Mismatched Shaping Schemes for Bit-Interleaved Coded Modulation

Li Peng
University of Cambridge
lp327@cam.ac.uk

Albert Guillén i Fàbregas
ICREA & Universitat Pompeu Fabra
University of Cambridge
guillen@ieee.org

Alfonso Martinez
Universitat Pompeu Fabra
alfonso.martinez@ieee.org

***Abstract*— We consider bit-interleaved coded modulation (BICM) schemes where, instead of the true bit or symbol probabilities and the constellation used at the transmitter, the decoder uses arbitrary probabilities or reference constellations. We study the corresponding low- and high- signal-to-noise-ratio regimes and show that even in the presence of this extra sources of mismatch, BICM has a negligible penalty with respect to coded modulation.**

I. Introduction

Bit-Interleaved Coded Modulation (BICM) was introduced by Zehavi [1] as a pragmatic coding scheme for combining coding and modulation to achieve high spectral efficiency. It was later extensively studied by Caire *et al.* [2] under the assumption of infinite depth interleaving, where it is argued that the system essentially behaves as a set of parallel independent memoryless binary-input output-symmetric channels.

In a recent work by Martinez *et al.* [3], [4], the BICM decoder has been cast as a mismatched decoder [5], [6]. The model is extended to account for shaping in [7], i.e., the bit or symbol probabilities are optimized. The restriction of infinite depth interleaving was lifted with this mismatched decoder framework and the generalized mutual information (GMI) is used to measure the achievable rates. When the classical BICM decoder as in [4], [7] is used, the mismatch solely comes from the assumption of independent bit metrics. It was found in [7], [8] that for the AWGN channel, this mismatch can be mostly recovered by shaping, virtually closing the gap that made BICM suboptimal compared to coded modulation (CM) in terms of information rates. In particular, in the low signal-to-noise ratio (SNR) regime, BICM with shaping becomes first- and second-order optimal [7], [8].

In this paper, we take a closer look at shaping and the BICM mismatched decoder. In particular, we consider a generalized BICM decoder where one decodes using arbitrary bit or symbol probabilities and reference constellations instead of the true ones that used at the transmitter. The resulting additional source of mismatch might reduce the complexity of the decoder of BICM shaping schemes in [7]. We also analytically study the behavior of this doubly-mismatched decoder in the low SNR regime, and show that our scheme can be made first- and second-order optimal. Finally, we also provide a framework for studying the behavior of BICM GMI in the high SNR regime, more specifically, in the saturation region.

This work has been supported by the European Research Council under ERC grant agreement 259663. A. Martinez received funding from the Ministry of Science and Innovation (Spain) under grant RYC-2011-08150 and from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant PCIG10-GA-2011-303633 MDITwACM.

II. Channel Model and Code Ensembles

A. Channel Model

We consider transmission of information with a block code of length N and rate R. At the transmitter, a message m is mapped into a codeword $\boldsymbol{x}(\mathsf{m}) = (x_1(\mathsf{m}), \ldots, x_N(\mathsf{m}))$, where $x_k(\mathsf{m}) \in \mathcal{X}$, and $\mathcal{X}$ is the channel input alphabet. Let $M \triangleq |\mathcal{X}|$ and $m \triangleq \log_2(M)$ denotes the cardinality of $\mathcal{X}$ and the number of bits per symbol, respectively.

The output $\boldsymbol{y} = (y_1, \ldots, y_N)$ is a random transformation of the input with transition probability distribution $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$. We assume memoryless channels, therefore

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{k=1}^{N} P_{Y|X}(y_k|x_k), \tag{1}$$

where $P_{Y|X}(y|x)$ is the channel transition probability, and X, Y denote the underlying random variables taking values on the input and output alphabet $\mathcal{X}, \mathcal{Y}$.

