Bit-Interleaved Coded Modulation Revisited: A Mismatched Decoding Perspective

Alfonso Martinez, Albert Guillén i Fàbregas, Giuseppe Caire and Frans Willems

***Abstract*— We revisit the information-theoretic analysis of bit-interleaved coded modulation (BICM) by modeling the BICM decoder as a mismatched decoder. The mismatched-decoding model is well-defined for finite, yet arbitrary, block lengths, and captures the channel memory among the bits belonging to the same symbol. The generalized mutual information of the mismatched decoder coincides with the infinite-interleaver BICM capacity, where BICM is modeled as a set of independent parallel binary-input channels whose output is the bitwise log-likelihood ratio. The error exponent —and hence the cutoff rate— of the BICM mismatched decoder is upper bounded by that of coded modulation and may thus be lower than in the infinite-interleaved model. For binary reflected Gray mapping in Gaussian channels the loss in error exponent is small.**

I. Introduction

The classical bit-interleaved coded modulation (BICM) decoder proposed by Zehavi in [1] uses metrics for each of the bits of a symbol based on the channel observation, rather than symbol metrics used in Ungerböck's coded modulation (CM) [2]. This decoder is sub-optimal and non-iterative, but offers very good performance and is interesting from a practical perspective due to its low implementation complexity.

Caire *et al.* [3] further elaborated on Zehavi's decoder and, under the assumption of an infinite-length interleaver, presented and analyzed a BICM channel model as a set of parallel independent binary-input output symmetric channels. Based on the data processing theorem [4], Caire *et al.* showed that the BICM mutual information cannot be larger than that of CM. However, and rather surprisingly a priori, they found that the cutoff rate of BICM can sometimes be larger than that of CM [5]. The error exponents for the parallel-channel model were studied in [6].

In this paper we take a closer look to the classical BICM decoder proposed by Zehavi and cast it as a mismatched decoder [7–9]. The observation that the classical BICM decoder treats the different bits in a given symbol as independent, even if they are clearly not, naturally leads to a simple model of the symbol mismatched decoding metric as the product of bit decoding metrics, in turn related to the log-likelihood ratios. Without requiring any assumptions on the interleaver length, we provide a per-symbol expression of the error exponent using the BICM mismatched metric.

We show that the generalized mutual information [7–9] of the BICM mismatched decoder yields the mutual information derived in [3] for the independent parallel channel model, giving a proof of achievability without assuming infinite interleaving; the interleaver is actually not necessary to achieve this rate. We also show that the BICM error exponent (and in particular the cutoff rate) is always upper-bounded by that of CM, as opposed to the corresponding exponent for the independent parallel channel model [3], which can sometimes be larger. To any extent, the loss in error exponent is negligible for binary reflected Gray mapping in Gaussian channels.

This paper is organized as follows. Section II introduces the system model and notation. Section III shows general results on the error exponents, including the generalized mutual information and cutoff rate as particular instances. Section IV studies the achievable rates of BICM under mismatched decoding. Numerical results and comparison with the parallel-channel models are presented in Section V. Finally, Section VI draws some concluding remarks.

II. Decoding Model

We consider transmission over a memoryless channel with transition probabilities $p(y|x)$, where $x \in \mathcal{X}, y \in \mathcal{Y}$ denote the channel input and output, respectively, and $\mathcal{X}, \mathcal{Y}$ are the corresponding alphabets. We denote by X, Y the underlying random variables. We consider a discrete input alphabet $\mathcal{X} = \{x_1, \ldots, x_M\}$, with $M \triangleq 2^m = |\mathcal{X}|$. For future use, we define the sequences $\boldsymbol{x} \triangleq (x_1, \ldots, x_N)$, $\boldsymbol{y} \triangleq (y_1, \ldots, y_N)$, the corresponding random vectors by $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively, and the input and output sets by $\boldsymbol{\mathcal{X}} \triangleq \mathcal{X}^N$, $\boldsymbol{\mathcal{Y}} \triangleq \mathcal{Y}^N$.

For coded modulation schemes, encoding and mapping to a modulation symbol are intertwined, so that the encoder output directly corresponds to a modulation symbol. In bit-interleaved coded modulation schemes, a binary labeling function $\mu : \{0,1\}^m \to \mathcal{X}$ maps a binary codeword $\boldsymbol{c}$ of a binary code $\mathcal{C}$ onto the signal constellation symbols. We define the sets $\mathcal{X}_b^j = \{x \in \mathcal{X} : b_j(x) = b\}$ as the set of signal constellation points with bit b in the j-th position of the binary label, where we have defined the inverse mapping function as for labeling position j as $b_j : \mathcal{X} \to \{0,1\}$.

