

On Additive Channels with Generalized Gaussian Noise

Alex Dytso*, Ronit Bustin**, H. Vincent Poor*, and Shlomo Shamai (Shitz)**

Abstract—This paper considers a problem of communication over an additive noise channel where the noise is distributed according to a Generalized Gaussian (GG) distribution.

In the first part of the paper, a number of properties of the family of GG distributions are derived which are of independent interest. For example, considerable attention is given to the properties of the characteristic function of the GG distribution.

In the second part of the paper, the capacity of an additive noise channel with GG noise is considered under p -th absolute moment constraints. It is shown that, even though Shannon's upper bound is achievable in some instances, in general such achievability is not possible. Moreover, it is shown that discrete inputs can achieve capacity within a constant gap or full degree of freedom for any p -th absolute moment constraint. Following the seminal work of Smith, the paper also gives a condition under which discrete inputs are exactly optimal.

I. INTRODUCTION

It is well known that the traditional assumption of Gaussian noise does not capture scenarios arising in modern applications. The goal of this work is to study a large family of probability distributions termed generalized Gaussian (GG) that have received attention in many communication settings. We shall refer to X with the GG distribution defined by the probability density function (pdf)

$$f_X(x) = \frac{c_p}{\alpha} e^{-\frac{|x-\mu|^p}{2\alpha^p}}, c_p = \frac{p}{2^{\frac{p+1}{p}} \Gamma\left(\frac{1}{p}\right)}, x \in \mathbb{R}, \quad (1)$$

as $X \sim \mathcal{N}_p(\mu, \alpha^p)$. Another commonly used name for this type of distribution, especially in economics, is the Generalized Error distribution. The flexible parametric form of the pdf of GG distributions allows for tails that are either heavier than Gaussian ($p < 2$) or lighter than Gaussian ($p > 2$) which makes it an excellent choice for many modeling scenarios. The origin of the GG family can be traced to the seminal work of Levy [1]. The best-known cases of this family of distributions include: the Laplace distribution for $p = 1$; the Gaussian distribution for $p = 2$; and the uniform distribution on $[-\beta, \beta]$ for $p = \infty$ and $\alpha = \lim_{p \rightarrow \infty} \left(\frac{1}{2}\right)^{\frac{1}{p}} \beta$.

*A. Dytso and H.V. Poor are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA (email: adytso, poor@princeton.edu).

**R. Bustin and S. Shamai (Shitz) are with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: bustin@technion.ac.il, sshlomo@ee.technion.ac.il).

The work of A. Dytso and H.V. Poor was supported by the National Science Foundation under Grants CCF-1420575 and CNS-1456793. The work of S. Shamai and R. Bustin was supported by the Unions Horizon 2020 Research and Innovation Programme Grant 694630. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the funding agencies.

Past Work. In communication theory, the GG distribution finds many modeling applications in *impulsive noise channels* which occur when the noise pdf has a longer tail than the Gaussian pdf. For example, in [2] it is shown that in ultrawideband systems with time-hopping (TH-UWB) the interference should be modeled with probability distributions that are longer-tailed than Gaussian. Moreover, it has been shown that for moderate and high signal-to-noise ratio (SNR) the TH-UWB interference is well modeled by the GG distribution with parameter $p \leq 1$. In [3] and [4] many atmospheric noises were shown to be impulsive and GG distributions with parameter values of $0.1 < p < 0.6$ were shown to provide good approximations to their distributions. In [5] GG has been recognized as a model for the underwater acoustic channel where values of $p = 2.2$ and $p = 1.6$ have been found to model ship transit noise and sea surface agitation noise, respectively. In [6] the GG distribution has been shown to model RF signals in medical applications that are involved in recovering echocardiogram images.

Other application of the GG distribution include modeling of Gabor coefficients in face-recognition problems [7]; load demand in power systems [8]; pixels forming fine-resolution synthetic aperture radar (SAR) images [9]; and texture retrieval problems [10].

As the pdf of the GG family has a very simple form, many quantities such as moments, entropy and Rényi entropy can be easily computed; see for example [10] and [11].

