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*—The capacity-achieving input distribution of non-coherent Rayleigh fading channels with average- and peak-power constraints is known to be discrete with a finite number of points. We sharpen this result by deriving upper and lower bounds on the number of amplitude levels. The upper bounds are based on two techniques from complex analysis: counting the number of maxima of a function that characterizes the Karush-Kuhn-Tucker conditions and an oscillation theorem. The latter result provides a stronger bound but applies only if the average power constraint is inactive.**

I. Introduction

Rayleigh fading is often used to model multipath propagation for wireless communication. Non-coherent fading has terminals lack access to channel state information (CSI) and is encountered in scenarios with low-power wireless devices, short-range wireless links, and when coherent detection is infeasible or undesirable. We study such channels subject to average- and peak-power constraints. Our primary goal is to provide bounds on the number of amplitude levels of the capacity-achieving input distribution.

A. Background

The capacity-achieving input of additive white Gaussian noise (AWGN) channels with an average power constraint is zero-mean Gaussian [1]. The optimal input becomes discrete under a peak power constraint, as shown by Smith [2] who linked the support of the input to the zeros of analytical functions. For Rayleigh fading channels, Richters [3] speculated that the optimal input with an average power constraint is discrete. This result was proved by Abou-Faycal et al. [4].

Similar results were demonstrated for other channels. Gursoy et al. [5] studied Rician fading with an additional fourth-moment constraint and showed that the optimal input is discrete with a finite number of points. Katz and Shamai [6] studied noncoherent AWGN channels with an average power constraint and showed that the optimal input amplitude is discrete with infinitely many support points. Further examples include symmetric coherent vector additive Gaussian channels [7]–[10], noncoherent block-independent fading AWGN channels [11], and Poisson channels [12]. Extensions of these findings to additive channels appear in [13]–[16]. Generalizations to multi-user channels, such as multiple access and wiretap channels, are discussed in [17] and [18]–[20], respectively. Please see [21] for a summary of discreteness results,

The proofs in the above papers are nonconstructive, relying on arguments by contradiction [2]. However, such proofs cannot limit the optimal input's support size, location, and probability values. Recently, alternative tools were developed. In [22], for AWGN channels with a peak-power constraint, the authors used Karlin's oscillation theorem and complex analysis for root counting to establish upper and lower bounds on the optimal input's support size. In [23], the authors introduced two new tools, one of which relies on the strong data-processing inequality, to bound the support size of the optimal input for Poisson noise channels. Another method to bound the values of probabilities appears in [24]. These techniques were extended to wiretap channels in [20].

Bounds on non-coherent channel capacity can be found in [25]–[27]; see also [28], [29]. The capacity scales linearly with the input power P as $\mathsf{P} \to 0$ and as $\log\log \mathsf{P}$ as $\mathsf{P} \to \infty$.

B. Organization

The paper is organized as follows. Sec. II defines notation, specifies the channel model, and reviews the Karush-Kuhn-Tucker (KKT) conditions and an oscillation theorem. Sec. III presents upper and lower bounds on the number of mass points of the capacity-achieving input. Sec. IV provides the proof of the main theorem; other proofs are relegated to the Appendix. Sec. V concludes the paper.

II. Preliminaries

A. 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 $\mathcal{A} \subseteq \mathbb{R}$ the probability distribution is written as $P_X(\mathcal{A}) = \mathbb{P}[X \in \mathcal{A}]$. The support set of P_X is

$$\mathsf{supp}(P_X) = \{x : P_X(\mathcal{D}) > 0 \text{ for every open set } \mathcal{D} \ni x\}. \tag{1}$$

When X is discrete, we write $P_X(x)$ for $P_X(\{x\})$, i.e., P_X is a probability mass function (pmf). The relative entropy of the distributions P and Q is $D(P \,\|\, Q)$.

Given a function $f : \mathbb{C} \mapsto \mathbb{C}$ and a set $\mathcal{A} \subseteq \mathbb{C}$, define the set of zeros of f in $\mathcal{A}$ as

$$\mathsf{Z}(\mathcal{A}; f) = \{z : f(z) = 0\} \cap \mathcal{A}. \tag{2}$$

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

B. Channel Model

The output of a Rayleigh fading channel with AWGN is

$$V = HU + W \tag{3}$$

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 realization of H, but both know the statistics of H. The input is subject to both average-power and peak-power constraints given by

$$\mathbb{E}\left[|U|^2\right] \le \tilde{\mathsf{P}}, \tag{4}$$

$$\mathbb{P}\left[|U| \le \tilde{\mathsf{A}}\right] = 1 \left(\text{ or equiv. } |U| \le \tilde{\mathsf{A}} \text{ a.s.} \right) \tag{5}$$

for $0 \le \tilde{\mathsf{P}}, \tilde{\mathsf{A}} \le \infty$. The capacity of the channel is

$$C_R(\tilde{\mathsf{P}}, \tilde{\mathsf{A}}) = \max_{P_U:\, \mathbb{E}[|U|^2] \le \tilde{\mathsf{P}},\, |U| \le \tilde{\mathsf{A}}} I(U; V). \tag{6}$$

An optimal input $U^\star$ has a discrete amplitude and uniform phase [4]. Moreover, the capacity is not reduced by amplitude modulation, scaling, and square-law detection with output

$$Y = \left| \tilde{H} X + \tilde{W} \right|^2 \tag{7}$$

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 (3) is the exponential model with

$$Y = \frac{1}{S} T \tag{8}$$

where $T \sim \text{Exp}(1)$ (i.e., exponential with mean one) and S are independent. It was demonstrated in [4] that

$$C_R(\tilde{\mathsf{P}}, \tilde{\mathsf{A}}) = C_E(\mathsf{P}, \mathsf{A}) = \max_{P_S: S \in (\frac{1}{1+\mathsf{A}^2}, 1], \mathbb{E}\left[\frac{1}{S}\right] \le 1 + \mathsf{P}} I(S; Y) \tag{9}$$

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 (6) and the optimal distribution in (9) is given by

$$S^\star = \frac{1}{1 + (X^\star)^2}. \tag{10}$$

In this paper, we use the exponential model (8). The results can be mapped to the original model via (10).

