Parallel Bit Interleaved Coded Modulation: BICM without Asymptotic Assumptions

Amir Ingber, *Member, IEEE,* and Meir Feder, *Fellow, IEEE*

***Abstract*—A new variant of bit interleaved coded modulation (BICM) is proposed. In the new scheme, called *parallel BICM*, L identical binary codes are used in parallel using a mapper, a newly proposed finite-length interleaver and a binary dither signal. As opposed to previous approaches, the scheme does not rely on any assumptions of an ideal, infinite-length interleaver.**

Over a memoryless channel, the new scheme is proven to be equivalent to a *binary* memoryless channel, for any blocklegth. Therefore the scheme enables one to easily design coded modulation schemes using a simple binary code that was designed for that binary channel. The overall performance of the coded modulation scheme is analytically evaluated based on the performance of the binary code over the binary channel. The new scheme is then analyzed from an information theoretic viewpoint, where the capacity, error exponent and channel dispersion are considered. The capacity of the scheme is identical to the BICM capacity. The error exponent of the scheme is numerically compared to a recently proposed mismatched-decoding exponent analysis of BICM.

I. Introduction

Bit interleaved coded modulation (BICM) is a pragmatic approach for coded modulation [1]. It enables the construction of nonbinary communication schemes from binary codes by using a long bit interleaver that separates the coding and the modulation. BICM has drawn much attention in recent years, because of its efficiency for wireless and fading channels.

The information-theoretic properties of BICM were first studied by Caire et. al. in [2]. BICM was modeled as a binary channel with a random state that is known at the receiver. The state determines how the input bit is mapped to the channel, along with the other bits that are assumed to be random. Under the assumption of an *infinite-length, ideal interleaver*, the BICM scheme is modeled by parallel uses of independent instances of this binary channel. This model is referred to as the *independent parallel channel model*. Using this model, the capacity of the BICM scheme could be calculated. It was further shown that BICM suffers from a gap from the full channel capacity, and that when Gray mapping is used this gap is generally small. In [2], methods for evaluating the error probability of BICM were proposed, which rely on the properties of the specific binary codes that were used (e.g. Hamming weight of error events).

A basic information-theoretic quantity other than the channel capacity is the error exponent [3], which quantifies the speed at which the error probability decreases to zero with the block length n. Another tool for evaluating the performance at finite block length is the channel dispersion, which was presented in 1962 [4] and was given more attention only in recent years [5]. It would therefore be interesting to analyze BICM at finite block length from the information-theoretic viewpoint.

Several attempts have been made to provide error exponent results for BICM. In their work on multilevel codes, Wachsmann et. al. [6] have considered the random coding error exponent of BICM, by relying on the independent parallel channels model. However, there were several flaws in the derivation: First, the independent parallel channels model is justified by an infinite-length interleaver so therefore it might be problematic to use its properties for evaluating the *finite length* performance of BICM. In the current paper we address this point and propose a scheme with a finite-length interleaver for that purpose. Second, there was a technical flaw in the derivation, which resulted in an inaccurate expression for the random coding error exponent. We discuss this point in detail in Theorem 4. Third, as was noticed in [7], the error exponent result obtained in [6] sometimes may even exceed that of unconstrained coding over the channel (called in [7] the "coded modulation exponent"). We therefore agree with [7] in the claim that "the independent parallel channel model fails to capture the statistics of the channel". However, by properly designing the communication scheme the model can become valid in a rigorous way, as we show in Theorem 1.

In [7] (see also [8]), Martinez et al. have considered the BICM decoder as a mismatched decoder, which has access only to the log-likelihood values (LLR) of each bit, where the LLR calculation assumes that the other bits mapped to the same symbol are random, independent and equiprobable (as in the classical BICM scheme [2]). By cleverly harnessing the mismatched decoding framework, the authors in [7] presented the generalized error exponent and the generalized mutual information, and pinpointed the loss of BICM that incurs from using the mismatched LLRs. While this result is valid for any block size and any interleaver length, achieving this error exponent in practice requires complex code design. For example, one cannot design a good binary code for a binary memoryless channel and have any guarantee that the BICM scheme will perform well with that code. In fact, the code design for this scheme requires taking into account the statistical dependencies between the levels, or equivalently, the usage of essentially nonbinary codes, which is what we wish to avoid when choosing BICM.

On the theoretical side, another drawback of existing approaches is the lack of converse results (for either capacity or

The material in this paper was presented in part at the 48^{th} Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, 2010.

error exponent). The initial discussion of BICM information theory in [2] assumes the model of independent channels which is justified by an infinite-length interleaver, so any converse result based on this model must assume that an infinite interleaver as well. Therefore the converse results (upper bound on the achievable rate with BICM) do not hold for finite-length interleavers. The authors in [7] provide no converse results for their model.

In this paper we propose the *parallel BICM* (PBICM) scheme, which has the following properties. First, the scheme includes an explicit, *finite length* interleaver. Second, in order to attain good performance on any memoryless channel, PBICM allows one to design a binary code for a binary memoryless channel, and guarantees good performance on the nonbinary channel. Third, because the scheme does not rely on the use of an infinite-length interleaver, any finite blocklength bounds on the performance of channel codes (e.g. [5]), can be used in order to evaluate the PBICM performance. We shall give specific examples, and calculate the error exponent and the dispersion of the scheme (both achievability and converse results)

The comparison between PBICM and the mismatched decoding approach [7] should be done with care. With PBICM, when the binary codeword length is n the scheme requires n channel uses. In [7], a binary code of length n requires n/L channel uses, where L is the number of bits per symbol. Therefore when the latency is kept equal for both schemes, PBICM uses a codeword length that is L times shorter than the codeword used in the mismatched decoder. A fair comparison must take this into account. For example, fixing the binary codeword length n for both schemes would result in different latency, but equal decoder complexity. If we were to keep the latency of both schemes fixed, we would get different decoder complexities. When we numerically compare the error exponent of PBICM to the mismatched-decoding exponent we address this issue. In our comparison the additive white Gaussian noise (AWGN) channel and the Rayleigh fading channel are considered. When the latency of both schemes is equal, the mismatched-decoding is generally better. However, when the complexity is equal (or where the codeword length of the underlying binary code is equal), the PBICM exponent is better in many cases.

The paper is organized as follows. In Section II we briefly review the classical BICM model and its properties, under the assumption of an infinite-length, ideal interleaver. In Section III the parallel BICM scheme is presented, and the equivalence to a memoryless binary channel is established. In Section IV parallel BICM is studied from an information-theoretical viewpoint. Numerical examples and summary follow in Sections V and VI respectively.

