Properties of the Capacity-Achieving Input of Non-Coherent Rayleigh Fading Channels

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 works studies non-coherent Rayleigh fading channels subject to average- and peak-power constraints. Several properties of the optimal input distribution are derived based on the Karush-Kuhn-Tucker conditions. In particular, the capacity-achieving distribution is characterized in the small peak and average power regimes, upper and lower bounds on the optimal input probabilities are presented, insights about the locations of the support points are provided, and bounds on the channel capacity are established.**

I. Introduction

Rayleigh fading models help to characterize the impact of multipath propagation in wireless communication. Wireless channel signals encounter multiple paths due to reflections, diffractions, and scattering, resulting in time-varying signal distortions at the receiver. In this work, we consider non-coherent models where the channel state information (CSI) is unavailable at the receiver and transmitter. Non-coherent Rayleigh fading is often observed for low-power or low-cost wireless devices, short-range wireless links, or cases where maintaining coherent detection is impractical or undesirable.

Although significant progress has been made in understanding non-coherent communication, characterizing capacity remains a formidable challenge. The objective of this study is to investigate properties of capacity-achieving input distributions when subjected to average-power and peak-power constraints. We thereby gain insight into the limits of communication. Besides the theoretical interest, properties of the optimal distributions guide the design 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 know the statistics of H. The input is subject to both average-and peak-power constraints given by

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

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

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

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

B. Contributions and Paper Outline

This paper is organized as follows. Sec. II surveys the literature and Sec. III derives equivalent channel models and reviews the Karush-Kuhn-Tucker (KKT) conditions. Sec. IV presents our main results: a characterization of the capacity-achieving distribution in the small (but non-vanishing) peak and average power regime, upper bounds on the values of probability masses of the optimal input distribution, lower bounds on the probabilities of the maximum point and zero point of the capacity achieving distribution, and results on the location of the support points. Sec. V concludes the paper.

C. Notation

All logarithms are to the base e. Deterministic scalar quantities are denoted by lower-case letters and random variables by uppercase letters. For a random variable X and any measurable set $\mathcal{A} \subseteq \mathbb{R}$ we write the probability distribution of X by $P_X(\mathcal{A}) = \mathbb{P}[X \in \mathcal{A}]$. The support set of P_X is

$$\begin{aligned}\mathsf{supp}(P_X) = \{x : \text{for every open set } \mathcal{D} \ni x \\ \text{we have that } P_X(\mathcal{D}) > 0\}.\end{aligned} \tag{5}$$

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 $D\left(P \,\|\, Q\right)$.

II. Background

The capacity-achieving input distribution of an AWGN channel with an average-power constraint is Gaussian [1]. Interestingly, if the average-power constraint is replaced by a peak-power constraint, then the capacity-achieving distribution is discrete. This result was shown by Smith in [2] who introduced a new approach connecting the support of the capacity-achieving distribution to the zeros of analytic functions. For the Rayleigh-fading channel, the capacity-achieving distribution with an average-power constraint was conjectured to be discrete by Richter in [3] by also appealing to properties of analytic functions. This conjecture was proved by Abou-Faycal *et al.* in [4] who also showed that there is positive-probability mass point at zero input.

The capacity-achieving distribution of other fading channels is also discrete. Gursoy *et al.* [5] considered a Rician channel with an additional fourth moment constraint and showed that the capacity-achieving distribution is discrete with a finite number of points. In [6], Katz and Shamai studied noncoherent AWGN channels with an average-power constraint and showed that the optimal input amplitude is discrete with an infinite number of mass points. Other channels for which either an input distribution is discrete or some component of the optimal input distribution (e.g., amplitude) is discrete include symmetric coherent vector additive Gaussian channels [7]–[10]; noncoherent block-independent AWGN channels [11]; and Poisson channels [12]. Attempts to generalize these results to additive channels are described in [13]–[16]. Generalizations to multi-user channels such as multiple access and wiretap channels can be found in [17] and [18]–[20], respectively. We refer the reader to [21] for a summary of related results.

Recently, alternative techniques were applied to refine our knowledge about the structure of the capacity-achieving distribution. For instance, [22] introduced two new techniques to upper and lower bound the probabilities of the support points; one of the methods relies on the strong data-processing inequality. Another method to bound the probabilities can be found in [23].

Characterizing the capacity-achieving input is often easier in the low-power regime. For example, several papers show that the optimal input distribution (or its magnitude for vector AWGN channel) in the low peak-power regime is supported on only two points [2], [9], [24]–[26]. We remark that all of these results hold in the small but non-vanishing peak-power regime. In contrast, for the non-coherent Rayleigh fading channel, to the best of our knowledge it has not been shown that two-point distributions are optimal at low power. Existing results show only that, as the power vanishes, two-point distributions attain the same first and second-order terms of the Taylor expansion of the capacity around zero power [27]. These binary distributions are called *flash-signals* in [27] since the zero signal has large probability and one extremely large signal value has a vanishing probability as the average power approaches infinity. In this work, we show that binary inputs are indeed optimal for small but non-vanishing power.

Bounds on the capacity of non-coherent fading channels have also been considered in [28]–[30]; the interested reader is referred to [31], [32] for a literature review. It is known that the capacity in low-power regime scales as P and in the large power regime scales as $\log\log \mathsf{P}$ where P denotes power.

III. Channel Models and KKT Conditions

This section presents two equivalent channel models that we use in our analysis. We also present auxiliary results and tools such as the KKT conditions, and we briefly argue why including an additional peak-power constraint is reasonable.

A. Equivalent Channel Models

We present a normalization of the Rayleigh channel model which lets us focus on the distribution of $X \triangleq |U|\frac{\sigma_H}{\sigma_W}$.

Proposition 1. *Let*

$$Y = |\tilde{H}X + \tilde{W}|^2 \tag{6}$$

where $\tilde{H} \sim \mathcal{CN}(0,1)$ and $\tilde{W} \sim \mathcal{CN}(0,1)$. Then, the capacity of the Rayleigh channel model in (1) with average-power and peak-power constraints as in (2) and (3) can be equivalently expressed as the capacity of the channel in (6):

$$C(\tilde{\mathsf{P}}, \tilde{\mathsf{A}}) = \max_{P_X:\, \mathbb{E}[X^2]\le \mathsf{P},\, X\le \mathsf{A}} I(X;Y) \triangleq C(\mathsf{P},\mathsf{A}), \tag{7}$$

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 capacity-achieving inputs and outputs of both channels can be related as follows:

$$|Y^\star| = \frac{|V^\star|^2}{\sigma_W^2}, \text{ and } |X^\star| = |U^\star|\frac{\sigma_H}{\sigma_W}. \tag{8}$$

Proof. See Appendix I. □

We now propose a channel model based on Prop. 1. This model will be useful for studying capacity indirectly.

Proposition 2. *The capacity of the channel model of (6) is equivalent to the capacity of the exponential model*

$$Y = \frac{1}{S}T \tag{9}$$

where $T \sim Exp(1)$ and S are independent random variables, with input constraints

$$\mathbb{E}\left[\frac{1}{S}\right] \le 1 + \mathsf{P}, \quad \frac{1}{1+\mathsf{A}^2} \le S \le 1. \tag{10}$$

The capacity is

$$C(\mathsf{P},\mathsf{A}) = \max_{P_S:\, \mathbb{E}\left[\frac{1}{S}\right]\le 1+\mathsf{P},\, S\in\left[\frac{1}{1+\mathsf{A}^2},1\right]} I(S;Y). \tag{11}$$

Proof. The channel transition law $f_{Y|X}$ is an exponential probability density function (pdf) of variable y with parameter $s = \frac{1}{1+x^2}$, i.e.,

$$f_{Y|S}(y|s) = s\exp\left(-sy\right), \qquad y \ge 0. \tag{12}$$

Hence, conditioned on $S = s$, we can write $Y = \frac{1}{s}T$ where $T \sim \mathrm{Exp}(1)$. The other results follow easily. □

B. KKT Conditions

We next introduce the Karush-Kuhn-Tucker (KKT) conditions for the optimality of the input distributions P_X and P_S of the Rayleigh and exponential models, respectively.

Lemma 1. *Let $P_{X^\star}$ be an input distribution of the Rayleigh channel model of Prop. 1, and let $f_{Y^\star}$ be the induced optimal output pdf. Then, $P_{X^\star}$ is capacity-achieving if and only if there exists $\lambda \ge 0$ such that*

$$D\left(f_{Y|X}(\cdot|x)\,\middle\|\, f_{Y^\star}\right) \le C(\mathsf{P},\mathsf{A}) + \lambda(x^2 - \mathsf{P}),\ x \in [0,\mathsf{A}], \tag{13}$$

$$D\left(f_{Y|X}(\cdot|x)\,\middle\|\, f_{Y^\star}\right) = C(\mathsf{P},\mathsf{A}) + \lambda(x^2 - \mathsf{P}),\ x \in \mathsf{supp}(P_{X^\star}) \tag{14}$$

Proof. See Proposition 2 and [4]. □

Lemma 2. *Let $P_{S^\star}$ be an input distribution of the exponential model of Prop. 2, 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*

$$D\left(f_{Y|S}(\cdot|s) \,\|\, f_{Y^\star}\right) \leq C(\mathsf{P},\mathsf{A}) + \lambda\left(\frac{1}{s} - 1 - \mathsf{P}\right),\ s \in \left[\frac{1}{1+\mathsf{A}^2}, 1\right], \quad (15)$$

$$D\left(f_{Y|S}(\cdot|s) \,\|\, f_{Y^\star}\right) = C(\mathsf{P},\mathsf{A}) + \lambda\left(\frac{1}{s} - 1 - \mathsf{P}\right),\ s \in \mathsf{supp}(P_{S^\star}). \quad (16)$$

Proof. See [4]. □

Note that the Lagrange multiplier λ that accounts for the average-power constraint is the same for the different formulations of the KKT conditions. Indeed, λ depends only on the capacity as

$$\lambda = \frac{\partial}{\partial \mathsf{P}} C(\mathsf{P},\mathsf{A}) \quad (17)$$

and it is a function of P and A.

Next, we introduce the following functions:

$$\Xi_R(x; f_{Y^\star}) \triangleq D\left(f_{Y|X}(\cdot|x) \,\|\, f_{Y^\star}\right) - C(\mathsf{P},\mathsf{A}) - \lambda(x^2 - \mathsf{P}), \quad (18)$$

$$\Xi_E(s; f_{Y^\star}) \triangleq D\left(f_{Y|S}(\cdot|s) \,\|\, f_{Y^\star}\right) - C(\mathsf{P},\mathsf{A}) - \lambda\left(\frac{1}{s} - 1 - \mathsf{P}\right), \quad (19)$$

that let us rewrite the KKT conditions as follows:

$$\text{Rayleigh Model: } \begin{cases} \Xi_R(x; f_{Y^\star}) \leq 0 & x \in [0, \mathsf{A}] \\ \Xi_R(x; f_{Y^\star}) = 0 & x \in \mathsf{supp}(P_{X^\star}) \end{cases} \quad (20)$$

$$\text{Exponential Model: } \begin{cases} \Xi_E(s; f_{Y^\star}) \leq 0 & s \in \left[\frac{1}{1+\mathsf{A}^2}, 1\right] \\ \Xi_E(s; f_{Y^\star}) = 0 & s \in \mathsf{supp}(P_{S^\star}) \end{cases} \quad (21)$$

While the KKT conditions refer to channels with the same capacity, certain proofs are easier with one or another model. We will convert and report the results in the Rayleigh model. We also distinguish between two different types of zeros.

Definition 1. *Suppose that $s^\star$ is a zero of $\Xi_E(s; f_{Y^\star})$. Then, $s^\star$ is called a* non-nodal zero *if and only if both of the following two conditions hold:*

- *$s^\star \in (0, 1)$ (i.e., $s^\star$ is not at the boundary of the support); and*
- *$\Xi'_E(s^\star; f_{Y^\star}) = 0$ (i.e., $s^\star$ is critical point) .*

Otherwise, $s^\star$ is referred to as nodal zero.

Remark 1. $s^\star = 1$ is a nodal zero.[1]

Remark 2. All internal global maxima of Ξ_E are non-nodal zeros, and we have $\Xi''_E(s^\star; f_{Y^\star}) \leq 0$.