The decoder decides on the estimate of the message $\widehat{\mathsf{m}}$ according to a decoding metric $q(\boldsymbol{x}, \boldsymbol{y})$. We assume a maximum metric decoder given by

$$\widehat{\mathsf{m}} = \arg\max_{\mathsf{m}} q(\boldsymbol{x}(\mathsf{m}), \boldsymbol{y}) = \arg\max_{\mathsf{m}} \prod_{k=1}^{N} q(x_k(\mathsf{m}), y_k). \tag{2}$$

When the metric $q(x, y)$ is a strictly increasing bijective function of $P_{Y|X}(y|x)$, the decoder will always select the maximum a posteriori (MAP) codeword. Otherwise, we have a mismatched decoder [5], [6].

B. Coded Modulation (CM)

The random coding ensemble corresponding to CM has channel inputs selected i.i.d. from $\mathcal{X}$ according to a probability distribution $P_X(x)$. The CM decoder employs the MAP metric. The largest information rate that can be achieved with CM under the constraint $x \in \mathcal{X}$ is

$$\mathsf{C}^{\mathrm{cm}} = \sup_{P_X} I(X; Y). \tag{3}$$

C. Bit-Interleaved Coded Modulation (BICM)

In a BICM scheme, the codewords are obtained as the serial concatenation of a binary codeword of length $n = mN$, a bit-level interleaver[1], and a binary labeling function $\mu : \{0,1\}^m \to \mathcal{X}$ which takes blocks of m bits and maps the m bits to signal constellation symbols x, such that $x_k = \mu\big(b_{(k-1)m+1}, \ldots, b_{km}\big), k = 1, \ldots, N$. We denote the inverse labeling function by $b_j : \mathcal{X} \to \{0,1\}$, so that $b_j(x)$ is the j-th bit in the binary label of modulation symbol x, for $j = 1, \ldots, m$. With slight abuse of notation, we let $B_1, \ldots, B_m$ and $b_1, \ldots, b_m$ denote the random variables and their corresponding realizations of the bits in a given label position $j = 1, \ldots, m$. In this paper, we consider the case where the modulation symbols x are used with probabilities

$$P_X^{\text{bicm}}(x) = \prod_{j=1}^{m} P_{B_j}(b_j(x)), \tag{4}$$

where $P_{B_j}(b)$ is the probability of the jth bit being equal to b. Finally, we denote the probability of symbols with bit b in the j-th position of the label by $P_{X|B_j}(x|b)$. By construction, $P_{X|B_j}(x|b)$ is zero if $b_j(x) = b$.

The main difference between CM and BICM is at the decoder end. The BICM decoder treats each of the m bits in a symbol as independent, yielding

$$q(x, y) = \prod_{j=1}^{m} q_j(b_j(x), y), \tag{5}$$

where the jth bit decoding metric $q_j(b, y)$ is given by

$$q_j(b, y) = \sum_{x' \in \hat{\mathcal{X}}} P_{Y|X}(y|x') Q_{X|B_j}(x'|b), \tag{6}$$

where $\hat{\mathcal{X}}$ and $Q_{X|B_j}(x|b)$ respectively denote the specific reference constellation and the symbol probabilities used for decoding at the receiver, that is not necessarily those used at the transmitter, i.e., $\mathcal{X}$ and $P_{X|B_j}(x|b)$. Mapping is also considered on $\hat{\mathcal{X}}$. We denote the reference inverse labeling function at the decoder by $\hat{b}_j : \hat{\mathcal{X}} \to \{0,1\}$. For the cases we considered, $Q_{X|B_j}(x|b)$ is non-zero only when $\hat{b}_j(x) = b$.

III. Achievable Rates

The largest achievable rate with mismatched decoding is not known in general. However, the GMI has been shown to be the largest achievable rate with i.i.d. codebooks [5], [6]. For BICM with fixed input symbol probability mass function (pmf) $P_X^{\text{bicm}}(x)$ the GMI is given by [7]

$$I_{\text{bicm}}(P_X^{\text{bicm}}) = \sup_{s>0} I_{\text{bicm}}(s, P_X^{\text{bicm}}), \tag{7}$$

with

$$I_{\text{bicm}}(s, P_X^{\text{bicm}}) = \sum_{j=1}^{m} \mathbb{E}\left[\log \frac{q_j(B, Y)^s}{\sum_{b'=0}^{1} P_{B_j}(b') q_j(b', Y)^s}\right], \tag{8}$$

[1]This interleaver can be safely ignored in our analysis as it has been absorbed in the description of the random coding ensemble.

