

# Threshold Optimization for Capacity-Achieving Discrete Input One-Bit Output Quantization

Rudolf Mathar

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

Meik Dörpinghaus

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

**Abstract**—In this paper, we consider one-bit output quantization of a discrete signal with  $m$  real signaling points subject to arbitrary additive noise. First, the capacity-achieving distribution is determined for the corresponding channel. For any fixed quantization threshold  $q$  it concentrates on the two most distant signaling points, hence leading to an interpretation as binary asymmetric channel. The direct proof of this result allows for an explicit form of the capacity as a function of threshold  $q$ . We characterize stationary points as candidates for optimal thresholds by a condition on the differential quotient of the derivative of the binary entropy function. In contrast to intuition, symmetry of the noise distribution does not ensure a unique optimum antipodal threshold.

## I. INTRODUCTION AND MOTIVATION

Modern digital receivers use analog-to-digital converters (ADC) of rather high precision, normally 8 to 12 bits. Increasing bandwidth of such systems requires sampling at extremely high speed, which is computationally demanding and energy costly. A drastic reduction of quantization depth to low precision one-bit quantizers allows for designing high-speed systems at tolerable energy consumption. Before rethinking the design of digital systems into this direction the fundamental limits of low precision ADC must be understood.

This paper aims at contributing to this objective. The question we ask is quite basic. Firstly, what is the capacity-achieving distribution for an additive noise channel with a one-bit output quantizer in the class of discrete input distributions with  $m$  support or signaling points. Discrete channel input is quite natural from a practical point of view, since for digital transmission systems usually finite sets of signalling points are used. Moreover, in [1] it is stated that the capacity-achieving distribution over all input distributions is discrete, once an average power constraint applies. In the work [2], the real, discrete-time additive white Gaussian noise (AWGN) channel with an average power constraint is considered. The authors show that for a  $K$ -bit quantizer with a precision of  $\log_2 K$  bits the capacity-achieving input distribution is discrete with at most  $K + 1$  mass points. For binary symmetric quantization

Meik Dörpinghaus is now with the Vodafone Chair Mobile Communications Systems, TU Dresden, Germany. His work was partially supported by the *Deutsche Forschungsgemeinschaft* (DFG) under grant DO 1568/1-1 and in the framework of the Collaborative Research Center 912 'Highly Adaptive Energy-Efficient Computing'.

this result is refined to demonstrating that antipodal signaling is optimum for arbitrary signal-to-noise ratios. The authors conjecture that symmetric quantizers are optimal, however, are not able to provide a proof. Moreover, the loss by low precision ADCs is numerically quantified.

One-bit quantization is also considered in [1], [3] for the same channel model. In the low signal-to-noise regime, as is relevant for spread-spectrum and ultra-wideband communications, it is shown that asymmetric signal constellations combined with asymmetric quantization are superior to the fully symmetric case. It is shown that with such asymmetric threshold quantizers the capacity per unit-energy of the Gaussian channel without output quantization is achieved. However, for this purpose flash-signaling input distributions are required, which are not within the class of peak power constrained input distributions as they will be considered in the present work.

In the recent paper [3], the work [1] is extended to including a peak power constraint. This leads to the result that the capacity-achieving input distribution is concentrated on two extreme mass points, cf. [3, Prop. 1]. Capacity is written as a maximization problem over all possible thresholds, but only numerical indications are given that the optimum threshold is 0. Moreover, for the Gaussian channel it is shown in [3] that a threshold quantizer is optimal.

In the work [4] one-bit quantization is interpreted as an asymmetric channel. Channel capacity and minimal error probability are investigated in parallel, and optimal threshold settings are determined numerically.

In [5] the closely related problem of optimal one-bit source quantization is studied. It is shown that for symmetric and log-concave source distributions the optimal one-bit quantizer is symmetric about the origin.

In the recent paper [6], for a complex-valued fading channel the ergodic capacity and outage probability of one-bit output quantization for discrete I/Q modulation schemes is determined.

