Iterative Nonlinear Detection of MIMO Signals using an MMSE-OSIC Detector and Adaptive Cancellation

Alexander Boronka, Tobias Rankl and Joachim Speidel

Institute of Telecommunications, University of Stuttgart, Pfaffenwaldring 47, D-70569 Stuttgart, Germany, {boronka,rankl,speidel}@inue.uni-stuttgart.de

Abstract

We present an improved algorithm for iterative detection of the received signal in a convolutionally encoded wireless MIMO system. The method is a combination of an iterative linear MMSE equalization and an enhanced version of the nonlinear ordered successive interference (OSIC) cancellation. To improve regular OSIC we introduce threshold-based adaptive cancellation. The threshold is switched and takes reliability of a precedent symbol decision into account. By this, the number of cancellations of wrongly decided symbols is reduced to a large extend. In addition, the MAP soft output demapper takes the so called Post Detection Interference and Noise (PDIN) into account. The variance of the PDIN is derived. As a result, we obtain an SNR gain of 3 dB at a BER of 10^{-4} compared to existing solutions for uncorrelated Rayleigh fading MIMO channels with additive white Gaussian noise. For a correlated channel we obtain a large improvement for the first pass of about 5 dB at a BER of 10^{-3}.

1 Introduction

In the initial publications [1], [2] the BLAST (Bell-Labs Layered Space-Time) algorithm emphasis is on zero-forcing (ZF) with ordered successive interference cancellation (OSIC). The minimum mean squared error (MMSE) solution is presented in e. g. [3] and extended to iterative linear MMSE detection in [4]. It is well known that OSIC reduces the bit error ratio (BER) of an *uncoded* BLAST system [3]. However, the behavior of a *coded* BLAST system normally 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 the coded bits $x'(t_x)$, interleaved resulting in $x(t_x)$, and mapped to $\mathbf{s}^{\mathrm{T}}(t_s)$ so that each of the M transmitters carries an independent data stream $s_j(t_s)$ ($j = 1, \ldots, M$). t_u, t_x, and t_s are the discrete time variables, 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 $G_{\mathrm{oct}} = 7$ are used. As usual, the polynomial is given in octal numbers. This structure was called *vertical coding* [5].

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

$$\mathbf{r} = H\mathbf{s} + \mathbf{n} \tag{1}$$

where $\mathbf{s} = (s_1, \ldots, s_M)^{\mathrm{T}}$ is the transmit symbol column vector. Each component s_j is a complex QAM symbol to be sent by antenna j ($j = 1, 2, \ldots, M$). $\mathbf{n} = (n_1, \ldots, n_N)^{\mathrm{T}}$ is an additive noise column vector with components n_j, which are complex AWGN at receive antenna j, each with zero mean and variance

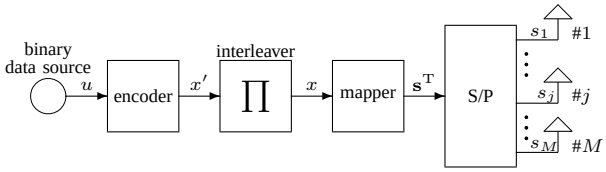

Fig. 1: Transmitter

σ_{n}^2 ($j = 1, 2, \ldots, N$). This means that the complex n_j is Gaussian with variance $\sigma_{\mathrm{n}}^2/2 = N_0/2$ per real component. N is the number of receive antennas. We consider $\mathbf{n}$ to be uncorrelated with expected value $\mathsf{E}\{\mathbf{n}\mathbf{n}^*\} = \sigma_{\mathrm{n}}^2 I_N$. I_N is the identity matrix of size N. Furthermore, it is assumed that the vectors $\mathbf{s}$ and $\mathbf{n}$ are uncorrelated. Together with $\mathsf{E}\{\mathbf{n}\} = 0$ follows $\mathsf{E}\{\mathbf{s}\mathbf{n}^*\} = 0$. $\mathbf{r} = (r_1, \ldots, r_N)^{\mathrm{T}}$ is the received symbol column vector, where r_j is received by antenna j ($j = 1, 2, \ldots, N$). The channel matrix H is given by $H = (\mathbf{h}_1 \ldots \mathbf{h}_M)$ where $\mathbf{h}_i = (h_{1i}, \ldots, h_{Ni})^{\mathrm{T}}$ is a column vector of H. The impulse response h_{ji} from transmitter i to receiver j is modelled as a zero-mean, complex Gaussian random variable satisfying $\mathsf{E}\{|h_{ij}|^2\} = 1$ (i. e. the channel is passive). All entries of H are i.i.d. $(\cdot)^{\mathrm{T}}$ 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 $E_b/N_0 =$

$\frac{E_s \cdot N}{R_c \cdot Q \cdot \sigma_n^2}$ where $\frac{E_s \cdot N}{R_c \cdot Q}$ is the transmitted energy per information bit at the receiver and Q is the number of bits per symbol. Using the column vectors of H we can write (1) as

$$\mathbf{r} = \sum_{i=1}^{M} \mathbf{h}_i s_i + \mathbf{n} \tag{2}$$

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

The paper is organized as follows. In Sections 2 and 3 we describe the OSIC receiver and introduce what we call adaptive cancellation. In Section 4 simulation results are presented and discussed. Section 5 concludes this paper.

2 The Receiver

We combine two principal detection methods, namely the iterative linear MMSE detection algorithm presented in [4] and the ordered successive interference cancellation (OSIC) introduced in [1]. We call this combination MMSE-OSIC. In addition, we introduce a variable threshold to control *nonlinear* OSIC. However, to our knowledge the algorithm in [4] is the best *linear* MMSE detection scheme for MIMO systems. We refer to the MMSE algorithm described in [4] as *original MMSE*.

