Chalmers Publication Library

Copyright Notice

(Article begins on next page)

On the performance of BICM with trivial interleavers in nonfading channels

Alex Alvarado[§], Leszek Szczecinski and Erik Agrell[§]

INRS-EMT, Montreal, Canada

[§]Department of Signals and Systems, Communication Systems Group

Chalmers University of Technology, Gothenburg, Sweden

{alex.alvarado,agrell}@chalmers.se, *leszek@emt.inrs.ca*

Abstract—Recent results have shown that the performance of bit-interleaved coded modulation (BICM) using convolutional codes in nonfading channels can be greatly improved if the bit-level interleaver takes a trivial form (BICM-T), i.e., if it does not interleave the bits at all. The reported gains reach a few decibels and are obtained using a less complex BICM system. In this paper, we give a formal explanation for these results and show that BICM-T is in fact the combination of a TCM transmitter and a BICM receiver. Analytical bounds that predict the performance of BICM-T are developed and a new type of distance spectrum for the convolutional code is introduced.

I. INTRODUCTION

Ungerboeck's trellis coded modulation (TCM) [1] and Imai and Hirakawa's multilevel coding [2] are probably the most popular coded modulation (CM) schemes for the AWGN channel. Bit-interleaved coded modulation (BICM) [3]–[5] appeared in 1992 as an alternative for CM in fading channels. One particularly appealing feature of BICM is that all the operations are bit-wise, i.e., off-the-shelf binary codes and Gray-mapped constellations are used at the transmitter's side and connected via a bit-level interleaver. At the receiver's side, reliability metrics for the coded bits (L-values) are calculated by the demapper, de-interleaved, and then fed to a binary decoder. This structure gives the designer the flexibility to choose the modulator independently of the encoder, which in turn allows, for example, for an easy adaptation of the transmission to the channel conditions (adaptive modulation and coding). This flexibility is arguably the main advantage of BICM over other CM schemes, and also the reason of why it is used in almost all of the current wireless communications standards, e.g., HSPA, IEEE 802.11a/g/n, and DVB [5, Ch. 1].

Bit-interleaving before modulation was introduced in Zehavi's original paper [3] on BICM. Bit-interleaving is indeed crucial in fading channels since it guarantees that consecutive coded bits to be sent over symbols affected by independent fades. This results in an increase (compared to TCM) of the so-called code diversity, and therefore, BICM is the preferred alternative for CM in fading channels. BICM can also be used in nonfading channels. However, in this scenario, and compared with TCM, BICM gives a smaller minimum Euclidean distance (the proper performance metric in nonfading channels),

and also a smaller constraint capacity [4]. If a Gray labeling is used, the capacity loss is small, and therefore, BICM is still considered valid option for CM over nonfading channels. However, the decrease in minimum Euclidean distance makes BICM less appealing than TCM in nonfading channels.

The use of a bit-level interleaver in nonfading channels has been inherited from the original works on BICM by Zehavi [3] and Caire *et al.* [4] because it simplifies the performance analysis. The use of the interleaver in nonfading channels is considered mandatory in most of the existing literature, however, Martinez *et al.* [6] recently showed from an information theory point of view that the interleaver in BICM is not required.

Previously, we have shown in [7] how—by using multiple interleavers—the performance of BICM can be improved in nonfading channels. Recently, however, it has been shown in [8] that in nonfading channels, considerably larger gains (a few decibels) can be obtained if the interleaver is *completely removed* from the tranceiver's configurations. In other words, it was shown that in nonfading channels BICM without an interleaver performs better than the conventional configurations of [3], [4]. The results presented in [8] are only numerical and an explanation behind such an improvement is not given (although some intuitive explanations and a bit labeling optimization are presented).

In this paper, we study the performance of BICM with trivial interleavers (BICM-T) in nonfading channels for a spectral efficiency of two bits per real channel use, i.e., the BICM system introduced in [8] where no interleaving is performed. We show that BICM-T is the combination of a TCM transmitter and a BICM receiver, and that the transmitter in fact corresponds to particular cases of the so-called pragmatic TCM [9, Ch. 8] introduced by Viterbi *et al.* in [10], Ungerboeck's 1D-TCM [1], and general TCM [11, Fig. 18.11]. We develop analytical bounds which explain why BICM-T with convolutional codes performs well in nonfading channels. We also introduce a new type of distance spectrum for the convolutional codes which allows us to analytically corroborate the results presented in [8]. The main contribution if this paper is to present an analytical model for BICM-T and to provide further understanding for the results in [8].
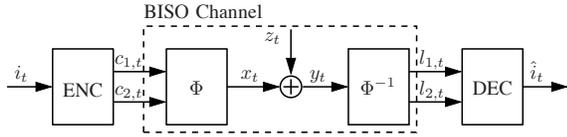
Fig. 1. BICM-T system analyzed in this paper for any time instant t.

II. SYSTEM MODEL AND PRELIMINARIES

Throughout this paper, we use boldface letters $\boldsymbol{c}_t = [c_{1,t}, \ldots, c_{N,t}]$ to denote row vectors and capital boldface letters $\mathbf{C} = [\boldsymbol{c}_1^T, \ldots, \boldsymbol{c}_M^T]$ to denote matrices, where $(\cdot)^T$ denotes transposition. We use $d_H(\mathbf{C})$ to denote the total Hamming weight of the matrix $\mathbf{C}$. We denote probability by $\Pr(\cdot)$ and the probability density function (PDF) of a random variable Λ by $p_\Lambda(\lambda)$. The convolution between two PDFs is denoted by $p_{\Lambda_1}(\lambda) * p_{\Lambda_2}(\lambda)$ and $\{p_\Lambda(\lambda)\}^{*w}$ denotes the w-fold self-convolution of the PDF $p_\Lambda(\lambda)$. A Gaussian distribution with mean value μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$, the Gaussian function with the same parameters by $\psi(\lambda; \mu, \sigma) \triangleq \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(\lambda-\mu)^2}{2\sigma^2})$, and the Q-function by $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{u^2}{2}\right) du$.

