Turbo Detection for MIMO Systems: Bit Labeling and Pre-Coding

Stephan Bäro
Institute for Communications Engineering (LNT), TU München (TUM)
Stephan.Baero@ei.tum.de, Arcisstr. 21, D-80290 München, Germany
Tel. +49 89 289-23485, Fax. +49 89 289-23490

Abstract

Bit-interleaved coded modulation with iterative detection for MIMO systems demands an efficient APP detector that delivers soft information about the coded bits. The list-sequential detector is a reduced-complexity approximation to an exhaustive APP detector. It uses a tree search employing the stack algorithm to find dominant terms of the log-likelihood ratio. The commonly used symbol-by-symbol Gray labeling is not necessarily the best approach to good performance of iterative systems. We therefore consider different bit labelings and differential pre-coding adapted to MIMO systems. With the help of EXIT charts and BER simulations, we show that symbol-by-symbol Gray labeling in combination with an outer Turbo code is a good compromise for fast convergence and low error floor of MIMO systems with iterative detection.

Part of the material in this article was published at the 5th International ITG Conference on Source and Channel Coding (SCC) at Erlangen, Germany (January 14th–16th, 2004).

I. INTRODUCTION

In the strive to exploit the available capacity of multi-antenna or multi-input multi-output (MIMO) systems, recent attention has turned to iterative detection and decoding at the receiver. The well-known bit-interleaved coded modulation technique for single-antenna channels [1] was used in [2] with iterative detection on the MIMO channel. Here, a modified version of the sphere decoder [3] has eased the detection complexity of the multi-antenna signals. In [4], a *LIS*t-*S*equential (LISS) detector was shown to be a good approximation to APP (a posteriori probability) detection at reduced complexity. The LISS detector uses the stack algorithm known from sequential decoding to find dominant terms in the calculation of the log-likelihood ratios (LLRs or L-values), which are exchanged between the detector and the decoder.

The inner block of this serially concatenated system consists of the MIMO signal mapper together with the channel. Although we cannot influence the channel itself, the mapper leaves space for modifications. Already in [5], anti-Gray labeling was mentioned as a good alternative to Gray labeling in iterative systems. Recently in [6], good bit labelings were found for single-antenna bit-interleaved coded modulation with iterative detection. The design principles can be extended to MIMO systems, where the design of multi-dimensional signal constellations lets us gain another degree of freedom.

Furthermore, for serially concatenated codes it was stated as a design rule in [7], that the inner encoder must be a convolutional recursive encoder of rate 1 if an interleaver gain (gain of the iterative system) is desired. We are going to verify this condition for MIMO systems, where in the standard design the MIMO mapper is not recursive, but rather block-oriented. A rate-1 differential pre-coder within the mapping will introduce the necessary dependence between the bits entering the MIMO mapper. This approach is similar to a recent description of iterative detection and decoding for vector channels by ten Brink and Kramer [8], [9]. In addition to differential pre-coding they employ nonsystematic repeat-accumulate codes, where we stick to standard convolutional and parallel concatenated convolutional (PCC or "Turbo") codes as the outer code.

After an introduction of the system model for multi-antenna bit-interleaved coded modulation with iterative detection in Section II, we will present the LISS detector as a low-complexity alternative to a full APP detector for MIMO signals. In Section III we will then propose the two modifications to the standard transmitter: Multi-dimensional bit labeling and differential pre-coding. Finally, in Section IV we present extrinsic information transfer (EXIT) charts [10] and bit error rate (BER) simulations to compare the different options of the MIMO mapping.

II. Bit-interleaved coded MIMO transmission and iterative detection

A. System Model

[Figure 1 about here.]

We consider a system according to Figure 1 with the multi-input multi-output (MIMO) channel consisting of n_T transmit and n_R receive antennas. First, the (long) information vector $\mathbf{u}$ is encoded and interleaved (depicted by the permutation Π). At this point, we pose no constraints on the frame length, but we know from standard results for systems with iterative detection that each code word (equal to one frame) should contain several thousand information bits. For the error-correcting code, we will consider both convolutional codes and parallel-concatenated convolutional (PCC or "Turbo") codes. The interleaver is a random interleaver. Other forms of interleaving are possible, see e.g., [11] for design rules of semi-random (S-random) interleavers for PCC codes. For the long frame lengths considered in this thesis, there are no large improvements to be expected from the design of specific interleavers. If a differential pre-coder is used, this sequence passes through a recursive rate-1 encoder (see Section III-B), and we obtain a modified sequence $\mathbf{x}$.

For the definition of the necessary notation and for most of this chapter, we will just consider a single channel use, where m is the number of bits per symbol, the constellation comprises therefore $M = 2^m$ distinct points. As part of the total coded sequence, we obtain an mn_T-dimensional binary vector of coded bits $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_{n_\mathrm{T}})^\mathrm{T}$, with $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,m})$. We assume the binary digits $x_{i,j}$ to take independent values from $\{+1, -1\}$, where $+1$ is the "null" element in GF(2).

This sequence is mapped onto an n_T-dimensional complex vector $\mathbf{s} = (s_1, \ldots, s_{n_\mathrm{T}})^\mathrm{T}$. Initially, each element $s_i = \mathrm{map}(\mathbf{x}_i)$ is taken from some complex constellation $\mathcal{X}$ (e.g., BPSK, QPSK or 16-QAM using Gray labeling). We will also use other bit labeling schemes, propose a generalized multi-dimensional bit labeling for MIMO systems, and we will introduce the aforementioned differential pre-coder within the MIMO mapper.

The transmit energy per antenna is normalized to $\mathrm{E}\{|s_i|^2\} = E_\mathrm{S}/n_\mathrm{T}$ such that E_S represents the total emitted energy per channel use.

Using the linear model of the ergodic MIMO channel, we describe the vector of n_R received complex symbols by $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}$, where $\mathbf{H}$ is a matrix of dimension $n_\mathrm{R} \times n_\mathrm{T}$ with each entry being an independent realization of a complex Gaussian random variable of zero mean and variance $1/2$ per real dimension. The channel is passive and changes with each channel use. $\mathbf{n}$ is a vector of independent noise values, each with zero mean and power spectral density $N_0/2$ per real dimension.

