Abstract—We present an improved algorithm for detection of the received signal of a convolutionally encoded wireless MIMO system. The method is based on BLAST (Bell-Labs Layered Space-Time) algorithm with successive interference cancellation. Compared to existing solutions, cancellation is done adaptively. The receiver uses a MAP soft output demapper combined with an iterative (Turbo-) decoder. Adaptive symbol cancellation is done using a threshold which takes reliability of a symbol decision into account. Cancellations of wrongly decided symbols are reduced to a large extend. As a result, we obtain an SNR gain of 1 to 1.5 dB for Rayleigh fading MIMO channels with additive white Gaussian noise.

I. INTRODUCTION

In the initial publications [1], [2] the BLAST (Bell-Labs Layered Space-Time) algorithm was mainly presented as zero-forcing (ZF) with ordered successive interference cancellation (OSIC). The minimum mean square error (MMSE) solution was mentioned there, but presented in detail e. g. in [3]. It is well known that OSIC reduces the bit error ratio (BER) of an uncoded BLAST system [3]. However, the behaviour of a coded BLAST system is quite different. A simple and straightforward implementation of a coded BLAST transmitter is shown in Fig. 1. The binary output of the data source (information bits) \( u(t_u) \) is encoded to yield \( x'(t_x) \) (coded bits), interleaved resulting in \( x(t_x) \), and mapped to \( s(t_x) \) so that each of the \( M \) transmitters carries an independent data stream \( s_i(t_x) \) where the index \( i \) denotes the number of the transmitter \( (i = 1, \ldots, M) \). \( t_u, t_x, \) and \( t_s \) are the discrete time instances of information bits, coded bits, and symbols, respectively, which are dropped in the following to simplify the notation. A random interleaver of size \( S = 96000 \) and a systematic convolutional (SC) code of rate \( R_c = 1/2 \), memory \( \nu = 2 \), and feedforward polynomial \( G_{oc_f} = 7 \) are used. As usual, the polynomial is given in octal numbers. This structure was called vertical coding [4].

A convenient way to describe a flat fading MIMO model is the use of the matrix notation

\[
\mathbf{r} = \mathbf{Hs} + \mathbf{n}
\]  

where \( \mathbf{s} = (s_1, \ldots, s_M)^T \) is the transmit symbol column vector. Each component \( s_i \) is a complex QAM symbol to be sent by antenna \( i \) \((i = 1, 2, \ldots, M)\). \( \mathbf{n} = (n_1, \ldots, n_N)^T \) is an additive noise column vector with components \( n_j \), which are complex AWGN at receive antenna \( j \), each with zero mean and variance \( 2\sigma_n^2 \) \((j = 1, 2, \ldots, N)\). This means that real and imaginary part of \( n_j \) are both Gaussian with variance \( \sigma_n^2 = N_0/2 \). \( N \) is the number of receive antennas. We consider \( n \) to be uncorrelated with expected value \( \mathbb{E}\{nn^*\} = 2\sigma_n^2 I_N \). \( I_N \) is the identity matrix of size \( N \). Furthermore, it is assumed that the vectors \( \mathbf{s} \) and \( \mathbf{n} \) are uncorrelated, i. e. \( \mathbb{E}\{\mathbf{s}\mathbf{n}^*\} = 0 \). \( \mathbf{r} = (r_1, \ldots, r_N)^T \) is the received symbol column vector, where \( r_j \) is received by antenna \( j \) \((j = 1, 2, \ldots, N)\). The channel matrix \( \mathbf{H} \) is given by \( \mathbf{H} = (\mathbf{h}_1 \ldots \mathbf{h}_M)^T \) with \( \mathbf{h}_i = (h_{i1}, \ldots, h_{iN})^T \) is a column vector of \( \mathbf{H} \). The impulse response \( h_{ij} \) from transmitter \( i \) to receiver \( j \) is modeled as a zero-mean, complex Gaussian random variable satisfying \( \mathbb{E}\{|h_{ij}|^2\} = 1 \) (i. e. the channel is passive). All entries of \( \mathbf{H} \) are i.i.d. \((\cdot)^T \) denotes the transpose and \((\cdot)^* \) the conjugate transpose. If not stated otherwise, bold characters are used for vectors in our text. (Complex) scalars are denoted by small and matrices by capital letters.