From the information theoretic perspective the GG distribution is interesting because it maximizes the entropy and Rényi entropy under a p -th absolute moment constraint [12], [13]. The maximum entropy property can serve as an important intermediate step in a number of proofs. For example, in [14] it has been used to generalize the Ozarow-Wyner bound [15] on the mutual information of discrete inputs over arbitrary channels.

While the number of applications of the GG distribution is large, many of its properties have been drawn from numerical studies, and few analytical properties of the GG family are known beyond the cases $p = 1, 2$ and $p = \infty$. For instance, very little is known about the characteristic function of the GG distribution and only expressions in terms of hypergeometric functions are known [16].

Paper Outline and Main Contributions. Our contributions are as follows: In Section II, Theorem 1 connects the pdf of GG distributions to *positive definite* functions. In particular, we show that for $p \leq 2$ the pdf of the GG distribution is a positive definite function and for $p > 2$ the pdf is not a positive definite function. Moreover, it is shown that for $p \leq 2$ the pdf of the GG distribution can be expressed as an integral of a Gaussian pdf with respect

to some non-negative finite Borel measure. In Section III, Proposition 1 gives inequalities and properties of moments and absolute moments of the GG distribution. In Section IV we study properties of the characteristic function of the GG distribution: in Proposition 2 we show conditions under which the characteristic function of the GG distribution is a real analytic function; in Theorem 3 we study the distribution of zeros of the characteristic function of the GG distribution. In particular, it is shown that for $p \leq 2$ the characteristic function of the GG distribution has no zeros and is always positive, and for $p > 2$ the characteristic function has at least one positive-to-negative zero crossing. In Theorem 4 we study the conditions under which a GG distribution of order p can be additively transformed into another GG distribution of order q . In Section V we study properties of information measures of the GG distribution: in Theorem 5 and Theorem 6 we review properties of the differential entropy of the GG distribution; in Proposition 5 we derive the derivative of the mutual information with respect to SNR of an additive channel with GG distributed noise; and in Proposition 6 some properties of the relative entropy of GG distribution are considered. In Section VI we study properties of an additive channel with GG distributed noise of order p under a q -th moment constraint: in Proposition 8 we show that Shannon's upper bound, besides being achievable for $(p, q) = (2, 2)$ (i.e., Gaussian noise and the second moment constraint), is also achievable for $(p, q) = (1, 1)$ (i.e., Laplace noise and the absolute first moment constraint); and in Proposition 9, using the deconvolution results from Theorem 4, it is shown that in general Shannon's upper bound is not achievable; in Proposition 10, it is shown that for any (p, q) Shannon's upper bound is achievable within a constant gap by a uniformly spaced equally distributed discrete input, implying that discrete inputs achieve the full *degrees of freedom*; in Proposition 11 we show that the GG distributed input can also be optimal within an additive gap; in Proposition 13 we show that when p and q are even integers then the optimal input distribution is discrete with finitely many points; and in Proposition 14 we study the case of $p = \infty$ (amplitude constraint) and give a condition under which the optimal input distribution is binary.

Due to space limitations, the proofs are omitted and can be found in an extended version of this paper [17].

II. RELATION TO POSITIVE DEFINITE FUNCTIONS

As will be observed throughout this paper the GG family exhibits different properties depending whether $p \leq 2$ or $p > 2$. At the heart of this behavior is the concept of positive-definite functions.

Definition 1. A function $f : \mathbb{R} \rightarrow \mathbb{C}$ is called positive definite if for all real numbers x_1, x_2, \dots, x_n the $n \times n$ matrix

$$A = (a_{i,j})_{i,j=1}^n, \quad a_{i,j} = f(x_i - x_j),$$

is positive semi-definite.

Our first result relates the pdf of the GG distribution to the class of positive definite functions.

Theorem 1. The function $e^{-\frac{|x|^p}{2}}$ is

- not positive definite for $p > 2$; and

- positive definite for $0 < p \leq 2$. Moreover, for all $x > 0$

$$e^{-\frac{x^p}{2}} = \int_0^\infty e^{-\frac{tx^2}{2}} d\mu(t), \quad (2)$$

where $\mu(x)$ is finite non-negative Borel measure on \mathbb{R}^+ .

The concept of positive-definite functions will also play an important role in examining properties of the characteristic function of the GG distribution.

