbf{X}; \mathbf{H}\mathbf{X} + \mathbf{Z}) \\
&= \max_{F_{\tilde{\mathbf{X}}}: \tilde{\mathbf{X}} \in \mathbf{H}\mathcal{X}} I(\tilde{\mathbf{X}}; \tilde{\mathbf{X}} + \mathbf{Z}) \qquad (7) \\
&\overset{a)}{\leq} \max_{F_{\tilde{\mathbf{X}}}: \tilde{\mathbf{X}} \in \mathcal{B}_{\mathbf{0}}(d), d := r_{\mathsf{max}}(\mathbf{H}\mathcal{X})} I(\tilde{\mathbf{X}}; \tilde{\mathbf{X}} + \mathbf{Z}) \\
&\overset{b)}{\leq} \log\left(c_{n_r}(d) + \frac{\mathrm{Vol}\big(\mathcal{B}_{\mathbf{0}}(d)\big)}{(2\pi\mathrm{e})^{\frac{n_r}{2}}}\right).
\end{aligned}
$$

Here, $a)$ follows from enlarging the optimization domain and $b)$ from using the upper bound in (6). This proves (4).

In order to show the upper bound in (5), we proceed with an alternative upper bound to (7):

$$
\begin{aligned}
C(\mathcal{X}, \mathbf{H}) &= \max_{F_{\tilde{\mathbf{X}}}: \tilde{\mathbf{X}} \in \mathbf{H}\mathcal{X}} I(\tilde{\mathbf{X}}; \tilde{\mathbf{X}} + \mathbf{Z}) \\
&\overset{a)}{\leq} \max_{F_{\tilde{\mathbf{X}}}: \tilde{\mathbf{X}} \in \mathrm{Box}(\mathbf{H}\mathcal{X})} I(\tilde{\mathbf{X}}; \tilde{\mathbf{X}} + \mathbf{Z}) \\
&\overset{b)}{\leq} \max_{F_{\tilde{\mathbf{X}}}: \tilde{\mathbf{X}} \in \mathrm{Box}(\mathbf{H}\mathcal{X})} \sum_{i=1}^{n_r} I(\tilde{X}_i; \tilde{X}_i + Z_i) \\
&\overset{c)}{=} \sum_{i=1}^{n_r} \max_{F_{\tilde{X}_i}: |\tilde{X}_i| \leq A_i} I(\tilde{X}_i; \tilde{X}_i + Z_i) \\
&\overset{d)}{\leq} \sum_{i=1}^{n_r} \log\left(1 + \frac{2A_i}{\sqrt{2\pi\mathrm{e}}}\right),
\end{aligned}
$$

where the (in)equalities follow from: $a)$ enlarging the optimization domain; $b)$ single-letterizing the mutual information; $c)$ choosing individual amplitude constraints $(A_1, \ldots, A_{n_r}) =: \mathbf{a} \in \mathbb{R}_+^{n_r}$ such that $\mathrm{Box}(\mathbf{a}) = \mathrm{Box}(\mathbf{H}\mathcal{X})$; and $d)$ using the upper bound in (6) for $n = 1$. This concludes the proof. ∎

B. Lower Bounds

A classical approach to bound a mutual information from below is to use the entropy power inequality (EPI).

Theorem 3. (EPI Lower Bounds) *For any fixed channel matrix $\mathbf{H}$ and any channel input space $\mathcal{X}$ with $\mathbf{X}$ absolutely continuous, we have*

$$C(\mathcal{X}, \mathbf{H}) \geq \underline{C}_{\mathsf{EPI}}(\mathcal{X}, \mathbf{H}) := \max_{F_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} \frac{n_r}{2} \log\left(1 + \frac{2^{\frac{2}{n_r}h(\mathbf{H}\mathbf{X})}}{2\pi\mathrm{e}}\right).$$
(8)

Moreover, if $n_t = n_r = n$, $\mathbf{H} \in \mathbb{R}^{n \times n}$ is invertible, and $\mathbf{X}$ is uniformly distributed over $\mathcal{X}$, then

$$C(\mathcal{X}, \mathbf{H}) \geq \underline{C}_{\mathsf{EPI}}(\mathcal{X}, \mathbf{H}) := \frac{n}{2} \log\left(1 + \frac{|\det(\mathbf{H})|^{\frac{2}{n}}\mathrm{Vol}(\mathcal{X})^{\frac{2}{n}}}{2\pi\mathrm{e}}\right).$$
(9)