Assuming that we use the same mapping for all transmit antennas and mean energy per symbol \( E_s \), the averaged total transmit power equals \( ME_s \). We define the signal to noise ratio (SNR) as \( E_s/N_0 = \frac{E_sN}{R_cQ\sigma_n^2} \) where \( R_cN \) is the transmitted energy per information bit at the receiver. Using the column vectors of \( \mathbf{H} \) we can write Eq. (1) as

\[
\mathbf{r} = \sum_{i=1}^{M} \mathbf{h}_i s_i + \mathbf{n}
\]  

The BLAST system needs to have at least as much receive antennas \( N \) as transmit antennas \( M \). While the receiver does not need to have any knowledge about the channel, we assume that the receiver has full knowledge of \( \mathbf{H} \).

The paper is organized as follows. In Section II we describe the receiver. The new detection scheme is presented in Section III. In Section IV simulation results are given and discussed follows by complexity considerations in Section V. Section VI concludes this paper.
II. THE RECEIVER

Fig. 2 shows the block diagram of the receiver. The MIMO detector takes the \(N\) received samples \(r_j\) and performs the so-called minimum mean squared error ordered successive interference cancellation (MMSE-OSIC) or just the MMSE algorithm. The latter operates without symbol cancellation [1]. Alternatively, the detector performs ZF (zero forcing) or ZF-OSIC. Further, in the following we will introduce adaptive cancellation to improve BER. As explained later, in our solution the MAP soft-output demapper is part of the detector rather than a separate block after detection.

Assume, that the detection order is given by the ordered sequence \(\mathcal{D} = \{d_1, d_2, \ldots, d_M\}\), i.e. symbol \(s_{d_1}\) is detected first out of the received vector \(r\), then \(s_{d_2}\) etc. until \(s_{d_M}\). The BLAST algorithm generates \(\mathcal{D}\) upon the evaluation of the signal to noise ratio during processing. The detection of the symbol \(s_{d_1}\) is based on its equalization row vector

\[
\mathbf{w}_{d_1} = (w_{d_1,1}, w_{d_1,2}, \ldots, w_{d_1,N})
\]

where \(\mathbf{w}_{d_1}\) is the \(d_1\)th row of the MMSE inverse of the channel matrix given by

\[
W = (\alpha I_M + H^*H)^{-1} H^*
\]

with \(\alpha = (2\sigma_n^2)/E_s\) and \(w_{i,j}\) is in general the \(j\)th element of \(\mathbf{w}_i\). The detection process for the first symbol \(s_{d_1}\) is given by

\[
y_{d_1} = \mathbf{w}_{d_1}^T \mathbf{r} = \mathbf{w}_{d_1}^T (H\mathbf{s} + \mathbf{n}) = \mathbf{w}_{d_1} H\mathbf{s} + \mathbf{w}_{d_1} \mathbf{n}.
\]

Introducing a so-called Post Detection Channel (PDC) \(\vec{h}_{d_1} = \mathbf{w}_{d_1} H\) and a Post Detection Noise (PDN) \(\vec{n}_{d_1} = \mathbf{w}_{d_1} \mathbf{n}\) results in

\[
y_{d_1} = \vec{h}_{d_1} \mathbf{s} + \vec{n}_{d_1}
\]

\[
= \vec{h}_{d_1,s} \mathbf{s} + \vec{h}_{d_1,t} \mathbf{s} + \vec{n}_{d_1}
\]

\[
= \left\{ \begin{array}{l}
\vec{h}_{d_1,s} \mathbf{s} + \vec{h}_{d_1,t} \mathbf{s} + \vec{n}_{d_1} \\
\end{array} \right.
\]

\[
\begin{array}{l}
\text{(A)} \\
\text{(B)} \\
\end{array}
\]

where \(\vec{h}_{i,j}\) is in general the \(j\)th element of \(\vec{h}_i\). Note, that \(\vec{h}_i\) is a row vector whereas \(\mathbf{h}_i\) is a column vector. Term (A) is the desired symbol \(s_{d_1}\) weighted with its PDC coefficient. Term (B) is the Post Detection Inter Symbol Interference (PD-ISI) which consists of all other symbols weighted with their corresponding PDC coefficients. Term (C) is the PDN for symbol \(s_{d_1}\).

