On symbol mappings in bit interleaved coded modulation with iterative detection

Adrian Florin PĂUN*, Octavian FRATU*, Iancu CEAPĂ*

Key word. *Iterative detection, EXIT chart, Turbo receiver*

Abstract. *In this paper we investigate bit-interleaved coded modulation with iterative decoding (BICM-ID) for bandwidth efficient transmission. For this receiver scheme the bit error rate is reduced through iterations between a multilevel demapper and the channel decoder. The assignment strategy of the binary indices to signal points is crucial in order to achieve a significant turbo-gain. We address the problem of finding the most suitable index assignments to arbitrary, high order signal constellations. The method based on the binary switching algorithm is studied to find optimized symbol mapping.*

Cuvinte cheie. *Detecție iterativă, diagrama EXIT, receptor Turbo*

Rezumat. *În acest articol analizăm tehnica modulației codate cu întrețeserea biților şi detecție iterativă (BICM-ID) pentru transmisiuni eficiente de bandă limitată. Pentru această schemă de recepție, rata de eronare binară se reduce de-a lungul iterațiilor dintre blocul de detecție de simbol şi decodorul de canal. Modul prin care biții codați se alocă punctelor din constelația simbolurilor este foarte importantă pentru a obține un câştig semnificativ prin prelucrare iterativă. S-a abordat problema găsirii regulii cea mai potrivită pentru alocarea indicilor simbolurilor pentru o modulație multinivel arbitrară.*

Introduction

Bit-interleaved coded modulation (BICM) [1], [2] is the well suited technique for bandwidth efficient transmission over fading channels. This modulation scheme consist in the concatenation of an encoder, an interleaver and a symbol mapper. Similar to iterative decoding of serial concatenated codes (SCCC) [3], the performance BICM receiver can be improved through iterative information exchange between the *inner* decoder, i.e. the demapper, and the *outer* channel decoder. This system was introduced in [4], and is usually referred to as BICM with iterative decoding (BICM-ID). For multilevel modulation, the choice of the mapping (labeling map) is the crucial design parameter to achieve a high coding gain over the iterations. In this paper, we study a low complexity method to find labeling maps with desired characteristics for arbitrary signal constellations. The optimization scheme is based on the *binary switching algorithm* (BSA) with a suitable cost functions for BICM-ID based on mutual information and error bounds.

Section 2 introduces the system model. Then, in Section 3, a *distance spectrum* for mappings and the EXIT charts [5] are used to characterize mappings. The BSA and appropriate cost functions are investigated in Section 4.

The analysis and simulation results of the new mappings show the performance gains of the optimization with the BSA in Section 5.

* „Politehnica" University, Bucharest.

System model

We consider the BICM-ID system depicted in figure 1. A block of data bits ***u*** is encoded by a convolutional encoder and bit-interleaved by the block interleaver Π. From the coded and interleaved sequence, **c'**, *m* consecutive bits are grouped to form the subsequences $\mathbf{c}'_k = \left[c'_k(1), c'_k(2) \ldots c'_k(m)\right]$. Each subsequence $\mathbf{c}'_k$ is mapped to a complex symbol $x_k = s_n = \mu(\mathbf{c}'_k)$, chosen from the 2^m-ary signal constellation *S* according to the labeling map *μ*.

The emitted symbol energy is set to unity, E_S=1.

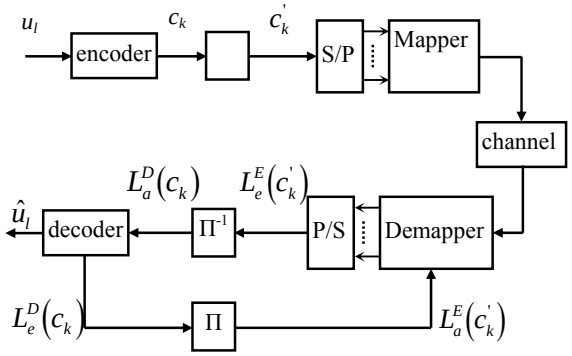


Fig. 1. Turbo demapper principle.

