On the Generalized Mutual Information of BICM Systems with Approximate Demodulation

Joakim Jaldén
Signal Processing Lab, School of Electrical Engineering
Royal Institute of Technology (KTH)
100 44 Stockholm, Sweden
Email: joakim.jalden@ee.kth.se

Peter Fertl, Gerald Matz
Inst. of Communications and Radio-Frequency Engineering
Vienna University of Technology
1040 Vienna, Austria
Email: {peter.fertl,gerald.matz}@nt.tuwien.ac.at

***Abstract*—We consider a generic bit-interleaved coded modulation (BICM) systems with an approximate demodulator or log-likelihood ratio (LLR) computer. The performance of a BICM system with optimal demodulation has previously been characterized by Caire et al. in terms of the capacity of an independent parallel-channel model with binary inputs and (continuous) LLRs as outputs, and by Martinez et al. in terms of the generalized mutual information (GMI) where the BICM decoder is viewed as a mismatched decoder. Whereas GMI and capacity of the parallel-channel model coincide under optimal demodulation, they differ in general for the case of an approximate demodulator. Herein we show (i) that augmenting approximate BICM demodulators with scalar LLR correction increases the GMI and (ii) that the GMI of the LLR-corrected system coincides with the capacity of the parallel-channel model with binary inputs and outputs given by the approximate LLRs.**

I. Introduction

Bit-interleaved coded modulation (BICM) [1, 2] provides a pragmatic approach to coded modulation and has received much attention in wireless communications due to its bandwidth and power efficiency. Key to the success of the BICM paradigm is its applicability in a wide range of scenarios, achieved by separating modulation and demodulation from channel coding and decoding. The BICM receiver demodulates groups of bits, each group mapped to a single data-symbol and transmitted over a memoryless channel, and passes the resulting soft reliability information to a subsequent binary decoder; the soft information is usually represented by log-likelihood ratios (LLRs). In the binary decoder, bits within such groups are treated as independent, an assumption typically motivated by the inclusion of a bitwise interleaver.

If many bits are simultaneously transmitted (like with MIMO-BICM), optimal demodulation often becomes prohibitively complex [3]. This has motivated a large number of approximations of the optimal demodulator, see [4] and references therein. To improve the performance of MIMO-BICM with approximate demodulation, *scalar LLR correction* was used in [5–8], building on previous work in [9]; it was observed that scalar LLR correction can significantly improve the BER performance. This work provides an information-theoretic justification for such scalar LLR correction.

This work has been funded by the European Research Council under the European Community's Seventh Framework Programme, ERC Grant Agreement No. 228044, and by the FWF Grant S10606 "Information Networks."

The information-theoretic properties of BICM were first studied in [2] under the assumption of an ideal (infinitely long and completely random) interleaver. This assumption provided a rigorous background for the independent bit assumption and allowed the BICM system to be modeled as a set of independent parallel channels with binary inputs and corresponding LLR outputs. The BICM capacity was accordingly defined as the sum of mutual informations for the parallel channels [2]. An alternative to the parallel-channel model was provided in [10, 11] by viewing the BICM decoder, which neglects the conditional dependence among bits mapped to the same symbol, as a mismatched decoder. Under the assumption of optimal demodulation, the generalized mutual information (GMI) of this decoder was obtained in [10] and shown to equal the capacity obtained with the parallel-channel model. This result does not straightforwardly generalize to the case of approximate demodulation.

In this work, we extend the result in [10] by showing that the GMI of a BICM system with approximate demodulators using scalar LLR correction equals the capacity obtained for an equivalent parallel-channel model with binary inputs and continuous outputs provided by the approximate demodulator. We further show that augmenting the BICM demodulator with scalar LLR correction can never decrease the GMI, and that LLR correction as suggested in [9] is optimal in the sense that the GMI cannot be further increased within the BICM paradigm. In particular, the GMI without LLR correction is upper bounded by the GMI with LLR correction, and is in most cases strictly smaller. The results are valid for arbitrary approximate demodulators but depend crucially on the scalar LLR correction being matched to the demodulator at hand. These results are made precise in Theorem 1 in Section III-B. Together, the results provide additional justification for using the mutual information between input bits and approximate LLRs as a code-independent performance measure for approximate demodulators, as was done in [4].

After a rudimentary description of BICM systems and existing work in Section II, we provide precise statements of our new results in Section III. In Section IV, a numerical example illustrates the gain in GMI provided by scalar LLR correction and its effect on BER in an explicit BICM system. Conclusions are given in Section V.

