Towards Fully Optimized BICM Transceivers

Md. Jahangir Hossain, Alex Alvarado, and Leszek Szczecinski

***Abstract*—Bit-interleaved coded modulation (BICM) transceivers often use equally spaced constellations and a random interleaver. In this paper, we propose a new BICM design, which considers hierarchical (nonequally spaced) constellations, a bit-level multiplexer, and multiple interleavers. It is shown that this new scheme increases the degrees of freedom that can be exploited in order to improve its performance. Analytical bounds on the bit error rate (BER) of the system in terms of the constellation parameters and the multiplexing rules are developed for the additive white Gaussian Noise (AWGN) and Nakagami-m fading channels. These bounds are then used to design the BICM transceiver. Numerical results show that, compared to conventional BICM designs, and for a target BER of 10^{-6}, gains up to 3 dB in the AWGN channel are obtained. For fading channels, the gains depend on the fading parameter, and reach 2 dB for a target BER of 10^{-7} and $m = 5$.**

***Index Terms*—Bit-interleaved coded modulation, bit error rate, interleaver design, multiple interleavers, L-values, nonequally spaced constellations, pulse amplitude modulation, quadrature amplitude modulation, trellis coded modulation.**

I. INTRODUCTION

Bit-interleaved coded modulation (BICM) [1]–[3] is used in most of the existing wireless communication standards, e.g., HSPA, IEEE 802.11a/g/n, DVB, etc. In BICM, the channel encoder and the modulator are separated by a bit-level interleaver which allows the designer to choose the code rate and the constellation independently. BICM maximizes the code diversity, and therefore, outperforms trellis coded modulation (TCM) in fading channels. Compared to TCM, BICM is suboptimal for the additive white Gaussian noise (AWGN) channel because it decreases the minimum Euclidean distance. Nevertheless, its simplicity and flexibility make it an attractive coded modulation scheme even when fading is not present.

BICM appears as a simple out-of-the box coded modulation scheme, however, its full potential is achieved when the design is optimized. In particular, it has been shown in [4], [5] that the interleaver and the code can be jointly designed to exploit the so-called unequal error protection (UEP) caused by the binary labeling of the equally spaced (ES) constellations. More generally speaking, if the channel offers UEP to the coded bits, gains in terms of bit error rate (BER) can be obtained, cf. [4, Sec. I] and references therein. UEP can also be intentionally introduced when designing the transceiver. For example, UEP can be imposed by allowing unequal power allocation for different bits, by deleting bits using certain patterns (puncturing), by changing the binary labeling of the constellation, or by using nonequally spaced (NES) constellations or nonequally probable symbols, known as *geometrical* and *probabilistic* shaping, respectively.

Shaping techniques for BICM have received some attention in the literature (e.g., geometrical in [6]–[8] and probabilistic in [9], [10, Sec. III-F]). All these works, however, are based on capacity maximization arguments, and therefore, the gains can be obtained if capacity-approaching codes are used. The approach we adopt in this paper is different, i.e., we opt for the joint optimization of the BICM building blocks: the code, the modulator, and the interleaver. This approach offer gains compared to previous designs because in practice, codes are sensitive to UEP, which is an inherent feature of BICM. Because of the UEP, the performance of BICM system will depend, in general, on how the modulator inputs are "matched" to the encoder's outputs. In this paper, we use a well designed interleaver to connect the modulator and the encoder, as an alternative to the random interleaver postulated by [2] which eliminates any code-modulator match. This matching technique can be used, in general, for any code that is UEP-sensitive, however, the analysis and the resulting design depend on the type of code used. In this paper we consider convolutional codes for which the analysis is tractable and also because these codes are simple and popular (e.g., BICM with convolutional codes is used in wireless standards such as HSPA [11], IEEE 802.11a/g [12], and IEEE 802.11n [13]). When convolutionally-encoded BICM is considered, capacity arguments like the ones in [6]–[10] are less relevant, and thus, in this paper we base our analysis on a different approach that aims at the minimization of the BER for a given signal to noise ratio (SNR).

In this paper, we propose not only to exploit the existing UEP but also to *control* it via geometrical shaping. We use the so-called hierarchical constellations [14] which have received a great deal of attention in many applications where independent data streams with different qualities must be sent at the

Manuscript submitted Dic. 2010; revised Apr. 2011.

Research supported by the Australian Research Council (ARC) Linkage Project (research grant #LP 0991370), NSERC, Canada (research grant #356807-07), by the Swedish Research Council, Sweden (research grant #2006-5599), and by the Department of Signals and Systems (Solveig and Karl G Eliasson Memorial Fund), Chalmers University of Technology.

The material in this paper was presented in part at the International Conference on Communications (ICC) 2010, Cape Town, South Africa, May 2010, and at the Global Communications Conference (GLOBECOM) 2011, Houston, Texas, USA, December 2011.

Md J. Hossain was the Institut National de la Recherche Scientifique, INRS-EMT, 800, Gauchetiere W. Suite 6900 Montreal, H5A 1K6, Canada and with the Institute for Telecommunications Research, University of South Australia, Australia. He is now with the School of Engineering, The University of British Columbia, Okanagan, BC, Canada (email: jahangir.hossain@ubc.ca).

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.

same time, e.g., in multi-resolution image transmission [15], [16], and simultaneous voice and multi-class data transmission [17]. Hierarchical quadrature amplitude modulation (HQAM) constellations are also used in QUALCOMM's MediaFLO [18] and have been standardized for the latest digital video broadcasting-terrestrial (DVB-T2) [19], [20]. In MediaFLO, the source encoder generates two information bit streams known as base and enhancement layers. These two independent streams are processed/coded independently and modulated onto the same HQAM symbol. The base layer bits are mapped to the more reliable bits of the constellation and the enhancement layer bits are mapped to the less reliable bits of the constellation. By doing this, the receivers with relatively good channel quality can decode both layers correctly and the receivers with relatively weak channel quality can decode only the base layer. In this paper, we use HQAM constellations in a different context than the one described above, i.e., we use them to transmit only one data stream with improved error performance. In particular, we consider that a single bit stream, generated by the source encoder, is encoded by the channel encoder. Then different coded bits are mapped to the different bit positions of the HQAM constellation with a particular assignment pattern. This allows to protect different coded bits differently against the channel impairments and the coded BER of the single information bit stream is improved.

In order to exploit the UEP offered by the constellation, the interleaver must be properly designed. The most commonly interleaver considered in the literature is the single interleaver (S-interleaver) of [2], which eliminates the UEP caused by the binary labeling. UEP in BICM was in fact considered an "undesired feature" in [2, Sec. II]. Recently, the the so-called multiple interleavers (M-interleavers) [4] were shown to improve the performance of the system by exploiting the UEP caused by the modulator. In fact, the use of BICM with M-interleavers (BICM-M) corresponds to the original BICM configuration proposed by Zehavi in [1], as well as the original BICM with iterative decoding (BICM-ID) scheme proposed by Li and Ritcey in [21]. M-interleavers have also been shown to outperform S-interleavers when BICM-ID is considered [5].

The BICM-M system in [4], [5] uses a random bit-level multiplexing (R-MUX) that connects the encoder and the M-interleavers, i.e., the M-interleavers assign the coded bits to a particular bit position in the modulator in a pseudo-random fashion (with predetermined probabilities). By doing this, the dependency of adjacent coded bits is ignored. In this paper, we propose an multiplexing/interleaving inspired by the well-known puncturing strategy based on the periodic elimination of the bits according to a prescribed pattern that matches the temporal structure of the code, cf. [22]. We show that such a deterministic multiplexing (D-MUX) of the coded bits (followed by random interleaving) notably outperforms the R-MUX used in [4].

The contributions of this paper can be summarized as follows. We propose and study a BICM scheme for fading and nonfading channels which considers the use of HQAM constellations (HQAM-BICM), a periodic (and deterministic) bit-level multiplexer, and M-interleavers. It is demonstrated that the degrees of freedom of such a scheme can be exploited to notably improve performance of the system in terms of BER. Using a generalization of the so-called consistent model (CoMod) introduced in [23], we develop closed-form expressions for the probability density function (PDF) of the L-values for HQAM-BICM, which are shown to be Gaussian mixtures that depend on the constellation parameters. These PDFs are used to develop union bounds (UB) on the BER of the system for fading and nonfading channels, which in turn are then used to optimize the transceiver's design. Presented numerical examples show that the proposed system offers gains over previous BICM configurations (ES-QAM and S-interleavers [2] or ES-QAM and M-interleavers [4]). For the particular cases analyzed in this paper, the gains can be up to 3 dB for a BER target of 10^{-6} in the AWGN channel, and for the Nakagami-m fading channel, the gains can reach 2 dB for a target of 10^{-7} and $m = 5$.

II. PROPOSED BICM TRANSCEIVER

Throughout this paper, we use boldface letters $\boldsymbol{c}_t = [c_{1,t}, \ldots, c_{N,t}]$ to denote row vectors and capital boldface letters $\boldsymbol{C} = [\boldsymbol{c}_1^{\mathrm{T}}, \ldots, \boldsymbol{c}_M^{\mathrm{T}}]^{\mathrm{T}}$ to denote a matrix of M rows, where $(\cdot)^{\mathrm{T}}$ denotes transposition. We denote probability by $\Pr(\cdot)$ and the PDF of a random variable X by $p_X(x)$. A Gaussian distribution with mean value μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$, the Gaussian PDF with the same parameters by $\psi(\lambda; \mu, \sigma) \triangleq \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(\lambda-\mu)^2}{2\sigma^2})$, and the Q-function by $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{u^2}{2}\right) du$. The combinations of i nonegative integers such that their sum is l is denoted by $\mathcal{W}_i(l)$, where $\mathcal{W}_i(l) \triangleq \{[w_1, \ldots, w_i] \in (\mathbb{Z}^+)^i : w_1 + \ldots + w_i = l\}$.

The HQAM-BICM system model under consideration is shown in Fig. 1. In what follows, we describe functionalities of various blocks of such transmission scheme.