Remark 3. The point $s^\star = \frac{1}{1+\mathsf{A}^2}$ may or may not be a support point; when it is, it may or may not be a nodal zero. For example, if $\mathsf{P} = \infty$, then $s^\star = \frac{1}{1+\mathsf{A}^2}$ is a nodal zero. In addition, if A^2 is much larger than P, then $s^\star = \frac{1}{1+\mathsf{A}^2}$ might not be a support point.

[1] In [4], it was shown that $s^\star = 1$ is a support point.

C. Why Peak Power is Needed

For non-coherent channels subject only to the average-power constraint, *flash signaling* achieves the capacity slope for $\mathsf{P} \to 0$ [27]. Flash signaling is not peak-power limited, i.e., the maximum support point $x^\star_{\max}$ converges to infinity as P goes to zero.

Proposition 3. *Let $P_{X^\star}$ be the capacity-achieving input distribution, and let*

$$x^\star_{\max}(\mathsf{P},\mathsf{A}) = \arg\max\{\mathsf{supp}(P_{X^\star})\} \quad (22)$$

be its largest amplitude. Then we have

$$\lim_{\mathsf{P}\to 0}\lim_{\mathsf{A}\to\infty} x^\star_{\max}(\mathsf{P},\mathsf{A}) = \infty, \quad \lim_{\mathsf{P}\to 0}\lim_{\mathsf{A}\to\infty} P_{X^\star}(x^\star_{\max}) = 0. \quad (23)$$

Proof. See Lemma 5 in Appendix II-A. □

On the other hand, for $\mathsf{A} \to \infty$ and for P as small as 10^{-6}, we compute $x^\star_{\max} \approx 2.6746$. A similar argument can be made for those P's at which a new mass point appears in $P_{X^\star}$. Since at those P's the evolution of the capacity-achieving input distribution is abrupt, considering a peak power constraint has several advantages. First, the constraint is physically reasonable and, for sufficiently large $\mathsf{A} < \infty$, the capacity is almost identical to the capacity with only the average-power constraint while being mathematically more tractable. However, many of our results will hold for $\mathsf{A} = \infty$.

IV. Main Results

The next theorem presents our main results. Let $x^\star_{\max} = \max \mathsf{supp}(P_{X^\star})$ and $\zeta = \frac{\mathbb{E}[(1+(X^\star)^2)^{-1}]}{\mathbb{E}[(1+(X^\star)^2)^{-2}]}$.

Theorem 1. *Suppose $\mathsf{P} \in (0, \infty]$ and $\mathsf{A} \in (0, \infty]$, and let $P_{X^\star}$ be the capacity-achieving input distribution. Then:*

1) (Small Average- and Peak-Power Regime)
- *If $\mathsf{A}^2 \leq \mathsf{P}$, then $\lambda = 0$ (i.e., the average-power constraint is not active) and $x^\star_{\max} = \mathsf{A}$;*
- *If $\mathsf{A}^2 < \sqrt{2}$, then $\lambda = 0$ or $\mathsf{supp}(P_{X^\star}) = \{0, \mathsf{A}\}$;*
- *If $\lambda = 0$, then $\mathsf{supp}(P_{X^\star}) = \{0, \mathsf{A}\}$ or $\mathsf{A}^2 \geq \frac{\mathrm{e}}{3-\mathrm{e}}$.*

2) (On the Location of Support Points)
- *$0 \in \mathsf{supp}(P_{X^\star})$;*
- *If the average-power constraint is active, then*

$$\mathsf{P} \leq \frac{\mathsf{P}}{1 - P_{X^\star}(0)} \leq (x^\star_{\max})^2 \leq \min\left\{\mathsf{A}^2, \frac{1+\mathsf{P}}{P_{X^\star}(x^\star_{\max})} - 1\right\}, \quad (24)$$

$$\zeta - 1 \leq (x^\star)^2, \qquad x^\star \in \mathsf{supp}(P_{X^\star}) \setminus \{0\}. \quad (25)$$

Also, if a non-nodal support point $x^\star \neq 0$ exists, then

$$\max\left\{\sqrt{2}, \frac{3}{2}\zeta - 1\right\} \leq (x^\star_{\max})^2, \quad (26)$$

$$\frac{\sqrt{2}}{1-\frac{1}{1+(x^\star_{\max})^2}}-1 \le (x^\star)^2, \quad x^\star \in \mathsf{supp}(P_{X^\star}) \setminus \{0\}. \tag{27}$$

- *If the average-power constraint is not active, then*

$$\zeta - 1 \le (x^\star)^2 \le \mathsf{A}^2, \qquad x^\star \in \mathsf{supp}(P_{X^\star}) \setminus \{0\}. \tag{28}$$

Also, if a non-nodal support point $x^\star \neq 0$ exists, then

$$(x^\star)^2 \le \frac{\mathsf{A}^2-1}{2}, \qquad x^\star \in \mathsf{supp}(P_{X^\star}) \setminus \{0\}. \tag{29}$$

3) (Upper Bounds on the Probabilities) *Let* $\tilde{\eta} = \frac{\mathsf{A}^2}{(1+\mathsf{A}^2)^{1+\frac{1}{\mathsf{A}^2}}}$. *Then:*

- *If the average-power constraint is not active:*

$$P_{X^\star}(x^\star) \le \mathrm{e}^{-\frac{C(\mathsf{P},\mathsf{A})}{\tilde{\eta}}}, \qquad x^\star \in \mathsf{supp}(P_{X^\star}); \tag{30}$$

- *If the average-power constraint is active:*

$$P_{X^\star}(x^\star_{\max}) \le \mathrm{e}^{-\frac{C(\mathsf{P},\mathsf{A})}{\tilde{\eta}}}. \tag{31}$$

Moreover, if $\mathsf{A} = \infty$, *then:*

$$P_{X^\star}(x^\star_{\max}) \le \mathrm{e}^{-\frac{C(\mathsf{P},\mathsf{A}) - \frac{(1+\mathsf{P})}{1+(x^\star_{\max})^2}+1}{\tilde{\eta}}} \le \mathrm{e}^{-\frac{C(\mathsf{P},\mathsf{A})}{\tilde{\eta}}}. \tag{32}$$

- *Location dependent bound:*

$$P_{X^\star}(x^\star) \le \frac{\min\{\mathsf{P},\mathsf{A}^2\}}{(x^\star)^2}, \qquad x^\star \in \mathsf{supp}(P_{X^\star}). \tag{33}$$

Moreover, if the average-power constraint is active, then

$$P_{X^\star}(0) \le 1 - \frac{\mathsf{P}}{(x^\star_{\max})^2} \le 1 - \frac{\mathsf{P}}{\mathsf{A}^2}. \tag{34}$$

4) (Lower Bounds on the Probabilities)

- *If the average-power constraint is active and there exists a non-nodal support point, then*

$$P_{X^\star}(0) \ge 1 - \frac{\min\{\mathsf{P},\mathsf{A}^2\}}{\sqrt{2}-1}; \tag{35}$$

- *If the average-power constraint is not active and there exists a non-nodal support point, then*

$$P_{X^\star}(x^\star_{\max}) \ge \frac{0.3}{(\mathsf{A}^2+1)\mathsf{A}^2\mathrm{e}^{\mathsf{A}^2+1}(1-\mathrm{e}^{-1})}. \tag{36}$$

Proof. See Appendix II. □

The results in Theorem 1 depend on the value $C(\mathsf{P},\mathsf{A})$ and on the Lagrange multiplier λ which are unknown. Therefore, it is useful to have upper and lower bounds.

Proposition 4. *Fix some* $\mathsf{P} \in (0,\infty]$ *and* $\mathsf{A} \in (0,\infty]$. *Then,*

$$\underline{C}(\mathsf{P_A}) \le C(\mathsf{P},\mathsf{A}) \le \overline{C}(\mathsf{P_A}), \tag{37}$$

where $\mathsf{P_A} := \min(\mathsf{P},\mathsf{A}^2)$ *and*

$$\overline{C}(\mathsf{P_A}) = \Big\{ \log(1+\mathsf{P_A}), \tag{38}$$

$$\log(1+\log(1+\mathsf{P_A})) + 1 + \Gamma\left(1 - \frac{1}{1+\log(1+\mathsf{P_A})}\right)\Big\},$$

$$\underline{C}(\mathsf{P_A}) = \log\left(\sqrt{\left(\mathrm{e}^{-\gamma-1}\log(\mathsf{P_A}+1)\right)^2+1}\right)$$
$$\ge \max\left\{0, \log(\log(1+\mathsf{P_A})) - \gamma - 1\right\}, \tag{39}$$

where Γ *is the Gamma function and* $\gamma \approx 0.577$ *is the Euler-Mascheroni constant. Moreover, the gap between the bounds is at most 4 nats.*

Proof. See Appendix II-E. □

We emphasize that Proposition 4 does not give the tightest possible bounds but shows how capacity scales in the large power regime.

Proposition 5. *If the average-power constraint is active, then:*

- $\mathsf{P} \mapsto \lambda$ *is a decreasing function of* P*;*
- *If* $\mathsf{A} < \infty$, *then*

$$0 \le \lambda \le \frac{C(\mathsf{P},\mathsf{A})}{\mathsf{P}} \le 1; \tag{40}$$

Also, if there exists a non-nodal support point $x^\star$, then

$$\lambda \ge \frac{3\mathrm{e}^{-1}}{1+(x^\star_{\max})^2} - \zeta(3\mathrm{e}^{-1}-1). \tag{41}$$

- *If* $\mathsf{A} = \infty$, *then*

$$0 \le \frac{1}{1+(x^\star_{\max})^2} \le \lambda \le \frac{C(\mathsf{P},\mathsf{A})}{\mathsf{P}} \le 1. \tag{42}$$

Proof. See Appendix II-B. □

A. Discussion

Fig. 1 subdivides the space spanned by P and A^2 into regions, each characterized by specific properties. The region $\mathcal{R}_0$ includes the pairs $(\mathsf{P},\mathsf{A}^2)$ outside of the Small Average- and Peak-Power Regime defined in Theorem 1. The union $\bigcup_{i=1}^3 \mathcal{R}_i$ constitutes the Small Average- and Peak-Power Regime region. To better understand when the average-power constraint is active, consider the following result.

Lemma 3. *Suppose* $\mathsf{supp}(P_{X^\star}) = \{0,\mathsf{A}\}$ *and define*

$$\rho := \mathbb{E}\left[\log f_Y(Y) \mid X=0\right] - \mathbb{E}\left[\log f_Y(Y) \mid X=\mathsf{A}\right]$$
$$- \log(1+\mathsf{A}^2) \tag{43}$$

where

$$f_Y(y) = \left(1-\frac{\mathsf{P}}{\mathsf{A}^2}\right)\mathrm{e}^{-y} + \frac{\mathsf{P}}{\mathsf{A}^2}\frac{1}{1+\mathsf{A}^2}\mathrm{e}^{-\frac{y}{1+\mathsf{A}^2}}, \qquad y \ge 0. \tag{44}$$

If $\rho > 0$, *then the average-power constraint is active with* $\lambda = \frac{\rho}{\mathsf{A}^2}$ *and* $P_{X^\star}(\mathsf{A}) = \frac{\mathsf{P}}{\mathsf{A}^2}$. *Otherwise, if* $\rho \le 0$, *the average-power constraint is not active* ($\lambda = 0$) *and* $P_{X^\star}(\mathsf{A})$ *is a solution to the equation in* (45).

Proof. See Appendix II-D4. □

By combining Theorem 1 and Lemma 3, the regions $\mathcal{R}_1$ to $\mathcal{R}_3$ are characterized by properties summarized in Table I.

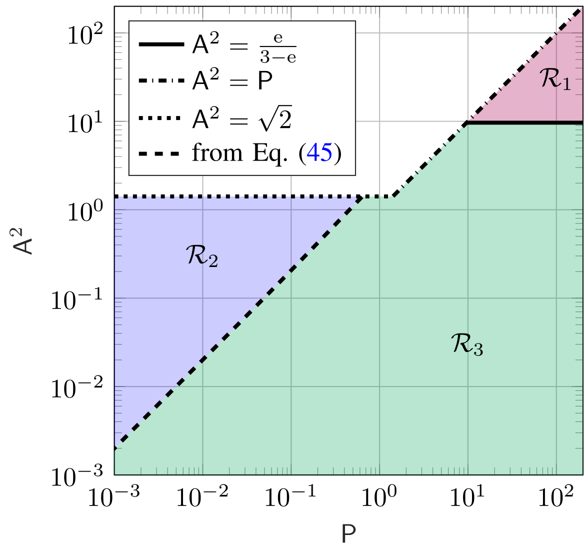

Fig. 1: Small average- and peak-power regime regions.

