

Copyright Notice

©2011 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

This document was downloaded from Chalmers Publication Library (<http://publications.lib.chalmers.se/>), where it is available in accordance with the IEEE PSPB Operations Manual, amended 19 Nov. 2010, Sec. 8.1.9 (<http://www.ieee.org/documents/opsmanual.pdf>)

(Article begins on next page)

On BICM receivers for TCM transmission

Alex Alvarado, Leszek Szczecinski, and Erik Agrell

Abstract—Recent results have shown that the performance of bit-interleaved coded modulation (BICM) using convolutional codes in nonfading channels can be significantly improved when the interleaver takes a trivial form (BICM-T), i.e., when it does not interleave the bits at all. 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. To predict the performance of BICM-T, a new type of distance spectrum for convolutional codes is introduced, analytical bounds based on this spectrum are developed, and asymptotic approximations are presented. It is shown that the free Hamming distance of the code is not the relevant optimization criterion for BICM-T. Asymptotically optimal convolutional codes for different constraint lengths are tabulated and BICM-T is shown to offer asymptotic gains of about 2 dB over traditional BICM designs based on random interleavers. The asymptotic gains over uncoded transmission are found to be the same as those obtained by Ungerboeck’s one-dimensional trellis-coded modulation (1D-TCM), and therefore, in nonfading channels, BICM-T is shown to be as good as 1D-TCM.

Index Terms—Bit-interleaved Coded Modulation, Binary Reflected Gray Code, Coded Modulation, Convolutional Codes, Interleaver, Quadrature Amplitude Modulation, Pulse Amplitude Modulation, Set Partitioning, Trellis Coded Modulation.

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 done at the bit-level, and thus, at the transmitter’s side, off-the-shelf binary codes are connected to the modulator 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 and the encoder independently, 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-T2/S2/C2.

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 are sent over symbols affected by independent fades. This results in an increase (compared to TCM) of the so-called code diversity (the suitable performance measure in fading channels), 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]. Despite that, if a Gray labeling is used, the capacity loss is small, and therefore, BICM is still considered as a valid option for CM over 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]. It simplifies the performance analysis of BICM and is implicitly considered mandatory in the literature. However, the reasons for its presence are seldom discussed. Previously, we have shown in [6] how the performance of BICM can be improved in nonfading channels by using multiple interleavers. Recently, however, it has been shown in [7] that in nonfading channels, considerably larger gains (a few decibels) can be obtained if the interleaver is *completely removed* from the transceiver’s configurations, i.e., BICM without an interleaver performs better than the conventional configurations of [3], [4]. The results presented in [7] 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 present a formal study of BICM with trivial interleavers (BICM-T) in nonfading channels, i.e., the BICM system introduced in [7] where no interleaving is performed. We recognize BICM-T as the combination of a TCM transmitter and a BICM receiver and we develop analytical bounds that give a formal explanation of why BICM-T with convolutional codes (CCs) performs well in nonfading channels. We also introduce a new type of distance spectrum for the CCs which allows us to analytically corroborate the results presented in [7]. Moreover, we search and tabulate optimum CCs for BICM-T, and we show that the asymptotic gains obtained by BICM-T are the same obtained by Ungerboeck’s one-dimensional TCM (1D-TCM) demonstrating that a properly designed BICM-T system performs asymptotically as well as

Manuscript submitted Aug. 2010; revised Dec. 2010 and Apr. 2011.

This work was partially supported by the European Commission under projects NEWCOM++ (216715) and FP7/2007-2013 (236068), and by the Swedish Research Council, Sweden (2006-5599).

The material in this paper was presented in part at the International Conference on Communications (ICC), Kyoto, Japan, June 2011.

A. Alvarado was with the Dept. of Signals and Systems, Chalmers Univ. of Technology, Sweden, and is now with the Dept. of Engineering, University of Cambridge, Cambridge CB2 1PZ, United Kingdom (email: alex.alvarado@ieee.org).

L. Szczecinski is with the Institut National de la Recherche Scientifique, INRS-EMT, 800, Gauchetiere W. Suite 6900 Montreal, H5A 1K6, Canada (e-mail: leszek@emt.inrs.ca). When this work was submitted for publication, L. Szczecinski was on sabbatical leave with CNRS, Laboratory of Signals and Systems, Gif-sur-Yvette, France.

E. Agrell is with the Dept. of Signals and Systems, Chalmers Univ. of Technology, SE-41296 Göteborg, Sweden (email: agrell@chalmers.se).

1D-TCM. The main contribution of this paper is to present an analytical model for BICM-T which is used to explain the results presented in [7] and also to design an asymptotically optimum BICM-T system in nonfading channels.

Throughout this paper, we use boldface letters \mathbf{c} to denote length- L row vectors $\mathbf{c}_t = [c_{t,1}, \dots, c_{t,L}]$ and also to denote matrices $\mathbf{c} = [c_1^T, \dots, c_N^T]$, where $(\cdot)^T$ denotes transposition. The distinction between a matrix and a vector is clear from the context in which they are used. Random variables are denoted by capital letters C and random vectors/matrices by capital boldface letters \mathbf{C} . We use $d_H(\mathbf{c})$ to denote the total Hamming weight of the binary 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 $p_\Lambda(\lambda)$. The Gaussian pdf with mean value μ and variance σ^2 is denoted by $\psi(\lambda; \mu, \sigma) \triangleq \frac{1}{\sqrt{2\pi}\sigma} \exp(-(\lambda - \mu)^2/2\sigma^2)$, and the Q-function by $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-u^2/2) du$. Sets are denoted using calligraphic letters \mathcal{C} . All the polynomial generators of the CC are given in octal notation and following the notation of [8], the constraint length of the codes K is defined such that the number of states in the trellis of the code is 2^{K-1} .

II. SYSTEM MODEL AND PRELIMINARIES

A. System Model

The BICM system model under consideration is presented in Fig. 1. We use a constraint length K , rate $R = \frac{1}{2}$ convolutional code, connected to a 16-ary quadrature amplitude modulation (16-QAM) labeled by the binary reflected Gray code (BRGC) [9]. This configuration is indeed very simple, yet practical, yielding a spectral efficiency of two information bits per complex channel use. This example simplifies the presentation of the main ideas related to the fact that the interleaver is removed. The generalization to other modulations and coding rates is possible but would increase the complexity of notation potentially hindering the main concepts of the analysis presented in this paper.

The input sequence of N bits $\mathbf{i} = [i_1, \dots, i_N]$ is fed to the encoder (ENC) which at each time instant $t = 1, \dots, N$ generates two coded bits $\mathbf{c}_t = [c_{1,t}, c_{2,t}]$. We use the matrix $\mathbf{c} = [c_1^T, \dots, c_N^T]$ of size 2 by N to represent the transmitted codeword. These coded bits are interleaved by Π , yielding $\mathbf{c}^\pi = \Pi(\mathbf{c})$, where the different interleaving alternatives will be discussed in detail in Sec. II-B. The coded and interleaved bits are then mapped to 16-QAM symbols, 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

$$\Delta \triangleq \frac{1}{\sqrt{5}} \quad (1)$$

so that the PAM constellation normalized to unit average symbol energy, i.e., $E_s = 1$.

A quick inspection of the BRGC for 4-PAM (cf. Fig. 3) reveals that the BRGC offers unequal error protection (UEP) to the transmitted bits depending on their position. In particular, the bit at the first position ($k = 1$) receives higher protection¹ than the bit at the second position $k = 2$. More details about this can be found in [6]. Moreover, for $k = 2$ a bit labeled by zero (inner constellation points) will receive a lower protection than a bit labeled by one transmitted in the same bit position (outer constellation points), and therefore, the binary-input soft-output (BISO) channel for $k = 2$ is nonsymmetric. To simplify the analysis, we “symmetrize” the channel by randomly inverting the bits before mapping them to the 4-PAM symbol, i.e., $\tilde{\mathbf{c}}^\pi = \mathbf{c}^\pi \oplus \mathbf{s} = [\tilde{c}_1^\pi, \dots, \tilde{c}_N^\pi]$, where \oplus represents modulo-2 element-wise addition and the elements of the matrix $\mathbf{s} = [s_1^T, \dots, s_N^T] \in \{0, 1\}^{2 \times N}$, with $\mathbf{s}_t = [s_{1,t}, s_{2,t}]$, are randomly generated vectors of bits. Such a scrambling symmetrizes the BISO channel but it does not eliminate the UEP. 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 numerical simulations.

