The Capacity Achieving Distribution for the Amplitude Constrained Additive Gaussian Channel: An Upper Bound on the Number of Mass Points

Alex Dytso, *Member, IEEE*, Semih Yagli, *Student Member, IEEE*, H. Vincent Poor, *Fellow, IEEE*, and Shlomo Shamai (Shitz), *Life Fellow, IEEE*

***Abstract*— This paper studies an n-dimensional additive Gaussian noise channel with a peak-power-constrained input. It is well known that, in this case, when $n = 1$ the capacity-achieving input distribution is discrete with finitely many mass points, and when $n > 1$ the capacity-achieving input distribution is supported on finitely many concentric shells. However, due to the previous proof technique, not even a bound on the exact number of mass points/shells was available. This paper provides an alternative proof of the finiteness of the number mass points/shells of the capacity-achieving input distribution while producing the first firm bounds on the number of mass points and shells, paving an alternative way for approaching many such problems. The first main result of this paper is an order tight implicit bound which shows that the number of mass points in the capacity-achieving input distribution is within a factor of two from the number of zeros of the downward shifted capacity-achieving output probability density function. Next, this implicit bound is utilized to provide a first firm upper on the support size of optimal input distribution, an $O(\mathsf{A}^2)$ upper bound where A denotes the constraint on the input amplitude. The second main result of this paper generalizes the first one to the case when $n > 1$, showing that, for each and every dimension $n \geq 1$, the number of shells that the optimal input distribution contains is $O(\mathsf{A}^2)$. Finally, the third main result of this paper reconsiders the case $n = 1$ with an additional average power constraint, demonstrating a similar $O(\mathsf{A}^2)$ bound.**

***Index Terms*— Amplitude constraint, power constraint, additive vector Gaussian noise channel, capacity, discrete distributions.**

I. Introduction

We consider an additive noise channel for which the input-output relationship is given by

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}, \tag{1}$$

where the input $\boldsymbol{X} \in \mathbb{R}^n$ is independent of the standard Gaussian noise $\boldsymbol{Z} \in \mathbb{R}^n$. We are interested in finding the capacity of the channel in (1) subject to the constraint that $\boldsymbol{X} \in \mathcal{B}_0(\mathsf{A})$ where $\mathcal{B}_0(\mathsf{A})$ is an n-ball centered at zero with radius A (i.e., amplitude or peak-power constrained input), that is

$$C_n(\mathsf{A}) = \max_{\boldsymbol{X}:\, \boldsymbol{X} \in \mathcal{B}_0(\mathsf{A})} I(\boldsymbol{X}; \boldsymbol{Y}). \tag{2}$$

In his seminal paper [1] (see also [2]), for the case of $n = 1$, Smith has shown that an optimizing distribution in (2) is unique, symmetric around the origin, and perhaps surprisingly, discrete with finitely many mass points. Using tools such as the Identity Theorem from complex analysis, Smith has proven that the cardinality of the support set of the optimal input distribution cannot be infinite, and, thus, must be finite. Employing this proof by contradiction, Shamai and Bar-David [3] have extended the method of Smith to $n = 2$, and showed that, in this setting, the maximizing input random variable is given by

$$\boldsymbol{X}^\star = R^\star \cdot \boldsymbol{U}^\star \tag{3}$$

where the magnitude $R^\star$ is discrete with finitely many points and the random unit vector $\boldsymbol{U}^\star$, which is independent of $R^\star$, has a uniform phase on $[0, 2\pi)$. In other words, the support is given by finitely many concentric shells, e.g., Fig. 1. As a matter of fact, this phenomena that the optimal input distribution lies on finitely many concentric spheres remains true for any $n \geq 2$, cf. [4], [5] and [6].

Regrettably, the method of proof by contradiction does not lead to a characterization of the number of spheres (number of mass points when $n = 1$) in the capacity-achieving input distribution. In fact, as of the writing of this paper, very little is known about the structure of that distribution, and a very simple question remains open about 50 years after Smith's contribution:

When $n = 1$, what is the cardinality of the support of the optimal input distribution as a function of A?

In this work, we provide the first firm upper bound on the number of points for $n = 1$ and the number of shells for every $n > 1$, partially answering the above question. Furthermore, for the case of $n = 1$, using similar methods, we also provide

Manuscript received February 13, 2019; revised August 27, 2019; accepted September 30, 2019. Date of publication October 21, 2019; date of current version March 17, 2020. The work of A. Dytso, S. Yagli, and H. V. Poor was supported in part by the U.S. National Science Foundation under Grant CCF-1908308 and in part by the United States–Israel Binational Science Foundation under Grant BSF-2018710. The work of S. Shamai was supported in part by the United States–Israel Binational Science Foundation under Grant BSF-2018710 and in part by the European Union's Horizon 2020 Research and Innovation Programme under Grant 694630. This article was presented in part at the 2019 IEEE International Symposium on Information Theory.

A. Dytso, S. Yagli, and H. V. Poor are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: adytso@princeton.edu; syagli@princeton.edu; poor@princeton.edu).

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 3200003, Israel (e-mail: sshlomo@ee.technion.ac.il).

Communicated by F. Alajaji, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2019.2948636

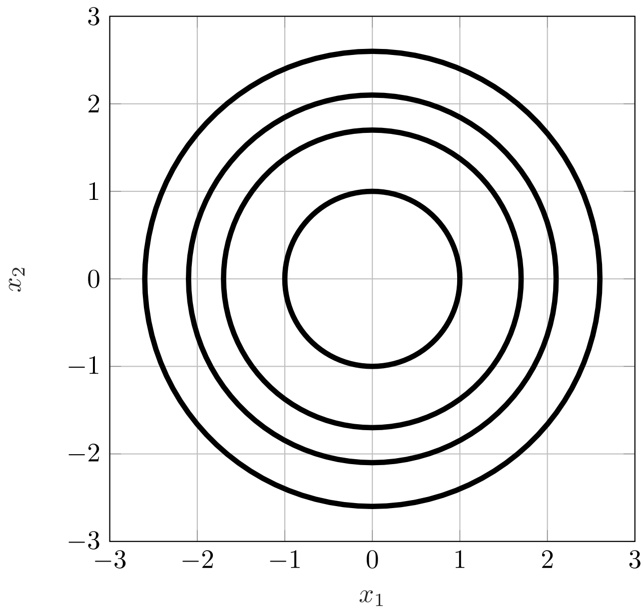

Fig. 1. An example of a support of an optimal input distribution for $n = 2$.

an upper bound on the cardinality of the support of the distribution achieving

$$C(\mathsf{A}, \mathsf{P}) = \max_{\substack{X:\,|X| \le \mathsf{A} \\ \mathbb{E}[X^2] \le \mathsf{P}}} I(X;Y). \tag{4}$$

A. Prior Work

The history of the problem begins with Shannon who was the first to consider an amplitude constraint on the input [7]. Shannon's original paper proposes both upper and lower bounds on the capacity and shows that peak-power capacity and average power capacity have the same asymptotic behavior at the low signal to noise ratio. The next major breakthrough is the seminal paper of Smith [1], where Smith proves the discreteness of the capacity-achieving input distribution and also shows the optimality of the equiprobable binary input on $\{\pm\mathsf{A}\}$ so long as $\mathsf{A} \le 0.1$. Sharma and Shamai [8] extend the result of Smith, and argue[1] that an equiprobable input on $\{\pm\mathsf{A}\}$ is optimal if and only if $\mathsf{A} \le \bar{\mathsf{A}} \approx 1.665$. The proof of the result in [8], which generalizes to vector channels, is shown in [9]. Also, based on numerical evidence, Sharma and Shamai [8] conjectured that the number of mass points increases by at most one and a new point always appears at zero. Based on this conjecture, in [8] it has been shown that a ternary input distributed on $\{-\mathsf{A}, 0, \mathsf{A}\}$ is optimal for all $\bar{\mathsf{A}} \le \mathsf{A} \le \bar{\bar{\mathsf{A}}} \approx 2.786$.

A progress on the algorithmic aspect of computing the optimal input distribution was made in [10] which proposed an iterative procedure that converges to the a capacity achieving distribution based on the cutting-plane method. The bound on the number of mass points found in our work is particularly relevant for numerical methods as it reduces the optimization space for algorithms such as the one contained in [10].

[1]The formal proof is incomplete as the argument in [8] uses a conjecture which presumes that the number of points in the optimal distribution increases by 1 as the amplitude constraint is relaxed. Proof of this fact was later shown to be true in [9].

A number of papers have also focused on upper and lower bounds on the capacity in (2). Broadly speaking, there are three types of capacity upper bounding approaches. The first approach uses the maximum entropy principle [11, Chapter 12] and upper bounds the output differential entropy, $h(Y)$, subject to some moment constraint [12]. The second approach uses a dual capacity characterization[2] where the maximization of the mutual information over the input distribution is replaced by minimization of the relative entropy over the output distribution. A suboptimal choice of an output distribution in the dual capacity expression results in an upper bound on the capacity [15]–[18]. The third approach uses a characterization of the mutual information as an integral of the minimum mean square error (MMSE) [19], and leads to an upper bound by replacing the optimal estimator in the MMSE term by a suboptimal one [9]. As for the lower bounds on the capacity, the first one relevant to our setting, as mentioned above, was proposed by Shanon in [7] which was based on the entropy power inequality. Other important lower bounds include Ozarow-Wyner bounds [20], [21], and bounds based on Jensen's inequality [22].

There is also a substantial literature that extends the proof recipe of Smith to the other channels. For example, the approach of Smith for showing discreteness of an optimal input distribution has been extended to complex Gaussian channels [3], additive noise channels where noise has a sufficiently regular pdf [23], Rayleigh fading channels [24], and Poisson channels [25]. For an overview of the literature on various optimization methods that show discreteness of a capacity-achieving distribution the interested reader is referred to [6]. Moreover, a comprehensive account of capacity results for point-to-point Gaussian channels can be found in [26].

One of the ingredients of our proof is the Oscillation Theorem of Karlin [27]. In the past, Karlin's theorem has been used to study extreme distributions; however, not to the same degree as it is used in this paper. For example, in the context of a Bayesian estimation problem [28], Oscillation Theorem has been used to show the necessary and sufficient conditions for a binary distribution to be the least favorable. In [9], in a vector version of the optimization in (2), Oscillation Theorem has been used to show the necessary and sufficient conditions for a uniform distribution on a single sphere to be optimal.

B. Contributions and Paper Outline

In what follows:

1) Section II presents our main results;
2) Section III provides the proof of the first part of our main result for the case of $n = 1$. There, it is shown that the number of zeros of the shifted optimal output probability density function (pdf) is within a factor of two from the number of mass points of the optimal input distribution. The main element of this part relies on Karlin's Oscillation Theorem;
3) Section IV provides the proof of the second part of the main result for the case of $n = 1$. Specifically, an explicit

[2]Also known as Redundancy-Capacity Theorem (see, e.g., [13], [14].)

upper bound on the number of extreme points of an arbitrary output pdf of the Gaussian channel described in (1) is derived. The proof of this result exploits the analyticity of the Gaussian density together with Tijdeman's Number of Zeros Lemma [29, Lemma 1]. The proof for the vector case ($n > 1$) follows along the same lines as the proof for the scalar case ($n = 1$), albeit with a more involved algebra, therefore it is relegated to the Appendix; and

4) Section V concludes the paper with some final remarks.

C. Notation

Throughout the paper, the deterministic scalar quantities are denoted by lower-case letters, deterministic vectors are denoted by bold lowercase letters, random variables are denoted by uppercase letters, and random vectors are denoted by bold uppercase (e.g., x, $\boldsymbol{x}$, X, $\boldsymbol{X}$). We denote the distribution of a random vector $\boldsymbol{X}$ by $P_{\boldsymbol{X}}$. Moreover, we say that a point $\boldsymbol{x}$ is in the support, denoted by $\mathsf{supp}(P_{\boldsymbol{X}})$,[3] of the distribution $P_{\boldsymbol{X}}$ if for every open set $\mathcal{O} \ni \boldsymbol{x}$ we have that $P_{\boldsymbol{X}}(\mathcal{O}) > 0$. We refer to symmetric random variables as those that are symmetric with respect to the origin.

The number of zeros of a function $f : \mathbb{R} \to \mathbb{R}$ on the interval $\mathcal{I}$ is denoted by[4] $\mathrm{N}(\mathcal{I}, f)$. Similarly, if $f : \mathbb{C} \to \mathbb{C}$ is a function on the complex domain, $\mathrm{N}(\mathcal{D}, f)$ denotes the number of its zeros within the region $\mathcal{D}$.

Finally, while the relative entropy between X and Y is denoted by $D(X\|Y)$, the entropy of a discrete random variable X is denoted by $H(X)$ and the differential entropy of a continuous random variable X is denoted by $h(X)$.

II. Main Results

Theorem 1, stated below, gives the first firm upper bound on the support size of the capacity-achieving input of the scalar additive Gaussian channel with an amplitude constraint.

Theorem 1. *Consider the amplitude constrained scalar additive Gaussian channel $Y = X + Z$ where the input X, satifying $|X| \leq \mathsf{A}$, is assumed to be independent from the noise $Z \sim \mathcal{N}(0, 1)$. Assuming $\mathsf{A} \geq 1$, let $P_{X^\star}$ be the optimizing input distribution for this channel. Then, $P_{X^\star}$ is a symmetric discrete distribution with*

$$\frac{1}{2}\mathrm{N}\left([-R, R], f_{Y^\star} - \kappa_1\right) \leq |\mathsf{supp}(P_{X^\star})| \tag{5}$$

$$\leq \mathrm{N}\left([-R, R], f_{Y^\star} - \kappa_1\right) \tag{6}$$

$$< \infty, \tag{7}$$

where $\kappa_1 = \mathrm{e}^{-C(\mathsf{A})-h(Z)}$ and[5] *$R = \mathsf{A} + \log^{\frac{1}{2}}\left(\frac{1}{2\pi\kappa_1^2}\right)$. Moreover,*

$$\sqrt{1 + \frac{2\mathsf{A}^2}{\pi \mathrm{e}}} \leq |\mathsf{supp}(P_{X^\star})| \tag{8}$$

$$\leq \mathrm{N}\left([-R, R], f_{Y^\star} - \kappa_1\right) \tag{9}$$

$$\leq a_2\mathsf{A}^2 + a_1\mathsf{A} + a_0, \tag{10}$$

with

$$a_2 = 9\mathrm{e} + 6\sqrt{\mathrm{e}} + 5, \tag{11}$$

$$a_1 = 6\mathrm{e} + 2\sqrt{\mathrm{e}}, \tag{12}$$

$$a_0 = \mathrm{e} + 2\log\left(4\sqrt{\mathrm{e}} + 2\right) + 1. \tag{13}$$

Since it consists of two parts, the proof of Theorem 1 is divided into two sections. While Section III proves the order tight bounds (5) and (6), Section IV finds the lower and upper bounds presented in (8) and (10).

Remark 1. Observe that the bounds in (5) and (6) are order tight. While the same cannot be said about the bounds in (8) and (10), we conjecture that the order of the lower bound in (8) is the one that is tight. A possible approach for tightening the upper bound is discussed in Section IV along with a figure that supports our conjecture, see Figure 2.

