Capacity Achieving Modulation for Fixed Constellations with Average Power Constraint

Georg Böcherer, Fabian Altenbach, and Rudolf Mathar

Institute for Theoretical Information Technology

RWTH Aachen University, 52056 Aachen, Germany

Email: {boecherer,altenbach,mathar}@ti.rwth-aachen.de

Abstract—The capacity achieving probability mass function (PMF) of a finite signal constellation with an average power constraint is in most cases non-uniform. A common approach to generate non-uniform input PMFs is Huffman shaping, which consists of first approximating the capacity achieving PMF by a sampled Gaussian density and then to calculate the Huffman code of the sampled Gaussian density. The Huffman code is then used as a prefix-free modulation code. This approach showed good results in practice, can however lead to a significant gap to capacity. In this work, a method is proposed that efficiently constructs optimal prefix-free modulation codes for any finite signal constellation with average power constraint in additive noise. The proposed codes operate as close to capacity as desired. The major part of this work elaborates an analytical proof of this property. The proposed method is applied to 64-QAM in AWGN and numeric results are given, which show that, opposed to Huffman shaping, by using the proposed method, it is possible to operate very close to capacity over the whole range of parameters.

I. Introduction

Reliable communication over a noisy channel at maximum rate is only possible if the input is distributed according to a capacity achieving distribution, i.e., a distribution for which the mutual information between channel input and channel output is maximum.

In digital communication systems, the input is not continuous but has to be chosen from a discrete and finite constellation of signal points. In addition, a modulator has to generate the probability mass function (PMF) of the signal points from equiprobable binary input data. The idea to do so by prefix-free modulation codes originates in [1, IV.A]. Based on this idea, Huffman Shaping was developed in [2]–[4]. Huffman Shaping consists of two steps. First, the PMF of the signal points that minimizes the average energy subject to fixed entropy is chosen. The solution of this optimization problem is a sampled Gaussian density [2], [5, Sec. 4.1.2]. An equivalent formulation of this approach is to look for the signal point PMF that maximizes entropy subject to an average power constraint. Then, in a second step, the Huffman code of the obtained sampled Gaussian density is used as a prefix-free modulation code. However, Huffman Shaping is sub-optimal and can lead to non-trivial gaps to capacity [6, Sec. VIII.A], [5, Sec. 4.2.6]. The reason is that maximizing input entropy is in

general not equivalent to maximizing the mutual information between input and output. Furthermore, the distance measure minimized by Huffman coding is not appropriate [7].

In this work, we propose a method to derive optimal prefix-free modulation codes for fixed signal constellations with an average power constraint and additive noise. We first show that for every fixed signal constellation, average power constraint, and additive noise density, the capacity achieving PMF is given by the solution of a convex optimization problem that can efficiently be solved numerically. We then use Geometric Huffman Coding [7] to find prefix-free modulation codes that approximate capacity achieving PMFs. We finally prove that our method approximates any capacity achieving PMF arbitrarily well both with respect to (w.r.t.) the resulting average power and mutual information. As an illustration of our results, we apply our method to 64-QAM in AWGN and observe that capacity can be approximated extremely well with prefix-free modulation over the whole range of the average power constraint. Our method differs from Huffman Shaping in two ways: first, we approximate the capacity achieving PMFs, which are very different from the sampled Gaussian density over a large range of the average power constraint. Second, the capacity achieving prefix-free modulation codes are not Huffman.

The remainder of this work is organized as follows. In Section II we state the problem of finding good prefix-free modulation codes. Capacity achieving PMFs are characterized in Section III. We then derive in Section IV the offset in mutual information that results from using a 'wrong' PMF. In Section V, we show how to find optimal prefix-free modulation codes. Finally, we present numerical results for 64-QAM in Section VI.

II. Problem Statement

Consider the discrete-time memoryless channel with additive noise given by $Y = X + Z$. The input X takes values in the finite signal constellation set $\mathcal{X} \subset \mathbb{C}$ of cardinality $|\mathcal{X}| = m$. Input X is subject to an average power constraint $\bar{E}$ in terms of energy per channel use. The additive noise term Z takes values in $\mathbb{C}$ and is distributed according to a density h. For each $i = 1, \ldots, m$, conditioned on $x_i \in \mathcal{X}$, the channel output Y is distributed according to $h_i(y) \triangleq h(y - x_i)$.