Proof: By means of the EPI

$$2^{\frac{2}{n_r}h(\mathbf{H}\mathbf{X}+\mathbf{Z})} \geq 2^{\frac{2}{n_r}h(\mathbf{H}\mathbf{X})} + 2^{\frac{2}{n_r}h(\mathbf{Z})},$$

we conclude

$$2^{\frac{2}{n_r}C(\mathcal{X}, \mathbf{H})} \geq 1 + (2\pi\mathrm{e})^{-1} 2^{\frac{2}{n_r}\max_{F_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} h(\mathbf{H}\mathbf{X})}, \quad (10)$$

which finalizes the proof of the lower bound in (8).

To show the lower bound in (9), all we need is to recall that

$$h(\mathbf{H}\mathbf{X}) = h(\mathbf{X}) + \log|\det(\mathbf{H})|,$$

which is maximized for $\mathbf{X}$ uniformly distributed over $\mathcal{X}$. But if $\mathbf{X}$ is uniformly drawn from $\mathcal{X}$, we have

$$2^{\frac{2}{n}h(\mathbf{H}\mathbf{X})} = \mathrm{Vol}(\mathbf{H}\mathcal{X})^{\frac{2}{n}} = |\det(\mathbf{H})|^{\frac{2}{n}}\mathrm{Vol}(\mathcal{X})^{\frac{2}{n}},$$

which completes the proof. ∎

The results in [2]–[4] suggest that the channel input distribution that maximizes (2) might be discrete. Therefore, there is a need for lower bounds that, unlike the bounds in Theorem 3, rely on discrete inputs.

Remark 2. We note that the problem of finding the optimal input distribution of a general MIMO channel with an amplitude constraint is still open. The technical difficulty relies on the fact that the identity theorem from complex analysis, a key tool in the method developed by Smith [2] for the scalar case, does not extend to $\mathbb{R}^n$ and $\mathbb{C}^n$. The interested reader is referred to [16] for a detailed discussion on this issue with examples of why the identity theorem fails in the MIMO setting.

Theorem 4. (Ozarow-Wyner Type Lower Bound) *Let $\mathbf{X}_D \in \mathsf{supp}(\mathbf{X}_D) \subset \mathbb{R}^{n_t}$ be a discrete random vector of finite entropy, $g : \mathbb{R}^{n_r} \to \mathbb{R}^{n_t}$ a measurable function, and $p > 0$. Furthermore, let $\mathcal{K}_p$ be a set of continuous random vectors,*

independent of $\mathbf{X}_D$, *such that for every* $\mathbf{U} \in \mathcal{K}_p$ *we have* $h(\mathbf{U}) < \infty$, $\|\mathbf{U}\|_p < \infty$, *and*

$$\mathrm{supp}(\mathbf{U} + \mathbf{x}_i) \cap \mathrm{supp}(\mathbf{U} + \mathbf{x}_j) = \varnothing \qquad (11)$$

for all $\mathbf{x}_i, \mathbf{x}_j \in \mathrm{supp}(\mathbf{X}_D)$, $i \neq j$. *Then,*

$$C(\mathcal{X}, \mathbf{H}) \geq \underline{C}_{\mathsf{OW}}(\mathcal{X}, \mathbf{H}) := [H(\mathbf{X}_D) - \mathsf{gap}]^+,$$

where

$$\mathsf{gap} := \inf_{\mathbf{U},g,p} \left(G_{1,p}(\mathbf{U}, \mathbf{X}_D, g) + G_{2,p}(\mathbf{U}) \right)$$

with

$$G_{1,p}(\mathbf{U}, \mathbf{X}_D, g) := n_t \log\left(\frac{\|\mathbf{U} + \mathbf{X}_D - g(\mathbf{Y})\|_p}{\|\mathbf{U}\|_p} \right),$$

$$G_{2,p}(\mathbf{U}) := n_t \log\left(\frac{k_{n_t,p} \, n_t^{\frac{1}{p}} \|\mathbf{U}\|_p}{2^{\frac{1}{n_t} h(\mathbf{U})}} \right),$$

and $k_{n_t,p}$ *as defined in Lemma 1, respectively.*

Proof: The proof is identical to the proof of [14, Th.2] and omitted due to space constraints. ∎

Interestingly, the bound in Theorem 4 holds for arbitrary channels and the interested reader is referred to [14] for details.