TABLE I
BICM Schemes of Interest

Schemes	$P_X^{\text{bicm}}(x)$	$\hat{\mathcal{X}}$	$Q_{X\|B_j}(x\|b)$
BICM$_0$	$\frac{1}{M}$	$\mathcal{X}$	$P_{X\|B_j}(x)$
BICM$_1$	optimized	$\mathcal{X}$	$P_{X\|B_j}(x)$
BICM$_2$	optimized	$\bar{\mathcal{X}}$	$\frac{2}{M}\mathbb{1}\left\{\hat{b}_j(x) = b\right\}$
BICM$_3$	optimized	$\mathcal{X}$	$\frac{2}{M}\mathbb{1}\left\{\hat{b}_j(x) = b\right\}$

$\mathbb{1}\{\cdot\}$ denotes the indicator function.

where the expectation is carried out according to $P_{B_j, Y|B_j}$.

Since Eq. (8) gives an achievable rate for the distribution P_X^{bicm}, one can find the input bit distribution with largest GMI, resulting in probabilistic shaping, by solving

$$\mathsf{C}^{\text{bicm}} = \sup_{\substack{s>0 \\ P_X^{\text{bicm}}}} I_{\text{bicm}}(s, P_X^{\text{bicm}}). \tag{9}$$

Note that maximization over $P_X^{\text{bicm}}(x)$ is equivalent to maximization over $P_{B_1}(b), \ldots, P_{B_m}(b)$.

Due to the non-convex, non-concave nature of the BICM GMI, there is no simple algorithmic way to find the suboptimal values of s and P_X^{bicm}. For discrete memoryless channels, we derived a Blahut-Arimoto type of algorithm that finds the optimal $P_X^{\text{bicm}}(x)$ for fixed s and guarantees local optimality [9]. As $I_{\text{bicm}}(s, P_X^{\text{bicm}})$ is a concave function of s, one can also optimize over s without resorting to exhaustive search.

IV. The AWGN Channel

In this section, we focus on BICM for the AWGN channel, and study both low- and high-SNR regimes.

A. Channel Model and Achievable Rate

We consider transmission over the complex ($\mathcal{X} \subset \mathbb{C}, \mathcal{Y} = \mathbb{C}$) AWGN channel for which

$$y_k = \sqrt{\mathsf{snr}}\, x_k + z_k \quad k = 1, \ldots, N, \tag{10}$$

where z_k are realizations of an i.i.d. circularly symmetric complex Gaussian random variable with zero mean and unit variance, and snr is the average SNR. Codewords are subject to a power constraint $\frac{1}{N}\sum_{k=1}^{N} |x_k|^2 = 1$. In the presence of a power constraint, the largest achievable rate with i.i.d. codebooks is now

$$\mathsf{C}^{\text{bicm}}(\mathsf{snr}) = \sup_{\substack{s>0 \\ P_X^{\text{bicm}}:\ \mathbb{E}_{P_X^{\text{bicm}}}[|X|^2]=1}} I_{\text{bicm}}(\mathsf{snr}) \tag{11}$$

where

$$I_{\text{bicm}}(\mathsf{snr}) = \sum_{j=1}^{m} \mathbb{E}_{P_{B_j, X|B_j, Z}}\left[f_j(\rho, X, B_j, s)\right], \tag{12}$$

where $\rho \triangleq \sqrt{\mathsf{snr}}$ and

$$f_j(\rho, x, b, s) \triangleq \log \frac{\left(\sum_{x' \in \hat{\mathcal{X}}} e^{-|\rho(x-x')+z|^2} Q_{X|B_j}(x'|b)\right)^s}{\sum_{b'=0}^{1} P_{B_j}(b') \left(\sum_{x' \in \hat{\mathcal{X}}} e^{-|\rho(x-x')+z|^2} Q_{X|B_j}(x'|b')\right)^s}. \quad (13)$$

We have omitted the dependence of $I_{\mathrm{bicm}}(\mathsf{snr})$ on s, P_X^{bicm} for the simplicity of the notation.

