Frank Schreckenbach and Patrick Henkel
Institute for Communications Engineering (LNT)
Munich University of Technology (TUM)
80290 Munich, Germany
frank.schreckenbach@tum.de

Norbert Görtz
School of Engineering and Electronics
The University of Edinburgh, Scotland, UK

Gerhard Bauch
DoCoMo Communications Laboratories Europe
GmbH, 80687 Munich, Germany

XI NATIONAL SYMPOSIUM OF RADIO SCIENCES
Poznan, April 7 – 8, 2005

ANALYSIS AND DESIGN OF MAPPINGS FOR ITERATIVE DECODING OF BICM[1]

Abstract: **We consider bit-interleaved coded modulation with iterative decoding (BICM-ID) for bandwidth efficient transmission, where the bit error rate is reduced through iterations between a multilevel demapper and a simple channel decoder. The iterations lead to a performance gain only if a suitable mapping of the bits to the complex transmit symbols is chosen. We investigate properties of mappings and compare different methods to find mappings optimized for arbitrary, high order signal constellations.**

1. INTRODUCTION

We investigate the combination of baseband modulation and channel coding. Bit-interleaved coded modulation (BICM) is a widely used promising scheme. BICM was suggested by Zehavi [1] and profoundly analyzed by Caire et al. [2]. The main advantages of BICM are the maximized time diversity order obtained by bit-wise interleaving and the simple and flexible implementation possibilities.

BICM was originally presented as a concatenation of a convolutional code, a bit-wise interleaver and a mapper. Subsequently, several BICM systems with different iterative decoding strategies have been proposed to improve the performance. One strategy is to use a strong channel code, e.g. serial or parallel concatenated convolutional codes or a Low Density Parity Check (LDPC) code with iterative decoding instead of a simple convolutional code. Another strategy is to include the demapper in the iterative decoding process and to perform iterative decoding between the demapper as an inner code and a simple convolutional channel decoder as an outer code, similar to iterative decoding of serial concatenated codes [3]. This system is usually referred to as BICM with iterative decoding (BICM-ID) [4][5] and is an attractive low complexity and highly flexible solution that is well suited for a combination with e.g. iterative equalization or MIMO detection.

If very low bit error rates are required, an additional recursive rate-1 inner encoder should be used [3]. Promising inner code structures have been proposed e.g. in [6],[7] and [8]. We will focus on the low complexity BICM-ID system without additional inner code.

It was soon recognized that in the BICM-ID system, the choice of the bit to symbol mapping, i.e. the assignment of the binary indices to the complex signal points, is the crucial design parameter to achieve a high coding gain over the iterations. It is well known that a Gray mapping is optimum when no feedback from the channel decoder to the demapper is implemented. With Gray mapping, the bit labels associated to neighboring symbols differ only in one bit. Thus, the most probable symbol errors result in only one bit error. The question arises now which mapping is optimum *with* feedback. With increasing constellation order, an exhaustive computer search becomes intractable due to high complexity. Several mappings for BICM-ID were proposed e.g. in [9][5][10][11]. In [12], a Binary Switching Algorithm has been proposed to easily find labeling maps with optimized characteristics for arbitrary signal constellations and results are given for QAM signal constellations. An analysis of optimized PSK constellations can be found in [13].

In this paper, we will thoroughly review the characteristics of labeling maps and compare different approaches to find optimized mappings. Section 2 reviews the BICM-ID system model. Then, in Section 3, properties and a description method of mappings are investigated. Different optimization possibilities are compared in Section 4. Finally, simulation results are given in Section 5.

[1] This work was sponsored by NEWCOM and DoCoMo Communications Laboratories Europe GmbH

2. SYSTEM MODEL

We consider the bit-interleaved coded modulation BICM system depicted in Fig. 1.

