The Capacity-Achieving Input of Non-Coherent Rayleigh Fading Channels: Bounds on the Number of Mass Points

Antonino Favano*, Luca Barletta*, Alex Dytso†, and Gerhard Kramer**

* Politecnico di Milano, Milano, 20133, Italy. Email: {antonino.favano, luca.barletta}@polimi.it

† Qualcomm Flarion Technologies, Bridgewater, NJ 08807, USA. Email: odytso2@gmail.com

** Technical University of Munich, Munich, 80333, Germany. Email: gerhard.kramer@tum.de

***Abstract*—This work studies non-coherent Rayleigh fading channels subject to average- and peak-power constraints. The amplitude of the capacity-achieving input is known to be discrete with a finite number of points. We sharpen this result by providing upper and lower bounds for the number of amplitude levels. The upper bounds are derived using two techniques: the first counts the number of maxima of the function that characterizes the Karush-Kuhn-Tucker conditions, and the second uses an oscillation theorem. The oscillation theorem is stronger, but can be applied only if the average power constraint is not active.**

I. Introduction

Rayleigh fading models are employed to describe how multipath propagation affects wireless communication. In this research, we focus on non-coherent models, where both the receiver and transmitter lack access to channel state information (CSI). Non-coherent Rayleigh fading is often encountered in scenarios involving low-power wireless devices, short-range wireless connections, or situations where maintaining coherent detection is unfeasible or undesirable.

The main goal of this study is to explore the characteristics of input distributions that maximize capacity under average-power and peak-power constraints. This investigation helps us gain valuable insights into the fundamental limits of communication. Beyond its theoretical significance, understanding the properties of these optimal distributions can inform the development of practical coding schemes.

A. Channel Model

The input-output relationship of a Rayleigh fading channel with additive white Gaussian noise (AWGN) is

$$V = HU + W \tag{1}$$

where $H \sim \mathcal{CN}(0, \sigma_H^2)$, $W \sim \mathcal{CN}(0, \sigma_W^2)$, and the input U are mutually statistically independent. Neither the transmitter nor receiver knows the value of H, but both are assumed to know the statistics of H. We assume that the input is subject to both average-power and peak-power constraints given by

$$\mathbb{E}\left[|U|^2\right] \le \tilde{\mathsf{P}}, \tag{2}$$

$$\mathbb{P}\left[|U| \le \tilde{\mathsf{A}}\right] = 1 \Big(\text{ or equiv. } |U| \le \tilde{\mathsf{A}} \text{ a.s.}\Big) \tag{3}$$

for some $0 \le \tilde{\mathsf{P}}, \tilde{\mathsf{A}} \le \infty$. We denote the information capacity of this channel by

$$C_R(\tilde{\mathsf{P}}, \tilde{\mathsf{A}}) = \max_{P_U:\, \mathbb{E}[|U|^2] \le \tilde{\mathsf{P}},\, |U| \le \tilde{\mathsf{A}}} I(U;V). \tag{4}$$

It was shown in [1] that the maximizing random variable $U^\star$ has a discrete amplitude and uniform phase. Moreover, the authors of [1] have shown that a model with the same capacity of (1) is given by

$$Y = \left|\tilde{H}X + \tilde{W}\right|^2, \tag{5}$$

where $Y = |V|^2/\sigma_W^2$, $\tilde{H} \sim \mathcal{CN}(0,1)$, $X = |U|\sigma_H/\sigma_W$, and $\tilde{W} \sim \mathcal{CN}(0,1)$. Another channel model with the same capacity of (1) is the exponential model given by

$$Y = \frac{1}{S}T \tag{6}$$

where $T \sim \text{Exp}(1)$ (i.e., exponential with mean one) and S are independent random variables. It was demonstrated in [1] that

$$C_R(\tilde{\mathsf{P}}, \tilde{\mathsf{A}}) = C_E(\mathsf{P}, \mathsf{A}) = \max_{P_S: S \in (\frac{1}{1+\mathsf{A}^2}, 1], \mathbb{E}\left[\frac{1}{S}\right] \le 1+\mathsf{P}} I(S;Y) \tag{7}$$

where $\mathsf{P} \triangleq \frac{\sigma_H^2}{\sigma_W^2}\tilde{\mathsf{P}}$ and $\mathsf{A} \triangleq \frac{\sigma_H^2}{\sigma_W^2}\tilde{\mathsf{A}}$. Moreover, the mapping between the optimal distribution in (4) and the optimal distribution in (7) is given by

$$S^\star = \frac{1}{1 + (X^\star)^2}. \tag{8}$$

In this paper, due to ease of analytical manipulation, we mostly rely on the exponential model in (6). Then, if needed, all the results can be mapped to the original model via (8).

B. Contributions and Paper Outline

This paper is organized as follows. Section II surveys the literature and Sec. III reviews the Karush-Kuhn-Tucker (KKT) conditions and introduces the oscillation theorem. Section IV presents our main result: a characterization of the capacity-achieving distribution that gives upper and lower bounds to the number of mass points. The proof of the main theorem is in Sec. V. Section VI concludes the paper.

C. Notation

All logarithms are to the base e. Deterministic scalar quantities are denoted by lower-case letters and random variables are denoted by uppercase letters.

For a random variable X and every measurable subset $\mathcal{A} \subseteq \mathbb{R}$ we denote the probability distribution of X by $P_X(\mathcal{A}) = \mathbb{P}[X \in \mathcal{A}]$. The support set of P_X is denoted and defined as

$$\mathsf{supp}(P_X) = \{x : \text{for every open set } \mathcal{D} \ni x \text{ we have that } P_X(\mathcal{D}) > 0\}. \quad (9)$$

When X is a discrete random variable, we write $P_X(x)$ for $P_X(\{x\})$, i.e., P_X is understood as a probability mass function (pmf). The relative entropy between distributions P and Q is denoted by $D\left(P \,\|\, Q\right)$.

Given a function $f : \mathbb{C} \mapsto \mathbb{C}$ and a set $\mathcal{A} \subseteq \mathbb{C}$, we define $\mathrm{Z}\left(\mathcal{A}; f\right)$ to be the set of zeros of f in $\mathcal{A}$, that is

$$\mathrm{Z}\left(\mathcal{A}; f\right) = \{z : f(z) = 0\} \cap \mathcal{A}. \quad (10)$$

We denote the cardinality of $\mathrm{Z}\left(\mathcal{A}; f\right)$ by $\mathsf{N}\left(\mathcal{A}; f\right)$.

II. Background

The input distribution that maximizes capacity in an AWGN channel with an average power restriction is a zero-mean Gaussian distribution, as initially described by Shannon [2]. Interestingly, when the average power constraint is replaced by a peak power constraint, the optimal distribution is *discrete*. This discovery was made by Smith in [3], who introduced a novel approach linking the support of the optimal distribution to the zeros of analytical functions.

For the Rayleigh fading channel, Richters in [4] speculated that the optimal distribution with an average power constraint is also discrete, based on properties of analytical functions. This conjecture was formalized by Abou-Faycal *et al.* in [1].

The fact that the distribution maximizing capacity is discrete has been demonstrated in various related fading channels. In the study by Gursoy et al. [5], a Rician channel with an additional constraint involving the fourth moment was considered, and it was shown that the distribution achieving capacity is discrete and composed of a finite number of points. In another work by Katz and Shamai [6], noncoherent AWGN channels with an average power constraint were examined, revealing that the optimal input amplitude is discrete with *infinitely* many support points.

