On optimal constellations for BICM at low SNR

Erik Agrell and Alex Alvarado
Department of Signals and Systems, Communication Systems Group
Chalmers University of Technology, Gothenburg, Sweden
{*agrell,alex.alvarado*}*@chalmers.se*

***Abstract*—In this paper we study the problem of finding capacity-maximizing constellations in BICM for asymptotically low signal-to-noise ratios (SNRs). We base our analysis on the so-called Hadamard transform and on a linear approximation of the BICM capacity for asymptotically low SNRs. We fully characterize the set of constellations, input distributions, and binary labelings that achieve Shannon's limit $E_\mathrm{b}/N_0 = -1.59$ dB. For equiprobable input distributions, a constellation achieves this limit if and only if it is a linear projection of a hypercube.**

I. Introduction

Bit-interleaved coded modulation (BICM) was first introduced in [1], and later analyzed in detail in [2], [3]. When compared to a coded modulation (CM) scheme, and from a capacity point of view, BICM is suboptimal [2, Sec. III-A], however, if BICM is used with an appropriate binary labeling, the difference becomes very small. BICM is nowadays a *de facto* standard, and it is used in many of the existing wireless systems, e.g., HSDPA, IEEE 802.11a/g or 802.16, etc. Unlike the CM capacity, the BICM capacity strongly depends on the binary labeling used. Caire *et al*. conjectured that Gray labelings maximize the BICM capacity [2, Sec. III-C], where the binary reflected Gray code (BRGC) and Ungerboeck's set partitioning where compared. This was recently disproved in [4], where it was shown that a non-Gray binary labeling maximizes the BICM capacity for M-ary pulse amplitude modulation (M-PAM) constellations and low signal-to-noise ratios (SNRs).

An analytical characterization of BICM for low SNRs was presented in [5], where an asymptotic linear approximation of the BICM capacity as function of the constellation and its binary labeling was developed. It was showed in [5] that there is a bounded loss between the BICM capacity with Gray-mapped M-PAM and the CM capacity for very low SNRs, i.e., that BICM with this configuration does not reach the Shannon limit (SL) -1.59 dB. Based on the results of [5], Stierstorfer and Fisher showed in [6] that BICM for low SNRs can achieve the SL for M-PAM constellations if the binary labeling is properly selected.

The results in [4]–[6] motivate the fundamental question about the optimal capacity-maximizing binary labelings for BICM, or more generally, optimal "constellations" (binary labelings, input distributions, and constellation points). Somehow surprisingly, this question has received very little attention in the literature apart from [4]–[6]. The analysis presented in [4] is based on a full search (therefore with obvious limitations), and the results in [6] are limited to regular M-PAM and M-ary quaternary amplitude modulation (M-QAM) constellations with equiprobable input distributions. In this paper, we analyze the problem of selecting optimum constellations for BICM and asymptotically low SNRs. We base our analysis on the linear approximation of the BICM capacity of [5] and on the so-called Hadamard transform [7, pp. 53–54]. The main contribution of this paper is to identify the set of constellations that make BICM achieve the SL $E_\mathrm{b}/N_0 = -1.59$ dB for asymptotically low SNRs.

Research supported by the Swedish Research Council, Sweden (under research grant #2006-5599).

II. Preliminaries

A. Binary Labelings

Definition 1 (Binary labeling): A binary labeling[1] $\mathbb{L}$ of order $m \in \mathbb{Z}^+$ is represented by a matrix of dimensions $M = 2^m$ by m, where each row corresponds to one of the M length-m distinct binary codewords, $\mathbb{L} = [\mathbf{c}_0^\mathrm{T}, \ldots, \mathbf{c}_{M-1}^\mathrm{T}]^\mathrm{T}$, where $\mathbf{c}_i = [c_{0,i}, c_{1,i}, \ldots, c_{m-1,i}] \in \{0,1\}^m$.

Definition 2 (± 1 labeling): For any labeling matrix $\mathbb{L}$, a modified labeling matrix $\mathbb{Q} = \mathbb{Q}(\mathbb{L})$ is defined by reversing the order of the columns and applying the mapping $(0 \to 1, 1 \to -1)$, i.e.,

$$q_{ki} \triangleq \begin{cases} -1, & \text{if } c_{m-1-k,i} = 1 \\ 1, & \text{if } c_{m-1-k,i} = 0 \end{cases} \qquad (1)$$

with $i = 0, \ldots, M-1$ and $k = 0, \ldots, m-1$.