Due to the matrix structure of the MIMO channel, interference between the transmitted symbols is introduced at the receiver. The probability density of a particular realization of the received vector, conditioned on the transmitted vector, is given by the multi-dimensional Gaussian distribution

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(\pi N_0)^{n_\mathrm{R}}} \exp\left(-\frac{1}{N_0}\left\|\mathbf{y} - \mathbf{H}\mathbf{s}(\mathbf{x})\right\|^2\right).$$

At the receiver with perfect channel knowledge, we consider an iterative system consisting of an outer soft-in-soft-out decoder and an inner APP MIMO detector. The MIMO detector takes information from the channel (the received vector $\mathbf{y}$ and the channel matrix $\mathbf{H}$) as well as a priori information in form of a vector of L-values $L(\mathbf{x})$ about the coded bits:

$$\begin{aligned} L(\mathbf{x}) &= \left(L(\mathbf{x}_1), \ldots, L(\mathbf{x}_{n_\mathrm{T}})\right)^\mathrm{T} \\ L(\mathbf{x}_i) &= \left(L(x_{i,1}), \ldots, L(x_{i,m})\right) \\ L(x_{i,j}) &= \ln \frac{P(x_{i,j} = +1)}{P(x_{i,j} = -1)} \end{aligned} \tag{1}$$

The MIMO detector generates a posteriori information $L(\mathbf{x}|\mathbf{y})$ about the coded bits. The detector's output is defined as a vector of conditional L-values

$$L(\mathbf{x}|\mathbf{y}) = \left(L(\mathbf{x}_1|\mathbf{y}), \ldots, L(\mathbf{x}_{n_\mathrm{T}}|\mathbf{y})\right)^\mathrm{T}, \tag{2}$$

$$L(\mathbf{x}_i|\mathbf{y}) = \left(L(x_{i,1}|\mathbf{y}), \ldots, L(x_{i,m}|\mathbf{y})\right),$$

$$L(x_{i,j}|\mathbf{y}) = \ln \frac{P(x_{i,j} = +1|\mathbf{y})}{P(x_{i,j} = -1|\mathbf{y})}.$$

By subtracting the a priori information, extrinsic output is created. This avoids feeding of the outer decoder with the information added by itself in the previous iteration — only "new" information is passed on.

After de-interleaving, this extrinsic information becomes the "channel" input $L(\mathbf{x}')$ of the outer decoder — there is no a priori input to the outer decoder as we assume the information bits to be independent and equally probable. The decoder generates decisions about the information bits as well as extrinsic information about the coded bits. This information is re-interleaved and fed back to the MIMO detector, where it is re-used as a priori knowledge. At the start of the iterative process (iteration 0), no a priori information is available. From the assumption of equiprobable information bits, we also assume the coded bits to take on values $+1$ and -1 with equal probability. Therefore, in iteration 0 all L-values take on zero value, $L(x_{i,j}) = 0$.

This receiver structure is generally referred to as a "Turbo" receiver for serially concatenated systems according to the "Turbo" principle [12].

B. APP detection of MIMO signals

In an iterative system we want the APP detector to generate likelihood information about the transmitted bits. The L-value of the bit $x_{q,r}$, knowing the received vector $\mathbf{y}$, is defined through standard manipulations as

$$L(x_{q,r}|\mathbf{y}) = \ln \frac{\sum_{\mathbf{x}\in\mathcal{X}_{q,r,+1}} p(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{\sum_{\mathbf{x}\in\mathcal{X}_{q,r,-1}} p(\mathbf{y}|\mathbf{x})P(\mathbf{x})}, \tag{3}$$

where the two sums describe expectations of $p(\mathbf{y}|\mathbf{x})$ over $\mathcal{X}_{q,r,\pm1} = \{\mathbf{x}|x_{q,r} = \pm 1\}$. Generally, in MIMO systems, $\mathcal{X}_{q,r,\pm1}$ contain too many entries to evaluate the two sums in a brute-force approach. Already for a system with $n_{\mathrm{T}} = 4$ transmit antennas employing 16-QAM, 32768 terms have to be summed up in both the enumerator and the denominator for each channel use.

Therefore, we want to restrict our evaluation to the dominant terms in the two sums of (3) corresponding to transmit vectors from a yet to be determined list $\mathcal{L}$ (much smaller than $\mathcal{X}_{q,r,\pm1}$), which is independent of particular bit values. At the same time we want to approach the true L-value as closely as possible, because a bad quality of the exchanged information between detector and decoder will deteriorate the overall performance of the receiver.

To avoid numerical problems, the evaluation of the L-value is normally done with logarithmic probabilities which simplify if we assume independence of the bits $\mathbf{x}$. It was shown in [4] how the decision metric Λ defined by the logarithmic probability

$(\ln p(\mathbf{y}|\mathbf{x}) + \ln P(\mathbf{x}))$ can be rewritten as a cumulative metric for $n_\mathrm{R} \geq n_\mathrm{T}$ if we have an initial unconstrained estimate $\hat{\mathbf{s}} = (\mathbf{H}^\mathrm{H}\mathbf{H})^{-1}\mathbf{H}^\mathrm{H}\mathbf{y}$:

$$\ln p(\mathbf{y}|\mathbf{x}) + \ln P(\mathbf{x}) = C + -\frac{1}{N_0}\left\|\mathbf{L}(\mathbf{s}-\hat{\mathbf{s}})\right\|^2 + \sum_{i,j} \ln P(x_{i,j}) \tag{4}$$

$$= C + \sum_{i=1}^{n_\mathrm{T}} \Lambda_i, \tag{5}$$

$$\Lambda_i = -\frac{1}{N_0}\left|\sum_{j=1}^{i} \ell_{i,j}(s_i - \hat{s}_i)\right|^2 + \sum_{j=1}^{M} \ln P(x_{i,j}),$$

where $\mathbf{L}$ with entries $\ell_{i,j}$ is a lower triangular matrix resulting from the QL decomposition of $\mathbf{H} = \mathbf{QL}$, and the constant

$$C = -n_\mathrm{R}\ln(\pi N_0) + \mathbf{y}^\mathrm{H}(\mathbf{I} - \mathbf{H}(\mathbf{H}^\mathrm{H}\mathbf{H})^{-1}\mathbf{H}^\mathrm{H})\mathbf{y}$$

does not depend on the vectors $\mathbf{x}$ and will therefore cancel out in the L-value. The increments Λ_i of the metric

$$\Lambda = \sum_{i=1}^{n_\mathrm{T}} \Lambda_i$$

depend only on the transmission symbols $\mathbf{x}_j$ for $j \leq i$ and the corresponding a priori values.

The metric values are calculated sequentially following a tree structure. We look at the bits $\mathbf{x}$ as a tree of total depth n_T, where each node has 2^m leaving branches, each branch labelled with $\mathbf{x}_i$ and the corresponding symbol $s_i(\mathbf{x}_i)$. In Fig. 2 we have depicted an example tree for $n_\mathrm{T} = 3$ transmit antennas employing BPSK. At each node, a decision has to be made for either a sent $+1$ or a sent -1, such that for example the uppermost path through the tree corresponds to a transmitted sequence of $(+1, +1, +1)$.