C. Benefits of a Peak Power Constraint

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, at low power the maximum amplitude support point $x_{\max}^\star$ approaches infinity as P approaches zero.

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

$$x_{\max}^\star(\mathsf{P}, \mathsf{A}) = \arg\max\{\mathsf{supp}(P_{X^\star})\} \tag{11}$$

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

$$\lim_{\mathsf{P} \to 0} \lim_{\mathsf{A} \to \infty} x_{\max}^\star(\mathsf{P}, \mathsf{A}) = \infty, \qquad \lim_{\mathsf{P} \to 0} \lim_{\mathsf{A} \to \infty} P_{X^\star}(x_{\max}^\star) = 0. \tag{12}$$

A similar behavior is observed numerically for the P's at which a new mass point appears in $P_{X^\star}$: the new mass point has a large amplitude. A peak power constraint is thus beneficial in several ways. First, the constraint defines a physically reasonable channel. Second, the model is more general: for large enough A, it is almost identical to having only the average-power constraint while being mathematically more tractable. However, many derived results hold for $\mathsf{A} = \infty$.

D. KKT Conditions

We review the KKT conditions for an optimal input distribution $P_{S^\star}$ that induces the output density $f_{Y^\star}$.

Lemma 1. *The input distribution $P_{S^\star}$ achieves capacity for the model (8) if and only if there exists $\lambda \ge 0$ such that*

$$\Xi_E(s; f_{Y^\star}) \le 0,\ s \in \left[\frac{1}{1+\mathsf{A}^2}, 1\right], \tag{13}$$

$$\Xi_E(s; f_{Y^\star}) = 0,\ s \in \mathsf{supp}(P_{S^\star}) \tag{14}$$

where, for $s > 0$, we have

$$\begin{aligned} \Xi_E(s; f_{Y^\star}) &\triangleq D\left(f_{Y|S}(\cdot|s) \,\middle\|\, f_{Y^\star}\right) - C - \lambda \left(\frac{1}{s} - 1 - \mathsf{P}\right) \\ &= -\mathbb{E}\left[\log(Y f_{Y^\star}(Y)) + \gamma + 1 + C \right. \\ &\qquad \left. + \lambda (Y - 1 - \mathsf{P}) \mid S = s\right] \end{aligned} \tag{15}$$

Proof. See [4]. The Lagrange multiplier λ accounts for the average-power constraint and is equal to $\lambda = \frac{\partial}{\partial \mathsf{P}} C(\mathsf{P}, \mathsf{A})$ which is a function of P and A. □

E. Oscillation Theorem

To upper bound the support size of $P_{X^\star}$ or $P_{S^\star}$, we follow steps in [22]. The key tool is the *variation-diminishing* based on Karlin's oscillation theorem [31]. To state this 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*

$$\mathscr{S}(\xi) = \sup_{m \in \mathbb{N}} \left\{ \sup_{y_1 < \cdots < y_m \subseteq \mathcal{X}} \mathscr{N}\{\xi(y_i)\}_{i=1}^m \right\} \tag{16}$$

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

The following theorem from [31, Thm. 3] (see also [32, Thm. 3.1, p. 21]) provides a key step in our proof.

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*

$$\Xi(x) = \int \xi(y)\, p(x, y)\, \mathrm{d}\mu(y). \tag{17}$$

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 $\Xi \equiv 0$.

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

$$\Xi(s) = \int_0^\infty \xi(y) f_{Y|S}(y|s)\mathrm{d}y, \qquad s > 0. \tag{18}$$

III. Main Results and Discussion

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

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

(i) Lower bound: For $\mathsf{P_A} = \min\{\mathsf{P}, \mathsf{A}^2\} \ge 0$, we have

$$\mathsf{K} \ge \underline{\mathsf{K}} = \max\left\{2, \left\lceil \sqrt{\left(\mathrm{e}^{-\gamma-1}\log(\mathsf{P_A}+1)\right)^2 + 1} \right\rceil \right\}. \tag{19}$$

(ii) Upper bound:

$$\mathsf{K} \le \overline{\mathsf{K}} = \left\lfloor \frac{\log \frac{(2\mathsf{a}+\frac{1}{2})(\frac{1}{2}+\lambda\mathsf{a})(2\mathsf{a}+1)}{(\mathsf{a}-\frac{1}{2})^2}}{\log\left(1+\frac{1}{2\mathsf{a}-1}\right)} + 1 \right\rfloor \tag{20}$$

where $\mathsf{a} = 1 + (x^\star_{\max})^2$. Moreover, if the average-power constraint is inactive, then the bound can be sharpened:

$$\mathsf{K} \le \left\lfloor \log\left(\mathrm{e}(1+\mathsf{A}^2)+1\right) + \mathrm{e}(1+\mathsf{A}^2) + 1 \right\rfloor. \tag{21}$$

Observe that $1 \le \mathsf{a} \le 1 + \mathsf{A}^2$ since $0 \le x^\star_{\max} \le \mathsf{A}$. Also, we have $\lambda \le 1$ from [4]. Therefore, we have

$$\overline{\mathsf{K}} \le \overline{\mathsf{K}}_2 = \left\lfloor \frac{\log\left(4(\frac{5}{2}+2\mathsf{A}^2)(\frac{3}{2}+\mathsf{A}^2)(3+2\mathsf{A}^2)\right)}{\log\left(1+\frac{1}{1+2\mathsf{A}^2}\right)} \right\rfloor. \tag{22}$$

We thus 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 shows a heatmap of $\log_2 \underline{\mathsf{K}}$ and also a plot of $\log_2 \overline{\mathsf{K}}_2$.