Theorem 2. *Consider the amplitude constrained vector additive Gaussian channel $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}$ where the input $\boldsymbol{X}$, satisfying $\|\boldsymbol{X}\| \leq \mathsf{A}$, is assumed to be independent from the white Gaussian noise $\boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. Let $\boldsymbol{X}^\star \sim P_{\boldsymbol{X}^\star}$ be the optimizing input for this channel. Then, $P_{\boldsymbol{X}^\star}$ is unique, radially symmetric, and the distribution of its amplitude, namely $P_{\|\boldsymbol{X}^\star\|}$, is a discrete distribution with*

$$\left|\mathsf{supp}\left(P_{\|\boldsymbol{X}^\star\|}\right)\right| \leq \mathfrak{a}_{n_2}\mathsf{A}^2 + \mathfrak{a}_{n_1}\mathsf{A} + \mathfrak{a}_{n_0}, \tag{14}$$

where, denoting the gamma function by Γ,

$$\mathfrak{a}_{n_2} = 4 + 4\mathrm{e} + \sqrt{8\mathrm{e} + 4}, \tag{15}$$

$$\mathfrak{a}_{n_1} = \left(3 + 4\mathrm{e} + \sqrt{2\mathrm{e} + 1}\right) n + \sqrt{\frac{32}{n - 1}}, \tag{16}$$

$$\mathfrak{a}_{n_0} = \log \frac{\mathrm{e}^2\sqrt{\pi}\,\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} + \left(3 + 4\mathrm{e} + \sqrt{2\mathrm{e} + 1}\right)\left(\frac{n}{2} + \log\frac{\sqrt{\pi}\,\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right). \tag{17}$$

The proof of Theorem 2 benefits from the same technique that is used in the proof of Theorem 1. For this reason, its presentation is postponed to Appendix A.

Remark 2. Note that when the vector channel is of dimension 2, Theorem 2 gives an upper bound on the number of shells of the optimal input distribution for the additive complex Gaussian channel with an amplitude constraint.

For the sake of demonstrating the versatility of our novel method, proven next is an upper bound on the support size of the optimal input distribution for the scalar additive Gaussian channel with both an amplitude and a power constraint.

Theorem 3. *Consider the amplitude and power constrained scalar additive Gaussian channel $Y = X + Z$ where the input X, satisfying $|X| \leq \mathsf{A}$ and $\mathbb{E}[|X|^2] \leq \mathsf{P}$, is assumed to be independent from the noise $Z \sim \mathcal{N}(0, 1)$. Assuming $\mathsf{A} \geq 1$, let $P_{X^\star}$ be the optimizing input distribution for this channel. Then, $P_{X^\star}$ is a symmetric discrete distribution with*

$$\sqrt{1 + \frac{2\min\left\{\mathsf{A}^2, 3\mathsf{P}\right\}}{\pi \mathrm{e}}} \leq |\mathsf{supp}(P_{X^\star})| \tag{18}$$

[3] Also known as "points of increase of $P_{\boldsymbol{X}}$" or "spectrum of $P_{\boldsymbol{X}}$."

[4] The definition $\mathrm{N}(\mathcal{I}, f)$ is blind to the multiplicities of the zeros.

[5] Unless otherwise stated, the logarithms in this paper are of base e.

$$\leq a_{\mathsf{P}_2}\mathsf{A}_{\mathsf{P}}^2 + a_{\mathsf{P}_1}\mathsf{A}_{\mathsf{P}} + a_{\mathsf{P}_0}, \tag{19}$$

where

$$\mathsf{A}_{\mathsf{P}} = \frac{\mathsf{AP}}{\mathsf{P} - \log(1+\mathsf{P})1\{\mathsf{P} < \mathsf{A}^2\}}, \tag{20}$$

$$a_{\mathsf{P}_2} = (1+2\lambda_{\mathsf{P}})(9\mathrm{e} + 6\sqrt{\mathrm{e}} + 1) + 2(2-\lambda_{\mathsf{P}})(1-2\lambda_{\mathsf{P}}), \tag{21}$$

$$a_{\mathsf{P}_1} = (1+2\lambda_{\mathsf{P}})(6\mathrm{e} + 2\sqrt{\mathrm{e}}), \tag{22}$$

$$a_{\mathsf{P}_0} = (1+2\lambda_{\mathsf{P}})\mathrm{e} + 2\log\left(\frac{2+4\sqrt{\mathrm{e}}(1+2\lambda_{\mathsf{P}})}{1-2\lambda_{\mathsf{P}}}\right) + 1, \tag{23}$$

$$\lambda_{\mathsf{P}} = \frac{\log(1+\mathsf{P})}{2\mathsf{P}} \cdot 1\left\{\mathsf{P} < \mathsf{A}^2\right\}. \tag{24}$$

With only small alterations, proof of Theorem 3 imitates that of Theorem 1 and is shown in Appendix C.

Remark 3. In the case when $\mathsf{P} \geq \mathsf{A}^2$, the power constraint becomes inactive and Theorem 3 recovers the result of Theorem 1.

III. Proof for the First Part of Theorem 1

This section proves the first part of our main result in Theorem 1, namely the bounds in (5) and (6).

A. On Equations Characterizing the Support of $P_{X^\star}$

The first ingredient of the proof is the following characterization of the optimal input distribution shown in [1, Corollary 1].

Lemma 1. *Consider the amplitude constrained scalar additive Gaussian channel $Y = X + Z$ where the input X, satisfying $|X| \leq \mathsf{A}$, is independent from the noise $Z \sim \mathcal{N}(0,1)$. Then, $P_{X^\star}$ is the capacity-achieving input distribution if and only if the following two equations are satisfied*:

$$i(x; P_{X^\star}) = C(\mathsf{A}), \quad x \in \mathsf{supp}(P_{X^\star}), \tag{25}$$

$$i(x; P_{X^\star}) \leq C(\mathsf{A}), \quad x \in [-\mathsf{A}, \mathsf{A}], \tag{26}$$

where $C(\mathsf{A})$ denotes the capacity of the channel, and

$$i(x; P_{X^\star}) = \int_{\mathbb{R}} \frac{\mathrm{e}^{-\frac{(y-x)^2}{2}}}{\sqrt{2\pi}} \log\frac{1}{f_{Y^\star}(y)}\mathrm{d}y - h(Z), \tag{27}$$

with $h(Z) = \log\sqrt{2\pi\mathrm{e}}$ denoting the differential entropy of the standard Gaussian distribution, and $f_{Y^\star}(y)$ denoting the output pdf induced by the input $P_{X^\star}$, that is, for $X \sim P_{X^\star}$,

$$f_{Y^\star}(y) = \frac{1}{\sqrt{2\pi}}\mathbb{E}\left[\mathrm{e}^{-\frac{(y-X)^2}{2}}\right]. \tag{28}$$

Remark 4. An immediate consequence of Lemma 1 is the fact that

$$\mathsf{supp}(P_{X^\star}) \subseteq \{x : i(x; P_{X^\star}) - C(\mathsf{A}) = 0\},$$

which leads to the following inequalities on the size of the support:

$$|\mathsf{supp}(P_{X^\star})| \leq \mathrm{N}([-\mathsf{A}, \mathsf{A}], \Xi_{\mathsf{A}}(\cdot; P_{X^\star})) \tag{29}$$

$$\leq \mathrm{N}(\mathbb{R}, \Xi_{\mathsf{A}}(\cdot; P_{X^\star})), \tag{30}$$

where the function $\Xi(\cdot; P_{X^\star})\colon \mathbb{R} \to \mathbb{R}$ is defined as

$$\Xi_{\mathsf{A}}(x; P_{X^\star}) = i(x; P_{X^\star}) - C(\mathsf{A}). \tag{31}$$

Note that, as it stands, the upper bound in (29) does not yet reveal any information on the discreteness of $P_{X^\star}$ as the right side might just as well be ∞.

B. Connecting the Number of Oscillations of $f_{Y^\star}$ to the Number of Masses in $P_{X^\star}$

This section gives an alternative proof that $P_{X^\star}$ is discrete by relating the cardinality of $\mathsf{supp}(P_{X^\star})$ to the number of zeros of the shifted output pdf $f_{Y^\star} - \mathrm{e}^{-C(\mathsf{A})-h(Z)}$. The following definition sets the stage.

Definition 1. *Sign Changes of a Function. The number of sign changes of a function ξ is given by*

$$\mathscr{S}(\xi) = \sup_{m\in\mathbb{N}}\left\{\sup_{y_1<\cdots<y_m} \mathscr{N}\{\xi(y_i)\}_{i=1}^m\right\}, \tag{32}$$

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

Proven in [27], the following theorem is the main tool in connecting the number of zeros of a shifted output pdf $f_{Y^\star}$ to the number of mass points of a capacity-achieving input distribution $P_{X^\star}$.

Theorem 4. *Oscillation Theorem [27]. Given open intervals $\mathbb{I}_1$ and $\mathbb{I}_2$, let $p\colon \mathbb{I}_1 \times \mathbb{I}_2 \to \mathbb{R}$ be a strictly totally positive kernel.*[6] *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 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*

$$\Xi(x) = \int \xi(y)p(x,y)\mathrm{d}\mu(y). \tag{33}$$

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

Note that Theorem 4 is applicable in our setting as the Gaussian distribution is a strictly totally positive kernel [27]. The following result shows the connection between the support size of $P_{X^\star}$ and the number of zeros of the shifted optimal output pdf $f_{Y^\star}$ and recovers the bounds in (5) and (6).

Theorem 5. *The support set of the capacity-achieving input distribution $P_{X^\star}$ satisfies*

$$\frac{1}{2}\mathrm{N}\left([-R, R], f_{Y^\star} - \kappa_1\right) \leq |\mathsf{supp}(P_{X^\star})| \tag{34}$$

$$\leq \mathrm{N}\left([-R, R], f_{Y^\star} - \kappa_1\right) \tag{35}$$

$$< \infty, \tag{36}$$

where $\kappa_1 = \mathrm{e}^{-C(\mathsf{A})-h(Z)}$ and[7] *$R > \mathsf{A} + \log^{\frac{1}{2}}\left(\frac{1}{2\pi\kappa_1^2}\right)$.*

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

[7] See Remark 8 and observe that $\kappa_1 \in \left(0, \frac{1}{\sqrt{2\pi}}\right)$.

Proof: To see (35) and (36), observe that $\Xi_{\mathsf{A}}(x;P_{X^\star})$, defined in (31), can be written as follows:

$$\Xi_{\mathsf{A}}(x;P_{X^\star}) = \int_{\mathbb{R}} \frac{\xi_{\mathsf{A}}(y)}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(y-x)^2}{2}} \mathrm{d}y, \tag{37}$$

where

$$\xi_{\mathsf{A}}(y) = \log \frac{1}{f_{Y^\star}(y)} - C(\mathsf{A}) - h(Z). \tag{38}$$

First, observe that it is impossible for $\Xi_{\mathsf{A}}(x;P_{X^\star}) = 0$ for all $x \in \mathbb{R}$ since otherwise $\xi_{\mathsf{A}}(y)$ would be zero for all $y \in \mathbb{R}$. Furthermore, using the fact that the Gaussian distribution is a strictly totally positive kernel,

$$\begin{aligned} |\mathsf{supp}(P_{X^\star})| &\le \mathrm{N}(\mathbb{R}, \Xi_{\mathsf{A}}(\cdot;P_{X^\star})) && (39)\\ &\le \mathscr{S}(\xi_{\mathsf{A}}) && (40)\\ &\le \mathrm{N}(\mathbb{R}, \xi_{\mathsf{A}}) && (41)\\ &= \mathrm{N}\left(\mathbb{R}, f_{Y^\star} - \kappa_1\right) && (42)\\ &= \mathrm{N}\left([-R,R], f_{Y^\star} - \kappa_1\right) && (43)\\ &< \infty, && (44) \end{aligned}$$

where (39) is a consequence of Lemma 1 (see Remark 4); (40) follows from Theorem 4; (41) follows because the number of zeros is an upper bound on the number of sign changes; (42) follows by observing that $\xi_{\mathsf{A}}(y) = 0$ if and only if $f_{Y^\star}(y) = \kappa_1$; and finally (43) and (44) follow from Lemma 2 in Section IV.

To see (34), using the fact that $\mathsf{supp}(P_{X^\star})$ is a finite set, suppose $|\mathsf{supp}(P_{X^\star})| = n$, let $x_1 < \ldots < x_n$ the elements of $\mathsf{supp}(P_{X^\star})$, and write

$$f_{Y^\star}(y) = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{n} P_{X^\star}(x_i) \exp\left(-\frac{1}{2}(y-x_i)^2\right), \tag{45}$$

where, by the definition of $\mathsf{supp}(P_{X^\star})$, the probabilities satisfy $P_{X^\star}(x_i) > 0$ for each $i = 1, \ldots, n$. Observing that the number of zeros of $f_{Y^\star}(y) - \kappa_1$ is the same as the number of zeros of the right-shifted function $f_{Y^\star}(y - |x_1| - 1) - \kappa_1$, let

$$\begin{aligned} f(y) &= f_{Y^\star}(y - |x_1| - 1) - \kappa_1 && (46)\\ &= \sum_{i=1}^{n} a_i \exp\left(-\frac{1}{2}(y-u_i)^2\right) - a_0, && (47) \end{aligned}$$

where for $i = 0, 1, \ldots, n$ both $a_i > 0$ and $u_i > 0$ as

$$u_i = x_i + |x_1| + 1 \quad \text{for } i = 1, \ldots n, \tag{48}$$

$$a_i = \begin{cases} \kappa_1 & i = 0 \\ \frac{1}{\sqrt{2\pi}} P_{X^\star}(x_i) & i = 1, \ldots, n. \end{cases} \tag{49}$$

Given arbitrary $0 < \epsilon_1 < \cdots < \epsilon_n$, consider the perturbed function

$$\widetilde{f}(y, \epsilon_1, \ldots, \epsilon_n) = \sum_{i=1}^{n} a_i \exp\left(-\frac{1}{2}(1+\epsilon_i)(y-u_i)^2\right) - a_0. \tag{50}$$

Note that

$$\mathrm{e}^{-\frac{1}{2}y^2} \widetilde{f}(y, \epsilon_1, \ldots, \epsilon_n) = \sum_{i=0}^{n} b_i \exp\left(-\frac{(2+\epsilon_i)}{2}(y-v_i)^2\right), \tag{51}$$

where

$$\epsilon_0 = -1, \tag{52}$$

$$b_i = \begin{cases} -a_0 & i = 0, \\ a_i \exp\left(-\frac{1+\epsilon_i}{2(2+\epsilon_i)} u_i^2\right) & i = 1, \ldots, n, \end{cases} \tag{53}$$

$$v_i = \begin{cases} 0 & i = 0, \\ \frac{1+\epsilon_i}{2+\epsilon_i} u_i & i = 1, \ldots, n, \end{cases} \tag{54}$$

is a linear combination of $n+1$ distinct Gaussians with distinct variances and therefore has at most $2n$ zeros [30, Proposition 7]. Since this holds for any arbitrary choice of ϵ_i's and since $\widetilde{f}(y, \epsilon_1, \ldots, \epsilon_n) \to f(y)$ as $(\epsilon_1, \ldots, \epsilon_n) \to (0, \ldots, 0)$, it follows that

$$\begin{aligned} 2|\mathsf{supp}(P_{X^\star})| &\ge \mathrm{N}(\mathbb{R}, f(y)) && (55)\\ &= \mathrm{N}(\mathbb{R}, f_{Y^\star}(y) - \kappa_1) && (56)\\ &= \mathrm{N}([R,R], f_{Y^\star}(y) - \kappa_1). && (57) \end{aligned}$$

■

Remark 5. With a different approach than the one taken in [1], observe that Theorem 5 recovers the result of Smith [1] showing that the support set of $P_{X^\star}$ is finite; hence, $P_{X^\star}$ is discrete with finitely many mass points. An advantage of the proof presented here is that the Fourier analysis required in the proof provided by [1] is now completely avoided. Another advantage is that, since the presented proof is of the constructive nature, one can indeed attempt at counting the zeros of $f_{Y^\star} - \kappa_1$, which is the topic of the next section.

