Improved Exponents and Rates 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*— Mismatched decoding theory is applied to study the error exponents (both random-coding and expurgated) and achievable rates for bit-interleaved coded modulation (BICM). The gains achieved by constant-composition codes with respect to the the usual random codes are highlighted.**

I. Introduction

Random-coding techniques are typically used to prove channel coding theorems for achievable rates and exponents under maximum-likelihood (ML) and maximum-metric mismatched decoding [1]–[5]. For these problems, different random-coding ensembles, where codewords are generated in an i.i.d. manner, have been considered in the literature. Depending on the specific form of the distribution, we may distinguish between the i.i.d. ensemble, the constant-composition ensemble and the cost-constrained ensemble [1]–[6]. The i.i.d. ensemble can be used to prove the achievability of the generalized mutual information (GMI) [2], whereas the latter two can be used to prove the achievability of the higher LM rate [1], [4].

A particularly relevant instance of mismatched decoding is bit-interleaved coded modulation (BICM), introduced by Zehavi [7] as a pragmatic coding scheme for combining coding and modulation to achieve high spectral efficiency. BICM has been extensively studied in terms of its achievable rate in [8]–[13]. In particular, its GMI has been considered, and it has been shown that the input probabilities may be optimized to close the gap that made the BICM information rate suboptimal compared to coded modulation (CM) in the AWGN channel [11]–[13]. However, the LM rate and error exponent of BICM have not yet been considered in the literature. In this paper, we study the LM rate and error exponents for BICM for the constant-composition (or cost-constrained) ensemble.

II. Preliminaries

A. System Setup

We consider discrete-time transmission of information with a block code $\mathcal{M}$ of length N and rate R, where $R = \frac{1}{N}\log|\mathcal{M}|$ nats per channel use. At the encoder, a message m drawn equiprobably from the set $\{1,\dots,|\mathcal{M}|\}$ is mapped to a codeword $\boldsymbol{x}(\mathsf{m}) = \big(x_1(\mathsf{m}),\dots,x_N(\mathsf{m})\big)$, where $x_k(\mathsf{m}) \in \mathcal{X}$, and $\mathcal{X}$ is the channel input alphabet. We consider a continuous-output alphabet $\mathcal{Y}$ such that the output $\boldsymbol{y} \in \mathcal{Y}^N$ is characterized by the probability density $P_{\mathsf{Y}|\mathsf{X}}(\boldsymbol{y}|\boldsymbol{x})$. We assume a memoryless channel, therefore $P_{\mathsf{Y}|\mathsf{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{k=1}^{N} P_{Y|X}(y_k|x_k)$.

The decoder decides on the estimate of the message $\hat{\mathsf{m}}$ according to a decoding metric $q(\boldsymbol{x},\boldsymbol{y})$,

$$\hat{\mathsf{m}} = \arg\max_{\mathsf{m}} q(\boldsymbol{x}(\mathsf{m}),\boldsymbol{y}) = \arg\max_{\mathsf{m}} \prod_{k=1}^{N} q(x_k(\mathsf{m}), y_k), \quad (1)$$

where $q(x,y)$ is the symbol decoding metric. When $q(\boldsymbol{x},\boldsymbol{y})$ is a strictly increasing bijective function of $P_{\mathsf{Y}|\mathsf{X}}(\boldsymbol{y}|\boldsymbol{x})$, the decoder will always select the ML codeword. Otherwise, we have a mismatched decoder [1]–[5].

We let P_{e} denote the average error probability of the code $\mathcal{M}$ under this metric. A rate R is said achievable if, for every $\epsilon > 0$ and for N sufficiently large, there exists an encoder and decoder pair such that $\frac{1}{N}\log|\mathcal{M}| \geq R-\epsilon$ and $P_{\mathrm{e}} \leq \epsilon$. An error exponent $E_{\mathrm{r}}(R)$ is said achievable if there exists a code of length N with rate R such that $\lim_{N\to\infty} -\frac{1}{N}\log P_{\mathrm{e}} \geq E_{\mathrm{r}}(R)$.