[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.

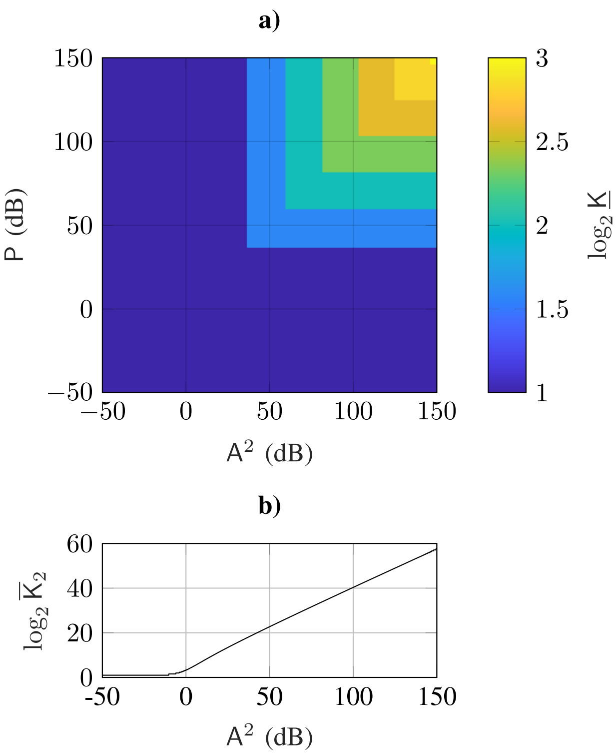

Fig. 1. For $10^{-5} \le \mathsf{P}, \mathsf{A}^2 \le 10^{15}$, logarithm of the: **a)** lower bound; and **b)** upper bound on the support size.

IV. Proof of Theorem 2

For the lower bound (19), for discrete $X^\star$ we may write

$$\begin{aligned} C(\mathsf{P}, \mathsf{A}) = I(X^\star; Y^\star) &= H(X^\star) - H(X^\star \,|\, Y^\star) \\ &\le H(X^\star) \le \log|\mathsf{supp}(P_{X^\star})|. \end{aligned} \tag{23}$$

Now apply the capacity lower bound in Appendix I.

For the upper bound (20), the KKT conditions imply

$$\mathsf{supp}(P_{S^\star}) \subseteq \mathrm{Z}\left((0,1]; \Xi_E(\cdot; f_{Y^\star})\right) \tag{24}$$

or

$$|\mathsf{supp}(P_{S^\star})| \le \mathrm{N}\left((0,1];\, \Xi_E(\cdot; f_{Y^\star})\right). \tag{25}$$

We use two approaches to evaluate (25): directly count the number of zeros of the function Ξ_E, and apply the oscillation theorem to Ξ_E. In both cases, we must upper bound the number of zeros of a function. A key tool is the following.

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 $f \ne 0$ which is analytic on $|z| \le (vt+v+t)\mathsf{R}$, its number of zeros $\mathrm{N}(\mathcal{D}_\mathsf{R}; f)$ within the disk $\mathcal{D}_\mathsf{R} = \{z\colon |z| \le \mathsf{R}\}$ satisfies*

$$\mathrm{N}(\mathcal{D}_\mathsf{R}; f) \le \frac{1}{\log v} \log \frac{\max_{|z| \le (vt+v+t)\mathsf{R}} |f(z)|}{\max_{|z| \le t\mathsf{R}} |f(z)|}. \tag{26}$$

1) Direct Counting of the Zeros of Ξ_E: We 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 have

$$
\begin{aligned}
&|\mathsf{supp}(P_{S^\star})| \le \mathrm{N}\left([\mathsf{r}, 1];\ \Xi_E(\cdot; f_{Y^\star}) \right)\\
&\stackrel{(a)}{\le} \mathrm{N}\left([\mathsf{r}, 1];\ \Xi'_E(\cdot; f_{Y^\star}) \right) + 1\\
&\stackrel{(b)}{=} \mathrm{N}\left([\mathsf{r}, 1];\ \frac{1}{s}\mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] \mid S = s\right] + \frac{\lambda}{s^2} \right) + 1\\
&\stackrel{(c)}{=} \mathrm{N}\left([\mathsf{r}, 1];\ \mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] + \lambda Y \mid S = s\right] \right) + 1\\
&\stackrel{(d)}{=} \mathrm{N}\left(\left[-\frac{1-\mathsf{r}}{2}, \frac{1-\mathsf{r}}{2}\right];\right.\\
&\qquad \left. \mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] + \lambda Y \,\middle|\, S = s + \frac{1+\mathsf{r}}{2}\right] \right) + 1
\end{aligned} \tag{27}
$$

where (a) follows by Rolle's theorem (see, [22, Lemma 3]); (b) follows by Proposition 3; (c) is 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 (d) follows from the change of variable $s \to s + \frac{1+\mathsf{r}}{2}$ which centers the interval at $s = 0$. Now consider the complex analytic extension $z \mapsto \breve{g}(z)$ of the real function

$$
s \mapsto g(s) = \mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] + \lambda Y \,\middle|\, S = s + \frac{1+\mathsf{r}}{2}\right] \tag{28}
$$

which is analytic on $|z| < \frac{1+\mathsf{r}}{2}$.

To apply Tijdeman's lemma, we must bound the maximum value of $\breve{g}$ in a disk of radius B.