Definition 3 (Labeling expansion): To expand a labeling $\mathbb{L}_m = [\mathbf{c}_0^\mathrm{T}, \ldots, \mathbf{c}_{M-1}^\mathrm{T}]^\mathrm{T}$ into a labeling $\mathbb{L}_{m+1}$, do the following. Repeat each codeword once to obtain a new matrix $[\mathbf{c}_0^\mathrm{T}, \mathbf{c}_0^\mathrm{T}, \ldots, \mathbf{c}_{M-1}^\mathrm{T}, \mathbf{c}_{M-1}^\mathrm{T}]^\mathrm{T}$, and then obtain $\mathbb{L}_{m+1}$ by appending one extra column $[0,1,1,0,0,1,1,0,\ldots,0,1,1,0]^\mathrm{T}$ of length $2M$.

Definition 4 (Labeling repetition): To generate a labeling $\mathbb{L}_{m+1}$ from a labeling $\mathbb{L}_m = [\mathbf{c}_0^\mathrm{T}, \ldots, \mathbf{c}_{M-1}^\mathrm{T}]^\mathrm{T}$ by repetition, do the following. Repeat the labeling $\mathbb{L}_m$ once to obtain a new matrix $[\mathbf{c}_0^\mathrm{T}, \ldots, \mathbf{c}_{M-1}^\mathrm{T}, \mathbf{c}_0^\mathrm{T}, \ldots, \mathbf{c}_{M-1}^\mathrm{T}]^\mathrm{T}$. Add an extra column from the left, consisting of M zeros followed by M ones.

[1]Hereafter we use lowercase letters x to denote a scalar and boldface letters $\mathbf{x}$ to denote a vector of scalars. Capital letters X denote random variables, $\mathbb{P}(\cdot)$ denotes probability, $\mathbb{E}[\cdot]$ denotes expectation, and $p_\mathbf{X}(\mathbf{x})$ denotes the probability density function of the random vector $\mathbf{X}$. Blackboard bold letters $\mathbb{X}$ represent matrices or vectors, $\mathbb{X}^\mathrm{T}$ denotes the transpose of $\mathbb{X}$, and $x_{i,j}$ represents its (i,j)th entry of $\mathbb{X}$ where all the indices start at zero.

In this paper we are particularly interested in the binary reflected Gray code (BRGC) [8] and the natural binary code (NBC).

Definition 5 (Binary reflected Gray code): The BRGC of order m, denoted by $\mathbb{G}_m$, is generated by $m-1$ recursive expansions of the trivial labeling $\mathbb{L}_1 = [0\,1]^{\mathrm{T}}$, for any $m \geq 1$.

Definition 6 (Natural binary code): The NBC of order m, denoted by $\mathbb{N}_m$, is generated by $m-1$ recursive repetitions of the trivial labeling $\mathbb{L}_1 = [0\,1]^{\mathrm{T}}$, for any $m \geq 1$.

Example 1 (Binary labelings $\mathbb{G}_3$ and $\mathbb{N}_3$):

$$\mathbb{G}_3 = \begin{bmatrix} 0\,0\,0 \\ 0\,0\,1 \\ 0\,1\,1 \\ 0\,1\,0 \\ 1\,1\,0 \\ 1\,1\,1 \\ 1\,0\,1 \\ 1\,0\,0 \end{bmatrix}, \mathbb{N}_3 = \begin{bmatrix} 0\,0\,0 \\ 0\,0\,1 \\ 0\,1\,0 \\ 0\,1\,1 \\ 1\,0\,0 \\ 1\,0\,1 \\ 1\,1\,0 \\ 1\,1\,1 \end{bmatrix}, \mathbb{Q}(\mathbb{N}_3) = \begin{bmatrix} +1 & +1 & +1 \\ -1 & +1 & +1 \\ +1 & -1 & +1 \\ -1 & -1 & +1 \\ +1 & +1 & -1 \\ -1 & +1 & -1 \\ +1 & -1 & -1 \\ -1 & -1 & -1 \end{bmatrix}.$$

We denote the base-2 representation of the integer $0 \leq i \leq M-1$ by the vector $\mathbf{b}(i) = [b_{m-1}(i), b_{m-2}(i), \ldots, b_0(i)]$, where $b_{m-1}(i)$ is the most significant bit of i and $b_0(i)$ the least significant. Using this notation, the NBC given in Definition 6 can be alternatively defined as the codewords $\mathbf{c}_i$ that are the base-2 representations of the integers $i = 0, \ldots, M-1$, which is in fact the reason for the name "natural binary code," i.e.,

$$\mathbb{N}_m = [\mathbf{b}(0)^{\mathrm{T}}, \ldots, \mathbf{b}(M-1)^{\mathrm{T}}]^{\mathrm{T}}. \quad (2)$$

B. The Hadamard Transform

The Hadamard transform (HT) is a discrete, linear, orthogonal transform, like for example the discrete Fourier transform, but its coefficients take values in ± 1 only. Among the different applications that the HT has, one that is often overlooked is as an analysis tool for binary labelings [9].