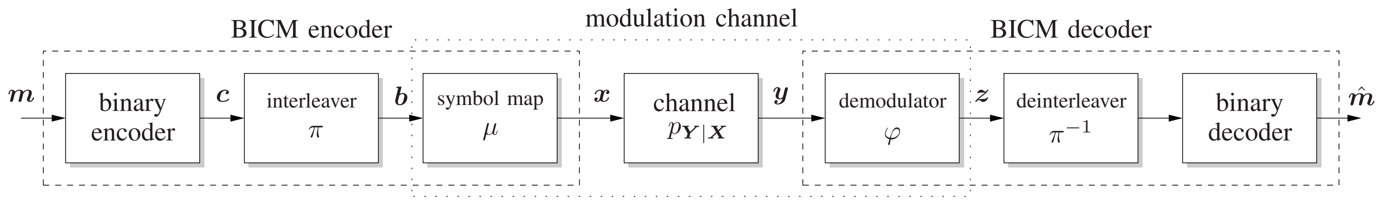


Fig. 1. The BICM system model consisting of BICM encoder, memoryless channel, and BICM decoder. The equivalent modulation channel seen by the binary encoder-decoder pair is indicated by the dotted rectangle.

II. Background

A. The BICM system model

We consider the BICM system model shown in Fig. 1, where a message $\boldsymbol{m}$ is mapped to a length-N codeword $\boldsymbol{x} = (x^1, \ldots, x^N) \in \mathcal{X}^N$ that induces a channel output $\boldsymbol{y} = (y^1, \ldots, y^N) \in \mathcal{Y}^N$ according to the memoryless transition probability

$$p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{n=1}^{N} p_{Y|X}(y^n|x^n)\,. \tag{1}$$

We assume that $\mathcal{X}$ is a finite set of cardinality $|\mathcal{X}| = 2^J$, i.e., $\mathcal{X}$ is capable of conveying up to J bits per channel use (bpcu). The output set $\mathcal{Y}$ can be either discrete or continuous, and $p_{Y|X}$ should accordingly be interpreted as a probability mass function (pmf) or probability density function (pdf) depending on the context. We let $\mathcal{M}$ be the set of possible messages and the rate of transmission is given by $R = N^{-1} \log |\mathcal{M}|$.

The message $\boldsymbol{m}$ is first encoded by a *binary encoder* into a sequence of JN bits which are interleaved and then mapped to the data symbols x^n in groups of J bits by a binary labeling function $\mu : \{0,1\}^J \to \mathcal{X}$. The J bits mapped to x^n are labeled $b_1^n, \ldots, b_J^n$, and we write $x^n(\boldsymbol{m})$ and $b_j^n(\boldsymbol{m})$ to make explicit the dependence on the transmitted message. Let $B \triangleq (B_1, \ldots, B_J)$ with B_j, $j = 1, \ldots, J$, i.i.d. random binary variables uniformly distributed over $\{0,1\}$ and let Y be the output of the memoryless channel $p_{Y|X}$ with random input $X \triangleq \mu(B)$. The joint distribution of (Y, B) is given by

$$p_{Y,B}(y, b) = p_{Y|X}(y|x = \mu(b)) p_B(b) \tag{2}$$

where $b \triangleq (b_1, \ldots, b_J) \in \{0,1\}^J$ and $p_B(b) = 2^{-J}$. We denote by $p_{B_j|Y}(b_j|y)$ the conditional distribution of B_j given Y and by $p_{Y|B_j}(y|b_j)$ the conditional distribution of Y given B_j. The optimal LLR computer or *soft demodulator* $\varphi_j : \mathcal{Y} \to \mathbb{R}$ for the jth bit position is defined by

$$\varphi_j(y) \triangleq \log \frac{p_{B_j|Y}(B_j = 1|Y = y)}{p_{B_j|Y}(B_j = 0|Y = y)}\,. \tag{3}$$