2.1 Principle Receiver Structure

The considered receiver is depicted in Fig. 2. The received signal, namely the N received samples r_j $(j = 1, 2, \ldots N)$ are detected with a soft-in soft-out MMSE Detector. The inherent soft-output MAP (maximum a posteriori) demapper delivers the a posteriori information D_1. Information D_1 becomes the extrinsic information E_1 after subtracting the a priori information A_1. After deinterleaving, A_2 is decoded with an APP decoder [6], [7]. The extrinsic information stemming from $E_2 = D_2 - A_2$ becomes after interleaving the new a priori information A_1 which is needed by the detector for the iterative detection process. After several iterations are processed, finally a hard decision on D_2' is done, resulting in the output information bits $\widehat{u}$. We use log-likelihood L-values.

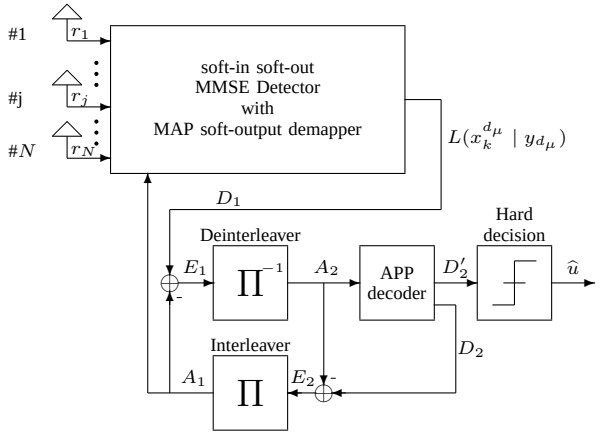

Fig. 2: Receiver structure

One iteration involves feedback, detection and decoding. Writing *detection* we also include the required *channel equalization* [8].

2.2 The MMSE-OSIC Detector

In general the soft-in soft-out detector in Fig. 2 consists of M separate and parallel detector branches, one for each component of the transmit vector $\mathbf{s}$. However, with OSIC algorithm the branches are processed *sequentially* during detection process, which simplifies hardware implementation. It is possible and sufficient to use only one branch M times per symbol vector. The detector branch for the $d_\mu{}^{\text{th}}$ detection step is depicted in Fig. 3. Besides the received signal $\mathbf{r}$, the detector processes the a priori information A_1 from the previous iteration provided by the decoder, and in addition the already detected symbols $y_{d_1}, \ldots, y_{d_{\mu-1}}$ during the current OSIC detection. To describe this accurately we introduce the ordered sequence $\mathcal{D} = \{d_1, d_2, \ldots, d_M\}$, which represents the OSIC detection order. The optimum order will be derived in Subsection 2.3. For the $d_\mu{}^{\text{th}}$ detection step we get $\mathcal{D}_{d_\mu} = \{d_1, d_2, \ldots, d_{\mu-1}\}$ the index set of all yet detected symbols. $\overline{\mathcal{D}_{d_\mu}} = \{d_{\mu+1}, d_{\mu+2}, \ldots, d_M\}$ is the index set of all not yet processed symbols excluding the actual symbol s_{d_μ}. Thus, we get

$$\mathcal{D}_{d_\mu} \cup \{d_\mu\} \cup \overline{\mathcal{D}_{d_\mu}} = \mathcal{D} \ . \tag{3}$$

To obtain the MMSE estimate y_{d_μ} of the $d_\mu{}^{\text{th}}$ symbol we first perform an interference cancellation. The estimates $\hat{s}_{d_\mu} = \mathcal{Q}\left(y_{d_\mu}\right)$ $(j \in \mathcal{D}_{d_\mu})$ of yet detected symbols are "hard cancelled". The remaining symbols are "soft cancelled". By "hard cancellation" we mean that we subtract the already estimated and quantized symbols from the symbol vector, as it is known from the original OSIC algorithm [1]. By "soft cancellation" [9] we mean that we use the soft information provided

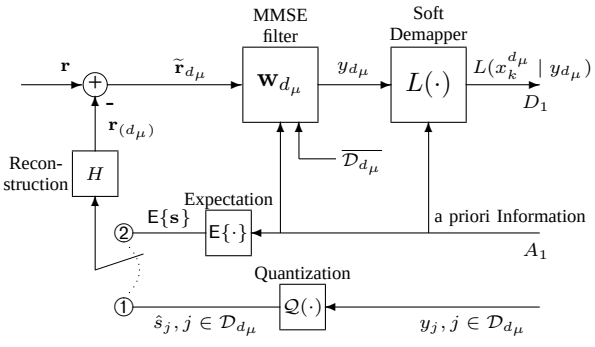

Fig. 3: Detector branch of detection step d_μ as part of the soft-in soft-out detector in Fig. 2

by the previous iteration to delete a soft rather than a quantized symbol. Using the notation (3), it is clear that symbols represented by $\mathcal{D}_{d_\mu}$ undergo hard cancellation, whereas symbols represented by $\overline{\mathcal{D}_{d_\mu}}$ are treated by soft cancellation. This is indicated by the switch in Fig. 3. Now, while processing the $d_\mu{}^{\text{th}}$ symbol, for the interfering symbols $\mathcal{D}_{d_\mu} = \{d_1, d_2, \ldots, d_{\mu-1}\}$ the switch is in position ① and for the interfering symbols $\overline{\mathcal{D}_{d_\mu}} = \{d_{\mu+1}, d_{\mu+2}, \ldots, d_M\}$ the switch is in position ②. After reconstruction of the hard or soft cancelled symbols with the channel matrix we subtract this value $\mathbf{r}_{(d_\mu)}$ from the observations $\mathbf{r}$. The interference reduced signal results in

$$\widetilde{\mathbf{r}}_{d_\mu} = \mathbf{r} - \underbrace{\sum_{j \in \overline{\mathcal{D}_{d_\mu}}} \mathbf{h}_j \mathsf{E}\{s_j\} - \sum_{j \in \mathcal{D}_{d_\mu}} \mathbf{h}_j \hat{s}_j}_{\mathbf{r}_{(d_\mu)}} \quad . \tag{4}$$