III. MOMENTS

The moments of the GG distribution are given as follows:

Proposition 1. (Moments [11].) For any $p > 0$ and $k > -1$ the moments of $X \sim \mathcal{N}_p(0, \alpha^p)$ are given by

$$\mathbb{E}[X^k] = \frac{2^{\frac{k}{p}} \alpha^k}{\Gamma\left(\frac{1}{p}\right)} \Gamma\left(\frac{k+1}{p}\right), \quad k = \text{even},$$

$$\mathbb{E}[X^k] = 0, \quad k = \text{odd}.$$

Moreover, for any $p > 0$ and $k > -1$ the absolute moments are given by

$$\mathbb{E}[|X|^k] = \frac{2^{\frac{k}{p}} \alpha^k}{\Gamma\left(\frac{1}{p}\right)} \Gamma\left(\frac{k+1}{p}\right). \quad (3)$$

Note that the p -th absolute moment of $X \sim \mathcal{N}_p(0, \alpha^p)$ is given by $\mathbb{E}[|X|^p] = \frac{2\alpha^p}{p}$.

The following corollary, which relates k -th moments of two GG distributions of different orders, is useful in many proofs.

Corollary 1. Let $Y \sim \mathcal{N}_q(0, 1)$ and $Z \sim \mathcal{N}_p(0, 1)$; then for $q \geq p > 0$

$$\mathbb{E}[|Y|^k] \leq \mathbb{E}[|Z|^k], \quad k \in \mathbb{R}^+.$$

Moreover, for $q > p$, $\lim_{k \rightarrow \infty} \left(\frac{\mathbb{E}[|Z|^k]}{\mathbb{E}[|Y|^k]}\right)^{\frac{1}{k}} = \infty$.

IV. THE CHARACTERISTIC FUNCTION

The focus of this section is the characteristic function of the GG distribution. This analysis will play an important role in analyzing the output distribution of additive channels with GG distributed noise.

Theorem 2. For $p > 0$ the characteristic function of $X \sim \mathcal{N}_p(0, \alpha^p)$ is given by

$$\psi_p(t) = 2c_p \int_0^\infty \cos(t\alpha x) e^{-\frac{x^p}{2}} dx.$$

Examples of the characteristic function of $X \sim \mathcal{N}_p(0, 1)$ for several values of p are given in Fig. 1.

a) *Analyticity Properties of the Characteristic Function:* An important question, in particular for numerical methods, is whether the characteristic function of a random variable can be represented by a power series of the form

$$\sum_{k=0}^{\infty} \frac{(it)^k}{k!} \mathbb{E}[X^k],$$

i.e., is real analytic on some domain. The above expression is especially useful since the moments of the GG distribution are known for every k ; see Proposition 1.

Proposition 2. $\psi_p(t)$ is a real analytic function for

- $t \in \mathbb{R}$ for $p > 1$;

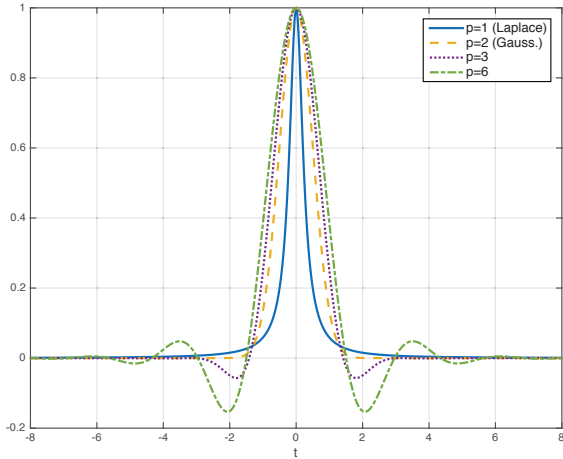


Fig. 1: Plot of the characteristic function of $X \sim \mathcal{N}_p(0, \alpha^p = 2)$ for several values of p .

- $|t| < \frac{1}{2}$ for $p = 1$; and
- $t = 0$ for $p < 1$.

The above results also imply that for $p > 1$ the moment generating function of $X \sim \mathcal{N}_p(0, 1)$ exists for all t .

b) On the Distribution of Zeros of the Characteristic Function: As can be seen from Fig. 1 the characteristic function of the GG distribution can have zeros. The following theorem gives a somewhat surprising result on the distribution of zeros of $\psi_{X,p}(t)$.

