Unequal Error Protection in BICM with QAM Constellations: Interleaver and Code Design

Alex Alvarado, Erik Agrell, Leszek Szczecinski[§] and Arne Svensson
Department of Signals and Systems, Communication Systems Group
Chalmers University of Technology, Gothenburg, Sweden
[§]INRS-EMT, University of Quebec, Montreal, Canada
{*alex.alvarado,agrell,arnes*}*@chalmers.se*, *leszek@emt.inrs.ca*

***Abstract*—In this paper we present a general methodology for the interleaver and code design for QAM-based BICM transmissions. We develop analytical bounds on the bit error rate and we use them to predict the performance of BICM when unequal error protection (UEP) is introduced by the constellation labeling. Based on these bounds, the optimum design of interleaver and code is presented. The improvements obtained reached 2 dB for the analyzed cases, and are obtained without complexity increase. Although previous works noted the influence of the interleaver design and the UEP, to the best of our knowledge, this paper is the first to analyze formally this problem for BICM transmissions.**

I. Introduction

Bit-interleaved coded modulation (BICM) was first introduced by Zehavi in [1], and later analyzed from an information theory point of view in the landmark paper of Caire *et al.* [2]. BICM owes its popularity to the fact that the channel encoder and the modulator are separated by a bit-level interleaver, which may be chosen independently allowing for a simple and flexible design [2, Sec. V]. BICM is a de facto standard, and it is used in most of the existing wireless systems, e.g., EDGE, HSDPA, IEEE 802.11a/g, etc.

When a systematic encoder is used, the coded bits can be classified into two groups: systematic and parity bits. When BICM is used with Gray-mapped quaternary phase shift keying both systematic and parity bits are treated equally by the modulator. If BICM is used with high-order constellations, the bit mapping causes the so-called unequal error protection (UEP) [2], [3], i.e., depending on the bits position within the symbol, the bits experience different "protection", which may be interpreted in terms of uncoded error probability or average mutual information. The obvious question is then how to protect the different bits using the mapping which inherently causes UEP, or how to select/design the code and the interleaver to take this effect into account.

Following the framework set in [2], (pseudo)random (RN) interleavers are most often applied in BICM. This simplifies the analysis of the resulting system, but leads to sub-optimality already noted in the literature [4]. For example, the original BICM paper of Zehavi [1] postulated the application of independent interleavers between each of the encoder's output and the corresponding modulator's input. Similar modular (MD) interleavers have been used for example in [4]–[7] in different contexts. MD interleavers have also been proposed in the 3GPP/HSDPA standard [8], [9] when 16-QAM is used. They are relevant from an implementation point of view since for example, one of the reasons for using 2 parallel interleavers in HSDPA with 16-QAM, is the fact that the already implemented interleaver for 4-QAM can be "re-used" [8, Sec. 4.5.6]. When such MD interleavers are used, the performance gains will strongly depend on the bit assignment between the encoder's output and the bit positions in the complex symbol.

In this paper we show that there is no unique answer about the protection of systematic/parity bits, and we tackle this problem by analyzing a general type of probabilistic interleaver, of which RN and MD are two special cases. We develop union bounds for the coded bit error rate (BER) of the system, and using these bounds, the optimum design of interleaver and code is presented, and an exhaustive computer search for the optimum interleaver design and code selection is performed. We also introduce the concept of generalized optimum distance spectrum (GODS) convolutional codes, which are the optimum codes for this scenario.

II. System Model

We consider the BICM system shown in Fig. 1. The vector of N information bits $\mathbf{b} = [b^1, \dots, b^N]$ is encoded by a rate $R = 1/n$ channel encoder where the superscripts have a meaning of discrete time. The vectors of coded bits $\mathbf{c}_1, \dots, \mathbf{c}_n$ are then fed to the interleaver units where the p-th output vector of the encoder is given by $\mathbf{c}_p = [c_p^1, \dots, c_p^N]$.

The interleavers $(\pi_1, \dots, \pi_n)$ in Fig. 1 are assumed to be ideal and independent interleavers, yielding randomly permuted sequences $\mathbf{c}'_p = \pi_p\{\mathbf{c}_p\}$ of the coded bits. The multiplexing unit (MUX) assigns the coded and interleaved bits to the different bit positions in the M^2-QAM symbol. The mapping considered here is based on the so-called binary reflected Gray code (BRGC)[1] [10], so each symbol is a superposition of independently modulated real/imaginary parts [11, Sec. X]. Consequently, we focus on the equivalent M-PAM constellation (cf. Fig. 1) where $M = 2^m$.

Research supported by the Swedish Research Council, Sweden (under research grant #2006-5599), and by NSERC, Canada (under research grant #249704-07).

[1]The selection of the BRGC for our analysis is based on its relevance in practical systems and its optimality in terms of uncoded BER [10].

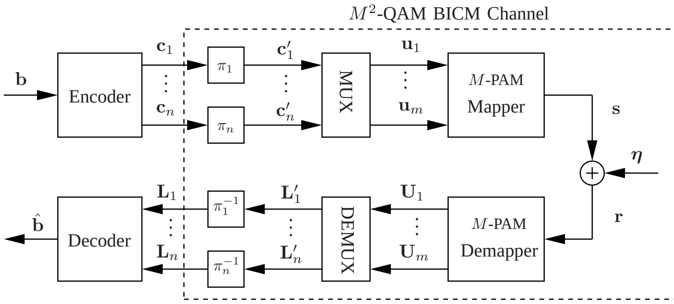

Figure 1. Model of BICM-QAM transmission.