We conclude this section by providing a lower bound that is based on Jensen's inequality and holds for arbitrary inputs.

Theorem 5. (Jensen's Inequality Lower Bound) *For any channel input space* $\mathcal{X}$ *and fixed channel matrix* $\mathbf{H}$, *we have*

$$C(\mathcal{X}, \mathbf{H}) \geq \underline{C}_{\mathsf{Jensen}}(\mathcal{X}, \mathbf{H})$$

$$:= \max_{F_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} \log^+\left(\left(\frac{2}{e}\right)^{\frac{n_r}{2}} \mathbb{E}\left[e^{-\frac{\|\mathbf{H}(\mathbf{X} - \mathbf{X}')\|^2}{4}} \right]^{-1} \right),$$

where $\mathbf{X}'$ *is an independent copy of* $\mathbf{X}$.

Proof: In order to show the lower bound, we follow an approach of [17]. Note that by Jensen's inequality

$$h(\mathbf{Y}) = -\mathbb{E}[\log f_{\mathbf{Y}}(\mathbf{Y})] \geq -\log \mathbb{E}[f_{\mathbf{Y}}(\mathbf{Y})]$$

$$= -\log \int_{\mathbb{R}^{n_r}} f_{\mathbf{Y}}(\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y}) \, d\mathbf{y}. \quad (12)$$

Now, evaluating the integral in (12) results in

$$\int_{\mathbb{R}^{n_r}} f_{\mathbf{Y}}(\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y}) \, d\mathbf{y}$$

$$= \frac{1}{(2\pi)^{n_r}} \int_{\mathbb{R}^{n_r}} \mathbb{E}\left[e^{-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{X}\|^2}{2}} \right] \mathbb{E}\left[e^{-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{X}'\|^2}{2}} \right] d\mathbf{y}$$

$$\overset{a)}{=} \frac{1}{(2\pi)^{n_r}} \mathbb{E}\left[\int_{\mathbb{R}^{n_r}} e^{-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{X}\|^2 + \|\mathbf{y} - \mathbf{H}\mathbf{X}'\|^2}{2}} d\mathbf{y} \right]$$

$$\overset{b)}{=} \frac{1}{(2\pi)^{n_r}} \mathbb{E}\left[e^{-\frac{\|\mathbf{H}\mathbf{X} - \mathbf{H}\mathbf{X}'\|^2}{4}} \int_{\mathbb{R}^{n_r}} e^{-\|\mathbf{y}\|^2} d\mathbf{y} \right]$$

$$\overset{c)}{=} \frac{1}{2^{\frac{n_r+1}{2}} \pi^{\frac{n_r}{2}}} \mathbb{E}\left[e^{-\frac{\|\mathbf{H}(\mathbf{X} - \mathbf{X}')\|^2}{4}} \right], \quad (13)$$

where $a)$ follows from the independence of $\mathbf{X}$ and $\mathbf{X}'$ and Tonelli's theorem, $b)$ from completing a square, and $c)$ from the fact that $\int_{\mathbb{R}^{n_r}} e^{-\|\mathbf{y}\|^2} d\mathbf{y} = (2\pi)^{\frac{n_r}{2}}$.
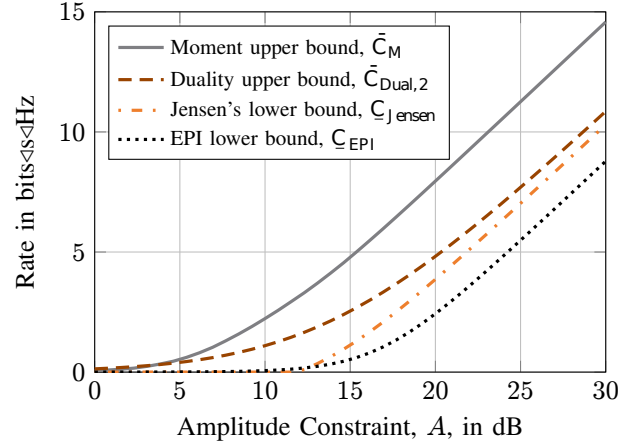

Fig. 1. Comparison of the upper and lower bounds of Theorems 1–3 and 5 evaluated for a 2×2 MIMO system with amplitude constraints $A_1 = A_2 = A$ (i.e., $\mathbf{a} = (A, A)$) and channel matrix $\mathbf{H} = \left(\begin{smallmatrix} 0.3 & 0 \\ 0 & 0.1 \end{smallmatrix} \right)$.