Let the input X be independent identically distributed (IID) according to a PMF $\mathbf{p} = (p_1, \ldots, p_m)^T$. We denote the energy of the ith signal point by $w_i = |x_i|^2$ and define the energy vector as $\mathbf{w} = (w_1, \ldots, w_m)^T$. The average power constraint $\bar{E}$ can now be written as $\mathbf{w}^T\mathbf{p} \leq \bar{E}$. A capacity achieving PMF maximizes the mutual information between X and Y. According to [8, Eq. (2.4.40)], the mutual information is given by

$$\mathcal{I}(\mathbf{p}) = \mathcal{H}(Y) - \mathcal{H}(Y|X) \tag{1}$$

$$= -\int_{\mathbb{C}} \left(\sum_i p_i h_i(y)\right) \log\left(\sum_i p_i h_i(y)\right) dy$$

$$+ \sum_i p_i \int_{\mathbb{C}} h_i(y) \log[h_i(y)] \, dy \tag{2}$$

where $\mathcal{H}(\cdot)$ denotes the entropy function.

In prefix-free modulation, the binary data stream $(B_k, k \in \mathbb{N})$ is parsed into words from a full prefix-free code, and each word is mapped to a signal point from $\mathcal{X}$ by a one-to-one mapping, see, e.g., [9]. The data bits B_k are equiprobable and jointly independent. As a consequence, the parsed words are independent and identically distributed according to $p_i = 2^{-\ell_i}$, i.e., the probability to parse word i is $2^{-\ell_i}$, where ℓ_i is the number of bits in the ith word. We define the set $\mathbb{D}$ of *dyadic PMFs* as

$$\mathbb{D} = \{\mathbf{p} \,|\, \mathbf{p} \text{ is a PMF},$$

$$\text{for } i = 1, \ldots, m : p_i = 2^{-\ell_i}, \ell_i \in \mathbb{N}\}. \tag{3}$$

By the Kraft inequality [10, Theorem 5.2.1], any $\mathbf{p} \in \mathbb{D}$ can be generated by parsing (B_k) by a prefix-free code with the corresponding codeword lengths. Searching for good prefix-free modulation codes is thus equivalent to searching for good dyadic input PMFs. The rest of this work is therefore about dyadic input PMFs, but keep in mind that for each dyadic input PMF, appropriate prefix-free modulation codes are at hand.

Restricting the input PMF $\mathbf{p}$ to the set of dyadic PMFs $\mathbb{D}$, the discrete optimization problem becomes

maximize $\quad \mathcal{I}(\mathbf{p})$

subject to $\quad p_i = 2^{-\ell_i}, \ell_i \in \mathbb{N}, i = 1, \ldots, m$

$$\mathbf{1}^T\mathbf{p} = 1, \quad \mathbf{w}^T\mathbf{p} \leq \bar{E} \tag{4}$$

where $\mathbf{1}$ is a $m \times 1$ vector with all components equal to one. This is a non-linear optimization problem with integer constraint and no efficient algorithm to solve it is known. Our approach is therefore to first drop the integer constraint and to calculate capacity achieving (in general non-dyadic) PMFs and then to approximate these capacity achieving PMFs by dyadic PMFs. We will do so in the following sections.

III. CAPACITY ACHIEVING PMFs

Mutual information is concave in $\mathbf{p}$, which can be seen as follows. The first term in (2) is an integral over a positively weighted sum of functions concave in p_i, and is thereby concave [11, Ch. 3.2]. The second term is a linear function in $\mathbf{p}$. Therefore, the mutual information is concave in $\mathbf{p}$.
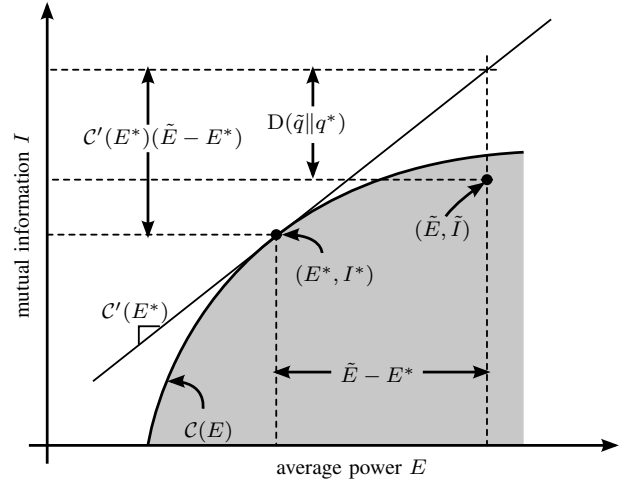

Fig. 1. The shaded area below the capacity curve $\mathcal{C}(E)$ is the region of valid operating points $\{(\mathbf{w}^T\mathbf{p}, \mathcal{I}(\mathbf{p})) \,|\, \mathbf{p} \text{ is an input PMF}\}$. Consider a capacity achieving PMF $\mathbf{p}^*$ and some PMF $\tilde{\mathbf{p}}$. If $\tilde{p}_i = 0$ whenever $p_i^* = 0$, then the corresponding operating points (E^*, I^*) and $(\tilde{E}, \tilde{I})$ relate as $\tilde{I} = I^* + \mathcal{C}'(E^*)(\tilde{E} - E^*) - \mathrm{D}(\tilde{q}\|q^*)$ where $\tilde{q}$ and q^* are the output densities that correspond to $\tilde{\mathbf{p}}$ and $\mathbf{p}^*$, respectively.

