*(Article begins on next page)*

# On BICM-ID with Multiple Interleavers

Alex Alvarado *Student Member, IEEE*, Leszek Szczecinski, *Senior Member, IEEE*, Erik Agrell, and Arne Svensson, *Fellow, IEEE*

***Abstract*—In this letter, we study the performance of BICM-ID with multiple interleavers (BICM-ID-M) in terms of bit-error rate (BER), and show that BICM-ID-M is well-suited to exploit the unequal error protection (UEP) caused by the binary labeling. We show that BICM-ID-M should *always* be the preferred alternative for BICM-ID and that the gains obtained appear even for the simplest configuration ($0.5$–$0.75$ dB for a BER of $10^{-7}$). It is found that conventional design paradigms such as maximizing the free distance of the code should be modified.**

***Index Terms*—BICM, BICM-ID, M-interleavers, UEP.**

## I. Introduction

BIT-INTERLEAVED coded modulation (BICM) was introduced in [1], analyzed in [2], [3], and is nowadays the preferred alternative for CM over the Gaussian and fading channels [3, Sec. 1]. Its flexibility makes it very attractive and it has made its way into a large number of communication standards [3, Sec. 1]. By recognizing BICM as a serial concatenation of codes, *BICM with iterative decoding* (BICM-ID) was introduced in [4]–[6]. BICM-ID exhibits a waterfall and an error floor region, and it has been well studied in the literature, cf. [7] and references therein. In BICM-ID, the binary labeling plays a key role and its optimization usually targets a decrease of the BER in the error floor region.

The original papers introducing BICM [1] and BICM-ID [4] postulated the application of multiple interleavers (M-interleavers) connecting each of the encoder's output to one modulator's input. However, most of the existing literature on BICM and BICM-ID follows the framework set in [2] and assumes the use of one single interleaver (S-interleavers). BICM with M-interleavers were analyzed in [8] and shown to offer gains when the modulation introduces UEP.

In this letter, we study the error floor of BICM-ID-M. We prove that BICM-ID-M asymptotically *always* outperforms BICM-ID with S-interleavers (BICM-ID-S) and that the gains obtained by using BICM-ID-M instead of BICM-ID-S appear even for the simplest configuration. We show that conventional design paradigms for the encoder, e.g., the use of optimum distance spectrum (ODS) codes [9], should be modified.

Manuscript received March 26, 2010. The associate editor coordinating the review of this letter and approving it for publication was Vladimir Stankovic.

A. Alvarado, E. Agrell, and A. Svensson are with the Department of Signals and Systems, Communication Systems Group, Chalmers University of Technology, Gothenburg, Sweden, e-mail: {alexa,agrell,arnes}@chalmers.se.

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.

This work was partially supported by the European Commission under projects NEWCOM++ (216715) and FP7/2007-2013 (236068), by the Swedish Research Council, Sweden (2006-5599), and by the FQRNT (Merit Scholarship Program for Foreign Students), Canada.

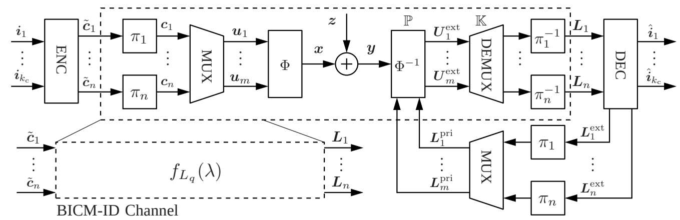


Fig. 1. BICM-ID model. The equivalent BICM channel is also shown.

TABLE I
The three most relevant configurations defined by $\mathbb{K}$.

| MUX | Configuration obtained | Analyzed in |
|---|---|---|
| $\mathbb{K} = m^{-1}\mathbb{1}_m$ | S-interleavers | [2] |
| $\mathbb{K} = \mathbb{I}_m$ | Original M-interleavers | [1], [4] |
| $\mathbb{K} = \Pi(\mathbb{I}_m)$ | Optimized M-interleavers | [8] |