Similar results of either a discrete input distribution or discrete components of the optimal input distribution (e.g., amplitude) have been observed in various other channels. These include symmetric coherent vector additive Gaussian channels [7]–[10], noncoherent block-independent AWGN channels [11], and Poisson channels [12]. Efforts to extend these findings to additive channels have been made in [13]–[15], and [16]. Generalizations to multi-user channels, such as multiple access and wiretap channels, are discussed in [17] and [18]–[20], respectively. For a comprehensive summary of these discreteness results, readers are directed to [21].

The techniques for obtaining these results are usually non-constructive, relying on arguments by contradiction as first introduced in [3]. For instance, such techniques cannot bound the support size, location, and probability values of the optimal input distribution. Recently, there have been efforts to employ alternative techniques to enhance our understanding of the capacity-achieving distribution's structure. In [22], the authors improved Smith's findings in [3] for the scalar AWGN channel. The authors of [22] established definitive upper and lower bounds on the optimal input distribution's support size based on the peak-power constraint. In [22], key tools included Karlin's oscillation theorem and complex analysis for root counting. Furthermore, in [23], the authors provided bounds on the size of the support of the optimal input distribution for the Poisson noise channel. Moreover, in [23], two new techniques were introduced for upper and lower bounding support probabilities, with one relying on the strong data-processing inequality. Another method for bounding the values of probabilities can be found in [24]. Recently, in [20], these techniques have been also extended to the wiretap channel and bounds on the size of the support have been provided.

Bounds on non-coherent fading channel capacity have been explored in [25]–[27]; the interested reader is referred to [28], [29] for a literature review. The capacity scales as P in low-power regimes and as $\log\log \mathsf{P}$ in high-power regimes.

III. Channel Model and Other Preliminaries

This section presents auxiliary results and tools such as the KKT conditions, and we briefly argue why including an additional peak-power constraint is reasonable.

A. KKT Conditions

We next introduce the KKT conditions for the optimality of the input distribution P_S.

Lemma 1. *Let $P_{S^\star}$ be an input distribution of the exponential model in* (6)*, and let $f_{Y^\star}$ be the induced optimal output pdf. Then, $P_{S^\star}$ is capacity-achieving if and only if there exists $\lambda \geq 0$ such that*

$$\Xi_E(s; f_{Y^\star}) \leq 0,\ s \in \left[\frac{1}{1+\mathsf{A}^2}, 1\right], \quad (11)$$

$$\Xi_E(s; f_{Y^\star}) = 0,\ s \in \mathsf{supp}(P_{S^\star}) \quad (12)$$

where, for $s > 0$, we have

$$\Xi_E(s; f_{Y^\star}) \triangleq D\left(f_{Y|S}(\cdot|s) \,\big\|\, f_{Y^\star}\right) - C - \lambda\left(\frac{1}{s} - 1 - \mathsf{P}\right) \quad (13)$$

$$= -\mathbb{E}\left[\log(Y f_{Y^\star}(Y)) + \gamma + 1 + C + \lambda\left(Y - 1 - \mathsf{P}\right) \mid S = s\right] \quad (14)$$

Proof. See [1] and the equivalent exponential model in (6). □

Note that the Lagrange multiplier λ accounts for the average-power constraint and is equal to $\lambda = \frac{\partial}{\partial \mathsf{P}} C(\mathsf{P}, \mathsf{A})$ and it is a function of both P and A.

B. Why Peak Power is Needed

For non-coherent channels subject only to the average-power constraint, the so-called *flash signaling* is necessary to achieve the capacity slope of $\mathsf{P} \to 0$ [30]. Flash signaling is not peak-power limited and has obvious practical limitations.

For instance, in the low-power regime, the maximum amplitude support point $x^\star_{\max}$ goes to infinity as P goes to zero.

Proposition 1. *Let $P_{X^\star}$ be the capacity-achieving input distribution, and let*

$$x^\star_{\max}(\mathsf{P},\mathsf{A}) = \arg\max\{\mathsf{supp}(P_{X^\star})\} \tag{15}$$

be its largest amplitude. Then flash signaling is optimal, i.e.,

$$\lim_{\mathsf{P}\to 0}\lim_{\mathsf{A}\to\infty} x^\star_{\max}(\mathsf{P},\mathsf{A}) = \infty, \qquad \lim_{\mathsf{P}\to 0}\lim_{\mathsf{A}\to\infty} P_{X^\star}(x^\star_{\max}) = 0. \tag{16}$$

A similar behaviour is observed numerically for those P's at which a new mass point appears in $P_{X^\star}$. In particular, the new mass is observed to have infinite amplitude at the moment of appearance. Since at those P's the evolution of the capacity-achieving input distribution is abrupt, considering a peak power constraint can be beneficial in several ways. First, introducing a peak-power constraint allows to define a channel that is more reasonable from a physical standpoint. Second, the model is more general: for a large enough $\mathsf{A} < \infty$, it is virtually identical to the channel with only the average-power constraint while being mathematically more tractable. However, many of the derived results will hold for $\mathsf{A} = \infty$.

C. Oscillation Theorem

To find an upper bound on the number of points in the support of $P_{X^\star}$ or $P_{S^\star}$, we will follow the proof technique developed in [22] for the Gaussian noise channel. The key step that we borrow from [22] is the use of the *variation-diminishing* property, which is captured by the oscillation theorem of Karlin [31]. To state the oscillation theorem, we need the following definition.

Definition 1 (Sign Changes of a Function)**.** *The number of sign changes of a function $\xi : \mathcal{X} \to \mathbb{R}$ is given by*

$$\mathscr{S}(\xi) = \sup_{m\in\mathbb{N}} \left\{ \sup_{y_1<\cdots<y_m\subseteq\mathcal{X}} \mathscr{N}\{\xi(y_i)\}_{i=1}^m \right\}, \tag{17}$$

where $\mathscr{N}\{\xi(y_i)\}_{i=1}^m$ is the number of sign changes of the sequence $\{\xi(y_i)\}_{i=1}^m$ (where a zero is treated as a positive number).

The following theorem, shown in [31, Thm. 3] (see also [32, Thm. 3.1, p. 21]), will be a key step in the proof of the upper bound on the number of mass points.

Theorem 1 (Oscillation Theorem)**.** *Given domains $\mathbb{I}_1$ and $\mathbb{I}_2$, let $p\colon \mathbb{I}_1 \times \mathbb{I}_2 \to \mathbb{R}$ be a strictly totally positive kernel.*[1] *For an arbitrary y, suppose $p(\cdot,y)\colon \mathbb{I}_1 \to \mathbb{R}$ is an n-times differentiable function. Assume that μ is a regular σ-finite measure on $\mathbb{I}_2$, and let $\xi\colon \mathbb{I}_2 \to \mathbb{R}$ be a function with $\mathscr{S}(\xi) = n$. For $x \in \mathbb{I}_1$, define*

$$\Xi(x) = \int \xi(y)\, p(x,y)\, \mathrm{d}\mu(y). \tag{18}$$

If $\Xi\colon \mathbb{I}_1 \to \mathbb{R}$ is an n-times differentiable function, then either $\mathrm{N}(\mathbb{I}_1, \Xi) \le n$, or $\Xi \equiv 0$.

[1] A function $f : \mathbb{I}_1 \times \mathbb{I}_2 \to \mathbb{R}$ is said to be a strictly totally positive kernel of order n if $\det\left([f(x_i,y_j)]_{i,j=1}^m\right) > 0$ for all $1 \le m \le n$, and for all $x_1 < \cdots < x_m \in \mathbb{I}_1$, and $y_1 < \cdots < y_m \in \mathbb{I}_2$. If f is a strictly totally positive kernel of order n for all $n \in \mathbb{N}$, then f is a strictly totally positive kernel.