For later use, we define the j-th marginal $p_j(y|b)$ as

$$p_j(y|b) \triangleq \frac{1}{|\mathcal{X}_b^j|} \sum_{x' \in \mathcal{X}_b^j} p(y|x'). \tag{1}$$

A. Guillén i Fàbregas is with the Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, UK, e-mail: guillen@ieee.org.

A. Martinez and F. Willems are with the Department of Electrical Engineering, Technische Universiteit Eindhoven, Postbus 513, 5600 MB Eindhoven,The Netherlands, e-mail: alfonso.martinez@ieee.org, f.m.j.willems@tue.nl.

G. Caire is with the Electrical Engineering Department, University of Southern California, 3740 McClintock Ave., Los Angeles, CA 90080, USA, e-mail: caire@usc.edu.

This work has been supported by the International Incoming Short Visits Scheme 2007/R2 of the Royal Society.

Assuming equally likely codewords, the decoder is a function from the set $\boldsymbol{\mathcal{Y}}$ to the space of codewords which estimates the transmitted codeword $\hat{\boldsymbol{x}}$ as

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} q(\boldsymbol{x}, \boldsymbol{y}) \tag{2}$$

where $q(\boldsymbol{x}, \boldsymbol{y})$ is the codeword decoding metric. Since the channel is memoryless, we assume a codeword metric of the form $q(\boldsymbol{x}, \boldsymbol{y}) = \prod_{n=1}^{N} q(x_n, y_n)$, namely the product of *symbol metrics*.

In CM constructions, such as Ungerböck's [2], the symbol decoding metric is proportional to the channel transition probability, that is $q(x, y) \propto p(y|x)$. The value of proportionality constant is irrelevant, as long as it is not zero.

In bit-interleaved CM, the natural model of the symbol decoding metric is given by

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

where for each $j = 1, \ldots, m$, we have a bit decoding metric

$$q_j\big(b_j(x) = b, y\big) = \sum_{x' \in \mathcal{X}_b^j} p(y|x'). \tag{4}$$

The BICM decoder treats thus each of the m bits in a symbol as if they were independent. In general, these bit metrics are statistically dependent and uses as a metric proportional to the product of the marginals $p_j(y|b)$. These marginals are the transition probabilities of m binary-input channels, with respective input the j-th bit in the binary label of the modulation symbol, and whose output is a log-likelihood ratio, or more exactly a log-metric ratio.

In practice, due to complexity limitations, we might be interested in the following lower-complexity version of (4),

$$q_j(b, y) = \max_{x \in \mathcal{X}_b^j} p(y|x). \tag{5}$$

In the log-domain this is known as the max-log approximation.

In either case, the explicit decoding rule of the BICM decoder is therefore given by

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \prod_{n=1}^{N} \prod_{j=1}^{m} q_j\big(b_j(x_n), y_n\big). \tag{6}$$

A special channel model of particular interest is the complex AWGN. The output is given by

$$Y = H\sqrt{\mathsf{snr}}X + Z, \tag{7}$$

where H is a complex-valued fading coefficient, Z a circularly symmetric unit-variance complex Gaussian random variable, and snr the signal-to-noise ratio (SNR). The input set $\mathcal{X}$ corresponds to unit energy PSK or QAM signal sets. Therefore,

$$p(y|x) = \frac{1}{\pi} e^{-|y - h\sqrt{\mathsf{snr}}x|^2}, \quad x \in \mathcal{X}.$$

The unfaded AWGN channel is obtained by letting $H = 1$ in the above. There are many other interesting examples, like orthogonal modulation, such as pulse-position modulation (PPM) or frequency-shift keying (FSK) with coherent or non-coherent detection, and the M-ary symmetric channel or the M-ary erasure channel, or a combination of the two.

