ISSN 0032-9460, Problems of Information Transmission, 2010, Vol. 46, No. 4, pp. 283–299. © Pleiades Publishing, Inc., 2010.
Original Russian Text © N. Sharma, S. Shamai (Shitz), 2010, published in Problemy Peredachi Informatsii, 2010, Vol. 46, No. 4, pp. 14–32.

INFORMATION THEORY

Transition Points in the Capacity-Achieving Distribution for the Peak-Power Limited AWGN and Free-Space Optical Intensity Channels[1]

N. Sharma[a] and S. Shamai (Shitz)[b2]

[a] *Tata Institute of Fundamental Research (TIFR), Mumbai, India*
nareshs@ieee.org
[b] *Technion, Haifa, Israel*
sshlomo@ee.technion.ac.il

Received March 10, 2010

Abstract—The capacity-achieving input distribution for many channels like the additive white Gaussian noise (AWGN) channel and the free-space optical intensity (FSOI) channel under the peak-power constraint is discrete with a finite number of mass points. The number of mass points is itself a variable, and figuring it out is a part of the optimization problem. We wish to understand the behavior of the optimal input distribution at the transition points where the number of mass points changes. To this end, we give a new set of necessary and sufficient conditions at the transition points, which offer new insights into the transition and make the computation of the optimal distribution easier. For the real AWGN channel case, we show that for the zero-mean unit-variance Gaussian noise, the peak amplitude A of 1.671 and 2.786 mark the points where the binary and ternary signaling, respectively, are no longer optimal. For the FSOI channel, we give transition points where binary gives way to ternary, and in some cases where ternary gives way to quaternary, in the presence of the peak-power constraint and with or without the average-power constraint.

DOI: 10.1134/S0032946010040022

1. INTRODUCTION

A basic problem in information theory is to find the capacity for a given channel model for different constraints on the input. Perhaps one of the most well studied channel models is the AWGN model given by a discrete memoryless channel

$$Y = X + V, \tag{1}$$

where V is the additive noise, X is the input, Y is the output of the channel, and V and X are independent. Y, X, and V could be real or complex. The noise V is modeled as Gaussian with zero mean and unit variance. The input is peak-power constrained, i.e.,

$$\Pr\{|X| > \mathcal{A}\} = 0. \tag{2}$$

There could be an additional constraint on the second moment of X, but as we shall see later, this constraint is easy to add with the method described in the paper. For the real AWGN case, it was

[1] Supported in part by the Israel Science Foundation and the Technion Fund for Research.
[2] The material in this paper was presented in part at the International Symposium on Information Theory and Applications (ISITA), Auckland, New Zealand, December 2008, and at the IEEE Information Theory Workshop (ITW), Cairo, Egypt, January 2010.

shown in [1] that the capacity-achieving distribution is discrete with a finite number of mass points. An immediate consequence of this is that optimization is needed over finite variables rather than continuous distributions.

how many ?

For the complex AWGN case with a peak-power constraint, it was shown in [2] that the same conclusion holds for the amplitude of the input, and the phase is uniformly distributed in $[0, 2\pi)$. The support of the distribution looks like a collection of concentric circles on the 2-dimensional plane.

The FSOI channel for short range communication is modeled as a discrete memoryless channel with input X and output Y, and

$$Y = X + V, \tag{3}$$

where V is the Gaussian noise with zero mean and unit variance and is independent of the input X. All the quantities are assumed to be real. We shall assume that the input signal X modulates the optical intensity of the emitted light, and the power (optical intensity) is proportional to X [3]. Hence, it is reasonable to assume that X is nonnegative, i.e.,

$$\Pr\{X < 0\} = 0. \tag{4}$$

We assume a constraint on the peak-power of the transmitted signal, i.e.,

$$\Pr\{X > \mathcal{A}\} = 0, \tag{5}$$

and a constraint on the average-power of the transmitted signal, i.e.,

$$\mathbf{E}\{X\} < \mathcal{E}. \tag{6}$$

Note that for the FSOI channel, the average-power constraint manifests as a constraint on the first moment of the input X rather than the commonly used second moment constraint. We denote the ratio between the average-power and peak-power as

$$\alpha \triangleq \frac{\mathcal{E}}{\mathcal{A}}. \tag{7}$$

More details on the above channel model can be found in [4–6].

In the presence of the peak-power constraint, the discreteness of the unique capacity-achieving distribution with a finite number of mass points is an obvious consequence of the analysis in [1] when applied to the above channel, and is also proved in [5]. In the presence of only the average-power constraint, it is not known if the capacity-achieving distribution is discrete.

For the fading channel, it was shown in [7] that with the second moment constraint and no channel state information (CSI) at the transmitter and the receiver, the optimal input distribution is again discrete.

For the noncoherent AWGN channel, it was shown in [8] that the optimal input distribution is discrete with an infinite number of mass points. In [9] it was shown that, with the second moment constraint and the CSI known at the receiver and/or at the transmitter, the optimal distribution is Gaussian.

For the Rician fading channel, with no CSI at the transmitter and the receiver, with peak-power constraint, the optimal distribution is discrete in amplitude and uniform in phase. The conclusion is the same [10] with the second moment constraint and phase noise in the specular component. A family of noise probability density functions for which the optimal input distribution is discrete was given in [11].

It is widely believed that the number of discrete levels should monotonically increase as the peak-power constraint is relaxed. But this is still an open problem. In [12] the conjecture was proved that the discrete distribution is binary at low signal-to-noise ratio (SNR) for the AWGN channel.

Channel with block fading changing correlatively within the block but independently across the block, with the peak-power constraint and/or the average-power (second moment) constraint, and no CSI, is studied in [13, 14]. Bounds on the capacity of fading channels in high SNR regime with the peak-power constraint with no CSI are given in [15].

In this paper, we are interested in understanding the behavior of the capacity-achieving distribution at the point of transition where the number of mass points increase. We propose a new set of necessary and sufficient conditions specifically tuned to the point of transition, which are helpful in understanding the transition behavior and are also useful in computation of the capacity-achieving distribution.

2. REVIEW OF [1]: WHY OPTIMAL DISTRIBUTION IS DISCRETE FOR THE PEAK-POWER LIMITED AWGN CHANNEL

We briefly review the necessary and sufficient conditions given in [1] for the real AWGN channel, since we shall need many of the ideas in his proof. Our goal is to maximize the mutual information between X and Y, denoted by $I(X;Y)$, subject to the constraints on the input. Since

$$\boxed{I(X;Y) = H(Y) - H(V),} \tag{8}$$

where $H(\cdot)$ denotes the differential entropy, the problem reduces to finding the input distribution that maximizes the differential entropy of output Y given by

$$\boxed{H(Y;F) = -\int\limits_{-\mathcal{A}}^{\mathcal{A}}\int\limits_{-\infty}^{\infty} p_V(y-x)\log[p_Y(y;F)]\,dy\,dF(x),} \tag{9}$$

where $\mathcal{A}$ is the peak-power constraint; F is the cumulative distribution function (CDF) of the input, and $F \in \mathcal{F}^{\mathcal{A}}$, the topological space of all CDFs of input X satisfying the peak-power constraint; $p_V(\cdot)$ and $p_Y(\cdot)$ are the probability density functions (PDF) of V and Y, respectively; and

$$\boxed{p_Y(y;F) = \int\limits_{-\mathcal{A}}^{\mathcal{A}} p_V(y-x)\,dF(x).} \quad \therefore H(Y,F) = -\int_{-\infty}^{\infty} p_Y \log p_Y\,dy \tag{10}$$

Note that $\mathcal{F}^{\mathcal{A}}$ is a metric space with the Lévy metric, and is convex and compact (compactness essentially follows from the bounded support). It was shown in [1] that $H(Y;F)\colon \mathcal{F}^{\mathcal{A}} \to \mathbb{R}$ is continuous and strictly concave for the AWGN channel. Concavity follows since for $F_\theta = (1-\theta)F_1 + \theta F_2$ we have

$$p_Y(y;F_\theta) = (1-\theta)p_Y(y;F_1) + \theta p_Y(y;F_2) \tag{11}$$

and

$$\begin{aligned} &H(Y;F_\theta) - (1-\theta)H(Y;F_1) - \theta H(Y;F_2) \\ &\qquad = (1-\theta)D[p_Y(y;F_1)\,\|\,p_Y(y;F_\theta)] + \theta D[p_Y(y;F_2)\,\|\,p_Y(y;F_\theta)] \ge 0, \end{aligned} \tag{12}$$

with equality if and only if $\theta = 0$ or 1, where $D(\cdot\,\|\,\cdot)$ is the relative entropy. The differential entropy $H(Y;F)$ has a weak derivative given by

$$H'_{F_1}(F_2) = \lim_{\theta\downarrow 0}\frac{H(Y;F_\theta) - H(Y;F_1)}{\theta} = \int\limits_{-\mathcal{A}}^{\mathcal{A}} h(x;F_1)\,dF_2(x) - H(Y;F_1), \tag{13}$$

if F^* is $\infty \Rightarrow h(x, F^*)$ has same value on ∞ $x \in [A, A]$ $\Rightarrow$ h has same value over whole complex plane (by Bolzano-Weistrass Thm & Identity Thm for $\mathbb{C}$) $\Rightarrow$ $P_Y(y, F^*)$ is a constant which is not possible

where

$$\boxed{h(x; F_1) = -\int_{-\infty}^{\infty} p_V(y - x) \log[p_Y(y; F_1)]\, dy.} \tag{14}$$