Remark 6. Theorem 5 proves that the number of zeros of the shifted optimal output pdf $f_{Y^\star} - \kappa_1$ gives an order tight upper bound on the support size of the optimal input pmf $|\mathsf{supp}(P_{X^\star})|$. This result can be considered as the main result of this paper.

IV. Proof of the Explicit Bounds in (8) and (10)

Section III demonstrates that the number of mass points of $P_{X^\star}$ is within a factor of two of the number of zeros of $f_{Y^\star} - \kappa_1$ where $\kappa_1 = \mathrm{e}^{-C(\mathsf{A})-h(Z)}$. In this section, we first provide an upper bound on the number of zeros of $f_{Y^\star} - \kappa_1$ and establish (10). Additionally, through the use of entropy-power inequality, we also provide a lower bound on the support size of $P_{X^\star}$, yielding (8).

Remark 7. A critical observation here is that, due to the lack of knowledge of the optimal input distribution $P_{X^\star}$ or the capacity expression $C(A)$, we do not know the optimal output distribution $f_{Y^\star}$ nor the constant $\kappa_1 = \mathrm{e}^{-C(A)-h(Z)}$. Therefore, we must instead work with generic f_Y and κ_1 throughout this section.

A. Bounds on the Number of Extreme Points of a Gaussian Convolution

The aim of this subsection of the paper is to study the following problem: given an *unknown constant* $0 \le \kappa_1 \le \max_b f_Y(b)$, find a worst-case upper bound on the number of zeros of the shifted output pdf

$$f_Y - \kappa_1,$$

where f_Y denotes the pdf of the random variable $Y = X + Z$, with X being an *arbitrary* zero mean[8] random variable at the input of the channel satisfying the amplitude constraint: $|X| \leq \mathsf{A}$; Z being the standard Gaussian random variable independent from X; and Y being the random variable induced by the input X at the output of this additive Gaussian channel.

As a starting point, before chasing after the number of zeros of $f_Y - \kappa_1$, the following lemma shows that the zeros of $f_Y - \kappa_1$ are always contained on an interval that is only "slightly" larger than $[-\mathsf{A}, \mathsf{A}]$.

Lemma 2. *On the Location and Finiteness of Zeros. For a fixed $\kappa_1 \in \left(0, \frac{1}{\sqrt{2\pi}}\right]$ there exists some $\mathsf{B}_{\kappa_1} = \mathsf{B}_{\kappa_1}(\mathsf{A}) < \infty$ such that*

$$\mathrm{N}(\mathbb{R}, f_Y - \kappa_1) = \mathrm{N}([-\mathsf{B}_{\kappa_1}, \mathsf{B}_{\kappa_1}], f_Y - \kappa_1) \tag{58}$$
$$< \infty. \tag{59}$$

In other words, there are finitely many zeros of $f_Y(y) - \kappa_1$ all of which are contained within the interval $[-\mathsf{B}_{\kappa_1}, \mathsf{B}_{\kappa_1}]$. Moreover, B_{κ_1} can be upper bounded as follows:

$$\mathsf{B}_{\kappa_1} \leq \mathsf{A} + \log^{\frac{1}{2}}\left(\frac{1}{2\pi\kappa_1^2}\right). \tag{60}$$

Proof: Using the monotonicity of e^{-u}, for all $|y| > \mathsf{A}$,

$$f_Y(y) = \frac{1}{\sqrt{2\pi}}\mathbb{E}\left[\mathrm{e}^{-\frac{(y-X)^2}{2}}\right] \tag{61}$$
$$\leq \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(y-\mathsf{A})^2}{2}}. \tag{62}$$

Since the right side of (62) is a decreasing function for all $|y| > A$, it follows that

$$f_Y(y) - \kappa_1 < 0 \tag{63}$$

for all

$$|y| > \mathsf{A} + \log^{\frac{1}{2}}\left(\frac{1}{2\pi\kappa_1^2}\right). \tag{64}$$

This means that there exists B_{κ_1} satisfying (60) such that all zeros of $f_Y - \kappa_1$ are located within the interval $[-\mathsf{B}_{\kappa_1}, \mathsf{B}_{\kappa_1}]$.

To see that there are finitely many zeros, recall the fact that a convolution with a Gaussian distribution preserves analyticity [31, Proposition 8.10]; hence f_Y is an analytic function on $\mathbb{R}$. Standard methods (e.g., invoking Bolzano-Weierstrass Theorem and the Identity Theorem) yield the fact that analytic functions have finitely many zeros on a compact interval, which is the desired result. ■

Since the exact value of the constant κ_1 is unknown, in counting the number of zeros of $f_Y - \kappa_1$, a worst-case approach needs to be taken. In an attempt at doing so, the following elementary result from calculus provides a bound on the number of zeros of a function in terms of the number of its extreme points. As simple as it is, Lemma 3 is one of the key steps in this paper. It states that, to find a bound on the number of zeros of $f_Y - \kappa_1$, it suffices to find a bound on that of f_Y', eliminating the dependence on the nuisance constant κ_1.

Lemma 3. *Suppose that f is continuous on $[-R, R]$ and differentiable on $(-R, R)$. If $\mathrm{N}([-R, R], f) < \infty$, then*

$$\mathrm{N}([-R, R], f) \leq \mathrm{N}([-R, R], f') + 1, \tag{65}$$

where f' denotes the derivative of f.

Proof: Let $x_1 < \ldots < x_{\mathrm{n}_0}$ denote the zeros of f. By Rolle's Theorem, each of the intervals (x_i, x_{i+1}) for $i = 1, \ldots, \mathrm{n}_0 - 1$ contains at least one extreme point. ■

Thanks to Lemma 3, to upper bound the number of zeros of $f_Y - \kappa_1$, all that is needed is to find an upper bound on the number of zeros of the derivative of f_Y, namely

$$f_Y'(y) = \frac{1}{\sqrt{2\pi}}\mathbb{E}\left[(X - y)\exp\left(-\frac{(y-X)^2}{2}\right)\right]. \tag{66}$$

At this point, there are several trajectories that one could follow. For example, using the fact that f_Y' is an analytic function (cf. [2, Appendix B]) and letting $\breve{f}_Y'$ denote its complex analytic extension,

$$\mathrm{N}([-R, R], f_Y') \leq \inf_{\epsilon > 0} \mathrm{N}(\mathcal{D}_{R+\epsilon}, \breve{f}_Y') \tag{67}$$
$$= \inf_{\epsilon > 0} \frac{1}{2\pi i} \oint_{\partial\mathcal{D}_{R+\epsilon}} \frac{\breve{f}_Y''(z)}{\breve{f}_Y'(z)} \mathrm{d}z \tag{68}$$
$$\leq \inf_{\epsilon > 0}(R + \epsilon) \max_{|z| = R+\epsilon} \left|\frac{\breve{f}_Y''(z)}{\breve{f}_Y'(z)}\right|, \tag{69}$$

where in (67) $\mathcal{D}_t \subset \mathbb{C}$ is an open disc[9] of radius t centered at the origin and the inequality follows because $f_Y'(y) = 0 \implies \breve{f}_Y'(y) = 0$; in (68) $\partial\mathcal{D}_t$ denotes the boundary of the disc $\mathcal{D}_t$ and equality follows from Cauchy's argument principle (e.g., [32, Corollary 10.9]); and finally (69) follows from the ML inequality for the contour integral [32, Chapter 4.10].

Unfortunately, due to the implicit definitions of the functions f_Y'' and f_Y', the maximization of the ratio $\breve{f}_Y''/\breve{f}_Y'$ in the right side of (69) does not seem to have a tractable explicit solution. Luckily, there are alternative, more tractable methods that yield an explicit upper bound on the number of zeros of $\breve{f}_Y'$. The method used in this paper is based on Tijdeman's Number of Zeros Lemma, which is presented next.

Lemma 4. *Tijdeman's Number of Zeros Lemma [29]. Let R, s, t be positive numbers such that $s > 1$. For the complex valued function $f \neq 0$ which is analytic on $|z| < (st+s+t)R$, its number of zeros $\mathrm{N}(\mathcal{D}_R, f)$ within the disk $\mathcal{D}_R = \{z \colon |z| \leq R\}$ satisfies*

$$\mathrm{N}(\mathcal{D}_R, f) \\ \leq \frac{1}{\log s}\left(\log \max_{|z| \leq (st+s+t)R} |f(z)| - \log \max_{|z| \leq tR} |f(z)|\right). \tag{70}$$

[8]Since the channel is symmetric, the capacity-achieving input is symmetric. Therefore, there is no loss of optimality in restricting attention to zero mean inputs.

[9]In fact, $\mathcal{D}_R$ can be any open connected set that contains the interval $[-R, R]$. For example, a rectangle of width $2(R + \epsilon)$ and arbitrary hight 2ϵ is a typical choice.

The following two lemmas find upper and lower bounds on absolute value of the complex analytic extension[10] of f'_Y over a disc of finite radius centered at the origin.

Lemma 5. *Suppose $f'_Y\colon \mathbb{R} \to \mathbb{R}$ is as in (66) and let $\breve{f}'_Y\colon \mathbb{C} \to \mathbb{C}$ denote its complex extension. Then,*

$$\max_{|z|\le \mathsf{B}} \left|\breve{f}'_Y(z)\right| \le \frac{1}{\sqrt{2\pi}}(\mathsf{A}+\mathsf{B})\exp\left(\frac{\mathsf{B}^2}{2}\right). \tag{71}$$

Proof: Using the standard rectangular representation of a complex number, let $z=\xi+i\eta\in\mathbb{C}$,

$$\begin{aligned}
&\max_{|z|\le \mathsf{B}} \left|\breve{f}'_Y(z)\right| \\
&= \max_{|z|\le \mathsf{B}} \left\{\frac{1}{\sqrt{2\pi}}\left|\mathbb{E}\left[(X-z)\exp\left(-\frac{(z-X)^2}{2}\right)\right]\right|\right\} && (72)\\
&\le \max_{|z|\le \mathsf{B}} \left\{\frac{1}{\sqrt{2\pi}}\mathbb{E}\left[|X-z|\left|\exp\left(-\frac{(z-X)^2}{2}\right)\right|\right]\right\} && (73)\\
&= \max_{|z|\le \mathsf{B}} \left\{\frac{1}{\sqrt{2\pi}}\mathbb{E}\left[|X-z|\exp\left(\frac{\eta^2-(\xi-X)^2}{2}\right)\right]\right\} && (74)\\
&\le \max_{|z|\le \mathsf{B}} \left\{\frac{1}{\sqrt{2\pi}}\mathbb{E}\left[(|X|+|z|)\exp\left(\frac{\eta^2}{2}\right)\right]\right\} && (75)\\
&\le \frac{1}{\sqrt{2\pi}}(\mathsf{A}+\mathsf{B})\exp\left(\frac{\mathsf{B}^2}{2}\right), && (76)
\end{aligned}$$

where (73) follows from Jensen's inequality; (75) follows from triangle inequality; and finally (76) is because $|X|\le\mathsf{A}$, and $|z|\le\mathsf{B}$ implies $|\eta|\le\mathsf{B}$. ■

Lemma 6. *Suppose $f'_Y\colon \mathbb{R} \to \mathbb{R}$ is as in (66) and let $\breve{f}'_Y\colon \mathbb{C} \to \mathbb{C}$ denote its complex extension. For any $|X|\le\mathsf{A}\le\mathsf{B}$,*

$$\max_{|z|\le \mathsf{B}} \left|\breve{f}'_Y(z)\right| \ge \frac{\mathsf{A}}{\sqrt{2\pi}}\exp\left(-2\mathsf{A}^2\right). \tag{77}$$

Proof: Thanks to the suboptimal choice of $z=\mathsf{A}\le\mathsf{B}$,

$$\begin{aligned}
\max_{|z|\le \mathsf{B}} \left|\breve{f}'_Y(z)\right| &\ge \frac{1}{\sqrt{2\pi}}\left|\mathbb{E}\left[(X-\mathsf{A})\exp\left(-\frac{(\mathsf{A}-X)^2}{2}\right)\right]\right| && (78)\\
&\ge \frac{1}{\sqrt{2\pi}}\mathbb{E}\left[(\mathsf{A}-X)\exp\left(-2\mathsf{A}^2\right)\right] && (79)\\
&= \frac{\mathsf{A}}{\sqrt{2\pi}}\exp\left(-2\mathsf{A}^2\right), && (80)
\end{aligned}$$

where (79) follows because $|X|\le\mathsf{A}$; and (80) is a consequence of $\mathbb{E}[X]=0$. ■

By assembling the results of Lemmas 4, 5 and 6, Theorem 6 below provides an upper bound on the number of oscillations of a Gaussian convolution.

Theorem 6. *Bound on the Number of Oscillations of f_Y. Let $|X|\le\mathsf{A}<R$ for some fixed R. Then, the number of extreme points of f_Y, namely the number of zeros of f'_Y, within the interval $[-R,R]$ satisfies*

$$\begin{aligned}
&\mathrm{N}([-R,R],f'_Y)\\
&\le \min_{s>1}\left\{\frac{\left(\frac{((\mathsf{A}+R)s+\mathsf{A})^2}{2}+2\mathsf{A}^2+\log\left(2+\frac{(\mathsf{A}+R)s}{\mathsf{A}}\right)\right)}{\log s}\right\}. && (81)
\end{aligned}$$

Proof: Let $\mathcal{D}_R\subset\mathbb{C}$ be a disk of radius R centered at $z_0=0$, and note that

$$\begin{aligned}
&\mathrm{N}([-R,R],f'_Y)\\
&\le \mathrm{N}(\mathcal{D}_R,\breve{f}'_Y) && (82)\\
&\le \min_{s>1,\,t\ge\frac{\mathsf{A}}{R}}\left\{\frac{\log\frac{\max_{|z|\le(st+s+t)R}|\breve{f}'_Y(z)|}{\max_{|z|\le tR}|\breve{f}'_Y(z)|}}{\log s}\right\} && (83)\\
&\le \min_{s>1,\,t\ge\frac{\mathsf{A}}{R}}\left\{\frac{\frac{(st+s+t)^2R^2}{2}+2\mathsf{A}^2+\log\left(1+\frac{(st+s+t)R}{\mathsf{A}}\right)}{\log s}\right\} && (84)\\
&= \min_{s>1}\left\{\frac{\frac{((\mathsf{A}+R)s+\mathsf{A})^2}{2}+2\mathsf{A}^2+\log\left(2+\frac{(\mathsf{A}+R)s}{\mathsf{A}}\right)}{\log s}\right\}, && (85)
\end{aligned}$$

where (82) follows because zeros of f'_Y are also zeros of its complex extension $\breve{f}'_Y$; (83) is a consequence of Lemma 4; (84) follows from Lemmas 5 and 6; and finally, in (85), we use the fact that $t=\frac{A}{R}$ is the minimizer in the right hand side of (84). ■

Finally, combining the results of Lemmas 2 and 3, and Theorem 6, the following corollary presents the desired result of this section.

Corollary 1. *Given an arbitrary constant $\kappa_1\in\left(0,\frac{1}{2\pi}\right)$, suppose $R>\mathsf{A}+\log^{\frac12}\left(\frac{1}{2\pi\kappa_1^2}\right)$. Then, the number of zeros of $f_Y-\kappa_1$ satisfies*

$$\begin{aligned}
&\mathrm{N}(\mathbb{R},f_Y-\kappa_1)\\
&= \mathrm{N}([-R,R],f_Y-\kappa_1) && (86)\\
&\le 1+\min_{s>1}\left\{\frac{\left(\frac{((\mathsf{A}+R)s+\mathsf{A})^2}{2}+2\mathsf{A}^2+\log\left(2+\frac{(\mathsf{A}+R)s}{\mathsf{A}}\right)\right)}{\log s}\right\}. && (87)
\end{aligned}$$

Remark 8. Observe that in presenting the main result of this section, a "worst-case scenario" approach is taken. Indeed, the result in (87) is independent of the choice of κ_1. If $\kappa_1\approx 0$, then $\mathrm{N}(\mathbb{R},f_Y-\kappa_1)\le 2$ and the bound above may be quite loose. In applying Corollary 1 in the next section, we let

$$\kappa_1=\frac{\mathrm{e}^{-C(\mathsf{A})}}{\sqrt{2\pi\mathrm{e}}} \tag{88}$$

where $C(\mathsf{A})$ denotes the capacity of the amplitude constrained additive Gaussian channel. In that case, it can be shown that

$$\left(2\pi\mathrm{e}\left(1+\mathsf{A}^2\right)\right)^{-\frac12}\le\kappa_1\le\left(2\pi\mathrm{e}+4\mathsf{A}^2\right)^{-\frac12}, \tag{89}$$

and the result presented above is more relevant.