III. Error Probability with Random Codes

A. Random coding exponent

The error probability of random codes under ML decoding, i. e., for a decoding metric satisfying $q(x, y) \propto p(y|x)$, was analyzed in [10] and used to prove the achievability part of the channel coding theorem. In particular, Gallager showed that there exist codes (in this case, random codes) whose error probability vanishes for sufficiently large block lengths provided that $R < C$. The error probability decreases exponentially with the block length according to a parameter called the *error exponent*. For memoryless channels Gallager found [10] that the average error probability over the random coding ensemble can be bounded as

$$\bar{P}_e \leq \exp\Big(-N\big(E_0(\rho) - \rho R\big)\Big) \tag{8}$$

where

$$E_0(\rho) \stackrel{\Delta}{=} -\log\left(\sum_y \left(\sum_x p(x) p(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right) \tag{9}$$

is the Gallager function, and $0 \leq \rho \leq 1$ is a free parameter. For continuous output, the summation over y is replaced by an integral. The tightest bound for a particular input distribution $p(X)$ is obtained by optimizing over ρ, which determines the *random coding exponent*

$$E_{\mathrm{r}}(R) = \max_{0 \leq \rho \leq 1} \; E_0(\rho) - \rho R. \tag{10}$$

Metric that the random coding exponent admits a symbolwise, or per letter, factorization when the channel is memoryless.

For memoryless channels and for codeword metrics decomposable as product of symbols metrics, that is $q(\boldsymbol{x}, \boldsymbol{y}) = \prod_{n=1}^{N} q(x_n, y_n)$, Gallager's derivation can easily be extended. Details can be found in [8]. The error probability is upper bounded by the expression

$$\bar{P}_e \leq \exp\Big(-N\big(E_0^q(\rho, s) - \rho R\big)\Big), \tag{11}$$

where the generalized Gallager function $E_0^q(\rho, s)$ is given by

$$E_0^q(\rho, s) = -\log \mathbb{E}_{X,Y}\left[\left(\sum_{x' \in \mathcal{X}} p(x') \frac{q(x', Y)^s}{q(X, Y)^s}\right)^{\rho}\right]. \tag{12}$$

For a particular input distribution $p(X)$, the random coding error exponent is then given by [8]

$$E_{\mathrm{r}}^q(R) = \max_{0 \leq \rho \leq 1} \max_{s > 0} E_0^q(\rho, s) - \rho R. \tag{13}$$

For a generic bit metric (3), Gallager's generalized function for BICM (assuming uniform inputs) is given by

$$E_0^{\mathrm{bicm}}(\rho, s) = -\log \mathbb{E}_{X,Y}\left[\left(\frac{1}{2^m} \sum_{x' \in \mathcal{X}} \prod_{j=1}^{m} \frac{q_j(b_j(x'), Y)^s}{q_j(b_j(X), Y)^s}\right)^{\rho}\right]. \tag{14}$$

B. Data processing inequality for error exponents

In [8], it was proved that the data-processing inequality holds for error exponents, in the sense that for a given input distribution we have that

$$E_0^q(\rho, s) \leq E_0(\rho) \qquad \text{for any } s > 0. \tag{15}$$

For the sake of completeness, we next rederive this result by extending Gallager's reasoning in [10] to the mismatched decoding regime. For a fixed channel observation y, the function inside the logarithm of the generalized Gallager function $E_0^q(\rho, s)$ can be rewritten as the summation over y of

$$\left(\sum_{x\in\mathcal{X}} p(x)p(y|x)q(x,y)^{-s\rho}\right)\left(\sum_{x'\in\mathcal{X}} p(x')q(x',y)^{s}\right)^{\rho}. \tag{16}$$

This form is reminiscent of the right-hand side of Hölder's inequality (see Exercise 4.15 of [10]), which can be written as $\left(\sum_i a_i b_i\right)^{1+\rho} \leq \left(\sum_i a_i^{1+\rho}\right)\left(\sum_i b_i^{\frac{1+\rho}{\rho}}\right)^{\rho}$. With the identifications $a_i = p(x)^{\frac{1}{1+\rho}} p(y|x)^{\frac{1}{1+\rho}} q(x,y)^{\frac{-s\rho}{1+\rho}}$ and $b_i = p(x)^{\frac{\rho}{1+\rho}} q(x,y)^{\frac{s\rho}{1+\rho}}$, we can lower bound Eq. (16) by the quantity

$$\sum_{x\in\mathcal{X}} p(x)p(y|x)^{\frac{1}{1+\rho}}. \tag{17}$$

Recovering the summation over the channel output y and the logarithm in Eq. (16), in general one has that $E_0^q(\rho, s) \leq E_0(\rho)$ for arbitrary $s > 0$ and any input distribution.