[Figure 2 about here.]

Using classical algorithms operating on trees like the stack or the Fano algorithm, we can find the most promising candidate sequence with largest corresponding metric values. While the sphere decoder in [2] can be interpreted as a variant of the Fano algorithm, we propose a list variant of the stack algorithm which proceeds as follows while keeping track of all visited open nodes in an ordered stack list:

1) Initialize the stack with the root node having metric zero and length zero.

2) Extend the path in the stack with the largest metric which has not yet reached its full length. Replace its entry with its $M = 2^m$ successors and their incremented metrics while keeping the order of the stack.
3) If the size of the stack exceeds a predefined maximum $L_{\max}$, remove the exceeding entries.
4) If a sufficient number of full-length paths has been found — or — if a predefined number of nodes has been extended, stop the algorithm.
5) Otherwise, go to step 2.

As in general we want to stop the algorithm before the maximum number of node extensions, we will end up with a partially extended tree, represented by a stack list with associated metric entries.

C. Path augmentation and generalized log-likelihood ratios

If we wanted to rely exclusively on the classical stack algorithm, the best path (maximum metric) would be the only transmission vector we could use. The resulting hard decisions imply too large absolute values of the L-values, which would make good performance of the iterative system impossible.

We could continue to run the stack algorithm until a larger number of full-length candidates with only hard detected bits along the path guarantees a better quality of the decisions. A different approach is to use the information contained in the remaining (incomplete) stack entries to improve the soft output. In order to use all entries from the stack list, we have to correct the metric values by making assumptions about the undetected part — otherwise the metric of the incomplete path would inevitably be too large compared to the true value. Two assumptions are readily available: the unconstrained (zero-forcing) estimate $(\hat{s}_{n_{\mathrm{d}}+1}, \ldots, \hat{s}_{n_{\mathrm{T}}})$ (ZF), and the soft bits $(\bar{x}_{n_{\mathrm{d}}+1,1}, \ldots, \bar{x}_{n_{\mathrm{T}},M})$ from the a priori information (AP) resulting in the estimates $\bar{s}_i = E\{s_i\}$. These two approaches have been compared in [13] where the AP path augmentation has proved to give a slight advantage, which might be outweighed by the simpler calculation of the ZF augmentation.

If we want to include these incomplete paths in the calculation of the log-likelihood ratio (3), we have to assure a special treatment of the unknown bits on each path. This can be done by generalizing the familiar log-likelihood ratio.

For the undetected bits we can use the available a priori information to derive the following "soft" bits:

$$\bar{x}_{q,r} = E\{x_{q,r}\} = \tanh\left(L(x_{q,r})/2\right). \tag{6}$$

They are not perfectly reliable, i.e., $|\bar{x}_{q,r}| < 1$. During the calculation of L-values on an incomplete tree, we will encounter incomplete binary vectors $\bar{\mathbf{x}} \in \mathcal{L}$. These are vectors where some "hard" bits $x_{q,r} \in \{+1; -1\}$ are assumed to be known perfectly (first branches of the tree), whereas for others we only have some less reliable "soft" bits based on the a priori L-values. The use of this list $\mathcal{L}$ generalizes the expectations in the enumerator and denominator of (3):

$$L(x_{q,r}|\mathbf{y}) = \ln \frac{\sum\limits_{\bar{\mathbf{x}} \in \mathcal{L}} p(\mathbf{y}|\bar{\mathbf{x}}) P(\bar{\mathbf{x}}) P(x_{q,r} = +1)}{\sum\limits_{\bar{\mathbf{x}} \in \mathcal{L}} p(\mathbf{y}|\bar{\mathbf{x}}) P(\bar{\mathbf{x}}) P(x_{q,r} = -1)} \tag{7}$$

If some path reaches the end of the search tree, all bits along this candidate will be hard bits — we assume well-defined decisions along this path. In the case where $\mathcal{L}$ enumerates all $2^{mn_{\mathrm{T}}}$ possible vectors $\mathbf{x}$, all bits $x_{q,r}$ are assumed perfectly known (on each path) and (7) simplifies again to (3).

III. Modifications of the MIMO mapping

Using an APP detector or the LISS detector as its simplified version, we can successfully implement an iterative receiver for coded MIMO systems. Its performance depends to a large degree on the accuracy (and therefore the complexity) of the two principal components at the receiver, i.e., the inner detector and the outer decoder.

It remains the question how the two corresponding serially concatenated components on the transmitter side should be designed in order to achieve the best possible performance of the whole system. Here, we will consider two standard choices of the outer code (PCC or convolutional) in combination with modified bit labeling and/or differential pre-coding at the inner MIMO mapper.

A. Multi-dimensional bit labeling

Although in most standard communication systems Gray labeling is used to assign bit patterns to signal points, recent results have shown that this is not necessarily optimum for systems with iterative detection ([5], [6], [14]).

[Figure 3 about here.]

For good performance in iterative systems with a priori information, the Euclidean distance between a symbol s and its counterpart z is important, where z has the same bit labeling except a single inverted bit. See Fig. 3 for a single-antenna example using QPSK with either Gray labeling (left) or natural labeling (right): If we assume ideal a priori knowledge of the second bit, the detector's choice becomes limited to a symbol pair instead of the four possible symbols (indicated by lines in Fig. 3). Clearly, with natural labeling, the Euclidean distance between the two remaining symbols is larger than with Gray labeling. This implies a more reliable decision in the presence of noise.

[Figure 4 about here.]

For MIMO systems, this can be generalized if arbitrary vector labelings $\mathbf{s} = \mathrm{map}(\mathbf{x})$ are possible instead of symbol-by-symbol labelings $s_i = \mathrm{map}(x_i)$. We look at the mn_T bits as one transmission vector that is to be transmitted via n_T complex constellation symbols. Consider Fig. 4 for an example using $n_\mathrm{T} = 2$ transmit antennas and QPSK: Symbol-by-symbol Gray labeling implies that the first two bits of the vector $\mathbf{x}$ are always used to define the symbol transmitted over the first antenna, and the second pair of bits defines the transmit symbol of the second antenna.

This condition is relaxed for the two-dimensional MIMO labeling. We clearly see that there is no one-to-one relationship between the bits of the vector $\mathbf{x}$ and the QPSK symbols transmitted over the two transmit antennas.