We consider transmissions over an additive white Gaussian noise (AWGN) channel. Assuming an ideal matched filter and perfect synchronization, the equivalent discrete-time baseband received signal is given by $y_t = x_t + z_t$, where z_t is a zero-mean Gaussian noise with variance $N_0/2$ (with $N_0/2$ being the power spectral density of the continuous-time AWGN), x_t represents the transmitted symbol, and $y_t, x_t, z_t \in \mathbb{R}$. At each discrete time $t = 1, \dots, N$, the coded, interleaved, and scrambled bits \tilde{c}_t^π are mapped to a symbol x_t , where $x_t = \Phi(\tilde{c}_t^\pi) \in \mathcal{X}$ and \mathcal{X} is the 4-PAM constellation. The signal-to-noise ratio is defined as $\gamma \triangleq E_s/N_0 = 1/N_0$.

At the receiver’s side, reliability metrics for the bits are calculated by the demapper Φ^{-1} in the form of logarithmic-likelihood ratios (L-values) as

$$\tilde{l}_{k,t}^\pi = \log \frac{p_{Y_t}(y_t | \tilde{C}_{k,t}^\pi = 1)}{p_{Y_t}(y_t | \tilde{C}_{k,t}^\pi = 0)}, \quad (2)$$

which allows us to write

$$p_{Y_t}(y_t | \tilde{C}_{k,t}^\pi = u) = \frac{\exp(u \tilde{l}_{k,t}^\pi)}{1 + \exp(\tilde{l}_{k,t}^\pi)}, \quad (3)$$

with $u \in \{0, 1\}$. Since $\tilde{c}_{k,t}^\pi = c_{k,t}^\pi \oplus s_{k,t}$, it can be shown that $l_{k,t}^\pi$ passed to the deinterleaver (cf. Fig. 1) can be written as

$$l_{k,t}^\pi = (-1)^{s_{k,t}} \tilde{l}_{k,t}^\pi, \quad (4)$$

i.e., after “descrambling”, the sign of the L-values is changed using $(-1)^{s_{k,t}}$. These L-values are deinterleaved by Π^{-1} yielding the matrix $\mathbf{l} = \Pi^{-1}(\mathbf{l}^\pi)$ of size 2 by N , which is then passed to the decoder which calculates an estimate of the information sequence $\hat{\mathbf{i}}$.

¹The protection may be defined in different ways, where arguably the simplest one is the bit error probability per bit position at the demapper’s output.

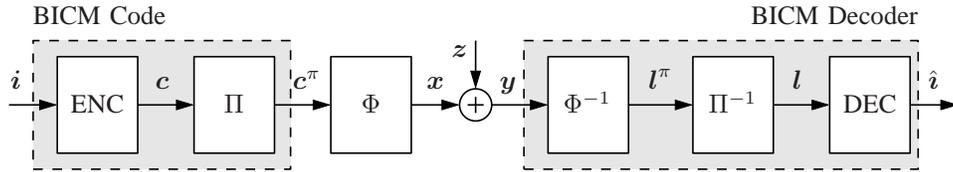


Fig. 1. Model of BICM transmission.

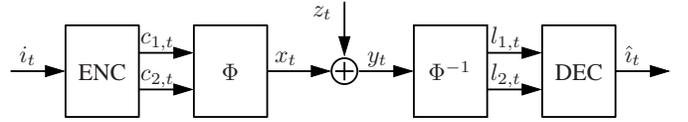
B. The Interleaver

Throughout this paper, three interleaving alternatives (cf. Π in Fig. 1) will be discussed: BICM with a single interleaver (BICM-S), BICM with multiple interleavers (M-interleavers, BICM-M), and BICM with a trivial interleaver (BICM-T). A brief description of the three interleaving alternatives is given below. This paper focuses on BICM-T.

- BICM-S was introduced in [4] and is the most common configuration analyzed in the literature. It corresponds to an interleaver that randomly permutes the bits in 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.
- BICM-M can be seen as a particularization of BICM-S with the following additional constraint: bits from the encoder's k th output must be assigned to the modulator's k th input. BICM-M was formally analyzed in [6] and in fact corresponds to the original model introduced by Zehavi in [3] (BICM) and Li in [10] (BICM with iterative decoding, BICM-ID). Recently, M-interleavers have also been proven to be asymptotically optimum for BICM-ID [11].
- BICM-T was introduced in [7]. In this configuration, the interleaver Π in Fig. 1 is not present, i.e., $c^\pi = c$ and $l^\pi = l$.

When BICM-T is considered, the resulting system is the one shown in Fig. 2. A careful examination of Fig. 2 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 [12, Fig. 4.17]. The transmitter of BICM-T can also be considered a particular case of the so-called general TCM [13, Fig. 18.11] when $k = \hat{k}$ (using the notation of [13]) and when the BRGC is used instead of Ungerboeck's set-partitioning (SP). The transmitter of BICM-T is also equivalent to the simplest configuration of the so-called pragmatic TCM [14, Ch. 8] [15] (see also [16], [17, Sec. 9.2.4]), i.e., when two bits per symbol are considered.

The receiver of BICM-T in Fig. 2 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 [7] 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


 Fig. 2. BICM-T system analyzed in this paper for any time instant t .

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).

As explained above, the main difference of the system in Fig. 2 and pragmatic TCM is the receiver used. Also, it is important to note that if BICM-T with larger constellations (e.g., 8-PAM) is considered, the binary labeling will still be the BRGC, which is different than the one used in pragmatic TCM, cf. [14, Fig. 8-30]. A detailed analysis/comparison of BICM-T and pragmatic TCM for larger constellations is, however, out of the scope of this paper.

C. The Decoder and the decoding errors

A maximum likelihood (ML) sequence decoder chooses the most likely coded sequence c using the vector of channel observations $\mathbf{y} = [y_1, \dots, y_N]$ (cf. Fig. 1), i.e.,

$$\hat{c}^{\text{ML}} = \underset{c \in \mathcal{C}}{\operatorname{argmax}} \{ \log p_{\mathbf{Y}}(\mathbf{y} | \mathbf{C} = c) \} \quad (5)$$

$$= \underset{c \in \mathcal{C}}{\operatorname{argmax}} \{ \log p_{\mathbf{Y}}(\mathbf{y} | \Pi(\mathbf{C}) = \Pi(c)) \} \quad (6)$$

$$= \underset{c \in \mathcal{C}}{\operatorname{argmax}} \left\{ \log \prod_{t=1}^N p_{Y_t}(y_t | C_{1,t}^\pi = c_{1,t}^\pi, C_{2,t}^\pi = c_{2,t}^\pi) \right\}, \quad (7)$$

where \mathbf{C} is the transmitted codeword, \mathcal{C} is the code, to pass from (5) to (6) we use the fact that the interleaver is a bijective mapping, and to pass from (6) to (7) we used the fact that the noise samples z_t affecting y_t are mutually independent.

The striking feature of the BICM decoder shown in Fig. 1 is that it replaces the decoding metric used in (7) with the

metric calculated at the bit-level, i.e.,

$$\hat{c}^{\text{BICM}} = \underset{c \in \mathcal{C}}{\operatorname{argmax}} \left\{ \log \prod_{t=1}^N p_{Y_t}(y_t | C_{1,t}^\pi = c_{1,t}^\pi) \cdot p_{Y_t}(y_t | C_{2,t}^\pi = c_{2,t}^\pi) \right\} \quad (8)$$

$$= \underset{c \in \mathcal{C}}{\operatorname{argmax}} \left\{ \sum_{t=1}^N \sum_{k=1}^2 \log p_{Y_t}(y_t | C_{k,t}^\pi = c_{k,t}^\pi) \right\}. \quad (9)$$

The optimal metric used in (7) is not the same as the one used in (9), and thus, it is sometimes referred to as a mismatched metric [18, Sec. II-B.2].