II. The BICM Communication Model

A. Notation and Channel Model

letters in bold ($\mathbf{x}, \mathbf{y}$...) denote row vectors, capital letters (X, Y...) denote random variables, and tilde denotes interleaved signals ($\tilde{\mathbf{b}}, \tilde{\mathbf{z}}$). $P_X(x)$ denotes the probability that the random variable (RV) X will get the value x, and similarly $P_{Y|X}(y|x)$ denotes the probability Y will get the value y given that the RV X is equal to x. $\mathbb{E}[\cdot]$ denotes statistical expectation. $\log$ means $\log_2$ and rates are given in bits. $f(n) = O(\varepsilon_n)$ shall mean that $|f(n)| \leq c\varepsilon_n$ for some $c > 0$ and sufficiently large n, and $f(n) \leq O(\varepsilon_n)$ shall mean that $f(n) \leq c\varepsilon_n$. $f_n = g_n + O(\varepsilon_n)$ shall mean that $f_n - g_n = O(\varepsilon_n)$, and $f(n) \geq O(\varepsilon_n)$ shall mean that $-f(n) \leq O(\varepsilon_n)$.

Let W denote a memoryless channel with input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ respectively. The transition probabilities are defined by $W(y|x)$ for $y \in \mathcal{Y}$ and $x \in \mathcal{X}$. We assume that $|\mathcal{X}| = 2^L$, and consider equiprobable signaling only over the channel W. An (n, R) code $\mathcal{C} \subseteq \mathcal{X}^n$ is a set of $M = 2^{nR}$ codewords $\mathbf{c} \in \mathcal{X}^n$. The encoder wishes to convey one of M equiprobable messages.

B. Classical BICM Encoding and Decoding

In BICM, a binary code is used to encode information messages $[m_1, m_2, ...]$ into binary codewords $[\mathbf{b}_1, \mathbf{b}_2, ...]$. The binary codewords are then interleaved using a long interleaver $\pi(\cdot)$, which applies a permutation on the coded bits. The interleaved bit stream $\tilde{\mathbf{b}}$ is partitioned into groups of L consecutive bits and inserted into a mapper $\mu : \{0,1\}^L \to \mathcal{X}$. The mapper output, denoted $\mathbf{x}$, is fed into the channel. The decoding process of BICM proceeds as follows. The channel output $\mathbf{y}$ is fed into a bit metric calculator, which calculates the log-likelihood ratio (LLR) of each input bit b given the corresponding output sample y (L LLR values for each output sample). These LLR values (or bit metrics) denoted $\tilde{\mathbf{z}}$ are de-interleaved and partitioned into bit metrics $[\mathbf{z}_1, \mathbf{z}_2, ...]$ that correspond to the binary input codewords. Finally, the binary decoder decodes the messages $[\hat{m}_1, \hat{m}_2, ...]$ from $[\mathbf{z}_1, \mathbf{z}_2, ...]$. The BICM encoding-decoding process is shown in Figure 1.

The LLR of the j^{th} bit in a symbol given the output value y is calculated as follows:

$$LLR_j(y) \triangleq \log \frac{P_{Y|B_j}(y|0)}{P_{Y|B_j}(y|1)}, \tag{1}$$

where $P_{Y|B_j}(y|b)$ is the conditional probability of the channel output getting the value y given that the j^{th} bit at the mapper input was b, and the other $(L-1)$ bits are equiprobable and independent binary random variables (RVs).

C. Classical BICM Analysis: Ideal Interleaving

In classical BICM (e.g. [2]) the LLR calculation is motivated by the assumption of a very long (*ideal*) interleaver π, so the coded bits go through essentially *independent* channels. These binary channels are defined as follows:

Definition 1: Let W_i be a binary channel with transition probability

$$\begin{aligned} W_i(y|b) &\triangleq \mathbb{E}\left[W(y|X = \mu(B_1, ..., B_L))|B_i = b\right] \\ &= \frac{1}{2^{L-1}} \sum_{\substack{b_j;\ i\neq j \\ b_i = b}} W(y|\mu(b_1, ..., b_L)). \end{aligned} \tag{2}$$

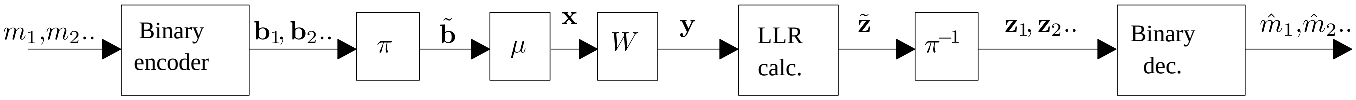


Fig. 1. BICM encoding and decoding for the channel W. π and μ are the interleaver and mapper, respectively.

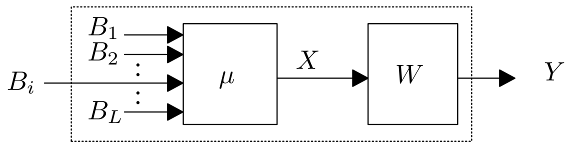


Fig. 2. The binary channel W_i with input B_i and output Y. The bits $\{B_j\}_{j\neq i}$ are equiprobable and independent RVs.

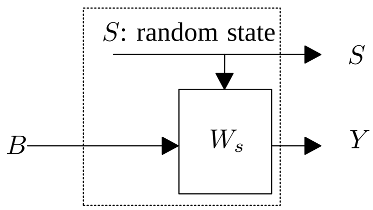


Fig. 3. The binary channel $\widetilde{W}$ with input B and output (S, Y).

The channel $W_i(y|b_i)$ can be thought of as the original channel W where the input is $x = \mu(b_1...b_L)$, where the bits $\{b_j\}_{j\neq i}$ are equiprobable and independent RVs (see Fig. 2).

In [2], Caire et al. have proposed the following channel model for BICM called the *independent parallel channel model*. In this model the channel has a binary input b. A channel state s is selected at random from $\mathcal{S} \triangleq \{1, ..., L\}$ with equal probability (and independently of b). Given a state s, the input bit b is fed into the channel W_s. The channel outputs are the state s and the output y of the channel W_s. The channel, denoted by $\widetilde{W}$, is depicted in Figure 3.

The transition probability function of $\widetilde{W}$ is given by

$$\widetilde{W}(y, s|b) = P_{Y,S|B}(y, s|b) = \frac{1}{L} W_s(y, b). \tag{3}$$

Note that both outputs can be combined into a single output, the LLR, which is a sufficient statistic for optimal decoding over any binary-input channel. The LLR calculation for the channel $\widetilde{W}$ is given by $LLR_{\widetilde{W}}(y, s) \triangleq LLR_s(y)$, where $LLR_s(\cdot)$ is given in (1). Therefore the independent parallel channel model transforms the original nonbinary channel W into a set of L independent parallel copies of the channel $\widetilde{W}$. Using a binary code that was designed for the simple binary channel $\widetilde{W}$, reliable communication for the original channel W can be attained.