Note that the expression in Eq. (17) is independent of s and of the specific decoding metric $q(x, y)$. Nevertheless, evaluation of Gallager's generalized function for the specific choices $s = \frac{1}{1+\rho}$ and $q(x, y) \propto p(y|x)$ gives the lower bound, which is also Eq. (9). This gives a sufficient condition for the lower bound to be attained.

C. Independent parallel channels

In their analysis of multilevel coding and successive decoding, Wachsmann *et al.* provided the error exponents of BICM modelled as a set of parallel channels [6]. The corresponding Gallager's function, which we denote by $E_0^{\mathrm{ind}}(\rho)$, can be written as

$$E_0^{\mathrm{ind}}(\rho) = -\sum_{j=1}^{m} \log \mathbb{E}_{B,Y}\left[\left(\frac{1}{2}\sum_{b'=0}^{1} \frac{p_j(Y|b')^{\frac{1}{1+\rho}}}{p_j(Y|B)^{\frac{1}{1+\rho}}}\right)^{\rho}\right] \tag{18}$$

as corresponds to a binary-input channel with output y and bit metric matched to the transition probability $p_j(y|b)$.

This channel is only directly related to the original channel $p(y|x)$ whenever the channel output can be decomposed into parallel, independent outputs. Otherwise, since all subchannels are affected by the same noise realization y, the parallel-channel model fails to capture the statistics of the true channel. We will later provide examples for which it is either larger or smaller than the original error exponent, i. e. $E_0^{\mathrm{ind}}(\rho, s) > E_0(\rho)$ or $E_0^{\mathrm{ind}}(\rho, s) < E_0(\rho)$ for suitable values of s. The sum can thus be seen as the error exponent of a different channel, obtained by replicating the original channel m times; this channel may be more or less noisy than the original one, depending on the form of the transition probability $p(y|x)$.

D. Generalized mutual information

The largest achievable rate with mismatched decoding is not known in general. Perhaps the easiest candidate to deal with is the generalized mutual information (GMI)[7–9], given by

$$I^{\mathrm{gmi}}(X;Y) = \sup_{s>0} I_s^{\mathrm{gmi}}(X;Y), \tag{19}$$

where

$$I_s^{\mathrm{gmi}}(X;Y) = \mathbb{E}_{X,Y}\left[\log \frac{q(X,Y)^s}{\sum_{x'\in\mathcal{X}} p(x')q(x',Y)^s}\right]. \tag{20}$$

As in the case of matched decoding, this definition can be recovered from the error exponent,

$$I_s^{\mathrm{gmi}}(X;Y) = \left.\frac{dE_0^q(\rho, s)}{d\rho}\right|_{\rho=0}. \tag{21}$$

For completeness, we define the generalized cutoff rate as

$$R_0 \stackrel{\Delta}{=} E_{\mathrm{r}}(R=0) = \max_{s>0} E_0(1, s). \tag{22}$$

From (15), the generalized cut-off rate is upperbounded by the cut-off rate of the matched decoder.

For uniform input distribution, we define the *coded modulation exponent* $E_0^{\mathrm{cm}}(\rho)$ as the exponent of a decoder which uses metrics $q(x, y) \propto p(y|x)$, namely

$$E_0^{\mathrm{cm}}(\rho) = -\log \mathbb{E}_{X,Y}\left[\left(\frac{1}{2^m}\sum_{x'}\left(\frac{p(x',Y)}{p(X,Y)}\right)^{\frac{1}{1+\rho}}\right)^{\rho}\right]. \tag{23}$$

This expression is (9). This decoder achieves the rate *coded modulation capacity* $I^{\mathrm{cm}}(X;Y)$, given by

$$I^{\mathrm{cm}}(X;Y) = \mathbb{E}_{X,Y}\left[\log \frac{p(Y|X)}{\frac{1}{2^m}\sum_{x'\in\mathcal{X}} p(Y|x')}\right]. \tag{24}$$

IV. Achievable Rates with BICM Decoding

In this section we study the achievable rates of the BICM decoder. By using mismatched decoding, we apply the results from the previous section, and elaborate on the generalized mutual information. In particular, we show that the generalized mutual information is equal to the BICM capacity of [3] when the metric (4) is used. As we have seen in the previous section, the BICM decoder uses a metric of the form given in Eq. (3). We have the following results.