For BICM, the codewords are obtained as the serial concatenation of a binary encoder $\mathcal{C}$ of length $n = mN$, a bit-level interleaver, and a binary labeling function $\mu : \{0,1\}^m \to \mathcal{X}$ which takes blocks of m bits and maps them to signal constellation symbols x, such that $x_k = \mu\big(b_{(k-1)m+1},\dots,b_{km}\big)$, $k = 1,\dots,N$. This interleaver can be safely ignored in our analysis as it has been absorbed in the description of the random coding ensemble. We let $M \triangleq |\mathcal{X}|$ denote the cardinality of $\mathcal{X}$ so that $m = \log_2 M$. 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 $j \in [1,m]$. With a slight abuse of notation, we let B_j and b_j denote the random variables and their corresponding realizations of the bits in a given label position j. In this paper, we consider the case where the modulation symbols x are used with probabilities

$$P_X^{\mathrm{bicm}}(x) = \prod_{j=1}^{m} P_{B_j}\big(b_j(x)\big), \quad (2)$$

where $P_{B_j}(b)$ is the probability of the j-th bit being equal to b. We denote the conditional probability of symbols given that bit B_j in the j-th position of the label is b by $P_{X|B_j}(x|b)$. By

This work has been funded in part by the European Research Council under ERC grant agreement 259663, by the European Unions 7th Framework Programme (PEOPLE-2011-CIG) under grant agreement 303633 and by the Spanish Ministry of Economy and Competitiveness under grants RYC-2011-08150 and TEC2012-38800-C03-03.

construction, $P_{X|B_j}(x|b)$ is zero if $b_j(x) \neq b$.

The BICM decoder treats each of the m bits in a symbol as independent, yielding [10]

$$q(x,y) = \prod_{j=1}^{m} q_j(b_j(x), y), \tag{3}$$

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

$$q_j(b,y) = \sum_{x' \in \hat{\mathcal{X}}} P_{Y|X}(y|x') Q_{X|B_j}(x'|b). \tag{4}$$

Here, $\hat{\mathcal{X}}$ and $Q_{X|B_j}(x|b)$ respectively denote the reference constellation and the conditional symbol probabilities used for decoding, not necessarily those used at the transmitter, namely $\mathcal{X}$ 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 in this paper, $Q_{X|B_j}(x|b)$ is non-zero only when $\hat{b}_j(x) = b$. The mapping is kept the same at both transmitter and receiver.

B. Random-coding Ensembles

In the mismatched-decoding analysis, it has proved convenient to use random coding ensembles where codewords are generated in an i.i.d. manner. Depending on the specific form of the distribution, we have the following three ensembles.

The i.i.d. ensemble: each codeword in the ensemble is composed of symbols generated i.i.d. according to, i.e.,

$$P_{\mathsf{X}}(\boldsymbol{x}) = \prod_{k=1}^{N} P_X(x_k). \tag{5}$$

The cost-constrained ensemble: each codeword satisfies a given pseudo-cost constraint, $N\bar{a} - \delta < a(\boldsymbol{x}) \leq N\bar{a}$, where $a(\boldsymbol{x}) = \sum_{k=1}^{N} a(x_k)$, $a : \mathcal{X} \to \mathbb{R}$, is the pseudo-cost function, $\bar{a} = \mathbb{E}[a(X)]$, and the constant $\delta > 0$ limits the shell on which codewords lie. Let X be an i.i.d. random vector with distribution $\tilde{P}_{\mathsf{X}}$ satisfying the pseudo-cost constraint. The codebook is constructed by [14]

$$\tilde{P}_{\mathsf{X}}(\boldsymbol{x}) = \zeta^{-1} \mathbb{1}\{N\bar{a} - \delta < a(\boldsymbol{x}) \leq N\bar{a}\} P_{\mathsf{X}}(\boldsymbol{x}), \tag{6}$$

where ζ is a normalization constant and $P_{\mathsf{X}}(\boldsymbol{x})$ is the i.i.d. distribution in (5). The pseudo-cost improves the performance of the random-coding ensemble [5], [15].

The constant-composition ensemble: each codeword is selected according to the distribution

$$\tilde{P}_{\mathsf{X}}(\boldsymbol{x}) = |T(P_N)|^{-1} \mathbb{1}\{\boldsymbol{x} \in T(P_N)\}, \tag{7}$$

where P_N is the most probable type under $P_X(x)$. Codewords have the same empirical distribution since they are generated uniformly over the type class $T(P_N)$. The type class $T(P)$ is defined as the set $T(P) = \{\boldsymbol{x} \in \mathcal{X}^N : P_{\boldsymbol{x}} = P\}$, where the type $P_{\boldsymbol{x}}$ of $\boldsymbol{x}$ is the relative proportion of occurrence of each symbol of $\mathcal{X}$, i.e. its empirical distribution [16].

The i.i.d. and cost-constrained ensembles apply to both discrete and continuous alphabets, whereas the constant-composition ensemble is only valid for discrete inputs.