Let $\mathbf{C}(\cdot)$ denote the Shannon capacity of a channel (with equiprobable input). According to [2], the BICM capacity (assuming an infinite-length ideal interleaver) is given by the capacity of L independent copies of the channel $\widetilde{W}$, which equals $L \cdot \mathbf{C}(\widetilde{W})$. Direct calculation gives

$$\begin{aligned} \mathbf{C}^{\text{BICM}}(W) &= L\mathbf{C}\left(\widetilde{W}\right) = L \cdot I(B; Y, S) = L \cdot I(B; Y|S) \\ &= L \cdot \mathbb{E}_S I(B; Y|S = s) = L \cdot \mathbb{E}_S \mathbf{C}(W_s) = \sum_{s=1}^{L} \mathbf{C}(W_s). \end{aligned} \tag{4}$$

It is known that $\mathbf{C}^{\text{BICM}}(W)$ is generally smaller than the full channel capacity $\mathbf{C}(W)$, as opposed to other schemes, most notably multilevel coding and multistage decoding (MLC-MSD) [6], in which $\mathbf{C}(W)$ can be achieved. However, for many channels, with Gray mapping the gap is small and can sometimes be tolerated. For example, for 8-PSK signaling over the AWGN channel with SNR = 5dB, $\mathbf{C}(W) = 1.86$bit where $\mathbf{C}^{\text{BICM}}(W) = 1.84$bit. For an elaborate discussion on the mapping, as well as examples for channels where Gray mapping is suboptimal, see [9] and references within.

III. The Parallel BICM Scheme

In this section we propose an explicit BICM-type communication scheme which we call *parallel BICM* (PBICM), which allows the usage of binary codes on nonbinary channels at finite blocklength. The main features of the scheme include the following: (1) Binary codewords are used *in parallel* to construct a codeword that enters the channel, (2) A new finite-length interleaver, (3) A random binary signal (binary dither) that is added to the binary codewords. With the proposed scheme, we rigorously show how the original channel W relates to the channel $\widetilde{W}$, thus allowing exact analysis and design of codes at finite block lengths.

A. Interleaver Design

We wish to design a finite length interleaver that will be as short as possible, that will be as simple as possible, and that will cause the binary codewords to go through a binary memoryless channel. In order to achieve the last goal, each binary codeword must be spread over n channel uses of W, so the interleaver output length cannot be less than n channel uses. The newly proposed interleaver has of output length of exactly n, which satisfies the above requirements.

Let ENC and DEC be an encoder-decoder pair for a binary code. Let $\mathbf{b}_1, ..., \mathbf{b}_L$ be L consecutive codewords from the output of ENC, bunched together to form a matrix $\mathbf{B}$:

$$\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_L \end{pmatrix} = \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & b_{lk} & \vdots \\ b_{L1} & \dots & b_{Ln} \end{pmatrix}. \tag{5}$$

Let $\mathbf{s}$ be a vector of i.i.d. random states drawn from $\mathcal{S}^n = \{1..L\}^n$. $\mathbf{s}$ shall be the interleaving signal. Each column in $\mathbf{B}$

shall be shifted cyclically by the corresponding element s_k, so the interleaved signal $\tilde{\mathbf{B}}$ is defined as

$$\tilde{\mathbf{B}} = \begin{pmatrix} b_{(1+s_1)_L 1} & \cdots & b_{(1+s_n)_L n} \\ \vdots & b_{(l+s_k)_L k} & \vdots \\ b_{(L+s_1)_L 1} & \cdots & b_{(1+s_n)_L n} \end{pmatrix}, \quad (6)$$

where $(\xi)_L \triangleq (\xi \text{ modulo } L)+1$. Each column vector of interleaved signal $\tilde{\mathbf{B}}$ is mapped to a single channel symbol $x_k = \mu(b_{(1+s_k)_L k}, \ldots, b_{(L+s_k)_L k})$, and we call $\mathbf{x} = [x_1, ..., x_n]$ the *channel codeword*.

At the decoder an LLR value is calculated for every bit b in $\tilde{\mathbf{B}}$ from $\mathbf{y}$. The LLR values are denoted by $\tilde{\mathbf{Z}}$. We assume that $\mathbf{s}$ is known at the decoder (utilizing common randomness), therefore the de-interleaving operation is simply sorting back the columns of $\tilde{\mathbf{Z}}$ according to $\mathbf{s}$ by reversing the modulo operation. The de-interleaver output is a vector of LLR values $\mathbf{z}$ for each transmitted codeword $\mathbf{b}$, according to (1). Each codeword is decoded independently by DEC.

B. Binary Dither

Since the decoder decodes each binary codeword independently, the communication scheme employing the above interleaver can be viewed as as set of parallel encoder-decoder pairs, which we denote by $ENC_1, ..., ENC_L$ and $DEC_1, ..., DEC_L$ (see Figures 4 and 5). Note that we cannot assume any independence between the effective channels between each encoder-decoder pair.

Consider the first encoder-decoder pair, ENC_1 and DEC_1. Since the input of DEC_1 depends on the codewords transmitted by ENC_2,...,ENC_L, the channel between ENC_1 and DEC_1 is not strictly memoryless. If, somehow, the decoders DEC_2,...,DEC_L were forced to send i.i.d. equiprobable binary codewords, then channel between ENC_1 and DEC_1 would be exactly the channel $\widetilde{W}$ (which is a binary memoryless channel) with the accurate LLR calculation (1).

In order to achieve the goal of L binary memoryless channels between each encoder-decoder pair simultaneously, we add a binary dither – an i.i.d. equiprobable binary signal – to each encoder-decoder pair as follows. Let the dither signals $\mathbf{d}_l = [d_{l1}, ..., d_{ln}]$, $l \in \{1, ..., L\}$ be L random vectors, each of length n, that are drawn independently from

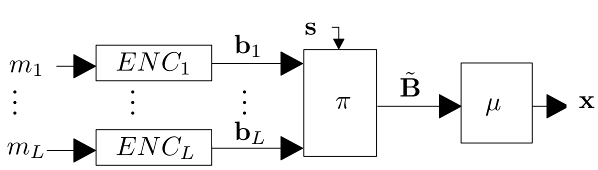


Fig. 4. Interleaving scheme viewed as parallel encoders

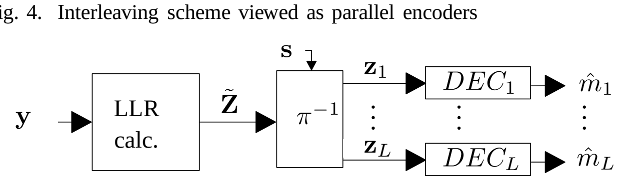


Fig. 5. De-interleaving scheme viewed as parallel decoders

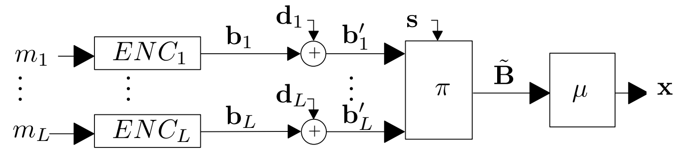


Fig. 6. PBICM encoding scheme. '+' denotes modulo-2 addition (XOR).