A. System Model

We consider a simple BICM configuration shown in Fig. 1. We use a constraint length K rate $R = \frac{1}{2}$ convolutional encoder connected to a 16-ary quadrature amplitude modulation (16-QAM) labeled by the binary reflected Gray code (BRGC) [12]. This configuration is indeed very simple yet practical yielding a spectral efficiency of two bits per real channel use. This example is not restrictive, yet simplifies the presentation of the main ideas. The generalization to other modulations and coding rate is possible but would obviously increase the complexity of notation potentially hindering the main concepts of the analysis presented in this paper.

The input sequence $\boldsymbol{i} = [i_1, \ldots, i_N]$ is fed to the encoder (ENC) which at each time instant $t = 1, \ldots, N$ generates two coded bits $\boldsymbol{c}_t = [c_{1,t}, c_{2,t}]$. We use the matrix $\mathbf{C} = [\boldsymbol{c}_1^T, \ldots, \boldsymbol{c}_N^T]$ of size $2 \times N$ to represent the transmitted codeword. These coded bits are interleaved by Π, where the different interleaving alternatives will be discussed in detail in Sec. II-B. The coded and interleaved bits are then mapped to a 16-QAM symbol, where the 16-QAM constellation is formed by the direct product of two 4-ary pulse amplitude modulation (4-PAM) constellations labeled by the BRGC. Therefore, we analyze the real part of the constellation only, i.e., one of the constituent 4-PAM constellations. The mapper is defined as $\Phi : \{[11], [10], [00], [01]\} \rightarrow \{-3\Delta, -\Delta, \Delta, 3\Delta\}$, where $\Delta = \frac{1}{\sqrt{5}}$ so that the PAM constellation is normalized to unit average symbol energy.

At each time $t = 1, \ldots, N$, the coded bits $\boldsymbol{c}_t$ are mapped to a symbol x_t, where $x_t = \Phi(\boldsymbol{c}_t) \in \mathcal{X}$ and $\mathcal{X}$ is the 4-PAM constellation. The symbols x_t are sent over an additive white Gaussian noise (AWGN) channel so the received signal is given by $y_t = x_t + z_t$, where z_t is a zero-mean Gaussian noise with variance $N_0/2$. The signal-to-noise ratio (SNR) is defined

as $\gamma \triangleq E_s/N_0 = 1/N_0$. At the receiver's side, reliability metrics for the coded bits are calculated by the demapper Φ^{-1} in the form of logarithmic-likelihood ratios (L-values). These L-values are deinterleaved and then passed to the decoder which calculates an estimate of the information sequence $\hat{\boldsymbol{i}}$. As shown in Fig. 1, the modulator, channel, and demodulator can be replaced by a binary-input soft-output (BISO) channel.

The BISO channel in Fig. 1 is nonsymmetric for $c_{2,t}$, i.e., it will give a lower protection to a bit $c_{2,t} = 0$ (two inner symbols in the constellation) compared to a bit $c_{2,t} = 1$ (two outer symbols). To simplify the analysis, we "symmetrize" the channel by randomly inverting the bits before mapping them to the 4-PAM symbol, i.e., we add a scrambler that gives $\tilde{\mathbf{C}} = \mathbf{C} \oplus \mathbf{S}$, where $\oplus$ represents modulo-2 element-wise addition and the elements of the matrix $\mathbf{S} = [\boldsymbol{s}_1^T, \ldots, \boldsymbol{s}_N^T] \in \{0,1\}^{2 \times N}$ where $\boldsymbol{s}_t = [s_{1,t}, s_{2,t}]$ are randomly generated vectors of bits. The L-values are then given by

$$\tilde{l}_{k,t} = \log \frac{\Pr(\tilde{c}_{k,t} = 1|y_t)}{\Pr(\tilde{c}_{k,t} = 0|y_t)}. \tag{1}$$

Since $\tilde{c}_{k,t} = c_{k,t} \oplus s_{k,t}$, it can be shown that $l_{k,t} = (-1)^{s_{k,t}} \tilde{l}_{k,t}$, i.e., after descrambling, the sign of the L-values is changed using $(-1)^{s_{k,t}}$.

We note that the scrambling is introduced only to simplify the analysis, and therefore, it is not shown in Fig. 1 nor used in the simulations. This symmetrization was in fact proposed in [4], and as we will see in Sec. IV, the bounds developed based on this symmetrization perfectly match the simulations.

B. The interleaver

Throughout this paper, three different interleaving alternatives will be analyzed. We put particular attention to the last one. The first interleaving alternative is BICM with a single interleaver (BICM-S). BICM-S was introduced in [4] and is the most commonly used in the literature. BICM-S corresponds to an interleaver that randomly permutes the bits $\mathbf{C}$ prior to modulation, where the permutation is random in two "dimensions," i.e., it permutes the bits over the bit positions and over time. The second alternative is BICM with multiple interleavers (BICM-M). In BICM-M, the interleaver permutes the bits randomly only over time (and not over the bit positions). This can be seen as a particularization of the interleaver of BICM-S, where the following extra constraint is added: bits from the kth encoder's output must be assigned to the kth modulator's input. BICM-M was formally analyzed in [7] and in fact corresponds to the original model introduced by Zehavi in [3] (BICM) and Li and Ritcey in [13] (BICM with iterative decoding). The last interleaving alternative is BICM with a trivial interleaver (BICM-T), i.e., when the interleaver is simply not present [8], cf. Fig. 1.