In this paper, we consider a real input channel with a finite number of signalling points, arbitrary additive noise and one-bit quantization. What discriminates this work from others, particularly [3], are the following contributions. We give a direct proof of the fact that the capacity-achieving input concentrates on the extreme signaling points only. This proof,

together with the concept of conjugate functions, leads to an explicit representation of the capacity as a function of threshold  $q$ . This is illustrated by the intuitive graphical approach in Fig. 3. We also consider arbitrary noise distributions, which allows for the interesting example that symmetry only does not ensure a unique optimal threshold under the capacity-achieving distribution. We furthermore characterize stationary points as candidates for optimal thresholds by a condition on the differential quotient of the derivative of the binary entropy function, see (26).

## II. CHANNEL MODEL

We assume an additive noise channel. Real input  $X$  with cumulative distribution function (cdf)  $F(x)$  is subject to additive noise  $W$  with density function  $\varphi(w)$  and corresponding cdf  $\Phi(w)$ .  $X$  and  $W$  are assumed to be stochastically independent. The noisy signal  $X+W$  is then quantized by a binary quantizer  $Q$  with threshold  $q$  as  $Q(s) = 1$ , if  $s \geq q$  and  $Q(s) = 0$ , otherwise. The system model is depicted in Fig. 1 and reads as

$$Y = Q(X + W). \quad (1)$$

The conditional distribution of  $Y$ , given  $X = x$  is

$$\begin{aligned} g_{0|x} &= P(Y = 0 | X = x) = \Phi(q - x) \\ g_{1|x} &= P(Y = 1 | X = x) = 1 - g_{0|x} \end{aligned} \quad (2)$$

and the distribution of  $Y$  may be written as

$$\begin{aligned} g_0 &= P(Y = 0) = \int_{-\infty}^{\infty} \Phi(q - x) dF(x) \\ g_1 &= P(Y = 1) = 1 - g_0. \end{aligned} \quad (3)$$

Let  $h(p)$  denote the binary entropy function

$$h(p) = -p \log p - (1 - p) \log(1 - p), \quad 0 \leq p \leq 1. \quad (4)$$

It is well known that  $h(p)$  is a strictly concave function of  $p \in [0, 1]$ .

Mutual information between input  $X$  and binary output  $Y$  may be written as

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y | X) \\ &= h\left(\int_{-\infty}^{\infty} \Phi(q - x) dF(x)\right) - \int_{-\infty}^{\infty} h(\Phi(q - x)) dF(x) \\ &= h(g_0) - \int_{-\infty}^{\infty} h(g_{0|x}) dF(x) \\ &= \int_{-\infty}^{\infty} D((g_{0|x}, g_{1|x}) || (g_0, g_1)) dF(x), \end{aligned} \quad (5)$$

where  $D$  denotes the Kullback-Leibler divergence.

Mutual information is hence a function of the input distribution  $F$  and the quantization threshold  $q$ . This motivates the notation  $I(X; Y) = I(F; q)$ . In the case that  $F$  corresponds to a discrete distribution with discrete density  $\mathbf{p} = (p_1, \dots, p_m)$  we also write  $I(\mathbf{p}; q)$ . It is well known that  $I(F; q)$  is a concave function of  $F$ .

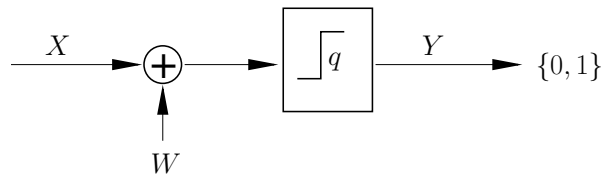


Fig. 1. The system model: some real input  $X$  is subject to additive noise  $W$  and is quantized with threshold  $q$  to yield binary output  $Y$ .

## III. CAPACITY-ACHIEVING INPUT DISTRIBUTION

In the following we assume that the input distribution is discrete with  $m$  probability mass points  $x_1, \dots, x_m \in \mathbb{R}$  in decreasing order  $x_1 > x_2 > \dots > x_m$ . Besides practical requirements on the signaling space also theoretical results imply that the input distribution should be discrete. For the special case of (1) with Gaussian noise it is shown in [2] that the capacity-achieving distribution is discrete if an average power constraint applies. In [3], Theorem 1 states that the capacity-achieving distribution for channel model (1) subject to an average power constraint has at most three mass points.

