Increasing Signaling Power not Necessarily Improves Channel Capacity

Rudolf Mathar
Inst. for Theoretical Information Technology
RWTH Aachen University
D-52056 Aachen, Germany
mathar@ti.rwth-aachen.de

Anke Schmeink
UMIC Research Center
RWTH Aachen University
D-52056 Aachen, Germany
schmeink@umic.rwth-aachen.de

***Abstract*— This work is inspired by the general question of how to choose signaling points from a bounded set such that capacity of the corresponding channel is maximized. Since subject to peak power constraints capacity achieving distributions become discrete, this question is most relevant for practical application. However, a solution seems to be difficult in general. In this paper we confine ourselves to determining optimum signaling points for selected schemes. As a key problem, determining the entropy of mixture distributions is identified, which is of interest in itself and has applications in many engineering applications. Even for the equiprobable mixture of two Gaussians no simple analytical expression is known. For 2-PAM signaling we investigate two simpler noise distributions, the triangular and chopped uniform one, and determine the capacity of the corresponding channels. While in the first case capacity increases monotonically as signaling points become further apart, in the second monotonicity does not hold. We finally conjecture that monotonicity can be concluded from the behavior of the equivalent binary asymmetric channel.**

I. Introduction and Motivation

The starting point of this paper is a rather general problem of optimal signaling over discrete input, continuous noise channels. Given that signaling points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_M \in \mathcal{S}$ may be chosen from a bounded set $\mathcal{S} \subset \mathbb{R}^N$, what is the optimum choice of $\boldsymbol{x}_1, \dots, \boldsymbol{x}_M \in \mathcal{S}$ and the optimum distribution $(p_1, \dots, p_M)$ such that mutual information is maximized over all discrete distributions with at most M support points?

This problem is motivated on one hand by practical requirements, present digital signaling is exactly done that way, like M-QAM or M-PSK. On the other hand, is turns out to be the key problem when dealing with peak power constraints. Considering average power constraints only, the classical result of Shannon for additive white Gaussian noise (AWGN) channels, and its extension to MIMO channels, see, e.g., [1], state that the capacity-achieving distribution is also Gaussian with infinite continuous support. However, if peak power constraints are added, then the capacity-achieving input distribution for the scalar Gaussian channel becomes discrete with finite support, as was demonstrated by [2]. Other channels like Poisson, quadrature Gaussian and additive vector Gaussian were also shown to possess a discrete capacity-achieving input distribution under average and peak power constraints as surveyed in [3]. This work and [4] generalize a number of previous papers on the subject by considering conditional Gaussian vector channels subject to bounded-input constraints described by a bounded and closed support $\mathcal{S} \subset \mathbb{R}^N$. Under certain conditions on $\mathcal{S}$ the capacity-achieving distribution is discrete, which includes the previously mentioned modulation schemes as special cases. In [5] the related topic of characterizing the optimum number of mass points is analyzed.

Once the position of signaling points is fixed, the optimum input distribution can be characterized by help of the Kullback-Leibler divergence, as is done in the work [6]. This does not solve the general problem of course, but gives valuable advise how to modify the bit mapping onto complex symbols if the noise distribution is not circularly symmetric, as is often the case for fibre-optical channels.

In this paper, we derive from the above problem interesting subproblems, which are important in themselves and are still hard to solve. We give partial solutions that provide valuable insight into the structure of the general problem.