Table I
UEP CAUSED BY THE BRGC: MODULATING CODEWORDS, 8-PAM SYMBOLS, DISTANCES $d_q(s)$, AND VIRTUAL CHANNELS Θ_j.

$[u_1,\ldots,u_m]$	000	001	011	010	110	111	101	100
s	-7Δ	-5Δ	-3Δ	$-\Delta$	Δ	3Δ	5Δ	7Δ
$d_1(s)$	8Δ	6Δ	4Δ	2Δ	2Δ	4Δ	6Δ	8Δ
Θ_j	Θ_4	Θ_3	Θ_2	Θ_1	Θ_1	Θ_2	Θ_3	Θ_4
$d_2(s)$	4Δ	2Δ	2Δ	4Δ	4Δ	2Δ	2Δ	4Δ
Θ_j	Θ_2	Θ_1	Θ_1	Θ_2	Θ_2	Θ_1	Θ_1	Θ_2
$d_3(s)$	2Δ	2Δ	2Δ	2Δ	2Δ	2Δ	2Δ	2Δ
Θ_j	Θ_1	Θ_1	Θ_1	Θ_1	Θ_1	Θ_1	Θ_1	Θ_1

For a fully general approach, we define the multiplexing unit using a matrix $\mathbf{K}_{n\times m} \equiv \mathbf{K}$ of dimensions $n \times m$, whose elements, $0 \leq \kappa_{p,q} \leq 1$, denote the fraction of bits $\mathbf{c}_p$ that will be assigned to the q-th output $\mathbf{u}_q$. As all the vectors $\mathbf{u}_q$ for $q = 1,\ldots,m$ have the same length, the constraint $\sum_{p=1}^{n} \kappa_{p,q} = \frac{n}{m}$ must be satisfied, and since all the bits in the vector $\mathbf{c}'_p$ must be assigned to one of the m outputs, the constraint $\sum_{q=1}^{m} \kappa_{p,q} = 1$ must also be satisfied.

The interleavers and multiplexing unit in Fig. 1 allow us to consider any interleaver configuration. For example, for $n = m$, if $\mathbf{K} = \mathbf{I}_n$ ($\mathbf{I}_n$ being the identity matrix), the system is transformed into the BICM with modular interleaver (BICM-MD) where all the bits from the same encoder's output are assigned to the same bit position. Exchanging the rows of this matrix allows us to consider all possible MD interleavers. On the other hand, if $\kappa_{p,q} = \frac{1}{m}\ \forall p, q$, the BICM with random interleaver (BICM-RN) of [2] is obtained. Obviously, there are many other combinations, specially for a non-square matrix $\mathbf{K}$. We will refer to "interleaver design" as the process of selecting the elements $\kappa_{p,q}$ defining $\mathbf{K}$.

At any time instant t, the coded and interleaved bits $[u_1^t,\ldots,u_m^t]$ are mapped to a M-PAM symbol $s^t \in \mathcal{X}$ using a binary memoryless mapping $\mathcal{M} : \{0,1\}^m \rightarrow \mathcal{X}$, where $\mathcal{X} = \{(1-M)\Delta, (3-M)\Delta, \ldots, (M-1)\Delta\}$ is the set of M-PAM symbols and where 2Δ is the minimum distance between them. The constellation is normalized to unit average energy so $\Delta = \sqrt{\frac{3}{2(M^2-1)}}$. The result of the transmission of N symbols is given by $\mathbf{r} = \mathbf{s}+\boldsymbol{\eta}$, where $\mathbf{s} = [s^1,\ldots,s^N]$, $\boldsymbol{\eta} \in \mathbb{R}^N$ are zero-mean and independent Gaussian variables with variance $N_0/2$. The signal-to-noise ratio (SNR) per complex symbol is given by $\gamma = \frac{1}{N_0}$.

At the receiver's side, the reliability metrics of the transmitted bits are calculated in the form of logarithmic likelihood ratios (L-values) for each bit position as [1], [2]

$$U_q^t = \gamma\left(\min_{a\in\mathcal{X}_{q,0}} \left\{(r^t-a)^2\right\} - \min_{a\in\mathcal{X}_{q,1}} \left\{(r^t-a)^2\right\}\right), \tag{1}$$

where $\mathcal{X}_{q,b}$ is the set of symbols labelled with the q-th bit equal to b. Since the mapping is memoryless, from now on we drop the time index t, e.g., $U_q^t \equiv U_q$.

The vector of soft information $\mathbf{U}_q$ is demultiplexed ($\mathbf{L}'_p$), deinterleaved ($\mathbf{L}_p$) and then passed to a channel decoder which produces an estimate of the transmitted bits $\hat{\mathbf{b}}$.