The BICM decoder decides in favor of a codeword $\hat{\boldsymbol{m}} \in \mathcal{M}$ according to the decision rule

$$\hat{\boldsymbol{m}} = \arg\max_{\boldsymbol{m} \in \mathcal{M}} \prod_{n=1}^{N} \prod_{j=1}^{J} p_{Y|B_j}(y^n|b_j^n(\boldsymbol{m})) \tag{4a}$$

$$= \arg\max_{\boldsymbol{m} \in \mathcal{M}} \sum_{n=1}^{N} \sum_{j=1}^{J} \sigma\big(b_j^n(\boldsymbol{m})\big) \varphi_j(y^n) \tag{4b}$$

where $\sigma \colon \{0,1\} \to \{\pm 1\}$ denotes the binary sign function for which $\sigma(0) = -1$ and $\sigma(1) = 1$. The formulation in (4b) is readily obtained from (4a) by using Bayes' rule, taking the logarithm, multiplying by 2 and subtracting $\log p_{B_j|Y}(B_j = 1|Y) + \log p_{B_j|Y}(B_j = 0|Y)$ (which is independent of $\boldsymbol{m}$) from each term in the sum. We will interchangeably use (4a) and (4b), depending on which is more convenient.

The decoder in (4b) separates the decoding process into the computation of $\varphi_j(y^n)$ for each n and j (demodulation) followed by binary decoding in which maximization over all possible codewords is performed. Note that the metric in (4a), being the product of marginal probabilities, effectively treats the bits as conditionally independent. This is suboptimal in general (the bits mapped to the same symbol *are* conditionally dependent), but tends to simplify the decoder which is one of the main motivations of BICM.

B. Parallel channels and generalized mutual information

The (memoryless) channel seen by the binary encoder-decoder pair is given by the conditional probability $p_{Z|B}(z|b)$, where $Z \triangleq (Z_1, \ldots, Z_J)$ and where $Z_j = \varphi_j(Y)$ is the LLR value corresponding to the jth bit position in X. However, as the BICM receiver treats the individual bits mapped to a symbol as independent, it makes sense to model the BICM system as a set of J parallel channels with transition probabilities $p_{Z_j|B_j}(z_j|b_j)$ for $j = 1, \ldots, J$. This approach was introduced and made rigorous in [2] by introducing the assumption of an *ideal interleaver* (i.e., infinite length and completely random). Accordingly, the BICM capacity was defined in [2] as the sum of the mutual informations over the parallel channels, i.e.,

$$C^{\text{bicm}} = \sum_{j=1}^{J} I(B_j; Z_j) \tag{5}$$

where ($\mathbb{E}$ denotes expectation)

$$I(B_j; Z_j) = \mathbb{E}\left[\log \frac{p_{B_j|Z_j}(B_j|Z_j)}{P_{B_j}(B_j)}\right]\,.$$

Note that Z_j is a sufficient statistic for B_j given Y [9] and hence $I(B_j; Z_j) = I(B_j; Y)$; the BICM capacity can thus equivalently be expressed in terms of $I(B_j; Y)$.

An alternative approach to studying the BICM system—not relying on the ideal interleaver assumption—was taken in [10] by viewing (4) as a mismatched decoder and obtaining the *generalized mutual information* [10–13] of this decoder. The GMI is here defined as the supremum of all rates for which, given a specific decoder metric, the random coding exponent [14] is strictly positive, see [10, 11] for details. It follows immediately by the standard random coding argument [14] that any rate below the GMI is achievable in the sense that there is a sequence of codes with vanishing probability of error for increasing block lengths.The GMI is however not necessarily the largest achievable rate [12]. Like [2, 10], we consider only binary random codebooks drawn from an i.i.d. equiprobable input distribution, i.e., we do not consider any optimization over the input distribution.

The GMI of an arbitrary decoder of the form

$$\hat{\boldsymbol{m}} = \arg\max_{\boldsymbol{m}\in\mathcal{M}} \prod_{n=1}^{N} \prod_{j=J}^{m} q_j(b_j^n(\boldsymbol{m}), y^n)\,, \tag{6}$$

is in this context shown to be given by [10]

$$I^{\mathrm{gmi}} = \sup_{s>0} I^{\mathrm{gmi}}(s) \tag{7}$$

where

$$I^{\mathrm{gmi}}(s) \triangleq \sum_{j=1}^{J} \mathbb{E}\left[\log \frac{q_j(B_j, Y)^s}{\frac{1}{2}\sum_{b_j=0}^{1} q_j(b_j, Y)^s}\right]. \tag{8}$$

A key result of [10] is that whenever $q_j(b_j^n, y^n) \propto p_{Y|B_j}(y^n|b_j^n)$ (cf. (4a)) it follows that $I^{\mathrm{gmi}} = C^{\mathrm{bicm}}$, i.e., the GMI equals the capacity of the parallel-channel model as defined in [2]. This implies that any rate below C^{bicm} is achievable with the decoder metric in (4), without the need for the ideal interleaver assumption. The GMI of (4) may alternatively be expressed in terms of the LLRs by inserting

$$q_j(b_j, y) = \exp\big(\tfrac{1}{2}\sigma(b_j)\varphi_j(y)\big) \tag{9}$$

into (8); it can be shown that this metric indeed satisfies $q_j(b_j, y) \propto p_{Y|B_j}(y|b_j)$.