Theorem 4.1: The generalized mutual information of the BICM mismatched decoder is equal to the sum of the generalized mutual informations of the independent binary-input parallel channel model of BICM,

$$I^{\mathrm{gmi}}(X;Y) = \sup_{s>0} \sum_{j=1}^{m} \mathbb{E}_{B,Y}\left[\log \frac{q_j(b,Y)^s}{\frac{1}{2}\sum_{b'=0}^{1} q_j(b',Y)^s}\right]. \tag{25}$$

Proof: For a given s, and uniform inputs, i.e., $p(x) = \frac{1}{2^m}$, Eq. (20) gives

$$I_s^{\mathrm{gmi}}(X;Y) = \mathbb{E}_{X,Y}\left[\log \frac{\prod_{j=1}^{m} q_j\big(b_j(X),Y\big)^s}{\sum_{x'\in\mathcal{X}} \frac{1}{2^m}\prod_{j=1}^{m} q_j\big(b_j(x'),Y\big)^s}\right]. \tag{26}$$

We now have a closer look at the denominator in the logarithm of (26). The key observation here is that the sum over the constellation points of the product over the binary label positions can be expressed as the product over the label position is the sum of the probabilities of the bits being zero and one, i.e.,

$$\sum_{x'\in\mathcal{X}} \frac{1}{2^m} \prod_{j=1}^{m} q_j\big(b_j(x'),Y\big)^s \tag{27}$$

$$= \frac{1}{2^m} \prod_{j=1}^{m} \big(q_j(b'=0,Y)^s + q_j(b'=1,Y)^s\big) \tag{28}$$

$$= \prod_{j=1}^{m} \left(\frac{1}{2}\big(q_j(b'=0,Y)^s + q_j(b'=1,Y)^s\big)\right). \tag{29}$$

Rearranging terms in (26) we obtain,

$$I_s^{\mathrm{gmi}}(X;Y) \tag{30}$$

$$= \mathbb{E}_{X,Y}\left[\sum_{j=1}^{m} \log \frac{q_j\big(b_j(x),Y\big)^s}{\frac{1}{2}\big(q_j(b'=0,Y)^s + q_j(b'=1,Y)^s\big)}\right] \tag{31}$$

$$= \sum_{j=1}^{m} \frac{1}{2^m} \sum_{b=0}^{1} \sum_{x\in\mathcal{X}_b^j} \mathbb{E}_{Y|B=b,J=j}\left[\log \frac{q_j\big(b_j(x),Y\big)^s}{\frac{1}{2}\sum_{b'=0}^{1} q_j(b',Y)^s}\right]. \tag{32}$$

■

There are a number of interesting particular cases of the above theorem.

Corollary 4.1: For the metric in Eq. (4),

$$I^{\mathrm{gmi}}(X;Y) = \sum_{j=1}^{m} \mathbb{E}_{B,Y}\left[\log \frac{p_j(Y|B)}{\frac{1}{2}\sum_{b'=0}^{1} p_j(Y|b')}\right] \tag{33}$$

Proof: For the metric in Eq. (4), $q_j(b,y)$ is proportional to $p_j(y|b)$, and we can therefore identify the quantity

$$\mathbb{E}_{B,Y}\left[\log \frac{q_j\big(B,Y\big)^s}{\frac{1}{2}\sum_{b'=0}^{1} q_j(b',Y)^s}\right] \tag{34}$$

as the generalized mutual information of a *matched* binary-input channel with transitions $p_j(y|b)$. Then, the supremum over s is achieved at $s=1$ and we get the desired result. ■

Expression (33) coincides with the result in [3], even though we have lifted the assumption of infinite interleaving.

Corollary 4.2: For the metric in Eq. (5),

$$I^{\mathrm{gmi}}(X;Y)$$

$$= \sup_{s>0} \sum_{j=1}^{m} \mathbb{E}_{B,Y}\left[\log \frac{\big(\max_{x\in\mathcal{X}_j^B} p(y|x)\big)^s}{\frac{1}{2}\sum_{b=0}^{1}\big(\max_{x'\in\mathcal{X}_b^j} p(y|x')\big)^s}\right]. \tag{35}$$

The fundamental difference between the mutual information given in [11] and the generalized mutual information given in (35) is the optimization over s. In particular, both expressions are equal when $s=1$. Therefore, the optimization over s may induce a larger achievable rate. Furthermore, as we shall see in the examples, letting $s=1$ in the mismatched error exponent yields significant degradation.