II. The Channel Model and its Mathematical Structure

We consider a channel model with a finite number M of input signaling points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_M \in \mathbb{R}^n$ which are used by the transmitter according to a certain input distribution $\boldsymbol{p} = (p_1, \dots, p_M) \in \mathcal{D}^M$, where the set of all probability distributions with M support points is denoted by

$$\mathcal{D}^M = \big\{\boldsymbol{p} = (p_1, \dots, p_M) \mid p_i \geq 0, \textstyle\sum_{i=1}^M p_i = 1\big\}.$$

Let random variable $\boldsymbol{X}$ denote the discrete channel input with support $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_M\}$ and distribution $\boldsymbol{p}$. The channel output $\boldsymbol{Y}$ is randomly distorted by noise. Throughout the paper we assume that the distribution of $\boldsymbol{Y}$ given input $\boldsymbol{X} = \boldsymbol{x}_i$ has (Lebesgue) density

$$f(\boldsymbol{y} \mid \boldsymbol{x}_i) = f_i(\boldsymbol{y}),\ \boldsymbol{y} \in \mathbb{R}^n.$$

The AWGN channel $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{n}$ is a special case hereof with $f_i(\boldsymbol{y}) = \varphi(\boldsymbol{y} - \boldsymbol{x}_i)$. Here, φ denotes the density of a Gaussian distribution $N_n(\boldsymbol{0}, \boldsymbol{\Sigma})$.

Mutual information between channel input and output as a function of $\boldsymbol{p} = (p_1, \dots, p_M)$ and $f_1, \dots, f_M$ may be written as

$$
\begin{aligned}
I(\boldsymbol{X};\boldsymbol{Y}) &= I\big(\boldsymbol{p};(f_1,\dots,f_M)\big)\\
&= H(\boldsymbol{Y}) - H(\boldsymbol{Y}\mid \boldsymbol{X})\\
&= H\Big(\sum_{i=1}^M p_i f_i\Big) - \sum_{i=1}^M p_i H(f_i) \qquad (1)\\
&= \sum_{i=1}^M p_i D\Big(f_i\Big\|\sum_{j=1}^M p_j f_j\Big),
\end{aligned}
$$

where $D(f\|g) = \int f\log\frac{f}{g}$ denotes the Kullback-Leibler divergence between densities f and g.

Let $\mathcal{F}$ denote the set of all Lebesgue densities $f:\mathbb{R}^n\to\mathbb{R}^+$. From the convexity of $t\log t$, $t\geq 0$, it is easily concluded that

$$H\Big(\sum_{i=1}^M p_i f_i\Big) \text{ is a concave function of } \boldsymbol{p}\in\mathcal{D}^M. \qquad (2)$$

By applying the log-sum inequality (cf. [7]) we also obtain

$$
\begin{aligned}
&\alpha f_1\log\frac{f_1}{g_1} + (1-\alpha)f_2\log\frac{f_2}{g_2}\\
&\quad\geq \big(\alpha f_1 + (1-\alpha)f_2\big)\log\frac{\alpha f_1+(1-\alpha)f_2}{\alpha g_1+(1-\alpha)g_2},
\end{aligned}
$$

pointwise for any pairs of densities $(f_1,g_1),(f_2,g_2)\in\mathcal{F}^2$. Integrating both sides of the above inequality shows that

$$D(f\|g) \text{ is a convex function of the pair } (f,g)\in\mathcal{F}^2. \qquad (3)$$

Applying (2) and (3) to the third and forth line of representation (1), respectively, gives the following.

Proposition 1: Mutual information $I\big(\boldsymbol{p};(f_1,\dots,f_M)\big)$ is a concave function of $\boldsymbol{p}\in\mathcal{D}^M$ and a convex function of $(f_1,\dots,f_M)\in\mathcal{F}^M$.

Hence, determining the capacity of the channel for fixed channel transfer densities $f_1,\dots,f_M$ leads to a concave optimization problem, namely

$$C = \max_{\boldsymbol{p}\in\mathcal{D}^M} I(\boldsymbol{p};f_1,\dots,f_M). \qquad (4)$$

III. Capaciyt-achieving Input Distributions

In the work [8], capacity-achieving input distributions are characterized by exploiting the KKT-conditions of problem (4). It turns out that distribution $\boldsymbol{p}^*$ is capacity-achieving if the mixture density $\sum_{j=1}^M p_j^* f_j(\boldsymbol{y})$ with weights p_j^* is placed in the center of all noise densities $f_j(\boldsymbol{y})$, where "distance" is measured by the Kullback-Leibler divergence.