Using the results presented in [12] it is possible to build an equivalent model for the M^2-QAM BICM channel shown in Fig. 1. In this model each bit u_q after the MUX can be seen as being sent over a *virtual channel* whose output L-value U_q has a distribution that depends on the bit's position q and the symbol sent, i.e., the value of the other bits $u_v, v \neq q$. We explain it briefly below while more details may be found in [12]. Let $d_q(s)$ denote the Euclidean distance between the symbol s and the closest symbol in the constellation with the opposite value of the bit labeling s at position q, i.e., if $s \in \mathcal{X}_{q,b}$, $b \in \{0,1\}$, $d_q(s) = \min_{a\in\mathcal{X}_{q,\bar{b}}} |s-a|$. Due to the properties of the BRGC, symbols with the q-th bit set to 0 or 1 are clustered so that $d_q(s)$ may be at a distance that varies from 2Δ to $2\Delta\cdot\frac{M}{2^q}$. That is, when $q = m$, there is always an adjacent symbol (at distance 2Δ) with the opposite value of the bit. On the other hand, for $q = 1$, the number of possible distances is $M/2$. To clarify this, in Table I the distances $d_q(s)$ for the specific case of $M = 8$ are shown. Since $d_q(s)$ determines the "protection" experienced by the bit, different values of $d_q(s)$ cause UEP. For $q = m$ the bits have always the same "low" protection but for $q = 1$, depending on the value of other bits in the modulating codeword, the protection may be relatively high.

According to the model in [12], there are $M/2$ different Gaussian distributions that can be used to model the L-values. By definition, a bit transmitted at position q passes through the virtual channel Θ_j when it is sent using a symbol s such that $d_j(s) = 2\Delta\cdot j$. Then, the L-value U_q has a distribution that may be modelled as Gaussian with mean μ_j and variance σ^2 where

$$(\mu_j, \sigma^2) = (4\gamma\Delta^2(2j-1), 8\gamma\Delta^2), \tag{2}$$

with $j = 1,\ldots,M/2$. In Table I the virtual channels associated with the different symbols and bit positions are shown.

The probability that an L-value at bit position q is distributed with parameters (μ_j, σ^2) is given by

$$\omega_{q,j} = \begin{cases} \dfrac{1}{2^{m-q}} & \text{if } j = 1,\ldots,2^{m-q} \\ 0 & \text{if } j = 2^{m-q}+1,\ldots,\dfrac{M}{2} \end{cases}, \tag{3}$$

that is, the virtual channel Θ_1 can be used by the bit for all

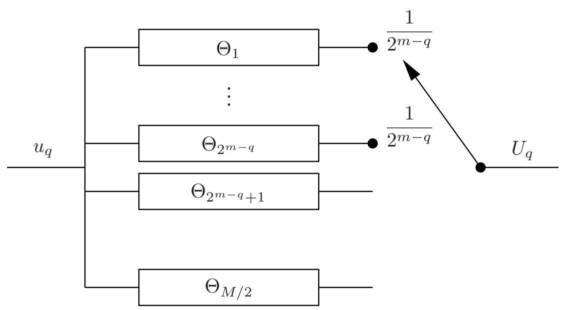

Figure 2. Equivalent channel model: the virtual channels $\Theta_j, j = 1\ldots, 2^{m-q}$ are selected with equal probability, while the channels $\Theta_j, j = 2^{m-q}+1, \ldots, M/2$ are not available for the bit at position q.

positions q, Θ_2 only for $q \leq m-1$, Θ_3 and Θ_4 only for $q \leq m-2$, $\Theta_5, \ldots, \Theta_8$ for $q \leq m-3$, and so on (cf. Table I).

To fully characterize the equivalent M^2-QAM BICM channel we define the matrix $\mathbf{O}_{m\times\frac{M}{2}} \equiv \mathbf{O}$ of dimensions $m \times M/2$ where each element $\omega_{q,j}$ in $\mathbf{O}$ is the probability that a transmitted bit at position q is transmitted using the channel Θ_j. The resulting equivalent channel model is schematically shown in Fig. 2.

Using the equivalent channel model of Fig. 2, the M^2-QAM BICM Channel of Fig. 1 can be replaced by a compound channel completely defined by the matrices $\mathbf{K}$ (interleaver) and $\mathbf{O}$ (mapping). If we define the $\mathbf{X} \triangleq \mathbf{K}\cdot\mathbf{O}$, then the p-th output $L_p \in \mathbb{R}$ of this channel is associated with the p-th binary input c_p, where L_p is a Gaussian mixture with density given by

$$\mathrm{p}_{L_p}(\lambda) = \sum_{j=1}^{M/2} \xi_{p,j} \cdot \Phi(\mu_j, \sigma^2; \lambda), \tag{4}$$

where $\Phi(\mu_j, \sigma^2; \lambda) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\lambda-\mu_j)^2}{2\sigma^2}\right)$ is a Gaussian function, and $\xi_{p,j}$ is the (p,j)-th element of $\mathbf{X}$ which denotes the probability that the p-th bit passes through Θ_j.

III. Interleaver and Code Design

A. Generalized weight distribution spectrum

For any linear code it is possible to define a *generalized transfer function* (GTF) which enumerates not only the number of non-zero output bits over a path, but the *location* of those bits, i.e., it indicates which branch the non-zero outputs are associated with [1]. For a rate-$1/n$ code we define the GTF of the code as $T(\mathbf{W}, I, L) = \sum_{\mathbf{w}} \sum_i \sum_l t_{\mathbf{w},i,l} I^i L^l \prod_{p=1}^n W_p^{w_p}$, where the *generalized weight* $\mathbf{w} = (w_1, \ldots, w_n)$ denotes the weights of each output of the encoder, and $\mathbf{W} = (W_1, \ldots, W_n)$, I, and L are dummy variables. The coefficient $t_{\mathbf{w},i,l}$ enumerates the number of paths diverging from the zero state and merging with the zero state after l steps, associated with an input sequence of weight i, and an output sequence of generalized weight $\mathbf{w}$. The coefficients $t_{\mathbf{w},i,l}$ can be calculated using standard techniques. Efficient methods for this calculation include the recursive algorithm of Divsalar *et al.* [13], or a breadth first search algorithm.