Based on the error bounds given for bit-interleaved coded modulation (BICM) it was derived in [6], [14] that the asymptotic performance of BICM with iterative detection is governed in the single-antenna case by the harmonic mean of all possible vector pairs $(\mathbf{s}, \mathbf{z})$. This is an expurgated union bound where only single bit-errors during one channel use are taken into account.

This bound can be extended to our system of bit-interleaved coded modulation over the ergodic MIMO channel. As an example, we can see from Fig. 4 by comparing the bit sequences $++++$ and $+++-$, that the exemplary 2D MIMO labeling gives a much larger Euclidean distance between the two symbol vectors, as *both* antenna symbols change, whereas for 1D Gray labeling only *one* symbol changes by the smallest possible amount.

By using a Chernoff bounding technique it can be shown that in the multi-dimensional case, the asymptotic bit error rate (for $E_{\mathrm{S}}/N_0 \to \infty$) is proportional to

$$P_b \sim \frac{W(d_f)}{k d_M^{2d_f}} \left(\frac{E_{\mathrm{S}}}{4N_0}\right)^{-d_f n_{\mathrm{R}}},$$

where d_f is the minimum free distance of the outer code, $W(d_f)$ is the corresponding average input weight (see e.g., [15]), and the harmonic mean of the Euclidean distance of the mapping writes

$$d_{\mathrm{h}}^2 = \left(\frac{1}{mn_{\mathrm{T}}2^{mn_{\mathrm{T}}}} \sum_{i=1}^{mn_{\mathrm{T}}} \sum_{b\in\pm1} \sum_{\mathbf{s}\in\mathcal{X}_{i,b}} \frac{1}{||\mathbf{s}-\mathbf{z}||^{2n_{\mathrm{R}}}}\right)^{-1}.$$

We strive to find the labeling that maximizes this harmonic mean, in order to minimize bit error rate. As we want to design our transmitter independent of the number of receive antennas, we will set $n_{\mathrm{R}} = 1$ for our search of good labelings.

In the absence of a priori information the distance has to take into account all possible pairs of symbols, not only those with one inverted bit [6]. Exhaustive design of good labelings is difficult due to the large number of possible labelings, and true multi-dimensional labelings prevent the straight-forward application of the LISS detector. As we will see, already simple choices like 1D natural labeling instead of Gray labeling yield substantial changes in the behavior of the concatenated system.

B. Differential Pre-Coding

Let us introduce a differential rate-1 pre-coder in the mapper (see Fig. 1). The mapping is no longer a symbol-by-symbol mapping, and the pre-coder operating in GF(2) is defined by the recursive state-space equations [16]

$$\xi_{k+1} = \mathbf{A}\xi_k + \mathbf{B}\mathbf{v}^{(k)}, \quad (8)$$

$$\mathbf{x}^{(k)} = \mathbf{C}\xi_k + \mathbf{D}\mathbf{v}^{(k)}, \quad (9)$$

transforming the length mn_{T} vectors $\mathbf{v}^{(k)}$ into vectors $\mathbf{x}^{(k)}$ of the same size. The state vector of the recursive pre-coder at time k is described by ξ_k. Without going into details of the choice of the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$, we can find a pre-coder that exhibits the best compromise of good performance without and with a priori information. For maximum mutual dependence of the bits, all bits contribute to the state of the memory-1 encoder, therefore

$$\mathbf{A} = 1, \quad \text{and}$$

$$\mathbf{B} = [1, \ldots, 1].$$

The pre-coder ist assumed to start in the zero state $\xi_1 = 0$. The pre-coder inevitably leads to a performance loss if no a priori information is available. If a pre-coder is used, we should therefore stick to Gray labeling and only one output bit should be affected by the encoder. This leads to the choice of the two remaining matrices

$$\mathbf{D} = \mathbf{I}, \quad \text{and}$$

$$\mathbf{C} = [1, 0, \ldots, 0]^{\mathrm{T}},$$

where $\mathbf{I}$ is the $mn_{\mathrm{T}} \times mn_{\mathrm{T}}$ identity matrix. The resulting pre-coder for a system with four input bits per channel use is shown in Fig. 5.

[Figure 5 about here.]

At the receiver the transmitted bits can be recovered using a standard APP decoder for trellis coded modulation based for example on the BCJR algorithm, whose metric is extended to take into account the multiple receive antennas. For larger systems (such as 16-QAM with 4 transmit antennas transmitting 16 coded bits per channel use) the number of parallel transitions in the decoder can become a problem and demands for efficient detector algorithms.

IV. SIMULATION RESULTS

We will use a transmission system with $n_{\mathrm{T}} = n_{\mathrm{R}} = 4$ antennas and QPSK modulation. The outer code is at first an $\nu = 3$ PCC code, punctured to code rate $R = 1/2$ (8 internal decoder iterations, interleaver length 9214 bits).

[Figure 6 about here.]

Fig. 6 shows two EXIT charts for the system operating at an SNR of 2.2 dB. EXIT charts are a tool to analyze the convergence behavior of iterative systems [10]. The thin lines show the mutual information of the extrinsic output of the detector as a function of the mutual information of the detector's a priori input. The thick lines show the corresponding characteristic of the outer decoder, but with flipped axes to depict the information exchange between detector and decoder. If the intersection between both characteristics is close to the right and top axes of the chart, there will be an open "tunnel" between the two graphs, and convergence will take place at low bit error rate.

To avoid confusion of different effects, exhaustive APP detection is used. The left-hand chart shows the combination of an outer PCC code and two different bit labelings. We see that for PCC/Gray there is a small gap between both lines which corresponds to the convergence of this system. If we use natural instead of Gray labeling, convergence will be prevented by the early intersection of the two characteristics.

Therefore, natural labeling should be employed with a simpler outer code that exhibits a less pronounced "knee". In the right-hand chart, we show the combination of natural labeling with an outer convolutional code ($R = 1/2$, constraint length $\nu = 4$, generator polynomial $G(D) = [1 + D^3 + D^4, 1 + D + D^3 + D^4]$). Now the characteristic of the APP detector clearly lies above the outer code which implies that convergence should take place even earlier than 2.2 dB. However, there is another intersection of both characteristics, which is still relatively far from the right axis. This will result in an error floor.

The typical error floor can be avoided if differential pre-coding is used. As predicted, the corresponding characteristic reaches the upper-right corner of the right-hand EXIT chart. However, due to the initial loss if no a priori information is present, there is no visible gap for the iterative detection to perform well at the investigated SNR of 2.2 dB. Therefore, we expect convergence only to take place at a higher SNR. If pre-coding was combined with natural labeling, convergence would take place even later.

[Figure 7 about here.]