The HT is defined by means of an $M \times M$ matrix, the Hadamard matrix, which is defined recursively as follows when M is a power of two [7, pp. 53–54].

$$\mathbb{H}_1 \triangleq 1 \qquad \mathbb{H}_{2M} \triangleq \begin{bmatrix} \mathbb{H}_M & \mathbb{H}_M \\ \mathbb{H}_M & -\mathbb{H}_M \end{bmatrix}, \qquad M \geq 1$$

Example 2 (Hadamard matrix $\mathbb{H}_8$):

$$\mathbb{H}_8 = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 \end{bmatrix}. \quad (3)$$

In the following, we will drop the index, letting $\mathbb{H}$ represent a Hadamard matrix of any size $M = 2^m$. Hadamard matrices have the following appealing properties.

$$\mathbb{H}^{\mathrm{T}} = \mathbb{H}, \quad \mathbb{H}^{-1} = \frac{1}{M}\mathbb{H}. \quad (4)$$

It can be shown that the elements of a Hadamard matrix are

$$h_{ij} = \prod_{k=0}^{m-1} (-1)^{b_k(i) b_k(j)},$$

from which we observe for future use that for all $i = 0, \ldots, M-1$ and $l = 0, \ldots, m-1$,

$$h_{i,0} = 1, \quad (5)$$

$$h_{i,2^l} = \prod_{k=0}^{m-1} (-1)^{b_k(i) b_k(2^l)} = (-1)^{b_l(i)}. \quad (6)$$

At this point it is interesting to note the close relation between the columns of the matrix $\mathbb{N}_3$ in Example 1 and the columns of $\mathbb{H}_8$ in (3). Generalizing, it follows from (5) that for any m, the columns of $\mathbb{Q}(\mathbb{N}_m)$ are simply the columns 2^l of $\mathbb{H}$ for $l = 0, \ldots, m-1$, which will be used later.

The HT operates on a vector of length $M = 2^m$, for any integer m, or in a more general case, on a matrix with $M = 2^m$ rows. The transform of a matrix $\mathbb{X}$ is denoted $\tilde{\mathbb{X}}$ and has the same dimensions as $\mathbb{X}$. It is defined as

$$\tilde{\mathbb{X}} \triangleq \frac{1}{M}\mathbb{H}\mathbb{X} \quad (7)$$

and the inverse transform is

$$\mathbb{X} = \mathbb{H}\tilde{\mathbb{X}}. \quad (8)$$

Equivalently,

$$\tilde{\mathbf{x}}_j = \frac{1}{M}\sum_{i=0}^{M-1} h_{ij}\mathbf{x}_i, \qquad \mathbf{x}_i = \sum_{j=0}^{M-1} h_{ij}\tilde{\mathbf{x}}_j, \quad (9)$$

where we have introduced the row vectors $\mathbf{x}_i$ and $\tilde{\mathbf{x}}_j$ such that

$$\mathbb{X} = \left[\mathbf{x}_0^{\mathrm{T}}, \ldots, \mathbf{x}_{M-1}^{\mathrm{T}}\right]^{\mathrm{T}}, \qquad \tilde{\mathbb{X}} = \left[\tilde{\mathbf{x}}_0^{\mathrm{T}}, \ldots, \tilde{\mathbf{x}}_{M-1}^{\mathrm{T}}\right]^{\mathrm{T}}.$$

Because of (5), the first element of the transform is simply the mean $\tilde{\mathbf{x}}_0 = \frac{1}{M}\sum_{i=0}^{M-1}\mathbf{x}_i$.