Using the GTF, it is possible to enumerate the sum of bit errors for error events of generalized weight $\mathbf{w}$ using a generalized weight distribution spectrum (GWDS) of the code. This n-dimensional GWDS takes into account not only the number of errors (conventional approach), but also the output of the encoder they are associated to. It can be calculated as

$$\beta(\mathbf{w}) = \frac{1}{\prod_{p=1}^n w_p!} \cdot \frac{\partial^w}{\partial \mathbf{W}^{\mathbf{w}}} \frac{\partial}{\partial I} T(\mathbf{W}, I, L)\Bigg|_{\mathbf{W}=\mathbf{0}, I=L=1}, \tag{5}$$

where $\frac{\partial^w}{\partial \mathbf{W}^{\mathbf{w}}} = \frac{\partial^{w_1}}{\partial W_1^{w_1}} \cdots \frac{\partial^{w_n}}{\partial W_n^{w_n}}$ and $w = w_1 + \ldots + w_n$.

To calculate the weight distribution spectrum of a turbo code (TC), the concept of *uniform interleaver* introduced by Benedetto *et al.* [14] can be used to calculate the spectrum of the code. The extension to a GWDS is straightforward; more details can be found in [13], [14].

B. Union bounds for BICM-QAM

In order to use the GWDS of the code to calculate union bounds for the coded BER, we define the set $\mathcal{Z}_i(l)$ as all the combinations of i nonnegative integers such that the sum of the elements is l, i.e., $\mathcal{Z}_i(l) \triangleq \{(z_1, \ldots, z_i) \in (\mathbb{Z}^+)^i : z_1 + \ldots + z_i = l\}$. Using the GWDS of the code, the union bound (UB) on the BER BICM is given by

$$\mathrm{BER} \leq \mathrm{UB} = \sum_{l=w_\mathrm{f}}^{\infty} \sum_{\mathbf{w}\in\mathcal{Z}_n(l)} \beta(\mathbf{w}) \cdot \mathrm{PEP}(\mathbf{w}), \tag{6}$$

where w_f is the free distance of the code, and $\mathrm{PEP}(\mathbf{w})$ is the pairwise error probability which represents the probability of detecting a codeword with generalized weight $\mathbf{w}$ instead of the transmitted all-one codeword.

To calculate the PEP we need to calculate the probability that the decoder selects a codeword with generalized weight $\mathbf{w}$ instead of the transmitted all-one codeword. To this end, we note that the decision is made based on the sum of $w_1 + \ldots + w_n$ L-values in the divergent path. Let Z be the decision variable where

$$Z = \sum_{i=1}^{w_1} L_1^{(i)} + \ldots + \sum_{i=1}^{w_n} L_n^{(i)} = \sum_{p=1}^{n} \sum_{i=1}^{w_p} L_p^{(i)}, \tag{7}$$

i.e., a sum of l independent random variables, where the random variable associated with the i-th output is a sum of i.i.d Gaussian mixtures given by (4). Consequently, for a given value of $\mathbf{w}$, the PEP can be calculated as the tail integral of the PDF of Z, i.e.,

$$\mathrm{PEP}(\mathbf{w}) = \mathrm{Prob}\{Z < 0\} = \int_{-\infty}^{0} \mathrm{p}_Z(\lambda)\, d\lambda. \tag{8}$$

To calculate $\mathrm{p}_Z(\lambda)$ we first define the j-fold self convolution operator as follows. Let L be a random variable with density $\mathrm{p}_L(\lambda)$, its j-fold self convolution is denoted by $[\mathrm{p}_L(\lambda)]^{*(j)} \triangleq \mathrm{p}_L(\lambda) * \ldots * \mathrm{p}_L(\lambda)$, which corresponds to the PDF of the sum of j i.i.d. random variables L.

Using the above notation and (4), we can calculate the PDF of the decision variable Z in (7) as

$$\mathrm{p}_Z(\lambda) = [\mathrm{p}_{L_1}(\lambda)]^{*(w_1)} * \ldots * [\mathrm{p}_{L_n}(\lambda)]^{*(w_n)}, \qquad (9)$$

where the p-th term in (9) can be approximated[2] by

$$\left[\mathrm{p}_{L_p}(\lambda)\right]^{*(w_p)} = \left[\sum_{j=1}^{\frac{M}{2}} \xi_{p,j} \Phi(\mu_j, \sigma^2; \lambda)\right]^{*(w_p)} \qquad (10)$$

$$= \sum_{j_1=1}^{\frac{M}{2}} \cdots \sum_{j_{w_p}=1}^{\frac{M}{2}} \Phi\left(\sum_{i=1}^{w_p} \mu_{j_i}, w_p\sigma^2; \lambda\right) \prod_{i=1}^{w_p} \xi_{p,j_i} \qquad (11)$$

$$= \sum_{\mathbf{r}\in\mathcal{Z}_{\frac{M}{2}}(w_p)} \binom{w_p}{\mathbf{r}} \Phi\left(\sum_{j=1}^{\frac{M}{2}} r_j\mu_j, w_p\sigma^2; \lambda\right) \prod_{j=1}^{\frac{M}{2}} \xi_{p,j}^{r_j}. \qquad (12)$$