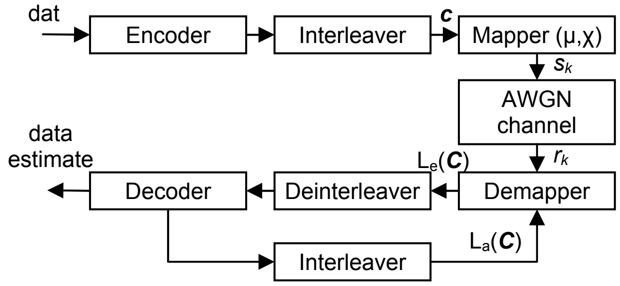


Fig. 1. System model.

A sequence of information bits is encoded by a binary code and bit-interleaved by a random interleaver Π. m consecutive bits of the coded and interleaved sequence $\mathbf{c}$ are grouped to form the subsequences $\mathbf{c}_k = (c_k[1],...,c_k[m])$. Each subsequence $\mathbf{c}_k$ is mapped to a complex symbol $s_k = \mu(\mathbf{c}_k)$ chosen from the 2^m-ary signal set χ according to the one-to-one binary labeling map $\mu : \{0,1\}^m \rightarrow \chi$.

We consider the transmission over an additive white Gaussian noise (AWGN) channel, for which the discrete-time received signal can be expressed as

$$r_k = s_k + n_k. \quad (1)$$

The noise samples n_k are independent and identically distributed (i.i.d.), with zero mean, complex-valued and with noise variance σ_n^2 for both real and imaginary part ($n_k \sim N_C(0,\sigma_n^2)$). The signal-to-noise ratio (SNR) at the receiver is $E_b/N_0 = E_s/(N_0 R m)$, where E_b is the average transmitted energy per information bit and N_0 is the noise power spectral density.

At the receiver, we use the log-likelihood ratio (LLR) notation [14]

$$L(C_k[i]) = \log(P(C_k[i]=0)/P(C_k[i]=1)) \quad (2)$$

to describe the estimated value and reliability of the bits $c_k[i]$. $C_k[i]$ denotes the binary random variable with realizations $c_k[i] \in \{0,1\}$.

The demapper performs APP detection and outputs the extrinsic LLRs $L_e(C_k[i])$:

$$L_e(C_k[i]) = \log \frac{\sum_{s_k \in \chi_0^i} p(r_k \mid s_k) \cdot \prod_{j=1, j\neq i}^{m} \mathrm{e}^{-L_a(C_k[j]) \cdot c_k[j]}}{\sum_{s_k \in \chi_1^i} \underbrace{p(r_k \mid s_k)}_{\text{channel}} \cdot \prod_{j=1, j\neq i}^{m} \underbrace{\mathrm{e}^{-L_a(C_k[j]) \cdot c_k[j]}}_{\text{a priori information}}}. \quad (3)$$

The sums in the numerator and denominator are taken over all symbols s_k whose bit labels have in position $i \in \{1,...,m\}$ the value $C_k[i]=0$ and $C_k[i]=1$, respectively. These subsets are depicted in the example of Fig. 2, where the shaded and unshaded regions include the symbols whose first bit in the binary label has the value 1 (subset χ_1^1) and value 0 (subset χ_0^1), respectively.

The first term $p(r_k \mid s_k)$ in (3) is given by the Gaussian channel transition probability for the AWGN channel:

$$p(r_k \mid s_k) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \cdot \exp\left(-\frac{|r_k - s_k|^2}{2\sigma_n^2}\right). \quad (4)$$

The second term in (3) describes the a priori information, where $L_a(C_k[i])$ denotes the a priori LLRs and $c_k[j]$ is the corresponding bit value at the j-th position of the bit label associated to the symbol s_k. During the initial demapping step or if the feedback from the channel decoder to the demapper is not implemented, the a priori LLRs $L_a(C_k[i])$ are equal to zero. Otherwise the extrinsic estimates of the coded bits from the decoder may be fed back and regarded as a priori information at the demapper.

This basic BICM system may be extended to e.g. iterative MIMO detection, where the demapper is extended to a MIMO detector or to iterative equalization, where the demapper is extended to an equalizer.

