Signal Shaping for BICM at Low SNR

Erik Agrell and Alex Alvarado

Abstract—The generalized mutual information (GMI) of bit-interleaved coded modulation (BICM) systems, sometimes called the BICM capacity, is investigated at low signal-to-noise ratio (SNR). The combinations of input alphabet, input distribution, and binary labeling that achieve the Shannon limit –1.59 dB are completely characterized. The main conclusion is that a BICM system with probabilistic shaping achieves the Shannon limit at low SNR if and only if it can be represented as a zero-mean linear projection of a hypercube. Hence, probabilistic shaping offers no extra degrees of freedom to optimize the low-SNR BICM-GMI, in addition to what is provided by geometrical shaping. The analytical conclusions are confirmed by numerical results, which also show that for a fixed input alphabet, probabilistic shaping can improve the BICM-GMI in the low and medium SNR range.

Index Terms—Binary labeling, bit-interleaved coded modulation, generalized mutual information, Hadamard transform, probabilistic shaping, Shannon limit, wideband regime.

I. INTRODUCTION

The most important breakthrough for coded modulation (CM) in fading channels came in 1992, when Zehavi introduced the so-called bit-interleaved coded modulation (BICM) [1], usually referred to as a pragmatic approach for CM [2], [3]. Despite not being fully understood theoretically, BICM has been rapidly adopted in commercial systems such as wireless and wired broadband access networks, 3G/4G telephony, and digital video broadcasting, making it the de facto standard for current telecommunications systems [3, Ch. 1].

Signal shaping refers to the use of non-equally spaced and/or non-equally likely symbols, i.e., *geometrical shaping* and *probabilistic shaping*, resp. Signal shaping has been studied during many years, cf. [4], [5] and references therein. In the context of BICM, geometrical shaping was studied in [6]–[8], and probabilistic shaping, i.e., varying the probabilities of the bit streams, was first proposed in [9], [10] and developed further in [11]–[14]. Probabilistic shaping offers another degree of freedom in the BICM design, which can be used to make the discrete input distribution more similar to the optimal distribution (which is in general unknown). This is particularly advantageous at low and medium SNR.

For the additive white Gaussian noise (AWGN) channel, the so-called Shannon Limit (SL) -1.59 dB represents the average bit energy-to-noise ratio needed to transmit information reliably when the signal-to-noise ratio (SNR) tends to zero [15], [16], i.e., in the wideband regime. When discrete input alphabets are considered at the transmitter and a BICM decoder is used at the receiver, the SL is not always achieved as first noticed in [17]. This was later shown to be caused by the selection of the binary labeling [18]. The behavior of BICM in the wideband regime was studied in [17]–[21] as a function of the alphabet ($\mathbb{X}$) and the binary labeling ($\mathbb{L}$), *assuming a uniform input distribution*. First-order optimal (FOO) constellations were defined in [21] as the triplet $[\mathbb{X}, \mathbb{P}, \mathbb{L}]$ that make a BICM system achieve the SL, where $\mathbb{P}$ represents the input distribution.

In this paper, the results of [21] are generalized to nonuniform input distributions and give a complete characterization of FOO constellations for BICM in terms of $[\mathbb{X}, \mathbb{P}, \mathbb{L}]$. More particularly, the geometrical and/or probabilistic shaping rules that should be applied to a constellation to make it FOO are found. The main conclusion is that probabilistic shaping offers no extra degrees of freedom in addition to what is provided by geometrical shaping for BICM in the wideband regime.

II. PRELIMINARIES

A. Notation

Bold italic letters $\boldsymbol{x}$ denote row vectors. Block letters $\mathbb{X}$ denote matrices or sometimes column vectors. The identity matrix is $\mathbb{I}$. The inner product between two row vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is denoted by $\langle \boldsymbol{a}, \boldsymbol{b} \rangle$ and their element-wise product by $\boldsymbol{a} \circ \boldsymbol{b}$. The Euclidean norm of the vector $\boldsymbol{a}$ is denoted by $\|\boldsymbol{a}\|$. Random variables are denoted by capital letters X and random vectors by boldface capital vectors $\boldsymbol{X}$. The probability density function (pdf) of the random vector $\boldsymbol{Y}$ is denoted by $p_{\boldsymbol{Y}}(\boldsymbol{y})$ and the conditional pdf by $p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$. A similar notation applies to probability mass functions of a random variable, which are denoted by $P_Y(y)$ and $P_{Y|X}(y|x)$. Expectations are denoted by $\mathbb{E}$.

The empty set is denoted by $\varnothing$ and the binary set by $\mathcal{B} \triangleq \{0, 1\}$. The negation of a bit b is denoted by $\bar{b} = 1 - b$. Binary addition (exclusive-OR) of two bits a and b is denoted by $a \oplus b$. The same notation $a \oplus b$ denotes the integer that results from taking the bitwise exclusive-or of two integers a and b.

B. System Model

We consider transmissions over a discrete-time memoryless vectorial fast fading channel. The received vector at any discrete time instant is

$$\boldsymbol{Y} = \boldsymbol{H} \circ \boldsymbol{X} + \boldsymbol{Z} \qquad (1)$$

where $\boldsymbol{X}$ is the channel input and $\boldsymbol{Z}$ is Gaussian noise with zero mean and variance $N_0/2$ in each dimension [1], [3,
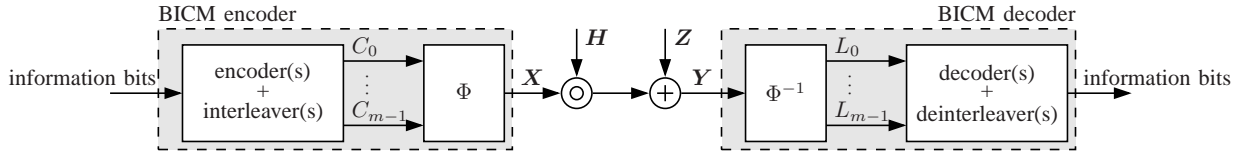
Fig. 1.   A generic BICM system, consisting of a BICM transmitter, the channel, and a BICM receiver.

App. 2.A]. The channel is represented by the N-dimensional vector $\boldsymbol{H}$. It contains the real fading coefficients H_i, which are random, possibly dependent, with the same pdf $p_H(h)$. We assume that $\boldsymbol{H}$ and N_0 are perfectly known at the receiver or can be perfectly estimated, and that the technical requirements on $\boldsymbol{X}$ and $\boldsymbol{H}$ in [21, Sec. I-D] are satisfied.

The conditional transition pdf of the channel in (1) is

$$p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{H}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h}) = \frac{1}{(N_0\pi)^{N/2}} \exp\left(-\frac{\|\boldsymbol{y}-\boldsymbol{h}\circ\boldsymbol{x}\|^2}{N_0}\right). \quad (2)$$

The SNR is defined as

$$\rho \triangleq \mathbb{E}[H^2]\frac{E_\text{s}}{N_0} = R_\text{c}\frac{E_\text{b}^\text{r}}{N_0} \quad (3)$$

where $E_\text{s} \triangleq \mathbb{E}[\|\boldsymbol{X}\|^2]$ is the average *transmitted* symbol energy, R_c is the transmission rate in information bits per symbol, and $E_\text{b}^\text{r} \triangleq \mathbb{E}[H^2]E_\text{s}/R_\text{c}$ is the average *received* energy per information bit.

The generic BICM scheme in Fig. 1 is considered. The transmitter is, in the simplest case, a single binary encoder concatenated with an interleaver and a memoryless mapper Φ. Multiple encoders and/or interleavers may be needed to achieve probabilistic shaping [10]–[13]. At the receiver, using the channel output $\boldsymbol{Y}$, the demapper Φ^{-1} computes metrics L_k for the individual coded bits C_k with $k = 0,\ldots,m-1$, usually in the form of logarithmic likelihood ratios. These metrics are then passed to the deinterleaver(s) and decoder(s) to obtain an estimate of the information bits.

The mapper Φ is defined via the input alphabet $\mathbb{X} = [\boldsymbol{x}_0^\text{T},\ldots,\boldsymbol{x}_{M-1}^\text{T}]^\text{T} \in \mathbb{R}^{M\times N}$, where m bits are used to index the $M = 2^m$ symbols vectors $\boldsymbol{x}_i \in \mathbb{R}^N$ for $i = 0,\ldots,M-1$. We associate with each symbol $\boldsymbol{x}_i$ the codeword (binary labeling) $\boldsymbol{c}_i \triangleq [c_{i,0},\ldots,c_{i,m-1}] \in \mathcal{B}^m$ and the probability $0 \leq P_i \leq 1$, where $P_i \triangleq P_{\boldsymbol{X}}(\boldsymbol{x}_i)$. The binary labeling is denoted by $\mathbb{L} = [\boldsymbol{c}_0^\text{T},\ldots,\boldsymbol{c}_{M-1}^\text{T}]^\text{T} \in \mathcal{B}^{M\times m}$ and the input distribution by $\mathbb{P} = [P_0,\ldots,P_{M-1}]^\text{T} \in [0,1]^M$.

In the following, the labeling used throughout this paper is defined. This can be done without loss of generality, as will be explained in Sec. II-C.

Definition 1 (Natural binary code): The natural binary code (NBC) is the binary labeling $\mathbb{N}_m \triangleq [\boldsymbol{n}(0)^\text{T},\ldots,\boldsymbol{n}(M-1)^\text{T}]^\text{T}$, where $\boldsymbol{n}(i) = [n_{i,0},\ldots,n_{i,m-1}] \in \mathcal{B}^m$ denotes the base-2 representation of the integer $0 \leq i \leq M-1$, with $n_{i,m-1}$ being the most significant bit.

This definition of the NBC is different from the one in [21]. The difference lies only in the bit ordering, i.e., in this paper we consider the last column of $\mathbb{N}_m$ to contain the most significant bits of the base-2 representation of the integers $i =$

$0,1,\ldots,M-1$. It follows from Definition 1 that

$$n_{2^l,k} = \begin{cases} 1, & k = l, \\ 0, & k \neq l \end{cases} \quad (4)$$

for $k = 0,\ldots,m-1$ and $l = 0,\ldots,m-1$, and

$$n_{i\oplus j,k} = n_{i,k} \oplus n_{j,k} \quad (5)$$

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

C. Probabilistic Shaping in BICM

Assuming independent, but possibly nonuniformly distributed, bits $C_0,\ldots,C_{m-1}$ at the input of the modulator (cf. Fig. 1), the symbol probabilities are given by [21, eq. (30)] [13, eq. (8)] [22, eq. (9)]

$$P_i = \prod_{k=0}^{m-1} P_{C_k}(c_{i,k})$$

for $i = 0,\ldots,M-1$, where $P_{C_k}(u)$ for $u \in \mathcal{B}$ is the probability of $C_k = u$. Since $P_{C_k}(1) = 1 - P_{C_k}(0)$, the distribution $\mathbb{P}$ is fully specified by the vector of bit probabilities $\boldsymbol{b} \triangleq [P_{C_0}(0),\ldots,P_{C_{m-1}}(0)]$.

Throughout this paper, we assume that $0 < P_{C_k}(0) < 1$ for all $k = 0,\ldots,m-1$; i.e., all constellation points are used with a nonzero probability. This can be done without loss of generality, because if $P_{C_k}(0) = 0$ or $P_{C_k}(0) = 1$ for some k, then half of the constellation points will never be transmitted. If this is the case, the corresponding branches in Fig. 1 are removed, m is reduced by one, and the mapper Φ is redefined accordingly.[1] The result is another BICM scheme with identical performance, which satisfies $0 < P_{C_k}(0) < 1$ for all k.

For any constellation $[\mathbb{X},\mathbb{P},\mathbb{L}]$, a set of equivalent constellations can be constructed by permuting the rows of $\mathbb{X}$, $\mathbb{L}$, and $\mathbb{P}$, provided that the same permutation is applied to all three matrices. Specifically, denote the permutation that maps the NBC into the desired labeling $\mathbb{L}$ by Π, i.e., $\Pi(\mathbb{N}_m) = \mathbb{L}$. The BICM system defined by the alphabet $\Pi(\mathbb{X})$, the distribution $\Pi(\mathbb{P})$, and the labeling $\Pi(\mathbb{N}_m) = \mathbb{L}$ is entirely equivalent to the system with alphabet $\mathbb{X}$, distribution $\mathbb{P}$, and labeling $\mathbb{N}_m$. Without loss of generality, the analysis in this paper is therefore restricted to the latter case.

Based on the previous discussion, from now on we use the name *constellation* to denote the pair $[\mathbb{X},\mathbb{P}]$, where the NBC labeling is implicit. Thus, $\mathbb{L} = \mathbb{N}_m$ and $c_{i,k} = n_{i,k}$ for all i

---

[1]Constellations with $P_{C_k} = 0$ for some k can yield counter-intuitive results, such as Gray-labeled constellations being FOO (see [13], [14] and Example 7.)

and k, which simplifies the analysis. Note that $\mathbb{P}$ cannot be chosen arbitrarily in BICM; only distributions that satisfy

$$P_i = \prod_{k=0}^{m-1} P_{C_k}(n_{i,k}) \qquad (6)$$

for some vector of bit probabilities $\boldsymbol{b}$ will be considered in the paper. An important special case is the *uniform distribution*, for which $\boldsymbol{b} = [1/2, \ldots, 1/2]$ and $\mathbb{P} = \mathbb{U}_m \triangleq [1/M, \ldots, 1/M]^{\mathrm{T}}$.

D. The Hadamard Transform

The Hadamard transform (HT), or Walsh–Hadamard transform, is a discrete, linear, orthogonal transform, whose coefficients take values in ± 1. It is popular in image processing [23] and can be used to analyze various aspects of binary labelings in digital communications and source coding [21], [24]–[26].