A. Encoder, Multiplexing, and Interleaving

The k_{c} vectors of information bits $\boldsymbol{i}_l = [i_{l,1}, \ldots, i_{l,N_{\mathrm{c}}}]$ with $l = 1, \ldots, k_{\mathrm{c}}$ are encoded by a rate $R = k_{\mathrm{c}}/n$ convolutional encoder (ENC) yielding the vectors of coded bits $\boldsymbol{c}_p = [c_{p,1}, \ldots, c_{p,N_{\mathrm{c}}}]$ with $p = 1, \ldots, n$. These are then fed to a deterministic multiplexing (D-MUX) unit which bijectively maps $\boldsymbol{C} = [\boldsymbol{c}_1^{\mathrm{T}}, \ldots, \boldsymbol{c}_n^{\mathrm{T}}]^{\mathrm{T}}$ onto $\boldsymbol{O} = [\boldsymbol{o}_1^{\mathrm{T}}, \ldots, \boldsymbol{o}_q^{\mathrm{T}}]^{\mathrm{T}}$ with $\boldsymbol{o}_k = [o_{k,1}, \ldots, o_{k,N_{\mathrm{s}}}]$ and $k = 1, \ldots, q$. Without loss of generality, we assume $N_{\mathrm{s}} q = N_{\mathrm{c}} n$. The vector of bits after the D-MUX are fed to q parallel interleavers π_k. The q interleavers are assumed to be independent and give randomly permuted sequences of the bits, i.e., $\boldsymbol{u}_k = \pi_k\{\boldsymbol{o}_k\}$. Each of the interleavers is connected to the qth bit positions in the hierarchical M-ary pulse amplitude modulation (HPAM) constellation, where $q = \log_2 M$.

In general, the D-MUX can be defined as a one-to-one mapping between the blocks of nN_{c} and qN_{s} bits, i.e., $\{0,1\}^{nN_{\mathrm{c}}} \leftrightarrow \{0,1\}^{qN_{\mathrm{s}}}$. We define it via an $n \times N_{\mathrm{c}}$ matrix $\tilde{\boldsymbol{K}}$, whose (p, t')th entry is a pair (k, t) where $k \in \{1, \ldots, q\}$ and $t \in \{1, \ldots, N_{\mathrm{s}}\}$. The entry (k, t) indicates that the bit $c_{p,t'}$ is assigned to the kth D-MUX's output at time instant t, i.e., $o_{k,t} = c_{p,t'}$.

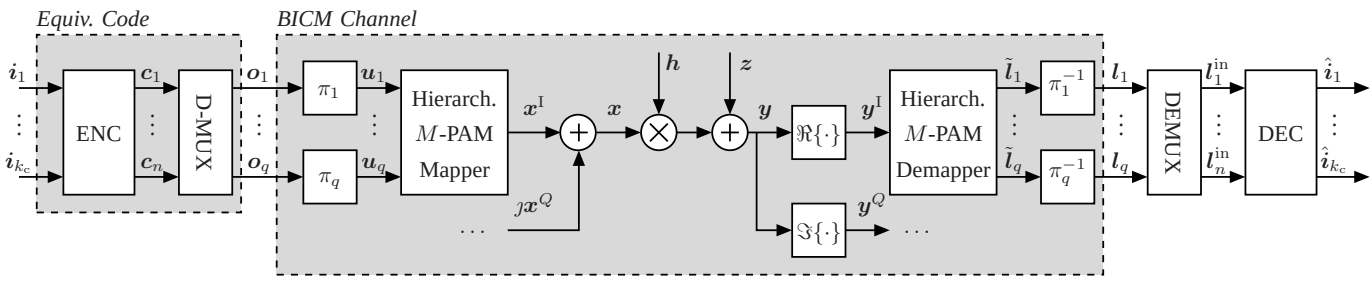

Fig. 1. Model of HQAM-BICM transmission: a channel encoder followed by the multiplexer (D-MUX), the interleavers ($\pi_1, \ldots, \pi_q$), the hierarchical M-PAM mapper, the channel, and the processing blocks at the receiver's side.

The previous definition of the D-MUX is entirely general but difficult to deal with, and thus, in this paper we only consider D-MUX configurations that operate periodically over blocks of nJ bits. We then represent $\tilde{\boldsymbol{K}}$ as a concatenation of N_c/J matrices $\boldsymbol{K}_\tau$, each of dimensions $n \times J$, i.e., $\tilde{\boldsymbol{K}} = [\boldsymbol{K}_0, \ldots, \boldsymbol{K}_{N_\mathrm{c}/J-1}]$, where the variable J is called the period of the D-MUX. The entries of $\boldsymbol{K}_\tau$ are pairs $(k, t + \tau nJ/q)$ with $k \in \{1, \ldots, q\}$ and $t \in \{1, \ldots, nJ/q\}$. Without loss of generality, we assume that $(N_\mathrm{c} \bmod J) = 0$ and that $(nJ \bmod q) = 0$. To clarify these definitions, consider the following example.

Example 1: Assume $k_\mathrm{c} = 1$ and $n = 2$ ($R = 1/2$), $J = 3$, and $q = 3$ (8-ary constellation). One possible D-MUX is defined by

$$\boldsymbol{K}_\tau = \begin{bmatrix} (1, 1+2\tau) & (2, 2+2\tau) & (2, 1+2\tau) \\ (1, 2+2\tau) & (3, 2+2\tau) & (3, 1+2\tau) \end{bmatrix}, \quad (1)$$

which results in

$$\tilde{\boldsymbol{K}} = \left[\begin{array}{ccc|ccc|c} (1,1) & (2,2) & (2,1) & (1,3) & (2,4) & (2,3) & \ldots \\ (1,2) & (3,2) & (3,1) & (1,4) & (3,4) & (3,3) & \ldots \end{array}\right].$$

The mapping between $\boldsymbol{C}$ and $\boldsymbol{O}$ is then

$$\boldsymbol{C} = \left[\begin{array}{ccc|ccc|c} c_{1,1} & c_{1,2} & c_{1,3} & c_{1,4} & c_{1,5} & c_{1,6} & \ldots \\ c_{2,1} & c_{2,2} & c_{2,3} & c_{2,4} & c_{2,5} & c_{2,6} & \ldots \end{array}\right]$$
$$\Longleftrightarrow \boldsymbol{O} = \left[\begin{array}{cc|cc|c} c_{1,1} & c_{2,1} & c_{1,4} & c_{2,4} & \ldots \\ c_{1,3} & c_{1,2} & c_{1,6} & c_{1,5} & \ldots \\ c_{2,3} & c_{2,2} & c_{2,6} & c_{2,5} & \ldots \end{array}\right]. \quad (2)$$

Since a matrix $\boldsymbol{K}_\tau$ is simply a permutation of the set $\{1, \ldots, q\} \times \{1, \ldots, nJ/q\}$, $(nJ)!$ different matrices $\boldsymbol{K}_\tau$ can be generated. However, this number can be reduced since *trivial operations* that do not affect the performance of the system can be applied to $\boldsymbol{K}_\tau$. For example, for the matrix in (1) with $\tau = 0$, consider the following two matrices:

$$\boldsymbol{K}_0' = \begin{bmatrix} (1,2) & (2,2) & (2,1) \\ (1,1) & (3,1) & (3,2) \end{bmatrix},$$
$$\boldsymbol{K}_0'' = \begin{bmatrix} (2,1) & (1,1) & (2,2) \\ (3,1) & (1,2) & (3,2) \end{bmatrix}.$$

The matrix $\boldsymbol{K}_0'$ is obtained by permuting the elements of $\boldsymbol{K}_0$ such that the first elements in the entries of $\boldsymbol{K}_0$ are not altered. Because M-interleavers are used after the D-MUX (cf. Fig. 1), the temporal structure of the sequences $\boldsymbol{o}_k$ is randomized, and thus, the second elements of the entries are not (which determines to which time the bit $c_{p,t'}$ is assigned) relevant. Consequently, the performance of the system using $\boldsymbol{K}_0$ or $\boldsymbol{K}_0'$ will be the same. The matrix $\boldsymbol{K}_0''$ is obtained by cyclically rotating the columns of the matrix $\boldsymbol{K}_0$, which will produce $\tilde{\boldsymbol{K}}$ with columns shifted to the left or right. For long coded sequences, i.e., $J \ll N_\mathrm{c}$, the original matrix and its shifted version will yield the same performance.

It is worth mentioning that while in general we are able to discard matrices that result from trivial operations, the only way to determine which of the available (and non-trivial) matrices $\boldsymbol{K}_\tau$ is the optimal one for a given code and modulation, is via the BER-criterion we develop in Sec. IV-A.

B. HPAM Constellations

In this paper, we consider HQAM constellations labeled by the binary reflected Gray code (BRGC) [24] presented in [14]. In HQAM constellations, each symbol is a superposition of independently modulated real/imaginary parts, which allows us to focus on the equivalent HPAM constellation, cf. Fig. 1. At any time instant t, the coded and interleaved bits $[u_{1,t}, \ldots, u_{q,t}]$ are mapped to an HPAM symbol $x^\mathrm{I}(t) \in \mathcal{X} = \{x_0^\mathrm{I}, \ldots, x_{M-1}^\mathrm{I}\}$ using a binary memoryless mapping $\mathcal{M} : \{0,1\}^q \to \mathcal{X}$. Since the mapper is memoryless, from now on we drop the time index t.

We analyze HPAM constellations as the one shown in Fig. 2 ($M = 8$), which are defined by the distances d_k with $k = 1, \ldots, q$. In this figure, the M constellation points are shown with black circles, where the white squares/triangles are "virtual" symbols that help to understand the construction of the HPAM constellation as explained below. We use $k = 1, \ldots, q$ to denote the bit position of the binary labeling, where $k = 1$ represents the left most bit position. The bit value of $k = 1$ selects one of the two squares in Fig. 2. Similarly, for a given value of the first bit, the bit value for the next position ($k = 2$) selects one of the two triangles that surround the previously selected square. Finally, given the bit values for $k = 1$ and $k = 2$, the bit value of bit position $k = 3$ selects one of the two black symbols that surround the previously selected triangle. This selected symbol (black circle) is finally transmitted by the modulator.

We denote the base-2 representation of the integer $0 \le j \le M - 1$ by the vector $\hat{\boldsymbol{b}}(j) = [\hat{b}_1(j), \ldots, \hat{b}_q(j)]$, where $\hat{b}_1(j)$ is the most significant bit of j and $\hat{b}_q(j)$ the least significant. This allows us to express the elements $x_j^\mathrm{I} \in \mathcal{X}$ of the HPAM

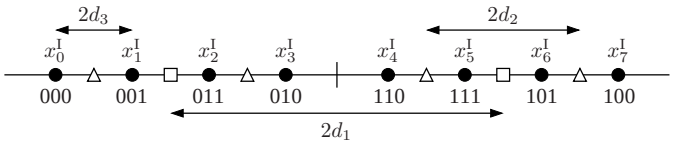

Fig. 2. HPAM ($M = 8$) constellation labeled with the BRGC.

constellation as

$$x_j^{\mathrm{I}} = \sum_{k=1}^{q} (-1)^{\hat{b}_k(j)-1} d_k. \tag{3}$$

We also define the normalized constellation parameters as

$$\alpha_k \triangleq \frac{d_{k+1}}{d_1}, \tag{4}$$

with $k = 1, \ldots, q-1$. Using (3), and for equiprobable symbol transmission, the average symbol energy is given by $E_{\mathrm{s}} = (1 + \sum_{k=1}^{q-1} \alpha_k^2) d_1^2$ with α_k is given by (4). Throughout this paper, we consider that the constellation is normalized to have unit energy, which translates into the relation $d_1 = (\alpha_1^2 + \alpha_2^2 + \ldots + \alpha_{q-1}^2 + 1)^{-1/2}$.

