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

Alex Dytsou, Member, IEEE, Semih Yagli, Student Member, IEEE, H. Vincent Poor, Fellow, IEEE, and Shlomo Shamai (Shitz), 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 $O(A_n^2)$ 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 $O(A_n^2)$. Finally, the third main result of this paper reconsiders the case $n = 1$ with an additional average power constraint, demonstrating a similar $O(A_n^2)$ bound.

Keywords: 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

$$Y = X + Z,$$

where the input $X \in \mathbb{R}^n$ is independent of the standard Gaussian noise $Z \in \mathbb{R}^n$. We are interested in finding the capacity of the channel in (1) subject to the constraint that $X \in B_0(A)$ where $B_0(A)$ is an $n$-ball centered at zero with radius $A$ (i.e., amplitude or peak-power constrained input), that is

$$C_n(A) = \max_{X : X \in B_0(A)} I(X; Y).$$

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

$$X^* = R^* \cdot U^*$$

where the magnitude $R^*$ is discrete with finitely many points and the random unit vector $U^*$, which is independent of $R^*$, 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 an upper bound on the cardinality of the support of the distribution achieving

$$C(A, P) = \max_{X : E[X^2] \leq P} I(X; Y).$$
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 \( \{ \pm A \} \) so long as \( A \leq 0.1 \). Sharma and Shamai [8] extend the result of Smith, and argue\(^1\) that an equiprobable input on \( \{ \pm A \} \) is optimal if and only if \( A \leq \bar{A} \approx 1.665 \). 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 \( \{-A, 0, A\} \) is optimal for all \( \bar{A} \leq A \leq \bar{A} \approx 2.786 \).

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 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].

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], [16], [17], [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 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 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], [16], [17], [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 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 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], [16], [17], [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 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

\(^1\)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].

\(^2\)Also known as Redundancy-Capacity Theorem (see, e.g., [13], [14].)
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., $x$, $X$, $X^*$). We denote the distribution of a random vector $X$ by $P_X$. Moreover, we say that a point $x$ is in the support, denoted by supp$(P_X)$, of the distribution $P_X$ if for every open set $O \ni x$ we have that $P_X(O) > 0$. 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 $I$ is denoted by $N(Z, f)$. Similarly, if $f : C \to C$ is a function on the complex domain, $N(D, f)$ denotes the number of its zeros within the region $D$.

Finally, while the relative entropy between $X$ and $Y$ is denoted by $D(X \parallel Y)$, 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. Consider the amplitude constrained scalar additive Gaussian channel $Y = X + Z$ where the input $X$, satisfying $|X| \leq A$, is assumed to be independent from the white Gaussian noise $Z \sim \mathcal{N}(0, I_a)$. Let $X^* \sim P_{X^*}$ be the optimizing input distribution for this channel. Then, $P_{X^*}$ is a symmetric discrete distribution with $1\over 2 N([-R, R], f_{Y^*} - \kappa_1) \leq |\text{supp}(P_{X^*})|$ \leq N([-R, R], f_{Y^*} - \kappa_1) \leq a_2 A^2 + a_1 A + a_0, \tag{10}

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

$$\sqrt{1 + \frac{2 A^2}{\pi e}} \leq |\text{supp}(P_{X^*})| \leq a_2 A^2 + a_1 A + a_0, \tag{10}$$

with

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

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

3) Also known as “points of increase of $P_{X^*}$” or “spectrum of $P_{X^*}$.”

4) The definition $N(Z, f)$ is blind to the multiplicities of the zeros.

5) Unless otherwise stated, the logarithms in this paper are of base $e$.

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 $Y = X + Z$ where the input $X$, satisfying $\|X\| \leq A$, is assumed to be independent from the white Gaussian noise $Z \sim \mathcal{N}(0, I_a)$. Let $X^* \sim P_{X^*}$ be the optimizing input distribution for this channel. Then, $P_{X^*}$ is unique, radially symmetric, and the distribution of its amplitude, namely $P_{||X^*||}$, is a discrete distribution with $|\text{supp}(P_{||X^*||})| \leq a_{n_2} A^2 + a_{n_1} A + a_{n_0}, \tag{14}$

where, denoting the gamma function by $\Gamma$, $a_{n_2} = 4 + 4e + \sqrt{8e + 4}, \tag{15}$

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

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

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 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 $|X| \leq A$ and $E[|X|^2] \leq P$, is assumed to be independent from the white Gaussian noise $Z \sim \mathcal{N}(0, 1)$. Assuming $A \geq 1$, let $P_{X^*}$ be the optimizing input distribution for this channel. Then, $P_{X^*}$ is a symmetric discrete distribution with $\sqrt{1 + \frac{2 \min\{A^2, 3P\}}{\pi e}} \leq |\text{supp}(P_{X^*})| \leq a_{p_2} A^2 + a_{p_1} A + a_{p_0}, \tag{19}$

where

$$A_p = \frac{AP}{P - \log(1 + P) I\{P < A^2\}}. \tag{20}$$
\[ a_{p_2} = (1 + 2 \lambda_p)(9e + 6 \sqrt{e} + 1) + 2(2 - \lambda_p)(1 - 2 \lambda_p). \]  
(21)

\[ a_{p_1} = (1 + 2 \lambda_p)(6e + 2 \sqrt{e}). \]  
(22)

\[ a_{p_0} = (1 + 2 \lambda_p)e + 2 \log \left( \frac{2 + 4 \sqrt{e}(1 + 2 \lambda_p)}{1 - 2 \lambda_p} \right) + 1. \]  
(23)

\[ \lambda_p = \frac{\log(1 + P)}{2P} \cdot 1 \{ P < A^2 \}. \]  
(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 \( P \geq A^2 \), 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 \( P_X \).

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 \( |X| \leq A \), is independent from the noise \( Z \sim \mathcal{N}(0, 1) \). Then, \( P_X \) is the capacity-achieving input distribution if and only if the following two equations are satisfied:

\[ i(x; P_X^*) = C(A), \quad x \in \text{supp}(P_X^*), \]  
(25)

\[ i(x; P_X^*) \leq C(A), \quad x \in [-A, A], \]  
(26)

where \( C(A) \) denotes the capacity of the channel, and

\[ i(x; P_X^*) = \int e^{-\frac{(x-y)^2}{2}} \sqrt{\frac{1}{2\pi}} \log \frac{1}{f_{Y^*}(y)} \, dy - h(Z), \]  
(27)

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

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

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

\[ \text{supp}(P_X^*) \subseteq \{ x : i(x; P_X^*) = C(A) = 0 \}, \]

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

\[ |\text{supp}(P_X^*)| \leq \mathbb{N}([-A, A], \Xi_A(\cdot; P_X^*)), \]  
(29)

\[ \leq \mathbb{N}([-A, A], \Xi_A(\cdot; X^*)), \]  
(30)

where the function \( \Xi_A(\cdot; P_X^*) : \mathbb{R} \to \mathbb{R} \) is defined as

\[ \Xi_A(x; P_X^*) = i(x; P_X^*) - C(A). \]  
(31)

Note that, as it stands, the upper bound in (29) does not yet reveal any information on the discreteness of \( P_X^* \), as the right side might just as well be \( \infty \).

B. Connecting the Number of Oscillations of \( f_{Y^*} \) to the Number of Masses in \( P_X^* \).

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

**Definition 1.** Sign Changes of a Function. The number of sign changes of a function \( \xi \) is given by

\[ \mathcal{J}(\xi) = \sup_{m \in \mathbb{N}} \left\{ \sup_{y_1 < \cdots < y_m} \mathcal{N}(\xi(y_i)) \right\}, \]  
(32)

where \( \mathcal{N}(\xi(y_i)) \) 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 \( f_{Y^*} \) to the number of mass points of a capacity-achieving input distribution \( P_X^* \).