## II. System Model and Preliminaries

### A. System Model

The BICM-ID system model is presented in Fig. 1, which can be considered a generalization of [8, Sec. II-A]. The vectors of information bits $\boldsymbol{i}_s$ with $s = 1, \ldots, k_c$ are encoded by a rate $R_c = k_c/n$ convolutional encoder. The vectors of coded bits $\tilde{\boldsymbol{c}}_i$ with $i = 1, \ldots, n$ are fed to the interleavers $\pi_1, \ldots, \pi_n$ which give statistically independent randomly permuted sequences $\boldsymbol{c}_i = \pi_i(\tilde{\boldsymbol{c}}_i)$. The multiplexing unit (MUX) assigns the bits $\boldsymbol{c}_i$ to the different bit positions in the symbol. We define the MUX using an $n \times m$ matrix $\mathbb{K}$, whose elements, $0 \le k_{iq} \le 1$, represent the probability that a bit from $\boldsymbol{c}_i$ is assigned to the $q$th output $\boldsymbol{u}_q$ with $q = 1, \ldots, m$. For simplicity, and since we are interested in the original BICM(-ID) configuration(s), from now on, we only consider $n = m$ and $k_{iq} \in \{0, 1\}$. The three most relevant configurations in this case are shown in Table I, where $\mathbb{1}_m$ and $\mathbb{I}_m$ are the all-ones and the identity matrices, respectively, and where $\Pi(\cdot)$ is a row permutation (see more details in [8]).

At any time instant $t$, the codeword $\boldsymbol{b} = [u_{1,t}, \ldots, u_{m,t}]$ is mapped to a complex constellation symbol $x \in \mathcal{X}$ using a memoryless mapping $\Phi : \{0,1\}^m \to \mathcal{X}$, where $\mathcal{X} \subset \mathbb{C}$, and $|\mathcal{X}| = 2^m$. We analyze phase shift keying ($M$-PSK) modulation, i.e., $x_j = \exp\left(-\frac{2\pi j\sqrt{-1}}{M}\right)$ with $j = 1, \ldots, M$; extension to other modulations is straightforward. The symbols are transmitted through an AWGN channel $y = x + z$, where $z$ is a circularly-symmetric complex Gaussian random variable with zero-mean and variance $N_0/2$ in real/imaginary parts. The bit energy-to-noise ratio is $\frac{E_b}{N_0} = \frac{1}{N_0 k_c}$. The demapper computes extrinsic L-values as $U_q^{\text{ext}}(y) = U_q^{\text{pos}}(y) - L_q^{\text{pri}}$, where $U_q^{\text{pos}}(y) \triangleq \log \frac{\Pr\{b_q=0|y\}}{\Pr\{b_q=1|y\}}$ and $L_q^{\text{pri}} \triangleq \log \frac{\Pr\{b_q=0\}}{\Pr\{b_q=1\}}$ are the a posteriori and a priori L-values for $b_q$, respectively.

Let $\mathcal{D} \triangleq \{d_1, \ldots, d_D\}$ be the set of squared Euclidean distances between the constellation points, where $D = \frac{M}{2}$ for $M$-PSK. For example, for 8-PSK $D = 4$, $d_1 = 0.58$, $d_2 = 2$, $d_3 = 3.41$, and $d_4 = 4$. The generalized Euclidean distance spectrum (GEDS) of a constellation (see also [7, Ch. 4]) is defined by the $m \times D$ matrix $\mathbb{P}$ whose entries $p_{ql}$ are the number of pairs (normalized by $M/2$) of constellation points at distance $d_l$ such that their binary labelings differ in all the bit position except in the $q$th one. We also define the generalized minimum Euclidean distance (GMED) of the constellation by $d_q^{\min}$ with $q = 1, \ldots, m$, which corresponds to the squared Euclidean distance associated to the first nonzero element in the $q$th row of $\mathbb{P}$. For example, for 4-PSK ($D = 2$, $d_1 = 2$, $d_2 = 4$) only two labelings with different GEDS exist: the Gray code (GC) and the anti-Gray code (AGC) [7, App. A]. The GEDS of the AGC is given by $p_{1,1} = p_{2,2} = 1$, $p_{1,2} = p_{2,1} = 0$, $d_1^{\min} = 2$ and $d_2^{\min} = 4$, and for the GC by $p_{1,1} = p_{2,1} = 1$, $p_{1,2} = p_{2,2} = 0$, $d_1^{\min} = d_2^{\min} = 2$.