III. Contributions

A. Approximate LLR computations

A common difficulty in the practical realization of BICM systems lies in the evaluation of $\varphi_j(y^n)$. Typically, the transition probability $p_{Y|X}(y|x)$ will have a simple closed form, whereas $p_{B_j|Y}(b_j|y)$ needs to be obtained according to

$$p_{B_j|Y}(b_j|y) \propto \sum_{x\in\mathcal{X}_j^{b_j}} p_{Y|X}(y|x)\,, \tag{10}$$

where $\mathcal{X}_j^{b_j}$ is the subset of all symbols in $\mathcal{X}$ whose jth bit equals b_j. As $|\mathcal{X}_j^{b_j}| = 2^{J-1}$ grows exponentially fast in J, the complexity of evaluating (10) quickly becomes practically unfeasible. This is especially the case when a large number of bits is transmitted per channel use such as in MIMO systems [3]. For this reason, approximations of $\varphi_j(y^n)$ are typically used. Particulary in the MIMO-BICM scenario, many different approximate soft demodulators (LLR computers) have been proposed, see [4] and references therein. A natural decoder metric, paralleling the optimal BICM decoder, is in this case given by (cf. (4b))

$$\hat{\boldsymbol{m}} = \arg\max_{\boldsymbol{m}\in\mathcal{M}} \sum_{n=1}^{N} \sum_{j=1}^{m} \sigma\big(b_j^n(\boldsymbol{m})\big)\hat{\varphi}_j(y^n)\,. \tag{11}$$

where $\hat{\varphi}_j$ denotes the approximate LLR computer. The GMI corresponding to (11), denoted $I_{\mathrm{A}}^{\mathrm{gmi}}$ in what follows, may be obtained according to (7) and (8) with (cf. (9))

$$q_j(b_j, y) = \exp\big(\tfrac{1}{2}\sigma(b_j)\hat{\varphi}_j(y)\big)\,. \tag{12}$$

Similar to Section II-B, the channel seen by the encoder-decoder pair in the case of approximate demodulation can be described by the transition probability $p_{\hat{Z}|B}(\hat{z}|b)$, where $\hat{Z} \triangleq (\hat{Z}_1, \ldots, \hat{Z}_J)$, $\hat{Z}_j = \hat{\varphi}_j(Y)$ for $j = 1, \ldots, J$, denotes the set of approximate LLR outputs. We define the *BICM capacity with approximate demodulation* as (cf. (5))

$$C_{\mathrm{A}}^{\mathrm{bicm}} \triangleq \sum_{j=1}^{J} I(B_j; \hat{Z}_j) \le C^{\mathrm{bicm}}\,, \tag{13}$$

where

$$I(B_j; \hat{Z}_j) = \mathbb{E}\left[\log \frac{p_{B_j|\hat{Z}_j}(B_j|\hat{Z}_j)}{P_{B_j}(B_j)}\right]. \tag{14}$$

The bound in (13) follows from the data processing inequality [15]. Unlike with exact LLRs (optimal demodulation), the equality $C_{\mathrm{A}}^{\mathrm{bicm}} = I_{\mathrm{A}}^{\mathrm{gmi}}$ *does not* hold in general.

B. Scalar LLR correction

It may be argued that the metric in (11) is twofold mismatched: on the one hand the bits are implicitly assumed to be independent given y^n, when in fact they are not; one the other hand the approximate LLRs are treated as though they were exact LLRs (cf. (4b) and (11)). However, the second source of mismatch can be eliminated [9].