Proposition 2: Input distribution $\boldsymbol{p}^*$ is capacity-achieving if and only if

$$D\Big(f_i\Big\|\sum_{j=1}^M p_j^* f_j\Big) = \zeta \qquad (5)$$

for some $\zeta>0$, for all i such that $p_i>0$. Furthermore, if $H(f_i)$ is independent of i, then $\boldsymbol{p}^*$ is capacity-achieving iff

$$\int f_i(\boldsymbol{y})\log\Big(\sum_{j=1}^M p_j^* f_j(\boldsymbol{y})\Big)d\boldsymbol{y} = \zeta \qquad (6)$$

for all i such that $p_i>0$.

We now turn to the problem of how to choose optimum signaling points if the input distribution is fixed, mostly the uniform in what follows. The entropy of mixture distributions will be the key to finding optimum constellations, usually hard to determine as will be outlined in the next section.

IV. Entropy of Mixtures

For a given bounded set $\mathcal{S}\subset\mathbb{R}^N$ and a maximum number of signaling points $M\in\mathbb{N}$ the full optimum signaling problem reads as

$$\max_{\boldsymbol{p}\in\mathcal{D}^M,\ \boldsymbol{x}_1,\dots,\boldsymbol{x}_M\in\mathcal{S}} I\big(\boldsymbol{p};(f_1,\dots,f_M)\big) \qquad (7)$$

with

$$f_i(\boldsymbol{y}) = f(\boldsymbol{y}\mid\boldsymbol{x}_i),\ \boldsymbol{y}\in\mathbb{R}^N.$$

In this section, we assume that the entropy of density $f_i(\boldsymbol{y}) = f(\boldsymbol{y}\mid\boldsymbol{x}_i)$ is independent of i with the same value $H(f_0)$, say, for all i. This holds true, e.g., for the additive noise model $\boldsymbol{Y}=\boldsymbol{X}+\boldsymbol{n}$ where $f_i(\boldsymbol{y}) = f_0(\boldsymbol{y}-\boldsymbol{x}_i)$ for some given noise density f_0. In this case, as can be seen from (1), mutual information only depends, up to the constant $H(f_0)$, on the entropy of the mixture $\sum_{i=1}^M f_i$. Hence, determining the entropy of a mixture distribution plays a prominent role for solving (7).

The entropy of mixture distributions is an important building block for different engineering problems. Gaussian mixtures are used as noise models in certain interference channels, see [9], [10]. An analytical representation of the corresponding entropy seems to be hard to achieve, as is demonstrated in the work [11]. Therein, the entropy of a simple equiprobable mixture of two scalar Guassians with expectation $-\mu$ and μ, respectively, is given involving a certain integral over $\ln$ of $\cosh$. Numerically it is demonstrated that the entropy of this mixture

$$H\big(\tfrac{1}{2}\,\varphi(y+\mu) + \tfrac{1}{2}\,\varphi(y-\mu)\big) \qquad (8)$$

is monotonically increasing as a function of $\mu\geq 0$, with $\varphi(y)$ denoting the scalar Gaussian density with zero expectation and variance σ^2. Although this result seems to be intuitively obvious, and fully meets the interpretation of differential entropy, its derivation is amazingly intricate.

The reduction of Gaussian mixtures is an important problem inter alia in multi-target tracking for radar systems. Mixture pairs of Gaussians are successively merged into a single Gaussian component whose moments match the first two moments of the pair. In the paper [12], the Kulback-Leibler divergence is proposed to minimize the dissimilarity between the single Gaussian and the mixture. The basic problem of computing the Kullback-Leibler divergence between mixtures is in the vein of the present optimum signaling question. No explicit solution is provided in [12], an easily computed upper bound is used instead.