### B. Perfect Feedback and the BICM-ID Channel

We use the so-called perfect feedback (PF) assumption to analyze the error-floor region. This assumption states that after a certain number of iterations, and for a sufficiently high signal-to-noise ratio (SNR), it can be assumed that the a-priori L-values are large enough so that the demapper knows all the bits except the one for which it is calculating the extrinsic L-value. This transforms the detection of the high-order modulation into the detection of binary symbols, and thus, the extrinsic L-values calculated by the demapper $\Phi^{-1}$ can be shown to be Gaussian-distributed [7, Ch. 4].

For a given transmitted symbol $x$ labeled by $\boldsymbol{b} = [b_1, \ldots, b_m]$, it can be shown that $U_q^{\mathrm{ext}}(y|x) \sim \mathcal{N}(\mu, 2|\mu|)$, where $\mu = N_0^{-1}(-1)^{b_q} d$ and where $d \in \mathcal{D}$. Therefore, there exist $D$ Gaussian distributions that can be used to model the extrinsic L-values, where $d$ depends on the transmitted symbol and the bit position, i.e., $d = d_l$ only if $p_{ql} \neq 0$. The probability density function (pdf) of $L_q$ is then given by

$$f_{L_q}(\lambda) = \sum_{l=1}^{D} g_{ql} \Phi(\lambda; \mu_l, 2\mu_l), \qquad (1)$$

where $\Phi(\lambda; \mu, \sigma^2)$ is a Gaussian function, $g_{ql}$ is the $(q, l)$th entry of the $m \times D$ matrix $\mathbb{G} \triangleq \mathbb{K}\mathbb{P}$ which represents the probability that the $q$th L-value is Gaussian distributed with parameters $(\mu_l, 2\mu_l)$, and where $\mu_l = N_0^{-1} d_l$ (assuming $b_q = 0$). Expression (1) states that the L-values passed to the decoder (cf. the output of the BICM-ID channel in Fig. 1) are modeled using a Gaussian mixture, where the structure of the matrix $\mathbb{K}$ determines the weights $g_{ql}$ of the Gaussian mixture in (1). Using (1), we replace the BICM-ID channel by a symmetric binary-input soft-output memoryless channel as shown in Fig. 1.

### C. Union Bound

Let $\beta^{\mathcal{C}}(\boldsymbol{w})$ be the generalized weight distribution spectrum of a convolutional encoder, where the *generalized weight* $\boldsymbol{w} = [w_1, \ldots, w_m]$ gathers the weights $w_q$ of each of the encoder's outputs, and where $\beta^{\mathcal{C}}(\boldsymbol{w})$ can be calculated as described in [8, Sec. III-A]. The (truncated) union bound (UB) is given by

$$\mathrm{BER} \leq \mathrm{UB} = \sum_{\boldsymbol{w} \in \mathcal{W}} \beta^{\mathcal{C}}(\boldsymbol{w}) \mathrm{PEP}(\boldsymbol{w}), \qquad (2)$$

where $\mathcal{W} \triangleq \{\boldsymbol{w} \in (\mathbb{N}_0)^m : \omega^{\mathrm{free}} \leq \sum_{q=1}^{m} w_q \leq \hat{\omega}\}$, $\mathbb{N}_0$ is the set of nonnegative integers, $w^{\mathrm{free}}$ is the free distance of the code, $\hat{w}$ is the truncation of the UB, and $\mathrm{PEP}(\boldsymbol{w})$ is the probability of detecting a sequence with generalized weight $\boldsymbol{w}$ instead of the transmitted all-zero sequence.

For a given $\boldsymbol{w}$, the decision variable passed to the decoder is $S(\boldsymbol{w}) = S(w_1) + \ldots + S(w_m)$ where $S(w_q) = \sum_{i=1}^{w_q} L_q^{(i)}$, and where $L_q^{(i)}$ are i.i.d. random variables with a pdf given by (1). Let $\binom{w_q}{\boldsymbol{r}_q} \triangleq \frac{w_q!}{r_{q1}! \ldots r_{qD}!}$ be the multinomial coefficients which represents the number of different ways of ordering $w_q$ bits in subsets of $r_{q1}, \ldots, r_{qD}$ elements, where $\boldsymbol{r}_q \triangleq [r_{q1}, \ldots, r_{qD}]$.