Definition 2: The HT $\tilde{\mathbb{X}} = [\tilde{\boldsymbol{x}}_0^{\mathrm{T}}, \ldots, \tilde{\boldsymbol{x}}_{M-1}^{\mathrm{T}}]^{\mathrm{T}}$ of a matrix (or vector) $\mathbb{X} = [\boldsymbol{x}_0^{\mathrm{T}}, \ldots, \boldsymbol{x}_{M-1}^{\mathrm{T}}]^{\mathrm{T}}$ with $M = 2^m$ rows is

$$\tilde{\boldsymbol{x}}_i \triangleq \frac{1}{M} \sum_{j=0}^{M-1} \boldsymbol{x}_j h_{i,j}, \qquad i = 0, \ldots, M-1 \qquad (7)$$

where for all $i = 0, \ldots, M-1$ and $j = 0, \ldots, M-1$

$$h_{i,j} = \prod_{k=0}^{m-1} (-1)^{n_{i,k} n_{j,k}}. \qquad (8)$$

Because $n_{0,k} = 0$ for $k = 0, \ldots, m-1$, setting $i = 0$ in (7)–(8) shows that the first HT vector

$$\tilde{\boldsymbol{x}}_0 = \frac{1}{M} \sum_{j=0}^{M-1} \boldsymbol{x}_j \qquad (9)$$

can be interpreted as the uniformly weighted mean of the alphabet. This is a property that the HT shares with, e.g., the discrete Fourier transform.

It can be shown from (8) that

$$\sum_{i=0}^{M-1} h_{i,l} h_{i,j} = \begin{cases} M, & j = l, \\ 0, & j \neq l \end{cases} \qquad (10)$$

for all $j = 0, \ldots, M-1$ and $l = 0, \ldots, M-1$. Therefore, the inverse transform is identical to the forward transform, apart from a scale factor:

$$\boldsymbol{x}_j = \sum_{i=0}^{M-1} \tilde{\boldsymbol{x}}_i h_{i,j}, \qquad j = 0, \ldots, M-1. \qquad (11)$$

E. A New Transform

In this section, we define a linear transform between vectors or matrices, which depends on the input distribution $\mathbb{P}$ via the bit probabilities $\boldsymbol{b}$. Its usage will become clear in Section III-C.

Definition 3: Given the bit probabilities $\boldsymbol{b} = [P_{C_0}(0), \ldots, P_{C_{m-1}}(0)]$, the transform $\mathring{\mathbb{X}} = [\mathring{\boldsymbol{x}}_0^{\mathrm{T}}, \ldots, \mathring{\boldsymbol{x}}_{M-1}^{\mathrm{T}}]^{\mathrm{T}}$ of a matrix (or vector) $\mathbb{X} = [\boldsymbol{x}_0^{\mathrm{T}}, \ldots, \boldsymbol{x}_{M-1}^{\mathrm{T}}]^{\mathrm{T}}$ is

$$\mathring{\boldsymbol{x}}_i \triangleq \sum_{j=0}^{M-1} \boldsymbol{x}_j g_{i,j} \sqrt{P_j}, \qquad i = 0, \ldots, M-1 \qquad (12)$$

where P_j is given by (6). The coefficients $g_{i,j}$ are defined as

$$g_{i,j} \triangleq \prod_{k=0}^{m-1} \left[(-1)^{\bar{n}_{i,k} n_{j,k}} \sqrt{P_{C_k}(0)} + (-1)^{n_{i,k} \bar{n}_{j,k}} \sqrt{P_{C_k}(1)} \right] \qquad (13)$$

for all $i = 0, \ldots, M-1$ and $j = 0, \ldots, M-1$, where the bars represent negation ($\bar{b} = 1 - b$, see Sec. II-A).

Remark 1: For equally likely symbols, i.e., $\mathbb{P} = \mathbb{U}_m$, the transform becomes the identity operation $\mathring{\mathbb{X}} = \mathbb{X}$, because then $g_{i,i} = \sqrt{M}$ for $i = 1, \ldots, M$ and $g_{i,j} = 0$ for $i \neq j$.

The transform coefficients $g_{i,j}$ are nonsymmetric in the sense that in general $g_{i,j} \neq g_{j,i}$. They have some appealing properties given by the following lemma, which will be used in the proofs of Theorems 3, 4, and 8.

Lemma 1: For any $j = 0, \ldots, M-1$ and $l = 0, \ldots, M-1$,

$$\sum_{i=0}^{M-1} g_{i,l} g_{i,j} = \begin{cases} M, & j = l, \\ 0, & j \neq l \end{cases} \qquad (14)$$

$$\sum_{i=0}^{M-1} h_{l,i} g_{i,j} = M h_{j,l} \sqrt{P_{j \oplus l}} \qquad (15)$$

where P_j is given by (6) and $h_{l,i}$ is defined in (8).

Proof: See the Appendix. $\qquad \square$

We pay particular attention to two important special cases of (15). First, if $l = 0$, then $h_{l,j} = h_{j,l} = 1$ and $P_{j \oplus l} = P_j$ for $j = 0, \ldots, M-1$. Second, if $l = 2^k$ for any integer $k = 0, \ldots, m-1$, then by (8), $h_{l,i} = h_{i,l} = (-1)^{n_{i,k}}$ for any $i = 0, \ldots, M-1$ and by (6)

$$P_{j \oplus l} = \prod_{k'=0}^{m-1} P_{C_k}(n_{j \oplus 2^k, k'}).$$

Using first (5) and then (4), we obtain

$$P_{j \oplus l} = \left(\prod_{k'=0}^{m-1} P_{C_k}(n_{j,k'}) \right) \frac{P_{C_k}(n_{j,k} \oplus 1)}{P_{C_k}(n_{j,k})}$$
$$= P_j \frac{P_{C_k}(\bar{n}_{j,k})}{P_{C_k}(n_{j,k})}.$$

Substituting these two cases ($l = 0$ and $l = 2^k$) into (15) proves the following corollary.

Corollary 2: For any $j = 0, \ldots, M-1$,

$$\sum_{i=0}^{M-1} g_{i,j} = M \sqrt{P_j} \qquad (16)$$

$$\sum_{i=0}^{M-1} (-1)^{n_{i,k}} g_{i,j} = M (-1)^{n_{j,k}} \sqrt{P_j \frac{P_{C_k}(\bar{n}_{j,k})}{P_{C_k}(n_{j,k})}}. \qquad (17)$$

The fact that the sums $\sum_{i=0}^{M-1} g_{i,l} g_{i,j}$ in (14) are zero whenever $j \neq l$, independently of the input distribution, implies that the coefficients $g_{i,j}$ form an orthogonal basis. As a consequence, the transform is invertible, as shown in the next theorem.

Theorem 3: The inverse transform $\mathbb{X} = [\boldsymbol{x}_0^{\mathrm{T}}, \ldots, \boldsymbol{x}_{M-1}^{\mathrm{T}}]^{\mathrm{T}}$ of a matrix (or vector) $\mathring{\mathbb{X}} = [\mathring{\boldsymbol{x}}_0^{\mathrm{T}}, \ldots, \mathring{\boldsymbol{x}}_{M-1}^{\mathrm{T}}]^{\mathrm{T}}$ is, given the bit

$$
\begin{array}{ccc}
\mathbb{X} & \overset{\text{Def. 3}}{\Longleftrightarrow} & \mathbb{S} = \mathring{\mathbb{X}} = \mathbb{G}\mathbb{D}^{1/2}\mathbb{X} \\[2pt]
 & \text{Theorem 3} & \\[2pt]
\text{HT} \Updownarrow & & \text{HT} \Updownarrow \\[4pt]
 & \text{Theorem 4} & \\[2pt]
\tilde{\mathbb{X}} = \tfrac{1}{M}\mathbb{H}\mathbb{X} & \Longleftrightarrow & \tilde{\mathbb{S}} = \tfrac{1}{M}\mathbb{H}\mathbb{S} = \mathbb{T}\tilde{\mathbb{X}}
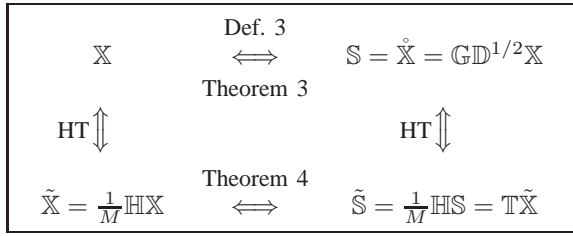\end{array}
$$

Fig. 2.   The relations between the alphabet $\mathbb{X}$, its transform $\mathbb{S}$, and their respective Hadamard transforms $\tilde{\mathbb{X}}$ and $\tilde{\mathbb{S}}$. The transform matrices $\mathbb{G}$, $\mathbb{D}$, $\mathbb{H}$, and $\mathbb{T}$ are defined in Examples 1 and 2.

probabilities $\boldsymbol{b} = [P_{C_0}(0), \ldots, P_{C_{m-1}}(0)]$,

$$
\boldsymbol{x}_j = \frac{1}{M\sqrt{P_j}} \sum_{i=0}^{M-1} \mathring{\boldsymbol{x}}_i g_{i,j}, \qquad j = 0, \ldots, M-1. \quad (18)
$$

Proof: For $j = 0, \ldots, M-1$,

$$
\sum_{i=0}^{M-1} \mathring{\boldsymbol{x}}_i g_{i,j} = \sum_{i=0}^{M-1} g_{i,j} \sum_{l=0}^{M-1} \boldsymbol{x}_l g_{i,l} \sqrt{P_l}
$$
$$
= \sum_{l=0}^{M-1} \boldsymbol{x}_l \sqrt{P_l} \sum_{i=0}^{M-1} g_{i,l} g_{i,j}.
$$

Applying (14) and dividing both sides by $M\sqrt{P_j}$, which by Sec. II-C is nonzero, completes the proof. $\qquad\square$

Example 1: If the bit probabilities are $\boldsymbol{b} = [0.35, 0.50]$, then the symbol probabilities (6) are $\mathbb{P} = [0.175, 0.325, 0.175, 0.325]^{\mathrm{T}}$. The transform coefficients $g_{i,j}$ in (13) are the elements at row i, column j of

$$
\mathbb{G} = \begin{bmatrix} 1.977 & 0.304 & 0 & 0 \\ -0.304 & 1.977 & 0 & 0 \\ 0 & 0 & 1.977 & 0.304 \\ 0 & 0 & -0.304 & 1.977 \end{bmatrix} \quad (19)
$$

It is readily verified that $\mathbb{G}^{\mathrm{T}}\mathbb{G} = M\mathbb{I}$, which is (14) in matrix notation. The mean values in each column of (19) are $[0.418, 0.570, 0.418, 0.570]^{\mathrm{T}}$, which in agreement with (16) are the square roots of the elements in $\mathbb{P}$. Similarly, it can be shown that $\mathbb{G}$ in (19) satisfies (15) and (17).

If the Gray-labeled 4-ary pulse amplitude modulation (PAM) constellation $[\mathbb{X}, \mathbb{P}]$ is considered, $\mathbb{X} = [-3, -1, 3, 1]^{\mathrm{T}}$. Rewriting (12) in matrix notation, the transform can be calculated as $\mathring{\mathbb{X}} = \mathbb{G}\mathbb{D}^{1/2}\mathbb{X} = [-2.654, -0.746, 2.654, 0.746]$, where $\mathbb{D} \triangleq \mathrm{diag}(\mathbb{P})$. This nonequally spaced 4-PAM alphabet will be illustrated and analyzed in Example 3. The inverse transform (18) can be written as $\mathbb{X} = (1/M)\mathbb{D}^{-1/2}\mathbb{G}^{\mathrm{T}}\mathring{\mathbb{X}}$. For a uniform distribution, $\mathbb{G} = \mathbb{D}^{-1/2} = \sqrt{M}\mathbb{I}$, which agrees with Remark 1.

In Sec. IV-B, we will need to apply the HT and the new transform after each other to the same alphabet. However, the two transforms do not commute, and the result will therefore depend on in which order the transforms are applied. Of particular interest for our analysis is the setup in Fig. 2, where $\mathbb{X}$ and $\mathbb{S}$ are related via the transform defined above. Their HTs $\tilde{\mathbb{X}}$ and $\tilde{\mathbb{S}}$ are however not related via the same transform. Instead, a relation between $\tilde{\mathbb{X}}$ and $\tilde{\mathbb{S}}$ can be established via the following theorem.

Theorem 4: If $\mathbb{S} = \mathring{\mathbb{X}}$, then their HTs $\tilde{\mathbb{S}}$ and $\tilde{\mathbb{X}}$ satisfy

$$
\tilde{\boldsymbol{s}}_i = \psi_i \sum_{j=0}^{M-1} \tilde{\boldsymbol{x}}_j \prod_{\substack{k=0 \\ n_{j,k} \neq n_{i,k}}}^{m-1} (P_{C_k}(0) - P_{C_k}(n_{j,k})),
$$
$$
i = 0, \ldots, M-1, \quad (20)
$$
$$
\tilde{\boldsymbol{x}}_j = \sum_{i=0}^{M-1} \frac{\tilde{\boldsymbol{s}}_i}{\psi_i} \prod_{\substack{k=0 \\ n_{i,k} \neq n_{j,k}}}^{m-1} (P_{C_k}(n_{i,k}) - P_{C_k}(0)),
$$
$$
j = 0, \ldots, M-1 \quad (21)
$$

where

$$
\psi_i \triangleq \prod_{\substack{k=0 \\ n_{i,k}=1}}^{m-1} 2\sqrt{P_{C_k}(0)P_{C_k}(1)}, \qquad i = 0, \ldots, M-1 \quad (22)
$$

and a product over $\varnothing$ is defined as 1.

Proof: See the Appendix. $\qquad\square$

Remark 2: The summation in (20) can be confined to $\sum_{j=i}^{M-1}$, because whenever $j < i$, there exists at least one bit position k for which $n_{i,k} \neq n_{j,k} = 0$. Analogously, the summation in (21) can be confined to $\sum_{i=j}^{M-1}$.

Example 2: Expression (20) can be written as $\tilde{\mathbb{S}} = \mathbb{T}\tilde{\mathbb{X}}$, or $\tilde{\mathbb{X}} = \mathbb{T}^{-1}\tilde{\mathbb{S}}$. The element at row i, column j of $\mathbb{T}$ and $\mathbb{T}^{-1}$ are given by (20)–(21) as, resp., $\psi_i \prod_{k:\, n_{j,k} \neq n_{i,k}} (P_{C_k}(0) - P_{C_k}(n_{j,k}))$ and $(1/\psi_j) \prod_{k:\, n_{j,k} \neq n_{i,k}} (P_{C_k}(n_{j,k}) - P_{C_k}(0))$. With $\boldsymbol{b}$, $\mathbb{X}$ and $\mathring{\mathbb{X}}$ from Example 1, we obtain $[\psi_0, \ldots, \psi_3] = [1, 0.954, 1, 0.954]$ and

$$
\mathbb{T} = \begin{bmatrix} 1 & -0.300 & 0 & 0 \\ 0 & 0.954 & 0 & 0 \\ 0 & 0 & 1 & -0.300 \\ 0 & 0 & 0 & 0.954 \end{bmatrix}, \quad \mathbb{T}^{-1} = \begin{bmatrix} 1 & 0.315 & 0 & 0 \\ 0 & 1.048 & 0 & 0 \\ 0 & 0 & 1 & 0.315 \\ 0 & 0 & 0 & 1.048 \end{bmatrix} \quad (23)
$$

which, as predicted by Remark 2, are upper triangular.