To further suppress the residual interference, $\widetilde{\mathbf{r}}_{d_\mu}$ is filtered with the equalization vector $\mathbf{w}_{d_\mu}$ yielding the MMSE estimate

$$y_{d_\mu} \quad = \quad \mathbf{w}_{d_\mu}^* \widetilde{\mathbf{r}}_{d_\mu} \tag{5}$$

where $\mathbf{w}_{d_\mu} = \left(w_{d_\mu,1}, w_{d_\mu,2}, \ldots, w_{d_\mu,N}\right)^{\mathrm{T}}$ is chosen to minimize the mean squared error $\mathsf{E}\{|s_{d_\mu} - y_{d_\mu}|^2\}$, as in [4] or [8]. Under the assumption that the variance of a quantized, and thus, exactly known symbol is zero, the MMSE filter coefficient vector for the given MIMO system results in

$$\mathbf{w}_{d_\mu} = \left[\sum_{\substack{j \in \overline{\mathcal{D}_{d_\mu}} \\ j = d_\mu}} \mathbf{h}_j \ \sigma_{\mathrm{ap},j}^2 \ \mathbf{h}_j^* + \sigma_{\mathrm{n}}^2 I_N \right]^{-1} E_s \mathbf{h}_{d_\mu} \tag{6}$$

where $\sigma_{\mathrm{ap},j}^2$ is the variance of the a priori knowledge. For $j = d_\mu$ the a priori variance $\sigma_{\mathrm{ap},j}^2$ is equal to E_s. To calculate (6), the MMSE filter also has to know $\overline{\mathcal{D}_{d_\mu}}$. This is also drawn in Fig. 3. If we insert (2) and (4) into (5) and introduce a so called *Post Detection Channel* (PDC) vector

$$\widetilde{\mathbf{h}}_{d_\mu} = \mathbf{w}_{d_\mu}^* H \tag{7}$$

and a *Post Detection Gaussian Noise* (PDGN) value $\widetilde{n}_{d_\mu} = \mathbf{w}_{d_\mu}^* \mathbf{n}$, the MMSE estimate results in

$$\begin{aligned} y_{d_\mu} = & \underbrace{\widetilde{h}_{d_\mu,d_\mu} s_{d_\mu}}_{(A)} + \underbrace{\sum_{j \in \overline{\mathcal{D}_{d_\mu}}} \widetilde{h}_{d_\mu,j} \left(s_j - \mathsf{E}\{s_j\}\right)}_{(B)} \\ & + \underbrace{\sum_{j \in \mathcal{D}_{d_\mu}} \widetilde{h}_{d_\mu,j} \left(s_j - \hat{s}_j\right)}_{(C)} + \underbrace{\widetilde{n}_{d_\mu}}_{(D)} \end{aligned} \tag{8}$$

where $\widetilde{h}_{i,j}$ is the j^{th} element of $\widetilde{\mathbf{h}}_i$. Note that $\widetilde{\mathbf{h}}_i$ is a row vector whereas $\mathbf{h}_i$ is a column vector.

Term (*A*) is the desired Symbol s_{d_μ} weighted with its *Post Detection Channel* (PDC) coefficient. Term (*B*) is the *Post Detection Inter Symbol Interference* (PD-ISI) which consists of all other not yet processed symbols weighted with their corresponding PDC coefficient. Term (*C*) represents the *Post Cancellation Interference* (PCI), i. e. the summation of all symbols that have been detected and cancelled. Term (*D*) is the PDGN of symbol s_{d_μ}. From (8) it can be seen, that all the transmitted symbols s_j $(j \neq d_\mu)$ disturb s_{d_μ} either as PD-ISI or as PCI if $s_j \neq \hat{s}_j$. Term (*C*) vanishes for correct decided symbols $s_j = \hat{s}_j$. Altogether, terms (*B*), (*C*), (*D*) can be regarded as *Post Detection Interference and Noise* (PDIN) and we may write in words

$$y_{d_\mu} = \text{PDC} \cdot s_{d_\mu} + \text{PD-ISI} + \text{PCI} + \text{PDGN} \tag{9}$$

The subsequent soft demapper in Fig. 3 uses both, the MMSE estimate of the transmitted symbol and the PDIN variance (to be derived below) for computation of the L-values. Since $\mathbf{n}$ is Gaussian (i.e. all entries have Gaussian pdf) $\widetilde{n}_{d_\mu}$ of term *(D)* can be seen as a realization of a Gaussian process with variance

$$\sigma_{\mathrm{PDGN},d_\mu}^2 = \sigma_{\mathrm{n}}^2 \sum_{j=1}^{N} |w_{d_\mu,j}|^2 \tag{10}$$

We can also compute the variance (var) of the PD-ISI (term *(B)*):

$$\sigma_{\mathrm{PD-ISI},d_\mu}^2 = \sum_{j \in \overline{\mathcal{D}_{d_\mu}}} |\widetilde{h}_{d_\mu,j}|^2 \sigma_{\mathrm{ap},j}^2 \tag{11}$$

The conditioned variance of the PCI (term *(C)*) is given by

$$\sigma_{\mathrm{PCI},d_\mu}^2 = \sum_{j \in \mathcal{D}_{d_\mu}} |\widetilde{h}_{d_\mu,j}|^2 \mathrm{var}\left(s_j - \hat{s}_j | y_{d_\mu}\right) \tag{12}$$