Optimizing solution F^* is unique (due to strict concavity), and using the result in [16], the necessary and sufficient conditions that F^* needs to satisfy to be optimal is

$$H'_{F^*}(F) \le 0, \quad \forall\, F \in \mathcal{F}^{\mathcal{A}}, \tag{15}$$

with equality if F is in the support of F^*. It follows by taking $F(t) = U(t - x)$ (the step function at $x \in [-\mathcal{A}, \mathcal{A}]$) that

$$h(x; F^*) \le H(Y; F^*). \tag{16}$$

If x is in the support of F^*, i.e., x denotes the point of increase of F^*, then

$$h(x; F^*) = H(Y; F^*). \tag{17}$$

Points of increase of F^* are finite. If they are infinite, then $h(x; F^*)$ has the same value for an infinite number of points $x \in [-\mathcal{A}, \mathcal{A}]$, and firstly they have a limit point according to the Bolzano–Weierstrass theorem, and secondly one can show that $h(\cdot)$ has an analytic extension to the complex plane. Contradiction

Both these things imply according to the identity theorem that $h(z; F^*)$ has the same value on the entire complex plane and in particular on the real line. This implies that

$$\underset{(14)}{h(x; F^*)} = \underset{(9)}{H(Y; F^*)}, \quad \forall\, x \in [-\mathcal{A}, \mathcal{A}] \tag{18}$$

This leads to

$$\int_{-\infty}^{\infty} p_V(y - x)\{\log[p_Y(y; F_1)] - H(Y; F^*)\}\, dy = 0, \tag{19}$$

which can be shown to lead to

$$p_Y(y; F_1) = e^{H(Y; F^*)} \tag{20}$$

or $p_Y(y; F_1)$ is constant over the real line, which is not possible.

3. PROBLEM STATEMENT AND OVERVIEW

In this section, we provide the problem statement and briefly give an overview of our proposed solution.

3.1. Problem Statement

In this section, we shall assume that the peak-power constraint is present, and hence the capacity-achieving distribution is discrete.

The CDF of X is determined by the vectors $\boldsymbol{q}$, $\boldsymbol{x}$, and n, where

$$\boldsymbol{x} = [x_1, \ldots, x_n] \tag{21}$$

denotes the location of the mass points,

$$\boldsymbol{q} = [q_1, \ldots, q_n] \tag{22}$$

is a probability vector that denotes the weights associated with the mass points, and n determines the number of mass points. We shall assume without loss of generality that $x_1 < \ldots < x_n$.

For the real AWGN channel, we shall assume that only the peak-power constraint is present, i.e.,

$$0 \le x_j \le \mathcal{A}, \quad j = 1, \dots, n. \tag{23}$$

For the FSOI channel, we shall additionally have a constraint on the first moment, and since for this channel the intensity is modulated, this constraint is also referred to as the average-power constraint and is given by

$$\sum_{j=1}^{n} q_j x_j \le \mathcal{E}. \tag{24}$$

The PDF of Y is given by

$$p_Y(y) = \sum_{j=1}^{n} q_j \varphi(y, x_j), \tag{25}$$

→ mixture

where for the FSOI channel,

$$\varphi(y, x) = \frac{1}{\sqrt{2\pi}} e^{-(y-x)^2/2}, \tag{26}$$

→ why no symm?

and for the real AWGN channel,

→ symmetric }

$$\varphi(y, x) = \frac{1}{2\sqrt{2\pi}} \left[e^{-(y-x)^2/2} + e^{-(y+x)^2/2}\right]. \tag{27}$$

Note that we have implicitly assumed that for the real AWGN channel the input distribution of X is symmetric around the origin, a consequence of the strict concavity of $H(\cdot)$ (see Proposition 1 for a proof). The odd number of mass points are accommodated by having one of the x_i's at the origin. Note that the actual number of mass points for the AWGN channel is either $2n-1$, if there is a mass point at the origin, or $2n$ otherwise.

We wish to maximize the differential entropy of Y, and the optimization problem is stated as

$$H^*(Y) = \max_{(\boldsymbol{q}, \boldsymbol{x}, n)} H(Y; F_n), \tag{28}$$

where

Substitute (10) in (9)

$$H(Y; F_n) = -\mathbf{E}\{\log[p_Y(y; F_n)]\} = -\int_{-\infty}^{\infty} p_Y(y; F_n) \log[p_Y(y; F_n)]\, dy. \tag{29}$$

Let the capacity-achieving distribution be characterized by $\boldsymbol{q}^*$, $\boldsymbol{x}^*$, and n^*; i.e.,

$$(\boldsymbol{q}^*, \boldsymbol{x}^*, n^*) = \arg\max_{(\boldsymbol{q}, \boldsymbol{x}, n)} H(Y; F_n). \tag{30}$$

We will denote the optimal input distribution by $F^*_{n^*}$, and the optimal n-point distribution by F^*_n (note that there is no * in the superscript of n). Clearly,

$$F^*_{n^*} = \arg\max_{n} H(Y; F^*_n). \tag{31}$$

It is well known that the capacity-achieving distribution for both of the chosen channel models is unique [5].

Similar to [1], using the Lagrangian theorem, the conditions to be satisfied by the capacity-achieving distribution for some $\lambda \ge 0$ are

$$h(x; F^*_{n^*}) \le H(Y; F^*_{n^*}) + \lambda(x - \mathcal{E}), \quad \forall\, x \in [0, \mathcal{A}], \tag{32}$$

$$h(x^*_i; F^*_{n^*}) = H(Y; F^*_{n^*}) + \lambda(x^*_i - \mathcal{E}), \tag{33}$$

where

$$h(x;F^*_{n^*}) = -\int\limits_{\infty}^{\infty} \varphi(y-x)\log[p^*_Y(y)]\,dy, \tag{34}$$

$$p^*_Y(y;F^*_{n^*}) = \sum_{j=1}^{n^*} q^*_j \varphi(y-x^*_j). \tag{35}$$

We shall assume that the constraint on the first moment of the input is inactive for the real AWGN channel; i.e., $\lambda = 0$.

Before we get to the proposed solution, it is of interest to know different approaches to the solution one might like to take. It is easily seen that the points of increase (except for $\mathcal{A}$) are local maxima of the function $h(x;F^*_{n^*}) - \lambda(x-\mathcal{E})$, $x \in [0,\mathcal{A}]$, and hence these are critical points, i.e., where

$$\frac{\partial h(w;F^*_{n^*})}{\partial w}\Big|_{w=x_i} - \lambda = 0. \tag{36}$$

If one shows that these points of increase are the *only* local maxima, and gets an estimate on the number of critical points where $h'(x;F^*_{n^*}) = 0$, one could perhaps get an estimate of n^*.

Morse theory [17] studies critical points on a manifold and their relation to the global topology of the manifold. Morse inequalities give bounds on the number of critical points in terms of Betti numbers. It is unclear whether the function is a Morse function to start with, i.e., the critical points are nondegenerate.

It is known by [7] and previous works of these authors that at some transition points there is a bifurcation (perhaps pitch-fork) of a mass point into two. But it is unclear how one could apply the results of catastrophe theory [18] for this function, which seems quite messy.

An approach to the same problem using the Lagrange multipliers was tried in [19]. But the Lagrange multiplier approach does not work well when the number of variables is not fixed, as in the problem at hand.

In the remainder of the paper, we denote the CDF of X by $F_{n,\mathcal{A}}$ and $F_{n,\mathcal{A},\mathcal{E}}$ for the real AWGN and FSOI channels, respectively. The subscripts indicate a number n that determines the number of mass points, the peak-power constraint for both channels, and the constraint on the first moment for the FSOI channel.

3.2. Overview of the Results

For our chosen channel models, our approach can be briefly described as follows. We first claim using Berge's maximum theorem (see [20]) that the capacity-achieving distribution is continuous as a function of the power constraints relaxed. Furthermore, we assume that the conjecture that the number of mass points increases at most by 1 is correct. These considerations, as we shall see, put restrictions on the weights and locations of the mass points and gives us necessary conditions for a new mass point to appear. Sufficient conditions for a new mass point to appear are obtained by concocting a distribution with a higher number of mass points and then finding conditions for which it beats the best distribution with a smaller number of mass points.