TABLE I: Properties of the small average- and peak-power regime regions.

Region	Average-power constraint	$P_{X^\star}$ properties
$\mathcal{R}_1$	OFF	$\max \mathsf{supp}(P_{X^\star}) = \mathsf{A}$
$\mathcal{R}_2$	ON	$\mathsf{supp}(P_{X^\star}) = \{0, \mathsf{A}\}$
$\mathcal{R}_3$	OFF	$\mathsf{supp}(P_{X^\star}) = \{0, \mathsf{A}\}$

B. Numerical Results for the Optimal Input Distribution

We provide further insights on $P_{X^\star}$ and its support for the regions $\mathcal{R}_0$–$\mathcal{R}_3$ of Fig. 1. For some values of A^2, we evaluated an estimate $\hat{P}_{X^\star}$ of the capacity-achieving distribution by adapting the numerical algorithm in [20] to the Rayleigh fading channel. The new algorithm iteratively updates a tentative pmf $\hat{P}_{X^\star}$ and stops when the KKT conditions in (20) are satisfied within a small tolerance ε, see [20, Sec. 5.1].

Fig. 2 shows the support of the pmf estimate $\hat{P}_{X^\star}$ vs. P, for $\mathsf{A}^2 = 1, 6.25, 100$. The size of the blue dots qualitatively shows the probability associated with each mass point. Comparing Fig. 2 to Fig. 1 shows that our definition of the small power regime regions $\mathcal{R}_i$'s is conservative, especially for $\mathcal{R}_1$, and that one could improve the characterization of $\mathcal{R}_1$ and $\mathcal{R}_3$. Moreover, as expected, for each A^2 the pmf estimates do not change in the regions with inactive average power constraint, i.e., in $\mathcal{R}_1$ and $\mathcal{R}_3$. Finally, for the wide range of points given by $\mathsf{P}, \mathsf{A}^2 \leq 200$, we have $|\mathsf{supp}(P_{X^\star})| = 2$.

V. Conclusion

We studied non-coherent Rayleigh channels and developed properties of the capacity-achieving input distribution based on the KKT conditions. Both average- and peak-power constraints were considered. Our main focus was on small, but non-vanishing, average- and peak powers. We showed that the capacity-achieving distribution is binary for sufficiently small values of the power constraints. We also derived upper and lower bounds on the optimal input probability values, provided insight into the mass point positions, and derived bounds on the channel capacity.

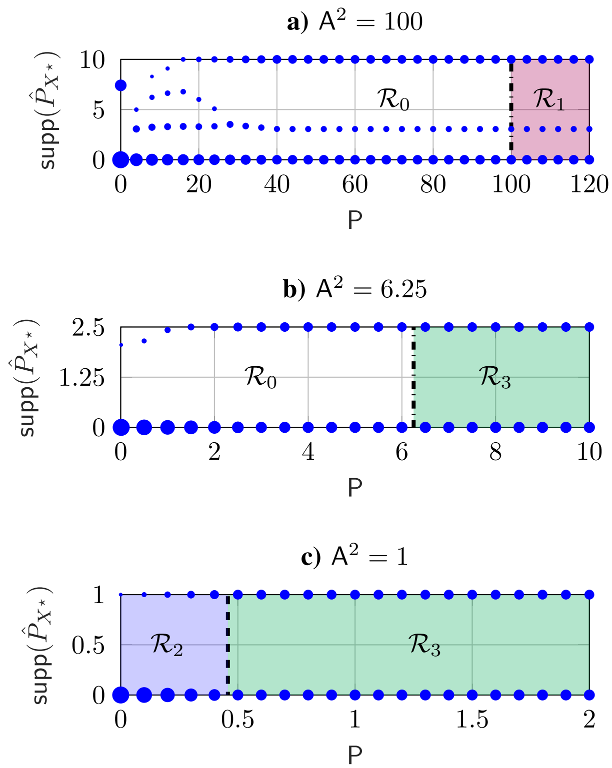

Fig. 2: Approximate optimal pmf support vs. P for $\mathsf{A}^2 = 100, 6.25, 1$.

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." 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.

$$\int_0^\infty \left(\frac{1}{1+\mathsf{A}^2}\mathrm{e}^{-\frac{y}{1+\mathsf{A}^2}} - \mathrm{e}^{-y}\right)\log\left((1-P_{X^\star}(\mathsf{A}))\mathrm{e}^{-y} + \frac{P_{X^\star}(\mathsf{A})}{(1+\mathsf{A}^2)}\mathrm{e}^{-\frac{y}{1+\mathsf{A}^2}}\right)\mathrm{d}y + \log(1+\mathsf{A}^2) = 0. \tag{45}$$

[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, 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.

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

[24] N. Sharma and S. Shamai, "Transition points in the capacity-achieving distribution for the peak-power limited AWGN and free-space optical intensity channels," *Problems of Inf. Transmiss.*, vol. 46, pp. 283–299, 2010.

[25] M. Raginsky, "On the information capacity of Gaussian channels under small peak power constraints," in *2008 46th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2008, pp. 286–293.

[26] A. Dytso, M. Goldenbaum, H. V. Poor, and S. Shamai, "Amplitude constrained MIMO channels: Properties of optimal input distributions and bounds on the capacity," *Entropy*, vol. 21, no. 2, 2019.

[27] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, 2002.

[28] G. Taricco and M. Elia, "Capacity of fading channel with no side information," *Electronics Letters*, vol. 33, no. 16, pp. 1368–1370, 1997.

[29] 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.

[30] 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.

[31] 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.

[32] G. Kramer, "Information rates for channels with fading, side information and adaptive codewords," *Entropy*, vol. 25, no. 5, 2023.

[33] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1997, vol. 11.

[34] J. Cohen, J. H. Kempermann, and G. Zbaganu, *Comparisons of stochastic matrices with applications in information theory, statistics, economics and population*. Springer Science & Business Media, 1998.

[35] A. Dytso, M. Cardone, and I. Zieder, "Meta derivative identity for the conditional expectation," *IEEE Trans. Inf. Theory*, 2023.

APPENDIX I
PROOF OF PROPOSITION 1

Following the lines of [5, Prop. 1], an input distribution with uniform phase is capacity-achieving. As a result, the capacity-achieving output distribution has a uniform phase as well. The mutual information can be written as

$$
\begin{aligned}
I(U;V) &= h(V) - h\left(V \,|\, U\right) && (46)\\
&= h\left(|V|^2\right) + \log(\pi) - h\left(V\mathrm{e}^{-jU} \,\middle|\, U\right) && (47)\\
&= h\left(\left|H|U| + W\right|^2\right) + \log(\pi) - h\left(H|U| + W \,|\, |U|\right) && (48)\\
&= h\left(\left|H|U| + W\right|^2\right) - h\left(\left|H|U| + W\right|^2 \,\middle|\, |U|\right) && (49)\\
&= I\left(|U|\,; |V|^2\right) && (50)
\end{aligned}
$$

where in (47) we used that for any circular symmetric complex random variable V we have $h(V) = h(|V|^2) + \log(\pi)$; in (48) we used the statistical independence between amplitude and phase of U; and in (49) we used the circular symmetry of the random variable $H|u| + W$.

By introducing $Y \triangleq \frac{|V|^2}{\sigma_W^2}$ and $X \triangleq |U|\frac{\sigma_H}{\sigma_W}$ we get the model

$$
Y = |\tilde{H}X + \tilde{W}|^2 \tag{51}
$$

where $\tilde{H} \triangleq \frac{H}{\sigma_H} \sim \mathcal{CN}(0,1)$ and $\tilde{W} \triangleq \frac{W}{\sigma_W} \sim \mathcal{CN}(0,1)$. From (51) we can see that $\sqrt{Y}$ given $X = x$ is a Rayleigh random variable, i.e.

$$
f_{\sqrt{Y}|X}(t|x) = \frac{2t}{1+x^2}\exp\left(-\frac{t^2}{1+x^2}\right), \qquad t \geq 0 \tag{52}
$$

hence we have that

$$
f_{Y|X}(y|x) = \frac{1}{1+x^2}\exp\left(-\frac{y}{1+x^2}\right), \qquad y \geq 0. \tag{53}
$$

The new average- and peak-power constraints read as

$$
\begin{aligned}
\mathbb{E}[X^2] &\leq \frac{\sigma_H^2}{\sigma_W^2}\tilde{\mathsf{P}} \triangleq \mathsf{P} && (54)\\
0 \leq X &\leq \frac{\sigma_H^2}{\sigma_W^2}\tilde{\mathsf{A}} \triangleq \mathsf{A} && (55)
\end{aligned}
$$

APPENDIX II
PROOF OF THEOREM 1

A. On the Location of Support Points

In this section, we show some results about the location of support points.

Lemma 4. *Let $P_{S^\star}$ be the optimal input of the exponential channel model. We have the following results:*

- *If the average-power constraint is active, then*

$$
\max\left\{\frac{1}{1+\mathsf{A}^2}, \frac{P_{S^\star}(s^\star_{\min})}{1+\mathsf{P}}\right\} \leq s^\star_{\min} \leq \frac{1}{1+\frac{\mathsf{P}}{1-P_{S^\star}(1)}} \leq \frac{1}{1+\mathsf{P}}, \tag{56}
$$

$$
s^\star \leq \frac{\mathbb{E}\left[(S^\star)^2\right]}{\mathbb{E}\left[S^\star\right]}, \qquad s^\star \neq 1. \tag{57}
$$

Moreover, if there exists a non-nodal support point $s^\star \neq 1$, then

$$
s^\star_{\min} \leq \min\left\{\frac{2}{3}\frac{\mathbb{E}\left[(S^\star)^2\right]}{\mathbb{E}\left[S^\star\right]}, \frac{1}{1+\sqrt{2}}\right\}, \tag{58}
$$

$$
s^\star \leq \frac{1-s^\star_{\min}}{\sqrt{2}}, \qquad s^\star \neq 1. \tag{59}
$$

- *If the average-power constraint is not active, then*

$$
\frac{1}{1+\mathsf{A}^2} = s^\star_{\min} \leq s^\star \leq \frac{\mathbb{E}\left[(S^\star)^2\right]}{\mathbb{E}\left[S^\star\right]}, \qquad s^\star \neq 1. \tag{60}
$$

Moreover, if there exists a non-nodal support point $s^\star \neq 1$, then

$$s^\star \geq \frac{2}{1+\mathsf{A}^2}, \qquad s^\star \neq 1. \tag{61}$$

Proof. First of all, note that it must be $s^\star_{\min} \geq \frac{1}{1+\mathsf{A}^2}$ in all cases, because the peak-power constraint cannot be violated. For the bounds in (56): Using the assumption that power constraint is active and, therefore, must be met with equality, we have that

$$1+\mathsf{P} = \mathbb{E}\left[\frac{1}{S^*}\right] \leq P_{S^\star}(1) + (1-P_{S^\star}(1))\frac{1}{s^\star_{\min}} \leq \frac{1}{s^\star_{\min}}. \tag{62}$$

Also, note that

$$1+\mathsf{P} = \mathbb{E}\left[\frac{1}{S^*}\right] \geq \frac{P_{S^\star}(s^\star_{\min})}{s^\star_{\min}}. \tag{63}$$

By using the result of Lemma 15 we have that

$$s^\star_{\min} \leq s^\star \leq \frac{\mathbb{E}\left[(S^\star)^2\right]}{\mathbb{E}\left[S^\star\right]} - \lambda \leq \frac{\mathbb{E}\left[(S^\star)^2\right]}{\mathbb{E}\left[S^\star\right]}. \tag{64}$$

For the first bound in (58): Suppose that there exists a non-nodal support point $s^\star \neq 1$ (i.e., such that $\Xi'_E(s^\star; f_{Y^\star}) = 0$ and $\Xi''_E(s^\star; f_{Y^\star}) \leq 0$).