Another relation between $\tilde{\mathbb{X}}$ and $\tilde{\mathbb{S}}$ can be deduced from Fig. 2. Defining the Hadamard matrix $\mathbb{H}$ as the matrix with elements $h_{i,j}$ for $i, j = 0, \ldots, M-1$, the HT relations (7) and (11) yield $\tilde{\mathbb{S}} = (1/M)\mathbb{H}\mathbb{S}$ and $\mathbb{X} = \mathbb{H}\tilde{\mathbb{X}}$. Since from Example 1 $\mathbb{S} = \mathring{\mathbb{X}} = \mathbb{G}\mathbb{D}^{1/2}\mathbb{X}$, we conclude that $\tilde{\mathbb{S}} = (1/M)\mathbb{H}\mathbb{G}\mathbb{D}^{1/2}\mathbb{H}\tilde{\mathbb{X}}$, which implies that $\mathbb{T} = (1/M)\mathbb{H}\mathbb{G}\mathbb{D}^{1/2}\mathbb{H}$. Because $\mathbb{H}^{-1} = (1/M)\mathbb{H}$ (see (10)) and $\mathbb{G}^{-1} = (1/M)\mathbb{G}^{\mathrm{T}}$, the inverse relation is $\mathbb{T}^{-1} = (1/M^2)\mathbb{H}\mathbb{D}^{-1/2}\mathbb{G}^{\mathrm{T}}\mathbb{H}$. It is straightforward to verify that $\mathbb{T}$ and $\mathbb{T}^{-1}$ calculated in this manner, using the numerical values of $\mathbb{G}$ and $\mathbb{D}$ in Example 1, indeed yield (23).

III. BICM AT LOW SNR

A. Mutual Information

The *mutual information* (MI) in bits per channel use between the random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ for an arbitrary channel parameter $\boldsymbol{H}$ perfectly known at the receiver is defined as

$$
I(\boldsymbol{X}; \boldsymbol{Y} | \boldsymbol{H}) \triangleq \mathbb{E}\left[\log_2 \frac{p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{H}}(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{H})}{p_{\boldsymbol{Y}|\boldsymbol{H}}(\boldsymbol{Y}|\boldsymbol{H})} \right]
$$

where the expectation is taken over the joint pdf $p_{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{H}}$, and $p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{H}}$ is given by (2).

The MI between $\boldsymbol{X}$ and $\boldsymbol{Y}$ conditioned on the value of the kth bit at the input of the modulator is defined as

$$I(\boldsymbol{X}; \boldsymbol{Y} | \boldsymbol{H}, C_k) \triangleq \mathbb{E}\left[\log_2 \frac{p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{H},C_k}(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{H},C_k)}{p_{\boldsymbol{Y}|\boldsymbol{H},C_k}(\boldsymbol{Y}|\boldsymbol{H},C_k)}\right]$$

where the expectation is taken over the joint pdf $p_{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{H},C_k}$.

Definition 4 (BICM Generalized Mutual Information): The BICM generalized mutual information (BICM-GMI) is defined as [2], [17], [18], [27]

$$I(\rho) \triangleq \sum_{k=0}^{m-1} I(C_k; \boldsymbol{Y}|\boldsymbol{H})$$

$$= mI(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{H}) - \sum_{k=0}^{m-1} I(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{H}, C_k) \qquad (24)$$

where the second line follows by the chain rule. We will analyze the right-hand side of (24) as a function of ρ, for a given pdf p_H. According to (3), ρ can be varied in two ways, either by varying N_0 for a fixed constellation $[\mathbb{X}, \mathbb{P}]$ or, equivalently, by rescaling the alphabet $\mathbb{X}$ linearly for fixed N_0 and input distribution $\mathbb{P}$.

Martinez *et al.* [27] recognized the BICM decoder in Fig. 1 as a mismatched decoder and showed that the BICM-GMI in (24) corresponds to an achievable rate of such a decoder. This means that reliable transmission using a BICM system at rate R_c is possible if $R_c \leq I(\rho)$. Since from (3) $E_b^r/N_0 = \rho/R_c$, the inequality $R_c \leq I(\rho)$ gives[2]

$$\frac{E_b^r}{N_0} \geq \frac{\rho}{I(\rho)} \qquad (25)$$

for any ρ. Focusing on the wideband regime, i.e., asymptotically low SNR, we make the following definition.

Definition 5 (Low-GMI Parameters): The *low-GMI parameters* of a constellation $[\mathbb{X}, \mathbb{P}]$ are defined as $[\boldsymbol{\mu}, E_s, \alpha]$, where

$$\boldsymbol{\mu} \triangleq \mathbb{E}[\boldsymbol{X}]$$
$$E_s \triangleq \mathbb{E}[\|\boldsymbol{X}\|^2]$$
$$\alpha \triangleq \left.\frac{dI(\rho)}{d\rho}\right|_{\rho=0}.$$

In the wideband regime, the average bit energy-to-noise ratio needed for reliable transmission is, using (25) and the definition of α, lower-bounded by

$$\frac{E_b^r}{N_0} \geq \lim_{\rho \to 0^+} \frac{\rho}{I(\rho)} = \frac{1}{\alpha}. \qquad (26)$$

Furthermore, since in the wideband regime $E_b^r/N_0 \geq \log_e 2 = -1.59$ dB [15], $\alpha^{-1} \geq -1.59$ dB.

The first-order behavior of the BICM-GMI in (24) is fully determined by α, which, as we shall see later (e.g., in (36)), in turn depends on $\boldsymbol{\mu}$ and E_s. This is why we designate this triplet as low-GMI parameters. The same definitions can be applied to other MI functions $I(\rho)$ such as the coded modulation MI (CM-MI) [21]. In this paper, however, we are only interested in the BICM-GMI.

---

[2]The definition of the related function $f(R_c)$ in [21, eq. (37)] is erroneous and should read "E_b/N_0 is bounded from below by $f(R_c)/\mathbb{E}_H[H^2]$, where $f(R_c) \triangleq \mathsf{C}^{-1}(R_c)/R_c$."

The main contributions of this paper are to characterize the low-GMI parameters for arbitrary constellations, including those with nonuniform distributions (Sec. III-C), and to identify the set of constellations for BICM that maximize α, i.e., minimize E_b^r/N_0 in the wideband regime (Sec. IV-B).

B. Low-GMI Parameters for Uniform Distributions

The low-GMI parameters $[\boldsymbol{\mu}, E_s, \alpha]$ have been analyzed in detail for arbitrary input alphabets $\mathbb{X}$ under the assumption of uniform probabilities [21]. Under this assumption, they can be expressed as given by the following theorem.

Theorem 5: For a constellation $[\mathbb{X}, \mathbb{U}_m]$, the low-GMI parameters are

$$\boldsymbol{\mu} = \frac{1}{M} \sum_{i=0}^{M-1} \boldsymbol{x}_i, \qquad (27)$$

$$E_s = \frac{1}{M} \sum_{i=0}^{M-1} \|\boldsymbol{x}_i\|^2, \qquad (28)$$

$$\alpha = \frac{\log_2 e}{M^2 E_s} \sum_{k=0}^{m-1} \left\| \sum_{i=0}^{M-1} (-1)^{n_{i,k}} \boldsymbol{x}_i \right\|^2. \qquad (29)$$

Proof: Expressions (27) and (28) follow directly from Definition 5, while (29) was proved in [21, eq. (50)]. $\square$

The low-GMI parameters can be conveniently expressed as functions of the HT $\tilde{\mathbb{X}}$ of the alphabet $\mathbb{X}$, as shown in the following theorem.

Theorem 6: The low-GMI parameters can be expressed as

$$\boldsymbol{\mu} = \tilde{\boldsymbol{x}}_0, \qquad (30)$$

$$E_s = \sum_{i=0}^{M-1} \|\tilde{\boldsymbol{x}}_i\|^2, \qquad (31)$$

$$\alpha = \frac{\log_2 e}{E_s} \sum_{k=0}^{m-1} \|\tilde{\boldsymbol{x}}_{2^k}\|^2. \qquad (32)$$

Proof: The expression (30) is obtained from (9), (31) from [21, eq. (16)], and (32) from [21, Th. 11]. $\square$

C. Low-GMI Parameters for Nonuniform Distributions

The next theorem is analogous to Theorem 5 but applies to an arbitrary input distribution.

Theorem 7: For a constellation $[\mathbb{X}, \mathbb{P}]$, the low-GMI parameters are

$$\boldsymbol{\mu} = \sum_{i=0}^{M-1} P_i \boldsymbol{x}_i, \qquad (33)$$

$$E_s = \sum_{i=0}^{M-1} P_i \|\boldsymbol{x}_i\|^2, \qquad (34)$$

$$\alpha = \frac{\log_2 e}{E_s} \sum_{i=0}^{M-1} P_i \sum_{j=0}^{M-1} P_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$$
$$\cdot \sum_{k=0}^{m-1} (-1)^{n_{i,k}+n_{j,k}} \frac{P_{C_k}(\bar{n}_{i,k})}{P_{C_k}(n_{j,k})}. \qquad (35)$$

TABLE I

LOW-GMI PARAMETERS AND FOO CONDITIONS FOR BICM USING UNIFORM AND NONUNIFORM INPUT DISTRIBUTIONS. THE RESULTS FOR $[\mathbb{X}, \mathbb{U}_m]$ ARE FROM [21] (CF. THEOREMS 6 AND 9) AND THE ONES FOR $[\mathbb{X}, \mathbb{P}]$ OR $[\mathring{\mathbb{X}}, \mathbb{U}_m]$ ARE FROM THEOREMS 7, 8, AND 11.

	$[\mathbb{X}, \mathbb{U}_m]$	$[\mathbb{X}, \mathbb{P}]$ or $[\mathring{\mathbb{X}}, \mathbb{U}_m]$
$\boldsymbol{\mu}$	$\tilde{\boldsymbol{x}}_0$	$\displaystyle\sum_{i=0}^{M-1} P_i \boldsymbol{x}_i$
E_s	$\displaystyle\sum_{i=0}^{M-1} \|\tilde{\boldsymbol{x}}_i\|^2$	$\displaystyle\sum_{i=0}^{M-1} P_i \|\boldsymbol{x}_i\|^2$
α	$\dfrac{\log_2 \mathrm{e}}{E_\mathrm{s}} \displaystyle\sum_{k=0}^{m-1} \|\tilde{\boldsymbol{x}}_{2^k}\|^2$	$\dfrac{\log_2 \mathrm{e}}{E_\mathrm{s}} \displaystyle\sum_{i=0}^{M-1} P_i \sum_{j=0}^{M-1} P_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \sum_{k=0}^{m-1} (-1)^{n_{i,k}+n_{j,k}} \dfrac{P_{C_k}(\bar{n}_{i,k})}{P_{C_k}(n_{j,k})}$
FOO Condition	$\tilde{\boldsymbol{x}}_j = \boldsymbol{0}, \ \forall j \notin \{1, 2, 4, \ldots, M/2\}$	$\boldsymbol{\mu} = \boldsymbol{0}$ and $\tilde{\boldsymbol{x}}_j = \boldsymbol{0}, \ \forall j \notin \{0\} \cup \{1, 2, 4, \ldots, M/2\}$

Proof: Again, (33) and (34) follow from Definition 5, while (35) requires some analysis. It was shown in [21, Th. 10] that

$$\alpha = \frac{\log_2 \mathrm{e}}{2E_\mathrm{s}} \sum_{k=0}^{m-1} \left[\left\| \sum_{i=0}^{M-1} (-1)^{n_{i,k}} \frac{P_i \boldsymbol{x}_i}{\sqrt{P_{C_k}(n_{i,k})}} \right\|^2 + \left\| \sum_{i=0}^{M-1} \frac{P_i \boldsymbol{x}_i}{\sqrt{P_{C_k}(n_{i,k})}} \right\|^2 - 2\|\boldsymbol{\mu}\|^2 \right]. \quad (36)$$

Substituting (33) and writing the squared norms as the inner products of two identical vectors yields

$$\alpha = \frac{\log_2 \mathrm{e}}{2E_\mathrm{s}} \sum_{k=0}^{m-1} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P_i P_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$$
$$\cdot \left[\frac{(-1)^{n_{i,k}}(-1)^{n_{j,k}}}{\sqrt{P_{C_k}(n_{i,k}) P_{C_k}(n_{j,k})}} + \frac{1}{\sqrt{P_{C_k}(n_{i,k}) P_{C_k}(n_{j,k})}} - 2 \right].$$

The expression in brackets can be simplified as

$$\frac{(-1)^{n_{i,k}+n_{j,k}} + 1}{\sqrt{P_{C_k}(n_{i,k}) P_{C_k}(n_{j,k})}} - 2 = \begin{cases} 2\frac{P_{C_k}(\bar{n}_{j,k})}{P_{C_k}(n_{j,k})}, & n_{j,k} = n_{i,k} \\ -2, & n_{j,k} \neq n_{i,k} \end{cases}$$
$$= 2(-1)^{n_{i,k}+n_{j,k}} \frac{P_{C_k}(\bar{n}_{i,k})}{P_{C_k}(n_{j,k})}$$

which completes the proof of (35). $\quad\square$

Theorem 7 shows that the low-GMI parameters depend on the input alphabet $\mathbb{X}$, the binary labeling (via $n_{i,k}$ in the expression for α), and the input distribution (via $P_{C_k}(u)$ and P_i). While the low-GMI parameters of an alphabet $\mathbb{X}$ with uniform probabilities are conveniently expressed in terms of its HT $\tilde{\mathbb{X}}$ (cf. Theorem 6), no similar expressions are known for the low-GMI parameters of a general constellation in (33)–(35). This has so far prevented the analytic optimization of such constellations. The new transform introduced in Section II-E, however, solves this problem by establishing an equivalence between an arbitrary constellation, possibly with nonuniform probabilities, and another constellation with uniform probabilities.

Theorem 8: The low-GMI parameters $[\boldsymbol{\mu}, E_\mathrm{s}, \alpha]$ of any constellation $[\mathbb{X}, \mathbb{P}]$ are equal to the low-GMI parameters of $[\mathring{\mathbb{X}}, \mathbb{U}_m]$.