For the multi-dimensional bit labelings, a random search was performed over $5.5 \cdot 10^5$ different bit labeling patterns, a small subset of all possible labelings. Only a small range of values is occupied by the values of d_h^2 for $n_{\mathrm{R}} = 1$: The minimum was determined as $d_{\min}^2 = 1.56$, and the maximum was $d_{\max}^2 = 1.82$.

Therefore, we are interested in the performance of these two extreme bit labelings. This is shown in the EXIT chart in Figure 7. There is only a slight difference between the two graphs. Compared to the detector characteristics from Figure 6 the slope is much higher than for the symbol-by-symbol labelings. We therefore see the symbol-by-symbol Gray and natural labelings as two extreme examples of multi-dimensional bit labelings.

Although the performance with a priori information is close to optimal, the intersection on the left-hand side of the chart will prevent convergence at an SNR of 2.5 dB. Therefore, we do not expect good performance from this combination. Experiences with higher-order modulation schemes such as 16-QAM modulation show even steeper detector characteristics, which prevent convergence over a larger range of SNR values.

[Figure 8 about here.]

These results are well verified through BER simulations as shown in Fig. 8, where the Shannon limit to capacity is situated at an SNR of approximately 1.6 dB. The system with natural labeling and a convolutional code reaches the performance of the PCC/Gray system after 8 iterations and even outperforms it after 20 iterations. However, the high error floor at more than 10^{-3} will be inacceptable in mobile communication systems. The large number of iterations, which is necessary in systems with a very narrow "tunnel" in the EXIT chart, also reduces their practical importance.

With Gray labeling and pre-coding, no error floor is visible. Combined with an outer memory-4 convolutional code, convergence takes place at an SNR of approximately 0.4 dB more than in the PCC/Gray system. If the pre-coder is combined with an outer memory-2 convolutional code there is an additional loss of 0.15 dB.

Therefore, for MIMO systems, Gray labeling without pre-coding and with an outer PCC code is a compromise for overall good performance, situated only at about 0.6 dB from capacity at a BER of 10^{-4}.

V. CONCLUSION

Iterative detection of bit-interleaved coded MIMO systems is a promising way of approaching the ergodic MIMO capacity. The LISS detector can be used to reduce the complexity of the APP detector needed in such systems. We have discussed several options in the transmitter design of the concatenated system consisting of an outer code and the MIMO mapping. Using an outer convolutional code with natural (or anti-Gray) QPSK mapping, we have early convergence at the price of an almost untolerable error floor. We can gain another degree of freedom by designing multi-dimensional bit labelings, but their steeper EXIT characteristic prevents early convergence of the iterative detection process. With a differential pre-coder and Gray mapping, no error floor appears at the price of later convergence. The combination of an outer PCC code and standard Gray mapping achieves a good compromise of fast convergence and low error floor at only 0.6 dB from the Shannon limit to capacity.

ACKNOWLEDGEMENT

The author would like to thank Sylvia Reitz for providing part of the simulation results.

REFERENCES

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

[2] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[3] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1639–1642, July 1999.

[4] S. Bäro, J. Hagenauer, and M. Witzke, "Iterative detection of MIMO transmission using a list-sequential (LISS) detector," in *Proc. ICC*. Anchorage (AK), USA: IEEE, May 2003.

[5] S. ten Brink, J. Speidel, and R.-H. Yan, "Iterative demapping and decoding for multilevel modulation," in *Proc. IEEE GLOBECOM*, Sydney, Australia, Nov. 1998, pp. 579–584.

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

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

[8] S. ten Brink and G. Kramer, "Turbo processing for scalar and vector channels," in *Proc. 3rd International Symposium On Turbo Codes & Related Topics*, Sept. 2003, pp. 23–30.

[9] ——, "Design of repeat-accumulate codes for iterative detection and decoding," *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 2764–2772, Nov. 2003.

[10] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[11] S. Dolinar and D. Divsalar, "Weight distributions for turbo codes using random and nonrandom permutations," JPL, TDA Progress Report 42-122, Aug. 1995.

[12] J. Hagenauer, "The turbo principle: Tutorial introduction and state of the art," in *Proc. International Symposium on Turbo Codes*. ENST de Bretagne, September 1997, pp. 1–11.

[13] S. Bäro, "Turbo detection for MIMO systems using a list-sequential detector: Improved soft output by path augmentation," in *2. Diskussionssitzung Angewandte Informationstheorie*. Dresden: ITG, June 2003, pp. 66–68.

[14] X. Li, A. Chindapol, and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8PSK signaling," *IEEE Trans. Commun.*, vol. 50, no. 8, pp. 1250–1257, Aug. 2002.

[15] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. McGraw-Hill, 1979.

[16] R. Otnes and M. Tüchler, "On iterative equalization, estimation, and decoding," in *Proc. ICC*. Anchorage (AK), USA: IEEE, May 2003.

BIOGRAPHY

Stephan Bäro studied electrical engineering and telecommunications in Aachen, Germany, and Paris and Toulouse, France. He graduated from RWTH Aachen in 1997. He has been with the Institute of Communications Engineering (LNT) of the Munich Institute of Technology (TUM) since 1997, where he obtained his PhD in 2004. His research interests are in multi-antenna systems and iterative receivers.