If the constellation points move freely, it is possible that they cross each other (by having for example $d_3 < 0$ in Fig. 2), and therefore, the binary labeling is not the BRGC anymore. Since in this paper we restrict the analysis to the BRGC, extra constraints on the values of α_k must be added, namely,

$$\alpha_k \geq \sum_{j=k+1}^{q-1} \alpha_j, \sum_{k=1}^{q-1} \alpha_k \leq 1, \text{ and } \alpha_{q-1} \geq 0, \tag{5}$$

where $k = 1, \ldots, q-1$. The inequalities in (5) are found by solving $(x_{j+1} - x_j) \geq 0$ with $j = 0, \ldots, M-2$.

Example 2 (Constellation parameters for $M = 8$): For $M = 8$, the constellation optimization space is formed by two variables, $\alpha_1 = d_2/d_1$ and $\alpha_2 = d_3/d_1$, cf. (4). From (5) we have the following constrains $\alpha_1 \geq \alpha_2$, $\alpha_1 + \alpha_2 \leq 1$, and $\alpha_2 \geq 0$, which result in a pair of constellation parameters (α_1, α_2) shown in Fig. 3. In this figure, the evolution of the constellation for different values of (α_1, α_2) are shown; the shadowed region represents the values of (α_1, α_2) that give a BRGC-labeled constellation. Particularly important cases are the equally spaced 2-PAM, 4-PAM, and 8-PAM constellations.

The result of the transmission of a complex symbol $x = x^{\mathrm{I}} + \jmath x^{Q}$ is given by $y = hx + z$, where $h = h^{\mathrm{I}} + \jmath h^{Q}$ is the complex channel gain, and z is a complex Gaussian noise with zero mean and variance $N_0/2$ in each dimension. The amplitude of the channel gain $|h|$ follows a Nakagami-m distribution [25], and thus, the instantaneous SNR, defined as $\gamma \triangleq \frac{|h|^2}{N_0}$, follows a Gamma distribution, i.e., [26, eq. (2.21)], [27, eq. (3)]

$$p_{\Gamma}(\gamma; \overline{\gamma}) = \frac{\gamma^{m-1}}{G(m)} \left(\frac{m}{\overline{\gamma}}\right)^m \exp\left(-\frac{m\gamma}{\overline{\gamma}}\right), \tag{6}$$

where $G(m)$ is the Gamma function, $\overline{\gamma} = \mathbb{E}_{\Gamma}[\gamma]$ is the average SNR, and Γ is the random variable that represents the instantaneous SNR. The AWGN channel is obtained when $|h| = 1$.

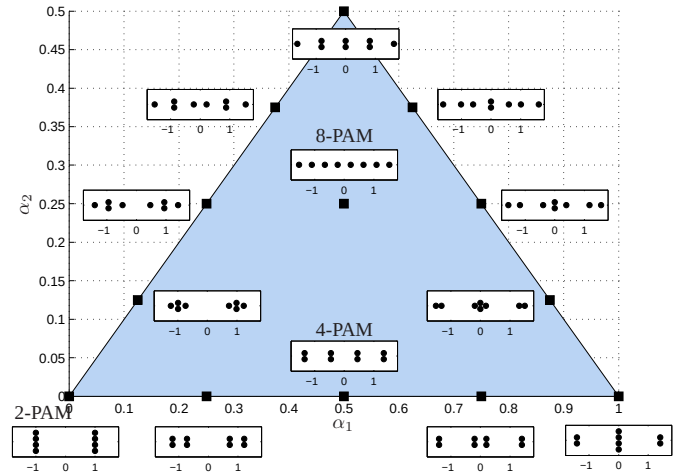

Fig. 3. Constellation parameters for an HPAM constellation with $M = 8$. The shadowed region shows the values of (α_1, α_2) that give a BRGC-labeled constellation and the 13 filled squared some particular constellations (not at scale). Any point outside this region corresponds to a non-BRGC constellation.

At the receiver's side, the real part of the received signal is normalized by the channel gain h ($y^{\mathrm{I}} = \Re\{y/h\}$) and passed to the demapper, which computes logarithmic likelihood ratios (L-values) for each bit in the transmitted symbol. The kth L-value given the transmitted symbol x_j and the channel gain h (or equivalently, γ) can be written as [1]–[4]

$$\tilde{l}_k(y^{\mathrm{I}}|x_j, \gamma) \triangleq \log \frac{\Pr(u_k = 0|y^{\mathrm{I}}, \gamma)}{\Pr(u_k = 1|y^{\mathrm{I}}, \gamma)} \tag{7}$$

$$\approx \gamma \left[\min_{a \in \mathcal{X}_{k,1}} \left\{ (y^{\mathrm{I}} - a)^2 \right\} - \min_{a \in \mathcal{X}_{k,0}} \left\{ (y^{\mathrm{I}} - a)^2 \right\} \right], \tag{8}$$

where $\mathcal{X}_{k,b}$ is the set of symbols labeled with the kth bit equal to b, and where we have used the so-called max-log approximation [1], [2], [28].

The vectors of L-values calculated by the demapper are then deinterleaved, generating the sequence $\boldsymbol{l}_k = \pi_k^{-1}(\tilde{\boldsymbol{l}}_k)$ with $k = 1, \ldots, q$, cf. Fig. 1. These L-values are reorganized by the demultiplexer unit (DEMUX), defined as $l_{p,t'}^{\mathrm{in}} = l_{k,t}$, which simply inverts the process done by the D-MUX at the transmitter. Finally, these L-values are passed to the the channel decoder which produces an estimate of the transmitted bits. In this paper, we consider convolutional codes and a soft-input Viterbi decoder.

III. EQUIVALENT CHANNEL MODEL

In order to predict the coded BER performance of the system, finding the PDF of the L-values passed to the channel decoder is crucial. In what follows, we develop closed-form expressions for the PDF of L-values for HQAM-BICM transmission as a function of the constellation parameters. These expressions will later be used to compute bounds on the BER of the systems, and then, used to optimize the design of the system. From now on, all the analysis is made for the constituent HPAM constellation (cf. Fig. 1), and thus, with a slight abuse of notation, we use x and y to denote the real

TABLE I
VALUES OF $\mu_{k,j}$ FOR HPAM WITH $M = 8$ IN (10).

x_j	$\mu_{1,j}$	$\mu_{2,j}$	$\mu_{3,j}$
x_0	$+4d_1^2$	$+4d_2^2$	$+4d_3^2$
x_1	$+4(d_1-d_3)^2$	$+4(d_2-d_3)^2$	$-4d_3^2$
x_2	$+4(d_1-d_2)^2$	$-4(d_2-d_3)^2$	$-4d_3^2$
x_3	$+4(d_1-d_2-d_3)^2$	$-4d_2^2$	$+4d_3^2$
x_4	$-4(d_1-d_2-d_3)^2$	$-4d_2^2$	$+4d_3^2$
x_5	$-4(d_1-d_2)^2$	$-4(d_2-d_3)^2$	$-4d_3^2$
x_6	$-4(d_1-d_3)^2$	$+4(d_2-d_3)^2$	$-4d_3^2$
x_7	$-4d_1^2$	$+4d_2^2$	$+4d_3^2$

part of the transmitted symbol and the real part of the received signal, respectively, i.e., we skip the superscript $(\cdot)^{\mathrm{I}}$.

The use of the max-log approximation in (8) transforms the nonlinear relation between the received signal and the L-values into a piecewise linear relation. Examples of this piece-wise linear relation can be found in literature, see for example [23, Fig. 3], [29, Fig. 2, Fig. 4], [30, Fig. 3], or [31, Table I]. This piece-wise linear relation has been used to develop expressions for the PDF of the L-values in (8) using arbitrary signal sets in [32] (based on an algorithmic approach), closed-form expressions for ES-QAM constellations labeled with the BRGC for the AWGN channel in [23], and for fading channels in [33]. Recently, closed-form approximations for the PDF of the L-values for arbitrary signal sets and binary labeling in fading channels have been developed in [27].

For a given transmitted symbol x_j and SNR γ, $Y \sim \mathcal{N}(x_j, 1/(2\gamma))$. Using a generalization of the so-called consistent model (CoMod) introduced in [23], the PDF of the L-values in (8) can be approximated by

$$\tilde{L}_k(x_j;\gamma) \sim \mathcal{N}(\gamma\mu_{k,j}, \gamma\sigma_{k,j}^2), \tag{9}$$

where

$$\begin{aligned} \mu_{k,j} &= (-1)^{2-\tilde{b}_{k,j}}(\hat{x}_{k,j}-x_j)^2, \\ \sigma_{k,j}^2 &= 2(\hat{x}_{k,j}-x_j)^2 = 2|\mu_{k,j}|, \end{aligned} \tag{10}$$

$\tilde{\boldsymbol{b}}_j = [\tilde{b}_{1,j},\ldots,\tilde{b}_{q,j}]$ is the binary label of the symbol x_j, and $\hat{x}_{k,j}$ is the closest symbol to x_j with the opposite bit value at bit position k. The model in (10) fulfills the consistency condition ($\sigma_{k,j}^2 = 2|\mu_{k,j}|$), and thus, the Gaussian distribution in (9) is completely determined by $\mu_{k,j}$.

The results in (9)–(10) can be considered as a particular case of the results in [27, "Case 1", Fig. 2, Tab. 1]). In what follows, we will develop generic expressions for $\mu_{k,j}$ in (10) in terms of the distances defining the HPAM constellation.

In Table I, we present the values of $\mu_{k,j}$ in (10) as a function of the constellation distances d_k for $M = 8$. These values are obtained by direct inspection of Fig. 2. From this table we can see that the symmetry of the constellation is reflected in the mean values. For example, for $k = 1$, $\mu_{1,3} = -\mu_{1,4}$, $\mu_{1,2} = -\mu_{1,5}$, $\mu_{1,1} = -\mu_{1,6}$, and $\mu_{1,0} = -\mu_{1,7}$. If we analyze the variances, cf. (10), we note that for $k = 1$ there are 4 different variances, for $k = 2$ two different variances, and for $k = 3$ only one. This idea can be generalized, i.e., from Table I, it is possible to infer that for a given k, there are $M_k \triangleq \frac{M}{2^k}$ different variances, which are determined by the first M_k values of j. This idea was previously used in [23].