Also for channels without quantization discrete distributions are capacity-achieving in certain cases. Additive Gaussian noise channels, but also others like Poissonian, quadrature Gaussian and additive vector Gaussian, have a discrete capacity-achieving input distribution if the input is subject to peak power and average power constraints. This was first shown by [7] for additive Gaussian channels and is extended in [8] to conditionally Gaussian channels.

We assume that the input  $X$  with support  $x_1 > x_2 > \dots > x_m$  is governed by a discrete distribution with probabilities  $\mathbf{p} = (p_1, \dots, p_m)$ . Let  $\mathcal{D}$  denote the set of stochastic vectors of length  $m$ , i.e.,

$$\mathcal{D} = \{\mathbf{p} = (p_1, \dots, p_m) \mid p_i \geq 0, \sum_{i=1}^m p_i = 1\}.$$

For a fixed quantization threshold  $q$  we search for the capacity-achieving distribution  $\mathbf{p}^*$ , the one which maximizes mutual information  $I(X; Y) = I(\mathbf{p}, q)$  over all  $\mathbf{p} \in \mathcal{D}$ .

From (3) it is clear that the channel may be represented as a discrete memoryless channel with  $m$  input symbols and binary output. The  $m \times 2$  channel matrix has the form

$$\begin{aligned} \mathbf{V} &= \begin{pmatrix} \Phi(q - x_1) & 1 - \Phi(q - x_1) \\ \vdots & \vdots \\ \Phi(q - x_m) & 1 - \Phi(q - x_m) \end{pmatrix} \\ &= \begin{pmatrix} \gamma_1 & 1 - \gamma_1 \\ \vdots & \vdots \\ \gamma_m & 1 - \gamma_m \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_m \end{pmatrix} \end{aligned} \quad (6)$$

with  $\gamma_i = \Phi(q - x_i)$  and  $m$  rows  $\mathbf{v}_1, \dots, \mathbf{v}_m$  of length two.

*Proposition 1:* For any quantization threshold  $q \in \mathbb{R}$  the capacity-achieving distribution  $\mathbf{p}^*$  of channel model (1) is concentrated on the extreme signaling points  $x_1$  and  $x_m$ . It holds that

$$p_1^* = \frac{z - \gamma_m}{\gamma_1 - \gamma_m}, \quad p_m^* = \frac{\gamma_1 - z}{\gamma_1 - \gamma_m}, \quad (7)$$

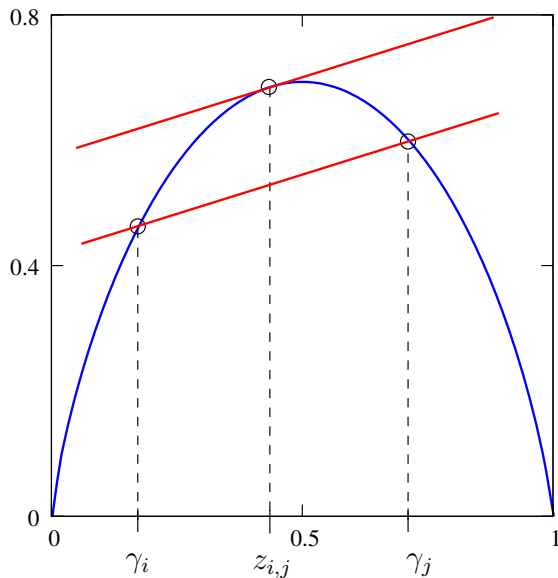


Fig. 2. The binary entropy function  $h$  and the tangent line at the point  $z$ .

where

$$z = \frac{1}{1 + e^d} \quad \text{and} \quad d = \frac{h(\gamma_1) - h(\gamma_2)}{\gamma_1 - \gamma_2}. \quad (8)$$

*Proof.* From [9], see also [10], Th. 4.5.1, we conclude that  $\mathbf{p}^*$  is capacity-achieving if and only if  $D(\mathbf{v}_i \| \mathbf{p}^* \mathbf{V}) = \zeta$  for some  $\zeta > 0$  and all  $i$  with  $p_i^* > 0$ , and furthermore,  $D(\mathbf{v}_j \| \mathbf{p}^* \mathbf{V}) \leq \zeta$  for all  $j$  with  $p_j^* = 0$ .