Since \(\mathbf{n}\) is Gaussian (i.e. its entries have Gaussian pdf) \(\vec{n}_{d_1}\) can be seen as a realization of a Gaussian process with variance \(\sigma_{\vec{n}_{d_1}}^2 = \sigma_n^2 \sum_{i=1}^N |w_{d_1,i}|^2\). We can also compute the variance of the PD-ISI: \(\sigma^2_{\text{PD-ISI}} = \sum_{i,j} \mathbb{E}[|s_i|^2]\). We assume that each antenna transmits symbols with the same energy \(E_s\), i.e. \(\mathbb{E}[|s_i|^2] = E_s\) for \(i = 1, \ldots, M\). The distribution of the PD-ISI is not Gaussian. However, to simplify the scheme, we summarize terms (B) and (C) into one single realization of a random process which we assume to be Gaussian as in [5] with variance \(\sigma^2_{\text{PD-ISI}} = \sigma_{\vec{n}_{d_1}}^2 + \sigma_{\text{PD-ISI}}^2\).

The MAP demapper in Fig. 2 performs bit-wise soft-output mapping described in [6]. Thus, the \(L\)-value of the \(k\)th bit \(x_k\) of symbol \(y_{d_1}\) is given by

\[
L(x_k | y_{d_1}) = L_\alpha (x_k) + \ln \frac{\sum_{x_{d_1} = 0}^{x_{d_1} = 1} p(y_{d_1} | x_{d_1} = 0, x_{j=0}^{j=k-1} x_j \neq \nu) \cdot \alpha_{x_k}}{\sum_{x_{d_1} = 0}^{x_{d_1} = 1} p(y_{d_1} | x_{d_1} = 1, x_{j=0}^{j=k-1} x_j \neq \nu) \cdot \alpha_{x_k}}
\]

The values of the bits \(x_{d_1}^j\) in (7) accomplish the following equation [6]: \(\sum_{j=0}^{j=k-1} x_{d_1}^j \cdot Q^j + \sum_{j=k+1}^{j=q-1} x_{d_1}^j \cdot 2^{j-k} = \nu\). Eq. (7) means that we set bit \(x_{d_1}^j\) to 1 in the numerator and to 0 in the denominator and permute over all other bits within the symbol. \(Q\) is the number of bits per symbol. The BLAST algorithm provides no a priori information on the coded or information bits. So the complete a priori information for the very first pass is set to zero: \(L_\alpha (x_{d_1}^j) = 0\) for all \(i\). Note, that the MAP demapper performs a 1x1 demapping as for systems with only 1 transmitter and 1 receiver. Thus, its complexity is independent of the number of transmit or receive antennas and depend only on the number of bits per symbol.

To detect the next symbol \(s_{d_2}\), in case of no cancellation, the detector goes back to (3) and replaces index \(d_1\) with \(d_2\). This procedure is done until all \(M\) transmitted symbols are detected. With cancellation one has to compute a reduced receive vector \(\mathbf{r}(\{d_1\}) = \mathbf{r} - \vec{h}_{d_1,s} \hat{s}_{d_1}\) where \(\hat{s}_{d_1}\) is the estimate of \(s_{d_1}\) based on \(y_{d_1}\) and the index \(\{d_1\}\) indicates that symbol \(s_{d_1}\) is deleted in \(\mathbf{r}\). In addition, one has to pull out the corresponding \(d_1\)th column of \(H\) to compute the detection row vector \(\mathbf{w}_{d_2}\). Now, in (5), \(d_1\) and \(\mathbf{r}\) have to be replaced by \(d_2\) and \(\mathbf{r}(\{d_1\})\), respectively, to compute \(y_{d_2}\). Following this detection principle [1] equation (6) rewrites for an arbitrary
A solution is, only to cancel out those symbols that are recognized as correct with a high reliability. As a consequence we can avoid error propagation and gain diversity if the cancellation is correct. But the problem is how to find out those symbols? The decoding and error correction process starts after all symbols are soft output demapped. However, this is too late, because the BLAST cancellation is already done. Horizontal decoding [4] could help, but an efficient decoder needs to have independent and thus interleaved input. But a per-layer-interleaver in combination with symbol cancellation introduces detection time delay and complexity.

For the sake of completeness, the scheme in Fig. 2 is already prepared for iterative processing. The a priori information $A_1$ is subtracted from the a posteriori information $D_1$ given by the L-values in (7). After deinterleaving of $E_1$, the result $A_2$ is passed on as new a priori information to the soft input soft output APP decoder. The decoder calculates the new a posteriori information $D_2$. The extrinsic information $E_2$ is interleaved resulting in new a priori information $A_1$ for the detector and demapper. The feedback information $A_1$ can be used for iterative schemes as e.g. presented in [5], [6], [15]. After some iterations the output $L$-values $D'_2$ of the decoder are hard decided to obtain estimates $\tilde{u}$ on the information bits. In this paper we focus on the improvement of the very first detection step (0th iteration). Note, that for ZF $\alpha = 0$ and $\bar{H} = I_M$ resulting in (B) $= 0$ and $\bar{h}_{d_\mu}$, $d_\mu = 1$ for symbol $s_{d_\mu}$ ($\mu = 1, 2, \ldots, M$), but still (D) $\neq 0$.