Finally, combining (12) and (13), subtracting $h(\mathbf{Z}) = \frac{n_r}{2} \log(2\pi e)$, and maximizing over $F_{\mathbf{X}}$ proves the result. ∎

In Fig. 1, the upper bounds of (5) and Theorem 1 and the lower bounds of Theorems 3 and 5 are evaluated for a diagonal 2×2 MIMO channel with per-antenna amplitude constraints. It turns out that the moment upper bound and the EPI lower bound perform well in the small amplitude regime while the duality upper bound and Jensen's inequality lower bound perform well in the high amplitude regime (note that the Jensen's inequality lower bound becomes strictly positive around $9\,\mathrm{dB}$).

IV. INVERTIBLE CHANNEL MATRICES

In this section, we consider the special case of $n_t = n_r = n$ antennas with $\mathbf{H} \in \mathbb{R}^{n \times n}$ being invertible. Our first result bounds the gap between (2) and the lower bound in (9).

Theorem 6. *Let* $\mathbf{H} \in \mathbb{R}^{n \times n}$ *be of full rank and*

$$\rho(\mathcal{X}, \mathbf{H}) := \frac{\mathrm{Vol}\big(\mathcal{B}_{\mathbf{0}}\left(r_{\max}(\mathbf{H}\mathcal{X})\right)\big)}{\mathrm{Vol}(\mathbf{H}\mathcal{X})}.$$

Then,

$$C(\mathcal{X}, \mathbf{H}) - \underline{C}_{\mathsf{EPI}}(\mathcal{X}, \mathbf{H}) \leq \frac{n}{2} \log\left((\pi n)^{\frac{1}{n}} \rho(\mathcal{X}, \mathbf{H})^{\frac{2}{n}} \right).$$

Proof: For notational convenience let the volume of an n-dimensional ball of radius $r > 0$ be denoted as

$$V_n(r) := \mathrm{Vol}\big(\mathcal{B}_{\mathbf{0}}(r)\big) = V_n(1) r^n = \frac{\pi^{\frac{n}{2}} r^n}{\Gamma\left(\frac{n}{2} + 1\right)}.$$

First, observe that by choosing $p = 2$, the upper bound of Theorem 1 can further be upper bounded as

$$\bar{C}_{\mathsf{M}}(\mathcal{X}, \mathbf{H}) \leq n \log\left(\frac{k_{n,2}}{(2\pi e)^{\frac{1}{2}}} n^{\frac{1}{2}} \|\tilde{\mathbf{x}} + \mathbf{Z}\|_2 \right)$$

$$\overset{a)}{=} \frac{n}{2} \log\left(\frac{1}{n} \mathbb{E}\big[\|\tilde{\mathbf{x}} + \mathbf{Z}\|^2 \big] \right)$$

$$\overset{b)}{=} \frac{n}{2} \log\left(1 + \frac{1}{n} \|\tilde{\mathbf{x}}\|^2 \right),$$

where $a)$ follows since $k_{n,2} = \sqrt{\frac{2\pi e}{n}}$ and $b)$ since $\mathbb{E}[\|\mathbf{Z}\|^2] = n$. Therefore, the gap between (9) and the moment upper bound of Theorem 1 can be upper bounded as follows:

