First-order Asymptotics of the BICM Mutual Information: Uniform vs. Nonuniform Distributions

Erik Agrell
Department of Signals and Systems
Chalmers University of Technology
SE-41296 Göteborg
Sweden
Email: agrell@chalmers.se

Alex Alvarado
Department of Engineering
University of Cambridge
Cambridge CB2 1PZ
United Kingdom
Email: alex.alvarado@ieee.org

***Abstract*— A linear, invertible transform is defined between two vectors or matrices as a tool for analyzing the bit-interleaved coded modulation (BICM) mutual information in the wideband regime. The transform coefficients depend on a set of real values, which can be interpreted as probabilities. The transform relates any BICM system with a nonuniform input distribution to another BICM system with a uniform distribution. Numerical evidence suggests that the two systems have the same first-order behavior, which would make possible to analyze nonuniform BICM systems based on known properties of uniform BICM systems.**

I. Introduction

In 1992, Zehavi introduced the so-called bit-interleaved coded modulation (BICM) [1], and it has since then been rapidly adopted in commercial systems such as wireless and wired broadband access networks, 3G/4G telephony, and digital video broadcasting [2]. An achievable rate for BICM systems is the so-called BICM mutual information (BICM-MI) defined as the sum of the mutual informations for each bit separately [3] [4]. The asymptotic behavior of the BICM-MI at low signal-to-noise ratio (SNR), i.e., in the wideband regime, was studied in [5]–[9] as a function of the alphabet and the binary labeling, assuming a uniform input distribution.

Probabilistic shaping for BICM, i.e., varying the probabilities of the bit streams, was first proposed in [10], [11] and developed further in [12]–[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 a Gaussian distribution. This is particularly advantageous at low and medium SNR.

In this paper, the first-order asymptotic behavior of the BICM-MI is analyzed for BICM systems with nonuniform distributions. A linear transform is introduced, which establishes an equivalence between an arbitrary nonuniform constellation and another constellation with uniform probabilities, in the sense that the BICM-MI of the two constellations have the same first-order asymptote. Since the uniform case has been investigated in detail and is now well understood (see references above, in particular [9]), the new transform offers an instrument to generalize such results into the nonuniform case.

Notation: Matrices are denoted by block letters $\mathbb{A}$ and vectors are row vectors. The Euclidean norm of $\boldsymbol{a}$ is denoted by $\|\boldsymbol{a}\|$. Random variables are denoted by capital letters A and random vectors by boldface capital vectors $\boldsymbol{A}$. The binary set is denoted by $\mathcal{B} \triangleq \{0,1\}$ and the negation of a bit b is denoted by $\bar{b} = 1 - b$. Expectations are denoted by $\mathbb{E}$.

II. System Model

We consider the generic BICM scheme in Fig. 1. 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 [11]–[14]. The mapper Φ is defined via the alphabet $\mathbb{X} \triangleq [\boldsymbol{x}_0^{\mathrm{T}}, \ldots, \boldsymbol{x}_{M-1}^{\mathrm{T}}]^{\mathrm{T}}$, where $M = 2^m$ and $\boldsymbol{x}_i \in \mathbb{R}^N$ for $i = 0, \ldots, M-1$. Each symbol $\boldsymbol{x}_i$ is labeled by the m bits that form the base-2 representation $[n_{i,0}, n_{i,1}, \ldots, n_{i,m-1}]$ of the integer i. As a consequence of this enumeration convention, the constellation $\mathbb{X}$ is labeled by the natural binary code. However, this is merely a notational convention, which does not limit the applicability of the results. An arbitrary binary labeling for the same alphabet can be analyzed simply by reordering the rows of $\mathbb{X}$.

Assuming independent, but possibly nonuniformly distributed, bits $C_0, \ldots, C_{m-1}$ at the input of the mapper (cf. Fig. 1), the probability that the symbol $\boldsymbol{x}_i$ will be transmitted is [9, eq. (30)] [14, eq. (8)] [15, eq. (9)]

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

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 set of bit probabilities $[P_{C_0}(0), \ldots, P_{C_{m-1}}(0)]$. A *constellation* is defined as the pair $[\mathbb{X}, \mathbb{P}]$, where $\mathbb{P} \triangleq [P_0, \ldots, P_{M-1}]^{\mathrm{T}}$ is the *input distribution*. An important special case is the *uniform distribution*, for which $P_{C_k}(0) = 1/2$ for $k = 0, \ldots, m-1$ and $\mathbb{P} = \mathbb{U}_m \triangleq [1/M, \ldots, 1/M]^{\mathrm{T}}$.