The channel is described by $y_k = a_k \cdot x_k + w_k$, where a_k denotes the fading coefficient, and w_k is the complex zero mean Gaussian noise with variance $\sigma_w^2 = N_0/2E_s$ in each real dimension.

At the receiver, the demapper processes the received complex symbols y_k and the corresponding a priori log-likelihood ratios (LLRs) of the coded bits and outputs the extrinsic LLRs [6]:

$$L_e(c'_k(i)) = \log\left(\frac{P\left(c'_k(i)=1 \middle| y_k, L_a(c'_k(i))\right)}{P\left(c'_k(i)=0 \middle| y_k, L_a(c'_k(i))\right)}\right) - L_a(c'_k(i)) \tag{1}$$

Let S_b^i denote the subset of symbols $s_n \in S$ whose bit labels have the value $b \in \{0,1\}$ in position *i*. Assuming an ideal bits interleaver, using Bayes' rule and taking the expectation of $p(y_k|x_k)$ over $P(s_n|c'_k(i)=b)$, $s_n \in S_b^i$ yields

$$L_e(c'_k(i)) = \log\left(\frac{\sum_{s_n \in S_1^i} p(y_k|s_n) P(s_n|c'_k(i)=1)}{\sum_{s_n \in S_0^i} p(y_k|s_n) P(s_n|c'_k(i)=0)}\right) - L_a(c'_k(i)) \tag{2}$$

The second term in numerator and denominator of (2) represents the a-priori knowledge fed to the detector from the decoder. This can be computed using LLR values for symbols

$$L_k(s_n) = \log\left(P(x_k = s_n)/P(x_k = s_1)\right)$$

$$P(x_k = s_n) = P(x_k = s_1) \cdot \exp\left[L_k(s_n)\right] \tag{3}$$

where s_1 is the symbol with all bits set to zero. The symbols L-values can be computed using bits L-values $L_a(c'_k(i))$:

$$L_k(s_n) = \sum_{i=1}^{m} b^i(s_n) \cdot L_a(c'_k(i)), \quad b^i(s_n) \in \{0,1\} \tag{4}$$

The conditioned probability density in (2) is therefore given by the complex Gaussian distribution:

$$p(y_k|s_n) = \frac{1}{\pi N_0} \exp\left(-\frac{|y_k - a_k \cdot s_n|^2}{N_0}\right) \tag{5}$$

However, for the L-value computation only the second term is relevant – the constant scaling factor can thus be omitted.

To evaluate the numerator and denominator of (2) it is useful to recursively apply the so called "Jacobian Logarithm" or one of its approximation [7] :

$$\log\left(e^{a_1} + e^{a_2}\right) = \max(a_1, a_2) + \log\left(1 + e^{-|a_1 - a_2|}\right) \tag{6}$$

The extrinsic estimates $L_e(c'_k(i))$ are deinterleaved and applied to the APP channel decoder.

Performing iterative decoding, extrinsic information about the coded bits from the decoder is fed back and regarded as a priori information $L_a\left(c_k^{'}(i)\right)$ at the demapper. During the initial demapping step, the a priori LLRs are set to zero.

Characteristics of labeling maps

The applied labeling map is the crucial design parameter for the considered BICM-ID system. The labeling map can be characterized through a *distance spectrum* introduced in [8].

Similar to the distance spectrum of Hamming weights for a channel code, the distance spectrum of Euclidean distances is the average number of bit errors made at a specific Euclidean distance d_E.

$$N(d_E)=\frac{1}{m2^m}\sum_{s_n\in S}N_H(d_E,s_n) \qquad (7)$$

where $N_H(d_E,s_n)$ is the total Hamming distance between the symbol s_n and the symbols at Euclidian distance d_E from s_n .