Proof: Let the low-GMI parameters of $[\mathring{\mathbb{X}}, \mathbb{U}_m]$ be denoted by $[\boldsymbol{\mu}', E_\mathrm{s}', \alpha']$. First, (27) and (12) yield

$$\boldsymbol{\mu}' = \frac{1}{M} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \boldsymbol{x}_j g_{i,j} \sqrt{P_j}$$
$$= \frac{1}{M} \sum_{j=0}^{M-1} \boldsymbol{x}_j \sqrt{P_j} \sum_{i=0}^{M-1} g_{i,j}. \quad (37)$$

Applying (16) to the inner sum in (37) reveals that

$$\boldsymbol{\mu}' = \sum_{j=0}^{M-1} P_j \boldsymbol{x}_j = \boldsymbol{\mu}.$$

Second, (28) and (12) yield

$$E_\mathrm{s}' = \frac{1}{M} \sum_{i=0}^{M-1} \|\boldsymbol{x}_i\|^2$$
$$= \frac{1}{M} \sum_{i=0}^{M-1} \left\langle \sum_{j=0}^{M-1} \boldsymbol{x}_j g_{i,j} \sqrt{P_j}, \sum_{l=0}^{M-1} \boldsymbol{x}_l g_{i,l} \sqrt{P_l} \right\rangle$$
$$= \frac{1}{M} \sum_{j=0}^{M-1} \sum_{l=0}^{M-1} \langle \boldsymbol{x}_j, \boldsymbol{x}_l \rangle \sqrt{P_j P_l} \sum_{i=0}^{M-1} g_{i,j} g_{i,l}.$$

Evaluating the inner sum using (14) gives

$$E_\mathrm{s}' = \sum_{j=0}^{M-1} P_j \|\boldsymbol{x}_j\|^2 = E_\mathrm{s}.$$

For the third and last part of the theorem, (29) yields

$$\alpha' = \frac{\log_2 \mathrm{e}}{M^2 E_\mathrm{s}} \sum_{k=0}^{m-1} \left\| \sum_{i=0}^{M-1} (-1)^{n_{i,k}} \mathring{\boldsymbol{x}}_i \right\|^2 \quad (38)$$

where $\mathring{\boldsymbol{x}}_i$ is given by (12). The inner sum can be expanded as

$$\sum_{i=0}^{M-1} (-1)^{n_{i,k}} \mathring{\boldsymbol{x}}_i = \sum_{i=0}^{M-1} (-1)^{n_{i,k}} \sum_{j=0}^{M-1} \boldsymbol{x}_j g_{i,j} \sqrt{P_j}$$
$$= \sum_{j=0}^{M-1} \boldsymbol{x}_j \sqrt{P_j} \sum_{i=0}^{M-1} (-1)^{n_{i,k}} g_{i,j}.$$

Applying (17) to the inner sum, we obtain

$$\sum_{i=0}^{M-1} (-1)^{n_{i,k}} \mathring{\boldsymbol{x}}_i = M \sum_{j=0}^{M-1} \boldsymbol{x}_j P_j (-1)^{n_{j,k}} \sqrt{\frac{P_{C_k}(\bar{n}_{j,k})}{P_{C_k}(n_{j,k})}}.$$
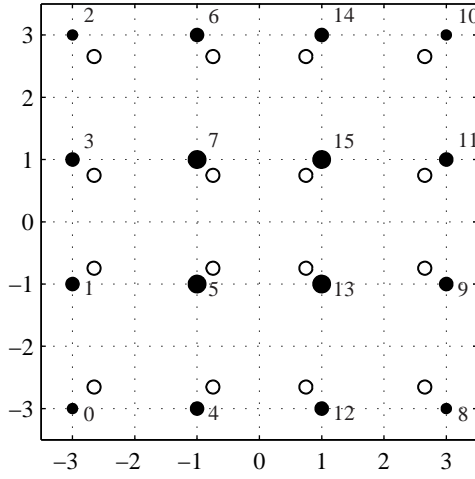
Fig. 3. 16-QAM constellation $[\mathbb{X}_{QAM}, \mathbb{P}_1]$ (black circles) with bit probabilities $\boldsymbol{b}_1 = [0.35, 0.50, 0.35, 0.50]$. Each symbol $\boldsymbol{x}_j$ is marked with its index j, and its probability P_j is proportional to the area of the corresponding circle. White circles represent the transformed constellation $[\mathring{\mathbb{X}}_{QAM}, \mathbb{U}_4]$, which has the same low-GMI parameters.
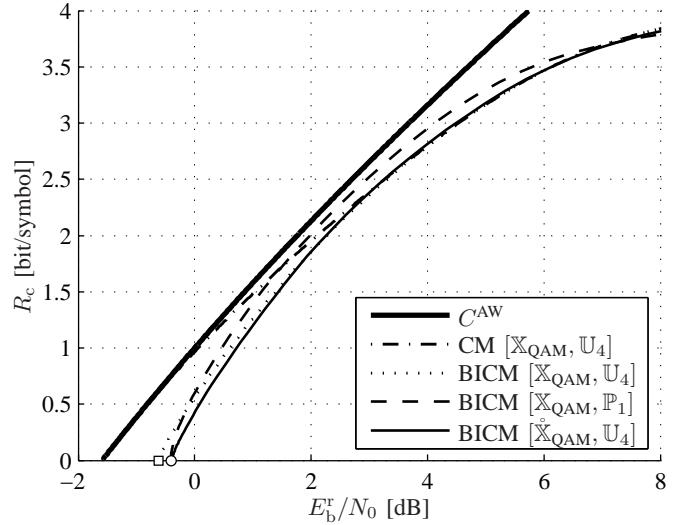

Fig. 4. BICM-GMI for probabilistically shaped 16-QAM and its transform. The CM-MI and BICM-GMI for uniform input distributions are also shown, as is the AWGN capacity C^{AW}. The white circle and square indicate (26) with α given by (41) and by [21, eq. (55)], resp.

We take the inner product of this vector with itself and substitute the obtained expression for the squared norm in (38). This yields, after rearranging terms,

$$\alpha' = \frac{\log_2 e}{E_s} \sum_{k=0}^{m-1} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$$

$$\cdot P_i P_j (-1)^{n_{i,k}+n_{j,k}} \sqrt{\frac{P_{C_k}(\bar{n}_{i,k})}{P_{C_k}(n_{i,k})} \frac{P_{C_k}(\bar{n}_{j,k})}{P_{C_k}(n_{j,k})}} \quad (39)$$

$$= \alpha. \quad (40)$$

The square root in (39) is $P_{C_k}(\bar{n}_{i,k})/P_{C_k}(n_{i,k})$ if $n_{j,k} = n_{i,k}$ or 1 if $n_{j,k} = \bar{n}_{i,k}$. In both cases, it can be expressed as $P_{C_k}(\bar{n}_{i,k})/P_{C_k}(n_{j,k})$ (or, equivalently, $P_{C_k}(\bar{n}_{j,k})/P_{C_k}(n_{i,k})$). Comparing this result with (35) gives (40). □

Theorem 8 shows that the constellation $[\mathbb{X}, \mathbb{P}]$ can be mapped to another constellation $[\mathring{\mathbb{X}}, \mathbb{U}_m]$ with the same low-GMI parameters, where $\mathring{\mathbb{X}}$ is related to $\mathbb{X}$ via (12) and (18). This relation between $[\mathbb{X}, \mathbb{P}]$ and $[\mathring{\mathbb{X}}, \mathbb{U}_m]$ will be applied in Sec. IV-B to prove Theorem 10, which is the main result of the paper. To summarize, Table I lists the low-GMI parameters for BICM given by Theorems 6 and 7. The equivalence of the parameters for $[\mathbb{X}, \mathbb{P}]$ or $[\mathring{\mathbb{X}}, \mathbb{U}_m]$ comes from Theorem 8.

D. Numerical Examples

In this Section, we show examples of how the transform defined in Sec. II-E works and we also present equivalent constellations $[\mathbb{X}, \mathbb{P}]$ and $[\mathring{\mathbb{X}}, \mathbb{U}_m]$. All results are for the AWGN channel. The (G)MIs are numerically evaluated using Gauss–Hermite quadratures following [28, Sec. III].

Example 3: Consider the equally spaced 16-ary square quadrature amplitude modulation (16-QAM) alphabet $\mathbb{X} = \mathbb{X}_{QAM}$ labeled by the binary reflected Gray code (BRGC) [29] with bit probabilities $\boldsymbol{b}_1 = [0.35, 0.5, 0.35, 0.5]$, shown with black circles in Fig. 3. The input distribution $\mathbb{P} = \mathbb{P}_1$

given by (6) is $P_0 = P_2 = P_8 = P_{10} = 0.031$, $P_1 = P_3 = P_4 = P_6 = P_9 = P_{11} = P_{12} = P_{14} = 0.057$, and $P_5 = P_7 = P_{13} = P_{15} = 0.106$. These symbol probabilities are indicated in Fig. 3, where the area of the circle representing $\boldsymbol{x}_j$ is proportional to the corresponding probability P_j.

Another alphabet $\mathring{\mathbb{X}}_{QAM}$ is obtained by applying the transform in (12) to the constellation $[\mathbb{X}_{QAM}, \mathbb{P}_1]$. The white circles in Fig. 3 represent the symbols in $\mathring{\mathbb{X}}_{QAM}$ using a uniform distribution $\mathbb{U}_4$. The alphabet $\mathring{\mathbb{X}}_{QAM}$ is still a rectangular 16-QAM constellation, but a nonuniformly spaced one. Every row in Fig. 3 can be regarded as a probabilistically shaped (black) or a geometrically shaped (white) 4-PAM constellation; in fact the same 4-PAM constellations as in Example 1.

The low-GMI parameters of the two constellations $[\mathbb{X}_{QAM}, \mathbb{P}_1]$ and $[\mathring{\mathbb{X}}_{QAM}, \mathbb{U}_4]$, given by Theorems 7 and 5, resp., are identical, as predicted by Theorem 8. These are

$$\boldsymbol{\mu} = \boldsymbol{0}, \qquad E_s = 7.60, \qquad \alpha = 1.10. \quad (41)$$

The BICM-GMI for the constellations $[\mathbb{X}_{QAM}, \mathbb{P}_1]$ and $[\mathring{\mathbb{X}}_{QAM}, \mathbb{U}_4]$ are shown in Fig. 4. In this figure, we also show the capacity of the AWGN channel C^{AW} [21, eq. (22)] and the CM-MI and BICM-GMI for 16-QAM using a uniform input distribution, i.e., $[\mathbb{X}_{QAM}, \mathbb{U}_4]$. The results show that the BICM-GMIs of the original constellation $[\mathbb{X}_{QAM}, \mathbb{P}_1]$ and the transformed constellation $[\mathring{\mathbb{X}}_{QAM}, \mathbb{U}_4]$ are in general different; however, they converge in the low-SNR regime. The endpoints of the BICM-GMI curves for $[\mathbb{X}_{QAM}, \mathbb{P}_1]$ and $[\mathring{\mathbb{X}}_{QAM}, \mathbb{U}_4]$ are shown with a white circle, whose value follows from (41) and (26). The endpoint for the BICM-GMI curve for the constellation $[\mathbb{X}_{QAM}, \mathbb{U}_4]$ is shown with a white square [17, eq. (18)].

Example 4: Consider the NBC-labeled M-ary phase-shift keying (PSK) alphabet $\mathbb{X} = \mathbb{X}_{PSK}$, where $\boldsymbol{x}_j = [\cos(2\pi j/M + \pi/M), \sin(2\pi j/M + \pi/M)]$ with $j = 0, \ldots, M - 1$. The constellation for $M = 8$ for two input distributions $\mathbb{P}$ are
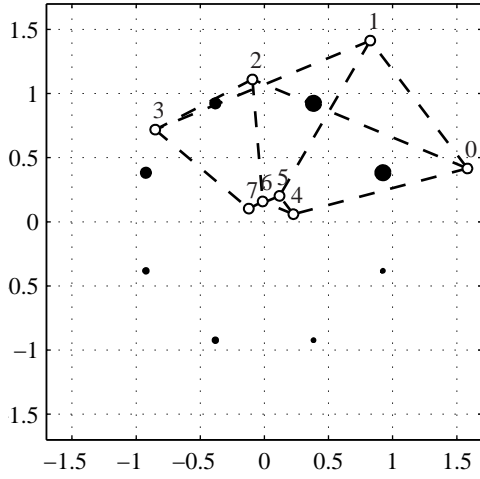
Fig. 5. 8-PSK constellation $[\mathbb{X}_{PSK}, \mathbb{P}_2]$ with bit probabilities $\boldsymbol{b}_2$ (black circles) and the transformed constellation $[\mathring{\mathbb{X}}_{PSK}, \mathbb{U}_3]$ (white circles), which have the same low-GMI parameters. The circle areas is proportional to the symbol probabilities P_j. Dashed lines join symbols whose labels differ in one bit only.
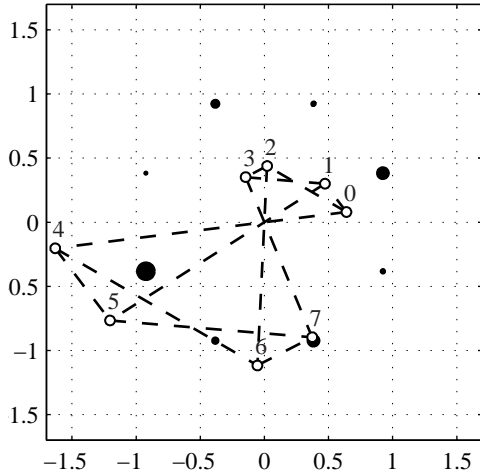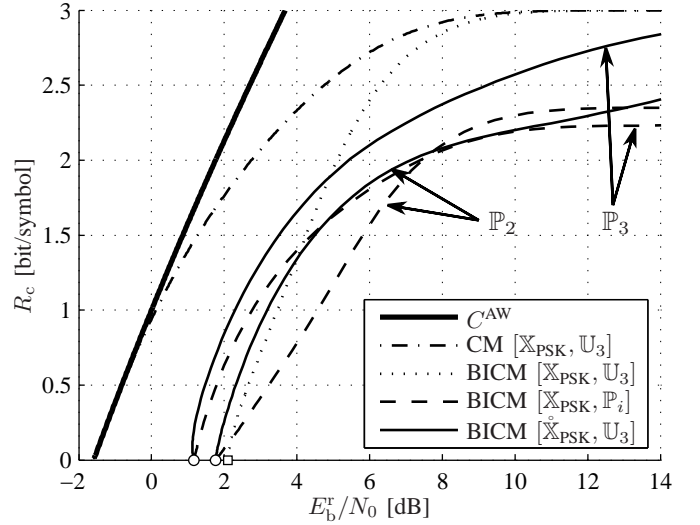

Fig. 7. BICM-GMI of the two probabilistically shaped 8-PSK constellations in Figs. 5 and 6 and their transforms $[\mathring{\mathbb{X}}_{PSK}, \mathbb{U}_3]$. The CM-MI and BICM-GMI for uniform input distributions are also shown. The white circles and square show the endpoints α^{-1}.
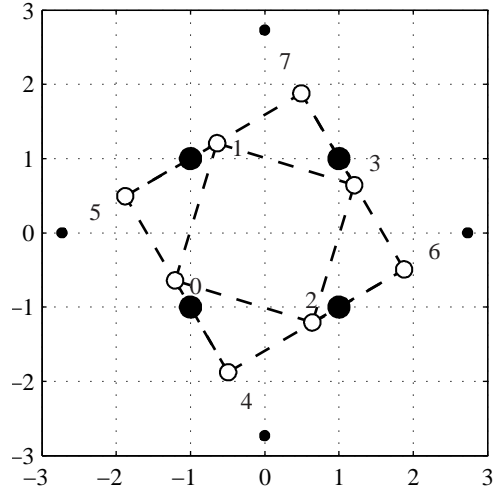

Fig. 6. 8-PSK constellation $[\mathbb{X}_{PSK}, \mathbb{P}_3]$ with bit probabilities $\boldsymbol{b}_3$ (black circles) and its transform $[\mathring{\mathbb{X}}_{PSK}, \mathbb{U}_3]$ (white circles).