A careful examination of Fig. 1 reveals that the structure of the transmitter of BICM-T is the same as the transmitter of Ungerboeck's 1D-TCM [1] or the TCM transmitter in [14, Fig. 4.17]. The transmitter of BICM-T can also be considered a particular case of the so-called "general TCM" [11, Fig. 18.11] when $k = \tilde{k}$ (using the notation of [11]) and when the BRGC is used instead of Ungerboeck's set-partitioning. The transmitter

of BICM-T is also equivalent to the simplest configuration of the so-called pragmatic TCM [9, Ch. 8] [10], i.e., when two bits per symbol are considered.

The receiver of BICM-T in Fig. 1 corresponds to a conventional BICM receiver, where L-values for each bit are computed and fed to a soft-input Viterbi decoder (VD). The difference between this receiver's structure and a TCM receiver (one-dimensional or pragmatic) is that bit-level processing is used instead of a symbol-by-symbol VD. In conclusion, the BICM-T system introduced in [8] is a simple TCM transmitter used in conjunction with a BICM receiver. Nevertheless, throughout this paper, we use the name BICM-T to reflect the fact that this transmitter/receiver structure can be considered as a particular case of BICM-S, where the interleaver takes a trivial form. Moreover, the concept behind BICM-T might be useful in adaptive modulation schemes where the interleaver design is adapted to the channel conditions, i.e., if fading is present, BICM-S is used, and if fading is not present, the interleaver is dropped (BICM-T).

C. The Decoder

A maximum likelihood sequence decoder (e.g., the VD) chooses the most likely coded sequence $\hat{\mathbf{C}}$ using the vector of channel observations $\mathbf{y} = [y_1, \ldots, y_N]$ as

$$\hat{\mathbf{C}} = \max_{\mathbf{C} \in \mathcal{D}} \{\log \Pr\{\mathbf{C}|\mathbf{y}\}\} = \max_{\mathbf{C} \in \mathcal{D}} \left\{\log \prod_{t=1}^{N} \Pr\{\mathbf{c}_t|y_t\}\right\}, \quad (2)$$

where $\mathcal{D}$ is the set of all codewords. If we assume that the bits $[c_{1,t}, c_{2,t}]$ are independent, we obtain

$$\log \prod_{t=1}^{N} \Pr\{\mathbf{c}_t|y_t\} = \log \prod_{k=1}^{2} \prod_{t=1}^{N} \Pr\{c_{k,t}|\mathbf{y}\}. \quad (3)$$

Under this independence assumption and by using the relation between an L-value l and the bit's probabilities of being $b \in \{0, 1\}$ $\Pr\{b|y\} = \frac{e^{bl}}{1+e^l}$, we obtain

$$\log \prod_{k=1}^{2} \prod_{t=1}^{N} \Pr\{c_{k,t}|\mathbf{y}\} = \sum_{k=1}^{2} \sum_{t=1}^{N} \log \Pr\{c_{k,t}|\mathbf{y}\}$$
$$= \sum_{k=1}^{2} \sum_{t=1}^{N} c_{k,t} l_{k,t} - \sum_{k=1}^{2} \sum_{t=1}^{N} \log(1 + \exp(l_{k,t})). \quad (4)$$

Since the second term in (4) is independent of $\mathbf{C}$, it is irrelevant to the decision of the decoder in (2). Therefore, the final decision of the decoder can be written as

$$\hat{\mathbf{C}} = \max_{\mathbf{C} \in \mathcal{D}} \left\{\sum_{k=1}^{2} \sum_{t=1}^{N} c_{k,t} l_{k,t}\right\}. \quad (5)$$

In a BICM system with convolutional codes, the decoder is implemented using an off-the-shelf soft-input VD, which assumes that the bits are independent, and thus, uses the relation in (3) (i.e., it uses the decision rule in (5)). The relation in (3) is in indeed valid when BICM-S [4] or BICM-M [3], [7], [15] configurations are used, since in those cases, the use of a random interleaver (cf. Sec. II-B) assure that the bits $[c_{1,t}, c_{2,t}]$ are transmitted in different symbols, and therefore, are affected by different noise realizations.

However, when BICM-T with a soft-input VD is considered, and since the bits $[c_{1,t}, c_{2,t}]$ are affected by the same noise realization, the relation in (3) does not hold, i.e., the two L-values passed to the decoder at any time instant t are not independent. Nevertheless, the decoder treats the bits as independent and still uses the decision rule in (5). A similar observation was made in [6], where the whole BICM decoder is modeled as a mismatched decoder.

III. PERFORMANCE EVALUATION

A. BER Performance

Because of the symmetrization of the channel, we can, without loss of generality, assume that the all-zero codeword was transmitted. We define $\mathcal{E}$ as the set of codewords corresponding to paths in the trellis of the code diverging from the zero-state at the arbitrarily chosen instant $t = t_0$, and remerging with it after T trellis stages. We also denote these codewords as $\mathbf{E} \triangleq [\mathbf{e}_1^{\mathrm{T}}, \ldots, \mathbf{e}_T^{\mathrm{T}}]$, where $\mathbf{e}_t = [e_{1,t}, e_{2,t}]$. Then, the bit error rate (BER) can be upper-bounded using a union bound (UB) as

$$\mathrm{BER} \leq \mathrm{UB} \triangleq \sum_{\mathbf{E} \in \mathcal{E}} \mathrm{PEP}(\mathbf{E}) d_{\mathrm{H}}(\boldsymbol{i}_{\mathbf{E}}), \quad (6)$$

where $d_{\mathrm{H}}(\boldsymbol{i}_{\mathbf{E}})$ is the Hamming weight of the input sequence $\boldsymbol{i}_{\mathbf{E}}$ corresponding to the codeword $\mathbf{E}$, and the pairwise error probability (PEP) is given by

$$\mathrm{PEP}(\mathbf{E}) = \Pr\left\{\sum_{t=t_0}^{t_0+T-1} \left(e_{1,t} l_{1,t} + e_{2,t} l_{2,t}\right) > 0\right\}. \quad (7)$$

The general expression for the PEP in (7) and the UB in (6) reduce to well-known particular cases if simplifying assumptions for the distribution of $l_{k,t}$ are adopted.