[10]The fact that the complex extension of f_Y, and hence that of f'_Y, is analytic on $\mathbb{C}$ is proven in [2, Appendix B].

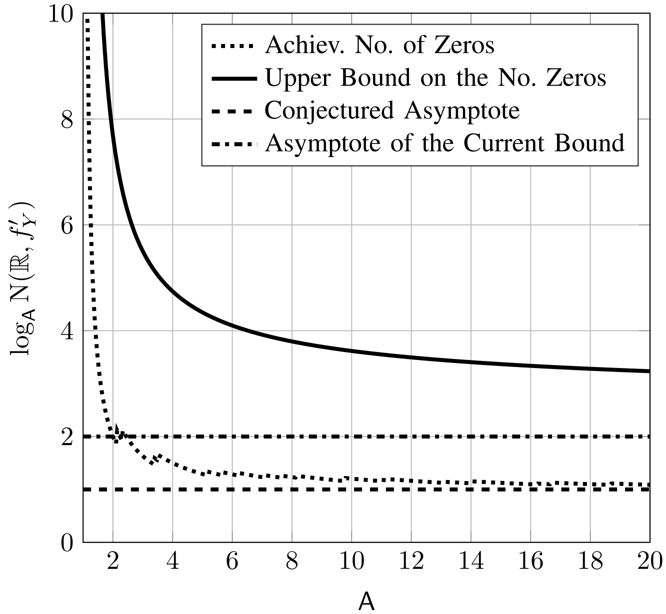

Fig. 2. Plot of the logarithm (in base A) of the number of zeros of f'_Y. The solid black line uses the upper bound on the number of zeros in Corollary 1 with the bound on κ_1 in (89) and the dashed-dotted line is the asymptote of this upper bound. The dashed line is the conjectured asymptote in (90). The dotted line is the number of zeros found through an exhaustive numerical search.

Remark 9. We believe that the bound can be further tightened. In fact, we conjecture that[11]

$$\max_{X \in [-\mathsf{A},\mathsf{A}]} \mathrm{N}(\mathbb{R}, f'_{Y^*}) = \Theta(\mathsf{A}). \tag{90}$$

Fig. 2 demonstrates a result of an extensive computer search that supports the claim in (90) and compares it to the current best upper bound in Corollary 1.

B. Proof of the Upper Bound in (10)

We begin by simplifying the previously provided upper bound on B_{κ_1}. Note that an amplitude constraint $|X| \le \mathsf{A}$ induces a second moment constraint $\mathbb{E}[X^2] \le \mathsf{A}^2$, and therefore

$$C(\mathsf{A}) = \max_{\substack{|X|\le \mathsf{A} \\ \mathbb{E}[X^2]\le \mathsf{A}^2}} I(X;Y) \tag{91}$$

$$\le \frac{1}{2}\log\left(1+\mathsf{A}^2\right). \tag{92}$$

Since the differential entropy of a standard normal distribution is $h(Z) = \frac{1}{2}\log(2\pi \mathrm{e})$, (92) implies that

$$\frac{1}{\kappa_1} = \exp(C(\mathsf{A}) + h(Z)) \tag{93}$$

$$\le \sqrt{2\pi\mathrm{e}\,(1+\mathsf{A}^2)}. \tag{94}$$

Capitalizing on the bound in (60),

$$\mathsf{B}_{\kappa_1} \le \mathsf{A} + \sqrt{1+\log(1+\mathsf{A}^2)} \tag{95}$$

$$\le 2\mathsf{A}+1, \tag{96}$$

where the last inequality follows because $\log(1+x) \le x$ and $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$.

As the finalizing step, letting $R \leftarrow (2\mathsf{A}+1)$ in Theorem 5 above, an application of Corollary 1 in Section IV yields

$$\mathrm{N}([-\mathsf{B}_{\kappa_1}, \mathsf{B}_{\kappa_1}], f_{Y^\star} - \kappa_1) - 1$$

$$= \mathrm{N}([-2\mathsf{A}-1, 2\mathsf{A}+1], f_{Y^\star} - \kappa_1) - 1 \tag{97}$$

$$\le \min_{s>1}\left\{ \frac{\left(\frac{((3\mathsf{A}+1)s+\mathsf{A})^2}{2} + 2\mathsf{A}^2 + \log\left(2+\frac{(3\mathsf{A}+1)s}{\mathsf{A}}\right)\right)}{\log s} \right\} \tag{98}$$

$$\le \min_{s>1}\left\{ \frac{1}{\log s}\left(\frac{((3s+1)\mathsf{A}+s)^2}{2} + 2\mathsf{A}^2 + \log(2+4s) \right) \right\} \tag{99}$$

$$\le \left(\mathrm{e} + 2\log\left(4\sqrt{\mathrm{e}}+2\right) + (4\mathrm{e}+2\sqrt{\mathrm{e}})\mathsf{A} + (5+4\sqrt{\mathrm{e}}+4\mathrm{e})\mathsf{A}^2\right) \tag{100}$$

$$= a_2\mathsf{A}^2 + a_1\mathsf{A} + a_0 - 1, \tag{101}$$

where (99) follows because $3\mathsf{A}+1 \le 4\mathsf{A}$ for[12] $\mathsf{A} \ge 1$; (100) follows by choosing a suboptimal value $s = \sqrt{\mathrm{e}}$ in the minimization; and (101) follows by letting $a_2 = 9\mathrm{e}+6\sqrt{\mathrm{e}}+5$, $a_1 = 6\mathrm{e}+2\sqrt{\mathrm{e}}$ and $a_0 = \mathrm{e} + 2\log\left(4\sqrt{\mathrm{e}}+2\right)+1$. ■

Remark 10. A more careful optimization of (100) over the parameter s would lead to better absolute constants a_0, a_1 and a_2. However, note that the order A^2 in (100) would not change.

C. Proof of the Lower Bound in (8)

Using the fact that the optimizing input distribution is discrete with finitely many points and denoting by $H(P_{X^\star})$ the entropy of the optimizing input distribution $P_{X^\star}$, it follows that

$$\frac{1}{2}\log\left(1+\frac{2\mathsf{A}^2}{\pi\mathrm{e}}\right) \le \max_{X:\,|X|\le\mathsf{A}} I(X;Y). \tag{102}$$

$$\le H(P_{X^\star}) \tag{103}$$

$$\le \log\left(|\mathsf{supp}(P_{X^\star})|\right), \tag{104}$$

where (102) is a lower bound due to Shannon [7, Section 25]. ■

V. Concluding Remarks

This paper has introduced several new tools for studying the capacity of the amplitude constrained additive Gaussian channels. Not only are the introduced tools strong enough to show that the optimal input distribution is discrete with finite support, but they are also able to provide concrete upper bounds on the number of elements in that support. The main result of this paper is that the number of zeros of the downward shifted optimal output density provides an implicit

[11]Let $f(x)$ and $g(x)$ be two nonnegative valued functions. Then, f is $\Theta(g(x))$ if and only if $c_1\, g(x) \le f(x) \le c_2\, g(x)$ for some $c_1, c_2 > 0$ and all $x > x_0$.

[12]The unessential assumption that $\mathsf{A} \ge 1$ is just for simplifying the presentation. In the case when $\mathsf{A} \le 1$, the optimality of $P_{X^\star}$ that is equiprobable on $\mathcal{X} = \{-\mathsf{A}, \mathsf{A}\}$ is known [8].

upper bound on the support size of the capacity-achieving input distribution. While this upper bound has been shown to be tight within a factor of one half, it can also be used as a means to get an explicit upper bound on the support size.

As a note on its flexibility, the novel method that is described in this paper has been demonstrated to be easily generalizable to other settings such as a scalar additive Gaussian channel with both peak and average power constraints. In addition to the scalar case, the method is shown to work for a vector Gaussian channel with an amplitude constraint A. In particular, for an optimal input $\boldsymbol{X}^\star$, it has been shown that its magnitude $\|\boldsymbol{X}^\star\|$ is a discrete random variable with at most $\mathrm{O}(\mathsf{A}^2)$ mass points for any fixed dimension n.

An interesting direction for further work in this area would be to sharpen the explicit bounds on the number of mass points. Indeed, it has been conjectured with sufficient supporting arguments that the correct order on the number of points should be $\mathrm{O}(\mathsf{A})$ rather than $\mathrm{O}(\mathsf{A}^2)$. Although, finding a better explicit upper bound will ultimately still be related to finding the maximum number of oscillations of a Gaussian convolution within a bounded region.

As has been argued by Smith in [2, p. 40], showing discreteness of the input distribution without providing bounds on the number of mass points does not reduce the maximization of the mutual information from an infinite dimensional optimization (i.e., over the space of all distributions) to the finite dimensional optimization (i.e., over $\mathbb{R}^n$ for some fixed n). This issue has also been pointed out in [10, p. 2346]. The results of this work, in fact, achieves this objective and reduce the infinite dimensional optimization over probability spaces to that in[13] $\mathbb{R}^{2n}$ where $n = \mathrm{O}(\mathsf{A}^2)$, and where $\mathbf{v} \in \mathbb{R}^{2n}$ consists $\mathbf{v} = [p_1, \dots, p_n, x_1, \dots, x_n]$ where p_i is the probability mass of the location x_i. This dimensionality reduction potentially enables applications of efficient optimization algorithms with convergence guarantees such as the gradient descent and is the topic of our current investigation.

It is highly likely that the presented approach generalizes to other (possibly non-additive) channels where channel transition probability is given by a strictly totally positive kernel (e.g., Poisson channel); the interested reader is referred to [34] for a preliminary work on utilizing the techniques of this paper to non-additive settings. The optimization technique used in this paper can also be adapted to other functionals over probability distributions such as the Bayesian minimum mean squared error; the interested reader is referred to [35] for this extension.

Finally, it would interesting to see if the results of this paper can be extended to multiuser channels such as a multiple access channel with an amplitude constraint on the inputs where it is known that the discrete inputs are sum-capacity optimal [36], yet there are no bounds on the number of mass points of the optimal inputs.

[13]Considering the symmetries and properties of a distribution, the dimension of the search space can be reduced to $\mathbb{R}^n$.

Appendix A
Proof of Theorem 2

The starting point is the following sufficient and necessary conditions that can be found in[14] [3], [5].

Lemma 7. *Consider the amplitude constrained vector additive Gaussian channel $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}$ where the input $\boldsymbol{X}$, satisfying $\|\boldsymbol{X}\| \le \mathsf{A}$, is independent from the white Gaussian noise $\boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. If $\boldsymbol{X}^\star$ is an optimal input, the distribution of its magnitude, namely $P_{R^\star} = P_{\|\boldsymbol{X}^\star\|}$, satisfies*

$$i_n(r; P_{R^\star}) = C_n(\mathsf{A}) + \nu_n, \quad r \in \mathsf{supp}(P_{R^\star}), \tag{105}$$
$$i_n(r; P_{R^\star}) \le C_n(\mathsf{A}) + \nu_n, \quad r \in [0, \mathsf{A}], \tag{106}$$

where $C_n(\mathsf{A})$ denotes the capacity of the channel, and

$$i_n(r; P_{R^\star}) = \int_0^\infty f_{\chi_n^2}(x|r) \log \frac{1}{g_n(x; P_{R^\star})} \mathrm{d}x, \tag{107}$$
$$f_{\chi_n^2}(x|r) = \frac{1}{2} \exp\left(-\frac{x + r^2}{2}\right) \left(\frac{\sqrt{x}}{r}\right)^{\frac{n}{2}-1} \mathrm{I}_{\frac{n}{2}-1}\big(r\sqrt{x}\big), \tag{108}$$
$$g_n(x; P_{R^\star}) = \int_0^A \frac{2 f_{\chi_n^2}(x|r)}{x^{\frac{n}{2}-1}} \mathrm{d}P_{R^\star}(r), \tag{109}$$
$$\nu_n = \frac{n}{2} + \log\left(2^{\frac{n}{2}-1}\,\Gamma\left(\frac{n}{2}\right)\right), \tag{110}$$

with $\mathrm{I}_n(x)$ denoting the modified Bessel function of the first kind of order n.

In a similar spirit to the proof of the scalar case, define

$$\kappa_n = \exp(-C_n(\mathsf{A}) - \nu_n), \tag{111}$$
$$\Phi_n(s; P_{R^\star}) = i_n(s; P_{R^\star}) + \log \kappa_n, \tag{112}$$
$$\phi_n(x; P_{R^\star}) = \log \frac{\kappa_n}{g_n(x; P_{R^\star})}, \tag{113}$$

and observe that

$$\Phi_n(r; P_{R^\star}) = \int_0^\infty \phi_n(x; P_{R^\star}) f_{\chi_n^2}(x|r) \mathrm{d}x, \tag{114}$$

where $f_{\chi_n^2}(x|r)$ is as defined in (108). Note that since $f_{\chi_n^2}(x|r)$ is the density of a non-central chi-squared distribution (with non-centrality parameter r^2, and degrees of freedom n), it is a strictly totally positive kernel [37]. Hence, following the footprints of (39)–(43),

$$\begin{aligned}
|\mathsf{supp}(P_{R^\star})| &\le \mathrm{N}\left([0, \mathsf{A}], \Phi_n(\cdot; P_{R^\star})\right) && (115)\\
&\le 1 + \mathrm{N}\left((0, \infty), \Phi_n(\cdot; P_{R^\star})\right) && (116)\\
&\le 1 + \mathrm{N}\left((0, \infty), \phi_n(\cdot; P_{R^\star})\right) && (117)\\
&= 1 + \mathrm{N}\left((0, \infty), g_n(\cdot; P_{R^\star}) - \kappa_n\right) && (118)\\
&\le 1 + \mathrm{N}\left([0, \mathsf{B}_{\kappa_n}], g_n(\cdot; P_{R^\star}) - \kappa_n\right) && (119)\\
&\le \mathfrak{a}_{n_2}\mathsf{A}^2 + \mathfrak{a}_{n_1}\mathsf{A} + \mathfrak{a}_{n_0}, && (120)
\end{aligned}$$

where (115) is a consequence of Lemma 7; the extra $+1$ in (116) is just to account for the possibility that $\Phi_n(0; P_{R^\star}) = 0$; (117) follows from Karlin's Oscillation Theorem, see Theorem 4; (118) follows since $\phi_n(\cdot; P_{R^\star})$

[14]The most general result is shown in [5]. However, a pleasing formulation such as the one in Lemma 7 is hidden behind the heavy notation of [5]. We apply change of variables to provide much simpler presentation.

has the same zeros as $g_n(\cdot;P_{R^\star})-\kappa_n$; (119) follows from Lemma 9 in Appendix B; and (120) is shown in Lemma 14 that can be found in Appendix B.

Appendix B
Additional Lemmas for the Upper Bound Proof of Theorem 2

This section contains several supplementary lemmas that are used in the upper bound proof of Theorem 2.