Theorem 3. (Distribution of Zeros.) *The characteristic function $\psi_{X,p}(t)$ has the following properties:*

- for $p > 2$, $\psi_p(t)$ has at least one positive to negative zero crossing; and
- for $0 \leq p \leq 2$, $\psi_p(t)$ is non-negative and

$$\psi_p(t) = c_p \int_0^\infty \sqrt{\frac{2\pi}{v}} e^{-\frac{t^2}{2v}} d\mu(v), \quad (4)$$

for some non-negative finite Borel measure μ .

The following result is immediate from Theorem 3.

Corollary 2. *For $0 < p \leq 2$, $\psi_p(t)$ is a strictly decreasing function for $t > 0$.*

c) Deconvolution Results: Next we seek to understand when

$$\psi_{(p,q,\alpha)}(t) = \frac{\psi_q(\alpha \cdot t)}{\psi_p(t)}, \quad (5)$$

is again a characteristic function for all $\alpha \geq 1$. Such a question is important since often we are interested in the existence of random variable X independent of Z such that

$$\alpha Y = X + Z, \quad (6)$$

where Y is the target probability distribution. Here we are interested in answering whether it is possible to transform $Z \sim \mathcal{N}_p(0, 1)$ into αY where $Y \sim \mathcal{N}_q(0, 1)$.

We first focus on the cases when $\psi_{(p,q,\alpha)}(t)$ is not a characteristic function.

Theorem 4. *For $(p, q) \in \mathbb{R}_+^2$ let*

$$\begin{aligned} \mathbb{S} = & \{(p, q) : q \in (0, 2), p > 0, p \neq q\} \\ & \cup \{(p, q) : q \geq 2, 0 < p < q\}. \end{aligned} \quad (7)$$

Then the function $\psi_{(p,q,\alpha)}(t)$ satisfies

- for $(p, q) \in \mathbb{S}$ it is not a characteristic function for $\alpha \geq 1$; and
- for $\{(p, q) : 2 < q \leq p\}$ there exists an $\alpha \geq 1$ such that $\psi_{(p,q,\alpha)}(t)$ is not a characteristic function.

We would like to point out that for $2 < q \leq p$ there are cases when $\psi_{(p,q,\alpha)}(t)$ is a characteristic function. The most trivial case is $p = q$ and $\alpha = 1$ in which $\psi_{(p,q,\alpha)}(t) = 1$, which is a trivial characteristic function. A less trivial example is when $p = q = \infty$ in which case $\psi_{X,\infty}(t) = \text{sinc}(t)$ and

$$\psi_{(\infty,\infty,\alpha)}(t) = \frac{\text{sinc}(\alpha t)}{\text{sinc}(t)}.$$

For example, when $\alpha = 2$ we have that $\psi_{(\infty,\infty,\alpha)}(t) = \frac{1}{2} \cos(2t)$ which corresponds to the characteristic function of random variable ± 1 equally likely.

In the above example since zeros of $\psi_p(t)$ occur periodically we can select an α such that for $\psi_{(p,q,\alpha)}(t)$ the poles and the zeros cancel out. However, we conjecture that such examples are only possible for $p = \infty$, and for $2 < p < \infty$ zeros of $\psi_p(t)$ do not appear periodically and the second statement of Theorem 4 can be improved to the following.

Conjecture 1. *For $2 < q \leq p < \infty$, $\psi_{(p,q,\alpha)}(t)$ is not a characteristic function for any $\alpha > 1$.*

It is not difficult to check, by using the property that convolution with an analytic function is again analytic, that Conjecture 1 is true if p is an even integer and q is any non-even real number.

One of the consequence of Theorem 4 is that the additive transformation from one GG distribution to another GG distribution is not always possible.

Proposition 3. *Let $Z \sim \mathcal{N}_p(0, 1)$. Then we have*

- for $(p, q) \in \mathbb{S}$ and any $\alpha \geq 1$ there exists no random variable X independent of Z such that

$$Y = X + Z,$$

where $Y \sim \mathcal{N}_q(0, \alpha^q)$; and

- for any $\{(p, q) : 2 < q \leq p\}$ there exists some $\alpha \geq 1$ such that there is no random variable X independent of Z such that

$$Y = X + Z,$$

where $Y \sim \mathcal{N}_q(0, \alpha^q)$.