Throughout this paper, we assume that $0 < P_{C_k}(0) < 1$ for all $k = 0, \ldots, m-1$, i.e., we assume that all constellation points are used with a nonzero probability. If that was not the case, the cardinality of the constellation should be reduced. Hence $P_i > 0$ for $i = 0, \ldots, M-1$.

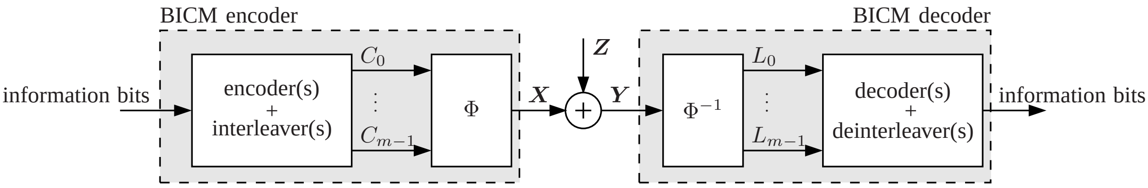


Fig. 1. A generic BICM system, consisting of a transmitter, an AWGN channel, and a receiver.

We consider transmissions over a discrete-time memoryless additive white Gaussian noise channel. The received vector at any discrete time instant is

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}, \tag{2}$$

where $\boldsymbol{X}$ is the channel input and $\boldsymbol{Z}$ is a Gaussian noise with zero mean and variance $N_0/2$ in each dimension. The SNR is defined as

$$SNR \triangleq \frac{E_{\mathrm{s}}}{N_0} = R_{\mathrm{c}}\frac{E_{\mathrm{b}}}{N_0}, \tag{3}$$

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

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 $k = 0, \dots, m-1$, 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.

III. Mutual Information

The *mutual information* (MI) in bits per channel use between the random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ is

$$I_{\boldsymbol{X}}(\boldsymbol{X};\boldsymbol{Y}) \triangleq \mathbb{E}\left[\log_2 \frac{p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})}{p_{\boldsymbol{Y}}(\boldsymbol{Y})}\right], \tag{4}$$

where the expectation is taken over the joint pdf $p_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y})$ and the conditional transition pdf is

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

The *conditional MI* is defined as the MI between $\boldsymbol{X}$ and $\boldsymbol{Y}$ conditioned on the value of the kth bit at the input of the mapper, i.e.,

$$I_{\boldsymbol{X}|C_k=u}(\boldsymbol{X};\boldsymbol{Y}) \triangleq \mathbb{E}\left[\log_2 \frac{p_{\boldsymbol{Y}|\boldsymbol{X},C_k}(\boldsymbol{Y}|\boldsymbol{X},u)}{p_{\boldsymbol{Y},C_k}(\boldsymbol{Y}|u)}\right], \tag{5}$$

where the expectation is taken over the conditional joint pdf $p_{\boldsymbol{X},\boldsymbol{Y}|C_k}(\boldsymbol{x},\boldsymbol{y}|u)$.

We are interested in the BICM-MI, defined as [3]–[5], [8]

$$I(SNR) \triangleq mI_{\boldsymbol{X}}(\boldsymbol{X};\boldsymbol{Y}) - \sum_{k=0}^{m-1}\sum_{u\in\mathcal{B}} P_{C_k}(u) I_{\boldsymbol{X}|C_k=u}(\boldsymbol{X};\boldsymbol{Y}). \tag{6}$$

We will analyze the right-hand side of (6) as a function of SNR, which means either varying N_0 for fixed constellation or, equivalently, rescaling the alphabet $\mathbb{X}$ linearly for fixed input distribution and N_0.