**Theorem 4.** Oscillation Theorem [27]. Given open intervals \( \mathbb{I}_1 \) and \( \mathbb{I}_2 \), let \( p : \mathbb{I}_1 \times \mathbb{I}_2 \to \mathbb{R} \) be a strictly totally positive kernel.\(^6\) For an arbitrary \( y \), suppose \( p(\cdot, y) : \mathbb{I}_1 \to \mathbb{R} \) is an \( n \)-times differentiable function. Assume that \( \mu \) is a measure on \( \mathbb{I}_2 \), and let \( \xi : \mathbb{I}_2 \to \mathbb{R} \) be a function with \( \mathcal{J}(\xi) = n \). For \( x \in \mathbb{I}_1 \), define

\[ \Xi(x) = \int \xi(y)p(x, y) \, d\mu(y). \]  
(33)

If \( \Xi : \mathbb{I}_1 \to \mathbb{R} \) is an \( n \)-times differentiable function, then either \( \mathbb{N}(\mathbb{I}_1, \Xi) \leq n \), 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 \( P_X^* \) and the number of zeros of the shifted optimal output pdf \( f_{Y^*} \) and recovers the bounds in (5) and (6).

**Theorem 5.** The support set of the capacity-achieving input distribution \( P_X^* \), satisfies

\[ \frac{1}{2} \mathbb{N}([-R, R], f_{Y^*} - \kappa_1) \leq |\text{supp}(P_X^*)| \]  
(34)

\[ \leq \mathbb{N}([-R, R], f_{Y^*} - \kappa_1) < \infty, \]  
(35)

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

**Proof.** To see (35) and (36), observe that \( \Xi_A(x; P_X^*) \), defined in (31), can be written as follows:

\[ \Xi_A(x; P_X^*) = \int_{\mathbb{R}} \xi_A(y) e^{-\frac{(x-y)^2}{2}} \, dy, \]  
(37)

where

\[ \xi_A(y) = \log \frac{1}{f_{Y^*}(y)} - C(A) - h(Z). \]  
(38)

\(^6\)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 \( \det \left[ \mathcal{J}(f(x_i, y_j)) \right]_{i,j=1}^n > 0 \) 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 \( \kappa_1 \in (0, \frac{1}{\sqrt{2\pi e}}) \).
First, observe that it is impossible for \( \Xi_A(x; P_X) = 0 \) for all \( x \in \mathbb{R} \) since otherwise \( \xi_A(y) \) would be zero for all \( y \in \mathbb{R} \). Furthermore, using the fact that the Gaussian distribution is a strictly totally positive kernel,

\[
|\text{supp}(P_X)\rangle \leq N(\mathbb{R}, \Xi_A(\cdot; P_X)) \quad (39)
\]

follows by observing that the number of zeros is an upper bound on the number of sign changes; (40) follows from Theorem 4; (41) follows because the number of zeros of \( f \) is a consequence of Lemma 1 (see Remark 4); (40) follows because the number of zeros of \( f \) is a consequence of Lemma 1 (see Remark 4); (41) follows from Theorem 4; (42) follows by observing that \( \xi_A(y) = 0 \) if and only if \( f_Y(y) = \kappa_1 \); and finally (43) and (44) follow from Lemma 2 in Section IV.

To see (34), using the fact that \( \text{supp}(P_X) \) is a finite set, suppose \( |\text{supp}(P_X)\rangle = n \), let \( x_1 < \ldots < x_n \) the elements of \( \text{supp}(P_X) \), and write

\[
f_Y(y) = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{n} P_X(x_i) \exp\left(-\frac{1}{2}(y-x_i)^2\right), \quad (45)
\]

where, by the definition of \( \text{supp}(P_X) \), the probabilities satisfy \( P_X(x_i) > 0 \) for each \( i = 1, \ldots, n \). Observing that the number of zeros of \( f_Y(y) = \kappa_1 \); is the same as the number of zeros of the right-shifted function \( f_Y(y - |x_1| = \kappa_1, \) let

\[
f(y) = f_Y(y - |x_1| = \kappa_1, \) \quad (46)
\]

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

\[
u_i = x_i + |x_1| + 1 \quad \text{for} \quad i = 1, \ldots, n, \quad (48)
\]

\[
a_i = \begin{cases} \kappa_1 & i = 0 \\ \frac{1}{\sqrt{2\pi}} P_X(x_i) & i = 1, \ldots, n. \end{cases} \quad (49)
\]

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

\[
\tilde{f}(y, \epsilon_1, \ldots, \epsilon_n) = \sum_{i=1}^{n} a_i \exp\left(-\frac{1}{2}(1+\epsilon_i)(y-u_i)^2\right) - a_0, \quad (50)
\]

Note that

\[
e^{-2y^2} \tilde{f}(y, \epsilon_1, \ldots, \epsilon_n) = \sum_{i=0}^{n} b_i \exp\left(-\frac{(2+\epsilon_i)}{2}(y-v_i)^2\right), \quad (51)
\]

where

\[
\epsilon_0 = -1, \quad (52)
\]

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

\[
u_i = \begin{cases} 0 & i = 0, \\ \frac{1+\epsilon_i}{2+\epsilon_i} u_i & i = 1, \ldots, n, \end{cases} \quad (54)
\]

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 \( \epsilon_i \)'s and since \( \tilde{f}(y, \epsilon_1, \ldots, \epsilon_n) \rightarrow f(y) \) as \( (\epsilon_1, \ldots, \epsilon_n) \rightarrow (0, \ldots, 0) \), it follows that

\[
2|\text{supp}(P_X)\rangle \geq N(\mathbb{R}, f(y)) \quad (55)
\]

\[
= N(\mathbb{R}, f_Y(y) - \kappa_1) \quad (56)
\]

\[
= N([R, R], f_Y(y) - \kappa_1). \quad (57)
\]

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 \( P_X \) is finite; hence, \( P_X \) 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 \( f_Y = \kappa_1 \), which is the topic of the next section.

Remark 6. Theorem 5 proves that the number of zeros of the shifted optimal output pdf \( f_Y = \kappa_1 \) gives an order tight upper bound on the support size of the optimal input pmf \( |\text{supp}(P_X)\rangle \). 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 \( P_X \) is within a factor of two of the number of zeros of \( f_Y = \kappa_1 \) where \( \kappa_1 = e^{-C(A) - h(Z)} \). In this section, we first provide an upper bound on the number of zeros of \( f_Y = \kappa_1 \) and establish (10). Additionally, through the use of entropy-power inequality, we also provide a lower bound on the support size of \( P_X \), yielding (8).

Remark 7. A critical observation here is that, due to the lack of knowledge of the optimal input distribution \( P_X \), or the capacity expression \( C(A) \), we do not know the optimal output distribution \( f_Y \), nor the constant \( \kappa_1 = e^{-C(A) - h(Z)} \). Therefore, we must instead work with generic \( f_Y \) and \( \kappa_1 \) 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 \( 0 \leq \kappa_1 \leq \max_{b} f_Y(b) \), find a worst-case upper bound on the number of zeros of the shifted output pdf

\[
f_Y = \kappa_1, \quad (58)
\]

where \( f_Y \) denotes the pdf of the random variable \( Y = X + Z \), with \( X \) being an arbitrary zero mean random variable at the input of the channel satisfying the amplitude constraint:

\[8\]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.
$|X| \leq A$; $Z$ 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 $[-A, A]$.