4. REAL AWGN CHANNEL

In this section, we look at the transition points for the real AWGN channel.

Proposition 1. *The capacity-achieving distribution is symmetric around the origin.*

Proof. The distribution $F_{n^*,\mathcal{A}}$ characterized by $(\boldsymbol{q}^*, \boldsymbol{x}^*, n^*)$ must be symmetric around the origin. To prove this, we first note that the input distributions $F_{n^*,\mathcal{A}}$ and $\acute{F}_{n^*,\mathcal{A}} = [f(\boldsymbol{q}^*), -f(\boldsymbol{x}^*), n^*]$

Since all x_i are positive

(satisfying the peak-power constraint), where $f([x_1, \ldots, x_n]) = [x_n, \ldots, x_1]$ is a function that flips the vector, give the same differential entropy of Y. It follows from the concavity of the differential entropy that a convex combination of $F_{n^*,\mathcal{A}}$ and $\acute{F}_{n^*,\mathcal{A}}$ must give the differential entropy of Y that is at least the same as the one given by $F_{n^*,\mathcal{A}}$. Now invoking the uniqueness of the capacity-achieving distribution, we conclude that $F_{n^*,\mathcal{A}}$ and $\acute{F}_{n^*,\mathcal{A}}$ must be the same, and hence $F_{n^*,\mathcal{A}}$ must be symmetric around the origin.

We now give the necessary and sufficient conditions that need to be satisfied by the capacity-achieving distribution at the transition point when the number of mass points increases.

4.1. Necessary Conditions at the Transition Points

We need the continuity of $F^*_{n^*,\mathcal{A}}$ as a function of $\mathcal{A}$ to prove the necessary conditions. The proof of continuity follows along similar lines as the proof of Berge's maximum theorem [20], which gives the continuity of the optimal distribution. The conditions for the proof are met by the strict concavity of $H(\cdot)$ and the compactness of the space $\mathcal{F}_\mathcal{A}$ (which follows from the peak-power constraint).

It follows from Berge's maximum theorem that $F^*_{n^*,\mathcal{A}}$ is continuous in the Lévy metric, which is the side length of the largest square that can be inscribed in two CDFs with sides parallel to the coordinate axes. This essentially puts an upper bound to the deviation in q_i^* or x_i^*, and the continuity of $\boldsymbol{q}^*$ and $\boldsymbol{x}^*$ as a function of $\mathcal{A}$ follows immediately.

We cannot claim the same, however, for n^*; i.e., we cannot claim that $n^*(\mathcal{A}+\Delta\mathcal{A}) \le n^*(\mathcal{A})+1$. This is because new mass points with weights close to zero will also be admissible under continuity.

As a consequence of the continuity of the capacity-achieving distribution as a function of $\mathcal{A}$, it follows that if a new mass point appears at say x with weight q at the peak-power constraint of $\mathcal{A}$, then it must have

since F^* is cont wrt A, new mass point will rise gradually. Either there will be a new peak or an existing peak will split into 2.

what about symmetry of F^*

necessary cond^n

$$\boxed{h(x; F^*_{n,\mathcal{A}}) = H(F^*_{n,\mathcal{A}}),} \quad \because (17) \tag{37}$$

$$q = 0, \tag{38}$$

$$\frac{dh(w; F^*_{n,\mathcal{A}})}{dw}\Big|_{w=x} = h'(x; F^*_{n,\mathcal{A}}) = 0. \tag{39}$$

If a mass point at x_i splits into two, then

$$\frac{d^2h(w; F^*_{n,\mathcal{A}})}{dw^2}\Big|_{w=x_i} = h''(x_i; F^*_{n,\mathcal{A}}) = 0. \tag{40}$$

This is because for the peak-power constraint of $\mathcal{A}+\Delta\mathcal{A}$, the new mass points are located at $x_i \pm \Delta x_i$, hence, $h'(x_i \pm \Delta x_i; F^*_{n,\mathcal{A}+\Delta\mathcal{A}}) = 0$, and hence, by taking the limit $\Delta\mathcal{A} \to 0$ and invoking continuity, it follows that $\Delta x_i \to 0$ and $h''(x_i; F^*_{n,\mathcal{A}}) = 0$. Note that $F^*_{n,\mathcal{A}}$ does not have any dependence on w in (39) and (40).

justification for this assumption

4.2. Sufficient Conditions at the Transition Points

Can # peaks decrease as A increases!

We assume that the number of mass points in the capacity-achieving distribution increases monotonically and at most by one as A increases. It follows from this assumption that the following two cases are the only possibilities: either a new mass point appears at the origin or an existing mass point at the origin splits into two. See also [2, 7] for a discussion on different channel models as well. Let us consider these two situations and give sufficient conditions for both cases.

New mass point appears at the origin. Let $\mathcal{A}$ be the point where the distribution goes from $2n$ to $2n+1$ mass points; i.e., for any $\Delta\mathcal{A} > 0$, the optimal distribution at $\mathcal{A}+\Delta\mathcal{A}$ has more is F^* convex wrt A? especially n^*

than $2n$ mass points. We know from the necessary conditions that for the peak-power constraint of $\mathcal{A}$ we have $h(0; F^*_{n,\mathcal{A}}) = H(F^*_{n,\mathcal{A}})$.

For the peak-power constraint of $\mathcal{A} + \Delta\mathcal{A}$, the optimal $2n$-point distribution is denoted by $F^*_{n,\mathcal{A}+\Delta\mathcal{A}}$, and we choose a $(2n+1)$-point distribution denoted by $F_{n+1,\mathcal{A}+\Delta\mathcal{A}}$ as a version of $F^*_{n,\mathcal{A}+\Delta\mathcal{A}}$ scaled down by $(1-\varepsilon)$ along with a mass point at the origin with weight ε. This results in the PDF of Y as

$$p_Y(y; F_{n+1,\mathcal{A}+\Delta\mathcal{A}}) = p_Y(y; F^*_{n,\mathcal{A}+\Delta\mathcal{A}}) + \varepsilon\big[\varphi(y,0) - p_Y(y; F^*_{n,\mathcal{A}+\Delta\mathcal{A}})\big]. \tag{41}$$

Note that it is not claimed that $F^*_{n,\mathcal{A}+\Delta\mathcal{A}}$ is the optimal distribution at $\mathcal{A} + \Delta\mathcal{A}$, and hence there is no * in the superscript of n. It is not difficult to show that

$$H(Y; F_{n+1,\mathcal{A}+\Delta\mathcal{A}}) = H(Y; F^*_{n,\mathcal{A}+\Delta\mathcal{A}}) + \varepsilon\big[h(0; F^*_{n,\mathcal{A}+\Delta\mathcal{A}}) - H(Y; F^*_{n,\mathcal{A}+\Delta\mathcal{A}})\big] + O(\varepsilon^2). \tag{42}$$

Hence, if, for small ε, $\Delta\mathcal{A} > 0$,

new mass point at origin has larger h than all other mass points in $A+\Delta A$

Sufficient condn →

$$\boxed{h(0; F^*_{n,\mathcal{A}+\Delta\mathcal{A}}) > h(x^*_i + \Delta x^*_i; F^*_{n,\mathcal{A}+\Delta\mathcal{A}})} \tag{43}$$

for all i, where $x^*_i + \Delta x^*_i$ are mass point locations for $F^*_{n,\mathcal{A}+\Delta\mathcal{A}}$, then for small enough $\varepsilon > 0$ and using (42) we get

$$H(Y; F_{n+1,\mathcal{A}+\Delta\mathcal{A}}) > H(Y; F^*_{n,\mathcal{A}+\Delta\mathcal{A}}). \tag{44}$$

In other words, we have a $2n+1$-point distribution that beats the optimal $2n$-point distribution, and hence the globally optimal distribution cannot have $2n$ mass points. Thus, (43) are the sufficient conditions.