In view of Theorem 4 it remains to answer what happens to $\psi_{(p,q,\alpha)}(t)$ when $0 < q = p < 2$.

Proposition 4. *For $\{(p, q) : 0 < q = p \leq 2\}$, if $\psi_{(p,q,\alpha)}(t)$ is monotonically decreasing for $\alpha > 1$ then it is a characteristic function.*

For example, in the case of $p = 1$ (Laplace) we have that

$$\psi_{(1,1,\alpha)}(t) = \frac{1 + 4t^2}{1 + 4\alpha^2 t^2},$$

which is a monotonically decreasing function for $\alpha > 1$ and corresponds to a random variable X with the pdf given by $f_X(x) = \frac{1}{\alpha^2} \delta(x) + (1 - \frac{1}{\alpha^2}) \frac{c_1}{\alpha} e^{-\frac{|x|}{\alpha}}$.

At this point we are unable to show that $\psi_{(p,q,\alpha)}(t)$ is monotonically decreasing for $\{(p, q) : 0 < q = p < 2\}$ even though the numerical simulations seem to suggest so.

V. INFORMATION MEASURES

a) *Entropy and Conditional Entropy*: From an information theoretic point of view the class of GG distributions is interesting since it maximizes the entropy subjects to an absolute p -th moment constraint.

Theorem 5. (Maximum Entropy Distribution [12].) Let $X \in \mathbb{R}$ be such that $\mathbb{E}[|X|^p] \leq \frac{2\alpha^p}{p}$. Then

$$h(X) \leq \frac{1}{p} \ln \left(\frac{pe}{2c_p^p} \cdot \mathbb{E}[|X|^p] \right) \leq \frac{1}{p} \ln \left(\frac{\alpha^p e}{c_p^p} \right). \quad (8)$$

The equality in (8) is attained iff $X \sim \mathcal{N}_p(0, \alpha^p)$.

The conditional version of Theorem 5 was show in [14] and is given below.

Theorem 6. For any U such that $h(U) < \infty$ and $\mathbb{E}[|U|^p] < \infty$ for some $p \in (0, \infty)$, and for any V , we have

$$h(U|V) \leq \frac{1}{p} \log \left(\frac{pe}{2c_p^p} \cdot \text{mmpe}(U|V; p) \right),$$

$$\text{mmpe}(U|V; p) = \inf_f \mathbb{E}[|U - f(V)|^p],$$

where $f(\cdot)$ is an arbitrary measurable function.

The inequality in Theorem 6 results in a sharper version of the continuous version of Fano's inequality [12] given by

$$h(U|V) \leq \frac{1}{2} \log(2\pi e \cdot \text{mmse}(U|V)),$$

where $\text{mmse}(U|V) = \mathbb{E}[(U - \mathbb{E}[U|V])^2]$.

Theorem 6 also plays a key role in the improvement of the Ozarow-Wyner bound that lower bounds the mutual information of discrete inputs over arbitrary channels [14].

b) *Mutual Information and Conditional Mean Estimation*: The I-MMSE relationship [18] has found many applications in the context of Gaussian noise channels. Next, we give a similar expression for channels with GG distributed noise.

Proposition 5. For any $p \geq 1$ and any $X \in L^p$ independent of $Z \sim \mathcal{N}_p(0, 1)$, and $Y = \sqrt{\text{snr}}X + Z$, we have

$$\frac{d}{d\text{snr}} I(X; \sqrt{\text{snr}}X + Z) = M(X, \text{snr}, p),$$

where

$$M(X, \text{snr}, p) = \frac{p}{4\sqrt{\text{snr}}} \mathbb{E} \left[X \cdot \mathbb{E} \left[\text{sign}(Z) |Z|^{p-1} |Y \right] \right].$$

Note that for $p = 2$, by using the orthogonality principle and the identity $(X - \mathbb{E}[X|Y]) = -\frac{1}{\sqrt{\text{snr}}}(Z - \mathbb{E}[Z|Y])$, we have

$$M(X, \text{snr}, 2) = \frac{1}{2} \mathbb{E}[(X - \mathbb{E}[X|Y])^2],$$

which recovers the I-MMSE result.