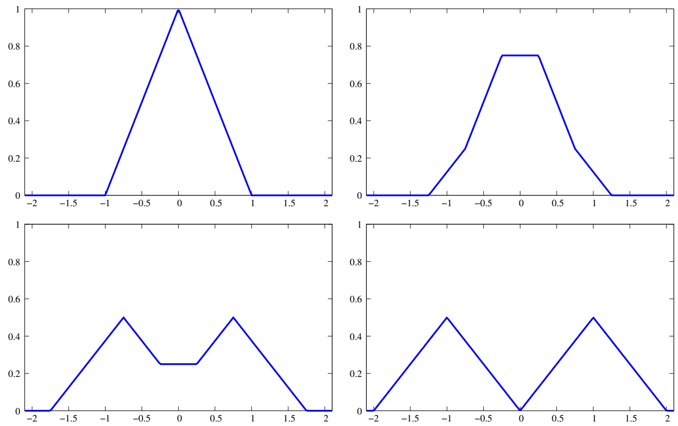

Fig. 1. Mixture densities of the triangular (9) for $\mu = 0, \mu = 0.25, \mu = 0.75$ and $\mu = 1$ (from upper left to lower right).

V. Increasing Power not Always Increases Capacity

Monotonicity of the entropy in (8) when equiprobably mixing two Gaussians with expectations $-\mu$ and μ may be interpreted as the fact that power increases capacity of a scalar AWGN channel with 2-PAM and signaling points $-\mu$ and μ. This seems to be obvious and matches intuition.

The following question reads as a special case of the full problem (7). Select $\mathcal{S}$ to be the intervall $\mathcal{S} = [-b, b]$. Assume two signaling point $-\mu, \mu$ placed symmetrically around zero within $\mathcal{S}$ and both used with the same probability $p_1 = p_2 = \frac{1}{2}$. What is the capacity-maximizing choice of $-\mu$ and μ? One might easily conjecture that maximum power solves the problem, i.e., selecting $-\mu = -b$ and $\mu = b$, which seems to be true in the case of Gaussian noise, as is numerically shown in [11].

The following example with triangular noise densities supports this conjecture, however, it will turn out to be false in general.

A. Mixture of triangular densities

Define the densitiy of a triangular distribution as

$$f_0(y) = \begin{cases} y+1, & -1 \le y \le 0 \\ -y+1, & 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases} . \tag{9}$$

Let

$$f_{-\mu}(y) = f_0(y+\mu) \text{ and } f_\mu(y) = f_0(y-\mu),\ \ \mu \ge 0,$$

hence applying signaling points $-\mu$ and μ for the corresponding channel with additive noise governed by the triangular distribution.

Corresponding mixture densities for $\mu = 0, \mu = 0.25, \mu = 0.75$ and $\mu = 1$ are shown in Figure 1

Tedious algebra leads to an amazingly simple analytical expression for the entropy of the mixture and the mutual

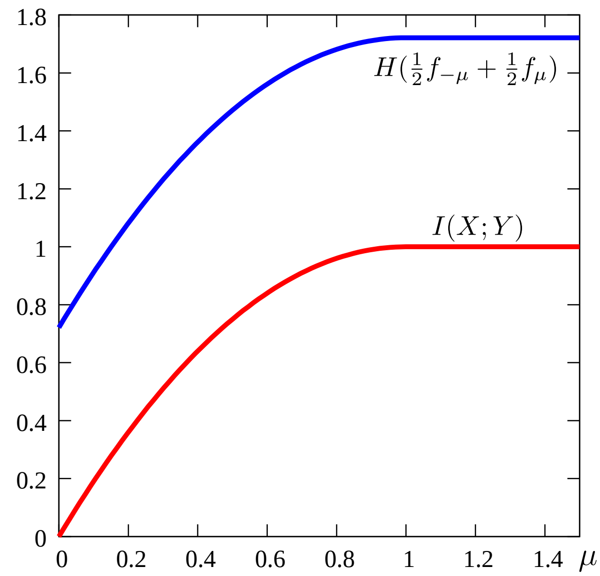

Fig. 2. Entropy and mutual information of a channel as a function of $\mu > 0$ with the triangular as noise distribution and 2-PAM signaling points $-\mu$ and μ.