The above theorem says that the number of zeros of a function $\Xi(x)$, which is the output of integral transformation, is less than the number of sign changes of the function $\xi(y)$, which is the input to the integral transformation. In our setting, μ would be the Lebesgue measure and $p(x,y)$ would be $f_{Y|S}(y|s) = s\mathrm{e}^{-sy}$ for $y \ge 0$ and the integral is given by

$$\Xi(s) = \int_0^\infty \xi(y) f_{Y|S}(y|s)\mathrm{d}y, \qquad s > 0. \tag{19}$$

IV. Main Results and Discussion

The next theorem presents our main results. Let $x^\star_{\max} = \max\mathsf{supp}(P_{X^\star})$ and $\mathsf{K} = |\mathsf{supp}(P_{X^\star})|$.

Theorem 2. *Assume that $\mathsf{P} \in (0,\infty]$ and $\mathsf{A} \in (0,\infty]$, let $P_{X^\star}$ be the capacity-achieving input distribution. Then,*

(i) (A Lower Bound on the Size of the Support) *For every* $\mathsf{P_A} = \min\{\mathsf{P},\mathsf{A}^2\} \ge 0$,

$$\mathsf{K} \ge \underline{\mathsf{K}} = \max\left\{2, \left\lceil \sqrt{\left(\mathrm{e}^{-\gamma-1}\log(\mathsf{P_A}+1)\right)^2 + 1} \right\rceil \right\}. \tag{20}$$

(ii) (An Upper Bound on the Size of the Support)

$$\mathsf{K} \le \overline{\mathsf{K}} = \left\lfloor \frac{\log \frac{(2\mathsf{a}+\frac{1}{2})(\frac{1}{2}+\lambda\mathsf{a})(2\mathsf{a}+1)}{\left(\mathsf{a}-\frac{1}{2}\right)^2}}{\log\left(1+\frac{1}{2\mathsf{a}-1}\right)} + 1 \right\rfloor, \tag{21}$$

where $\mathsf{a} = 1 + (x^\star_{\max})^2$. Moreover, if the average-power constraint is not active, the bound can be tightened to:

$$\mathsf{K} \le \left\lfloor \log\left(\mathrm{e}(1+\mathsf{A}^2)+1\right) + \mathrm{e}(1+\mathsf{A}^2) + 1 \right\rfloor. \tag{22}$$

A. Numerical Results for the Optimal Input Distribution

We provide further insights on $P_{X^\star}$ and its support via some numerical results. Given the bounds in Theorem 2, in Fig. 1 we show a heatmap of $\log_2 \underline{\mathsf{K}}$ and a further upper bound on $\log_2 \overline{\mathsf{K}}$. To numerically evaluate (21) we consider upper bounds on both a and λ. Since $0 \le x^\star_{\max} \le \mathsf{A}$, we have that $1 \le \mathsf{a} \le 1 + \mathsf{A}^2$. Moreover, from [1] we have that $\lambda \le 1$. Therefore, a further upper bound on $\overline{\mathsf{K}}$ is given by

$$\overline{\mathsf{K}} \le \overline{\mathsf{K}}_2 = \left\lfloor \frac{\log\left(4(\frac{5}{2}+2\mathsf{A}^2)(\frac{3}{2}+\mathsf{A}^2)(3+2\mathsf{A}^2)\right)}{\log\left(1+\frac{1}{1+2\mathsf{A}^2}\right)} \right\rfloor. \tag{23}$$

Notice that, asymptotically we have $\underline{\mathsf{K}} \sim \log \mathsf{P_A}$ for $\mathsf{P_A} \to \infty$, while $\overline{\mathsf{K}}_2 \sim \mathsf{A}^2 \log \mathsf{A}$ for $\mathsf{A} \to \infty$.

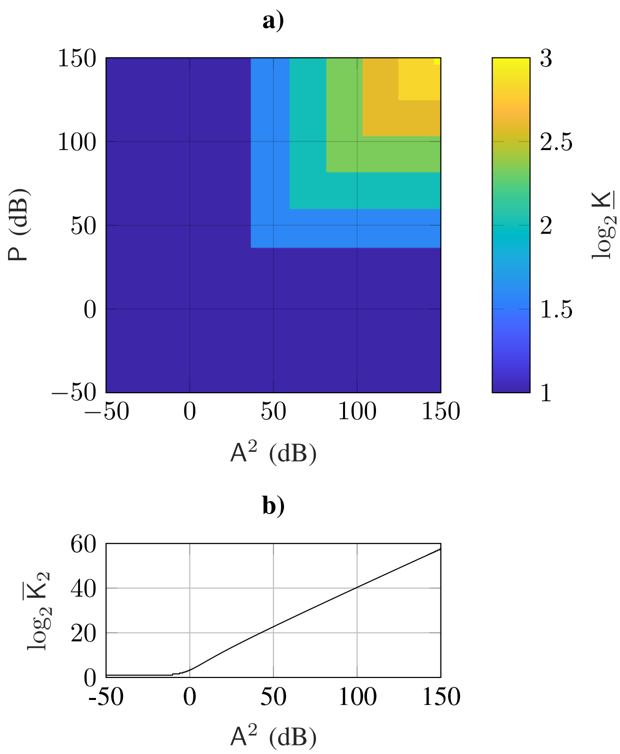

Fig. 1. For $10^{-5} \leq \mathsf{P}, \mathsf{A}^2 \leq 10^{15}$, logarithm of the: **a)** lower bound; and **b)** upper bound on the size of the support.

V. Proof of Theorem 2

A. Lower Bound on the Size of the Support

Use the fact that $X^\star$ is discrete and write

$$C(\mathsf{P},\mathsf{A}) = I(X^\star;Y^\star) = H(X^\star) - H(X^\star \,|\, Y^\star) \tag{24}$$
$$\leq H(X^\star) \leq \log|\mathsf{supp}(P_{X^\star})|. \tag{25}$$

Finally, use the capacity lower bound derived in Appendix I to get the result in (20).

B. Upper Bound on the Size of the Support

The main observation we rely on to count the number of mass points of $P_{S^\star}$ is that

$$\mathsf{supp}(P_{S^\star}) \subseteq \mathrm{Z}\left((0,1];\Xi_E(\cdot;f_{Y^\star})\right), \tag{26}$$

or

$$|\mathsf{supp}(P_{S^\star})| \leq \mathrm{N}\left((0,1];\,\Xi_E(\cdot;f_{Y^\star})\right) \tag{27}$$

which is a consequence of the KKT conditions.

To evaluate (27) we will follow two approaches: the first is based on direct counting the number of zeros of the function Ξ_E, while the second is based on the application of the oscillation theorem to the function Ξ_E. In both cases, we need to upper-bound the number of zeros of a function: A key tool is the following result.