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

$$
\max_{|z|\le\mathsf{B}} |\breve{g}(z)| \le \frac{(\frac{1+\mathsf{r}}{2} + \mathsf{B})(\frac{1+\mathsf{r}}{2} - \mathsf{B} + 1 + \lambda)}{\left(\frac{1+\mathsf{r}}{2} - \mathsf{B}\right)^2} \tag{29}
$$

and for $\frac{1}{2} \le \mathsf{B} < \frac{1+\mathsf{r}}{2}$ we have

$$
\max_{|z|\le\mathsf{B}} |\breve{g}(z)| \ge 1 + \frac{\lambda}{1+\frac{\mathsf{r}}{2}} - \frac{1}{1+\frac{\mathsf{r}}{2}} \ge \frac{1}{\frac{2}{\mathsf{r}}+1}. \tag{30}
$$

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

$$
\begin{aligned}
&\max_{|z|\le\mathsf{B}} |\breve{g}(z)| \le \max_{|z|\le\mathsf{B}} \qquad (31)\\
&\int_0^\infty \left|(1 - y\mathbb{E}\left[S^\star \mid Y = y\right] + \lambda y)\left(z + \frac{1+\mathsf{r}}{2}\right) \mathrm{e}^{-zy - \frac{1+\mathsf{r}}{2}y}\right| \mathrm{d}y \qquad (32)\\
&\le \int_0^\infty (1 + y\mathbb{E}\left[S^\star \mid Y = y\right] + \lambda y)\left(\mathsf{B} + \frac{1+\mathsf{r}}{2}\right) \mathrm{e}^{\mathsf{B}y - \frac{1+\mathsf{r}}{2}y}\mathrm{d}y \qquad (33)\\
&\le \int_0^\infty (1 + y(1+\lambda))\left(\mathsf{B} + \frac{1+\mathsf{r}}{2}\right) \mathrm{e}^{\mathsf{B}y - \frac{1+\mathsf{r}}{2}y}\mathrm{d}y \qquad (34)\\
&= \frac{(\frac{1+\mathsf{r}}{2} + \mathsf{B})(\frac{1+\mathsf{r}}{2} - \mathsf{B} + 1 + \lambda)}{\left(\frac{1+\mathsf{r}}{2} - \mathsf{B}\right)^2}, \qquad (35)
\end{aligned}
$$

where (32) follows from Jensen's inequality; (33) follows from triangle inequality, from $|\mathrm{e}^{-zy}| = \mathrm{e}^{-\Re\{z\}y} \le \mathrm{e}^{\mathsf{B}y}$, and from $|z| \le \mathsf{B}$; and (34) follows from $S^\star \le 1$.

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

$$
\begin{aligned}
&\max_{|z|\le\mathsf{B}} |\breve{g}(z)| \ge \left|\breve{g}\left(\frac{1}{2}\right)\right| \qquad (36)\\
&= \left|\int_0^\infty (1 - y\mathbb{E}\left[S^\star \mid Y = y\right] + \lambda y)\left(1 + \frac{\mathsf{r}}{2}\right)\mathrm{e}^{-(1+\frac{\mathsf{r}}{2})y}\mathrm{d}y\right| \qquad (37)\\
&= \left|1 + \frac{\lambda}{1+\frac{\mathsf{r}}{2}} - \int_0^\infty y\mathbb{E}\left[S^\star \mid Y = y\right]\left(1 + \frac{\mathsf{r}}{2}\right)\mathrm{e}^{-(1+\frac{\mathsf{r}}{2})y}\mathrm{d}y\right|. \qquad (38)
\end{aligned}
$$

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

$$
\begin{aligned}
&\int_0^\infty y\mathbb{E}\left[S^\star \mid Y = y\right]\left(1 + \frac{\mathsf{r}}{2}\right)\mathrm{e}^{-(1+\frac{\mathsf{r}}{2})y}\mathrm{d}y \qquad (39)\\
&\le \int_0^\infty y\left(1 + \frac{\mathsf{r}}{2}\right)\mathrm{e}^{-(1+\frac{\mathsf{r}}{2})y}\mathrm{d}y = \frac{1}{1+\frac{\mathsf{r}}{2}} \qquad (40)
\end{aligned}
$$

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

$$
\max_{|z|\le\mathsf{B}} |\breve{g}(z)| \ge 1 + \frac{\lambda}{1+\frac{\mathsf{r}}{2}} - \frac{1}{1+\frac{\mathsf{r}}{2}} \ge \frac{1}{\frac{2}{\mathsf{r}}+1}, \tag{41}
$$

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

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

$$
\begin{aligned}
&\mathrm{N}\left(\mathcal{D}_{\frac{1-\mathsf{r}}{2}};\ \breve{g}\right)\\
&\le \min_{1<v<\frac{1+\mathsf{r}}{1-\mathsf{r}}} \left\{ \frac{\log \frac{(1+\frac{\mathsf{r}}{2}+v(1-\frac{\mathsf{r}}{2}))(\frac{\mathsf{r}}{2}-v(1-\frac{\mathsf{r}}{2})+1+\lambda)}{\left(\frac{\mathsf{r}}{2}-v(1-\frac{\mathsf{r}}{2})\right)^2} - \log\frac{1}{\frac{2}{\mathsf{r}}+1}}{\log v} \right\}\\
&\le \frac{\log \frac{(2+\frac{\mathsf{r}}{2})(\frac{\mathsf{r}}{2}+\lambda)(\frac{2}{\mathsf{r}}+1)}{\left(1-\frac{\mathsf{r}}{2}\right)^2}}{\log\frac{1}{1-\frac{\mathsf{r}}{2}}} = \frac{\log \frac{(\frac{2}{\mathsf{r}}+\frac{1}{2})(\frac{1}{2}+\frac{\lambda}{\mathsf{r}})(\frac{2}{\mathsf{r}}+1)}{\left(\frac{1}{\mathsf{r}}-\frac{1}{2}\right)^2}}{\log\left(1+\frac{1}{\frac{2}{\mathsf{r}}-1}\right)}
\end{aligned} \tag{42}
$$

where in (42) we chose $v = \frac{1}{1-\frac{\mathsf{r}}{2}} = 1 + \frac{1}{\frac{2}{\mathsf{r}}-1} < \frac{1+\mathsf{r}}{1-\mathsf{r}}$ for $\mathsf{r} > 0$. Defining $\mathsf{a} := 1/\mathsf{r}$ and collecting the results, we have