Then, by using the result of Proposition 6, we have

$$0 = s^\star \mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] \mid S = s^\star\right] + \lambda \tag{65}$$

which gives a relationship between λ and $s^\star$. By using the condition on the second derivative, we have

$$\begin{aligned}
0 &\geq (s^\star)^2 \Xi''_E(s^\star; f_{Y^\star}) &(66)\\
&= -1 + s^\star \mathbb{E}\left[Y^2 \mathbb{E}\left[S^\star \mid Y\right] \,\middle|\, S = s^\star\right] - \frac{2\lambda}{s^\star} &(67)\\
&= -1 + s^\star \mathbb{E}\left[Y^2 \mathbb{E}\left[S^\star \mid Y\right] \,\middle|\, S = s^\star\right] + 2 - 2\mathbb{E}\left[Y\mathbb{E}\left[S^\star \mid Y\right] \mid S = s^\star\right] &(68)
\end{aligned}$$

where (67) follows from Proposition 6; and step (68) from the equality in (65). Solving (68) for $s^\star$ yields

$$\begin{aligned}
s^\star &\leq \frac{2\mathbb{E}\left[Y\mathbb{E}\left[S^\star \mid Y\right] \mid S = s^\star\right] - 1}{\mathbb{E}\left[Y^2\mathbb{E}\left[S^\star \mid Y\right] \mid S = s^\star\right]} &(69)\\
&\leq \frac{2\frac{\mathbb{E}\left[(S^\star)^2\right]}{\mathbb{E}[S^\star]}\mathbb{E}\left[Y \mid S = s^\star\right] - 1}{s^\star_{\min}\mathbb{E}\left[Y^2 \mid S = s^\star\right]} &(70)\\
&= \frac{2\frac{\mathbb{E}\left[(S^\star)^2\right]}{\mathbb{E}[S^\star]}\frac{1}{s^\star} - 1}{s^\star_{\min}\frac{2}{(s^\star)^2}} &(71)
\end{aligned}$$

where (70) follows from Lemma 16; and the last step holds by $\mathbb{E}\left[Y \mid S = s^\star\right] = \frac{1}{s^\star}$ and $\mathbb{E}\left[Y^2 \,\middle|\, S = s^\star\right] = \frac{2}{(s^\star)^2}$. Solving (71) for $s^\star$ gives

$$s^\star + 2s^\star_{\min} \leq 2\frac{\mathbb{E}\left[(S^\star)^2\right]}{\mathbb{E}\left[S^\star\right]}. \tag{72}$$

Next, using that $s^\star \geq s^\star_{\min}$, we get

$$s^\star_{\min} \leq \frac{2}{3}\frac{\mathbb{E}\left[(S^\star)^2\right]}{\mathbb{E}\left[S^\star\right]}. \tag{73}$$

This concludes the proof of the first bound in (58).

To show the bound in (59), we use the alternative expression of $\Xi''_E(s^\star; f_{Y^\star})$ from Proposition 6:

$$\begin{aligned}
0 &\geq (s^\star)^2 \Xi''_E(s^\star; f_{Y^\star}) &(74)\\
&= -\mathbb{E}\left[1 - 2Y\mathbb{E}\left[S^\star \mid Y\right] + Y^2 \mathsf{Var}\left[S^\star \mid Y\right] \,\middle|\, S = s^\star\right] - \frac{2\lambda}{s^\star} &(75)\\
&= -\mathbb{E}\left[1 - 2Y\mathbb{E}\left[S^\star \mid Y\right] + Y^2 \mathsf{Var}\left[S^\star \mid Y\right] \,\middle|\, S = s^\star\right] + 2\mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] \mid S = s^\star\right] &(76)\\
&= \mathbb{E}\left[1 - Y^2 \mathsf{Var}\left[S^\star \mid Y\right] \,\middle|\, S = s^\star\right] &(77)\\
&\geq 1 - \frac{(1-s^\star_{\min})^2}{4}\mathbb{E}\left[Y^2 \,\middle|\, S = s^\star\right] &(78)
\end{aligned}$$

$$= 1 - \frac{(1-s^\star_{\min})^2}{2(s^\star)^2} \tag{79}$$

where (75) follows from Proposition 6; step (76) from the equality in (65); step (78) from Popoviciu's inequality $\mathsf{Var}\left[S^\star \mid Y\right] \leq \frac{(1-s^\star_{\min})^2}{4}$ because $S^\star \in [s^\star_{\min}, 1]$; and the last step holds by $\mathbb{E}\left[Y^2 \mid S = s\right] = \frac{2}{s^2}$. Solving inequality (79) gives

$$s^\star \leq \frac{1-s^\star_{\min}}{\sqrt{2}}. \tag{80}$$

To show the second bound in (58), just use $s^\star = s^\star_{\min}$ in (80).

If the average-power constraint is not active (*i.e.*, $\lambda = 0$), then the input is peak-power limited and by using the same argument to prove part iii) of Lemma 9, we can conclude that $s^\star_{\min} = \frac{1}{1+\mathsf{A}^2}$. Using the result of Lemma 15 we have that

$$s^\star_{\min} \leq s^\star \leq \frac{\mathbb{E}\left[(S^\star)^2\right]}{\mathbb{E}\left[S^\star\right]}. \tag{81}$$

To show the bound in (61): If there exists a non-nodal support point $s^\star \neq 1$, then by using $\lambda = 0$ we can write

$$\begin{aligned} 0 &\geq (s^\star)^2 \Xi''_E(s^\star; f_{Y^\star}) && (82)\\ &= -1 + s^\star \mathbb{E}\left[Y^2 \mathbb{E}\left[S^\star \mid Y\right] \,\middle|\, S = s^\star\right] && (83)\\ &\geq -1 + \frac{2}{s^\star}\frac{1}{1+\mathsf{A}^2} && (84)\end{aligned}$$

where in the last inequality we have used $\mathbb{E}\left[S^\star \mid Y\right] \geq s^\star_{\min} = \frac{1}{1+\mathsf{A}^2}$ and then $\mathbb{E}\left[Y^2 \mid S = s\right] = \frac{2}{s^2}$. By solving inequality (84) in $s^\star$ we get (61). □

Next, we show that flash signaling is optimal. First, we restate Proposition 3 under the exponential channel model and then prove the claim.

Lemma 5. *In absence of a peak-power constraint, flash signaling is optimal in the limit of small* P*, i.e.,*

$$\lim_{\mathsf{P}\to 0}\lim_{\mathsf{A}\to\infty} s^\star_{\min}(\mathsf{P},\mathsf{A}) = 0, \qquad \lim_{\mathsf{P}\to 0}\lim_{\mathsf{A}\to\infty} P_{S^\star}(s^\star_{\min}) = 0. \tag{85}$$

Proof. Using the bound of Lemma 6 together with the bound $s^\star \leq \frac{1}{\sqrt{2}}$ shown in Lemma 4, we have

$$\lim_{\mathsf{P}\to 0} P_{S^\star}(s^\star) \leq \lim_{\mathsf{P}\to 0} \frac{\mathsf{P}}{\sqrt{2}-1} = 0, \qquad \forall s^\star \in \mathsf{supp}(P_{S^\star}) \setminus \{1\}, \tag{86}$$

which, in turn, implies $\lim_{\mathsf{P}\to 0} P_{S^\star}(1) = 1$.

Consider now the values $\Xi_E(s^\star; f_{Y^\star})$ for $s^\star = 1$ and $s^\star = s^\star_{\min}$ and equate them to find:

$$\lambda\left(\frac{1}{s^\star_{\min}} - 1\right) + \log\frac{1}{s^\star_{\min}} \overset{!}{=} \int_0^\infty (\mathrm{e}^{-y} - s^\star_{\min}\mathrm{e}^{-s^\star_{\min}y}) \log f_{Y^\star}(y)\mathrm{d}y. \tag{87}$$

Towards a contradiction, assume that $s^\star_{\min} \geq a > 0$ as $\mathsf{P} \to 0$. Then, by taking the limit for $\mathsf{P} \to 0$ of (87), we have

$$\begin{aligned} \frac{1}{s^\star_{\min}} - 1 + \log\frac{1}{s^\star_{\min}} &\overset{!}{=} \int_0^\infty (\mathrm{e}^{-y} - s^\star_{\min}\mathrm{e}^{-s^\star_{\min}y}) \log \mathrm{e}^{-y}\mathrm{d}y && (88)\\ &= \frac{1}{s^\star_{\min}} - 1, && (89)\end{aligned}$$

which implies $\lim_{\mathsf{P}\to 0} s^\star_{\min} = 1$. This clearly contradicts the fact that $s^\star_{\min} \leq \frac{1}{\sqrt{2}}$ shown in Lemma 4. Hence, we must have $\lim_{\mathsf{P}\to 0} s^\star_{\min} = 0$. This concludes the proof. □

B. Bounds on the Lagrange Multiplier λ: Proof of Proposition 5

Proof. By the envelope theorem [33], we know that

$$\lambda = \frac{\partial}{\partial \mathsf{P}} C(\mathsf{P},\mathsf{A}). \tag{90}$$

Furthermore, since the capacity is a strictly concave function, when average power is active, we have that $\mathsf{P} \mapsto \lambda$ is a decreasing function of P.

To show the upper bound, we know that there is a point mass at $s^\star = 1$ and that the KKT condition (16) can be written as

$$\lambda\mathsf{P} = C(\mathsf{P},\mathsf{A}) - D\left(f_{Y|S}(\cdot|1) \,\middle\|\, f_{Y^\star}\right) \leq C(\mathsf{P},\mathsf{A}). \tag{91}$$

As for the lower bound when $\mathsf{A} < \infty$: from the expressions (219) and (221) of the first and second derivatives of function Ξ_E presented in Proposition 6, we can also write that

$$s^2\Xi'_E(s; f_{Y^\star}) = -s^3\Xi''_E(s, f_{Y^\star}) - \lambda + \mathbb{E}\left[\left((sY)^2 - sY\right)\mathbb{E}\left[S^\star \mid Y\right] \,\middle|\, S = s\right]. \tag{92}$$

Now, suppose there exists a non-nodal support point $s^\star$ (see Definition 1), then by using $\Xi'_E(s^\star; f_{Y^\star}) = 0$ and $\Xi''_E(s^\star, f_{Y^\star}) \leq 0$ we can write

$$\begin{aligned}
0 &\geq -\lambda + \mathbb{E}\left[s^\star Y\left(s^\star Y - 1\right)\mathbb{E}\left[S^\star \mid Y\right] \mid S = s^\star\right] & (93)\\
&= -\lambda + \mathbb{E}\left[s^\star Y\left(s^\star Y - 1\right)\mathbb{E}\left[S^\star \mid Y\right]\left(\mathbb{1}(s^\star Y - 1 \geq 0) + \mathbb{1}(s^\star Y - 1 < 0)\right) \mid S = s^\star\right] & (94)\\
&\geq -\lambda + s^\star_{\min}\mathbb{E}\left[s^\star Y\left(s^\star Y - 1\right)\mathbb{1}(s^\star Y - 1 \geq 0) \mid S = s^\star\right] + \frac{\mathbb{E}\left[(S^\star)^2\right]}{\mathbb{E}\left[S^\star\right]}\mathbb{E}\left[s^\star Y\left(s^\star Y - 1\right)\mathbb{1}(s^\star Y - 1 < 0) \mid S = s^\star\right] & (95)\\
&= -\lambda + s^\star_{\min}3\mathrm{e}^{-1} - \frac{\mathbb{E}\left[(S^\star)^2\right]}{\mathbb{E}\left[S^\star\right]}(3\mathrm{e}^{-1} - 1), & (96)
\end{aligned}$$

where in (95) we used the bounds $s^\star_{\min} \leq \mathbb{E}\left[S^\star \mid Y\right] \leq \frac{\mathbb{E}\left[(S^\star)^2\right]}{\mathbb{E}\left[S^\star\right]}$ of Lemma 16.