The problem is, that s_j is unknown to the receiver. Just the estimate $\hat{s}_j$ is available. Thus, the variance in (12) is calculated by the use of the expected values

$$E\left\{s_j - \hat{s}_j | y_{d_\mu}\right\} = \sum_{r=1}^{2^Q} (a_r - \hat{s}_j)\, p(a_r | y_{d_\mu}) \tag{13}$$

and

$$E\left\{|s_j - \hat{s}_j|^2 | y_{d_\mu}\right\} = \sum_{r=1}^{2^Q} |a_r - \hat{s}_j|^2 p(a_r | y_{d_\mu}) \tag{14}$$

resulting in

$$\begin{aligned} \mathrm{var}\left(s_j - \hat{s}_j | y_{d_\mu}\right) = & \sum_{r=1}^{2^Q} |a_r - \hat{s}_j|^2 p(a_r | y_{d_\mu}) \\ & - \left| \sum_{r=1}^{2^Q} (a_r - \hat{s}_j) \ p(a_r | y_{d_\mu}) \right|^2 \end{aligned} \tag{15}$$

where a_r $(r = 1 \ldots 2^Q)$ are the symbols of the used QAM constellation diagram $\mathcal{A}$. Applying the Bayes'

rule and assuming initially equally distributed transmit symbols the pdf in (15) results in

$$p\ (a_r|y_{d_\mu}) = \frac{\exp\left(-\frac{|y_{d_\mu}-\mu_{d_\mu,r}|^2}{\widetilde{\sigma}^2_{d_\mu}}\right)}{\sum_{r=1}^{2Q}\exp\left(-\frac{|y_{d_\mu}-\mu_{d_\mu,r}|^2}{\widetilde{\sigma}^2_{d_\mu}}\right)} \qquad (16)$$

where $\mu_{d_\mu,r} = \widetilde{h}_{d_\mu,d_\mu}a_r$ is a symbol affected by the PDC.

To yield the hard decided estimate $\hat{s}_j$, (17) is applied:

$$\hat{s}_{d_\mu} = \mathcal{Q}\left(y_{d_\mu}\right) = \arg\underbrace{\min}_{a_r\in\mathcal{A}}\left\{|y_{d_\mu}-\mu_{d_\mu,r}|^2\right\} \qquad (17)$$

Now, with (16), (17), and (15), equation (12) can be computed. To avoid numerical instabilities in practical implementations this computation should be executed in the log domain.

The distribution of the PD-ISI and the PCI is not Gaussian. However, to simplify the scheme, we summarize interference terms (*B*), (*C*) and (*D*) into one single realization of a random process with Gaussian distribution (Gaussian assumption [10]) and variance

$$\widetilde{\sigma}^2_{d_\mu} = \sigma^2_{\mathrm{PD-ISI},d_\mu} + \sigma^2_{\mathrm{PCI},d_\mu} + \sigma^2_{\mathrm{PDGN},d_\mu} \qquad (18)$$

which is the PDIN variance to be used for L-value computation.

The soft demapper in Fig. 3 performs bit-wise soft-output demapping described in [11]. Thus, the L-value of the k^{th} bit $x_k^{d_\mu}$ of symbol y_{d_μ} is given by

$$L(x_k^{d_\mu}\mid y_{d_\mu}) = L_a\left(x_k^{d_\mu}\right) + \ln\frac{\sum_{\nu=0}^{2^{Q-1}-1} p\left(y_{d_\mu}|x_k^{d_\mu}=1,x_{j,j=0,\ldots,Q-1,j\neq k}^{d_\mu}\right)\mathrm{e}^{\sum_{j=0,j\neq k}^{Q-1} x_j^{d_\mu}L_a\left(x_j^{d_\mu}\right)}}{\sum_{\nu=0}^{2^{Q-1}-1} p\left(y_{d_\mu}|x_k^{d_\mu}=0,x_{j,j=0,\ldots,Q-1,j\neq k}^{d_\mu}\right)\mathrm{e}^{\sum_{j=0,j\neq k}^{Q-1} x_j^{d_\mu}L_a\left(x_j^{d_\mu}\right)}} \qquad (19)$$

The values of the bits $x_j^{d_\mu}$ in (19) accomplish the following equation (see [11] for details):

$$\sum_{j=0}^{k-1} x_j^{d_\mu}\cdot 2^j + \sum_{j=k+1}^{Q-1} x_j^{d_\mu}\cdot 2^{j-1} = \nu\ . \qquad (20)$$

Eq. (19) means that we set bit $x_k^{d_\mu}$ to 1 in the numerator and to 0 in the denominator and permute over all other bits within the symbol. The a priori information $L_a\left(x_k^{d_\mu}\right)$ is provided by the previous iteration. For the very first iteration (0^{th} iteration) this information is set to zero. Note, that the soft 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.

2.3 Detection Order

In the previous Section we introduced the detection order as $\mathcal{D} = \{d_1, d_2, \ldots, d_M\}$. Now we show how to determine the best detection order throughout the OSIC detection process. Therefor, the minimum mean squared error is evaluated, from which the detection order is generated.

The mean squared error between the transmitted symbol vector $\mathbf{s}$ and the detected signal vector $\mathbf{y}$ is given by

$$\begin{aligned}\varepsilon^2 &= \mathsf{E}\{(\mathbf{s}-\mathbf{y})^*(\mathbf{s}-\mathbf{y})\}\\ &= \sum_{i=1}^{M}\mathsf{E}\{(s_i-y_i)^*(s_i-y_i)\}\\ &= \sum_{i=1}^{M}\varepsilon_i^2 \end{aligned} \qquad (21)$$

with $\varepsilon_i^2 = \mathsf{E}\{(s_i-y_i)^*(s_i-y_i)\}$. With (8) we obtain after some algebra

$$\begin{aligned}\varepsilon_i^2 &= \sigma^2_{\mathrm{ap},i}(1-\widetilde{h}_{i,i})^2\\ &\quad + \sigma^2_{\mathrm{PD-ISI},i} + \sigma^2_{\mathrm{PCI},i} + \sigma^2_{\mathrm{PDGN},i}\\ &= \sigma^2_{\mathrm{ap},i}(1-\widetilde{h}_{i,i})^2 + \widetilde{\sigma}_i^2\end{aligned} \qquad (22)$$

We point out that $\widetilde{h}_{i,i}$ is a real number as proven in the appendix. From (22) we also observe the following: For high SNR the last term becomes very small or almost zero ($\sigma^2_{\mathrm{PDGN},i}\approx 0$). If we have genie symbol cancellation then in addition $\sigma^2_{\mathrm{PCI},i}=0$. For this case the error ε_i^2 depends on the variance of the interferers ($\sigma^2_{\mathrm{PD-ISI},i}$) and the difference $1-\widetilde{h}_{i,i}$. Thus, in the noiseless case the ZF filter is the optimum where $\widetilde{h}_{i,j}=1$ for $i=j$ and 0 for $i\neq j$. Also, it is clear from (22) that ε_i^2 increases if the PDC $\widetilde{h}_{i,i}$ differs from 1.