For example, we will consider three characteristic 16-QAM mappings proposed in the literature and depicted in Fig.2: Gray, where symbols at minimum Euclidean distance differ in one bit, Modified Set Partitioning (MSP) [9] and Maximum Squared Euclidean Weight (MSEW) mapping [10].

The shaded regions in Fig. 2 correspond to the decision regions for bit *i* having the value 1 (subset S_1^i), the unshaded regions correspond to the decision regions for bit *i* = 1 having the value 0 (subset S_0^i). If a symbol error occurs within one decision region, no error will be made on the corresponding bit. Large decision regions provide a high protection for the corresponding bit, since the number of *nearest neighbors*, i.e. the number of symbols $\hat{s}_n\in S_{\bar{b}}^i$ at minimum Euclidean distance from $s_n\in S_b^i$ is minimized. In other words, the Hamming distances between symbols at small Euclidean distance should be minimized in order to minimize the number of bit errors for one symbol error.

Table 1 shows the distance spectrum in the case where *no a priori* information about the coded bits is available at the demapper (e.g. during the initial demapping step).

The performance is expected to be the best for Gray labeling, where the average number of nearest neighbors is minimized, followed by MSP and MSEW mapping.

The performance with *ideal a priori* information at the demapper represents the achievable gain over the iterations. In this case, all the bits are perfectly known at the demapper, except for the bit to be detected, since only the *extrinsic* information is used.

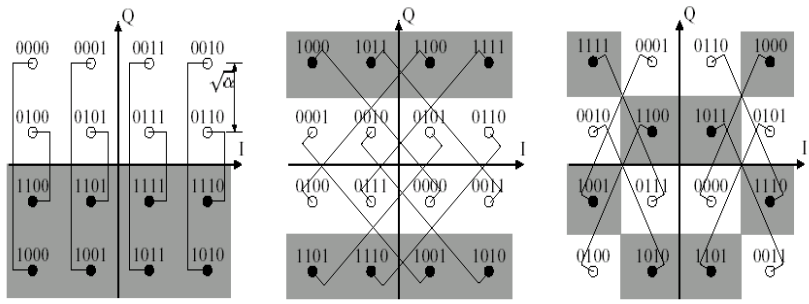
The a priori known bits select a pair of symbols which differ only in the bit *i* to be detected. Possible symbol pairs for bit *i* = 1 are shown in Fig. 2.

The distance spectrum, shown in Table 2. With Gray mapping, the number of distances at minimum Euclidean distance is not reduced through a priori knowledge, but is reduced at other Euclidian distances. Thus, only very small performance improvement is expected over the iterations. With MSP, designed in [9] to provide a good trade-off between performance with and without a priori information, the number of small distances is reduced and with MSEW mapping, the minimum squared Euclidean distance between the symbol pairs is maximized.

Table 1

Distance spectrum with no apriori information

d_E^2	$2E_S/5$	$4E_S/5$	$8E_S/5$	$10E_S/5$	$16E_S/5$	$18E_S/5$	$20E_S/5$	$26E_S/5$	$36E_S/5$
Gray	0.75	1.12	1	2.25	1	0.25	0.75	0.75	0.12
MSP	1.62	1.19	0.75	2.25	0.25	0.62	0.75	0.5	0.06
MSEW	2.25	1.12	1	0.75	1	0.75	0.75	0.25	0.12



(a) Gray (b) MSP (c) MSEW

Fig. 2. 16-QAM mappings with decision regions regions without a priori information for the first bit and by the ideal apriori information of the last 3 bits.