$$
\begin{aligned}
\bar{C}_{\mathsf{M}}(\mathcal{X}, \mathbf{H}) - \underline{C}_{\mathsf{EPI}}(\mathcal{X}, \mathbf{H}) &= \frac{n}{2} \log\left(\frac{1 + \frac{1}{n}\|\tilde{\mathbf{x}}\|^2}{1 + \frac{\mathrm{Vol}(\mathbf{H}\mathcal{X})^{\frac{2}{n}}}{2\pi e}} \right) \\
&\stackrel{a)}{=} \frac{n}{2} \log\left(\frac{1 + \frac{1}{n}\left(\frac{V_n(\|\tilde{\mathbf{x}}\|)}{V_n(1)} \right)^{\frac{2}{n}}}{1 + \frac{\mathrm{Vol}(\mathbf{H}\mathcal{X})^{\frac{2}{n}}}{2\pi e}} \right) \\
&= \frac{n}{2} \log\left(\frac{1 + \frac{1}{n}\left(\frac{\rho(\mathcal{X},\mathbf{H})\mathrm{Vol}(\mathbf{H}\mathcal{X})}{V_n(1)} \right)^{\frac{2}{n}}}{1 + \frac{\mathrm{Vol}(\mathbf{H}\mathcal{X})^{\frac{2}{n}}}{2\pi e}} \right) \\
&\stackrel{b)}{\leq} \frac{n}{2} \log\left(\frac{1}{n} 2\pi e \left(\frac{\rho(\mathcal{X},\mathbf{H})}{V_n(1)} \right)^{\frac{2}{n}} \right) \\
&\stackrel{c)}{\leq} \frac{n}{2} \log\left((\pi n)^{\frac{1}{n}} \rho(\mathcal{X},\mathbf{H})^{\frac{2}{n}} \right).
\end{aligned}
$$

Here, $a)$ is due to the fact that $\|\tilde{\mathbf{x}}\|$ is the radius of an n-dimensional ball, $b)$ follows from the inequality $\frac{1+cx}{1+x} \leq c$ for $c \geq 1$ and $x \in \mathbb{R}_+$, and $c)$ follows from using Stirling's approximation to obtain $\left(\frac{1}{V_n(1)} \right)^{\frac{2}{n}} \leq \frac{1}{2e\pi^{1-\frac{1}{n}}} n^{1+\frac{1}{n}}$. ∎

We conclude this section by characterizing the gap to the capacity when $\mathbf{H}$ is diagonal and the channel input space is the Cartesian product of n PAM constellations. In this context, $\mathsf{PAM}(N, A)$ refers to the set of $N \in \mathbb{N}$ equidistant PAM-constellation points with amplitude constraint $A \in \mathbb{R}_+$, whereas $X \sim \mathsf{PAM}(N, A)$ means that X is uniformly distributed over $\mathsf{PAM}(N, A)$ [14].

Theorem 7. *Let* $\mathbf{H} = \mathrm{diag}(h_{11}, \ldots, h_{nn}) \in \mathbb{R}^{n \times n}$ *be fixed and* $\mathbf{X} = (X_1, \ldots, X_n)$. *Then, if* $X_i \sim \mathsf{PAM}(N_i, A_i)$, $i = 1, \ldots, n$, *for some given* $\mathbf{a} = (A_1, \ldots, A_n) \in \mathbb{R}_+^n$, *it holds that*

$$\bar{C}_{\mathsf{Dual},2}(\mathrm{Box}(\mathbf{a}), \mathbf{H}) - \underline{C}_{\mathsf{OW}}(\mathrm{Box}(\mathbf{a}), \mathbf{H}) \leq c \cdot n \text{ bits}, \quad (14)$$

where $N_i := \left\lfloor 1 + \frac{2A_i|h_{ii}|}{\sqrt{2\pi e}} \right\rfloor$ *and*

$$c := \log(2) + \frac{1}{2}\log\left(\frac{\pi e}{6} \right) + \frac{1}{2}\log\left(1 + \frac{6}{\pi e} \right) \approx 1.64.$$

Moreover, if $X_i \sim \mathsf{PAM}(N_i, A)$, $i = 1, \ldots, n$, *for some given* $A \in \mathbb{R}_+$, *it holds that*

$$\bar{C}_{\mathsf{Dual},2}(\mathcal{B}_0(A), \mathbf{H}) - \underline{C}_{\mathsf{OW}}(\mathcal{B}_0(A), \mathbf{H}) \leq c \cdot n \text{ bits}, \quad (15)$$

where $N_i := \left\lfloor 1 + \frac{2A|h_{ii}|}{\sqrt{n}\sqrt{2\pi e}} \right\rfloor$.