The performance evaluation in Sec. IV is based on the transmission of the all-zero sequence, and thus, here we only need to consider positive values of $\mu_{k,j}$ (from (10), $\mu_{k,j} > 0$ if $\tilde{b}_{k,j} = 0$). From the evolution of $\mu_{k,j}$ in Table I, we can write a generic closed-form expression for $\mu_{k,j}$ (for any M) in terms of constellation distance parameters as follows

$$\mu_{k,j} = 4\left(d_k - \sum_{k'=k+1}^{q} \check{b}_{k'-k}(j) d_{k'}\right)^2, \tag{11}$$

where $j = 0,1,\ldots,M_k-1$, $k = 1,2,\ldots,q$, and $\check{\boldsymbol{b}}(j) = [\check{b}_q(j),\ldots,\check{b}_1(j)]$ is the binary representation of the integer j where $\check{b}_1(j)$ is the least significant bit.

Using the approximation for the PDF of the L-values given in (9), it is possible to build an equivalent model for the *BICM channel* shown in Fig. 1 [23], [27], [34]. This model considers M_k *virtual channels*, each of them determined by $\mu_{k,j}$ in (11). A given bit $o_{k,t} = 0$ can be transmitted through the jth virtual channel with a probability given by $\xi_{k,j} = \frac{1}{M_k}$. Then, the PDF of the L-values at the output of the kth interleaver L_k can be expressed as a Gaussian mixture with density given by

$$\begin{aligned} p_{L_k}(\lambda;\gamma) &= \sum_{j=0}^{M_k-1} \xi_{k,j}\psi(\lambda;\gamma\mu_{k,j},2\gamma\mu_{k,j}) \\ &= \frac{2^k}{M}\sum_{j=0}^{M/2^k-1} \psi(\lambda;\gamma\mu_{k,j},2\gamma\mu_{k,j}). \end{aligned} \tag{12}$$

IV. PERFORMANCE ANALYSIS

In this section, we develop union bounds (UBs) on the BER of the HQAM-BICM system proposed in Sec. II using the PDF of the L-values developed in Sec. III.

A. Union Bound

We define a *remerging sequence* as a path in the trellis of the code (ENC) that leaves the zero state and remerge with it after certain number of trellis stages. The ENC and the D-MUX are grouped into an "equivalent code" (as shown in Fig. 1) and characterized by an equivalent weight distribution spectrum (EWDS) $\beta_{\boldsymbol{K}}(\boldsymbol{w})$ with $\boldsymbol{w} = [w_1,\ldots,w_q] \in (\mathbb{Z}^+)^q$.[1] We use the notation $\beta_{\boldsymbol{K}}(\boldsymbol{w})$ to emphasize that the EWDS depends on the D-MUX configuration determined by $\boldsymbol{K}$. This EWDS counts the Hamming weights of all the input sequences that generate remerging sequences with weight $\boldsymbol{w}$ at the D-MUX's output.

Using the previous definitions, we can express the (truncated) UB on the BER as

$$\mathrm{BER} \le \mathrm{UB} \approx \frac{1}{k_{\mathrm{c}}} \sum_{w=w^{\mathrm{free}}}^{\hat{w}} \sum_{\boldsymbol{w}\in\mathcal{W}_q(w)} \beta_{\boldsymbol{K}}(\boldsymbol{w})\mathrm{PEP}(\boldsymbol{w};\overline{\gamma}), \tag{13}$$

[1]To alleviate the notation, from now on we will refer to the matrix $\boldsymbol{K}_\tau$ as $\boldsymbol{K}$.

where $\mathrm{PEP}(\boldsymbol{w};\overline{\gamma})$ is the pairwise error probability which represents the probability that the decoder selects a codeword with weight $\boldsymbol{w}$ instead of the transmitted all-zero codeword. The PEP can be expressed in terms of the decision variable $D(\boldsymbol{w})$ as

$$\mathrm{PEP}(\boldsymbol{w};\overline{\gamma}) = \Pr\{D(\boldsymbol{w}) > 0\}, \tag{14}$$

where

$$D(\boldsymbol{w}) = \sum_{l=1}^{w_1} L_1^{(l)} + \ldots + \sum_{l=1}^{w_q} L_q^{(l)}, \tag{15}$$

and where $L_k^{(i)}$ are independent samples of the random variables representing the L-values whose PDF is given by (12). In the following subsections, we will show how to compute $\beta_{\boldsymbol{K}}(\boldsymbol{w})$ and $\mathrm{PEP}(\boldsymbol{w};\overline{\gamma})$ required to evaluate the UB in (13).

B. Equivalent Weight Distribution Spectrum

The vector $\boldsymbol{w}$ corresponds to the Hamming weights of the rows of the matrix $\boldsymbol{O}$ generated by remerging sequences represented by the matrix $\boldsymbol{C}$. The correspondence between $\boldsymbol{C}$ and $\boldsymbol{O}$ is determined by the matrix $\boldsymbol{K}$ (which defines the D-MUX) as well as by the time at which the remerging sequence starts to diverge. However, due to the periodic structure of $\tilde{\boldsymbol{K}}$, only J time instants must be considered. Based on this, the EWDS can be expressed as

$$\beta_{\boldsymbol{K}}(\boldsymbol{w}) = \frac{1}{J}\sum_{j=1}^{J} \beta_{\boldsymbol{K}}^{(j)}(\boldsymbol{w}), \tag{16}$$

where $\beta_{\boldsymbol{K}}^{(j)}(\boldsymbol{w})$ represents the EWDS when the decoder starts to diverge at time $t+j$ with arbitrary t. We note that the similarities between the EWDS in (16) and the computation of the WDS of punctured convolutional codes [22]. The following example clarifies the main principle behind (16), while more details can be found in [22, Sec. II-B].

Example 3 (EWDS of the code $(5,7)_8$): Consider the constraint length $K = 3$ convolutional code with polynomial generators $(5,7)_8$, with $w^{\mathrm{free}} = 5$ and the D-MUX in Example 1. For this code, there is one divergent sequence generated by an input sequence with Hamming weight one and output weight w^{free}. The $J = 3$ possible input sequences are $\boldsymbol{i}_1^{(1)} = [\ldots,0,1,0,0,0,0,\ldots]$, $\boldsymbol{i}_1^{(2)} = [\ldots,0,0,1,0,0,0,\ldots]$, and $\boldsymbol{i}_1^{(3)} = [\ldots,0,0,0,1,0,0,\ldots]$, which result in the following matrices $\boldsymbol{C}$

$$\boldsymbol{C}^{(1)} = \begin{bmatrix} \ldots\,1\;0\;1\;0\;0\,\ldots \\ \ldots\,1\;1\;1\;0\;0\,\ldots \end{bmatrix},$$
$$\boldsymbol{C}^{(2)} = \begin{bmatrix} \ldots\,0\;1\;0\;1\;0\,\ldots \\ \ldots\,0\;1\;1\;1\;0\,\ldots \end{bmatrix},$$
$$\boldsymbol{C}^{(3)} = \begin{bmatrix} \ldots\,0\;0\;1\;0\;1\,\ldots \\ \ldots\,0\;0\;1\;1\;1\,\ldots \end{bmatrix},$$

which by using (2) yield

$$\boldsymbol{O}^{(1)} = \begin{bmatrix} \ldots\,1\;1\;0\,\ldots \\ \ldots\,1\;0\;0\,\ldots \\ \ldots\,1\;1\;0\,\ldots \end{bmatrix},$$
$$\boldsymbol{O}^{(2)} = \begin{bmatrix} \ldots\,0\;0\;1\;1\,\ldots \\ \ldots\,0\;1\;0\;0\,\ldots \\ \ldots\,1\;1\;0\;0\,\ldots \end{bmatrix},$$
$$\boldsymbol{O}^{(3)} = \begin{bmatrix} \ldots\,0\;0\;0\;1\;0\,\ldots \\ \ldots\,1\;0\;0\;1\;0\,\ldots \\ \ldots\,1\;0\;0\;1\;0\,\ldots \end{bmatrix}.$$

If we consider only this event at minimum Hamming distance (with input weight one), the final EWDS given by (16) is obtained by computing the Hamming weights of the rows of $\boldsymbol{O}^{(j)}$ with $j = 1,2,3$, i.e.,

$$\beta_{\boldsymbol{K}}(\boldsymbol{w}) = \begin{cases} \frac{2}{3}, & \text{if } \boldsymbol{w} = [2,1,2] \\ \frac{1}{3}, & \text{if } \boldsymbol{w} = [1,2,2] \end{cases}.$$

The spectrum $\beta_{\boldsymbol{K}}(\boldsymbol{w})$ can be numerically calculated using a breadth-first search algorithm [35]. Clearly, the spectrum must be truncated so that only diverging sequences with total Hamming weight $w_1 + \ldots + w_q \leq \hat{w}$ are considered, cf. (13).

C. Computation of $\mathrm{PEP}(\boldsymbol{w};\overline{\gamma})$

A common approach for computing the PEP in (14) is through the use of the Laplace transform of the PDF of the decision variable (see for example [27] and the references therein). However, due to its simplicity and accuracy, the saddlepoint approximation (SPA) [36] has recently attracted considerable interest. In this subsection, we use a generalization of the PEP computation based on the SPA used in [27], [33], [36], and we apply it to BICM systems based on M-interleavers and HQAM constellations.

Let $\Phi_{L_k}(s;\overline{\gamma})$ be the two-sided Laplace transform of the PDF of the L-value L_k in (12), and let $\Phi'_{L_k}(s;\overline{\gamma})$ and $\Phi''_{L_k}(s;\overline{\gamma})$ be its first and second derivative with respect to s, respectively. Let also denote the so-called saddlepoint by $\hat{s}$, where $\hat{s}$ is the solution of $\Phi'_{L_k}(\hat{s};\overline{\gamma}) = 0$.

Theorem 1: The PEP in (14) can be approximated using the SPA as

$$\mathrm{PEP}(\boldsymbol{w};\overline{\gamma}) \approx \frac{1}{\hat{s}\sqrt{2\pi}}\left[\sum_{k=1}^{q} w_k \frac{\Phi''_{L_k}(\hat{s};\overline{\gamma})}{\Phi_{L_k}(\hat{s};\overline{\gamma})}\right]^{-1/2} \cdot \prod_{k=1}^{q}\left[\Phi_{L_k}(\hat{s};\overline{\gamma})\right]^{w_k}. \tag{17}$$

Proof: The proof is given in Appendix A. ■

The PEP in Theorem 1 allows us to compute UBs on the BER for the proposed HQAM-BICM for the AWGN and Nakagami-m fading channels, as stated in the following two theorems.

Theorem 2: The UB for HQAM-BICM for the AWGN channel ($\gamma = \overline{\gamma}$) using the SPA is given by (18) (shown at the bottom of the page)

Proof: The proof is given in Appendix B. ■

Theorem 3: The UB for HQAM-BICM for the Nakagami-m fading channel using the SPA is given by (19) (shown at the bottom of the page)

Proof: The proof is given in Appendix C. ■

The expressions in Theorems 2 and 3 explicitly show the mean values $\mu_{k,j}$ (which depend on the constellation parameters) and the EWDS of the code, and thus, they can be used to optimize the performance of the system. Moreover, we recognize that the PEP computation in (14) for both AWGN and Nakagami-m fading channels can be done directly using the PDF of the L-values in (12) (cf. [4] for the AWGN channel). Nevertheless, we used the SPA because it results in ready-to-use formulas, cf. (18) and (19).