information of this channel. Logarithms are considered to the base 2.

$$H\big(\tfrac{1}{2}f_{-\mu} + \tfrac{1}{2}f_\mu\big) = \begin{cases} 1 + \frac{\log e}{2} - (1-\mu)^2, & 0 \le \mu \le 1 \\ 1 + \frac{\log e}{2}, & \mu \ge 1 \end{cases},$$

and

$$I\big(X;Y\big) = \begin{cases} 1 - (1-\mu)^2, & 0 \le \mu \le 1 \\ 1, & \mu \ge 1 \end{cases}.$$

Both, entropy and mutual information are depicted as functions of μ in Figure 2, clearly indicating that monotonicity holds. The conclusion is that maximum power, $-\mu = -b$ and $\mu = b$, yields maximum capacity.

Entropy starts at $\mu = 0$ with $\log e/2 = 1/(2 \ln 2) = 0.72135$, the entropy of the standard triangular distribution (9). It increases to $H(\frac{1}{2}f_{-1} + \frac{1}{2}f_1) = 1 + \log e/2 = 1.72135$ at $\mu = 1$, where from it remains constant over all arguments exceeding 1.

Zero capacity is achieved at $\mu = 0$, where signaling points are indiscriminable at the receiver. Capacity increases to one bit per channel use at $\mu = 1$, and stays constant at value 1 for any values $\mu \ge 1$. This is because the support of the error densities does not overlap for $\mu \ge 1$, and hence signaling points can be discriminated without error at the receiver.

The following example shows that capacity not necessarily increases with power. Although it may look artifical, the example shows that monotonicity is strictly connected to the shape of the noise distribution.

B. Mixture of chopped uniforms

Now let

$$f_0(y) = \begin{cases} \frac{1}{2}, & \frac{1}{2} \le |y| \le \frac{3}{2} \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

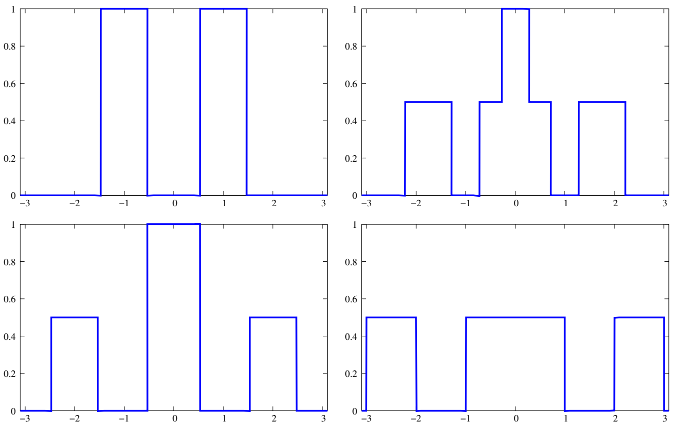

Fig. 3. Mixture densities of the chopped uniform (10) for $\mu = 0, \mu = 0.75, \mu = 1.25$ and $\mu = 1.5$ (from upper left to lower right).

denote the density of a chopped uniform distribution. Again let

$$f_{-\mu}(y) = f_0(y+\mu) \text{ and } f_\mu(y) = f_0(y-\mu), \;\; \mu \geq 0,$$

be the additive noise densities corresponding to signaling points $-\mu$ and μ.