Finally, using $\sum_{j=0}^{M-1}\|\tilde{\mathbf{x}}_j\|^2 = \mathrm{trace}\big(\tilde{\mathbb{X}}^{\mathrm{T}}\tilde{\mathbb{X}}\big)$, (7), and (4), we note that an analogy of Parseval's theorem holds, i.e.,

$$\sum_{j=0}^{M-1}\|\tilde{\mathbf{x}}_j\|^2 = \frac{1}{M}\sum_{i=0}^{M-1}\|\mathbf{x}_i\|^2. \quad (10)$$

III. A Linear Approximation of the Capacity

A. System Model

In this section, we analyze the BICM scheme shown in Fig. 1. The information sequence is passed to a rate-k_c binary channel encoder (ENC) which generates a vector of coded bits. This vector is interleaved and then partitioned into length-m codewords $\mathbf{c} = [c_0, \ldots, c_{m-1}]$. The codewords are mapped to constellation points in an N-dimensional Euclidean space using a memoryless mapping rule $\mathcal{M} : \{0,1\}^m \to \mathcal{X}$, where the input alphabet $\mathcal{X} = \{\mathbf{x}_0, \ldots, \mathbf{x}_{M-1}\}$. The probability of transmitting a symbol $\mathbf{x} \in \mathcal{X}$ is denoted $\mathbb{P}(\mathbf{x})$.

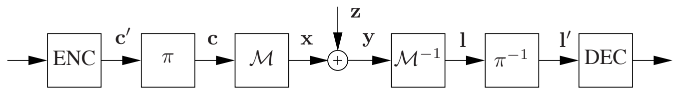

Figure 1. A BICM scheme: A channel encoder, an interleaver, a mapper, an AWGN channel, and the inverse processes at the receiver's side.

We assume a zero-mean constellation, i.e., $\mathbb{E}[\mathbf{X}] = \mathbf{0}$, and a given average symbol energy $E_s = \mathbb{E}[||\mathbf{X}||^2]$; no other constraints are imposed on the constellation. Moreover, we emphasize that the constellation $\mathbb{X}$, in our notation, is a list, not a set. The mapper $\mathcal{M}$ is then defined as a one-to-one mapping rule that associates each codeword with one symbol, i.e., $(\mathbf{c}_i \in \mathbb{L}) \Leftrightarrow (\mathbf{x}_i \in \mathbb{X})$. Based on these definitions, changing the binary labeling of the system can be seen as permuting the rows of $\mathbb{X}$.

We consider transmission over the discrete-time memoryless additive white Gaussian noise (AWGN) channel

$$\mathbf{y} = \mathbf{x} + \mathbf{z}, \tag{11}$$

where $\mathbf{y} \in \mathbb{R}^N$ is the output vector and $\mathbf{z}$ are samples of independent Gaussian random variables with zero mean and variance $N_0/2$. The signal-to-noise ratio (SNR) is given by $\gamma \triangleq E_\mathrm{s}/N_0$. At the receiver's side, the demapper ($\mathcal{M}^{-1}$) computes soft information on the coded bits, which are then deinterleaved and passed to the channel decoder (DEC), which generates an estimate of the information bits.

The capacity of a BICM system can be shown to be [4, Sec. II-B], [3, Sec. 3.2]

$$\mathcal{C}_{\mathbb{L}}^{\mathrm{BI}}(\gamma) = \sum_{k=0}^{m-1} I(C_k; \mathbf{Y}),$$

where $\mathbf{C} = [C_0, \ldots, C_{m-1}]$ are the binary random variables representing the bits in the codewords $\mathbf{c}$ in Fig. 1, and $I(\mathbf{X};\mathbf{Y})$ is the average mutual information between $\mathbf{X}$ and $\mathbf{Y}$.

With a BICM scheme and a good code, it is possible to transmit information with arbitrarily small error probability at a bit rate of $k_c < \mathcal{C}_{\mathbb{L}}^{\mathrm{BI}}$ but not $k_c > \mathcal{C}_{\mathbb{L}}^{\mathrm{BI}}$. The BICM capacity depends on both the labeling and the SNR.

B. The coefficients of the linear approximation

Based on the results of [10] and a Taylor expansion of $\mathcal{C}$ around $\gamma = 0$, $\mathcal{C}(\gamma) = (\alpha\gamma + \beta\gamma^2 + \mathcal{O}(\gamma^2))/\log 2$, it was shown in [5] that when transmitting at the BICM capacity,

$$\frac{E_\mathrm{b}}{N_0} = \frac{\gamma}{\alpha\gamma + \mathcal{O}(\gamma)} \log 2, \tag{12}$$

where[2]

$$\alpha = \frac{2}{E_s} \sum_{k=0}^{m-1} \sum_{u=0}^{1} \left\| \sum_{i \in \mathcal{I}_u^k} \mathbf{x}_i \mathbb{P}(\mathbf{x}_i) \right\|^2 \tag{13}$$

[2]The original expression was given as a function of conditional probabilities $P_X(x)$ (using the notation in [5]), whose values are twice our $\mathbb{P}(\mathbf{x}_i)$. Also, we need an expression for constellations without energy normalization. A factor of $4/E_s$ was included in (13) to implement these changes.