LIST OF FIGURES

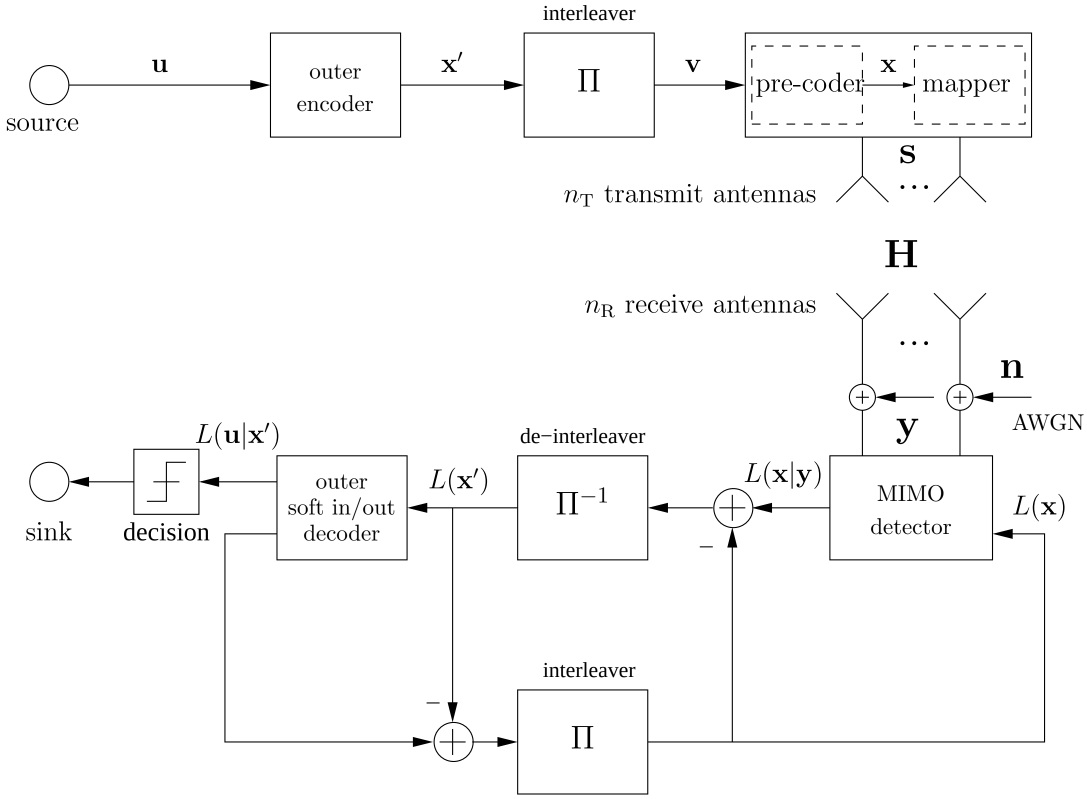

Fig. 1. System model of MIMO bit-interleaved coded modulation with iterative detection.

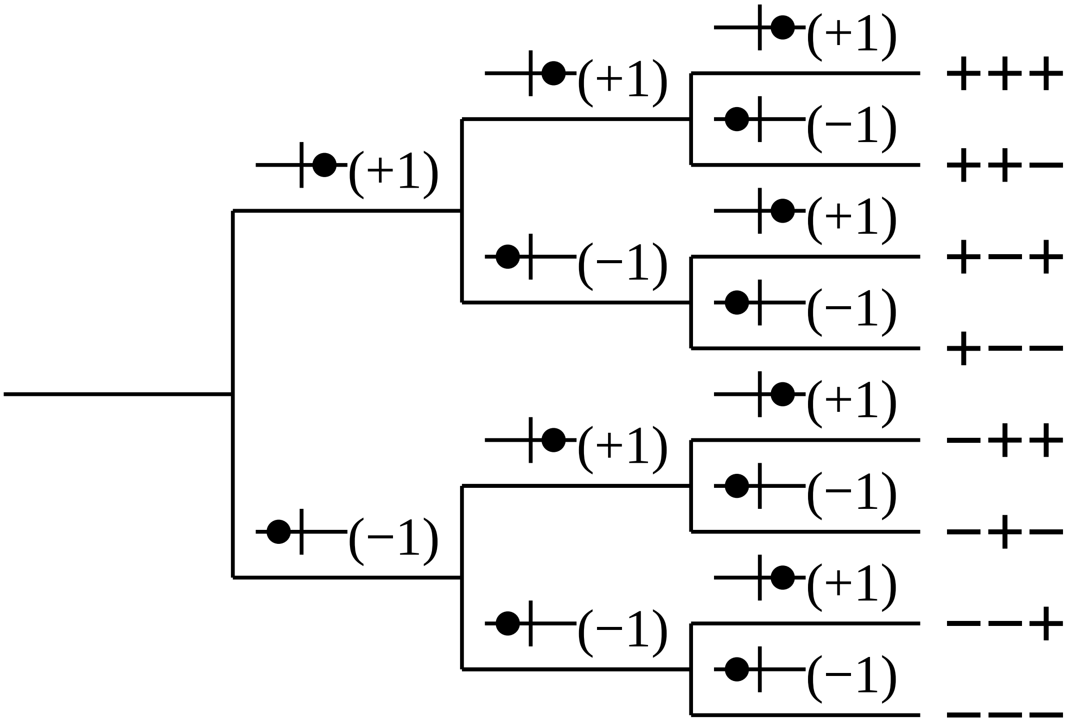

Fig. 2. Decision tree for the stack algorithm. Example with $n_{\mathrm{T}} = 3$ transmit antennas and BPSK modulation.

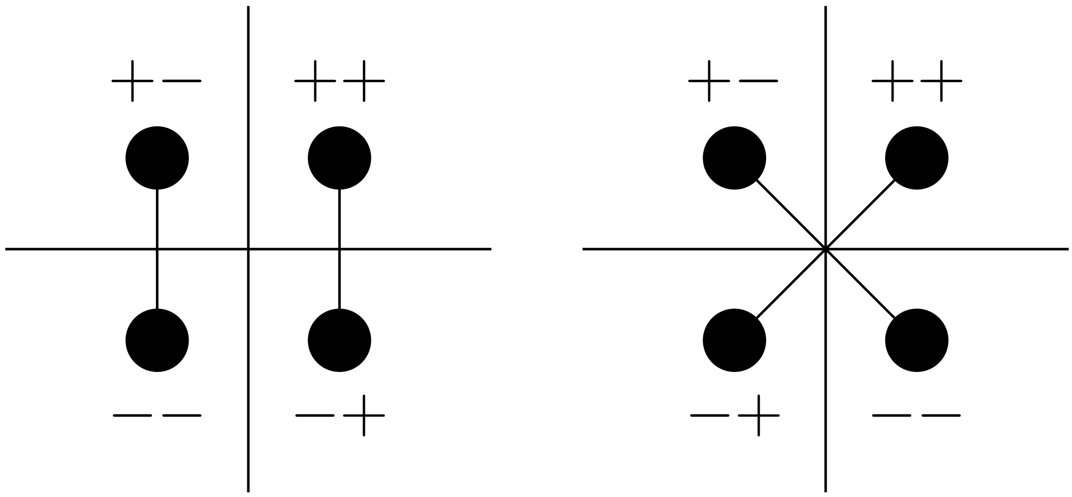

Fig. 3. Different labelings for QPSK transmission (left: Gray labeling, right: natural labeling)