C. Achievable Rates

The authors in [2], [5] have proved the achievability of the GMI using the i.i.d. ensemble. The GMI has also been proved to be the largest achievable rate for the i.i.d. ensemble. The GMI for a given input distribution P_X is given by

$$I_0(P_X) \triangleq \sup_{s \geq 0} \mathbb{E}\left[\log \frac{q(X,Y)^s}{\sum_{x'} P_X(x') q(x',Y)^s}\right]. \tag{8}$$

The cost-constrained ensemble and constant-composition ensemble are used to prove a higher achievable rate. This rate is known as the LM rate, and is given by [1], [16]

$$I_1(P_X) \triangleq \sup_{s \geq 0, a(\cdot)} \mathbb{E}\left[\log \frac{q(X,Y)^s e^{a(X)}}{\sum_{x'} P_X(x') q(x',Y)^s e^{a(x')}}\right]. \tag{9}$$

The LM rate is also the largest achievable rate for the constant-composition [3] and cost-constrained ensembles with discrete alphabets [6]. Since the GMI can be recovered from the LM rate by choosing $a(x) = 0$ for $x \in \mathcal{X}$, we have that

$$I_0(P_X) \leq I_1(P_X). \tag{10}$$

D. Error Exponents

We denote the random-coding error exponents at rate R for the ensembles defined in (5)–(6) by $E_{\mathrm{r},0}(R)$ and $E_{\mathrm{r},1}(R)$ respectively. They are given by given by [6]

$$E_{\mathrm{r},0}(R) \triangleq \sup_{\substack{0 \leq \rho \leq 1 \\ s \geq 0}} E_0(\rho, s) - \rho R \tag{11}$$

$$E_{\mathrm{r},1}(R) \triangleq \sup_{\substack{0 \leq \rho \leq 1 \\ s \geq 0, r, \bar{r}, a(\cdot)}} E_1\big(\rho, s, r, \bar{r}, a(\cdot)\big) - \rho R, \tag{12}$$

where the functions E_0 and E_1 are respectively given by

$$E_0(\rho, s) \triangleq -\log \mathbb{E}\left[\left(\sum_{x'} P_X(x') \frac{q(x',Y)^s}{q(X,Y)^s}\right)^{\rho}\right] \tag{13}$$

$$E_1\big(\rho, s, r, \bar{r}, a(\cdot)\big) \triangleq -\log \mathbb{E}\left[\frac{e^{r a(X)}}{e^{r\bar{a}}} \left(\sum_{x'} P_X(x') \frac{q(x',Y)^s e^{\bar{r} a(x')}}{q(X,Y)^s e^{\bar{r}\bar{a}}}\right)^{\rho}\right]. \tag{14}$$

For discrete memoryless channels, both exponents are tight for the respective code ensembles under consideration [6]. The exponent $E_{\mathrm{r},1}(R)$ in (12) is a refinement of the following exponent with cost constraint by Shamai and Sason [15]

$$E_{\mathrm{r},1'}(R) \triangleq \sup_{\substack{0 \leq \rho \leq 1 \\ s \geq 0, a(\cdot)}} E_{1'}\big(\rho, s, a(\cdot)\big) - \rho R, \tag{15}$$

where

$$E_{1'}\big(\rho, s, a(\cdot)\big) \triangleq -\log \mathbb{E}\left[\left(\sum_{x'} P_X(x') \frac{q(x',Y)^s e^{a(x')}}{q(X,Y)^s e^{a(X)}}\right)^{\rho}\right] \tag{16}$$

$$= E_1\big(\rho, s, -\rho, 1, a(\cdot)\big). \tag{17}$$

Thus, $E_{\mathrm{r},1'}(R) \leq E_{\mathrm{r},1}(R)$ always holds. The GMI and LM rates can be recovered from the exponents as

$$I_0(P_X) = \sup_{0\le\rho\le1, s\ge0} \frac{E_0(\cdot)}{\rho} = \frac{d \sup_{s\ge0} E_0(\cdot)}{d\rho}\bigg|_{\rho=0} \tag{18}$$

$$I_1(P_X) = \sup_{\substack{s\ge0, a(\cdot)\\ 0\le\rho\le1}} \frac{E_{1'}(\cdot)}{\rho} = \sup_{\substack{0\le\rho\le1\\ s\ge0, a(\cdot)}} \frac{E_1(\cdot)}{\rho}\bigg|_{\substack{r=-\rho\\ \bar{r}=1}} \tag{19}$$

$$= \frac{d \sup_{s\ge0, a(\cdot)} E_{1'}(\cdot)}{d\rho}\bigg|_{\rho=0} = \frac{d \sup_{s\ge0, a(\cdot)} E_1(\cdot)}{d\rho}\bigg|_{\substack{r=-\rho\\ \bar{r}=1\\ \rho=0}}. \tag{20}$$

The exponent $E_{\mathrm{r},0}(R)$ is usually called the GMI exponent. We also have the following properties for the above exponents.