Lemma 1. $M(X, \text{snr}, p) > 0$ for all $p \geq 1$ and therefore $I(X; \sqrt{\text{snr}}X + Z)$ is an increasing function of snr .

c) *Relative Entropy*: Another question that often arises in information theoretic applications is the relative entropy distance between $Y = x + Z$ and $Y = X + Z$ where x is a realization of some random variable X .

Proposition 6. Let $D(x) = D(\mathcal{N}_p(x, 1) \| f_Y)$. Then for any $p > 0$

$$\frac{D(x_1) - D(x_2)}{x_1 - x_2} = \frac{p}{2} \mathbb{E} \left[\mathbb{E} \left[\text{sign}(Z) \cdot |Z|^{p-1} |Y = \bar{Y} \right] \right],$$

where $\bar{Y} \sim f_{\bar{Y}}(y) = \frac{F(y-x_2) - F(y-x_1)}{x_1 - x_2}$ with $F(\cdot)$ is the cumulative distribution function (CDF) of $\mathcal{N}_p(0, 1)$. Moreover,

$$D'(x) = \frac{p}{2} \mathbb{E} \left[\mathbb{E} \left[\text{sign}(Z) \cdot |Z|^{p-1} |Y = \hat{Y} \right] \right],$$

where $\hat{Y} \sim \mathcal{N}_p(x, 1)$.

Proposition 6 gives the local behavior of relative entropy and will be useful later when we demonstrate under which conditions binary communication is optimal in channels with amplitude constraint.

VI. CHANNEL CAPACITY

a) *Shannon's Upper Bound*: Using the maximum entropy principle under a p -th moment constraint we arrive at an upper bound on the channel capacity when the input is subject to a p -th moment constraint.

Proposition 7. (Shannon's Bound.) For any $p, q > 0$ let

$$C(\alpha; p, q) = \sup_{X \in \mathcal{F}_p(\alpha)} I(X; X + Z), \quad (9a)$$

where $Z \sim \mathcal{N}_q(0, 1)$ and $\mathcal{F}_p(\alpha) = \left\{ X : \mathbb{E}[|X|^p] \leq \frac{2\alpha^p}{p} \right\}$. Then the set $\mathcal{F}_p(\alpha)$ is convex and compact, and the supremum in (9a) is achievable for all $p > 0$. Moreover,

$$C(\alpha; p, q) \leq \max_{X \in \mathcal{F}_p(\alpha)} \frac{1}{p} \log(a_{p,q} \cdot \mathbb{E}[|X + Z|^p]), \quad (9b)$$

where $a_{p,q} = \frac{p}{2} \left(\frac{c_q}{c_p} \right)^p e^{\frac{q-p}{q}}$ and the inequality holds iff $Y = X + Z$ is GG distributed of order p .

In the above, the case of $p = \infty$ corresponds to an amplitude constraint; that is $\mathcal{F}_\infty(\alpha) = \{X : |X| \leq \alpha \text{ a.s.}\}$.

The upper bound in (9b) is often referred to as Shannon's upper bound. There are several instances when the upper bound in (9b) is tight, the most famous of which is $(p, q) = (2, 2)$, i.e., Gaussian noise with a second moment constraint. Next, we look at the case of $(p, q) = (1, 1)$; that is Laplace noise and the absolute moment constraint.

Proposition 8. For $(p, q) = (1, 1)$ Shannon's upper bound in (9b) is achievable. Moreover, the pdf of the optimal input distribution is given by $f_X(x) = \frac{1}{\beta^2} \delta(x) + \left(1 - \frac{1}{\beta^2}\right) \frac{c_1}{\beta} e^{-\frac{|x|}{\beta}}$ for some $\beta \in (\alpha, \alpha + 1)$.

The Laplace distribution has many connections to the exponential distribution. The achievability of Shannon's upper bound for channels with exponential noise and the first moment constraint was shown in [19].

However, in general Shannon's bound cannot be achieved and the inequality in (9b) is actually strict.

Proposition 9.