Fig. 8. Star-shaped 8-QAM constellation $[\mathbb{X}_{8QAM}, \mathbb{P}_4]$ with bit probabilities $\boldsymbol{b}_4$ (black circles) and its transform $[\mathring{\mathbb{X}}_{8QAM}, \mathbb{U}_3]$ (white circles).

shown in Figs. 5 and 6, where again the circle areas are proportional to the symbol probabilities P_j. We denote these input distributions by $\mathbb{P}_2$ and $\mathbb{P}_3$, which are generated by $\boldsymbol{b}_2 = [0.5, 0.7, 0.9]$ and $\boldsymbol{b}_3 = [0.9, 0.7, 0.3]$, resp. The transforms $\mathring{\mathbb{X}}_{PSK}$ are irregular, not resembling a PSK alphabet. Nevertheless, the low-GMI parameters for pairs of constellations $[\mathbb{X}_{PSK}, \mathbb{P}_i]$ and $[\mathring{\mathbb{X}}_{PSK}, \mathbb{U}_3]$ are again equal. Particularly, $\alpha = 0.67$ for $[\mathbb{X}_{PSK}, \mathbb{P}_2]$ and $\alpha = 0.76$ for $[\mathbb{X}_{PSK}, \mathbb{P}_3]$.

The BICM-GMI for the two 8PSK constellations in Figs. 5 and 6 is shown in Fig. 7. Also shown are the BICM-GMI and CM-MI for 8-PSK with uniform input distributions, for which $\alpha = 0.62$ [21, eq. (56)] and $\alpha = \log_2 e$, resp., and the capacity of the AWGN channel. Again, the results show that the BICM-GMIs of the original constellation $[\mathbb{X}_{PSK}, \mathbb{P}_i]$ and the transformed constellation $[\mathring{\mathbb{X}}_{PSK}, \mathbb{U}_3]$ are in general different but converge in the low-SNR regime. The endpoints

of the BICM-GMI curves are obtained from (26).

Example 5: Consider the eight-level star-shaped QAM alphabet shown in Fig. 8 (black circles), which we denote by $\mathbb{X}_{8QAM}$. This alphabet is used with bit probabilities $\boldsymbol{b}_4 = [0.5, 0.5, 0.85]$, giving an input symbol probability $\mathbb{P}_4$. In this figure we also show the transformed constellation $[\mathring{\mathbb{X}}_{8QAM}, \mathbb{U}_3]$, which according to Theorem 8 has the same low-GMI parameters as $[\mathbb{X}_{8QAM}, \mathbb{P}_4]$. This can be appreciated in Fig. 9, where the corresponding BICM-GMIs are shown. Fig. 9 also shows how probabilistic shaping improves the BICM-GMI considerably over a wide range of SNRs.

The results in Figs. 4, 7, and 9 also show other interesting properties of probabilistic shaping for BICM. In the high-SNR regime, the use of a nonuniform distribution results in a loss in GMI, i.e., the curves flatten out at a value below m [bit/symbol], but for a wide range of moderately high SNR,
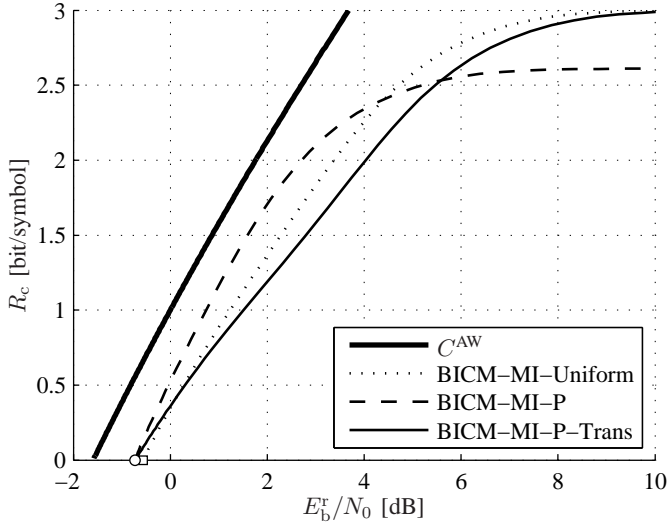
Fig. 9. BICM-GMI for $[\mathbb{X}_{8\text{QAM}}, \mathbb{U}_3]$, $[\mathbb{X}_{8\text{QAM}}, \mathbb{P}_4]$ and $[\mathring{\mathbb{X}}_{8\text{QAM}}, \mathbb{U}_3]$. The AWGN capacity is also shown. The white circle and square show the endpoints α^{-1} with $\alpha = 1.14$ and 1.18.

the BICM-GMI is higher with probabilistic shaping. For the 16-QAM alphabet in Example 3, in the medium SNR regime, the use of nonequally likely symbols even gives a larger GMI than the CM-MI obtained with and a uniform distribution.

IV. FIRST-ORDER OPTIMAL CONSTELLATIONS

Having characterized the low-SNR behavior of the BICM-GMI of an arbitrary constellation, the next step is to search for optimal constellations in terms of the BICM-GMI at low SNR. The following definition formally defines BICM systems that achieve the SL.

Definition 6 (FOO constellation): The constellation $[\mathbb{X}, \mathbb{P}]$ is said to be first-order optimal (FOO) if a BICM system using $[\mathbb{X}, \mathbb{P}]$ achieves the SL -1.59 dB, i.e., $\alpha = \log_2 e$.

As discussed in Sec. II-C, we have fixed the labeling to be the NBC, and thus, an FOO constellation is fully characterized by only two parameters, the input alphabet $\mathbb{X}$ and its input distribution $\mathbb{P}$, where the latter satisfies (6). The analysis is nevertheless, without loss of generality, applicable to an arbitrary labeling by permuting the constellation, see Sec. II-C.

A. FOO Constellations for Uniform Distributions

In this section we review results on FOO constellations for BICM for uniform input distributions. The next theorem gives necessary and sufficient conditions for an input alphabet to be FOO if the binary labeling is the NBC and input distribution is uniform.

Theorem 9: The constellation $[\mathbb{X}, \mathbb{U}_m]$ is FOO if and only if

$$\tilde{\boldsymbol{x}}_j = \boldsymbol{0}, \qquad \forall j \notin \{1, 2, 4, \ldots, M/2\}. \tag{42}$$

Proof: From (31) and (32), $\alpha = \log_2 e$ if and only if

$$\sum_{i=0}^{M-1} \|\tilde{\boldsymbol{x}}_i\|^2 = \sum_{k=0}^{m-1} \|\tilde{\boldsymbol{x}}_{2^k}\|^2$$

which gives (42). $\qquad\square$

This theorem was given in [21, Th. 12], where it was used to find FOO constellations for BICM when $\mathbb{P} = \mathbb{U}_m$. It offers an appealing intuitive geometrical interpretation: An input alphabet is FOO for a uniform input distribution if and only if it is a zero-mean linear projection of a hypercube. This behavior is illustrated in Example 6 and also in [21, Fig. 4].

B. FOO Constellations for Nonuniform Distributions

In this section, we derive necessary and sufficient conditions for a BICM system, with an arbitrary input alphabet and probability distribution, to achieve the SL, i.e., we find FOO constellations for BICM. The conditions are derived by transforming an arbitrary constellation into another constellation with uniform probabilities using Theorem 8 and applying Theorem 9 to this transformed constellation. Since Theorem 9 is expressed in terms of the HT, a relation between the HTs of $\mathbb{X}$ and $\mathring{\mathbb{X}}$ is needed, which is illustrated by the bottom arrow in Fig. 2. Such a relation is provided by Theorem 4 and will be applied in the proofs of Theorems 10 and 11.

Theorem 10: The constellation $[\mathbb{X}, \mathbb{P}]$ is FOO if and only if the HT $\tilde{\mathring{\mathbb{X}}}$ of $\mathring{\mathbb{X}}$ satisfies both the following conditions:

$$\tilde{\boldsymbol{x}}_0 = \sum_{k=0}^{m-1} \tilde{\boldsymbol{x}}_{2^k} \left(P_{C_k}(1) - P_{C_k}(0) \right) \tag{43}$$

$$\tilde{\boldsymbol{x}}_j = \boldsymbol{0}, \qquad \forall j \notin \{0\} \cup \{1, 2, 4, \ldots, M/2\}. \tag{44}$$

Proof: We will prove the theorem in two steps. First, we prove the "if" part by showing that (43) and (44) imply that $[\mathbb{X}, \mathbb{P}]$ is FOO. Second, we prove the "only if" part by showing that if $[\mathbb{X}, \mathbb{P}]$ is FOO, then (43) and (44) hold.

For the "if" part, suppose that (43) and (44) hold for a given constellation $[\mathbb{X}, \mathbb{P}]$. Applying (44) in (20) yields for the HT $\tilde{\mathbb{S}}$ of $\mathbb{S} = \mathring{\mathbb{X}}$

$$\tilde{\boldsymbol{s}}_j = \psi_j \Bigg[\tilde{\boldsymbol{x}}_0 \prod_{\substack{k=0 \\ n_{0,k} \neq n_{j,k}}}^{m-1} \left(P_{C_k}(0) - P_{C_k}(n_{0,k}) \right)$$
$$+ \sum_{l=0}^{m-1} \tilde{\boldsymbol{x}}_{2^l} \prod_{\substack{k=0 \\ n_{2^l,k} \neq n_{j,k}}}^{m-1} \left(P_{C_k}(0) - P_{C_k}(n_{2^l,k}) \right) \Bigg],$$
$$j = 0, \ldots, M - 1. \tag{45}$$

Since the bits $n_{0,k} = 0$ for $k = 0, \ldots, m-1$, the first product in (45) is always zero, except when $j = 0$. (Recall that a product over $\varnothing$ in (20) is defined as 1.) Furthermore, because of (4), the second product in (45) is zero whenever $j \notin \{0, 2^l\}$ for some integer l. We can therefore identify three cases for (45). First,

$$\tilde{\boldsymbol{s}}_j = \boldsymbol{0}, \quad j \notin \{0\} \cup \{1, 2, \ldots, 2^{m-1}\}. \tag{46}$$

Second, $j = 0$ yields

$$\tilde{\boldsymbol{s}}_0 = \psi_0 \Bigg[\tilde{\boldsymbol{x}}_0 + \sum_{l=0}^{m-1} \tilde{\boldsymbol{x}}_{2^l} \left(P_{C_l}(0) - P_{C_l}(n_{2^l,l}) \right) \Bigg]$$
$$= \boldsymbol{0} \tag{47}$$

because of (43). And third, letting $j = 2^i$ for an integer i,

$$\tilde{\boldsymbol{s}}_{2^i} = \psi_{2^i} \sum_{l=0}^{m-1} \tilde{\boldsymbol{x}}_{2^l} \prod_{\substack{k=0 \\ n_{2^i,k} \neq n_{2^l,k}}}^{m-1} \left(P_{C_k}(0) - P_{C_k}(n_{2^l,k}) \right),$$
$$i = 0, \ldots, m-1. \quad (48)$$

When $l \neq i$, the product in (48) includes two factors, $k = l$ and $k = i$. For $k = i$, $n_{2^l,k} = 0$ and the whole product is 0. Therefore, only $l = i$ contributes to the sum in (48). When $l = i$, the product in (48) is again over $\varnothing$ and (48) becomes

$$\tilde{\boldsymbol{s}}_{2^i} = \psi_{2^i} \tilde{\boldsymbol{x}}_{2^i}. \quad (49)$$

Combining the three cases (46), (47), and (49) yields

$$\tilde{\boldsymbol{s}}_j = \begin{cases} \psi_j \tilde{\boldsymbol{x}}_j, & j \in \{1, 2, \ldots, 2^{m-1}\}, \\ \boldsymbol{0}, & \text{otherwise.} \end{cases}$$

Using Theorem 9, we conclude that the constellation $[\mathring{\mathbb{X}}, \mathbb{U}_m]$ is FOO. Finally, Theorem 8 implies that $[\mathbb{X}, \mathbb{P}]$ is also FOO, which completes the proof of the "if" part.

For the "only if" part, assume that $[\mathbb{X}, \mathbb{P}]$ is FOO. By Theorem 8, $[\mathring{\mathbb{X}}, \mathbb{U}_m]$ is also FOO, and by Theorem 9, $\tilde{\boldsymbol{s}}_i = \boldsymbol{0}$ for any i that is not a power of two, where $\tilde{\mathbb{S}}$ is the HT of $\mathbb{S} = \mathring{\mathbb{X}}$. We will now use Theorem 4 to translate the condition on $\tilde{\mathbb{S}}$ into conditions on $\mathring{\mathbb{X}}$.