Lemma 2 (Tijdeman's Number of Zeros Lemma [33])**.** *Let R, v, t be positive numbers such that $v > 1$. For the complex valued function $f \neq 0$ which is analytic on $|z| \leq (vt+v+t)\mathsf{R}$, its number of zeros $\mathrm{N}(\mathcal{D}_\mathsf{R}; f)$ within the disk $\mathcal{D}_\mathsf{R} = \{z\colon |z| \leq \mathsf{R}\}$ satisfies*

$$\mathrm{N}(\mathcal{D}_\mathsf{R}; f) \leq \frac{1}{\log v} \log \frac{\max_{|z|\leq(vt+v+t)\mathsf{R}} |f(z)|}{\max_{|z|\leq t\mathsf{R}} |f(z)|}. \tag{28}$$

1) Direct Counting of the Zeros of Function Ξ_E: Note that we can count the zeros of the function $s \mapsto \Xi_E(s;f_{Y^\star})$ in the interval $[\mathsf{r},1]$, where $\mathsf{r} = s^\star_{\min}$. We can upper-bound the number of support points as follows:

$$|\mathsf{supp}(P_{S^\star})| \leq \mathrm{N}\left([\mathsf{r},1];\,\Xi_E(\cdot;f_{Y^\star})\right) \tag{29}$$
$$\leq \mathrm{N}\left([\mathsf{r},1];\,\Xi_E'(\cdot;f_{Y^\star})\right) + 1 \tag{30}$$
$$= \mathrm{N}\left([\mathsf{r},1];\, \frac{1}{s}\mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] \mid S = s\right] + \frac{\lambda}{s^2}\right) + 1 \tag{31}$$
$$= \mathrm{N}\left([\mathsf{r},1];\, \mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] + \lambda Y \mid S = s\right]\right) + 1 \tag{32}$$
$$= \mathrm{N}\left(\left[-\frac{1-\mathsf{r}}{2}, \frac{1-\mathsf{r}}{2}\right];\; \mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] + \lambda Y \,\middle|\, S = s + \frac{1+\mathsf{r}}{2}\right]\right) + 1, \tag{33}$$

where (30) is a direct consequence of Rolle's theorem (see, [22, Lemma 3]); (31) holds by Proposition 3; (32) holds because multiplying the function by s does not change the number of zeros and because $\mathbb{E}\left[Y \mid S = s\right] = \frac{1}{s}$; and the last step follows from the change of variable $s \to s + \frac{1+\mathsf{r}}{2}$, such that the interval is centered at $s = 0$. Now consider the complex analytic extension

$$z \mapsto \breve{g}(z) = \mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] + \lambda Y \,\middle|\, S = z + \frac{1+\mathsf{r}}{2}\right], \tag{34}$$

of the real function $s \mapsto g(s) = \mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] + \lambda Y \,\middle|\, S = s + \frac{1+\mathsf{r}}{2}\right]$, which is analytic on $|z| < \frac{1+\mathsf{r}}{2}$.

In order to apply Tijdeman's lemma, we need to compute upper and lower bounds on the maximum value of $\breve{g}$ in a disk of radius B.

Lemma 3. *For $\mathsf{B} < \frac{1+\mathsf{r}}{2}$,*

$$\max_{|z|\leq\mathsf{B}} |\breve{g}(z)| \leq \frac{\left(\frac{1+\mathsf{r}}{2} + \mathsf{B}\right)\left(\frac{1+\mathsf{r}}{2} - \mathsf{B} + 1 + \lambda\right)}{\left(\frac{1+\mathsf{r}}{2} - \mathsf{B}\right)^2}, \tag{35}$$

and for $\frac{1}{2} \leq \mathsf{B} < \frac{1+\mathsf{r}}{2}$

$$\max_{|z|\leq\mathsf{B}} |\breve{g}(z)| \geq 1 + \frac{\lambda}{1+\frac{\mathsf{r}}{2}} - \frac{1}{1+\frac{\mathsf{r}}{2}} \geq \frac{1}{\frac{2}{\mathsf{r}}+1}. \tag{36}$$

Proof. For the upper bound, for $\mathsf{B} < \frac{1+\mathsf{r}}{2}$ we have

$$\max_{|z|\leq\mathsf{B}} |\breve{g}(z)| \leq \max_{|z|\leq\mathsf{B}} \tag{37}$$
$$\int_0^\infty \left|(1 - y\mathbb{E}\left[S^\star \mid Y = y\right] + \lambda y)\left(z + \frac{1+\mathsf{r}}{2}\right) \mathrm{e}^{-zy - \frac{1+\mathsf{r}}{2}y}\right| \mathrm{d}y \tag{38}$$

$$\leq \int_0^{\infty}(1+y\mathbb{E}[S^\star \mid Y=y]+\lambda y)\left(\mathsf{B}+\frac{1+\mathsf{r}}{2}\right)\mathrm{e}^{\mathsf{B}y-\frac{1+\mathsf{r}}{2}y}\mathrm{d}y \tag{39}$$

$$\leq \int_0^{\infty}(1+y(1+\lambda))\left(\mathsf{B}+\frac{1+\mathsf{r}}{2}\right)\mathrm{e}^{\mathsf{B}y-\frac{1+\mathsf{r}}{2}y}\mathrm{d}y \tag{40}$$

$$= \frac{(\frac{1+\mathsf{r}}{2}+\mathsf{B})(\frac{1+\mathsf{r}}{2}-\mathsf{B}+1+\lambda)}{\left(\frac{1+\mathsf{r}}{2}-\mathsf{B}\right)^2}, \tag{41}$$

where (38) follows from Jensen's inequality; (39) follows from triangle inequality, from $|\mathrm{e}^{-zy}| = \mathrm{e}^{-\Re\{z\}y} \leq \mathrm{e}^{\mathsf{B}y}$, and from $|z| \leq \mathsf{B}$; and (40) follows from $S^\star \leq 1$.

For the lower bound, for $\frac{1}{2} \leq \mathsf{B} < \frac{1+\mathsf{r}}{2}$, we have

$$\max_{|z|\leq \mathsf{B}} |\breve{g}(z)| \geq \left|\breve{g}\left(\frac{1}{2}\right)\right| \tag{42}$$

$$= \left|\int_0^{\infty}(1-y\mathbb{E}[S^\star \mid Y=y]+\lambda y)\left(1+\frac{\mathsf{r}}{2}\right)\mathrm{e}^{-\left(1+\frac{\mathsf{r}}{2}\right)y}\mathrm{d}y\right| \tag{43}$$

$$= \left|1+\frac{\lambda}{1+\frac{\mathsf{r}}{2}}-\int_0^{\infty}y\mathbb{E}[S^\star \mid Y=y]\left(1+\frac{\mathsf{r}}{2}\right)\mathrm{e}^{-\left(1+\frac{\mathsf{r}}{2}\right)y}\mathrm{d}y\right|. \tag{44}$$

Now, by using $S^\star \leq 1$, note that

$$\int_0^{\infty}y\mathbb{E}[S^\star \mid Y=y]\left(1+\frac{\mathsf{r}}{2}\right)\mathrm{e}^{-\left(1+\frac{\mathsf{r}}{2}\right)y}\mathrm{d}y \tag{45}$$

$$\leq \int_0^{\infty}y\left(1+\frac{\mathsf{r}}{2}\right)\mathrm{e}^{-\left(1+\frac{\mathsf{r}}{2}\right)y}\mathrm{d}y = \frac{1}{1+\frac{\mathsf{r}}{2}} \tag{46}$$

which is strictly smaller than 1; Therefore, we have that

$$\max_{|z|\leq \mathsf{B}} |\breve{g}(z)| \geq 1+\frac{\lambda}{1+\frac{\mathsf{r}}{2}}-\frac{1}{1+\frac{\mathsf{r}}{2}} \geq \frac{1}{\frac{2}{\mathsf{r}}+1}, \tag{47}$$

for $\frac{1}{2} \leq \mathsf{B} < \frac{1+\mathsf{r}}{2}$, where in the last step we used $\lambda \geq 0$. ☐