While processing the OSIC algorithm, it is necessary to compute the best symbol for detection within each detection step. This means, in detection step μ we have to compute the mean squared error ε_i^2 for all not yet detected symbol and choose the best, meaning the minimum element, as shown in [1] and proven in [2]. The chosen symbol is processed as number d_μ. In other words, we calculate ε_i^2 for all not yet processed symbols and choose the smallest one. The computational complexity increases, but using the wrong detection order leads to a huge decrease of performance especially in the later iterations.

2.4 Adaptive Cancellation

In this subchapter we will introduce a new method for symbol cancellation which we call adaptive cancellation. The method results in reduced BER.

[12] describes this adaptive cancellation in detail for a non-iterative system. The idea is to omit cancellation if it is likely that the quantized symbol is wrong. This means we only cancel out those symbols that

are recognized to be correct with high probability. As a measure for this cancellation decision we use the minimum L-value of a symbol

$$\mathcal{R}(s) = \underbrace{\min}_{j=1\ldots Q} |L(b_j)| \tag{23}$$

where b_j is the j^{th} bit of symbol s and $L(\cdot)$ is the L-value operator. One could use the a posteriori L-value or the extrinsic L-value. Note, that for the very first pass, this is the same. Simulation results show that the use of extrinsic L-values achieves slightly lower BER in the first iteration. However, this improvement is negligible.

The rule for adaptive cancellation is as follows:

- Choose proper threshold $\mathcal{R}_{\text{th}}$
- Cancel the detected and quantized symbol $\hat{s}_{d_\mu}$ if $\mathcal{R}\left(y_{d_\mu}\right) > \mathcal{R}_{\text{th}}$.
- Otherwise use soft cancellation.

As described in [12], the choice of the threshold $\mathcal{R}_{\text{th}}$ is subject to optimization with respect to the resulting BER. If we set $\mathcal{R}_{\text{th}} = 0$ hard cancellation of all symbols $\hat{s}_j$ ($j = 1, \ldots, M$) is performed. For $\mathcal{R}_{\text{th}} = \infty$ we do not hard cancel at all but perform the original MMSE algorithm [4]. The use of an optimized threshold for each SNR step and each iteration leads to the best results. A brief description how to obtain these optimized threshold is given in Section 4.2.

The algorithm described in Subsection 2.2 can be easily adapted to this adaptive cancellation method. To do so, we split $\mathcal{D}_{d_\mu}$ into two subsets

$$\mathcal{D}_{d_\mu} = \mathcal{D}'_{d_\mu} \cup \mathcal{D}''_{d_\mu} \tag{24}$$

where $\mathcal{D}'_{d_\mu}$ is the set of all detected and hard cancelled symbols and $\mathcal{D}''_{d_\mu}$ is the set of all detected but *not* hard cancelled symbols. As before $\overline{\mathcal{D}_{d_\mu}}$ the set of all not yet processed symbols. As a consequence we get

$$\mathcal{D} = \mathcal{D}'_{d_\mu} \cup \{d_\mu\} \cup \overline{\mathcal{D}_{d_\mu}} \cup \mathcal{D}''_{d_\mu} \ . \tag{25}$$

Now, we set

$$\mathcal{C}_{d_\mu} = \mathcal{D}'_{d_\mu} \tag{26}$$

$$\overline{\mathcal{C}_{d_\mu}} = \overline{\mathcal{D}_{d_\mu}} \cup \mathcal{D}''_{d_\mu} \tag{27}$$

and replace

$$\mathcal{D}_{d_\mu} \longrightarrow \mathcal{C}_{d_\mu} \qquad \text{and} \qquad \overline{\mathcal{D}_{d_\mu}} \longrightarrow \overline{\mathcal{C}_{d_\mu}} \tag{28}$$

in (8), (11), (12), and in Fig. 3.

3 OSIC in Iterative Receivers

This chapter presents a detailed description of what happens during execution of the OSIC algorithm and during the iterations. [1] describes OSIC by using operations like nulling, subtracting the detected symbol from the received signal, and permutating columns of matrices for a subsequent computation of the matrix inverse. These operations are also inherent in the new MMSE-OSIC algorithm but do not have to be executed explicitly. Instead, the use of a priori information during processing automatically performs the mentioned operations. The extension of iterative original MMSE [4] to iterative MMSE-OSIC is a migration from parallel to serial processing and an override (or update) of the a priori information by the use of detected and quantized symbols. For example, if the detected symbol is exactly known (i.e. the ideal case of hard cancellation), then its variance is set zero and the expected values is set to be the quantized symbol itself. In the iterations, we have a priori information stemming from the decoder. With this additional information, a mixture of soft and hard cancellation is performed. All yet detected symbols are hard cancelled if the threshold decision permits to do so. The remaining symbols undergo soft cancellation.

3.1 The Detection Process

We consider a complete symbol vector to be detected. Besides the observation $\mathbf{r}$, we assume to have full knowledge about the channel matrix H. In addition, we can use a priori information of the symbols respective their underlying bits which is of course zero at the beginning but contains useful information in the following iterations.

At the beginning the sequence $\mathcal{D}$ is unknown. Thus, we have to determine the first element to be detected. Therefore, each ε_i^2 $(i = 1 \ldots M)$ for every transmitted symbol has to be determined. Here we should mention, that it is necessary to compute all filter vectors $\mathbf{w}_i$ $(i = 1, \ldots, M)$ according to (6) to be able to calculate the mean squared error ε_i^2. If all ε_i^2 are computed, we chose the smallest one. Now we know the symbol to be detected first.