Proposition 2.1: For a fixed input distribution, $s \ge 0$, $\rho \in [0,1]$, r, $\bar{r}$ and $a(x)$ for $x \in \mathcal{X}$,

1) $E_0(\rho, s)$, is concave in s and in ρ.
2) $E_{1'}\big(\rho, s, a(\cdot)\big)$ is jointly concave in s and $a(x)$, and is a concave function of ρ.
3) $E_1\big(\rho, s, r, \bar{r}, a(\cdot)\big)$ is jointly concave in s and $a(x)$, jointly concave in s, r and $\bar{r}$, and concave in ρ.

Proof: The proposition can be simply proved by using the Hölder's inequality with the definition of concavity. ■

In [6], the authors have introduced the cost-constrained ensemble with L constraints to improve on $E_{\mathrm{r},1}(R)$. For a distribution of the cost-constrained codewords given by

$$\tilde{P}_{\mathsf{X}}(\boldsymbol{x}) = \frac{\mathbb{1}\{N\bar{a}_k - \delta_k < a_k(\boldsymbol{x}) \le N\bar{a}_k, k=1,\dots,L\}\, P_{\mathsf{X}}(\boldsymbol{x})}{\zeta}$$
$$\le \zeta^{-1} P_{\mathsf{X}}(\boldsymbol{x}) e^{\sum_{k=1}^L r_k(a_k(\boldsymbol{x}) - N\bar{a}_k + \delta_k)}, \tag{21}$$

they use the fact that ζ decays polynomially in N to show that the corresponding error exponent is given by

$$E_{\mathrm{r},L}(R) \triangleq \sup_{\substack{0\le\rho\le1, s\ge0\\ \{r_k\},\{\bar{r}_k\},\{a_k(\cdot)\}}} E_L\big(\rho, s, \{r_k\}, \{\bar{r}_k\}, \{a_k(\cdot)\}\big) - \rho R, \tag{22}$$

where

$$E_L\big(\rho, s, \{r_k\}, \{\bar{r}_k\}, \{a_k(\cdot)\}\big) \triangleq$$
$$-\log \mathbb{E}\left[\left(\frac{\mathbb{E}\left[q(X',Y)^s e^{\sum_{k=1}^L \bar{r}_k\left(a_k(X') - \bar{a}_k\right)} | Y\right]}{q(X,Y)^s e^{-\sum_{k=1}^L \frac{r_k}{\rho}(a_k(X) - \bar{a}_k)}}\right)^{\rho}\right], \tag{23}$$

$\bar{a}_k = \mathbb{E}[a_k(X)]$, $\{a_k(\cdot)\}$ denotes $\{a_1(\cdot), \dots, a_L(\cdot)\}$, and similarly for $\{r_k\}$ and $\{\bar{r}_k\}$. The authors in [6] have shown that the L cost-constrained ensemble contains the constant-composition ensemble as a special case, and that the constant-composition ensemble exponent can be recovered using the cost-constrained ensemble with at most two cost constraints.

A known drawback of the random-coding ensembles described above is that at low rates they suffer from the effect of some exceedingly bad codes having two or more code vectors which are identical. Expurgation often is used to mitigate this effect and to improve the bound at low rates. By following similar steps to those used by Gallager [14, Section 5.7], we can derive mismatched-decoding expurgated exponents for the i.i.d. and the L-cost-constrained ensembles, as summarized in the following theorem.