As for the lower bound on λ when there is no peak-power constraint, *i.e.*, $\mathsf{A} = \infty$: Thanks to the KKT conditions for the exponential channel model we necessarily have $\Xi_E(s; f_{Y^\star}) \leq \Xi_E(1; f_{Y^\star}) = 0$ for all $s \in (0, 1]$. Then, we can write

$$\begin{aligned}
0 &\geq \lim_{s \to 0^+} \Xi_E(s; f_{Y^\star}) & (97)\\
&= \lim_{s \to 0^+} -\log\frac{1}{s} - \frac{\lambda}{s} + \lambda(1 + \mathsf{P}) - C - 1 - \mathbb{E}\left[\log f_{Y^\star}(Y) \mid S = s\right] & (98)\\
&= \lim_{s \to 0^+} -\log\frac{1}{s} - \frac{\lambda}{s} + \lambda(1 + \mathsf{P}) - C - 1 - \log f_{Y^\star}(0) + \frac{1}{s}\mathbb{E}\left[\mathbb{E}\left[S^\star \mid Y\right] \mid S = s\right] & (99)\\
&\geq \lim_{s \to 0^+} -\log\frac{1}{s} + \frac{s^\star_{\min} - \lambda}{s} + \lambda(1 + \mathsf{P}) - C - 1 - \log f_{Y^\star}(0) & (100)
\end{aligned}$$

where in (99) we used the first identity of Lemma 14 and the first identity of Lemma 13; and in (100) we used $\mathbb{E}\left[S^\star \mid Y\right] \geq s^\star_{\min}$. The limit (100) is negative if and only if $s^\star_{\min} \leq \lambda$. $\square$

C. Bounds on the Probability Values

In this section, by using different techniques, we propose bounds on probability values. Let us first introduce a result that gives tight upper bounds in the small average- and peak-power regime.

Lemma 6. *For $s^\star \in \mathsf{supp}(P_{S^\star}) \setminus \{1\}$ we have*

$$P_{S^\star}(s^\star) \leq \frac{\min\{\mathsf{P}, \mathsf{A}^2\}}{\frac{1}{s^\star} - 1}. \tag{101}$$

If the average-power constraint is active, then

$$P_{S^\star}(1) \leq 1 - \frac{\mathsf{P}}{\frac{1}{s^\star_{\min}} - 1} \leq 1 - \frac{\mathsf{P}}{\mathsf{A}^2}. \tag{102}$$

In addition, if the average-power constraint is active and there exists a non-nodal (see Definition 1) support point, then

$$P_{S^\star}(1) \geq 1 - \frac{\min\{\mathsf{P}, \mathsf{A}^2\}}{\sqrt{2} - 1}. \tag{103}$$

Proof. To prove the upper bound for $s^\star < 1$, starting from the peak- and average-power constraints, we have

$$\begin{aligned}
1 + \min\{\mathsf{P}, \mathsf{A}^2\} &\geq \mathbb{E}\left[\frac{1}{S^\star}\right] & (104)\\
&\geq 1 - P_{S^\star}(s^\star) + \frac{1}{s^\star}P_{S^\star}(s^\star) & (105)
\end{aligned}$$

where we used the bound $s \leq 1$ for all $s \neq s^\star$.

The upper bound on $P_{S^\star}(1)$ is a direct consequence of (62), the average-power constraint $\mathbb{E}\left[\frac{1}{S^\star}\right] = \mathsf{P} + 1$, and of $s^\star_{\min} \geq \frac{1}{1+\mathsf{A}^2}$.

The lower bound on $P_{S^\star}(1)$ can be obtained as follows:

$$1 + \min\{\mathsf{P}, \mathsf{A}^2\} \geq \mathbb{E}\left[\frac{1}{S^\star}\right] \tag{106}$$

$$\geq P_{S^\star}(1) + \sqrt{2}\,(1 - P_{S^\star}(1)) \tag{107}$$

where we used the bound $s^\star \leq \frac{1}{\sqrt{2}}$ for all $s^\star < 1$ from (59). Solving the inequality for $P_{S^\star}(1)$ gives the sought result. $\square$

Next, we introduce two more techniques for bounding probability values. Both of these techniques generalize the methods proposed in [22] to the case of the input average-power constraint. The first technique will rely on the exact characterization of the probability masses, and the second technique will rely on the strong data processing inequality.

Theorem 2. *Suppose that $s^\star \in \mathsf{supp}(P_{S^\star})$. Then,*

$$P_{S^\star}(s^\star) = \mathrm{e}^{-C(\mathsf{P},\mathsf{A}) + \lambda\left(1+\mathsf{P}-\frac{1}{s^\star}\right) + \mathbb{E}\left[\log P_{S^\star|Y^\star}(s^\star|Y)\;\middle|\;S=s^\star\right]}. \tag{108}$$

Proof. By using the KKT equality condition (16), write

$$C(\mathsf{P},\mathsf{A}) = \mathbb{E}\left[\log \frac{f_{Y|S}(Y|s^\star)}{f_{Y^\star}(Y)}\;\middle|\;S = s^\star\right] + \lambda\left(1 + \mathsf{P} - \frac{1}{s^\star}\right) \tag{109}$$

$$= \mathbb{E}\left[\log P_{S^\star|Y^\star}(s^\star|Y)\;\middle|\;S = s^\star\right] - \log P_{S^\star}(s^\star) + \lambda\left(1 + \mathsf{P} - \frac{1}{s^\star}\right) \tag{110}$$

which gives result (108). This concludes the proof. $\square$

We now show how the above results can be used to bound the probabilities.

Corollary 1. *Suppose that the average-power constraint is not active, then*

$$P_{S^\star}(s^\star) \leq \mathrm{e}^{-C(\mathsf{P},\mathsf{A})}, \qquad s^\star \in \mathsf{supp}(P_{S^\star}). \tag{111}$$

Suppose that the average-power constraint is active, then

$$P_{S^\star}(s^\star_{\min}) \leq \mathrm{e}^{-C(\mathsf{P},\mathsf{A})}. \tag{112}$$

In addition, if $\mathsf{A} = \infty$, then

$$P_{S^\star}(s^\star_{\min}) \leq \mathrm{e}^{-C(\mathsf{P},\infty) + s^\star_{\min}\left(1+\mathsf{P}-\frac{1}{s^\star_{\min}}\right)}. \tag{113}$$

Proof. First, note that by using (108) and the bound $P_{S^\star|Y^\star}(s^\star|Y) \leq 1$, we arrive at

$$P_{S^\star}(s^\star) \leq \mathrm{e}^{-C(\mathsf{P},\mathsf{A}) + \lambda\left(1+\mathsf{P}-\frac{1}{s^\star}\right)}. \tag{114}$$

Now, if the average-power constraint is not active, then $\lambda = 0$ and the proof follows.

If the average-power constraint is active, then $\lambda > 0$, and the thesis follows from the inequality

$$\lambda\left(1 + \mathsf{P} - \frac{1}{s^\star_{\min}}\right) \leq 0 \tag{115}$$

which is a consequence of $s^\star_{\min} \leq \frac{1}{1+\mathsf{P}}$ from Lemma 4. The last result in (113) holds by additionally using the inequality $s^\star_{\min} \leq \lambda$ in Prop. 5 which uses the assumption $\mathsf{A} = \infty$:

$$\lambda\left(1 + \mathsf{P} - \frac{1}{s^\star_{\min}}\right) \leq s^\star_{\min}\left(1 + \mathsf{P} - \frac{1}{s^\star_{\min}}\right). \tag{116}$$

$\square$

Next, we show an approach that relies on strong-data processing inequality. Recall, that the classical *data-processing inequality* for the relative entropy states that for $Q_X \to Q_{Y|X} \to Q_Y$ and $P_X \to Q_{Y|X} \to P_Y$, we have

$$D\left(P_Y \,\|\, Q_Y\right) \leq D\left(P_X \,\|\, Q_X\right). \tag{117}$$

The *strong data-processing inequality* is the sharpening of (117) where one seeks to find a coefficient $0 < \eta \leq 1$ such that

$$D\left(P_Y \,\|\, Q_Y\right) \leq \eta D\left(P_X \,\|\, Q_X\right). \tag{118}$$

It is not difficult to see that

$$\eta \leq \eta_{\mathsf{KL}}(\mathcal{X}, Q_{Y|X}) := \sup_{\substack{Q_X, P_X:\\ \mathsf{supp}(Q_X)\subseteq\mathcal{X}, \mathsf{supp}(P_X)\subseteq\mathcal{X}\\ 0<D(P_X\|Q_X)<\infty\\ Q_X\to Q_{Y|X}\to Q_Y\\ P_X\to Q_{Y|X}\to P_Y}} \frac{D\left(P_Y \,\|\, Q_Y\right)}{D\left(P_X \,\|\, Q_X\right)}. \tag{119}$$

The quantity $0 < \eta_{\mathsf{KL}}(\mathcal{X}, Q_{Y|X}) \leq 1$ is known as the contraction coefficients.

The next result provides an upper bound on the contraction coefficient for the exponential channel.

Lemma 7. *Let $P_{Y|S}$ be an exponential channel. Then, for all $\mathsf{A} \geq 0$*

$$\eta_{\mathsf{KL}}\left(\left[\frac{1}{1+\mathsf{A}^2}, 1\right]; P_{Y|S}\right) \leq \frac{\mathsf{A}^2}{(1+\mathsf{A}^2)^{1+\frac{1}{\mathsf{A}^2}}}. \tag{120}$$

Proof. As was shown in [34, Proposition II.4.10], the contraction coefficient is upper-bounded by

$$\eta_{\mathsf{KL}}\left(\left[\frac{1}{1+\mathsf{A}^2}, 1\right]; P_{Y|S}\right) \leq \sup_{s,s':\ \frac{1}{1+\mathsf{A}^2} \leq s \leq 1,\ \frac{1}{1+\mathsf{A}^2} \leq s' \leq 1} \mathsf{TV}(P_{Y|S=s} \| P_{Y|S=s'}) \tag{121}$$

where in (121) $\mathsf{TV}(P_{Y|X=x} \| P_{Y|X=x'})$ is the total variation distance.

Now note that

$$\mathsf{TV}(P_{Y|S=s} \| P_{Y|S=s'}) = \frac{1}{2}\int_0^\infty |s\mathrm{e}^{-ys} - s'\mathrm{e}^{-ys'}| \mathrm{d}y \tag{122}$$

$$= a^{-\frac{1}{a-1}} \left|1 - \frac{1}{a}\right| \tag{123}$$

where $a = \frac{s'}{s}$. The function $a \mapsto a^{-\frac{1}{a-1}} \left|1 - \frac{1}{a}\right|$ takes the maximum value at the boundaries of the support given by $a = \frac{1}{\frac{1}{1+\mathsf{A}^2}} = 1+\mathsf{A}^2$ or $a = \frac{\frac{1}{1+\mathsf{A}^2}}{1} = \frac{1}{1+\mathsf{A}^2}$ and the maximum is given by

$$\max_{\frac{1}{1+\mathsf{A}^2} \leq a \leq 1+\mathsf{A}^2} a^{-\frac{1}{a-1}} \left|1 - \frac{1}{a}\right| = \frac{\mathsf{A}^2}{(1+\mathsf{A}^2)^{1+\frac{1}{\mathsf{A}^2}}}. \tag{124}$$

This concludes the proof. □

We are now ready to show the new upper bound.

Theorem 3. *Suppose that $s \in \mathsf{supp}(P_{S^\star})$, then*

$$P_{S^\star}(s) \leq \mathrm{e}^{-\frac{C(\mathsf{P},\mathsf{A}) + \lambda\left(\frac{1}{s} - 1 - \mathsf{P}\right)}{\tilde{\eta}}}, \tag{125}$$

where $\tilde{\eta} = \frac{\mathsf{A}^2}{(1+\mathsf{A}^2)^{1+\frac{1}{\mathsf{A}^2}}}$. Moreover,

- *If the average-power constraint is not active:*

$$P_{S^\star}(s) \leq \mathrm{e}^{-\frac{C(\mathsf{P},\mathsf{A})}{\tilde{\eta}}} \tag{126}$$

- *If the average-power constraint is active:*

$$P_{S^\star}(s^\star_{\min}) \leq \mathrm{e}^{-\frac{C(\mathsf{P},\mathsf{A})}{\tilde{\eta}}} \tag{127}$$

  Moreover, if $\mathsf{A} = \infty$, then

$$P_{S^\star}(s^\star_{\min}) \leq \mathrm{e}^{-\frac{C(\mathsf{P},\mathsf{A}) - s^\star_{\min}\left(1+\mathsf{P} - \frac{1}{s^\star_{\min}}\right)}{\tilde{\eta}}} \tag{128}$$

Proof. Let Y_s be the output of the channel $P_{Y|S}$ when the input is $P_S = \delta_s$, where δ_s is the Dirac delta measure centered in s. Suppose that $s \in \mathsf{supp}(P_{S^\star})$ we have that

$$C(\mathsf{P},\mathsf{A}) + \lambda\left(\frac{1}{s} - 1 - \mathsf{P}\right) = D\left(f_{Y|S}(\cdot|s) \,\|\, f_{Y^\star}\right) \tag{129}$$

$$= D\left(f_{Y_s} \,\|\, f_{Y^\star}\right) \tag{130}$$

$$\leq \eta_{\mathsf{KL}}\, D\left(\delta_s \,\|\, P_{S^\star}\right) \tag{131}$$

$$= \eta_{\mathsf{KL}} \log \frac{1}{P_{S^\star}(s)} \tag{132}$$

where in (131) we have noted that $\delta_s \to P_{Y|S} \to P_{Y_s}$ and $P_{S^\star} \to P_{Y|S} \to P_{Y^\star}$ and used the strong data-processing inequality in (118) with $\eta_{\mathsf{KL}} = \eta_{\mathsf{KL}}\left(\left[\frac{1}{1+\mathsf{A}^2}, 1\right]; P_{Y|S}\right)$.