We set the index of the first symbol to be detected to d_1. Now, after the calculation of the expected values and the variance, the interference reduction can be performed, yielding the interference reduced signal $\widetilde{\mathbf{r}}_{d_1}$. After filtering with filter $\mathbf{w}_{d_1}$ we get y_{d_1} as the MMSE estimate of the transmitted symbol s_{d_1}. To operate the subsequent soft demapping we determine the PDIN variance $\widetilde{\sigma}_{d_1}^2$ according to (10) - (18) and execute the demapping.

In case of adaptive cancellation, we decide if the symbol is reliable enough for cancellation. If necessary, we update the expected values and the variance stemming from the a priori information of the current symbol with the new values. Then, the detection of the first symbol s_{d_1} is completed.

The detection of the following symbol starts again with determination of the best choice of detection, i.e. to select the symbol with minimum ε_i^2. The choice has to be done among all s_{d_j} with $d_j \in \overline{\mathcal{D}_{d_1}}$. The calculation of all remaining ε_i^2 and the choice of the smallest element leads to the next symbol i.e. s_{d_2}. A special case for the calculation of the ε_i^2 occurs if we do not perform cancellation within the previous detection step due to rejected cancellation. This case occurs, if the reliability of the detected symbol was too small for

cancellation and thus, no cancellation was done. Then the detection order does not have to be calculated again. The ε_i^2 will not be changed and the choice of the next element for detection can be based upon the remaining ε_i^2 (excluding the already detected values) stemming from the previous detection step. Now interference reduction, filtering and L-value determination has to be executed. Depending on the result of the L-value analysis the cancellation is performed or not. This procedure is repeated until all M symbols are detected and holds for all iterations.

4 Simulation Results

We show simulation results of a MIMO system with $M = N = 4$ transmitters and receivers, QPSK, and Gray mapping.

4.1 Original MMSE vs. MMSE-OSIC

In Fig. 4 different receiver schemes are compared. The solid curve with lowest BER holds for the interference free transmission. Such a system owns the maximum gain of diversity. It is not possible to reach better performance with any MMSE or ZF detection algorithm. The interference free transmission equals genie cancellation [9] or genie a priori knowledge. Thus, it can be regarded as a lower bound for BER. The uppermost curve represents the original *uncoded* V-BLAST with MMSE-OSIC which was just placed between encoder and decoder without considering any variance calculations. Especially, the PCI is not considered. As can be seen, BER for high SNR is rather bad. Therefore, it can be regarded as an upper bound. The BER curves with *unfilled* markers are simulated by the use of the iterative original MMSE detection algorithm presented in [4] to our best knowledge. We have plotted 3 iterations. Note, that we refer to the very first pass as the 0^{th} iteration. At last, the curves with the *filled* markers belong to our iterative MMSE-OSIC with adaptive cancellation. For this simulation we have set the threshold to $\mathcal{R}_{\text{th}} = 0$ for the 0^{th} iteration and $\mathcal{R}_{\text{th}} = \infty$ for all other iterations. As a result we see a large gain of about 2 dB for BER of about $3 \cdot 10^{-4}$ of the iterative MMSE-OSIC algorithm over the iterative original MMSE algorithm in the 0^{th} iteration. This is an important gain for hard- or software implementation with reduced complexity for practical implementations, where time-consuming iterations are not feasible. In the first iteration the MMSE-OSIC algorithm also outperforms the original MMSE algorithm. However, the gain is only small. At last, for the second iteration both detection schemes show almost the same performance. This holds for all further iterations as well, where MMSE-OSIC and original MMSE have practically identical BER. Both, original MMSE and MMSE-OSIC reach the lower bound BER given by the interference free transmission.

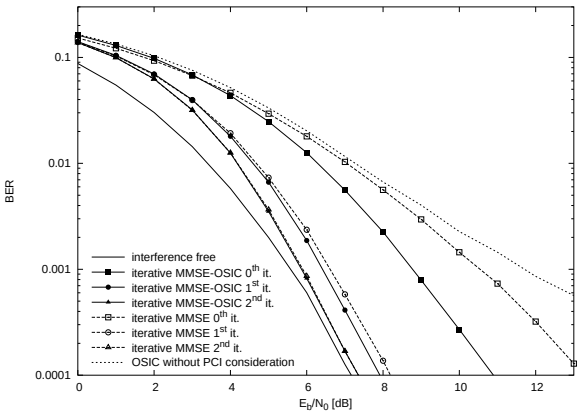

Fig. 4: BER $M = N = 4$, QPSK, Gray mapping, original MMSE and MMSE-OSIC

4.2 Optimized Thresholds

As mentioned above, similarly as in [12], adaptive cancellation in conjunction with optimized thresholds should further improve BER. The threshold $\mathcal{R}_{\text{th}}$ has to be optimized independently for all iteration steps and different SNR values. Thus, $\mathcal{R}_{\text{th}} = f(\#\text{it., SNR})$. For the 0^{th} iteration this can easily be done using the EXIT chart. In detail, one has to maximize extrinsic information E_1 for no a priori input, i. e.

$$\max_{\mathcal{R}_{\text{th}}} E_1(A_1 = 0) \ . \tag{29}$$

For the following iterations, we used the brute search to determine the optimized thresholds.

Table I gives the optimized thresholds we found by simulations. The first column is the SNR. The second column gives the optimal thresholds for the 0^{th} iteration. The third and forth column give the optimal thresholds in the 1^{st} iteration when applied to the a posteriori L-value and extrinsic L-value, respectively.

TABLE I: Optimized Thresholds

SNR	0^{th} it.	1^{st} it. (a post.)	1^{st} it. (extr.)
6 dB	1.0	13.0	5.0
7 dB	1.0	19.0	9.0
8 dB	0.95	23.0	12.5
9 dB	0.9	25.0	22.0
10 dB	0.9	n.d.[a)]	n.d.
11 dB	0.9	n.d.	n.d.
12 dB	0.9	n.d.	n.d.