We conclude this section by noting that the BICM models in [4] and [2] can be regarded as a particular cases of the model we introduced in this paper. The model in [4] can be obtained using our model if $J \to N_{\rm c}$ and the first elements of the entries (k,t) of $\tilde{\boldsymbol{K}}$, which represent the assignement of the coded bits to a particular interleaver, are randomly selected with predetermined probabilities. The BICM with S-interleavers (BICM-S) configuration of [2] can be obtained by letting $J \to N_{\rm c}$ and by selecting a matrix $\tilde{\boldsymbol{K}}$ with elements randomly permuted. By doing this, we assure that the coded bits are uniformly assigned over time and also over the bit positions.

V. NUMERICAL RESULTS

In this section, we present numerical examples that illustrate the gains that can be obtained by using an optimized HQAM-BICM system. In particular, we analyze two practically relevant spectral efficiencies: 1 bit/dimension and 1.5 bit/dimension. We use a rate $R = 1/2$ optimum distance spectrum convolutional code with constraint length $K = 3$ and generator polynomial $(5,7)_8$. The decoding is based on the soft-input Viterbi algorithm without memory truncation, and the block length used for simulation is $N_{\rm c} = 24000$. In the following subsections, we use the names 2-PAM, 4-PAM, and 8-PAM to refer to the ES-PAM constellations.

The UB depends on the average SNR, the constellation parameter, the D-MUX configuration, and m (in fading channels). Therefore, to obtain the optimal design, one needs to jointly optimize $\boldsymbol{K}$ and $\alpha_1,\ldots,\alpha_{q-1}$ for each value of $\overline{\gamma}$ and m (in fading channels), i.e., the optimal design is obtained by selecting the constellation (via $\alpha_1,\ldots,\alpha_{q-1}$) and the MUX (via $\boldsymbol{K}$) that minimize the UB for a given m and $\overline{\gamma}$. The values that minimize the UB for a given $\overline{\gamma}$ are defined as $\alpha_1^*(\overline{\gamma}),\ldots,\alpha_{q-1}^*(\overline{\gamma})$ and $\boldsymbol{K}^*(\overline{\gamma})$, i.e.,

$$[\alpha_1^*(\overline{\gamma}),\ldots,\alpha_{q-1}^*(\overline{\gamma}),\boldsymbol{K}^*(\overline{\gamma})] = \underset{\alpha_1,\ldots,\alpha_{q-1},\boldsymbol{K}}{\operatorname{argmin}} \{\mathrm{UB}(\alpha_1,\ldots,\alpha_{q-1},\boldsymbol{K})\}, \quad (18)$$

where UB is a function of the constellation parameter $\alpha_1,\ldots,\alpha_{q-1}$ and the MUX $\boldsymbol{K}$.

Because of the nature of the UB, the bound is tight only for BER below certain value (typically 10^{-3} or 10^{-4}), and thus, the optimization will be valid only for BER below this limit. The optimization of the UB presented in the following subsections was carried out numerically via an exhaustive search over the valid range of constellation parameter, with a step size of 0.01. We note that the objective function (UB) is potentially non-convex and the optimization space is one- or two-dimensional. Because of this, an exhaustive seach is a feasible and robust alternative compared to other (more complex) optimization approaches.

A. Spectral efficiency 1 bit/dimension

For this particular case ($n = q = 2$) there is only one constellation parameter, i.e., α_1. From Fig. 3 (with $\alpha_2 = 0$), we observe three cases of particular interest: $\alpha_1 = 0$ gives a 2-PAM constellation, $\alpha_1 = 1/2$ gives a 4-PAM constellation, and $\alpha_1 = 1$ gives a three-point constellation. We consider a D-MUX with period $J = 2$ which result in only four different matrices:

$$\boldsymbol{K}^{(1)} = \begin{bmatrix} (1,1) & (1,2) \\ (2,1) & (2,2) \end{bmatrix}, \boldsymbol{K}^{(2)} = \begin{bmatrix} (1,1) & (2,1) \\ (1,2) & (2,2) \end{bmatrix},$$

$$\boldsymbol{K}^{(3)} = \begin{bmatrix} (2,1) & (1,1) \\ (1,2) & (2,2) \end{bmatrix}, \text{and } \boldsymbol{K}^{(4)} = \begin{bmatrix} (2,1) & (2,2) \\ (1,1) & (1,2) \end{bmatrix}.$$

$$\mathrm{UB}(\alpha_1,\ldots,\alpha_{q-1},\boldsymbol{K}) \approx \frac{1}{k_{\rm c}} \sum_{w=w^{\rm free}}^{\hat{w}} \sum_{\boldsymbol{w}\in\mathcal{W}_q(w)} \beta_{\boldsymbol{K}}(\mathbf{w}) \left[\pi\gamma \sum_{k=1}^{q} w_k \frac{\sum_{j=0}^{M_k-1} \xi_{k,j}\mu_{k,j} \exp\left(-\mu_{k,j}\gamma/4\right)}{\sum_{j=0}^{M_k-1} \xi_{k,j} \exp\left(-\mu_{k,j}\gamma/4\right)}\right]^{-1/2} \cdot \prod_{k=1}^{q} \left[\sum_{j=0}^{M_k-1} \xi_{k,j} \exp\left(-\mu_{k,j}\gamma/4\right)\right]^{w_k}. \quad (18)$$

$$\mathrm{UB}(\alpha_1,\ldots,\alpha_{q-1},\boldsymbol{K}) \approx \frac{1}{k_{\rm c}} \sum_{w=w^{\rm free}}^{\hat{w}} \sum_{\boldsymbol{w}\in\mathcal{W}_q(w)} \beta_{\boldsymbol{K}}(\mathbf{w}) \left[\pi\overline{\gamma} \sum_{k=1}^{q} w_k \frac{\sum_{j=0}^{M_k-1} \xi_{k,j}\mu_{k,j} \left(\frac{4m}{4m+\overline{\gamma}\mu_{k,j}}\right)^{(m+1)}}{\sum_{j=0}^{M_k-1} \xi_{k,j} \left(\frac{4m}{4m+\overline{\gamma}\mu_{k,j}}\right)^{m}}\right]^{-1/2} \prod_{k=1}^{q} \left[\sum_{j=0}^{M_k-1} \xi_{k,j} \left(\frac{4m}{4m+\overline{\gamma}\mu_{k,j}}\right)^{m}\right]^{w_k}. \quad (19)$$

For this specific example, the D-MUX configurations given by $\boldsymbol{K}^{(1)}$, and $\boldsymbol{K}^{(4)}$ are identical to the two possible R-MUX configurations [4], i.e., when all the coded bits from one encoder's output are assigned to one modulator's input.

In order to justify the requirement of joint optimization of the D-MUX configuration and the signal constellation, we first study the behavior of the UBs given by Theorem 2 and Theorem 3, respectively[2], as a function of the constellation parameter α_1. The UB is shown in Fig. 4 for the four different D-MUX configurations and different values of $\overline{\gamma}$. This figure shows that for a given constellation parameter, different D-MUX configurations give different BER performances. In particular, for all the 3 cases in Fig. 4, when $\alpha_1 = 1/2$ (4-PAM) is considered, the lowest UB performance is obtained $\boldsymbol{K} = \boldsymbol{K}^{(4)}$ (but this changes if another value of α is chosen). This is equivalent to the R-MUX when all the bits from $(7)_8$ are assigned to $k = 1$ and the coded bits from $(5)_8$ to $k = 2$, which was shown in [4] (only for the AWGN case). From this figure, it is also clear that the 4-PAM constellation is suboptimal, i.e., by selecting another value of α_1, gains are obtained. More particularly, for the AWGN case, by changing the value of α_1 from $\alpha_1 = 1/2$ to $\alpha_1 = 0.12$, the UB decreases from $\text{UB} \approx 0.8 \cdot 10^{-6}$ to $\text{UB} \approx 0.7 \cdot 10^{-7}$. The gains for the Nakagami-m fading channel are smaller but still visible.

We performed a numerical optimization over $\boldsymbol{K}$ and α_1 for $m = 1$, $m = 5$, $m = 20$, and for the AWGN channel for the values of $\overline{\gamma}$ that give a BER of interest ($\text{BER} \leq 10^{-3}$). For both the AWGN channel and Nakagami-m fading channels, The results obtained showed that in such a case, the optimum D-MUX is always given by $\boldsymbol{K}^*(\overline{\gamma}) = \boldsymbol{K}^{(4)}$. The values of $\alpha_1^*(\overline{\gamma})$ obtained in the optimization are shown in Fig. 5 (left). This results show that when the fading is severe, the optimal constellation is close to a 4-PAM constellation, and in fact does not depend much on the value of $\overline{\gamma}$. On the other hand, for the AWGN channel, the dependency on the average SNR is notable and a 4-PAM constellation is far from the optimum. In Fig. 5 (right), we show the behavior of the UB as a function of α_1 for a given average SNR $\overline{\gamma} = 9$ dB and different channel conditions. This figure shows how the value $\alpha_1^*(\overline{\gamma})$ evolves from $\alpha_1^*(\overline{\gamma}) \approx 0.5$ (for $m = 2$) to $\alpha_1^*(\overline{\gamma}) \approx 0.16$ for the AWGN case.

We conclude this subsection by presenting the BER performance obtained by using the proposed HQAM-BICM system, where the constellation and the D-MUX are optimized for each SNR. The results are presented in Fig. 6, where we also show the results obtained by the conventional BICM-S system of [2], and the one studied in [4] (R-MUX), both of them using a 4-PAM constellation. The results in this figure confirm the tightness of the UBs developed in this paper (for $\text{BER} \leq 10^{-3}$)[3]. They also confirm that a joint optimization of the D-MUX and the constellation outperforms the previous designs. More particularly, for a BER target of 10^{-7}, the obtained gains are approximately 1 dB for the AWGN channel and 0.4 dB for $m = 5$. These gains decrease when the fading in the channel increases, and for $m = 1$ (Rayleigh fading channel), they are marginal.

[2]All the results presented in this subsection were obtained using $\hat{w} = 125$ for the AWGN channel and $\hat{w} = 30$ for Nakagami-m fading channels, cf. (13).

[3]We note a slight mismatch between the simulations and the bounds for $m = 1$ (for the three configurations shown in Fig. 6). We conjecture this is caused by the approximation used to model the PDF of the L-values, cf. Sec. III. A similar mismatch is observed in Fig. 8, cf. Sec. V-B.