$$
\begin{aligned}
|\mathsf{supp}(P_{S^\star})| &\le \mathrm{N}\left(\left[-\frac{1-\mathsf{r}}{2}, \frac{1-\mathsf{r}}{2}\right];\ g\right) + 1\\
&\le \mathrm{N}\left(\mathcal{D}_{\frac{1-\mathsf{r}}{2}};\ \breve{g}\right) + 1\\
&\le \frac{\log\frac{(2\mathsf{a}+\frac{1}{2})(\frac{1}{2}+\lambda\mathsf{a})(2\mathsf{a}+1)}{\left(\mathsf{a}-\frac{1}{2}\right)^2}}{\log\left(1+\frac{1}{2\mathsf{a}-1}\right)} + 1
\end{aligned} \tag{43}
$$

which grows as a $\log(\mathsf{a})$ for $\mathsf{a} \to \infty$. This concludes the proof.

2) Application of the Oscillation Theorem: For $\lambda = 0$ we improve the upper bound on $|\mathsf{supp}(P_{S^\star})|$ by using the oscillation theorem. We can write

$$
\begin{aligned}
&|\mathsf{supp}(P_{S^\star})| \le \mathrm{N}\left((0, 1];\ \Xi_E(\cdot; f_{Y^\star}) \right)\\
&\stackrel{(a)}{=} \mathrm{N}\left((0, 1];\ \mathbb{E}\left[\log(Y f_{Y^\star}(Y)) + \gamma + 1 + C \mid S = s\right] \right)\\
&\stackrel{(b)}{\le} \mathrm{N}\left((0, \infty);\ -\log(y f_{Y^\star}(y)) - \gamma - 1 - C \right)\\
&\stackrel{(c)}{\le} \mathrm{N}\left((0, \infty);\ \mathbb{E}\left[S^\star \mid Y = y\right] - y^{-1} \right) + 1\\
&\stackrel{(d)}{=} \mathrm{N}\left((0, \infty);\ y\mathbb{E}\left[S^\star \mid Y = y\right] - 1 \right) + 1
\end{aligned} \tag{44}
$$

where in (a) we used (15); (b) follows from the oscillation theorem (Th. 1); (c) follows from Rolle's theorem [22, Lemma 3] and Lemma 5; and (d) holds because multiplying the function by y does not change the number of zeros. Next, note that

$$y\mathbb{E}\left[S^\star \mid Y=y\right]-1 \geq y s_{\min}^\star - 1 \tag{45}$$

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

$$y\mathbb{E}\left[S^\star \mid Y=y\right]-1 \leq y - 1 \tag{46}$$

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

$$\begin{aligned} |\mathsf{supp}(P_{S^\star})| &\leq \mathrm{N}\left([1,\mathsf{R}]\,;\, y\mathbb{E}\left[S^\star \mid Y=y\right]-1\right)+1 \\ &\overset{(a)}{=} \mathrm{N}\left([1,\mathsf{R}]\,;\, y\mathbb{E}\left[(S^\star)^2\mathrm{e}^{-S^\star y}\right]-\mathbb{E}\left[S^\star\mathrm{e}^{-S^\star y}\right]\right)+1 \\ &\leq \mathrm{N}\left(\mathcal{D}_{\mathsf{R}};\, \breve{g}(z)\right)+1, \end{aligned} \tag{47}$$

where (a) follows from $\mathbb{E}\left[S^\star \mid Y=y\right] = \frac{\mathbb{E}\left[(S^\star)^2\mathrm{e}^{-S^\star y}\right]}{\mathbb{E}\left[S^\star\mathrm{e}^{-S^\star y}\right]}$ and from multiplying the function by $\mathbb{E}\left[S^\star\mathrm{e}^{-S^\star y}\right]$; and where $\breve{g}(z)$ is the complex analytic extension of

$$g(y) = y\mathbb{E}\left[(S^\star)^2\mathrm{e}^{-S^\star y}\right]-\mathbb{E}\left[S^\star\mathrm{e}^{-S^\star y}\right] \tag{48}$$

which is analytic on $z \in \mathbb{C}$. To apply Tijdeman's lemma, we must bound the maximum value of $\breve{g}$ in a disk of radius B. For the upper bound, we have

$$\begin{aligned} \max_{|z|\leq\mathsf{B}} |\breve{g}(z)| &= \max_{|z|\leq\mathsf{B}} \left|\mathbb{E}\left[((S^\star)^2 z - S^\star)\mathrm{e}^{-S^\star z}\right]\right| \\ &\overset{(a)}{\leq} \max_{|z|\leq\mathsf{B}} \mathbb{E}\left[|(S^\star)^2 z - S^\star|\mathrm{e}^{-S^\star \Re\{z\}}\right] \\ &\overset{(b)}{\leq} \mathbb{E}\left[((S^\star)^2\mathsf{B} + S^\star)\mathrm{e}^{S^\star\mathsf{B}}\right]. \end{aligned} \tag{49}$$

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

$$\begin{aligned} \max_{|z|\leq\mathsf{B}} |\breve{g}(z)| \geq |\breve{g}(-\mathsf{B})| &= \left|\mathbb{E}\left[(-(S^\star)^2\mathsf{B} - S^\star)\mathrm{e}^{S^\star\mathsf{B}}\right]\right| \\ &= \mathbb{E}\left[((S^\star)^2\mathsf{B} + S^\star)\mathrm{e}^{S^\star\mathsf{B}}\right]. \end{aligned} \tag{50}$$

Comparing (49) and (50), we obtain

$$\max_{|z|\leq\mathsf{B}} |\breve{g}(z)| = \mathbb{E}\left[((S^\star)^2\mathsf{B} + S^\star)\mathrm{e}^{S^\star\mathsf{B}}\right]. \tag{51}$$