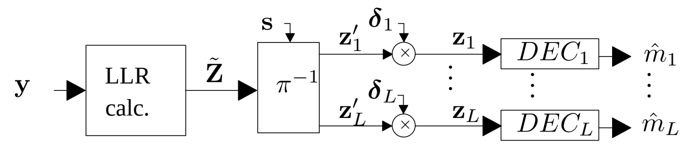


Fig. 7. PBICM decoding scheme. $\boldsymbol{\delta}_l \triangleq 1 - 2 \cdot \mathbf{d}_l$, '$\times$' denotes element-wise multiplication.

a memoryless equiprobable binary source. The output of each encoder ENC_l, $\mathbf{b}_l$, goes through a component-wise XOR operation with the dither vector $\mathbf{d}_l$. The output of the XOR operation, denoted $\mathbf{b}'_l$, is fed into the interleaver π. The full PBICM encoding scheme is shown in Fig. 6.

We let each decoder DEC_l know the value of the dither used by its corresponding encoder ENC_l, $\mathbf{d}_l$ (in practice the dither signals are generated using a pseudo-random generator). In order to compensate for the dither at the decoder, the LLR values are modified by flipping their sign for each dither value of 1 (and maintaining the sign where the dither is 0). Formally, denote the LLR values at the de-interleaver output by $\mathbf{z}'_l = [z'_{l1} \ ... \ z'_{ln}]$. The LLR values at the decoders input shall be denoted by $\mathbf{z}_l = [z_{l1} \ ... \ z_{ln}]$ and given by $z_{lj} = z'_{lj}(1 - 2d_{lj}), \quad j = 1, ..., n$. The PBICM decoding scheme is shown in Fig. 7.

C. Model Equivalence

Before we analyze the channel between each encoder-decoder pair in PBICM, let us define a binary memoryless channel that is related to $\widetilde{W}$, which will prove useful in the analysis of PBICM.

Definition 2: Let $\overline{W}$ be a memoryless binary channel with input B and output $\langle Y, S, D \rangle$: S is drawn at random from $\{1, ..., L\}$, D is drawn at random from $\{0, 1\}$ (S and D are independent, and both do not depend on the input B). Y is the output of the channel W_S with input $B \oplus D$ ($\oplus$ is the XOR operation). Note that the channel $\overline{W}$ is the channel $\widetilde{W}$ where the input is XORed with a binary RV D (see Fig. 8) and that the LLR calculation for the channel $\overline{W}$ is given by $LLR_{\overline{W}}(y, s, d) = (-1)^d LLR_{\widetilde{W}}(y, s) = (-1)^d LLR_s(y)$.

Theorem 1: In parallel BICM, the channel between every encoder-decoder pair is exactly the binary memoryless channel $\overline{W}$, with its exact LLR output.

Proof: Consider the pair ENC_1 and DEC_1. Let $\mathbf{b}_1$ be the codeword sent from ENC_1. After adding the dither $\mathbf{d}_1$, the dithered codeword $\mathbf{b}'_1$ enters the interleaver. The other codewords $\mathbf{b}_2, ..., \mathbf{b}_L$ are dithered using $\mathbf{d}_2, ..., \mathbf{d}_L$. Since the dither of these codewords is unknown at DEC_1, the dithered codewords $\mathbf{b}'_2, ..., \mathbf{b}'_L$ are truly random i.i.d. signals.

The interleaving signal $\mathbf{s}$ interleaves the dithered codewords according to (5). The interleaved signal enters the mapper μ and the channel W, resulting in an output $\mathbf{y}$. Since the dithered codewords $\mathbf{b}'_2, ..., \mathbf{b}'_L$ are i.i.d., the equivalent channel from $\mathbf{b}'_1$ to $\langle \mathbf{y}, \mathbf{s} \rangle$ is exactly the channel $\widetilde{W}$. The LLR calculation at the PBICM receiver along with the interleaver produce $\underset{\smile}{\mathbf{z}'_1}$, which is exactly the LLR calculation that fits the channel $\widetilde{W}$.

Recalling that the channel $\overline{W}$ is nothing but the channel $\widetilde{W}$ with its input XORed with a binary RV, and that the LLR of the channel $\widetilde{W}$ can be easily modified by the dither to produce the LLR of the channel $\overline{W}$, we conclude that the channel between $\mathbf{b}_1$ to $\mathbf{z}_1$ is exactly the channel $\overline{W}$ with LLR calculation.

Since by symmetry the above holds for any encoder-decoder pair ENC_l-DEC_l, the proof is concluded. ■

An important note should be made: Parallel BICM allows the decomposition of the nonbinary channel W to L binary channels of the type $\overline{W}$. These L channels are *not* independent. For example, if W is an additive noise channel, and at some point the noise instance is very strong, this will affect all the decoders and they will fail in decoding together. However, since in the PBICM scheme the channels are used independently, the operation of each decoder is not changed. The outputs of these decoders will inevitably be dependent, and we take this into consideration when analyzing the performance of coding using PBICM in the following.

D. Error Probability Analysis

We wish to analyze the performance of PBICM, and specifically, we are interested in the overall codeword error probability. Let $\mathcal{C}$ be a binary (n, R) code, used in the PBICM scheme. To assure a fair comparison, we regard each L consecutive information messages $(m_1, ..., m_L)$ as a single message m, and regard the scheme as a code of length n on the channel input alphabet $\mathcal{X}$. We define the following error events: Let $\mathcal{E}_l$ be the event of a codeword error in DEC_l, and let $\mathcal{E}$ be the event of an error in *any* of the messages $\{m_1, ..., m_L\}$, i.e. $\mathcal{E} = \bigcup_l \mathcal{E}_l$. Denote the corresponding error probabilities by p_{e_l} and p_e respectively.

Corollary 1: Let $p_e(\overline{W})$ be the codeword error probability of a binary code $\mathcal{C}$ over the channel $\overline{W}$. Then the overall error probability p_e of the code $\mathcal{C}$ when used with PBICM can be bounded by

$$p_e(\overline{W}) \leq p_e \leq L \cdot p_e(\overline{W}). \tag{7}$$

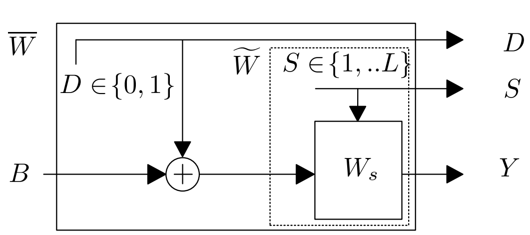


Fig. 8. The binary channel $\overline{W}$. The random state S and the dither D are known at the receiver.

Proof: First note that

$$\min\{p_{e_1}, ..., p_{e_L}\} \leq p_e \leq \sum_l p_{e_l}, \tag{8}$$

where the left inequality follows by definition and the right inequality follows from the union bound. Following Theorem 1 we have $p_{e_1} = p_{e_2} = ... = p_{e_L} = p_e(\overline{W})$ and the proof is concluded. ■

Note that due to similar arguments, the average bit error rate (BER) with PBICM will be equal to the BER attained by the code $\mathcal{C}$ over the channel $\overline{W}$.