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

$$N(\mathbb{R}, f_Y - \kappa_1) = N([-B_{\kappa_1}, B_{\kappa_1}], f_Y - \kappa_1) < \infty.$$  \hspace{1cm} (58)

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

$$B_{\kappa_1} \leq A + \log^{\frac{1}{2}} \left( \frac{1}{2\pi \kappa_1^2} \right). \hspace{1cm} (60)$$

**Proof.** Using the monotonicity of $e^{-x^2}$, for all $|y| > A$,

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

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

$$f_Y(y) - \kappa_1 < 0 \hspace{1cm} (63)$$

for all $|y| > A + \log^{\frac{1}{2}} \left( \frac{1}{2\pi \kappa_1^2} \right). \hspace{1cm} (64)$

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

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. \hfill \blacksquare

Since the exact value of the constant $\kappa_1$ 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 $\kappa_1$.

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

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

where $f'$ denotes the derivative of $f$.

**Proof.** Let $x_1 < \ldots < x_n$ denote the zeros of $f$. By Rolle’s Theorem, each of the intervals $(x_i, x_{i+1})$ for $i = 1, \ldots, n-1$ contains at least one extreme point. \hfill \blacksquare

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

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

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 $\bar{f}_Y'$ denote its complex analytic extension,

$$N([-R, R], \bar{f}_Y') \leq \inf_{\epsilon > 0} N(D_{\epsilon+}, \bar{f}_Y') \leq \inf_{\epsilon > 0} \sup_{\partial D_{\epsilon+}} f_Y'(z) dz = \inf_{\epsilon > 0} (R + \epsilon) \frac{1}{\pi} \max_{|z| = R + \epsilon} |f_Y'(z)|. \hspace{1cm} (69)$$

where in (67) $D_t \subseteq \mathbb{C}$ is an open disc of radius $t$ centered at the origin and the inequality follows because $f_Y'(y) = 0 \implies \bar{f}_Y'(y) = 0$; in (68) $\partial D_t$ denotes the boundary of the disc $D_t$ 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 $\bar{f}_Y'$, the maximization of the ratio $\frac{f_Y'}{\bar{f}_Y'}$ 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 $\bar{f}_Y'$. 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 $N(D_R, f)$ within the disk $D_R = \{ z : |z| \leq R \}$ satisfies

$$N(D_R, f) \leq \frac{1}{\log s} \left( \log \max_{|z| \leq (st+s+t)R} |f(z)| - \log \max_{|z| \leq tR} |f(z)| \right). \hspace{1cm} (70)$$

The following two lemmas find upper and lower bounds on absolute value of the complex analytic extension9 of $f_Y'$ over a disc of finite radius centered at the origin.

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

10The 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].
Lemma 5. Suppose \( f'_Y : \mathbb{R} \to \mathbb{R} \) is as in (66) and let \( \tilde{f}'_Y : \mathbb{C} \to \mathbb{C} \) denote its complex extension. Then,
\[
\max_{|z| \leq B} |\tilde{f}'_Y(z)| \leq \frac{1}{\sqrt{2\pi}} (A + B) \exp \left( \frac{B^2}{2} \right). \tag{71}
\]

Proof. Using the standard rectangular representation of a complex number, let \( z = \xi + i\eta \in \mathbb{C} \),
\[
\max_{|z| \leq B} |\tilde{f}'_Y(z)| = \max_{|z| \leq B} \left\{ \frac{1}{\sqrt{2\pi}} \mathbb{E} \left[ (X - z) \exp \left( -\frac{(z - X)^2}{2} \right) \right] \right\} \tag{72}
\]
\[
\leq \max_{|z| \leq B} \left\{ \frac{1}{\sqrt{2\pi}} \mathbb{E} \left[ |X - z| \exp \left( -\frac{(z - X)^2}{2} \right) \right] \right\} \tag{73}
\]
\[
= \max_{|z| \leq B} \left\{ \frac{1}{\sqrt{2\pi}} \mathbb{E} \left[ |X - \xi - i\eta|^2 \exp \left( -\frac{(\eta^2 - (\xi - X)^2)}{2} \right) \right] \right\} \tag{74}
\]
\[
\leq \max_{|z| \leq B} \left\{ \frac{1}{\sqrt{2\pi}} \mathbb{E} \left[ (|X| + |z|)^2 \exp \left( -\frac{\eta^2}{2} \right) \right] \right\} \tag{75}
\]
\[
\leq \frac{1}{\sqrt{2\pi}} (A + B) \exp \left( \frac{B^2}{2} \right), \tag{76}
\]
where (73) follows from Jensen’s inequality; (75) follows from triangle inequality; and finally (76) is because \( |X| \leq A \), and \( |z| \leq B \) implies \( |\eta| \leq B \).

Lemma 6. Suppose \( f'_Y : \mathbb{R} \to \mathbb{R} \) is as in (66) and let \( \tilde{f}'_Y : \mathbb{C} \to \mathbb{C} \) denote its complex extension. For any \( |X| \leq A \leq B \),
\[
\max_{|z| \leq B} |\tilde{f}'_Y(z)| \geq \frac{A}{\sqrt{2\pi}} \exp \left( -2A^2 \right). \tag{77}
\]

Proof. Thanks to the suboptimal choice of \( z = A \leq B \),
\[
\max_{|z| \leq B} |\tilde{f}'_Y(z)| \geq \frac{1}{\sqrt{2\pi}} \mathbb{E} \left[ (X - A) \exp \left( -\frac{(A - X)^2}{2} \right) \right] \tag{78}
\]
\[
\geq \frac{1}{\sqrt{2\pi}} \mathbb{E} \left[ (A - X) \exp \left( -2A^2 \right) \right] \tag{79}
\]
\[
= \frac{A}{\sqrt{2\pi}} \exp \left( -2A^2 \right), \tag{80}
\]
where (79) follows because \( |X| \leq A \); 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 \( |X| \leq A < R \) for some fixed \( R \). Then, the number of extreme points of \( f_Y \), namely the number of zeros of \( \tilde{f}'_Y \), within the interval \([-R, R]\) satisfies
\[
N([-R, R], f'_Y) \leq \min_{s > 1} \left\{ \frac{\left( \frac{(A+R)+A)^2}{2} + 2A^2 + \log \left( 2 + \frac{(A+R)s}{A} \right) \right)}{\log s} \right\}. \tag{81}
\]

Proof. Let \( \mathcal{D}_R \subset \mathbb{C} \) be a disk of radius \( R \) centered at \( z_0 = 0 \), and note that
\[
N([-R, R], f'_Y) \leq N(D_R, \tilde{f}'_Y) \tag{82}
\]
\[
\leq \min_{s > 1, t > \frac{R}{4}} \left\{ \frac{\log \max_{x \leq \max(s+x+R)} |\tilde{f}'_Y(z)|}{\log s} \right\} \tag{83}
\]
\[
\leq \min_{s > 1, t > \frac{R}{4}} \left\{ \frac{(s+R)^2 + 2A^2 + \log \left( 2 + \frac{(A+R)s}{A} \right)}{\log s} \right\} \tag{84}
\]
\[
= \min_{s > 1} \left\{ \frac{(A+R)s)^2 + 2A^2 + \log \left( 2 + \frac{(A+R)s}{A} \right)}{\log s} \right\}. \tag{85}
\]
where (82) follows because zeros of \( f'_Y \) are also zeros of its complex extension \( \tilde{f}'_Y \); (83) is a consequence of Lemma 4; (84) follows from Lemmas 5 and 6; and finally, in (85), we use the fact that \( t = \frac{R}{4} \) is the minimizer in the right 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 \( \kappa_1 \in (0, \frac{1}{\pi^2}) \), suppose \( R > A + \log \frac{1}{2\pi \kappa_1^2} \). Then, the number of zeros of \( f_Y - \kappa_1 \) satisfies
\[
N(R, f_Y - \kappa_1) \leq N([-R, R], f'_Y - \kappa_1) \tag{86}
\]
\[
\leq 1 + \min_{s > 1} \left\{ \frac{\left( \frac{(A+R)+A)^2}{2} + 2A^2 + \log \left( 2 + \frac{(A+R)s}{A} \right) \right)}{\log s} \right\}. \tag{87}
\]