Lemma 8. *For $n\in\mathbb{N}$ and $z\in\mathbb{C}$*

$$|\mathrm{I}_n(z)|\le \frac{\sqrt{\pi}|z|^n}{2^n\,\Gamma(n+\frac{1}{2})}\mathrm{e}^{|\Re e(z)|}. \tag{121}$$

Proof: Thanks to the integral representation of the modified Bessel function of the first kind, see [38, 9.6.18],

$$\mathrm{I}_n(z)=\frac{(\frac{1}{2}z)^n}{\sqrt{\pi}\,\Gamma(n+\frac{1}{2})}\int_0^{\pi}\mathrm{e}^{z\cos(\theta)}\sin^{2n}(\theta)\mathrm{d}\theta, \tag{122}$$

it follows from the modulus inequality that

$$|\mathrm{I}_n(z)|\le \frac{(\frac{1}{2}|z|)^n}{\sqrt{\pi}\,\Gamma(n+\frac{1}{2})}\int_0^{\pi}|\mathrm{e}^{z\cos(\theta)}||\sin^{2n}(\theta)|\mathrm{d}\theta \tag{123}$$

$$\le \frac{\sqrt{\pi}|z|^n}{2^n\,\Gamma(n+\frac{1}{2})}\mathrm{e}^{|\Re e(z)|}, \tag{124}$$

where (124) follows because $|\mathrm{e}^{z\cos(\theta)}||\sin^{2n}(\theta)|\le \mathrm{e}^{|\Re e(z)|}$. ∎

Similar to its counterpart in Lemma 2, the next lemma provides a bound on the interval for zeros of the function $g_n(\cdot;P_{R^\star})-\kappa_n$.

Lemma 9. *On the Location and Finiteness of Zeros of $g_n(\cdot;P_R)-\kappa_n$. Given an arbitrary distribution P_R, for a fixed $\kappa_n\in(0,1]$ there exists some $\mathsf{B}_{\kappa_n}<\infty$ such that*

$$\mathrm{N}\left([0,\infty),g_n(\cdot;P_R)-\kappa_n\right)=\mathrm{N}\left([0,\mathsf{B}_{\kappa_n}],g_n(\cdot;P_R)-\kappa_n\right) \tag{125}$$

$$<\infty. \tag{126}$$

In particular, there are finitely many zeros of $g_n(\cdot;P_R)-\kappa_n$ all of which are contained within the interval $[0,\mathsf{B}_{\kappa_n}]$. Moreover, B_{κ_n} can be upper bounded as follows:

$$\mathsf{B}_{\kappa_n}\le\left(\mathsf{A}+\sqrt{\mathsf{A}^2+2\log\left(\frac{\gamma_n}{\kappa_n}\right)}\right)^2, \tag{127}$$

where

$$\gamma_n=\frac{\sqrt{\pi}}{2^{\frac{n}{2}-1}\,\Gamma(\frac{n-1}{2})}. \tag{128}$$

Proof: From the definition of the pdf $g_n(\cdot;P_R)$ in (109),

$$g_n(x;P_R)$$
$$=\int_0^{\mathsf{A}}\exp\left(-\frac{x+r^2}{2}\right)\frac{\mathrm{I}_{\frac{n}{2}-1}\big(r\sqrt{x}\big)}{(r\sqrt{x})^{\frac{n}{2}-1}}\mathrm{d}P_R(r) \tag{129}$$

$$\le\int_0^{\mathsf{A}}\frac{\sqrt{\pi}}{2^{\frac{n}{2}-1}\Gamma(\frac{n-1}{2})}\exp\left(-\frac{1}{2}\left(\sqrt{x}-r\right)^2\right)\mathrm{d}P_R(r) \tag{130}$$

$$\le\frac{\sqrt{\pi}}{2^{\frac{n}{2}-1}\,\Gamma(\frac{n-1}{2})}\exp\left(-\frac{x}{2}+\mathsf{A}\sqrt{x}\right), \tag{131}$$

where (130) follows from Lemma 8; and (131) utilizes $r\in[0,\mathsf{A}]$. Since the right side of (131) is decreasing for all $x>\mathsf{A}^2$, it follows that

$$g_n(x;P_R)-\kappa_n<0 \tag{132}$$

for all

$$x>\left(\mathsf{A}+\sqrt{\mathsf{A}^2+2\log\left(\frac{\gamma_n}{\kappa_n}\right)}\right)^2. \tag{133}$$

This means that there exists a B_{κ_n} satisfying (127) such that all zeros of $g_n(\cdot;P_R)-\kappa_n$ are contained within the interval $[0,\mathsf{B}_{\kappa_n}]$.

To see that there are finitely many zeros of $g_n(\cdot;P_R)-\kappa_n$, using the fact that $g_n(\cdot;P_R)$ is analytic[15] suffices, because non-zero analytic functions can only have finitely many zeros on a compact interval. ∎

Following the footsteps of the upper bound proof in the scalar case, the evaluation of the derivative of the function $g_n(\cdot;P_R)$ is established next.

Lemma 10. *For $x\in(0,\infty)$*

$$\frac{\mathrm{d}}{\mathrm{d}x}g_n(x;P_R)$$
$$=\mathbb{E}\left[\frac{\exp\left(-\frac{x+R^2}{2}\right)}{2(R\sqrt{x})^{\frac{n}{2}-1}}\left(\frac{R}{\sqrt{x}}\mathrm{I}_{\frac{n}{2}}(R\sqrt{x})-\mathrm{I}_{\frac{n}{2}-1}(R\sqrt{x})\right)\right], \tag{134}$$

where $R\sim P_R$.

Proof: First of all, given $r>0$, observe that for $u\in(0,\infty)$

$$t_n(u|r)=2\left(\frac{r}{u}\right)^{n-2}\exp\left(\frac{r^2}{2}\right)f_{\chi_n^2}\left(\frac{u^2}{r^2}\middle|r\right) \tag{135}$$

$$=\mathrm{e}^{-\frac{u^2}{2r^2}}\frac{\mathrm{I}_{\frac{n}{2}-1}(u)}{u^{\frac{n}{2}-1}} \tag{136}$$

is a differentiable function and

$$\frac{\mathrm{d}}{\mathrm{d}u}t_n(u|r)=\frac{\mathrm{e}^{-\frac{u^2}{2r^2}}}{u^{\frac{n}{2}-1}}\left(\mathrm{I}_{\frac{n}{2}}(u)-\frac{u}{r^2}\mathrm{I}_{\frac{n}{2}-1}(u)\right), \tag{137}$$

where we have employed the fact that [38, Eqn. (9.6.26)]

$$\frac{\mathrm{d}}{\mathrm{d}u}\mathrm{I}_{\frac{n}{2}-1}(u)=\mathrm{I}_{\frac{n}{2}}(u)+\frac{n-2}{2u}\mathrm{I}_{\frac{n}{2}-1}(u). \tag{138}$$

The desired result then follows from the chain rule as

$$g_n(x;P_R)=\int_0^{\mathsf{A}}\mathrm{e}^{-r^2/2}t_n(r\sqrt{x}|r)\mathrm{d}P_R(r). \tag{139}$$

∎

As was the case in the scalar Gaussian channel, we shall analyze the complex extension of the derivative of $g_n(x;P_R)$. For this reason, in what follows, we denote the complex extension of the derivative of $g_n(x;P_R)$ by $\breve{g}_n'(x;P_R)$.

[15]For a proof, refer to [3, Appendix I] and [5, Propositions 1 and 2] for the respective cases of $n=2$, and $n\ge 2$.

Lemma 11. *Given $r > 0$ and $\mathsf{D} > 0$*

$$\begin{aligned}&\mathrm{I}_{\frac{n}{2}-1}(\mathsf{D}r) - \frac{r}{\mathsf{D}}\mathrm{I}_{\frac{n}{2}}(\mathsf{D}r)\\ &\geq (\mathsf{D}r)^{\frac{n}{2}-1}\frac{2^{1-\frac{n}{2}}}{\Gamma(\frac{n}{2})}\left(1-\frac{2r^2}{n-1+\sqrt{(n-1)^2+(2\mathsf{D}r)^2}}\right) \quad (140)\\ &> 0. \quad (141)\end{aligned}$$

Proof: Using the fact that $I_n(x) > 0$ for $x > 0$

$$\begin{aligned}&\mathrm{I}_{\frac{n}{2}-1}(\mathsf{D}r)\left(1-\frac{r}{\mathsf{D}}\frac{\mathrm{I}_{\frac{n}{2}}(\mathsf{D}r)}{\mathrm{I}_{\frac{n}{2}-1}(\mathsf{D}r)}\right)\\ &\geq \mathrm{I}_{\frac{n}{2}-1}(\mathsf{D}r)\left(1-\frac{2r^2}{n-1+\sqrt{(n-1)^2+(2\mathsf{D}r)^2}}\right) \quad (142)\\ &\geq (\mathsf{D}r)^{\frac{n}{2}-1}\frac{2^{1-\frac{n}{2}}}{\Gamma(\frac{n}{2})}\left(1-\frac{2r^2}{n-1+\sqrt{(n-1)^2+(2\mathsf{D}r)^2}}\right), \quad (143)\end{aligned}$$

where (142) follows from (see [39, Theorem 1])

$$\frac{\mathrm{I}_{\frac{n}{2}}(x)}{\mathrm{I}_{\frac{n}{2}-1}(x)} \leq \frac{2x}{n-1+\sqrt{(n-1)^2+(2x)^2}}; \quad (144)$$

and (143) follows from the fact that $x^{-n}\mathsf{I}_n(x)$ is monotonically increasing for $x > 0$ and that[16]

$$\lim_{x\to 0} x^{-n}\mathsf{I}_n(x) = 2^{-n}\,\Gamma^{-1}(n+1). \quad (145)$$

■

To be plugged into the Tijdeman's Number of Zeros Lemma, Lemmas 12 and 13 find useful suboptimal lower and upper bounds for the maximum value of $\breve{g}_n'(\cdot; P_R)$ on a disc centered at $z_0 \in \mathbb{C}$ where

$$z_0 = \frac{\mathsf{B}_{\kappa_n}}{2} + i\;0. \quad (146)$$

Lemma 12. *Suppose $\mathsf{D} > 0$. For $\mathsf{B}_{\kappa_n} \leq 2\mathsf{D}^2$*

$$\begin{aligned}&\max_{|z|\leq \mathsf{D}^2}\left|\breve{g}_n'\left(z+\frac{\mathsf{B}_{\kappa_n}}{2}; P_R\right)\right| 2^{\frac{n}{2}}\,\Gamma\left(\frac{n}{2}\right)\exp\left(\frac{\mathsf{D}^2+\mathsf{A}^2}{2}\right)\\ &\geq \left(1-\frac{2\mathsf{A}^2}{n-1+\sqrt{(n-1)^2+(2\mathsf{D}\mathsf{A})^2}}\right). \quad (147)\end{aligned}$$

Proof: Observe that for $R \sim P_R$

$$\begin{aligned}&\max_{|z|\leq \mathsf{D}^2}\left|\breve{g}_n'\left(z+\frac{B_{\kappa_n}}{2}; P_R\right)\right| \geq \left|\breve{g}_n'\left(\mathsf{D}^2; P_R\right)\right| \quad (148)\\ &= \left|\mathbb{E}\left[\frac{\exp\left(-\frac{\mathsf{D}^2+R^2}{2}\right)}{2(\mathsf{D}R)^{\frac{n}{2}-1}}\left(\frac{R}{\mathsf{D}}\mathrm{I}_{\frac{n}{2}}(\mathsf{D}R)-\mathrm{I}_{\frac{n}{2}-1}(\mathsf{D}R)\right)\right]\right| \quad (149)\\ &\geq \mathbb{E}\left[\exp\left(-\frac{\mathsf{D}^2+R^2}{2}\right)\frac{2^{-\frac{n}{2}}}{\Gamma(\frac{n}{2})}\right.\\ &\qquad\left.\cdot\left(1-\frac{2R^2}{n-1+\sqrt{(n-1)^2+(2\mathsf{D}R)^2}}\right)\right] \quad (150)\\ &\geq \frac{2^{-\frac{n}{2}}}{\Gamma(\frac{n}{2})}\exp\left(-\frac{\mathsf{D}^2+\mathsf{A}^2}{2}\right)\\ &\qquad\cdot\left(1-\frac{2\mathsf{A}^2}{n-1+\sqrt{(n-1)^2+(2\mathsf{D}\mathsf{A})^2}}\right), \quad (151)\end{aligned}$$

[16]See [38, Eqn. (9.6.28)], and [38, Eqn. (9.6.7)], respectively.

where (148) follows by choosing a suboptimal value of $z = \mathsf{D}^2 - \frac{\mathsf{B}_{\kappa_n}}{2}$; (150) follows from Lemma 11; and (151) follows because $R \leq \mathsf{A}$. ■

Lemma 13. *Suppose $\mathsf{M} > 0$. For $\mathsf{B}_{\kappa_n} \leq 2\mathsf{M}^2$*

$$\begin{aligned}&\max_{|z|\leq \mathsf{M}^2}\left|\breve{g}_n'\left(z+\frac{\mathsf{B}_{\kappa_n}}{2}; P_R\right)\right|\\ &\leq \frac{\gamma_n}{2}\left(\frac{\mathsf{A}^2}{n-1}+1\right)\exp\left(\frac{1}{2}\left(\mathsf{A}+\sqrt{2}\mathsf{M}\right)^2\right), \quad (152)\end{aligned}$$

where γ_n is as defined in (128).

Proof: Capitalizing on the result of Lemma 10, the complex extension of the derivative of $g_n(x; P_R)$ satisfies

$$\begin{aligned}&\left|\breve{g}_n'(z; P_R)\right|\\ &= \left|\mathbb{E}\left[\frac{\exp\left(-\frac{z+R^2}{2}\right)}{2(R\sqrt{z})^{\frac{n}{2}-1}}\left(\frac{R}{\sqrt{z}}\mathrm{I}_{\frac{n}{2}}(R\sqrt{z})-\mathrm{I}_{\frac{n}{2}-1}(R\sqrt{z})\right)\right]\right| \quad (153)\\ &\leq \mathbb{E}\left[\left|\frac{\exp\left(-\frac{z+R^2}{2}\right)}{2(R\sqrt{z})^{\frac{n}{2}-1}}\right|\left(\left|\frac{R}{\sqrt{z}}\mathrm{I}_{\frac{n}{2}}(R\sqrt{z})\right|+\left|\mathrm{I}_{\frac{n}{2}-1}(R\sqrt{z})\right|\right)\right] \quad (154)\\ &\leq \mathbb{E}\left[\frac{\sqrt{\pi}}{2^{\frac{n}{2}}\,\Gamma(\frac{n-1}{2})}\left(\frac{R^2}{n-1}+1\right)\mathrm{e}^{\Re e\left(-\frac{(R\pm\sqrt{z})^2}{2}\right)}\right], \quad (155)\end{aligned}$$

where (154) follows from subsequent applications of modulus and triangular inequalities; (155) is a consequence of Lemma 8. To finalize the proof, using the fact that $R \in [0, \mathsf{A}]$, we simply observe that

$$\begin{aligned}&\max_{|z|\leq \mathsf{M}^2}\left|\breve{g}_n'\left(z+\frac{\mathsf{B}_{\kappa_n}}{2}; P_R\right)\right|\\ &\leq \max_{|z|\leq \mathsf{M}^2}\mathbb{E}\left[\frac{\sqrt{\pi}\left(\frac{R^2}{n-1}+1\right)}{2^{\frac{n}{2}}\,\Gamma(\frac{n-1}{2})}\mathrm{e}^{\Re e\left(-\frac{1}{2}\left(R\pm\sqrt{z+\frac{\mathsf{B}_{\kappa_n}}{2}}\right)^2\right)}\right] \quad (156)\\ &\leq \max_{|z|\leq \mathsf{M}^2}\mathbb{E}\left[\frac{\sqrt{\pi}\left(\frac{R^2}{n-1}+1\right)}{2^{\frac{n}{2}}\,\Gamma(\frac{n-1}{2})}\mathrm{e}^{\frac{1}{2}\left(|R|+\left(|z|+\frac{\mathsf{B}_{\kappa_n}}{2}\right)^{\frac{1}{2}}\right)^2}\right] \quad (157)\\ &\leq \frac{\sqrt{\pi}}{2^{\frac{n}{2}}\,\Gamma(\frac{n-1}{2})}\left(\frac{\mathsf{A}^2}{n-1}+1\right)\exp\left(\frac{1}{2}\left(\mathsf{A}+\sqrt{2}\mathsf{M}\right)^2\right), \quad (158)\end{aligned}$$

where (157) follows after realizing $\Re e(z) \leq |z|$, and applying the triangle inequality twice. ■

Assembling the results of Lemmas 9, 12, and 13, togerher with Tijdeman's Number of Zeros Lemma, i.e., Lemma 4, the following result establishes a suboptimal upper bound on the number of zeros of the function $g_n(\cdot; P_R) - \kappa_n$.