To pass from (10) to (11) we have expanded the convolution of sums as sums of convolutions and then applied $\Phi(\mu_i, \sigma_i^2; \lambda) * \Phi(\mu_j, \sigma_j^2; \lambda) = \Phi(\mu_i + \mu_j, \sigma_i^2 + \sigma_j^2; \lambda)$. To pass from (11) to (12) we note that one Gaussian with parameters $(r_1\mu_1 + \ldots + r_{\frac{M}{2}}\mu_{\frac{M}{2}}, w_p\sigma^2)$ can be generated by different combinations of $(j_1, \ldots, j_{w_p})$. Furthermore, the number of combinations (multiplicities) for a given value of $\mathbf{r} = (r_1, \ldots, r_{\frac{M}{2}})$ are the multinomial coefficients given by $\binom{w_p}{\mathbf{r}} \triangleq \frac{w_p!}{r_1!\cdot\ldots\cdot r_{\frac{M}{2}}!}$.

Using (12) in (9) we get the final and exact expression for the density of Z shown in (14) (next page), where

$$\mathcal{W}(\mathbf{r}_1, \ldots, \mathbf{r}_n) = \prod_{p=1}^{n} \left[\binom{w_p}{\mathbf{r}_p} \prod_{j=1}^{\frac{M}{2}} \xi_{p,j}^{r_{pj}}\right]. \qquad (15)$$

Based on the previous discussion, we present one theorem and two corollaries which are the main results of this section.

Theorem 1: The union bound on the coded BER for BICM-QAM can be approximated as

$$\mathrm{UB} \approx \sum_{l=w_\mathrm{f}}^{\infty} \sum_{\mathbf{w}\in\mathcal{Z}_n(l)} \beta(\mathbf{w}) \sum_{\mathbf{r}_1,\ldots,\mathbf{r}_n} \mathcal{W}(\mathbf{r}_1, \ldots, \mathbf{r}_n) \cdot \mathrm{Q}\big(\mathcal{A}(\mathbf{r}_1, \ldots, \mathbf{r}_n)\big), \qquad (16)$$

where

$$\mathcal{A}(\mathbf{r}_1, \ldots, \mathbf{r}_n) = \frac{\sum_{p=1}^{n} \sum_{j=1}^{\frac{M}{2}} r_{pj}\mu_j}{\sqrt{l\sigma^2}}, \qquad (17)$$

$\mathcal{W}(\mathbf{r}_1, \ldots, \mathbf{r}_n)$ is given by (15), and $\mathrm{Q}(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp(-t^2/2)\,dt$ is the error function.

Proof: From (6), (8), and (14). ∎

Analyzing the expression in (16) we note that the channel properties defined by $\mathbf{O}$ are fixed for a given value of M, and that the optimum performance of the system will be achieved by a joint design of the interleaver *and* the code. We also note that all combinations in (13) are in general tedious to evaluate (especially for large values of n and/or m), thus we seek further approximations.

[2]The approximation refers to the fact that the Gaussian model for the L-values is used instead of the exact densities.

The simplification presented in the following corollary is based on considering, for each l, only the Gaussian density with the smallest mean-to-standard deviation ratio. The intuition behind this approximation is that the error coefficients generated by other Gaussian densities are less important.

Corollary 1: The UB in (16) can be approximated by

$$\mathrm{UB}' = \sum_{l=w_\mathrm{f}}^{\infty} \mathrm{Q}\left(\sqrt{2l\gamma\Delta^2}\right) \sum_{\mathbf{w}\in\mathcal{Z}_n(l)} \beta(\mathbf{w}) \prod_{p=1}^{n} \xi_{p,1}^{w_p}. \qquad (18)$$

Proof: Approximate $\mathcal{Z}_{\frac{M}{2}}(w_p)$ in the third sum of (16) by its leading element $\mathbf{r}_p = (w_p, 0, \ldots, 0)$. Then $\mathcal{W}(\mathbf{r}_1, \ldots, \mathbf{r}_n) = \prod_{p=1}^{n} \xi_{p,1}^{w_p}$ from (15) and $\mathcal{A}(\mathbf{r}_1, \ldots, \mathbf{r}_n) = \sqrt{l}\mu_1/\sigma = \sqrt{2l\gamma}\Delta$ from (17) and (2). Now (18) follows from (16). ∎

We emphasize here that (18) is quite simple to evaluate compared with the original bound (16), and it still takes into account the parameters to optimize the transmission (interleaver and code).

The following corollary presents an even simpler asymptotic bound, i.e., when the SNR goes to infinity. It confirms our initial intuition about the sub-optimality of the codes that were designed for binary transmissions. This result will provide us with the new criteria to select the optimum code and interleaver design (cf. Sec. IV-B).