Using an expression analogous to (3), (9) can be expressed as

$$\hat{c}^{\text{BICM}} = \underset{c \in \mathcal{C}}{\operatorname{argmax}} \left\{ \sum_{t=1}^N \sum_{k=1}^2 c_{k,t}^\pi l_{k,t}^\pi - \sum_{t=1}^N \sum_{k=1}^2 \log(1 + \exp(l_{k,t}^\pi)) \right\} \quad (10)$$

$$= \underset{c \in \mathcal{C}}{\operatorname{argmax}} \left\{ \sum_{t=1}^N \sum_{k=1}^2 c_{k,t}^\pi l_{k,t}^\pi \right\}, \quad (11)$$

where the second term in (10) being independent of c is irrelevant to the decision of the decoder.

The decision of the decoder based on the rule (11) is erroneous if it detects a codeword \hat{c} instead of the transmitted codeword c . The probability of this event, the so-called *pairwise error probability* (PEP), is defined by

$$\operatorname{PEP}(c \rightarrow \hat{c}) \triangleq \Pr \left\{ \sum_{t=1}^N (\hat{c}_{1,t} l_{1,t} + \hat{c}_{2,t} l_{2,t}) \geq \sum_{t=1}^N (c_{1,t} l_{1,t} + c_{2,t} l_{2,t}) \right\} \quad (12)$$

$$= \Pr \left\{ \sum_{t=1}^N (e_{1,t} l_{1,t} + e_{2,t} l_{2,t}) \geq 0 \right\}, \quad (13)$$

where $e_{k,t}$ are the elements of the ‘‘error’’ codeword $e = \hat{c} - c$.

The PEP in (13) depends on $2N$ L-values l_1, l_2, \dots, l_N and its evaluation is, in general, difficult because this sequence contains pairs of dependent L-values that were calculated from the same channel outcome (same noise realization). However, when convolutional codes are considered, the most relevant events are those involving a relatively small number of errors $e_t, t_0 \leq t \leq t_0 + T - 1$, which means that PEP is affected by the consecutive L-values $l_{t_0}, l_{t_0+1}, \dots, l_{t_0+T-1}$. These $2T$ L-values will be independent if we ensure that all the bits $c_{t_0}, c_{t_0+1}, \dots, c_{t_0+T-1}$ are transmitted in $2T$ different time instants (and thus, affected by different noise realizations) after interleaving. This condition of *local independence* can be obtained by an appropriate design of the interleaver, and is very likely to be satisfied when BICM-S or BICM-M are used.

When BICM-T is considered, for each $t = 1, \dots, N$, two bits c_t will be transmitted using the same symbol x_t , and thus, the two corresponding L-values l_t will be mutually dependent

(because they were obtained from the same observation y_t). In the following subsection, we show how to analyze the performance of such a system.

III. PERFORMANCE EVALUATION

A. BER Analysis

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 $e \triangleq [e_1^T, \dots, e_T^T]$, where $e_t = [e_{1,t}, e_{2,t}]$. Then, the bit error rate (BER) can be upper-bounded using a union bound (UB) as

$$\operatorname{BER} \leq \operatorname{UB} \triangleq \sum_{e \in \mathcal{E}} \operatorname{PEP}(e) d_H(\mathbf{i}_e), \quad (14)$$

where $d_H(\mathbf{i}_e)$ is the Hamming weight of the input sequence \mathbf{i}_e corresponding to the codeword $c = e$, and the pairwise error probability (PEP) is given by (cf. (13))

$$\operatorname{PEP}(e) = \Pr \left\{ \sum_{t=t_0}^{t_0+T-1} (e_{1,t} l_{1,t} + e_{2,t} l_{2,t}) > 0 \right\}. \quad (15)$$

The general expression for the PEP in (15) and the UB in (14) reduce to well-known cases if simplifying assumptions for the distribution of $l_{k,t}$ are adopted. To clarify the main differences between BICM-T and BICM-S/BICM-M, in the following, we briefly analyze these well-known cases.

1) *Independent and identically distributed L-values (BICM-S)*: In BICM-S, and because of the interleaver, the L-values $l_{k,t}$ passed to the decoder are locally independent (cf. Sec. II-C) and identically distributed (i.i.d.). They can be described using the conditional pdf $p_L(\lambda|B)$ with $B \in \{0, 1\}$ and where the pdf is independent of k and t . In this case, the PEP in (15) depends only on the Hamming weight of the codeword e , i.e., the PEP is given by (16) (at the bottom of next page). The UB in (14) can be expressed as

$$\operatorname{UB}_S = \sum_w \operatorname{PEP}_S(w) \sum_{e \in \mathcal{C}_w} d_H(\mathbf{i}_e) \quad (17)$$

$$= \sum_w \operatorname{PEP}_S(w) \beta_w^{\mathcal{C}}, \quad (18)$$

where \mathcal{C}_w represents the set of codewords with Hamming weight w , i.e., $\mathcal{C}_w \triangleq \{e \in \mathcal{E} : d_H(e) = w\}$, and \mathcal{C} denotes the convolutional code used for transmission. To pass from (17) to (18) we group the codewords e that have the same Hamming weight and add their contributions, which results in the well-known (input-output) weight distribution spectrum of the code $\beta_w^{\mathcal{C}}$. The expression in (18) 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 (BICM-M)*: In BICM-M, the L-values passed to the each decoder’s input are locally independent, however, their distributions depend on the bit’s position $k = 1, 2$. Therefore, we need to use the conditional pdfs $p_{L_1}(\lambda|B_1)$ and $p_{L_2}(\lambda|B_2)$, where L_1 and L_2 are the random variables representing the

L-values at the decoder's inputs. The PEP in this case is given by (19) (shown at the bottom of the page), where $\bar{w}_{e,k}$ is the Hamming weight of the k th row of e . The UB in (14) can be expressed as

$$\begin{aligned} \text{UB}_M &= \sum_{w_1, w_2} \text{PEP}_M(w_1, w_2) \sum_{e \in \mathcal{C}_{w_1, w_2}} d_H(\mathbf{i}_e) \\ &= \sum_{w_1, w_2} \text{PEP}_M(w_1, w_2) \beta_{w_1, w_2}^{\mathcal{C}}, \end{aligned} \quad (20)$$

where \mathcal{C}_{w_1, w_2} is the set of codewords with *generalized* Hamming weight $[w_1, w_2]$ (w_k in its k th row), i.e., $\mathcal{C}_{w_1, w_2} \triangleq \{e \in \mathcal{E} : w_1 = \bar{w}_{e,1}, w_2 = \bar{w}_{e,2}\}$, and $\beta_{w_1, w_2}^{\mathcal{C}}$ is the generalized weight distribution spectrum of the code that takes into account the errors at each encoder's output separately. The UB in (20) was shown in [6] to be useful when analyzing the UEP introduced by the binary labeling and also to optimize the interleaver and the code.

3) *BICM without bit-interleaving (BICM-T)*: For BICM-T, yet a different particularization of (15) must be adopted. Let Λ_e be the metric associated to the codeword e and assume without loss of generality that $t_0 = t$, cf. (15). This metric is a sum of independent random variables, i.e.,

$$\Lambda_e \triangleq \Lambda^{(t)} + \Lambda^{(t+1)} + \Lambda^{(t+2)} + \dots, \quad (21)$$

where $\Lambda^{(t)} = e_{1,t}l_{1,t} + e_{2,t}l_{2,t}$ which corresponds to the L-values defining the PEP in (15). Because the interleaver is removed, $l_{k,t} = \tilde{l}_{k,t}^\pi$, and thus, by using (4), we express each of these metrics as

$$\Lambda^{(t)} \equiv \Lambda(e_t, \mathbf{s}_t) = \begin{cases} 0, & \text{if } e_t = [0, 0] \\ (-1)^{s_{1,t}} \tilde{l}_{1,t}^\pi, & \text{if } e_t = [1, 0] \\ (-1)^{s_{2,t}} \tilde{l}_{2,t}^\pi, & \text{if } e_t = [0, 1] \\ \sum_{k=1}^2 (-1)^{s_{k,t}} \tilde{l}_{k,t}^\pi, & \text{if } e_t = [1, 1] \end{cases}, \quad (22)$$

where we use $\Lambda(e_t, \mathbf{s}_t)$ to show that $\Lambda^{(t)}$ depends on the scrambling's outcome \mathbf{s}_t (through $\tilde{l}_{k,t}^\pi$) and the error pattern at time t , e_t .