Theorem 2.1: For an arbitrary memoryless channel, there exist codes from the expurgated cost-constrained random coding ensemble with length N, rate R and pseudo-cost function $a_k(\cdot)$ with $k = 1, \dots, L$ for which, for a given input distribution P_X and all messages

$$P_{\mathrm{e}} \le e^{-N E^{\mathrm{x}}_{\mathrm{r},L}\left(R + \frac{1}{N}\log\frac{4}{\zeta^2} e^{\sum_{k=1}^L \delta_k(r_k + \bar{r}_k)}\right)}, \tag{24}$$

where the function $E^{\mathrm{x}}_{\mathrm{r},L}(\cdot)$ is given by

$$E^{\mathrm{x}}_{\mathrm{r},L}(R) \triangleq \sup_{\substack{s\ge0, \rho\ge1,\\ \{r_k\},\{\bar{r}_k\},\{a_k(\cdot)\}}} E^{\mathrm{x}}_L\big(\rho, s, \{r_k\}, \{\bar{r}_k\}, \{a_k(\cdot)\}\big) - \rho R \tag{25}$$

$$E^{\mathrm{x}}_L\big(\rho, s, \{r_k\}, \{\bar{r}_k\}, \{a_k(\cdot)\}\big) \triangleq -\rho \log \sum_x \sum_{x'} P_X(x)$$
$$\cdot P_X(x') \frac{e^{\sum_{k=1}^L r_k a_k(x) - r_k \bar{a}_k}}{e^{-\sum_{k=1}^L \bar{r}_k a_k(x) - \bar{r}_k \bar{a}_k}} \left(\mathbb{E}_{P_{Y|X=x}}\left[\frac{q(x',Y)^s}{q(x,Y)^s}\right]\right)^{\frac{1}{\rho}}. \tag{26}$$

The corresponding result for the i.i.d. ensemble can be obtained by setting $\delta_k = 0, \zeta = 1, a_k(\cdot) = 0$ in $E^{\mathrm{x}}_{\mathrm{r},L}(R)$. Hence

$$E^{\mathrm{x}}_{\mathrm{r},0}(R) \triangleq \sup_{s\ge0, \rho\ge1} E^{\mathrm{x}}_0(\rho, s) - \rho R \tag{27}$$

$$E^{\mathrm{x}}_0(\rho, s) \triangleq -\rho \log \mathbb{E}\left[\left(\mathbb{E}\left[\frac{q(X',Y)^s}{q(X,Y)^s}\bigg| X, X'\right]\right)^{\frac{1}{\rho}}\right], \tag{28}$$

where the triplet of random variables (X, X', Y) are distributed according to $P_X(x)P_X(x')P_{Y|X}(y|x)$.

We have that

$$E^{\mathrm{x}}_{\mathrm{r},L}(R) \ge E^{\mathrm{x}}_{\mathrm{r},0}(R). \tag{29}$$

Theorem 2.2: At $R \to 0$, the expurgated exponents satisfy

$$\lim_{R\to0} E^{\mathrm{x}}_{\mathrm{r},0}(R) = \lim_{R\to0} E^{\mathrm{x}}_{\mathrm{r},L}(R)$$
$$= \sup_{s\ge0} -\mathbb{E}\left[\log\left(\mathbb{E}\left[\frac{q(X',Y)^s}{q(X,Y)^s}\bigg| X, X'\right]\right)\right]. \tag{30}$$

Proof: It is not difficult to see that

$$\lim_{R\to0} E^{\mathrm{x}}_{\mathrm{r},0}(R) = \sup_{s\ge0, \rho\ge1} E^{\mathrm{x}}_0(\rho, s) \tag{31}$$

$$\lim_{R\to0} E^{\mathrm{x}}_{\mathrm{r},L}(R) = \sup_{\substack{s\ge0, \rho\ge1,\\ \{r_k\},\{\bar{r}_k\},\{a_k(\cdot)\}}} E^{\mathrm{x}}_L\big(\rho, s, \{r_k\}, \{\bar{r}_k\}, \{a_k(\cdot)\}\big). \tag{32}$$

According to [14, Appendix 5B], $E^{\mathrm{x}}_0(\rho, s)$ for fixed s is a non-decreasing concave function of ρ for $\rho \ge 0$. Hence the supremum of $E^{\mathrm{x}}_0(\rho, s)$ for fixed s over $\rho \ge 1$ is achieved at $\rho \to \infty$. Evaluating this limit by L'Hôpital's rule, we obtained the desired expression in (30). For $E^{\mathrm{x}}_L\big(\rho, s, \{r_k\}, \{\bar{r}_k\}, \{a_k(\cdot)\}\big)$, we apply a change of variables for the L-cost case and have that

$$\begin{aligned}&\sup_{s\geq 0,\rho\geq 1,\{r_k\},\{\bar{r}_k\},\{a_k(\cdot)\}} E_L^{\mathrm{x}}\big(\rho,s,\{r_k\},\{\bar{r}_k\},\{a_k(\cdot)\}\big)\\ =&\sup_{s\geq 0,\rho\geq 1,\{r'_k\},\{\bar{r}'_k\},\{a_k(\cdot)\}} G_L\big(\rho,s,\{r'_k\},\{\bar{r}'_k\},\{a_k(\cdot)\}\big),\end{aligned} \tag{33}$$

where the function G_L is obtained from (26) as

$$\begin{aligned}G_L\big(\rho,s,\{r'_k\},&\{\bar{r}'_k\},\{a_k(\cdot)\}\big)\triangleq\\ &=E_L^{\mathrm{x}}\left(\rho,s,\left\{\frac{r'_k}{\rho}\right\},\left\{\frac{\bar{r}'_k}{\rho}\right\},\{a_k(\cdot)\}\right).\end{aligned} \tag{34}$$