Now, rearranging the terms we have that

$$P_{S^\star}(s) \le \mathrm{e}^{-\frac{C(\mathsf{P},\mathsf{A})+\lambda\left(\frac{1}{s}-1-\mathsf{P}\right)}{\eta_{\mathsf{KL}}}}, \qquad s \in \mathsf{supp}(P_{S^\star}). \tag{133}$$

The proof is now concluded by following the same steps as in Corollary 1 and also using the bound on η_{KL} in Lemma 7. $\square$

Finally, we propose a technique for computing a lower bound to $P_{S^\star}(s^\star_{\min})$.

Lemma 8. *Let $\lambda = 0$ and suppose that there exists a non-nodal support point. Then,*

$$P_{S^\star}(s^\star_{\min}) \ge \frac{0.3}{(\mathsf{A}^2+1)\mathsf{A}^2\mathrm{e}^{\mathsf{A}^2+1}(1-\mathrm{e}^{-1})}. \tag{134}$$

Proof. Start from the fact that the second derivative of function $s \mapsto \Xi_E(s; f_{Y^\star})$ must be negative for a non-nodal support point $s^\star$, i.e.,

$$\begin{aligned}
0 &\ge (s^\star)^2 \Xi_E''(s^\star; f_{Y^\star}) && (135)\\
&= -1 + s^\star \mathbb{E}\left[Y^2 \mathbb{E}\left[S^\star \mid Y\right] \,\middle|\, S = s^\star\right] && (136)\\
&= -1 + s^\star \mathbb{E}\left[Y^2 \mathbb{E}\left[S^\star \mid Y\right] (\mathbb{1}(Y < c) + \mathbb{1}(Y \ge c)) \,\middle|\, S = s^\star\right] && (137)\\
&\ge -1 + s^\star s^\star_{\min} \mathbb{E}\left[Y^2 \mathbb{1}(Y \ge c) \,\middle|\, S = s^\star\right] + s^\star \mathbb{E}\left[Y^2 \mathbb{E}\left[S^\star \mid Y\right] \mathbb{1}(Y < c) \,\middle|\, S = s^\star\right] && (138)\\
&= -1 + \frac{s^\star_{\min}}{s^\star} \mathrm{e}^{-cs^\star}(2 + cs^\star(cs^\star+2)) + s^\star \mathbb{E}\left[Y^2 \mathbb{E}\left[S^\star \mid Y\right] \mathbb{1}(Y < c) \,\middle|\, S = s^\star\right], && (139)
\end{aligned}$$

where $c > 0$ is a real number; and (138) follows from $\mathbb{E}\left[S^\star \mid Y\right] \ge s^\star_{\min}$. Let $s^\star_2 := \min\{\mathsf{supp}(P_{S^\star}) \setminus \{s^\star_{\min}\}\}$ be the second smallest support point. Then, we lower-bound the conditional expectation as follows:

$$\begin{aligned}
\mathbb{E}\left[S^\star \mid Y\right] &= \mathbb{E}\left[S^\star(\mathbb{1}(S = s^\star_{\min}) + \mathbb{1}(S \ge s^\star_2)) \mid Y\right] && (140)\\
&\ge s^\star_{\min} P_{S^\star}(s^\star_{\min}) \frac{s^\star_{\min}\mathrm{e}^{-s^\star_{\min}Y}}{f_{Y^\star}(Y)} + s^\star_2 \Pr(S \ge s^\star_2 | Y) && (141)\\
&= s^\star_2 - (s^\star_2 - s^\star_{\min}) P_{S^\star}(s^\star_{\min}) \frac{s^\star_{\min}\mathrm{e}^{-s^\star_{\min}Y}}{f_{Y^\star}(Y)} && (142)
\end{aligned}$$

where in the last step we used $\Pr(S \ge s^\star_2 | Y) = 1 - \Pr(S = s^\star_{\min} | Y)$. Using (142) into (139) yields

$$\begin{aligned}
0 \ge -1 &+ \frac{s^\star_{\min}}{s^\star} \mathrm{e}^{-cs^\star}(2 + cs^\star(cs^\star+2)) + \frac{s^\star_2}{s^\star}\left(2 - \mathrm{e}^{-cs^\star}(2 + cs^\star(cs^\star+2))\right)\\
&- P_{S^\star}(s^\star_{\min}) s^\star (s^\star_2 - s^\star_{\min}) \mathbb{E}\left[Y^2 \frac{s^\star_{\min}\mathrm{e}^{-s^\star_{\min}Y}}{f_{Y^\star}(Y)} \mathbb{1}(Y < c) \,\middle|\, S = s^\star\right]. \qquad (143)
\end{aligned}$$

An upper bound to the expectation is obtained by letting

$$f_{Y^\star}(Y)\mathbb{1}(Y < c) \ge \mathbb{E}\left[S^\star \mathrm{e}^{-S^\star c}\right] \mathbb{1}(Y < c) \ge s^\star_{\min}\mathrm{e}^{-c}\mathbb{1}(Y < c), \tag{144}$$

thus obtaining

$$\begin{aligned}
\mathbb{E}\left[Y^2 \frac{s^\star_{\min}\mathrm{e}^{-s^\star_{\min}Y}}{f_{Y^\star}(Y)} \mathbb{1}(Y < c) \,\middle|\, S = s^\star\right] &\le \mathbb{E}\left[Y^2 \mathrm{e}^{c - s^\star_{\min}Y} \mathbb{1}(Y < c) \,\middle|\, S = s^\star\right] && (145)\\
&\le c^2 \mathrm{e}^c \Pr(Y < c | S = s^\star) && (146)\\
&= c^2 \mathrm{e}^c (1 - \mathrm{e}^{-s^\star c}). && (147)
\end{aligned}$$

By choosing $c = \frac{1}{s^\star_{\min}}$ and $s^\star = s^\star_2$ and by solving inequality (143) for $P_{S^\star}(s^\star_{\min})$, we get

$$P_{S^\star}(s^\star_{\min}) \ge \frac{1 - \left(1 - \frac{s^\star_{\min}}{s^\star_2}\right) \mathrm{e}^{-\frac{s^\star_2}{s^\star_{\min}}} (2 + \frac{s^\star_2}{s^\star_{\min}}(\frac{s^\star_2}{s^\star_{\min}} + 2))}{\frac{s^\star_2}{s^\star_{\min}}(\frac{s^\star_2}{s^\star_{\min}} - 1)\mathrm{e}^{\frac{1}{s^\star_{\min}}}(1 - \mathrm{e}^{-1})}. \tag{148}$$

Now, it is easy to show that the function $x \mapsto 1 - \left(1 - \frac{1}{x}\right) \mathrm{e}^{-x}(2 + x(x+2))$ is lower-bounded by 0.3 for $x > 0$. As a result, we have

$$P_{S^\star}(s^\star_{\min}) \ge \frac{0.3}{\frac{s^\star_2}{s^\star_{\min}}(\frac{s^\star_2}{s^\star_{\min}} - 1)\mathrm{e}^{\frac{1}{s^\star_{\min}}}(1 - \mathrm{e}^{-1})} \tag{149}$$

$$\geq \frac{0.3}{(\mathsf{A}^2+1)\mathsf{A}^2 \mathrm{e}^{\mathsf{A}^2+1}(1-\mathrm{e}^{-1})}, \tag{150}$$

where in the last step we used $s_2^\star \leq 1$ and $s_{\min}^\star = \frac{1}{1+\mathsf{A}^2}$.

□

D. Small Average- and Peak-Power Regime

In this section, we prove the first part of Theorem 1. We identify the following cases, which will be shown separately.

- Case 1: $\mathsf{A}^2 \leq \mathsf{P}$;
- Case 2: $\mathsf{A}^2 < \sqrt{2}$;
- Case 3: $\lambda = 0$.

1) Proof for Case 1:

Lemma 9. *If* $\mathsf{A}^2 \leq \mathsf{P}$, *then*

i) the peak-power constraint is active,

ii) the average-power constraint is inactive ($\lambda = 0$), and

iii) $\mathsf{A} = \max \mathsf{supp}(P_{X^\star})$.

Proof. To prove part i) we use a contradiction argument by assuming that the peak-power constraint is not active. Then we necessarily have $\mathbb{E}\left[(X^\star)^2\right] = \mathsf{P}$, which implies that there exists an $x^\star \in \mathsf{supp}(P_{X^\star})$ such that $x^\star \geq \sqrt{\mathsf{P}} \geq \mathsf{A}$, which contradicts the assumption.

To prove part ii), we can see that $\mathbb{E}\left[(X^\star)^2\right] \leq \mathsf{A}^2 \leq \mathsf{P}$ where the first inequality holds because the peak-power constraint is active.

To show part iii), we proceed again by contradiction, by assuming that $X^\star \leq \mathsf{B} < \mathsf{A}$. Define $\bar{X} = \frac{\mathsf{A}}{\mathsf{B}} X^\star$, then

$$I\left(\bar{Y}; \bar{X}\right) = I\left(\widetilde{H}\bar{X} + \widetilde{W}; \bar{X}\right) \tag{151}$$

$$= I\left(\widetilde{H}\frac{\mathsf{A}}{\mathsf{B}}X^\star + \widetilde{W}; X^\star\right) \tag{152}$$

$$= I\left(\widetilde{H}X^\star + \frac{\mathsf{B}}{\mathsf{A}}\widetilde{W}; X^\star\right) \tag{153}$$

$$\geq I\left(\widetilde{H}X^\star + \widetilde{W}; X^\star\right) = C \tag{154}$$

which is a contradiction. Note that the inequality in (154) follows from $\mathsf{B} < \mathsf{A}$. □

2) Proof for Case 2:

Lemma 10. *If* $\mathsf{A}^2 < \sqrt{2}$, *then* $\lambda = 0$ *or* $\mathsf{supp}(P_{X^\star}) = \{0, \mathsf{A}\}$.

Proof. Note that the lower bound in (56) and the upper bound in (58) of Lemma 4 say that if $\lambda > 0$ and if there exists a support point $s^\star \neq 1$ such that $\Xi'_E(s^\star; f_{Y^\star}) = 0$ and $\Xi''_E(s^\star; f_{Y^\star}) \leq 0$, then $\frac{1}{1+\mathsf{A}^2} \leq \frac{1}{1+\sqrt{2}}$. A corresponding contrapositive sentence is as follows: if $\frac{1}{1+\mathsf{A}^2} > \frac{1}{1+\sqrt{2}}$, then $\lambda = 0$ or there exists only a second support point with $\Xi'_E(s_{\min}^\star; f_{Y^\star}) < 0$. However, this latter condition can happen only if the peak-power constraint is active, which implies that $s_{\min}^\star = \frac{1}{1+\mathsf{A}^2}$, hence $\mathsf{supp}(P_{X^\star}) = \{0, \mathsf{A}\}$.

□

3) Proof for Case 3:

Lemma 11. *If* $\lambda = 0$, *then* $\mathsf{supp}(P_{X^\star}) = \{0, \mathsf{A}\}$ *or* $\mathsf{A}^2 \geq \frac{\mathrm{e}}{3-\mathrm{e}}$.