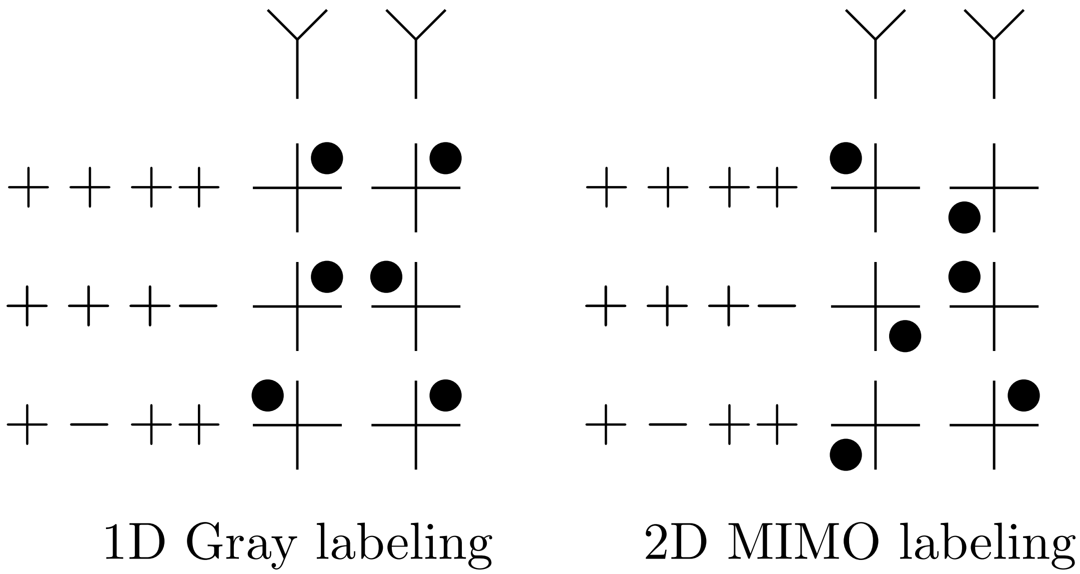

Fig. 4. Example for multi-dimensional MIMO labeling, QPSK, $n_{\mathrm{T}} = 2$ transmit antennas

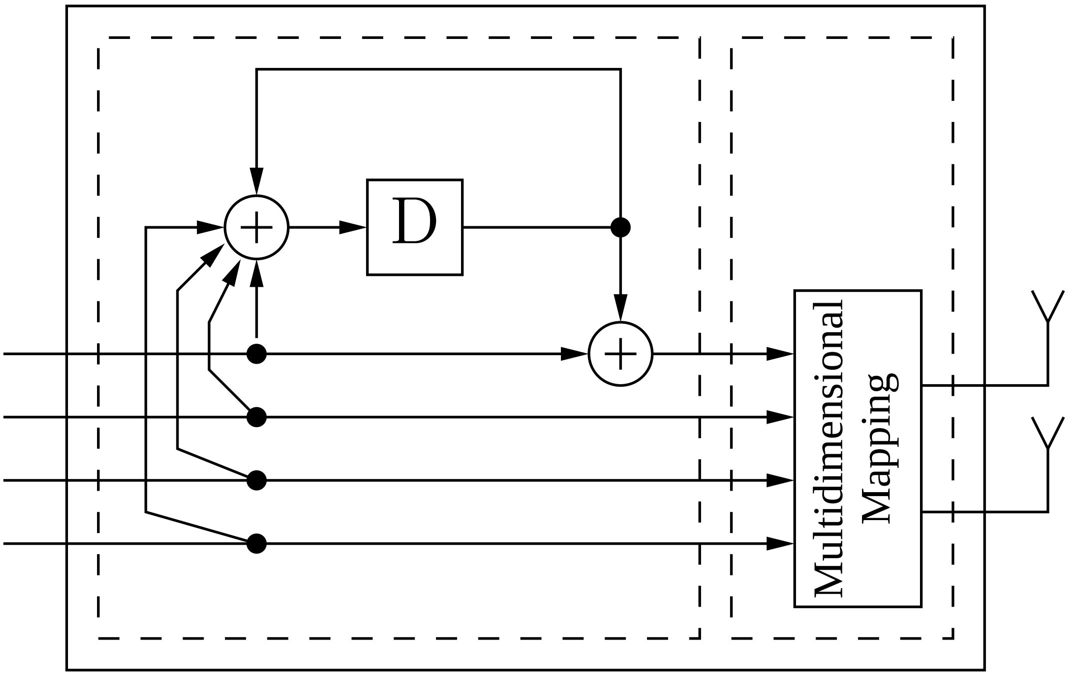

Fig. 5. MIMO mapper with differential rate-1 pre-coder for $n_{\mathrm{T}} = 2$ antennas and QPSK

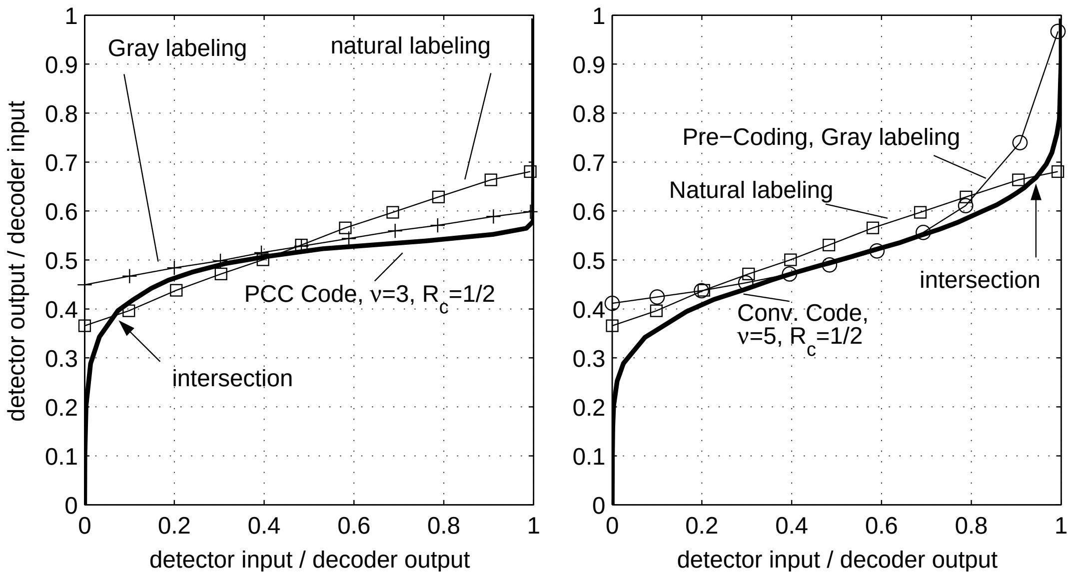

Fig. 6. EXIT charts for different labeling/pre-coding strategies, $n_{\mathrm{T}} = n_{\mathrm{R}} = 4$, QPSK, E_{S}/N_0 at 2.2 dB.