Since $\tilde{l}_{k,t}^\pi$ are random variables (that depend on k and x_t), we need three pdfs: $p_{\Lambda_1}(\lambda|B_1)$, $p_{\Lambda_2}(\lambda|B_2)$, and $p_{\Lambda_\Sigma}(\lambda|\mathbf{B})$, for the three relevant cases defined in (22). We note that $p_{\Lambda_\Sigma}(\lambda|\mathbf{B})$ is conditioned not only on one bit, but on the pair of transmitted bits $\mathbf{B} = [B_1, B_2]$, where B_1, B_2 , and \mathbf{B} represent the bits $C_{1,t}^\pi, C_{2,t}^\pi$, and C_t^π , respectively. From (21), and due to the independence of the individual L-values, the PEP in (15) can be expressed as (23) (shown at the bottom of the page) where $w_{e,1}$, $w_{e,2}$, and $w_{e,\Sigma}$ are, respectively, the number of

columns in e being equal to $e_t = [1, 0]^T$, $e_t = [0, 1]^T$, and $e_t = [1, 1]^T$.² Then, the UB expression in (14) becomes

$$\begin{aligned} \text{UB}_T &= \sum_{w_1, w_2, w_\Sigma} \text{PEP}_T(w_1, w_2, w_\Sigma) \sum_{e \in \mathcal{C}_{w_1, w_2, w_\Sigma}} d_H(\mathbf{i}_e) \\ &= \sum_{w_1, w_2, w_\Sigma} \text{PEP}_T(w_1, w_2, w_\Sigma) \beta_{w_1, w_2, w_\Sigma}^{\mathcal{C}}, \end{aligned} \quad (24)$$

where $\mathcal{C}_{w_1, w_2, w_\Sigma} \triangleq \{e \in \mathcal{E} : w_1 = w_{e,1}, w_2 = w_{e,2}, w_\Sigma = w_{e,\Sigma}\}$ and $\beta_{w_1, w_2, w_\Sigma}^{\mathcal{C}}$ is a new weight distribution spectrum of the code \mathcal{C} that takes into account the generalized weight $[w_1, w_2, w_\Sigma]$ of the codewords, i.e., it considers separately the case when $e_t = [1, 1]$. This is different from $\beta_{w_1, w_2}^{\mathcal{C}}$, where such a case will increase w_1 and w_2 in the generalized weight $[w_1, w_2]$. Clearly, the following relationships hold

$$d_H(e) = w_{e,1} + w_{e,2} + 2w_{e,\Sigma} \quad (25)$$

$$\bar{w}_{e,1} = w_{e,1} + w_{e,\Sigma} \quad (26)$$

$$\bar{w}_{e,2} = w_{e,1} + w_{e,\Sigma}. \quad (27)$$

Example 1: Consider the constraint length $K = 3$ optimum distance spectrum convolutional code (ODSCC) with polynomial generators (5, 7) [19, Table I]. The free distance of the code is $d_H^{\text{free}} = 5$, and $\beta_5^{\mathcal{C}} = 1$, i.e., there is one divergent path at Hamming distance five from the all-zero codeword, and the Hamming weight of that path is $d_H(\mathbf{i}_e) = 1$. Moreover, it is possible to show that this codeword is given by

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

i.e., $d_H(e) = 5$, $w_{e,1} = 0$, $w_{e,2} = 1$, and $w_{e,\Sigma} = 2$. If BICM-M is considered, $\bar{w}_{e,1} = 2$ and $\bar{w}_{e,2} = 3$.

B. PDF of the L-values

In order to calculate the PEP for BICM-T in (23) we need to compute the conditional pdfs $p_{\Lambda_1}(\lambda|B_1)$, $p_{\Lambda_2}(\lambda|B_2)$, and $p_{\Lambda_\Sigma}(\lambda|\mathbf{B})$. In this subsection we show how to find approximations for these PDFs.

The L-values in (2) can be expressed as

$$\tilde{l}_{k,t}^\pi = \log \frac{\sum_{x \in \mathcal{X}_{k,1}} p_{Y_t}(y_t|X_t = x)}{\sum_{x \in \mathcal{X}_{k,0}} p_{Y_t}(y_t|X_t = x)}, \quad (28)$$

where $\mathcal{X}_{k,b}$ is the set of constellation symbols labeled with b at bit position k . Using the fact that the channel is Gaussian and if

²We note that the three arguments $w_{e,1}$, $w_{e,2}$ and $w_{e,\Sigma}$ are a consequence of the code rate $R = 1/2$ considered in this paper. For other code rates, more arguments will be needed.

$$\text{PEP}(e) = \text{PEP}_S(d_H(e)) = \int_0^\infty \{p_L(\lambda|b=0)\}^{*d_H(e)} d\lambda. \quad (16)$$

$$\text{PEP}(e) = \text{PEP}_M(\bar{w}_{e,1}, \bar{w}_{e,2}) = \int_0^\infty \{p_{L_1}(\lambda|B_1=0)\}^{*\bar{w}_{e,1}} * \{p_{L_2}(\lambda|B_2=0)\}^{*\bar{w}_{e,2}} d\lambda, \quad (19)$$

$$\text{PEP}(e) = \text{PEP}_T(w_{e,1}, w_{e,2}, w_{e,\Sigma}) = \int_0^\infty \{p_{\Lambda_1}(\lambda|B_1=0)\}^{*w_{e,1}} * \{p_{\Lambda_2}(\lambda|B_2=0)\}^{*w_{e,2}} * \{p_{\Lambda_\Sigma}(\lambda|\mathbf{B}=[0,0])\}^{*w_{e,\Sigma}} d\lambda, \quad (23)$$

the so-called max-log approximation $\log(\exp(a) + \exp(b)) \approx \max\{a, b\}$ is used, the L-values can be expressed as³

$$\tilde{l}_{k,t}^\pi(y_t | \mathbf{s}_t) \approx \gamma \left[\min_{x \in \mathcal{X}_{k,0}} (y_t - x)^2 - \min_{x \in \mathcal{X}_{k,1}} (y_t - x)^2 \right], \quad (29)$$

where from now on we use the notation $\tilde{l}_{k,t}^\pi(y_t | \mathbf{s}_t)$ to emphasize that the L-values depend on the received signal and the scrambler's outcome \mathbf{s}_t . In fact, the L-values depend on the transmitted symbol x_t , however, since the all-zero codeword is transmitted ($[c_{1,t}^\pi, c_{2,t}^\pi] = [0, 0]$) and no interleaving is performed, x_t is completely determined by \mathbf{s}_t , i.e., $x_t = \Phi(\tilde{c}_t^\pi) = \Phi(\mathbf{s}_t)$.

The L-value in (29) is a *piece-wise linear* function of y_t . Moreover, the L-values $\Lambda(e_t, \mathbf{s}_t)$ in (22) are linear combinations of $\tilde{l}_{k,t}^\pi(y_t | \mathbf{s}_t)$ in (29), and therefore, they are also piece-wise linear functions of y_t . Two cases are of particular interest, namely, when $e_t = [1, 0]$ or $e_t = [0, 1]$, and when $e_t = [1, 1]$. The piece-wise linear relationships for the first case are shown in Fig. 3 a). In this figure we also show the constellation symbols (or equivalently, \mathbf{s}_t) with their binary labelings and we use the notation $\mathbf{s}_t = [0/1, :]$ and $\mathbf{s}_t = [:, 0/1]$ to show that for $e_t = [1, 0]$ and $e_t = [0, 1]$ the L-values $\Lambda(e_t, \mathbf{s}_t)$ are independent of $s_{2,t}$ and $s_{1,t}$, respectively. In Fig. 3 b), the four possible cases when $e_t = [1, 1]$ are shown.