*Theorem 1:* The pdf of $S(\boldsymbol{w})$ can be expressed as

$$f_{S(\boldsymbol{w})}(\lambda) = \sum_{\mathbb{R} \in \mathcal{R}(\boldsymbol{w})} \prod_{q=1}^{m} \binom{w_q}{\boldsymbol{r}_q} \prod_{l=1}^{D} (g_{ql})^{r_{ql}} \Phi\left(\lambda; \Delta(\mathbb{R}), 2\Delta(\mathbb{R})\right),$$

where $\mathbb{R} \triangleq [\boldsymbol{r}_1^{\mathrm{T}}, \ldots, \boldsymbol{r}_m^{\mathrm{T}}]^{\mathrm{T}}$, $\Delta(\mathbb{R}) \triangleq N_0^{-1} \sum_{q=1}^{m} \sum_{l=1}^{D} d_l r_{ql}$, $\mathcal{R}(\boldsymbol{w}) \triangleq \{\mathbb{R} \in \mathbb{N}_0^{m \times D} : \sum_{l=1}^{D} r_{ql} = w_q, q = 1, \ldots, m\}$, and we interpret $0^0$ as 1.

*Proof:* Because of the interleaving, the L-values are independent, and thus, the pdf of $S(w_q)$ is the convolution of $w_q$ copies of the Gaussian mixture in (1), i.e.,

$$f_{S(w_q)}(\lambda) = \sum_{\boldsymbol{r}_q \in \mathcal{V}(w_q)} \binom{w_q}{\boldsymbol{r}_q} \prod_{l=1}^{D} (g_{ql})^{r_{ql}} \Phi\left(\lambda; \delta_q, 2\delta_q\right),$$

where $\mathcal{V}(w_q) \triangleq \{\boldsymbol{r}_q \in (\mathbb{N}_0)^D : \sum_{l=1}^{D} r_{ql} = w_q\}$, where the $l$th element in $\boldsymbol{r}_q$ represents the number of bits transmitted using the $l$th Gaussian distribution and $\delta_q = N_0^{-1} \sum_{l=1}^{D} d_l r_{ql}$. The pdf of $S(\boldsymbol{w})$ is obtained by convolving the densities $f_{S(w_q)}(\lambda)$, $q = 1, \ldots, m$, which completes the proof. ∎

By computing $\mathrm{PEP}(\boldsymbol{w})$ as the tail integral of the pdf given by Theorem 1, the following UB expression is obtained.

*Corollary 2:* The UB in (2) can be expressed as

$$\mathrm{UB} = \sum_{\boldsymbol{w} \in \mathcal{W}'} \sum_{\mathbb{R} \in \mathcal{R}'(\boldsymbol{w})} W^{\mathcal{C}}(\mathbb{R}, \mathbb{G}, \boldsymbol{w}) \mathrm{Q}\left(\sqrt{\Delta(\mathbb{R})/2}\right). \qquad (3)$$

In (3) $\mathcal{W}' \triangleq \{\boldsymbol{w} \in \mathcal{W} : \beta^{\mathcal{C}}(\boldsymbol{w}) \neq 0\}$, $\mathcal{R}'(\boldsymbol{w}) \triangleq \{\mathbb{R} \in \mathcal{R}(\boldsymbol{w}) : (g_{ql})^{r_{ql}} \neq 0 \ \forall q, l\}$, and

$$W^{\mathcal{C}}(\mathbb{R}, \mathbb{G}, \boldsymbol{w}) = \beta^{\mathcal{C}}(\boldsymbol{w}) \prod_{q=1}^{m} \binom{w_q}{\boldsymbol{r}_q} \prod_{l=1}^{D} (g_{ql})^{r_{ql}}. \qquad (4)$$

The definitions of $\mathcal{W}'$ and $\mathcal{R}'(\boldsymbol{w})$ guarantee $W^{\mathcal{C}}(\mathbb{R}, \mathbb{G}, \boldsymbol{w}) \neq 0$ for any $\boldsymbol{w} \in \mathcal{W}'$ and $\mathbb{R} \in \mathcal{R}'(\boldsymbol{w})$. Clearly, the multiplexing $\mathbb{K}$ affects only the inner product in (4).