1) Independent and identically distributed L-values: In BICM-S [4], the L-values $l_{k,t}$ passed to the decoder are independent and identically distributed. They can be described using the conditional PDF $\mathrm{p}(\lambda|b)$ with $b \in \{0, 1\}$ and where the PDF is independent of k and t. In this case, the PEP in (7) depends only on the Hamming weight of the codeword $\mathbf{E}$, i.e.,

$$\mathrm{PEP}(\mathbf{E}) = \mathrm{PEP}_{\mathrm{S}}(d_{\mathrm{H}}(\mathbf{E})) = \int_0^\infty \{\mathrm{p}(\lambda|b=0)\}^{* d_{\mathrm{H}}(\mathbf{E})} \, \mathrm{d}\lambda. \quad (8)$$

The UB in (6) can be expressed as

$$\mathrm{UB}_{\mathrm{S}} = \sum_w \mathrm{PEP}_{\mathrm{S}}(w) \sum_{\mathbf{E} \in \mathcal{D}_w} d_{\mathrm{H}}(\boldsymbol{i}_{\mathbf{E}}) = \sum_w \mathrm{PEP}_{\mathrm{S}}(w) \beta_w, \quad (9)$$

where $\mathcal{D}_w$ represents the set of codewords with Hamming weight w, i.e., $\mathcal{D}_w \triangleq \{\mathbf{E} \in \mathcal{E} : d_{\mathrm{H}}(\mathbf{E}) = w\}$. In (9), we grouped the codewords $\mathbf{E}$ that have the same Hamming weight and add their contributions, which results in the well-known weight distribution spectrum of the code β_w. The expression in the r.h.s. of (9) is the most common expression for the UB for BICM, cf. [4, eq. (26)], [5, eq. (4.12)].

2) Independent but not identically distributed L-values: In BICM-M [7], the L-values passed to the each decoder's input are independent, however, their conditional PDF depends on k. Thus, the L-values are modeled by the set of conditional PDFs $\{p_1(\lambda|b), p_2(\lambda|b)\}$. The PEP is given by

$$
\begin{aligned}
\text{PEP}(\mathbf{E}) &= \text{PEP}_\text{M}(\overline{w}_{\mathbf{E},1}, \overline{w}_{\mathbf{E},2}) \\
&= \int_0^\infty \{p_1(\lambda|b_1=0)\}^{*\overline{w}_{\mathbf{E},1}} * \{p_2(\lambda|b_2=0)\}^{*\overline{w}_{\mathbf{E},2}}\, d\lambda,
\end{aligned}
$$

where $\overline{w}_{\mathbf{E},k}$ is the Hamming weight of the kth row of $\mathbf{E}$. The UB in (6) can be expressed as

$$
\begin{aligned}
\text{UB}_\text{M} &= \sum_{w_1,w_2} \text{PEP}_\text{M}(w_1,w_2) \sum_{\mathbf{E}\in\mathcal{D}_{w_1,w_2}} d_\text{H}(\boldsymbol{i}_\mathbf{E}) \\
&= \sum_{w_1,w_2} \text{PEP}_\text{M}(w_1,w_2)\beta_{w_1,w_2}, \quad (10)
\end{aligned}
$$

where $\mathcal{D}_{w_1,w_2}$ is the set of codewords with *generalized* Hamming weight $[w_1,w_2]$ (w_k in its kth row), i.e., $\mathcal{D}_{w_1,w_2} \triangleq \{\mathbf{E}\in\mathcal{E} : w_1 = \overline{w}_{\mathbf{E},1}, w_2 = \overline{w}_{\mathbf{E},2}\}$, and β_{w_1,w_2} is the generalized weight distribution spectrum of the code that takes into account the errors at each encoder's output separately. The UB in (10) was shown in [7] to be useful when optimizing the design of the interleaver and the code.

3) BICM with a trivial interleaver: For BICM-T, yet a different particularization of (7) must be adopted. Let $\Lambda_\mathbf{E}$ be the metric associated to the codeword $\mathbf{E}$ and assume without loss of generality that $t_0 = t$. This metric is a sum of independent random variables, i.e.,

$$
\Lambda_\mathbf{E} \triangleq \Lambda_t + \Lambda_{t+1} + \Lambda_{t+2} + \ldots, \quad (11)
$$

where $\Lambda_t = e_{1,t}l_{1,t} + e_{2,t}l_{2,t}$ corresponds to the elements defining the PEP in (7). We then express the tth metric as

$$
\Lambda_t(\boldsymbol{e}_t, \boldsymbol{s}_t) = \begin{cases} 0, & \text{if } \boldsymbol{e}_t = [0,0] \\ (-1)^{s_{1,t}}\tilde{l}_{1,t}, & \text{if } \boldsymbol{e}_t = [1,0] \\ (-1)^{s_{2,t}}\tilde{l}_{2,t}, & \text{if } \boldsymbol{e}_t = [0,1] \\ \sum_{k=1}^2 (-1)^{s_{k,t}}\tilde{l}_{k,t}, & \text{if } \boldsymbol{e}_t = [1,1] \end{cases}, \quad (12)
$$

where we use the notation $\Lambda_t(\boldsymbol{e}_t, \boldsymbol{s}_t)$ to show that Λ_t depends on the scrambling's outcome $\boldsymbol{s}_t$ (through $\tilde{l}_{k,t}$) and also on $\boldsymbol{e}_t$.