Scalar LLR correction has been used previously as a means to correct approximate LLRs prior to decoding, thereby providing the binary decoder with accurate bit reliability information [5–9]. To this end, let $\phi_j : \mathbb{R} \to \mathbb{R}$, $j = 1, \ldots, J$, be a set of scalar maps used to process the LLRs prior to decoding. Specifically, as proposed in [9], let

$$\phi_j(\hat{z}_j) \triangleq \log \frac{p_{B_j|\hat{Z}_j}(B_j = 1|\hat{Z}_j = \hat{z}_j)}{p_{B_j|\hat{Z}_j}(B_j = 0|\hat{Z}_j = \hat{z}_j)}\,, \tag{15}$$

i.e., $\phi_j(\hat{z}_j)$ computes the LLR of an input bit, given the output $\hat{z}_j$, which is itself an approximate LLR. In practice, ϕ_j can be implemented via a one-dimensional lookup table that has previously been constructed offline [8, 9]. Observe that the deterministic LLR correction does not change the capacity in the parallel channel model, i.e., replacing $\hat{Z}_j$ in (13) with $\phi_j(\hat{Z}_j)$ again results in the approximate BICM capacity $C_{\mathrm{A}}^{\mathrm{bicm}}$. We note that our model and results extend naturally to the case (not discussed further for reasons of space) where $\hat{z}_j$ consist of an approximate LLR plus some additional side information provided by the demodulator [5, 8].

The BICM decoder, applied to the corrected LLRs, reads

$$\hat{\boldsymbol{m}} = \arg\max_{\boldsymbol{m}\in\mathcal{M}} \sum_{n=1}^{N}\sum_{j=1}^{J} \sigma\big(b_j^n(\boldsymbol{m})\big)\phi_j(\hat{\varphi}_j(y^n))\,. \tag{16}$$

The corresponding GMI, denoted $I_{\mathrm{C}}^{\mathrm{gmi}}$ in what follows, is given by (7) and (8) with (cf. (9))

$$q_j(b_j, y) = \exp\big(\tfrac{1}{2}\sigma(b_j)\phi_j(\hat{\varphi}_j(y))\big)\,. \tag{17}$$

C. Main result and discussion

The main contribution of this paper is the following theorem, whose proof is provided in the appendix.

THEOREM 1. *The GMI of the LLR-corrected decoder* (16) *equals the approximate BICM capacity obtained under the parallel channel model, i.e.,*

$$I_{\mathrm{C}}^{\mathrm{gmi}} = C_{\mathrm{A}}^{\mathrm{bicm}}.$$

Furthermore, scalar LLR correction according to (15) *is optimal in the sense that replacing* ϕ_j *in* (16) *by some other scalar map* $\hat{z}_j \mapsto f_j(\hat{z}_j)$ *can not further increase the GMI. For the particular case* $f_j(\hat{z}_j) = \hat{z}_j$ *(no LLR correction), it thus holds that*

$$I_{\mathrm{A}}^{\mathrm{gmi}} \leq I_{\mathrm{C}}^{\mathrm{gmi}}.$$

We note that the GMI could potentially be increased by more general (non-scalar) mappings. Since such mappings would take into account more information than contained in $\hat{Z}_j$ in the decoder, this would leave the realm of the BICM paradigm and of *scalar* LLR correction.

Given that any rate below the GMI is achievable, Theorem 1 shows that any rate below $C_{\mathrm{A}}^{\mathrm{bicm}}$ is achievable as well with the decoder metric in (16) and appropriate binary codes. Apart from its theoretical significance, this result is also practically useful since $C_{\mathrm{A}}^{\mathrm{bicm}}$ is easier to evaluate numerically than $I_{\mathrm{C}}^{\mathrm{gmi}}$, specifically since obtaining $C_{\mathrm{A}}^{\mathrm{bicm}}$ does not require the LLR correction stage while obtaining $I_{\mathrm{C}}^{\mathrm{gmi}}$ requires computation of $I_{\mathrm{C}}^{\mathrm{gmi}}(s)$ with LLR correction for any $s > 0$.

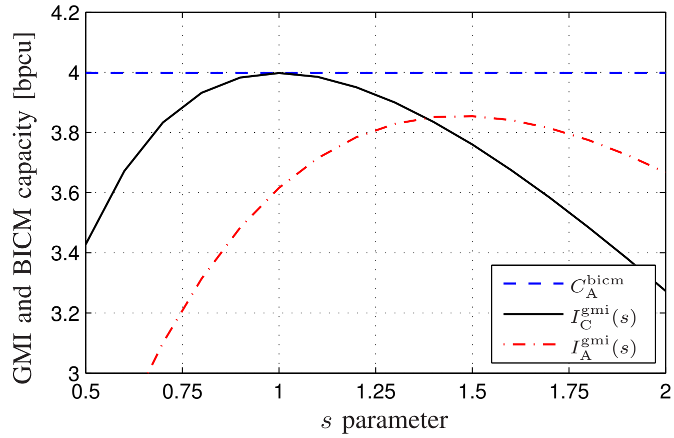
IV. NUMERICAL EXAMPLE

In order to illustrate Theorem 1, we provide a numerical example inspired by [6]. The example illustrates the performance of a 2×2 16-QAM Rayleigh fading MIMO-BICM system with Gray labeling, a max-log demodulator, i.e.,