Proof: The proof can be found in [10]. ∎

V. ARBITRARY CHANNEL MATRICES

For a MIMO channel with an arbitrary channel matrix and an average power constraint, the capacity is achieved by a singular value decomposition (SVD) of the channel matrix (i.e., $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$) and considering the equivalent channel

$$\tilde{\mathbf{Y}} = \mathbf{\Lambda}\tilde{\mathbf{X}} + \tilde{\mathbf{Z}},$$
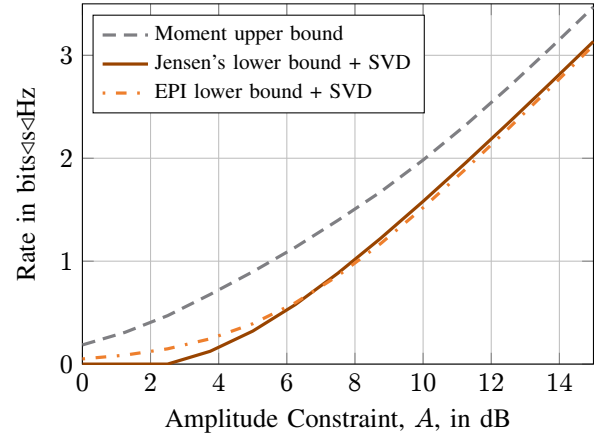

Fig. 2. Comparison of the upper bound in Theorem 1 to the lower bounds in Theorem 8 for a 3×1 MIMO system with amplitude constraints $A_1 = A_2 = A_3 = A$ (i.e., $\mathbf{a} = (A, A, A)$) and channel matrix $\mathbf{H} = (0.6557, 0.0357, 0.8491)$.

where $\tilde{\mathbf{Y}} := \mathbf{U}^T\mathbf{Y}$, $\tilde{\mathbf{X}} := \mathbf{V}^T\mathbf{X}$, and $\tilde{\mathbf{Z}} := \mathbf{U}^T\mathbf{Z}$, respectively. The following theorem provides lower bounds for this channel with amplitude constraints.

Theorem 8. (Lower Bounds with SVD Precoding) *Let* $\mathbf{H} \in \mathbb{R}^{n_r \times n_t}$ *be fixed,* $n_{\min} := \min(n_r, n_t)$, *and* $\mathcal{X} = \mathrm{Box}(\mathbf{a})$ *for some* $\mathbf{a} = (A_1, \ldots, A_{n_t}) \in \mathbb{R}_+^{n_t}$. *Furthermore, let* σ_i, $i = 1, \ldots, n_{\min}$, *be the* i-*th singular value of* $\mathbf{H}$. *Then,*

$$\underline{C}_{\mathsf{Jensen}}(\mathrm{Box}(\mathbf{a}), \mathbf{H}) = \log^+\left(\left(\frac{2}{e} \right)^{\frac{n_{\min}}{2}} \frac{1}{\psi(\mathbf{H}, \mathbf{b}^\star)} \right) \quad (16)$$

and

$$\underline{C}_{\mathsf{EPI}}(\mathrm{Box}(\mathbf{a}), \mathbf{H}) = \frac{n_{\min}}{2} \log\left(1 + \frac{|\prod_{i=1}^{n_{\min}} A_i \sigma_i|^{\frac{2}{n_{\min}}}}{2\pi e} \right),$$

where

$$\psi(\mathbf{H}, \mathbf{b}^\star) := \min_{\mathbf{b} \in \mathrm{Box}(\mathbf{a})} \prod_{i=1}^{n_{\min}} \varphi(\sigma_i B_i)$$

with $\mathbf{b} := (B_1, \ldots, B_{n_t})$ *and* $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ *given by*

$$\varphi(x) := \frac{1}{x^2}\left(e^{-x^2} - 1 + \sqrt{\pi}x\left(1 - 2Q(\sqrt{2}x) \right) \right).$$

Proof: The proof is deferred to [10]. ∎

Notice that choosing the optimal $\mathbf{b}$ for the lower bound (16) is an *amplitude allocation problem*, which is reminiscent of waterfilling in the average power constraint case. It would be interesting to study whether the bound in (16) is connected to what is known as *mercury waterfilling* [18], [19].

In Fig. 2, the lower bounds of Theorem 8 are compared to the moment upper bound of Theorem 1 for the special case of a 3×1 MIMO channel. Similarly to the example presented in Fig. 1, the EPI lower bound performs well in the low amplitude regime while Jensen's inequality lower bound performs well in the high amplitude regime.

We conclude this section by showing that for an arbitrary channel input space $\mathcal{X}$, in the large amplitude regime the capacity pre-log is given by $\min(n_r, n_t)$.