We can now apply Tijdeman's lemma. We compute

$$\begin{aligned} &\frac{\max_{|z|\leq(vt+v+t)\mathsf{R}} |\breve{g}(z)|}{\max_{|z|\leq t\mathsf{R}} |\breve{g}(z)|} \\ &\overset{(a)}{=} \frac{\mathbb{E}\left[((S^\star)^2(vt+v+t)\mathsf{R} + S^\star)\mathrm{e}^{S^\star(vt+v+t)\mathsf{R}}\right]}{\mathbb{E}\left[((S^\star)^2 t\mathsf{R} + S^\star)\mathrm{e}^{S^\star t\mathsf{R}}\right]} \\ &\overset{(b)}{\leq} \max_{s\in[\frac{1}{1+\mathsf{A}^2},1]} \frac{(s^2(vt+v+t)\mathsf{R} + s)\mathrm{e}^{s(vt+v+t)\mathsf{R}}}{(s^2 t\mathsf{R} + s)\mathrm{e}^{st\mathsf{R}}} \\ &= \max_{s\in[\frac{1}{1+\mathsf{A}^2},1]} \frac{(s(vt+v+t)\mathsf{R} + 1)\mathrm{e}^{sv(t+1)\mathsf{R}}}{(st\mathsf{R} + 1)} \\ &\overset{(c)}{=} \frac{((vt+v+t)\mathsf{R} + 1)\mathrm{e}^{v(t+1)\mathsf{R}}}{(t\mathsf{R} + 1)}, \end{aligned} \tag{52}$$

where (a) follows from (51); (b) follows from the bound on a moment ratio derived in Lemma 7 and from the peak-power constraint $s \geq \frac{1}{1+\mathsf{A}^2}$; and (c) gives the maximum value obtained for $s = 1$. By choosing $t = 0$, we have

$$\begin{aligned} \mathrm{N}\left(\mathcal{D}_{\mathsf{R}};\, \breve{g}(z)\right) &\leq \min_{v>1,\, t\geq 0} \left\{ \frac{\log \frac{\max_{|z|\leq(vt+v+t)\mathsf{R}} |\breve{g}(z)|}{\max_{|z|\leq t\mathsf{R}} |\breve{g}(z)|}}{\log v} \right\} \\ &\leq \min_{v>1} \left\{ \frac{\log(v\mathsf{R}+1) + v\mathsf{R}}{\log v} \right\} \\ &\overset{(a)}{\leq} \log\left(\mathrm{e}\mathsf{R}+1\right) + \mathrm{e}\mathsf{R} \\ &\overset{(b)}{\leq} \log\left(\mathrm{e}(1+\mathsf{A}^2)+1\right) + \mathrm{e}(1+\mathsf{A}^2) \end{aligned} \tag{53}$$

where in (a) we have used $v = \mathrm{e}$; and in (b) we used $\mathsf{R} = \frac{1}{s_{\min}^\star} \leq 1 + \mathsf{A}^2$. Putting everything together, we obtain

$$|\mathsf{supp}(P_{S^\star})| \leq \log\left(\mathrm{e}(1+\mathsf{A}^2)+1\right) + \mathrm{e}(1+\mathsf{A}^2) + 1. \tag{54}$$

V. Conclusion

This work investigated non-coherent Rayleigh channels and presented new results on the structure of the capacity-achieving input based on the Karush-Kuhn-Tucker (KKT) conditions. We considered both average and peak-power constraints to make the model more general and practical.

We provided bounds on the number of amplitude levels of the capacity-achieving input. Specifically, the upper bound was based on bounding the number of maxima of a function characterizing the KKT conditions. In the limit of a large peak-power A^2, the upper bound scales as $\mathsf{A}^2\log(\mathsf{A})$. Furthermore, we used the oscillation theorem to refine this bound to scale as A^2 when the average power constraint is inactive.

The asymptotic growth rate of the optimal number of amplitude levels remains open, however, given that the asymptotic order of the lower bound is $\log(\mathsf{A})$. We expect that both the lower and upper bounds need further refinement.

References

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

[2] 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.

[3] J. S. Richters, "Communication over fading dispersive channels," Massachusetts Institute of Technology, Research Laboratory of Electronics, Tech. Rep. 464, Nov. 1967.

[4] 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.

[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 Int. Symp. Inf. Theory*, 2003, pp. 373–373.

[12] S. Shamai, "Capacity of a pulse amplitude modulated direct detection photon channel," *IEE Proc. 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* $\mathsf{P}_\mathsf{A} = \min(\mathsf{P}, \mathsf{A}^2)$*, a lower bound on the channel capacity is given by*

$$C(\mathsf{P}, \mathsf{A}) \geq \underline{C}(\mathsf{P}_\mathsf{A}) = \log\left(\sqrt{\left(\mathrm{e}^{-\gamma-1}\log(\mathsf{P}_\mathsf{A}+1)\right)^2 + 1}\right) \tag{55}$$

$$\geq \max\left\{0, \log(\log(1+\mathsf{P}_\mathsf{A}))) - \gamma - 1\right\}, \tag{56}$$

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

$$\mathbb{E}\left[\frac{1}{S}\right] = \mathbb{E}[\mathrm{e}^{\log\frac{1}{S}}] = \frac{\mathrm{e}^{\log(\mathsf{P}_\mathsf{A}+1)} - \mathrm{e}^0}{\log(\mathsf{P}_\mathsf{A}+1)} = \frac{\mathsf{P}_\mathsf{A}}{\log(\mathsf{P}_\mathsf{A}+1)} \leq \mathsf{P}_\mathsf{A} + 1 \leq \mathsf{P} + 1 \tag{57}$$