Table 2

Distance spectrum with ideal apriori information

d_E^2	$2E_S/5$	$4E_S/5$	$8E_S/5$	$10E_S/5$	$16E_S/5$	$18E_S/5$	$20E_S/5$	$26E_S/5$	$36E_S/5$
Gray	0.75	0	0	0	0	0.25	0	0	0
MSP	0	0.06	0.25	0.12	0.25	0	0.12	0.12	0.06
MSEW	0	0	0	0.75	0	0	0	0.25	0

Optimization of the index assignments

The goal is to have a method, where weights for the performance without and with ideal a priori knowledge can be set and the algorithm finds an optimal mapping for the selected weight distribution.

Those mappings could be found through exhaustive search, which becomes intractable higher order constellations since $2^m!$ different possibilities have to be checked.

To overcome these complexity problems it can be used binary switching algorithm (BSA)[9]. This algorithm finds a local optimum on a given cost function.We investigate two possible cost functions.

A. Binary switching algorithm

The binary switching algorithm (BSA) is started with an initial mapping. Using a suitable cost function defined in the next sections, the cost of each symbol and the total cost are calculated. An ordered list of symbols, sorted by decreasing costs, is generated. The idea is to pick the symbol with the highest cost in the list (which has the strongest contribution to a ”bad” performance), and to try to switch the index of this symbol with the index of another symbol. The latter is selected such, that the decrease of the total cost due to the switch is as large as possible. If no switch partner can be found

for the symbol with the highest cost, the symbol with the second-highest cost will be tried to switch next. This process continues for symbols in the list with decreasing costs until a symbol is found that allows a switch that lowers the total cost. After an accepted switch, a new ordered list of symbols is generated, and the algorithm continues as described above until no further reduction of the total cost is possible. Several algorithm executions with random initial mappings yield to the presumed global optimum, as the BSA finds a local optimum.

B. Cost function based on mutual information

One possible cost function for the BSA can be the mutual information $I(C,Y)\in[0,1]$ between the coded bits and the received channel. This mutual information, also used in the EXIT charts [5], can be evaluated by numerical integration over the signal space C.

$$I(C,Y)=\frac{1}{2m}\sum_{i=1}^{m}\sum_{b=0}^{1}\int_{\mathrm{C}}p(a)\int_{\mathrm{C}}p\left(y\middle|C(i)=b\right)\cdot \cdot\log_2\frac{2\,p\left(y\middle|C(i)=b\right)}{p\left(y\middle|C(i)=0\right)+p\left(y\middle|C(i)=1\right)}\,\mathrm{d}r\mathrm{d}a \tag{8}$$

with

$$p\left(y\middle|C(i)=b\right)=\frac{1}{2^{m-1}}\sum_{s\in S_b^i}p(y|s) \tag{9}$$

for independent and uniformly distributed code bits. For AWGN channels, the integration over the probability density function $p(a)$ of the fading coefficient a can be omitted as a = 1 and $p(y|s)$ is given by the Gaussian distribution (3). For the BSA cost function it is used the mutual information only for the case of ideal a priori information, where the subset S_b^i in the summation of () is reduced to one symbol

$$D^{I_r}=1-I(C,Y) \quad \text{and} \quad D^{I_a}=1-I(C,Y)\big|_{a=1} \tag{10}$$

for the Rayleigh and AWGN channel respectively.

C. Cost functions based on error bounds

For evitation of numerical integration computation in (8) the cost function can be definet using simple terms that characterize the influence of the mapping in error bounds. It is not necessary to use tight error bounds, but to have a *qualitative* measure to define costs for mappings.

Let $P(s_n\to\hat{s}_n)$ denote the probability of choosing the symbol $\hat{s}_n$ instead of the transmitted symbol s_n. In the general case of a Rician fading channel with Rice-Factor K, the Chernoff upper bound of $P(s_n\to\hat{s}_n)$ is given by [8]:

$$P(s_n\to\hat{s}_n)\le\frac{1+K}{1+K+\frac{E_S}{4N_0}|s_n-\hat{s}_n|^2}\cdot \cdot\exp\left(\frac{K\frac{E_S}{4N_0}|s_n-\hat{s}_n|^2}{1+K+\frac{E_S}{4N_0}|s_n-\hat{s}_n|^2}\right) \tag{11}$$