We can now apply Tijdeman's lemma with $t = \frac{1}{1-\mathsf{r}}$:

$$\mathrm{N}\left(\mathcal{D}_{\frac{1-\mathsf{r}}{2}};\, \breve{g}\right) \tag{48}$$

$$\leq \min_{1<v<\frac{1+\mathsf{r}}{1-\mathsf{r}}}\left\{\frac{\log\frac{(1+\frac{\mathsf{r}}{2}+v(1-\frac{\mathsf{r}}{2}))(\frac{\mathsf{r}}{2}-v(1-\frac{\mathsf{r}}{2})+1+\lambda)}{\left(\frac{\mathsf{r}}{2}-v(1-\frac{\mathsf{r}}{2})\right)^2}-\log\frac{1}{\frac{2}{\mathsf{r}}+1}}{\log v}\right\} \tag{49}$$

$$\leq \frac{\log\frac{(2+\frac{\mathsf{r}}{2})(\frac{\mathsf{r}}{2}+\lambda)(\frac{2}{\mathsf{r}}+1)}{\left(1-\frac{\mathsf{r}}{2}\right)^2}}{\log\frac{1}{1-\frac{\mathsf{r}}{2}}} = \frac{\log\frac{(\frac{2}{\mathsf{r}}+\frac{1}{2})(\frac{1}{2}+\frac{\lambda}{\mathsf{r}})(\frac{2}{\mathsf{r}}+1)}{\left(\frac{1}{\mathsf{r}}-\frac{1}{2}\right)^2}}{\log\left(1+\frac{1}{\frac{2}{\mathsf{r}}-1}\right)}, \tag{50}$$

where in (50) we have chosen $v = \frac{1}{1-\frac{\mathsf{r}}{2}} = 1+\frac{1}{\frac{2}{\mathsf{r}}-1} < \frac{1+\mathsf{r}}{1-\mathsf{r}}$ for $\mathsf{r} > 0$. Now, by letting $\mathsf{a} := \frac{1}{\mathsf{r}}$ and putting everything together, we have

$$|\mathsf{supp}(P_{S^\star})| \leq \mathrm{N}\left(\left[-\frac{1-\mathsf{r}}{2},\frac{1-\mathsf{r}}{2}\right];\, g\right)+1 \tag{51}$$

$$\leq \mathrm{N}\left(\mathcal{D}_{\frac{1-\mathsf{r}}{2}};\, \breve{g}\right)+1 \tag{52}$$

$$\leq \frac{\log\frac{(2\mathsf{a}+\frac{1}{2})(\frac{1}{2}+\lambda\mathsf{a})(2\mathsf{a}+1)}{\left(\mathsf{a}-\frac{1}{2}\right)^2}}{\log\left(1+\frac{1}{2\mathsf{a}-1}\right)}+1, \tag{53}$$

whose asymptotic growth is of the order of $\mathsf{a}\log(\mathsf{a})$ as $\mathsf{a} \to \infty$. This concludes the proof.

2) Application of the Oscillation Theorem: For the case $\lambda = 0$ we can get a tighter upper bound on $|\mathsf{supp}(P_{S^\star})|$ by using the oscillation theorem. First of all, we can write

$$|\mathsf{supp}(P_{S^\star})| \leq \mathrm{N}\left((0,1];\, \Xi_E(\cdot; f_{Y^\star})\right) \tag{54}$$

$$= \mathrm{N}\left((0,1];\, \mathbb{E}[\log(Yf_{Y^\star}(Y))+\gamma+1+C \mid S=s]\right) \tag{55}$$

$$\leq \mathrm{N}\left((0,\infty);\, -\log(yf_{Y^\star}(y))-\gamma-1-C\right) \tag{56}$$

$$\leq \mathrm{N}\left((0,\infty);\, \mathbb{E}[S^\star \mid Y=y]-y^{-1}\right)+1 \tag{57}$$

$$= \mathrm{N}\left((0,\infty);\, y\mathbb{E}[S^\star \mid Y=y]-1\right)+1, \tag{58}$$

where in (55) we used (14); (56) follows from the oscillation theorem (Th. 1); (57) follows from Rolle's theorem [22, Lemma 3] and from Lemma 5; and the last equality holds because multiplying the function by y does not change the number of zeros. Next, note that

$$y\mathbb{E}[S^\star \mid Y=y]-1 \geq ys^\star_{\min}-1 \tag{59}$$

which is always positive for $y > \mathsf{R} := \frac{1}{s^\star_{\min}}$. Similarly,

$$y\mathbb{E}[S^\star \mid Y=y]-1 \leq y-1 \tag{60}$$

which is always negative for $y < 1$. Hence, we have

$$|\mathsf{supp}(P_{S^\star})| \leq \mathrm{N}\left([1,\mathsf{R}];\, y\mathbb{E}[S^\star \mid Y=y]-1\right)+1 \tag{61}$$

$$= \mathrm{N}\left([1,\mathsf{R}];\, y\mathbb{E}\left[(S^\star)^2\mathrm{e}^{-S^\star y}\right]-\mathbb{E}\left[S^\star\mathrm{e}^{-S^\star y}\right]\right)+1 \tag{62}$$

$$\leq \mathrm{N}\left(\mathcal{D}_{\mathsf{R}};\, \breve{g}(z)\right)+1, \tag{63}$$

where (62) follows from $\mathbb{E}[S^\star \mid Y=y] = \frac{\mathbb{E}\left[(S^\star)^2\mathrm{e}^{-S^\star y}\right]}{\mathbb{E}\left[S^\star\mathrm{e}^{-S^\star y}\right]}$ and from multiplying the function by $\mathbb{E}\left[S^\star\mathrm{e}^{-S^\star y}\right]$; and where

$$\breve{g}(z) = z\mathbb{E}\left[(S^\star)^2\mathrm{e}^{-S^\star z}\right]-\mathbb{E}\left[S^\star\mathrm{e}^{-S^\star z}\right] \tag{64}$$

is the complex analytic extension of $g(y) = y\mathbb{E}\left[(S^\star)^2\mathrm{e}^{-S^\star y}\right]-\mathbb{E}\left[S^\star\mathrm{e}^{-S^\star y}\right]$, which is analytic on $z \in \mathbb{C}$. In order to apply Tijdeman's lemma, we need to compute upper and lower bounds on the maximum value of $\breve{g}$ in a disk of radius B. For the upper bound, we have

$$\max_{|z|\leq \mathsf{B}} |\breve{g}(z)| = \max_{|z|\leq \mathsf{B}}\left|\mathbb{E}\left[((S^\star)^2z-S^\star)\mathrm{e}^{-S^\star z}\right]\right| \tag{65}$$

$$\leq \max_{|z|\leq \mathsf{B}}\mathbb{E}\left[|(S^\star)^2z-S^\star|\mathrm{e}^{-S^\star\Re\{z\}}\right] \tag{66}$$

$$\leq \mathbb{E}\left[((S^\star)^2\mathsf{B}+S^\star)\mathrm{e}^{S^\star\mathsf{B}}\right]. \tag{67}$$

where (66) follows from Jensen's inequality; and (67) follows from triangle inequality and from $|z| \leq \mathsf{B}$. For the lower bound, we have

$$\max_{|z|\leq \mathsf{B}} |\breve{g}(z)| \geq |\breve{g}(-\mathsf{B})| = \left|\mathbb{E}\left[(-(S^\star)^2\mathsf{B}-S^\star)\mathrm{e}^{S^\star\mathsf{B}}\right]\right| \tag{68}$$

$$= \mathbb{E}\left[((S^\star)^2\mathsf{B}+S^\star)\mathrm{e}^{S^\star\mathsf{B}}\right]. \tag{69}$$

By comparing (67) and (69), we get that

$$\max_{|z|\leq \mathsf{B}} |\breve{g}(z)| = \mathbb{E}\left[((S^\star)^2\mathsf{B}+S^\star)\mathrm{e}^{S^\star\mathsf{B}}\right]. \tag{70}$$