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

$$h(L) = h\left(\log\frac{1}{S} - Z\right) \tag{58}$$

$$\geq \log\left(\sqrt{\mathrm{e}^{2h\left(\log\frac{1}{S}\right)} + \mathrm{e}^{2h(Z)}}\right) \tag{59}$$

$$= \log\left(\sqrt{\mathrm{e}^{2\log\log(\mathsf{P}_\mathsf{A}+1)} + \mathrm{e}^{2(\gamma+1)}}\right) \tag{60}$$

$$= \log\left(\sqrt{(\log(\mathsf{P}_\mathsf{A}+1))^2 + \mathrm{e}^{2(\gamma+1)}}\right) \tag{61}$$

$$\geq \max\left\{\log(\log(\mathsf{P}_\mathsf{A}+1))), \gamma+1\right\}. \tag{62}$$

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

$$C(\mathsf{P}, \mathsf{A}) \geq \max\left\{\log(\log(\mathsf{P}_\mathsf{A}+1))) - \gamma - 1, 0\right\} \tag{63}$$

$$= [\log(\log(\mathsf{P}_\mathsf{A}+1))) - \gamma - 1]^+ \tag{64}$$

$$:= \underline{C}(\mathsf{P}_\mathsf{A}). \tag{65}$$

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*

$$f_{Y|S}(y|s) = s\exp(-sy), \qquad y \geq 0. \tag{66}$$

We have

$$\frac{\mathrm{d}}{\mathrm{d}s}f_{Y|S}(y|s) = -\frac{1-sy}{s^2}\frac{\mathrm{d}}{\mathrm{d}y}f_{Y|S}(y|s) \tag{67}$$

$$= \frac{1-sy}{s}f_{Y|S}(y|s). \tag{68}$$

Proof. First of all, compute

$$\frac{\mathrm{d}}{\mathrm{d}s}f_{Y|S}(y|s) = \frac{\mathrm{d}}{\mathrm{d}s}\exp(-sy + \log s) \tag{69}$$

$$= \frac{1-sy}{s}\exp(-sy + \log s) \tag{70}$$

$$= \frac{1-sy}{s}f_{Y|S}(y|s). \tag{71}$$

Then, compute

$$\frac{\mathrm{d}}{\mathrm{d}y}f_{Y|S}(y|s) = \frac{\mathrm{d}}{\mathrm{d}y}\exp(-sy + \log s) \tag{72}$$

$$= -s\exp(-sy + \log s) \tag{73}$$

$$= -s f_{Y|S}(y|s) \tag{74}$$

$$= -\frac{s^2}{1-sy}\frac{\mathrm{d}}{\mathrm{d}s} f_{Y|S}(y|s) \tag{75}$$

$\square$

Lemma 5. *Let us consider the exponential channel model of (66) and an output distribution f_Y. Then, we have*

$$\frac{\mathrm{d}}{\mathrm{d}y}\log f_Y(y) = -\mathbb{E}\left[S \mid Y=y\right], \tag{76}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2}\log f_Y(y) = \mathsf{Var}\left[S \mid Y=y\right]. \tag{77}$$

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

$$f_{Y|S}(y|s) = \exp\left(-sy + \log s\right) \tag{78}$$

$$= h(y)\mathrm{e}^{sT(y)-\phi(s)} \tag{79}$$

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

$$\mathbb{E}\left[S \mid Y=y\right] = D_y \log\frac{f_Y(y)}{h(y)} = -\frac{\mathrm{d}}{\mathrm{d}y}\log f_Y(y). \tag{80}$$

As for the second derivative, write

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2}\log f_Y(y) = -\frac{\mathrm{d}}{\mathrm{d}y}\mathbb{E}\left[S \mid Y=y\right] \tag{81}$$

$$= D_y\mathbb{E}\left[S \mid Y=y\right] \tag{82}$$

$$= \mathsf{Var}\left[S \mid Y=y\right] \tag{83}$$

where the last step is due to [34, Prop. 4]. $\square$

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

$$\mathbb{E}\left[f(Y) \mid S=s\right] = f(0) + \frac{1}{s}\mathbb{E}\left[f'(Y) \mid S=s\right], \tag{84}$$

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\left[f(Y) \mid S=s\right] = -\frac{1}{s}\mathbb{E}\left[Yf'(Y) \mid S=s\right], \tag{85}$$

$$\frac{\mathrm{d}}{\mathrm{d}s}\frac{1}{s}\mathbb{E}\left[f(Y) \mid S=s\right] = -\frac{1}{s}\mathbb{E}\left[Yf(Y) \mid S=s\right]. \tag{86}$$

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

$$\mathbb{E}\left[f(Y) \mid S=s\right] = \int_0^\infty f(y)se^{-sy}\mathrm{d}y \tag{87}$$

$$= \left(-e^{-sy}f(y)\right)\Big|_0^\infty - \int_0^\infty f'(y)\left(-e^{-sy}\right)\mathrm{d}y \tag{88}$$

$$= f(0) + \frac{1}{s}\int_0^\infty se^{-sy}f'(y)\mathrm{d}y \tag{89}$$

$$= f(0) + \frac{1}{s}\mathbb{E}\left[f'(Y) \mid S=s\right], \tag{90}$$

where in (89) 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

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\left[f(Y) \mid S=s\right] = \frac{\mathrm{d}}{\mathrm{d}s}\int_0^\infty f(y)se^{-sy}\mathrm{d}y \tag{91}$$

$$= \int_0^\infty f(y)e^{-sy}\mathrm{d}y - \int_0^\infty f(y)yse^{-sy}\mathrm{d}y \tag{92}$$

$$= \left(ye^{-sy}f(y)\right)\Big|_0^\infty - \int_0^\infty y\left(-se^{-sy}f(y) + e^{-sy}f'(y)\right)\mathrm{d}y - \int_0^\infty sye^{-sy}f(y)\mathrm{d}y \tag{93}$$

$$= -\int_0^\infty ye^{-sy}f'(y)\mathrm{d}y \tag{94}$$

$$= -\frac{1}{s}\mathbb{E}\left[Yf'(Y) \mid S = s\right], \tag{95}$$

where in (94) we used that $\lim_{y\to\infty} f(y)e^{-sy} = 0$. To prove the third identity, write