Proof. Note that the bound (41) of Proposition 5 says that if there exists a non-nodal support point, then $\lambda \geq s_{\min}^\star 3\mathrm{e}^{-1} - \frac{\mathbb{E}\left[(S^\star)^2\right]}{\mathbb{E}[S^\star]}(3\mathrm{e}^{-1} - 1) \geq \frac{1}{1+\mathsf{A}^2} 3\mathrm{e}^{-1} - (3\mathrm{e}^{-1} - 1)$. This last quantity is strictly positive if $\mathsf{A}^2 < \frac{\mathrm{e}}{3-\mathrm{e}}$. Therefore, an equivalent statement is: if there exists a non-nodal support point and $\mathsf{A}^2 < \frac{\mathrm{e}}{3-\mathrm{e}}$, then $\lambda > 0$. The corresponding contrapositive sentence is: if $\lambda = 0$, then there does not exist any non-nodal support point or $\mathsf{A}^2 \geq \frac{\mathrm{e}}{3-\mathrm{e}}$. However, notice that the only case without non-nodal support points and $\lambda = 0$ is when $\mathsf{supp}(P_{X^\star}) = \{0, \mathsf{A}\}$. □

4) Proof of Lemma 3:

Proof. To see if the average-power constraint is active, we need to determine whether λ is strictly positive. To do so, thanks to the KKT conditions in (20) we can write

$$\Xi_R(0; f_{Y^\star}) = \Xi_R(\mathsf{A}; f_{Y^\star}) = 0 \tag{155}$$

that, combined with (18), gives the relationship

$$\lambda = \frac{\rho}{\mathsf{A}^2}. \tag{156}$$

When the expression in (156) is strictly positive, the average-power constraint is active, hence we must have $\mathsf{P} = \mathbb{E}\left[(X^\star)^2\right] = \mathsf{A}^2 P_{X^\star}(\mathsf{A})$, which gives $P_{X^\star}(\mathsf{A}) = \frac{\mathsf{P}}{\mathsf{A}^2}$. On the contrary, when (156) is not strictly positive, the average-power constraint is not active and the value of $P_{X^\star}(\mathsf{A})$ is solution to the equation

$$\int_0^\infty \left(\frac{1}{1+\mathsf{A}^2}\mathrm{e}^{-\frac{y}{1+\mathsf{A}^2}} - \mathrm{e}^{-y}\right) \log\left((1-P_{X^\star}(\mathsf{A}))\mathrm{e}^{-y} + \frac{P_{X^\star}(\mathsf{A})}{(1+\mathsf{A}^2)}\mathrm{e}^{-\frac{y}{1+\mathsf{A}^2}}\right) \mathrm{d}y + \log(1+\mathsf{A}^2) = 0. \tag{157}$$

which is obtained by setting $\lambda = 0$ into conditions (155). $\square$

E. Bounds on Capacity: Proof of Proposition 4

Proof. We begin by showing the upper bound. First, note that

$$C(\mathsf{P}, \mathsf{A}) = \max_{P_X:\, \mathbb{E}[X^2]\le \mathsf{P},\, X\le \mathsf{A}} I(X;Y) \le \max_{P_X:\, \mathbb{E}[X^2]\le \min(\mathsf{P},\mathsf{A}^2)} I(X;Y) = C(\mathsf{P}_\mathsf{A}, \infty) \tag{158}$$

where we let $\mathsf{P}_\mathsf{A} = \min(\mathsf{P}, \mathsf{A}^2)$. Second, we have that

$$\begin{aligned}
I(U; HU+W) &\le I(U; HU+W|H) && (159)\\
&= h(HU+W|H) - h(W) && (160)\\
&\le \mathbb{E}\left[\log(\pi\mathrm{e}(|H|^2\mathbb{E}\left[|U|^2\right] + \sigma_W^2))\right] - \log(\pi\mathrm{e}\sigma_W^2) && (161)\\
&\le \log\left(\frac{\sigma_H^2}{\sigma_W^2}\tilde{\mathsf{P}} + 1\right) && (162)\\
&= \log(\mathsf{P}+1), && (163)
\end{aligned}$$

so we can conclude that $C(\mathsf{P}_\mathsf{A}, \infty) \le \log(1+\mathsf{P}_\mathsf{A})$.

For the second capacity upper bound, the mutual information is

$$\begin{aligned}
I(S;Y) = I\left(S; -\log(S) - Z\right) &= h(L) - h(Z) && (164)\\
&= h(L) - (\gamma+1) && (165)
\end{aligned}$$

where we have used $h(Z) = \gamma + 1$. Consider a Gumbel pdf $q(z) = \frac{1}{\sigma}\mathrm{e}^{-\frac{z}{\sigma} - \mathrm{e}^{-\frac{z}{\sigma}}}$ for $z \in \mathbb{R}$. Then, by Gibb's inequality, we have

$$\begin{aligned}
h(L) &\le -\mathbb{E}[\log q(L)] && (166)\\
&= -\mathbb{E}\left[\log q\left(-\log(S) - Z\right)\right] && (167)\\
&= -\mathbb{E}\left[\log\frac{1}{\sigma}\exp\left(-\frac{1}{\sigma}\log\frac{1}{S} + \frac{1}{\sigma}Z - \mathrm{e}^{-\frac{1}{\sigma}\log\frac{1}{S} + \frac{1}{\sigma}Z}\right)\right] && (168)\\
&= \log\sigma + \mathbb{E}\left[\frac{1}{\sigma}\log\frac{1}{S}\right] - \mathbb{E}\left[\frac{1}{\sigma}Z\right] + \mathbb{E}\left[\mathrm{e}^{-\frac{1}{\sigma}\log\frac{1}{S} + \frac{1}{\sigma}Z}\right] && (169)\\
&= \log\sigma + \mathbb{E}\left[\frac{1}{\sigma}\log\frac{1}{S}\right] - \mathbb{E}\left[\frac{1}{\sigma}Z\right] + \mathbb{E}\left[\left(\frac{1}{S}\right)^{-\frac{1}{\sigma}}\right]\mathbb{E}\left[\mathrm{e}^{\frac{1}{\sigma}Z}\right] && (170)\\
&= \log\sigma + \mathbb{E}\left[\frac{1}{\sigma}\log\frac{1}{S}\right] - \frac{\gamma}{\sigma} + \mathbb{E}\left[\left(\frac{1}{S}\right)^{-\frac{1}{\sigma}}\right]\Gamma\left(1 - \frac{1}{\sigma}\right) && (171)\\
&\le \log\sigma + \frac{1}{\sigma}\log(\mathsf{P}_\mathsf{A} + 1) - \frac{\gamma}{\sigma} + \Gamma\left(1 - \frac{1}{\sigma}\right) && (172)
\end{aligned}$$

where in (171) we have used $\mathbb{E}[Z] = \gamma$ and $\mathbb{E}[\mathrm{e}^{\frac{1}{\sigma}Z}] = \Gamma\left(1 - \frac{1}{\sigma}\right)$ for $\sigma > 1$; and in (172) we used Jensen's inequality, the average power constraint (10), and that $S \leq 1$. By choosing $\sigma = 1 + \log(\mathsf{P_A} + 1)$ for $\mathsf{P_A} > 0$, we finally get

$$C(\mathsf{P_A}, \infty) \leq \log(1 + \log(\mathsf{P_A} + 1)) + \frac{\log(\mathsf{P_A} + 1) - \gamma}{1 + \log(\mathsf{P_A} + 1)} + \Gamma\left(1 - \frac{1}{1 + \log(\mathsf{P_A} + 1)}\right) \tag{173}$$

$$\leq \log(1 + \log(\mathsf{P_A} + 1)) + 1 + \Gamma\left(1 - \frac{1}{1 + \log(\mathsf{P_A} + 1)}\right) \tag{174}$$

$$:= \overline{C}(\mathsf{P_A}). \tag{175}$$

We now show the lower bound. 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

$$\mathbb{E}\left[\frac{1}{S}\right] = \mathbb{E}[\mathrm{e}^{\log\frac{1}{S}}] = \frac{\mathrm{e}^{\log(\mathsf{P_A}+1)} - \mathrm{e}^0}{\log(\mathsf{P_A} + 1)} = \frac{\mathsf{P_A}}{\log(\mathsf{P_A} + 1)} \leq \mathsf{P_A} + 1 \leq \mathsf{P} + 1 \tag{176}$$

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

$$h(L) = h\left(\log\frac{1}{S} - Z\right) \tag{177}$$

$$\geq \log\left(\sqrt{\mathrm{e}^{2h\left(\log\frac{1}{S}\right)} + \mathrm{e}^{2h(Z)}}\right) \tag{178}$$

$$= \log\left(\sqrt{\mathrm{e}^{2\log\log(\mathsf{P_A}+1)} + \mathrm{e}^{2(\gamma+1)}}\right) \tag{179}$$

$$= \log\left(\sqrt{(\log(\mathsf{P_A} + 1))^2 + \mathrm{e}^{2(\gamma+1)}}\right) \tag{180}$$

$$\geq \max\left\{\log(\log(\mathsf{P_A} + 1))), \gamma + 1\right\}. \tag{181}$$

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

$$C(\mathsf{P}, \mathsf{A}) \geq \max\left\{\log(\log(\mathsf{P_A} + 1))) - \gamma - 1, 0\right\} \tag{182}$$

$$= [\log(\log(\mathsf{P_A} + 1))) - \gamma - 1]^+ \tag{183}$$

$$:= \underline{C}(\mathsf{P_A}). \tag{184}$$

This concludes the proof of the capacity bounds.

The gap between capacity upper and lower bounds is:

$$\overline{C}(\mathsf{P_A}) - \underline{C}(\mathsf{P_A}) = \min\left\{\log(\mathsf{P_A} + 1), \log\left(\frac{1}{\log(\mathsf{P_A} + 1)} + 1\right) + 2 + \gamma + \Gamma\left(1 - \frac{1}{1 + \log(\mathsf{P_A} + 1)}\right)\right\}. \tag{185}$$

The minimum in (185) is upper-bounded by $\log(\xi + 1)$, where $\xi \approx 51.8618$ is the value of $\mathsf{P_A}$ where the two expressions in the minimum of (185) have the same value. Then, it turns out that

$$\overline{C}(\mathsf{P_A}) - \underline{C}(\mathsf{P_A}) \leq \log(\xi + 1) \approx 3.9677 < 4. \tag{186}$$

This concludes the proof. □

Appendix III
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 12. *Let us consider the exponential channel model of* (12). *Then, we have*

$$\frac{\mathrm{d}}{\mathrm{d}s} f_{Y|S}(y|s) = -\frac{1 - sy}{s^2}\frac{\mathrm{d}}{\mathrm{d}y} f_{Y|S}(y|s) \tag{187}$$

$$= \frac{1 - sy}{s} f_{Y|S}(y|s). \tag{188}$$

Proof. First of all, compute

$$\frac{\mathrm{d}}{\mathrm{d}s} f_{Y|S}(y|s) = \frac{\mathrm{d}}{\mathrm{d}s}\exp\left(-sy + \log s\right) \tag{189}$$

$$= \frac{1 - sy}{s}\exp\left(-sy + \log s\right) \tag{190}$$

$$= \frac{1 - sy}{s} f_{Y|S}(y|s). \tag{191}$$

Then, compute

$$\frac{\mathrm{d}}{\mathrm{d}y} f_{Y|S}(y|s) = \frac{\mathrm{d}}{\mathrm{d}y} \exp\left(-sy + \log s\right) \tag{192}$$
$$= -s \exp\left(-sy + \log s\right) \tag{193}$$
$$= -s f_{Y|S}(y|s) \tag{194}$$
$$= -\frac{s^2}{1 - sy} \frac{\mathrm{d}}{\mathrm{d}s} f_{Y|S}(y|s) \tag{195}$$

□

Lemma 13. *Let us consider the exponential channel model of* (12) *and an output distribution* f_Y*. Then, we have*

$$\frac{\mathrm{d}}{\mathrm{d}y} \log f_Y(y) = -\mathbb{E}\left[S \mid Y = y\right], \tag{196}$$
$$\frac{\mathrm{d}^2}{\mathrm{d}y^2} \log f_Y(y) = \mathsf{Var}\left[S \mid Y = y\right]. \tag{197}$$

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

$$f_{Y|S}(y|s) = \exp\left(-sy + \log s\right) \tag{198}$$
$$= h(y)\mathrm{e}^{sT(y) - \phi(s)} \tag{199}$$

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 [35, Remark 7] to write

$$\mathbb{E}\left[S \mid Y = y\right] = D_y \log \frac{f_Y(y)}{h(y)} = -\frac{\mathrm{d}}{\mathrm{d}y} \log f_Y(y). \tag{200}$$