Martinez *et al.* [3] recognized the BICM decoder in Fig. 1 as a mismatched decoder and showed that the BICM-MI in (6) 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 and only if $R_{\mathrm{c}} \le I(SNR)$. Since from (3) $SNR/R_{\mathrm{c}} = E_{\mathrm{b}}/N_0$, the inequality $R_{\mathrm{c}} \le I(SNR)$ gives

$$\frac{E_{\mathrm{b}}}{N_0} \ge \frac{SNR}{I(SNR)}$$

for any SNR. Specifically, for asymptotically low SNR (or equivalently asymptotically low rates R_{c}, which means in the wideband regime), the average bit energy needed for reliable transmission is lower-bounded by $E_{\mathrm{b}}/N_0 \ge \alpha^{-1}$, where

$$\alpha \triangleq \lim_{SNR\to 0^+} \frac{I(SNR)}{SNR}.$$

Furthermore, $\alpha^{-1} \ge \log_{\mathrm{e}} 2 = -1.59$ dB, since no communication system can surpass the Shannon limit (SL).

IV. A Distribution-Dependent 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 $[P_{C_0}(0), \dots, P_{C_{m-1}}(0)]$.

For all $i = 0, \dots, M-1$ and $j = 0, \dots, M-1$, we define the transform coefficients

$$\gamma_{i,j} \triangleq \prod_{k=0}^{m-1}\left[(-1)^{\bar{n}_{i,k}n_{j,k}}\sqrt{P_{C_k}(0)} + (-1)^{n_{i,k}\bar{n}_{j,k}}\sqrt{P_{C_k}(1)}\right]. \tag{7}$$

Given probabilities $[P_{C_0}(0), \dots, P_{C_{m-1}}(0)]$, the transform $\mathring{\mathbb{X}} = [\mathring{\boldsymbol{x}}_0^{\mathrm{T}}, \dots, \mathring{\boldsymbol{x}}_{M-1}^{\mathrm{T}}]^{\mathrm{T}}$ of a matrix (or vector) $\mathbb{X} = [\boldsymbol{x}_0^{\mathrm{T}}, \dots, \boldsymbol{x}_{M-1}^{\mathrm{T}}]^{\mathrm{T}}$ with $M = 2^m$ rows is now defined as

$$\mathring{\boldsymbol{x}}_i \triangleq \sum_{j=0}^{M-1} \boldsymbol{x}_j \gamma_{i,j}\sqrt{P_j}, \qquad i = 0, \dots, M-1, \tag{8}$$

with P_j given by (1). For equally likely symbols, i.e., $\mathbb{P} = \mathbb{U}_m$, the new transform becomes the identity operation $\mathring{\mathbb{X}} = \mathbb{X}$, because then $\gamma_{i,i} = \sqrt{M}$ for $i = 1, \dots, M$ and $\gamma_{i,j} = 0$ for $i \ne j$.

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

$$\boldsymbol{x}_j = \frac{1}{M\sqrt{P_j}}\sum_{i=0}^{M-1} \mathring{\boldsymbol{x}}_i \gamma_{i,j}, \qquad j = 0, \dots, M-1. \tag{9}$$

Before proving Theorem 1, we need to establish a lemma.

Lemma 1: 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

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

□

This Lemma will now be used in the proof of Theorem 1.

Proof of Theorem 1:

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

$$\begin{aligned}
\sum_{i=0}^{M-1} \mathring{\boldsymbol{x}}_i \gamma_{i,j} &= \sum_{i=0}^{M-1} \gamma_{i,j} \sum_{l=0}^{M-1} \boldsymbol{x}_l \gamma_{i,l} \sqrt{P_l} \\
&= \sum_{l=0}^{M-1} \boldsymbol{x}_l \sqrt{P_l} \sum_{i=0}^{M-1} \gamma_{i,l} \gamma_{i,j}. \qquad (10)
\end{aligned}$$