Let **c** and $\hat{\mathbf{c}}$ denote two coded bit sequences which differ in *d* consecutive positions. The bits c are transmitted within distinct symbols. $P(\mathbf{c}\to\hat{\mathbf{c}})$ is the *pairwise error probability* (PEP), i.e. the probability of choosing the sequence $\hat{\mathbf{c}}$ instead of the transmitted sequence **c**. Assuming perfect interleaving and averaging over all symbols and bit positions, the PEP of the two sequences **c** and $\hat{\mathbf{c}}$ is given by [2]:

$$P(\mathbf{c}\to\hat{\mathbf{c}})=\left(\frac{1}{m2^m}\sum_{i=1}^{m}\sum_{b=0}^{1}\sum_{s_k\in S_b^i}\sum_{s_k\in S_{\bar{b}}^i}P(s_n\to\hat{s}_n)\right)^d \tag{12}$$

For small values of *K* (e.g. *K*=0, Rayleigh fading) and high SNR $\left((1+K)<<E_s|s_n-\hat{s}_n|^2/(4N_0)\right)$, using

(10) and (11), the influence of the mapping on the PEP is described by

$$D^r = \frac{1}{m2^m} \sum_{i=1}^{m} \sum_{b=0}^{1} \sum_{s_k \in S_b^i} \sum_{s_k \in S_{\bar{b}}^i} \frac{1}{\left| s_n - \hat{s}_n \right|^2} \qquad (13)$$

The inverse of D^r is interpreted in [2] as the *harmonic mean* of the Euclidean distance.

For large values of K (e.g. $K \to \infty$, AWGN channel) or low SNR $\left((1+K) >> E_s \left| s_n - \hat{s}_n \right|^2 / (4N_0) \right)$, the influence of the mapping on the PEP is described by

$$D^a = \frac{1}{m2^m} \sum_{i=1}^{m} \sum_{b=0}^{1} \sum_{s_k \in S_b^i} \sum_{s_k \in S_{\bar{b}}^i} \exp\left(-\frac{E_s}{4N_0} \left| s_n - \hat{s}_n \right|^2 \right) \qquad (14)$$

Depending on the channel, D^r or D^a can be used as cost functions for the BSA. Separate costs for each symbol, required by the BSA, can be easily determined by considering only the selected symbol s_n in the summation in (12) and (13). We distinguish between the case without a priori knowledge (D_0^r, D_0^a) and with perfect a priori knowledge (D_1^r, D_1^a). In the latter case, the signal subsets S_b^i and $S_{\bar{b}}^i$ in the summations (12) and (13) are reduced to one symbol.

The weighted combination (15) has been proposed in [8] to be used as a cost function for the Rayleigh and AWGN channel respectively; λ_0 and λ_1 denote the weights for the performance without and with perfect a priori knowledge.

$$D^{E_r} = \lambda_0 D_0^r + \lambda_1 D_1^r \text{ and } D^{E_a} = \lambda_0 D_0^a + \lambda_1 D_1^a \qquad (15)$$

Simulation results

EXIT chart [5] is a powerful tool for analysis of the iterative exchange of mutual information between the demapper and the decoder.

Fig. 3 depicts the EXIT chart of iterative demapping and 4-state, rate-1/2 convolutional code APP decoding. It is used the three 16-QAM mappings presented in Section 3, over an AWGN channel at three SNR values. Let us denote the values of the demapper functions with no a priori knowledge and ideal a priori knowledge by I_0 and I_1 respectively. High values of I_0 and I_1 are desirable in order to avoid an early crossing of the transfer functions, which would cause the iterative process to stop, and to reach low error rates respectively. As expected, Gray mapping has the highest I_0, MSEW mapping the highest I_1, while MSP is a good trade off, allowing both high I_0 and I_1.