In many cases the bit error rate (BER) is of interest. Suppose that each of the messages $(m_1, ..., m_L)$ represents k information bits and the entire message m represents $L \cdot k$ information bits. Let $\mathcal{E}^b_{lk'}$ denote the error in the k'-th bit of the information message m_l. The average BER for the encoder-decoder pair ENC_l-DEC_l is defined by

$$p^b_{e_l} \triangleq \frac{1}{n'} \sum_{k'=1}^{n'} \Pr\{\mathcal{E}^b_{lk'}\}. \tag{9}$$

Similarly, define the overall average BER as

$$p^b_e \triangleq \frac{1}{L \cdot n'} \sum_{l=1}^{L} \sum_{k'=1}^{n'} \Pr\{\mathcal{E}^b_{lk'}\} = \frac{1}{L} \sum_{l=1}^{L} p^b_{e_l}. \tag{10}$$

Corollary 2: Let $p^b_e(\overline{W})$ be the average BER of a binary code $\mathcal{C}$ over the channel $\overline{W}$. Then the average BER p^b_e of the code $\mathcal{C}$ used with PBICM is equal to $p^b_e(\overline{W})$.

Proof: Follows directly from Theorem 1 and from the definition of the average BER in (10). ■

IV. Parallel BICM: Information Theoretical Analysis

In the previous section we defined the PBICM scheme and analyzed its basic error probability properties. The equivalence of the channel between each encoder-decoder pair that was established in Theorem 1 enables a full information-theoretical analysis of the scheme. We show that the highest achievable rate by PBICM (the PBICM capacity) is equal to the BICM capacity as in Equation (4), which should not be a surprise. At the finite-length regime, we derive error exponent and channel dispersion results as information-theoretical measures for optimal PBICM performance at finite blocklength. Note that the definition of error exponent and dispersion are meaningless when the BICM scheme relies on an infinitely long interleaver.

A. Capacity

Let the PBICM capacity of W, $\mathbf{C}^{\text{PBICM}}(W)$, be the highest achievable rate for reliable communication over the channel W with PBICM and a given mapping μ.

Theorem 2: The PBICM capacity is given by

$$\mathbf{C}^{\text{PBICM}}(W) = L \cdot \mathbf{C}(\overline{W}) = \sum_{s=1}^{L} \mathbf{C}(W_s) = \mathbf{C}^{\text{BICM}}(W). \tag{11}$$

Proof: First, by Theorem 1, the performance of a binary code with PBICM and its performance on the binary channel are tightly coupled as in Equation (7). It therefore follows that the PBICM capacity is equal to L times the capacity of the channel $\mathbf{C}(\overline{W})$. Straightforward calculation of the capacity gives $\mathbf{C}(\overline{W}) = \mathbf{C}(\widetilde{W}) = \frac{1}{L}\sum_{s=1}^{L}\mathbf{C}(W_s)$, as required. ■

A note regarding the capacity proof: one might me tempted to try and prove the capacity theorem for PBICM without dither, since with random coding, the code $\mathcal{C}$ is merely an i.i.d. binary random vector. This approach fails because of the following. In the decoding of each codeword, the correctness of the model $\widetilde{W}$ relies on the fact that the *other* codewords are i.i.d. signals. Since PBICM requires a single code for all the L levels, such a condition can never be met. It is possible to prove the achievability without dither when using a different random code at each level, but such an approach will not guarantee the existence of a single code, as required by PBICM.

B. Error Exponent

The error exponent of a channel W is defined by $\mathbf{E}(R) \triangleq \lim_{n\to\infty} -\frac{1}{n}\log\left(p_e(n)\right)$, where $p_e(n)$ is the average codeword error probability for the best code of length n. A lower bound on $\mathbf{E}(R)$ for memoryless channels is the *random coding* error exponent [3], which is given by $\mathbf{E}_r(R) = \max_{\rho\in[0,1]}\max_{P_X(\cdot)}\{\mathbf{E}_0(\rho, P_X) - \rho R\}$, where

$$\mathbf{E}_0(\rho) \triangleq -\log\left[\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}}P_X(x)W(y|x)^{1/(1+\rho)}\right)^{1+\rho}\right]. \tag{12}$$

Since we consider equiprobable inputs only we omit the dependence of $\mathbf{E}_0(\rho)$ in P_X, and omit the maximization w.r.t. P_X in the definition of $\mathbf{E}_r(\cdot)$. Others known bounds on the error exponent include the *expurgation* error exponent lower bound, the *sphere packing* error exponent (an upper bound) and others [3]. We now extend the error exponent definition to the PBICM scheme:

Definition 3: For a given channel W and a mapping μ, let $\mathbf{E}^{\text{PBICM}}(R)$ be defined as[1]

$$\mathbf{E}^{\text{PBICM}}(R) \triangleq \lim_{n\to\infty} -\frac{1}{n}\log\left(p_e(n)\right), \tag{13}$$

where $p_e(n)$ is the average codeword error probability for the best PBICM scheme with block length of n.

Using Corollary 1, we can calculate the PBICM exponent using the error exponent of $\overline{W}$:

Theorem 3: The PBICM error exponent of a channel W is given by

$$\mathbf{E}^{\text{PBICM}}(R) = \mathbf{E}(R/L), \tag{14}$$

where $\mathbf{E}(\cdot)$ is the error exponent of the binary channel $\overline{W}$.

Proof: It follows from (7) that

$$-\frac{1}{n}\log(L\cdot p_e^{(n)}(\overline{W})) \le -\frac{1}{n}\log p_e^{(n)} \le -\frac{1}{n}\log p_e^{(n)}(\overline{W}),$$

where $p_e^{(n)}$ is the PBICM error probability and $p_e^{(n)}(\overline{W})$ is the error probability of the same (n,R) binary code over $\overline{W}$. By taking $n\to\infty$ the factor of L vanishes and we get that for any series of codes,

$$\lim_{n\to\infty} -\frac{1}{n}\log p_e^{(n)} = \lim_{n\to\infty} -\frac{1}{n}\log p_e^{(n)}(\overline{W}). \tag{15}$$

Since the rate for PBICM is L times the rate for coding on $\overline{W}$, the proof is concluded. ■

The channel $\overline{W}$ has a special structure, and is related to the binary sub-channels W_i. We now calculate two basic bounds for the error exponent of $\overline{W}$ in terms of the sub-channels W_i. By Theorem 3, the PBICM error exponent of the channel W can be bounded accordingly.

Theorem 4: Let $\mathbf{E}(R)$ be the error exponent of the channel $\overline{W}$. It can be bounded as follows:

Random coding:

$$\mathbf{E}(R) \ge \mathbf{E}_r(R) = \max_{\rho\in[0,1]}\{\mathbf{E}_0(\rho) - \rho R\}, \tag{16}$$

where

$$\mathbf{E}_0(\rho) = -\log\mathbb{E}\left[2^{-\mathbf{E}_0^{(S)}(\rho)}\right], \tag{17}$$

$\mathbf{E}_0^{(s)}(\rho)$ is the $\mathbf{E}_0$ function for the channel W_s, and the expectation is w.r.t. S which is drawn uniformly from $\{1..L\}$.