Since $\tilde{l}_{k,t}$ are random variables (that depend on k and x_t), according to (12), there exist three PDFs that can be used to model the individual metrics in (11). We denote the set of these three conditional PDFs by $\{p_1(\lambda|b_1), p_2(\lambda|b_2), p_\Sigma(\lambda|\boldsymbol{b})\}$, for the three relevant cases defined in (12), respectively. We note that $p_\Sigma(\lambda|\boldsymbol{b})$ is conditioned not only on one bit, but on the pair of transmitted bits $\boldsymbol{b} = [b_1, b_2]$, where b_1, b_2, and $\boldsymbol{b}$ represent the bits $c_{1,t}$, $c_{2,t}$, and $\boldsymbol{c}_t$, respectively. From (11), and due to the independence of the individual metrics, the PEP in (7) can be expressed as

$$
\begin{aligned}
\text{PEP}(\mathbf{E}) &= \text{PEP}_\text{T}(w_{\mathbf{E},1}, w_{\mathbf{E},2}, w_{\mathbf{E},\Sigma}) \\
&= \int_0^\infty \{p_1(\lambda|b_1=0)\}^{*w_{\mathbf{E},1}} * \{p_2(\lambda|b_2=0)\}^{*w_{\mathbf{E},2}} \\
&\quad * \{p_\Sigma(\lambda|\boldsymbol{b}=[0,0])\}^{*w_{\mathbf{E},\Sigma}}\, d\lambda, \quad (13)
\end{aligned}
$$

where $w_{\mathbf{E},k}$ is the number of columns in $\mathbf{E}$ where only the kth row of $\mathbf{E}$ is one, and $w_{\mathbf{E},\Sigma}$ is the number columns in $\mathbf{E}$

where both entries are equal to one. Clearly, $d_\text{H}(\mathbf{E}) = w_{\mathbf{E},1} + w_{\mathbf{E},2} + 2w_{\mathbf{E},\Sigma}$.

Example 1: Consider the constraint length $K = 3$ optimum distance spectrum convolutional code (ODSCC) with polynomial generators $(5,7)_8$ [16, Table I]. The free distance of the code is $d_\text{H}^\text{free} = 5$, and $\beta_5 = 1$, i.e., there is one divergent path at Hamming distance five from the all-zero codeword, and the Hamming weight of that path is $d_\text{H}(\boldsymbol{i}_\mathbf{E}) = 1$. Moreover, it is possible to show that this codeword is

$$
\mathbf{E} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix},
$$

i.e., $d_\text{H}(\mathbf{E}) = 5$, $w_{\mathbf{E},1} = 0$, $w_{\mathbf{E},2} = 1$, and $w_{\mathbf{E},\Sigma} = 2$. Also, $\overline{w}_{\mathbf{E},1} = 2$ and $\overline{w}_{\mathbf{E},2} = 3$.

We define $\mathcal{D}_{w_1,w_2,w_\Sigma}$ as the set of codewords $\mathbf{E}$ with w_1 columns such that $\boldsymbol{e}_t = [1,0]$, w_2 columns with $\boldsymbol{e}_t = [0,1]$, and w_Σ columns with $\boldsymbol{e}_t = [1,1]$, i.e., $\mathcal{D}_{w_1,w_2,w_\Sigma} \triangleq \{\mathbf{E}\in\mathcal{E} : w_1 = w_{\mathbf{E},1}, w_2 = w_{\mathbf{E},2}, w_\Sigma = w_{\mathbf{E},\Sigma}\}$. Using this, the UB expression in (6) for BICM-T is given by

$$
\begin{aligned}
\text{UB}_\text{T} &= \sum_{w_1,w_2,w_\Sigma} \text{PEP}_\text{T}(w_1,w_2,w_\Sigma) \sum_{\mathbf{E}\in\mathcal{D}_{w_1,w_2,w_\Sigma}} d_\text{H}(\boldsymbol{i}_\mathbf{E}) \\
&= \sum_{w_1,w_2,w_\Sigma} \text{PEP}_\text{T}(w_1,w_2,w_\Sigma)\beta_{w_1,w_2,w_\Sigma}, \quad (14)
\end{aligned}
$$

where β_{w_1,w_2,w_Σ} is a weight distribution spectrum of the code that not only considers the generalized weight $[w_1, w_2]$ of the codewords, but takes into account the temporal behavior, i.e., it considers the case when $\boldsymbol{e}_t = [1,1]$ as a different kind of event. This differs from β_{w_1,w_2}, where such an event will be considered as an extra contribution to the total weight.

B. PDF of the L-values

In order to calculate the PEP for BICM-T in (13) we need the compute the set of conditional PDFs $\{p_1(\lambda|b_1), p_2(\lambda|b_2), p_\Sigma(\lambda|\boldsymbol{b})\}$. In this subsection we show how to find approximations for these PDFs.

The L-values in (1) can be expressed as

$$
\tilde{l}_{k,t}(y_t|\boldsymbol{s}_t) \approx \gamma \left[\min_{x\in\mathcal{X}_{k,0}} (y_t - x)^2 - \min_{x\in\mathcal{X}_{k,1}} (y_t - x)^2 \right], \quad (15)
$$

where $\mathcal{X}_{k,b}$ is the set of constellation symbols labeled with b at bit position k and where we used the max-log approximation $\log(e^a + e^b) \approx \max\{a,b\}$ and the fact that the channel is Gaussian. We use the notation $\tilde{l}_{k,t}(y_t|\boldsymbol{s}_t)$ to emphasize that the L-values depend on the received signal and the scrambler's outcome $\boldsymbol{s}_t$. In fact, the L-values depend on the transmitted symbol x_t, however, and since $\boldsymbol{c}_t = \boldsymbol{0}$ and no interleaving is performed, x_t is determined by $\boldsymbol{s}_t$.