For convenience, we replace the maximization in (2) by minimization. This leads to the optimization problem

minimize $\quad -\mathcal{I}(\mathbf{p})$

subject to $\quad \mathbf{1}^T\mathbf{p} = 1, \quad \mathbf{p} \geq 0$

$$\mathbf{w}^T\mathbf{p} \leq \bar{E}. \tag{5}$$

Since the objective function and the inequality constraints are convex and the equality constraint is affine, the above optimization problem is convex. Therefore, an optimal solution can be calculated efficiently by numerical optimization methods [11]. We now explicitly evaluate the Karush-Kuhn-Tucker (KKT) conditions for (5). We refer to these conditions later in this work. The Lagrangian of the optimization problem (5) is given by

$$L(\mathbf{p}, \boldsymbol{\mu}, \nu, \lambda) = -\mathcal{I}(\mathbf{p}) - \boldsymbol{\mu}^T\mathbf{p} + \nu(\mathbf{w}^T\mathbf{p} - \bar{E})$$

$$+ \lambda(\mathbf{1}^T\mathbf{p} - 1) \tag{6}$$

with dual variables $\boldsymbol{\mu} \in \mathbb{R}^m$ and $\nu, \lambda \in \mathbb{R}$. Assuming primal feasibility, for each $1 \leq i \leq m$, the KKT conditions are

$$\frac{\partial L(\mathbf{p}, \boldsymbol{\mu}, \nu, \lambda)}{\partial p_i} = -\frac{\partial \mathcal{I}(\mathbf{p})}{\partial p_i} - \mu_i + \nu w_i + \lambda = 0 \tag{7}$$

$$\mu_i p_i = 0, \quad \nu(\mathbf{w}^T\mathbf{p} - \bar{E}) = 0 \tag{8}$$

$$\mu_i \geq 0, \quad \nu \geq 0. \tag{9}$$

Denote by $\mathbf{p}^*, \boldsymbol{\mu}^*, \lambda^*, \nu^*$ a tuple that fulfills the KKT conditions. From dual feasibility (9) it follows that $\mu_i^* \geq 0$. For every $p_i^* > 0$, by complementary slackness (8), we have $\mu_i^* = 0$. Using these observations in (7) and rearranging the terms, we get

$$\frac{\partial \mathcal{I}(\mathbf{p}^*)}{\partial p_i^*} \leq \lambda^* + \nu^* w_i, \quad \text{with equality if } p_i^* > 0. \tag{10}$$

IV. USING THE 'WRONG' PMF

Suppose our average power constraint is E^*. Solving (5) for $\bar{E} = E^*$ yields the corresponding capacity achieving PMF $\mathbf{p}^*$ and the mutual information $I^* = \mathcal{I}(\mathbf{p}^*)$. Our target operating point is thus (E^*, I^*). If we use some other PMF $\tilde{\mathbf{p}}$ as an approximation of $\mathbf{p}^*$, the effective operating point is $(\tilde{E}, \tilde{I})$, where $\tilde{E} = \mathbf{w}^T\tilde{\mathbf{p}}$ and $\tilde{I} = \mathcal{I}(\tilde{\mathbf{p}})$. We denote by $\mathcal{C}(E)$ the capacity curve, i.e., the maximum mutual information that is achievable under the average power constraint $\bar{E} = E$. Formally,

$$\mathcal{C}(E) = \mathcal{I}(\mathbf{p}) : \mathbf{p} \text{ is a solution of (5) for } \bar{E} = E. \qquad (11)$$

Note that $\mathcal{I}(\mathbf{p}^*) = \mathcal{C}(E^*)$ but in general $\mathcal{I}(\tilde{\mathbf{p}}) \neq \mathcal{C}(\tilde{E})$. The following proposition shows how the effective operating point $(\tilde{E}, \tilde{I})$ relates to the target operating point (E^*, I^*) in terms of $\tilde{\mathbf{p}}$ and $\mathbf{p}^*$. A visualization of the proposition is given in Fig. 1.