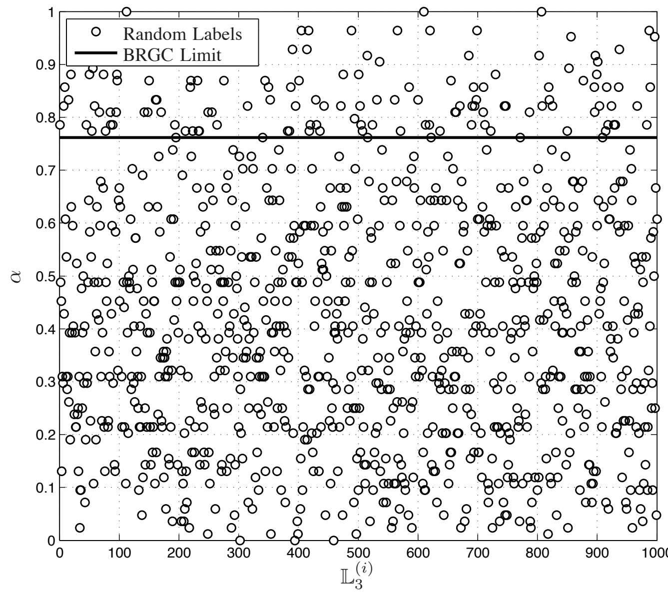

Figure 2. Coefficient α for 10^3 randomly generated labels of order $m = 3$ and an M-PAM constellation. $\alpha_{\mathbb{G}_3}^{\mathrm{BI-PAM}}$ is shown for comparison.

and $\mathcal{I}_u^k$ is the subset of indices $i \in \{0, \ldots, M-1\}$ such that $\mathbf{x}_i$ is labeled with $u \in \{0, 1\}$ at bit position $k = 0, \ldots, m-1$.

At low SNR (low capacity), we obtain

$$\lim_{\gamma \to 0} \frac{E_\mathrm{b}}{N_0} = \frac{\log 2}{\alpha}. \tag{14}$$

Since for reliable transmission we require $E_\mathrm{b}/N_0 \geq \log 2$ [10, Sec. I], the constant α can be interpreted as the penalty of a certain BICM system over an optimal CM scheme. Thus,

$$\alpha \leq 1. \tag{15}$$

In this paper we are interested in the analysis of α in (13) for BICM with different constellations and binary labelings.

For the special case of nonbinary M-PAM constellations, it was shown in [5] that the BRGC does not achieve the SL, even if the number of points grows to infinity, because

$$\alpha = \alpha_{\mathbb{G}_m}^{\mathrm{BI-PAM}} \triangleq \frac{3M^2}{4(M^2-1)} < 1 \tag{16}$$

if $m > 1$. In contrast, the NBC achieves $\alpha = 1$ for all M-PAM constellations [6]. The following example shows that many binary labelings are better than $\mathbb{G}_3$ (higher value of α). It also shows labelings with very low values of α, which translates into a very high E_b/N_0 limit in (14).

Example 3 (Random labelings for 8-PAM): In Fig. 2 the coefficients α for 10^3 randomly generated labelings of order $m = 3$ are presented. The limit for $\mathbb{G}_3$ in (16) is also shown.

In the following section, we find all combinations of constellations, input distributions, and labelings that yield $\alpha = 1$.

IV. Asymptotically optimum constellations

Shannon stated in 1959, "There is a curious and provocative duality between the properties of a source with a distortion measure and those of a channel" [11]. Many instances of this duality have been observed during the last 50 years of communications research. In this context, we point out that the coefficient α is mathematically similar to the so-called *linearity index* [9], which was used to indicate the approximative performance of labelings in a source coding application at high SNR. The usage of the HT in this section was inspired by the analysis in [9].