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 \( \kappa_1 \). If \( \kappa_1 \approx 0 \), then \( N(R, f_Y - \kappa_1) \leq 2 \) and the bound above may be quite loose. In applying Corollary 1 in the next section, we let
\[
\kappa_1 = \frac{e^{-C(A)}}{\sqrt{2\pi \varepsilon}} \tag{88}
\]
where \( C(A) \) denotes the capacity of the amplitude constrained additive Gaussian channel. In that case, it can be shown that
\[
(2\pi e \left( 1 + A^2 \right))^{-\frac{1}{2}} \leq \kappa_1 \leq (2\pi e + 4A^2)^{-\frac{1}{2}}, \tag{89}
\]
and the result presented above is more relevant.

Remark 9. We believe that the bound in Theorem 6, and hence the one in Corollary 1, can be further tightened. In fact, we conjecture that\(^{11}\)
\[
\max_{X \in [-A, A]} N(R, f'_Y) = \Theta(A). \tag{90}
\]

\(^{11}\)Let \( f(x) \) and \( g(x) \) be two nonnegative valued functions. Then, \( f \) is \( \Theta(g(x)) \) if and only if \( c_1 g(x) \leq f(x) \leq c_2 g(x) \) for some \( c_1, c_2 > 0 \) and all \( x > x_0 \).
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.

\[ \log_A N(B, f_Y) \]

![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 \( \kappa_1 \) 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 a worst-case exhaustive numerical search.](image)

A possible bottleneck in our proof above is the bound in (82), where the function is extended to the complex plane, and the number of zeros are counted over a disk rather than over a finite interval. Doing so is effectively doubling the dimension of the problem. In other words, the produced order \( A^2 \) bound follows because extending \( f_Y \) to the complex domain potentially creates a ton of zeros that our analysis counts even though the function in the real domain cannot possibly realize those zeros. To work around this issue, one might consider using another version of Tijdeman’s Lemma [33, Lemma 1] which works with arbitrary open sets, unlike the bound in Lemma 4 which works only over discs in the complex plane. Improvement is left for the future work.

B. Proof of the Upper Bound in (10)

We begin by simplifying the previously provided upper bound on \( B_{\kappa_1} \). Note that an amplitude constraint \( |X| \leq A \) induces a second moment constraint \( E[X^2] \leq A^2 \), and therefore

\[ C(A) = \max_{|X| \leq A, E[X^2] \leq A^2} I(X; Y) \]

\[ \leq \frac{1}{2} \log (1 + A^2). \]  

(92)

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

\[ \frac{1}{\kappa_1} = \exp(C(A) + h(Z)) \leq \sqrt{2\pi e} (1 + A^2). \]  

(93)

(94)

Capitalizing on the bound in (60),

\[ B_{\kappa_1} \leq A + \sqrt{1 + \log(1 + A^2)} \]

\[ \leq 2A + 1, \]  

(95)

(96)

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

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

\[ N([-B_{\kappa_1}, B_{\kappa_1}], f_{Y^*} - \kappa_1) - 1 \]

\[ \leq N([-2A - 1, 2A + 1], f_{Y^*} - \kappa_1) - 1 \]

\[ \leq \min_{s > 1} \left\{ \frac{((3A+1)s+A)^2}{2} + 2A^2 + \log \left( 2 + \frac{(3A+1)s}{A} \right) \right\} \]

\[ \leq \min_{s > 1} \left\{ \frac{1}{\log s} \left( \frac{(3s+1)s+A}{2} + 2A^2 + \log(2 + 4s) \right) \right\} \]

\[ \leq \left( e + 2 \log(4\sqrt{e} + 2) + (4e + 2\sqrt{e})A + (5 + 4\sqrt{e} + 4)eA^2 \right) \]

\[ = a_2A^2 + a_1A + a_0 - 1, \]

(97)

(98)

(99)

(100)

where (99) follows because \( 3A + 1 \leq 4A \) for\(^{12} \) \( A \geq 1 \); (100) follows by choosing a suboptimal value \( s = \sqrt{e} \) in the minimization; and (101) follows by letting \( a_2 = 9e + 6\sqrt{e} + 5, \)

\[ a_1 = 6e + 2\sqrt{e} \]  

\[ a_0 = e + 2 \log(4\sqrt{e} + 2) + 1. \]

This completes the proof of Corollary 2.

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 \( H(P_{X^*}) \) the entropy of the optimizing input distribution \( P_{X^*} \), it follows that

\[ \frac{1}{2} \log \left( 1 + \frac{2A^2}{\pi e} \right) \leq \max_{X: |X| \leq A} I(X; Y). \]  

(102)

(103)

(104)

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

\(^{12}\)The unessential assumption that \( A \geq 1 \) is just for simplifying the presentation. In the case when \( A \leq 1 \), the optimality of \( P_{X^*} \) that is equiprobable on \( \mathcal{X} = \{-A, A\} \) is known [8].
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 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 \( X^\star \), it has been shown that its magnitude \( ||X^\star|| \) is a discrete random variable with at most \( O(A^2) \) 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 \( O(A) \) rather than \( O(A^2) \). 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 the space of all distributions). 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 \( \mathbb{R}^{2n} \) where \( n = O(A^2) \), and where \( v \in \mathbb{R}^{2n} \) consists of \( \{p_1, \ldots, p_n, x_1, \ldots, x_n\} \) 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 functions 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.

APPENDIX A

PROOF OF THEOREM 2

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

**Lemma 7.** Consider the amplitude constrained vector additive Gaussian channel \( Y = X + Z \) where the input \( X \), satisfying \( \|X\| \leq A \), is independent from the white Gaussian noise \( Z \sim \mathcal{N}(0, I_n) \). If \( X^\star \) is an optimal input, the distribution of its magnitude, namely \( P_{R^\star} = \hat{P}_r(\cdot ||X^\star||) \), satisfies

\[
i_n(r; P_{R^\star}) = C_n(A) + n_r, \quad r \in \text{supp}(P_{R^\star}),
\]

\[
i_n(r; P_{R^\star}) \leq C_n(A) + \nu_n, \quad r \in [0, A],
\]

where \( C_n(A) \) denotes the capacity of the channel, and

\[
i_n(r; P_{R^\star}) = \int_0^\infty f_{X_n^\star}(x|r) \log \frac{1}{g_n(x; P_{R^\star})} \, dx,
\]

\[
f_{X_n^\star}(x|r) = \frac{1}{2} \exp \left( -\frac{x^2 + r^2}{2} \right) \left( \frac{\sqrt{r}}{2} \right)^{r-1} I_{r-1}(r \sqrt{x}),
\]

\[g_n(x; P_{R^\star}) = \int_0^A 2f_{X_n^\star}(x|r) \frac{1}{x^{r-1}} \, dP_{R^\star}(r),\]

\[\nu_n = \frac{n}{2} + \log \left( \frac{2^{\frac{1}{2}} - 1}{\Gamma \left( \frac{n}{2} \right)} \right),
\]

with \( I_n(x) \) denoting the modified Bessel function of the first kind of order \( n \).