Though one can specify $\hat{\mathcal{X}}$ and $Q_{X|B_j}(x|b)$ arbitrarily, we limit our interest to the schemes shown in Table I where $\bar{\mathcal{X}}$ satisfies

$$\sum_{x \in \bar{\mathcal{X}}} \frac{1}{M} |x|^2 = 1. \quad (14)$$

In summary, BICM_0 is classical BICM with equiprobable symbols, BICM_1 is BICM with optimized symbol probabilities solving (11). In BICM_2 and BICM_3, the symbol probabilities at the transmitter are optimized based on a mismatched receiver that assumes equiprobable symbols. The main difference between BICM_2 and BICM_3 is that BICM_3 knows the energy of the transmitted constellation (hence the reference constellation is exactly $\mathcal{X}$), while BICM_2 does not (hence the reference constellation and the transmitted constellation are mismatched).

Figure 1 shows the performance comparison among the four BICM schemes featured in Table I. The capacity of CM is also shown for reference. The results illustrate that the mismatched symbol pmf $Q_{X|B_j}(x|b)$ and/or mismatched constellation $\hat{\mathcal{X}}$ are efficiently compensated by shaping at low SNR. On the other hand, in the mid-to-high range of SNR, BICM_2 performs only slightly better than BICM_0. We also observe that as soon as the reference constellation matches the transmit constellation (As in BICM_3) while keeping equiprobable symbol probabilities at the decoder, the gap can be recovered almost in full.

B. Low-SNR Regime

The GMI of BICM at mid-to-high-SNR can only be evaluated by numerical experiments. On the other hand, at low SNR, the GMI can be analyzed in closed form. The GMI admits a Taylor expansion series in terms of snr,

$$I_{\mathrm{bicm}}(\mathsf{snr}) = c_1 \mathsf{snr} + c_2 \mathsf{snr}^2 + \mathrm{o}(\mathsf{snr}^2). \quad (15)$$

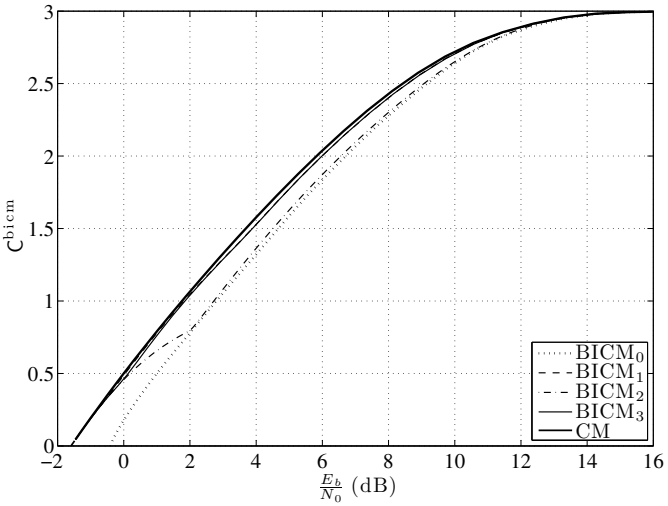

Fig. 1. Comparison of $\mathsf{C}^{\mathrm{bicm}}$ among different BICM schemes with 8PAM modulation with Gray labeling. Schemes are shown as in Table I.

A scheme is said to be first- and second-order optimal if $c_1 = 1$ and $c_2 = -\frac{1}{2}$ [10].