Now, for any  $\mathbf{p}$  let  $z = \sum_{j=1}^m p_j \gamma_j = g_0$  denote the probability of output 0. We use the representation

$$\begin{aligned} D(\mathbf{v}_i \| \mathbf{p}^* \mathbf{V}) &= D((\gamma_i, 1 - \gamma_i) \| (z, 1 - z)) \\ &= \gamma_i \log \frac{\gamma_i}{z} + (1 - \gamma_i) \log \frac{1 - \gamma_i}{1 - z} \\ &= -h(\gamma_i) - \gamma_i \log \frac{z}{1 - z} - \log(1 - z) \\ &= -h(\gamma_i) + \gamma_i h'(z) - \log(1 - z) \end{aligned} \quad (9)$$

since  $h'(z) = \log \frac{1-z}{z}$ ,  $z \in \{0, 1\}$ , where  $\cdot'$  denotes the derivative.

Now assume that

$$D(\mathbf{v}_i \| \mathbf{p}^* \mathbf{V}) = D(\mathbf{v}_j \| \mathbf{p}^* \mathbf{V}) \quad (10)$$

for some  $\mathbf{p}$  with  $p_i > 0$ ,  $p_j > 0$ . Then from (9)

$$h(\gamma_i) - \gamma_i h'(z) = h(\gamma_j) - \gamma_j h'(z) \quad (11)$$

which entails that

$$\frac{h(\gamma_i) - h(\gamma_j)}{\gamma_i - \gamma_j} = h'(z). \quad (12)$$

$h'(z)$  denotes the derivative of  $h$ . Hence,  $z = P(Y = 0)$  must be chosen in such a way that the slope of the tangent line of the binary entropy function  $h$  at the point  $z$  equals the differential quotient through the point  $h(\gamma_i)$  and  $h(\gamma_j)$ . This condition is depicted in Fig. 2.

Since  $\gamma_1 < \dots < \gamma_m$ , there cannot be three points such that the slope of the tangent line at  $z$  equals the differential quotient between each pair. Hence, there are exactly two elements in  $\{x_1, \dots, x_m\}$  with positive probability, all others carry probability mass zero.

For each pair of points  $\gamma_i < \gamma_j$  the corresponding  $z_{i,j} \in (\gamma_i, \gamma_j)$  is determined by (12) giving an input distribution  $\mathbf{p}^{(i,j)}$  with all components zero except of the  $i$ th and  $j$ th. The Kullback-Leibler divergence reads as

$$\begin{aligned} D(\mathbf{v}_j \| \mathbf{p}^{(i,j)} \mathbf{V}) &= -h(\gamma_j) + \gamma_j h'(z_{i,j}) - \log(1 - z_{i,j}) \\ &= u(z_{i,j}), \text{ say.} \end{aligned} \quad (13)$$

The final step of the proof consists of showing that the pair of extreme points  $x_1$  and  $x_m$  maximizes (13). For this purpose we consider the Kullback-Leibler divergence  $u(z)$  as a function of  $z$  with  $\gamma_j$  fixed. Since  $h''(z) = -\frac{1}{1-z} - \frac{1}{z}$ ,  $0 < z < 1$ , the derivative is obtained as

$$\begin{aligned} u'(z) &= \gamma_j h''(z) + \frac{1}{1-z} \\ &= -\frac{\gamma_j}{1-z} - \frac{\gamma_j}{z} + \frac{1}{1-z} \\ &= \frac{1 - \gamma_j}{1-z} - \frac{\gamma_j}{z} < 0, \end{aligned} \quad (14)$$

since  $z < \gamma_j$  and thus  $1 - z > 1 - \gamma_j$ .