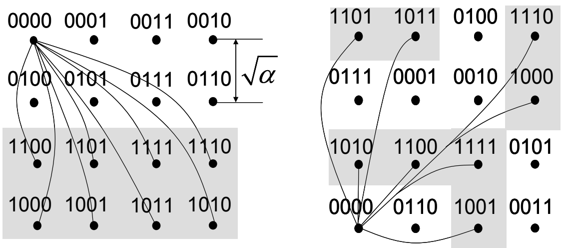
3. CHARACTERISTICS OF MAPPINGS

An exact way to characterize a binary labeling map is to use its Euclidean distance spectrum [15], similar to the characterization of a channel code through its Hamming distance spectrum. Let us define the three sets D_{ex}, D and Λ:

The set $D_{ex} = \{d_1^{ex},...,d_n^{ex}\}$ is defined as the set of all possible distinct (*ex*purgated) Euclidean distances between any two distinct signal points of the signal set χ. Let $D = \{d_1,...,d_l\} \supset D_{ex}$ denote the set of all (not necessarily distinct) Euclidean distances between the signal points s_k and $\hat{s}_k$, where $\hat{s}_k \in \chi_{\bar{b}}^i$ is defined as a symbol whose bit-label differs in the i-th bit position from the bit-label of $\hat{s}_k \in \chi_b^i$. Finally, let λ_j denote the frequency of the distance d_j^{ex} in the set D and $\Lambda = \{\lambda_1,...,\lambda_n\}$. D_{ex} depends only on the signal constellation, whereas D and Λ characterize the bit mapping.

The Euclidean distance spectrum lists the frequency λ_j of the Euclidean distances d_j^{ex}.

We consider as example several 16QAM mappings proposed in the literature: Gray [16], Set Partitioning (SP) [17], Modified Set Partitioning (MSP) [5], Maximum Squared Euclidean Weight (MSEW) [11], mappings optimized for error free feedback from the channel decoder, i.e. ideal a priori information (mapping M16^a for AWGN and M16^r for fading channel) [12] and a mapping with a good trade-off between performance without and with a priori information (I16) [18]. A Gray and the optimized M16^a mapping are shown in Fig. 2.



a) Gray b) Optimized (M16^a)

Fig. 2. 16QAM mappings. with example for distances $d_j \in D$ listed in the distance spectrum for the first bit and symbol with bit-label 0000. Shaded region: first bit = 1; Unshaded region: first bit = 0.

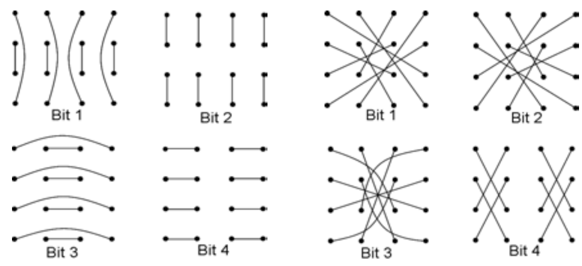
For the analysis of the mappings, we distinguish between the scenarios of no a priori and ideal a priori information at the demapper.

3.1. No a priori information at the demapper

During the initial demapping step or if the feedback from the channel decoder to the demapper is not implemented, the demapper has no a priori information. The distance spectrum with no a priori information is given in Table 1 and lists the frequency λ_j of the Euclidean distances d_j^{ex} in the set D. All the Euclidean distances between symbols where the i-th bit of the corresponding bit-label is different are considered. For example, for the signal point with bit label 0000 and the bit position $i=1$, the Euclidean distances to all signal points whose first bit is equal to 1 should be listed. The relevant distances for this example are shown in Fig. 2 for Gray and M16^a mapping.
The error probability is dominated by the value λ_1, i.e. the frequency of the minimum Euclidean distance d_1^{ex}. For no a priori information at the demapper, it is well known that Gray mapping is optimal [2]. This is reflected in the smallest value of $\lambda_1 = 24$ in Table 1. Since not all distances are relevant for the error probability [2], we may use an *expurgated* distance spectrum.