We can now apply Tijdeman's lemma: Let us first compute

$$\begin{aligned}
&\frac{\max_{|z|\leq(vt+v+t)\mathsf{R}}|\breve{g}(z)|}{\max_{|z|\leq t\mathsf{R}}|\breve{g}(z)|}\\
&=\frac{\mathbb{E}\left[((S^\star)^2(vt+v+t)\mathsf{R}+S^\star)\mathrm{e}^{S^\star(vt+v+t)\mathsf{R}}\right]}{\mathbb{E}\left[((S^\star)^2t\mathsf{R}+S^\star)\mathrm{e}^{S^\star t\mathsf{R}}\right]} &(71)\\
&\leq\max_{s\in[\frac{1}{1+\mathsf{A}^2},1]}\frac{(s^2(vt+v+t)\mathsf{R}+s)\mathrm{e}^{s(vt+v+t)\mathsf{R}}}{(s^2t\mathsf{R}+s)\mathrm{e}^{st\mathsf{R}}} &(72)\\
&=\max_{s\in[\frac{1}{1+\mathsf{A}^2},1]}\frac{(s(vt+v+t)\mathsf{R}+1)\mathrm{e}^{sv(t+1)\mathsf{R}}}{(st\mathsf{R}+1)} &(73)\\
&=\frac{((vt+v+t)\mathsf{R}+1)\mathrm{e}^{v(t+1)\mathsf{R}}}{(t\mathsf{R}+1)}, &(74)
\end{aligned}$$

where (71) follows from (70); (72) follows from Lemma 7 and from the peak-power constraint $s\geq\frac{1}{1+\mathsf{A}^2}$; and the final step gives the maximum value obtained for $s=1$. Then, by choosing $t=0$, we have

$$\begin{aligned}
\mathrm{N}\left(\mathcal{D}_{\mathsf{R}};\,\breve{g}(z)\right)&\leq\min_{v>1,\,t\geq0}\left\{\frac{\log\frac{\max_{|z|\leq(vt+v+t)\mathsf{R}}|\breve{g}(z)|}{\max_{|z|\leq t\mathsf{R}}|\breve{g}(z)|}}{\log v}\right\} &(75)\\
&\leq\min_{v>1}\left\{\frac{\log(v\mathsf{R}+1)+v\mathsf{R}}{\log v}\right\} &(76)\\
&\leq\log\left(\mathrm{e}\mathsf{R}+1\right)+\mathrm{e}\mathsf{R} &(77)\\
&\leq\log\left(\mathrm{e}(1+\mathsf{A}^2)+1\right)+\mathrm{e}(1+\mathsf{A}^2), &(78)
\end{aligned}$$

where in (77) we have used $v=\mathrm{e}$; and in the last step we have used $\mathsf{R}=\frac{1}{s^\star_{\min}}\leq1+\mathsf{A}^2$.

Putting everything together, we obtain

$$|\mathsf{supp}(P_{S^\star})|\leq\log\left(\mathrm{e}(1+\mathsf{A}^2)+1\right)+\mathrm{e}(1+\mathsf{A}^2)+1. \quad (79)$$

VI. Conclusion

This work investigated a non-coherent Rayleigh channel and presented new results on the structure of the capacity-achieving input distribution based on the Karush-Kuhn-Tucker (KKT) conditions. In addition to an average-power constraint, we considered a peak-power constraint, making the model more general and practical.

We provided bounds on the number of optimal amplitude levels for the capacity-achieving input distribution. Specifically, we obtained an upper bound on the optimal amplitude levels by upper-bounding the number of maxima of the function characterizing the KKT conditions. In the limit of a large peak-power constraint A^2, the upper bound is of the order of $\mathsf{A}^2\log(\mathsf{A})$. Furthermore, utilizing the oscillation theorem, we demonstrated how to asymptotically refine this bound to the order of A^2 in cases where the average-power constraint is inactive.

However, an open question remains regarding the correct asymptotic growth of the optimal number of amplitude levels, given that the asymptotic order of the proposed lower bound is $\log(\mathsf{A})$.

References

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

[2] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

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

[4] J. S. Richters, "Communication over fading dispersive channels." 1967.

[5] M. C. Gursoy, H. V. Poor, and S. Verdú, "The noncoherent rician fading channel-part i: structure of the capacity-achieving input," *IEEE Trans. on Wirel. Commun.*, vol. 4, no. 5, pp. 2193–2206, 2005.

[6] M. Katz and S. Shamai, "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2257–2270, 2004.

[7] S. Shamai 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, 1995.

[8] 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, 2016.

[9] A. Dytso, M. Al, H. V. Poor, and S. Shamai, "On the capacity of the peak power constrained vector Gaussian channel: An estimation theoretic perspective," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3907–3921, 2019.

[10] J. Eisen, R. R. Mazumdar, and P. Mitran, "Capacity-achieving input distributions of additive vector Gaussian noise channels: Even-moment constraints and unbounded or compact support," *Entropy*, vol. 25, no. 8, 2023.

[11] R. Nuriyev and A. Anastasopoulos, "Capacity characterization for the noncoherent block-independent AWGN channel," in *IEEE International Symposium on Information Theory*, 2003, pp. 373–373.

[12] S. Shamai, "Capacity of a pulse amplitude modulated direct detection photon channel," *IEE Proceedings I (Communications, Speech and Vision)*, vol. 137, no. 6, pp. 424–430, 1990.

[13] A. Das, "Capacity-achieving distributions for non-Gaussian additive noise channels," in *IEEE Intern. Symp. Inf. Theory*. IEEE, 2000.

[14] A. Tchamkerten, "On the discreteness of capacity-achieving distributions," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2773–2778, 2004.

[15] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2073–2088, 2005.

[16] J. Fahs and I. Abou-Faycal, "On properties of the support of capacity-achieving distributions for additive noise channel models with input cost constraints," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1178–1198, 2017.

[17] B. Mamandipoor, K. Moshksar, and A. K. Khandani, "On the sum-capacity of gaussian mac with peak constraint," in *IEEE Intern. Symp. on Inf. Theory*. IEEE, 2012, pp. 26–30.

[18] O. Ozel, E. Ekrem, and S. Ulukus, "Gaussian wiretap channel with amplitude and variance constraints," *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5553–5563, 2015.

[19] A. Dytso, M. Egan, S. M. Perlaza, H. V. Poor, and S. Shamai, "Optimal inputs for some classes of degraded wiretap channels," in *2018 IEEE Information Theory Workshop (ITW)*. IEEE, 2018, pp. 1–5.

[20] A. Favano, L. Barletta, and A. Dytso, "Amplitude constrained vector Gaussian wiretap channel: Properties of the secrecy-capacity-achieving input distribution," *Entropy*, vol. 25, no. 5, 2023.

[21] A. Dytso, M. Goldenbaum, H. V. Poor, and S. Shamai, "When are discrete channel inputs optimal? — optimization techniques and some new results," in *52nd Annual Conf. on Inf. Sciences and Systems*, 2018.

[22] A. Dytso, S. Yagli, H. V. Poor, and S. Shamai, "The capacity achieving distribution for the amplitude constrained additive Gaussian channel: An upper bound on the number of mass points," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2006–2022, 2020.

[23] A. Dytso, L. Barletta, and S. Shamai, "Properties of the support of the capacity-achieving distribution of the amplitude-constrained Poisson noise channel," *IEEE Trans. Inf. Theory*, vol. 67, no. 11, pp. 7050–7066, 2021.

[24] N. Shulman and M. Feder, "The uniform distribution as a universal prior," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1356–1362, 2004.