The inner sum can be expanded using the definition of $\gamma_{i,j}$ in (7) as

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

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

$$\begin{aligned}
f_{k,u} = (-1)^{u(n_{l,k}+n_{j,k})} &\Big[(-1)^{n_{l,k}+n_{j,k}} P_{C_k}(0) + P_{C_k}(1) \\
&+ (-1)^u \left[(-1)^{n_{l,k}} + (-1)^{n_{j,k}}\right] \sqrt{P_{C_k}(0) P_{C_k}(1)}\Big]
\end{aligned}$$

and obtain

$$\begin{aligned}
\sum_{i=0}^{M-1} \gamma_{i,l}\gamma_{i,j} = \prod_{k=0}^{m-1} &\Big[(-1)^{n_{l,k}+n_{j,k}} P_{C_k}(0) + P_{C_k}(1) \\
&+ \left[(-1)^{n_{l,k}} + (-1)^{n_{j,k}}\right] \sqrt{P_{C_k}(0) P_{C_k}(1)} \\
&+ (-1)^{n_{l,k}+n_{j,k}} \Big((-1)^{n_{l,k}+n_{j,k}} P_{C_k}(0) + P_{C_k}(1) \\
&\quad - \left[(-1)^{n_{l,k}} + (-1)^{n_{j,k}}\right] \sqrt{P_{C_k}(0) P_{C_k}(1)}\Big)\Big] \\
= \prod_{k=0}^{m-1} &\Big[\left((-1)^{n_{l,k}+n_{j,k}} + 1\right) P_{C_k}(0) \\
&+ \left(1 + (-1)^{n_{l,k}+n_{j,k}}\right) P_{C_k}(1) \\
&+ \left[(-1)^{n_{l,k}} + (-1)^{n_{j,k}} - (-1)^{n_{j,k}} - (-1)^{n_{l,k}}\right] \\
&\quad \cdot \sqrt{P_{C_k}(0) P_{C_k}(1)}\Big] \\
= \prod_{k=0}^{m-1} &\left(1 + (-1)^{n_{l,k}+n_{j,k}}\right), \qquad (11)
\end{aligned}$$

where the last step follows because $P_{C_k}(0) + P_{C_k}(1) = 1$. The factors in (11) 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, the product will be zero unless *all* bits of l and j are equal, and

$$\sum_{i=0}^{M-1} \gamma_{i,l}\gamma_{i,j} = \begin{cases} M, & j = l, \\ 0, & j \neq l, \end{cases} \qquad l = 0, \ldots, M-1. \qquad (12)$$

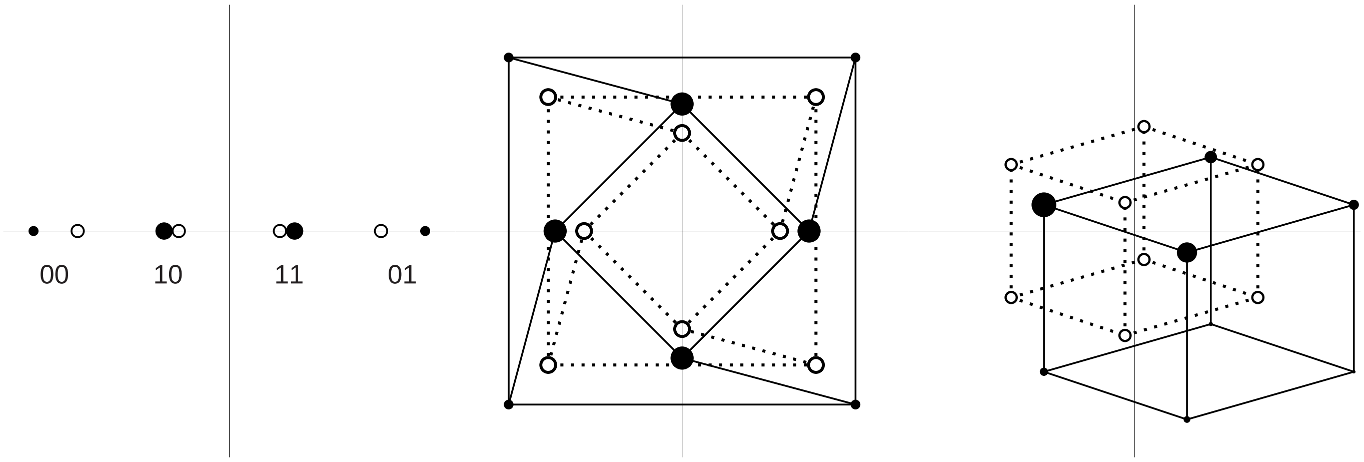
Applying (12) to the inner sum of (10) and dividing both sides by $M\sqrt{P_j}$, which by Sec. II is nonzero, completes the proof.