Hence, if  $z_{\ell,j} < z_{i,j}$ , as is the case if  $\gamma_\ell < \gamma_i$ , then

$$\begin{aligned} D(\mathbf{v}_\ell \| \mathbf{p}^{(\ell,j)} \mathbf{V}) &= -h(\gamma_j) + \gamma_j h'(z_{\ell,j}) - \log(1 - z_{\ell,j}) \\ &> -h(\gamma_j) + \gamma_j h'(z_{i,j}) - \log(1 - z_{i,j}) \\ &= D(\mathbf{v}_j \| \mathbf{p}^{(i,j)} \mathbf{V}). \end{aligned} \quad (15)$$

Analogously, moving from any point  $\gamma_j$  to the extreme point  $\gamma_m$  increases the corresponding Kullback-Leibler divergence.

Since  $\Phi$  is strictly increasing, the extreme signaling points  $x_m$  and  $x_1$  are identified as the ones to be exclusively used for maximizing mutual information  $I(X; Y)$ . The corresponding capacity-achieving distribution may be computed from  $h'(z_{1,m}) = \log((1 - z_{1,m})/z_{1,m}) = d$ , such that  $z_{1,m} = \frac{1}{1 + e^d}$ . Since  $z_{1,m} = p_1 \gamma_1 + p_m \gamma_m$  the capacity-achieving distribution has the form stated in Prop. 1. This completes the proof. ■

The result of Proposition 1 holds for any channel matrix of the form (6), provided that  $\gamma_1 < \dots < \gamma_m$ . The capacity-achieving distribution of the asymmetric binary channel is obtained as a special case, cp. [4], [9].

So far we have solved the problem

$$\max_{\mathbf{p} \in \mathcal{D}} I(\mathbf{p}; q)$$

when the quantization threshold  $q$  is fixed.

#### IV. OPTIMUM QUANTIZATION THRESHOLD FOR THE CAPACITY-ACHIEVING INPUT

In the previous section we have shown that the capacity-achieving distribution has only two support points, namely  $x_1$  and  $x_m$ , generating  $\gamma_k = \Phi(q - x_k)$ ,  $k = 1, m$ . The

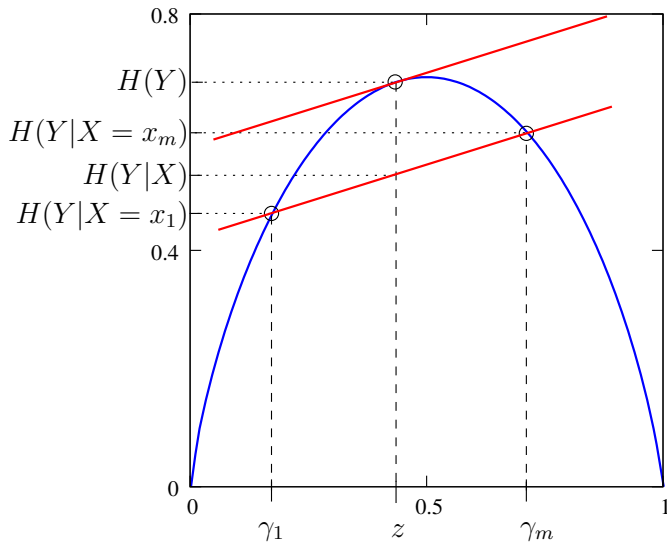


Fig. 3. The binary entropy function  $h$ , the tangent line at  $z$  and relations to entropy and mutual information.

corresponding  $z = P(Y = 0)$  is implicitly defined by equation (12).

We now aim at characterizing the optimum threshold  $q^*$ . For notational convenience the argument  $q$  is suppressed in  $\gamma_1$  and  $\gamma_m$ , however, both  $\gamma_i(q) = \Phi(q - x_i)$ ,  $i = 1, m$ , are complicated functions of  $q \in (-\infty, \infty)$ .