B. Reliability measure for symbol decision

A measure for the reliability of a bit $b_k$ is the magnitude of its likelihood value (L-value) defined as $L(b_k) = \ln \frac{P[b_k=1]}{P[b_k=0]}$, where $P[\cdot]$ is the probability. As an extension, the reliability value $\mathcal{R}(s)$ of a symbol $s$ consisting of $Q$ bits can be based upon the L-values of these underlying bits. Two simple (different) measures are proposed as

$$\mathcal{R}(s) = \sum_{j=1}^{Q} |L(b_j)| \quad (9)$$

or

$$\mathcal{R}(s) = \min_{j=1,\ldots,Q} |L(b_j)| \quad (10)$$

During our simulations we found that (10) is a conservative approach and suited to achieve low BER. So we have taken this metric in our algorithm. Suppose we have just detected the symbol $y_{d_\mu}$, and have computed the L-values $L(x_{d_\mu}^k)$ ($k = 1, \ldots, Q$). Then we compute $\mathcal{R}(y_{d_\mu})$ as in (10). The probability density functions (pdf) of $\mathcal{R}(y_{d_\mu})$ is $p(\mathcal{R}(y_{d_\mu}))$. The key is, that we split this pdf into two parts and define

$$p_{c,y_{d_\mu}} = p(\mathcal{R}(y_{d_\mu}) | \hat{s}_{d_\mu} = s_{d_\mu})$$

and

$$p_{w,y_{d_\mu}} = p(\mathcal{R}(y_{d_\mu}) | \hat{s}_{d_\mu} \neq s_{d_\mu})$$

where the indices c and w stand for correct and wrong decision, respectively. The receiver of course does not know whether $\hat{s}_{d_\mu} = s_{d_\mu}$ or $\hat{s}_{d_\mu} \neq s_{d_\mu}$. But as can be seen from Fig. 3 the two distributions are quite different. In this figure we show the simulated distributions of the first symbol $y_{d_1}$ to be detected in a system with $M = N = 2$, QPSK, Gray mapping at an operation point of $E_b/N_0 = 10$ dB. During our investigation we observed that, if we change the number of bits per symbol, use anti-Gray mappings or operate at different $E_b/N_0$, the distributions obtained are basically of the same shape. Thus, the depicted distributions in Fig. 3 are typical to explain the scheme. The pdfs $p_{c,y_{d_1}}$ are significantly larger than $p_{w,y_{d_1}}$. In case of wrong estimation the underlying histogram has almost no entries for $\mathcal{R}(y_{d_1}) > 14$. The
simulated pdfs of the MMSE case differ a little bit from the ZF case, especially for low \( R = 0 \). In particular, for MMSE the two distributions \( p_{c,.y_{d_1}} \) and \( p_{w,.y_{d_1}} \) are better ‘separated’, i.e.:

\[
\int_0^\infty p_c[y_{d_1}] p_w[y_{d_1}] dR(y_{d_1}) < \int_0^\infty p_c[y_{d_1}] p_w[y_{d_1}] dR(y_{d_1})
\]

This leads us to the conclusion that our presented scheme is more advantageous for MMSE than for ZF algorithm.

In fact, if \( R(y_{d_1}) \) is high (e.g. \( R(y_{d_1}) = 15 \)) then it is very likely that \( \hat{s}_{d_1} = s_{d_1} \) because \( p[R(y_{d_1})|\hat{s}_{d_1} \neq s_{d_1}] \approx 0 \). On the other hand, if \( R(y_{d_1}) \) is low we are not sure whether \( \hat{s}_{d_1} = s_{d_1} \) or not. But with this knowledge we can set a threshold \( R_{th} \) and apply the following rule:

- cancel the detected symbol \( \hat{s}_{d_1} \) if \( R(y_{d_1}) > R_{th} \)
- do not cancel otherwise

**C. Definition of threshold**

The higher we choose the threshold the more conservative the cancellation process is. That means that if we only want to cancel those symbols which are extremely save, we have to choose a high threshold. For \( R_{th} \to \infty \) we do not cancel at all, whereas for \( R_{th} = 0 \) we have permanent cancellation. We need the demapper providing the L-values to be part of the detector to be able to decide whether to cancel or not.