The L-values in (15) are a piece-wise linear function of y_t. Moreover, the L-values $\Lambda_t(\boldsymbol{e}_t, \boldsymbol{s}_t)$ in (12) are linear combinations of $\tilde{l}_{k,t}(y_t|\boldsymbol{s}_t)$ in (15), and therefore, they are also piece-wise linear functions of y_t. Two cases are of particular interest, namely, when $\boldsymbol{e}_t = [1,0]$ or $\boldsymbol{e}_t = [0,1]$, and when $\boldsymbol{e}_t = [1,1]$. The piece-wise linear relationships for the first case are shown in Fig. 2(a) for 4-PAM. In this figure we also show the constellation symbols and we use the notation $\boldsymbol{s}_t = [0/1, :]$ and $\boldsymbol{s}_t = [:, 0/1]$ to show that for $\boldsymbol{e}_t = [1,0]$ and
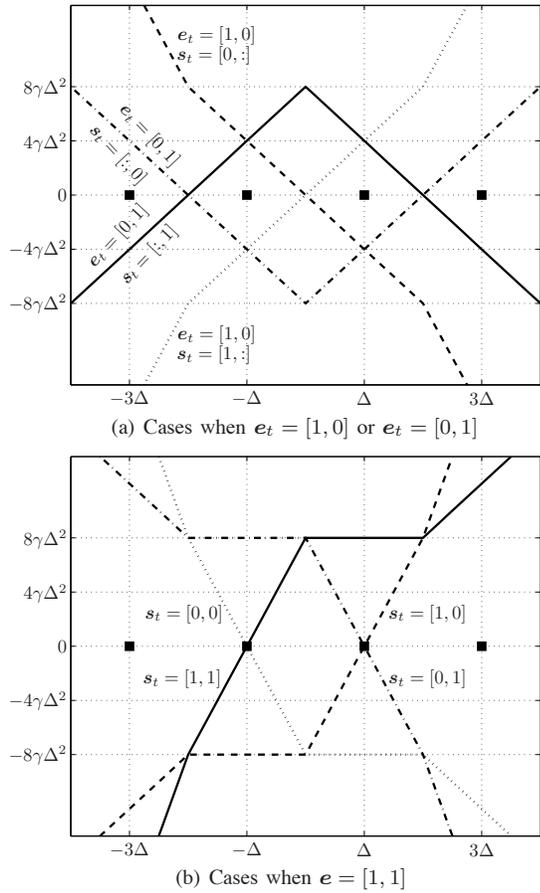
Fig. 2. Piece-wise relation between $\Lambda_t(\boldsymbol{e}_t, \boldsymbol{s}_t)$ and y_t in (12) for 4-PAM and all possible values of $\boldsymbol{s}_t$. The transmitted symbols are shown with squares.

$\boldsymbol{e}_t = [0, 1]$ the L-values $\Lambda_t(\boldsymbol{e}_t, \boldsymbol{s}_t)$ are independent of $s_{2,t}$ and $s_{1,t}$, respectively. In Fig. 2(b), the four possible cases when $\boldsymbol{e}_t = [1, 1]$ are shown.

For a given transmitted symbol x_t (determined by $\boldsymbol{s}_t$), we have that $y_t \sim \mathcal{N}(x_t, N_0/2)$. Therefore, each L-value $\Lambda_t(\boldsymbol{e}_t, \boldsymbol{s}_t)$ in (12) is a sum of piece-wise Gaussian functions[1]. In order to obtain expressions that are easy to work with, we use the so-called zero-crossing approximation of the L-values proposed in [17, Sec. III-C] which replaces all the Gaussian pieces required in the max-log model of L-values by a single Gaussian function. Intuitively, this approximation states that

$$\Lambda_t(y_t | \boldsymbol{e}_t, \boldsymbol{s}_t) \approx \hat{a}(\boldsymbol{e}_t, \boldsymbol{s}_t) y_t + \hat{b}(\boldsymbol{e}_t, \boldsymbol{s}_t), \qquad (16)$$

where $\hat{a}(\boldsymbol{e}_t, \boldsymbol{s}_t)$ and $\hat{b}(\boldsymbol{e}_t, \boldsymbol{s}_t)$ are the slope and the intercept of the closest linear piece to the transmitted symbol x_t.

In Table I we show the values of $\hat{a}(\boldsymbol{e}_t, \boldsymbol{s}_t)$ and $\hat{b}(\boldsymbol{e}_t, \boldsymbol{s}_t)$ defining (16) for 4-PAM, where for notation simplicity we have defined $\alpha \triangleq 4\gamma\Delta^2$. To clarify how these coefficients are obtained, consider for example $\boldsymbol{e}_t = [0, 1]$. In this case, for $\boldsymbol{s}_t = [1, 1]$, which corresponds to $x_t = -3\Delta$, the closest linear piece intersecting the x-axis is the left-most part of the curve labeled in Fig. 2(a) by $\boldsymbol{e}_t = [0, 1]$ and $\boldsymbol{s}_t = [:, 1]$. All the other

[1]Closed-form expressions for these PDFs of $\Lambda_t(\boldsymbol{e}_t, \boldsymbol{s}_t)$ when $\boldsymbol{e}_t = [1, 0]$ and $\boldsymbol{e}_t = [1, 0]$ (cf. Fig. 2(a)) were presented in [17].

TABLE I
VALUES OF $\hat{a} \equiv \hat{a}(\boldsymbol{e}_t, \boldsymbol{s}_t)$ AND $\hat{b} \equiv \hat{b}(\boldsymbol{e}_t, \boldsymbol{s}_t)$ IN (16) FOR 4-PAM FOUND BY DIRECT INSPECTION OF FIG. 2(A) AND FIG. 2(B).