- The inequality in (9b) is strict for all $\alpha > 0$ and $(p, q) \in \mathbb{S}$.
- For every $\{(p, q) : 2 < q \leq p\}$ there exists an $\alpha > 0$ such that the inequality (9b) is strict.

Even though Shannon's bound is not achievable in general, we can show that it is asymptotically tight. Moreover, we show a remarkable fact that equally spaced and uniformly distributed input (PAM) achieves the full degree of freedom for any p, q .

Proposition 10. For any $p, q > 0$

$$\lim_{\alpha \rightarrow \infty} \frac{C(\alpha; p, q)}{\frac{1}{p} \log(a_{p,q} \cdot \mathbb{E}[|X + Z|^p])} = 1,$$

achieved by PAM with $\mathbb{E}[|X|^p] = \frac{2\alpha^p}{p}$ and with the number of points given by $N = \left\lfloor \left(\alpha + \frac{p}{2} \mathbb{E}[|Z|^p] \right)^{\frac{1}{p}} \right\rfloor$.

b) *Achievability with GG Input:* It is also interesting to note that discrete inputs are not the only inputs that can get close to the capacity; as shown next, this can also be done with inputs from the GG family.

Proposition 11. (GG input is not too bad.) For any $p, q > 0$

$$I(X'; X' + Z) \leq C(\alpha; p, q) \leq I(X'; X' + Z) + C_p,$$

where $X' \sim \mathcal{N}_p(0, \alpha^p)$ and

$$C_p = \begin{cases} \frac{1}{p} \log(2), & 0 \leq p \leq 1, \\ \frac{1}{2} \log(2), & p = 2, \\ \log(2), & \text{else} . \end{cases}$$

c) *On Capacity Achieving Distributions:* Next we examine the structure of the optimal input distribution.

Proposition 12. (On the Optimal Input Distribution [20].) For $C(\alpha; p, q)$ the support of the optimal input distribution is

- unbounded for $q < p$; and
- compact for $p > q$.

Proof. The proof follows by verifying [20, Theorem 7]. \square An important question is: When does the optimal input distribution in addition to being compactly supported have finite support? Using the method of Smith [21] we have the following.

Proposition 13. For any even integers p and q such that $p > q$ (including the case of $p = \infty$) the optimal input distribution for $C(\alpha; p, q)$ is discrete with finitely many points.

d) *Optimal Input Distribution in a Small Amplitude Regime:* Next we investigate a case of $C(\alpha; \infty, q)$, i.e., an amplitude constraint. Specifically, we are interested in identifying when binary communication is optimal. In other words, for what values of α is the input $X = \{\pm\alpha\}$ equally likely optimal. The next result gives necessary and sufficient conditions for such optimality.

Proposition 14. For q even, the binary input is optimal in $C(\alpha; \infty, q)$ if and only if $\alpha \leq \bar{\alpha}_q$ where

$$\bar{\alpha}_q = \sup \{ \alpha : \mathbb{E}[\mathbb{E}[Z^{q-1} | Y = \bar{Y}]] \geq 0 \}, \quad (10)$$

where $\bar{Y} \sim f_{\bar{Y}}(y) = \frac{F(y) - F(y - \alpha)}{\alpha}$ and $F(\cdot)$ is the CDF of $\mathcal{N}_q(0, 1)$.

For $p = 2$ the condition in (10) is equivalent to solving

$$\int \tanh(\bar{\alpha}_2 y) (Q(y - \bar{\alpha}_2) - Q(y)) dy = \frac{\bar{\alpha}_2}{2},$$

for which the solution is given by $\bar{\alpha}_2 \approx 1.671$ which recovers the result of [22]. A weaker sufficient condition than that in (10) is given by forcing $\mathbb{E}[Z^{p-1} | Y = y]$ to be non-negative for all $y \in \mathbb{R}$ which is equivalent to

$$\bar{\alpha}_p = \sup \left\{ \alpha : \frac{(y - \alpha)^{p-1}}{(y + \alpha)^{p-1}} e^{-\frac{(y - \alpha)^p + (y + \alpha)^p}{2}} \leq 1, \forall y > 0 \right\}.$$