3.2. Ideal a priori information at the demapper

If the demapper acquires the correct extrinsic information from the decoder (ideal a priori information, so-called genie or error free feedback case), the demapper knows all the relevant bits except the bit $c_k[i]$ to be detected. Thus, since the demapper has only to decide between the two symbols with bit-labels differing solely in the a priori unknown i-th bit, the 2^m-ary signal set is reduced to a binary signal set. If, in a 16QAM constellation for example, the demapper a priori knows that the last 3 bits are one, it has to decide only between the symbols with labels 0111 and 1111. It is obvious that the genie performance corresponds to a lower bound on the actual error rate.
The distance spectrum simply lists the frequencies λ_j of the distances between all possible symbol pairs depicted in Fig. 3. From the set of distances D used for the case without a priori information, only those distances between symbols that differ in only one position are considered.
+



a) Gray b) Optimized (M16^a)

Fig. 3. Distances $d_j \in D$ listed in the distance spectrum for ideal a priori information. See Fig. 2. for the bit-mapping.

D_{ex}	d_1^{ex} α	d_2^{ex} 2α	d_3^{ex} 4α	d_4^{ex} 5α	d_5^{ex} 8α	d_6^{ex} 9α	d_7^{ex} 10α	d_8^{ex} 13α	d_9^{ex} 18α
Λ	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9
Gray	24	36	32	72	32	8	24	24	4
SP	56	32	24	60	8	16	28	28	4
MSP	52	38	24	72	8	20	24	16	2
MSEW	72	36	32	24	32	24	24	8	4
M16^a	56	42	40	40	16	24	24	8	6
M16^r	56	42	32	56	8	24	16	16	6
I16	52	42	40	48	8	20	32	8	6

Table 1. Distance spectrum (frequency λ_j of the Euclidean distances d_j^{ex}) with no a priori information. α is the minimum squared Euclidean distance between any two signal points (see Fig. 2).

D_{ex}	d_1^{ex} α	d_2^{ex} 2α	d_3^{ex} 4α	d_4^{ex} 5α	d_5^{ex} 8α	d_6^{ex} 9α	d_7^{ex} 10α	d_8^{ex} 13α	d_9^{ex} 18α
Λ	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9
Gray	24	0	0	0	0	8	0	0	0
SP	4	8	8	0	8	4	0	0	0
MSP	0	2	8	4	8	0	4	4	2
MSEW	0	0	0	24	0	0	0	8	0
M16^a	0	0	0	16	4	0	4	8	0
M16^r	0	0	4	8	8	0	8	4	0
I16	0	0	0	16	8	0	0	8	0

Table 2. Distance spectrum (frequency λ_j of the Euclidean distances d_j^{ex}) with ideal a priori information. α is the minimum squared Euclidean distance between any two signal points (see Fig. 2).

A high coding gain over the demapping and decoding iterations may be achieved if the *minimum Euclidean distance*, i.e. the minimum index *j* of d_j^{ex} where $\lambda_j \neq 0$ is maximized.

With Gray mapping, the minimum Euclidean distance and the value of $\lambda_1 = 24$ in Table 2 are the same as in Table 1. Thus, no large gains are expected if a priori information is used. The optimum mapping with ideal a priori information with respect to its distance spectrum is the M16^a mapping. Both a good performance without and with a priori information are desirable in order to allow the iterative process to start and to reach low error rates, respectively. The I16 mapping is both close to optimum with a priori information and has a reasonable performance without a priori information ($\lambda_1 = 52$ in Table 1).

4. OPTIMIZATION OF MAPPINGS

Mappings with optimized distance spectra could be found through exhaustive search, which becomes intractable for higher order constellations ($m \geq 4$) since a maximum number of $2^m!$ different possibilities have to be checked. In [12], a binary switching algorithm (BSA), previously used for index optimization in vector quantization [19], has been proposed to overcome the complexity problems of the brute-force approach. This algorithm finds a local optimum on a given cost function. The problem of finding optimal labeling maps can also be formulated as a quadratic assignment problem (QAP) [20], for which several methods have been proposed [21]. We will only consider the *tabu search* algorithm [22] and a solution based on *integer programming*. We describe and compare the results of the different approaches.