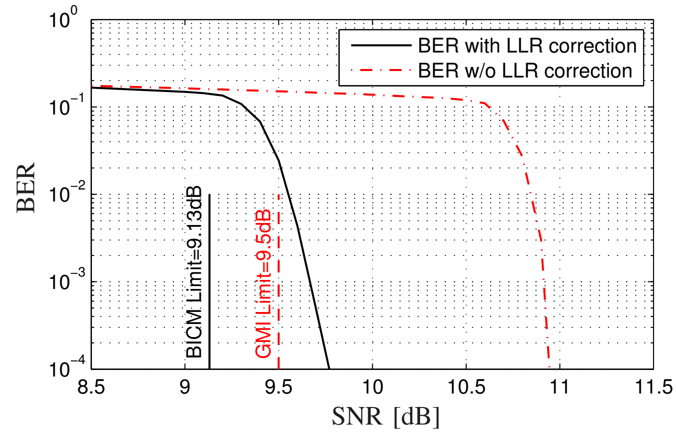
$$\hat{\varphi}_j(y) = \max_{x\in\mathcal{X}_j^1} \log p_{Y|X}(y|x) - \max_{x\in\mathcal{X}_j^0} \log p_{Y|X}(y|x)\,,$$

and output LLRs clipped (truncated) to a range of $[-2, 2]$. We note that by limiting the range of the output LLRs, the computational complexity of the max-log demodulator may be significantly reduced [8]. In terms of BER, it was previously observed in [6] that LLR correction can provide significant performance improvements for small clipping thresholds.

Fig. 2(a) shows the approximate BICM capacity $C_{\mathrm{A}}^{\mathrm{bicm}}$, the GMI of the clipped max-log demodulator $I_{\mathrm{A}}^{\mathrm{gmi}}(s)$ (cf. (8) and (12)), and the GMI after scalar LLR correction $I_{\mathrm{C}}^{\mathrm{gmi}}(s)$ (cf. (8) and (17)) versus the parameter s at a signal-to-noise



(a) GMI vs. parameter s at an SNR of 9.13 dB.



(b) LDPC code BER vs. SNR with and without scalar LLR correction.

Fig. 2. Performance of a 2×2 MIMO system with 16-QAM, Gray labeling, and max-log demodulation with LLR clipping.

ratio (SNR) of 9.13 dB. In this example, the LLR correction was implemented using an 8 bit (256 bin) lookup table with nonuniform quantization. It is seen that $I_{\mathrm{A}}^{\mathrm{gmi}} \leq I_{\mathrm{C}}^{\mathrm{gmi}} = \sup_{s>0} I_{\mathrm{C}}^{\mathrm{gmi}}(s) = C_{\mathrm{A}}^{\mathrm{bicm}}$ as stated by Theorem 1. For the example at hand, scalar LLR correction increases the GMI by about 0.15 bpcu. Fig. 2(b) illustrates how the increase in GMI translates into an improved BER for an irregular rate-1/2 LDPC code[1] of block length 64000. The binary code rate of 1/2 corresponds to an overall rate of 4 bpcu. The minimum SNR required to achieve an approximate BICM capacity (with LLR correction) and a GMI (without LLR correction) of 4 bpcu is indicated by vertical lines. The performance of the LDPC code with scalar LLR correction is within 0.6 dB of approximate BICM capacity, outperforming the case without LLR correction by roughly 1 dB.

V. CONCLUSION

In this work, the GMI of a BICM system with approximate demodulation and scalar LLR correction was considered. It was shown that the GMI of a decoder applied to the corrected

[1]The code design was performed using the EPFL web-application at `http://lthcwww.epfl.ch/research/ldpcopt`.