We start from the equation  $I(X; Y) = H(Y) - H(Y | X)$  and observe that

$$\begin{aligned} H(Y) &= h(z), \\ H(Y | X) &= h(\gamma_1) + (z - \gamma_1)h'(z), \\ &= h(\gamma_m) + (z - \gamma_m)h'(z), \end{aligned} \quad (16)$$

where  $z$  is implicitly defined by  $h'(z) = \frac{h(\gamma_1) - h(\gamma_m)}{\gamma_1 - \gamma_m}$ . This implies that

$$\begin{aligned} I(X; Y) &= h(z) - h(\gamma_1) - (z - \gamma_1)h'(z) \\ &= h(z) - h(\gamma_m) - (z - \gamma_m)h'(z). \end{aligned} \quad (17)$$

The substance of formula (17) is visualized in Fig. 3.

From (17) we derive the representation

$$\begin{aligned} I(X; Y) &= \sup_{x \in [0, 1]} \left\{ h(x) - h(\gamma_1) - \frac{h(\gamma_m) - h(\gamma_1)}{\gamma_m - \gamma_1} (x - \gamma_1) \right\} \\ &= \sup_{x \in [0, 1]} \left\{ h(x) - x d(\gamma_1, \gamma_m) \right\} + \gamma_1 d(\gamma_1, \gamma_m) - h(\gamma_1), \end{aligned} \quad (18)$$

where  $d(\gamma_1, \gamma_m) = \frac{h(\gamma_m) - h(\gamma_1)}{\gamma_m - \gamma_1}$  denotes the differential quotient of  $h$  at  $\gamma_1$  and  $\gamma_m$ .

We use the fact that the conjugate function of the negative binary entropy  $-h(x) = x \log x + (1 - x) \log(1 - x)$  is given by

$$(-h)^*(y) = \sup_x \{yx + h(x)\} = \log(1 + e^y) \quad (19)$$

cf. [11], Chap. 3.3, to obtain the representation

$$\begin{aligned} I(X; Y) &= (-h)^*(-d(\gamma_1, \gamma_m)) + \gamma_1 d(\gamma_1, \gamma_m) - h(\gamma_1) \\ &= \log(1 + e^{-d(\gamma_1, \gamma_m)}) + \gamma_1 d(\gamma_1, \gamma_m) - h(\gamma_1). \end{aligned} \quad (20)$$

Analogously it holds that

$$\begin{aligned} I(X; Y) &= \log(1 + e^{-d(\gamma_1, \gamma_m)}) + \gamma_m d(\gamma_1, \gamma_m) - h(\gamma_m). \end{aligned} \quad (21)$$

We continue with representation (20) to investigate maximum points of the mutual information

$$I(X; Y) = \log(1 + e^{-d(\gamma_1, \gamma_m)}) + \gamma_1 d(\gamma_1, \gamma_m) - h(\gamma_1) \quad (22)$$

as a function of  $q$ . As a real function of the argument  $q \in \mathbb{R}$  points of derivative zero are candidates for generating extrema. Recall that  $\cdot'$  denotes the derivative.

Since

$$\begin{aligned} \frac{\partial d(\gamma_1, \gamma_m)}{\partial q} &= \\ \frac{1}{\gamma_1 - \gamma_m} (h'(\gamma_1)\gamma_1' - h'(\gamma_m)\gamma_m' - (\gamma_1' - \gamma_m') d(\gamma_1, \gamma_m)), \end{aligned} \quad (23)$$

the derivative  $\frac{\partial I(X; Y)}{\partial q}$  of (22) equals zero if and only if

$$\begin{aligned} (h'(\gamma_1) - d)\gamma_1'(\gamma_m - \frac{e^{-d}}{1 + e^{-d}}) \\ = (h'(\gamma_m) - d)\gamma_m'(\gamma_1 - \frac{e^{-d}}{1 + e^{-d}}), \end{aligned} \quad (24)$$

as is obtained after some algebra. Using the fact that  $z$  and  $d$  correspond via, see (12)

$$\begin{aligned} h'(z) &= \log \frac{1 - z}{z} = d \\ h'^{-1}(d) &= \frac{e^{-d}}{1 + e^{-d}} = z \end{aligned} \quad (25)$$

the condition for stationarity finally reads as

$$\frac{h'(\gamma_1) - h'(z)}{\gamma_1 - z} \gamma_1' = \frac{h'(\gamma_m) - h'(z)}{\gamma_m - z} \gamma_m'. \quad (26)$$

Equation (26) balances the differential quotients of  $h'$  at  $\gamma_1$  and  $z$ , and  $\gamma_m$  and  $z$ , respectively, against the slope of  $\gamma_1$  and  $\gamma_m$ . Although condition (26) looks rather symmetric it is hard to evaluate explicitly.