V. Numerical Results

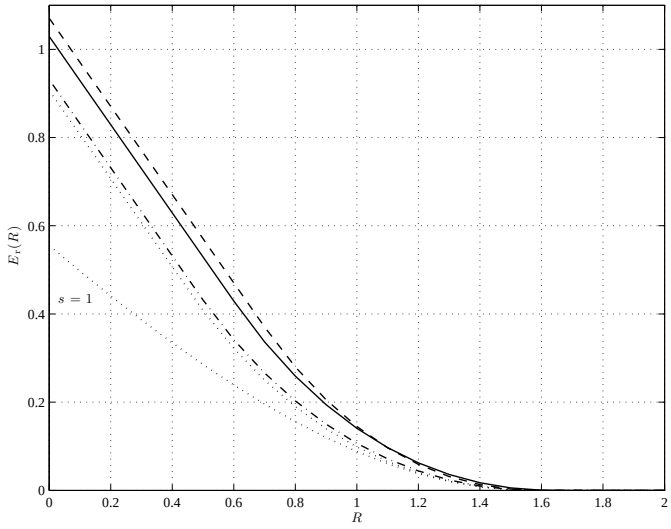
In this section we show a number of examples illustrating the error exponents of BICM and CM in AWGN channels with and without fading, at different SNR values. Figures 1(a), 1(b) and 1(c) show the error exponents for CM (solid), BICM with independent parallel channels (dashed), BICM using mismatched metric (4) (dash-dotted), and BICM using mismatched metric (5) (dotted) for 16-QAM with Gray mapping, Rayleigh fading and $\mathsf{snr} = 5, 15, -25$ dB, respectively. Dotted lines labeled with $s=1$ correspond to the error exponent of BICM using mismatched metric (5) letting $s=1$, i.e., without optimizing over s. As we observe, there is a large penalty for not optimizing over s. We further observe that the parallel channel model gives a larger exponent than the CM, in agreement with the cutoff rate results of [3]. Instead, assuming a finite-length interleaver without neglecting the dependency between the different bits yields a lower exponent than CM.

We also remark that using the metric (5) yields marginal loss in the exponent for mid-to-large SNR, while both models (independent parallel channels and finite-length interleaver with mismatched decoding) yield the same capacity. When the SNR is low, i.e., in the wideband regime [12], we observe that the independent parallel channel model and BICM using the mismatched metric (4) have the same exponent, while we observe a significant penalty when metrics (5) are used. We denote with crosses the corresponding information rates.

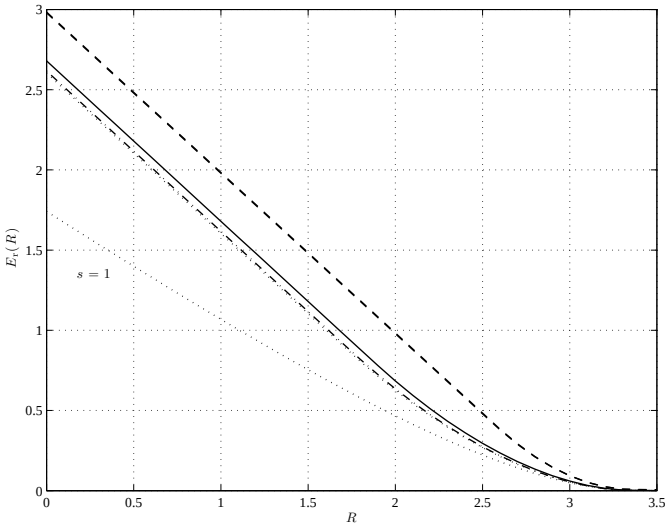
An interesting question is whether the independent parallel channel error exponent is always larger than that of BICM using mismatched decoding. To illustrate this point, Figure 1(d) shows the error exponents for CM (solid), BICM with independent parallel channels (dashed), BICM using mismatched metric (4) (dash-dotted), and BICM using mismatched metric (5) (dotted) for 8-PSK with Gray mapping in the unfaded AWGN channel. As we can see, the error exponent obtained from the parallel channel model can indeed be smaller.