Proposition 1. *Consider a capacity achieving PMF $\mathbf{p}^*$ and some input PMF $\tilde{\mathbf{p}}$. If*

$$\tilde{p}_i = 0 \text{ whenever } p_i^* = 0 \qquad (12)$$

then the corresponding operating points $(\tilde{E}, \tilde{I})$ and (E^*, I^*) relate as follows:

$$\tilde{E} = \mathbf{w}^T\tilde{\mathbf{p}}, \qquad E^* = \mathbf{w}^T\mathbf{p}^* \qquad (13)$$
$$\tilde{I} = I^* + \mathcal{C}'(E^*)(\tilde{E} - E^*) - D(\tilde{q}\|q^*) \qquad (14)$$

where $\tilde{q}$ and q^* are the output densities that correspond to $\tilde{\mathbf{p}}$ and $\mathbf{p}^*$, respectively and where $\mathcal{C}'(E)$ is the derivative of $\mathcal{C}(E)$ w.r.t. E. $D(\cdot\|\cdot)$ denotes the Kullback-Leibler (KL) distance as defined in [10, Sec. 8.5].

Proof: The output density q that results from using the input PMF $\mathbf{p}$ is given by

$$q(y) = \sum_i p_i h_i(y), \qquad y \in \mathbb{C}. \qquad (15)$$

Denote by $\tilde{q}$ and q^* the output densities that result from using the input PMF $\tilde{\mathbf{p}}$ and $\mathbf{p}^*$, respectively. We now have

$$\tilde{I} = \mathcal{I}(\tilde{\mathbf{p}}) = \sum_i \tilde{p}_i \int_{\mathbb{C}} h_i(y) \log \frac{h_i(y)}{\tilde{q}(y)} \, dy \qquad (16)$$

$$= \sum_i \tilde{p}_i \int_{\mathbb{C}} h_i(y) \log \frac{h_i(y) q^*(y)}{\tilde{q}(y) q^*(y)} \, dy \qquad (17)$$

$$= \sum_x \tilde{p}_i \int_{\mathbb{C}} h_i(y) \log \frac{h_i(y)}{q^*(y)} \, dy$$
$$+ \sum_i \tilde{p}_i \int_{\mathbb{C}} h_i(y) \log \frac{q^*(y)}{\tilde{q}(y)} \, dy. \qquad (18)$$

For the second summand in (18), we further get

$$\sum_i \tilde{p}_i \int_{\mathbb{C}} h_i(y) \log \frac{q^*(y)}{\tilde{q}(y)} \, dy = \int_{\mathbb{C}} \left(\sum_i \tilde{p}_i h_i(y)\right) \log \frac{q^*(y)}{\tilde{q}(y)} \, dy$$
$$= \int_{\mathbb{C}} \tilde{q}(y) \log \frac{q^*(y)}{\tilde{q}(y)} \, dy = -D(\tilde{q}\|q^*). \qquad (19)$$

By simple calculus, we get in accordance with [8, Eq. (4.5.5)] for the partial derivatives of $\mathcal{I}(\mathbf{p})$ with respect to p_i

$$\frac{\partial \mathcal{I}(\mathbf{p})}{\partial p_i} = \int_{\mathbb{C}} h_i(y) \log \frac{h_i(y)}{q(y)} \, dy - \log e \qquad (20)$$

Using (10) and (20), we get for the integral term in the first summand in (18)

$$\int_{\mathbb{C}} h_i(y) \log \frac{h_i(y)}{q^*(y)} \, dy = \frac{\partial \mathcal{I}(\mathbf{p}^*)}{\partial p_i^*} + \log e \qquad (21)$$

$$\leq \lambda^* + \nu^* w_i + \log e \qquad (22)$$

with equality if $p_i^* > 0$. Expectation w.r.t. $\tilde{\mathbf{p}}$ gives

$$\sum_i \tilde{p}_i \int_{\mathbb{C}} h_i(y) \log \frac{h_i(y)}{q^*(y)} \, dy = \sum_i \tilde{p}_i(\lambda^* + \nu^* w_i + \log e) \qquad (23)$$

$$= \lambda^* + \nu^* \sum_i \tilde{p}_i w_i + \log e \qquad (24)$$

$$= \lambda^* + \nu^* \sum_i (\tilde{p}_i - p_i^* + p_i^*) w_i + \log e \qquad (25)$$

$$= \lambda^* + \nu^* \sum_i p_i^* w_i + \log e + \nu^* \sum_i (\tilde{p}_i - p_i^*) w_i \qquad (26)$$