The function G_L is a nondecreasing concave function of ρ for $\rho\geq 0$ and fixed s. Therefore, (33) can be evaluated similarly using L'Hôpital's rule. ■

III. BICM Rates and Exponents in the Gaussian Channel

In this section, we evaluate the achievable rates when the input probabilities are optimized and error exponents for the BICM transmission over the AWGN channel for which

$$y_k=\sqrt{\mathsf{snr}}x_k+z_k \quad k=1,\ldots,N, \tag{35}$$

where z_k are realizations of an i.i.d. Gaussian random variable with zero mean and unit variance, and snr is the average signal-to-noise ratio (SNR). Codewords are subject to a power constraint $N^{-1}\sum_{k=1}^{N}|x_k|^2=1$.

We are interested in solving the following problem

$$\mathsf{C}_k^{\mathrm{bicm}}=\sup_{P_X^{\mathrm{bicm}}:\mathbb{E}_{P_X^{\mathrm{bicm}}}[|X|^2]\leq 1} I_k^{\mathrm{bicm}}(P_X^{\mathrm{bicm}}),\ k=0,1 \tag{36}$$

where $I_k^{\mathrm{bicm}}(P_X^{\mathrm{bicm}})$, $k=0,1$ is obtained by substituting the BICM decoding metric in (3) and (4) into the GMI or LM rate expression in (8) and (9). Moreover, the channel input distribution P_X^{bicm} satisfies (2).

Table I shows the schemes of interest. $\mathcal{X}$ is the constellation used at the transmitter, and we have $\sum_{x\in\bar{\mathcal{X}}}M^{-1}|x|^2=1$. In short, BICM_1 is the classical BICM in [8] with equiprobable symbols. When the input probabilities are optimized for BICM_1, we get the optimized BICM scheme in [11], [12]. In BICM_2 and BICM_3, both mismatched receivers assume equiprobable $Q_{X|B_j}(x|b)$. The difference between them is that the reference constellation at the receiver $\hat{\mathcal{X}}=\mathcal{X}$ for BICM_3, while for BICM_2 $\hat{\mathcal{X}}$ is normalized. Note that BICM_2 and BICM_3 are the same only when equiprobable inputs are used.

Fig. 1 shows the rates for the schemes in Table I when the input probabilities are optimized. The LM rate has virtually no improvement over the GMI for BICM_1 with equiprobable inputs and BICM_3 with optimized $P_X^{\mathrm{bicm}}(x)$ over all SNR rage. However, for BICM_2 with optimized $P_X^{\mathrm{bicm}}(x)$, at moderate SNRs the LM rate improves by as much as the equivalent of 0.5 dB (the decoder in BICM_2 is the most mismatched among the schemes in Table I).

TABLE I
BICM Schemes of Interest

Schemes	$\hat{\mathcal{X}}$	$Q_{X\|B_j}(x\|b)$
BICM_1	$\mathcal{X}$	$P_{X\|B_j}(x\|b)=\frac{P_X^{\mathrm{bicm}}(x)}{P_{B_j}(b)}\mathbb{1}\left\{\hat{b}_j(x)=b\right\}$
BICM_2	$\bar{\mathcal{X}}$	$\frac{2}{M}\mathbb{1}\left\{\hat{b}_j(x)=b\right\}$
BICM_3	$\mathcal{X}$	$\frac{2}{M}\mathbb{1}\left\{\hat{b}_j(x)=b\right\}$