[25] G. Taricco and M. Elia, "Capacity of fading channel with no side information," *Electronics Letters*, vol. 33, no. 16, pp. 1368–1370, 1997.
[26] A. Lapidoth and S. Shamai, "Fading channels: how perfect need "perfect side information" be?" *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, 2002.
[27] 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, 2003.
[28] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 117–132, 2013.
[29] G. Kramer, "Information rates for channels with fading, side information and adaptive codewords," *Entropy*, vol. 25, no. 5, 2023.
[30] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, 2002.
[31] S. Karlin, "Pólya type distributions, ii," *The Annals of Mathematical Statistics*, vol. 28, no. 2, pp. 281–308, 1957.
[32] ——, *Total Positivity*. Stanford University Press, 1968, vol. I.
[33] R. Tijdeman, "On the number of zeros of general exponential polynomials," in *Indagationes Mathematicae (Proceedings)*, vol. 74. North-Holland, 1971, pp. 1–7.
[34] A. Dytso, M. Cardone, and I. Zieder, "Meta derivative identity for the conditional expectation," *IEEE Trans. Inf. Theory*, 2023.

APPENDIX I
LOWER BOUND ON CAPACITY

Proposition 2. *Fix some* $\mathsf{P} \in (0, \infty]$ *and* $\mathsf{A} \in (0, \infty]$. *Given* $\mathsf{P_A} = \min(\mathsf{P}, \mathsf{A}^2)$*, a lower bound on the channel capacity is given by*

$$C(\mathsf{P}, \mathsf{A}) \geq \underline{C}(\mathsf{P_A}) = \log\left(\sqrt{\left(\mathrm{e}^{-\gamma-1}\log(\mathsf{P_A}+1)\right)^2 + 1}\right) \tag{80}$$

$$\geq \max\left\{0, \log(\log(1+\mathsf{P_A}))) - \gamma - 1\right\}, \tag{81}$$

Proof. Let us pick the input distribution such that $\log\frac{1}{S} \sim \mathcal{U}[0, \log(\mathsf{P_A}+1)]$, so that the average power constraint

$$\mathbb{E}\left[\frac{1}{S}\right] = \mathbb{E}[\mathrm{e}^{\log\frac{1}{S}}] = \frac{\mathrm{e}^{\log(\mathsf{P_A}+1)} - \mathrm{e}^0}{\log(\mathsf{P_A}+1)} = \frac{\mathsf{P_A}}{\log(\mathsf{P_A}+1)} \leq \mathsf{P_A} + 1 \leq \mathsf{P} + 1 \tag{82}$$

is satisfied for $\mathsf{P} > 0$. Next, we lower bound the output entropy by using the entropy power inequality:

$$h(L) = h\left(\log\frac{1}{S} - Z\right) \tag{83}$$

$$\geq \log\left(\sqrt{\mathrm{e}^{2h\left(\log\frac{1}{S}\right)} + \mathrm{e}^{2h(Z)}}\right) \tag{84}$$

$$= \log\left(\sqrt{\mathrm{e}^{2\log\log(\mathsf{P_A}+1)} + \mathrm{e}^{2(\gamma+1)}}\right) \tag{85}$$

$$= \log\left(\sqrt{(\log(\mathsf{P_A}+1))^2 + \mathrm{e}^{2(\gamma+1)}}\right) \tag{86}$$

$$\geq \max\left\{\log(\log(\mathsf{P_A}+1))), \gamma+1\right\}. \tag{87}$$

By subtracting the entropy $h(Z) = \gamma + 1$, we get

$$C(\mathsf{P}, \mathsf{A}) \geq \max\left\{\log(\log(\mathsf{P_A}+1))) - \gamma - 1, 0\right\} \tag{88}$$

$$= [\log(\log(\mathsf{P_A}+1))) - \gamma - 1]^+ \tag{89}$$

$$:= \underline{C}(\mathsf{P_A}). \tag{90}$$

This concludes the proof. □

APPENDIX II
DERIVATIVES OF $\Xi_E(\cdot; f_{Y^\star})$

To find the derivatives of the function $\Xi_E(\cdot; f_{Y^\star})$, we will need the following auxiliary results.

Lemma 4. *Let us consider the exponential channel model characterized by a pdf*

$$f_{Y|S}(y|s) = s\exp(-sy), \qquad y \geq 0. \tag{91}$$

We have

$$\frac{\mathrm{d}}{\mathrm{d}s} f_{Y|S}(y|s) = -\frac{1-sy}{s^2}\frac{\mathrm{d}}{\mathrm{d}y} f_{Y|S}(y|s) \tag{92}$$

$$= \frac{1-sy}{s} f_{Y|S}(y|s). \tag{93}$$

Proof. First of all, compute

$$\frac{\mathrm{d}}{\mathrm{d}s} f_{Y|S}(y|s) = \frac{\mathrm{d}}{\mathrm{d}s}\exp(-sy + \log s) \tag{94}$$

$$= \frac{1-sy}{s}\exp(-sy + \log s) \tag{95}$$

$$= \frac{1-sy}{s} f_{Y|S}(y|s). \tag{96}$$

Then, compute

$$\frac{\mathrm{d}}{\mathrm{d}y} f_{Y|S}(y|s) = \frac{\mathrm{d}}{\mathrm{d}y}\exp(-sy + \log s) \tag{97}$$

$$= -s\exp(-sy + \log s) \tag{98}$$

$$= -s f_{Y|S}(y|s) \tag{99}$$

$$= -\frac{s^2}{1-sy}\frac{\mathrm{d}}{\mathrm{d}s} f_{Y|S}(y|s) \tag{100}$$

□

Lemma 5. *Let us consider the exponential channel model of* (91) *and an output distribution* f_Y*. Then, we have*

$$\frac{\mathrm{d}}{\mathrm{d}y}\log f_Y(y) = -\mathbb{E}\left[S \mid Y=y\right], \tag{101}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2}\log f_Y(y) = \mathsf{Var}\left[S \mid Y=y\right]. \tag{102}$$

Proof. The channel transition density $f_{Y|S}$ belongs to the family of exponential distributions. Indeed,

$$f_{Y|S}(y|s) = \exp\left(-sy + \log s\right) \tag{103}$$

$$= h(y)\mathrm{e}^{sT(y)-\phi(s)} \tag{104}$$

where $h(y)=1$, $T(y)=-y$, and $\phi(s)=-\log s$. By introducing the operator $D_y := \frac{1}{T'(y)}\frac{\mathrm{d}}{\mathrm{d}y} = -\frac{\mathrm{d}}{\mathrm{d}y}$, we can apply the result of [34, Remark 7] to write

$$\mathbb{E}\left[S \mid Y=y\right] = D_y \log\frac{f_Y(y)}{h(y)} = -\frac{\mathrm{d}}{\mathrm{d}y}\log f_Y(y). \tag{105}$$

As for the second derivative, write

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2}\log f_Y(y) = -\frac{\mathrm{d}}{\mathrm{d}y}\mathbb{E}\left[S \mid Y=y\right] \tag{106}$$

$$= D_y\mathbb{E}\left[S \mid Y=y\right] \tag{107}$$

$$= \mathsf{Var}\left[S \mid Y=y\right] \tag{108}$$

where the last step is due to [34, Prop. 4]. □

Lemma 6. *Given* $s>0$*, suppose that* $\lim_{y\to\infty} f(y)e^{-sy}=0$*. Then,*

$$\mathbb{E}\left[f(Y) \mid S=s\right] = f(0) + \frac{1}{s}\mathbb{E}\left[f'(Y) \mid S=s\right], \tag{109}$$

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\left[f(Y) \mid S=s\right] = -\frac{1}{s}\mathbb{E}\left[Yf'(Y) \mid S=s\right], \tag{110}$$

$$\frac{\mathrm{d}}{\mathrm{d}s}\frac{1}{s}\mathbb{E}\left[f(Y) \mid S=s\right] = -\frac{1}{s}\mathbb{E}\left[Yf(Y) \mid S=s\right]. \tag{111}$$