$$= \sum_i p_i^*(\lambda^* + \nu^* w_i + \log e) + \nu^*(\tilde{E} - E^*) \qquad (27)$$

$$= \sum_i p_i^* \int_{\mathbb{C}} h_i(y) \log \frac{h_i(y)}{q^*(y)} \, dy + \nu^*(\tilde{E} - E^*) \qquad (28)$$

$$= I^* + \nu^*(\tilde{E} - E^*) \qquad (29)$$

where we have equality in (23) since, according to the assumption of the proposition, $\tilde{p}_i = 0$ whenever $p_i^* = 0$, and for all i with $p_i^* > 0$, we have because of (10) equality in (22). In (27), we used $\tilde{E} = \mathbf{w}^T\tilde{\mathbf{p}}$ and $E^* = \mathbf{w}^T\mathbf{p}^*$, respectively. Using (29) and (19) in (18), we get

$$\tilde{I} = I^* + \nu^*(\tilde{E} - E^*) - D(\tilde{q}\|q^*). \qquad (30)$$

By [11, Ch. 5.6.3] it holds that

$$\nu^* = \left.\frac{\partial \mathcal{C}(E)}{\partial E}\right|_{E=E^*} \triangleq \mathcal{C}'(E^*). \qquad (31)$$

Using (31) in (30) gives the statement of the proposition. ∎

Proposition 1 is formulated for the case where consecutive signal points are generated independently according to the same PMF $\mathbf{p}$. We now consider the case where n consecutive signal points are generated according to a joint PMF $\mathbf{p}^{(n)}$. The resulting average power and mutual information per block become respectively

$$E^{(n)} = \mathcal{E}(\mathbf{p}^{(n)}), \qquad I^{(n)} = \mathcal{I}(\mathbf{p}^{(n)}) \qquad (32)$$

where $\mathcal{E}(\mathbf{p}^{(n)})$ is defined as

$$\mathcal{E}(\mathbf{p}^{(n)}) = \sum_{\mathbf{x} \in \mathcal{X}^n} p^{(n)}(\mathbf{x})\|\mathbf{x}\|^2 \qquad (33)$$

where $\mathcal{X}^n$ denotes the Cartesian product of n copies of $\mathcal{X}$ and where $\|\cdot\|$ denotes the Euclidean norm. Since the channel is memoryless, a capacity achieving joint PMF $\mathbf{p}^{(n)*}$ is the

product of n copies of some $\mathbf{p}^*$. As a consequence, $I^{(n)*} = nI^*$ and $E^{(n)*} = nE^*$. The capacity curve is $\mathcal{C}^{(n)}(E^{(n)*}) = n\mathcal{C}(E^{(n)*}/n)$. Consequently, we get for the derivative

$$\frac{\partial \mathcal{C}^{(n)}(E^{(n)*})}{\partial E^{(n)*}} = \frac{\partial n \mathcal{C}(\frac{E^{(n)*}}{n})}{\partial E^{(n)*}} \tag{34}$$

$$= n \cdot \mathcal{C}'\Big(\frac{E^{(n)*}}{n}\Big) \cdot \frac{1}{n} = \mathcal{C}'(E^*). \tag{35}$$

For blocks of n symbols, Proposition 1 now becomes

$$\tilde{I}^{(n)} = \tilde{I}^{(n)*} + \mathcal{C}'(E^*)(\tilde{E}^{(n)} - nE^*) - \mathrm{D}(\tilde{q}^{(n)} \| q^{(n)*}) \tag{36}$$

where $\tilde{q}^{(n)}$ and $q^{(n)*}$ are the output densities that result from using the input PMFs $\tilde{\mathbf{p}}^{(n)}$ and $\mathbf{p}^{(n)*}$, respectively. Dividing by n we get for the mutual information per channel use $\tilde{I}_n = \tilde{I}^{(n)}/n$ and the energy per channel use $\tilde{E}_n = \tilde{E}^{(n)}/n$

$$\tilde{I}_n = I^* + \mathcal{C}'(E^*)(\tilde{E}_n - E^*) - \frac{\mathrm{D}(\tilde{q}^{(n)} \| q^{(n)*})}{n}. \tag{37}$$

V. Optimal Dyadic PMFs

For a given average power constraint E^*, we want to find an operating point $(\tilde{E}, \tilde{I})$ that is close to the target operating point $(E^*, I^* = \mathcal{C}(E^*))$ both in terms of average power and mutual information. The following proposition gives sufficient conditions to accomplish this.