REFERENCES

- [1] P. Lévy, *Calcul des Probabilités*. Gauthier-Villars, 1925.
- [2] N. C. Beaulieu and D. J. Young, "Designing time-hopping ultrawide bandwidth receivers for multiuser interference environments," *Proc. IEEE*, vol. 97, no. 2, pp. 255–284, 2009.
- [3] V. R. Algazi and R. M. Lerner, "Binary detection in white non-Gaussian noise," *M.I.T. Lincoln Lab*, vol. 18, no. Res. DS-2138, pp. 241–250.
- [4] J. H. Miller and J. B. Thomas, "Detectors for discrete-time signals in non-Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 18, no. 2, pp. 241–250, 1972.
- [5] S. Banerjee and M. Agrawal, "Underwater acoustic noise with generalized Gaussian statistics: Effects on error performance," in *Proc. OCEANS - Bergen, 2013 MTS/IEEE*, June 2013, pp. 1–8.
- [6] O. Bernard, J. D'hooge, and D. Fribouler, "Statistical modeling of the radio-frequency signal in echocardiographic images based on generalized Gaussian distribution," in *Proc. 3rd IEEE Int. Symp. on Biomed. Image.: Nano to Macro, 2006*. IEEE, 2006, pp. 153–156.
- [7] D. Gonzalez-Jimenez, F. Perez-Gonzalez, P. Comesana-Alfaro, L. Perez-Freire, and J. L. Alba-Castro, "Modeling Gabor coefficients via generalized Gaussian distributions for face recognition," in *Proc. IEEE Int. Conf. Image Proc.*, vol. 4, Sept 2007, pp. IV – 485–IV – 488.
- [8] M. O. M. Mahmoud, M. Jaidane-Saidane, J. Souissi, and N. Hizaoui, "Modeling of the load duration curve using the asymmetric generalized Gaussian distribution: Case of the Tunisian power system," in *Proc. Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*. IEEE, 2008, pp. 1–7.
- [9] M. S. Davis, P. Bidigare, and D. Chang, "Statistical modeling and ML parameter estimation of complex SAR imagery," in *Proc. Forty-First Asilomar Conference on Signals, Systems and Computers*, Nov 2007, pp. 500–502.
- [10] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Proc.*, vol. 11, no. 2, pp. 146–158, 2002.
- [11] S. Nadarajah, "A generalized normal distribution," *Journal of Applied Statistics*, vol. 32, no. 7, pp. 685–694, 2005.
- [12] T. Cover and J. Thomas, *Elements of Information Theory: Second Edition*. Wiley, 2006.
- [13] E. Lutwak, D. Yang, and G. Zhang, "Moment-entropy inequalities for a random vector," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1603–1607, April 2007.
- [14] A. Dytso, R. Bustin, D. Tuninetti, N. Devroye, S. Shamai, and H. V. Poor, "On the minimum mean p -th error in Gaussian noise channels and its applications." Submitted to *IEEE Trans. Inf. Theory*, <https://arxiv.org/abs/1607.01461>, 2016.
- [15] L. Ozarow and A. 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.
- [16] T. K. Pogány and S. Nadarajah, "On the characteristic function of the generalized normal distribution," *Comptes Rendus Mathématique*, vol. 348, no. 3, pp. 203–206, 2010.
- [17] A. Dytso, R. Bustin, H. V. Poor, and S. Shamai, "Generalized Gaussian distribution in information theory," www.princeton.edu/~adytso/papers/GGInfoTheory.pdf, 2017.
- [18] 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, April 2005.
- [19] S. Verdú, "The exponential distribution in information theory," *Problemy Peredachi Informatsii*, vol. 32, no. 1, pp. 100–111, 1996.
- [20] J. Fahs and I. Abou-Faycal, "Input constraints and noise density functions: A simple relation for bounded-support and discrete capacity-achieving inputs," *arXiv preprint arXiv:1602.00878*, 2016.
- [21] J. G. Smith, "The information capacity of amplitude-and variance-constrained scalar Gaussian channels," *Information and Control*, vol. 18, no. 3, pp. 203–219, 1971.
- [22] N. Sharma and S. Shamai, "Transition points in the capacity-achieving distribution for the peak-power limited AWGN and free-space optical intensity channels," *Problems of Information Transmission*, vol. 46, no. 4, pp. 283–299, 2010.