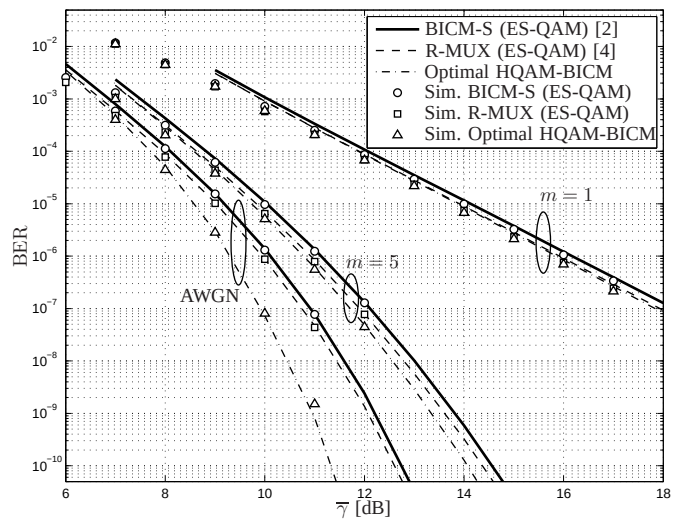

Fig. 6. UBs (lines) for $n = q = 2$ (1 bit/dimension), for Nakagami-m fading channels and the AWGN channel given by Theorem 2 and Theorem 3, respectively. Numerical simulations are also included (markers). The proposed HQAM-BICM system uses the optimal $\boldsymbol{K}^*(\overline{\gamma}) = \boldsymbol{K}^{(4)}$ and $\alpha_1^*(\overline{\gamma})$ shown in Fig. 5 (left). The BICM-S system of [2] with a 4-PAM constellation and the BICM system with R-MUX and a 4-PAM constellation of [4] are shown for comparison.

B. Spectral efficiency 1.5 bit/dimension

We consider 8-ary constellations ($q = 3$), which together with the rate $R = 1/2$ code ($n = 2$) gives a spectral efficiency of 1.5 bits/dimension. In this case, the constellation is defined by the pair (α_1, α_2). From Fig. 3, we see that the 8-PAM constellation is obtained with $\alpha_2 = 1/4$ and $\alpha_1 = 1/2$ and that 4-ary constellations are obtained with $\alpha_2 = 0$ (4-PAM with $\alpha_1 = 1/2$). We consider D-MUX configurations with the shortest possible period, i.e., $J = 3$, for which there will be a total of thirty different D-MUX configurations. The UBs presented in this subsection are evaluated for $\hat{w} = 30$ for both fading and the AWGN channel.

For the AWGN channel, and an average SNR $\overline{\gamma}$ [dB] $\in \{10, 11, \ldots, 15\}$ (which give a UB below 10^{-3}), we obtained the optimal D-MUX configuration $\boldsymbol{K}^*(\overline{\gamma})$ and constellation parameters $(\alpha_1^*(\overline{\gamma}), \alpha_2^*(\overline{\gamma}))$ using Theorem 2 and an exhaustive search. The optimal matrix for all $\overline{\gamma}$ [dB] $\in \{10, 11, \ldots, 15\}$ was found to be

$$\boldsymbol{K}^*(\overline{\gamma}) = \begin{bmatrix} (1,1) & (2,1) & (3,1) \\ (3,2) & (2,2) & (1,2) \end{bmatrix} \tag{19}$$

and the optimal constellation parameters are

$$(\boldsymbol{\alpha}_1^*(\overline{\gamma}), \boldsymbol{\alpha}_2^*(\overline{\gamma})) = [(0.46, 0)\,(0.45, 0)\,(0.44, 0)\,(0.43, 0)\,(0.43, 0)\,(0.43, 0)]. \tag{20}$$

The results in (20) indicate that the optimal constellation is a 4-ary constellation ($\alpha_2 = 0$) with $\alpha_1 \approx 0.45$, which translates

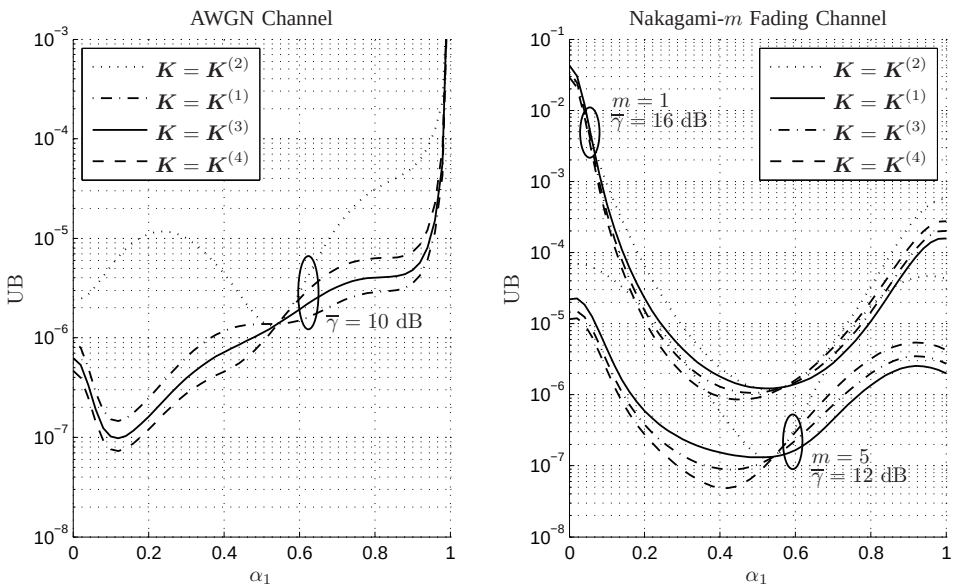

Fig. 4. UB versus the constellation parameter α_1 for the AWGN channel (left) and for Nakagami-m fading channels (right) for different D-MUX configurations and different values of $\overline{\gamma}$. The UBs are given by Theorem 2 and Theorem 3, respectively.

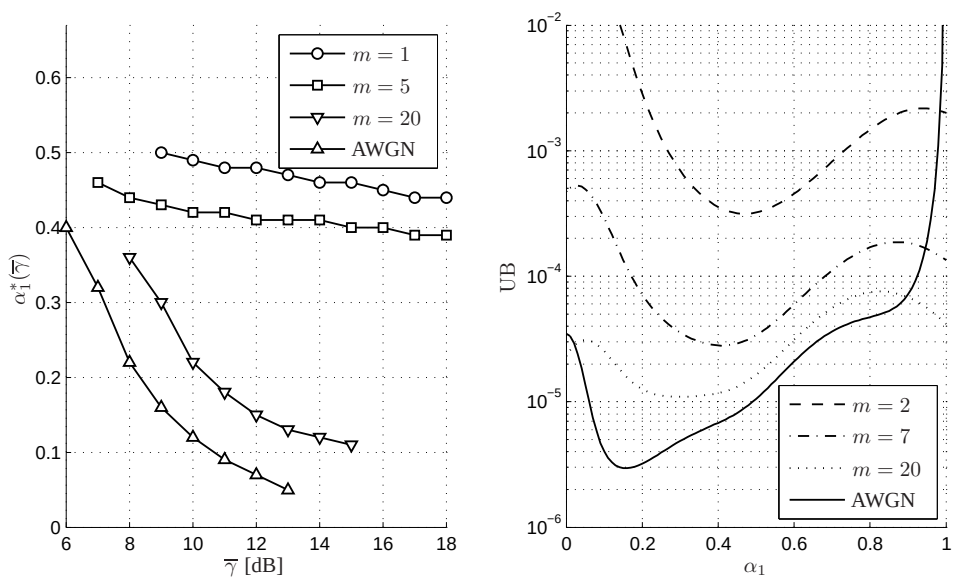

Fig. 5. The optimal constellation parameter $\alpha_1^*(\overline{\gamma})$ versus average SNR $\overline{\gamma}$ for $m = 1$, $m = 5$, $m = 20$, and for the AWGN channel (left) and the UB versus α_1 for $\overline{\gamma} = 9$ dB and $m = 2$, $m = 7$, $m = 20$ and the AWGN channel (right). The UBs are given by Theorem 2 and Theorem 3, respectively.

into a system where the third output of the D-MUX (cf. (19)) is completely eliminated.

Another way of interpreting the results in (19)–(20) is that for this code, the minimum BER is obtained when the original rate 1/2 code is punctured (giving a rate $R = 3/4$) and transmitted with a 4-ary constellation (with $\alpha_1 \approx 0.45$), and a puncturing pattern given by

$$\boldsymbol{P} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \tag{21}$$

where following the notation of [22], the columns of $\boldsymbol{P}$ have a meaning of time and a 0 denotes a puncture.

An intuitive explanation of the previous results is the following. In a coded modulation system, and for high SNR values, there is a trade-off between the minimum Euclidian distance of the constellation and the minimum Hamming distance of the code. By puncturing this code, its minimum Hamming distance $w^{\text{free}} = 5$ will decrease. On the other hand, by reducing the constellation size (from 8-ary to 4-ary), the minimum Euclidian distance increases. For this particular code, the improvement due to an increased minimum Euclidian distance is larger than the degradation due to a decrease minimum Hamming distance, and thus, the optimal solution is given by (19)–(20). It is important to mention that, that in general, a reduction of the constellation size (by completely

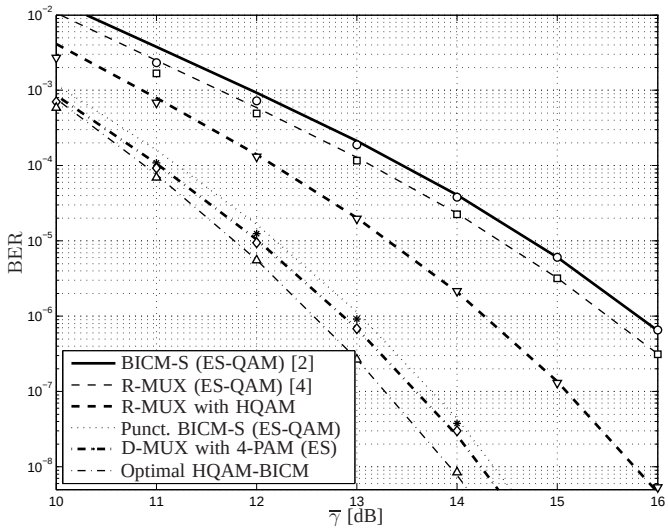

Fig. 7. UB (lines) for $n = 2$ and $q = 3$ (1.5 bit/dimension) for the AWGN channel given by Theorem 2. The simulation results are shown with markers. The proposed HQAM-BICM system uses the optimized parameters in (19)–(20). The BICM-S system of [2] with an 8-PAM constellation and the BICM system with R-MUX and an 8-PAM constellation of [4] are shown for comparison. The R-MUX system of [4] with optimized constellation and the BICM-S system with a (punctured) code with $R = 3/4$ as well as a BICM system with the optimal D-MUX and a 4-PAM (ES) are also shown.

puncturing some output of the encoder) does not necessarily lead to the optimal solution. The solution depends on the spectral efficiency targeted, the SNR under consideration, and the constraint length of the code.