Theorem 9. *Let $\mathcal{X}$ be arbitrary and $\mathbf{H} \in \mathbb{R}^{n_r \times n_t}$ fixed. Then,*

$$\lim_{r_{\min}(\mathcal{X}) \to \infty} \frac{C(\mathcal{X}, \mathbf{H})}{\log\left(1 + \frac{2r_{\min}(\mathcal{X})}{\sqrt{2\pi e}}\right)} = \min(n_r, n_t) \ .$$

Proof: Notice that there always exists $\mathbf{a} \in \mathbb{R}_+^{n_t}$ and $c \in \mathbb{R}_+$ such that $\mathrm{Box}(\mathbf{a}) \subseteq \mathcal{X} \subset c\mathrm{Box}(\mathbf{a})$. Thus, without loss generality we can consider $\mathcal{X} = \mathrm{Box}(\mathbf{a})$, $\mathbf{a} = (A, \ldots, A)$, for sufficiently large $A \in \mathbb{R}_+$. To prove the result we therefore start with enlarging the constraint set of the bound in (5):

$$\begin{aligned}
\mathrm{Box}\big(\mathbf{H}\mathrm{Box}(\mathbf{a})\big) &\subseteq \mathcal{B}_{\mathbf{0}}\big(r_{\max}(\mathbf{H}\mathrm{Box}(\mathbf{a}))\big) \\
&\subseteq \mathcal{B}_{\mathbf{0}}\big(r_{\max}(\mathbf{H}\mathcal{B}_{\mathbf{0}}(\sqrt{n_t}A))\big) \\
&= \mathcal{B}_{\mathbf{0}}\big(r_{\max}(\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \mathcal{B}_{\mathbf{0}}(\sqrt{n_t}A))\big) \\
&= \mathcal{B}_{\mathbf{0}}\big(r_{\max}(\mathbf{U}\mathbf{\Lambda}\mathcal{B}_{\mathbf{0}}(\sqrt{n_t}A))\big) \\
&= \mathcal{B}_{\mathbf{0}}\big(r_{\max}(\mathbf{\Lambda}\mathcal{B}_{\mathbf{0}}(\sqrt{n_t}A))\big) \\
&\subseteq \mathcal{B}_{\mathbf{0}}(r) \\
&\subseteq \mathrm{Box}(\mathbf{a}') \ ,
\end{aligned}$$

where $r := \sqrt{n_t}A\sqrt{\sum_{i=1}^{n_{\min}} \sigma_i^2}$ and $\mathbf{a}' := \left(\frac{r}{\sqrt{n_{\min}}}, \ldots, \frac{r}{\sqrt{n_{\min}}}\right) \in \mathbb{R}_+^{n_{\min}}$. Thus, by using the upper bound in (5) it follows that

$$\begin{aligned}
C(\mathrm{Box}(\mathbf{a}), \mathbf{H}) &\leq \sum_{i=1}^{n_r} \log\left(1 + \frac{2A_i}{\sqrt{2\pi e}}\right) \\
&\leq n_{\min} \log\left(1 + \frac{2}{\sqrt{2\pi e}} \frac{\sqrt{n_t}A\sqrt{\sum_{i=1}^{n_{\min}}\sigma_i^2}}{\sqrt{n_{\min}}}\right) \ .
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\lim_{A \to \infty} \frac{C(\mathrm{Box}(\mathbf{a}), \mathbf{H})}{\log\left(1 + \frac{2A}{\sqrt{2\pi e}}\right)} \\
&\leq n_{\min} \lim_{A \to \infty} \frac{\log\left(1 + \frac{2}{\sqrt{2\pi e}} \frac{\sqrt{n_t}A\sqrt{\sum_{i=1}^{n_{\min}}\sigma_i^2}}{\sqrt{n_{\min}}}\right)}{\log\left(1 + \frac{2A}{\sqrt{2\pi e}}\right)} = n_{\min} \ .
\end{aligned}$$