Sphere packing:

$$\mathbf{E}(R) \le \mathbf{E}_{sp}(R) = \max_{\rho>0}\{\mathbf{E}_0(\rho) - \rho R\}, \tag{18}$$

where $\mathbf{E}_0(\rho)$ is given in (17).

Proof: The bounds in the theorem are the original random coding and sphere packing exponents [3]. The proof, therefore, boils down to the calculation of $\mathbf{E}_0$ of $\overline{W}$. By writing the definition of $\mathbf{E}_0$ for the channel $\overline{W}$, and noting that the transition probabilities are given by

$$\overline{W}(y,s,d|b) = \frac{1}{2}\widetilde{W}(y,s|b\oplus d) = \frac{1}{2L}W_s(y|b\oplus d), \tag{19}$$

$\mathbf{E}_0$ simplifies to the desired form of (17). ■

Several notes can be made. Since the random coding and sphere packing exponents coincide at rates above the critical rate of $\overline{W}$ (denoted $R_{cr}^{\overline{W}}$), the exact PBICM exponent is known for rates above the *PBICM critical rate* $R_{cr}^{\text{PBICM}} \triangleq L\cdot R_{cr}^{\overline{W}}$. For lower rates we may use any of the other bounds on the error exponent of $\overline{W}$ (although they may not be able to be presented as compactly as in (17)).

The function $\mathbf{E}_0$ of the channel $\widetilde{W}$ is equal to the error exponent of the channel $\overline{W}$. In [6], the authors offered the model of $\widetilde{W}$ for calculating the error exponent of BICM. It is claimed that $\mathbf{E}_0$ of the channel $\widetilde{W}$ is given by [6, Eq. (37)]:

$$\mathbb{E}\left[\mathbf{E}_0^{(S)}(\rho)\right] = \frac{1}{L}\sum_{s=1}^{L}\mathbf{E}_0^{(s)}(\rho). \tag{20}$$

By Theorem 4, this is not the exact expression. In fact, it can be shown using the Jensen inequality that $\mathbf{E}_0(\rho) \le \mathbb{E}\left[\mathbf{E}_0^{(S)}(\rho)\right]$, and therefore the resulting $\mathbf{E}_r(R)$ expression overestimates the true random coding of $\widetilde{W}$.

[1] In general, when defining error exponents, the definitions hold only when the limit exist.

C. Channel Dispersion

An alternative information theoretical measure for quantifying coding performance with finite block lengths is the *channel dispersion*. Suppose that a fixed codeword error probability p_e and a codeword length n are given. We can then seek the maximal achievable rate R given p_e and n. It appears that for fixed p_e and n, the gap to the channel capacity is approximately proportional to $Q^{-1}(p_e)/\sqrt{n}$ (where $Q(\cdot)$ is the complementary Gaussian cumulative distribution function). The proportion constant (squared) is called the channel dispersion. Formally, define the (operational) channel dispersion as follows [5]:

Definition 4: The dispersion $\mathbf{V}(W)$ of a channel W with capacity C is defined as

$$\mathbf{V}(W) = \lim_{p_e\to 0} \limsup_{n\to\infty} \; n \cdot \left(\frac{C - R(n,p_e)}{Q^{-1}(p_e)}\right)^2, \qquad (21)$$

where $R(n, p_e)$ is the highest achievable rate for codeword error probability p_e and codeword length n.

In 1962 , Strassen [4] used the Gaussian approximation to derive the following result for DMCs:

$$R(n,p_e) = C - \sqrt{V/n}Q^{-1}(p_e) + O\left(\frac{\log n}{n}\right), \qquad (22)$$

where C is the channel capacity, and the new quantity V is the (information-theoretic) dispersion , which is given by $V \triangleq VAR(i(X;Y))$, where $i(x;y)$ is the information density, given by $i(x;y) \triangleq \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}$, and the distribution of X is the capacity-achieving distribution that minimizes V. Strassen's result proves that the dispersion of DMCs is equal to $VAR(i(X;Y))$. This result was recently tightened (and extended to the power-constrained AWGN channel) in [5]. It is also known that the channel dispersion and the error exponent are related as follows. For a channel with capacity C and dispersion V, the error exponent can be approximated by $\mathbf{E}(R) \cong \frac{(C-R)^2}{2V \ln 2}$. See [5] for details on the early origins of this approximation by Shannon. We now extend the dispersion definition for PBICM.

Definition 5: The PBICM dispersion $\mathbf{V}^{\text{PBICM}}(W)$ of a channel W and PBICM capacity $\mathbf{C}^{\text{PBICM}}(W)$ is defined as

$$\mathbf{V}^{\text{PBICM}}(W) = \lim_{p_e\to 0} \limsup_{n\to\infty} \; n \cdot \left(\frac{\mathbf{C}^{\text{PBICM}}(W) - R(n,p_e)}{Q^{-1}(p_e)}\right)^2,$$

where $R(n, p_e)$ is the highest achievable rate for any PBICM scheme with a given n and p_e.

Relying on the relationship between the PBICM scheme and the binary channel $\overline{W}$, we can show the following:

Theorem 5: Let n be a given block length and let p_e be a given codeword error probability. Define the highest achievable rate attained using PBICM by $R^{\text{PBICM}}(n, p_e)$, and the gap to the PBICM capacity by $\Delta R = \mathbf{C}^{\text{PBICM}}(W) - R^{\text{PBICM}}(n, p_e)$. Then,

$$\Delta R \le \sqrt{\frac{L^2\mathbf{V}(\overline{W})}{n}}Q^{-1}\left(\frac{p_e}{L}\right) + O\left(\frac{1}{n}\right), \qquad (23)$$

$$\Delta R \ge \sqrt{\frac{L^2\mathbf{V}(\overline{W})}{n}}Q^{-1}(p_e) + O\left(\frac{\log n}{n}\right). \qquad (24)$$

As a result, the PBICM dispersion is given by $\mathbf{V}^{\text{PBICM}}(W) = L^2\mathbf{V}(\overline{W})$.

Proof: The achievability bound (23) follows from [5, Theorem 45] and from Corollary 1. The factor of $1/L$ in the inverse Q function comes from the unavoidable use of the union bound in (7). The converse (24) follows from [5, Theorem 49] and Corollary 1. The PBICM dispersion follows from (23) and (24), and from the fact that $\lim_{\varepsilon\to 0^+} \frac{Q^{-1}(\varepsilon)^2}{2\ln\frac{1}{\varepsilon}} = 1$. ∎

As in the error exponent case, the PBICM dispersion of a channel is related to the dispersion of the binary channel $\overline{W}$. We now calculate it explicitly from the dispersions of the binary sub-channels W_i.