The low-snr performance of BICM was studied in [11], where expressions of c_1 and c_2 were given for general mapping rules and equiprobable signaling with the classical BICM decoder (5). The result was exploited in [7] to show that BICM with Gray labeling and shaping is first- and second-order optimal in the wideband regime. To generalize the result in [11] and to determine how the multiple sources of mismatch introduced in this paper affect c_1 and c_2, we start directly with the expression of $I_{\mathrm{bicm}}(\mathsf{snr})$ defined in Eq. (12). For simplicity, we use natural logarithms. In order to find c_1 and c_2, we need the second- and forth-order derivatives of $f_j(\rho, x, b, s)$ with respect to ρ, respectively.

The first-order derivative of $f_j(\rho, x, b, s)$ with respect to ρ is given by [9]

$$f_j'(\rho, x, b, s) = \frac{-s\alpha_j(\rho, x, z, b)}{\beta_j(\rho, x, z, b)} + \frac{\sum_{b'=0}^{1} P_{B_j}(b') s \beta_j^{s-1}(\rho, x, z, b') \alpha_j(\rho, x, z, b')}{\sum_{b'=0}^{1} P_{B_j}(b') \beta_j^s(\rho, x, z, b')}, \quad (16)$$

$$f_j''(\rho, x, b, s) = \frac{s\lambda_j(\rho, x, z, b)\beta_j(\rho, x, z, b) - s\alpha_j^2(\rho, x, z, b)}{\beta_j^2(\rho, x, z, b)} + \left(\frac{\sum_{b'=0}^{1} P_{B_j}(b') s \beta_j^{s-1}(\rho, x, z, b') \alpha_j(\rho, x, z, b')}{\sum_{b'=0}^{1} P_{B_j}(b') \beta_j^s(\rho, x, z, b')}\right)^2 - \frac{\left(\sum_{b'=0}^{1} P_{B_j}(b') \beta_j^s(\rho, x, z, b')\right)\left(\sum_{b'=0}^{1} P_{B_j}(b')(s^2 - s)\beta_j^{s-2}(\rho, x, z, b')\alpha_j^2(\rho, x, z, b')\right)}{\left(\sum_{b'=0}^{1} P_{B_j}(b') \beta_j^s(\rho, x, z, b')\right)^2} - \frac{\left(\sum_{b'=0}^{1} P_{B_j}(b') \beta_j^s(\rho, x, z, b')\right)\left(\sum_{b'=0}^{1} P_{B_j}(b') s \beta_j^{s-1}(\rho, x, z, b')\lambda_j(\rho, x, z, b')\right)}{\left(\sum_{b'=0}^{1} P_{B_j}(b') \beta_j^s(\rho, x, z, b')\right)^2} \quad (21)$$

where,

$$\alpha_j(\rho,x,z,b) \triangleq \sum_{x'\in\hat{\mathcal{X}}} e^{-|\rho(x-x')+z|^2} Q_{X|B_j}(x'|b)\gamma(\rho,x',x,z), \tag{17}$$

$$\beta_j(\rho,x,z,b) \triangleq \sum_{x'\in\hat{\mathcal{X}}} e^{-|\rho(x-x')+z|^2} Q_{X|B_j}(x'|b), \tag{18}$$

$$\gamma(\rho,x',x,z) \triangleq 2z_r(x_r-x'_r)+2z_i(x_i-x'_i)+2\rho\kappa(x,x'), \tag{19}$$

$$\kappa(x,x') \triangleq -2|x-x'|^2, \tag{20}$$

x_r, x_i and z_r, z_i denote the real and imaginary parts of x and z respectively. The second-order derivative of $f_j(\rho,x,b,s)$ with respect to ρ is given in Eq. (21) at the bottom of the previous page, with

$$\lambda_j(\rho,x,z,b) \triangleq \sum_{x'\in\hat{\mathcal{X}}} e^{-|\rho(x-x')+z|^2} Q_{X|B_j}(x'|b) \times \big(\gamma(\rho,x',x,z)^2+\kappa(x,x')\big). \tag{22}$$