In general, the optimum threshold is not unique as the following example demonstrates. Let  $\mathcal{N}(\mu, \sigma^2)$  denote the distribution function of the Gaussian with expectation  $\mu$  and variance  $\sigma^2$ . The signalling points are set to  $x_1 = 1$  and  $x_m = -1$ . The noise distribution  $\Phi$  is formed as

$$\Phi = \frac{1}{2}\mathcal{N}(-\mu, 1) + \frac{1}{2}\mathcal{N}(\mu, 1). \quad (27)$$

for  $\mu = 2, 1.25, 0$  The corresponding channel capacity is shown as a function of threshold  $q$  in Fig. 4. Except of the case  $\mu = 0$  there are two optimum quantization thresholds.

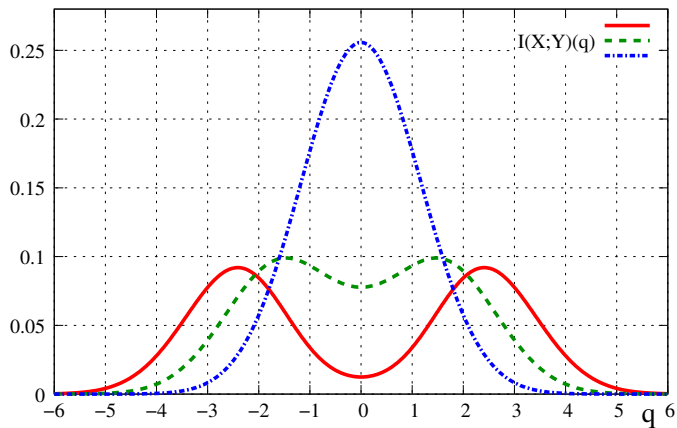


Fig. 4. The capacity of the binary quantization channel as a function of threshold  $q$  with signalling points  $x_1 = 1$ ,  $x_m = -1$  and a mixture of Gaussians from (27) as noise distribution. Solid line:  $\mu = 2$ , dashed:  $\mu = 1.25$ , dash-dotted:  $\mu = 0$ .

## V. SYMMETRY OF MUTUAL INFORMATION

If  $\Phi$  is symmetric with density  $\varphi(x) = \varphi(-x)$ , as is the case for Gaussian noise, the following identities are derived for  $q^* = (x_1 + x_m)/2$ . The notation 'superscript star' refers to values attained at argument  $q^*$ .

$$\begin{aligned} \gamma_1^* &= 1 - \gamma_m^*, \quad d^* = 0, \quad z^* = 1/2, \quad h'(z^*) = 0, \\ h'(\gamma_1^*) &= -h'(\gamma_m^*), \quad \gamma_1'^* = \gamma_m'^* \end{aligned}$$

It follows that condition (26) is satisfied such that in case of symmetry  $q^* = (x_1 + x_m)/2$  forms a stationary point and is a clear candidate for the optimum threshold.

In the sequel the dependence of  $I(X;Y)$  on the threshold  $q$  is emphasized by the notation  $I(X;Y)(q)$ , and correspondingly by writing  $\gamma_i(q)$ ,  $i = 1, m$ . We will demonstrate that  $I(X;Y)(q)$  is symmetric whenever the noise distribution is. For this purpose let

$$\bar{x} = (x_1 + x_m)/2 \quad (28)$$

and assume that  $\Phi(-x) = 1 - \Phi(x)$  for all  $x \in \mathbb{R}$ , as is the case, e.g., for Gaussian noise. Symmetry is understood as

$$I(X;Y)(\bar{x} + q) = I(X;Y)(\bar{x} - q) \quad (29)$$

for all  $q \in \mathbb{R}$ .