Theorem 1 (Coefficient α for BICM): For any zero-mean constellation $\mathbb{X}$, probability distribution $\mathbb{P}(\cdot)$, and labeling $\mathbb{L}$,

$$\alpha = \frac{1}{E_s}\sum_{k=0}^{m-1}\left\|\sum_{i=0}^{M-1} q_{ki}\mathbf{x}_i\mathbb{P}(\mathbf{x}_i)\right\|^2.$$

Proof: Expanding the second sum of (13) and using the identity $a^2+b^2=\frac{1}{2}(a+b)^2+\frac{1}{2}(a-b)^2$, we obtain

$$\alpha = \frac{1}{E_s}\sum_{k=0}^{m-1}\left(\|\mathbb{E}[\mathbf{X}]\|^2 + \left\|\sum_{i\in\mathcal{I}_0^k}\mathbf{x}_i\mathbb{P}(\mathbf{x}_i) - \sum_{i\in\mathcal{I}_1^k}\mathbf{x}_i\mathbb{P}(\mathbf{x}_i)\right\|^2\right),$$

where the first term vanishes because of the zero-mean assumption and the second term can be simplified using the definition of q_{ki} in (1). ■

The problem of finding asymptotically optimum constellations has three degrees of freedom: the labeling $\mathbb{L}$, the constellation points $\mathbb{X}$, and the input distribution $\mathbb{P}(\mathbf{x})$. As mentioned before, permuting the labeling $\mathbb{L}$ is equivalent to permuting the rows of $\mathbb{X}$. Hence, we can, without loss of generality, assume a fixed value for $\mathbb{L}$. We find it convenient to let $\mathbb{L}=\mathbb{N}_m$. Similarly, it simplifies the analysis to assume an equiprobable input distribution, which can also be done without loss of generality, as specified in the following Lemma.

Lemma 1: For a given zero-mean constellation $\mathbb{X}'$, labeling $\mathbb{L}$, and input distribution $\mathbb{P}(\mathbf{x}')$, there exists a constellation $\mathbb{X}$ that, if used with NBC labeling and equiprobable input distribution ($\mathbb{P}(\mathbf{x}_i)=1/M, \forall i$), has the same α. The constellation is

$$(\mathbf{x}_0,\dots,\mathbf{x}_{M-1}) = \Pi(\mathbb{P}(\mathbf{x}'_0)\mathbf{x}'_0,\dots,\mathbb{P}(\mathbf{x}'_{M-1})\mathbf{x}'_{M-1}),$$

where Π is the permutation that satisfies $\mathbb{N}_m=\Pi(\mathbb{L})$.

The existence and uniqueness of the permutation Π in Lemma 1 follows directly from Definition 1, while the proof of both constellations having the same α is trivial based on Theorem 1. The significance of this Lemma is that a joint optimization over the three degrees of freedom $[\mathbb{X}',\mathbb{L},\{\mathbb{P}\mathbf{x}\}]$ without loss of generality can be reduced to $[\mathbb{X},\mathbb{N}_m,\{1/M,\dots,1/M\}]$, which has only one degree of freedom: the constellation points themselves.

The expression for α in Theorem 1 can be simplified further using the HT, as elaborated on in the next Theorem. In view of Lemma 1, we confine the analysis to the NBC and equiprobable input distributions.

Theorem 2 (The HT and α): Consider an arbitrary constellation $\mathbb{X}$, whose points are equally likely and labeled by the NBC. The coefficient α is given by

$$\alpha = \frac{1}{E_s}\sum_{k=0}^{m-1}\|\tilde{\mathbf{x}}_{2^k}\|^2,$$

where $\tilde{\mathbf{x}}_{2^k}$ are elements of the HD of $\mathbb{X}$ defined by (7).

Proof: For the NBC, we conclude from (1), (2), and (6) that

$$q_{ki} = (-1)^{b_k(i)} = h_{i,2^k}. \tag{17}$$

Letting $\mathbb{P}(\mathbf{x}_i)=1/M$, $\forall i$ in Theorem 1 and using (9) yields

$$\alpha = \frac{1}{E_s}\sum_{k=0}^{m-1}\left\|\frac{1}{M}\sum_{i=0}^{M-1}h_{i,2^k}\mathbf{x}_i\right\|^2 = \frac{1}{E_s}\sum_{k=0}^{m-1}\|\tilde{\mathbf{x}}_{2^k}\|^2.$$

■

It follows from Theorem 2 and (10) that

$$\alpha \le \frac{1}{E_s}\sum_{j=0}^{M-1}\|\tilde{\mathbf{x}}_j\|^2 = \frac{1}{ME_s}\sum_{i=0}^{M-1}\|\mathbf{x}_i\|^2 = 1 \tag{18}$$

independently of the labeling. In view of Lemma 1, this upper bound holds for any zero-mean constellation, any labeling, and any input distribution, which is in perfect agreement with (15). We now proceed to determine the class of constellations and labelings for which the bound (18) is tight.