Expanding eq. (12) using Taylor series with eqs. (13)-(22) and letting $\rho \to 0$ we have that for general, constellations (both transmitted and reference), input pmfs (both transmitted and reference), mapping and value of s

$$c_1 = a_1 s^2 - a_2(s^2-s) - a_3 s, \tag{23}$$

where

$$a_1 \triangleq \sum_{j=1}^{m} \mathbb{E}_{P_X}\left[\left|X-\mathbb{E}_{P_{B_j}Q_{X|B_j}}[X]\right|^2\right], \tag{24}$$

$$a_2 \triangleq \sum_{j=1}^{m} \mathbb{E}_{P_X,P_{B_j}}\left[\left|X-\mathbb{E}_{Q_{X|B_j}}[X]\right|^2\right], \tag{25}$$

$$a_3 \triangleq \sum_{j=1}^{m} \mathbb{E}_{P_{B_j,X|B_j}}\left[\left|X-\mathbb{E}_{Q_{X|B_j}}[X]\right|^2\right]. \tag{26}$$

By maximizing c_1 over s we find that

$$c_1^* = -\frac{(a_2-a_3)^2}{4(a_1-a_2)}. \tag{27}$$

In order to find c_2 in expansion (15), we need the fourth-order derivative of $f_j(\rho,x,b,s)$ with respect to ρ. Due to page limitations, the derivation and expression of c_2 are omitted. For BICM_0, our expressions of c_1 (23) and c_2 recover the respective expressions provided in [11].

Figure 2 shows C^{bicm} and the corresponding wideband regime expansion (15) for 4 BICM_3 configurations (one with optimal and 3 with randomly picked values of s and bit probabilities). The results confirm the accuracy of our wideband regime analysis of the general BICM schemes.

C. High-SNR Regime

Both mutual information and GMI are upper bounded by $H(X)$, the input entropy. Furthermore, for high SNR the equivocation $H(X|Y)$ tends to zero, and therefore, the mutual information of CM saturates at $H(X)$, making equiprobable inputs optimal at high SNR. In this section, we analyze the high-SNR regime of the BICM schemes described previously. In particular, we will pay particular attention to BICM_2, where the reference constellation is different from the transmitted one. To do this, we rewrite the expression of $I_{\text{bicm}}(\mathsf{snr})$ in (11) as

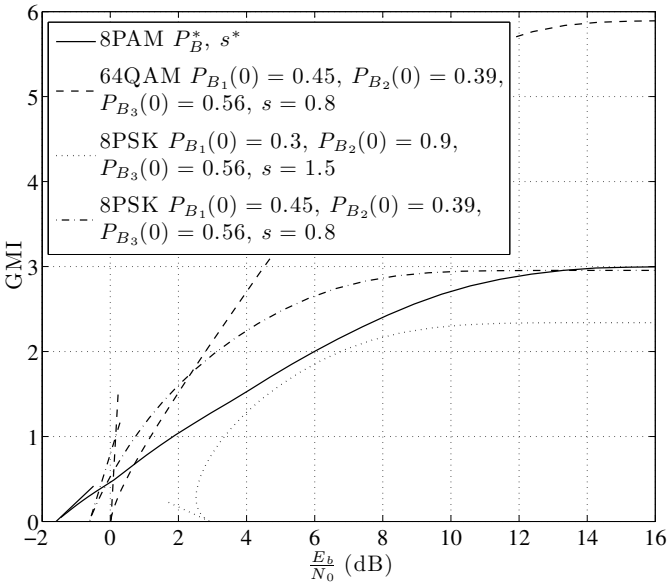

Fig. 2. BICM performance in the wideband regime. 4 different BICM configurations with Gray labeling are considered. P_B^* and s^* denote the optimal bit probabilities and s. For 64QAM, we assume symmetry between in-phase and quadrature, i.e., $P_{B_1}(0) = P_{B_4}(0)$, $P_{B_2}(0) = P_{B_5}(0)$, $P_{B_3}(0) = P_{B_6}(0)$.