Proposition 2. *Consider a capacity achieving PMF $\mathbf{p}^*$ and a sequence of PMFs $(\tilde{\mathbf{p}}_n, n \in \mathbb{N})$ where each PMF in the sequence fulfills condition* (12). *Then*

$$(\tilde{E}_n, \tilde{I}_n) \xrightarrow{n \to \infty} (E^*, I^*) \tag{38}$$

if one of the following two properties holds:

$$\text{Property 1:} \qquad \mathrm{D}(\tilde{q}_n \| q_n^*) \xrightarrow{n \to \infty} 0 \tag{39}$$

$$\text{Property 2:} \qquad \mathrm{D}(\tilde{\mathbf{p}}_n \| \mathbf{p}_n^*) \xrightarrow{n \to \infty} 0. \tag{40}$$

Proof: The sufficiency of Property 1 follows from the strict concavity of $\mathcal{C}(E)$ in E. This can best be seen by considering the visualization of Proposition 1 in Fig. 1. As $\mathrm{D}(\tilde{q} \| q^*)$ becomes smaller, $(\tilde{E}, \tilde{I})$ is approaching the tangent in (E^*, I^*) of the boundary. However, because the tangent is linear and the boundary is strictly concave, as $\mathrm{D}(\tilde{q} \| q^*)$ is getting smaller, $(\tilde{E}, \tilde{I})$ has to walk in the direction of (E^*, I^*). Otherwise, $(\tilde{E}, \tilde{I})$ would go above the boundary, which is impossible since by the definition of the boundary, $\tilde{I} \leq \mathcal{C}(\tilde{E})$.

Sufficiency of Property 2 holds because the KL-distance between the output densities is upper bounded by the KL-distance between the input PMFs, i.e,

$$\mathrm{D}(\tilde{q} \| q^*) \leq \mathrm{D}(\tilde{\mathbf{p}} \| \mathbf{p}^*) \tag{41}$$

This can easily be shown along the lines of [10, Sec. 4.4]. Thus, Property 2 implies Property 1, and the sufficiency of Property 1 was shown in the first part of this proof. ∎

We now come to the central point of this work, namely to approximate a target operating point (E^*, I^*) by a dyadic PMF $\tilde{\mathbf{p}} \in \mathbb{D}$. By Proposition 1, we know that minimizing the KL-distance $\mathrm{D}(\tilde{q} \| q^*)$ between the corresponding output

densities maximizes mutual information in the sense that $(\tilde{E}, \tilde{I})$ approaches the boundary $\mathcal{C}(E)$, and furthermore, by Proposition 2, we know that if $\mathrm{D}(\tilde{q} \| q^*)$ approaches zero, then $(\tilde{E}, \tilde{I})$ converges to (E^*, I^*) both in terms of average power and mutual information. By Property 2 in Proposition 2, we know that both effects can also be achieved by minimizing $\mathrm{D}(\tilde{\mathbf{p}} \| \mathbf{p}^*)$. No algorithm is known that finds the dyadic PMF that minimizes $\mathrm{D}(\tilde{q} \| q^*)$ with complexity polynomial in m. We therefore minimize the KL-distance between the input PMFs, i.e., the aim is to solve

$$\tilde{\mathbf{p}} = \operatorname*{argmin}_{\mathbf{p} \in \mathbb{D}} \mathrm{D}(\mathbf{p} \| \mathbf{p}^*). \tag{42}$$

As shown in [7], $\tilde{\mathbf{p}}$ can efficiently be found by *Geometric Huffman Coding* (GHC). For the definition of GHC and an implementation see [7], [12]. The complexity of GHC is $m \log m$. In the following, $\tilde{\mathbf{p}} = \mathrm{GHC}(\mathbf{p}^*)$, i.e., $\tilde{\mathbf{p}}$ denotes the optimal dyadic approximation of the capacity achieving PMF $\mathbf{p}^*$. Because of the discrete nature of $\mathbb{D}$, $\mathrm{D}(\tilde{\mathbf{p}} \| \mathbf{p}^*) > 0$ in most cases, and as a consequence, there is a non-zero gap between the reached operating point $(\tilde{E}, \tilde{I})$ and the target operating point (E^*, I^*). This gap can be made arbitrarily small by using GHC to approximate the capacity achieving joint PMF $\mathbf{p}^{(n)*}$ of n consecutive signal points. Since $\tilde{\mathbf{p}}^{(n)} = \mathrm{GHC}(\mathbf{p}^{(n)*})$, by [7, Proposition 2], we have

$$\frac{\mathrm{D}(\tilde{\mathbf{p}}^{(n)} \| \mathbf{p}^{(n)*})}{n} \xrightarrow{n \to \infty} 0 \tag{43}$$

Plugging this into (37), we get by the same concavity argument as in the proof of Proposition 2 the following.