Lemma 14. *Suppose that* $\mathsf{supp}(P_R) \in [0, \mathsf{A}]$ *and* B_{κ_n} *is as defined in Lemma 9. The number of zeros of* $g_n(\cdot; P_R) - \kappa_n$ *within* $[0, \mathsf{B}_{\kappa_n}]$ *satisfies*

$$\mathrm{N}\left([0, \mathsf{B}_{\kappa_n}], g_n(\cdot; P_R) - \kappa_n\right) \leq \mathfrak{a}_{n_2}\mathsf{A}^2 + \mathfrak{a}_{n_1}\mathsf{A} + \mathfrak{a}_{n_0} - 1, \tag{159}$$

where $\mathfrak{a}_{n_2}$, $\mathfrak{a}_{n_1}$, *and* $\mathfrak{a}_{n_0}$ *are as defined in* (15), (16), *and* (17), *respectively*

Proof: In light of Lemma 9, let

$$\overline{\mathsf{B}}_{\kappa_n} = \left(\mathsf{A} + \sqrt{\mathsf{A}^2 + 2\log\left(\frac{\gamma_n}{\kappa_n}\right)}\right)^2, \tag{160}$$

and note that

$$\begin{aligned}
&\mathrm{N}\left([0, \mathsf{B}_{\kappa_n}], g_n(\cdot; P_R) - \kappa_n\right) \\
&\leq 1 + \mathrm{N}\left([0, \mathsf{B}_{\kappa_n}], g_n'(\cdot; P_R)\right) && (161)\\
&= 1 + \mathrm{N}\left(\left[-\frac{\mathsf{B}_{\kappa_n}}{2}, \frac{\mathsf{B}_{\kappa_n}}{2}\right], g_n'\left(\cdot + \frac{\mathsf{B}_{\kappa_n}}{2}; P_R\right)\right) && (162)\\
&\leq 1 + \mathrm{N}\left(\mathcal{D}_{\frac{\mathsf{B}_{\kappa_n}}{2}}, \breve{g}_n'\left(\cdot + \frac{\mathsf{B}_{\kappa_n}}{2}; P_R\right)\right) && (163)\\
&\leq 1 + \min_{s>1,\, t>0}\left\{\frac{\log\frac{\max_{|2z|\leq(st+s+t)\mathsf{B}_{\kappa_n}}\left|\breve{g}_n'\left(z+\frac{\mathsf{B}_{\kappa_n}}{2}; P_R\right)\right|}{\max_{|2z|\leq t\mathsf{B}_{\kappa_n}}\left|\breve{g}_n'\left(z+\frac{\mathsf{B}_{\kappa_n}}{2}; P_R\right)\right|}}{\log s}\right\} && (164)\\
&\leq 1 + \max_{|2z|\leq(2\mathrm{e}+1)\overline{\mathsf{B}}_{\kappa_n}} \log\left|\breve{g}_n'\left(z + \frac{\mathsf{B}_{\kappa_n}}{2}; P_R\right)\right| \\
&\quad - \max_{|2z|\leq\overline{\mathsf{B}}_{\kappa_n}} \log\left|\breve{g}_n'\left(z + \frac{\mathsf{B}_{\kappa_n}}{2}; P_R\right)\right| && (165)\\
&\leq \log\frac{\mathrm{e}\sqrt{\pi}\,\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} + \left(\frac{3}{4} + \mathrm{e}\right)\overline{\mathsf{B}}_{\kappa_n} + \mathsf{A}^2 + \sqrt{2\mathrm{e}+1}\mathsf{A}\overline{\mathsf{B}}_{\kappa_n}^{\frac{1}{2}} \\
&\quad + \log\left(\frac{\mathsf{A}^2}{n-1} + 1\right) \\
&\quad - \log\left(1 - \frac{2\mathsf{A}^2}{n - 1 + \sqrt{(n-1)^2 + 2\overline{\mathsf{B}}_{\kappa_n}\mathsf{A}^2}}\right) && (166)\\
&\leq \log\frac{\mathrm{e}\sqrt{\pi}\,\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} + \left(\frac{3}{4} + \mathrm{e}\right)\overline{\mathsf{B}}_{\kappa_n} + \mathsf{A}^2 + \sqrt{2\mathrm{e}+1}\mathsf{A}\overline{\mathsf{B}}_{\kappa_n}^{\frac{1}{2}} \\
&\quad + \sqrt{\frac{32}{n-1}}\mathsf{A} && (167)\\
&\leq \mathfrak{a}_{n_2}\mathsf{A}^2 + \mathfrak{a}_{n_1}\mathsf{A} + \mathfrak{a}_{n_0} - 1, && (168)
\end{aligned}$$

where (161) follows from Rolle's Theorem; in (163) $\mathcal{D}_r \subset \mathbb{C}$ denotes a disk of radius r centered at the origin and the bound follows because zeros of g_n' are also zeros of its complex extension $\breve{g}_n'$; (164) follows from Tijdeman's Number of Zeros Lemma, see Lemma 4; (165) follows from the suboptimal choices:

$$s = \mathrm{e}, \tag{169}$$

$$t = \frac{\overline{\mathsf{B}}_{\kappa_n}}{\mathsf{B}_{\kappa_n}}; \tag{170}$$

(166) follows from Lemmas 12 and 13 with

$$\mathsf{D}^2 \leftarrow \frac{1}{2}\overline{\mathsf{B}}_{\kappa_n}, \tag{171}$$

$$\mathsf{M}^2 \leftarrow \frac{2\mathrm{e}+1}{2}\overline{\mathsf{B}}_{\kappa_n}; \tag{172}$$

(167) follows from a tedious algebra where we first note, from their definitions in (111) and (128), that the ratio $\gamma_n/\kappa_n > 1$, implying $\overline{\mathsf{B}}_{\kappa_n} > 2\mathsf{A}^2$, implying

$$1 - \frac{2\mathsf{A}^2}{n - 1 + \sqrt{(n-1)^2 + 2\overline{\mathsf{B}}_{\kappa_n}\mathsf{A}^2}} \geq \left(\frac{2\mathsf{A}^2}{n-1} + 1\right)^{-1}, \tag{173}$$

and allowing us to upper bound the last two "log" terms in the right side of (166) by

$$2\log\left(\frac{2\mathsf{A}^2}{n-1} + 1\right) \leq \left(\frac{32}{n-1}\right)^{\frac{1}{2}}\mathsf{A}; \tag{174}$$

finally (168) follows from the definitions of κ_n, γ_n (in (111), and (128), respectively) and the facts that

$$\mathsf{A}\sqrt{\mathsf{A}^2 + 2\log\left(\frac{\gamma_n}{\kappa_n}\right)} \leq \mathsf{A}^2 + \log\left(\frac{\gamma_n}{\kappa_n}\right), \tag{175}$$

$$C_n(\mathsf{A}) \leq \frac{n}{2}\log(1 + \mathsf{A}^2) \leq n\mathsf{A}. \tag{176}$$

■

APPENDIX C
PROOF OF THEOREM 3

A. Proof of the Upper Bound in Theorem 3

The first ingredient of the upper bound proof is once again due to Smith [1, Corollary 2] who characterizes the optimal input distribution as follows.

Lemma 15. *Consider the amplitude and power constrained scalar additive Gaussian channel* $Y = X + Z$ *where the input* X, *satisfying* $|X| \leq \mathsf{A}$ *and* $\mathbb{E}[|X|^2] \leq \mathsf{P}$, *is independent from the noise* $Z \sim \mathcal{N}(0, 1)$. *Then,* $P_{X^\star}$ *is the capacity-achieving input distribution if and only if the following conditions are satisfied*:

$$i(x; P_{X^\star}) = C(\mathsf{A}, \mathsf{P}) + \lambda(x^2 - \mathsf{P}), \quad x \in \mathsf{supp}(P_{X^\star}), \tag{177}$$

$$i(x; P_{X^\star}) \leq C(\mathsf{A}, \mathsf{P}) + \lambda(x^2 - \mathsf{P}), \quad x \in [-\mathsf{A}, \mathsf{A}], \tag{178}$$

$$0 = \lambda(\mathsf{P} - \mathbb{E}[X^2]), \tag{179}$$

where $C(\mathsf{A}, \mathsf{P})$ *denotes the capacity of the channel, and* $i(x; P_{X^\star})$ *is as defined in* (27).

Remark 11. Hidden in our notation for typographic reasons, the Lagrange multiplier λ in fact depends on amplitude and power constraints, namely A and P. Indeed, since $|X| \leq \mathsf{A}$, if $\mathsf{P} > \mathsf{A}^2$, the power constraint is inactive, implying $\lambda = 0$. In this case, the problem reduces to additive Gaussian channel with only amplitude constraint, and we recover Lemma 1.

As a corollary to above lemma, note that if x is a point of support of $P_{X^\star}$ (i.e., $x \in \mathsf{supp}(P_{X^\star})$), then x is a zero of the function

$$\Xi_{\mathsf{A},\mathsf{P}}(x; P_{X^\star}) = i(x; P_{X^\star}) - C(\mathsf{A}, \mathsf{P}) - \lambda(x^2 - \mathsf{P}). \tag{180}$$

In other words,

$$|\mathsf{supp}(P_{X^\star})| \le \mathrm{N}([-\mathsf{A},\mathsf{A}], \Xi_{\mathsf{A},\,\mathsf{P}}(\cdot; P_{X^\star})) \tag{181}$$
$$\le \mathrm{N}(\mathbb{R}, \Xi_{\mathsf{A},\,\mathsf{P}}(\cdot; P_{X^\star})). \tag{182}$$

Observe that, since

$$x^2 = \int_{\mathbb{R}} \frac{y^2-1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(y-x)^2}{2}} \mathrm{d}y, \tag{183}$$

we can write

$$\Xi_{\mathsf{A},\,\mathsf{P}}(x; P_{X^\star}) = \int_{\mathbb{R}} \frac{\xi_{\mathsf{A},\,\mathsf{P}}(y)}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(y-x)^2}{2}} \mathrm{d}y, \tag{184}$$

where

$$\xi_{\mathsf{A},\,\mathsf{P}}(y) = \log \frac{1}{f_{Y^\star}(y)} - h(Z) - C(\mathsf{A},\mathsf{P}) + \lambda\mathsf{P} - \lambda(y^2-1). \tag{185}$$

Keeping the steps (39)–(43) in mind, define

$$\mathfrak{F}^\star_{\mathsf{A},\,\mathsf{P}}(y) = \mathrm{e}^{\lambda y^2} f_{Y^\star}(y) - \kappa_{\mathsf{A},\,\mathsf{P}}, \tag{186}$$

with[17]

$$\kappa_{\mathsf{A},\,\mathsf{P}} = \exp(-h(Z) - C(\mathsf{A},\mathsf{P}) + \lambda(\mathsf{P}+1)). \tag{187}$$

Using the fact that the Gaussian distribution is a strictly totally positive kernel, and resuming from (182)

$$|\mathsf{supp}(P_{X^\star})| \le \mathrm{N}\left(\mathbb{R}, \Xi_{\mathsf{A},\,\mathsf{P}}(\cdot; P_{X^\star})\right) \tag{188}$$
$$\le \mathrm{N}\left(\mathbb{R}, \xi_{\mathsf{A},\,\mathsf{P}}\right) \tag{189}$$
$$= \mathrm{N}\left(\mathbb{R}, \mathfrak{F}^\star_{\mathsf{A},\,\mathsf{P}}\right) \tag{190}$$
$$= \mathrm{N}\left([-\mathsf{B}_{\kappa_{\mathsf{A},\,\mathsf{P}}}, \mathsf{B}_{\kappa_{\mathsf{A},\,\mathsf{P}}}], \mathfrak{F}^\star_{\mathsf{A},\,\mathsf{P}}\right) \tag{191}$$
$$\le a_{\mathsf{P}_2}\mathsf{A}_{\mathsf{P}}^2 + a_{\mathsf{P}_1}\mathsf{A}_{\mathsf{P}} + a_{\mathsf{P}_0}, \tag{192}$$

where (189) follows from Karlin's Oscillation Theorem, see Theorem 4; (190) follows because $\xi_{\mathsf{A},\,\mathsf{P}}(y) = 0$ if and only if $\mathfrak{F}^\star_{\mathsf{A},\,\mathsf{P}}(y) = 0$; (191) is a consequence Lemma 17 in Appendix D; finally (192) follows from Lemma 21 in Appendix D. ■

B. Proof of the Lower Bound in Theorem 3

Invoking entropy-power inequality,

$$I(X;Y) = h(X+Z) - h(Z) \tag{193}$$
$$\ge \frac{1}{2}\log\left(\mathrm{e}^{2h(X)} + \mathrm{e}^{2h(Z)}\right) - h(Z) \tag{194}$$
$$= \frac{1}{2}\log\left(\frac{1}{2\pi\mathrm{e}}\mathrm{e}^{2h(X)} + 1\right). \tag{195}$$

Therefore,

$$\log|\mathsf{supp}(P_{X^\star})| \ge H(P_{X^\star}) \tag{196}$$
$$\ge \max_{X:|X|\le\mathsf{A},\,\mathbb{E}[X^2]\le\mathsf{P}} I(X;Y) \tag{197}$$
$$\ge \max_{X:|X|\le\mathsf{A},\,\mathbb{E}[X^2]\le\mathsf{P}} \frac{1}{2}\log\left(\frac{\mathrm{e}^{2h(X)}}{2\pi\mathrm{e}} + 1\right) \tag{198}$$
$$\ge \max_{|a|\le\mathsf{A},\,\frac{a^2}{3}\le\mathsf{P}} \frac{1}{2}\log\left(\frac{2a^2}{\pi\mathrm{e}} + 1\right) \tag{199}$$
$$= \frac{1}{2}\log\left(\frac{2\min\left\{\mathsf{A}^2, 3\mathsf{P}\right\}}{\pi\mathrm{e}} + 1\right), \tag{200}$$

where (199) follows by sub-optimally choosing X to be uniform on $[-a,a]$. ■

Appendix D
Additional Lemmas for the Upper Bound Proof of Theorem 3

Crucial to the proofs that follow, the next lemma provides a bound on the value of the Lagrange multiplier λ in Smith's result [1, Corollary 2].