If $\tilde{\boldsymbol{s}}_i = \boldsymbol{0}$ for $i \notin \{1, 2, \ldots, 2^{m-1}\}$, then the summation over $i = 0, \ldots, M-1$ in (21) can be reduced to a summation over $i = 2^l$ for $l = 0, \ldots, m-1$,

$$\tilde{\boldsymbol{x}}_j = \sum_{l=0}^{m-1} \frac{\tilde{\boldsymbol{s}}_{2^l}}{\psi_{2^l}} \prod_{\substack{k=0 \\ n_{j,k} \neq n_{2^l,k}}}^{m-1} \left(P_{C_k}(n_{2^l,k}) - P_{C_k}(0) \right),$$
$$j = 0, \ldots, M-1. \quad (50)$$

Due to (4), the product in (50) is nonzero only if the product is over $k \in \varnothing$ or over the single-element set $k \in \{l\}$, i.e., if $j = 0$ or $j = 2^l$ for some integer l, resp. Again, three cases can be identified. First, if $j \notin \{0\} \cup \{1, 2, \ldots, 2^{m-1}\}$, then the product in (50) includes at least one $k \neq l$ for every l and

$$\tilde{\boldsymbol{x}}_j = \boldsymbol{0}, \qquad j \notin \{0\} \cup \{1, 2, \ldots, 2^{m-1}\}. \quad (51)$$

Second, for $j = 0$, the product comprises only one factor, $k = l$, and

$$\tilde{\boldsymbol{x}}_0 = \sum_{l=0}^{m-1} \frac{\tilde{\boldsymbol{s}}_{2^l}}{\psi_{2^l}} \left(P_{C_l}(1) - P_{C_l}(0) \right). \quad (52)$$

And third, setting $j = 2^i$,

$$\tilde{\boldsymbol{x}}_{2^i} = \sum_{l=0}^{m-1} \frac{\tilde{\boldsymbol{s}}_{2^l}}{\psi_{2^l}} \prod_{\substack{k=0 \\ n_{2^i,k} \neq n_{2^l,k}}}^{m-1} \left(P_{C_k}(n_{2^l,k}) - P_{C_k}(0) \right),$$
$$i = 0, \ldots, m-1. \quad (53)$$

As explained after (48), the product in (53) is 1 if $l = i$ (product over $\varnothing$) and 0 otherwise. Thus, the summation in (53) can be reduced to just one term, $l = i$, and

$$\tilde{\boldsymbol{x}}_{2^i} = \frac{\tilde{\boldsymbol{s}}_{2^i}}{\psi_{2^i}}, \qquad i = 0, \ldots, m-1. \quad (54)$$

Combining the three cases (51), (52), and (54) yields

$$\tilde{\boldsymbol{x}}_j = \begin{cases} \sum_{l=0}^{m-1} \tilde{\boldsymbol{s}}_{2^l} \frac{P_{C_l}(1) - P_{C_l}(0)}{\psi_{2^l}}, & j = 0, \\ \frac{\tilde{\boldsymbol{s}}_j}{\psi_j}, & j \in \{1, 2, \ldots, 2^{m-1}\} \\ \boldsymbol{0}, & \text{otherwise} \end{cases}$$

which satisfies (43) and (44). This completes the proof of the "only if" part. $\square$

Remark 3: Only (43) depends on the input distribution, not (44). In view of Theorem 9, the only difference between FOO constellations with uniform and nonuniform distributions lies in $\tilde{\boldsymbol{x}}_0$. The final theorem gives this statement a more intuitive interpretation.

Theorem 11: The constellation $[\mathbb{X}, \mathbb{P}]$ is FOO if and only if both the following conditions hold:

$$\boldsymbol{\mu} = \boldsymbol{0} \quad (55)$$
$$\tilde{\boldsymbol{x}}_j = \boldsymbol{0}, \qquad \forall j \notin \{0\} \cup \{1, 2, 4, \ldots, M/2\}. \quad (56)$$

Proof: We wish to prove that if (44) (or equivalently (56)) holds, then (43) and (55) are equivalent.

For any constellation $[\mathbb{X}, \mathbb{P}]$, the mean $\boldsymbol{\mu} = \tilde{\boldsymbol{s}}_0$, where $\tilde{\mathbb{S}}$ is the HT of $\mathbb{S} = \mathring{\mathbb{X}}$. This follows from Theorem 8 and (30). The mean $\boldsymbol{\mu}$ can be calculated by letting $i = 0$ in (20) and (22) as

$$\boldsymbol{\mu} = \psi_0 \sum_{i=0}^{M-1} \tilde{\boldsymbol{x}}_i \prod_{\substack{k=0 \\ n_{i,k}=1}}^{m-1} \left(P_{C_k}(0) - P_{C_k}(1) \right) \quad (57)$$

where $\psi_0 = 1$.

In this theorem, we are only interested in constellations that satisfy (44) or equivalently (56). For such constellations, the sum in (57) includes at most $m+1$ nonzero terms, namely,

$$\boldsymbol{\mu} = \tilde{\boldsymbol{x}}_0 + \sum_{l=0}^{m-1} \tilde{\boldsymbol{x}}_{2^l} \left(P_{C_l}(0) - P_{C_l}(1) \right).$$

This expression makes (43) and (55) equivalent. $\square$

Based on Theorem 9, the result in Theorem 11 can be understood as follows. If a constellation with a uniform input distribution is FOO, it will still be FOO for any other input distribution $\boldsymbol{b}$ provided that the input alphabet is translated to be zero mean. In view of the geometrical interpretation of Theorem 9 given in [21, Th. 12], the result in Theorem 11 also states that a constellation is FOO if and only if its input alphabet is a linear projection of a hypercube *and* it has zero mean.

We also note that the zero-mean condition in Theorem 11 is the same that guarantees FOO for the CM-MI [21, Footnote 12][3]. This implies that the only difference between FOO constellations for the CM-MI and the BICM-GMI lies on the extra constraint on the input alphabet to be a linear projection of a hypercube.

C. Numerical Examples

In this subsection we give numerical examples to illustrate the analytical results presented in this paper.

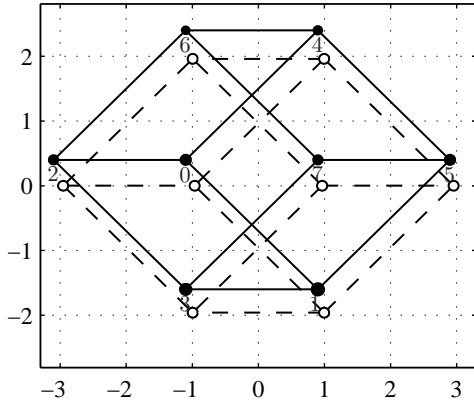[3]The parameter α for the CM-MI is [21, Th. 7] $\alpha = \log_2 \mathrm{e}(1 - \|\boldsymbol{\mu}\|^2 / E_\mathrm{s})$.

Fig. 10. The FOO constellations $[\mathbb{X}'_{8AMPM}, \mathbb{P}_5]$ (black circles) and $[\mathring{\mathbb{X}}'_{8AMPM}, \mathbb{U}_3]$ (white circles).
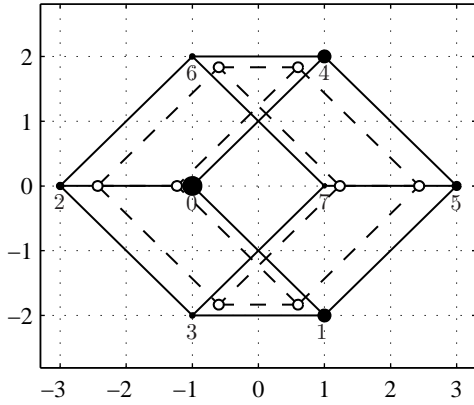

Fig. 11. The constellations $[\mathbb{X}_{8AMPM}, \mathbb{P}_6]$ (black circles) and $[\mathring{\mathbb{X}}_{8AMPM}, \mathbb{U}_3]$ (white circles). They both have zero mean and look like cubes, which indicates that they are FOO.

Example 6: Consider the so-called 8-AMPM alphabet [30]

$$\mathbb{X}_{8AMPM} = \begin{bmatrix} -1 & 1 & -3 & -1 & 1 & 3 & -1 & 1 \\ 0 & -2 & 0 & -2 & 2 & 0 & 2 & 0 \end{bmatrix}^{\mathrm{T}} \quad (58)$$

which corresponds to a projected hypercube. The constellation $[\mathbb{X}_{8AMPM}, \mathbb{P}]$ was shown to be FOO for $\mathbb{P} = \mathbb{U}_3$ in [21, Example 4]. In view of Theorem 11, it is FOO for any $\mathbb{P}$ if it has zero mean. Using (58) and (6) in (33), we find (after some algebra) that

$$\boldsymbol{\mu} = \begin{bmatrix} 1 + 2\left(P_{C_1}(0) - P_{C_0}(0) - P_{C_2}(0)\right) \\ 2\left(P_{C_0}(0) - P_{C_2}(0)\right) \end{bmatrix}^{\mathrm{T}}. \quad (59)$$

For example, the bit probabilities $\boldsymbol{b}_5 = [0.40, 0.55, 0.60]$ give an input distribution $\mathbb{P}_5$ for which the mean (59) is $\boldsymbol{\mu} = [0.10, -0.40]$. We define another alphabet $\mathbb{X}'_{8AMPM}$ by subtracting $\boldsymbol{\mu}$ from each element in $\mathbb{X}_{8AMPM}$. Fig. 10 shows the translated constellation $[\mathbb{X}'_{8AMPM}, \mathbb{P}_5]$ along with $[\mathring{\mathbb{X}}'_{8AMPM}, \mathbb{U}_3]$, where $\mathring{\mathbb{X}}'_{8AMPM}$ is the transform of $\mathbb{X}'_{8AMPM}$ for the distribution $\mathbb{P}_5$. They are both zero-mean projected hypercubes and thus FOO according to Theorem 11. This can be observed from the BICM-GMI curve for $[\mathbb{X}'_{8AMPM}, \mathbb{P}_5]$ in Fig. 12.

The exemplified method holds in full generality: Any alphabet that is FOO with a uniform input distribution is FOO also
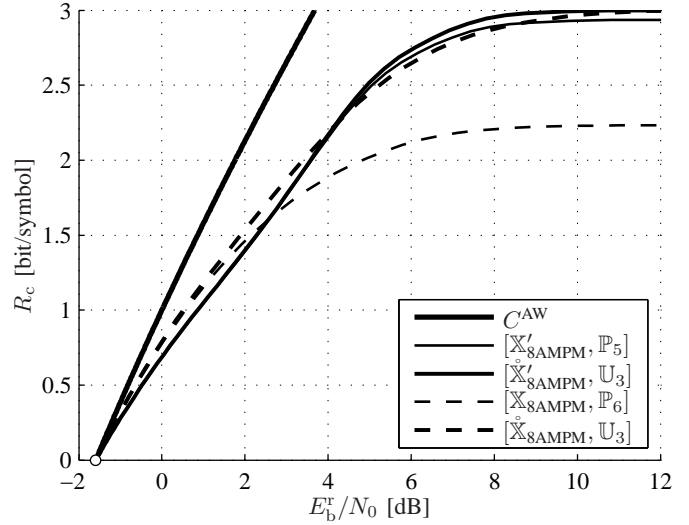

Fig. 12. BICM-GMI for the four FOO constellations in Figs. 10 and 11. The SL is shown with a white circle.

with an arbitrary nonuniform distribution, if it is translated to zero mean. Furthermore, *all* nonuniform FOO constellations can be constructed in this manner.

For certain distributions, the mean (59) is zero without translation. Specifically, $\boldsymbol{\mu} = \mathbf{0}$ if and only if

$$P_{C_0}(0) = P_{C_2}(0) = P_{C_1}(0)/2 + 1/4. \quad (60)$$

Clearly, the uniform case ($P_{C_0}(0) = P_{C_1}(0) = P_{C_2}(0) = 1/2$) analyzed in [21] fulfills (60). More interestingly, when any other vector of bit probabilities fulfilling (60) is used, the resulting constellation will be FOO. This is the case for instance with $\boldsymbol{b}_6 = [0.70, 0.90, 0.70]$. The obtained constellation, denoted by $[\mathbb{X}_{8AMPM}, \mathbb{P}_6]$ is illustrated in Fig. 11 along with its transform $[\mathring{\mathbb{X}}_{8AMPM}, \mathbb{U}_3]$. Graphically, both alphabets look like cubes, although viewed from different angles, which is precisely what Theorems 9 and 11 predict.

In Fig. 12, we show the BICM-GMI for the zero-mean constellations $[\mathbb{X}'_{8AMPM}, \mathbb{P}_5]$ and $[\mathbb{X}_{8AMPM}, \mathbb{P}_6]$ as well as for the constellations $[\mathring{\mathbb{X}}'_{8AMPM}, \mathbb{U}_3]$ and $[\mathring{\mathbb{X}}_{8AMPM}, \mathbb{U}_3]$. As expected, all the GMIs converge at the SL for low SNR. For these two cases, the transformed alphabets with uniform input distributions give larger GMI for all SNRs compared to the corresponding nonuniform ones. This is however not always the case, cf. Fig. 7 with $\mathbb{P}_2$.

Example 7: M-PAM alphabets have been shown to be FOO if the NBC is used with a uniform input distribution, i.e, the constellation $[\mathbb{X}_{PAM}, \mathbb{U}_m]$ with $\mathbb{X}_{PAM} = [-(M-1), -(M-3), \ldots, M-1]$ is FOO [18] [21, Th. 14]. In this example, we study the first-order behavior of the 8-PAM alphabet $\mathbb{X}'_{PAM} = [-7, 7, -1, 1, -5, 5, -3, 3]$, where the order of the points represents the BRGC. The constellation $[\mathbb{X}'_{PAM}, \mathbb{P}]$ is known not to be FOO for $\mathbb{U}_3$ [17, Th. 3].