Corollary 2: The asymptotic performance of BICM-QAM is given by

$$\mathrm{UB}'' \approx \mathrm{Q}\big(\sqrt{2\gamma\Delta^2 w_\mathrm{f}}\big) \sum_{\mathbf{w}\in\mathcal{Z}_n(w_\mathrm{f})} \beta(\mathbf{w}) \prod_{p=1}^{n} \xi_{p,1}^{w_p}. \qquad (19)$$

Proof: The bound (16) is a sum of weighted Q-functions, whose argument $\mathcal{A}(\mathbf{r}_1, \ldots, \mathbf{r}_n)$ depends on the number of bits that were transmitted using the different virtual channels. If $\gamma \to \infty$, only one of those Q-functions will dominate the bound, i.e., the Q-function with the smallest argument. For a given value of $\mathbf{w}$ we need to choose the combination of $(\mathbf{r}_1, \ldots, \mathbf{r}_n)$ that minimizes $\mathcal{A}(\mathbf{r}_1, \ldots, \mathbf{r}_n)$, i.e., $\min_{\mathbf{r}_1,\ldots,\mathbf{r}_n}\left\{\sum_{j=1}^{\frac{M}{2}} r_{1j}\mu_j + \ldots + \sum_{j=1}^{\frac{M}{2}} r_{nj}\mu_j\right\}$. Since $\mu_j > 0, j = 1, \ldots, \frac{M}{2}$ and $\mu_j > \mu_1, j = 2, \ldots, \frac{M}{2}$, it is clear that $\mathbf{r}_p = (w_p, 0, \ldots, 0)\ \forall p$ minimizes the previous expression.

Using the previous result and the definitions of μ_j and σ^2 in (2), it can be seen that the function $\mathcal{A}(\mathbf{r}_1, \ldots, \mathbf{r}_n)$ has a minimum value of $\sqrt{2\gamma\Delta^2 l}$. Moreover, if l is increased, the argument of the dominant Q-function will increase and consequently, the minimum is obtained when $l = w_\mathrm{f}$, i.e., when all the w_f bits were transmitted using the least protected channel Θ_1. The weighting coefficient in (19) can be obtained using the definition of $\mathbf{X}$. By combining the results presented above, (19) can be obtained. ∎

$$\mathrm{p}_Z(\lambda) = \sum_{\mathbf{r}_1\in\mathcal{Z}_{\frac{M}{2}}(w_1)} \binom{w_1}{\mathbf{r}_1} \Phi\left(\sum_{j=1}^{\frac{M}{2}} r_{1j}\mu_j, w_1\sigma^2; \lambda\right) \prod_{j=1}^{\frac{M}{2}} \xi_{1,j}^{r_{1j}} * \ldots * \sum_{\mathbf{r}_n\in\mathcal{Z}_{\frac{M}{2}}(w_n)} \binom{w_n}{\mathbf{r}_n} \Phi\left(\sum_{j=1}^{\frac{M}{2}} r_{nj}\mu_j, w_n\sigma^2; \lambda\right) \prod_{j=1}^{\frac{M}{2}} \xi_{n,j}^{r_{nj}} \quad (13)$$

$$= \sum_{\mathbf{r}_1\in\mathcal{Z}_{\frac{M}{2}}(w_1)} \cdots \sum_{\mathbf{r}_n\in\mathcal{Z}_{\frac{M}{2}}(w_n)} \mathcal{W}(\mathbf{r}_1,\ldots,\mathbf{r}_n)\cdot\Phi\left(\sum_{p=1}^{n}\sum_{j=1}^{\frac{M}{2}} r_{pj}\mu_j, \sigma^2\sum_{p=1}^{n} w_p; \lambda\right), \quad (14)$$

IV. Numerical Results

A. UB for BICM-QAM

We analyze two different configurations that give a spectral efficiency of 1 [bit/s/Hz]: a rate-1/2 TC[3] used with 16QAM ($n = 2$ and $m = 2$), and a rate-1/3 TC used with 64-QAM ($n = 3$ and $m = 3$). In general the optimization space is formed by the variables $\kappa_{p,q}$, with $p = 1,\ldots,n-1$ and $q = 1,\ldots,m-1$ constrained to $0 \leq \kappa_{pq} \leq 1$, $\forall p,q$. These are in general continuous, however, we only analyze the cases that produce MD interleavers ($\kappa_{p,q} \in \{0,1\}$) and the RN interleaver ($\kappa_{p,q} = 1/m$).

For $n = m = 2$ there is only one degree of freedom ($\kappa_{1,1}$). In Fig. 3 the bound (16) is compared with the simulation results[4] where the perfect match in the error floor region can be clearly appreciated. The best interleaver—denoted by $\mathbf{K}_\mathrm{B}$—is achieved setting $\kappa_{1,1} = 0$, i.e., when the parity bits are more protected than the systematic bits (and the worst interleaver—denoted by $\mathbf{K}_\mathrm{W}$—if $\kappa_{1,1} = 1$). This result directly contradicts [9, Sec. 9.3.2] and [3], where it is claimed that systematic bits should always be sent to the more reliable positions. Based on the developed bound we see that this is a property of the code, which is completely captured by its GWDS, and consequently, it is not possible to draw general conclusions about the protection of the systematic/parity bits. In this case, and for a target BER of 10^{-6}, the difference between $\mathbf{K}_\mathrm{B}$ and $\mathbf{K}_\mathrm{W}$ is 1 dB, which is obtained without complexity increase but only by properly assigning the coded bits to the bit positions in the QAM symbol.

In Fig. 4 we present the results for $n = 3$ and $m = 3$, where we have four degrees of freedom ($\kappa_{1,1}$, $\kappa_{1,2}$, $\kappa_{2,1}$, and $\kappa_{2,2}$). The results presented in Fig. 4 are for the best and worst MD interleaver found, and also for the RN case. The best (worst) MD interleaver was found by selecting the matrix $\mathbf{K}$ that minimizes (maximizes) the UB at a target BER of 10^{-6}, yielding a difference between $\mathbf{K}_\mathrm{B}$ and $\mathbf{K}_\mathrm{W}$ of about 2 dB.