Lemma 16. *Bound on the Value of λ. The Lagrange multiplier λ that appears in Lemma 15 satisfies*

$$\lambda \le \frac{\log(1+\mathsf{P})}{2\mathsf{P}} \cdot 1\left\{\mathsf{P} < \mathsf{A}^2\right\}. \tag{201}$$

Proof: If $\mathsf{P} \ge \mathsf{A}^2$, the power constraint in (4) is not active, implying that the Lagrange multiplier $\lambda = 0$. Suppose $\mathsf{P} < \mathsf{A}^2$. It follows from Lemma 15 that

$$\lambda\mathsf{P} \le C(\mathsf{A},\mathsf{P}) - i(0; P_{X^\star}) \tag{202}$$
$$\le \frac{1}{2}\log(1+\mathsf{P}), \tag{203}$$

where (203) is because $C(\mathsf{A},\mathsf{P}) \le C(\infty,\mathsf{P}) = \frac{1}{2}\log(1+\mathsf{P})$, and $i(0; P_{X^\star}) = D(Z\|Y) \ge 0$. ■

Similar to its counterpart in Lemma 2, the next lemma provides a bound on the interval for the zeros of the function $\mathfrak{F}^\star_{\mathsf{A},\,\mathsf{P}}$.

Lemma 17. *Location and Finiteness of Zeros of $\mathrm{e}^{\lambda y^2} f_Y(y) - \kappa_{\mathsf{A},\,\mathsf{P}}$. For a fixed $\kappa_{\mathsf{A},\,\mathsf{P}} \in \left(0, \frac{1}{\sqrt{2\pi}}\right]$, there exists some $\mathsf{B}_{\kappa_{\mathsf{A},\,\mathsf{P}}} = \mathsf{B}_{\kappa_{\mathsf{A},\,\mathsf{P}}}(\mathsf{A},\mathsf{P}) < \infty$ such that*

$$\mathrm{N}\left(\mathbb{R}, \mathrm{e}^{\lambda y^2} f_Y(y) - \kappa_{\mathsf{A},\,\mathsf{P}}\right)$$
$$= \mathrm{N}\left([-\mathsf{B}_{\kappa_{\mathsf{A},\,\mathsf{P}}}, \mathsf{B}_{\kappa_{\mathsf{A},\,\mathsf{P}}}], \mathrm{e}^{\lambda y^2} f_Y(y) - \kappa_{\mathsf{A},\,\mathsf{P}}\right) \tag{204}$$
$$< \infty. \tag{205}$$

In other words, there are finitely many zeros of $\mathrm{e}^{\lambda y^2} f_Y(y) - \kappa_{\mathsf{A},\,\mathsf{P}}$ which are contained within the interval $[-\mathsf{B}_{\kappa_{\mathsf{A},\,\mathsf{P}}}, \mathsf{B}_{\kappa_{\mathsf{A},\,\mathsf{P}}}]$. Moreover,

$$\mathsf{B}_{\kappa_{\mathsf{A},\,\mathsf{P}}} \le \frac{\mathsf{A}}{1-2\lambda} + \left(\frac{1}{1-2\lambda}\log\frac{1}{2\pi\kappa^2_{\mathsf{A},\,\mathsf{P}}} + \frac{2\lambda\mathsf{A}^2}{(1-2\lambda)^2}\right)^{\frac{1}{2}}. \tag{206}$$

Proof: Using the monotonicity of e^{-u}, for $|y| > \mathsf{A}$,

$$\mathrm{e}^{\lambda y^2} f_Y(y)$$
$$= \frac{\mathrm{e}^{\lambda y^2}}{\sqrt{2\pi}} \mathbb{E}\left[\exp\left(-\frac{(y-X)^2}{2}\right)\right] \tag{207}$$
$$\le \frac{\mathrm{e}^{\lambda y^2}}{\sqrt{2\pi}} \exp\left(-\frac{(y-\mathsf{A})^2}{2}\right) \tag{208}$$

[17]Note that $0 \le i(0; P_{X^\star}) \le C(\mathsf{A},\mathsf{P}) - \lambda\mathsf{P}$, and $\lambda < 1/2$, cf. Lemma 16. This implies that $\kappa_{\mathsf{A},\,\mathsf{P}} < 1/\sqrt{2\pi}$.

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1-2\lambda}{2}\left(y-\frac{\mathsf{A}}{1-2\lambda}\right)^2+\frac{\lambda \mathsf{A}^2}{1-2\lambda}\right). \tag{209}$$

Since $\lambda \in [0, 1/2)$, cf. Lemma 16, the right side of (209) is a decreasing function for all $|y| > \frac{A}{1-2\lambda}$ and we have

$$\mathrm{e}^{\lambda y^2} f_Y(y) - \kappa_{\mathsf{A},\,\mathsf{P}} < 0 \tag{210}$$

for all

$$|y| > \frac{\mathsf{A}}{1-2\lambda} + \left(\frac{2}{1-2\lambda}\log\frac{1}{\kappa_{\mathsf{A},\,\mathsf{P}}\sqrt{2\pi}} + \frac{2\lambda\mathsf{A}^2}{(1-2\lambda)^2}\right)^{\frac{1}{2}}. \tag{211}$$

This means that there exists $\mathsf{B}_{\kappa_{\mathsf{A},\,\mathsf{P}}}$ satisfying (206) such that all zeros of $\mathfrak{F}_{\mathsf{A},\,\mathsf{P}}(y) = \mathrm{e}^{\lambda y^2} f_Y(y) - \kappa_{\mathsf{A},\,\mathsf{P}}$ are contained within the interval $[-\mathsf{B}_{\kappa_{\mathsf{A},\,\mathsf{P}}}, \mathsf{B}_{\kappa_{\mathsf{A},\,\mathsf{P}}}]$.

To see the finiteness of the number of zeros of $\mathfrak{F}_{\mathsf{A},\,\mathsf{P}}(y)$, it suffices to show that $\mathfrak{F}_{\mathsf{A},\,\mathsf{P}}$ is analytic on $\mathbb{R}$ as analytic functions have finitely many zeros on a compact interval. However, it is easy to see that $\mathfrak{F}_{\mathsf{A},\,\mathsf{P}}$ is analytic because convolution with a Gaussian preserves analyticity [31, Proposition 8.10]. ■

Lemma 18. *For $\kappa_{\mathsf{A},\,\mathsf{P}}$ as defined in* (187)*, the bound on the location of the zeros in Lemma 17 can be loosened as*

$$\mathsf{B}_{\kappa_{\mathsf{A},\,\mathsf{P}}} < 2\mathsf{A}_{\mathsf{P}} + 1, \tag{212}$$

where

$$\mathsf{A}_{\mathsf{P}} = \frac{\mathsf{AP}}{\mathsf{P} - \log(1+\mathsf{P})\cdot 1\{\mathsf{P} < \mathsf{A}^2\}}. \tag{213}$$

Proof: We may assume that $\mathsf{P} < \mathsf{A}^2$, otherwise see (96). In that case, observe that

$$C(\mathsf{A},\mathsf{P}) \le \frac{1}{2}\log(1+\mathsf{P}), \tag{214}$$

and hence,

$$\begin{aligned}\kappa_{\mathsf{A},\,\mathsf{P}} &= \exp(-h(Z) - C(\mathsf{A},\mathsf{P}) + \lambda(\mathsf{P}+1)) &(215)\\ &\ge \frac{\exp(\lambda(\mathsf{P}+1))}{\sqrt{2\pi\mathrm{e}(1+\mathsf{P})}}. &(216)\end{aligned}$$

Combining (206) and (216)

$$\begin{aligned}&\mathsf{B}_{\kappa_{\mathsf{A},\,\mathsf{P}}}\\ &\le \frac{\mathsf{A}}{1-2\lambda} + \left(1 + \frac{\log(1+\mathsf{P})}{1-2\lambda} + \frac{2\lambda}{1-2\lambda}\left(\frac{\mathsf{A}^2}{1-2\lambda} - \mathsf{P}\right)\right)^{\frac{1}{2}} &(217)\\ &\le \mathsf{A}_{\mathsf{P}} + (1+2\lambda\mathsf{A}_{\mathsf{P}}^2)^{\frac{1}{2}} &(218)\\ &< 2\mathsf{A}_{\mathsf{P}} + 1, &(219)\end{aligned}$$

where (218) follows from Lemma 16 as the right side of (217) is increasing in $\lambda \le \frac{\log(1+\mathsf{P})}{2\mathsf{P}}$; and (219) follows because $\lambda < \frac{1}{2}$, cf. Lemma 16. ■

Lemma 19. *Suppose $\mathfrak{F}_{\mathsf{A},\,\mathsf{P}}\colon \mathbb{R} \to \mathbb{R}$ is such that $\mathfrak{F}_{\mathsf{A},\,\mathsf{P}}(y) = \mathrm{e}^{\lambda y^2} f_Y(y) - \kappa_{\mathsf{A},\,\mathsf{P}}$. The complex extension of its derivative $\breve{\mathfrak{F}}'_{\mathsf{A},\,\mathsf{P}}\colon \mathbb{C} \to \mathbb{C}$ satisfies*

$$\begin{aligned}&\max_{|z|\le\mathsf{B}} \left|\breve{\mathfrak{F}}'_{\mathsf{A},\,\mathsf{P}}(z)\right|\\ &\le \frac{1}{\sqrt{2\pi}}(\mathsf{A} + (1+2\lambda_{\mathsf{P}})\mathsf{B})\exp\left(\frac{(1+2\lambda_{\mathsf{P}})\mathsf{B}^2}{2}\right) &(220)\\ &< \frac{1}{\sqrt{2\pi}}(\mathsf{A}+2\mathsf{B})\exp\left(\mathsf{B}^2\right), &(221)\end{aligned}$$

where in (220) $\lambda_{\mathsf{P}} = \frac{\log(1+\mathsf{P})}{2\mathsf{P}} \cdot 1\{\mathsf{P} < \mathsf{A}^2\}$.

Proof: Denote by $\breve{f}_Y$ and $\breve{f}'_Y$ the analytic complex extensions of the probability density function f_Y and its derivative f'_Y, respectively. Then,

$$\begin{aligned}&\max_{|z|\le\mathsf{B}} \left|\breve{\mathfrak{F}}'_{\mathsf{A},\,\mathsf{P}}(z)\right|\\ &= \max_{|z|\le\mathsf{B}} \left|\mathrm{e}^{\lambda z^2}\left(\breve{f}'_Y(z) + 2\lambda z \breve{f}_Y(z)\right)\right| &(222)\\ &\le \mathrm{e}^{\lambda\mathsf{B}^2} \max_{|z|\le\mathsf{B}} \left|\breve{f}'_Y(z) + 2\lambda z\breve{f}_Y(z)\right| &(223)\\ &\le \mathrm{e}^{\lambda\mathsf{B}^2}\left(\max_{|z|\le\mathsf{B}}\left|\breve{f}'_Y(z)\right| + \max_{|z|\le\mathsf{B}}\left|2\lambda z\breve{f}_Y(z)\right|\right) &(224)\\ &\le \frac{\mathrm{e}^{\lambda\mathsf{B}^2}}{\sqrt{2\pi}}\left((\mathsf{A}+\mathsf{B})\exp\left(\frac{\mathsf{B}^2}{2}\right)\right. &(225)\\ &\quad \left. + \max_{|z|\le\mathsf{B}}\left|2\lambda z\mathbb{E}\left[\exp\left(-\frac{(z-X)^2}{2}\right)\right]\right|\right) &(226)\\ &\le \frac{\mathrm{e}^{\lambda\mathsf{B}^2}}{\sqrt{2\pi}}\left((\mathsf{A}+\mathsf{B})\exp\left(\frac{\mathsf{B}^2}{2}\right)\right. &(227)\\ &\quad \left. + 2\lambda\mathsf{B}\max_{|z|\le\mathsf{B}}\mathbb{E}\left[\left|\exp\left(-\frac{(z-X)^2}{2}\right)\right|\right]\right) &(228)\\ &\le \frac{\mathrm{e}^{\lambda\mathsf{B}^2}}{\sqrt{2\pi}}\left((\mathsf{A}+\mathsf{B})\exp\left(\frac{\mathsf{B}^2}{2}\right) + 2\lambda\mathsf{B}\max_{|z|\le\mathsf{B}}\exp\left(\frac{\mathfrak{Im}^2(z)}{2}\right)\right) &(229)\\ &\le \frac{1}{\sqrt{2\pi}}(\mathsf{A}+(1+2\lambda)\mathsf{B})\exp\left(\frac{(1+2\lambda)\mathsf{B}^2}{2}\right), &(230)\end{aligned}$$

where (222) follows from definitions of the functions involved; (223) is because $|z| \le \mathsf{B}$ implies $\left|\mathrm{e}^{\lambda z^2}\right| \le \mathrm{e}^{\lambda\mathsf{B}^2}$; (224) follows from the triangle inequality; (226) follows from Lemma 5; (228) follows from the modulus inequality.

The desired result is a consequence of the fact that the Lagrange multiplier satisfies $\lambda \le \lambda_{\mathsf{P}} < 1/2$, cf. Lemma 16. ■

Lemma 20. *Let $\mathsf{B} \ge \mathsf{A}/(1-2\lambda)$. Suppose $\mathfrak{F}_{\mathsf{A},\,\mathsf{P}}\colon \mathbb{R} \to \mathbb{R}$ is such that $\mathfrak{F}_{\mathsf{A},\,\mathsf{P}}(y) = \mathrm{e}^{\lambda y^2} f_Y(y) - \kappa_{\mathsf{A},\,\mathsf{P}}$. The complex extension of its derivative $\breve{\mathfrak{F}}'_{\mathsf{A},\,\mathsf{P}}\colon \mathbb{C} \to \mathbb{C}$ satisfies*

$$\max_{|z|\le\mathsf{B}}\left|\breve{\mathfrak{F}}'_{\mathsf{A},\,\mathsf{P}}(z)\right| \ge \frac{\mathsf{A}}{\sqrt{2\pi}}\exp\left(-\frac{2-\lambda_{\mathsf{P}}}{1-2\lambda_{\mathsf{P}}}\mathsf{A}^2\right), \tag{231}$$

where $\lambda_{\mathsf{P}} = \frac{\log(1+\mathsf{P})}{2\mathsf{P}} \cdot 1\{\mathsf{P} < \mathsf{A}^2\}$.