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

\[\kappa_n = \exp(-C_n(A) - \nu_n),\]

\[\Phi_n(s; P_{R^\star}) = \eta_n(s; P_{R^\star}) + \log \kappa_n,\]

\[\phi_n(x; P_{R^\star}) = \log \frac{\kappa_n}{g_n(x; P_{R^\star})},\]

and observe that

\[\Phi_n(r; P_{R^\star}) = \int_0^\infty \phi_n(x; P_{R^\star}) f_{X_n^\star}(x|r) \, dx,
\]

where \( f_{X_n^\star}(x|r) \) is as defined in (108). Note that since \( f_{X_n^\star}(x|r) \) 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),

\[|\text{supp}(P_{R^\star})| \leq N \left[ (0, A], \text{\Phi}_n(\cdot; P_{R^\star}) \right) \]

\[\leq 1 + N \left( (0, \infty), \Phi_n(\cdot; P_{R^\star}) \right) \]

\[\leq 1 + N \left( (0, \infty), \phi_n(\cdot; P_{R^\star}) \right) \]

\[= 1 + N \left( (0, \infty), g_n(\cdot; P_{R^\star}) - \kappa_n \right)
\]

\[14\text{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.}
\[ \begin{align*}
&\leq 1 + N \left( [0, B_{\kappa_n}], g_n(\cdot; P_{R^*}) - \kappa_n \right) \\
&\leq a_n A^2 + a_n A + a_n, \\
\end{align*} \]
where (115) is a consequence of Lemma 7; the extra +1 in (116) is just to account for the possibility that \( \Phi_n(0; P_{R^*}) = 0 \); (117) follows from Karlin’s Oscillation Theorem, see Theorem 4; (118) follows since \( \phi_n(\cdot; P_{R^*}) \) has the same zeros as \( g_n(\cdot; P_{R^*}) - \kappa_n \); (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 \( z \in \mathbb{C} \)
\[ \left| I_n(z) \right| \leq \frac{\sqrt{\pi} |z|^n}{2^n \Gamma(n + \frac{1}{2})} e^{\Re(z)}. \]  
(121)

*Proof.* Thanks to the integral representation of the modified Bessel function of the first kind, see [38, 9.6.18],
\[ I_n(z) = \left( \frac{1}{2} z \right)^n \int_0^\infty e^{\cos(\theta)} \sin^{2n}(\theta) d\theta, \]
(122)
it follows from the modulus inequality that
\[ |I_n(z)| \leq \frac{\left( \frac{1}{2} |z| \right)^n}{\sqrt{\pi} \Gamma(n + \frac{1}{2})} \int_0^\infty |e^{\cos(\theta)}| \sin^{2n}(\theta) d\theta \]
\[ \leq \frac{\sqrt{\pi} |z|^n}{2^n \Gamma(n + \frac{1}{2})} e^{\Re(z)}, \]
(124)
where (124) follows because \( |e^{\cos(\theta)}| \sin^{2n}(\theta) \leq e^{\Re(z)} \).

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

**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 \( B_{\kappa_n} < \infty \) such that
\[ N \left( [0, \infty), g_n(\cdot; P_R) - \kappa_n \right) = N \left( [0, B_{\kappa_n}], g_n(\cdot; P_R) - \kappa_n \right) < \infty. \]
(125)

In particular, there are finitely many zeros of \( g_n(\cdot; P_R) - \kappa_n \) all of which are contained within the interval \( [0, B_{\kappa_n}] \). Moreover, \( B_{\kappa_n} \) can be upper bounded as follows:
\[ B_{\kappa_n} \leq \left( A + \sqrt{A^2 + 2 \log \left( \frac{\gamma_n}{\kappa_n} \right)} \right)^2, \]
(127)
where
\[ \gamma_n = \frac{\sqrt{\pi}}{2^{n-1} \Gamma(n - \frac{1}{2})}. \]
(128)

*Proof.* From the definition of the pdf \( g_n(\cdot; P_R) \) in (109),
\[ g_n(x; P_R) \]
\[ = \int_0^A \exp \left( -\frac{x+r^2}{2} \right) \frac{I_{\frac{n}{2}-1}(r \sqrt{x})}{(r \sqrt{x})^{\frac{1}{2}-1}} dP_R(r) \]
(129)
\[ \leq \int_0^A \sqrt{\pi} \frac{\sqrt{\pi}}{2^{n-1} \Gamma(n - \frac{1}{2})} \exp \left( -\frac{1}{2} (\sqrt{x} - r)^2 \right) dP_R(r), \]
(130)
\[ \leq \sqrt{\pi} \frac{\sqrt{\pi}}{2^{n-1} \Gamma(n - \frac{1}{2})} \exp \left( -\frac{x}{2} + A \sqrt{x} \right), \]
(131)
where (130) follows from Lemma 8; and (131) utilizes \( r \in [0, A] \). Since the right side of (131) is decreasing for all \( x > A^2 \), it follows that
\[ g_n(x; P_R) - \kappa_n < 0 \]
(132)
for all
\[ x > \left( A + \sqrt{A^2 + 2 \log \left( \frac{\gamma_n}{\kappa_n} \right)} \right)^2. \]
(133)
This means that there exists a \( B_{\kappa_n} \), satisfying (127) such that all zeros of \( g_n(\cdot; P_R) - \kappa_n \) are contained within the interval \([0, B_{\kappa_n}]\).

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 \( x \in (0, \infty) \)
\[ \frac{d}{dx} g_n(x; P_R) = E \left[ \exp \left( -\frac{x+R^2}{2} \right) \frac{R}{2(R \sqrt{x})^{\frac{1}{2}-1}} \left( \frac{2}{\sqrt{x}} I_{\frac{1}{2}}(R \sqrt{x}) - I_{\frac{1}{2}-1}(R \sqrt{x}) \right) \right], \]
(134)
where \( R \sim P_R \).

*Proof.* First of all, given \( r > 0 \), observe that for \( u \in (0, \infty) \)
\[ t_n(u|r) = 2 \left( \frac{r}{u} \right)^{n-2} \exp \left( \frac{r^2}{2} \right) f_{\frac{1}{2}} \left( \frac{u^2}{r^2} \right), \]
(135)
\[ = e^{-\frac{u^2}{2r^2}} I_{\frac{1}{2}-1}(u) \]
(136)
is a differentiable function and
\[ \frac{d}{du} t_n(u|r) = e^{-\frac{u^2}{2r^2}} \left( I_{\frac{1}{2}}(u) - \frac{u}{r^2} I_{\frac{1}{2}-1}(u) \right), \]
(137)
where we have employed the fact that [38, Eqn. (9.6.26)]
\[ \frac{d}{du} I_{\frac{1}{2}-1}(u) = I_{\frac{1}{2}}(u) + \frac{n-2}{2u} I_{\frac{1}{2}-1}(u). \]
(138)
The desired result then follows from the chain rule as
\[ g_n(x; P_R) = \int_0^A e^{-r^2/2} t_n(r \sqrt{x}) dP_R(r). \]
(139)

\(^{15}\)For a proof, refer to [3, Appendix I] and [5, Propositions 1 and 2] for the respective cases of \( n = 2 \), and \( n \geq 2 \).
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 \( \tilde{g}_n'(x; P_R) \).