In order to calculate the bound (16) for $n = m = 3$ (cf. Fig. 4), only weights up to 50 were counted. As mentioned before, when m and/or n increase, counting all the combinations in (16) becomes tedious, and consequently, the maximum value of l considered must be relatively small. In Fig. 4 we also present results for the (asymptotic) simplifications presented in Sec. III-B. We calculate the UB' using (18) counting weights up to 100. This simplification is very simple to evaluate compared with the exact bound (16), and yet it predicts the asymptotic performance of the system as shown in Fig. 4. It is worth mentioning that the bounds for $\mathbf{K}_\mathrm{W}$ and $\mathbf{K}_\mathrm{R}$ cross at $\mathrm{BER} \approx 10^{-9}$, however, $\mathbf{K}_\mathrm{B}$ still offers a performance gain of more than 1 dB.

[3]Two identical rate-1/2 RSCs $(1, 5/7)$ are concatenated in parallel separated by a random interleaver of length N yielding an overall rate of 1/3. To obtain the rate-1/2 TC alternate puncturing of the parity bits is performed. All the polynomials are given in octal notation.

[4]To calculate the bound in (16) numerically, only weights up to 100 were counted ($w_\mathrm{f} \leq l \leq 100$).

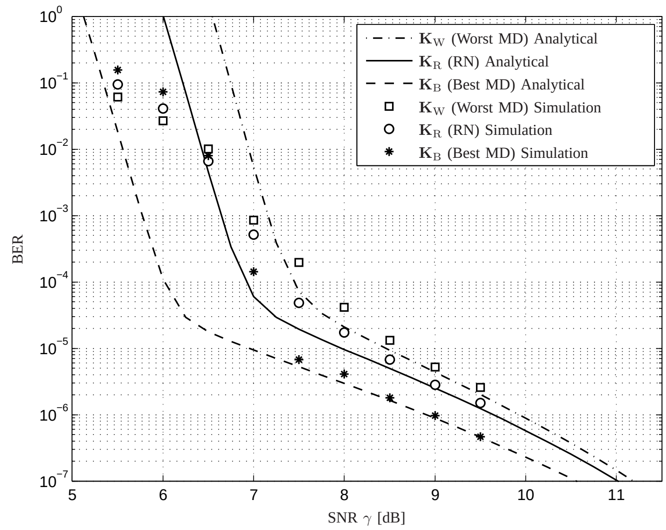

Figure 3. UB (16) and simulated BER for BICM-QAM for $R = 1/2$ and 16-QAM and different interleaver configurations. $N = 1000$ and 10 iterations.

Analyzing the results presented in Fig. 3 and Fig. 4, we can draw some interesting conclusions. For a given target BER of 10^{-6}, the SNR gains between the best and the worst interleaver configuration can be up to 2 dB. For the analyzed cases, RN interleavers were never optimum. Structured but improperly designed interleavers ($\mathbf{K}_\mathrm{W}$) can degrade the system performance compared with $\mathbf{K}_\mathrm{R}$. Thus, when using structured interleavers, the optimization of $\mathbf{K}$ becomes a mandatory step.

B. GODS Codes

It is well known that ODS codes—tabulated for example in [15]—are the optimum convolutional codes for binary transmissions. However, according to (19), when UEP is introduced by the channel, the optimization criterion is different to [15, Sec. II], namely, the interleaver and the GWDS of the code must be taken into account. In this section we define the generalized optimum distance spectrum (GODS) codes.

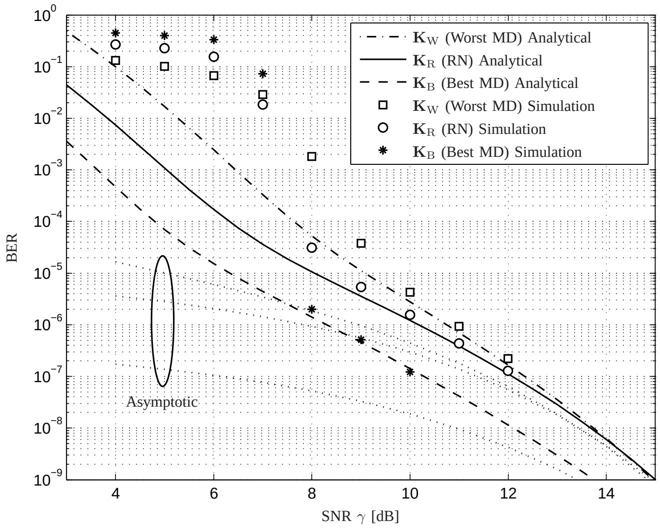

Figure 4. UB (16) and simulated BER for BICM-QAM for $R = 1/3$ and 64-QAM and different interleaver configurations. $N = 1500$ and 10 iterations. The asymptotic bounds based on (18) are also shown.

For a given constraint length K, code rate R, constellation size m, and assuming that the optimum free distance w_{f} is known (cf. for example [15, Table I, II or III]), any combination of code and interleaver will produce an asymptotic BER given by (19).