Proof: Note that

$$\begin{aligned}&\max_{|z|\le\mathsf{B}}\left|\breve{\mathfrak{F}}'_{\mathsf{A},\,\mathsf{P}}(z)\right|\\ &= \max_{|z|\le\mathsf{B}}\left|\mathrm{e}^{\lambda z^2}\left(\breve{f}'_Y(z) + 2\lambda z\breve{f}_Y(z)\right)\right| &(232)\\ &= \max_{|z|\le\mathsf{B}}\frac{1}{\sqrt{2\pi}}\left|\mathbb{E}\left[(X-(1-2\lambda)z)\mathrm{e}^{-\frac{1-2\lambda}{2}\left(z-\frac{X}{1-2\lambda}\right)^2+\frac{\lambda X^2}{1-2\lambda}}\right]\right| &(233)\end{aligned}$$

$$
\begin{aligned}
&\mathrm{N}\left([-\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}},\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}}],\mathfrak{F}_{\mathsf{A},\mathsf{P}}\right)\\
&\le 1+\mathrm{N}\left([-\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}},\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}}],\mathfrak{F}'_{\mathsf{A},\mathsf{P}}\right) && (239)\\
&\le 1+\mathrm{N}\left(\mathcal{D}_{\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}}},\breve{\mathfrak{F}}'_{\mathsf{A},\mathsf{P}}\right) && (240)\\
&\le 1+\min_{s>1,\,t\ge\frac{\mathsf{A}_\mathsf{P}}{\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}}}}\left\{\frac{1}{\log s}\left(\log\max_{|z|\le(st+s+t)\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}}}\left|\breve{\mathfrak{F}}'_{\mathsf{A},\mathsf{P}}\right|-\log\max_{|z|\le t\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}}}\left|\breve{\mathfrak{F}}'_{\mathsf{A},\mathsf{P}}\right|\right)\right\} && (241)\\
&\le 1+\min_{s>1,\,t\ge\frac{\mathsf{A}_\mathsf{P}}{\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}}}}\left\{\frac{1}{\log s}\left(\frac{(st+s+t)^2\mathsf{B}^2_{\kappa_{\mathsf{A},\mathsf{P}}}}{2/(1+2\lambda_\mathsf{P})}+\frac{2-\lambda_\mathsf{P}}{1-2\lambda_\mathsf{P}}\,\mathsf{A}^2+\log\left(1+\frac{(st+s+t)\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}}}{\mathsf{A}/(1+2\lambda_\mathsf{P})}\right)\right)\right\} && (242)\\
&= 1+\min_{s>1}\left\{\frac{1}{\log s}\left(\frac{((\mathsf{A}_\mathsf{P}+\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}})s+\mathsf{A}_\mathsf{P})^2}{2/(1+2\lambda_\mathsf{P})}+\frac{2-\lambda_\mathsf{P}}{1-2\lambda_\mathsf{P}}\,\mathsf{A}^2+\log\left(\frac{2}{1-2\lambda_\mathsf{P}}+\frac{(\mathsf{A}_\mathsf{P}+\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}})s}{\mathsf{A}/(1+2\lambda_\mathsf{P})}\right)\right)\right\} && (243)\\
&\le 1+\min_{s>1}\left\{\frac{1}{\log s}\left(\frac{((3\mathsf{A}_\mathsf{P}+1)s+\mathsf{A}_\mathsf{P})^2}{2/(1+2\lambda_\mathsf{P})}+\frac{2-\lambda_\mathsf{P}}{1-2\lambda_\mathsf{P}}\,\mathsf{A}^2+\log\left(\frac{2}{1-2\lambda_\mathsf{P}}+\frac{(3\mathsf{A}_\mathsf{P}+1)s}{\mathsf{A}/(1+2\lambda_\mathsf{P})}\right)\right)\right\} && (244)\\
&\le 1+2\left(\frac{((3\sqrt{\mathrm{e}}+1)\mathsf{A}_\mathsf{P}+\sqrt{\mathrm{e}})^2}{2/(1+2\lambda_\mathsf{P})}+\frac{2-\lambda_\mathsf{P}}{1-2\lambda_\mathsf{P}}\,\mathsf{A}^2+\log\left(\frac{2}{1-2\lambda_\mathsf{P}}+\frac{(3\mathsf{A}_\mathsf{P}+1)\sqrt{\mathrm{e}}}{\mathsf{A}/(1+2\lambda_\mathsf{P})}\right)\right) && (245)\\
&\le 1+2\left(\frac{((3\sqrt{\mathrm{e}}+1)\mathsf{A}_\mathsf{P}+\sqrt{\mathrm{e}})^2}{2/(1+2\lambda_\mathsf{P})}+(2-\lambda_\mathsf{P})(1-2\lambda_\mathsf{P})\,\mathsf{A}^2_\mathsf{P}+\log\left(\frac{2+4\sqrt{\mathrm{e}}(1+2\lambda_\mathsf{P})}{1-2\lambda_\mathsf{P}}\right)\right), && (246)
\end{aligned}
$$

$$
\begin{aligned}
&\ge\frac{1}{\sqrt{2\pi}}\left|\mathbb{E}\left[(X-\mathsf{A})\exp\left(\frac{-\mathsf{A}^2+2\mathsf{A}X-(1-2\lambda)X^2}{2-4\lambda}\right)\right]\right| && (234)\\
&\ge\frac{1}{\sqrt{2\pi}}\mathbb{E}\left[(\mathsf{A}-X)\exp\left(\frac{-3\mathsf{A}^2}{2-4\lambda}-\frac{1}{2}\mathsf{A}^2\right)\right] && (235)\\
&=\frac{\mathsf{A}}{\sqrt{2\pi}}\exp\left(-\left(\frac{1}{2}+\frac{3}{2-4\lambda}\right)\mathsf{A}^2\right) && (236)\\
&\ge\frac{\mathsf{A}}{\sqrt{2\pi}}\exp\left(-\frac{2-\lambda_\mathsf{P}}{1-2\lambda_\mathsf{P}}\,\mathsf{A}^2\right), && (237)
\end{aligned}
$$

where (234) follows from the suboptimal choice of $z=\frac{\mathsf{A}}{1-2\lambda}\le\mathsf{B}$; (235) follows because $|X|\le\mathsf{A}$; (236) is a consequence of $\mathbb{E}[X]=0$; and finally, (237) follows because $\lambda\le\lambda_\mathsf{P}$, see Lemma 16. ■

Lemma 21. *Bound on the Number of Oscillations of $\mathfrak{F}^\star_{\mathsf{A},\mathsf{P}}$. Suppose that $\mathfrak{F}^\star_{\mathsf{A},\mathsf{P}}$ is as defined in* (186) *and $\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}}$ be as defined in Lemma 17. The number of zeros of $\mathfrak{F}^\star_{\mathsf{A},\mathsf{P}}$ within the interval $[-\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}},\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}}]$ satisfies*

$$\mathrm{N}\left([-\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}},\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}}],\mathfrak{F}_{\mathsf{A},\mathsf{P}}\right)\le a_{\mathsf{P}_2}\mathsf{A}^2_\mathsf{P}+a_{\mathsf{P}_1}\mathsf{A}_\mathsf{P}+a_{\mathsf{P}_0},\quad(238)$$

where A_P, a_{P_2}, a_{P_1}, a_{P_0}, and λ_P are as defined in (20)*,* (21)*,* (22)*,* (23)*, and* (24)*.*

Proof: We may assume $\mathsf{P}<\mathsf{A}^2$, otherwise see the proof of the upper bound in Theorem 1. For an arbitrary output density f_Y define $\mathfrak{F}_{\mathsf{A},\mathsf{P}}\colon\mathbb{R}\to\mathbb{R}$ such that $\mathfrak{F}_{\mathsf{A},\mathsf{P}}(y)=\mathrm{e}^{\lambda y^2}f_Y(y)-\kappa_{\mathsf{A},\mathsf{P}}$ and let $\mathfrak{F}'_{\mathsf{A},\mathsf{P}}\colon\mathbb{R}\to\mathbb{R}$ be the derivative of $\mathfrak{F}_{\mathsf{A},\mathsf{P}}$. Consider the disk $\mathcal{D}_R\subset\mathbb{C}$ of radius R centered at the origin and note the following sequence of inequalities (239)–(246), shown at the top of this page.

There, (239) follows from Rolle's Theorem, see Lemma 3; (240) follows because the zeros of $\mathfrak{F}'_{\mathsf{A},\mathsf{P}}\colon\mathbb{R}\to\mathbb{R}$ are also the zeros of its complex extension $\breve{\mathfrak{F}}'_{\mathsf{A},\mathsf{P}}\colon\mathbb{C}\to\mathbb{C}$; (241) is a consequence of Tijdeman's Number of Zeros Lemma, namely Lemma 4; (242) follows by invoking Lemmas 19 and 20 above; (243) follows because $t=\frac{\mathsf{A}_\mathsf{P}}{\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}}}$ is the minimizer in the right side of (242); (244) follows from the fact that $\mathsf{B}_{\kappa_{\mathsf{A},\mathsf{P}}}<2\mathsf{A}_\mathsf{P}+1$, see Lemma 18; (245) is a consequence of the suboptimal choice $s=\sqrt{\mathrm{e}}$; and finally, (246) follows from the assumption that $\mathsf{A}>1$.

Algebraic manipulations in the right side of (246) yield the desired result in (238). ■

Acknowledgment

Authors of this work acknowledge with gratitude the counter-example that N. Kashyap and M. Krishnapur has sent regarding a conjecture we had in an earlier version of the paper. They have shown that the maximum number of modes that a constrained Gaussian mixture can have is indeed of order $\Theta(A^2)$, showing that Theorem 6 gives an order tight bound.

References

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

[2] J. G. Smith, "On the information capacity of peak and average power constrained Gaussian channels," Ph.D. dissertation, Univ. California, Berkeley, Berkeley, CA, USA, 1969.

[3] 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, Jul. 1995.

[4] 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, Jun. 2005.

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

[6] A. Dytso, M. Goldenbaum, H. V. Poor, and S. Shamai (Shitz), "When are discrete channel inputs optimal?—Optimization techniques and some new results," in *Proc. IEEE 52nd Annu. Conf. Inf. Sci. Syst. (CISS)*, Princeton, NJ, USA, Mar. 2018, pp. 1–6.

[7] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423 and 623–656, 1948.

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

[9] A. Dytso, M. Al, H. V. Poor, and S. Shamai (Shitz), "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. Huang and S. P. Meyn, "Characterization and computation of optimal distributions for channel coding," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2336–2351, Jul. 2005.

[11] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[12] A. Dytso, M. Goldenbaum, S. Shamai (Shitz), and H. V. Poor, "Upper and lower bounds on the capacity of amplitude-constrained MIMO channels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.

[13] R. G. Gallager. (Sep. 1976). *Source Coding With Side Information and Universal Coding*. [Online]. Available: http://web.mit.edu/gallager/www/papers/paper5.pdf

[14] B. Ryabko, "Coding of a source with unknown but ordered probabilities," *Problems Inf. Transmiss.*, vol. 15, no. 2, pp. 134–138, 1979.

[15] A. L. McKellips, "Simple tight bounds on capacity for the peak-limited discrete-time channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Chicago, IL, USA, Jun. 2004, p. 348.

[16] A. Thangaraj, G. Kramer, and G. Böcherer, "Capacity bounds for discrete-time, amplitude-constrained, additive white Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4172–4182, Apr. 2017.

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

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

[19] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[20] L. H. Ozarow and A. D. Wyner, "On the capacity of the Gaussian channel with a finite number of input levels," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1426–1428, Nov. 1990.

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

[22] A. Dytso, M. Goldenbaum, H. V. Poor, and S. Shamai (Shitz), "Amplitude constrained mimo channels: Properties of optimal input distributions and bounds on the capacity," *Entropy*, vol. 21, no. 2, p. 200, 2019. [Online]. Available: https://www.mdpi.com/1099-4300/21/2/200

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

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

[25] S. Shamai (Shitz), "Capacity of a pulse amplitude modulated direct detection photon channel," *IEE Proc. I Commun., Speech Vis.*, vol. 137, no. 6, pp. 424–430, Dec. 1990.

[26] S. Verdú, "Fifty years of Shannon theory," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2057–2078, Oct. 1998.

[27] S. Karlin, "Pólya type distributions, II," *Ann. Math. Statist.*, vol. 28, no. 2, pp. 281–308, 1957.

[28] G. Casella and W. E. Strawderman, "Estimating a bounded normal mean," *Ann. Statist.*, vol. 9, no. 4, pp. 870–878, 1981.

[29] R. Tijdeman, "On the number of zeros of general exponential polynomials," in *Indagationes Mathematicae*, vol. 74. Amsterdam, The Netherlands: North Holland, 1971, pp. 1–7.

[30] A. T. Kalai, A. Moitra, and G. Valiant, "Efficiently learning mixtures of two Gaussians," in *Proc. 42nd ACM Symp. Theory Comput.*, 2010, pp. 553–562.

[31] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*. Hoboken, NJ, USA: Wiley, 2013.

[32] J. Bak, D. J. Newman, and D. J. Newman, *Complex Analysis*. New York, NY, USA: Springer, 1982.

[33] Y. Il'Yashenko and S. Yakovenko, "Counting real zeros of analytic functions satisfying linear ordinary differential equations," *J. Differential Equ.*, vol. 126, no. 1, pp. 87–105, 1996.

[34] S. Yagli, A. Dytso, H. V. Poor, and S. Shamai (Shitz), "Some aspects of totally positive kernels useful in information theory," in *Proc. IEEE Inf. Theory Workshop*, Marrakech, Morocco, 2019, pp. 1–6.

[35] S. Yagli, A. Dytso, H. V. Poor, and S. Shamai (Shitz), "Estimation of bounded normal mean: An alternative proof for the discreteness of the least favorable prior," in *Proc. IEEE Inf. Theory Workshop*, Visby, Sweden, Jun. 2019, pp. 1–6.

[36] B. Mamandipoor, K. Moshksar, and A. K. Khandani, "Capacity-achieving distributions in Gaussian multiple access channel with peak power constraints," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6080–6092, Oct. 2014.

[37] S. Karlin, "Decision theory for Pólya type distributions. Case of two actions, i," in *Proc. 3rd Berkeley Symp. Math. Statist. Probab.*, vol. 1, 1956, pp. 115–128. [Online]. Available: https://projecteuclid.org/euclid.bsmsp/1200501651

[38] M. Abramowitz and I. Stegun, "Handbook of mathematical functions," *Amer. J. Phys.*, vol. 34, no. 2, p. 177, 1966.

[39] J. Segura, "Bounds for ratios of modified Bessel functions and associated Turán-type inequalities," *J. Math. Anal. Appl.*, vol. 374, no. 2, pp. 516–528, 2011.

Alex Dytso is currently a Postdoctoral Researcher in the Department of Electrical Engineering at Princeton University. In 2016, he received a Ph.D. degree from the Department of Electrical and Computer Engineering at the University of Illinois, Chicago. He received his B.S. degree in 2011 from the University of Illinois, Chicago, where he also received the International Engineering Consortium's William L. Everitt Student Award of Excellence for outstanding seniors. His current research interest are in the areas of multi-user information theory and estimation theory, and their applications in wireless networks.

Semih Yagli received his Bachelor of Science degree in Electrical and Electronics Engineering in 2013, his Bachelor of Science degree in Mathematics in 2014 both from Middle East Technical University and his Master of Arts degree in Electrical Engineering in 2016 from Princeton University. Currently, he is pursuing his Ph.D. degree in Electrical Engineering at Princeton University under the supervision of H. Vincent Poor. His research interest include information theory, optimization, statistical modeling, privacy and unsupervised machine learning.

H. Vincent Poor (S'72–M'77–SM'82–F'87) received the Ph.D. degree in electrical engineering and computer science from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor of Electrical Engineering. During 2006 to 2016, he served as Dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other institutions, most recently at Berkeley and Cambridge. His research interests are in the areas of information theory and signal processing, and their applications in wireless networks, energy systems and related fields. Among his publications in these areas is the recent book *Information Theoretic Security and Privacy of Information Systems* (Cambridge University Press, 2017).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and is a foreign member of the Chinese Academy of Sciences, the Royal Society, and other national and international academies. Recent recognition of his work includes the 2017 IEEE Alexander Graham Bell Medal, the 2019 ASEE Benjamin Garver Lamme Award, a D.Sc. *honoris causa* from Syracuse University, awarded in 2017, and a D.Eng. *honoris causa* from the University of Waterloo, awarded in 2019.

Shlomo Shamai (Shitz) is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is now a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. His research interests encompasses a wide spectrum of topics in information theory and statistical communications. Dr. Shamai (Shitz) is an IEEE Life Fellow, an URSI Fellow, a member of the Israeli Academy of Sciences and Humanities and a foreign member of the US National Academy of Engineering. He is the recipient of the 2011 Claude E. Shannon Award, the 2014 Rothschild Prize in Mathematics/Computer Sciences and Engineering and the 2017 IEEE Richard W. Hamming Medal. He is a co-recipient of the 2018 Third Bell Labs Prize for Shaping the Future of Information and Communications Technology.

He is the recipient of numerous technical and paper awards and recognitions and is listed as a Highly Cited Researcher (Computer Science) for the years 2013/4/5/6/7/8. He has served as Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY, and has also served twice on the Board of Governors of the Information Theory Society. He has also served on the Executive Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY, the IEEE Information Theory Society Nominations and Appointments Committee and the IEEE Information Theory Society, Shannon Award Committee.