**Lemma 11.** Given \( r > 0 \) and \( D > 0 \)

\[
I_{2-1}(Dr) - \frac{r}{D} I_{2-1}(Dr) \\
\geq (Dr)^{\frac{21}{2} - \frac{7}{\Gamma(\frac{n}{2})}} \left( 1 - \frac{2r^2}{n - 1 + \sqrt{(n-1)^2 + (2Dr)^2}} \right) \tag{140}
\]

\[
> 0.
\tag{141}
\]

**Proof.** Using the fact that \( I_n(x) > 0 \) for \( x > 0 \)

\[
I_{2-1}(Dr) \left( 1 - \frac{r}{D} \frac{I_{2-1}(Dr)}{I_{2-1}(x)} \right) \\
\geq I_{2-1}(Dr) \left( 1 - \frac{2r^2}{n - 1 + \sqrt{(n-1)^2 + (2Dr)^2}} \right) \tag{142}
\]

\[
\geq (Dr)^{\frac{21}{2} - \frac{7}{\Gamma(\frac{n}{2})}} \left( 1 - \frac{2r^2}{n - 1 + \sqrt{(n-1)^2 + (2Dr)^2}} \right) \tag{143}
\]

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

\[
\frac{I_2(x)}{I_{2-1}(x)} \leq \frac{2x}{n - 1 + \sqrt{(n-1)^2 + (2x)^2}};
\tag{144}
\]

and (143) follows from the fact that \( x^{-n}I_n(x) \) is monotonically increasing for \( x > 0 \) and that

\[
\lim_{x \to 0} x^{-n}I_n(x) = 2^{-n} \Gamma^{-1}(n+1). \tag{145}
\]

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 \( \tilde{g}_n'(z; P_R) \) on a disc centered at \( z_0 \in \mathbb{C} \) where

\[
z_0 = \frac{B_{\kappa_n}}{2} + i0. \tag{146}
\]

**Lemma 12.** Suppose \( D > 0 \). For \( B_{\kappa_n} \leq 2D^2 \)

\[
\max_{|z| \leq M^2} \left| \tilde{g}_n'(z + \frac{B_{\kappa_n}}{2}; P_R) \right| \\
\geq \left\{ \begin{array}{l}
\frac{2D^2 + A^2}{2} \Gamma \left( \frac{n}{2} \right) \exp \left( \frac{2r^2}{2} \right) \\
\geq \left( 1 - \frac{2A^2}{n - 1 + \sqrt{(n-1)^2 + (2DA)^2}} \right). \tag{147}
\end{array} \right.
\]

**Proof.** Observe that for \( R \sim P_R \)

\[
\max_{|z| \leq D^2} \left| \tilde{g}_n'(z + \frac{B_{\kappa_n}}{2}; P_R) \right| \geq \left| \tilde{g}_n'(D^2; P_R) \right| \geq \left| \tilde{g}_n'(D^2; P_R) \right| \geq \left| \tilde{g}_n'(D^2; P_R) \right| \geq \left( 1 - \frac{2A^2}{n - 1 + \sqrt{(n-1)^2 + (2DA)^2}} \right). \tag{148}
\]

**Lemma 13.** Suppose \( M > 0 \). For \( \kappa_n \leq 2M^2 \)

\[
\max_{|z| \leq M^2} \left| \tilde{g}_n'(z + \frac{B_{\kappa_n}}{2}; P_R) \right| \\
\leq \gamma_n \left( \frac{A^2}{n - 1} + 1 \right) \exp \left( \frac{1}{2} (A + \sqrt{2M})^2 \right), \tag{152}
\]

where \( \gamma_n \) 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

\[
\left| \tilde{g}_n'(z; P_R) \right| \\
= \left| \mathbb{E} \left[ \frac{\exp \left( -\frac{z+R^2}{2} \right)}{2(R\sqrt{z})^{\frac{n}{2} - 1}} \left( \frac{R}{\sqrt{z}} I_{\frac{2}{2}}(R\sqrt{z}) - I_{\frac{2}{2-1}}(R\sqrt{z}) \right) \right] \right| \tag{153}
\]

\[
\leq \mathbb{E} \left[ \left| \exp \left( -\frac{z+R^2}{2} \right) \frac{R}{2(R\sqrt{z})^{\frac{n}{2} - 1}} \left( I_{\frac{2}{2}}(R\sqrt{z}) \right) \right| \left| I_{\frac{2}{2-1}}(R\sqrt{z}) \right| \right] \tag{154}
\]

\[
\leq \mathbb{E} \left[ \frac{\sqrt{\pi}}{2^{\frac{n}{2}} \Gamma \left( \frac{n}{2} - 1 \right)} \left( \frac{R^2}{n - 1} + 1 \right) e^{\frac{\gamma_n}{2} \left( \frac{R\sqrt{z} + \frac{B_{\kappa_n}}{2}}{2} \right)^2} \right], \tag{155}
\]

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 \( R \in [0, A] \), we simply observe that

\[
\max_{|z| \leq M^2} \left| \tilde{g}_n'(z + \frac{B_{\kappa_n}}{2}; P_R) \right| \\
\leq \max_{|z| \leq M^2} \mathbb{E} \left[ \frac{\sqrt{\pi}}{2^{\frac{n}{2}} \Gamma \left( \frac{n}{2} - 1 \right)} \left( \frac{R^2}{n - 1} + 1 \right) e^{\frac{\gamma_n}{2} \left( |R| + \frac{B_{\kappa_n}}{2} \right)^2} \right] \tag{156}
\]

\[
\leq \max_{|z| \leq M^2} \mathbb{E} \left[ \frac{\sqrt{\pi}}{2^{\frac{n}{2}} \Gamma \left( \frac{n}{2} - 1 \right)} \left( A^2 \exp \left( \frac{1}{2} (A + \sqrt{2M})^2 \right) \right) \right] \tag{157}
\]

\[
\leq \frac{\sqrt{\pi}}{2^{\frac{n}{2}} \Gamma \left( \frac{n}{2} - 1 \right)} \left( A^2 \exp \left( \frac{1}{2} (A + \sqrt{2M})^2 \right) \right), \tag{158}
\]

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, together 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 \( \text{supp}(P_R) \in [0, A] \) and \( B_{\kappa_n} \) is as defined in Lemma 9. The number of zeros of \( g_n(\cdot; P_R) - \kappa_n \) within \([0, B_{\kappa_n}]\) satisfies

\[
N ([0, B_{\kappa_n}], g_n(\cdot; P_R) - \kappa_n) \leq a_n z^2 + a_n A + a_n - 1,
\]

where \( a_n, a_{n_1}, \) and \( a_{n_0} \) are as defined in (15), (16), and (17), respectively.