LLR values equals the natural definition for the BICM capacity of a parallel-channel model with approximate demodulation similar to [2], which is easier to obtain numerically. This provides further justification for using the BICM capacity with approximate demodulation as a performance measure for demodulator comparisons, as was done in [4]. This work also provides an interesting interpretation of the mismatch in (11). For BICM systems with approximate demodulation, the decoder is twofold mismatched: it treats bits as if they were independent and it treats approximate LLRs as if they were exact. While the first mismatch is inherent to the BICM paradigm, our results show that the second mismatch is effectively eliminated by scalar LLR correction.

Appendix: Proof of Theorem 1

We begin by establishing that $I_{\mathrm{C}}^{\mathrm{gmi}}(s=1) = C_{\mathrm{A}}^{\mathrm{bicm}}$ where $I_{\mathrm{C}}^{\mathrm{gmi}}(s)$ is given by (8) with $q_j(b_j, y)$ chosen according to (17). Inserting (15) into (17) shows that

$$q_j(b_j, y) = \left[\frac{p_{B_j|\hat{Z}_j}(1|\hat{z}_j)}{p_{B_j|\hat{Z}_j}(0|\hat{z}_j)}\right]^{\frac{1}{2}\sigma(b_j)}, \tag{18}$$

where $\hat{z}_j = \hat{\varphi}_j(y)$, $\sigma(0) = -1$, and $\sigma(1) = 1$. With (18) and $s = 1$, the argument of the expectation in (8) becomes

$$S_j(B_j, Y) \triangleq \frac{\left[\frac{p_{B_j|\hat{Z}_j}(1|\hat{Z}_j)}{p_{B_j|\hat{Z}_j}(0|\hat{Z}_j)}\right]^{\frac{1}{2}\sigma(B_j)}}{\frac{1}{2}\left[\frac{p_{B_j|\hat{Z}_j}(0|\hat{Z}_j)}{p_{B_j|\hat{Z}_j}(1|\hat{Z}_j)}\right]^{\frac{1}{2}} + \frac{1}{2}\left[\frac{p_{B_j|\hat{Z}_j}(1|\hat{Z}_j)}{p_{B_j|\hat{Z}_j}(0|\hat{Z}_j)}\right]^{\frac{1}{2}}}, \tag{19}$$

where $\hat{Z}_j = \hat{\varphi}_j(Y)$. Multiplying the numerator and denominator in (19) by $(p_{B_j|\hat{Z}_j}(1|\hat{Z}_j)p_{B_j|\hat{Z}_j}(0|\hat{Z}_j))^{\frac{1}{2}}$ yields

$$S_j(B_j, Y) = \frac{p_{B_j|\hat{Z}_j}(B_j|\hat{Z}_j)}{\frac{1}{2}p_{B_j|\hat{Z}_j}(0|\hat{Z}_j) + \frac{1}{2}p_{B_j|\hat{Z}_j}(1|\hat{Z}_j)},$$

which is easily verified by inserting $B_j = 1$ and $B_j = 0$ into (19). Because $p_{B_j|\hat{Z}_j}(0|\hat{Z}_j) + p_{B_j|\hat{Z}_j}(1|\hat{Z}_j) = 1$ and $p_{B_j}(B_j) = \frac{1}{2}$ it follows that

$$S_j(B_j, Y) = \frac{p_{B_j|\hat{Z}_j}(B_j|\hat{Z}_j)}{p_{B_j}(B_j)}$$

and hence that $I_{\mathrm{C}}^{\mathrm{gmi}}(1) = C_{\mathrm{A}}^{\mathrm{bicm}}$ (cf. (14)).

We proceed with proving that any other $s > 0$ or any other choice of BICM decoding metric q_j, which depends on Y only through $\hat{Z}_j$, i.e., $q_j(B_j, Y) = r_j(B_j, \hat{Z}_j)$ for $j = 1, \ldots, J$, leads to a GMI no larger than $I_{\mathrm{C}}^{\mathrm{gmi}}(1)$. The expression in (8) can be written as

$$I^{\mathrm{gmi}}(s) \triangleq \sum_{j=1}^{J} \mathbb{E}\left[\mathbb{E}\left[\log \frac{r_j(B_j, \hat{Z}_j)^s}{\frac{1}{2}\sum_{b_j=0}^{1} r_j(b_j, \hat{Z}_j)^s}\Big|\hat{Z}_j\right]\right]. \tag{20}$$