$$\frac{\mathrm{d}}{\mathrm{d}s}\frac{1}{s}\mathbb{E}\left[f(Y) \mid S = s\right] = \frac{\mathrm{d}}{\mathrm{d}s}\int_0^\infty \mathrm{e}^{-sy} f(y)\mathrm{d}y \tag{96}$$

$$= -\int_0^\infty y\mathrm{e}^{-sy} f(y)\mathrm{d}y \tag{97}$$

$$= -\mathbb{E}\left[Yf(Y) \mid S = s\right]. \tag{98}$$

□

Proposition 3. *The first and second derivatives of function Ξ_E given in (15) are as follows*

$$\Xi'_E(s; f_{Y^\star}) = \frac{1}{s}\mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] \mid S = s\right] + \frac{\lambda}{s^2} \tag{99}$$

$$\Xi''_E(s; f_{Y^\star}) = -\frac{1}{s^2}\mathbb{E}\left[1 - 2Y\mathbb{E}\left[S^\star \mid Y\right] + Y^2\mathsf{Var}\left[S^\star \mid Y\right] \,\middle|\, S = s\right] - \frac{2\lambda}{s^3}, \tag{100}$$

for $s > 0$, or alternatively

$$\Xi''_E(s; f_{Y^\star}) = -\frac{1}{s^2} + \frac{1}{s}\mathbb{E}\left[Y^2\mathbb{E}\left[S^\star \mid Y\right] \,\middle|\, S = s\right] - \frac{2\lambda}{s^3}, \qquad s > 0. \tag{101}$$

Proof. By taking the derivative of (15) with respect to s, we get

$$\Xi'_E(s; f_{Y^\star}) = -\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\left[\log f_{Y^\star}(Y) \mid S = s\right] + \frac{1}{s} + \frac{\lambda}{s^2} \tag{102}$$

$$= \frac{1}{s}\mathbb{E}\left[Y\left(\frac{\mathrm{d}}{\mathrm{d}y}\log f_{Y^\star}(Y)\right) \,\middle|\, S = s\right] + \frac{1}{s} + \frac{\lambda}{s^2} \tag{103}$$

$$= -\frac{1}{s}\mathbb{E}\left[Y\mathbb{E}\left[S^\star \mid Y\right] \mid S = s\right] + \frac{1}{s} + \frac{\lambda}{s^2} \tag{104}$$

$$= \frac{1}{s}\mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] \mid S = s\right] + \frac{\lambda}{s^2} \tag{105}$$

where in (103) we used Lemma 6; and in (104) we used Lemma 5.

As for the second derivative of Ξ_E, we compute the derivative of (103) to get

$$\Xi''_E(s; f_{Y^\star}) = -\frac{1}{s^2}\mathbb{E}\left[Y\left(\frac{\mathrm{d}}{\mathrm{d}y}\log f_{Y^\star}(Y)\right) \,\middle|\, S = s\right] + \frac{1}{s}\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\left[Y\left(\frac{\mathrm{d}}{\mathrm{d}y}\log f_{Y^\star}(Y)\right) \,\middle|\, S = s\right] - \frac{1}{s^2} - \frac{2\lambda}{s^3} \tag{106}$$

$$= -\frac{1}{s^2}\mathbb{E}\left[1 + 2Y\left(\frac{\mathrm{d}}{\mathrm{d}y}\log f_{Y^\star}(Y)\right) + Y^2\left(\frac{\mathrm{d}^2}{\mathrm{d}y^2}\log f_{Y^\star}(Y)\right) \,\middle|\, S = s\right] - \frac{2\lambda}{s^3} \tag{107}$$

$$= -\frac{1}{s^2}\mathbb{E}\left[1 - 2Y\mathbb{E}\left[S^\star \mid Y\right] + Y^2\mathsf{Var}\left[S^\star \mid Y\right] \,\middle|\, S = s\right] - \frac{2\lambda}{s^3} \tag{108}$$

where in (106) we used Lemma 6; and in (107) we used Lemma 5. The alternative expression of Ξ''_E is obtained by directly taking the derivative of (105) with respect to s and by applying the second identity of Lemma 6:

$$\Xi''_E(s; f_{Y^\star}) = -\frac{1}{s^2} + \frac{1}{s}\mathbb{E}\left[Y^2\mathbb{E}\left[S^\star \mid Y\right] \mid S = s\right] - \frac{2\lambda}{s^3}. \tag{109}$$

□

Appendix III
Bound on a Moment Ratio

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

$$\sup_{X\in[a,b]}\left|\frac{\mathbb{E}[g(X)]}{\mathbb{E}[f(X)]}\right| \le \max_{x\in[a,b]}\left|\frac{g(x)}{f(x)}\right|. \tag{110}$$

Proof. Let

$$M := \max_{x\in[a,b]}\left|\frac{g(x)}{f(x)}\right|. \tag{111}$$

Then,

$$|\mathbb{E}[g(X)]| \leq \mathbb{E}[|g(X)|] \tag{112}$$

$$= \mathbb{E}\left[\left|\frac{g(X)}{f(X)}\right| |f(X)|\right] \tag{113}$$

$$\leq M\mathbb{E}[|f(X)|], \tag{114}$$

therefore,

$$\sup_{X \in [a,b]} \frac{|\mathbb{E}[g(X)]|}{\mathbb{E}[|f(X)|]} = \sup_{X \in [a,b]} \left|\frac{\mathbb{E}[g(X)]}{\mathbb{E}[f(X)]}\right| \leq \max_{x \in [a,b]} \left|\frac{g(x)}{f(x)}\right| \tag{115}$$

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

Note that equality in (115) is achievable by choosing X to be a point mass concentrated at x that attains the maximum. $\square$