As for the second derivative, write

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2} \log f_Y(y) = -\frac{\mathrm{d}}{\mathrm{d}y} \mathbb{E}\left[S \mid Y = y\right] \tag{201}$$
$$= D_y \mathbb{E}\left[S \mid Y = y\right] \tag{202}$$
$$= \mathsf{Var}\left[S \mid Y = y\right] \tag{203}$$

where the last step is due to [35, Prop. 4]. □

Lemma 14. *Given* $s > 0$*, suppose that* $\lim_{y\to\infty} f(y)e^{-sy} = 0$*. Then,*

$$\mathbb{E}\left[f(Y) \mid S = s\right] = f(0) + \frac{1}{s} \mathbb{E}\left[f'(Y) \mid S = s\right], \tag{204}$$
$$\frac{\mathrm{d}}{\mathrm{d}s} \mathbb{E}\left[f(Y) \mid S = s\right] = -\frac{1}{s} \mathbb{E}\left[Y f'(Y) \mid S = s\right], \tag{205}$$
$$\frac{\mathrm{d}}{\mathrm{d}s} \frac{1}{s} \mathbb{E}\left[f(Y) \mid S = s\right] = -\frac{1}{s} \mathbb{E}\left[Y f(Y) \mid S = s\right]. \tag{206}$$

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

$$\mathbb{E}\left[f(Y) \mid S = s\right] = \int_0^\infty f(y) s e^{-sy} \mathrm{d}y \tag{207}$$
$$= \left(-e^{-sy} f(y)\right)\big|_0^\infty - \int_0^\infty f'(y) \left(-e^{-sy}\right) \mathrm{d}y \tag{208}$$
$$= f(0) + \frac{1}{s} \int_0^\infty s e^{-sy} f'(y) \mathrm{d}y \tag{209}$$
$$= f(0) + \frac{1}{s} \mathbb{E}\left[f'(Y) \mid S = s\right], \tag{210}$$

where in (209) 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

$$\frac{\mathrm{d}}{\mathrm{d}s} \mathbb{E}\left[f(Y) \mid S = s\right] = \frac{\mathrm{d}}{\mathrm{d}s} \int_0^\infty f(y) s e^{-sy} \mathrm{d}y \tag{211}$$

$$
\begin{aligned}
&= \int_0^\infty f(y)e^{-sy}\mathrm{d}y - \int_0^\infty f(y)yse^{-sy}\mathrm{d}y && (212)\\
&= (ye^{-sy}f(y))|_0^\infty - \int_0^\infty y\left(-se^{-sy}f(y) + e^{-sy}f'(y)\right)\mathrm{d}y - \int_0^\infty sye^{-sy}f(y)\mathrm{d}y && (213)\\
&= -\int_0^\infty ye^{-sy}f'(y)\mathrm{d}y && (214)\\
&= -\frac{1}{s}\mathbb{E}\left[Yf'(Y) \mid S=s\right], && (215)
\end{aligned}
$$

where in (214) we used that $\lim_{y\to\infty} f(y)e^{-sy} = 0$. To prove the third identity, write

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s}\frac{1}{s}\mathbb{E}\left[f(Y) \mid S=s\right] &= \frac{\mathrm{d}}{\mathrm{d}s}\int_0^\infty \mathrm{e}^{-sy}f(y)\mathrm{d}y && (216)\\
&= -\int_0^\infty y\mathrm{e}^{-sy}f(y)\mathrm{d}y && (217)\\
&= -\mathbb{E}\left[Yf(Y) \mid S=s\right]. && (218)
\end{aligned}
$$

$\square$

Proposition 6. *The first and second derivatives of function Ξ_E given in (19) are as follows*

$$
\begin{aligned}
\Xi_E'(s; f_{Y^\star}) &= \frac{1}{s}\mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] \mid S=s\right] + \frac{\lambda}{s^2} && (219)\\
\Xi_E''(s; f_{Y^\star}) &= -\frac{1}{s^2}\mathbb{E}\left[1 - 2Y\mathbb{E}\left[S^\star \mid Y\right] + Y^2\mathsf{Var}\left[S^\star \mid Y\right] \,\middle|\, S=s\right] - \frac{2\lambda}{s^3}, && (220)
\end{aligned}
$$

for $s > 0$, or alternatively

$$
\Xi_E''(s; f_{Y^\star}) = -\frac{1}{s^2} + \frac{1}{s}\mathbb{E}\left[Y^2\mathbb{E}\left[S^\star \mid Y\right] \,\middle|\, S=s\right] - \frac{2\lambda}{s^3}, \qquad s > 0. \tag{221}
$$

Proof. By taking the derivative of (19) with respect to s, we get

$$
\begin{aligned}
\Xi_E'(s; f_{Y^\star}) &= -\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\left[\log f_{Y^\star}(Y) \mid S=s\right] + \frac{1}{s} + \frac{\lambda}{s^2} && (222)\\
&= \frac{1}{s}\mathbb{E}\left[Y\left(\frac{\mathrm{d}}{\mathrm{d}y}\log f_{Y^\star}(Y)\right) \,\middle|\, S=s\right] + \frac{1}{s} + \frac{\lambda}{s^2} && (223)\\
&= -\frac{1}{s}\mathbb{E}\left[Y\mathbb{E}\left[S^\star \mid Y\right] \mid S=s\right] + \frac{1}{s} + \frac{\lambda}{s^2} && (224)\\
&= \frac{1}{s}\mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] \mid S=s\right] + \frac{\lambda}{s^2} && (225)
\end{aligned}
$$

where in (223) we used Lemma 14; and in (224) we used Lemma 13.

As for the second derivative of Ξ_E, we compute the derivative of (223) to get

$$
\begin{aligned}
\Xi_E''(s; f_{Y^\star}) &= -\frac{1}{s^2}\mathbb{E}\left[Y\left(\frac{\mathrm{d}}{\mathrm{d}y}\log f_{Y^\star}(Y)\right) \,\middle|\, S=s\right] + \frac{1}{s}\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\left[Y\left(\frac{\mathrm{d}}{\mathrm{d}y}\log f_{Y^\star}(Y)\right) \,\middle|\, S=s\right] - \frac{1}{s^2} - \frac{2\lambda}{s^3} && (226)\\
&= -\frac{1}{s^2}\mathbb{E}\left[1 + 2Y\left(\frac{\mathrm{d}}{\mathrm{d}y}\log f_{Y^\star}(Y)\right) + Y^2\left(\frac{\mathrm{d}^2}{\mathrm{d}y^2}\log f_{Y^\star}(Y)\right) \,\middle|\, S=s\right] - \frac{2\lambda}{s^3} && (227)\\
&= -\frac{1}{s^2}\mathbb{E}\left[1 - 2Y\mathbb{E}\left[S^\star \mid Y\right] + Y^2\mathsf{Var}\left[S^\star \mid Y\right] \,\middle|\, S=s\right] - \frac{2\lambda}{s^3} && (228)
\end{aligned}
$$

where in (226) we used Lemma 14; and in (227) we used Lemma 13. The alternative expression of Ξ_E'' is obtained by directly taking the derivative of (225) with respect to s and by applying the second identity of Lemma 14:

$$
\Xi_E''(s; f_{Y^\star}) = -\frac{1}{s^2} + \frac{1}{s}\mathbb{E}\left[Y^2\mathbb{E}\left[S^\star \mid Y\right] \,\middle|\, S=s\right] - \frac{2\lambda}{s^3}. \tag{229}
$$

$\square$

APPENDIX IV
ON THE SECOND LARGEST SUPPORT POINT OF $P_{S^\star}$

The following result shows that the second largest value of $\mathsf{supp}(P_{S^\star})$ is bounded away from 1.

Lemma 15. *Let $s^\star \in \mathsf{supp}(P_{S^\star}) \setminus \{1\}$. Then,*

$$s^\star \le \frac{\mathbb{E}[(S^\star)^2]}{\mathbb{E}[S^\star]} - \lambda \le \mathbb{E}\left[S\right] - \lambda. \tag{230}$$

Proof. From the KKT conditions, we have that $\Xi'_E(s^\star; f_{Y^\star}) = 0$ for $s^\star \in \mathsf{supp}(P_{S^\star}) \setminus \{1, s^\star_{\min}\}$ and that $\Xi'_E(s^\star_{\min}; f_{Y^\star}) \le 0$. By using the result of Proposition 6, for $s^\star \in \mathsf{supp}(P_{S^\star}) \setminus \{1\}$ we have

$$\begin{aligned}
0 &\ge \Xi'_E(s^\star; f_{Y^\star}) && (231)\\
&= s^\star \mathbb{E}\left[1 - Y\mathbb{E}\left[S^\star \mid Y\right] \mid S = s^\star\right] + \lambda && (232)\\
&\ge s^\star \mathbb{E}\left[1 - Y\frac{\mathbb{E}[(S^\star)^2]}{\mathbb{E}[S^\star]} \,\middle|\, S = s^\star\right] + \lambda && (233)\\
&= s^\star\left(1 - \frac{1}{s^\star}\frac{\mathbb{E}[(S^\star)^2]}{\mathbb{E}[S^\star]}\right) + \lambda && (234)\\
&= s^\star - \frac{\mathbb{E}[(S^\star)^2]}{\mathbb{E}[S^\star]} + \lambda && (235)
\end{aligned}$$

where (233) follows from Lemma 16; and (234) follows from $\mathbb{E}\left[Y \mid S = s\right] = \frac{1}{s}$. The inequality can be further loosened by using Cauchy-Schwarz. $\square$

For large values of P we expect λ to tend to 0. Lemma 15 shows that as P increases we do not observe splits of $\mathsf{supp}(P_{S^\star})$ for $s^\star = 1$.

APPENDIX V
CONNECTIONS TO CUMULANTS

The following result is a consequence of [35, Proposition 4]: for $\ell \in \mathbb{N}$

$$\kappa_{S|Y=y}(\ell) = (-1)^\ell \frac{\mathrm{d}^\ell}{\mathrm{d}y^\ell}\log\left(f_Y(y)\right), \qquad y \in (0,\infty), \tag{236}$$

where $\kappa_{S|Y=y}(\ell)$ is the conditional cumulant. In particular, note that

$$\begin{aligned}
\kappa_{S|Y=y}(1) &= -\frac{\mathrm{d}}{\mathrm{d}y}\log\left(f_Y(y)\right) && (237)\\
&= \mathbb{E}\left[S \mid Y = y\right] && (238)
\end{aligned}$$

where the last step follows from Lemma 13, and that

$$\kappa_{S|Y=y}(2) = \mathbb{E}[(S - \mathbb{E}[S|Y])^2|Y = y] = \mathsf{Var}\left[S \mid Y = y\right]. \tag{239}$$

As a consequence of the above discussion, we have that

Lemma 16.

$$\frac{\mathrm{d}}{\mathrm{d}y}\mathbb{E}[S|Y = y] = -\mathsf{Var}\left[S \mid Y = y\right]. \tag{240}$$

Consequently, $y \mapsto \mathbb{E}[S|Y = y]$ is a decreasing function. Furthermore,

$$\begin{aligned}
\lim_{y\to 0}\mathbb{E}[S|Y = y] &= \frac{\mathbb{E}[S^2]}{\mathbb{E}[S]} && (241)\\
\lim_{y\to\infty}\mathbb{E}[S|Y = y] &= \min\{\mathsf{supp}(S)\} && (242)
\end{aligned}$$

Proof. The first result follows by putting together (236), (238), and (239). Next, note that

$$\lim_{y\to 0}\mathbb{E}[S|Y = y] = \lim_{y\to 0}\frac{\mathbb{E}[S^2\mathrm{e}^{-Sy}]}{\mathbb{E}[S\mathrm{e}^{-Sy}]} = \frac{\mathbb{E}[S^2]}{\mathbb{E}[S]}. \tag{243}$$

Next, note

$$\mathbb{E}[S|Y=y] = \sum_s \frac{s^2 \mathrm{e}^{-sy} P_S(s)}{f_Y(y)} \tag{244}$$

$$\leq \sum_s \frac{s^2 \mathrm{e}^{-sy} P_S(s)}{s_{\min} \mathrm{e}^{-s_{\min} y} P_S(s_{\min})} \tag{245}$$

$$= s_{\min} + \sum_{s > s_{\min}} \frac{s^2 \mathrm{e}^{-sy} P_S(s)}{s_{\min} \mathrm{e}^{-s_{\min} y} P_S(s_{\min})}. \tag{246}$$

Taking the limit as $y \to +\infty$, and using that $s_{\min} \leq \mathbb{E}[S|Y=y]$, we arrive at

$$\lim_{y \to \infty} \mathbb{E}[S|Y=y] = s_{\min}. \tag{247}$$

□