Proposition 3. *Consider a capacity achieving PMF $\mathbf{p}^*$ and the corresponding capacity achieving joint PMFs $\mathbf{p}^{(n)*}$. For*

$$\tilde{\mathbf{p}}^{(n)} \triangleq \mathrm{GHC}(\mathbf{p}^{(n)*}) \tag{44}$$

$$\text{and } (\tilde{E}_n, \tilde{I}_n) \triangleq \Big(\frac{\mathcal{E}(\tilde{\mathbf{p}}^{(n)})}{n}, \frac{\mathcal{I}(\tilde{\mathbf{p}}^{(n)})}{n}\Big), \tag{45}$$

we have

$$(\tilde{E}_n, \tilde{I}_n) \xrightarrow{n \to \infty} (E^*, I^*). \tag{46}$$

VI. Numerical Results: 64-QAM

For illustrative purpose, we apply our algorithm to 64-QAM. The additive noise is zero-mean circular symmetric white Gaussian of unit variance. The scaling of 64-QAM is specified through the highest signal point energy $\max |x|^2$.

For $\max |x|^2 = 20$, and for the average power constraints of $\bar{E} = 2.5, 5, 10, 20$, the capacity achieving PMFs are displayed in Fig. 2. The PMFs are obtained by solving (5). For $\bar{E} = 2.5, 5$, the PMFs resemble the sampled Gaussian density, but for $\bar{E} = 10$, the signal point probabilities follow no longer a monotonic function of the signal point energy. For $\bar{E} = 20$, the average power constraint is no longer active and the resulting average power of the capacity achieving PMF is $E = 11.91$.

We now calculate the dyadic approximations of the capacity achieving operating points for $\bar{E} = 2.5, 2.6, 2.7, \ldots, 12$. For $\bar{E} > 11.91$, the average power constraint is no longer active. In Fig. 3, the dyadic operating points are displayed. The dyadic
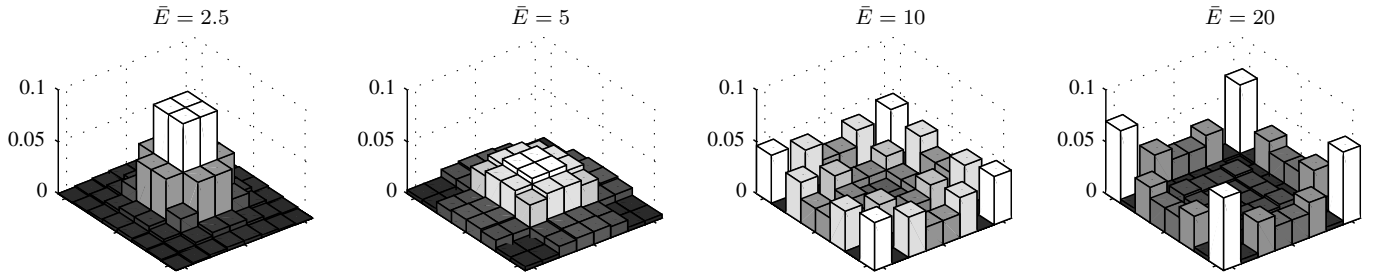
Fig. 2. Capacity achieving PMFs for the 8×8 signal points of 64-QAM for the average power constraints $\bar{E} = 2.5, 5, 10, 20$. The probabilities of the signal points are given by the heights of the vertical bars.
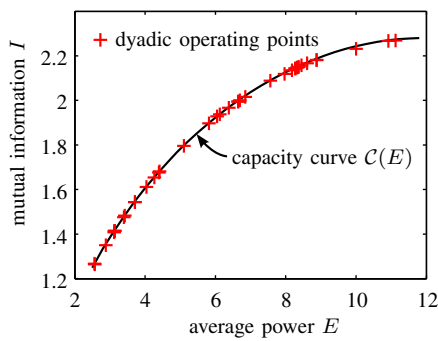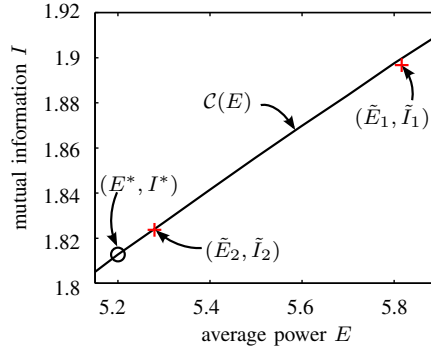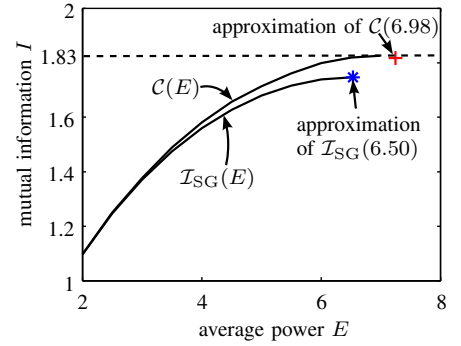

Fig. 3.