**Proof.** In light of Lemma 9, let

\[
B_{\kappa_n} = \left( A + \sqrt{A^2 + 2 \log \frac{\gamma_n}{\kappa_n}} \right)^2,
\]

and note that

\[
N ([0, B_{\kappa_n}], g_n(\cdot; P_R) - \kappa_n)
\leq 1 + N (0, B_{\kappa_n}], g_n'(\cdot; P_R))
\leq 1 + \max \left\{ \log \frac{\max_{[x \leq (z + \frac{B_{\kappa_n}}{2}; P_R]} g_n'(z + \gamma_n; P_R)}{\max_{[x \leq \frac{B_{\kappa_n}}{2}; P_R]} g_n'(z + \gamma_n; P_R)} \right\}
\leq 1 + \min_{s > 1, t > 0} \left\{ \log \frac{\sqrt{\max_{[x \leq (z + \frac{B_{\kappa_n}}{2}; P_R]} g_n'(z + \gamma_n; P_R)}}{\max_{[x \leq \frac{B_{\kappa_n}}{2}; P_R]} g_n'(z + \gamma_n; P_R)} \right\}
\leq 1 + \max_{[2z \leq (2e+1)B_{\kappa_n}]} \log \left| g_n'(z + \frac{B_{\kappa_n}}{2}; P_R) \right|
\quad - \max_{[2z \leq B_{\kappa_n}]} \log \left| g_n'(z + \frac{B_{\kappa_n}}{2}; P_R) \right|
\leq \log \left( \frac{\exp 1}{\Gamma \left( \frac{3}{4} + e \right)} \right) + \left( \frac{3}{4} + e \right) \left( \frac{1}{\Gamma \left( \frac{2}{3} \right)} \right) \log \left( \frac{A^2}{n - 1} + 1 \right)
\quad - \log \left( 1 - \frac{2A^2}{n - 1 + \sqrt{(n - 1)^2 + 2B_{\kappa_n}A^2}} \right)
\leq \log \left( \frac{\exp 1}{\Gamma \left( \frac{3}{4} + e \right)} \right) + \left( \frac{3}{4} + e \right) \left( \frac{1}{\Gamma \left( \frac{2}{3} \right)} \right) \log \left( \frac{A^2}{n - 1} + 1 \right)
\quad + \sqrt{\frac{32}{n - 1}} A
\leq a_n z^2 + a_n A + a_n - 1,
\]

(167) follows from Rolle’s Theorem; in (163) \( D_r \subset \mathbb{C} \) 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 \( g_n'' \); (164) follows from Tijdeman’s Number of Zeros Lemma, see Lemma 4; (165) follows from the suboptimal choices:

\[
s = e,
\]

\[
t = \frac{B_{\kappa_n}}{\kappa_n};
\]

(166) follows from Lemmas 12 and 13 with

\[
D^2 \leftarrow \frac{1}{2} \Sigma_{\kappa_n},
\quad M^2 \leftarrow \frac{2e + 1}{2} \Sigma_{\kappa_n}.
\]

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

\[
1 - \frac{2A^2}{n - 1 + \sqrt{(n - 1)^2 + 2B_{\kappa_n}A^2}} \geq \left( \frac{2A^2}{n - 1 + 1} \right)^{-1},
\]

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

\[
2 \log \left( \frac{2A^2}{n - 1 + 1} \right) \leq \left( \frac{32}{n - 1} \right)^{\frac{1}{2}} A;
\]

finally (168) follows from the definitions of \( \kappa_n, \gamma_n \) in (111), and (128), respectively and the facts that

\[
\lim A \sqrt{A^2 + 2 \log \frac{\gamma_n}{\kappa_n}} \leq A + \log \left( \frac{\gamma_n}{\kappa_n} \right),
\]

\[
C_n(A) \leq \frac{n}{2} \log (1 + A^2) \leq nA.
\]

**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 \( |X| \leq A \) and \( E[X^2] \leq P \), is independent from the noise \( Z \sim \mathcal{N}(0, 1) \). Then, \( P_X \) is the capacity-achieving input distribution if and only if the following conditions are satisfied:

\[
i(x; P_X) = C(A, P) + \lambda(x^2 - P), \quad x \in \text{supp}(P_X),
\]

(177)

\[
i(x; P_X) \leq C(A, P) + \lambda(x^2 - P), \quad x \in [-A, A],
\]

(178)

\[
0 = \lambda(P - E[X^2]),
\]

(179)

where \( C(A, P) \) denotes the capacity of the channel, and \( i(x; P_X) \) is as defined in (27).

**Remark 11.** Hidden in our notation for typographic reasons, the Lagrange multiplier \( \lambda \) in fact depends on amplitude and power constraints, namely \( A \) and \( P \). Indeed, since \( |X| \leq A \), if \( P > A^2 \), 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 \( P_{X^*} \) (i.e., \( x \in \text{supp}(P_{X^*}) \)), then \( x \) is a zero of the function
\[
\Xi_{A,P}(x; P_{X^*}) = i(x; P_{X^*}) - C(A,P) - \lambda(x^2 - P).
\] (180)

In other words,
\[
|\text{supp}(P_{X^*})| \leq N([-A,A], \Xi_{A,P}(\cdot; P_{X^*})).
\] (181)
\[
\leq N(\mathbb{R}, \Xi_{A,P}(\cdot; P_{X^*})).
\] (182)

Observe that, since
\[
x^2 = \int_{\mathbb{R}} \frac{y^2 - 1}{\sqrt{2\pi}} e^{-\frac{(y-x)^2}{2}} dy,
\] (183)
we can write
\[
\Xi_{A,P}(x; P_{X^*}) = \int_{\mathbb{R}} \xi_{A,P}(y) e^{-\frac{(y-x)^2}{2}} dy,
\] (184)
where
\[
\xi_{A,P}(y) = \log \frac{1}{f_Y(y)} - h(Z) - C(A,P) + \lambda P - \lambda(y^2 - 1).
\] (185)

Keeping the steps (39)–(43) in mind, define
\[
\tilde{\delta}_{A,P}(y) = e^{\lambda y^2} f_Y(y) - \kappa_{A,P},
\] (186)
with
\[
\kappa_{A,P} = \exp(-h(Z) - C(A,P) + \lambda(P + 1)).
\] (187)

Using the fact that the Gaussian distribution is a strictly totally positive kernel, and resuming from (182)
\[
|\text{supp}(P_{X^*})| \leq N(\mathbb{R}, \Xi_{A,P}(\cdot; P_{X^*})),
\] (188)
\[
\leq N(\mathbb{R}, \xi_{A,P}),
\] (189)
\[
= N(\mathbb{R}, \tilde{\delta}_{A,P}),
\] (190)
\[
= N([-B_{\kappa_{A,P}}, B_{\kappa_{A,P}}], \tilde{\delta}_{A,P}),
\] (191)
\[
\leq a_P A^2_p + a_P A_P + a_P,
\] (192)
where (189) follows from Karlin’s Oscillation Theorem, see Theorem 4; (190) follows because \( \xi_{A,P}(y) = 0 \) if and only if \( \tilde{\delta}_{A,P}(y) = 0 \); (191) is a consequence Lemma 17 in Appendix D; finally (192) follows from Lemma 21 in Appendix D.

**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 \( \lambda \) in Smith’s result [1, Corollary 2].

**Lemma 16. Bound on the Value of \( \lambda \).** The Lagrange multiplier \( \lambda \) that appears in Lemma 15 satisfies
\[
\lambda \leq \frac{\log(1 + P)}{2P} \cdot 1 \{ P < A^2 \}.
\] (201)

**Proof.** If \( P \geq A^2 \), the power constraint in (4) is not active, implying that the Lagrange multiplier \( \lambda = 0 \). It follows from Lemma 15 that
\[
\lambda P \leq C(A,P) - i(0; P_{X^*}) \leq \frac{1}{2} \log(1 + P),
\] (202)
where (203) is because \( C(A,P) \leq C(\infty,P) = \frac{1}{2} \log(1 + P) \), and \( i(0; P_{X^*}) = D(Z|Y) \geq 0 \).

Similar to its counterpart in Lemma 2, the next lemma provides a bound on the interval for the zeros of the function \( \tilde{\delta}_{A,P} \).