$$I_{\text{bicm}}(\mathsf{snr}) = H(X) - \hat{H}_{\mathsf{snr}}(X|Y), \tag{28}$$

where we isolated the snr-dependent term, $\hat{H}_{\mathsf{snr}}(X|Y)$,

$$\hat{H}_{\mathsf{snr}}(X|Y) \triangleq -\sum_{j=1}^{m} \mathbb{E}_{P_{B_j,X|B_j,Z}}\left[\log\big(P_{B_j}(B)\big)+f_j(\rho,X,B,s)\right]. \tag{29}$$

We now focus on the high-SNR behavior of $\hat{H}_{\mathsf{snr}}(X|Y)$. When $\mathsf{snr}\to\infty$, for fixed b and y (or x), the channel tends to a noiseless channel, and the limiting value of the log term in (29) can be calculated as

$$\lim_{\mathsf{snr}\to\infty} f_j(\rho,x,b,s) = \log \frac{Q_{X|B_j}\big(\hat{x}_{j,b}(x)|b\big)^s e^{-s|x-\hat{x}_{j,b}(x)|^2}}{\sum_{b'=0}^{1} P_{B_j}(b') Q_{X|B_j}\big(\hat{x}_{j,b'}(x)|b'\big)^s e^{-s|x-\hat{x}_{j,b'}(x)|^2}}, \tag{30}$$

where

$$\hat{x}_{j,b}(x) \triangleq \underset{x'\in\hat{\mathcal{X}},b_j(x')=b}{\arg\min} |x-x'|^2. \tag{31}$$

When $\mathsf{snr}\to\infty$, the term $\hat{H}_{\mathsf{snr}}(X|Y)$ in (29) converges to (32) at the bottom of the page. This characterizes the saturating value of the GMI for high SNR for a general reference constellation. In particular, for schemes with equal transmit and reference constellation, i.e., $\hat{\mathcal{X}} = \mathcal{X}$, Eq. (30) tends to 0 at high SNR, hence Eq. (32) also tends to 0 at high SNR,