The BICM-GMI of $\mathbb{X}'_{PAM}$ is shown in Fig. 13 for the set of bit probabilities $\boldsymbol{b}_7 = [0.5, p, p]$ for different values of p. For $p = 0.5$, the uniform distribution is obtained. As p decreases, the Gray-labeled constellation approaches a zero-mean binary alphabet, which is FOO [13, Fig. 2] [21,
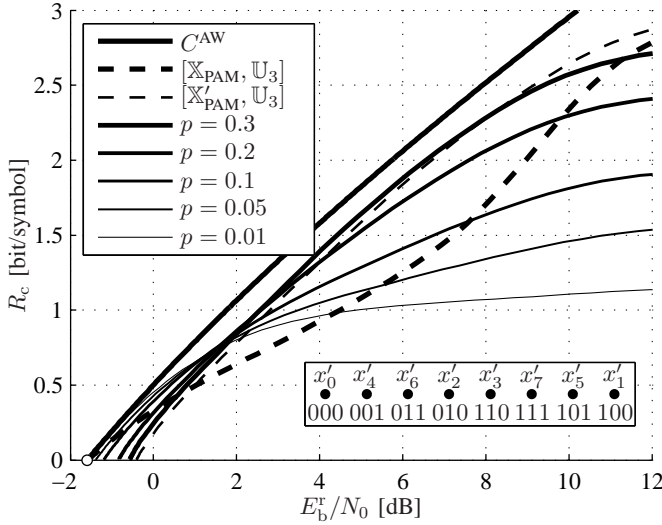
Fig. 13. The BICM-GMI for the 8-PAM alphabet $\mathbb{X}'_{\text{PAM}}$ labeled by the BRGC with uniform input distribution and bit probabilities $\boldsymbol{b}_7 = [0.5, p, p]$ for different values of p. The constellations approach a binary FOO constellation as $p \to 0$. The NBC-labeled 8-PAM alphabet $\mathbb{X}_{\text{PAM}}$, although FOO with a uniform distribution, is considerably weaker than $\mathbb{X}'_{\text{PAM}}$ for a wide range of SNRs.

Fig. 3 (b)]. Fig. 13 illustrates the tradeoff between the low- and high-SNR regimes: The SL can be approached by decreasing p, but this causes a decrease in GMI in the high-SNR regime. Alternatively, the SL can be attained by switching from the BRGC to the NBC, but this also comes with a heavy penalty at higher SNRs.

V. Conclusions

There exists a closed-form mapping between any probabilistically shaped constellation and a constellation with uniform input distribution, such that the two systems have the same low-SNR first-order behavior (Definition 3 and Theorem 8). Thus, the combination of probabilistic and geometric shaping is equivalent to pure geometric shaping at low SNR.

We are particularly interested in BICM systems that attain the SL -1.59 dB at asymptotically low SNR, i.e., in FOO constellations. Somewhat disappointingly, the set of probabilistically shaped FOO constellation is no larger than the set of FOO constellations with uniform distributions, disregarding translations of the whole input alphabet. Both sets can be fully characterized as the set of linear projections of a hypercube, translated to have zero mean for the considered input distribution (Theorems 9 and 10; cf. Figs. 10 and 11). Although non-FOO constellations for BICM can be improved by probabilistic shaping (Fig. 13), it is impossible to make them FOO except in degenerate cases (by setting some probabilities equal to zero).

Appendix
Properties of the Transform

In this Appendix, some theoretical properties of the new transform defined in Section II-E are proved. In the first section, the "sum-product lemma" is established, which will be used extensively throughout the appendix. In the following two sections, Lemma 1 and Theorem 4 are proved.

A. The Sum-Product Lemma

Many properties of NBC-labeled constellations and their transforms can be expressed as the sum of products of certain functions, where each function depends on one bit position. Such expressions, which occur frequently in the two subsequent sections, can be resolved using this general lemma.

Lemma 12: Let $f_{k,u}$ for $k = 0, \ldots, m-1$ and $u \in \mathcal{B}$ be any real numbers. Then

$$\sum_{i=0}^{M-1} \prod_{k=0}^{m-1} f_{k,n_{i,k}} = \prod_{k=0}^{m-1} (f_{k,0} + f_{k,1}).$$

Proof: A summation over $i = 0, \ldots, 2^m - 1$ is equivalent to m sums over $i_k \in \mathcal{B}$, where $k = 0, \ldots, m-1$ and $i = i_0 + 2i_1 + \cdots + 2^{m-1} i_{m-1}$. With this notation, $n_{i,k} = i_k$ and

$$\sum_{i=0}^{M-1} \prod_{k=0}^{m-1} f_{k,n_{i,k}} = \sum_{i_0 \in \mathcal{B}} \sum_{i_1 \in \mathcal{B}} \cdots \sum_{i_{m-1} \in \mathcal{B}} \prod_{k=0}^{m-1} f_{k,i_k}$$

$$= \sum_{i_0 \in \mathcal{B}} \sum_{i_1 \in \mathcal{B}} \cdots \sum_{i_{m-1} \in \mathcal{B}} f_{0,i_0} f_{1,i_1} \cdots f_{m-1,i_{m-1}}$$

$$= \sum_{i_0 \in \mathcal{B}} f_{0,i_0} \sum_{i_1 \in \mathcal{B}} f_{1,i_1} \cdots \sum_{i_{m-1} \in \mathcal{B}} f_{m-1,i_{m-1}}$$

$$= \prod_{k=0}^{m-1} \sum_{i_k \in \mathcal{B}} f_{k,i_k}. \qquad \square$$

B. The Transform Coefficients

Lemma 1, which lists two fundamental properties of the transform coefficients $g_{i,j}$, was given in Section II-E.

Proof of Lemma 1: The two parts of Lemma 1 will now be proved separately. First, the definition of $g_{i,j}$ in (13) yields

$$\sum_{i=0}^{M-1} g_{i,l} g_{i,j} = \sum_{i=0}^{M-1} \prod_{k=0}^{m-1} \left[(-1)^{\bar{n}_{i,k} n_{l,k}} \sqrt{P_{C_k}(0)} \right.$$
$$\left. + (-1)^{n_{i,k} \bar{n}_{l,k}} \sqrt{P_{C_k}(1)} \right]$$
$$\cdot \left[(-1)^{\bar{n}_{i,k} n_{j,k}} \sqrt{P_{C_k}(0)} \right.$$
$$\left. + (-1)^{n_{i,k} \bar{n}_{j,k}} \sqrt{P_{C_k}(1)} \right]$$

$$= \sum_{i=0}^{M-1} \prod_{k=0}^{m-1} \left[(-1)^{\bar{n}_{i,k}(n_{l,k}+n_{j,k})} P_{C_k}(0) \right.$$
$$+ (-1)^{\bar{n}_{i,k} n_{l,k} + n_{i,k} \bar{n}_{j,k}} \sqrt{P_{C_k}(0) P_{C_k}(1)}$$
$$+ (-1)^{n_{i,k} \bar{n}_{l,k} + \bar{n}_{i,k} n_{j,k}} \sqrt{P_{C_k}(0) P_{C_k}(1)}$$
$$\left. + (-1)^{n_{i,k}(\bar{n}_{l,k}+\bar{n}_{j,k})} P_{C_k}(1) \right]$$

$$= \sum_{i=0}^{M-1} \prod_{k=0}^{m-1} (-1)^{n_{i,k}(n_{l,k}+n_{j,k})}$$
$$\cdot \left[(-1)^{n_{l,k}+n_{j,k}} P_{C_k}(0) + P_{C_k}(1) \right.$$
$$\left. + (-1)^{n_{i,k}} \left[(-1)^{n_{l,k}} + (-1)^{n_{j,k}} \right] \sqrt{P_{C_k}(0) P_{C_k}(1)} \right]$$

where the last equality follows by repeatedly using the identities $\bar{u} = 1 - u$ and $(-1)^u = (-1)^{-u}$ for $u \in \mathcal{B}$. Lemma 12 now yields

$$
\begin{aligned}
\sum_{i=0}^{M-1} g_{i,l} g_{i,j} &= \prod_{k=0}^{m-1} \Big[(-1)^{n_{l,k}+n_{j,k}} P_{C_k}(0) + P_{C_k}(1) \\
&\quad + [(-1)^{n_{l,k}} + (-1)^{n_{j,k}}] \sqrt{P_{C_k}(0) P_{C_k}(1)} \\
&\quad + (-1)^{n_{l,k}+n_{j,k}} \Big((-1)^{n_{l,k}+n_{j,k}} P_{C_k}(0) + P_{C_k}(1) \\
&\quad - [(-1)^{n_{l,k}} + (-1)^{n_{j,k}}] \sqrt{P_{C_k}(0) P_{C_k}(1)} \Big) \Big] \\
&= \prod_{k=0}^{m-1} \Big[\big((-1)^{n_{l,k}+n_{j,k}} + 1 \big) P_{C_k}(0) \\
&\quad + \big(1 + (-1)^{n_{l,k}+n_{j,k}} \big) P_{C_k}(1) \\
&\quad + [(-1)^{n_{l,k}} + (-1)^{n_{j,k}} - (-1)^{n_{l,k}} - (-1)^{n_{j,k}}] \\
&\quad \cdot \sqrt{P_{C_k}(0) P_{C_k}(1)} \Big] \\
&= \prod_{k=0}^{m-1} \big(1 + (-1)^{n_{l,k}+n_{j,k}} \big)
\end{aligned}
\tag{61}
$$

where the last step follows because $P_{C_k}(0) + P_{C_k}(1) = 1$. The factors in (61) are either 2 or 0, depending on whether $n_{l,k} = n_{j,k}$ or $n_{l,k} \neq n_{j,k}$ for the particular bit position k. Thus, $\sum_{i=0}^{M-1} g_{i,l} g_{i,j}$ is either $2^m = M$ or 0, depending on whether l and j have *all* bits equal or not. This completes the proof of (14).

To prove the second part of Lemma 1, which is (15), we observe from (8) and (13) that

$$
\begin{aligned}
\sum_{i=0}^{M-1} h_{l,i} g_{i,j} &= \sum_{i=0}^{M-1} \prod_{k=0}^{m-1} (-1)^{n_{l,k} n_{i,k}} \\
&\quad \cdot \prod_{k=0}^{m-1} \Big[(-1)^{\bar{n}_{i,k} n_{j,k}} \sqrt{P_{C_k}(0)} + (-1)^{n_{i,k} \bar{n}_{j,k}} \sqrt{P_{C_k}(1)} \Big] \\
&= \sum_{i=0}^{M-1} \prod_{k=0}^{m-1} \phi_{k, n_{i,k}}
\end{aligned}
\tag{62}
$$

where

$$
\begin{aligned}
\phi_{k,u} &\triangleq (-1)^{u n_{l,k} + \bar{u} n_{j,k}} \sqrt{P_{C_k}(0)} \\
&\quad + (-1)^{u n_{l,k} + u \bar{n}_{j,k}} \sqrt{P_{C_k}(1)}.
\end{aligned}
$$

Intending to apply Lemma 12 to (62), we first calculate the quantity

$$
\begin{aligned}
\phi_{k,0} + \phi_{k,1} &= (-1)^{n_{j,k}} \sqrt{P_{C_k}(0)} + \sqrt{P_{C_k}(1)} \\
&\quad + (-1)^{n_{l,k}} \sqrt{P_{C_k}(0)} + (-1)^{n_{l,k}+\bar{n}_{j,k}} \sqrt{P_{C_k}(1)}
\end{aligned}
$$

for $k = 0, \ldots, m - 1$. For reasons that will soon become clear, we extract a common factor $(-1)^{n_{j,k} n_{l,k}}$ from all terms, obtaining

$$
\begin{aligned}
\phi_{k,0} + \phi_{k,1} &= (-1)^{n_{j,k} n_{l,k}} \\
&\quad \cdot \Big[\big((-1)^{n_{j,k} \bar{n}_{l,k}} + (-1)^{\bar{n}_{j,k} n_{l,k}} \big) \sqrt{P_{C_k}(0)} \\
&\quad + \big((-1)^{n_{j,k} n_{l,k}} + (-1)^{\bar{n}_{j,k} \bar{n}_{l,k}} \big) \sqrt{P_{C_k}(1)} \Big].
\end{aligned}
$$

The coefficient in front of $\sqrt{P_{C_k}(0)}$ is 2 if $n_{j,k} = n_{l,k}$ and 0 otherwise. Similarly, the coefficient in front of $\sqrt{P_{C_k}(1)}$ is 0 if $n_{j,k} = n_{l,k}$ and 2 otherwise. Thus, for every k, the expression depends on either $P_{C_k}(0)$ or $P_{C_k}(1)$ but not both, namely,

$$
\begin{aligned}
\phi_{k,0} + \phi_{k,1} &= \begin{cases} (-1)^{n_{j,k} n_{l,k}} \cdot 2 \sqrt{P_{C_k}(0)}, & n_{j,k} = n_{l,k} \\ (-1)^{n_{j,k} n_{l,k}} \cdot 2 \sqrt{P_{C_k}(1)}, & n_{j,k} \neq n_{l,k} \end{cases} \\
&= 2 (-1)^{n_{j,k} n_{l,k}} \sqrt{P_{C_k}(n_{j,k} \oplus n_{l,k})} \\
&= 2 (-1)^{n_{j,k} n_{l,k}} \sqrt{P_{C_k}(n_{j \oplus l, k})}
\end{aligned}
\tag{63}
$$

according to (5).

Using (63), we are now ready to apply Lemma 12 to (62), which yields

$$
\sum_{i=0}^{M-1} h_{i,l} g_{i,j} = \prod_{k=0}^{m-1} 2 (-1)^{n_{j,k} n_{l,k}} \sqrt{P_{C_k}(n_{j \oplus l, k})}.
$$

Applying (6) and (8), we obtain finally

$$
\begin{aligned}
\sum_{i=0}^{M-1} h_{i,l} g_{i,j} &= M \prod_{k=0}^{m-1} (-1)^{n_{j,k} n_{l,k}} \cdot \prod_{k=0}^{m-1} \sqrt{P_{C_k}(n_{j \oplus l, k})} \\
&= M h_{j,l} \sqrt{P_{j \oplus l}}
\end{aligned}
$$

which completes the proof of (15). $\qquad \square$

C. Two Consecutive Transforms

Theorem 4, which characterizes the vectors obtained by applying the new transform and the HT sequentially to a given alphabet, was given in Section II-E.