It remains to find the optimal threshold that yields low BER. For this reason we have to find the optimum between two cases: If the threshold is too high we hardly cancel anything and cannot benefit from the gain of genie cancellation. Contrary, if the threshold is too low we run the risk of cancelling wrong symbols resulting in a worse performance. Fig. 4 shows the optimum threshold points \( R_{th} \) over \( E_b/N_0 \) found by simulation for QPSK, Gray mapping, and \( M = N = 2 \). In contrast to OSIC it is now very likely to happen that there are already detected symbols which will be deleted and others will not. This modifies the BLAST algorithm. In the calculation of the equalization row vectors \( \mathbf{w}_{d_1} \) we only have to null out those columns of the channel matrix \( \mathbf{H} \) whose corresponding symbols are actually detected and deleted. Following this principle, (8) becomes

\[
y_{d_1} = \tilde{h}_{d_1,d_1} s_{d_1} + \left( \sum_{j \in C_{d_1}} \tilde{h}_{d_1,j} \cdot s_j \right) + \tilde{n}_{d_1}
\]

\[
+ \sum_{j \in C_{d_1}} \mathbf{w}_{d_1,j} (s_j - \hat{s}_j)
\]

where \( C_{d_1} \) is the set of all cancelled symbols until this detection step and \( C_{d_1} \) are all not yet detected symbols plus the symbols that have been detected but not cancelled excluding the symbol to be detected in this step. Note that \( C_{d_1} \) is a subset of \( D_{d_1} \).

**IV. Simulation results**

Fig. 5 shows the simulation results for a system with \( M = N = 2 \) antennas. We included both, MMSE and ZF, with (a)
OSIC, (b) no cancellation at all, (c) genie cancellation, and (d) adaptive threshold-based cancellation. It can be seen that OSIC performs worse than all other schemes: the advantage of OSIC in an uncoded BLAST system is no longer visible due to too optimistic L-values as input signal to the decoder. We also recognize that at a BER of $10^{-4}$ there is a gain of about 3.5 dB between the genie cancellation scheme and the scheme where no cancellation is done. Our proposed adaptive threshold-based cancellation method at BER of $10^{-4}$ outperforms the no-cancellation-case by approximately 1 dB for ZF and 1.5 dB for MMSE. This is achieved without reordering the detection order and without any iteration. Thus, further iteration steps have got an improved starting point. For MMSE our proposal is just the execution of (10), which is almost negligible. If we omit cancellation due to a low minimum distance $d_{mn}$ the complexity reduces compared to full OSIC. It has to be pointed out that there is no iteration, reordering or additional interference cancellation required. To use the optimal threshold by applying a lookup table, the knowledge of the SNR is required. However, with MMSE equalization according to (4), $\alpha$ is also required. Thus, it has to be estimated in both cases.

VI. CONCLUSION

The BLAST algorithm is an effective way to detect output signals of a MIMO transmission system. BLAST can be operated with zero forcing (ZF) or minimum mean square error (MMSE). Both schemes can be combined with ordered successive interference cancellation (OSIC). In this paper we have presented an improved method to reduce the BER of a convolutionally encoded MIMO system. The problem of conventional BLAST with OSIC is that it works fine, if cancellation of correctly decided symbols is done. A relatively high SNR is a prerequisite of that. If the SNR is low, symbol decision errors occur which give rise to error propagation, which reduces the performance. In our proposal we follow the idea to cancel only symbols, which are decided correctly. As a consequence, we reduce the error propagation effect of the conventional OSIC. To do so, we define a reliability measure, which indicates the correctness of the symbol decision. We propose the minimal L-value $R(s)$ of the underlying bits of the symbol $s$ as a measure. From statistical evaluations of the conditional probabilities of $R(s)$ given in Fig. 3 we derive a threshold $R_{th}$ for cancellation, which depends on the SNR of the channel (Fig. 4). The proposed receiver is composed of a conventional BLAST detector with an adaptive threshold for cancellation, an MAP soft output demapper and an iterative (Turbo-) decoder. We have tested our algorithm by extensive computer simulations using Rayleigh fading MIMO channels with additive white Gaussian noise. As a result, we get an SNR gain of about 1 dB with zero forcing and 1.5 dB with minimum mean square error, both compared to the conventional BLAST algorithm. We have also shown that the hardware overhead of our proposal is very small.

REFERENCES