	$\boldsymbol{s}_t = [1,1]$		$\boldsymbol{s}_t = [1,0]$		$\boldsymbol{s}_t = [0,0]$		$\boldsymbol{s}_t = [0,1]$	
	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$
$\boldsymbol{e}_t = [1,0]$	$\frac{\alpha}{\Delta}$	0	$\frac{\alpha}{\Delta}$	0	$-\frac{\alpha}{\Delta}$	0	$-\frac{\alpha}{\Delta}$	0
$\boldsymbol{e}_t = [0,1]$	$\frac{\alpha}{\Delta}$	2α	$-\frac{\alpha}{\Delta}$	-2α	$\frac{\alpha}{\Delta}$	-2α	$-\frac{\alpha}{\Delta}$	2α
$\boldsymbol{e}_t = [1,1]$	$\frac{2\alpha}{\Delta}$	2α	$2\frac{\alpha}{\Delta}$	-2α	$-\frac{2\alpha}{\Delta}$	-2α	$-\frac{2\alpha}{\Delta}$	2α

TABLE II
VALUES OF $\hat{\mu} \equiv \hat{\mu}(\boldsymbol{e}_t, \boldsymbol{s}_t)$ AND $\hat{\sigma}^2 \equiv \hat{\sigma}^2(\boldsymbol{e}_t, \boldsymbol{s}_t)$ GIVEN IN (18) AND (19).

	$\boldsymbol{s}_t = [1,1]$		$\boldsymbol{s}_t = [1,0]$		$\boldsymbol{s}_t = [0,0]$		$\boldsymbol{s}_t = [0,1]$	
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$
$\boldsymbol{e}_t = [1,0]$	-3α	2α	$-\alpha$	2α	$-\alpha$	2α	-3α	2α
$\boldsymbol{e}_t = [0,1]$	$-\alpha$	2α	$-\alpha$	2α	$-\alpha$	2α	$-\alpha$	2α
$\boldsymbol{e}_t = [1,1]$	-4α	8α	-4α	8α	-4α	8α	-4α	8α

values in Table I can be found by a similar direct inspection of Fig. 2(a) and Fig. 2(b).

Using the approximation in (16), the L-values can be modeled as Gaussian random variables where their mean and variance depend on $\boldsymbol{s}_t$, γ, and $\boldsymbol{e}_t$, i.e.,

$$p_{\Lambda_t}(\lambda | \boldsymbol{e}_t, \boldsymbol{s}_t) = \psi(\lambda; \hat{\mu}(\boldsymbol{e}_t, \boldsymbol{s}_t), \hat{\sigma}^2(\boldsymbol{e}_t, \boldsymbol{s}_t)), \qquad (17)$$

where the mean value and variance are given by

$$\hat{\mu}(\boldsymbol{e}_t, \boldsymbol{s}_t) = x_t \hat{a}(\boldsymbol{e}_t, \boldsymbol{s}_t) + \hat{b}(\boldsymbol{e}_t, \boldsymbol{s}_t) \qquad (18)$$

$$\hat{\sigma}^2(\boldsymbol{e}_t, \boldsymbol{s}_t) = [\hat{a}(\boldsymbol{e}_t, \boldsymbol{s}_t)]^2 \frac{N_0}{2}. \qquad (19)$$

In Table II we show the obtained mean values and variances for the same cases presented in Table I.

To obtain the PDF of Λ_t in (12), we simply average (17) over the symbols, which are assumed to be equiprobable. This results in the following expression

$$p_{\Lambda_t}(\lambda) = \begin{cases} \frac{1}{2}[\psi(\lambda; -3\alpha, 2\alpha) + \psi(\lambda; -\alpha, 2\alpha)], & \text{if } \boldsymbol{e}_t = [1, 0] \\ \psi(\lambda; -\alpha, 2\alpha), & \text{if } \boldsymbol{e}_t = [0, 1] \\ \psi(\lambda; -4\alpha, 8\alpha), & \text{if } \boldsymbol{e}_t = [1, 1] \end{cases}. \qquad (20)$$

IV. DISCUSSION AND APPLICATIONS

Expression (20) show the PDF of the L-values needed to compute the UB of BICM-T, cf. (13) and (14). Moreover, the results in (20) only involve Gaussian PDFs, which greatly simplifies the PEP computation in (13).

Theorem 1: The UB for BICM-T is

$$\text{UB}_\text{T} = \sum_{w_1, w_2, w_\Sigma} \beta_{w_1, w_2, w_\Sigma} \left(\frac{1}{2}\right)^{w_1} \sum_{j=0}^{w_1} \binom{w_1}{j} \cdot$$
$$Q\left(\sqrt{\frac{(w_1 + w_2 + 4w_\Sigma + 2j)^2}{(w_1 + w_2 + 4w_\Sigma)} \frac{2\gamma}{5}}\right). \qquad (21)$$

*Proof:* By using (20) in (13), by expressing the convolution of sums as sums of convolutions, and by using
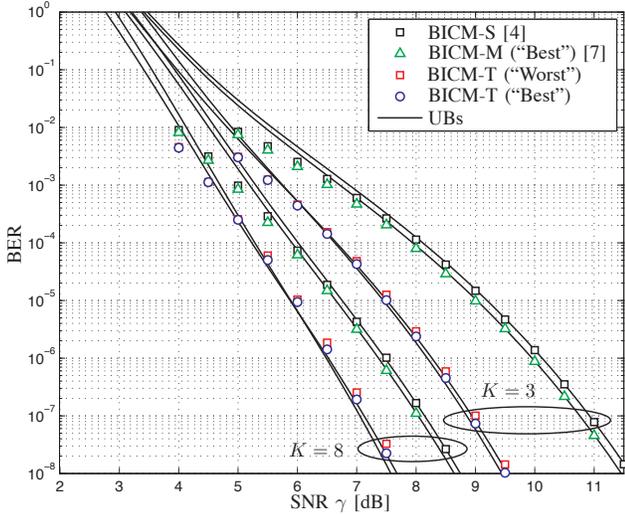
Fig. 3. BER for BICM using the $(5,7)_8$ and $(247,371)_8$ ODSCCs [16] and 4-PAM labeled with the BRGC, and for BICM-S [4], BICM-M [7], and BICM-T. The simulations are shown with markers and the UB with lines.