Proof. For the first result, by using integration by parts we have that

$$\mathbb{E}\left[f(Y) \mid S=s\right] = \int_0^\infty f(y)se^{-sy}\mathrm{d}y \tag{112}$$

$$= \left(-e^{-sy}f(y)\right)\Big|_0^\infty - \int_0^\infty f'(y)\left(-e^{-sy}\right)\mathrm{d}y \tag{113}$$

$$= f(0) + \frac{1}{s}\int_0^\infty se^{-sy}f'(y)\mathrm{d}y \tag{114}$$

$$= f(0) + \frac{1}{s}\mathbb{E}\left[f'(Y) \mid S=s\right], \tag{115}$$

where in (114) we used that $\lim_{y\to\infty} f(y)e^{-sy}=0$. To prove the second identity, by using again integration by parts, we have that

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\left[f(Y) \mid S=s\right] = \frac{\mathrm{d}}{\mathrm{d}s}\int_0^\infty f(y)se^{-sy}\mathrm{d}y \tag{116}$$

$$= \int_0^\infty f(y)e^{-sy}\mathrm{d}y - \int_0^\infty f(y)yse^{-sy}\mathrm{d}y \tag{117}$$

$$= \left(ye^{-sy}f(y)\right)\Big|_0^\infty - \int_0^\infty y\left(-se^{-sy}f(y)+e^{-sy}f'(y)\right)\mathrm{d}y - \int_0^\infty sye^{-sy}f(y)\mathrm{d}y \tag{118}$$

$$= -\int_0^\infty ye^{-sy}f'(y)\mathrm{d}y \tag{119}$$

$$= -\frac{1}{s}\mathbb{E}\left[Yf'(Y) \mid S = s\right], \tag{120}$$

where in (119) we used that $\lim_{y\to\infty} f(y)e^{-sy} = 0$. To prove the third identity, write

$$\frac{\mathrm{d}}{\mathrm{d}s}\frac{1}{s}\mathbb{E}\left[f(Y) \mid S = s\right] = \frac{\mathrm{d}}{\mathrm{d}s}\int_0^\infty \mathrm{e}^{-sy}f(y)\mathrm{d}y \tag{121}$$

$$= -\int_0^\infty y\mathrm{e}^{-sy}f(y)\mathrm{d}y \tag{122}$$

$$= -\mathbb{E}\left[Yf(Y) \mid S = s\right]. \tag{123}$$

□

Proposition 3. *The first and second derivatives of function Ξ_E given in* (14) *are as follows*

$$\Xi'_E(s; f_{Y^\star}) = \frac{1}{s}\mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] \mid S = s\right] + \frac{\lambda}{s^2} \tag{124}$$

$$\Xi''_E(s; f_{Y^\star}) = -\frac{1}{s^2}\mathbb{E}\left[1 - 2Y\mathbb{E}\left[S^\star \mid Y\right] + Y^2\mathsf{Var}\left[S^\star \mid Y\right] \,\middle|\, S = s\right] - \frac{2\lambda}{s^3}, \tag{125}$$

for $s > 0$, or alternatively

$$\Xi''_E(s; f_{Y^\star}) = -\frac{1}{s^2} + \frac{1}{s}\mathbb{E}\left[Y^2\mathbb{E}\left[S^\star \mid Y\right] \,\middle|\, S = s\right] - \frac{2\lambda}{s^3}, \qquad s > 0. \tag{126}$$

Proof. By taking the derivative of (14) with respect to s, we get

$$\Xi'_E(s; f_{Y^\star}) = -\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\left[\log f_{Y^\star}(Y) \mid S = s\right] + \frac{1}{s} + \frac{\lambda}{s^2} \tag{127}$$

$$= \frac{1}{s}\mathbb{E}\left[Y\left(\frac{\mathrm{d}}{\mathrm{d}y}\log f_{Y^\star}(Y)\right) \,\middle|\, S = s\right] + \frac{1}{s} + \frac{\lambda}{s^2} \tag{128}$$

$$= -\frac{1}{s}\mathbb{E}\left[Y\mathbb{E}\left[S^\star \mid Y\right] \mid S = s\right] + \frac{1}{s} + \frac{\lambda}{s^2} \tag{129}$$

$$= \frac{1}{s}\mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] \mid S = s\right] + \frac{\lambda}{s^2} \tag{130}$$

where in (128) we used Lemma 6; and in (129) we used Lemma 5.

As for the second derivative of Ξ_E, we compute the derivative of (128) to get

$$\Xi''_E(s; f_{Y^\star}) = -\frac{1}{s^2}\mathbb{E}\left[Y\left(\frac{\mathrm{d}}{\mathrm{d}y}\log f_{Y^\star}(Y)\right) \,\middle|\, S = s\right] + \frac{1}{s}\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\left[Y\left(\frac{\mathrm{d}}{\mathrm{d}y}\log f_{Y^\star}(Y)\right) \,\middle|\, S = s\right] - \frac{1}{s^2} - \frac{2\lambda}{s^3} \tag{131}$$

$$= -\frac{1}{s^2}\mathbb{E}\left[1 + 2Y\left(\frac{\mathrm{d}}{\mathrm{d}y}\log f_{Y^\star}(Y)\right) + Y^2\left(\frac{\mathrm{d}^2}{\mathrm{d}y^2}\log f_{Y^\star}(Y)\right) \,\middle|\, S = s\right] - \frac{2\lambda}{s^3} \tag{132}$$

$$= -\frac{1}{s^2}\mathbb{E}\left[1 - 2Y\mathbb{E}\left[S^\star \mid Y\right] + Y^2\mathsf{Var}\left[S^\star \mid Y\right] \,\middle|\, S = s\right] - \frac{2\lambda}{s^3} \tag{133}$$

where in (131) we used Lemma 6; and in (132) we used Lemma 5. The alternative expression of Ξ''_E is obtained by directly taking the derivative of (130) with respect to s and by applying the second identity of Lemma 6:

$$\Xi''_E(s; f_{Y^\star}) = -\frac{1}{s^2} + \frac{1}{s}\mathbb{E}\left[Y^2\mathbb{E}\left[S^\star \mid Y\right] \,\middle|\, S = s\right] - \frac{2\lambda}{s^3}. \tag{134}$$

□

Appendix III
Bound on a Moment Ratio

Lemma 7. *Suppose that $f > 0$ and g is arbitrary. Then,*

$$\sup_{X\in[a,b]}\left|\frac{\mathbb{E}[g(X)]}{\mathbb{E}[f(X)]}\right| \le \max_{x\in[a,b]}\left|\frac{g(x)}{f(x)}\right|. \tag{135}$$

Proof. Let

$$M := \max_{x\in[a,b]}\left|\frac{g(x)}{f(x)}\right|. \tag{136}$$

Then,

$$|\mathbb{E}[g(X)]| \leq \mathbb{E}[|g(X)|] \tag{137}$$

$$= \mathbb{E}\left[\left|\frac{g(X)}{f(X)}\right| |f(X)|\right] \tag{138}$$

$$\leq M\mathbb{E}[|f(X)|], \tag{139}$$

therefore,

$$\sup_{X\in[a,b]} \frac{|\mathbb{E}[g(X)]|}{\mathbb{E}[|f(X)|]} = \sup_{X\in[a,b]} \left|\frac{\mathbb{E}[g(X)]}{\mathbb{E}[f(X)]}\right| \leq \max_{x\in[a,b]} \left|\frac{g(x)}{f(x)}\right| \tag{140}$$

where the first equality holds due to the assumption $f > 0$.

Note that equality in (140) is achievable by choosing X to be a point mass concentrated at x that attains the maximum. $\square$