VI. Conclusions

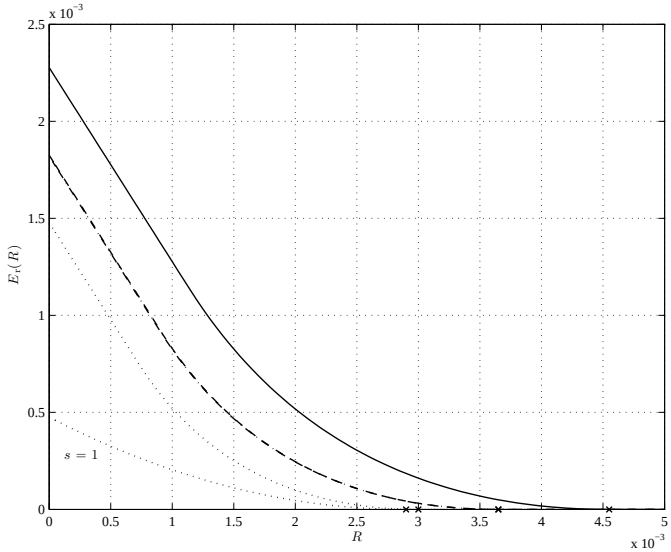
We have presented an analysis of BICM by casting the BICM decoder as a mismatched decoder. While the independent parallel channel models give the BICM mutual information [3, 6], they fail to capture the dependency among the different bits of a symbol. The characterization of the BICM decoder as a mismatched decoder captures the dependency between the different bits in a symbol, and to some extent (for random codes at least), of finite-length interleaving. We have shown that the error exponent cannot be larger than that of CM, contrary to the analysis of BICM as a set of independent parallel channels. As a consequence our analysis, the cutoff rate cannot be larger than that of CM. Note that the conclusions of [3] are purely based on the analysis of the parallel channel model which inherently ignores the dependency between the m bits of a symbol, and for which, the cutoff rate could be larger than that of CM. For Gaussian channels with binary reflected Gray mapping, the gap between the BICM and CM error exponents is small, as found by Caire *et al.* for the capacity.



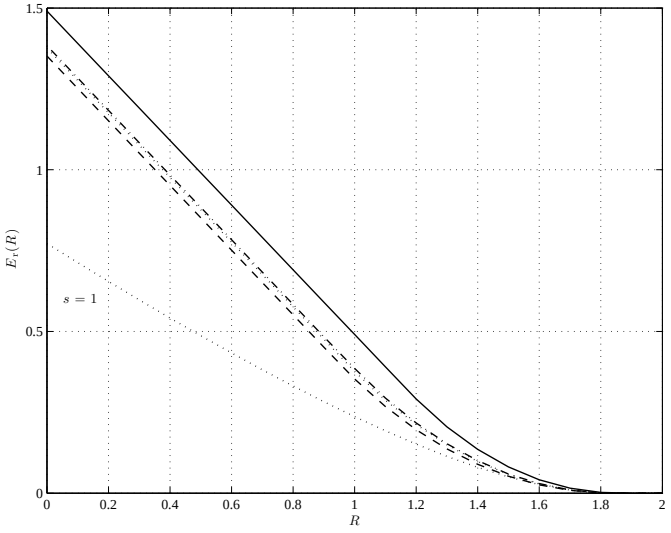
(a) 16-QAM, Rayleigh fading, and $\mathsf{snr} = 5$ dB.



(b) 16-QAM, Rayleigh fading and $\mathsf{snr} = 15$ dB.



(c) 16-QAM, Rayleigh fading and $\mathsf{snr} = -25$ dB. Crosses correspond to (from right to left) CM, BICM with metric (4), BICM with metric (5) and BICM with metric (5) with $s = 1$.



(d) 8-PSK, AWGN without fading, and $\mathsf{snr} = 5$ dB.

Fig. 1. Error exponents for CM (solid), BICM with independent parallel channels (dashed), BICM using mismatched metric (4) (dash-dotted), and BICM using mismatched metric (5) (dotted) for various modulations with Gray mapping, in AWGN channels with and without Rayleigh fading.

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. Ungerböck, "Channel Coding With Multilevel/Phase Signals.," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–66, Jan. 1982.

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

[4] T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley New York, 1991.

[5] J. M. Wozencraft and I. M. Jacobs, *Principles of communication engineering*, Wiley, 1965.

[6] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

[7] 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, 1994.

[8] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *AEU Int. J. Electron. Commun.*, vol. 47, no. 4, pp. 228–239, 1993.

[9] 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.

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

[11] L. Szczecinski, A. Alvarado, R. Feick, and L. Ahumada, "On the distribution of L-values in gray-mapped M^2-QAM signals: Exact expressions and simple approximations," in *IEEE Global Commun. Conf., Washington, DC, USA, Nov.*, 2007.

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