Proof of Theorem 4: To prove (20), we first write $\tilde{\mathbb{S}}$ as a function of $\mathbb{X}$ using (7) and (12), as illustrated in Fig. 2. For $i = 0, \ldots, M - 1$,

$$
\begin{aligned}
\tilde{\boldsymbol{s}}_i &= \frac{1}{M} \sum_{j=0}^{M-1} \boldsymbol{s}_j h_{i,j} \\
&= \frac{1}{M} \sum_{j=0}^{M-1} h_{i,j} \sum_{l=0}^{M-1} \boldsymbol{x}_l g_{j,l} \sqrt{P_l} \\
&= \frac{1}{M} \sum_{l=0}^{M-1} \boldsymbol{x}_l \sqrt{P_l} \sum_{j=0}^{M-1} h_{i,j} g_{j,l} \\
&= \sum_{l=0}^{M-1} \boldsymbol{x}_l h_{l,i} \sqrt{P_l P_{i \oplus l}}
\end{aligned}
$$

using (15) for the last equality. To obtain $\tilde{\mathbb{S}}$ as a function of $\tilde{\mathbb{X}}$, we express $\boldsymbol{x}_l$ in terms of its inverse HT in (11), to obtain

$$
\begin{aligned}
\tilde{\boldsymbol{s}}_i &= \sum_{l=0}^{M-1} h_{l,i} \sqrt{P_l P_{i \oplus l}} \sum_{j=0}^{M-1} \tilde{\boldsymbol{x}}_j h_{j,l} \\
&= \sum_{j=0}^{M-1} \tilde{\boldsymbol{x}}_j \eta_{i,j}
\end{aligned}
\tag{64}
$$

where

$$
\eta_{i,j} \triangleq \sum_{l=0}^{M-1} h_{l,i} h_{j,l} \sqrt{P_l P_{i \oplus l}}.
$$

The coefficient $\eta_{i,j}$ can be expressed using (8) and (5) as

$$
\eta_{i,j} = \sum_{l=0}^{M-1} \prod_{k=0}^{m-1} (-1)^{n_{l,k}(n_{i,k}+n_{j,k})}
$$
$$
\cdot \sqrt{P_{C_k}(n_{l,k}) P_{C_k}(n_{i,k} \oplus n_{l,k})} \tag{65}
$$

$$
= \prod_{k=0}^{m-1} \left[\sqrt{P_{C_k}(0) P_{C_k}(n_{i,k})} \right.
$$
$$
\left. + (-1)^{n_{i,k}+n_{j,k}} \sqrt{P_{C_k}(1) P_{C_k}(\bar{n}_{i,k})} \right] \tag{66}
$$

where to pass from (65) to (66) we used Lemma 12 with $f_{k,u} = (-1)^{u(n_{i,k}+n_{j,k})} \sqrt{P_{C_k}(u) P_{C_k}(n_{i,k} \oplus u)}$. Depending on the value of $n_{i,k}$, this product can be partitioned into two subproducts

$$
\eta_{i,j} = \prod_{\substack{k=0 \\ n_{i,k}=0}}^{m-1} [P_{C_k}(0) + (-1)^{n_{j,k}} P_{C_k}(1)]
$$
$$
\cdot \prod_{\substack{k=0 \\ n_{i,k}=1}}^{m-1} \left[\sqrt{P_{C_k}(0) P_{C_k}(1)} + (-1)^{\bar{n}_{j,k}} \sqrt{P_{C_k}(1) P_{C_k}(0)} \right]
$$
$$
= \psi_i \prod_{\substack{k=0 \\ n_{i,k}=0}}^{m-1} [P_{C_k}(0) + (-1)^{n_{j,k}} P_{C_k}(1)]
$$
$$
\cdot \prod_{\substack{k=0 \\ n_{i,k}=1}}^{m-1} \left[\frac{1 + (-1)^{\bar{n}_{j,k}}}{2} \right] \tag{67}
$$

making use of ψ_i defined in (22). Since the two factors in (67) depend on $n_{j,k}$ as

$$
P_{C_k}(0) + (-1)^{n_{j,k}} P_{C_k}(1) = \begin{cases} 1, & n_{j,k}=0, \\ P_{C_k}(0) - P_{C_k}(1), & n_{j,k}=1, \end{cases}
$$
$$
\frac{1 + (-1)^{\bar{n}_{j,k}}}{2} = \begin{cases} 0, & n_{j,k}=0, \\ 1, & n_{j,k}=1, \end{cases}
$$

they can be expanded into four factors as

$$
\eta_{i,j} = \psi_i \prod_{\substack{k=0 \\ n_{i,k}=0 \\ n_{j,k}=0}}^{m-1} 1 \cdot \prod_{\substack{k=0 \\ n_{i,k}=0 \\ n_{j,k}=1}}^{m-1} (P_{C_k}(0) - P_{C_k}(1))
$$
$$
\cdot \prod_{\substack{k=0 \\ n_{i,k}=1 \\ n_{j,k}=0}}^{m-1} 0 \cdot \prod_{\substack{k=0 \\ n_{i,k}=1 \\ n_{j,k}=1}}^{m-1} 1. \tag{68}
$$

The two products of ones can be omitted. The product of zeros may look strange but is perfectly legitimate; its value is by definition 1 if the set $\{k : [n_{i,k}, n_{j,k}] = [1,0]\}$ is empty and 0 otherwise. Furthermore, since $P_{C_k}(0) - P_{C_k}(n_{j,k}) = 0$ for all k in this set, the second and third factors of (68) can be merged into

$$
\eta_{i,j} = \psi_i \prod_{\substack{k=0 \\ n_{i,k} \neq n_{j,k}}}^{m-1} (P_{C_k}(0) - P_{C_k}(n_{j,k})) \tag{69}
$$

which together with (64) completes the proof of (20).

To prove the reverse relationship (21), we first multiply both sides of (64) with $h_{i,M-1} \eta_{l,i}/(\psi_i \psi_l)$ and sum over i:

$$
\sum_{i=0}^{M-1} \tilde{s}_i \frac{h_{i,M-1} \eta_{l,i}}{\psi_i \psi_l} = \sum_{j=0}^{M-1} \tilde{x}_j \sum_{i=0}^{M-1} \eta_{i,j} \frac{h_{i,M-1} \eta_{l,i}}{\psi_i \psi_l}. \tag{70}
$$

The Hadamard coefficient $h_{i,M-1}$ can be factorized using (8) with $n_{M-1,k} = 1, \forall k$ as

$$
h_{i,M-1} = \prod_{k=0}^{m-1} (-1)^{n_{i,k}}. \tag{71}
$$

Using (71) in (70) and replacing the ratios $\eta_{i,j}/\psi_i$ and $\eta_{l,i}/\psi_l$ with their factorizations according to (69), we obtain for the inner sum on the right-hand side of (70)

$$
\sum_{i=0}^{M-1} h_{i,M-1} \frac{\eta_{i,j}}{\psi_i} \frac{\eta_{l,i}}{\psi_l} = \sum_{i=0}^{M-1} \prod_{k=0}^{m-1} \phi_{k,n_{i,k}} \tag{72}
$$

where

$$
\phi_{k,u} \triangleq \begin{cases} (-1)^u, & u = n_{j,k} = n_{l,k}, \\ (-1)^u (P_{C_k}(0) - P_{C_k}(u)), & u = n_{j,k} \neq n_{l,k}, \\ (-1)^u (P_{C_k}(0) - P_{C_k}(n_{j,k})), & u = n_{l,k} \neq n_{j,k}, \\ (-1)^u (P_{C_k}(0) - P_{C_k}(n_{j,k}))(P_{C_k}(0) - P_{C_k}(u)), \\ \qquad u \neq n_{j,k} = n_{l,k}. \end{cases}
$$

The fourth case is always 0, because either $P_{C_k}(0) - P_{C_k}(n_{j,k})$ or $P_{C_k}(0) - P_{C_k}(u)$ is 0. Furthermore, the second and third cases can be combined into

$$
\phi_{k,u} \triangleq \begin{cases} (-1)^u, & u = n_{j,k} = n_{l,k}, \\ (-1)^u (P_{C_k}(0) - P_{C_k}(n_{j,k})), & n_{j,k} \neq n_{l,k}, \\ 0, & u \neq n_{j,k} = n_{l,k}. \end{cases} \tag{73}
$$

We wish to apply Lemma 12 to the right-hand side of (72). In order to do so, we first calculate from (73)

$$
\phi_{k,0} + \phi_{k,1} = \begin{cases} 1+0, & n_{j,k}=n_{l,k}=0, \\ (P_{C_k}(0) - P_{C_k}(n_{j,k})) \\ \quad -(P_{C_k}(0) - P_{C_k}(n_{j,k})), & n_{j,k} \neq n_{l,k}, \\ 0-1, & n_{j,k}=n_{l,k}=1 \end{cases}
$$
$$
= \begin{cases} (-1)^{n_{j,k}}, & n_{j,k}=n_{l,k}, \\ 0, & n_{j,k} \neq n_{l,k}. \end{cases} \tag{74}
$$

Now by Lemma 12, (72) can be expressed as

$$
\sum_{i=0}^{M-1} h_{i,M-1} \frac{\eta_{i,j}}{\psi_i} \frac{\eta_{l,i}}{\psi_l} = \prod_{k=0}^{m-1} (\phi_{k,0} + \phi_{k,1}).
$$

By (74), this product will be nonzero only if j and l match in *all* bit positions $k = 0, \ldots, m-1$, i.e., if $j = l$. Thus, again utilizing (71),

$$
\sum_{i=0}^{M-1} h_{i,M-1} \frac{\eta_{i,j}}{\psi_i} \frac{\eta_{l,i}}{\psi_l} = \begin{cases} \prod_{k=0}^{m-1} (-1)^{n_{l,k}}, & j=l, \\ 0, & j \neq l, \end{cases}
$$
$$
= \begin{cases} h_{l,M-1}, & j=l, \\ 0, & j \neq l. \end{cases} \tag{75}
$$

Finally, we combine (70) and (75) into

$$\sum_{i=0}^{M-1} \tilde{\boldsymbol{s}}_i \frac{h_{i,M-1}\eta_{l,i}}{\psi_i \psi_l} = h_{l,M-1}\tilde{\boldsymbol{x}}_l. \tag{76}$$

Dividing both sides by $h_{l,M-1}$ yields on the left-hand side the coefficent $h_{i,M-1}\eta_{l,i}/(h_{l,M-1}\psi_l)$, which can be expressed using (8) and (69) as

$$\frac{h_{i,M-1}\eta_{l,i}}{h_{l,M-1}\psi_l} = \prod_{k=0}^{m-1} (-1)^{n_{i,k}-n_{l,k}}$$
$$\cdot \prod_{\substack{k=0 \\ n_{l,k}\neq n_{i,k}}}^{m-1} \left(P_{C_k}(0) - P_{C_k}(n_{i,k})\right)$$
$$= \prod_{\substack{k=0 \\ n_{i,k}\neq n_{l,k}}}^{m-1} (-1) \left(P_{C_k}(0) - P_{C_k}(n_{i,k})\right).$$

This expression, substituted into (76), completes the proof of (21). $\square$

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] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable signaling on the Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 36, no. 4, pp. 726–740, July 1990.

[5] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. John Wiley & Sons, 2002.

[6] D. Sommer and G. P. Fettweis, "Signal shaping by non-uniform QAM for AWGN channels and applications using turbo coding," in *International ITG Conference on Source and Channel Coding (SCC)*, Munich, Germany, Jan. 2000.

[7] S. Y. Le Goff, "Signal constellations for bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 307–313, Jan. 2003.

[8] M. Barsoum, C. Jones, and M. Fitz, "Constellation design via capacity maximization," in *IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 2007.

[9] S. Y. Le Goff, B. S. Sharif, and S. A. Jimaa, "A new bit-interleaved coded modulation scheme using shaping coding," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Dallas, TX, Nov.–Dec. 2004.

[10] D. Raphaeli and A. Gurevitz, "Constellation shaping for pragmatic turbo-coded modulation with high spectral efficiency," *IEEE Trans. Commun.*, vol. 52, no. 3, pp. 341–345, Mar. 2004.

[11] S. Y. Le Goff, B. S. Sharif, and S. A. Jimaa, "Bit-interleaved turbo-coded modulation using shaping coding," *IEEE Commun. Lett.*, vol. 9, no. 3, pp. 246–248, Mar. 2005.

[12] S. Y. Le Goff, B. K. Khoo, C. C. Tsimenidis, and B. S. Sharif, "Constellation shaping for bandwidth-efficient turbo-coded modulation with iterative receiver," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2223–2233, June 2007.

[13] A. Guillén i Fàbregas and A. Martinez, "Bit-interleaved coded modulation with shaping," in *IEEE Information Theory Workshop (ITW)*, Dublin, Ireland, Aug.–Sep. 2010.

[14] L. Peng, A. Guillén i Fàbregas, and A. Martinez, "Mismatched shaping schemes for bit-interleaved coded modulation," in *IEEE International Symposium on Information Theory (ISIT)*, Cambridge, MA, July 2012.

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

[16] V. V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.

[17] 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.

[18] C. Stierstorfer and R. F. H. Fischer, "Asymptotically optimal mappings for BICM with M-PAM and M^2-QAM," *IET Electronics Letters*, vol. 45, no. 3, pp. 173–174, Jan. 2009.

[19] A. Alvarado, E. Agrell, A. Guillén i Fàbregas, and A. Martinez, "Corrections to 'Bit-interleaved coded modulation in the wideband regime'," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, p. 6513, Dec. 2010.

[20] C. Stierstorfer and R. F. H. Fischer, "Mappings for BICM in UWB scenarios," in *International ITG Conference on Source and Channel Coding (SCC)*, Ulm, Germany, Jan. 2008.

[21] E. Agrell and A. Alvarado, "Optimal alphabets and binary labelings for BICM at low SNR," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6650–6672, Oct. 2011.

[22] T. Nguyen and L. Lampe, "Bit-interleaved coded modulation with mismatched decoding metrics," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 437–447, Feb. 2011.

[23] W. K. Pratt, J. Kane, and H. C. Andrews, "Hadamard transform image coding," *Proceedings of the IEEE*, vol. 57, no. 1, pp. 58–72, Jan. 1969.

[24] R. Calderbank and J. E. Mazo, "A new description of trellis codes," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 6, pp. 784–791, Nov. 1984.

[25] 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.

[26] F. A. Khan, E. Agrell, and M. Karlsson, "Electronic dispersion compensation by Hadamard transformation," in *Optical Fiber Conference (OFC)*, San Diego, CA, Mar. 2010.

[27] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2756–2765, June 2009.

[28] A. Alvarado, F. Brännström, and E. Agrell, "High SNR bounds for the BICM capacity," in *IEEE Information Theory Workshop (ITW)*, Paraty, Brazil, Oct. 2011.

[29] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "On the optimality of the binary reflected Gray code," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3170–3182, Dec. 2004.

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