[a)] n.d. = not determined.

We observe that for the 0^{th} iteration, the optimal threshold decreases for increasing SNR. This is clear since higher SNR means smaller L-values of wrong symbols [12]. For the 1^{st} iteration the optimum threshold increases for increasing SNR. This is also clear since the coding gain is higher for higher SNR. The thresholds for the a posteriori case are higher than for the extrinsic case because the a posteriori L-value

consist of the summation of extrinsic and a priori value.

Fig. 5 shows that adaptive cancellation with optimized thresholds can further improve BER performance. The lower bound is given by the solid line. The two curves with the black triangles stem from the adaptive MMSE-OSIC with optimized thresholds (extrinsic case) and have best performance. The curves with unfilled markers represent the MMSE-OSIC with thresholds chosen as in Fig. 4. The two remaining curves represent the original MMSE.

As expected, optimized thresholds give additional gains in the 0^{th} as well as in the 1^{st} iteration. We also observe that the gain increases for increasing SNR. For the 1^{st} iteration we outperform original MMSE by about 0.5 dB.

4.3 Correlated Channel

Finally, we investigate the behavior of our scheme in a correlated channel. Using a widely accepted channel model [13], the MIMO channel with transmit correlation can be described by the matrix product

$$H_{\text{corr}} = HB \tag{30}$$

where H is the uncorrelated MIMO channel matrix and B introduces the transmit correlation. Assuming B hermitian and matrix R to be the long-term stable (normalized) transmit correlation matrix [13] we define B to fulfill

$$BB^* = B^*B = R\ . \tag{31}$$

With an angular spread of 10° we get

$$R = \begin{pmatrix} 1 & 0.43-0.78\imath & -0.36-0.55\imath & -0.45+0.05\imath \\ 0.43+0.78\imath & 1 & 0.43-0.78\imath & -0.36-0.55\imath \\ -0.36+0.55\imath & 0.43+0.78\imath & 1 & 0.43-0.78\imath \\ -0.45-0.05\imath & -0.36+0.55\imath & 0.43+0.78\imath & 1 \end{pmatrix} \tag{32}$$

where $\imath$ is the imaginary unit. The simulation result can be seen in Fig. 6. We used full cancellation for the 0^{th} iteration and *no* cancellation for all other iterations. Thus, no threshold optimization was executed for the correlated channel case. Our scheme has an advantage of about 5 dB at a BER of 10^{-3} in the first pass. However, after one iteration this gain is gone and both original MMSE and MMSE-OSIC have about the same performance. Moreover, after 5 iterations original MMSE outperforms MMSE-OSIC. In this case, successive interference cancellation has a negative effect within the iterations.

4.4 Some Remarks on the Computational Complexity

To compare the computational effort we only focus on matrix inversions. The original MMSE algorithm has to compute one inverse of an $N \times N$-matrix in the 0^{th} iteration (similar to (6)) and M inverses of different $N \times N$-matrices in every following iteration. Due to the search of the optimal ordering, MMSE-OSIC with full cancellation has to compute 4 inverses of different $N \times N$-matrices in the 0^{th} iteration and $\frac{M(M+1)}{2}$ inverses of different $N \times N$-matrices in every following iteration.

Applying adaptive threshold-based cancellation reduces the amount of matrix inversions. The reason is that we can re-use an already computed inverse again if cancellation is omitted. This of course depends on the SNR, on the threshold, and on the actual threshold decision in each step. Thus, it is not possible to give a closed form of the computational complexity in case of adaptive MMSE-OSIC. However, by simulation we found that for an SNR of 6 dB and an optimum threshold of 1.0 for a system with $M = N = 4$, about 14 % of all symbols are not cancelled in the 0^{th} iteration. This means, that we save about 14 % in terms of computational complexity when we use adaptive cancellation instead of MMSE-OSIC with full cancellation. Though we save processing time, the BER is lower.

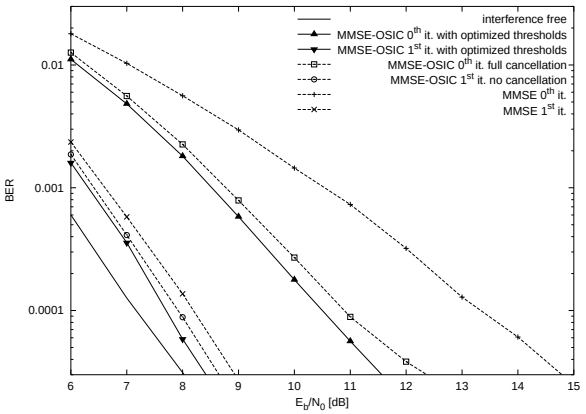

Fig. 5: BER $M = N = 4$, QPSK, Gray mapping, adaptive cancellation with optimized thresholds (extr.), 0^{th} and 1^{st} iteration

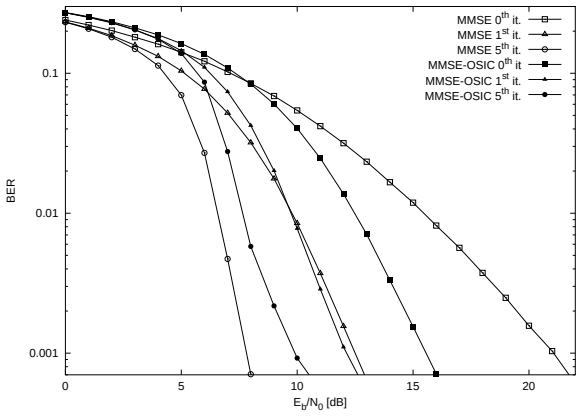

Fig. 6: BER $M = N = 4$, QPSK, Gray mapping, correlated channel, original MMSE and MMSE-OSIC without threshold optimization

5 Conclusion

Based on an iterative linear MMSE MIMO detection scheme [4], we presented an improved iterative nonlinear MMSE-OSIC algorithm, which employs the well-known ordered successive interference cancellation. To do so, it was essential to take all possible interference terms into account. Namely, the Post Detection Gaussian Noise (PDGN), the Post Detection Inter Symbol Interference (PD-ISI), and the Post Cancellation Interference (PCI). Their variances have been computed and applied in the soft output MAP demapper. During processing an optimum detection ordering was also calculated.