Because scaling of r_j does not affect (20), we assume without loss of generality that $\sum_{b_j=0}^{1} r_j(b_j, \hat{Z}_j)^s = 1$. Since

$$\mathbb{E}\left[\log p_{B_j|\hat{Z}_j}(B_j|\hat{Z}_j)\big|\hat{Z}_j\right] - \mathbb{E}\left[\log r_j(B_j, \hat{Z}_j)^s\big|\hat{Z}_j\right]$$

$$= \sum_{b_j=0}^{1} p_{B_j|\hat{Z}_j}(b_j|\hat{Z}_j) \log \frac{p_{B_j|\hat{Z}_j}(b_j|\hat{Z}_j)}{r_j(b_j, \hat{Z}_j)^s} \geq 0$$

by the nonnegativity of relative entropy [15], it follows that

$$\mathbb{E}\left[\log \frac{p_{B_j|\hat{Z}_j}(B_j|\hat{Z}_j)}{p_{B_j}(B_j)}\Big|\hat{Z}_j\right] \geq \mathbb{E}\left[\log \frac{r_j(B_j, \hat{Z}_j)^s}{\frac{1}{2}\sum_{b_j=0}^{1} r_j(b_j, \hat{Z}_j)^s}\Big|\hat{Z}_j\right] \tag{21}$$

for any $r_j(b_j, \hat{Z}_j)$ and $s > 0$.

For the particular choice of $r_j(B_j, \hat{Z}_j) = q_j(B_j, Y)$ where q_j is given by (17), it follows by (20) and (21) that $\sup_{s>0} I_{\mathrm{C}}^{\mathrm{gmi}}(s) = I_{\mathrm{C}}^{\mathrm{gmi}}(1) = I_{\mathrm{C}}^{\mathrm{gmi}} = C_{\mathrm{A}}^{\mathrm{bicm}}$, i.e., the supremum is attained for $s = 1$. Further, for arbitrary $q_j(B_j, Y) = r_j(b_j, \hat{Z}_j)$, it holds that $I^{\mathrm{gmi}} = \sup_{s>0} I^{\mathrm{gmi}}(s) \leq I_{\mathrm{C}}^{\mathrm{gmi}}$. The inequality $I_{\mathrm{A}}^{\mathrm{gmi}} \leq I_{\mathrm{C}}^{\mathrm{gmi}}$ is obtained as special case with $q_j(B_j, Y)$ chosen according to (12).

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. Information Theory*, vol. 44, no. 5, pp. 927–945, May 1998.

[3] A. Stefanov and T. M. Duman, “Turbo-coded modulation for systems with transmit and receive antenna diversity over block fading channels: System model, decoding approaches, and practical considerations,” *IEEE J. Selected Areas in Communications*, vol. 19, no. 5, pp. 958–968, May 2001.

[4] P. Fertl, J. Jaldén, and G. Matz, “Capacity-based performance comparison of MIMO-BICM demodulators,” in *Proc. IEEE SPAWC*, Recife, Brazil, July 2008, pp. 166–170.

[5] A. Burg, M. Wenk, and W. Fichtner, “VLSI implementation of pipelined sphere decoding with early termination,” in *Proc. IEEE EUSIPCO*, Florence, Italy, Sept. 2006.

[6] S. Schwandter, P. Fertl, C. Novak, and G. Matz, “Log-likelihood ratio clipping in MIMO-BICM systems: Information geometric analysis and impact on system capacity,” in *Proc. IEEE ICASSP*, Taipei, Taiwan, 2009, pp. 2433–2436.

[7] C. Novak, P. Fertl, and G. Matz, “Quantization for soft-output demodulators in bit-interleaved coded modulation systems,” in *Proc. IEEE ISIT*, Seoul, Korea, June 2009, pp. 1070–1074.

[8] C. Studer and H. Bölcskei, “Soft-input soft-output single tree-search sphere decoding,” 2009, submitted to *IEEE Trans. Inform. Theory*, available at arXiv:0906.0840v1 [cs.IT].

[9] M. van Dijk, A. J. E. M. Janssen, and A. G. C. Koppelaar, “Correcting systematic mismatches in computed log-likelihood ratios,” *Europ. Trans. Telecomm.*, vol. 14, no. 3, pp. 227–244, July 2003.

[10] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. Willems, “Bit-interleaved coded modulation revisited: A mismatched decoding perspective,” in *Proc. IEEE ISIT*, Toronta, Canada, July 2008.

[11] 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.

[12] A. Ganti, A. Lapidoth, and I. E. Telatar, “Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit,” *IEEE Trans. Information Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[13] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, “On information rates for mismatched decoders,” *IEEE Trans. Information Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[14] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.