In Fig. 7, we present the results obtained using the proposed system based on the optimum parameters in (19)–(20) and we compare them against five different BICM designs. The first one is the BICM-S system of [2] with an 8-PAM constellation and the second one the BICM system with R-MUX and an 8-PAM constellation of [4][4]. The other three will be explained below. When comparing the proposed system and the BICM-S system of [2], gains of about 3 dB are observed for a BER target of 10^{-6}. The gains compared to the system in [4] are about 2.75 dB.

The performance difference between the proposed system and the BICM system with R-MUX and an 8-PAM constellation of [4] are quite large (2.75 dB). However, the comparison is unfair since our system allows HPAM constellations while the results for the system in [4] are given for an 8-PAM constellation. In order to make a fair comparison, we have optimized the constellation for the system with R-MUX of [4], i.e., we selected the optimum constellation for each average SNR[5]. The obtained results are shown in Fig. 7, where we see that the performance improves, however, the gap to the entirely optimized design we proposed stays at 1.7 dB for a BER of 10^{-6}. We should not be surprised by this results as the HPAM constellation after optimization degenerates to a 4-PAM constellation. Consequently, the resulting assignment of the bits to the removed modulator's input (i.e., puncturing) is random, which is not the approach that should be used when doing puncturing [22].

Since the puncturing seems to appear as the solution in the case we study, we analyze also a BICM-S system with a 4-PAM constellation and a puncturing pattern that yields a rate $R = 3/4$. This configuration is simply another way of obtaining an spectral efficiency of 1.5 bit/dimension using the same encoder and decoder. While this may seem a "natural" solution for the BICM design, we note that in general, it cannot be assumed a priori that the puncturing combined with low-order modulation is the best way to transmit with a high spectral efficiency. We performed an exhaustive search over puncturing patterns (with a puncturing period of three bits) and found that for the SNR of interest, the optimal puncturing pattern is the one given by $\boldsymbol{P}$ in (21). The results are shown in Fig. 7. In this case, and for a target BER of 10^{-6}, the proposed system still offers gains of about 0.4 dB, which results uniquely from the fact that we use HQAM and D-MUX.

To complete the analysis of the AWGN channel, in Fig. 7 we also include the results obtained by a BICM system with the D-MUX in (19) and an equally spaced 4-PAM constellation (generated from an 8-PAM constellation with $\alpha_1 = 1/2$ and $\alpha_2 = 0$). The rationale behind presenting these results is to try to quantify the gains offered by the D-MUX compared to a joint optimization of the D-MUX and the HQAM constellation ("Optimal HQAM-BICM" in Fig. 7). The results in Fig. 7 show that approximately one third of the gap between "Best punctured BICM-S (ES-QAM)" and "Optimal HQAM-BICM" comes from the use of the D-MUX and two thirds from the use of the optimized HQAM constellation.

Now we turn our attention to Nakagami-m fading channels. Since the UB in Theorem 3 depends on $\boldsymbol{K}$, (α_1, α_2), m, and $\overline{\gamma}$, in general, the optimization must be done jointly over all these parameters. However, we have observed that for a given value of m, the optimal constellation and D-MUX do not change significantly for the SNR range of interest. Motivated by this observation, we have found the optimal constellation for an average SNR that gives a BER of approximately 10^{-7}, and we have used these values for all the range of average SNR. The obtained values are

$$\boldsymbol{K}^*_{m=1}(\overline{\gamma}) = \boldsymbol{K}^*_{m=2}(\overline{\gamma}) = \begin{bmatrix} (2,1) & (3,1) & (3,2) \\ (2,2) & (1,1) & (1,2) \end{bmatrix}, \quad (22)$$

$$\boldsymbol{K}^*_{m=5}(\overline{\gamma}) = \begin{bmatrix} (1,1) & (2,1) & (3,1) \\ (3,2) & (2,2) & (1,2) \end{bmatrix}, \quad (23)$$

and

$$(\alpha_1^*(\overline{\gamma}), \alpha_2^*(\overline{\gamma}))|_{m=1} = (0.48, 0.20) \quad (24)$$

$$(\alpha_1^*(\overline{\gamma}), \alpha_2^*(\overline{\gamma}))|_{m=2} = (0.47, 0.17) \quad (25)$$

$$(\alpha_1^*(\overline{\gamma}), \alpha_2^*(\overline{\gamma}))|_{m=5} = (0.42, 0.01). \quad (26)$$

We note that by selecting one set of parameters for the range of average SNR and a given m is relevant from a practical point of view. This is simply because in practice it would be more difficult to change the constellation parameters and the MUX for each value of $\overline{\gamma}$.

[4]The optimum R-MUX for this code is such that 2/3 of the coded bits form the second encoder's output are sent to $k = 1$, 1/3 of the bits from the first and the second encoder's outpus are sent to $k = 2$, and 2/3 of the bits from the first encoder's output to $k = 3$.

[5]We obtained $\alpha_1^*(\overline{\gamma}) = 0.49$, and $\alpha_2^*(\overline{\gamma}) = 0.0$ for all $\overline{\gamma}$ [dB] $\in \{10, 11, \ldots, 16\}$.

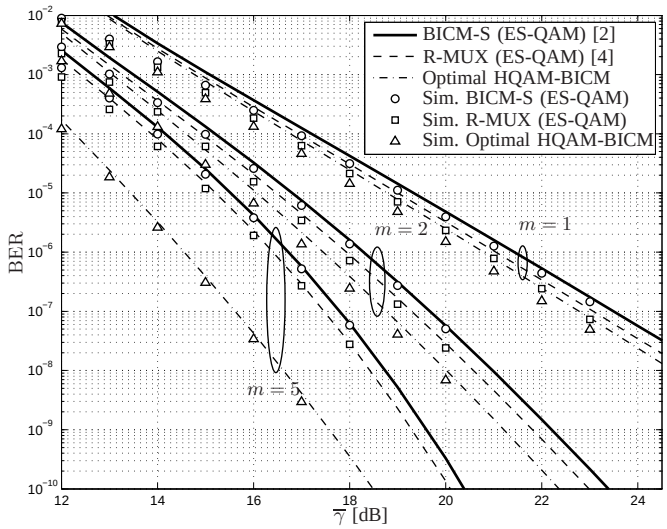

Fig. 8. UB (lines) for $n = 2$ and $q = 3$ (1.5 bit/dimension) for Nakagami-m fading channels given by Theorem 3 for $m = 1, 2, 5$. Numerical simulations are also included (markers). The simulations for the proposed HQAM-BICM system uses the values in (22)–(26). The BICM-S system of [2] with an 8-PAM constellation and the BICM system with R-MUX and an 8-PAM constellation of [4] are shown for comparison.

In Fig. 8 we present the simulated BER obtained by HQAM-BICM using (22)–(26) and the UB with optimized parameters for each SNR. We compare the results obtained by the proposed system against two previous BICM designs (as in Fig. 6): the BICM-S system of [2] with an 8-PAM constellation and the BICM system with an R-MUX in [4]. From this figure we observe that in fading channels, the proposed HQAM-BICM again outperforms previous BICM designs. When compared to BICM-S, the proposed system offers gains up to 2 dB for $m = 5$ and a BER target of 10^{-7}. The gains compared to the configuration in [4] are less than when compared to BICM-S, but still quite large. This figure also shows that the achievable gains increase when the fading is less severe (m increases).

VI. CONCLUSIONS

In this paper we proposed and studied a new BICM transmission framework that uses HQAM constellations in conjunction with a deterministic bit-level multiplexer and M-interleavers. It was shown that a number of degrees of freedom can be exploited, which in turn gives performance improvements of a few decibels compared to previous BICM designs. The gains were shown to depend on the fading parameter, the BER target, and the spectral efficiency, and in general, they increase when the fading is less severe.

There are a number of degrees of freedom that can be exploited in BICM transmission which may improve its performance even further. In particular, in this paper we only studied HQAM constellations labeled by the BRGC. The performance of BICM with other binary labelings and fully asymmetric constellation is still unknown. Moreover, the period of the MUX gives another degree of freedom not fully exploited in this paper (only short periods were considered).

In this paper, we analyzed a new BICM configuration based on convolutional codes. The gains offered by a similar BICM system using capacity approaching codes (turbo or low-density parity-check) are still unknown. The analysis for such a system is indeed interesting but left for further investigation.

APPENDIX A
PROOF OF THEOREM 1

The Laplace transform of the PDF of the decision variable $D(\boldsymbol{w})$ in (15) is

$$\Phi_{D(\boldsymbol{w})}(s;\overline{\gamma}) = \prod_{k=1}^{q} \left[\Phi_{L_k}(s;\overline{\gamma})\right]^{w_k}, \tag{27}$$

and its cumulant transform $\kappa_{D(\boldsymbol{w})}(s;\overline{\gamma})$ is

$$\begin{aligned}\kappa_{D(\boldsymbol{w})}(s;\overline{\gamma}) &= \log[\Phi_{D(\boldsymbol{w})}(s;\overline{\gamma})] \\ &= \sum_{k=1}^{q} w_k \log\left[\Phi_{L_k}(s;\overline{\gamma})\right]. \end{aligned}\tag{28}$$

The PEP can be approximated using the saddlepoint approximation [36] as

$$\mathrm{PEP}(\boldsymbol{w};\overline{\gamma}) \approx \frac{1}{\hat{s}\sqrt{2\pi\kappa''_{D(\boldsymbol{w})}(\hat{s};\overline{\gamma})}} \exp(\kappa_{D(\boldsymbol{w})}(\hat{s};\overline{\gamma})), \tag{29}$$

where $\hat{s}$ is the saddlepoint that can be found by solving $\kappa'_{D(\boldsymbol{w})}(\hat{s};\overline{\gamma}) = 0$. $\kappa''_{D(\boldsymbol{w})}(\hat{s};\overline{\gamma})$ is the second derivative of the cummulant generating function evaluated at the saddlepoint $\hat{s}$ and can be expressed as

$$\kappa''_{D(\boldsymbol{w})}(\hat{s};\overline{\gamma}) = \sum_{k=1}^{q} w_k \left[\frac{\Phi''_{L_k}(\hat{s};\overline{\gamma})}{\Phi_{L_k}(\hat{s};\overline{\gamma})}\right]. \tag{30}$$

Using (28) and (30) in (29) completes the proof.