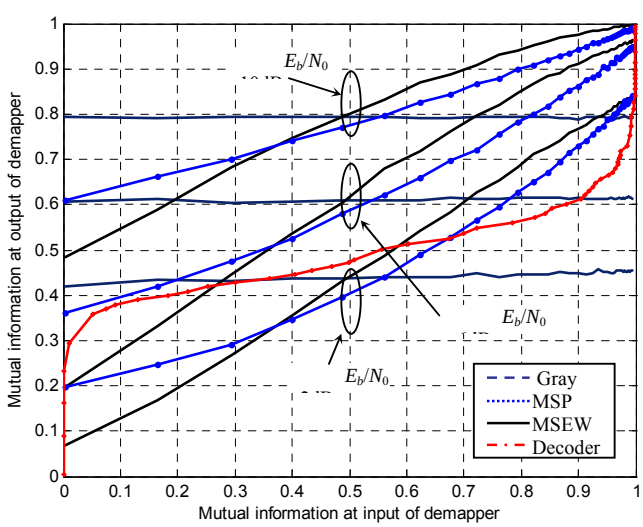


Fig. 3. Transfer functions of different mappings in the EXIT chart.

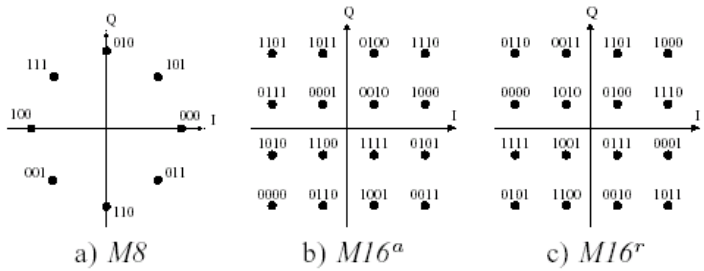


Fig. 4. Optimized mappings for ideal a priori information.

The cost functions D^{I_r}, D^{I_a} (10), D^{E_r} and D^{E_a} (15) can be determined for arbitrary signal constellations. The BSA can output for different initial mappings and cost functions different optimized mappings with the same distance spectrum.

Setting λ_0 = 1 and λ_1 = 0 in (15) maximizes the performance without a priori knowledge. As expected, the BSA outputs Gray or quasi-Gray mappings for all signal constellations. By setting λ_0 = 0 and λ_1 = 1 in (15) or using (10), the performance with ideal a priori information is maximized. For 8-PSK, all cost functions result in the same optimum mapping *M*8 shown in Fig. 4a). For 16-QAM, the BSA finds two new mappings shown in Fig. 4b) and 4c). $M16^a$ results from the cost functions D^{I_a} and D^{E_a} at high SNR. Due to the exponential cost decrease with the Euclidean distance and the SNR in AWGN channels, the minimum of the squared Euclidean distance is maximized. $M16^r$ results from the cost function D^{I_r}, D^{E_r} as well as from D^{I_a} and D^{E_a} at low SNR.

Table 3 and 4 show the distance spectrum in the case where *no a priori* and with *ideal a priori* information at the demapper, respectiv, for $M16^a$ and $M16^r$ mapping scheme'

The BER performance of the investigated 16-QAM mappings is shown in Fig. 7 after 1 and 10 iterations, where the convolutional code is a 4-state, rate-1/2 code, the channel is an AWGN channel, and the interleaver length is 8192 bits.