Fig. 4.

Fig. 5.

operating points are very close to the capacity curve $\mathcal{C}(E)$. This illustrates that minimizing $\mathrm{D}(\mathbf{p}\|\mathbf{p}^*)$ gives good results in practice. However, the placement of dyadic operating points is irregular. So, for some optimal operating points, there is no close dyadic operating point. This problem is discussed next.

We now illustrate how a specific (capacity achieving) target operating point can be approximated closely by block modulation. The results are displayed in Fig. 4. We choose as a target operating point $(E^*, I^*) = (5.20, 1.81)$. Denote the corresponding capacity achieving PMF by $\mathbf{p}^*$. Using the dyadic PMF $\tilde{\mathbf{p}} = \mathrm{GHC}(\mathbf{p}^*)$, the resulting dyadic operating point is $(\tilde{E}_1, \tilde{I}_1) = (5.82, 1.90)$, which corresponds to approximation errors of 10.65% and 4.74% w.r.t. average power and mutual information, respectively. Exceeding the power constraint by more than 10% may be critical. Jointly modulating two consecutive signal points, i.e., using the joint dyadic PMF $\tilde{\mathbf{p}}^{(2)} = \mathrm{GHC}(\mathbf{p}^{(2)*})$ results in the dyadic operating point $(\tilde{E}_2, \tilde{I}_2) = (5.28, 1.82)$, which corresponds to a power exceed of 1.52% and 0.55% of mutual information. This is a significant improvement and in accordance with Proposition 3.

For $\max |x|^2 = 10$, we display in Fig. 5 the capacity curve $\mathcal{C}(E)$ and the mutual information curve $\mathcal{I}_{\mathrm{SG}}(E)$ that results from using sampled Gaussian densities for the input PMFs. For small E, both curves lie close together. However, as E increases, there is an increasing gap. The capacity curve $\mathcal{C}(E)$ reaches its maximum $I = 1.83$ for $E = 6.98$. For $E > 6.98$, the average power constraint is no longer active. Thus, $\mathcal{C}(\infty) = 1.83$ is the capacity of the signal constellation without average power constraint. The curve $\mathcal{I}_{\mathrm{SG}}(E)$ reaches its maximum for $E = 6.50$. The corresponding mutual information differs from the capacity $\mathcal{C}(\infty)$ by -4.55%. While

this gap is rather small, it implies a bottleneck when the aim is to communicate at rates very close to capacity. Also, this gap may be much larger for other signal constellations and/or other noise densities. The dyadic approximation of $\mathcal{C}(6.98)$ is within -0.39% of capacity $\mathcal{C}(\infty)$, while the dyadic approximation of $\mathcal{I}_{\mathrm{SG}}(6.50)$ is within -4.52% of capacity $\mathcal{C}(\infty)$.

REFERENCES

[1] J. Forney, G., R. Gallager, G. Lang, F. Longstaff, and S. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Sel. Areas Commun.*, vol. 2, no. 5, pp. 632–647, 1984.
[2] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 913–929, 1993.
[3] G. Ungerboeck, "Huffman shaping," in *Codes, Graphs, and Systems*, R. Blahut and R. Koetter, Eds. Springer, 2002, ch. 17, pp. 299–313.
[4] G. Ungerboeck and A. J. Carlson, "System and method for Huffman shaping in a data communication system," U.S. Patent US 7,460,595 B2, Dec. 2, 2008.
[5] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission.* John Wiley & Sons, Inc., 2002.
[6] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, 1999.
[7] G. Böcherer and R. Mathar, "Matching dyadic distributions to channels," in *Proc. Data Compression Conf.*, 2010. [Online]. Available: http://arxiv.org/abs/1009.3751
[8] R. G. Gallager, *Information Theory and Reliable Communication.* John Wiley & Sons, Inc., 1968.
[9] J. Abrahams, "Variable-length unequal cost parsing and coding for shaping," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1648–1650, 1998.
[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., 2006.
[11] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.
[12] G. Böcherer, "Geometric huffman coding," http://www.georg-boecherer.de/ghc, Dec. 2010.