Furthermore, we introduced threshold-based adaptive cancellation to further reduce BER. The advantage of soft symbol cancellation is that it can mitigate the effects of error propagation. On the other hand it can not fully employ the positive effects of diversity gain. In contrast, hard symbol cancellation can employ the diversity gain if the estimates on the symbols are correct. However, wrong estimates enforce error propagation and lead to higher BER. The presented adaptive cancellation takes the reliability of a symbol decision into account to decide for soft or hard cancellation to combine the positive aspects of both methods. A threshold is a measure for adaptive cancellation. We presented simulation results on optimum thresholds.

Simulation results for a MIMO system with $M = N = 4$ transmitters and receivers, QPSK with Gray mapping, and Rayleigh fading channel with AWGN show that the gain of MMSE-OSIC over original MMSE is about 2 dB for BER of about $3 \cdot 10^{-4}$ for the 0^{th} iteration. The first iteration can still provide a small gain. For the second and all further iterations MMSE-OSIC and original MMSE have almost the same performance. However, for practical implementation the performance of the first few iterations is most important.

Using optimized threshold, we can further improve BER though the computational complexity decreases. The additional gain from MMSE-OSIC to adaptive MMSE-OSIC is up to about 0.5 dB.

In case of a correlated channel, the proposed scheme has an advantage of about 5 dB at a BER of 10^{-3} for the first pass. In further iterations the performance of MMSE-OSIC is below MMSE. No threshold optimization was executed for MMSE-OSIC in this case.

Appendix

We now proof that $\widetilde{h}_{i,i}$ is real.

Proof: $\widetilde{h}_{i,j}$ is the j^{th} element of $\widetilde{\mathbf{h}}_i$. With $d_\mu = i$, (6) and (7) we get

$$\widetilde{h}_{i,i} = \Bigg(\underbrace{\Big[\sum_{\substack{j \in \overline{\mathcal{D}_{d_\mu}} \\ j = d_\mu}} \mathbf{h}_j \; \sigma^2_{\text{ap},j} \; \mathbf{h}_j^* + \sigma^2_{\text{n}} I_N \Big]^{-1} E_s}_{=:F} \mathbf{h}_i \Bigg)^* \mathbf{h}_i$$

It is easy to check that the matrix F is hermitian since hermitian matrices are closed under addition. The matrix F^{-1} is also hermitian since hermitian matrices are also closed under inversion. Thus,

$$\widetilde{h}_{i,i} = E_s \mathbf{h}_i{}^* F^{-1} \mathbf{h}_i \tag{33}$$

Applying

Theorem 1: Matrix F^{-1} is hermitian if and only if $\mathbf{h}_i{}^* F^{-1} \mathbf{h}_i$ is real for any complex $\mathbf{h}_i$.

completes the proof. ■

References

[1] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *ELECTRONIC LETTERS*, vol. 35, no. 1, pp. 14–15, January 7, 1999.

[2] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *1998 International Symposium on Signals, Systems, and Electronics ISSSE'98*, Pisa, Italy, September 30, 1998.

[3] A. van Zelst, "Space division multiplexing algorithms," in *The 12th IEEE Mediterranean Electrotechnical Conference 2000 (MELECON)*, vol. 3, Limassol, Cyprus, Greece, May 29-31, 2000, pp. 1218–1221.

[4] M. Witzke, S. Bäro, F. Schreckenbach, and J. Hagenauer, "Iterative detection of MIMO signals with linear detectors," in *36th Asilomar Conference On Signals, Systems And Computers ACSSC 2002*, Pacific Grove, CA, USA, November 3-6, 2002.

[5] X. Li, H. Huang, G. J. Foschini, and R. A. Valenzuela, "Effects of iterative detection and decoding on the performance of BLAST," in *IEEE Global Telecommunications Conference (Globecom) 2000*, San Francisco, USA, November 27 - December 1, 2000, pp. 1061–1066.

[6] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, pp. 284–287, March 1974.

[7] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the Log domain," in *IEEE International Conference on Communications, 1995*, vol. 2, Seattle, USA, June 18-22, 1995, pp. 1009–1013.

[8] H. V. Poor, *An Introduction to Signal Detection and Estimation*, zweite ed., ser. Springer Texts in Electrical Engineering. New York: Springer-Verlag, 1994, ISBN 0-387-94173-8.

[9] S. Bäro, G. Bauch, A. Pavlic, and A. Semmler, "Improving BLAST performance using space-time block codes and turbo decoding," in *IEEE Global Telecommunications Conference (Globecom) 2000*, November 2000, pp. 1067–1071.

[10] W.-J. Choi, K.-W. Cheong, and J. M. Cioffi, "Iterative soft interference cancellation for multiple antenna systems," in *IEEE Wireless Communications and Networking Conference WCNC 2002*, 2002.

[11] S. ten Brink, J. Speidel, and R.-H. Yan, "Iterative demapping and decoding for multilevel modulation," in *IEEE Global Telecommunications Conference (Globecom) 1998*, November 1998, pp. 579–584.

[12] A. Boronka, D. Efinger, and J. Speidel, "Improving MIMO detection by L-value analysis and adaptive threshold based cancellation," in *IEEE Global Telecommunications Conference (Globecom) 2003*, San Francisco, USA, December 1-5, 2003.

[13] M. Kiessling, J. Speidel, I. Viering, and M. Reinhardt, "Statistical prefiltering for MIMO systems with linear receivers in the presence of transmit correlation," in *The 57th Semiannual IEEE Vehicular Technology Conference (VTC) 2003 Spring*, vol. 1, Jeju, Korea, April 22-25, 2003, pp. 267–271.

4. ITG-Fachtagung, Leipzig, Mai 2003, Seite 169-176