□

V. Numerical results

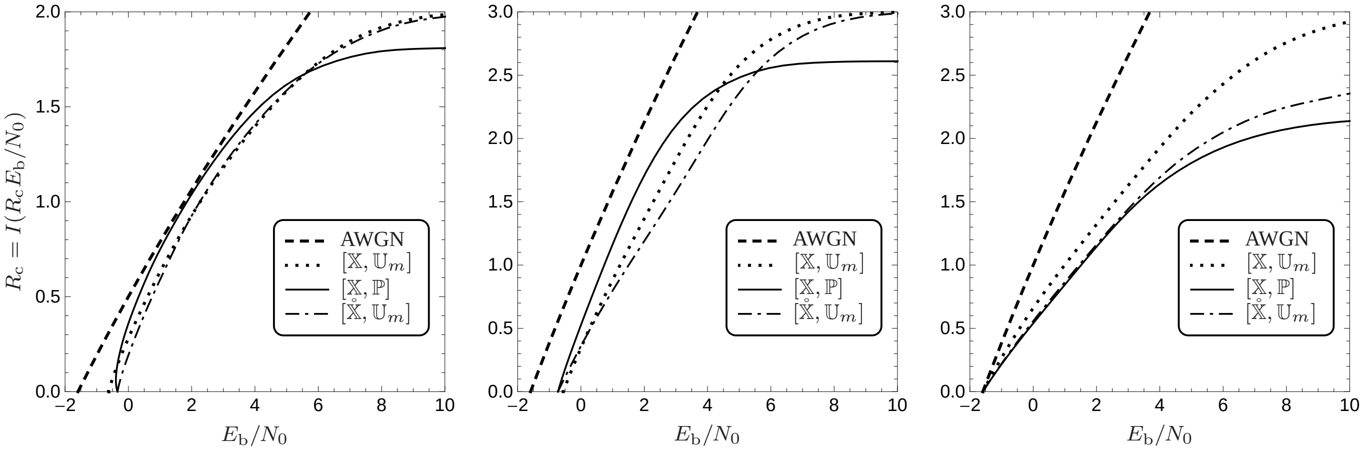
Fig. 2 illustrates three alphabets: quaternary pulse amplitude modulation (4-PAM), eight-level star-shaped quaternary amplitude modulation (8-QAM), and an eight-level irregular constellation, which achieves the SL -1.59 dB [9, Fig. 4 (a)]. Each alphabet is shown with two different input distributions, one uniform and one nonuniform, scaled to the the same average energy and translated to zero mean. The apparent three-dimensional structure in Fig. 2 (c) is an illusion; the alphabet consists of eight points in the plane, just like Fig. 2 (b). An alphabet that admits this particular illusion can be described as a linear projection of a hypercube and achieves the SL at low SNR [9, Theorem 12] when used with a uniform distribution.

The BICM-MI of these constellations are is shown in Fig. 3. For 4-PAM and 8-QAM, probabilistic shaping improves the BICM-MI considerably over a wide range of SNRs, which is reasonable since the shaped constellations in Fig. 2 resemble