Definition 1: A GODS convolutional code ($\mathcal{C}_{\mathrm{GODS}}$) is a code that—using an optimized interleaver configuration $\mathbf{K}_{\mathrm{GODS}}$—produces an asymptotic BER which is a minimum compared with the values that any other encoder and interleaver combination can generate, i.e.,

$$[\mathcal{C}_{\mathrm{GODS}}, \mathbf{K}_{\mathrm{GODS}}] = \arg\min_{\mathcal{C},\mathbf{K}} \left\{ \sum_{\mathbf{w} \in \mathcal{Z}_n(w_{\mathrm{f}})} \beta(\mathbf{w}) \prod_{p=1}^{n} \xi_{p,1}^{w_p} \right\}, \quad (20)$$

where $\mathcal{C}$ belongs to the set of all codes with optimum w_{f}.

Using the previous definition, an exhaustive search for pairs $[\mathcal{C}_{\mathrm{GODS}}, \mathbf{K}_{\mathrm{GODS}}]$ with constraint length up to $K = 10$ was performed. Two different configurations were tested: code rate $R = 1/2$ ($n = 2$) with 64-QAM ($m = 3$), or with 256-QAM ($m = 4$). The optimization space for $\mathbf{K}$ in these cases was $\kappa_{1,1}, \kappa_{1,2} \in \{0, 1/3, 2/3\}$ for $m = 3$, and $\kappa_{1,1}, \kappa_{1,2}, \kappa_{1,3} \in \{0, 1/2, 1\}$ for $m = 4$. The results are presented in Table II, where the asterisks denote codes found that are different from the ODS codes listed in [15]. Among the 16 combinations studied, 6 resulted in new optimal codes. Extension to any other combination of code rate and modulation order is straightforward.

V. Conclusions

In this paper we developed analytical bounds to predict the performance of BICM with QAM schemes when UEP is introduced by the constellation labeling. Together with the original union bound, two asymptotic expressions which are simple to evaluate were developed. The analytical developments were supported by simulation results yielding accurate results. We quantified the attainable gains when using optimized MD interleavers over unstructured random interleavers. These improvements can be up to 2 dB for the analyzed cases, and they can be obtained without complexity increase but only if the assignment of the coded bits to the bit positions in the complex symbol is optimized. We also introduced the concept of GODS codes, which are the optimum codes for the analyzed scenario.

Table II

Optimum interleavers and codes for $n = 2$ and $m = 3$ or $m = 4$.

	64-QAM ($m=3$)			256-QAM ($m=4$)			
K	$\mathcal{C}_{\mathrm{GODS}}$	κ_{11}	κ_{12}	$\mathcal{C}_{\mathrm{GODS}}$	κ_{11}	κ_{12}	κ_{13}
3	$(5,7)$	0	$1/3$	$(5,7)$	$1/2$	$1/2$	0
4	$(15,17)$	$2/3$	$1/3$	$(15,17)$	$1/2$	$1/2$	0
5	$(27,31)^*$	0	$1/3$	$(23,35)$	$1/2$	$1/2$	0
6	$(53,75)$	0	$1/3$	$(53,75)$	$1/2$	$1/2$	0
7	$(135,147)^*$	0	$1/3$	$(135,147)^*$	0	0	$1/2$
8	$(225,373)^*$	0	$1/3$	$(247,371)$	$1/2$	$1/2$	0
9	$(557,751)^*$	0	$1/3$	$(457,755)^*$	$1/2$	$1/2$	0
10	$(1151,1753)$	0	$1/3$	$(1151,1753)$	$1/2$	$1/2$	0

References

[1] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 873–884, May 1992.

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

[3] S. Le Goff, "Bandwidth-efficient turbo coding over Rayleigh fading channels," *Int. J. Commun. Syst.*, vol. 15, pp. 621–633, 2002.

[4] I. Abramovici and S. Shamai, "On turbo encoded BICM," *Annales des Telecommunications*, vol. 54, no. 3-4, pp. 225–234, Mar. - Apr. 1999.

[5] A. Nilsson and T. M. Aulin, "On in-line bit interleaving for serially concatenated systems," in *International Conference on Communications ICC 2005*, vol. 1, May 2005, pp. 547–552.

[6] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Lett.*, vol. 1, no. 6, pp. 169–171, Nov. 1997.

[7] C. Stierstorfer and R. Fischer, "Intralevel interleaving for BICM in OFDM scenarios," in *12th International OFDM Workshop*, Hamburg, Germany, Aug. 2007.

[8] 3GPP, "Universal mobile telecommunications system (UMTS); multiplexing and channel coding (FDD)," 3GPP, Tech. Rep. TS 25.212, V6.0.0 Release 6, Dec. 2003.

[9] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband*, 1st ed. Academic Press, 2007.

[10] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "Gray coding for multilevel constellations in Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 224–235, Jan. 2007.

[11] 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.

[12] A. Alvarado, L. Szczecinski, R. Feick, and L. Ahumada, "Distribution of L-values in Gray-mapped M^2-QAM signals: Exact expressions and simple approximations," in *IEEE Global Telecommunications Conference, GLOBECOM 2007*, Washington, USA, Nov. 2007.

[13] D. Divsalar, S. Dolinar, R. J. McEliece, and F. Pollara, "Transfer function bounds on the performance of turbo codes," JPL, Cal. Tech., TDA Progr. Rep. 42-121, Aug. 1995.

[14] S. Benedetto and G. Montorsi, "Unveiling turbo codes: Some results on parallel concatenated coding schemes," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 409–428, Mar. 1996.

[15] 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.