Theorem 6: The dispersion of the channel $\overline{W}$ is given by

$$\begin{aligned}\mathbf{V}(\overline{W}) &= \mathbf{V}(\widetilde{W}) = \mathbb{E}[\mathbf{V}(W_S)] + VAR\left[\mathbf{C}(W_S)\right] = \\ &= \left[\frac{1}{L}\sum_{s=1}^{L}\mathbf{V}(W_s)\right] + VAR(\mathbf{C}(W_S)),\end{aligned} \qquad (25)$$

where $VAR(\mathbf{C}(W_S))$ is the statistical variance of the capacity of W_s, i.e.

$$VAR(\mathbf{C}(W_S)) \triangleq \mathbb{E}[\mathbf{C}^2(W_S)] - \mathbb{E}^2[\mathbf{C}(W_S)]. \qquad (26)$$

Proof: Recall that

$$\overline{W}(y,s,d|b) = \frac{1}{2}\widetilde{W}(y,s|b\oplus d) = \frac{1}{2L}W_s(y|b\oplus d). \qquad (27)$$

The dispersion of all memoryless channels is given by the variance of the information density. For B being the input to the channel $\widetilde{W}$ we it can be shown that the information densities are related as follows:

$$i_{B;Y,S}(b;y,s) = i_{B;Y|S}(b;y|s). \qquad (28)$$

Therefore we have

$$\begin{aligned}\mathbf{V}(W_s) &= VAR(i_{B;Y,S}(B;Y|s)|S=s) \\ &= \mathbb{E}\left[i^2_{B;Y|S}(B;Y|s)|S=s\right] - \mathbf{C}(W_s)^2,\end{aligned}$$

$$\begin{aligned}\mathbf{V}(\widetilde{W}) &= VAR(i_{B;Y|S}(B;Y|S)) \\ &\overset{(a)}{=} \mathbb{E}\left[VAR[i_{B;Y|S}(B;Y|s)|S=s]\right] \\ &\quad + VAR\left[\mathbb{E}[i_{B;Y|S}(B;Y|S)|S=s]\right] \\ &= \mathbb{E}[\mathbf{V}(W_S)] + VAR\left[\mathbf{C}(W_S)\right] \\ &= \left[\frac{1}{L}\sum_{s=1}^{L}\mathbf{V}(W_s)\right] + VAR(\mathbf{C}(W_S)).\end{aligned} \qquad (29)$$

(a) follows from the law of total variance. The dispersion of $\overline{W}$ is calculated similarly, resulting in (26). ∎

Note that since large dispersion means higher backoff from the capacity (see (22)), the term $VAR(\mathbf{C}(W_S))$ can be thought of as a *penalty factor* for the dispersion, over the expected dispersion over the channels W_s, $\mathbb{E}[\mathbf{V}(W_S)]$. This factor grows as the capacities of the sub-channels W_i are more spread. It is interesting to note that the term $\mathbb{E}[\mathbf{V}(W_S)]$ appears in the context of multilevel codes and parallel independent decoding, a scheme which is related to parallel BICM (see [10, Ch. 5]).

V. NUMERICAL RESULTS

In this section we evaluate numerically the PBICM random coding error exponent (see Theorems 3 and 4) in order to compare it with the mismatched decoding exponent [7]. We consider the AWGN and the Rayleigh fading channels.

A. Normalization: Latency vs. Complexity

When the latency of both schemes is fixed to n channel uses, the the PBICM error exponent is generally inferior to that of the mismatched decoding. This can also be seen by observing that the PBICM random coding exponent has a slope of $-1/L$ (in its straight-line region), where the mismatched decoding exponent has a slope of -1. However, it should be taken into consideration that when the block length is n, the mismatched decoder is working with a binary code of length $n \cdot L$. The complexity of the maximum-metric decoder is proportional to the number of codewords $2^{n \cdot L \cdot R}$ [7], where R is the rate of the binary code. On the other hand, the number of codewords in the PBICM scheme is $L \cdot 2^{n \cdot R}$ only. In order to assure a fair comparison *from the complexity point of view*, the PBICM scheme shall use a block length that is L times the block length of the mismatched decoding scheme. Comparing the error probabilities of both schemes gives $nL\mathbf{E}_r^{\text{PBICM}} = n\mathbf{E}_r^{\text{Mismatched}}$. We therefore define the normalized PBICM error exponent as L times the PBICM error exponent (the concept of normalized error exponent was introduced by Arikan [11, Sec. VII] in the context of multilevel codes). We conclude that when the complexity is more important (and the latency is less important), the normalized PBICM exponent is the quantity of interest.

It could be claimed, of course, that practical codes used today (such as low-density parity check (LDPC) codes) will be used and they do not have exponential decoding complexity. On the other hand, such codes do not guarantee an exponentially decaying error probability.

B. Comparison with the Mismatched Decoding Exponent

In the following we show the comparison between the PBICM error exponent and the mismatched decoding error exponent [7]. We show the (unconstrained) random coding error exponent of the channel, along with the mismatched error exponent and the PBICM random coding error exponent (both normalized and un-normalized).

Figure 9 compares the exponents of 16QAM signaling over the Rayleigh fading channel at SNR = 5dB. Throughout the entire range of rates between zero and the BICM capacity, the normalized PBICM random coding exponent is higher (better) than the mismatched decoding exponent. Both BICM exponents are zero for rates above the BICM capacity, and the unconstrained random coding exponent reaches zero at the full channel capacity, as expected. A fact that might be somewhat surprising at first glance is that the normalized PBICM exponent is better than the unconstrained random coding exponent for some rates. While this may seem contradictory, recall that the normalized PBICM scheme essentially takes $n \cdot L$ channel uses so it cannot be considered a scheme that uses a block of n channel uses. The mismatched decoder never attains higher values than the unconstrained exponent, a fact that is known as the data processing inequality for exponents [7, Proposition 3.2]. Fig. 10 shows a similar picture for the case of 16QAM and SNR of 20dB (this behavior was observed over the Rayleigh fading channel for all practical ranges of SNR and 8PSK, 16QAM and 64QAM signaling). For the AWGN channel it cannot be claimed that the normalized PBICM exponent outperforms the mismatched exponent, and the other way around is also not generally true. Examples for both cases are shown in Figures 11(a) and 11(b).

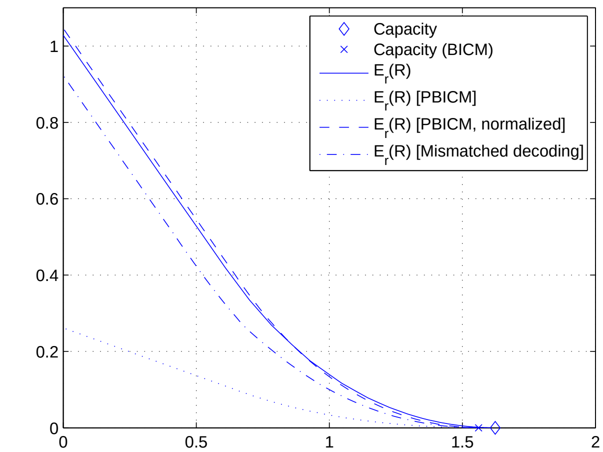


Fig. 9. Random coding exponents over the Rayleigh fading channel with 16-QAM signaling and SNR of 5dB