For a given scrambler outcome \mathbf{s}_t (or equivalently, a given transmitted symbol x_t), the received signal y_t is a Gaussian random variable with mean x_t and variance $N_0/2$. Therefore, each L-value $\Lambda(e_t, \mathbf{s}_t)$ in (22) is a sum of piece-wise Gaussian functions.⁴ 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 [21, 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

$$\Lambda(y_t | e_t, \mathbf{s}_t) \approx \hat{a}(e_t, \mathbf{s}_t) y_t + \hat{b}(e_t, \mathbf{s}_t), \quad (30)$$

where $\hat{a}(e_t, \mathbf{s}_t)$ and $\hat{b}(e_t, \mathbf{s}_t)$ 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 $\hat{a}(e_t, \mathbf{s}_t)$ and $\hat{b}(e_t, \mathbf{s}_t)$ defining (30) for 4-PAM, where for notation simplicity we have defined

$$\alpha \triangleq 4\gamma\Delta^2. \quad (31)$$

To clarify how these coefficients are obtained, consider for example $e_t = [0, 1]$. In this case, for $\mathbf{s}_t = [1, 1]$, 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. 3 by $e_t = [0, 1]$ and $\mathbf{s}_t = [:, 1]$ (dashed-dotted line). This gives $\hat{a}([0, 1], [1, 1]) = +\alpha/\Delta$ (slope) and $\hat{b}([0, 1], [1, 1]) = +2\alpha$ (intercept), as shown in Table I. If for example $e_t = [0, 1]$ and

³The max-log metric in (29) is suboptimal in terms of BER, however, it is very popular in practical implementations because of its low complexity. Moreover, when low order constellations are used, the use of this simplification results in a negligible impact on the receiver's performance [20, Fig. 9]. The impact of this approximation, however, becomes more important when higher order constellations are considered, as shown in [20, Fig. 9].

⁴Closed-form expressions for the pdfs of $\Lambda(e_t, \mathbf{s}_t)$ when $e_t = [1, 0]$ and $e_t = [0, 1]$ (cf. Fig. 3 a)) were presented in [21].

$\mathbf{s}_t = [0, 0]$ ($x_t = \Delta$), the closest linear piece is the right-most piece labeled by $e_t = [0, 1]$ and $\mathbf{s}_t = [:, 0]$ (dashed line), which gives $\hat{a}([0, 1], [0, 0]) = +\alpha/\Delta$ and $\hat{b}([0, 1], [0, 0]) = -2\alpha$. All the other values in Table I can be found by a similar direct inspection of Fig. 3.

Using the approximation in (30), the conditional L-values (conditioned on \mathbf{s}_t) can be modeled as Gaussian random variables where their mean and variance depend on \mathbf{s}_t , γ , and e_t , i.e.,

$$p_{\Lambda}(\lambda | \mathbf{S}_t = \mathbf{s}_t) = \psi(\lambda; \hat{\mu}(e_t, \mathbf{s}_t), \hat{\sigma}^2(e_t, \mathbf{s}_t)), \quad (32)$$

where the mean value and variance are given by

$$\hat{\mu}(\mathbf{s}_t) = x_t \hat{a}(e_t, \mathbf{s}_t) + \hat{b}(e_t, \mathbf{s}_t) \quad (33)$$

$$\hat{\sigma}^2(e_t, \mathbf{s}_t) = [\hat{a}(e_t, \mathbf{s}_t)]^2 \frac{N_0}{2}. \quad (34)$$

In Table II we show the mean values and variances in (33)–(34) for the same cases presented in Table I.

To obtain the (unconditional) pdf of Λ in (22), we average (32) over the scrambling outcomes \mathbf{s}_t (cf. Table II), which are assumed to be equiprobable. This results in the following expression

$$p_{\Lambda}(\lambda) = \begin{cases} \frac{1}{2} [\psi(\lambda; -3\alpha, 2\alpha) + \psi(\lambda; -\alpha, 2\alpha)], & \text{if } e_t = [1, 0] \\ \psi(\lambda; -\alpha, 2\alpha), & \text{if } e_t = [0, 1] \\ \psi(\lambda; -4\alpha, 8\alpha), & \text{if } e_t = [1, 1] \end{cases}. \quad (35)$$

IV. DISCUSSION AND APPLICATIONS

In the previous section, we developed approximations for the pdf of the L-values passed to the decoder in BICM-T, cf. (35). In this section we use them to quantify the gains offered by BICM-T over BICM-S, to define asymptotically optimum CCs, and to compare BICM-T with Ungerboeck's 1D-TCM.

A. Performance of BICM-T

Theorem 1: The UB for BICM-T is

$$\text{UB}_T = \sum_{w_1, w_2, w_\Sigma} \beta_{w_1, w_2, w_\Sigma}^C \left(\frac{1}{2}\right)^{w_1} \sum_{j=0}^{w_1} \binom{w_1}{j} \cdot Q\left(\sqrt{\frac{(w_1 + w_2 + 4w_\Sigma + 2j)^2 2\gamma}{(w_1 + w_2 + 4w_\Sigma) 5}}\right). \quad (36)$$

Proof: Using the pdf of the L-values in (35), the elements in the integral defining the PEP in (23) can be expressed as

$$\{p_{\Lambda_1}(\lambda | B_1 = 0)\}^{*w_1} = \left(\frac{1}{2}\right)^{w_1} \sum_{j=0}^{w_1} \binom{w_1}{j} \cdot \psi(\lambda; -\alpha(2j + w_1), 2\alpha w_1) \quad (37)$$

$$\{p_{\Lambda_2}(\lambda | B_2 = 0)\}^{*w_2} = \psi(\lambda; -\alpha w_2, 2\alpha w_2) \quad (38)$$

$$\{p_{\Lambda_\Sigma}(\lambda | \mathbf{B} = [0, 0])\}^{*w_\Sigma} = \psi(\lambda; -4\alpha w_\Sigma, 8\alpha w_\Sigma), \quad (39)$$

where we used $\psi(\lambda; \mu_1, \sigma_1^2) * \dots * \psi(\lambda; \mu_J, \sigma_J^2) = \psi(\lambda; \sum_{j=1}^J \mu_j, \sum_{j=1}^J \sigma_j^2)$ and where the binomial coefficient in (37) comes from the sum of the two Gaussian functions in

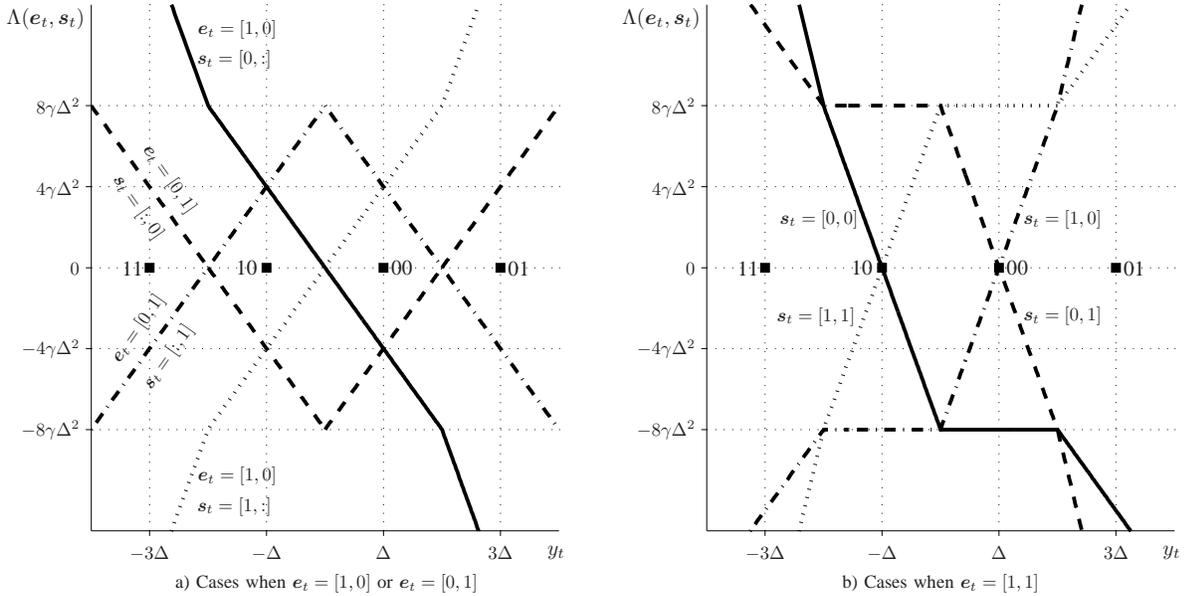


Fig. 3. Piece-wise relation between the L-values $\Lambda(e_t, s_t)$ in (22) and the received signal y_t for 4-PAM for all the possible values of e_t and s_t . The relation for the case when $e_t = [1, 0]$ or $e_t = [0, 1]$ is shown in a) and the relation when $e_t = [1, 1]$ is shown in b). The transmitted symbols and their corresponding binary labelings are shown with black squares. The notation $s_t = [0/1, :]$ is used to show that for $e_t = [1, 0]$ the L-values $\Lambda(e_t, s_t)$ are independent of $s_{2,t}$.