Corresponding mixture densities for $\mu = 0, \mu = 0.25, \mu = 0.75$ and $\mu = 1$ are shown in Figure 3

By studying all possible intersection cases of both densities, entropy of the mixture and mutual information of the corresponding binary input channel are obtained as piecewise linear functions.

$$H\left(\tfrac{1}{2}f_{-\mu} + \tfrac{1}{2}f_\mu\right) = \begin{cases} 2\mu + 1, & 0 \leq \mu \leq \frac{1}{2} \\ \frac{5}{2} - \mu, & \frac{1}{2} \leq \mu \leq 1 \\ \mu + \frac{1}{2}, & 1 \leq \mu \leq \frac{3}{2} \\ 2, & \mu \geq \frac{3}{2} \end{cases},$$

and

$$I(X;Y) = H\left(\tfrac{1}{2}f_{-\mu} + \tfrac{1}{2}f_\mu\right) - 1,$$

applying logarithms to the base 2.

Both are depicted as functions of μ in Figure 4. Obviously, capacity is not an increasing function of μ, and hence of power. Zero capacity is obviously achieved if $\mu = 0$, capacity increases to 1 at the point $\mu = \frac{1}{2}$ where both error densities have disjoint support. With again overlapping support for $\frac{1}{2} \leq \mu \leq \frac{3}{2}$, capacity first decreases linearly to $\frac{1}{2}$ and then increases back to its maximum value 1 at $\mu = \frac{3}{2}$. Since the support overlaps never again for signaling points $|\mu| \geq \frac{3}{2}$, capacity remains constant at value 1.

VI. A Conjecture

We conjecture that in the case of symmetric 2-PAM with additive noise, as used in the two examples above, monotonicity of mutual information holds whenever

$$r(\mu) = \int_0^\infty f_{-\mu}(y)dy + \int_{-\infty}^0 f_\mu(y)dy$$

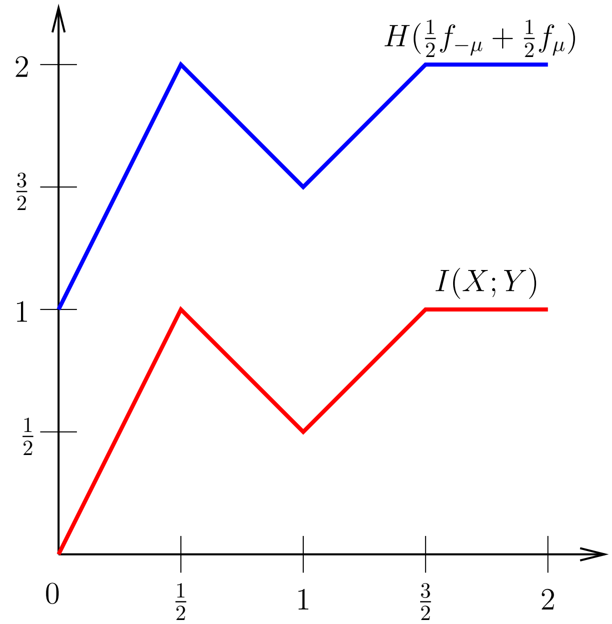

Fig. 4. Entropy and mutual information of a channel as a function of $\mu \geq 0$ with segmented uniforms as noise distribution and 2-PAM signaling points $-\mu$ and μ.

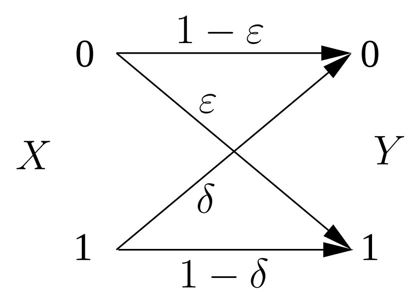

Fig. 5. The binary asymmetric channel with error probabilities $\varepsilon, \delta \in [0,1]$.

is a decreasing function of $\mu \geq 0$.

This is supported by the corresponding equivalent binary asymmetric channel with error probabilities $\varepsilon, \delta \in [0,1]$, see Figure 5. With log-likelihood decoding the error probilities ε and δ correspond to the integral of the noise density over the interval $(-\infty, 0)$ and $(0, \infty)$, respectively.