Theorem 3 (Linear projection of a hypercube): For any labeling $\mathbb{L}$ and an equiprobable input distribution, $\alpha=1$ if and only if there exists an $m\times N$ matrix $\mathbb{V}$ such that

$$\mathbb{X} = \mathbb{Q}(\mathbb{L})\mathbb{V}. \tag{19}$$

Proof: Consider first the NBC. Equality holds in (18) if and only if $\tilde{\mathbf{x}}_j=0$ for all $j=0,\dots,M-1$ except $j=1,2,4,\dots,2^{m-1}$. For such constellations, (9) yields

$$\mathbf{x}_i = \sum_{k=0}^{m-1}h_{i,2^k}\tilde{\mathbf{x}}_{2^k}.$$

Letting $\mathbf{v}_k \triangleq \tilde{\mathbf{x}}_{2^k}$ for $k=0,\dots,m-1$ and using (17), we obtain

$$\mathbf{x}_i = \sum_{k=0}^{m-1}q_{ki}\mathbf{v}_k, \qquad i=0,\dots,M-1. \tag{20}$$

Letting $\mathbb{V}=[\mathbf{v}_0^{\mathrm{T}},\dots,\mathbf{v}_{m-1}^{\mathrm{T}}]^{\mathrm{T}}$ completes the proof for the NBC. That the theorem also holds for an arbitrary labeling follows from Lemma 1. ■

Example 4 (NBC for M-PAM): Let $\mathbb{L}=\mathbb{N}_m$ and let $\mathbb{V}=[v_0,v_1,\dots,v_{m-1}]^{\mathrm{T}}=[-1,-2,-4,\dots,-2^{m-1}]^{\mathrm{T}}$. With q_{ki} given by (17), we obtain from (19) the constellation $\mathbb{X}_{M-\mathrm{PAM}}\triangleq[-M+1,-M+3,\dots,M-1]^{\mathrm{T}}$, which shows that M-PAM with NBC labeling achieves the SL. In view of Theorem 2, the optimality of M-PAM constellations comes from the fact that the HT of $\mathbb{X}_{M-\mathrm{PAM}}$ has its only nonzero elements in the m positions $1,2,4,\dots,2^{m-1}$.

Example 5 (Asymmetric 4-PAM): Let $\mathbb{L}_2=\mathbb{N}_2$ and let $\mathbb{V}=[(a-b)/2,(a+b)/2]^{\mathrm{T}}$ for any $a,b\in\mathbb{R}$. We obtain from (19) that $\mathbb{X}=[a,b,-b,-a]^{\mathrm{T}}$. This shows that any non equally

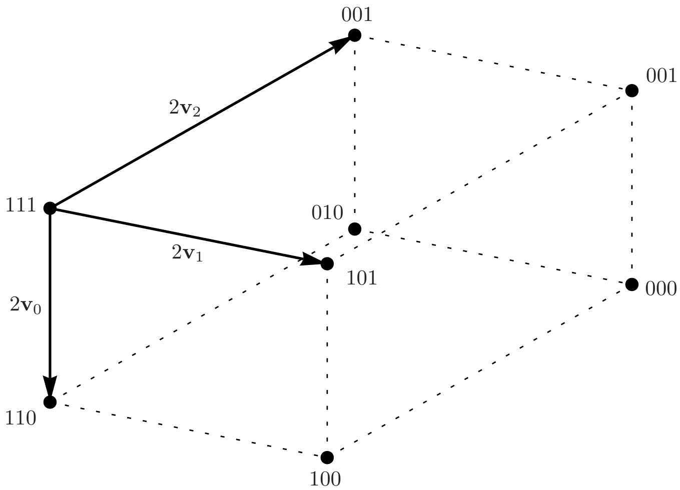

Figure 3. A two-dimensional ($N = 2$) constellation with $m = 3$ that fulfills Theorem 3 and therefore asymptotically achieves the SL. Graphically, it gives the impression of a stretched and projected cube.

spaced but symmetric 4-PAM constellation achieves the SL if NBC labeling is used. This holds even in the special case when $b = 0$. The duplicate points would make the constellation catastrophic for high-rate or uncoded transmission, but it is nevertheless good at very low rates.