TABLE I
VALUES OF $\hat{a}(e_t, s_t)$ AND $\hat{b}(e_t, s_t)$ IN (30) FOR 4-PAM FOUND BY DIRECT INSPECTION OF FIG. 3.

	$s_t = [1, 1]$		$s_t = [1, 0]$		$s_t = [0, 0]$		$s_t = [0, 1]$	
	$\hat{a}(e_t, s_t)$	$\hat{b}(e_t, s_t)$						
$e_t = [1, 0]$	$+\alpha/\Delta$	0	$+\alpha/\Delta$	0	$-\alpha/\Delta$	0	$-\alpha/\Delta$	0
$e_t = [0, 1]$	$+\alpha/\Delta$	$+2\alpha$	$-\alpha/\Delta$	-2α	$+\alpha/\Delta$	-2α	$-\alpha/\Delta$	$+2\alpha$
$e_t = [1, 1]$	$+2\alpha/\Delta$	$+2\alpha$	$+2\alpha/\Delta$	-2α	$-2\alpha/\Delta$	-2α	$-2\alpha/\Delta$	$+2\alpha$

TABLE II
VALUES OF $\hat{\mu}(e_t, s_t)$ AND $\hat{\sigma}^2(e_t, s_t)$ GIVEN IN (33) AND (34) FOR 4-PAM.

	$s_t = [1, 1]$		$s_t = [1, 0]$		$s_t = [0, 0]$		$s_t = [0, 1]$	
	$\hat{\mu}(e_t, s_t)$	$\hat{\sigma}^2(e_t, s_t)$						
$e_t = [1, 0]$	-3α	2α	$-\alpha$	2α	$-\alpha$	2α	-3α	2α
$e_t = [0, 1]$	$-\alpha$	2α	$-\alpha$	2α	$-\alpha$	2α	$-\alpha$	2α
$e_t = [1, 1]$	-4α	8α	-4α	8α	-4α	8α	-4α	8α

(35) when $e_t = [1, 0]$. Using the relations (37)–(39) in (23) gives

$$\text{PEP}_T(w_1, w_2, w_\Sigma) = \int_0^\infty \left(\frac{1}{2}\right)^{w_1} \sum_{j=0}^{w_1} \binom{w_1}{j} \cdot \psi(\lambda; \mu_{1,2,\Sigma,j}, \sigma_{1,2,\Sigma}^2) d\lambda, \quad (40)$$

where

$$\mu_{1,2,\Sigma,j} = -(w_1 + w_2 + 4w_\Sigma + 2j)\alpha \quad (41)$$

$$\sigma_{1,2,\Sigma}^2 = 2(w_1 + w_2 + 4w_\Sigma)\alpha. \quad (42)$$

By using the definition of α in (31) and Δ in (1), and (41) and (42) in (40), as well as the UB expression in (24), the

expression (36) is obtained. \blacksquare

In Fig. 4, numerical results for BICM-T with 4-PAM labeled with the BRGC and using the ODSCCs (5, 7) ($K = 3$) and (247, 371) ($K = 8$) [19, Table I] are shown. For BICM-M, two configurations exist for each code. The first one is when all the bits from the first encoder's output are interleaved and then assigned to the first modulator's input ($k = 1$) and all the bits from the second encoder's output are interleaved and then assigned to the second modulator's input ($k = 2$). The second alternative 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

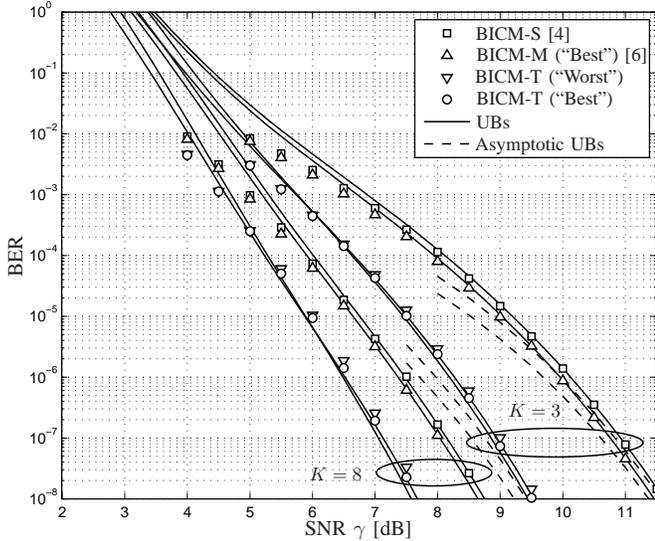


Fig. 4. BER for BICM using the (5, 7) and (247, 371) ODSCCs [19] and 4-PAM labeled with the BRGC, and for BICM-S [4], BICM-M [6], and BICM-T. The simulations are shown with markers and the UB with solid lines. The asymptotic UB is shown with dashed lines.

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) or (371) are sent over $k = 1$ and all the bits generated by the polynomial (5) or (247) are sent over $k = 2$.⁵ We denote the configuration that minimizes (or maximizes) the BER by “Best” (or “Worst”). In Fig. 4, we show BICM-T with the best and worst configurations, however, for BICM-M, we only show the best configuration.

To compute the UB for BICM-S and BICM-M, we use the expressions in [6, eq. (22)–(23)], and for BICM-T we use Theorem 1. All the UB computations were carried out considering a truncated spectrum of the code, i.e., $\{w, w_1, w_2, w_\Sigma\} \leq 30$ which is calculated numerically using a breadth-first search algorithm [22]. The results in Fig. 4 show that the UBs developed in this paper 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 $K = 8$. Moreover, these gains are obtained by decreasing the complexity of the system, i.e., by not doing interleaving/de-interleaving.

B. Asymptotic Performance

In this subsection, we analyze the performance of BICM-T for asymptotically high SNR and we compare it with BICM-S.

Theorem 2: For a given code \mathbb{C} and asymptotically high SNR, the leading term of the UB for BICM-T is

$$UB'_T = M^{\mathbb{C}} Q \left(\sqrt{\frac{2\gamma A^{\mathbb{C}}}{5}} \right), \quad (43)$$

⁵These configurations were found in [6] based on a UB for BICM-M similar to (20). It was shown in [6] that different configurations result in different BER performance because of the UEP introduced by the binary labeling.

where

$$A^{\mathbb{C}} \triangleq \min_{\substack{w_1, w_2, w_\Sigma \\ \beta_{w_1, w_2, w_\Sigma}^{\mathbb{C}} \neq 0}} (w_1 + w_2 + 4w_\Sigma) \quad (44)$$

$$M^{\mathbb{C}} = \sum_{\substack{w_1, w_2, w_\Sigma \\ w_1 + w_2 + 4w_\Sigma = A^{\mathbb{C}}}} \left(\frac{1}{2} \right)^{w_1} \beta_{w_1, w_2, w_\Sigma}^{\mathbb{C}}. \quad (45)$$

Proof: The UB in (36) is a sum of weighted Q-functions, and thus, for high SNR, only the Q-functions with the smallest argument should be considered. For each (w_1, w_2, w_Σ) there is a Q-function that dominates the inner sum in (36), which is obtained for $j = 0$. ■

For comparison purposes, we present here the performance of BICM-S at asymptotically high SNR. This can be obtained from [4, eqs. (55) and (64)] (or alternatively from [6, eq. (25)]), which gives

$$UB'_S = \left(\frac{3}{4} \right)^{d_H^{\text{free}}} \beta_{d_H^{\text{free}}}^{\mathbb{C}} Q \left(\sqrt{\frac{2d_H^{\text{free}}\gamma}{5}} \right), \quad (46)$$

where d_H^{free} is the free Hamming distance of the code which can be expressed as $d_H^{\text{free}} = w_1^{\text{free}} + w_2^{\text{free}} + 2w_\Sigma^{\text{free}}$, cf. (25). The expression (46) shows that the ODSCCs are the optimal choice for BICM-S because they have maximum free Hamming distance, cf. [19, Sec. I]. More details are given in Sec. IV-C.