4.1. Cost functions

The optimization algorithms require a cost function. We can either directly optimize the Euclidean distance spectrum by e.g. maximizing the minimum Euclidean distance with ideal a priori information, use error bounds or the mutual information as described in [12]. Here, we will only consider the relevant term ω_j in the error bounds that characterize the influence of the signal mapping. We use

$$\omega_j = \exp\left(-\frac{E_s}{4N_0}(d_j^{ex})^2\right), \qquad (5)$$

for the AWGN channel and

$$\omega_j = \frac{1}{(d_j^{ex})^2}. \qquad (6)$$

for the fully interleaved Rayleigh fading channel [12]. Using the values λ_j from the distance spectrum, the optimization of the mapping μ is done by minimizing the total cost Ω:

$$\min_{\mu}(\Omega) = \min_{\mu}\left(\sum_{j=1}^{n} \lambda_j \cdot \omega_j\right). \qquad (7)$$

4.2. Binary switching algorithm

The binary switching algorithm (BSA) [19][12] is started with an initial mapping. 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 which has the strongest contribution to a ”bad” performance in the list 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. As the BSA finds a local optimum, it is likely that several algorithm executions with random initial mappings yield to the presumed global optimum.

4.3. Reactive Tabu Search

The Reactive Tabu Search (RTS) [22] is initialized with a random mapping and is based on index switching like the BSA: The first step of each iteration is the computation of the reduction of the total costs for all possible binary index switches. The switch that results in the lowest total cost is chosen and made tabu, i.e. the switch between the indices of those two signal points will be forbidden as long as it remains in the tabu list. Note that the "best" switch will not necessarily reduce the total cost. In addition, a switch that results in an often previously passed through mapping is not allowed to avoid cycles. There are two exceptions: If all switches are tabu, the best switch is done neglecting all tabu in the list. If cycles are detected, some random switches are performed. In both cases, the tabu list is reduced for the next step by deleting the earliest tabu. The algorithm stops after a given number of switches.

4.4. Quadratic Integer Programming

Integer Programming (IP) is based on a different notation of the Quadratic Assignment Problem (QAP) and finds the optimum solution. With ideal a priori information, only the symbol pairs where $d_H(\mu^{-1}(s_i), \mu^{-1}(s_j)) = 1$ contribute to the costs, i.e. where the Hamming distance between the labels of the signal points s_i and s_j is one. These symbol pairs are selected with the function

$$f_{i,j}(\mu) = \begin{cases} 1, & \text{if} \quad d_H(\mu^{-1}(s_i), \mu^{-1}(s_j)) = 1 \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Furthermore, let ω_{ij} be the term as defined in (5) and (6) for an Euclidean distance $d_{ij} = |s_i - s_j|$ instead of d_j^{ex}. The problem of finding the mapping μ that minimizes the cost function defined in (7) can be stated as follows:

$$\min_{\mu} \sum_{i=1}^{2^m} \sum_{j=1}^{2^m} \omega_{ij} f_{ij}(\mu), \quad (9)$$

which is the standard formulation of the QAP. Quadratic Integer Programming (QIP) is based on an alternative formulation of the QAP, where the cost function is defined as a function of a permutation matrix $\mathbf{X}$ and no more as a function of the mapping μ itself:

$$\min_{\mathbf{X}} \sum_{i=1}^{2^m} \sum_{j=1}^{2^m} \sum_{k=1}^{2^m} \sum_{l=1}^{2^m} \omega_{kl} f_{ij} x_{ik} x_{jl},$$

$$\text{where} \quad \sum_{i=1}^{n} x_{ij} = 1, \forall j, \quad \sum_{j=1}^{n} x_{ij} = 1, \forall i, \quad x_{ij} \in \{0,1\}. \quad (10)$$

$\mathbf{X}$ describes the index assignment, i.e. a one in the i-th column and j-th row means that the j-th symbol corresponds to the i-th bit label. The QIP is a two-fold optimization: In the first step, we compute the optimum non-integer solution of $\mathbf{X}$ using the Simplex Algorithm. The second step of IP is the application of the Branch and Bound algorithm to get an integer solution of $\mathbf{X}$: We branch on a non-integer entry x_{ij} of $\mathbf{X}$ and obtain two sub problems: one for $x_{ij} = 0$ and one for $x_{ij} = 1$. The value of the cost function is updated and the Branch and Bound procedure repeated for an other non-integer entry $\mathbf{X}$. The global solution is obtained if all sub problems are solved and a feasible integer solution for $\mathbf{X}$ is found.