## III. Main Results and Conclusions

### A. BICM-ID-M with 4-PSK

In Fig. 2 we show the BER performance of BICM-ID-M and BICM-ID-S for one of the simplest configurations one could think of, i.e., 4-PSK with the AGC and a rate $R_{\mathrm{c}} = 1/2$ ODS

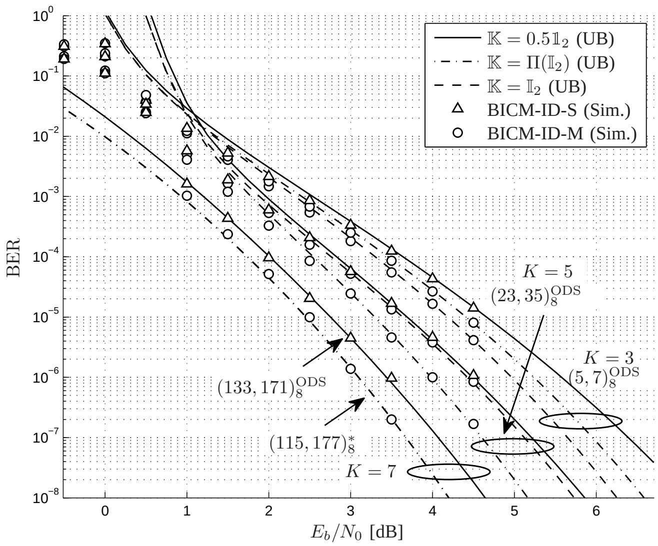


Fig. 2. BER performance of BICM-ID-M and BICM-ID-S. The simulation results are shown with markers and the bound in (3) with lines.

convolutional code [9] of constraint lengths $K = 3, 5$ (the results for $K = 7$ will be discussed in Sec. III-C). The bound in (3) is shown to agree with the simulations results, and gains of 0.5–0.75 dB are obtained for a BER of $10^{-7}$ if $\mathbb{K}$ is properly selected. We note that the optimum $\mathbb{K}$ depends on the code, and that for each $K$, the two BICM-ID-M configurations give a lower BER than BICM-ID-S. In the following subsection, we will prove that this is asymptotically always the case.

### B. Optimality of BICM-ID-M

The UB given by Corollary 2 is a sum of weighted Q-functions. We are interested in the behavior of (3) for high SNR, and thus, the arguments of the Q-functions become relevant. We consider constellations with a GEDS such that $d_q^{\min} \neq d_{q'}^{\min}$ for some $q, q'$, i.e., constellations that introduce UEP (e.g., 4-PSK with the AGC). We define $\overline{d}$ as the smallest element in the GMED of the constellation, i.e., $\overline{d} \triangleq \min_{q\in\{1,\ldots,m\}}\{d_q^{\min}\}$.

*Lemma 3:* The arguments of the dominant Q-functions in the UB (3) for a given $\boldsymbol{w} \in \mathcal{W}'$ are

$$\Delta_{\mathrm{M}}^* \triangleq N_0^{-1}\sum_{q=1}^{m} d_q^{\min} w_q, \quad \Delta_{\mathrm{S}}^* \triangleq N_0^{-1}\overline{d}\sum_{q=1}^{m} w_q,$$

for BICM-ID-M and BICM-ID-S, respectively.

*Proof:* For BICM-ID-M, $\mathbb{G}_{\mathrm{M}} = \mathbb{P}$, and therefore, the solution of $\min_{\mathbb{R}\in\mathcal{R}'(\boldsymbol{w})}\{\Delta_{\mathrm{M}}(\mathbb{R})\}$ is obtained when $\mathbb{R}$ is such that all the $w_q$ bits are transmitted using the Gaussian distribution associated to $d_q^{\min}$, $\forall q$. With this, we obtain the expression for $\Delta_{\mathrm{M}}^*$, which holds for any $\mathbb{G}'_{\mathrm{M}} = \Pi(\mathbb{G}_{\mathrm{M}})$. For BICM-ID-S, $\mathbb{G}_{\mathrm{S}} = m^{-1}\mathbb{1}_m\mathbb{P}$. This matrix has a first column with a nonzero entry determined by $\overline{d}$. Moreover, all the elements in this column are identical (and nonzero), and therefore, the solution of $\min_{\mathbb{R}\in\mathcal{R}'(\boldsymbol{w})}\{\Delta_{\mathrm{S}}(\mathbb{R})\}$ is obtained when all the $w_q$ bits are transmitted using the Gaussian distribution associated to $\overline{d}$, $\forall q$. Using this, we obtain the expression for $\Delta_{\mathrm{S}}^*$, which concludes the proof. ■

*Corollary 4:* For high SNR and a given code $\mathcal{C}$, the UB for BICM-ID-M is always smaller than the UB for BICM-ID-S.