In Fig. 4, we show asymptotic UBs for $K = 3$. For BICM-T we used Theorem 2, for BICM-S we use (46), and for BICM-M we use [6, eq. (25)]. All of them are shown to follow the simulation results quite well. Similar results can be obtained for the code with $K = 8$, however, we do not show these results not to overcrowd the figure.

The asymptotic gain (AG) provided by using BICM-T instead of BICM-S is obtained directly from Theorem 2 and (46), as stated in the following corollary.

Corollary 3: The AG provided by BICM-T with a given code \mathbb{C} with respect to BICM-S using an ODSCC is

$$AG_{S \rightarrow T} = 10 \log_{10} \left(\frac{A^{\mathbb{C}}}{d_H^{\text{free}}} \right), \quad (47)$$

where d_H^{free} is the free Hamming distance of the ODSCC.

Example 2 (AG for the (5, 7) code): For the particular code (5, 7), it is possible to see that the solution of (44) corresponds to the event at free Hamming distance⁶, i.e., $d_H^{\text{free}} = 5$, $w_{e,1} = 0$, $w_{e,2} = 1$, $w_{e,\Sigma} = 2$ (cf. Example 1), and therefore, $A^{\mathbb{C}} = 9$. This result in an AG of $10 \log_{10} \left(\frac{9}{5} \right) \approx 2.55$ dB, cf. (47). Moreover, since the input sequence that generates the codeword at free Hamming distance has Hamming weight one ($\beta_{0,1,2}^{\mathbb{C}} = 1$), we obtain $M_{0,1,2}^{\mathbb{C}} = 1$ for the configuration “Worst”. If the polynomials are swapped (which corresponds to swapping the rows of e), i.e., if we consider the code (7, 5), we obtain $w_{e,1} = 1$, $w_{e,2} = 0$, $w_{e,\Sigma} = 2$ and the same $A^{\mathbb{C}}$ (since $A^{\mathbb{C}}$ does not depend on the order of the polynomials). However, in this case $M_{1,0,2}^{\mathbb{C}} = 1/2$. These two asymptotic bounds are shown in Fig. 4, where the influence of the coefficient $M^{\mathbb{C}}$

⁶However, this is not always true for other codes.

($M_{0,1,2}^{\mathbb{C}} = 1$ for (5, 7) and $M_{1,0,2}^{\mathbb{C}} = 1/2$ for (7, 5)) can be observed as a weighting coefficient applied to the Q-function in (43) (this is also visible in the simulation results).

Corollary 4: The AG provided by BICM-T with respect to BICM-S is bounded as

$$\text{AG}_{\text{S} \rightarrow \text{T}} \leq 3 \text{ dB}. \quad (48)$$

Proof: For any constraint length K , and any code \mathbb{C} , we have

$$A^{\mathbb{C}} = \min_{\substack{w_1, w_2, w_{\Sigma} \\ \beta_{w_1, w_2, w_{\Sigma}}^{\mathbb{C}} \neq 0}} (w_1 + w_2 + 4w_{\Sigma}) \leq w_1^{\text{free}} + w_2^{\text{free}} + 4w_{\Sigma}^{\text{free}} \quad (49)$$

$$\leq 2w_1^{\text{free}} + 2w_2^{\text{free}} + 4w_{\Sigma}^{\text{free}}, \quad (50)$$

where we use $[w_1^{\text{free}}, w_2^{\text{free}}, w_{\Sigma}^{\text{free}}]$ to denote the generalized weight of the event(s) at free Hamming distance. The inequality in (49) holds because the event(s) at free Hamming distance belong to the elements in the minimization. The inequality in (50) holds because $w_1^{\text{free}} + w_2^{\text{free}} \geq 0$. Recognizing $2w_1^{\text{free}} + 2w_2^{\text{free}} + 4w_{\Sigma}^{\text{free}}$ in (50) as twice the free Hamming distance of the code \mathbb{C} , cf. (25), we obtain $2w_1^{\text{free}} + 2w_2^{\text{free}} + 4w_{\Sigma}^{\text{free}} < 2\hat{d}_{\text{H}}^{\text{free}}$, where $\hat{d}_{\text{H}}^{\text{free}}$ is the free Hamming distance of the ODSCC. Thus, $A^{\mathbb{C}} \leq 2\hat{d}_{\text{H}}^{\text{free}}$, which combined with (47) completes the proof. ■

C. Asymptotically Optimum Convolutional Codes

Optimum CCs—in the sense of minimizing the BER for asymptotically high SNR—are usually defined in terms of free Hamming distance, i.e., good CCs are those, which for a given rate and constraint length, have the maximum free distance (MFD) [23, Sec. 8.2.5]. The MFD criterion can be refined if the multiplicities associated to the different weights are considered [19], [13, Sec. 12.3]. Based on this refined optimality criterion, the ODSCCs were defined, which are optimal in both binary transmission and in BICM-S, cf. (46). For BICM-M, we have shown in [6] that $\hat{d}_{\text{H}}^{\text{free}}$ is still a good indicator of the optimality of the code. If BICM-T is considered, and as a direct consequence of Theorem 2, asymptotically optimum convolutional codes (AOCCs) can be defined.

Definition 1 (Asymptotically optimum CCs for BICM-T):

For a given K and R , a CC is said to be an AOCC if it has the lowest multiplicity $M^{\mathbb{C}}$ among the codes with the highest $A^{\mathbb{C}}$.

We have performed an exhaustive numerical search for AOCCs based on Definition 1, which from now on we denote by \mathbb{C}^* . We considered for constraint lengths $K = 3, 4, \dots, 8$ and all codes with free distance $0 < \hat{d}_{\text{H}}^{\text{free}} \leq \hat{d}_{\text{H}}^{\text{free}}$. The spectrum was truncated as $w_1 + w_2 + 4w_{\Sigma} \leq \hat{d}_{\text{H}}^{\text{free}} + 8$ and the search was performed in lexicographic order. The results are shown in Table III, where for comparison we also include the ODSCCs (with their respective coefficients $A^{\mathbb{C}}$ and $M^{\mathbb{C}}$). If there is more than one AOCC for a given K , we present the first one in the list. These results show that in general the free Hamming distance of the code is not the proper criterion in BICM-T, i.e., codes that are not MFD codes perform better

than the ODSCCs, cf. $K = 6, 7$. In fact, only for $K = 3$ the ODSCC is also optimum for BICM-T⁷. The results in Table III also show that the gains offered by the AOCCs in four out of six cases come from a reduced multiplicity $M^{\mathbb{C}}$, i.e., both AOCCs and ODSCCs have the same $A^{\mathbb{C}}$. In these cases, BICM-T with an AOCC and BICM-T with an ODSCC will give a similar BER performance.

In Table III, we also present the AG that BICM-T with \mathbb{C}^* offers with respect to BICM-S with an ODSCC, cf. (47). The values obtained are around 2 dB, which are consistent with Corollary 4.

D. BICM-T vs. TCM

As mentioned in Sec. II-B, the transmitter of BICM-T is identical to the transmitter of Ungerboeck's 1D-TCM and to pragmatic TCM. In this subsection we compare the asymptotic performance of BICM-T and Ungerboeck's 1D-TCM.

We have previously defined in (47) the AG of BICM-T over BICM-S. It is also possible to define the AG of BICM-T compared to uncoded transmission with the same spectral efficiency (uncoded 2-PAM). Since the minimum squared Euclidean distance of the 2-PAM constellation (normalized to $E_s = 1$) is 4, the AG is given by

$$\text{AG}_{\text{UC} \rightarrow \text{T}} = 10 \log_{10} \left(\frac{A^{\mathbb{C}}}{5} \right). \quad (51)$$

The AG in (51) is shown in Table III. For $K = 3$, $\text{AG}_{\text{UC} \rightarrow \text{T}}$ is equal to 2.55 dB, which is the same as $\text{AG}_{\text{S} \rightarrow \text{T}}$. This is because BICM-S with $K = 3$ does not offer any AG compared to uncoded 2-PAM. Analyzing the values of $\text{AG}_{\text{UC} \rightarrow \text{T}}$ in Table III, we find that they are the same as those obtained by 1D-TCM, cf. [1, Table I]. This states that if BICM-T is used with the correct CC, it performs asymptotically as well as 1D-TCM, and therefore, it should be considered as good alternative for CM in nonfading channels (however, this is not the case if BICM-S is used, or if BICM-T is used with the ODSCCs). While this asymptotic equivalence between BICM-T and 1D-TCM is very interesting, a systematic comparison of their performance when changing the constellation size coding rate or binary labeling requires a deeper analysis.