4.5. Comparison of optimization procedure

The main difference between both algorithms is that the BSA performs a random initialization once a local optimum was found. This is clearly illustrated in Fig. 4 and Fig. 5 for the BSA and RTS algorithm respectively, where the evolution of the costs for an AWGN channel from (5) and (7) is depicted as a function of the performed number of switches. Both algorithms find the mapping with the global minimum costs after a short computation time.

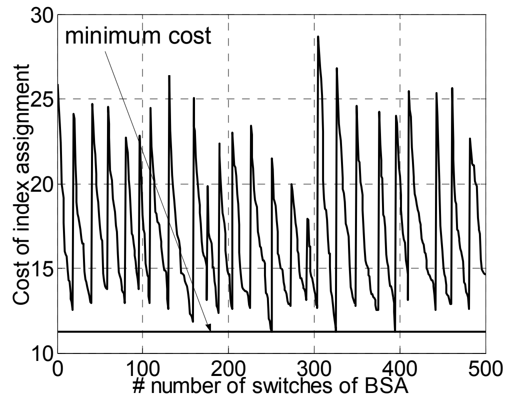


Fig. 4. Cost of mappings for 500 switches of BSA (25 initializations).

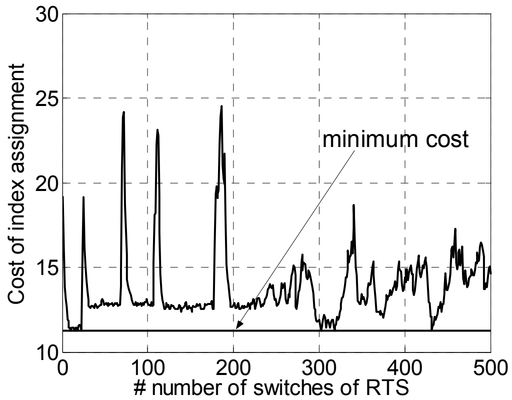


Fig. 5. Cost of mappings for 500 switches of RTS (1 initialization).

5. SIMULATION RESULTS

The BER performance of the investigated 16-QAM mappings is shown in Fig. 6 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 10000 bits. The analytic error bounds for ideal a priori information are computed as described in [2][23] based on Gauss-Chebyshev quadrature rules. The optimized mappings M16^a, M16^r and I16 clearly outperform the other mappings (Gray, SP, MSP, MSEW) once the a priori feedback information has a certain reliability. Since the new mappings are only optimized for ideal a priori information, other mappings can converge at lower signal-to-noise ratio (SNR). The MSEW mapping is outperformed in the whole SNR range by the optimized mappings. The I16 mapping offers a very good trade-off between early convergence and performance at high SNR.

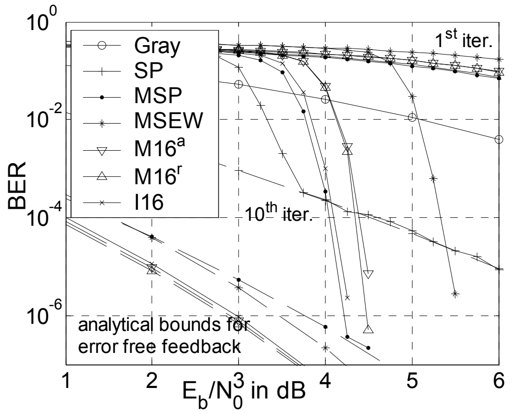


Fig. 6. BER of BICM-ID, 16QAM mappings, AWGN channel, rate $1/2$ 4-state conv. code.