Table 3

Distance spectrum with ideal apriori information

d_E^2	$2E_S/5$	$4E_S/5$	$8E_S/5$	$10E_S/5$	$16E_S/5$	$18E_S/5$	$20E_S/5$	$26E_S/5$	$36E_S/5$
$M16^a$	1.75	1.31	1.25	1.25	0.5	0.75	0.75	0.25	0.19
$M16^r$	1.75	1.31	1	1.75	0.25	0.75	0.5	0.5	0.19

Table 4

Distance spectrum with ideal apriori information

d_E^2	$2E_S/5$	$4E_S/5$	$8E_S/5$	$10E_S/5$	$16E_S/5$	$18E_S/5$	$20E_S/5$	$26E_S/5$	$36E_S/5$
$M16^a$	0	0	0	0.5	0.12	0	0.12	0.25	0
$M16^r$	0	0	0.12	0.25	0.25	0	0.25	0.12	0

The MSEW mapping is outperformed other mapping at high SNR range by the optimized mappings (the greatest slope). The two optimized mappings $M16^a$ and $M16^r$ have similar perromnaces with MSEW, but present a lower convergence treshold. Since these mappings are only optimized for ideal a priori information, other mappings can converge at lower SNR

The MSP mapping seems to be a good tradeoff beteewn convergence threshold and performance at high SNR.

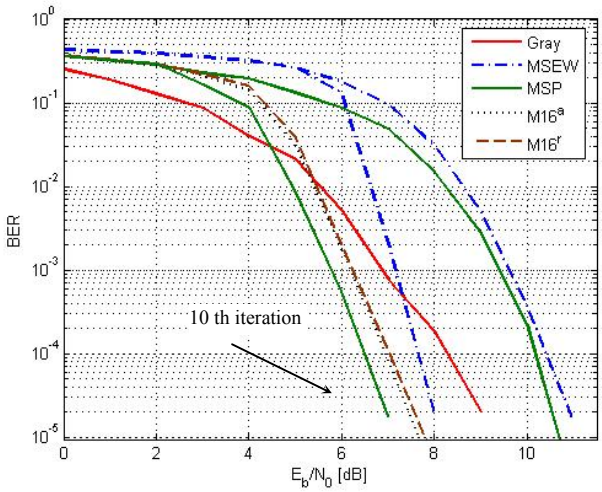


Fig. 5. BER of BICM-ID with 16-QAM mappings, AWGN channel, R=1/2, 4-state convolutional code.

Even though the $M16^r$ mapping is optimized for the fading channel, its performance is similar to the $M16^a$ mapping in the considered SNR range. All those relations are similar for a fading channel.

Conclusions

In this paper we investigare a low complexity method for finds optimized mappings for BICM-ID. This technique, based on a binary switching algorithm, proves its functionality for 8-PSK and 16-QAM modulation, but can be applied to any arbitrary signal constellation. Simulation results show that mappings optimized for fading channels are also well suited for AWGN

References

[1] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Transactions on Communications*, vol. 40, no. 5, pp. 873–884, May 1992

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

[3] G. Montorsi S. Benedetto, D. Divsalar and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 909–926, May 1998.

[4] 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.

[5] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Transactions on Communications*, vol. 49, no. 10, pp. 1727–1737, October 2001.

[6] J. Speidel S. ten Brink and R. Yan, "Iterative demapping and decoding for multilevel modulation," in *Proc. IEEE Globcom Conference*, Sydney, November 1998, pp. 579–584.

[7] A.F.Păun, C. Vlădeanu, Ş.G. Obreja, "Simplifications of Log-MAP algorithm for turbo decoder and turbo equalization", Telecomunicatii., vol 2, pp75-83,2006 .

[8] F. Schreckenbach, N. G¨ortz, J. Hagenauer G. Bauch, "Optimized Symbol Mappings for Bit-Interleaved Coded Modulation with Iterative Decoding" , IEEE International Conference Globecom 2003, San Francisco, Dec 2003

[9] A. Chindapol and J. Ritcey, "Design, analysis and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 5, pp. 944–957, May 2001.

[10] J. Tan and G.L. Stuber, "Analysis and design of interleaver mappings for iteratively decoded BICM," in *IEEE International Conference on Communications (ICC)*, New York, May 2002, vol. 3, pp. 1403–1407.