$\psi(\lambda;\mu_1,\sigma_1^2) * \ldots * \psi(\lambda;\mu_J,\sigma_J^2) = \psi(\lambda;\sum_{j=1}^{J}\mu_j, \sum_{j=1}^{J}\sigma_j^2)$, the PEP in (13) can be expressed as

$$\text{PEP}_\text{T}(w_1,w_2,w_\Sigma) = \int_0^\infty \left(\frac{1}{2}\right)^{w_1} \sum_{j=0}^{w_1}\binom{w_1}{j}\psi(\lambda;\mu,\sigma^2)\,d\lambda, \tag{22}$$

where

$$\mu \equiv \mu_{1,2,\Sigma,j} = -(w_1 + w_2 + 4w_\Sigma + 2j)\alpha \tag{23}$$
$$\sigma^2 \equiv \sigma_{1,2,\Sigma}^2 = 2(w_1 + w_2 + 4w_\Sigma)\alpha. \tag{24}$$

By using the definition of α and Δ^2, and (23) and (24) in (22), and the UB definition in (14), (21) is obtained. ∎

In Fig. 3, numerical results for BICM with 4-PAM labeled with the BRGC and using the ODSCCs $(5,7)_8$ ($K=3$) and $(247,371)_8$ ($K=8$) [16, Table I] are shown. For BICM-M, two configurations are considered for each code. The first one is when all the bits from the first encoder's output are assigned to the first modulator's input and all the bits from the second encoder's output are sent to the second modulator's input. The second alternative simply corresponds to the opposite, i.e., all the bits from the first encoder's output are sent over $k=2$ and the bits from the second encoder's output are sent over $k=1$. This is equivalent to defining the code by swapping the order of the polynomial generators. For these two particular codes, the configuration that minimizes the BER for medium to high SNR is the second one, i.e., when all the bits generated by the polynomial $(7)_8$ or $(371)_8$ are sent over $k=1$ and all the bits generated by the polynomial $(5)_8$ or $(247)_8$ are sent over $k=2$. We denote the configuration that minimizes (or maximizes) the BER by "Best" (or "Worst").

To compute the UB for BICM-S and BICM-M, we use [7, eq. (22)–(23)], and for BICM-T we use Theorem 1[2]. The results in Fig. 3 show that the UB developed in this paper

---

[2] A truncated spectrum of the code $\{w,w_1,w_2,w_\Sigma\} \leq 30$ was calculated using a breadth first search algorithm.

for BICM-T predict well the simulation results. The gains by using BICM-T instead of BICM-S for a BER target of 10^{-7} are approximately 2 dB for $K=3$ and 1 dB for $K=8$.

V. CONCLUSIONS

In this paper, we gave a formal explanation of why gains can be obtained when BICM-T is used in nonfading channels. BICM-T was shown to be a TCM transmitter used with a BICM receiver. An analytical model was developed and a new type of distance spectrum for the code was introduced, which is the relevant characteristic to optimize convolutional codes for BICM-T. The analytical model was used to validate the numerical results and could be used in the future to improve the design of BICM-T.

To have a concise explanation of the mechanisms behind BICM-T, the analysis presented in this paper was done only for a simple BICM configuration. A more general analysis is possible, however, it might require some non-trivial generalizations of the analysis presented in this paper.

REFERENCES

[1] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.
[2] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 3, pp. 371–377, May 1977.
[3] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 873–884, May 1992.
[4] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
[5] 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.
[6] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2756–2765, June 2009.
[7] A. Alvarado, E. Agrell, L. Szczecinski, and A. Svensson, "Exploiting UEP in QAM-based BICM: Interleaver and code design," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 500–510, Feb. 2010.
[8] C. Stierstorfer, R. F. H. Fischer, and J. B. Huber, "Optimizing BICM with convolutional codes for transmission over the AWGN channel," in *International Zurich Seminar on Communications*, Zurich, Switzerland, Mar. 2010.
[9] G. C. Clark Jr. and J. B. Cain, *Error-correction coding for digital communications*, 2nd ed. Plenum Press, 1981.
[10] A. J. Viterbi, J. K. Wolf, E. Zehavi, and R. Padovani, "A pragmatic approach to trellis-coded modulation," *IEEE Commun. Mag.*, vol. 27, no. 7, pp. 11–19, July 1989.
[11] S. Lin and D. J. Costello, Jr., *Error Control Coding*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice Hall, 2004.
[12] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "On the optimality of the binary reflected Gray code," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3170–3182, Dec. 2004.
[13] X. Li and J. Ritcey, "Bit-interleaved coded modulation with iterative decoding using soft feedback," *Electronic Letters*, vol. 34, no. 10, pp. 942–943, May 1998.
[14] S. H. Jamali and T. Le-Ngoc, *Coded-Modulation Techniques for Fading Channels*. Kluwer Academic Publishers, 1994.
[15] A. Alvarado, L. Szczecinski, E. Agrell, and A. Svensson, "On BICM-ID with multiple interleavers," *IEEE Commun. Lett.*, vol. 14, no. 9, pp. 785–787, Sep. 2010.
[16] P. Frenger, P. Orten, and T. Ottosson, "Convolutional codes with optimum distance spectrum," *IEEE Trans. Commun.*, vol. 3, no. 11, pp. 317–319, Nov. 1999.
[17] A. Alvarado, L. Szczecinski, R. Feick, and L. Ahumada, "Distribution of L-values in Gray-mapped M^2-QAM: Closed-form approximations and applications," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 2071–2079, July 2009.