Mass point splits into 2 at $x = 0$. Assume that the mass point at 0 splits into two at $\mathcal{A}$; then at $\mathcal{A} + \Delta\mathcal{A}$, two new mass points are located at $\pm\Delta x$. The input distribution for $\mathcal{A} + \Delta\mathcal{A}$ with $2n+2$ mass points is denoted by $F_{n+1,\mathcal{A}+\Delta\mathcal{A}}$, and the optimal $(2n-1)$-point distribution (not globally optimal) is denoted by $F^*_{n,\mathcal{A}+\Delta\mathcal{A}}$. We obtain $F_{n+1,\mathcal{A}+\Delta\mathcal{A}}$ from $F^*_{n,\mathcal{A}+\Delta\mathcal{A}}$ by scaling down the weight of the mass point at the origin by $(1-\varepsilon)$ and adding a mass point at Δx with weight ε. This gives the PDF of Y as

$$p_Y(y; F_{n+1,\mathcal{A}+\Delta\mathcal{A}}) = p_Y(y; F^*_{n,\mathcal{A}+\Delta\mathcal{A}}) + \varepsilon[\varphi(y,\Delta x) - \varphi(y,0)]. \tag{45}$$

After discarding the integral of odd functions, we get

$$H(Y; F_{n+1,\mathcal{A}+\Delta\mathcal{A}}) = H(Y; F^*_{n,\mathcal{A}+\Delta\mathcal{A}}) + \varepsilon\Delta x^2 h''(0; F^*_{n,\mathcal{A}+\Delta\mathcal{A}}) + O(\varepsilon^2). \tag{46}$$

Hence, for small $\varepsilon > 0$, if for some $\Delta\mathcal{A} > 0$,

$$h''(0, \acute{F}^*_{n,\mathcal{A}+\Delta\mathcal{A}}) > 0, \tag{47}$$

then the above $2n+2$ mass point distribution beats the optimal $2n-1$-point distribution. Note that since mass point at the origin is part of the odd constellation and is a extremal point, this condition implies that it is a local minima and is no longer a maxima.

4.3. Computation of the Optimal Distribution

We first note that a complete check over $x \in (-\mathcal{A}, \mathcal{A})$ that

$$h(x; F^*_{n^*,\mathcal{A}}) \le H(Y; F^*_{n^*}) \tag{48}$$

is true is unnecessary, as is done in [1], since the symmetry of the capacity-achieving distribution along with conjecture that the number of mass points increases at most by one implies that one needs to keep monitoring the origin as the peak-power constraint is relaxed.

An obvious criticism of the above necessary and sufficient conditions is that they are not the same, and it could be that there is a gap between the two. It turns out for the following examples that the necessary conditions imply the sufficient conditions, and hence there is no gap between the two. A proof that this is true in general is an open problem. The main difficulty in solving this problem is that one needs to get greater understanding of the manifold on which F_n^* lies.

4.4. Binary to Ternary Transition

For $\mathcal{A}$ small, $n = 1$ is optimal [12]. Thus,

$$F_{1,\mathcal{A}}^* = \varphi(y, \mathcal{A}). \tag{49}$$

Let us assume that at $\mathcal{A} + \Delta\mathcal{A}$, $n = 2$ is optimal, a new mass point appears at $x = 0$, and

$$F_2^*(\mathcal{A} + \Delta\mathcal{A}) = \varphi(y, \mathcal{A} + \Delta\mathcal{A}) + \varepsilon[\varphi(y, 0) - \varphi(y, \mathcal{A} + \Delta\mathcal{A})]. \tag{50}$$

Condition (43) can be further simplified to

$$\begin{aligned} &h(0; F_{1,\mathcal{A}+\Delta\mathcal{A}}^*) - h(\mathcal{A} + \Delta\mathcal{A}; F_{1,\mathcal{A}+\Delta\mathcal{A}}^*) \\ &\qquad = h(0; F_{1,\mathcal{A}}^*) - h(\mathcal{A}; F_{1,\mathcal{A}}^*) + \int\limits_{-\infty}^{\infty} R(y, \mathcal{A})\, dy\, \Delta\mathcal{A}^2 + O(\Delta\mathcal{A})^3, \end{aligned} \tag{51}$$

where

$$\begin{aligned} R(y, \mathcal{A}) = [-\varphi(y, \mathcal{A}) + \varphi(y, 0)] \left[\frac{\varphi_2(y, \mathcal{A})}{\varphi(y, \mathcal{A})} - \frac{\varphi_1(y, \mathcal{A})^2}{2\varphi(y, \mathcal{A})^2}\right] \\ - \varphi_2(y, \mathcal{A}) \log[\varphi(y, \mathcal{A})] - \frac{\varphi_1(y, \mathcal{A})^2}{\varphi(y, \mathcal{A})}, \end{aligned} \tag{52}$$

$\varphi_1(y, x) = \partial\varphi(y, x)/\partial x$, and $\varphi_2(y, x) = 2\partial^2\varphi(y, x)/\partial x^2$. One can show that $\int\limits_{-\infty}^{\infty} R(y, \mathcal{A})\, dy > 0$, $\forall\, \mathcal{A} \neq 0$. We omit the proof. Hence, by choosing $\Delta\mathcal{A}$ small enough, the necessary condition implies the sufficient condition.

For $n = 2$, (37) can be written as

$$-\frac{1}{\sqrt{2\pi}} \int\limits_{\infty}^{\infty} \left[e^{-y^2/2} - \frac{e^{-(y-\mathcal{A})^2/2}}{2} - \frac{e^{-(y+\mathcal{A})^2/2}}{2}\right] \log\left(e^{-\mathcal{A}y} + e^{\mathcal{A}y}\right) dy = \frac{\mathcal{A}^2}{2}. \tag{53}$$

Solving this, we get $\mathcal{A} \approx 1.671$, the point where binary signaling is no longer optimal. Note that (38) is obviously true, and (39) is satisfied since (37) implies that $h(w; F_{2,\mathcal{A}}^*)$ achieves the global maximum at $w = 0$ at the transition point.

4.5. Ternary to Quaternary Transition

Let the optimal 3-point distribution at $\mathcal{A}$ be

$$p_Y(y; F_{2,\mathcal{A}}^*) = (1 - q)\varphi(y, \mathcal{A}) + q\varphi(y, 0). \tag{54}$$

The only unknown in the above distribution is q, which satisfies $\partial H(Y; F_{2,\mathcal{A}}^*)/\partial q = 0$, or

$$\int\limits_{-\infty}^{\infty} [\varphi(y, \mathcal{A}) - \varphi(y, 0)] \log[p_Y(y; F_{2,\mathcal{A}}^*)]\, dy = 0. \tag{55}$$

By differentiating the above equation with respect to $\mathcal{A}$ and getting an equation with $dq/d\mathcal{A}$, one can show that q is an increasing function of $\mathcal{A}$. Let $\mathcal{A}$ be the point of transition, and let the best 3-point distribution at $\mathcal{A}+\Delta\mathcal{A}$ be

$$p_Y(y; F^*_{2,\mathcal{A}+\Delta\mathcal{A}}) = (1-q-\Delta q)\varphi(y, \mathcal{A}+\Delta\mathcal{A}) + (q+\Delta q)\varphi(y,0), \tag{56}$$

where $\Delta q > 0$. We consider the following 5-point distribution:

$$p_Y(y; F_{3,\mathcal{A}+\Delta\mathcal{A}}) = p_Y(y; F^*_{2,\mathcal{A}+\Delta\mathcal{A}}) + \varepsilon[\varphi(y,\Delta x) - \varphi(y,0)]. \tag{57}$$

A sufficient condition for $F_{3,\mathcal{A}+\Delta\mathcal{A}}$ to beat $F^*_{2,\mathcal{A}+\Delta\mathcal{A}}$ is given by (47), which simplifies using (40) to

$$h''(0; F^*_{2,\mathcal{A}+\Delta\mathcal{A}}) = \frac{\Delta q}{q}\int\limits_{-\infty}^{\infty}\frac{(y^2-1)\varphi(y,0)\varphi(y,\mathcal{A})}{(1-q)\varphi(y,\mathcal{A})+q\varphi(y,0)}\,dy + O[(\Delta q)^2]. \tag{58}$$

One can show that the first term on the right-hand side is positive, i.e.,

$$\int\limits_{-\infty}^{\infty}\frac{(y^2-1)\varphi(y,0)\varphi(y,\mathcal{A})}{(1-q)\varphi(y,\mathcal{A})+q\varphi(y,0)}\,dy > 0, \quad q\in(0,1). \tag{59}$$

Thus, by choosing Δq small enough, which can be done by making $\Delta\mathcal{A}$ small, the necessary condition implies the sufficient condition. For $n=3$, (37) and (40) can be written as

$$\int\limits_{-\infty}^{\infty}[\varphi(y,0)-\varphi(y,\mathcal{A})]\log[p_Y(y;F^*_{2,\mathcal{A}})]\,dy = 0, \tag{60}$$

$$\int\limits_{-\infty}^{\infty}(y^2-1)\varphi(y,0)\log[p_Y(y;F^*_{2,\mathcal{A}})]\,dy = 0. \tag{61}$$

Solving for q and $\mathcal{A}$ in the above two equations, we get $\mathcal{A}\approx 2.786$, the point where the ternary signaling is no longer optimal.