We start from the following set of equations

$$\begin{aligned} \gamma_1(\bar{x} + q) &= \Phi\left(q - \frac{x_1 - x_m}{2}\right) = 1 - \gamma_m(\bar{x} - q) \\ \gamma_1(\bar{x} - q) &= \Phi\left(-q - \frac{x_1 - x_m}{2}\right) = 1 - \gamma_m(\bar{x} + q) \\ \gamma_m(\bar{x} + q) &= \Phi\left(q + \frac{x_1 - x_m}{2}\right) = 1 - \gamma_1(\bar{x} - q) \\ \gamma_m(\bar{x} - q) &= \Phi\left(-q + \frac{x_1 - x_m}{2}\right) = 1 - \gamma_1(\bar{x} + q) \end{aligned} \quad (30)$$

for all  $q \in \mathbb{R}$ , and further

$$h(\Phi(x)) = h(1 - \Phi(x)) = h(\Phi(-x)) \quad (31)$$

for all  $x \in \mathbb{R}$ . To emphasize dependence on  $q$  we denote the differential quotient as

$$d(q) = \frac{h(\gamma_m(q)) - h(\gamma_1(q))}{\gamma_m(q) - \gamma_1(q)}. \quad (32)$$

It is easy to see that

$$d(\bar{x} + q) = -d(\bar{x} - q) \quad (33)$$

and

$$h(\gamma_1(\bar{x} + q)) = h(\gamma_m(\bar{x} - q)). \quad (34)$$

These equations together with (20) and (21) yield

$$\begin{aligned} I(X;Y)(\bar{x} + q) &= \log(1 + e^{-d(\bar{x}+q)}) + \gamma_1(\bar{x} + q) d(\bar{x} + q) \\ &\quad - h(\gamma_1(\bar{x} + q)) \\ &= \log(1 + e^{d(\bar{x}-q)}) - (1 - \gamma_m(\bar{x} - q)) d(\bar{x} - q) \\ &\quad - h(\gamma_m(\bar{x} - q)) \\ &= \log(1 + e^{d(\bar{x}-q)}) + \gamma_m(\bar{x} - q) d(\bar{x} - q) - d(\bar{x} - q) \\ &\quad - h(\gamma_m(\bar{x} - q)) \\ &= \log(1 + e^{-d(\bar{x}-q)}) + \gamma_m(\bar{x} - q) d(\bar{x} - q) \\ &\quad - h(\gamma_m(\bar{x} - q)) \\ &= I(X;Y)(\bar{x} - q) \end{aligned} \quad (35)$$

which proves (29).

## REFERENCES

- [1] T. Koch and A. Lapidoth, "Asymmetric quantizers are better at low SNR," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, St. Petersburg, Russia, Jul./Aug. 2011, pp. 2592–2596.
- [2] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3629–3639, Dec. 2009.
- [3] T. Koch and A. Lapidoth, "At low SNR asymmetric quantizers are better," *IEEE Trans. Inf. Theory*, submitted. [Online]. Available: <http://arxiv.org/abs/1201.2523v3>.
- [4] P. Amblard, O. Michel, and S. Morfu, "Revisiting the asymmetric binary channel: joint noise enhanced detection and information transmission through threshold devices," in *Proc. SPIE 5845, Noise in Complex Systems and Stochastic Dynamics III*, Austin, TX, USA, June 2005, pp. 50–60.
- [5] A. Magnani, A. Ghosh, and R. M. Gray, "Optimal one-bit quantization," in *Proc. Data Compression Conference, DCC'05*, Snowbird, UT, USA, Mar. 2005, pp. 270–278.
- [6] S. Krone and G. Fettweis, "Fading channels with 1-bit output quantization: optimal modulation, ergodic capacity and outage probability," in *Proc. IEEE Inform. Theory Workshop, ITW'10*, Dublin, Ireland, Aug./Sep. 2010, pp. 50–60.
- [7] 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.
- [8] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2073–2088, 2005.
- [9] 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.
- [10] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.