**Lemma 17. Location and Finiteness of Zeros of** \( e^{\lambda y^2} f_Y(y) - \kappa_{A,P} \). For a fixed \( \kappa_{A,P} \in (0, \sqrt{2\pi}) \), there exists some \( B_{\kappa_{A,P}} = B_{\kappa_{A,P}}(A,P) < \infty \) such that
\[
N \left[ \mathbb{R}, e^{\lambda y^2} f_Y(y) - \kappa_{A,P} \right] < \infty.
\] (204)

**Proof.** Using the monotonicity of \( e^{-\nu} \), for \( |y| < A \),
\[
e^{\lambda y^2} f_Y(y)
\] (205)
\[
= \frac{e^{\lambda y^2}}{\sqrt{2\pi}} E \left[ \exp \left( -\frac{(y - X)^2}{2} \right) \right]
\] (206)
\[
\geq \max_{X : |X| \leq A, E[X^2] \leq P} \frac{1}{2} \log \left( \frac{e^{2h(X)}}{2\pi e} + 1 \right).
\] (207)
Lemma 18. For $\kappa_{A,p}$ as defined in (187), the bound on the location of the zeros in Lemma 17 can be loosened as

\[ \frac{1}{2\sqrt{\pi}} \exp \left( -\frac{(y - A)^2}{2} \right) \]

(208)

\[ = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1 - 2\lambda}{2} \left( y - \frac{A}{1 - 2\lambda} \right)^2 + \frac{\lambda A^2}{1 - 2\lambda} \right) . \]

(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

\[ e^{\lambda y^2} f_Y(y) - \kappa_{A,p} < 0 \]

(210)

for all

\[ |y| > \frac{A}{1 - 2\lambda} + \left( \frac{2}{1 - 2\lambda} \log \frac{1 + 1}{\sqrt{2\pi} \kappa_{A,p} \sqrt{2\pi}} + \frac{2\lambda A^2}{(1 - 2\lambda)^2} \right)^{\frac{1}{2}} . \]

(211)

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

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

Lemma 19. Suppose $\tilde{F}_{A,p} : \mathbb{R} \to \mathbb{R}$ is such that $\tilde{F}_{A,p}(y) = e^{\lambda y^2} f_Y(y) - \kappa_{A,p}$. The complex extension of its derivative $\tilde{F}_{A,p} : \mathbb{C} \to \mathbb{C}$ satisfies

\[ \max_{|z| \leq B} |\tilde{F}_{A,p}(z)| \]

\[ \leq \frac{1}{\sqrt{2\pi}} (A + (1 + 2\lambda_p)B) \exp \left( \frac{1 + 2\lambda_p B^2}{2} \right) \]

(220)

\[ < \frac{1}{\sqrt{2\pi}} (A + 2B) \exp (B^2) . \]

(221)

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

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

\[ \max_{|z| \leq B} |\tilde{F}_{A,p}(z)| \]

\[ = \max_{|z| \leq B} \left| e^{\lambda z^2} \left( \tilde{f}_Y(z) + 2\lambda z \tilde{f}_Y(z) \right) \right| \]

(222)

\[ \leq e^{\lambda B^2} \max_{|z| \leq B} \left| \tilde{f}_Y(z) + 2\lambda z \tilde{f}_Y(z) \right| \]

(223)

\[ \leq e^{\lambda B^2} \left[ \max_{|z| \leq B} \left| \tilde{f}_Y(z) \right| + \max_{|z| \leq B} \left| 2\lambda z \tilde{f}_Y(z) \right| \right] \]

(224)

\[ \leq e^{\lambda B^2} \left( \frac{A + B}{\sqrt{2\pi}} \exp \left( \frac{B^2}{2} \right) \right) \]

(225)

\[ + 2AB \max_{|z| \leq B} \left[ \left| \exp \left( -\left( \frac{z - X}{2} \right)^2 \right) \right| \right] \]

(226)

\[ \leq e^{\lambda B^2} \left( \frac{A + B}{\sqrt{2\pi}} \exp \left( \frac{B^2}{2} \right) + 2AB \max_{|z| \leq B} \exp \left( -\frac{3m^2(z)}{2} \right) \right) \]

(227)

\[ \leq \frac{1}{\sqrt{2\pi}} (A + (1 + 2\lambda)B) \exp \left( \frac{1 + 2\lambda B^2}{2} \right) . \]

(228)

where (222) follows from definitions of the functions involved; (223) is because $|z| \leq B$ implies $|e^{\lambda z^2}| \leq e^{\lambda B^2}$; (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 $\lambda \leq \lambda_p < 1/2$, cf. Lemma 16.

Lemma 20. Let $B \geq A/(1 - 2\lambda)$. Suppose $\tilde{F}_{A,p} : \mathbb{R} \to \mathbb{R}$ is such that $\tilde{F}_{A,p}(y) = e^{\lambda y^2} \tilde{f}_Y(y) - \kappa_{A,p}$. The complex extension of its derivative $\tilde{F}_{A,p} : \mathbb{C} \to \mathbb{C}$ satisfies

\[ \max_{|z| \leq B} |\tilde{F}_{A,p}(z)| \]

\[ \geq \frac{A}{\sqrt{2\pi}} \exp \left( -\frac{2 - \lambda_p}{1 - 2\lambda_p} A^2 \right) , \]

(231)

where $\lambda_p = \frac{\log(1 + p)}{2p} \cdot 1\{p < A^2\}$.

Proof. Note that
\[
N([-B_{\kappa_A}, B_{\kappa_A}], \tilde{\gamma}_{A,p}) \\
\leq 1 + N([-B_{\kappa_A}, B_{\kappa_A}], \tilde{\gamma}_{A,p}') \\
\leq 1 + N(D_{B_{\kappa_A}, \delta_{A,p}}) \\
\leq 1 + \min_{s > 1, t \geq \pi_{\kappa_A}, P} \left\{ \frac{1}{\log s} \left( \log \max_{|z| \leq (s + t)/B_{\kappa_A, P}} \left| \tilde{\gamma}_{A,p} - \log \max_{|z| \leq B_{\kappa_A, P}} \left| \tilde{\gamma}_{A,p}' \right| \right) \right\} \\
\leq 1 + \min_{s > 1, t \geq \pi_{\kappa_A}, P} \left\{ \frac{1}{\log s} \left( \frac{(s + t)^2 B_{\kappa_A, P}^2}{2(1 + 2\lambda P)^2} + \frac{2 - \lambda P}{1 - 2\lambda P} A^2 + \log \left( \frac{2}{1 - 2\lambda P} + \frac{(A_P + B_{\kappa_A, P})s}{A/(1 + 2\lambda P)} \right) \right) \right\} \\
= 1 + \min_{s > 1} \left\{ \frac{1}{\log s} \left( \frac{(s + 1)s + A^2}{2(1 + 2\lambda P)} + \frac{2 - \lambda P}{1 - 2\lambda P} A^2 + \log \left( \frac{2}{1 - 2\lambda P} + \frac{(s + 1)s}{A/(1 + 2\lambda P)} \right) \right) \right\} \\
\leq 1 + \min_{s > 1} \left\{ \frac{1}{\log s} \left( \frac{(s + 1)s + A^2}{2(1 + 2\lambda P)} + \frac{2 - \lambda P}{1 - 2\lambda P} A^2 + \log \left( \frac{2}{1 - 2\lambda P} + \frac{(s + 1)e}{A/(1 + 2\lambda P)} \right) \right) \right\} \\
\leq 1 + \log s \left( \frac{(s + 1)s + A^2}{2(1 + 2\lambda P)} + \frac{2 - \lambda P}{1 - 2\lambda P} A^2 + \log \left( \frac{2}{1 - 2\lambda P} + \frac{(s + 1)e}{A/(1 + 2\lambda P)} \right) \right), \\
\right.
\]
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.
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.