4.6. Transition Points for Higher Constellations

But this approach becomes increasingly messy as n increases. For example, for a transition from quaternary to a 5-point constellation, the necessary conditions imply the sufficient conditions if the following equation is true for all $0\le a_1, a_2\le 1$, $a_1+a_2=1$:

$$\begin{aligned}\int\limits_{-\infty}^{\infty}\Bigg\{&-[\varphi(y,0)-a_1\varphi(y,x_1)-a_2\varphi(y,\mathcal{A})]\bigg[\frac{q_1\varphi_2(y,x_1)}{q_1\varphi(y,x_1)+q_2\varphi(y,\mathcal{A})}\\
&\qquad -\frac{1}{2}\frac{q_1^2\varphi_1(y,x_1)^2}{(q_1\varphi(y,x_1)+q_1\varphi(y,\mathcal{A}))^2}\bigg]+\frac{a_1\varphi_1(y,x_1)^2q_1}{q_1\varphi(y,x_1)+q_2\varphi(y,\mathcal{A})}\Bigg\}\Delta x_1^2\\
&\qquad +\Bigg\{-\frac{-a_2\varphi_1(y,\mathcal{A})q_1\varphi_1(y,x_1)-a_1\varphi_1(y,x_1)q_2\varphi_1(y,\mathcal{A})}{q_1\varphi(y,x_1)+q_2\varphi(y,\mathcal{A})}\\
&\qquad +\frac{(\varphi(y,0)-a_1\varphi(y,x_1)-a_2\varphi(y,\mathcal{A}))q_2\varphi_1(y,\mathcal{A})q_1\varphi_1(y,x_1)}{(q_1\varphi(y,x_1)+q_2\varphi(y,\mathcal{A}))^2}\Bigg\}\Delta\mathcal{A}\Delta x_1\\
&\qquad +\Bigg\{-[\varphi(y,0)-a_1\varphi(y,x_1)-a_2\varphi(y,\mathcal{A})]\bigg[\frac{q_2\varphi_2(y,\mathcal{A})}{q_1\varphi(y,x_1)+q_2\varphi(y,\mathcal{A})}\\
&\qquad -\frac{1}{2}\frac{q_2^2\varphi_1(y,\mathcal{A})^2}{(q_1\varphi(y,x_1)+q_2\varphi(y,\mathcal{A}))^2}\bigg]+\frac{a_2\varphi_1(y,\mathcal{A})^2q_2}{q_1\varphi(y,x_1)+q_2\varphi(y,\mathcal{A})}\\
&\qquad +a_2\varphi_2(y,\mathcal{A})\log[q_1\varphi(y,x_1)+q_2\varphi(y,\mathcal{A})]\Bigg\}\Delta\mathcal{A}^2\,dy>0,\end{aligned} \tag{62}$$

where $[x_1^*, \mathcal{A}]$, are the mass point locations of the optimal 4-point distribution with weights as $[q_1^*, q_2^*]$ respectively.

5. FREE-SPACE OPTICAL INTENSITY (FSOI) CHANNEL

Before we get to the transition points for the FSOI channel, we first prove some results for the capacity-achieving distribution.

Proposition 2. *There is a mass point at the origin; i.e.,* $x_1^* = 0$.

Proof. Assume that $x_1^* \neq 0$. Then we can define a new vector for mass point locations as

$$\widehat{\boldsymbol{x}}^* = [\widehat{x}_1^*, \ldots, \widehat{x}_n^*], \tag{63}$$

such that $\widehat{x}_i^* = (x_i^* - x_1^*)$, $i = 1, \ldots, n$. The new distribution, characterized by $(\boldsymbol{q}^*, \widehat{\boldsymbol{x}}^*, n^*)$ and denoted by $\acute{F}_{n^*,\mathcal{A},\mathcal{E}}$, satisfies the constraints. Since the differential entropy is unchanged by the translations, we have

$$H(Y; F_{n^*,\mathcal{A},\mathcal{E}}) = H(Y; \acute{F}_{n^*,\mathcal{A},\mathcal{E}}). \tag{64}$$

But since the capacity-achieving distribution is unique, this results in a contradiction, which can be avoided only if $x_1^* = 0$.

Proposition 3. *At least one of the peak or average-power constraints is active.*

Proof. Assume the contrary. Then, similarly to Proposition 2, we shall shift each of the mass points as

$$\widehat{x}_i^* = x_i^* + \alpha, \tag{65}$$

where

$$\alpha = \min\biggl(\mathcal{E} - \sum_{i=1}^{n} q_i x_i^*, \mathcal{A} - \max_i x_i^*\biggr). \tag{66}$$

Let this shifted distribution be denoted by $\acute{F}_{n^*,\mathcal{A},\mathcal{E}}$. We have $H(Y; F_{n^*,\mathcal{A},\mathcal{E}}) = H(Y; \acute{F}_{n^*,\mathcal{A},\mathcal{E}})$. Invoking the uniqueness of the capacity-achieving distribution, we conclude that α must be zero, which is possible only if at least one of the constraints is active.

Proposition 4. *For* $\mathcal{E} \in (\mathcal{A}/2, \mathcal{A}]$, *the average-power constraint is inactive, and the peak-power constraint is active.*

Proof. See [6] for the proof that the average-power constraint is inactive. Invoking Proposition 3, we conclude that the peak-power constraint must be active.

Proposition 5. *For* $\mathcal{E} \in (0, \mathcal{A}/2)$, *the average-power constraint is active, and is inactive for* $\mathcal{E} = \mathcal{A}/2$.

Proof. We say that the average power constraint is inactive if $\lambda = 0$ in (32).

Assume first that the average-power constraint is inactive for $\mathcal{E} \in (0, \mathcal{A}/2)$. Hence, $\lambda = 0$ in (32) and (33) (by invoking the Karush–Kuhn–Tucker necessary conditions [21]), and the conditions coincide with the case considered in [1] with only the peak-power constraint except that the permissible interval for the transmission is $[-\mathcal{A}/2, \mathcal{A}/2]$, and in our problem it is $[0, \mathcal{A}]$.

The distribution $D_1 = (\boldsymbol{q}^*, \boldsymbol{x}^*, n^*)$ must be symmetric around the point $\mathcal{A}/2$. To prove this, we first note that the input distributions D_1 and $D_2 = [f(\boldsymbol{q}^*), \mathcal{A} - f(\boldsymbol{x}^*), n^*]$ give the same differential entropy of Y, where $f([x_1, \ldots, x_n]) = [x_n, \ldots, x_1]$ is a function that flips the vector. Further, D_2 also satisfies the peak-power constraint, though violates the average-power constraint, which as per the assumption is inactive ($\lambda = 0$). Invoking the concavity of the output differential entropy in input distribution, a convex combination of D_1 and D_2 must give the differential entropy of the output Y that is at least the same as the one given by D_1. Now invoking the uniqueness of the capacity-achieving distribution, we conclude that D_1 and D_2 must be the same, and hence D_1 must be symmetric around $\mathcal{A}/2$, which is also its mean. But this is a contradiction, since $\mathcal{E} \in [0, \mathcal{A}/2)$. Hence, we must have the average-power constraint as active.

The case of $\mathcal{E} = \mathcal{A}/2$ is dealt by first assuming $\lambda = 0$ and finding D_1, which we are assured of finding, since there is a one-to-one correspondence with the case considered in [1] for the peak-power constraint of $\mathcal{A}/2$. The distribution D_1 has the average-power of $\mathcal{E} = \mathcal{A}/2$. Now invoking the uniqueness, we conclude that this must be the solution.

In this paper, we only consider the case where the average-power constraint is active, since the other case, as is mentioned above, is easy to consider by the results in [1]. Hence, in what follows, we assume that $\mathcal{E} \in [0, \mathcal{A}/2]$.