Next, using the EPI lower bound in (8), we have that

$$\begin{aligned}
&\lim_{A \to \infty} \frac{\underline{C}_{\mathsf{EPI}}(\mathrm{Box}(\mathbf{a}), \mathbf{\Lambda})}{\log\left(1 + \frac{2A}{\sqrt{2\pi e}}\right)} \\
&= n_{\min} \lim_{A \to \infty} \frac{\frac{1}{2}\log\left(1 + \frac{A\left|\prod_{i=1}^{n_{\min}}\sigma_i\right|^{\frac{2}{n_{\min}}}}{2\pi e}\right)}{\log\left(1 + \frac{2A}{\sqrt{2\pi e}}\right)} = n_{\min} \ .
\end{aligned}$$

This concludes the proof. ∎

VI. Conclusion

In this work, we have focused on studying the capacity of MIMO systems with bounded channel input spaces. Several new upper and lower bounds have been proposed and it has been shown that the lower and upper bounds are tight in the high amplitude regime. An interesting direction for future work is to determine the exact scaling in the massive MIMO regime (i.e., $n_{\min} \to \infty$). Another interesting future direction is to study generalizations of our techniques to MIMO wireless optical channels [20].

References

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommu.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[2] J. G. Smith, "The information capacity of amplitude- and variance-constrained scalar Gaussian channels," *Inf. Control*, vol. 18, no. 3, pp. 203–219, Apr. 1971.

[3] S. Shamai (Shitz) and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1060–1071, Jul. 1995.

[4] B. Rassouli and B. Clerckx, "On the capacity of vector Gaussian channels with bounded inputs," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6884–6903, Dec. 2016.

[5] N. Sharma and S. Shamai (Shitz), "Transition points in the capacity-achieving distribution for the peak-power limited AWGN and free-space optical intensity channels," *Probl. Inf. Transm.*, vol. 46, no. 4, pp. 283–299, 2010.

[6] A. L. McKellips, "Simple tight bounds on capacity for the peak-limited discrete-time channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Chicago, IL, Jul. 2004, pp. 348–348.

[7] A. Thangaraj, G. Kramer, and G. Böcherer. (2015) Capacity bounds for discrete-time, amplitude-constrained, additive white Gaussian noise channels. [Online]. Available: https://arxiv.org/abs/1511.08742v1

[8] B. Rassouli and B. Clerckx, "An upper bound for the capacity of amplitude-constrained scalar AWGN channel," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 1924–1926, Oct. 2016.

[9] A. ElMoslimany and T. M. Duman, "On the capacity of multiple-antenna systems and parallel Gaussian channels with amplitude-limited inputs," *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 2888–2899, Jul. 2016.

[10] A. Dytso, M. Goldenbaum, S. Shamai (Shitz), and H. V. Poor. (2017) "Upper and lower bounds on the capacity of amplitude-constrained MIMO channels". [Online]. Available: https://arxiv.org/pdf/1708.09517.pdf

[11] M. Vu, "MISO capacity with per-antenna power constraint," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1268–1274, May 2011.

[12] D. Tuninetti, "On the capacity of the AWGN MIMO channel under per-antenna power constraints," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Sydney, Australia, Jun. 2014, pp. 2153–2157.

[13] S. Loyka, "The capacity of Gaussian MIMO channels under total and per-antenna power constraints," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1035–1043, Mar. 2017.

[14] A. Dytso, M. Goldenbaum, H. V. Poor, and S. Shamai (Shitz), "A generalized Ozarow-Wyner capacity bound with applications," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 1058–1062.

[15] H. S. Witsenhausen, "Some aspects of convexity useful in information theory," *IEEE Trans. Inf. Theory*, vol. 26, no. 3, pp. 265–271, May 1980.

[16] J. Sommerfeld, I. Bjelaković, and H. Boche, "On the boundedness of the support of optimal input measures for Rayleigh fading channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Toronto, Canada, Jul. 2008, pp. 1208–1212.

[17] A. Dytso, D. Tuninetti, and N. Devroye, "Interference as noise: Friend or foe?" *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3561–3596, Jun. 2016.

[18] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, Jul. 2006.

[19] F. Pérez-Cruz, M. R. Rodrigues, and S. Verdú, "MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1084, Mar. 2010.

[20] S. M. Moser, M. Mylonakis, L. Wang, and M. Wigger. (2017) Capacity results for MIMO wireless optical communication. Submitted to *IEEE Int. Symp. Inf. Theory (ISIT)*. [Online]. Available: http://perso.telecom-paristech.fr/~Ewigger/mosermylonakiswangwigger_2017-1.pdf