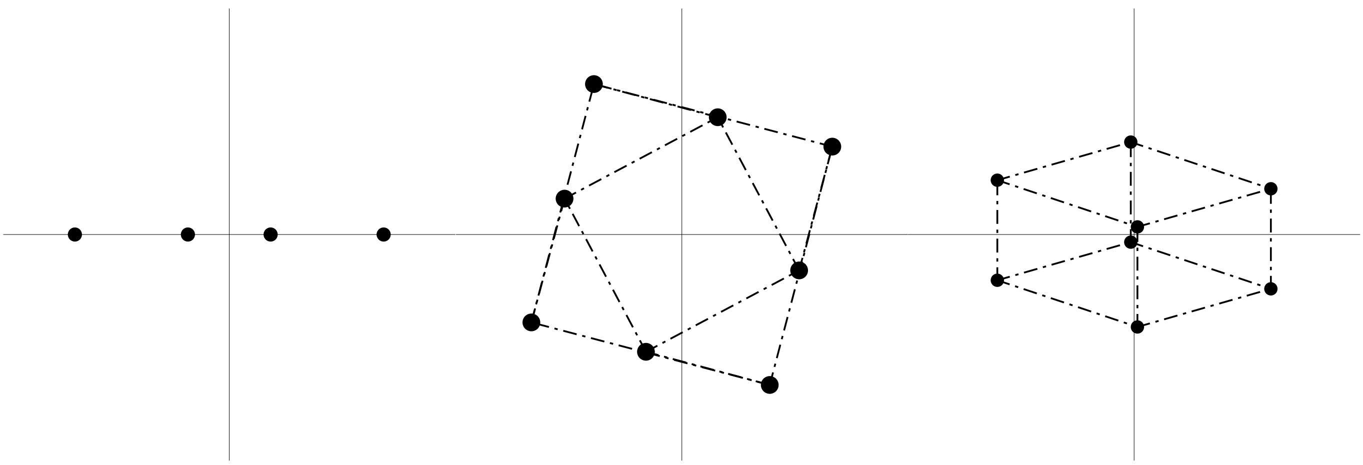
(a) (b) (c)

Fig. 2. Uniform $[\mathbb{X}, \mathbb{U}_m]$ (empty circles) and nonuniform (filled circles) constellations $[\mathbb{X}, \mathbb{P}]$ based on (a) 4-PAM, (b) star 8-QAM, and (c) an 8-level irregular alphabet. Lines connect symbols whose codewords have a Hamming distance of one. The symbol probabilities are proportional to the area of the corresponding circles. The constellations in (c) create an illusion of perspective, although they are purely two-dimensional just like those in (b).



(a) (b) (c)

Fig. 3. The BICM-MI of the six constellations in Fig. 2 and the three transformed constellations in Fig. 4. The AWGN bound $(N/2)\log_2(1+(2/N)SNR)$ is included for reference, where the dimension $N=1$ for (a) and $N=2$ for (b) and (c). In all cases, the curves for $[\mathbb{X}, \mathbb{P}]$ and $[\mathring{\mathbb{X}}, \mathbb{U}_m]$ meet at their lower endpoints, where $E_b/N_0 = \alpha^{-1} = -0.34$ (a), -0.73 (b), and -1.59 dB (c).



(a) (b) (c)

Fig. 4. The transforms $[\mathring{\mathbb{X}}, \mathbb{U}_m]$ of the three constellations $[\mathbb{X}, \mathbb{P}]$ in Fig. 2.

a Gaussian distribution better. This is not the case for the third alphabet, whose BICM-MI is not improved by shaping, at least not with this particular distribution.

The transforms of the three nonuniform constellations are calculated according to (8) and evaluated with a uniform distribution. Graphically, the transformed constellations (shown in Fig. 4) look quite different from the original constellations in Fig. 2. The 4-PAM constellation with equal spacing and nonuniform probabilities is converted into a 4-PAM constellation with unequal spacing and uniform probabilities. The BICM-MIs of the transformed constellations, which are also shown in Fig. 3, are different from the original ones in general, with one exception: their lower endpoints coincide, indicating that their α parameters are the same. This equivalence between constellations and their transforms has been observed for every studied nonuniform constellation.

VI. Conclusions

The numerical results in Sec. V provide evidence that the BICM-MI of two constellations related by the transform have the same low-SNR behavior, quantified by the parameter α. Analytical evidence will be provided in a future publication. The significance of this relation lies in the abundance of existing analytical results for the first-order asymptotics of the BICM-MI with uniform distributions [5]–[9], [16] and the absence of similar results for nonuniform distributions.

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] 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.

[3] 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.

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

[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] 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.

[7] 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.

[8] ——, "Asymptotically optimal mappings for BICM with M-PAM and M^2-QAM," *IET Electronics Letters*, vol. 45, no. 3, pp. 173–174, Jan. 2009.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

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