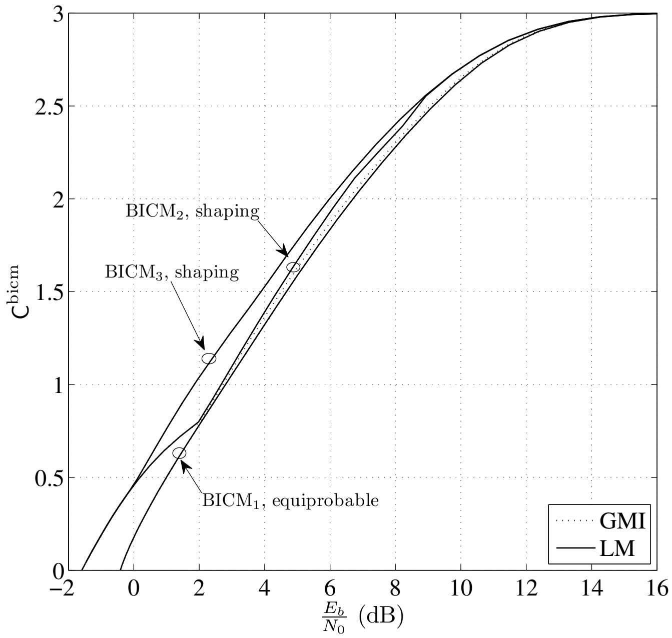

Fig. 1. Comparison of $\mathsf{C}_1^{\mathrm{bicm}}$ and $\mathsf{C}_0^{\mathrm{bicm}}$ among different BICM schemes with 8PAM modulation and Gray labeling. Schemes are shown as in Table I.

In Fig. 2, we plot the three exponents for BICM_1 with equiprobable channel inputs for $\mathsf{snr}=5$ dB. The exponents of CM transmission are also shown for reference. The markers on the x axis show $I(P_X)$, $I_0^{\mathrm{bicm}}(P_X^{\mathrm{bicm}})$ and $I_1^{\mathrm{bicm}}(P_X^{\mathrm{bicm}})$. We also see that though the LM rate improves on the GMI only marginally, the change in exponents $E_{\mathrm{r},1}(R)$ and $E_{\mathrm{r},1'}(R)$ is more pronounced. Similar to the observation by Scarlett [6], the exponent $E_{\mathrm{r},1}(R)$ improves over the exponent $E_{\mathrm{r},1'}(R)$.

In Fig. 3, we show the error exponents of BICM 8PAM transmission for $\mathsf{snr}=5$dB. We observe an improvement of $E_{\mathrm{r},L}(R)$ ($L=2$) over $E_{\mathrm{r},1}(R)$. The BICM scheme we simulated in this example has the decoding metric calculated from the following conditional symbol probability,

$$Q_{X|B_j}(x|b)=\frac{P_X^{\mathrm{bicm}}(x)}{P_{B_j}(\bar{b})}\mathbb{1}\left\{\hat{b}_j(x)=b\right\}, \tag{37}$$

where $\bar{b}$ denotes the binary complement of b. This is an artificially designed mismatched decoder whose purpose is to highlight the relative merits of the different random-coding ensembles. We have also used a sub-optimal non-equiprobable distribution $P_{B_1}(0)=0.5, P_{B_2}(0)=0.6, P_{B_3}(0)=0.4$, in order to further illustrate the difference in performance, because the closer the input distribution to the optimal one the less the gain one could have with multiple cost constraints.

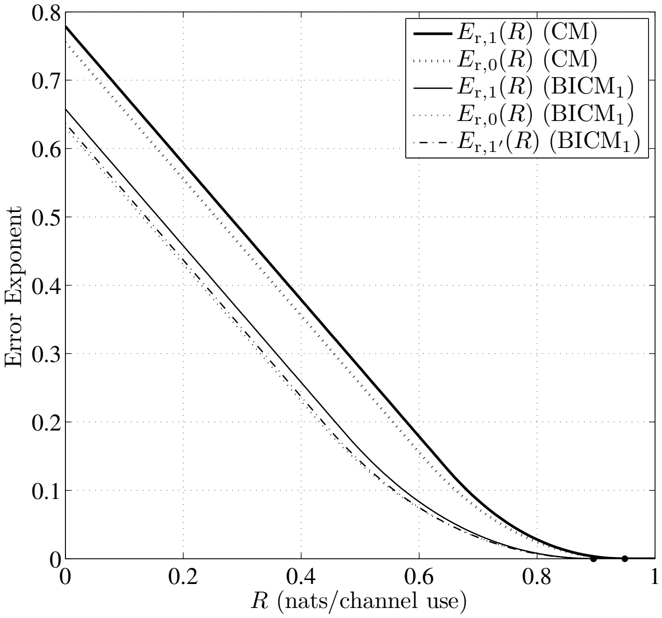

Fig. 2. Error exponents for BICM$_1$ and CM with 8PAM (AWGN channel with equiprobable input distritution at $\mathsf{snr} = 5$ dB). The rates $I(P_X)$, $I_0^{\mathrm{bicm}}(P_X^{\mathrm{bicm}})$ and $I_1^{\mathrm{bicm}}(P_X^{\mathrm{bicm}})$ are marked on the R axis.