5.1. Transition Points in the Capacity-Achieving Distribution

Transition from binary to ternary for $\boldsymbol{\mathcal{E} \in [0, \mathcal{A}/2)}$. Similarly to [1], for small $\mathcal{E}$ or $\mathcal{A}$, we shall assume that the capacity-achieving distribution is binary. Since the peak-power constraint is active, the two mass points are $\boldsymbol{x}^* = [0, \acute{\mathcal{A}}]$ with weights $\boldsymbol{q}^* = [1-\mathcal{E}/\acute{\mathcal{A}}, \mathcal{E}/\acute{\mathcal{A}}]$. We cannot, however, claim at the moment that $\acute{\mathcal{A}} = \mathcal{A}$. Let the distribution characterized by $\boldsymbol{q}^*$ and $\boldsymbol{x}^*$ be denoted by $F^*_{2,\acute{\mathcal{A}},\mathcal{E}}$. The derivative of $H(Y; F^*_{2,\acute{\mathcal{A}},\mathcal{E}})$ with respect to $\acute{\mathcal{A}}$ is given by

$$\frac{\partial H(Y; F^*_{2,\acute{\mathcal{A}}+\Delta\acute{\mathcal{A}},\mathcal{E}})}{\partial \acute{\mathcal{A}}} = -\int\limits_{-\infty}^{\infty} \left[\varphi(y-\acute{\mathcal{A}})(y-\acute{\mathcal{A}}) - \frac{\varphi(y-\acute{\mathcal{A}})-\varphi(y)}{\acute{\mathcal{A}}}\right] \log[p_Y(y)]\, dy. \tag{67}$$

If $\partial H(F^*_{2,\acute{\mathcal{A}},\mathcal{E}})/\partial\acute{\mathcal{A}} > 0$, then, clearly, increasing $\acute{\mathcal{A}}$ will result in an increase in $H(Y)$, since

$$H(Y; F^*_{2,\acute{\mathcal{A}}+\Delta\acute{\mathcal{A}},\mathcal{E}}) = H(Y; F^*_{2,\acute{\mathcal{A}},\mathcal{E}}) + (\Delta\acute{\mathcal{A}})\frac{\partial H(Y; F^*_{2,\acute{\mathcal{A}},\mathcal{E}})}{\partial\acute{\mathcal{A}}} + O\big[(\Delta\acute{\mathcal{A}})^2\big]. \tag{68}$$

Hence, if $\partial H(Y; F^*_{2,\acute{\mathcal{A}},\mathcal{E}})/\partial\acute{\mathcal{A}} > 0$, then $\acute{\mathcal{A}} = \mathcal{A}$.

Figure 1 divides the $(\mathcal{E}, \acute{\mathcal{A}})$ plane into two regions depending on the sign of $\partial H(Y; F^*_{2,\acute{\mathcal{A}},\mathcal{E}})/\partial\acute{\mathcal{A}}$, which is computed using (67).

Arguing similarly as in Section 4.1, we conclude that at the point of transition where binary gives way to ternary, the additional necessary condition apart from (32) and (33) to be satisfied is that for some $x_2^* \in (0, \acute{\mathcal{A}})$,

$$h(x_2^*; F_{2,\acute{\mathcal{A}},\mathcal{E}}) = H(Y; F_{2,\acute{\mathcal{A}},\mathcal{E}}) + \lambda(x_2^* - \mathcal{E}), \tag{69}$$

where $\boldsymbol{q}^* = [1-\mathcal{E}/\acute{\mathcal{A}}, \mathcal{E}/\acute{\mathcal{A}}]$, $\boldsymbol{x}^* = [0, \acute{\mathcal{A}}]$.

To arrive at the sufficiency condition, we look at the distribution $F_{3,q,\acute{\mathcal{A}},\Delta\acute{\mathcal{A}},\mathcal{E}}$ characterized by

$$\boldsymbol{q}^* = \left[1 - \frac{\mathcal{E}}{\acute{\mathcal{A}}+\Delta\acute{\mathcal{A}}} - \left(1 - \frac{x_2^*}{\acute{\mathcal{A}}+\Delta\acute{\mathcal{A}}}\right)q, q, \frac{\mathcal{E}}{\acute{\mathcal{A}}+\Delta\acute{\mathcal{A}}} - \frac{x_2^* q}{\acute{\mathcal{A}}+\Delta\acute{\mathcal{A}}}\right], \tag{70}$$

$$\boldsymbol{x}^* = [0, x_2^*, \acute{\mathcal{A}}+\Delta\acute{\mathcal{A}}], \tag{71}$$

where x_2^* is given by (69). It is easy to check that

$$\left.\frac{\partial H(Y; F_{3,q,\acute{\mathcal{A}},\Delta\acute{\mathcal{A}},\mathcal{E}})}{\partial\acute{\mathcal{A}}}\right|_{\mathcal{E},\acute{\mathcal{A}},\Delta\acute{\mathcal{A}}=0,q=0} = h(x_2) - h(0) + \frac{x_2}{\acute{\mathcal{A}}}[h(0) - h(\acute{\mathcal{A}})] = 0, \tag{72}$$

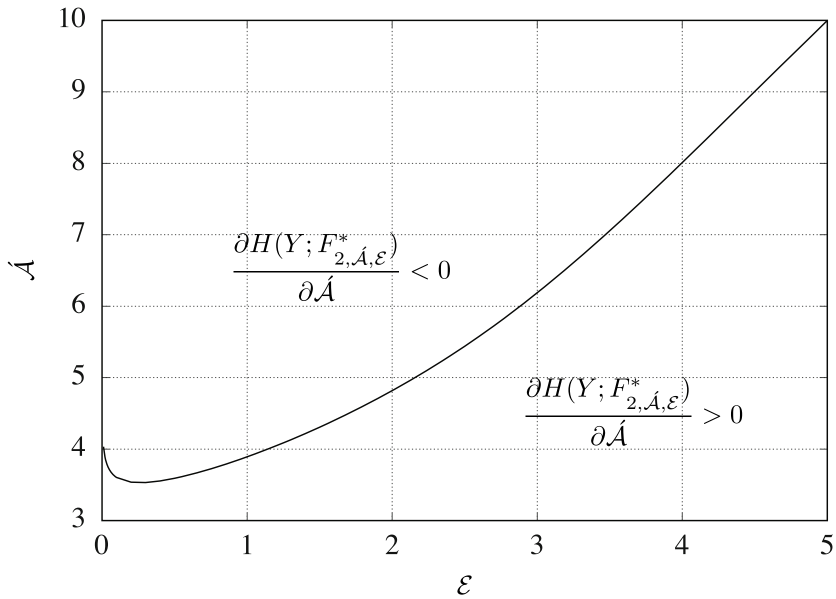

Fig. 1. Sign of $\partial H(Y;F^*_{2,\acute{\mathcal{A}},\mathcal{E}})/\partial\acute{\mathcal{A}}$ divides the $(\mathcal{E},\acute{\mathcal{A}})$ plane into two regions.

$$\frac{\partial H(Y;F_{3,q,\acute{\mathcal{A}},\Delta\acute{\mathcal{A}},\mathcal{E}})}{\partial q}\bigg|_{\mathcal{E},\acute{\mathcal{A}},\Delta\acute{\mathcal{A}}=0,q=0} = 0 \tag{73}$$

and

$$\begin{aligned}\frac{\partial^2 H(Y;F_{3,q,\acute{\mathcal{A}},\Delta\acute{\mathcal{A}},\mathcal{E}})}{\partial q\,\partial\acute{\mathcal{A}}}\bigg|_{\mathcal{E},\acute{\mathcal{A}},\Delta\acute{\mathcal{A}}=0,q=0}&\\ = -\int\limits_{\infty}^{\infty}\frac{\mathcal{E}\left[\dfrac{\varphi(y)-\varphi(y-\acute{\mathcal{A}})}{\acute{\mathcal{A}}}+(y-\acute{\mathcal{A}})\varphi(y-\acute{\mathcal{A}})\right]}{\acute{\mathcal{A}}p_Y^*(y)}&\left[\left(-1+\frac{x_2^*}{\acute{\mathcal{A}}}\right)\varphi(y)\right.\\ \left.+\varphi(y-x)-\frac{x_2^*}{\acute{\mathcal{A}}}\varphi(y-\acute{\mathcal{A}})\right]dy+\frac{x_2^*}{\acute{\mathcal{A}}}&\left[\frac{h(\acute{\mathcal{A}})-h(0)}{\acute{\mathcal{A}}}-\acute{h}(\acute{\mathcal{A}})\right],\end{aligned} \tag{74}$$

where

$$\acute{h}(x;F_{3,q,\acute{\mathcal{A}},\Delta\acute{\mathcal{A}},\mathcal{E}}) = \int\limits_{\infty}^{\infty}(y-x)\varphi(y-x)\log[p_Y^*(y;F_{3,q,\acute{\mathcal{A}},\Delta\acute{\mathcal{A}},\mathcal{E}})]\,dy. \tag{75}$$

Taking the Taylor series expansion, we get, using (72) and (73),

$$\begin{aligned}H(Y;F_{3,q,\acute{\mathcal{A}},\Delta\acute{\mathcal{A}},\mathcal{E}})|_{\mathcal{E},\acute{\mathcal{A}},\Delta\acute{\mathcal{A}},q} &= H(Y;F_{3,q,\acute{\mathcal{A}},\Delta\acute{\mathcal{A}},\mathcal{E}})|_{\mathcal{E},\acute{\mathcal{A}},\Delta\acute{\mathcal{A}}=0,q=0}+O[(\Delta\acute{\mathcal{A}})^2]\\ &+\left[\frac{\partial H(Y;F_{3,q,\acute{\mathcal{A}},\Delta\acute{\mathcal{A}},\mathcal{E}})}{\partial\acute{\mathcal{A}}}\bigg|_{\mathcal{E},\acute{\mathcal{A}},\Delta\acute{\mathcal{A}}=0,q=0}\right.\\ &\left.+\frac{\partial^2 H(Y;F_{3,q,\acute{\mathcal{A}},\Delta\acute{\mathcal{A}},\mathcal{E}})}{\partial q\,\partial\acute{\mathcal{A}}}\bigg|_{\mathcal{E},\acute{\mathcal{A}},\Delta\acute{\mathcal{A}}=0,q=0}q+O(q^2)\right]\Delta\acute{\mathcal{A}}.\end{aligned} \tag{76}$$