6. CONCLUSION

In this paper we have thoroughly reviewed the characteristics of labeling maps and compared different approaches to find optimized mappings. The optimized 16-QAM mappings clearly outperform previously proposed ones at high signal-to-noise ratio. The optimization methods can be applied and extended to any arbitrary signal constellation.

ACKNOWLODGEMENT

The authors would like to thank Jean Cardinal for his comments on the Quadratic Assignment Problem.

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

[3] S. Benedetto, D. Divsalar, G. Montorsi, 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] S. ten Brink, J. Speidel, and R. Yan, ‘‘Iterative demapping and decoding for multilevel modulation,” in *Proc. IEEE Globecom Conference*, Sydney, November 1998, pp. 579–584.

[5] 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.

[6] M. Tüchler, ‘‘Design of serially concatenated systems depending on the block length,” *IEEE Transactions on Communications*, vol. 52, no. 2, pp. 209–218, February 2004.

[7] F. Sanzi, A. Slama, and J. Speidel, ‘‘Multi code division multiplex with iterative map symbol-by-symbol estimation,” in *Proc. IEEE Globecom Conference*, San Antonio, USA, November 2001.

[8] S. Pfletschinger and F. Sanzi, ‘‘Iterative demapping for OFDM with zero-padding or cyclic prefix,” in *IEEE International Conference on Communications (ICC)*, Paris, France, June 2004.

[9] S. ten Brink, ‘‘Designing iterative decoding schemes with the extrinsic information transfer chart,” *AEÜ International Journal of Electronics and Communications*, vol. 54, no. 6, pp. 389–398, November 2000.

[10] L. Zhao, L. Lampe and J. Huber, “Study of Bit-Interleaved Coded Space-Time Modulation with Different Labeling”, *Information Theory Workshop (ITW)*, Paris, France, March-April 2003

[11] J. Tan and G.L. Stüber, ‘‘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.

[12] F. Schreckenbach, N. Görtz, J. Hagenauer, and G. Bauch, ‘‘Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding,” *IEEE Communications Letters*, vol. 7, no. 12, pp. 593–595, December 2003.

[13] T. Clevorn, S. Godtmann, and P. Vary, ‘‘PSK versus QAM for iterative decoding of bit-interleaved coded modulation,” in *Proc. IEEE Globecom Conference*, Dallas, USA, December 2004.

[14] J. Hagenauer, E. Offer, and L. Papke, ‘‘Iterative decoding of binary block and convolutional codes,” *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 429–445, March 1996.

[15] F. Schreckenbach, N. Görtz, J. Hagenauer, and G. Bauch, ‘‘Optimized symbol mappings for bit-interleaved coded modulation with iterative decoding,” in *Proc. IEEE Globecom Conference*, San Francisco, December 2003.

[16] E. Agrelland, J. Lassingand, E. Stromand, and T. Ottosson, ‘‘On the optimality of the binary reflected gray code,” *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3170–3182, December 2004.

[17] G. Ungerboeck, ‘‘Channel coding with multilevel/phase signals,” *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 55–67, January 1982.

[18] F. Schreckenbach and G. Bauch, ‘‘EXIT charts for iteratively decoded multilevel modulation,” in *th European Signal Processing Conference (EUSIPCO)*, Vienna, Austria, September 2004.

[19] K. Zeger and A. Gersho, ‘‘Pseudo-Gray coding,” *IEEE Transactions on Communications*, vol. 38, pp. 2147–2158, December 1990.

[20] Y. Huang and J. Ritcey, ‘‘Optimal constellation labeling for iteratively decoded bit-interleaved space-time coded modulation,” IEEE Transactions on Information Theory, accepted for publication, November 2004.

[21] E. Cela, *The Quadratic Assignment Problem, Theory and Algorithms*, Kluwer Academic, 1998.

[22] R. Battiti and G. Tecchiolli, ‘‘The reactive tabu search,” *ORSA Journal on Computing*, pp. 126-140, 1994.

[23] E. Biglieri, G. Caire, G. Taricco, and J. Ventur-Traveset, ‘‘Simple method for evaluating error probabilities,” *Electronic Letters*, vol. 32, no. 3, pp. 191–192, February 1996.