In [13], the capacity-achieving input distributon $\boldsymbol{p} = (p_0, p_1)$ of this channel is derived as

$$p_0^* = \frac{1}{1+b}, \quad p_1^* = \frac{b}{1+b},$$

with

$$b = \frac{a\varepsilon - (1-\varepsilon)}{\delta - a(1-\delta)} \text{ and } a = \exp\left(\frac{h(\delta) - h(\varepsilon)}{1-\varepsilon-\delta}\right),$$

and $h(\varepsilon) = H(\varepsilon, 1-\varepsilon)$, the entropy of $(\varepsilon, 1-\varepsilon)$.

The corresponding capacity is monotonically increasing as ε and δ decreases, which gives rise to conjecture the same behavior for the continuous channel model.

It seems to be extremely difficult to find a necessary and sufficient condition, which characterizes monotonicity of the capacity of general channels and, furthermore, has an intuitive interpretation.

VII. Conclusions

We have shown interesting differences in the behavior of channel capacity as signaling points become further apart, and hence transmission power increases. Two examples, triangular and chopped uniform noise distributions have been investigated in detail. The material developed in this paper is a special case of the general question of how to find optimum signaling points in a bounded set. A general solution even to this special case seems to be extremely hard. In the future, we will investigate general methods to find optimum signaling constellation for arbitrary noise distributions.

Acknowledgement. This work was supported by the UMIC Research Centre at RWTH Aachen University. UMIC (Ultra-High Speed Mobile Information and Communication) is a research cluster in the framework of the German government excellence initiative.

References

[1] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommuncations*, vol. 10, no. 6, pp. 585–595, 1999.

[2] J. G. Smith, "The information capacity of amplitude and variance-constrained scalar Gaussian channels," *Inf. Contr.*, vol. 18, no. 6, pp. 203–219, 1971.

[3] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2073–2088, June 2005.

[4] T. H. Chan and F. R. Kschischang, "On the discreteness of the capacity-achieving probability measure of conditional Gaussian channels," in *Proceedings IEEE International Symposium on Information Theory, ISIT 2004*, Chicago, June 2004, p. 347.

[5] S. Sharma and S. Shamai, "Characterizing the discrete capacity achieving distribution with peak power constraint at the transition points," in *IEEE International Symposium on Information Theory and Its Applications (ISITA)*, Dec. 2008, pp. 1–6.

[6] A. Feiten and R. Mathar, "Capacity-achieving discrete signaling over additive noise channels," in *Proceedings ICC 2007*, Glasgow, June 2007.

[7] I. Csiszar and J. Körner, *Information Theory, Coding theorems for discrete memoryless sources*. London: Academic Press, 1981.

[8] R. Mathar, A. Schmeink, and M. Zivkovic, "Optimum discrete signaling over channels with arbitrary noise distribution," in *2nd International Conference on Signal Processing and Communication Systems (ICSPCS)*, Gold Coast, Australia, Dec. 2008.

[9] V. Bhatia and B. Mulgrew, "Non-parametric likelihood based channel estimator for Gaussian mixture noise," *Signal Processing*, vol. 87, pp. 2569–2586, 2007.

[10] Y. Wang and L. Wu, "Nonlinear signal detection from an array of threshold devices for non-Gaussian noise," *Digital Signal Processing*, vol. 17, no. 1, pp. 76–89, 2007.

[11] J. Michalowicz, J. Nichols, and F. Buchholz, "Calculation of differential entropy for a mixed Gaussian distribution," *Entropy*, vol. 10, pp. 200–206, 2008.

[12] A. Runnals, "Kullback-Leibler approach to Gaussian mixture reduction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 989–999, July 2007.

[13] R. Mathar and A. Schmeink, "Saddle point properties and Nash equilibria for channel games," *EURASIP Journal on Advances in Signal Processing, special issue on: Game Theory in Signal Processing and Communications*, vol. 2009, pp. 1–9, 2009.