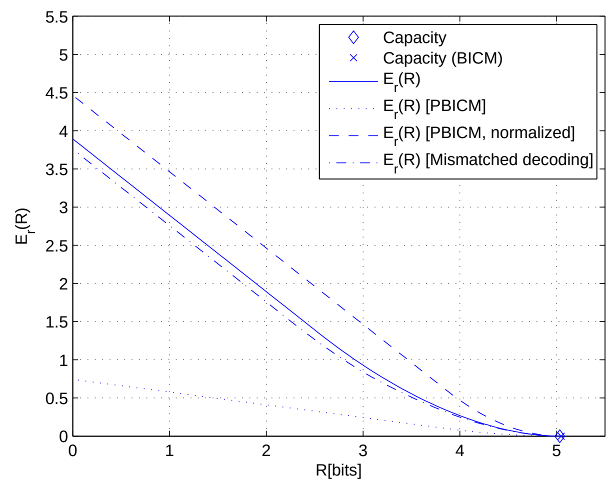
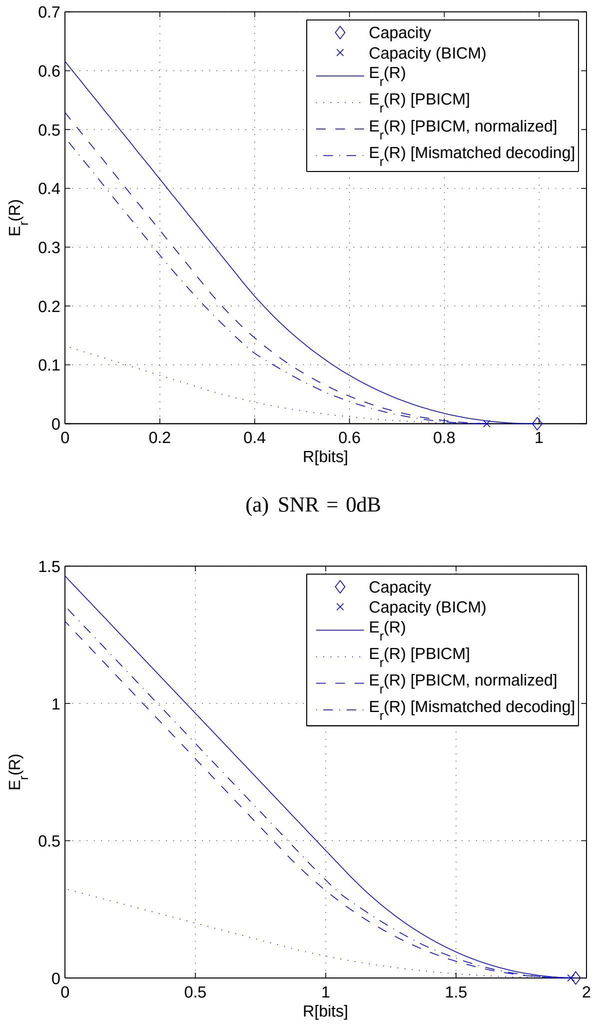


Fig. 10. Random coding exponents over the Rayleigh fading channel with 64-QAM signaling and SNR of 20dB.

VI. SUMMARY AND DISCUSSION

In this paper we have presented *parallel bit-interleaved coded modulation* (PBICM). The scheme is based on a finite-length interleaver and on adding binary dither to the binary



(a) SNR = 0dB

(b) SNR = 5dB

Fig. 11. Random coding exponents over the AWGN channel with 16QAM signaling for SNR values of 0dB and 5dB.

codewords. The scheme is shown to be equivalent to a binary memoryless channel, so it allows easy code design and exact analysis. The scheme was analyzed from an information-theoretical viewpoint, and the capacity, error exponent and the dispersion of the PBICM scheme were calculated.

Throughout the paper, the channel is assumed to be memoryless. This captures many interesting channels, including the AWGN channel, and the memoryless fading channel with (and without) state known at the receiver (ergodic fading). For slow-fading channels, another interleaver (symbol interleaver) is required in order to transform the slowly fading channel into a fast-fading channel (cf. [2]).

Another approach for analyzing BICM at finite block length was proposed in [7], where BICM is thought of as a mismatched decoder. Since this BICM setting uses finite length, the random coding error exponent of the scheme can be calculated. In the previous section we have compared the error exponents of PBICM and of the mismatched decoding approach. When the two schemes have the same latency (same block length) the PBICM exponent is inferior to that of the mismatched decoding approach. However, when the complexity of the scheme is considered (or equivalently, when codeword length of the underlying code is the same), PBICM becomes comparable, and generally better over the Rayleigh fading channel.

An important merit of the PBICM scheme is that it allows an easy code design. In PBICM, one has to design a binary code for a memoryless binary channel. When designing efficient binary codes such as LDPC [12], a desired property of a channel is that its output will be symmetric. It appears that no matter what channel W we have at hand, the resulting binary channel $\overline{W}$ is always output-symmetric (when the output is the LLR).

The PBICM scheme is composed of, among other things, binary dither. Dither is used in some cases as a theoretical tool for achievability proofs. In PBICM, it is an essential part of the scheme itself, and even the random coding capacity proof becomes impossible without it. The main role of the dither is to validate the equivalence of the PBICM scheme to a binary memoryless channel. In addition, the binary dither is the element that symmetrizes the binary channel, which makes the code design easier. This symmetrization property was also noticed by [13] where a similar dither is used with BICM (and termed 'channel adapters'). The code design proposed in [13] rely on the assumption of an ideal interleaver.

Because of its simplicity and easy code design, we conclude that PBICM is an attractive practical communication scheme, which also allows exact theoretical analysis.

Acknowledgment

Interesting discussions with A. G. i Fàbregas are acknowledged.

References

[1] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. on Communications*, vol. 40, no. 5, pp. 873–884, May 1992.
[2] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. on Information Theory*, vol. 44, no. 3, pp. 927–946, 1998.
[3] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons, Inc., 1968.
[4] V. Strassen, "Asymptotische abschätzungen in shannons informationstheorie," *Trans. Third Prague Conf. Information Theory, Czechoslovak Academy of Sciences*, pp. 689–723, 1962.
[5] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. on Information Theory*, vol. 56, no. 5, pp. 2307 –2359, May 2010.
[6] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. on Information Theory*, vol. 45, no. 5, pp. 1361–1391, 1999.
[7] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. Willems, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Trans. on Information Theory*, vol. 55, no. 6, pp. 2756–2765, June 2009.
[8] 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.
[9] E. Agrell and A. Alvarado, "Optimal alphabets and binary labelings for BICM at low SNR," *Information Theory, IEEE Transactions on*, vol. 57, no. 10, pp. 6650 –6672, oct. 2011.
[10] A. Ingber, "Finite blocklength coded modulation," Ph.D. dissertation, Tel Aviv Univeresity, 2011.

[11] E. Arikan, "Channel combining and splitting for cutoff rate improvement," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 628 –639, Feb. 2006.

[12] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 619–637, 2001.

[13] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. on Information Theory*, vol. 49, no. 9, pp. 2141–2155, 2003.