The proof of Corollary 4 follows directly from the inequality $\Delta_{\mathrm{M}}^* > \Delta_{\mathrm{S}}^*$ which holds for any $\boldsymbol{w}$. Corollary 4 states that, for high SNR, BICM-ID-M should always be preferred over BICM-ID-S, even if the MUX is not optimized. This conclusion does not hold for (noniterative) BICM, cf. [8].

### C. Optimal Convolutional Codes

Corollary 2 allows us to express the asymptotic behavior of the UB for the pair $[\mathcal{C}, \mathbb{K}]$ as $\mathrm{UB} \approx \mathrm{Q}\left(\sqrt{\frac{A^{\mathrm{d}}}{2N_0}}\right)$, where $A^{\mathrm{d}}$ is the argument of the dominant Q-function in the UB, i.e., the smallest $\Delta(\mathbb{R})$ for the pair $[\mathcal{C}, \mathbb{K}]$. In the following, we define the optimum convolutional codes (OCC).

*Definition 1 (OCC for BICM-ID):* A convolutional code $\mathcal{C}^*$ is said to be optimal if there exists a $\mathbb{K}^*$ such that, among all the other codes with the same constraint length and MUX configurations, the pair $[\mathcal{C}^*, \mathbb{K}^*]$ gives the largest $A^{\mathrm{d}}$.

Definition 1 considers both the MUX and the code as one entity, and does not assume $\mathcal{C}^*$ to belong to the set of codes with maximum free distance, which we denote by $\omega_{\max}^{\mathrm{free}}$. An exhaustive search showed that for $K = 5, 6$ ($\omega_{\max}^{\mathrm{free}} = 7, 8$) there exist many codes with $\omega^{\mathrm{free}} = \omega_{\max}^{\mathrm{free}} - 1$ that perform equally good as the ODS codes, i.e., they give the same $A^{\mathrm{d}}$. For $K = 7, 9$ ($\omega_{\max}^{\mathrm{free}} = 10, 12$), this happens for codes with $\omega^{\mathrm{free}} = \omega_{\max}^{\mathrm{free}} - 2$, which shows that maximizing $\omega^{\mathrm{free}}$ is not the criterion that defines optimal codes in this scenario.

The OCCs are defined asymptotically, which does not assure their optimality for a finite SNR. Alternatively, we can use (3) for a given SNR and search for a good pair $[\mathcal{C}, \mathbb{K}]$. As an example, we performed an exhaustive search for the optimal $[\mathcal{C}^*, \mathbb{K}^*]$ at $\frac{E_{\mathrm{b}}}{N_0} = 3.5$ dB for $K = 7$ and $\hat{\omega} = \omega^{\mathrm{free}} + 5$, cf. (2). We found the code $(115, 177)_8$ ($\omega^{\mathrm{free}} = 8$) and $\mathbb{K}^* = \Pi(\mathbb{I}_2)$ to be optimal. Its performance is presented in Fig. 2. Gains of 0.5 dB for $\mathrm{BER} = 10^{-6}$ are obtained when compared with the most common configuration, i.e., BICM-ID-S and the ODS code with $\omega^{\mathrm{free}} = 10$.

## 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] 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.

[5] S. ten Brink, J. Speidel, and R.-H. Yan, "Iterative demapping for QPSK modulation," *IEE Electronics Letters*, vol. 34, no. 15, pp. 1459–1460, July 1998.

[6] S. Benedetto, G. Montorsi, D. Divsalar, and F. Pollara, "Soft-input soft-output modules for the construction and distributed iterative decoding of code networks," *Eur. Trans. on Telecommun.*, vol. 9, no. 2, pp. 155–172, Mar.–Apr. 1998.

[7] F. Schreckenbach, "Iterative decoding of bit-interleaved coded modulation," Ph.D. dissertation, Munich University of Technology, Munich, Germany, 2007.

[8] 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.

[9] 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.