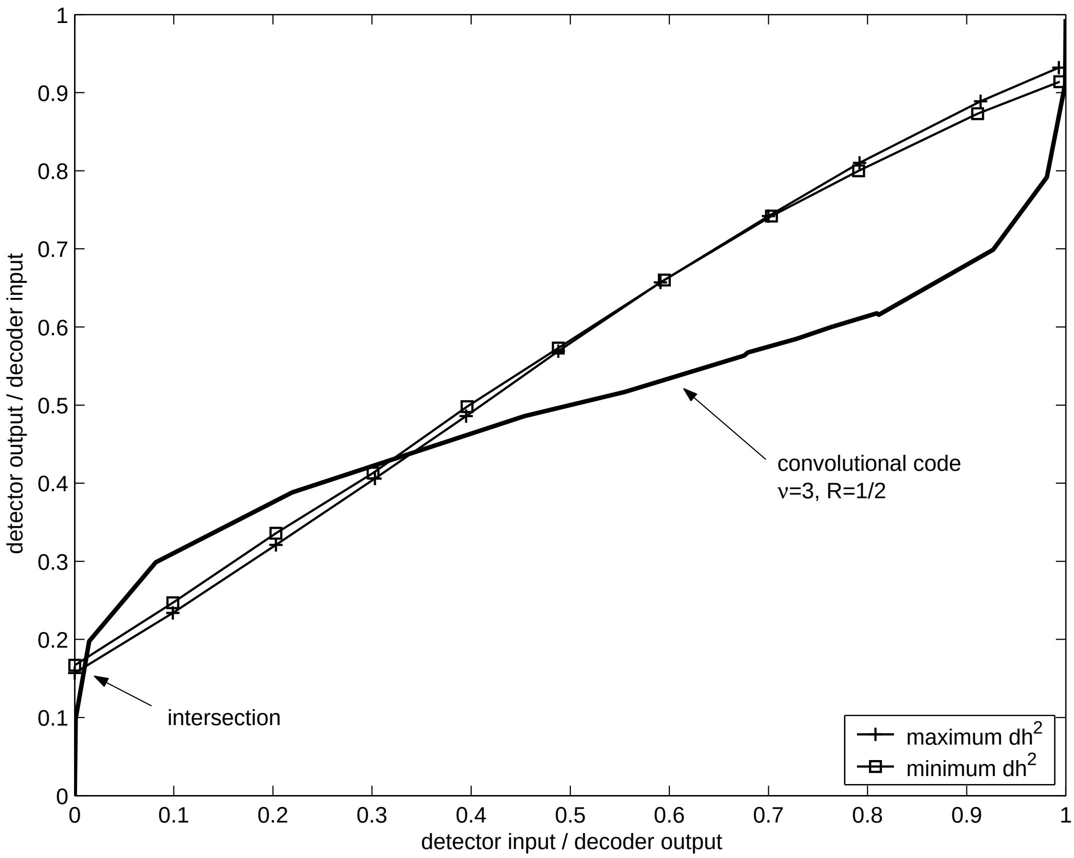

Fig. 7. EXIT chart for outer convolutional code ($\nu = 3$, $R = 1/2$) and multi-dimensional bit labeling, $n_{\mathrm{T}} = n_{\mathrm{R}} = 4$, QPSK, E_{S}/N_0 at 2.5 dB. Detector characteristics for bit labelings with minimum and maximum harmonic mean, determined from $5.5 \cdot 10^5$ random labelings

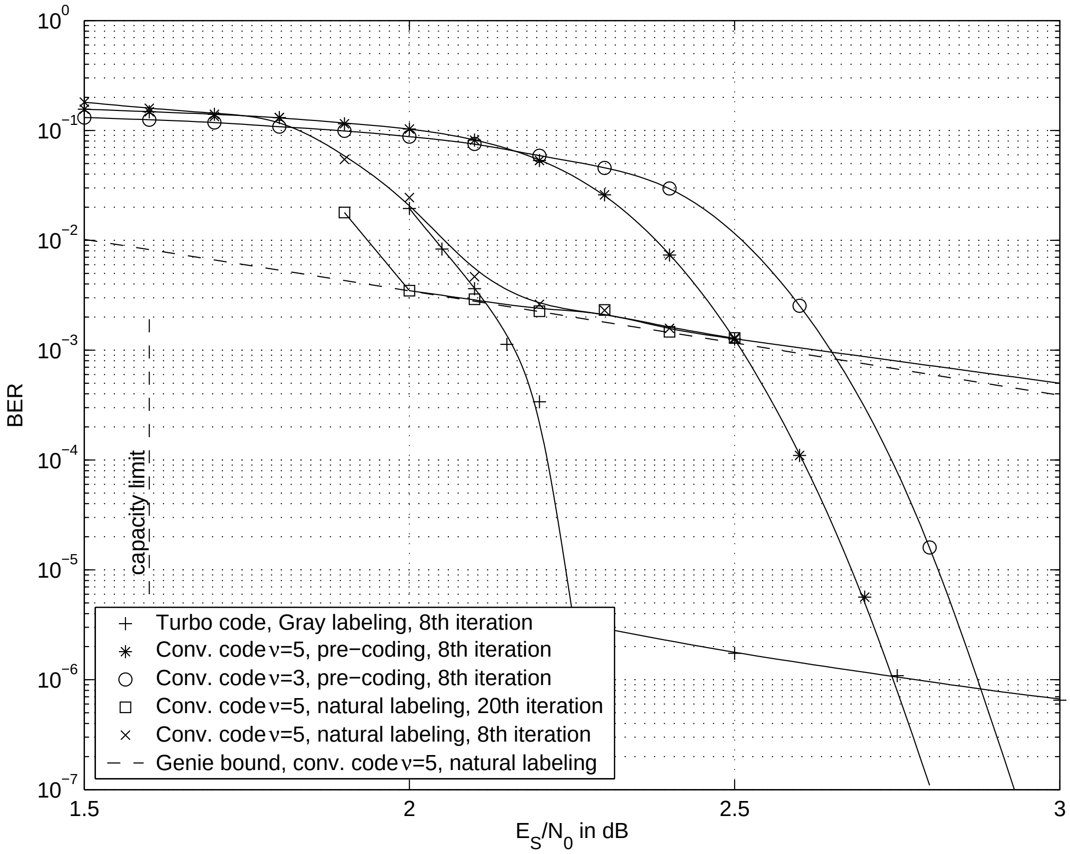

Fig. 8. BER simulations for different labeling/pre-coding strategies, $n_{\mathrm{T}} = n_{\mathrm{R}} = 4$, QPSK.