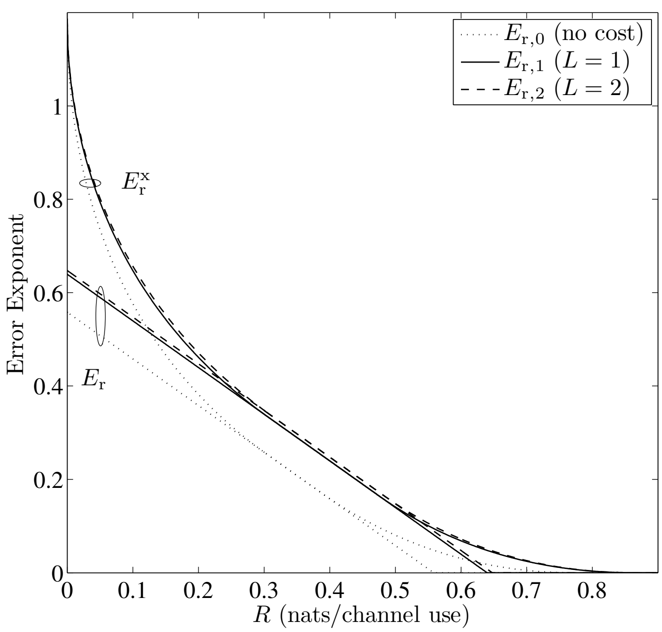

Fig. 3. Error exponents $E_{r,0}(R)$, $E_{r,L}(R)$ $(L = 1, 2)$, $E^x_{r,0}(R)$, and $E^x_{r,L}(R)$ $(L = 1, 2)$ for BICM with 8PAM (AWGN channel at $\mathsf{snr} = 5$ dB).

We also show in Fig. 3 the expurgated error exponents $E^x_{r,0}(R)$, $E^x_{r,L}(R)$ $(L = 1, 2)$ for the same BICM 8PAM transmission over AWGN channel at $\mathsf{snr} = 5$ dB. For rates where the optimal ρ is $\rho^\star = 1$, the random-coding and the expurgated error exponents coincide. For smaller rates, where we have that $\rho^\star > 1$, there is a significant improvement when the bad codes are expurgated from both code ensembles. In line with Theorem 2.2, the cost-constraint on the code ensemble helps us improve the expurgated error exponent for almost all rates, though this improvement vanishes as $R \to 0$.

IV. Acknowledgements

We thank Jonathan Scarlett for discussions and comments.

References

[1] J. Hui, "Fundamental Issues of Multiple Accessing," Ph.D. dissertation, MIT, 1983.

[2] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *AEU. Archiv für Elektronik und Übertragungstechnik*, vol. 47, no. 4, pp. 228–239, 1993.

[3] 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.

[4] I. Csiszár and P. Narayan, "Channel Capacity for a Given Decoding Metric," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 35–43, Jan. 1995.

[5] 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.

[6] J. Scarlett, A. Martinez, and A. G. i Fàbregas, "Ensemble-Tight Error Exponents for Mismatched Decoders," in *Fiftieth Annual Allerton Conference on Communication, Control, and Computing, Allerton, IL, 10/2012*, 2012.

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

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

[9] A. Guillén i Fàbregas, A. Martinez, and G. Caire, *Bit-Interleaved Coded Modulation*. Foundations and Trends on Communications and Information Theory, Now Publishers, 2008, vol. 5, no. 1-2.

[10] 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.

[11] A. Guillén i Fàbregas and A. Martinez, "Bit-Interleaved Coded Modulation with Shaping," in *2010 IEEE Inf. Theory Workshop*, Dublin, Ireland, Sep. 2010.

[12] 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.

[13] L. Peng, A. Guillén i Fàbregas, and A. Martinez, "Mismatched Shaping Schemes for Bit-Interleaved Coded Modulation," in *IEEE Int. Symp. Inf. Theory, Cambridge, MA, USA, July*, 2012.

[14] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, Inc. New York, NY, USA, 1968.

[15] S. S. (Shitz) and I. Sason, "Variations on the Gallager Bounds, Connections, and Applications," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3029–3051, Dec. 2002.

[16] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York-San Francisco-London: Academic Press, Inc.,, 2008.