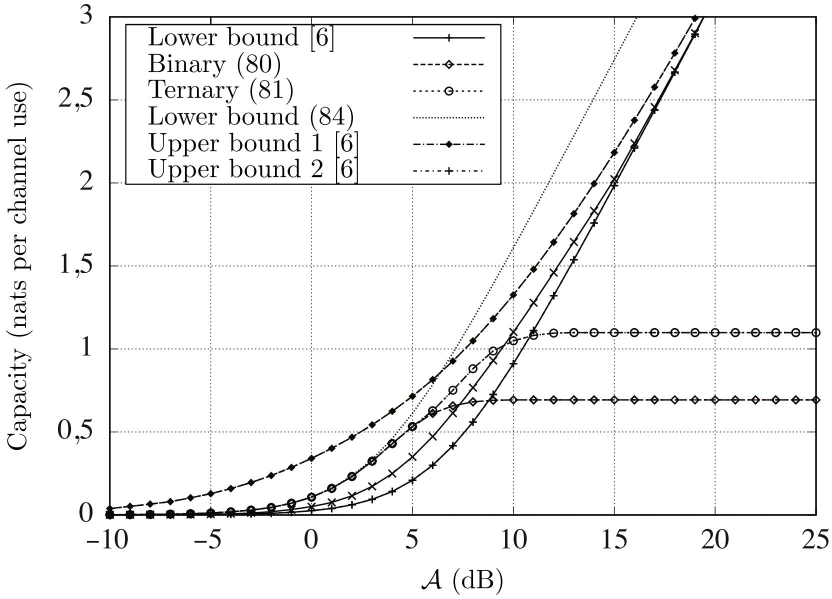

Fig. 2. Bounds on the capacity for $\mathcal{E} = 0.4\mathcal{A}$.

Hence, the sufficient conditions are

$$\left.\frac{\partial H(Y; F_{3,q,\acute{\mathcal{A}},\Delta\acute{\mathcal{A}},\mathcal{E}})}{\partial \acute{\mathcal{A}}}\right|_{\mathcal{E},\acute{\mathcal{A}},\Delta\acute{\mathcal{A}}=0,q=0} \geq 0, \tag{77}$$

$$\left.\frac{\partial^2 H(Y; F_{3,q,\acute{\mathcal{A}},\Delta\acute{\mathcal{A}},\mathcal{E}})}{\partial q\, \partial \acute{\mathcal{A}}}\right|_{\mathcal{E},\acute{\mathcal{A}},\Delta\acute{\mathcal{A}}=0,q=0} > 0. \tag{78}$$

It is not obvious that the necessary condition implies the sufficient conditions. But the numerical simulations have not yielded an exception where there is a gap between the necessary and the sufficient conditions.

Transition points for $\boldsymbol{\mathcal{E} \in [\mathcal{A}/2, \mathcal{A}]}$. Since this case is similar to the one considered in [1] where the channel is AWGN and the input is peak-power constrained as

$$\Pr\{|X| > \mathcal{A}/2\} = 0, \tag{79}$$

using the results in [22], the transition from binary to ternary occurs at $\mathcal{A} \approx 3.342$ and from ternary to quaternary at $\mathcal{A} \approx 5.572$.

5.2. Numerical Results

Figure 2 plots the capacity bounds that are given in [6] as a function of $\mathcal{A}$ with $\mathcal{E} = 0.4\mathcal{A}$. We also plot the mutual information obtained from binary signaling. For the binary case, we assume the distribution as $\boldsymbol{q} = [1-\alpha, \alpha]$ and $\boldsymbol{x} = [0, \acute{\mathcal{A}}]$. For binary signaling, the distribution of the output Y is given by

$$p_Y(y) = (1-q)\varphi(y) + q\varphi(y - \acute{\mathcal{A}}), \tag{80}$$

where $0 < q < 1$ and $q\acute{\mathcal{A}} < \mathcal{E}$. For the ternary case,

$$p_Y(y) = q_1\varphi(y) + q_2\varphi(y - x_2) + (1 - q_1 - q_2)\varphi(y - x_3), \tag{81}$$

where $q_1, q_2, (1-q_1-q_2) \geq 0$, $x_2 < x_3 \leq \mathcal{A}$, and $q_2x_2+(1-q_1-q_2)x_3 \leq \mathcal{E}$. The mutual information is computed as

$$I(X;Y) = -\int_{-\infty}^{\infty} p_Y(y)\log[p_Y(y)]\,dy - \frac{\log(2\pi e)}{2}. \tag{82}$$

The optimization over the parameter(s) $\acute{\mathcal{A}}$ for the binary signaling, and q_1, q_2, x_2, x_3 for the ternary signaling is performed using numerical optimization based on the iterative cutting-plane method outlined in [12]. We also consider the following distribution of X:

$$p_X(x) = \begin{cases} \beta, & 0 \leq x \leq x_1, \\ \gamma, & x_1 < x \leq x_2, \end{cases} \tag{83}$$

where $\beta, \gamma \geq 0$, $\beta x_1 + \gamma(x_2 - x_1) = 1$, and $0 \leq x_1 < x_2 \leq \mathcal{A}$, $\beta x_1^2/2 + \gamma(x_2^2 - x_1^2)/2 \leq \mathcal{E}$. We compute the mutual information $I(X;Y)$ obtained using the input distribution in (83) by maximizing over x_1 and x_2. Values of β and γ are obtained from x_1 and x_2 and used only if $\beta, \gamma \geq 0$. It is not difficult to show that the output distribution is given by

$$p_Y(y) = \beta[Q(y-x_1) - Q(y)] + \gamma[Q(y-x_2) - Q(y-x_1)]. \tag{84}$$

It is interesting to note that the "upper bound 2" in [6] is quite close to the mutual information obtained by binary signaling for low $\mathcal{A}$, where $\acute{\mathcal{A}} = \mathcal{A}$ in (80). The "upper bound 2" in [6] is given by

$$\mathcal{C}_{\mathrm{up}-2}(\mathcal{A}, \alpha\mathcal{A}) = \frac{1}{2}\log\left[1 + \alpha(1-\alpha)\mathcal{A}^2\right], \tag{85}$$

which simplifies for low $\mathcal{A}$ as

$$\mathcal{C}_{\mathrm{up}-2}(\mathcal{A}, \alpha\mathcal{A}) = \frac{\alpha(1-\alpha)}{2}\mathcal{A}^2 + O(\mathcal{A}^4). \tag{86}$$

If the mutual information obtained by binary signaling is denoted by $\mathcal{C}_{\mathrm{bin}}(\mathcal{A}, \alpha\mathcal{A})$, then it is not difficult to show (as also observed in [6]) that

$$\mathcal{C}_{\mathrm{bin}}(\mathcal{A}, \alpha\mathcal{A}) = \frac{\alpha(1-\alpha)}{2}\mathcal{A}^2 + O(\mathcal{A}^4). \tag{87}$$

We note that (84) betters the lower bound given in [6]. Figure 3 plots the results for $\alpha = 0.7$. The average-power constraint is inactive for this value of α. The binary signaling for this case is given by $\boldsymbol{q} = [0.5, 0.5]$ and $\boldsymbol{x} = [0, \mathcal{A}]$, and the associated mutual information is denoted by $\acute{\mathcal{C}}_{\mathrm{bin}}(\mathcal{A}, \alpha\mathcal{A})$.

Next, we plot the capacity bounds for the case where the peak-power constraint is absent in Fig. 4. We also plot the mutual information obtained by the binary and ternary signaling by assuming $\mathcal{A} = 10\mathcal{E}$. This large choice of $\mathcal{A}$ in relation to $\mathcal{E}$ is to ensure that $\mathcal{A}$ is not as restrictive as $\mathcal{E}$ in the choice of input distribution. The plot shows that the discrete distributions give a better lower bound, but it is difficult to get any closed-form expressions.

6. CONCLUSIONS

In conclusion, we characterize the capacity-achieving distribution for the real AWGN channel with the peak-power constraint at the point of transition where the number of mass points increase. The necessary and sufficient conditions seem different as they are presented, but when we look at the transition points for binary to ternary and ternary to quaternary, we find that the necessary conditions imply the sufficient conditions, and hence the necessary conditions are also the sufficient ones. But proving this for transition points for higher constellations is an open problem.