Theorem 3 has an appealing geometrical interpretation. Writing the set of constellation points as in (19), each row of $\mathbb{Q}$ can be interpreted as a vertex of an m-dimensional hypercube, and $\mathbb{V}$ as an $m \times N$ projection matrix. Hence, a BICM constellation achieves the SL if and only if its constellation is a *linear projection of a hypercube*. The constellation in Fig. 3, where $\mathbb{V}$ was arbitrarily chosen as $\mathbf{v}_0 = [0, -2.8]^{\mathrm{T}}, \mathbf{v}_1 = [4, -0.8]^{\mathrm{T}}, \mathbf{v}_2 = [4.4, 2.5]^{\mathrm{T}}$, exemplifies the concept. The figure illustrates that the minimum Euclidean distance, which is an important figure-of-merit at high SNR, plays no role at all when constellations are optimized for low SNR.

In Fig. 4, we present various capacity curves: the AWGN capacity, the CM capacity [12] with 8-PAM, and the BICM capacities for 8-PAM with NBC and BRGC. We observe that the CM capacity and the BICM capacity using the NBC achieve the SL. The intersection between the curves for $\mathbb{G}_3$ and $\mathbb{N}_3$ can be appreciated at approximately $k_c = 0.5$ bits/dimension. In this figure, we also include the BICM capacity for the constellation in Fig. 3, and the BICM capacity for symmetric and asymmetric 4-PAM constellations (cf. Example 5), all of them achieving the SL.

V. Conclusions

We have derived necessary and sufficient conditions for a BICM system to achieve the Shannon limit –1.59 dB in the wideband regime [5], i.e., at low code rates or low SNR. Interpreting the codewords of a binary labeling as the vertices of a hypercube, a system achieves the BICM capacity if and only if the constellation points form a linear projection of this hypercube. Important special cases of this result are that

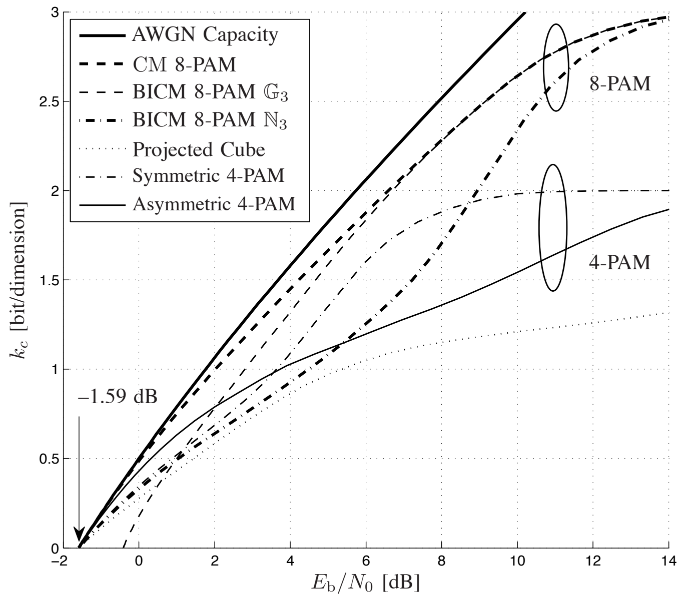

Figure 4. AWGN capacity, CM capacity for 8-PAM, BICM capacities for 8-PAM with the BRGC and the NBC, BICM capacity for the constellation in Fig. 3 ("Projected Cube"), and BICM capacities for symmetric ($a = 3b$) and asymmetric ($2a = 3b$) 4-PAM.

regular M-PAM and $(M_1 \times M_2)$-QAM constellations with natural binary codes achieve the Shannon limit. The result is generalized to non-equiprobable input distributions by rescaling the amplitude of each constellation point proportional to its probability.

References

[1] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 873–884, May 1992.

[2] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[3] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 1-2, pp. 1–153, 2008.

[4] C. Stierstorfer and R. Fischer, "(Gray) Mappings for bit-interleaved coded modulation," in *IEEE Vehicular Technology Conference 2007, VTC-2007 Spring*, Dublin, Ireland, Apr. 2007.

[5] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "Bit-interleaved coded modulation in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5447–5455, Dec. 2008.

[6] C. Stierstorfer and R. Fischer, "Asymptotically optimal mappings for BICM with M-PAM and M^2-QAM," *IET Electronics Letters*, no. 45, pp. 173–174, Jan. 2009.

[7] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1977.

[8] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "Gray coding for multilevel constellations in Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 224–235, Jan. 2007.

[9] P. Knagenhjelm and E. Agrell, "The Hadamard transform—a tool for index assignment," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1139–1151, July 1996.

[10] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[11] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE National Convention Record*, New York, NY, Mar. 1959, pp. 142–163, vol. 7, pt. 4.

[12] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.