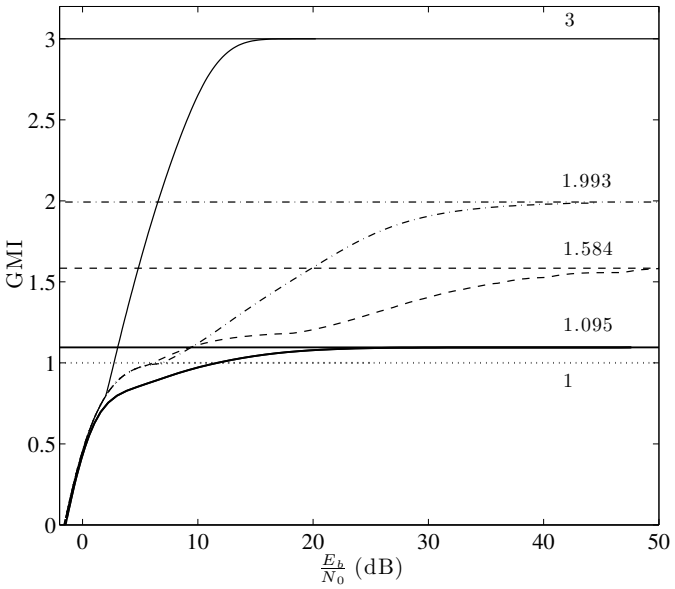

Fig. 3. High SNR analysis of different BICM_2-8PAM schemes with Gray labeling. The decoding reference constellation $\hat{\mathcal{X}} = \bar{\mathcal{X}}$ satisfies (14) for all schemes. The solid line corresponds to BICM_2. The thick solid line corresponds to fixing $P_{B_1}(0) = 0.5, P_{B_2}(0) = 1, P_{B_3}(0) = 0.5$ maximizing s. The dotted line corresponds to fixing $P_{B_2}(0) = 1, P_{B_3}(0) = 1$ and maximizing $P_{B_1}(0), s$. The dashed line corresponds to fixing $P_{B_2}(0) = 1$ and maximizing $P_{B_1}(0), P_{B_3}(0), s$. The dash-dotted line corresponds to fixing $P_{B_2}(0) = 0$ and maximizing $P_{B_1}(0), P_{B_3}(0), s$. The horizontal lines plot $\sup_s \lim_{\mathsf{snr}\to\infty}\big(H(X) - \hat{H}_{\mathsf{snr}}(X|Y)\big)$, and indicates the saturating value of the GMI for $\mathsf{snr} \to \infty$.

and $I_{\text{bicm}}(\mathsf{snr})$ saturates at $H(X)$. However, when $\hat{\mathcal{X}} \neq \mathcal{X}$, Eq. (30) may not be 0, and $I_{\text{bicm}}(\mathsf{snr})$ in Eq. (28) does not necessarily converge to $H(X)$.

Using Hölder's inequality, one can easily verify that $\lim_{\mathsf{snr}\to\infty} \hat{H}_{\mathsf{snr}}(X|Y)$ is a convex function of s. Hence, simple numerical methods can be used to optimize Eq. (28) for fixed $P_X^{\text{bicm}}(x)$. Optimization over $P_X^{\text{bicm}}(x)$ can then be performed through exhaustive search.

Figure 3 shows numerical evidence of the above analysis. In particular, we consider 8-PAM, with $\hat{\mathcal{X}} = \bar{\mathcal{X}}$. As we can see, in some cases, the resulting GMI does not saturate at the integer bits, but rather to the values predicted by (32). We can interpret the GMI for BICM_2 (thin solid line) as the envelope of all possible GMI values with a fixed input distribution. It turns out that this interpretation also gives an accurate estimation of the irregular behavior of BICM_2 for $\frac{E_b}{N_0}$ ranging from 0 to 2 dB.

V. Conclusion

We have studied the performance in terms of achievable rates of BICM with mismatched shaping strategies. We have considered schemes where the decoder employs a reference constellation that assumes that the symbols have been used with equal probability and we have studied both low- and high-SNR regimes. We have seen that the loss of this scheme with respect to a fully optimized BICM scheme is marginal at low SNR. In the case the decoder uses the same constellation as the transmitter while still assuming equiprobable symbols, the loss is negligible in all the range of SNRs.

References

[1] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, 1992.

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

[3] A. Guillén i Fàbregas, A. Martinez, and G. Caire, *Bit-Interleaved Coded Modulation*, vol. 5, Foundations and Trends on Communications and Information Theory, Now Publishers, 2008.

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

[5] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[6] A. Ganti, A. Lapidoth, and I. E. Telatar, "Mismatched decoding revisited: general alphabets, channels with memory, and the wideband limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, 2000.

[7] A. Guillén i Fàbregas and A. Martinez, "Bit-interleaved coded modulation with shaping," in *2010 IEEE Inf. Theory Workshop*, Dublin, Ireland, Sep. 2010.

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

[9] L. Peng, A. Guillén i Fàbregas, and A. Martinez, "Bit-interleaved coded modulation with shaping," *in preparation*, 2012.

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

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

$$\lim_{\mathsf{snr}\to\infty} \hat{H}_{\mathsf{snr}}(X|Y) = -\sum_{j=1}^{m}\sum_{b=0}^{1} P_{B_j}(b) \sum_{x\in\mathcal{X}} P_{X|B_j}(x) \log \frac{P_{B_j}(b) Q_{X|B_j}\big(\hat{x}_{j,b}(x)|b\big)^s e^{-s|x-\hat{x}_{j,b}(x)|^2}}{\sum_{b'=0}^{1} P_{B_j}(b') Q_{X|B_j}\big(\hat{x}_{j,b'}(x)|b'\big)^s e^{-s|x-\hat{x}_{j,b'}(x)|^2}} \tag{32}$$