APPENDIX B
PROOF OF THEOREM 2

The two-sided Laplace transform of $p_{L_k}(\lambda;\gamma)$ in (12), can be written as

$$\Phi_{L_k}(s;\gamma) = \int_{-\infty}^{\infty} \mathrm{e}^{-s\lambda} \sum_{j=0}^{M_k-1} \xi_{k,j}\psi(\lambda;\mu_{k,j}\gamma, 2|\mu_{k,j}|\gamma)\,\mathrm{d}\lambda \tag{31}$$

$$= \sum_{j=0}^{M_k-1} \xi_{k,j} \exp\left(\mu_{k,j}\gamma(s^2 - s)\right), \tag{32}$$

where to pass from (31) to (32) we used the transform pair $\mathcal{N}(\lambda;\mu,2\mu) \Leftrightarrow \exp(\mu(s^2 - s))$. The saddlepoint can be found by solving $\kappa'_{D(\boldsymbol{w})}(\hat{s};\overline{\gamma}) = 0$ (equivalently, $\Phi'_{L_k}(s;\gamma) = 0$) , which from (32) gives $\hat{s} = 1/2$.

From (32), the second derivative of $\Phi_{L_k}(s;\gamma)$ with respect to s at the saddlepoint ($\hat{s} = 1/2$) is

$$\Phi''_{L_k}(1/2;\gamma) = \sum_{j=0}^{M_k-1} 2\xi_{k,j}\mu_{k,j}\gamma \exp\left(-\mu_{k,j}\gamma/4\right). \tag{33}$$

Substituting (32) and (33) in (17), the PEP over the AWGN channel is given by (34) (shown at the bottom of the page). Using (34) in (13) completes the proof.

APPENDIX C
PROOF OF THEOREM 3

Due to the linearity property of the Laplace transform, we have

$$\Phi_{L_k}(s;\overline{\gamma}) = \int_0^\infty p_\Gamma(\gamma;\overline{\gamma})\Phi_{L_k}(s;\gamma)\,\mathrm{d}\gamma. \tag{35}$$

Using (6) and (32) in (35), we obtain

$$\Phi_{L_k}(s;\overline{\gamma}) = \sum_{j=0}^{M_k-1} \xi_{k,j}\left(\frac{m}{m-\overline{\gamma}\mu_{k,j}(s^2-s)}\right)^m. \tag{36}$$

From (36), the saddlepoint is $\hat{s} = 1/2$, and the second derivative of $\Phi_{L_k}(s;\overline{\gamma})$ with respect to s at the saddlepoint can be written as

$$\Phi''_{L_k}(1/2;\overline{\gamma}) = \sum_{j=0}^{M_k-1} 2\xi_{k,j}\overline{\gamma}\mu_{k,j}\left(\frac{4m}{4m+\overline{\gamma}\mu_{k,j}}\right)^{(m+1)}. \tag{37}$$

Substituting (36) and (37) in (17), the PEP over Nakagami-m fading is given by (38) (shown at the bottom of the page). Using (38) in (13) completes the proof.

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] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 1–2, pp. 1–153, 2008.

[4] 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.

[5] 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.

[6] D. Sommer and G. P. Fettweis, "Signal shaping by non-uniform QAM for AWGN channels and applications using turbo coding," in *International ITG Conference on Source and Channel Coding (SCC)*, Munich, Germany, Jan. 2000.

[7] S. Y. L. Goff, "Signal constellations for bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 307–313, Jan. 2003.

[8] M. Barsoum, C. Jones, and M. Fitz, "Constellation design via capacity maximization," in *IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 2007.

[9] A. Guillén i Fàbregas and A. Martinez, "Bit-interleaved coded modulation with shaping," in *IEEE Information Theory Workshop (ITW)*, Dublin, Ireland, Aug.–Sep. 2010.

[10] E. Agrell and A. Alvarado, "Optimal signal sets and binary labelings for BICM at low SNR," *IEEE Trans. Inf. Theory*, 2011 (to appear), available at http://arxiv.org/abs/1001.4548.

[11] 3GPP, "Universal mobile telecommunications system (UMTS); multiplexing and channel coding (FDD)," 3GPP, Tech. Rep. TS 125.212, V7.11.0 Release 7, Sep. 2009.

[12] IEEE 802.11, "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: High-speed physical layer in the 5GHz band," IEEE Std 802.11a-1999(R2003), Tech. Rep., July 1999.

[13] IEEE 802.11n, "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. Amendment 5: Enhancements for higher throughout," IEEE Std 802.11n-2009, Tech. Rep., Oct. 2009.

[14] P. K. Vitthaladevuni and M.-S. Alouini, "A recursive algorithm for the exact BER computation of generalized hierarchical QAM constellations," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 297–307, Jan. 2003.

[15] M. Morimoto, M. Okada, and S. Komaki, "A hierarchical image transmission system in fading channel," in *IEEE International Conference on Universal Personal Communications (ICUPC)*, Tokyo, Japan, Oct. 1995.

[16] Md. J. Hossain, M.-S. Alouini, and V. K. Bhargava, "Hierarchical constellation for multi-resolution data transmission over block fading channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 849–857, Apr. 2006.

[17] Md. J. Hossain, P. K. Vitthaladevuni, M.-S. Alouini, V. K. Bhargava, and A. J. Goldsmith, "Adaptive hierarchical constellations for simultaneous voice and multi-class data transmission over fading channels," *IEEE Trans. Veh. Technol.*, vol. 55, no. 4, pp. 1181–1194, July 2006.

[18] M. R. Chari, F. Ling, A. Mantravadi, R. Krishnamoorthi, R. Vijayan, G. K. Walker, and R. Chandhok, "FLO physical layer: an overview," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 107–145, Mar. 2007.

[19] ETSI, "Digital video broadcasting (DVB); Frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2)," ETSI, Tech. Rep. ETSI EN 302 755 V1.1.1 (2009-09), Sep. 2009.

[20] ——, "Digital video broadcasting (DVB); framing structure, channel coding and modulation for digital terrestrial television," ETSI, Tech. Rep. ETSI EN 300 744 V1.6.1 (2009-01), Jan. 2009.

[21] 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.

[22] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, no. 4, pp. 389–400, Apr. 1988.

[23] 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.

[24] 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.

[25] M. Nakagami, "The m-distribution—A general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*, W. G. Hoffman, Ed. Oxford, U.K.: Pergamon, 1960.

[26] M. K. Simon and M. S. Alouini, *Digital Communications over fading channels: A unified approach to perfomance analysis*, 1st ed. New York, NY: John Wiley & Sons, 2000.

[27] A. Kenarsari-Anhari and L. Lampe, "An analytical approach for performance evaluation of BICM over Nakagami-m fading channels," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1090–1101, Apr. 2010.

[28] A. J. Viterbi, "An intuitive justification and a simplified implementation of the MAP decoder for convolutional codes," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 260–264, Feb. 1998.

[29] M. Benjillali, L. Szczecinski, and S. Aissa, "Probability density functions of logarithmic likelihood ratios in rectangular QAM," in *Twenty-Third Biennial Symposium on Communications*, Kingston, ON, Canada, May 2006.

[30] K. Hyun and D. Yoon, "Bit metric generation for Gray coded QAM signals," *IEE Proc.-Commun.*, vol. 152, no. 6, pp. 1134–1138, Dec. 2005.

$$\mathrm{PEP}(\boldsymbol{w};\gamma) = \left[\pi\gamma\sum_{k=1}^{q} w_k \frac{\sum_{j=0}^{M_k-1}\xi_{k,j}\mu_{k,j}\exp\left(-\mu_{k,j}\gamma/4\right)}{\sum_{j=0}^{M_k-1}\xi_{k,j}\exp\left(-\mu_{k,j}\gamma/4\right)}\right]^{-1/2} \prod_{k=1}^{q}\left[\sum_{j=0}^{M_k-1}\xi_{k,j}\exp\left(-\mu_{k,j}\gamma/4\right)\right]^{w_k}. \tag{34}$$

$$\mathrm{PEP}(\boldsymbol{w};\overline{\gamma}) = \left[\pi\overline{\gamma}\sum_{k=1}^{q} w_k \frac{\sum_{j=0}^{M_k-1}\xi_{k,j}\mu_{k,j}\left(\frac{4m}{4m+\overline{\gamma}\mu_{k,j}}\right)^{(m+1)}}{\sum_{j=0}^{M_k-1}\xi_{k,j}\left(\frac{4m}{4m+\overline{\gamma}\mu_{k,j}}\right)^{m}}\right]^{-1/2} \prod_{k=1}^{q}\left[\sum_{j=0}^{M_k-1}\xi_{k,j}\left(\frac{4m}{4m+\overline{\gamma}\mu_{k,j}}\right)^{m}\right]^{w_k}. \tag{38}$$

[31] M. S. Raju, R. Annavajjala, and A. Chockalingam, "BER analysis of QAM on fading channels with transmit diversity," *IEEE Trans. Wireless Commun.*, vol. 5, no. 3, pp. 481–486, Mar. 2006.
[32] L. Szczecinski, R. Bettancourt, and R. Feick, "Probability density function of reliability metrics in BICM with arbitrary modulation: Closed-form through algorithmic approach," *IEEE Trans. Commun.*, vol. 56, no. 5, pp. 736–742, May 2008.
[33] L. Szczecinski, A. Alvarado, and R. Feick, "Distribution of max-log metrics for QAM-based BICM in fading channels," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2558–2563, Sep. 2009.
[34] A. Alvarado, E. Agrell, L. Szczecinski, and A. Svensson, "Unequal error protection in BICM with QAM constellations: Interleaver and code design," in *IEEE International Conference on Communications (ICC)*, Dresden, Germany, June 2009.
[35] 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.
[36] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "Error probability analysis of bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 262–271, Jan. 2006.

Md Jahangir Hossain (S'04, M'08) received the B.Sc. degree in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh where he served as a Lecturer. He received his Ph.D. degree from the University of British Columbia, Vancouver, Canada in October 2007 and his M.A.Sc. degree from the University of Victoria, Victoria, Canada, in August 2003. He was a Research Fellow at Institut national de la recherche scientifique, Montreal, Canada and at the Institute for Telecommunications Research, University of South Australia, Mawson Lakes, Australia. Currently, he is working as an Assistant Professor at the School of Engineering, University of British Columbia, Okanagan, Canada.

Dr. Hossain held Natural Science Engineering and Research Council's Postdoctoral Fellowship at McGill University, Montreal, Canada. He also received the University of British Columbia Graduate Fellowship: Hui memorial Fellowship (2004-2006 Canada), the University of Victoria Fellowship (2002-2003 Canada), and Petch Research Award (2003 Canada). His industrial experiences include a senior Systems Engineer position at Redline Communications, Canada and a research Intern position at Communication Technology Lab, Intel, Inc., USA. His research interests include, bit-interleaved coded modulation, spectrally and power efficient modulation schemes, cognitive radio and cooperative communication systems. He is serving as an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He was also a member of the Technical Program Committee of the IEEE International Conference on Communications (ICC' 2010).

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.