For completeness, in Table III, we also show the AG offered by BICM-S over uncoded transmission, defined as $\text{AG}_{\text{UC} \rightarrow \text{S}} = \text{AG}_{\text{UC} \rightarrow \text{T}} - \text{AG}_{\text{S} \rightarrow \text{T}}$. The performance of uncoded transmission does not depend on K , but on the other hand, the performance of BICM-T and BICM-S in general increases when K increases. This is reflected in an increase of the AGs $\text{AG}_{\text{UC} \rightarrow \text{T}}$ and $\text{AG}_{\text{UC} \rightarrow \text{S}}$ in Table III when K increases.⁸ Finally, we also note that for $K = 7$, the use of BICM-S with the ODS results in an asymptotic gain of 3.01 dB compared to uncoded transmission (cf. $\text{AG}_{\text{UC} \rightarrow \text{S}}$ in Table III). Moreover, for the same code, $\text{AG}_{\text{UC} \rightarrow \text{T}}$ is the same as the one obtained by pragmatic TCM [15, Table II].

⁷For $K = 4$ the AOCC (13, 17) has, in fact, the same spectrum $\beta_{w_1, w_2, w_{\Sigma}}^{\mathbb{C}}$ as the ODSCC (15, 17). The AOCC appears in the list because of the lexicographic order search.

⁸The fact that $\text{AG}_{\text{UC} \rightarrow \text{S}}$ is the same for $K = 7$ and $K = 8$ can be explained from the fact that the corresponding ODSCCs have the same $\hat{d}_{\text{H}}^{\text{free}} = 10$.

TABLE III
COEFFICIENTS A^C AND M^C FOR AOCCs AND ODSCCs IN BICM-T. THE ASYMPTOTIC GAINS AND THE HAMMING FREE DISTANCES ARE ALSO SHOWN.

K	AOCCs				ODSCCs				AG [dB]		
	(g_1, g_2)	d_H^{free}	A^{C^*}	M^{C^*}	(g_1, g_2)	d_H^{free}	A^C	M^C	$AG_{S \rightarrow T}$	$AG_{UC \rightarrow T}$	$AG_{UC \rightarrow S}$
3	(7, 5)	5	9	0.50	(7, 5)	5	9	0.50	2.55	2.55	0
4	(13, 17)	6	10	0.50	(15, 17)	6	10	0.50	2.22	3.01	0.79
5	(23, 33)	7	11	0.38	(23, 35)	7	11	0.88	1.96	3.42	1.46
6	(45, 55)	7	13	1.62	(53, 75)	8	12	0.50	2.11	4.15	2.04
7	(107, 135)	9	14	0.50	(133, 171)	10	14	3.09	1.46	4.47	3.01
8	(313, 235)	10	16	8.02	(371, 247)	10	15	0.61	2.04	5.05	3.01

V. CONCLUSIONS

In this paper, we formally explained why the recently proposed BICM-T system offers gains over regular BICM 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 CCs for BICM-T. The analytical model was used to validate the numerical results and to show that the use of the ODSCCs, which rely on the regular free Hamming distance criterion, is suboptimal.

The model presented in this paper was also used to analyze the asymptotic behavior of BICM-T. Optimal convolutional codes for BICM-T were tabulated and it was shown that a properly designed BICM-T system performs asymptotically as well as Ungerboeck's TCM. Moreover, it was shown that the asymptotic loss caused by using BICM-S instead of BICM-T in nonfading channels is never larger than 3 dB.

For notation simplicity and 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 ($R = 1/2$ and 4-PAM). A more general analysis is possible, and very interesting indeed. For example, it is still unknown what the performance gains will be in a more general setup, e.g., for different code rates, when the number of encoder outputs is not the same as the number of modulator inputs, or for different spectral efficiencies. Also, it is unclear what the gains offered by the AOCCs for finite SNR values are. All these questions, as well as a general comparison between BICM-T and pragmatic TCM, are left for further investigation.

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. Martínez, 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. 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.

- [7] 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.
- [8] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Prentice Hall, 1995.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] S. H. Jamali and T. Le-Ngoc, *Coded-Modulation Techniques for Fading Channels*. Kluwer Academic Publishers, 1994.
- [13] S. Lin and D. J. Costello, Jr., *Error Control Coding*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice Hall, 2004.
- [14] G. C. Clark Jr. and J. B. Cain, *Error-correction coding for digital communications*, 2nd ed. Plenum Press, 1981.
- [15] 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.
- [16] J. K. Wolf and E. Zehavi, " p^2 codes: Pragmatic trellis codes utilizing punctured convolutional codes," *IEEE Commun. Mag.*, vol. 33, no. 2, pp. 94–99, Feb. 1995.
- [17] R. H. Morelos-Zaragoza, *The Art of Error Correcting Coding*, 2nd ed. John Wiley & Sons, 2002.
- [18] A. Martínez, 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.
- [19] 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.
- [20] B. Classon, K. Blankenship, and V. Desai, "Channel coding for 4G systems with adaptive modulation and coding," *IEEE Wireless Commun. Mag.*, vol. 9, no. 2, pp. 8–13, Apr. 2002.
- [21] 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.
- [22] J. Belzile and D. Haccoun, "Bidirectional breadth-first algorithms for the decoding of convolutional codes," *IEEE Trans. Commun.*, vol. 41, no. 2, pp. 370–380, Feb. 1993.
- [23] J. G. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, 2000.

Alex Alvarado (S'06) was born in 1982 in Quellón, on the island of Chiloé, Chile. He received his Electronics Engineer degree (Ingeniero Civil Electrónico) and his M.Sc. degree (Magíster en Ciencias de la Ingeniería Electrónica) from Universidad Técnica Federico Santa María, Valparaíso, Chile, in 2003 and 2005, respectively. He obtained the degree of Licentiate of Engineering (Teknologie Licentiatexamen) in 2008 and his PhD degree in 2011, both of them from Chalmers University of Technology, Göteborg, Sweden.

Since 2006, he has been involved in a research collaboration with the Institut National de la Recherche Scientifique (INRS), Montreal, QC, Canada, on the analysis and design of BICM systems. In 2008, he was holder of the Merit Scholarship Program for Foreign Students, granted by the Ministère de l'Éducation, du Loisir et du Sports du Québec. His research interests are in the areas of digital communications, coding, and information theory.

A. Alvarado is currently a Newton International Fellow at the University of Cambridge, United Kingdom, funded by The British Academy and The Royal Society.

Leszek Szczecinski (M'98-SM'07), obtained M.Eng. degree from the Technical University of Warsaw in 1992, and Ph.D. from INRS-Telecommunications, Montreal in 1997. From 1998 to 2001, he held position of Assistant Professor at the Department of Electrical Engineering, University of Chile. He is now Associate Professor at INRS-EMT, University of Quebec, Canada, and Adjunct Professor at Electrical and Computer Engineering Department of McGill University. His research interests are in the area of modulation and coding, communication theory, wireless communications, and digital signal processing.

Erik Agrell received the M.S. degree in electrical engineering in 1989 and the Ph.D. degree in information theory in 1997, both from Chalmers University of Technology, Sweden.

From 1988 to 1990, he was with Volvo Technical Development as a Systems Analyst, and from 1990 to 1997, with the Department of Information Theory, Chalmers University of Technology, as a Research Assistant. In 1997–1999, he was a Postdoctoral Researcher with the University of Illinois at Urbana-Champaign and the University of California, San Diego. In 1999, he joined the faculty of Chalmers University of Technology, first as an Associate Professor and since 2009 as a Professor in Communication Systems. His research interests belong to the fields of information theory, coding theory, and digital communications, and his favorite applications are found in optical communications. More specifically, his current research includes bit-interleaved coded modulation and multilevel coding, polarization-multiplexed coding and modulation, optical intensity modulation, bit-to-symbol mappings in coded and uncoded systems, lattice theory and sphere decoding, and multidimensional geometry.

Prof. Agrell served as Publications Editor for IEEE Transactions on Information Theory from 1999 to 2002.