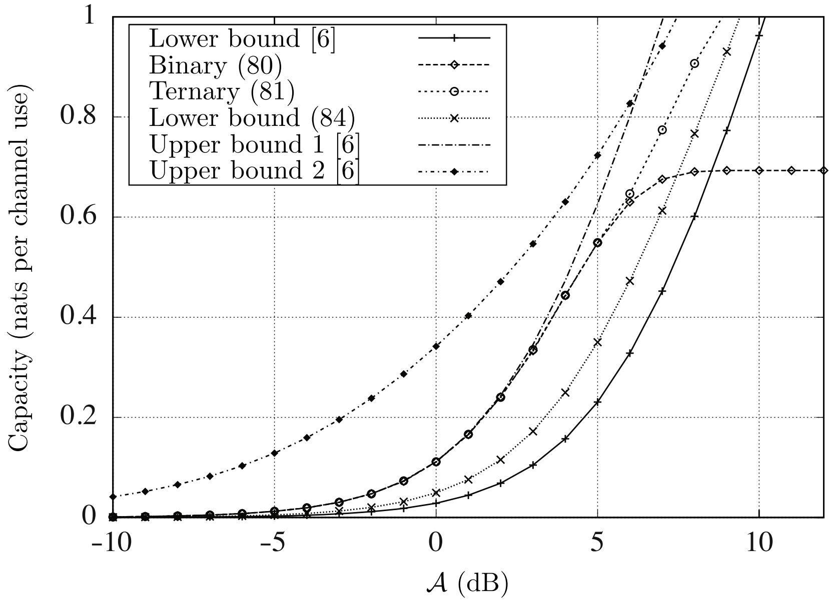

Fig. 3. Bounds on the capacity for $\mathcal{E} \in [\mathcal{A}/2, \mathcal{A}]$.

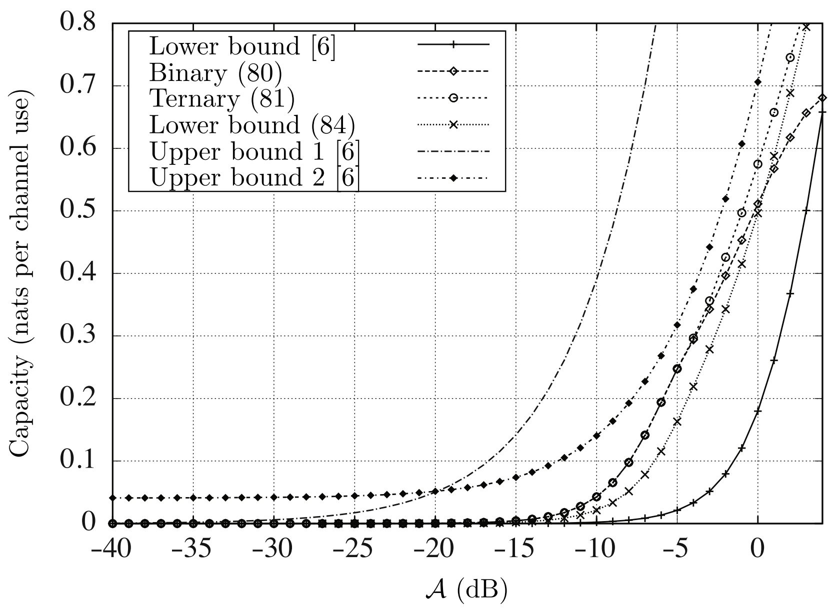

Fig. 4. Bounds on the capacity with only the average-power constraint $\mathcal{E}$.

For the FSOI channel, we gave the transition point where binary gives way to ternary signaling for $\mathcal{E} \in [0, \mathcal{A}/2)$. As an example, for $\mathcal{E} = 1$, $\mathcal{A} = 3.46$ marks the point where binary gives way to ternary. For $\mathcal{E} \in [\mathcal{A}/2, \mathcal{A})$, we use the results from the real AWGN channel to give the transition points where binary gives way to ternary and ternary to quaternary. We also show that the lower bounds on the channel capacity given by the binary signaling are tight for low peak power.

REFERENCES

1. Smith, J.G., The Information Capacity of Amplitude- and Variance-Constrained Scalar Gaussian Channels, *Inform. Control*, 1971, vol. 18, pp. 203–219.

2. Shamai (Shitz), S. and Bar-David, I., The Capacity of Average and Peak-Power-Limited Quadrature Gaussian Channels, *IEEE Trans. Inform. Theory*, 1995, vol. 41, no. 4, pp. 1060–1071.

3. Kahn, J.M. and Barry, J.R., Wireless Infrared Communications, *Proc. IEEE*, 1997, vol. 85, no. 2, pp. 265–298.

4. Hranilovic, S. and Kschischang, F.R., Capacity Bounds for Power- and Band-Limited Optical Intensity Channels Corrupted by Gaussian Noise, *IEEE Trans. Inform. Theory*, 2004, vol. 50, no. 5, pp. 784–795.

5. Chan, T.H., Hranilovic, S., and Kschischang, F.R., Capacity-Achieving Probability Measure for Conditionally Gaussian Channels with Bounded Inputs, *IEEE Trans. Inform. Theory*, 2005, vol. 51, no. 6, pp. 2073–2088.

6. Lapidoth, A., Moser, S.M., and Wigger, M.A., On the Capacity of Free-Space Optical Intensity Channels, *Proc. 2008 IEEE Int. Sympos. on Information Theory (ISIT'2008), Toronto, Canada*, 2008, pp. 2419–2423.

7. Abou-Faycal, I.C., Trott, M.D., and Shamai (Shitz), S., The Capacity of Discrete-Time Memoryless Rayleigh-Fading Channels, *IEEE Trans. Inform. Theory*, 2001, vol. 47, no. 3, pp. 1290–1301.

8. Katz, M. and Shamai (Shitz), S., On the Capacity-Achieving Distribution of the Discrete-Time Noncoherent and Partially Coherent AWGN Channels, *IEEE Trans. Inform. Theory*, 2004, vol. 50, no. 10, pp. 2257–2270.

9. Goldsmith, A.J. and Varaiya, P.P., Capacity of Fading Channels with Channel Side Information, *IEEE Trans. Inform. Theory*, 1997, vol. 43, no. 6, pp. 1986–1992.

10. Gursoy, M.C., Poor, H.V., and Verdú, S., The Noncoherent Rician Fading Channel—Part I: Structure of the Capacity-Achieving Input, *IEEE Trans. Wireless Commun.*, 2005, vol. 4, no. 5, pp. 2193–2206.

11. Tchamkerten, A., On the Discreteness of Capacity-Achieving Distributions, *IEEE Trans. Inform. Theory*, 2004, vol. 50, no. 11, pp. 2773–2778.

12. Huang, J. and Meyn, S.P., Characterization and Computation of Optimal Distributions for Channel Coding, *IEEE Trans. Inform. Theory*, 2005, vol. 51, no. 7, pp. 2336–2351.

13. Liang, Y. and Veeravalli, V.V., Capacity of Noncoherent Time-Selective Rayleigh-Fading Channels, *IEEE Trans. Inform. Theory*, 2004, vol. 50, no. 12, pp. 3095–3110.

14. Chen, J. and Veeravalli, V.V., Capacity Results for Block-Stationary Gaussian Fading Channels with a Peak Power Constraint, *IEEE Trans. Inform. Theory*, 2007, vol. 53, no. 12, pp. 4498–4520.

15. Lapidoth, A., On the Asymptotic Capacity of Stationary Gaussian Fading Channels, *IEEE Trans. Inform. Theory*, 2005, vol. 51, no. 2, pp. 437–446.

16. Luenberger, D.G., *Optimization by Vector Space Methods*, New York: Wiley, 1969.

17. Milnor, J., *Morse Theory*, Princeton: Princeton Univ. Press, 1963.

18. Arnold, V.I., *Teoriya katastrof*, Moscow: Nauka, 1990, 3rd ed. Translated under the title *Catastrophe Theory*, Berlin: Springer, 1992, 3rd ed.

19. Rajpurohit, P., Rawat, A., and Sharma, N., On the Capacity Achieving Distribution for a Peak Power Constrained Channel, in *Proc. National Conf. on Communications (NCC'2007), Kanpur, India, 2007*, pp. 79–83.

20. Sundaram, R.K., *A First Course in Optimization Theory*, Cambridge: Cambridge Univ. Press, 1996.

21. Bertsekas, D., *Nonlinear Programming*, Belmont: Athena Scientific, 2003.

22. Sharma, N. and Shamai (Shitz), S., Characterizing the Discrete Capacity Achieving Distribution with Peak Power Constraint at the Transition Points, in *Proc. IEEE Int. Sympos. on Information Theory and Its Applications (ISITA'2008), Auckland, New Zealand*, 2008.